

*Pacific
Journal of
Mathematics*

Volume 292 No. 2

February 2018

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

EDITORS

Don Blasius (Managing Editor)
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
blasius@math.ucla.edu

Paul Balmer
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
balmer@math.ucla.edu

Kefeng Liu
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
liu@math.ucla.edu

Jie Qing
Department of Mathematics
University of California
Santa Cruz, CA 95064
qing@cats.ucsc.edu

Vyjayanthi Chari
Department of Mathematics
University of California
Riverside, CA 92521-0135
chari@math.ucr.edu

Jiang-Hua Lu
Department of Mathematics
The University of Hong Kong
Pokfulam Rd., Hong Kong
jhlu@maths.hku.hk

Daryl Cooper
Department of Mathematics
University of California
Santa Barbara, CA 93106-3080
cooper@math.ucsb.edu

Sorin Popa
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
popa@math.ucla.edu

Paul Yang
Department of Mathematics
Princeton University
Princeton NJ 08544-1000
yang@math.princeton.edu

PRODUCTION

Silvio Levy, Scientific Editor, production@msp.org

SUPPORTING INSTITUTIONS

ACADEMIA SINICA, TAIPEI
CALIFORNIA INST. OF TECHNOLOGY
INST. DE MATEMÁTICA PURA E APLICADA
KEIO UNIVERSITY
MATH. SCIENCES RESEARCH INSTITUTE
NEW MEXICO STATE UNIV.
OREGON STATE UNIV.

STANFORD UNIVERSITY
UNIV. OF BRITISH COLUMBIA
UNIV. OF CALIFORNIA, BERKELEY
UNIV. OF CALIFORNIA, DAVIS
UNIV. OF CALIFORNIA, LOS ANGELES
UNIV. OF CALIFORNIA, RIVERSIDE
UNIV. OF CALIFORNIA, SAN DIEGO
UNIV. OF CALIF., SANTA BARBARA

UNIV. OF CALIF., SANTA CRUZ
UNIV. OF MONTANA
UNIV. OF OREGON
UNIV. OF SOUTHERN CALIFORNIA
UNIV. OF UTAH
UNIV. OF WASHINGTON
WASHINGTON STATE UNIVERSITY

These supporting institutions contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its contents or policies.


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

The subscription price for 2018 is US \$475/year for the electronic version, and \$640/year for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

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

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

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2018 Mathematical Sciences Publishers

LOCALLY HELICAL SURFACES HAVE BOUNDED TWISTING

DAVID BACHMAN, RYAN DERBY-TALBOT AND ERIC SEDGWICK

A topologically minimal surface may be isotoped into a normal form with respect to a fixed triangulation. If the intersection with each tetrahedron is simply connected, then the pieces of this normal form are triangles, quadrilaterals, and helicoids. Helical pieces can have any number of positive or negative twists. We show here that the net twisting of the helical pieces of any such surface in a given triangulated 3-manifold is bounded.

1. Introduction

In [Bachman 2010], the first author introduced the notion of a *topologically minimal* surface, as a generalization of incompressible [Haken 1968], strongly irreducible [Casson and Gordon 1987], and critical [Bachman 2002] surfaces. Such surfaces have a well-defined *index*, where incompressible, strongly irreducible, and critical surfaces have indices 0, 1, and 2, respectively.

The term “topologically minimal” was chosen because in many ways such surfaces behave like geometrically minimal surfaces, i.e., surfaces that represent critical points for the area function. Analogous properties of the two types of surfaces were made explicit in, e.g., [Bachman 2010; 2012b]. In this paper we show that these two types of surfaces actually *look* the same as well.

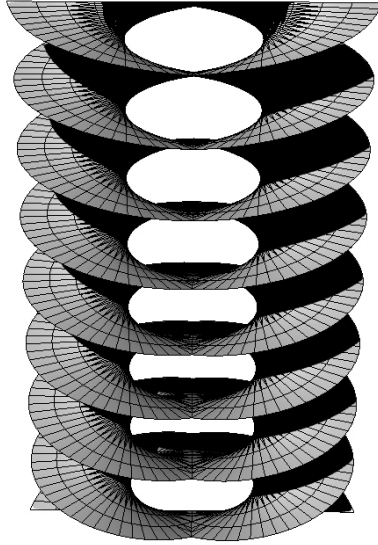
Certain geometrically minimal surfaces are described by the following theorem:

Theorem [Colding and Minicozzi 2006]. *Any nonsimply connected embedded minimal planar domain without small necks can be obtained from gluing together two oppositely oriented double spiral staircases [i.e., helicoids]. Note that because the two double spiral staircases are oppositely oriented, then one remains at the same level if one circles both axes.*

MSC2010: primary 57M99; secondary 53A10.

Keywords: normal surface, minimal surface, topologically minimal surface.

In other words, such geometrically minimal surfaces are depicted (topologically) by the following figure:



Here we show that topologically minimal surfaces (also without small necks) are generally comprised of pairs of oppositely oriented helicoids as well.

A useful fact about topologically minimal surfaces is that they can be isotoped into a standard normal form with respect to a triangulation. This was first done by Kneser [1929] and Haken [1961] in the index 0 case, Rubinstein [1997] and Stocking [2000] for closed index 1 surfaces, and Bachman et al. [2013] for index 1 surfaces with boundary. The general case of arbitrary index was addressed by the first author in [Bachman 2012a; 2012b; 2013]. The following theorem summarizes these results:

Theorem 1.1. *Let M be a compact, orientable, irreducible, triangulated 3-manifold with incompressible boundary. Then for each n there exists a finite, constructible set of surfaces in each tetrahedron of M from which one can build any index n topologically minimal surface in M (up to isotopy).*

The pieces from which index n surfaces can be built by Theorem 1.1 can be quite complicated. However, in [Bachman 2012b] the first author gave a relatively simple characterization of those components that are simply connected: such pieces are either triangles or helicoids¹ (see Figure 1). We say any surface built entirely from such pieces is *locally helical*.

Helical pieces are classified by their *axis* (see Section 3) and *twisting*. If H_* is a helicoid then the number of normal arcs comprising ∂H_* is $4(n + 1)$, for

¹We regard quadrilaterals as untwisted helicoids.

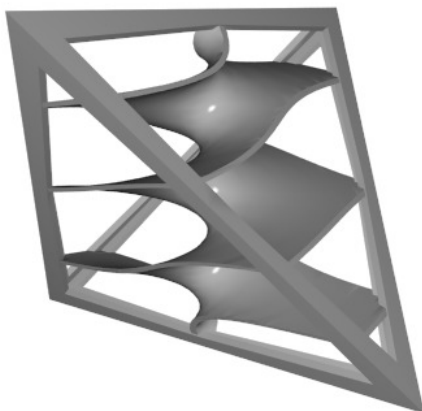


Figure 1. A helicoid whose boundary has length 16. Note that it meets one pair of opposite edges in single points, a second pair in three points, and a third pair in four points. The twisting of this helicoid is 3.

some n . The *twisting* of H_* , denoted $t(H_*)$, is the number $\pm n$, where the sign is determined by the handedness of the helicoid and the orientation of the manifold (see Definition 3.3). If H is a locally helical surface in a triangulated 3-manifold M , then the *net twisting* of H is the sum of the twisting of all of its helical pieces (see Definition 3.4 for a more precise definition). The *total absolute twisting* is the sum of the absolute values of the twisting of its helical pieces. Note that if a surface has bounded total absolute twisting, then each helical piece has a bounded number of twists. If, on the other hand, the net twisting is bounded then there may be helical pieces with an arbitrarily large number of, say, positive twists, as long as there are also pieces with large numbers of negative twists.

The results of [Bachman 2012a] and [Bachman 2012b], taken together, imply the following:

Theorem 1.2. *Any topologically minimal surface with index n that is isotopic to a locally helical surface is isotopic to one with total absolute twisting at most n .*

The results mentioned above give a direct generalization of Haken's normalization of incompressible surfaces [1968]. To see this, first note that by definition, an incompressible surface is index 0. By Theorem 1.1 such a surface can be isotoped to be locally topologically minimal. By incompressibility, we may assume that in this position it is locally simply connected. Finally, by Theorem 1.2 we conclude that the total absolute twisting must be 0, which means that it is a collection of triangles and quadrilaterals.

For higher index locally helical surfaces, the situation may be more complicated, as there may be helicoids distributed across tetrahedra in M . The main result of this paper is the following theorem, which says that the total absolute twisting outside of some prescribed set of tetrahedra Δ constrains the net twisting inside Δ .

Theorem 1.3. *Let M be a closed, oriented, triangulated 3-manifold, and let Δ be a set of tetrahedra in the triangulation of M . Let H be a locally helical surface in M such that the total absolute twisting of $H - \Delta$ is at most n . Then the net twisting of $H \cap \Delta$ is bounded, where the bound depends only on M and n .*

Three corollaries of this theorem are worth noting: where Δ is a single tetrahedron of M , where Δ is exactly two tetrahedra, and where Δ is the set of all tetrahedra in M .

Corollary 1.4. *Let M be a closed, oriented, triangulated 3-manifold, and let Δ be a tetrahedron of the triangulation. Let H be a locally helical surface in M such that the total absolute twisting of $H - \Delta$ is at most n . Then the total absolute twisting of H is bounded, where the bound depends only on M and n .*

In particular, in a given triangulated 3-manifold M , the total (absolute or net) twisting of any locally helical surface with a single helicoid is bounded, where the bound depends only on M . Corollary 1.4 follows from Theorem 1.3 by making the observation that a bound on the net twisting of a surface in a single tetrahedron serves as a bound for its absolute value.

Corollary 1.5. *Let M be a closed, oriented, triangulated 3-manifold, and let Δ_1, Δ_2 be a pair of tetrahedra in the triangulation of M . Let H be a locally helical surface in M such that the total absolute twisting of $H - (\Delta_1 \cup \Delta_2)$ is at most n . Then $t(H \cap \Delta_1) = -t(H \cap \Delta_2) + m$, where m is bounded by a function of M and n .*

In other words, if, in a sequence of surfaces with bounded total absolute twisting outside of $\Delta_1 \cup \Delta_2$, the number of left-handed twists in Δ_1 is growing, then the number of right-handed twists in Δ_2 must be growing at the same rate (asymptotically). This corollary is what establishes our claim that topologically minimal surfaces look like the geometrically minimal surfaces described by Colding and Minicozzi, as discussed above.

The last corollary of Theorem 1.3 is when Δ is the set of all tetrahedra in M . In this case, our result makes no mention of total absolute twisting.

Corollary 1.6. *Let M be a closed, oriented, triangulated 3-manifold, and let H be a locally helical surface in M . Then the net twisting of H is bounded, where the bound depends only on M .*

In contrast, note that one cannot prove that the total absolute twisting is bounded. A simple example can be seen by gluing a left-handed helicoid in a tetrahedron

to the right-handed helicoid in its mirror image, forming a closed locally helical sphere in S^3 . Such spheres have unbounded total absolute twisting, and zero net twisting.

In the next section we characterize normal curves by their *type*. In Section 3 we characterize helical disks by their *axis*. Finally, in Section 4 we define the *compatibility class* of a locally helical surface. Those familiar with normal surface theory will find several of these notions familiar. The layering of these definitions parses the set of locally helical surfaces in $(M; \Delta)$ more and more finely, imposing increasingly greater restrictions on how surfaces in the same class can intersect. Taken all at once, these characterizations produce a finite set of *consistency classes* for locally helical surfaces in $(M; \Delta)$, which have just the properties needed to prove Theorem 1.3.

2. The *type* of a normal curve on a tetrahedron

In this section we consider the combinatorics of normal loops on the boundary of a tetrahedron. For a basic reference on normal surface theory, we refer the reader to [Hass 1998].

Lemma 2.1. *Let σ be a tetrahedron, and α a normal loop of length at least four on $\partial\sigma$. Let ϕ denote a 180-degree rotation of σ about a line connecting the midpoints of opposite edges of σ . Then α is normally parallel to a loop that is preserved by ϕ .*

Proof. To begin, we claim that a normal loop of length at least four meets each pair of opposite edges of $\partial\sigma$ in the same number of points. One way to see this is by noting that the double cover of $\partial\sigma$, branched over the vertex set, is a torus (see Figure 2). Each edge of $\partial\sigma$ lifts to an essential loop on the torus, and each pair of opposite edges lifts to two parallel loops. Now, as a loop α of length at least four on $\partial\sigma$ also lifts to two essential loops on the torus, it must be the case that α intersects opposite edges of $\partial\sigma$ in an equal number of points.

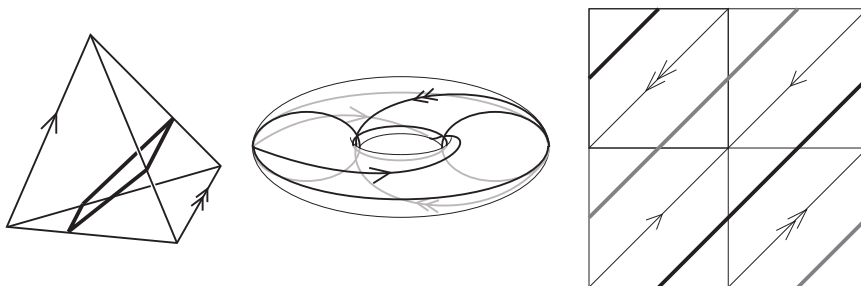


Figure 2. The torus as a double branched cover of the boundary of a tetrahedron, and components of a lift of a length four curve in its unfolded version.

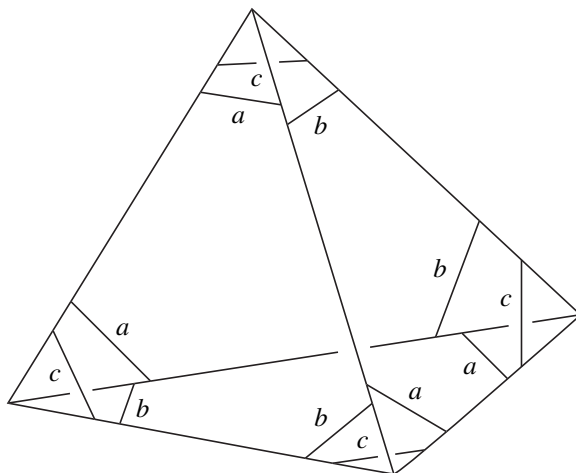


Figure 3. Labeling the normal arc types on the boundary of a tetrahedron, σ .

Now note that the rotation ϕ preserves the two edges that its axis intersects, and swaps the other two pairs of opposite edges. Hence, both α and $\phi(\alpha)$ will meet each edge in the same number of points. As these numbers completely determine the intersection of α with each face of σ (up to normal isotopy), the result follows. \square

This lemma gives us a way to classify normal curves on the boundary of a tetrahedron. Label the normal arc types on each face of a tetrahedron σ as in Figure 3. These labels are arranged so as to be preserved by 180-degree rotations about axes that connect the midpoints of opposite edges. Any normal loop α of length at least four on $\partial\sigma$ meets each face in a collection of normal arcs. By Lemma 2.1, the number of these arcs that are parallel to an arc with one label in one face will be the same as the number that are parallel to an arc with the same label in any other face. Hence, if we fix one face δ of σ and let $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$ be the number of arcs of $\alpha \cap \delta$ parallel to the labeled arcs a , b , and c of the figure, then these three functions will be independent of the choice of δ .

Note furthermore that for any loop α of length at least four, at least one of the three numbers $a(\alpha)$, $b(\alpha)$ or $c(\alpha)$ will be zero (otherwise α would have length three components). This motivates the following definition.

Definition 2.2. Let σ be a tetrahedron with labeled normal arc types as in Figure 3, and let α be a normal loop on $\partial\sigma$ of length at least four. We say α is *type a* if $a(\alpha) = 0$. Define *type b* and *type c* similarly.

Note that normal loops of length exactly four will be of two types. The notion of type constrains how two normal curves can intersect on the boundary of a tetrahedron, as seen in the following two lemmas.

Lemma 2.3. *Let α and β be normal loops of length at least four on $\partial\sigma$ of the same type. Let $\alpha + \beta$ be the normal loop(s) obtained by resolving all intersection points. Then $\alpha + \beta$ does not contain any components of length three.*

Proof. Suppose α and β are type a . Then $a(\alpha) = a(\beta) = 0$. As $a(\alpha + \beta) = a(\alpha) + a(\beta)$ for any two normal loops, we conclude $a(\alpha + \beta) = 0$. Thus, there is a missing arc type around each vertex of σ (see Figure 3). We conclude $\alpha + \beta$ does not have any components of length three. \square

Definition 2.4. Let α_0 and β_0 be normal arcs in an oriented triangle δ . Then α_0 and β_0 can be isotoped, keeping their boundaries fixed, so that they intersect transversely in at most one point. We define the (*normal*) *sign* of the point $\alpha_0 \cap \beta_0$, if it exists, as follows. Orient α_0 and β_0 so that the ordering (α_0, β_0) agrees with the orientation of δ . There are now two possibilities. If the regular exchange at $\alpha_0 \cap \beta_0$ attaches the tail of α_0 to the tip of β_0 then we say intersection point $\alpha_0 \cap \beta_0$ is *positive*. Otherwise we say it is *negative* (see Figure 4).

Note that with a fixed orientation on δ , the sign of $\alpha_0 \cap \beta_0$ is opposite the sign of $\beta_0 \cap \alpha_0$.

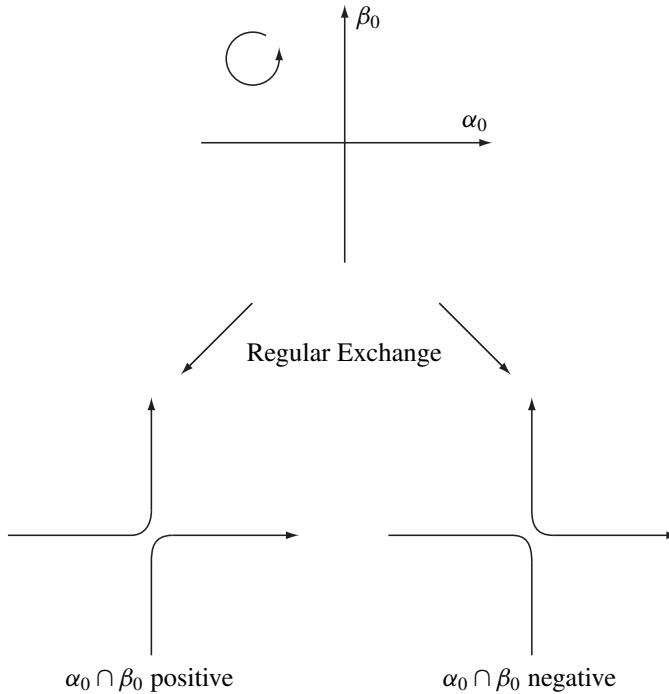


Figure 4. The sign of $\alpha_0 \cap \beta_0$, as determined by the regular exchange.

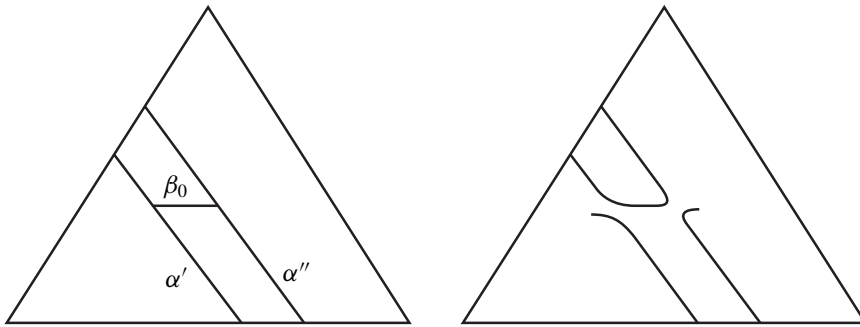


Figure 5. Resolving intersections of opposite signs produces a nonnormal arc.

Lemma 2.5. *Let α and β be collections of normal loops on $\partial\sigma$ whose non-length-three components are all of the same type, that have been normally isotoped to intersect minimally. Then each point of $\alpha \cap \beta$ has the same sign.*

Proof. If either α or β contains a component of length three, then it will be disjoint from the other collection. Thus, we may assume that all points of $\alpha \cap \beta$ lie on loops of length at least four. We will call such loops *long*. The long loops of α will all be parallel, as will the long loops of β . Thus, if there are any intersection points at all, then no long loop of α can be parallel to a long loop of β .

By way of contradiction, we now assume that two points of $\alpha \cap \beta$ are of opposite sign. We claim that then there is a subarc of α or β that connects two points of $\alpha \cap \beta$ of opposite sign. If not, then we may choose a component α_+ of α with only positive intersection points, and a component β_- of β with only negative intersection points. However, it then follows that α_+ is disjoint from β_- , which cannot happen for two nonparallel long loops. We proceed, then, without loss of generality assuming there is a subarc of β that connects points of opposite sign. It follows that there is such a subarc, β_0 , which does not meet α in its interior.

There are now two cases. Suppose first that the points of $\partial\beta_0$ lie on different components α' and α'' and of α . As all long components of α are normally parallel, α' and α'' cobound an annulus of $\partial\sigma$, with β_0 a spanning arc. By making this annulus thin, we may assume that β_0 lies in a face of σ . However, resolving the two intersections at each end of β_0 then produces a nonnormal arc. (See Figure 5.)

The second case is when the points of $\partial\beta_0$ lie on the same component α' of α . Note that α' is a loop that divides $\partial\sigma$ into two hemispheres, each intersecting the boundary of the tetrahedron in one of three ways, as seen in Figure 6. The loop α' cannot be as depicted in Figure 6 (left), where one of these hemispheres contains a single vertex of $\partial\sigma$, since it is long. Thus, we may assume both hemispheres contain two vertices of $\partial\sigma$. Let D be the hemisphere that contains β_0 . Note that

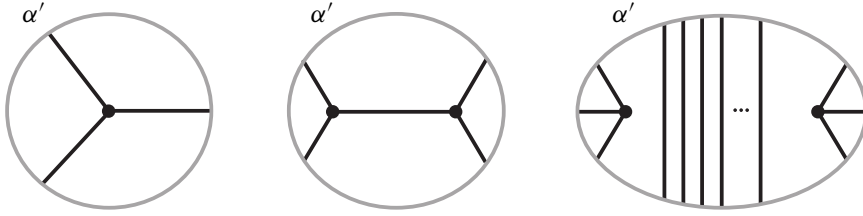


Figure 6. The three possibilities for one hemisphere of $\partial\sigma$, bounded by α' . The black points and edges indicate vertices and subarcs of edges of $\partial\sigma$, respectively.

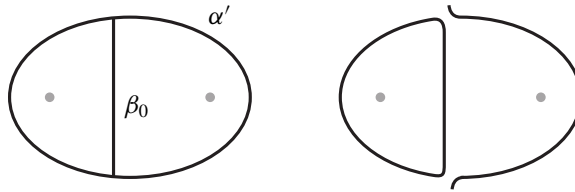


Figure 7. Resolving the intersection points of opposite sign of α' and β_0 produces a length three curve or nonnormal arc.

β_0 then divides D into two subdisks, and by the minimality of $|\alpha \cap \beta|$, each such subdisk will contain a vertex of σ . Resolving the intersections at each end of β_0 then produces a vertex linking loop. (See Figure 7.) We will leave it as an exercise for the reader that such a loop will then persist after all further resolutions, producing a length three normal loop. By Lemma 2.3, it follows that the long loops of α and β could not have been the same type. \square

3. Helicoids with the same axis

Definition 3.1. Let H_* be a disk properly embedded in a tetrahedron, whose boundary is a normal loop. If ∂H_* meets some pair of opposite edges e, e' in single points then we say H_* is a *helicoid*, and $\{e, e'\}$ is an *axis* of H_* .

Note that both quadrilaterals and octagons are helicoids with two axes, and all other helicoids have a unique axis. (See Figure 8.) However, the boundary of each helicoid with axis $\{e, e'\}$ meets e in a unique normal arc type as in Figure 3.

Definition 3.2. Given a helicoid H_* with axis $\{e, e'\}$ in a tetrahedron σ , there is an orientation-preserving simplicial homeomorphism from σ to the tetrahedron pictured in Figure 3 (equipped with the standard orientation on \mathbb{R}^3), where e and e' are taken to the edges that meet arc types a and b . We say H_* is *right-handed with respect to* $\{e, e'\}$ if $a(\partial H_*) = 0$ and *left-handed with respect to* $\{e, e'\}$ if $b(\partial H_*) = 0$.

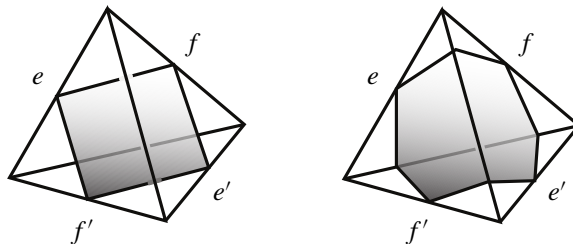


Figure 8. Quadrilaterals and octagons are the only two locally helical surfaces with more than one choice of axis. Note here that the quadrilateral is left-handed and the octagon is right-handed with respect to $\{e, e'\}$, with the opposite being the case with respect to $\{f, f'\}$.

Definition 3.3. Let H_* be a helicoid with $4(n+1)$ normal arcs comprising ∂H_* , and with axis $\{e, e'\}$ in a tetrahedron σ . We say the *twisting* of H_* , $t(H_*)$, is $+n$ if H_* is right-handed with respect to $\{e, e'\}$ and $-n$ if it is left-handed with respect to $\{e, e'\}$.

Note that an octagon can be regarded as having $+1$ or -1 twisting, depending on the choice of its axis. The handedness of the twisting of a quadrilateral is also dependent on a choice of axis, but in either case the value of the twisting is zero. Thus, a helical surface with no octagons has a well-defined net twisting. When there are octagons present, however, the net twisting will depend on choices of axes, motivating the following definition:

Definition 3.4. Let M be a triangulated 3-manifold containing a locally helical surface H , and let Δ be a set of tetrahedra in the triangulation of M . We say the *net twisting of H in Δ is bounded by n* if

$$-n \leq \sum_{\sigma \in \Delta} t(H \cap \sigma) \leq n$$

for all choices of axes of the components of $H \cap \sigma$, for each $\sigma \in \Delta$.

Definition 3.5. Let σ be an oriented tetrahedron. For any two normal curves α and β on $\partial\sigma$ in general position, let $\eta_\sigma(\alpha \cap \beta)$ denote the difference between the total number of positive and negative intersection points of $\alpha \cap \beta$ on the 2-simplices of $\partial\sigma$.

Lemma 3.6. Let H_* and G_* be helicoids with the same handedness with respect to the same choice of axis. Then

$$\eta_\sigma(\partial H_* \cap \partial G_*) = 2(t(H_*) - t(G_*)).$$

Proof. As H_* and G_* are helicoids with the same handedness with respect to some choice of axis, it follows that their boundaries are loops of the same type. It thus

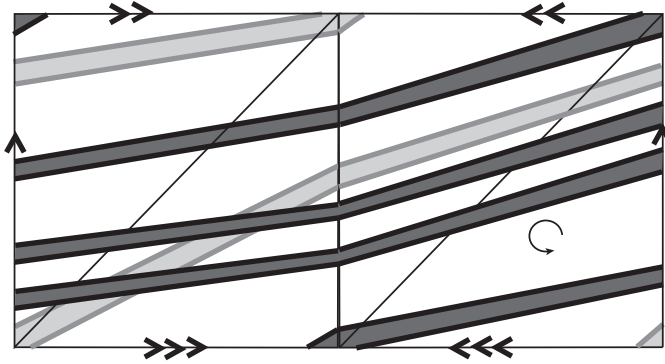


Figure 9. Here ∂H_* is the black curve, with a neighborhood of h being the dark gray band, and ∂G_* is depicted in lighter gray. Here, $|h \cap g| = 2$, $\eta_\sigma(\partial H_* \cap \partial G_*) = 8$, $t(H_*) = 6$, $t(G_*) = 2$ and thus $t(H_*) - t(G_*) = 4$.

follows immediately from Lemma 2.5 that $\eta_\sigma(\partial H_* \cap \partial G_*) = \pm |\partial H_* \cap \partial G_*|$, where the sign is determined by the normal intersection sign of the intersection points. Without loss of generality, assume this sign is positive. Our goal is to show

$$|\partial H_* \cap \partial G_*| = 2(t(H_*) - t(G_*)).$$

Notice that ∂H_* is a loop on $\partial\sigma$ dividing it into two hemispheres, where each hemisphere contains two of the vertices of σ . Let v and w be the vertices in one such hemisphere, and let h be an arc in this hemisphere connecting them. Note that the arc h can be chosen so that ∂H_* is normally parallel to a neighborhood of h .

Similarly, ∂G_* is parallel to the boundary of a neighborhood of an arc g connecting two vertices of σ . The arc g may be chosen so that at least one of its endpoints is distinct from the endpoints of h .

There are now two cases. If both endpoints of g are distinct from the endpoints of h then the curves can be arranged as in Figure 9. Note that $\partial H_* \cap \partial G_*$ contains four intersection points for each crossing of h and g . Furthermore, the difference in the twisting, $t(H_*) - t(G_*)$, is twice the number of crossings of h and g . Thus, the desired equation holds.

All intersection points depicted in the figure are positive, as is the twisting. Note that switching the orientation and keeping the ordering of the curves the same changes the sign of both the intersection points and the twisting. Alternatively, keeping the orientation fixed but changing the ordering of the curves will also change the sign of the intersection points, and reverse the order of the operands on the right side of the desired equation. Thus the equation still holds.

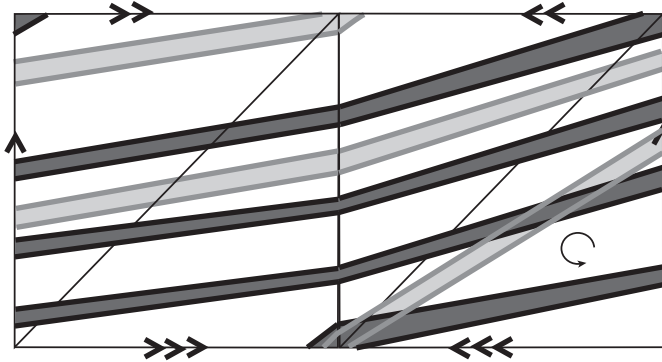


Figure 10. In this case, h and g share an endpoint. Here, $|h \cap g| = 2$, $\eta_\sigma(\partial H_* \cap \partial G_*) = 6$, $t(H_*) = 6$, $t(G_*) = 3$ and thus $t(H_*) - t(G_*) = 3$.

In the second case, h and g have an endpoint in common, as in Figure 10. In this case $|\partial H_* \cap \partial G_*| = 4|h \cap g| - 2$ and $t(H_*) - t(G_*) = 2|h \cap g| - 1$. Thus we still obtain the desired relationship between $\eta_\sigma(\partial H_* \cap \partial G_*)$ and $t(H_*) - t(G_*)$. \square

4. Compatibility classes of surfaces

The results of this section extend previous results that restrict intersections of boundary curves realized by compatibility classes of surfaces, found, e.g., in [Bachman et al. 2016; Jaco and Sedgwick 2003; Hatcher 1982].

Definition 4.1. Two surfaces in a triangulated 3-manifold are *compatible* if they meet the boundary of each tetrahedron in a collection of normal curves that can be normally isotoped to be disjoint.²

Henceforth we will assume that if α_0 and β_0 are contained in a 2-simplex $\delta \subset \partial M$, then the orientation on δ is induced by the orientation on M . Hence, for such curves we may reference the sign of each point of $\alpha_0 \cap \beta_0$ without mention of the orientation of the 2-simplex that contains it.

In the next lemma, we show that two compatible surfaces have a symmetric relationship between the signs of their normal intersections on the boundary of a subcomplex, Δ .

Lemma 4.2. Let M be a closed, oriented, triangulated 3-manifold. Let Δ be a set of tetrahedra in the triangulation of M . Suppose A and B are two locally helical surfaces in M that are compatible outside Δ . Let $\partial_\Delta A = \partial(A \cap \Delta)$ and $\partial_\Delta B =$

²We are allowing pseudotriangulations; i.e., M is realized as a collection of tetrahedra with face-pairings. Hence, for each 3-cell σ in M there is a map $\pi : \Sigma \rightarrow \sigma$, where Σ is a 3-simplex. Here we consider two surfaces to be compatible if they meet $\partial\sigma$ in curves whose preimages can be isotoped to be disjoint on $\partial\Sigma$.

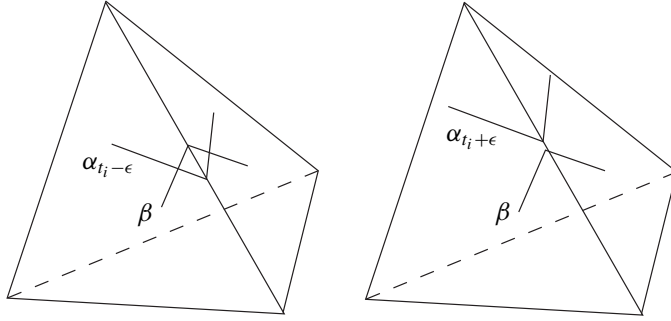


Figure 11. Positive and negative intersections cancel as t increases through t_i .

$\partial(B \cap \Delta)$. Suppose A and B have been normally isotoped so that $|\partial_\Delta A \cap \partial_\Delta B|$ is minimal. Then the number of points of $\partial_\Delta A \cap \partial_\Delta B$ with positive normal sign equals the number of points with negative normal sign.

Proof. Consider a tetrahedron σ of M that is not in Δ . Let α denote a component of $A \cap \partial\sigma$, and β a component of $B \cap \partial\sigma$. Orient each 2-simplex of $\partial\sigma$ by the induced orientation from σ , so that each point of $\alpha \cap \beta$ has a well-defined sign. Since A and B intersect minimally, we may assume each normal arc of α and β is a straight line segment. Recall from Definition 3.5 that $\eta_\sigma(\alpha \cap \beta)$ denotes the difference between the total number of positive and negative intersection points of $\alpha \cap \beta$ on the 2-simplices of $\partial\sigma$.

As A and B are compatible, there is an isotopy from α to a normal loop α' , also consisting of straight normal arcs, in $\partial\sigma$ that is disjoint from β . We can choose such an isotopy, α_t , so that for all t , each normal arc of α_t is a straight line segment and $\alpha_t \cap \beta$ contains at most one point of the 1-skeleton. Let $\{t_i\}$ denote the critical values of $\alpha_t \cap \beta$, i.e., the values of t such that α_t and β do not intersect transversely on $\partial\sigma$. It follows that for each i , $\alpha_{t_i} \cap \beta$ includes a point of the 1-skeleton.

Just before (or after) t_i , α_t meets β as in Figure 11. Here we see two intersections, one of each normal sign, of $\alpha_t \cap \beta$ which cancel as t increases through t_i . It follows that $\eta_\sigma(\alpha_{t_i - \epsilon} \cap \beta) = \eta_\sigma(\alpha_{t_i + \epsilon} \cap \beta)$. As $\alpha' \cap \beta = \emptyset$, we conclude $\eta_\sigma(\alpha_t \cap \beta)$ is zero for all noncritical t . In particular, it must have been the case that $\eta_\sigma(\alpha \cap \beta) = 0$.

Let $\eta_\sigma(A \cap B)$ now denote the sum, over all curves α of $A \cap \partial\sigma$ and β of $B \cap \partial\sigma$ of $\eta_\sigma(\alpha \cap \beta)$. It follows from the above argument that $\eta_\sigma(A \cap B) = 0$. Thus, the sum over all tetrahedra σ not in Δ of $\eta_\sigma(A \cap B)$ is also zero.

Now note that if δ is an interior 2-simplex, then the normal sign of any intersection point of $A \cap \delta$ and $B \cap \delta$ is opposite from the perspective of the tetrahedra on either side of δ . Hence, the sum of $\eta_\sigma(A \cap B)$, over all tetrahedra σ , must be equal to the

difference of the number of positive and negative intersection points of $\partial_\Delta A \cap \partial_\Delta B$. As we have reasoned above that this total is zero, the result follows. \square

5. Main proof

We now put the three notions of axis, twisting, and compatibility together.

Definition 5.1. Let M be a closed, oriented, triangulated 3-manifold, and let Δ be a set of tetrahedra in the triangulation of M . Two locally helical surfaces H and G are said to be *consistent* in $(M; \Delta)$ if they are compatible outside of Δ , and if for all $\sigma \in \Delta$, $H \cap \sigma$ and $G \cap \sigma$ have the same handedness with respect to the same choice of axis.

Theorem 1.3 is a consequence of the following lemma.

Lemma 5.2. *Let M be a closed, oriented, triangulated 3-manifold, and let Δ be a set of tetrahedra in the triangulation of M . If H and G are consistent, locally helical surfaces in $(M; \Delta)$, then the net twisting of $H \cap \Delta$ is the same as the net twisting of $G \cap \Delta$.*

Proof. For each $\sigma \in \Delta$, let $H_\sigma = H \cap \sigma$ and $G_\sigma = G \cap \sigma$. As noted in the proof of Lemma 3.6, for each $\sigma \in \Delta$, ∂H_σ and ∂G_σ must be normal loops of the same type. Thus, by Lemma 2.5, for each $\sigma \in \Delta$, all points of $\partial H_\sigma \cap \partial G_\sigma$ have the same sign. Let Δ_+ be the subset of Δ where this sign is positive, and Δ_- the subset of Δ where it is negative. Thus, on each $\sigma \in \Delta_+$,

$$\eta_\sigma(\partial H_\sigma \cap \partial G_\sigma) = |\partial H_\sigma \cap \partial G_\sigma|,$$

and for all $\sigma \in \Delta_-$,

$$\eta_\sigma(\partial H_\sigma \cap \partial G_\sigma) = -|\partial H_\sigma \cap \partial G_\sigma|.$$

Consider the sum $\sum_{\sigma \in \Delta} \#(\partial H_\sigma \cap \partial G_\sigma)$, where $\#(\partial H_\sigma \cap \partial G_\sigma)$ denotes the signed intersection number of ∂H_σ and ∂G_σ . Suppose σ_1 and σ_2 are adjacent tetrahedra in Δ , $p_1 \in \partial H_{\sigma_1} \cap \partial G_{\sigma_1}$, $p_2 \in \partial H_{\sigma_2} \cap \partial G_{\sigma_2}$, and p_1 is identified with p_2 in M .³ Then the sign of p_1 will be opposite the sign of p_2 , and thus p_1 and p_2 will cancel in $\sum_{\sigma \in \Delta} \#(\partial H_\sigma \cap \partial G_\sigma)$. If, on the other hand, p is a point of $\partial H_\sigma \cap \partial G_\sigma$ that is on a unique $\sigma \in \Delta$, then $p \in \partial(H - \Delta) \cap \partial(G - \Delta)$. By hypothesis, $H - \Delta$ and $G - \Delta$ are compatible surfaces; thus by Lemma 4.2 the number of positive and negative points of $\partial(H - \Delta) \cap \partial(G - \Delta)$ are equal. We conclude that $\sum_{\sigma \in \Delta} \#(\partial H_\sigma \cap \partial G_\sigma) = 0$, or equivalently,

$$\sum_{\sigma \in \Delta_+} |\partial H_\sigma \cap \partial G_\sigma| = \sum_{\sigma \in \Delta_-} |\partial H_\sigma \cap \partial G_\sigma|,$$

³Here we are allowing σ_1 to be equal to σ_2 when there are self-identifications, but in this case p_1 must be distinct from p_2 .

and thus,

$$\sum_{\sigma \in \Delta_+} \eta_\sigma(\partial H_\sigma \cap \partial G_\sigma) = - \sum_{\sigma \in \Delta_-} \eta_\sigma(\partial H_\sigma \cap \partial G_\sigma).$$

Applying Lemma 3.6 to this equality now yields

$$\sum_{\sigma \in \Delta_+} 2(t(H_\sigma) - t(G_\sigma)) = - \sum_{\sigma \in \Delta_-} 2(t(H_\sigma) - t(G_\sigma)),$$

which implies

$$\begin{aligned} 0 &= \sum_{\sigma \in \Delta_+} 2(t(H_\sigma) - t(G_\sigma)) + \sum_{\sigma \in \Delta_-} 2(t(H_\sigma) - t(G_\sigma)) \\ &= \sum_{\sigma \in \Delta} 2(t(H_\sigma) - t(G_\sigma)) \\ &= \sum_{\sigma \in \Delta} t(H_\sigma) - \sum_{\sigma \in \Delta} t(G_\sigma). \end{aligned}$$

Therefore, $\sum_{\sigma \in \Delta} t(H_\sigma) = \sum_{\sigma \in \Delta} t(G_\sigma)$; i.e., the net twisting is the same for all surfaces in the chosen consistency class. \square

We are now ready to prove Theorem 1.3.

Proof. Let n be a positive integer, and consider the set of all locally helical surfaces (up to normal isotopy) that have total absolute twisting $\leq n$ in $M - \Delta$. The number of compatibility classes of surfaces in $M - \Delta$ is finite, since there are only a finite number of normal loops on each tetrahedron of length $\leq 4(n + 1)$. Moreover, there are only three possible axes for each tetrahedron in Δ , and two choices of handedness for each. Thus, the number of consistency classes for $(M; \Delta)$ is finite. Theorem 1.3 thus immediately follows from Lemma 5.2. \square

References

- [Bachman 2002] D. Bachman, “Critical Heegaard surfaces”, *Trans. Amer. Math. Soc.* **354**:10 (2002), 4015–4042. MR Zbl
- [Bachman 2010] D. Bachman, “Topological index theory for surfaces in 3-manifolds”, *Geom. Topol.* **14**:1 (2010), 585–609. MR Zbl
- [Bachman 2012a] D. Bachman, “Normalizing topologically minimal surfaces, I: Global to local index”, preprint, 2012. arXiv
- [Bachman 2012b] D. Bachman, “Normalizing topologically minimal surfaces, II: Disks”, preprint, 2012. arXiv
- [Bachman 2013] D. Bachman, “Normalizing topologically minimal surfaces, III: Bounded combinatorics”, preprint, 2013. arXiv
- [Bachman et al. 2013] D. Bachman, R. Derby-Talbot, and E. Sedgwick, “Almost normal surfaces with boundary”, pp. 177–194 in *Geometry and topology down under*, edited by C. D. Hodgson et al., Contemp. Math. **597**, American Mathematical Society, Providence, RI, 2013. MR Zbl

- [Bachman et al. 2016] D. Bachman, R. Derby-Talbot, and E. Sedgwick, “Heegaard structure respects complicated JSJ decompositions”, *Math. Ann.* **365**:3-4 (2016), 1137–1154. MR Zbl
- [Casson and Gordon 1987] A. J. Casson and C. M. Gordon, “Reducing Heegaard splittings”, *Topology Appl.* **27**:3 (1987), 275–283. MR Zbl
- [Colding and Minicozzi 2006] T. H. Colding and W. P. Minicozzi, II, “Shapes of embedded minimal surfaces”, *Proc. Natl. Acad. Sci. USA* **103**:30 (2006), 11106–11111. MR Zbl
- [Haken 1961] W. Haken, “Theorie der Normalflächen”, *Acta Math.* **105** (1961), 245–375. MR Zbl
- [Haken 1968] W. Haken, “Some results on surfaces in 3-manifolds”, pp. 39–98 in *Studies in modern topology*, edited by P. J. Hilton, Math. Assoc. Amer., Washington, DC, 1968. MR Zbl
- [Hass 1998] J. Hass, “Algorithms for recognizing knots and 3-manifolds: knot theory and its applications”, *Chaos Solitons Fractals* **9**:4-5 (1998), 569–581. MR Zbl
- [Hatcher 1982] A. E. Hatcher, “On the boundary curves of incompressible surfaces”, *Pacific J. Math.* **99**:2 (1982), 373–377. MR Zbl
- [Jaco and Sedgwick 2003] W. Jaco and E. Sedgwick, “Decision problems in the space of Dehn fillings”, *Topology* **42**:4 (2003), 845–906. MR Zbl
- [Kneser 1929] H. Kneser, “Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten”, *Jahresber. Dtsch. Math.-Ver.* **38** (1929), 248–260. JFM
- [Rubinstein 1997] J. H. Rubinstein, “Polyhedral minimal surfaces, Heegaard splittings and decision problems for 3-dimensional manifolds”, pp. 1–20 in *Geometric topology* (Athens, GA, 1993), edited by W. H. Kazez, AMS/IP Stud. Adv. Math. **2**, American Mathematical Society, Providence, RI, 1997. MR Zbl
- [Stocking 2000] M. Stocking, “Almost normal surfaces in 3-manifolds”, *Trans. Amer. Math. Soc.* **352**:1 (2000), 171–207. MR Zbl

Received April 8, 2016. Revised March 2, 2017.

DAVID BACHMAN
PITZER COLLEGE
CLAREMONT, CA
UNITED STATES
bachman@pitzer.edu

RYAN DERBY-TALBOT
QUEST UNIVERSITY
SQUAMISH, BC
CANADA
rdt@questu.ca

ERIC SEDGWICK
DEPAUL UNIVERSITY
CHICAGO, IL
UNITED STATES
esedgwick@cdm.depaul.edu

SUPERCONVERGENCE TO FREELY INFINITELY DIVISIBLE DISTRIBUTIONS

HARI BERCOVICI, JIUN-CHAU WANG AND PING ZHONG

We prove superconvergence results for all freely infinitely divisible distributions. Given a nondegenerate freely infinitely divisible distribution ν , let μ_n be a sequence of probability measures and let k_n be a sequence of integers tending to infinity such that $\mu_n^{\boxplus k_n}$ converges weakly to ν . We show that the density $d\mu_n^{\boxplus k_n}/dx$ converges uniformly, as well as in all L^p -norms for $p > 1$, to the density of ν except possibly in the neighborhood of one point. Applications include the global superconvergence to freely stable laws and that to free compound Poisson laws over the whole real line.

1. Introduction

Consider a sequence $\{X_i\}_{i=1}^\infty$ of independent identically distributed random variables with zero mean and unit variance. The classical central limit theorem states that variables

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$$

converge in distribution to the standard normal law. Note that the variables S_n might always be discrete, even though their limit is absolutely continuous. This means that the convergence of S_n to a normal law must be expressed in terms of distribution functions, rather than densities.

Assume now that, instead of being independent, the variables $\{X_i\}_{i=1}^\infty$ are *freely* independent in the sense of [Voiculescu et al. 1992]. We still assume them identically distributed with zero mean and unit variance. Under the additional condition that the variables are bounded, it was shown in [Bercovici and Voiculescu 1995] that the distribution of S_n is absolutely continuous for sufficiently large n , and these densities converge uniformly, along with all of their derivatives, to the density of the semicircle law

$$\frac{1}{2\pi} \sqrt{4 - t^2}$$

on any interval $[a, b] \subset (-2, 2)$. This phenomenon was called *superconvergence* in that paper. In [Wang 2010], the assumption that X_i be bounded was removed. Even

MSC2010: 46L54.

Keywords: freely infinitely divisible law, free convolution, superconvergence.

when the variables X_i are not identically distributed, but are uniformly bounded, the support of S_n was shown by Kargin [2007] to converge to the interval $[-2, 2]$ as $n \rightarrow \infty$. See also [Anshelevich et al. 2014] for multiplicative superconvergence results.

The purpose of this paper is to demonstrate that the phenomenon of superconvergence is not limited to convergence to the semicircle law. Consider a nondegenerate probability measure ν on \mathbb{R} , which is infinitely divisible in the free sense (that is, \boxplus -infinitely divisible). It is known that its Cauchy transform,

$$(1-1) \quad G_\nu(z) = \int_{-\infty}^{+\infty} \frac{1}{z-t} d\nu(t),$$

defined for $\Im z > 0$, extends continuously to all points $z \in \mathbb{R}$ with at most one exception t_ν . The measure ν is absolutely continuous on $\mathbb{R} \setminus \{t_\nu\}$ and its density is locally analytic when strictly positive. To formulate our result, assume that for every positive integer n , we are given k_n freely independent, identically distributed random variables $X_{n1}, X_{n2}, \dots, X_{nk_n}$ such that $\lim_{n \rightarrow \infty} k_n = \infty$ and the sums

$$S_n = X_{n1} + X_{n2} + \dots + X_{nk_n}$$

converge in distribution to the measure ν . (Necessary and sufficient conditions for such a convergence to take place are found in [Bercovici and Pata 1999].) Our main result, Theorem 4.1, implies the following statement. For convenience, we denote by D_ν the singleton $\{t_\nu\}$ if this point exists. Otherwise, $D_\nu = \emptyset$.

Theorem 1.1. *Given any open set $U \supset D_\nu$, the distribution ν_n of S_n is absolutely continuous on $\mathbb{R} \setminus U$ for sufficiently large n , and the density of ν_n converges to the density of ν uniformly and in L^p -norms for $p > 1$ on $\mathbb{R} \setminus U$.*

Note that U can be taken to be empty if $D_\nu = \emptyset$.

In Proposition 5.1, we provide the necessary and sufficient conditions for the existence of the singularity t_ν , as well as a formula to compute it when this point exists. These conditions and the formula are further used to investigate the quality of convergence to freely stable and free compound Poisson densities.

To prove this result, we first approximate ν_n by a closely related \boxplus -infinitely divisible measure ρ_n and we use the fact that G_{ρ_n} is a conformal map. Related considerations appear in the work of Chistyakov and Götze [2013].

The remainder of this paper is organized as follows. In Section 2, we review some relevant preliminaries on free convolution and freely infinitely divisible distributions. Section 3 is devoted to describing the subordination function appearing in free convolution powers. Section 4 contains the proof of our main result, and some examples and applications are given in Section 5.

2. Free convolution and freely infinitely divisible distributions

Let $\mathbb{C}^+ = \{z \in \mathbb{C} : \Im z > 0\}$ be the complex upper half-plane, and let ν be a probability measure on \mathbb{R} . Recall that the Cauchy transform $G_\nu(z)$ of ν is defined by (1-1) for $z \in \mathbb{C}^+$. The measure ν can be recovered as the weak limit of the measures

$$d\nu_y(x) = -\frac{1}{\pi} \Im G_\nu(x + iy) dx, \quad x \in \mathbb{R}, y > 0,$$

as $y \rightarrow 0$, and the atoms of ν can be calculated as follows:

$$(2-1) \quad \lim_{y \rightarrow 0} iy G_\nu(\alpha + iy) = \nu(\{\alpha\}), \quad \alpha \in \mathbb{R}.$$

The reciprocal $F_\nu = 1/G_\nu$ is an analytic self-map of \mathbb{C}^+ and plays a role in the calculation of free convolution. More precisely, for any $\eta > 0$ there exists a positive constant $M = M(\eta, \nu)$ such that the function F_ν has an analytic right inverse F_ν^{-1} (relative to the composition) defined in the truncated cone

$$\Gamma_{\eta, M} = \{x + iy : y > M \text{ and } |x| < \eta y\}.$$

The Voiculescu transform φ_ν of ν is then defined as $\varphi_\nu(z) = F_\nu^{-1}(z) - z$, and for any probability law μ on \mathbb{R} , we have

$$\varphi_{\mu \boxplus \nu}(z) = \varphi_\mu(z) + \varphi_\nu(z)$$

for all z in a region of the form $\Gamma_{\eta, M}$ where all three transforms are defined (see [Bercovici and Voiculescu 1993] for the proof). In this sense, the Voiculescu transform linearizes the free convolution \boxplus .

The set of all finite Borel measures on \mathbb{R} is equipped with the topology of weak convergence from duality with continuous bounded functions. Denoting by \mathcal{M} the class of all Borel probability measures on \mathbb{R} , we can translate weak convergence of measures in \mathcal{M} into convergence properties of the corresponding Voiculescu transforms. We recall the following result from [Bercovici and Pata 1999].

Proposition 2.1. *Let μ, μ_1, μ_2, \dots be measures in \mathcal{M} . Then the sequence μ_n converges weakly to the law μ if and only if there exist $\eta, M > 0$ such that the functions φ_{μ_n} are defined on $\Gamma_{\eta, M}$ for every n , $\lim_{n \rightarrow \infty} \varphi_{\mu_n}(iy) = \varphi_\mu(iy)$ for every $y > M$, and $\varphi_{\mu_n}(iy) = o(y)$ uniformly in n as $y \rightarrow \infty$.*

A measure $\nu \in \mathcal{M}$ is said to be \boxplus -infinitely divisible if for every positive integer n , there exists a measure $\nu_n \in \mathcal{M}$ such that

$$\nu = \underbrace{\nu_n \boxplus \nu_n \boxplus \dots \boxplus \nu_n}_{n \text{ times}}.$$

We denote by $\mathcal{ID}(\boxplus)$ the set of all \boxplus -infinitely divisible measures in \mathcal{M} . It was shown in [Bercovici and Voiculescu 1993] that $\nu \in \mathcal{ID}(\boxplus)$ if and only if the function

φ_ν extends analytically to a map from \mathbb{C}^+ into $\mathbb{C}^- \cup \mathbb{R}$, in which case there exist a real constant γ and a finite Borel measure σ on \mathbb{R} such that φ_ν has the following *free Lévy–Khintchine representation*:

$$\varphi_\nu(z) = \gamma + \int_{\mathbb{R}} \frac{1+tz}{z-t} d\sigma(t).$$

The pair (γ, σ) is uniquely determined. Conversely, given such a pair (γ, σ) , there exists a unique probability law $\nu = \nu_{\boxplus}^{\gamma, \sigma} \in \mathcal{ID}(\boxplus)$ satisfying the above integral formula. We shall call the pair (γ, σ) the *free generating pair* for $\nu_{\boxplus}^{\gamma, \sigma}$. Weak convergence of \boxplus -infinitely divisible laws can be characterized in terms of their free generating pairs; namely, $\nu_{\boxplus}^{\gamma_n, \sigma_n} \rightarrow \nu_{\boxplus}^{\gamma, \sigma}$ weakly if and only if $\gamma_n \rightarrow \gamma$ and $\sigma_n \rightarrow \sigma$ weakly [Barndorff-Nielsen et al. 2006, Theorem 5.13].

We review some useful results related to the F -transforms of freely infinitely divisible distributions, which were proved in [Belinschi and Bercovici 2005; Huang 2015], and are closely related to Biane’s work [1997]. Given $\nu = \nu_{\boxplus}^{\gamma, \sigma}$ in $\mathcal{ID}(\boxplus)$, the function F_ν is a conformal map, and its inverse is the function

$$H_\nu(z) = z + \varphi_\nu(z) = z + \gamma + \int_{\mathbb{R}} \frac{1+tz}{z-t} d\sigma(t), \quad z \in \mathbb{C}^+.$$

This means that $H_\nu(F_\nu(z)) = z$ for all $z \in \mathbb{C}^+$. Note that $H_\nu : \mathbb{C}^+ \rightarrow \mathbb{C}$ is an analytic function satisfying $\Im H_\nu(z) \leq \Im z$ for all $z \in \mathbb{C}^+$. The following result is a consequence of [Belinschi and Bercovici 2005, Theorem 4.6].

Proposition 2.2. *The function F_ν has a one-to-one continuous extension to $\mathbb{C}^+ \cup \mathbb{R}$, and it satisfies*

$$(2-2) \quad |F_\nu(z_1) - F_\nu(z_2)| \geq \frac{1}{2}|z_1 - z_2|, \quad z_1, z_2 \in \mathbb{C}^+ \cup \mathbb{R}.$$

If $\alpha \in \mathbb{R}$ is a point such that $\Im F_\nu(\alpha) > 0$, then F_ν can be continued analytically to a neighborhood of α .

The inequality (2-2) implies that

$$|H_\nu(z_1) - H_\nu(z_2)| \leq 2|z_1 - z_2|, \quad z_1, z_2 \in \Omega_\nu,$$

where $\Omega_\nu = F_\nu(\mathbb{C}^+)$. The function H_ν has a one-to-one continuous extension to the closure $\bar{\Omega}_\nu$. This extension is still denoted H_ν . Thus, we have the following inversion relationships:

$$H_\nu(F_\nu(z)) = z, \quad z \in \mathbb{C}^+ \cup \mathbb{R}, \quad \text{and} \quad F_\nu(H_\nu(z)) = z, \quad z \in \bar{\Omega}_\nu.$$

We describe now the boundary set $\partial\Omega_\nu$. Given $x \in \mathbb{R}$ and $y > 0$, observe

$$\Im H_\nu(x + iy) = y \left(1 - \int_{\mathbb{R}} \frac{1+t^2}{(t-x)^2 + y^2} d\sigma(t) \right).$$

It follows that

$$\Im H_\nu(x + iy) = 0$$

if and only if

$$(2-3) \quad \int_{\mathbb{R}} \frac{1+t^2}{(t-x)^2+y^2} d\sigma(t) = 1.$$

On the other hand, note that for any $x \in \mathbb{R}$, the positive function

$$y \mapsto \int_{\mathbb{R}} \frac{1+t^2}{(t-x)^2+y^2} d\sigma(t)$$

is continuous and strictly decreasing in y , provided that $\sigma \neq 0$; the case $\sigma = 0$ corresponds to a measure ν which is a point mass. Thus, for any $x \in \mathbb{R}$, there exists at most one value $y > 0$ satisfying (2-3). It is natural to introduce two sets

$$A_\nu = \{x \in \mathbb{R} : g(x) > 1\}$$

and

$$B_\nu = \mathbb{R} \setminus A_\nu = \{x \in \mathbb{R} : g(x) \leq 1\},$$

where the function

$$g(x) = \int_{\mathbb{R}} \frac{1+t^2}{(t-x)^2} d\sigma(t) = \sup_{y>0} \int_{\mathbb{R}} \frac{1+t^2}{(t-x)^2+y^2} d\sigma(t), \quad x \in \mathbb{R},$$

is a lower semicontinuous function of x , so that A_ν is an open set. For $x \in A_\nu$, define $u_\nu(x)$ to be the unique y in $(0, \infty)$ satisfying (2-3); for $x \in B_\nu$, set $u_\nu(x) = 0$.

Proposition 2.3 [Huang 2015]. *The function F_ν maps \mathbb{R} bicontinuously to the graph γ_ν of the function u_ν , that is,*

$$F_\nu(\mathbb{R}) = \gamma_\nu = \{x + iu_\nu(x) : x \in \mathbb{R}\}.$$

In particular, the function u_ν is continuous on \mathbb{R} .

We note for further reference that the set A_ν is merely the collection of all $x \in \mathbb{R}$ such that $u_\nu(x) > 0$. Moreover, for any $t \in \mathbb{R}$, we have $\Im F_\nu(t) > 0$ if and only if $\Re F_\nu(t) \in A_\nu$. The graph γ_ν is precisely the boundary set $\partial\Omega_\nu$, and one has $\Omega_\nu = \{z \in \mathbb{C}^+ : H_\nu(z) \in \mathbb{C}^+\}$. The following result now follows easily from these facts; see also [Biane 1997; Huang 2015].

Proposition 2.4. *The function $t \mapsto \Re F_\nu(t)$ is a strictly increasing homeomorphism from \mathbb{R} to \mathbb{R} .*

As shown in [Bercovici and Voiculescu 1993], the measure ν has at most one atom. From (2-1), we see that α is an atom of ν if and only if $F_\nu(\alpha) = 0$ (which gives us the uniqueness of the atom by Proposition 2.2) and the Julia–Carathéodory derivative $F'_\nu(\alpha)$ is finite. (See [Shapiro 1993] for the definition, existence, and

properties of the Julia–Carathéodory derivative.) The value of this derivative is given by

$$F'_v(\alpha) = \frac{1}{v(\{\alpha\})}.$$

By the Stieltjes inversion formula, the density of v (relative to Lebesgue measure) is given by

$$\frac{dv}{dx}(t) = -\frac{1}{\pi} \Im G_v(t) = \frac{1}{\pi} \frac{\Im F_v(t)}{|F_v(t)|^2},$$

at points other than the possible atom α . (This uses the continuous extension of F_v to \mathbb{R} .)

Lemma 2.5. *Consider a measure $v \in \mathcal{ID}(\boxplus)$, and denote by s_v the density of the absolutely continuous part of v . We have $\lim_{|t| \rightarrow \infty} s_v(t) = 0$.*

Proof. Inequality (2-2) implies that

$$|F_v(t) - F_v(i)| \geq \frac{1}{2}|t - i| > \frac{1}{2}|t|, \quad t \in \mathbb{R},$$

so that $|F_v(t)| > \frac{1}{3}|t|$ for $|t| > 6|F_v(i)|$. Then the value of density s_v at such t can be estimated as follows:

$$(2-4) \quad s_v(t) = \frac{1}{\pi} \frac{\Im F_v(t)}{|F_v(t)|^2} \leq \frac{1}{\pi} \frac{1}{|F_v(t)|} < \frac{1}{\pi} \frac{3}{|t|}, \quad |t| > 6|F_v(i)|.$$

The conclusion follows. \square

The preceding result shows that if $F_v(t_v) = 0$, then we must have $|t_v| \leq 6|F_v(i)|$. Moreover, for any $p > 1$ and any neighborhood U of the point t_v , the estimate (2-4) implies that the p -th power $|s_v|^p$ is continuous and integrable over $\mathbb{R} \setminus U$. If such a zero t_v does not exist, then the density s_v is a continuous function which belongs to the space $L^p(\mathbb{R}, dx)$ for all $p > 1$.

The next result follows from the proof of Theorem 4.6 in [Belinschi and Bercovici 2005]. Here we offer a more direct argument.

Lemma 2.6. *The derivative of H_v is nonzero at $z = x + iu_v(x)$, for any $x \in A_v$.*

Proof. We have

$$H'_v(z) = 1 - \int_{\mathbb{R}} \frac{1+t^2}{(z-t)^2} d\sigma(t), \quad z \in \mathbb{C}^+.$$

When $x \in A_v$ and $z = x + iu_v(x)$, a straightforward calculation and the definition of u_v lead to

$$\begin{aligned} \left| \int_{\mathbb{R}} \frac{1+t^2}{(z-t)^2} d\sigma(t) \right| &< \int_{\mathbb{R}} \frac{1+t^2}{|z-t|^2} d\sigma(t) \\ &= \int_{\mathbb{R}} \frac{1+t^2}{(t-x)^2 + u_v(x)^2} d\sigma(t) = 1, \end{aligned}$$

which implies the desired conclusion. \square

Lemma 2.7. *Consider measures $\nu, \nu_n \in \mathcal{ID}(\boxplus)$, $n \in \mathbb{N}$, such that $\nu_n \rightarrow \nu$ weakly as $n \rightarrow \infty$, and let $I \subset \mathbb{R}$ be a compact interval such that the limiting density $d\nu/dx$ is bounded away from zero on I . Then the density $d\nu_n/dx$ converges uniformly on I to $d\nu/dx$ as $n \rightarrow \infty$.*

Proof. Let $(\gamma, \sigma), (\gamma_n, \sigma_n)$ be the free generating pairs of ν and ν_n , respectively. As seen earlier, $\gamma_n \rightarrow \gamma$ and $\sigma_n \rightarrow \sigma$ weakly as $n \rightarrow \infty$. Thus, the sequence H_{ν_n} converges to the function H_ν uniformly on compact subsets of \mathbb{C}^+ .

It is clear that $\Re F_\nu(I) \subset A_\nu$. Thus, by Lemma 2.6, $H'_\nu(z) \neq 0$ for $z \in F_\nu(I)$, and its inverse function F_ν has a conformal continuation to a neighborhood of I . Expressing inverse functions using the Cauchy integral formula, we conclude that, for large n , F_{ν_n} also has a conformal continuation to a neighborhood of I . Moreover, these continuations converge uniformly on I to the continuation of F_ν . Since $0 \notin F_\nu(I)$, the lemma follows from the Stieltjes inversion formula. \square

3. Free convolution powers and subordination functions

Given two probability measures μ_1 and μ_2 on \mathbb{R} , there exist two unique analytic functions $\omega_1, \omega_2 : \mathbb{C}^+ \rightarrow \mathbb{C}^+$ such that $F_{\mu_1 \boxplus \mu_2}(z) = F_{\mu_1}(\omega_1(z)) = F_{\mu_2}(\omega_2(z))$ and

$$F_{\mu_1 \boxplus \mu_2}(z) = \omega_1(z) + \omega_2(z) - z$$

for all $z \in \mathbb{C}^+$ (see [Voiculescu 1993; Biane 1998; Bercovici and Voiculescu 1998]).

Consider now a sequence $\{\mu_n\}_{n=1}^\infty$ in \mathcal{M} and positive integers $k_n \geq 2$, and denote by $\mu_n^{\boxplus k_n}$ the k_n -fold free convolution power of μ_n . Belinschi and Bercovici [2005] showed that $\mu_n^{\boxplus k_n}$ has at most one atom and otherwise $\mu_n^{\boxplus k_n}$ is absolutely continuous, and they studied the analytic subordination for these free convolution powers. Thus, let $\omega_n : \mathbb{C}^+ \rightarrow \mathbb{C}^+$ be the subordination function of $F_{\mu_n^{\boxplus k_n}}$ with respect to F_{μ_n} , that is,

$$F_{\mu_n^{\boxplus k_n}}(z) = F_{\mu_n}(\omega_n(z)).$$

Then we have

$$(3-1) \quad F_{\mu_n^{\boxplus k_n}}(z) = F_{\mu_n}(\omega_n(z)) = \omega_n(z) + \frac{1}{k_n - 1}(\omega_n(z) - z), \quad z \in \mathbb{C}^+.$$

Equation (3-1) implies that the inverse function

$$\omega_n^{-1}(z) = z + (k_n - 1)(z - F_{\mu_n}(z))$$

for $z \in \Gamma_{\eta, M}$, where η, M are positive constants. On the other hand, the function ω_n can be regarded as the F -transform of a unique probability measure on \mathbb{R} by the characterization of F -transforms (see [Bercovici and Voiculescu 1993, Proposition 5.2]). Let ρ_n be the probability measure on \mathbb{R} such that $\omega_n(z) = F_{\rho_n}(z)$, so

$$(3-2) \quad \varphi_{\rho_n}(z) = (k_n - 1)(z - F_{\mu_n}(z)).$$

This implies that the measure ρ_n is \boxplus -infinitely divisible. In particular, the function ω_n extends continuously to $\mathbb{C}^+ \cup \mathbb{R}$ and so, too, does the function $F_{\mu_n^{\boxplus k_n}}$ by (3-1).

Denote by $E_\mu(z) = z - F_\mu(z)$ the self-energy of μ . Given two measures $\mu_1, \mu_2 \in \mathcal{M}$, their Boolean convolution $\mu_1 \uplus \mu_2$, introduced in [Speicher and Woroudi 1997], is the unique probability measure on \mathbb{R} satisfying

$$E_{\mu_1 \uplus \mu_2}(z) = E_{\mu_1}(z) + E_{\mu_2}(z), \quad z \in \mathbb{C}^+.$$

Every probability measure on \mathbb{R} is \uplus -infinitely divisible. Given a measure $\nu \in \mathcal{M}$, the function E_ν is a map from \mathbb{C}^+ to $\mathbb{C}^- \cup \mathbb{R}$ and satisfies $E_\nu(iy)/iy \rightarrow 0$ as $y \rightarrow \infty$. (The latter limit actually holds uniformly for ν in any tight family of probability measures [Bercovici and Voiculescu 1993].) Thus, E_ν admits a unique *Nevanlinna representation*:

$$E_\nu(z) = \gamma + \int_{\mathbb{R}} \frac{1+tz}{z-t} d\sigma(t), \quad z \in \mathbb{C}^+.$$

Conversely, every such formula defines an analytic function which is of the form E_ν for a unique probability measure ν . We will write $\nu = \nu_{\boxplus}^{\gamma, \sigma}$ to indicate this correspondence. Note $E_{\nu_{\boxplus}^{\gamma, \sigma}}(z) = \varphi_{\nu_{\boxplus}^{\gamma, \sigma}}(z)$, and that the map $\nu_{\boxplus}^{\gamma, \sigma} \rightarrow \nu_{\boxplus}^{\gamma, \sigma}$ is a bijective map from the set $\mathcal{ID}(\boxplus)$ into the set \mathcal{M} . Finally, it is easy to verify that if a sequence ν_n converges weakly to a law ν in \mathcal{M} , then the limit $\lim_{n \rightarrow \infty} E_{\nu_n}(z) = E_\nu(z)$ holds for $z \in \mathbb{C}^+$.

We record for further use the following result from [Bercovici and Pata 1999, Theorem 6.3].

Theorem 3.1. *Fix a free generating pair (γ, σ) , a sequence $\{\mu_n\}_{n=1}^\infty$ in \mathcal{M} , and a sequence $\{k_n\}_{n=1}^\infty$ of unbounded positive integers. Then the sequence $\mu_n^{\boxplus k_n}$ converges weakly to $\nu_{\boxplus}^{\gamma, \sigma}$ if and only if the sequence $\mu_n^{\uplus k_n}$ converges weakly to $\nu_{\boxplus}^{\gamma, \sigma}$.*

Boolean limit theorems are used in the proof of the following result.

Proposition 3.2. *Let $\{\mu_n\}_{n=1}^\infty \subset \mathcal{M}$ and let $\{k_n\}_{n=1}^\infty \subset \mathbb{N}$ such that $\lim_{n \rightarrow \infty} k_n = \infty$. Suppose $\mu_n^{\boxplus k_n}$ converges weakly to a law $\nu \in \mathcal{ID}(\boxplus)$. For each n , choose $\rho_n \in \mathcal{ID}(\boxplus)$, such that*

$$F_{\mu_n^{\boxplus k_n}}(z) = F_{\mu_n}(F_{\rho_n}(z)), \quad z \in \mathbb{C}^+.$$

Then $\rho_n \rightarrow \nu$ weakly.

Proof. Assume that (γ, σ) is the free generating pair of ν . By Proposition 2.1, the weak convergence $\mu_n^{\boxplus k_n} \rightarrow \nu_{\boxplus}^{\gamma, \sigma}$ implies the existence of $M > 0$ such that

$$\lim_{n \rightarrow \infty} k_n \varphi_{\mu_n}(iy) = \varphi_{\nu_{\boxplus}^{\gamma, \sigma}}(iy)$$

for all $y > M$, and $k_n \varphi_{\mu_n}(iy) = o(y)$ uniformly in n as $y \rightarrow \infty$. In particular, it follows that the sequence μ_n converges weakly to the unit point mass at 0. On the other hand, Theorem 3.1 shows that $\mu_n^{\uplus k_n} \rightarrow \nu_{\boxplus}^{\gamma, \sigma}$ weakly.

By (3-2), we have

$$\varphi_{\rho_n}(z) = E_{\mu_n^{\boxplus k_n}}(z) - E_{\mu_n}(z), \quad z \in \mathbb{C}^+.$$

Since the two sequences $\{\mu_n^{\boxplus k_n}\}_{n=1}^\infty$ and $\{\mu_n\}_{n=1}^\infty$ are both tight, the last formula implies that $\varphi_{\rho_n}(iy) = o(y)$ uniformly in n as $y \rightarrow \infty$. To determine the limit of $\{\rho_n\}_{n=1}^\infty$, we calculate

$$\lim_{n \rightarrow \infty} \varphi_{\rho_n}(iy) = \lim_{n \rightarrow \infty} [E_{\mu_n^{\boxplus k_n}}(iy) - E_{\mu_n}(iy)] = E_{\nu_{\boxplus}^{\gamma, \sigma}}(iy) = \varphi_{\nu_{\boxplus}^{\gamma, \sigma}}(iy)$$

for every $y > M$. The desired conclusion follows from Proposition 2.1. \square

4. The main result

In the following statement, F_ν is viewed as a continuous function defined on $\mathbb{C}^+ \cup \mathbb{R}$.

Theorem 4.1. *Consider a nondegenerate \boxplus -infinitely divisible distribution ν on \mathbb{R} , a sequence $\{\mu_n\}_{n=1}^\infty$ of probability measures on \mathbb{R} , and a sequence $\{k_n\}_{n=1}^\infty$ of positive integers tending to infinity such that the measures $\mu_n^{\boxplus k_n}$ converge weakly to ν .*

- (1) *If $0 \notin F_\nu(\mathbb{R})$, then the measure ν has no atom and there exists $N > 0$ such that the measure $\mu_n^{\boxplus k_n}$ is Lebesgue absolutely continuous with a continuous density on \mathbb{R} for every $n \geq N$. Moreover, the density of the measure $\mu_n^{\boxplus k_n}$ converges uniformly on \mathbb{R} to the density of the measure ν .*
- (2) *If $0 \in F_\nu(\mathbb{R})$, and $U \subset \mathbb{R}$ is an open interval containing the singleton $F_\nu^{-1}(\{0\})$, then there exists $N > 0$ such that the restriction of the measure $\mu_n^{\boxplus k_n}$ to $\mathbb{R} \setminus U$ is absolutely continuous with a continuous density on $\mathbb{R} \setminus U$ for $n \geq N$. Moreover, the density of the measure $\mu_n^{\boxplus k_n}$ converges uniformly on $\mathbb{R} \setminus U$ to the density of the measure ν .*
- (3) *In all cases, the limit*

$$\lim_{n \rightarrow \infty} \left\| \frac{d\mu_n^{\boxplus k_n}}{dx} - \frac{d\nu}{dx} \right\|_{L^p(\mathbb{R} \setminus U)} = 0$$

holds for $p > 1$, with $U = \emptyset$ in case (1).

Remark. The condition that $0 \in F_\nu(\mathbb{R})$ is necessary for ν to have an atom, but it is not sufficient (see Proposition 5.1). If $F_\nu(t_\nu) = 0$, then the function G_ν extends continuously to all points $t \in \mathbb{R} \setminus \{t_\nu\}$. Theorem 1.1 follows from Theorem 4.1 and this observation.

Proof. As seen earlier, there exist measures $\rho_n \in \mathcal{ID}(\boxplus)$ satisfying

$$F_{\mu_n^{\boxplus k_n}}(z) = F_{\mu_n}(F_{\rho_n}(z)), \quad z \in \mathbb{C}^+.$$

To each n , denote by s_n and s the density of the absolutely continuous part of $\mu_n^{\boxplus k_n}$ and that of ν , respectively. Relation (3-1) shows that $|F_{\mu_n^{\boxplus k_n}} - F_{\rho_n}|$ is small relative to $|F_{\rho_n}|$. Thus, it suffices to focus on the asymptotic behavior of F_{ρ_n} .

Given $\varepsilon > 0$, we first prove that there exists $M > 0$ such that $|s_n(t) - s(t)| < \varepsilon$ for $|t| > M$ and for sufficiently large n . Since the measures ρ_n converge weakly to ν by Proposition 3.2, we have $|F_{\rho_n}(i)| \rightarrow |F_\nu(i)|$ as $n \rightarrow \infty$. In the sequel, we shall only consider the integers n which satisfy the following two conditions:

$$k_n > 13 \quad \text{and} \quad 9|F_\nu(i)| > 6|F_{\rho_n}(i)|.$$

Applying the estimate (2-4) to ρ_n , we get $|F_{\rho_n}(t)| > \frac{1}{3}|t|$ for all such n and for $|t| > 9|F_\nu(i)|$. It follows from (3-1) that $|F_{\mu_n \boxplus k_n}(t)| > \frac{1}{4}|t|$ for the same n and t . Combining this with another application of (2-4) to the density s , we get

$$(4-1) \quad |s_n(t) - s(t)| < \frac{7}{\pi} \frac{1}{|t|}, \quad |t| > 9|F_\nu(i)|,$$

for these large n . Therefore, the desired cutoff constant M can be chosen as

$$M = \max \left\{ 9|F_\nu(i)|, \frac{7}{\varepsilon\pi} \right\}.$$

We conclude that it suffices to prove the uniform convergence of s_n to s on a set of the form $I \setminus U$, where $I = [-M, M]$. To this purpose, fix $I = [-M, M]$ with $M > 0$, and let $\delta > 0$ be arbitrary but fixed. Recall that the map

$$t \mapsto \Re F_\nu(t)$$

is an increasing homeomorphism of \mathbb{R} . Thus, the set

$$J = \Re F_\nu(I) = \{x \in \mathbb{R} : \Re F_\nu(-M) \leq x \leq \Re F_\nu(M)\}$$

is a compact interval. Set

$$\Gamma = \{x \in J : u_\nu(x) \geq \delta\}$$

and

$$\Delta = \left\{ x \in J : u_\nu(x) > \frac{\delta}{2} \right\}.$$

We have $\Gamma \subset \Delta \subset J$, where Γ is closed, and Δ is relatively open in J . We conclude that Γ is contained in the union of finitely many connected components of Δ . Taking the closure of those components, we find a finite family J_1, J_2, \dots, J_K of pairwise disjoint, closed intervals such that

$$\Gamma \subset \bigcup_{1 \leq \ell \leq K} J_\ell \subset \bar{\Delta}.$$

We have $u_\nu \geq \delta/2$ on the union $\bigcup_{1 \leq \ell \leq K} J_\ell$ and $u_\nu \leq \delta$ on the complement $J' = J \setminus (\bigcup_{1 \leq \ell \leq K} J_\ell)$.

Denote $I_\ell = \{t \in I : \Re F_\nu(t) \in J_\ell\}$ for each $1 \leq \ell \leq K$. Note that

$$\Im F_\nu(t) \geq \delta/2$$

for each $t \in \bigcup_{1 \leq \ell \leq K} I_\ell$. Thus, the density of ν is bounded away from zero on $\bigcup_{1 \leq \ell \leq K} I_\ell$. From Lemma 2.7, we see that the functions F_ν and F_{ρ_n} both extend

analytically to a neighborhood of the set $\bigcup_{1 \leq \ell \leq K} I_\ell$ for sufficiently large n . These extensions are injective. Moreover, the convergence $F_{\rho_n} \rightarrow F_\nu$ holds uniformly in that neighborhood. By virtue of (3-1), we conclude that the functions $F_{\mu_n^{\boxplus k_n}}$ will have the same behavior on the set $\bigcup_{1 \leq \ell \leq K} I_\ell$ as $n \rightarrow \infty$. It follows that the measure $\mu_n^{\boxplus k_n}$ has no atom in the union $\bigcup_{1 \leq \ell \leq K} I_\ell$ for large n and $s_n \rightarrow s$ uniformly on this set by the Stieltjes inversion formula.

We prove next the uniform convergence on the set I' (or on $I' \setminus U$), where

$$(4-2) \quad I' = \{t \in I : \Re F_\nu(t) \in J'\} = I \setminus \left(\bigcup_{1 \leq \ell \leq K} I_\ell \right).$$

We claim that

$$(4-3) \quad \sup_{x \in J'} u_{\rho_n}(x) \leq 2\delta$$

for sufficiently large n . Assume, to get a contradiction, that there exist positive integers $n_1 < n_2 < \dots \rightarrow \infty$ and points $x_1, x_2, \dots \in J'$ such that $u_{\rho_{n_k}}(x_k) > 2\delta$. By the definition of u_{ρ_n} given in Section 2, we have

$$(4-4) \quad \int_{\mathbb{R}} \frac{1+t^2}{(t-x_k)^2 + u_{\rho_{n_k}}(x_k)^2} d\sigma_{n_k}(t) = 1, \quad k \geq 1,$$

where σ_{n_k} is the free generating measure of ρ_{n_k} . By passing to a subsequence if necessary, we assume that $x_k \rightarrow x_0 \in \bar{J}'$ as $k \rightarrow \infty$. Then, denoting $\nu_{\boxplus}^{\gamma, \sigma}$ by ν , the identity (4-4) and the fact that $\sigma_n \rightarrow \sigma$ weakly imply

$$1 \leq \int_{\mathbb{R}} \frac{1+t^2}{(t-x_k)^2 + (2\delta)^2} d\sigma_{n_k}(t) \rightarrow \int_{\mathbb{R}} \frac{1+t^2}{(t-x_0)^2 + (2\delta)^2} d\sigma(t)$$

as $k \rightarrow \infty$. We conclude that $u_\nu(x_0) \geq 2\delta$, which is in contradiction to the fact that $x_0 \in \bar{J}'$. Thus, the estimate (4-3) is proved.

The rest of the proof is divided into two cases according to whether $U = \emptyset$ or $U \neq \emptyset$. By translating the measure ν if necessary, we may assume that $\Re F_\nu(0) = 0$.

Case (1): $0 \notin F_\nu(\mathbb{R})$ and $U = \emptyset$. In this case, $u_\nu(0) > 0$ and thus $0 \in A_\nu$. Since the set A_ν is open, there exists a small number $a > 0$ such that the interval $[-4a, 4a]$ is contained in A_ν . By considering a smaller δ if necessary, we may assume further that

$$(4-5) \quad [-4a, 4a] \subset \bigcup_{1 \leq \ell \leq K} J_\ell.$$

Since the map $t \mapsto \Re F_\nu(t)$ is an increasing homeomorphism of \mathbb{R} , the uniform convergence of $F_{\rho_n} \rightarrow F_\nu$ on $\bigcup_{1 \leq \ell \leq K} I_\ell$ implies that there exists some integer $N > 0$ such that

$$[-2a, 2a] \subset \left\{ \Re F_{\rho_n}(t) : t \in \bigcup_{1 \leq \ell \leq K} I_\ell \right\}, \quad n \geq N.$$

Since the map $t \mapsto \Re F_{\rho_n}(t)$ is also a homeomorphism of the same nature, we have

$$\inf_{t \in I'} |\Re F_{\rho_n}(t)| \geq 2a, \quad n \geq N,$$

by recalling the definition (4-2) of the complement I' . Using (3-1) and enlarging N if necessary, we conclude that

$$(4-6) \quad \inf_{t \in I'} |\Re F_{\mu_n^{\boxplus k_n}}(t)| \geq a, \quad n \geq N.$$

Further enlarging N , the inequality (4-3) and the relation (3-1) imply that

$$(4-7) \quad \Im F_{\mu_n^{\boxplus k_n}}(t) \leq 3\delta, \quad t \in I', \quad n \geq N.$$

From (4-6) and (4-7), we see that

$$0 \leq s_n(t) \leq \frac{3\delta}{a^2\pi}$$

for $t \in I'$ and $n \geq N$. On the other hand, the relation (4-5) and the fact that $u_v \leq \delta$ on J' yield

$$0 \leq s(t) \leq \frac{\delta}{16a^2\pi}$$

for $t \in I'$. As the parameter δ can be arbitrarily small, we have proved the uniform convergence of $s_n \rightarrow s$ on I' . This finishes the proof of Theorem 4.1(1).

Case (2): $0 \in F_v(\mathbb{R})$. In this case, $u_v(0) = 0$ and $F_v(0) = 0 = H_v(0)$ by our normalization. Let a_n be the unique real number such that $\Re F_{\rho_n}(a_n) = 0$ (and hence $F_{\rho_n}(a_n) = iu_{\rho_n}(0)$). We first show that a_n is small for large n . To this end, we write $U = (-2b, 2b)$ where $b > 0$ and set $c = b/5$. Observe that $H_v(ic) \in \mathbb{C}^+$, and that the Lipschitz property of H_v yields

$$|H_v(ic)| = |H_v(ic) - H_v(0)| \leq 2c.$$

Since $\lim_{n \rightarrow \infty} H_{\rho_n}(ic) = H_v(ic)$, there exists an integer $N > 0$ such that $H_{\rho_n}(ic) \in \mathbb{C}^+$ for all $n \geq N$. Consequently, we have $u_{\rho_n}(0) < c$ for such n ; for if $u_{\rho_n}(0) \geq c > 0$, we will get

$$\begin{aligned} 1 &= \int_{\mathbb{R}} \frac{1+t^2}{t^2 + u_{\rho_n}(0)^2} d\sigma_n(t) \leq \int_{\mathbb{R}} \frac{1+t^2}{t^2 + c^2} d\sigma_n(t) \\ &= 1 - \frac{1}{c} \Im H_{\rho_n}(ic) < 1, \end{aligned}$$

a contradiction. Note further that

$$|H_{\rho_n}(ic) - a_n| = |H_{\rho_n}(ic) - H_{\rho_n}(iu_{\rho_n}(0))| \leq 2(c - u_{\rho_n}(0)) \leq 2c$$

for all $n \geq N$. (We have used the inversion relationship $a_n = H_{\rho_n}(F_{\rho_n}(a_n))$ here.) Therefore, by enlarging N if necessary, we conclude that $|a_n| < 5c = b$ for $n \geq N$.

Now, (2-2) shows that for any $t \in I' \setminus U$ and $n \geq N$, we have

$$|F_{\rho_n}(t) - F_{\rho_n}(a_n)| \geq \frac{1}{2}|t - a_n| > \frac{b}{2}.$$

This implies further that

$$|F_{\rho_n}(t)| > \frac{b}{2} - |F_{\rho_n}(a_n)| = \frac{b}{2} - |u_{\rho_n}(0)| > \frac{b}{4}, \quad t \in I' \setminus U, \quad n \geq N.$$

In other words, for such values of t and n , $|F_{\rho_n}(t)|$ is always bounded away from zero. Then an argument similar to the proof of Case (1) yields the absolute continuity of the free convolution $\mu_n^{\boxplus k_n}$ and the uniform convergence $s_n \rightarrow s$ on $I' \setminus U$, finishing the proof of Theorem 4.1(2).

Finally, the L^p -convergence result in Theorem 4.1(3) follows from the estimate (4-1) and the dominated convergence theorem. \square

Remark (Local analyticity and approximation). An important feature of superconvergence are the analyticity properties of the distributions in the limiting process. Indeed, under the weak convergence assumption of Theorem 4.1, if I is a finite interval on which the limit density $d\nu/dx$ is bounded away from zero (and hence it admits an analytic continuation to a neighborhood of I), then the restriction of the free convolution $\mu_n^{\boxplus k_n}$ on I becomes absolutely continuous in finite time and its density continues analytically to a neighborhood of I . Moreover, these extensions can be approximated uniformly by the analytic continuation of $d\nu/dx$ on I , thanks to Lemma 2.7 and the identity (3-1).

5. Applications

In this section, we apply our main result to some of the most important limit theorems in free probability. We begin by examining the geometric condition: $0 \in F_\nu(\mathbb{R})$. Note that the singular integral in the following result takes values in $(0, \infty]$.

Proposition 5.1. *Let $\nu = \nu_{\boxplus}^{\gamma, \sigma}$ be a nondegenerate law in $\mathcal{ID}(\boxplus)$. We have:*

(1) $0 \in F_\nu(\mathbb{R})$ if and only if

$$(5-1) \quad L = \sup_{\varepsilon > 0} \frac{-\Im \varphi_\nu(i\varepsilon)}{\varepsilon} = \int_{\mathbb{R}} \frac{1+t^2}{t^2} d\sigma(t) \leq 1.$$

In this case, the value of the unique zero t_ν of F_ν is given by

$$t_\nu = \gamma - \int_{\mathbb{R}} \frac{1}{t} d\sigma(t).$$

(2) $\nu(\{t_\nu\}) > 0$ if and only if $L < 1$, and we have $\nu(\{t_\nu\}) = 1 - L$ in this case.

Proof. The identity

$$\sup_{\varepsilon > 0} (-\Im \varphi_\nu(i\varepsilon))/\varepsilon = \int_{\mathbb{R}} \frac{1+t^2}{t^2} d\sigma(t)$$

follows from the free Lévy–Khintchine formula

$$-\Im \varphi_\nu(i\varepsilon) = \varepsilon \int_{\mathbb{R}} \frac{1+t^2}{\varepsilon^2+t^2} d\sigma(t)$$

and the monotone convergence theorem, and we see that the supremum here is in fact a genuine limit:

$$\sup_{\varepsilon>0} (-\Im \varphi_\nu(i\varepsilon))/\varepsilon = \lim_{\varepsilon \rightarrow 0^+} (-\Im \varphi_\nu(i\varepsilon))/\varepsilon.$$

Next, recall from [Belinschi and Bercovici 2005, Proposition 4.7] that $0 \in F_\nu(\mathbb{R})$ if and only if the limit

$$t_\nu = H_\nu(0) = \lim_{\varepsilon \rightarrow 0^+} H_\nu(i\varepsilon)$$

exists, $t_\nu \in \mathbb{R}$, and the Julia–Carathéodory derivative $H'_\nu(0) \geq 0$. Note that if the limit t_ν exists and is real, then the derivative

$$(5-2) \quad H'_\nu(0) = \lim_{\varepsilon \rightarrow 0^+} \frac{\Im H_\nu(i\varepsilon)}{\varepsilon}$$

always exists and belongs to the interval $[-\infty, 1)$. Moreover, if $0 \in F_\nu(\mathbb{R})$ and $H'_\nu(0) > 0$ then we have the Julia–Carathéodory derivative $F'_\nu(t_\nu) = 1/H'_\nu(0)$.

Now, if $0 \in F_\nu(\mathbb{R})$, then we know the limit $t_\nu \in \mathbb{R}$. Hence, (5-2) implies $H'_\nu(0) = 1 - L$. Since $H'_\nu(0) \geq 0$ in this case, we conclude that $1 \geq L$. On the other hand, since $F_\nu(\mathbb{R}) = \partial\Omega_\nu$, the inversion formula shows that

$$F_\nu(t_\nu) = F_\nu(H_\nu(0)) = 0.$$

Conversely, if the singular integral L converges and $1 \geq L$, then we have $\Im H_\nu(i\varepsilon) \rightarrow 0 \cdot (1 - L) = 0$ as $\varepsilon \rightarrow 0^+$. On the other hand, the estimate

$$\frac{|t|}{\varepsilon^2+t^2} \leq \frac{1+t^2}{\varepsilon^2+t^2} \leq \frac{1+t^2}{t^2} \in L^1(\sigma), \quad t \in \mathbb{R}, \varepsilon > 0,$$

and the dominated convergence theorem imply that the function $t \mapsto 1/t$ belongs to $L^1(\sigma)$ and

$$\Re H_\nu(i\varepsilon) = \gamma + (\varepsilon^2 - 1) \int_{\mathbb{R}} \frac{t}{\varepsilon^2+t^2} d\sigma(t) \rightarrow \gamma - \int_{\mathbb{R}} \frac{1}{t} d\sigma(t)$$

as $\varepsilon \rightarrow 0^+$. It follows that the vertical limit t_ν is equal to

$$\gamma - \int_{\mathbb{R}} \frac{1}{t} d\sigma(t) \in \mathbb{R}.$$

As seen earlier, this fact and the formula (5-2) imply that $H'_\nu(0) = 1 - L$. Therefore, we have $H'_\nu(0) \geq 0$, and the proof of (1) is finished.

The statement (2) follows from the fact that the derivative $F'_\nu(t_\nu) = 1/\nu(\{t_\nu\})$. \square

We remark that the results in [Belinschi and Bercovici 2005] were proved using Denjoy–Wolff analysis for boundary fixed points of analytic self-maps on \mathbb{C}^+ . A different approach to the same results has been used in [Huang and Wang 2015], which yields a more general description for the points on the boundary set $\partial\Omega_\nu$.

Stable approximation. Recall that two measures $\mu, \nu \in \mathcal{M}$ are said to have the same *type* (and we write $\mu \sim \nu$) if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that $\mu(E) = \nu(aE + b)$ for all Borel sets $E \subset \mathbb{R}$. The relation \sim is an equivalence relationship among all probability laws, and hence the set \mathcal{M} is partitioned into a union of distributions with inequivalent types. A nondegenerate distribution $\nu \in \mathcal{M}$ is said to be \boxplus -stable if $\nu \sim \nu_1 \boxplus \nu_2$ whenever $\nu_1 \sim \nu \sim \nu_2$. Clearly, within one type either all distributions are stable or else none of them is stable.

Each \boxplus -stable law ν is associated with a unique *stability index* $\alpha \in (0, 2]$, so that if X and Y are free random variables drawn from the same law ν and $a, b > 0$, then the distribution of the sum $aX + bY$ is a translate of the distribution of the scaled variable $(a^\alpha + b^\alpha)^{1/\alpha} X$. Stable laws of the same type share the same index.

Freely stable laws are \boxplus -infinitely divisible and absolutely continuous, and they can be classified using the stability index α . Following [Bercovici and Voiculescu 1993], every \boxplus -stable law has the same type as a unique distribution whose Voiculescu transform falls into the following list:

- (1) $\varphi(z) = 1/z$ for $\alpha = 2$.
- (2) $\varphi(z) = bz^{1-\alpha}$ for $1 < \alpha < 2$, where $|b| = 1$ and $\arg b \in [(\alpha - 2)\pi, 0]$.
- (3) $\varphi(z) = bz^{1-\alpha}$ for $0 < \alpha < 1$, where $|b| = 1$ and $\arg b \in [\pi, (1 + \alpha)\pi]$.
- (4) $\varphi(z) = -2bi + [2(2b - 1)/\pi] \log z$ for $\alpha = 1$, where $b \in [0, 1]$.

Here, the complex power and logarithmic functions are given by their principal value in \mathbb{C}^+ . One can also find a formula for the density of the \boxplus -stable laws in [Bercovici and Pata 1999]. Above all, we mention that the case $\alpha = 2$ corresponds to the stable type of the standard semicircular law.

The interest in the class of freely stable laws arises from the fact that a measure ν is \boxplus -stable if and only if there exist a sequence $\{X_i\}_{i=1}^\infty$ of identically distributed free random variables and constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that the distribution of the normalized sum $S_n = \sum_{i=1}^n (X_i - b_n)/a_n$ converges weakly to the law ν . In this case, the common distribution of the sequence X_i is said to belong to the *free domain of attraction* of the stable law ν . Thus, up to a change of scale and location, the distributional behavior of a large free convolution $\mu^{\boxplus n}$ for a measure μ in a free domain of attraction can be estimated using the corresponding freely stable law.

Free domains of attraction for \boxplus -stable laws are determined in [Bercovici and Pata 1999], showing that these domains of attraction coincide with their classical counterparts relative to the classical convolution. In the semicircular case, the free

domain of attraction consists of all nondegenerate measures $\mu \in \mathcal{M}$ such that the truncated variance function

$$H_\mu(x) = \int_{-x}^x t^2 d\mu(t), \quad x > 0,$$

satisfies $\lim_{x \rightarrow \infty} H_\mu(cx)/H_\mu(x) = 1$ for any given $c > 0$. This is in parallel to the classical theory of central limit theorems, that is, convergence to a Gaussian law.

With that being said, the following result shows that the quality of freely stable approximation is in fact much better than its classical counterpart. This result is stated in the general framework of triangular arrays with identical rows.

Proposition 5.2. *Let ν be a \boxplus -stable law for which the weak approximation $\mu_n^{\boxplus k_n} \rightarrow \nu$ holds. Then the measure $\mu_n^{\boxplus k_n}$ superconverges to the law ν on \mathbb{R} .*

Proof. This is a direct consequence of Theorem 4.1 and the criterion (5-1). Indeed, one has $L = \infty$ in all cases of the index α , which implies that $0 \notin F_\nu(\mathbb{R})$. \square

In particular, the preceding result generalizes the superconvergence for measures with finite variance in [Wang 2010] to the entire free domain of attraction of the semicircular law.

Notice that stable approximation to the free sum S_n could fail for any choice of constants a_n and b_n if the common distribution μ of the summands X_i does not belong to any free domain of attraction, but even in this case one may still have weak convergence along some subsequence S_{k_n} . The limit ν in this situation is necessarily \boxplus -infinitely divisible, and hence Theorem 4.1 still applies to this case. The law μ in this case is said to belong to the *free domain of partial attraction* of the law ν . In fact, a probability distribution is \boxplus -infinitely divisible if and only if its free domain of partial attraction is nonempty. It is also well known that the domain of partial attraction of a stable law is strictly larger than its domain of attraction in both free and classical theories. We refer to [Bercovici and Pata 1999] for the details of these results.

Poisson approximation. Here we study an example of freely infinitely divisible approximation relative to Poisson type limit theorems. Let μ be an arbitrary probability measure on \mathbb{R} , $\mu \neq \delta_0$, and let $\lambda > 0$ be a given parameter. Recall that the *compound free Poisson distribution* $\nu_{\lambda, \mu}$ with rate λ and jump distribution μ is the weak limit of

$$[(1 - \lambda/n)\delta_0 + (\lambda/n)\mu]^{\boxplus n}$$

as $n \rightarrow \infty$ [Voiculescu et al. 1992]. The law $\nu_{\lambda, \mu}$ is \boxplus -infinitely divisible, and its free generating pair is given by

$$\gamma = \lambda \int_{\mathbb{R}} \frac{t}{1+t^2} d\mu(t), \quad d\sigma(t) = \lambda \frac{t^2}{1+t^2} d\mu(t).$$

Thus, we see immediately that $L = \lambda$ and $t_{v_{\lambda,\mu}} = 0$ in this case, which leads further to the following result:

Proposition 5.3. *The origin is an atom of mass $1 - \lambda$ for the law $v_{\lambda,\mu}$ if and only if the parameter λ is less than 1. If $\lambda > 1$, then the superconvergence phenomenon in any weak approximation $\mu_n^{\boxplus k_n} \rightarrow v_{\lambda,\mu}$ holds globally on \mathbb{R} .*

Note the case $\mu = \delta_1$ corresponds to the approximation by Marčenko–Pastur law:

$$dv_{\lambda,\delta_1}(t) = \begin{cases} \frac{1}{2\pi t} \sqrt{4\lambda - (t - 1 - \lambda)^2} \chi(t) dt & \text{if } \lambda \geq 1; \\ (1 - \lambda)\delta_0 + \frac{1}{2\pi t} \sqrt{4\lambda - (t - 1 - \lambda)^2} \chi(t) dt & \text{if } 0 < \lambda < 1, \end{cases}$$

where χ stands for the indicator function of the open interval $((1 - \sqrt{\lambda})^2, (1 + \sqrt{\lambda})^2)$. Clearly, the measure v_{1,δ_1} has no atom and yet $F_{v_{1,\delta_1}}(0) = 0$.

Acknowledgements

Bercovici was supported in part by a grant of the National Science Foundation. Wang was supported by the NSERC Canada Discovery Grant RGPIN-2016-03796. Zhong acknowledges support provided by a Leverhulme Trust Research Project Grant when he was a postdoctoral fellow at Lancaster University where part of this work was done. He was also supported in part by NSFC (no. 11501423, no. 71301164), China Postdoctoral Science Foundation (no. 2015M570662) and Fundamental Research Funds for the Central Universities (no. 201410500069). The authors would like to thank an anonymous referee for providing comments and suggestions which improve the quality of this paper.

References

- [Anshelevich et al. 2014] M. Anshelevich, J.-C. Wang, and P. Zhong, “Local limit theorems for multiplicative free convolutions”, *J. Funct. Anal.* **267**:9 (2014), 3469–3499. MR Zbl
- [Barndorff-Nielsen et al. 2006] O. E. Barndorff-Nielsen, U. Franz, R. Gohm, B. Kümmerer, and S. Thorbjørnsen, *Quantum independent increment processes, II*, edited by M. Schürmann and U. Franz, Lecture Notes in Mathematics **1866**, Springer, Berlin, 2006. MR Zbl
- [Belinschi and Bercovici 2005] S. T. Belinschi and H. Bercovici, “Partially defined semigroups relative to multiplicative free convolution”, *Int. Math. Res. Not.* **2005**:2 (2005), 65–101. MR Zbl
- [Bercovici and Pata 1999] H. Bercovici and V. Pata, “Stable laws and domains of attraction in free probability theory”, *Ann. of Math. (2)* **149**:3 (1999), 1023–1060. MR Zbl
- [Bercovici and Voiculescu 1993] H. Bercovici and D. Voiculescu, “Free convolution of measures with unbounded support”, *Indiana Univ. Math. J.* **42**:3 (1993), 733–773. MR Zbl
- [Bercovici and Voiculescu 1995] H. Bercovici and D. Voiculescu, “Superconvergence to the central limit and failure of the Cramér theorem for free random variables”, *Probab. Theory Related Fields* **103**:2 (1995), 215–222. MR Zbl

- [Bercovici and Voiculescu 1998] H. Bercovici and D. Voiculescu, “Regularity questions for free convolution”, pp. 37–47 in *Nonselfadjoint operator algebras, operator theory, and related topics*, edited by H. Bercovici and C. Foias, Oper. Theory Adv. Appl. **104**, Birkhäuser, Basel, 1998. MR Zbl
- [Biane 1997] P. Biane, “On the free convolution with a semi-circular distribution”, *Indiana Univ. Math. J.* **46**:3 (1997), 705–718. MR Zbl
- [Biane 1998] P. Biane, “Processes with free increments”, *Math. Z.* **227**:1 (1998), 143–174. MR Zbl
- [Chistyakov and Götze 2013] G. Chistyakov and F. Götze, “Free infinitely divisible approximations of n -fold free convolutions”, pp. 225–237 in *Prokhorov and contemporary probability theory*, edited by A. N. Shiryaev et al., Springer Proc. Math. Stat. **33**, Springer, Heidelberg, 2013. MR Zbl
- [Huang 2015] H.-W. Huang, “Supports of measures in a free additive convolution semigroup”, *Int. Math. Res. Not.* **2015**:12 (2015), 4269–4292. MR Zbl
- [Huang and Wang 2015] H.-W. Huang and J.-C. Wang, “Regularization by free Lévy process”, preprint, 2015.
- [Kargin 2007] V. Kargin, “On superconvergence of sums of free random variables”, *Ann. Probab.* **35**:5 (2007), 1931–1949. MR Zbl
- [Shapiro 1993] J. H. Shapiro, *Composition operators and classical function theory*, Springer, New York, 1993. MR Zbl
- [Speicher and Woroudi 1997] R. Speicher and R. Woroudi, “Boolean convolution”, pp. 267–279 in *Free probability theory* (Waterloo, ON, 1995), edited by D. Voiculescu, Fields Inst. Commun. **12**, American Mathematical Society, Providence, RI, 1997. MR Zbl
- [Voiculescu 1993] D. Voiculescu, “The analogues of entropy and of Fisher’s information measure in free probability theory, I”, *Comm. Math. Phys.* **155**:1 (1993), 71–92. MR Zbl
- [Voiculescu et al. 1992] D. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, CRM Monograph Series **1**, American Mathematical Society, Providence, RI, 1992. MR Zbl
- [Wang 2010] J.-C. Wang, “Local limit theorems in free probability theory”, *Ann. Probab.* **38**:4 (2010), 1492–1506. MR Zbl

Received October 9, 2015. Revised July 1, 2017.

HARI BERCOVICI
DEPARTMENT OF MATHEMATICS
RAWLES HALL
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405
UNITED STATES
bercovic@indiana.edu

JIUN-CHAU WANG
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF SASKATCHEWAN
SASKATOON, S7N 5E6
CANADA
jcwang@math.usask.ca

PING ZHONG

DEPARTMENT OF PURE MATHEMATICS

UNIVERSITY OF WATERLOO

WATERLOO ONTARIO N2L 3G1

CANADA

and

SCHOOL OF MATHEMATICS AND STATISTICS

WUHAN UNIVERSITY

HUBEI

CHINA 430072

ping.zhong@uwaterloo.ca

NORM CONSTANTS IN CASES OF THE CAFFARELLI–KOH–NIRENBERG INEQUALITY

AKSHAY L. CHANILLO, SAGUN CHANILLO AND ALI MAALAOUI

Based on elementary linear algebra, we provide radically simplified proofs using quasiconformal changes of variables to obtain sharp constants and optimizers in cases of the Caffarelli–Kohn–Nirenberg inequality. Some of our results were obtained earlier by Lam and Lu.

1. Introduction

The aim of this elementary note is to obtain sharp constants and also the minimizer for special cases of the Caffarelli–Kohn–Nirenberg (CKN) inequality via change of variables employing a suitable quasiconformal map. We start by recalling the main inequality that we are interested in, proved first in [Caffarelli et al. 1984]:

$$(1-1) \quad \left(\int_{\mathbb{R}^n} |f(x)|^r |x|^{\alpha n} dx \right)^{1/r} \leq C \left(\int_{\mathbb{R}^n} |f(x)|^s |x|^{\alpha n} dx \right)^{(1-t)/s} \left(\int_{\mathbb{R}^n} |\nabla f(x)|^p |x|^{\alpha(n-p)} dx \right)^{t/p},$$

where $1 \leq p < n$ and $1/r = (1-t)/s + t(n-p)/(np)$ with $0 \leq t \leq 1$ and $\alpha > -1$.

In our investigation of the best constant C for which inequality (1-1) holds, we will use the constant $M(s, r, p)$, for s, r and p as above, appearing in the inequality

$$(1-2) \quad \|f\|_r \leq M(s, r, p) \|f\|_s^{1-t} \|\nabla f\|_p^t.$$

The constant and optimizers of this last inequality are the subjects of this result:

Theorem 1.1 [del Pino and Dolbeault 2013]. *Let $1 < p \leq n$ and $r > 1$ such that $r \leq p(n-1)/(n-p)$, and $s = p(r-1)/(p-1)$. There exists a positive constant $M(s, r, p)$ such that for every $f \in L^r(\mathbb{R}^n)$ with $\nabla f \in L^p(\mathbb{R}^n)$, we have*

$$\begin{cases} \|f\|_r \leq M(s, r, p) \|f\|_s^{1-t} \|\nabla f\|_p^t & \text{if } r > p, \\ \|f\|_s \leq M(s, r, p) \|f\|_r^{1-t} \|\nabla f\|_p^t & \text{if } r < p, \end{cases}$$

MSC2010: 46E35.

Keywords: Caffarelli–Kohn–Nirenberg inequality, weighted Sobolev inequality, quasiconformal maps, sharp constants, optimizers.

where t is taken as above. Moreover the best constant is achieved by the function:

$$(1-3) \quad f_0(x) = A(1 + B|x - x_0|^{p/(p-1)})^{-(p-1)/(r-p)},$$

with A and $B \in \mathbb{R}$, $x_0 \in \mathbb{R}^n$ and B has the sign of $a - p$.

The proof of (1-2) follows by noticing that

$$\|f\|_r \leq \|f\|_s^{1-t} \|f\|_{np/(n-p)}^t.$$

We then apply the Sobolev inequality to the second term on the right

$$\|f\|_{np/(n-p)} \leq C_p \|\nabla f\|_p,$$

to conclude. In this paper, we show that the optimization problem for the best constant of (1-1) exhibit different behavior for the two cases $\alpha > 0$ and $0 > \alpha > -1$. In the case $0 > \alpha > -1$, we compute the best constant and we show that the optimizer is radial. In the case $\alpha > 0$, we also compute the best constant and we show that in this case there is a break in the symmetry of the optimizers, since the best constant cannot be obtained by radial functions anymore. More importantly we establish that there is no optimizer. We mainly rely on a study of the eigenvectors and eigenvalues of the differential of the quasiconformal change of variable that we will use. In the sequel we set $\phi(x) = x|x|^\alpha$. The results of this paper can be stated as follows:

Theorem 1.2. *The sharp constant in the CKN inequality (1-1) for $-1 < \alpha < 0$ and any function f , radial or otherwise, is given by*

$$(1 + \alpha)^{t/n-t} M(s, r, p),$$

where $M(s, r, p)$ is the constant in (1-2). Moreover the optimizer of (1-1) is then a radial function and can be taken to be $f \circ \phi$, where f can be taken to be the radial optimizers in the cases investigated in [del Pino and Dolbeault 2013], namely (1-3).

The result in Theorem 1.2 was established earlier by Nguyen Lam and Guozhen Lu [2017] using the quasiconformal map that we use in our work. Our proof in part is a radically simplified approach. A very comprehensive list of references on this topic is found in [Lam and Lu 2017]. In [Dolbeault et al. 2016], the authors also investigate the symmetry and symmetry breaking of the optimizers of the CKN, in the case $p = 2$ and $t = 1$ using a nonlinear flow approach. In the case $t \neq 1$ and $p = 2$, the best constants of the CKN inequality were investigated in [Dolbeault and Esteban 2012]. In addition to Theorem 1.2 we prove a symmetry breaking phenomenon for the case $\alpha > 0$:

Theorem 1.3. *The sharp constant for the CKN inequality for $\alpha > 0$ is given by*

$$(1 + \alpha)^{t/n} M(s, r, p).$$

Moreover, there is no optimizer for the inequality and this constant is strictly bigger than the one obtained for radial functions.

Our proofs also show the following theorem for radial functions for all $\alpha > -1$:

Theorem 1.4. *The sharp constant for inequality (1-1) restricted to radial functions for $\alpha > -1$ is given by*

$$(1 + \alpha)^{t/n-t} M(s, r, p).$$

2. Proofs of the theorems

Our aim is now to make quasiconformal changes of variables in (1-2) with explicit information on the Jacobian and eigenvalues of the quasiconformal changes of variables. To this end, the eigenvalues and Jacobians can be calculated explicitly using a linear algebra trick in [Chanillo and Torchinsky 1986, Lemma 2.23, p. 14]. This method was introduced there to calculate the Hessian that appears in stationary phase calculations, but the linear algebra idea goes back to the calculation of what are called permanents and bordered matrices and may be found in classical books on algebra.

If we set $y = \phi(x)$ in (1-2), then

$$Dy = D\phi(x)$$

and so if we set $A = D\phi$ and use the chain rule, (1-2) becomes

$$(2-1) \quad \left(\int_{\mathbb{R}^n} |f \circ \phi(x)|^r |J_\phi(x)| dx \right)^{1/r} \leq M(s, r, p) \left(\int_{\mathbb{R}^n} |f \circ \phi(x)|^s |J_\phi(x)| dx \right)^{(1-t)/s} \\ \times \left(\int_{\mathbb{R}^n} |(A^{-1})^* \nabla(f \circ \phi)|^p |J_\phi(x)| dx \right)^{t/p}.$$

Here J_ϕ is the Jacobian of the map $y = \phi(x)$, that is $|J_\phi(x)| = |\det D\phi(x)|$ and B^* denotes the transpose of the matrix B . To obtain the Caffarelli–Kohn–Nirenberg inequalities in some cases we simply choose the explicit quasiconformal map (see also [Chanillo and Wheeden 1992]):

$$(2-2) \quad \phi(x) = x|x|^\alpha, \quad \alpha > -1.$$

Thus the goal is to calculate explicitly the eigenvalues of the differential of (2-2) and thus we have full information of the matrix A above and in particular the Jacobian of (2-2).

Remark. We remark that using [del Pino and Dolbeault 2013] we can also consider the case of $p = n$ and the Onofri inequality and Moser–Trudinger type inequalities.

Lemma 2.1. *Given the map $\phi(x)$ as in (2-2), the differential $A = D\phi(x)$ is unitarily diagonalizable and the eigenvalues of A are given by*

$$\lambda_1 = (1 + \alpha)|x|^\alpha, \quad \lambda_2 = \cdots = \lambda_n = |x|^\alpha.$$

Thus as a corollary we obtain that the Jacobian $J_\phi(x)$ is

$$|J_\phi(x)| = (1 + \alpha)|x|^{\alpha n}.$$

Proof. The proof of Lemma 2.1 involves implementing the elementary proof of Lemma 2.23, in [Chanillo and Torchinsky 1986] in this special situation. Since $D\phi$ is a symmetric matrix, $D\phi$ is unitarily diagonalizable, that is one can write $A = QRQ^t$, where Q is a rotation matrix and R a diagonal matrix. It is enough to compute the eigenvalues for $D\phi$; the Jacobian formula follows by multiplying the eigenvalues. First note

$$D\phi(x) = A = \begin{pmatrix} |x|^\alpha + \alpha x_1^2 |x|^{\alpha-2} & \cdots & \alpha x_1 x_j |x|^{\alpha-2} & \cdots & \alpha x_1 x_n |x|^{\alpha-2} \\ \vdots & & \vdots & & \vdots \\ \alpha x_1 x_n |x|^{\alpha-2} & \cdots & \alpha x_n x_j |x|^{\alpha-2} & \cdots & |x|^\alpha + \alpha x_n^2 |x|^{\alpha-2} \end{pmatrix}.$$

Next

$$A - \lambda I = \begin{pmatrix} |x|^\alpha + \alpha x_1^2 |x|^{\alpha-2} - \lambda & \cdots & \alpha x_1 x_j |x|^{\alpha-2} & \cdots & \alpha x_1 x_n |x|^{\alpha-2} \\ \vdots & & \vdots & & \vdots \\ \alpha x_1 x_n |x|^{\alpha-2} & \cdots & \alpha x_n x_j |x|^{\alpha-2} & \cdots & |x|^\alpha + \alpha x_n^2 |x|^{\alpha-2} - \lambda \end{pmatrix}.$$

It follows that

$$\det(A - \lambda I) = |x|^{n(\alpha-2)} \det C,$$

where

$$C = \begin{pmatrix} |x|^2 + \alpha x_1^2 - \lambda |x|^{2-\alpha} & \cdots & \alpha x_1 x_j & \cdots & \alpha x_1 x_n \\ \vdots & & \vdots & & \vdots \\ \alpha x_1 x_n & \cdots & \alpha x_n x_j & \cdots & |x|^2 + \alpha x_n^2 - \lambda |x|^{2-\alpha} \end{pmatrix}.$$

To compute the characteristic polynomial of $D\phi$ we simply compute $\det C$. It is now that we use the trick in [Chanillo and Torchinsky 1986]. We simply add an extra row and column to C such that the new matrix now with $n+1$ rows and $n+1$ columns has the same determinant as C . Thus we form the matrix D given by

$$D = \begin{pmatrix} 1 & x_1 & \cdots & x_j & \cdots & x_n \\ 0 & |x|^2 + \alpha x_1^2 - \lambda |x|^{2-\alpha} & \cdots & \alpha x_1 x_j & \cdots & \alpha x_1 x_n \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & \alpha x_1 x_n & \cdots & \alpha x_n x_j & \cdots & |x|^2 + \alpha x_n^2 - \lambda |x|^{2-\alpha} \end{pmatrix}.$$

Note $\det C = \det D$. Now we perform elementary row operations in D that preserve the determinant. We replace row R_j , $j \geq 2$ by $R_j - \alpha x_{j-1} R_1$ where R_1 is row 1.

The new matrix we get is

$$E = \begin{pmatrix} 1 & x_1 & \cdots & x_j & \cdots & x_n \\ -\alpha x_1 & |x|^2 - \lambda|x|^{2-\alpha} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -\alpha x_j & 0 & \cdots & |x|^2 - \lambda|x|^{2-\alpha} & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -\alpha x_n & 0 & \cdots & 0 & \cdots & |x|^2 - \lambda|x|^{2-\alpha} \end{pmatrix}.$$

Note $\det D = \det E$ and the matrix we get if we remove the first row and first column of E is a diagonal matrix with $|x|^2 - \lambda|x|^{2-\alpha}$ on the diagonal. To compute $\det E$ we simply expand for the determinant using the first row of E and then expanding by the j -th row of the subsequent cofactor matrices we get for the entry $a_{1,j+1}$. We get

$$\det E = (|x|^2 - \lambda|x|^{2-\alpha})^n + \alpha(|x|^2 - \lambda|x|^{2-\alpha})^{n-1} \sum_{j=1}^n x_j^2,$$

which is

$$(|x|^2 - \lambda|x|^{2-\alpha})^n + \alpha(|x|^2 - \lambda|x|^{2-\alpha})^{n-1} |x|^2.$$

The expression above obviously factors as

$$(2-3) \quad (|x|^2 - \lambda|x|^{2-\alpha})^{n-1} ((1 + \alpha)|x|^2 - \lambda|x|^{2-\alpha}).$$

From (2-3) the conclusion of our Lemma follows because

$$\det(A - \lambda I) = 0 = (|x|^2 - \lambda|x|^{2-\alpha})^{n-1} ((1 + \alpha)|x|^2 - \lambda|x|^{2-\alpha}). \quad \square$$

First note that $A = D\phi$ is a symmetric matrix and thus there exist rotation matrices Q such that,

$$A^{-1} = (A^{-1})^* = Q\hat{D}Q^t,$$

where \hat{D} is diagonal. Using the eigenvalues of A computed from Lemma 2.1 above we may write

$$\hat{D} = |x|^{-\alpha} D,$$

where

$$D = \text{diag}((1 + \alpha)^{-1}, 1, 1, \dots, 1).$$

Thus, by Lemma 2.1,

$$\int_{\mathbb{R}^n} |(A^{-1})^* \nabla f|^p |J_\phi(x)| dx = (1 + \alpha) \int_{\mathbb{R}^n} |QDQ^t(\nabla f)|^p |x|^{\alpha(n-p)} dx.$$

We now apply Lemma 2.1 to (2-1) and we get with $M(s, r, p)$ the constant that occurs in (1-2),

$$(2-4) \quad \left(\int_{\mathbb{R}^n} |f(x)|^r |x|^{\alpha n} dx \right)^{1/r} \leq (1+\alpha)^{t/n} A_\alpha M(s, r, p) \left(\int_{\mathbb{R}^n} |f(x)|^s |x|^{\alpha n} dx \right)^{(1-t)/s} \\ \times \left(\int_{\mathbb{R}^n} |\nabla f(x)|^p |x|^{\alpha(n-p)} dx \right)^{t/p},$$

where we define

$$A_\alpha = \sup_f \left[\frac{\int_{\mathbb{R}^n} |QDQ^t(\nabla f)(x)|^p |x|^{\alpha(n-p)} dx}{\int_{\mathbb{R}^n} |\nabla f(x)|^p |x|^{\alpha(n-p)} dx} \right]^{t/p}.$$

The supremum is taken over those functions f where the denominator in the definition above is finite.

Lemma 2.2. *For $\alpha > -1$,*

$$B_\alpha \leq A_\alpha \leq C_\alpha = \begin{cases} 1, & \alpha \geq 0, \\ (1+\alpha)^{-t}, & -1 < \alpha < 0, \end{cases}$$

with

$$B_\alpha = \left[\frac{\int_{S^{n-1}} \left(\left[\frac{1}{(1+\alpha)^2} - 1 \right] \cos^2 \psi(\sigma) + 1 \right)^{p/2} d\sigma}{\int_{S^{n-1}} d\sigma} \right]^{t/p},$$

where

$$\cos \psi(\sigma) = \langle \sigma, v \rangle,$$

where v is a unit eigenvector for the eigenvalue $1/(1+\alpha)$ at $\sigma \in S^{n-1}$. Moreover, on radial functions f we may take $A_\alpha = B_\alpha$ for any $\alpha > -1$ and the ratio defining A_α is identically B_α for all radial functions without the supremum.

Proof. We now verify the assertions made about A_α . We note that pointwise

$$|QDQ^t(\nabla f)(x)| = |DQ^t \nabla f(x)| \leq C_\alpha^{1/t} |\nabla f(x)|.$$

This establishes $A_\alpha \leq C_\alpha$.

Next we establish the lower bound on A_α . Here we assume f is radial. Now note

$$(2-5) \quad |x|^{-\alpha} A = QD^{-1}Q^t.$$

The coefficients of A are homogeneous of degree α and thus the coefficients of the left side of (2-5) are homogeneous of degree 0. Since D is a constant matrix, it follows that the coefficients of Q , Q^t are functions of $\sigma \in S^{n-1}$. We now wish to consider for f radial the expression

$$\frac{\int_{\mathbb{R}^n} |QDQ^t(\nabla f)|^p |x|^{\alpha(n-p)} dx}{\int_{\mathbb{R}^n} |\nabla f|^p |x|^{\alpha(n-p)} dx}.$$

Using the fact that the coefficients of Q depend only on σ we see when f is radial, the expression above when converted to polar coordinates is identical to

$$(2-6) \quad \frac{\int_{S^{n-1}} |Q(\sigma) D Q^t(\sigma) \nabla r|^p d\sigma}{\int_{S^{n-1}} d\sigma}.$$

Now let $\{e_i(\sigma)\}_{i=1}^n$ be an orthonormal basis of eigenvectors for $Q D Q^t$. Then since $\nabla r = \sigma$, we see that

$$|Q D Q^t \nabla r|^2 = \frac{1}{(1+\alpha)^2} \langle e_1, \sigma \rangle^2 + \sum_{j=2}^n \langle \sigma, e_j \rangle^2.$$

The expression above can be rearranged as

$$\left(\frac{1}{(1+\alpha)^2} - 1 \right) \cos^2 \psi + 1.$$

Substituting this expression into (2-6) we readily establish that $B_\alpha \leq A_\alpha$. Since we have equality at every step in the computation above, we also obtain that $A_\alpha = B_\alpha$ when f is radial. \square

Lemma 2.3. *For $\alpha > -1$, we have,*

$$B_\alpha = (1+\alpha)^{-t}.$$

Proof. From the formula for $A = D\phi$, the eigenvalue equation for $A - (1+\alpha)I$ is

$$(2-7) \quad (\sigma_j^2 - 1)y_j + \sum_{k \neq j} \sigma_k \sigma_j y_k = 0, \quad j = 1, 2, \dots, n,$$

where the eigenvector is $v = y = (y_1, y_2, \dots, y_n)$. Now we set $y = \sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ and we get the left side of (2-7) is

$$\sigma_j(\sigma_j^2 - 1) + \sigma_j \sum_{k \neq j} \sigma_k^2 = \sigma_j(\sigma_j^2 - 1) + \sigma_j(1 - \sigma_j^2) = 0.$$

Thus σ is the unit eigenvector for the eigenvalue $(1+\alpha)$, which we already know has a 1-dimensional eigenspace. Thus,

$$\cos \psi = \langle \sigma, \sigma \rangle = 1,$$

and it follows from the expression for B_α in the statement of Lemma 2.2 that for any $\alpha > -1$,

$$B_\alpha = (1+\alpha)^{-t}. \quad \square$$

This lemma shows in particular that if we restrict (1-1) to radial functions, then the sharp constant is $(1+\alpha)^{t/n-t} M(s, r, p)$, as stated in Theorem 1.4.

Corollary 2.4. *When $-1 < \alpha < 0$, then*

$$B_\alpha = A_\alpha = C_\alpha = (1 + \alpha)^{-t}.$$

Proof. The proof follows in an obvious manner by combining the conclusions of Lemmas 2.2 and 2.3, which yields $B_\alpha = C_\alpha$ when $-1 < \alpha < 0$. \square

Thus the sharp constant in (2-4) is established when $-1 < \alpha < 0$.

3. The case $\alpha > 0$

For the case $\alpha > 0$, so far we have the upper and lower bounds for the sharp constant that is $(1 + \alpha)^{-t} \leq A_\alpha \leq 1$, in (2-4). Also if we restrict to radial functions then

$$A_\alpha(\text{radial}) = (1 + \alpha)^{-t}.$$

Now if we check closely the computations in Lemma 2.3, we obtain

$$|QDQ^t(\nabla f)(x)|^2 = \left[\left(\frac{1}{(1 + \alpha)^2} - 1 \right) \cos^2 \psi + 1 \right] |\nabla f(x)|^2,$$

where $\cos \psi = \langle v, w \rangle$, v is a unit vector in the direction of the eigenvector for $(1 + \alpha)$ and w the unit vector in the direction of ∇f . In particular v is radial at all points $x \in \mathbb{R}^n$. But now for $\alpha > 0$, $(1 + \alpha)^{-2} - 1 < 0$, and so it is advantageous to arrange $\cos \psi = 0$ as opposed to $\alpha < 0$ when $(1 + \alpha)^{-2} - 1 > 0$ and so there it is advantageous to have $\cos \psi = 1$ or functions to be radial. So the idea of proving Theorem 1.3 is to choose a function gradient having a big angular component that dominates the radial component. If one wants an optimizer for the case $\alpha > 0$, ∇f needs to be orthogonal to v , that is tangent to the sphere at all points. Notice the tangential directions to the sphere are eigenvectors to the eigenvalue 1 for $A = D\phi$. But in particular in 3D, $\nabla \times \nabla f = 0$; no such functions exist, or if f has some smoothness the vector field ∇f on S^2 will be smooth and tangential to S^2 which cannot happen by the hairy ball theorem. In fact, we show the following:

Lemma 3.1. *If $\alpha > 0$, then $A_\alpha = 1$.*

Proof. Indeed, based on the computations above, we have that the matrix $\tilde{A} = QDQ^t$ has two eigenvalues. The first one is $1/(1 + \alpha) < 1$, corresponding to the radial direction ∇r and the second eigenvalue is 1 with multiplicity $(n - 1)$ corresponding to the angular directions (tangential to S^{n-1}). We would like to estimate the quantity

$$F(f) = \frac{\int_{\mathbb{R}^n} |\tilde{A} \nabla f|^p |x|^{\alpha(n-p)} dx}{\int_{\mathbb{R}^n} |\nabla f|^p |x|^{\alpha(n-p)} dx}$$

for some choice of function f knowing that $A_\alpha = \sup_f F(f)$. We use spherical coordinates $(r, \phi_1, \dots, \phi_{n-1})$, and we form

$$f_k(r, \phi_1, \dots, \phi_{n-1}) = h(r) \sin(\phi_1) \cdots \sin(\phi_{n-2}) \cos(k\phi_{n-1}),$$

where $h : [0, \infty) \rightarrow \mathbb{R}$ is smooth and $h(t) = 0$ for $t < 1$ and $t > 4$ and $k \in \mathbb{N}$. For the sake of simplicity, we do the computation in $n = 3$; the higher dimensional case is similar.

So $f_k = h(r) \sin \phi \cos(k\theta)$, and thus

$$\nabla f_k = h'(r) \sin \phi \cos(k\theta) \mathbf{u}_r + \frac{h(r)}{r} \cos(\phi) \cos(k\theta) \mathbf{u}_\phi - \frac{h(r)}{r} k \sin(k\theta) \mathbf{u}_\theta,$$

where $(\mathbf{u}_r, \mathbf{u}_\phi, \mathbf{u}_\theta)$ is the standard orthonormal base defining the spherical coordinate system. Thus

$$\tilde{A} \nabla f_k = \frac{h'(r) \sin \phi \cos(k\theta)}{(1 + \alpha)} \mathbf{u}_r + \frac{h(r)}{r} \cos(\phi) \cos(k\theta) \mathbf{u}_\phi - \frac{h(r)}{r} k \sin(k\theta) \mathbf{u}_\theta.$$

We compute then

$$\begin{aligned} |\tilde{A} \nabla f_k|^p &= \left[\frac{h'(r)^2 \sin^2 \phi \cos^2(k\theta)}{(1 + \alpha)^2} \right. \\ &\quad \left. + \left(\frac{h(r)}{r} \right)^2 \cos^2 \phi \cos^2(k\theta) + \left(\frac{h(r)}{r} \right)^2 k^2 \sin^2(k\theta) \right]^{\frac{p}{2}} \\ &= \left[\cos^2(k\theta) \left(\frac{h'(r)^2 \sin^2 \phi}{(1 + \alpha)^2} + \left(\frac{h(r)}{r} \right)^2 \cos^2 \phi \right) + k^2 \left(\frac{h(r)}{r} \right)^2 \sin^2(k\theta) \right]^{\frac{p}{2}}. \end{aligned}$$

Hence,

$$\begin{aligned} \int_{\mathbb{R}^3} |\tilde{A} \nabla f_k|^p |x|^{\alpha(n-p)} dx &= \int_1^4 \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \left[\cos^2(k\theta) \left(\frac{h'(r)^2 \sin^2 \phi}{(1 + \alpha)^2} + \left(\frac{h(r)}{r} \right)^2 \cos^2 \phi \right) \right. \\ &\quad \left. + k^2 \left(\frac{h(r)}{r} \right)^2 \sin^2(k\theta) \right]^{\frac{p}{2}} r^{\alpha(n-p)+2} \sin \phi d\theta d\phi dr \\ &= k^p \int_1^4 \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \left[\frac{\cos^2(k\theta)}{k^2} \left(\frac{h'(r)^2 \sin^2 \phi}{(1 + \alpha)^2} + \left(\frac{h(r)}{r} \right)^2 \cos^2 \phi \right) \right. \\ &\quad \left. + \left(\frac{h(r)}{r} \right)^2 \sin^2(k\theta) \right]^{\frac{p}{2}} r^{\alpha(n-p)+2} \sin \phi d\theta d\phi dr \\ &= k^p \int_1^4 \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \left[\frac{\cos^2(u)}{k^2} \left(\frac{h'(r)^2 \sin^2 \phi}{(1 + \alpha)^2} + \left(\frac{h(r)}{r} \right)^2 \cos^2 \phi \right) \right. \\ &\quad \left. + \left(\frac{h(r)}{r} \right)^2 \sin^2(u) \right]^{\frac{p}{2}} r^{\alpha(n-p)+2} \sin \phi du d\theta dr. \end{aligned}$$

Therefore we have

$$\begin{aligned} & \int_{\mathbb{R}^3} |\tilde{A} \nabla f_k|^p |x|^{\alpha(n-p)} dx \\ &= k^p \left[\int_1^4 \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \left[\frac{h(r)}{r} \sin(u) \right]^p r^{\alpha(n-p)+2} \sin \phi \, du \, d\phi \, dr + o(1) \right]. \end{aligned}$$

A similar computation yields

$$\begin{aligned} & \int_{\mathbb{R}^3} |\nabla f_k|^p |x|^{\alpha(n-p)} dx \\ &= k^p \left[\int_1^4 \int_0^{\frac{\pi}{2}} \int_0^{2\pi} \left[\frac{h(r)}{r} \sin(u) \right]^p r^{\alpha(n-p)+2} \sin \phi \, du \, d\phi \, dr + o(1) \right]. \end{aligned}$$

Therefore

$$F(f_k) = 1 + o(1), \quad \text{as } k \rightarrow \infty.$$

Combining this last estimate with Lemma 2.2, we get the conclusion. \square

Notice that with this lemma, we have the proof of Theorem 1.3.

Remark. One can see that this sequence of functions f_k always satisfies

$$F(f_k) \rightarrow 1 \quad \text{as } k \rightarrow \infty$$

for all $\alpha > -1$, but if $\alpha < 0$, we have that $A_\alpha = (1 + \alpha)^{-t} > 1$, thus the sequence f_k in the case $\alpha < 0$ is not optimizing and as we saw earlier, the optimizer is radially symmetric. The sequence f_k gains importance in the case $\alpha > 0$ since $(\alpha + 1)^{-t} < 1$. Hence there is a symmetry breaking phenomenon and the radially symmetric functions cannot be optimizers anymore.

On the other hand, by the Riemann–Lebesgue lemma, $f_k \rightarrow 0$ as $k \rightarrow \infty$, hence we do not obtain an optimizer in this case.

Acknowledgement

S. Chanillo was supported in part by NSF grant DMS-1201474.

References

- [Caffarelli et al. 1984] L. Caffarelli, R. Kohn, and L. Nirenberg, “First order interpolation inequalities with weights”, *Compositio Math.* **53**:3 (1984), 259–275. MR Zbl
- [Chanillo and Torchinsky 1986] S. Chanillo and A. Torchinsky, “Sharp function and weighted L^p estimates for a class of pseudodifferential operators”, *Ark. Mat.* **24**:1 (1986), 1–25. MR Zbl
- [Chanillo and Wheeden 1992] S. Chanillo and R. L. Wheeden, “Poincaré inequalities for a class of non- A_p weights”, *Indiana Univ. Math. J.* **41**:3 (1992), 605–623. MR Zbl

- [Dolbeault and Esteban 2012] J. Dolbeault and M. J. Esteban, “Extremal functions for Caffarelli–Kohn–Nirenberg and logarithmic Hardy inequalities”, *Proc. Roy. Soc. Edinburgh Sect. A* **142**:4 (2012), 745–767. MR Zbl
- [Dolbeault et al. 2016] J. Dolbeault, M. J. Esteban, and M. Loss, “Rigidity versus symmetry breaking via nonlinear flows on cylinders and Euclidean spaces”, *Invent. Math.* **206**:2 (2016), 397–440. MR Zbl
- [Lam and Lu 2017] N. Lam and G. Lu, “Sharp constants and optimizers for a class of Caffarelli–Kohn–Nirenberg inequalities”, *Adv. Nonlinear Stud.* **17**:3 (2017), 457–480. MR Zbl
- [del Pino and Dolbeault 2013] M. del Pino and J. Dolbeault, “The Euclidean Onofri inequality in higher dimensions”, *Int. Math. Res. Not.* **2013**:15 (2013), 3600–3611. MR Zbl

Received January 8, 2017. Revised July 15, 2017.

AKSHAY L. CHANILLO
SKILLMAN, NJ
UNITED STATES
achanillo3@gmail.com

SAGUN CHANILLO
DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
PISCATAWAY, NJ
UNITED STATES
chanillo@math.rutgers.edu

ALI MAALAOUI
DEPARTMENT OF MATHEMATICS AND NATURAL SCIENCES
AMERICAN UNIVERSITY OF RAS AL KHAIMAH
RAS AL KHAIMAH
UNITED ARAB EMIRATES
ali.maalaoui@aurak.ac.ae

NONCOMMUTATIVE GEOMETRY OF HOMOGENIZED QUANTUM $\mathfrak{sl}(2, \mathbb{C})$

ALEX CHIRVASITU, S. PAUL SMITH AND LIANG ZE WONG

We examine the relationship between certain noncommutative analogues of projective 3-space, \mathbb{P}^3 , and the quantized enveloping algebras $U_q(\mathfrak{sl}_2)$. The relationship is mediated by certain noncommutative graded algebras S , one for each $q \in \mathbb{C}^\times$, having a degree-two central element c such that $S[c^{-1}]_0 \cong U_q(\mathfrak{sl}_2)$. The noncommutative analogues of \mathbb{P}^3 are the spaces $\text{Proj}_{\text{nc}}(S)$. We show how the points, fat points, lines, and quadrics, in $\text{Proj}_{\text{nc}}(S)$, and their incidence relations, correspond to finite-dimensional irreducible representations of $U_q(\mathfrak{sl}_2)$, Verma modules, annihilators of Verma modules, and homomorphisms between them.

1. Introduction	305
2. Preliminary notions	312
3. Point and line modules for D , a Zhang twist of S	318
4. Points, lines, and quadrics in $\text{Proj}_{\text{nc}}(S)$	327
5. Relation to $U_q(\mathfrak{sl}_2)$ -modules	338
6. Relation to the nondegenerate Sklyanin algebras	346
Acknowledgements	353
References	353

1. Introduction

This paper concerns the interplay between the geometry of some noncommutative analogues of \mathbb{P}^3 and the representation theory of the quantized enveloping algebras, $U_q(\mathfrak{sl}_2)$, of $\mathfrak{sl}(2, \mathbb{C})$. We always assume that q is not a root of unity.

1A. $\text{Proj}_{\text{nc}}(S)$ and $U_q(\mathfrak{sl}_2)$. In Section 2D, we define a family of noncommutative graded algebras $S = \mathbb{C}[E, F, K, K']$ depending on a parameter $q \in \mathbb{C} - \{0, \pm 1, \pm i\}$ that have the same Hilbert series and the “same” homological properties as the polynomial ring in four variables. For these reasons the noncommutative spaces $\text{Proj}_{\text{nc}}(S)$ have much in common with \mathbb{P}^3 . The element KK' belongs to the center of S and $S[(KK')^{-1}]_0 \cong U_q(\mathfrak{sl}_2)$. Thus, $U_q(\mathfrak{sl}_2)$ is the coordinate ring of the “open

MSC2010: 14A22, 16S38, 16W50, 17B37.

Keywords: noncommutative algebraic geometry, quantum groups, quantum \mathfrak{sl}_2 .

complement” to the union of the “hyperplanes” $\{K = 0\}$ and $\{K' = 0\}$ in $\text{Proj}_{\text{nc}}(S)$. This analogy can be formalized: there is an abelian category $\text{QCOH}(\cdot)$, defined below, that plays the role of the category of quasicoherent sheaves and an adjoint pair of functors

$$(1-1) \quad \text{QCOH}(\text{Proj}_{\text{nc}}(S)) \xrightleftharpoons[j_*]{j^*} \text{Mod}(U_q(\mathfrak{sl}_2))$$

that behave like the inverse and direct image functors for an open immersion $j : \mathbb{P}^3 - \{\text{two planes}\} \rightarrow \mathbb{P}^3$.

1A1. By definition, $\text{QCOH}(\text{Proj}_{\text{nc}}(S))$ is the quotient category

$$\text{QGr}(S) := \frac{\text{Gr}(S)}{\text{Fdim}(S)},$$

where $\text{Gr}(S)$ denotes the category of \mathbb{Z} -graded left S -modules and $\text{Fdim}(S)$ denotes the full subcategory of $\text{Gr}(S)$ consisting of those modules that are the sum of their finite-dimensional submodules. If S were the polynomial ring on four variables, then the category $\text{QGr}(S)$ would be equivalent to $\text{QCOH}(\mathbb{P}^3)$, the category of quasicoherent sheaves on \mathbb{P}^3 , and this equivalence would send a graded module M to the $\mathcal{O}_{\mathbb{P}^3}$ -module that Hartshorne denotes by \tilde{M} .

1A2. There is an exact functor $\pi^* : \text{Gr}(S) \rightarrow \text{QGr}(S)$ that sends a graded S -module M to M viewed as an object in $\text{QGr}(S)$. The composition

$$(1-2) \quad \text{Gr}(S) \xrightarrow{\pi^*} \text{QGr}(S) = \text{QCOH}(\text{Proj}_{\text{nc}}(S)) \xrightarrow{j^*} \text{Mod}(U_q(\mathfrak{sl}_2))$$

sends a graded S -module M to $M[(KK')^{-1}]_0$.

1A3. The main theme of this paper is the interaction between noncommutative geometry (where $\text{QGr}(S)$ belongs) and representation theory (where $\text{Mod}(U_q(\mathfrak{sl}_2))$ belongs). We show how the points, fat points, lines, and quadrics, in $\text{Proj}_{\text{nc}}(S)$, and their incidence relations, correspond to finite-dimensional irreducible representations of $U_q(\mathfrak{sl}_2)$, Verma modules, annihilators of Verma modules, and homomorphisms between them.

Just as passing from affine to projective geometry provides a more elegant picture that unifies seemingly different objects (affine vs. projective conic sections, for example), passing from the “affine” category $\text{Mod}(U_q(\mathfrak{sl}_2))$ to the “projective” category $\text{QCOH}(\text{Proj}_{\text{nc}}(S))$ results in a more complete picture of $\text{Mod}(U_q(\mathfrak{sl}_2))$.

1B. Lines and Verma modules, fat points and finite-dimensional irreducible representations. The most important $U_q(\mathfrak{sl}_2)$ -modules are its finite-dimensional irreducible representations and its Verma modules. In Section 5, we show that for each Verma module $V \in \text{Mod}(U_q(\mathfrak{sl}_2))$ there is a graded S -module M such that

- (1) $V \cong j^* \pi^* M$;
- (2) $M \cong S/S\ell^\perp$, where $\ell^\perp \subseteq S_1$ is a codimension-two subspace;
- (3) $\dim(M_i) = i + 1$ for all $i \geq 0$, i.e., M has the same Hilbert series as the polynomial ring on two variables;
- (4) M is a line module for S ;
- (5) M is a Cohen–Macaulay S -module.

In Section 5, we also show that for each finite-dimensional irreducible representation L of $U_q(\mathfrak{sl}_2)$ there is a graded S -module F such that

- (1) $L \cong j^* \pi^* F$;
- (2) $\dim(F_i) = \dim(L)$ for all $i \geq 0$ and $\dim(F_i) = 0$ for all $i < 0$;
- (3) every proper quotient of F is finite-dimensional, whence F is a simple object in $\text{QCOH}(\text{Proj}_{\text{nc}}(S))$;
- (4) F is a fat point module for S ;
- (5) F is a Cohen–Macaulay S -module.

Items (4) are, essentially, definitions; see Section 2B.

1B1. Point modules and line modules. Let R be the polynomial ring on four variables with its standard grading. The points in $\mathbb{P}^3 = \text{Proj}(R)$ are in bijection with the graded modules R/I such that $\dim(R_i/I_i) = 1$ for all $i \geq 0$. The lines in \mathbb{P}^3 are in bijection with the modules R/I such that $\dim(R_i/I_i) = i + 1$ for all $i \geq 0$.

If S is one of the algebras in Section 2D, a graded S -module M is called a *point module*, resp. a *line module*, if it is isomorphic to S/I for some left ideal I such that $\dim(S_i/I_i) = 1$, resp. $\dim(S_i/I_i) = i + 1$, for all $i \geq 0$.

There are fine moduli spaces that parametrize the point modules and line modules for S . These fine moduli spaces are called the *point scheme* and *line scheme* respectively. The point scheme for S is a closed subscheme of $\mathbb{P}^3 = \mathbb{P}(S_1^*)$ and the line scheme for S is a closed subscheme of the Grassmannian $\mathbb{G}(1, 3)$ consisting of the lines in \mathbb{P}^3 .

In Section 4, we determine the line modules and the point modules for S .

1B2. The point modules for S . The point scheme, \mathcal{P}_S , for S is $C \cup C' \cup L \cup \{p_1, p_2\} \subseteq \mathbb{P}(S_1^*) = \mathbb{P}^3$, the union of two plane conics, C and C' , meeting at two points, the line L through those two points, and two additional points (Theorem 4.2). If $M_p = S/Sp^\perp$ is the point module corresponding to $p \in \mathcal{P}_S$, then $(M_p)_{\geq 1}$ is a shifted point module; i.e., $(M_p)_{\geq 1}(1)$ is a point module and therefore isomorphic to $M_{p'}$ for some point $p' \in \mathcal{P}_S$. General results show there is an automorphism $\sigma : \mathcal{P}_S \rightarrow \mathcal{P}_S$ such that $p' = \sigma^{-1}p$. Thus, $(M_p)_{\geq 1} \cong M_{\sigma^{-1}p}(-1)$. We determine \mathcal{P}_S and σ in Section 4.

1B3. *The line modules for S .* Theorem 4.5 says that the lines $\ell \subseteq \mathbb{P}^3 = \mathbb{P}(S_1^*)$ for which $S/S\ell^\perp$ is a line module are precisely those lines that meet $C \cup C'$ with multiplicity two; i.e., the secant lines to $C \cup C'$. These are exactly the lines lying on a certain pencil of quadrics $Q(\lambda) \subseteq \mathbb{P}^3$, $\lambda \in \mathbb{P}^1$. This should remind the reader of the analogous result for the 4-dimensional Sklyanin algebras in which the lines in \mathbb{P}^3 that correspond to line modules are exactly the secant lines to the quartic elliptic curve E .

The labeling of the line modules is such that the Verma module $M(\lambda)$ is isomorphic to $j^*\pi^*(S/S\ell^\perp)$ for a unique line $\ell \subseteq Q(\lambda)$.

1B4. *Incidence relations.* If $(p) + (p')$ is a degree-two divisor on $C \cup C'$, we write $M_{p,p'}$ for the line module $S/S\ell^\perp$, where ℓ^\perp is the subspace of S_1 that vanishes on the line $\ell \subseteq \mathbb{P}^3 = \mathbb{P}(S_1^*)$ whose scheme-theoretic intersection with $C \cup C'$ is $(p) + (p')$. Proposition 4.8 shows there is an exact sequence

$$0 \rightarrow M_{\sigma p, \sigma^{-1} p'}(-1) \rightarrow M_{p,p'} \rightarrow M_p \rightarrow 0.$$

Proposition 4.9 shows that if the line ℓ just referred to meets the line $\{K = K' = 0\} \subseteq \mathcal{P}_S$ at a point p'' , there is an exact sequence

$$0 \rightarrow M_{\sigma^{-1} p, \sigma^{-1} p'}(-1) \rightarrow M_{p,p'} \rightarrow M_{p''} \rightarrow 0.$$

1B5. *Finite-dimensional simple modules.* Let $n \in \mathbb{N}$. If q is not a root of unity there are two simple $U_q(\mathfrak{sl}_2)$ -modules of dimension $n + 1$. We label them $L(n, \pm)$ in such a way that there are exact sequences

$$(1-3) \quad 0 \rightarrow M(\pm q^{-n-2}) \rightarrow M(\pm q) \rightarrow L(n, \pm) \rightarrow 0$$

in which $M(\lambda)$ denotes the Verma module of highest weight λ .

In Section 5 we show there are S -modules $V(n, \pm)$ that are also $S[(KK')^{-1}]$ -modules, and hence modules over $S[(KK')^{-1}]_0 \cong U_q(\mathfrak{sl}_2)$ and, as such, $V(n, \pm) \cong L(n, \pm)$. We define graded S -modules $F(n, \pm)$ such that $F(n, \pm)[(KK')^{-1}]_0 \cong L(n, \pm)$; i.e., if we view $F(n, \pm)$ as an object in $\text{QGr}(S)$, then

$$j^*F(n, \pm) \cong L(n, \pm).$$

Furthermore, we show there are exact sequences

$$(1-4) \quad 0 \rightarrow M_{\ell_\pm}(-n-1) \rightarrow M_{\ell_\pm} \rightarrow F(n, \pm) \rightarrow 0$$

in $\text{QCOH}(\text{Proj}_{\text{nc}}(S))$ and that (1-3) is obtained from (1-4) by applying the functor j^* , i.e., by restricting the exact sequence (1-4) in $\text{QCOH}(\text{Proj}_{\text{nc}}(S))$ to the “open affine subscheme” $\{KK' \neq 0\}$. Here M_{ℓ_\pm} denotes the line module $S/S\ell_\pm^\perp$ corresponding to a line $\ell_\pm \subseteq \mathbb{P}(S_1^*) = \mathbb{P}^3$.

1B6. Heretical Verma modules. The connections we establish between Verma modules and line modules highlights one way in which the q -deformation $U_q(\mathfrak{sl}_2)$ is “more rigid” or “less symmetric” than the enveloping algebra $U(\mathfrak{sl}_2)$: there is a \mathbb{P}^1 -family of Borel subalgebras of \mathfrak{sl}_2 , but there are only two reasonable candidates for the role of the quantized enveloping algebra of a “Borel subalgebra” of quantum \mathfrak{sl}_2 .

Fix one of the two “Borel subalgebras”, $U_q(\mathfrak{b}) \subseteq U_q(\mathfrak{sl}_2)$. It gives rise by induction to Verma modules $M_b(\lambda) = U_q(\mathfrak{sl}_2) \otimes_{U_q(\mathfrak{b})} \mathbb{C}_\lambda$, $\lambda \in \mathbb{C}^\times$. Thus, one obtains two 1-parameter families of Verma modules for $U_q(\mathfrak{sl}_2)$. In sharp contrast, by varying both the Borel subalgebra and the highest weight one obtains a 2-parameter family of Verma modules for $U(\mathfrak{sl}_2)$. Our perspective on $U_q(\mathfrak{sl}_2)$ as a noncommutative open subscheme of a noncommutative \mathbb{P}^3 allows us to fit the two 1-parameter families of Verma modules for $U_q(\mathfrak{sl}_2)$ into a single 2-parameter family of modules, thus undoing the rigidification phenomenon alluded to in the previous paragraph. It is these additional Verma-like modules that we call “heretical” in the title of this subsection.

For simplicity of discussion, fix a finite-dimensional simple module $L(n, +)$ and the corresponding fat point module $F(n, +)$ for which $j^*F(n, +) \cong L(n, +)$. The module $L(n, +)$ appears in exactly two sequences of the form (1-3), one for each “Borel subalgebra” of $U_q(\mathfrak{sl}_2)$; in contrast, $F(n, +)$ appears in a 1-parameter family of sequences of the form (1-4), one for each line in one of the rulings on the quadric $Q(q^n)$. Likewise, a fixed finite-dimensional simple $U(\mathfrak{sl}_2)$ -module fits into a 1-parameter family of sequences of the form (1-3). If we broadened the definition of a Verma module for $U_q(\mathfrak{sl}_2)$ so as to include j^*M_ℓ for all line modules M_ℓ one would then obtain a 1-parameter family of sequences of the form (1-3).

1B7. Annihilators of Verma modules and quadrics in $\text{Proj}_{\text{nc}}(S)$. When q is not a root of unity, the center of $U_q(\mathfrak{sl}_2)$ is generated by a single central element C called the Casimir element. A Verma module is annihilated by $C - \nu$ for a unique $\nu \in \mathbb{C}$ and given ν there are, usually, four Verma modules annihilated by $C - \nu$.

There is a nonzero central element $\Omega \in S_2$ such that $C = \Omega(KK')^{-1}$ under the isomorphism $U_q(\mathfrak{sl}_2) \cong S[(KK')^{-1}]_0$. A line module for S is annihilated by $\Omega - \nu KK'$ for a unique $\nu \in \mathbb{C} \cup \{\infty\}$ and given ν there are, usually, two 1-parameter families of line modules annihilated by $\Omega - \nu KK'$. There is an isomorphism

$$\frac{S}{(\Omega - \nu KK')}[(KK')^{-1}]_0 \cong \frac{U_q(\mathfrak{sl}_2)}{(C - \nu)}$$

and an adjoint pair of functors

$$(1-5) \quad \text{QCOH}\left(\text{Proj}_{\text{nc}}\left(\frac{S}{(\Omega - \nu KK')}\right)\right) \xrightleftharpoons[j_*]{j^*} \text{Mod}\left(\frac{U_q(\mathfrak{sl}_2)}{(C - \nu)}\right).$$

We think of $\text{Proj}_{\text{nc}}(S/(\Omega - \nu K K'))$ as a noncommutative quadric hypersurface in $\text{Proj}_{\text{nc}}(S)$ and think of $U_q(\mathfrak{sl}_2)/(C - \nu)$ as the coordinate ring of a noncommutative affine quadric. Noncommutative quadrics in noncommutative analogues of \mathbb{P}^3 were examined in [Smith and Van den Bergh 2013]. The results there apply to the present situation. The line modules for S that are annihilated by $\Omega - \nu K K'$ provide rulings on the noncommutative quadric and the noncommutative quadric is smooth if and only if it has two rulings. We note that $\text{Proj}_{\text{nc}}(S/(\Omega - \nu K K'))$ is smooth if and only if $U_q(\mathfrak{sl}_2)/(C - \nu)$ has finite global dimension.

In Section 1B2, we mentioned the pencil of quadrics $Q(\lambda) \subseteq \mathbb{P}^3$, $\lambda \in \mathbb{P}^1$, that contain $C \cup C'$. The $Q(\lambda)$'s are commutative quadrics and should not be confused with the noncommutative ones in the previous paragraph. If ℓ is a line on $Q(\lambda)$, then $M_\ell = S/S\ell^\perp$ is a line module so is annihilated by $\Omega - \nu K K'$ for some $\nu \in \mathbb{C} \cup \infty$.

1B8. *What happens for $U(\mathfrak{sl}_2)$?* Le Bruyn and Smith [1993] considered a graded algebra $H(\mathfrak{sl}_2)$ that has a central element t in H_1 such that $H[t^{-1}]_0$ is isomorphic to the enveloping algebra $U(\mathfrak{sl}_2)$. They call $H(\mathfrak{sl}_2)$ a *homogenization* of $U(\mathfrak{sl}_2)$,

Since the Hilbert series of H equals that of the polynomial ring in four variables with its standard grading, and since H has “all” the good homological properties the polynomial ring does, they view H as a homogeneous coordinate ring of a noncommutative analogue of \mathbb{P}^3 , denoted $\text{Proj}_{\text{nc}}(H)$. Because $H[t^{-1}]_0 \cong U(\mathfrak{sl}_2)$, there is an adjoint pair of functors j^* and j_* fitting into diagrams like those in (1-1) and (1-2). Because t has degree one, j^* and j_* behave like the inverse and direct image functors associated to the open complement to the hyperplane at infinity in \mathbb{P}^3 . Le Bruyn and Smith examined the point and line modules for H and showed that these modules are related to the finite-dimensional irreducible representations and Verma modules for \mathfrak{sl}_2 . The situation for $U(\mathfrak{sl}_2)$ is simpler than that for $U_q(\mathfrak{sl}_2)$.

1B9. *Richard Chandler's results.* We are not the first to compute the point modules and line modules for S . Richard Chandler did this in his Ph.D. thesis [Chandler 2016]. His approach differs from ours. Following a method introduced by Shelton and Vancliff [2002b], he used *Mathematica* to compute a system of 45 quadratic polynomials in the Plücker coordinates on the Grassmannian $\mathbb{G}(1, 3)$, the common zero locus of which is the line scheme for S . In contrast, we use the results on central extensions in [Le Bruyn et al. 1996] to determine which lines in \mathbb{P}^3 correspond to line modules. The two approaches are complementary.

1C. *The structure of the paper.* In Section 2, we define the algebra S , the central focus of our paper, and discuss its position as a degenerate version of the 4-dimensional Sklyanin algebra and a homogenization of $U_q(\mathfrak{sl}_2)$. We introduce the category $\text{QGr}(S)$ and its noncommutative geometry. We focus on point, line, and fat point modules.

$\text{Gr}(S)$	$\text{Proj}_{\text{nc}}(S)$	$\text{Mod}(U_q(\mathfrak{sl}_2))$
Point modules	Points	Finite-dimensional irreducible modules
Line modules	Lines	Verma modules

Table 1. Relation to $U_q(\mathfrak{sl}_2)$ -modules.

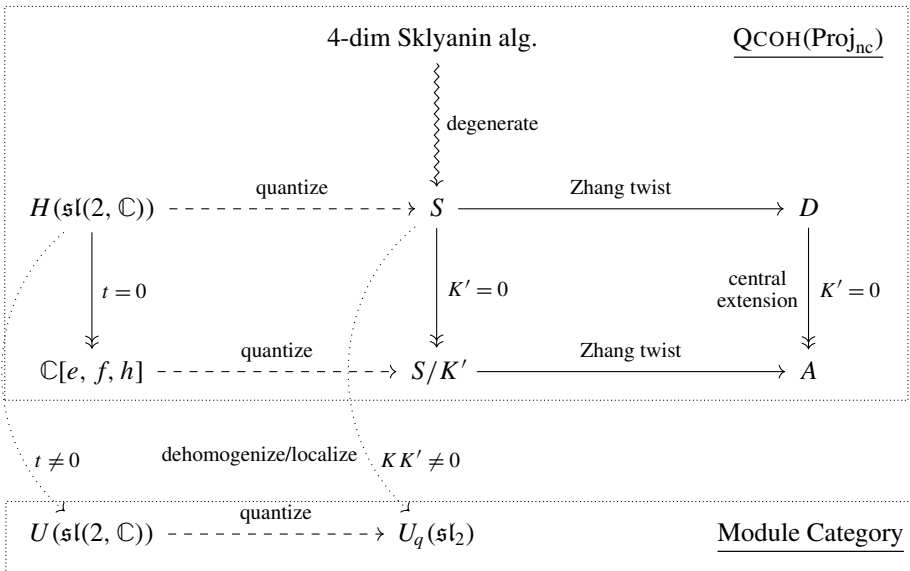
In Section 3, we examine a Zhang twist D of S . It has the property that $\text{Gr}(D) \equiv \text{Gr}(S)$. Moreover, D has a central element $z \in D_1$ such that $A = D/(z)$ is a 3-dimensional Artin–Schelter regular algebra, thereby making D a central extension of A . This allows us to use the results in [Le Bruyn et al. 1996] to determine the point and line modules of D in terms of those for A .

In Section 4, we use the equivalence $\text{Gr}(D) \equiv \text{Gr}(S)$ to transfer the results about D back to S .

In Section 5, we relate our results about point and line modules for S to results about the finite-dimensional irreducible representations and Verma modules of $U_q(\mathfrak{sl}_2)$. Table 1 summarizes some of these relations.

In Section 6, we show that some of our results can be obtained as “degenerations” of results in [Smith and Stafford 1992; Chirvasitu and Smith 2017; Smith and Staniszkis 1993] about the 4-dimensional Sklyanin algebra.

Figure 1 summarizes the algebras in this paper and their relationships to S .

**Figure 1.** Algebras in this paper, their relationship to S , and their associated categories.

2. Preliminary notions

2A. The category \mathbf{QGr} . Let \mathbb{k} be a field and R a \mathbb{Z} -graded \mathbb{k} -algebra. The category $\mathbf{QGr}(R)$ is defined to be the quotient category

$$\mathbf{QGr}(R) := \frac{\mathbf{Gr}(R)}{\mathbf{Fdim}(R)},$$

where $\mathbf{Gr}(R)$ denotes the category of \mathbb{Z} -graded left R -modules with degree-preserving homomorphisms and $\mathbf{Fdim}(R)$ denotes the full subcategory of $\mathbf{Gr}(R)$ consisting of those modules that are the sum of their finite-dimensional submodules.

The categories $\mathbf{QGr}(R)$ and $\mathbf{Gr}(R)$ have the same objects but different morphisms. There is an exact functor $\pi^* : \mathbf{Gr}(R) \rightarrow \mathbf{QGr}(R)$ that is the identity on objects. In the situations considered in this paper π^* has a right adjoint π_* . A morphism $f : M \rightarrow M'$ becomes an isomorphism in $\mathbf{QGr}(R)$, i.e., $\pi^* f$ is an isomorphism, if and only if $\ker(f)$ and $\operatorname{coker}(f)$ are in $\mathbf{Fdim}(R)$. In particular, a graded R -module is isomorphic to 0 in $\mathbf{QGr}(R)$ if and only if it is the sum of its finite-dimensional modules. Two modules in $\mathbf{Gr}(R)$ are *equivalent* if they are isomorphic in $\mathbf{QGr}(R)$.

If $M \in \mathbf{Gr}(R)$ and $n \in \mathbb{Z}$ we write $M(n)$ for the graded R -module that is M as a left R -module but with new homogeneous components, $M(n)_i = M_{n+i}$. The rule $M \rightsquigarrow M(n)$ extends to an autoequivalence of $\mathbf{Gr}(R)$. Because it sends finite-dimensional modules to finite-dimensional modules, it induces an autoequivalence of $\mathbf{QGr}(R)$ that we denote by $\mathcal{M} \rightsquigarrow \mathcal{M}(n)$.

If $M \in \mathbf{Gr}(R)$ we define $M_{\geq n} := M_n + M_{n+1} + \cdots$. If $R = R_{\geq 0}$, then $M_{\geq n}$ is a submodule of M .

2B. Linear modules. The importance of linear modules for noncommutative analogues of \mathbb{P}^n was first recognized by Artin, Tate, and Van den Bergh. We recall a few notions from their papers [Artin et al. 1990; 1991]. Let $M \in \mathbf{Gr}(R)$. If $M_n = 0$ for $n \ll 0$ and $\dim M_n < \infty$ for all n , the *Hilbert series* of M is the formal Laurent series

$$H_M(t) = \sum_n (\dim M_n) t^n.$$

We are particularly interested in cyclic modules M with Hilbert series having the form

$$H_M(t) = \frac{n}{(1-t)^d}$$

for some $n, d \in \mathbb{N}$. The *Gelfand–Kirillov (GK) dimension* of such a module is $d(M) = d$ and its *multiplicity* is n . If $d(M) = d$ and $d(M/N) < d$ for all nonzero submodules N , then M is called *d-critical*. Equivalent modules (in the sense of Section 2A) have the same GK dimension, and also have the same multiplicity if

they are not equivalent to 0, so the notions of GK dimension and multiplicity carry over to $\text{QGr}(R)$ as well.

We call M a *linear module* if it is cyclic and its Hilbert series is $(1-t)^{-d}$. The cases $d = 1$ and $d = 2$ play a key role: we call a linear module M a

- *point module* if it is cyclic, 1-critical, and $H_M(t) = (1-t)^{-1}$;
- *line module* if it is cyclic, 2-critical, and $H_M(t) = (1-t)^{-2}$.

We are also interested in modules of higher multiplicity: we call M a

- *fat point module* if it is 1-critical, generated by M_0 , and $H_M(t) = n(1-t)^{-1}$ for some $n > 1$.

Point modules and fat point modules are important because, as objects in $\text{QGr}(R)$, they are simple (or irreducible): all proper quotient modules of a 1-critical module are finite-dimensional and therefore zero in $\text{QGr}(R)$. The following result illustrates the relationship between finite-dimensional simple modules and fat point modules.

Lemma 2.1. *Let V be a simple left R -module of dimension $n < \infty$. Let $\mathbb{C}[z]$ be the polynomial ring generated by a degree-one indeterminate, z . Let $V \otimes \mathbb{C}[z]$ be the graded left R -module whose degree- j component is $V \otimes z^j$ with $a \in R_i$ acting as $a(v \otimes z^j) := (av) \otimes z^{i+j}$. Let $\pi : V \otimes \mathbb{C}[z] \rightarrow V$ be the R -module homomorphism $v \otimes z^j \mapsto v$.*

- (1) $V \otimes \mathbb{C}[z]$ is a fat point module of multiplicity n .
- (2) If M is a graded left R -module such that $M = M_{\geq 0}$ and $\psi : M \rightarrow V$ is a homomorphism in $\text{Mod}(R)$, then there is a unique homomorphism $\tilde{\psi} : M \rightarrow V \otimes \mathbb{C}[z]$ in $\text{Gr}(R)$ such that $\psi = \pi \tilde{\psi}$, namely $\tilde{\psi}(m) = \psi(m) \otimes z^n$ for $m \in M_n$.

2C. Geometry in $\text{Proj}_{\text{nc}}(R)$. The “noncommutative scheme” $\text{Proj}_{\text{nc}}(R)$ is defined implicitly by declaring that the category of “quasicoherent sheaves” on it is $\text{QGr}(R)$,

$$\text{QCOH}(\text{Proj}_{\text{nc}}(R)) := \text{QGr}(R).$$

The isomorphism class of a (fat) point module in $\text{QGr}(R)$ is called a (fat) point of $\text{Proj}_{\text{nc}}(R)$. Likewise, the isomorphism class of a line module in $\text{QGr}(R)$ is called a line in $\text{Proj}_{\text{nc}}(R)$.

2C1. Origin of the terminology. Let \mathbb{k} be an algebraically closed field, and let $R = \mathbb{k}[x_0, \dots, x_n]$ be the commutative polynomial ring with its standard grading, $\deg(x_j) = 1$ for all j . Then $\text{Proj}(R)$ is \mathbb{P}^n , projective n -space, and there is a bijection between closed points in \mathbb{P}^n and isomorphism classes of point modules for R : a point module is isomorphic to R/I for a unique ideal I , and I is generated by a codimension-1 subspace of $\mathbb{C}x_0 + \dots + \mathbb{C}x_n$; conversely, if I is such an ideal, then R/I is a point module. Under the equivalence $\text{QGr}(R) \xrightarrow{\sim} \text{QCOH}(\mathbb{P}^n)$, $M \mapsto \tilde{M}$,

the point module R/I corresponds to the skyscraper sheaf \mathcal{O}_p at the point $p \in \mathbb{P}^n$ where I vanishes. Similarly, if M is a line module for R , then $M \cong R/I$ for an ideal I that is generated by a codimension-2 subspace of $\mathbb{C}x_0 + \cdots + \mathbb{C}x_n$ and the zero locus of I is a line in \mathbb{P}^n , and this sets up a bijection between the lines in \mathbb{P}^n and the isomorphism classes of line modules. Indeed, there is a bijection between linear subspaces of \mathbb{P}^n and isomorphism classes of linear modules over the polynomial ring R .

Theorem 2.2 [Levasseur and Smith 1993, Theorem 1.13]. *Let $R = \mathbb{k}[x_0, \dots, x_n]$ be a polynomial ring in $n+1$ variables, graded by setting $\deg(x_j) = 1$ for all j . Let M be a finitely generated graded R -module. The following conditions on a graded R -module M are equivalent:*

- (1) M is cyclic with Hilbert series $(1-t)^{-d}$;
- (2) $M \cong R/R\ell^\perp$ for some codimension- d subspace $\ell \subseteq R_1$ or, equivalently, for some $(d-1)$ -dimensional linear subspace $\ell \subseteq \mathbb{P}(R_1^*)$;
- (3) M is a Cohen–Macaulay R -module having GK dimension d and multiplicity 1.

Thus, linear modules of GK dimension d correspond to linear subspaces of \mathbb{P}^n having dimension $d-1$.

2C2. Points, fat points, and lines in $\text{Proj}_{\text{nc}}(R)$. For the noncommutative graded algebras R in this paper, the points and lines in $\text{Proj}_{\text{nc}}(R)$ are parametrized by genuine (commutative) varieties [Artin et al. 1990; 1991].

Let R be any \mathbb{N} -graded \mathbb{k} -algebra such that $R_0 = \mathbb{k}$ and R is generated by R_1 as a \mathbb{k} -algebra. Let $\mathbb{P}(R_1^*)$ denote the projective space whose points are the 1-dimensional subspaces of $R_1^* = \text{Hom}_{\mathbb{k}}(R_1, \mathbb{k})$.

For V a linear subspace of R_1^* , define $V^\perp := \{x \in R_1 \mid \xi(x) = 0 \text{ for all } \xi \in V\}$. Let

$$\mathcal{P}_R := \{p \in \mathbb{P}(R_1^*) \mid R/Rp^\perp \text{ is a point module}\},$$

$$\mathcal{L}_R := \{\text{lines } \ell \text{ in } \mathbb{P}(R_1^*) \mid R/R\ell^\perp \text{ is a line module}\}.$$

For the algebras in this paper there are moduli problems for which \mathcal{P}_R and \mathcal{L}_R are fine moduli spaces; see [Artin et al. 1990, Corollary 3.13] and [Shelton and Vancliff 2002a, Corollary 1.5]. We call \mathcal{P}_R and \mathcal{L}_R the *point scheme* and *line scheme* for R .

Clearly, a line module $R/R\ell^\perp$ surjects onto a point module R/Rp^\perp if and only if p lies on the line ℓ . Thus, the incidence relations between points and lines in $\text{Proj}_{\text{nc}}(R)$ coincides with the incidence relations between *certain* points and lines in $\mathbb{P}(R_1^*)$. In such a situation the phrase “ p lies on ℓ ” is a statement about points and lines in $\mathbb{P}(R_1^*)$ and *also* a statement about points and lines in $\text{Proj}_{\text{nc}}(R)$. If a line module $R/R\ell^\perp$ surjects onto a fat point module F in $\text{QGr}(R)$ we say that the corresponding fat point lies on the line ℓ and understand this as a statement about the geometry of $\text{Proj}_{\text{nc}}(R)$.

Proposition 2.3 [Levasseur and Smith 1993]. *The kernel of a surjective homomorphism $\psi : M_\ell \rightarrow M_p$ in $\text{Gr}(S)$ from a line module to a point module is isomorphic to a shifted line module $M_{\ell'}(-1)$.*

Proof. There are elements $u, v, w \in S_1$ for which there is a commutative diagram

$$\begin{array}{ccc} M_\ell & \longrightarrow & S/Su + Sv \\ \psi \downarrow & & \downarrow \psi' \\ M_p & \longrightarrow & S/Su + Sv + Sw \end{array}$$

in which the horizontal arrows are isomorphisms and ψ' is the obvious map. The kernel of ψ' is isomorphic to the submodule $S\bar{w} = Su + Sv + Sw/Su + Sv$. Because M_ℓ is a critical Cohen–Macaulay module of GK dimension 2 and multiplicity 1, and M_p has GK dimension 1, the kernel is a Cohen–Macaulay module of GK dimension 2 and multiplicity 1. By [Levasseur and Smith 1993, Proposition 2.12], the kernel of ψ' is isomorphic to a shifted line module. \square

The associated exact sequence $0 \rightarrow M_{\ell'}(-1) \rightarrow M_\ell \rightarrow M_p \rightarrow 0$ is the analogue of an exact sequence $0 \rightarrow M(\lambda') \rightarrow M(\lambda) \rightarrow L \rightarrow 0$ in which $M(\lambda')$ and $M(\lambda)$ are Verma modules.

2C3. Noncommutative analogues of quadrics and \mathbb{P}^3 . Let S be one of the algebras in Section 2D. The Hilbert series of S is $(1-t)^{-4}$, the same as that of the polynomial ring on four variables. Furthermore, S has the “same” homological properties as that polynomial ring and, as a consequence, it is a domain [Artin et al. 1991, Theorem 3.9]. For these reasons we think of $\text{Proj}_{\text{nc}}(S)$ as a noncommutative analogue of \mathbb{P}^3 .

If Ω is a homogeneous, degree-two, central element in S we call $\text{Proj}_{\text{nc}}(S/(\Omega))$ a *quadric hypersurface* in $\text{Proj}_{\text{nc}}(S)$ and sometimes denote it by the symbols $\{\Omega = 0\}$. A line module $S/S\ell^\perp$ is annihilated by Ω if and only if there is a surjective map $S/(\Omega) \rightarrow S/S\ell^\perp$. If so we say that “the line ℓ lies on the quadric $\{\Omega = 0\}$ ” and interpret this as a statement about the geometry of $\text{Proj}_{\text{nc}}(S)$.

2D. The algebras S . The algebras of interest to us are the noncommutative \mathbb{C} -algebras S with generators x_0, x_1, x_2, x_3 subject to the relations

$$(2-1) \quad \begin{aligned} [x_0, x_1] &= 0, & \{x_0, x_1\} &= 2x_0x_1 = [x_2, x_3], \\ [x_0, x_2] &= b^2\{x_1, x_3\}, & \{x_0, x_2\} &= [x_3, x_1], \\ [x_0, x_3] &= -b^2\{x_1, x_2\}, & \{x_0, x_3\} &= [x_1, x_2], \end{aligned}$$

where $\{x, x'\} = xx' + x'x$, $[x, x'] = xx' - x'x$, and $b \in \mathbb{C} - \{0, \pm i\}$.

The algebras S occupy an interesting position between the nondegenerate 4-dimensional Sklyanin algebras and the quantized enveloping algebras $U_q(\mathfrak{sl}_2)$. We

now introduce these algebras and, in Proposition 2.4 below, describe their relation to S .

2D1. S is a degenerate Sklyanin algebra. A nondegenerate Sklyanin algebra is a \mathbb{C} -algebra $S(\alpha, \beta, \gamma)$ with generators x_0, x_1, x_2, x_3 subject to the relations

$$(2-2) \quad \begin{aligned} [x_0, x_1] &= \alpha\{x_2, x_3\}, & \{x_0, x_1\} &= [x_2, x_3], \\ [x_0, x_2] &= \beta\{x_1, x_3\}, & \{x_0, x_2\} &= [x_3, x_1], \\ [x_0, x_3] &= \gamma\{x_1, x_2\}, & \{x_0, x_3\} &= [x_1, x_2], \end{aligned}$$

where $\alpha, \beta, \gamma \in \mathbb{C}$ are such that $\alpha + \beta + \gamma + \alpha\beta\gamma = 0$, and further satisfy the nondegeneracy condition

$$(2-3) \quad \{\alpha, \beta, \gamma\} \cap \{0, 1, -1\} = \emptyset.$$

With this notation, $S = S(0, b^2, -b^2)$, and is degenerate.

For the rest of the paper, S will denote $S(0, b^2, -b^2)$ and $S(\alpha, \beta, \gamma)$ will denote a nondegenerate Sklyanin algebra.

The noncommutative space $\text{Proj}_{\text{nc}}(S(\alpha, \beta, \gamma))$ is well understood. Its point scheme was computed in [Smith and Stafford 1992], its lines and the incidence relations between its points and lines were determined in [Levasseur and Smith 1993], and its fat points and the incidence relations between fat points and lines were determined in [Smith and Staniszkis 1993]. A short account of these and related results can be found in the survey article [Smith 1994]. In this paper we carry out the same computations for S and compare them to what has been obtained for nondegenerate $S(\alpha, \beta, \gamma)$. This is the subject of Section 6.

2D2. S as a homogenization of $U_q(\mathfrak{sl}(2, \mathbb{C}))$. The quantized enveloping algebra $U_q(\mathfrak{sl}_2)$ is the \mathbb{C} -algebra with generators e, f, k^{\pm} subject to the relations

$$(2-4) \quad ke = q^2 ek, \quad kf = q^{-2} fk, \quad kk^{-1} = k^{-1}k = 1, \quad \text{and} \quad [e, f] = \frac{k - k^{-1}}{q - q^{-1}},$$

where $q \neq 0, \pm 1, \pm i$.

The representation theory of $U_q(\mathfrak{sl}_2)$ is the subject of books by Brown and Goodearl [2002], Jantzen [1996], Kassel [1995], Klimyk and Schmüdgen [1997], and others.¹

¹A slightly different algebra was studied by Jimbo [1985] and by Lusztig [1988]: they replace the last of the above relations by $[e, f] = (k - k^{-1})/(q^2 - q^{-2})$. Lusztig [1990] replaced that relation by the one in (2-4) and that seems to have become the “official” quantized enveloping algebra of $\mathfrak{sl}(2, \mathbb{C})$ used by subsequent authors. We call the algebra studied in [Jimbo 1985] and [Lusztig 1988] the “unofficial” quantized enveloping algebra of $\mathfrak{sl}(2, \mathbb{C})$. That unofficial version is a quotient of the algebra S in Proposition 2.4.

Before showing that S is a homogenization of $U_q(\mathfrak{sl}_2)$, we introduce notation that will be used throughout the paper:

$$(2-5) \quad \begin{aligned} q &= \frac{1-ib}{1+ib}, & E &= \frac{i}{2}(1-ib)(x_2 + ix_3), & K &= x_0 + bx_1, \\ \kappa &= \frac{1}{q^{-1}-q}, & F &= \frac{i}{2}(1+ib)(x_2 - ix_3), & K' &= x_0 - bx_1. \end{aligned}$$

Proposition 2.4. *The algebra S is the \mathbb{C} -algebra generated by E, F, K, K' modulo the relations*

$$(2-6) \quad \begin{aligned} KE &= qEK, & KF &= q^{-1}FK, & KK' &= K'K, \\ K'E &= q^{-1}EK', & K'F &= qFK', & [E, F] &= \frac{K^2 - K'^2}{q - q^{-1}}. \end{aligned}$$

Further, $S[(KK')^{-1}]_0$ is isomorphic to $U_q(\mathfrak{sl}_2)$ via

$$(2-7) \quad EK^{-1} \mapsto \sqrt{q}e, \quad F(K')^{-1} \mapsto \sqrt{q}f, \quad K(K')^{-1} \mapsto k,$$

where \sqrt{q} is a fixed square root of q .

Proof. A few tedious but straightforward calculations show that E, F, K, K' satisfy the relations in (2-6). For example, $KE = qEK$ because

$$\begin{aligned} (1+ib)KE - (1-ib)EK &= [K, E] + ib\{K, E\} \\ &= \frac{i}{2}(1-ib)([x_0, x_2] + i[x_0, x_3] + b[x_1, x_2] + ib[x_1, x_3] \\ &\quad + ib\{x_0, x_2\} - b\{x_0, x_3\} + ib^2\{x_1, x_2\} - b^2\{x_1, x_3\}) \\ &= \frac{i}{2}(1-ib)([x_0, x_2] - b^2\{x_3, x_1\} + i[x_0, x_3] + ib^2\{x_1, x_2\} \\ &\quad + ib\{x_0, x_2\} - ib[x_3, x_1] - b\{x_0, x_3\} + b[x_1, x_2]) \\ &= 0. \end{aligned}$$

Similar calculations show $KF = q^{-1}FK$, $K'E = q^{-1}EK'$ and $K'F = qFK'$.

Since $K^2 - K'^2 = 4bx_0x_1 = 2b\{x_0, x_1\}$ and $\frac{i}{2}(1-ib) \cdot \frac{i}{2}(1+ib) = -\frac{1}{4}(1+b^2)$, we have

$$-\frac{4}{1+b^2}[E, F] + ib^{-1}(K^2 - K'^2) = 2i[x_3, x_2] + 2i\{x_0, x_1\} = 0.$$

However,

$$q - q^{-1} = \frac{1-ib}{1+ib} - \frac{1+ib}{1-ib} = -\frac{4ib}{1+b^2},$$

so

$$[E, F] = -\frac{1+b^2}{4ib}(K^2 - K'^2) = \frac{K^2 - K'^2}{q - q^{-1}}.$$

For the second part of the proposition, it is clear that $S[(KK')^{-1}]_0$ is generated by

$$e := \frac{1}{\sqrt{q}}EK^{-1}, \quad f := \frac{1}{\sqrt{q}}F(K')^{-1}, \quad k := K(K')^{-1}, \quad \text{and} \quad k^{-1}.$$

Similar straightforward calculations then show that these elements satisfy the relations in (2-4). Hence $S[(KK')^{-1}]_0 \cong U_q(\mathfrak{sl}_2)$. \square

Since S is an Artin–Schelter regular algebra of global dimension 4 and has Hilbert series $(1-t)^{-4}$ we think of it as a homogeneous coordinate ring of a noncommutative analogue of \mathbb{P}^3 . Since $SK = KS$ and $SK' = K'S$ we think of $S/(K)$ and $S/(K')$ as homogeneous coordinate rings of noncommutative analogues of \mathbb{P}^2 .

Further, we think of $S[(KK')^{-1}]_0$, i.e., $U_q(\mathfrak{sl}_2)$, as the coordinate ring of the noncommutative affine scheme that is the “open complement” of the “union” of the “hyperplanes” $\{K = 0\}$ and $\{K' = 0\}$. These “hyperplanes” are effective divisors in the sense of Van den Bergh [2001, §3.6]. From this perspective, $U_q(\mathfrak{sl}_2)$ can be considered an “affine piece” of S . As explained in Section 1A2, this point of view can be formalized in terms of an adjoint pair of functors $j^* \dashv j_*$.

The left adjoint $j^* : \text{QGr}(S) \rightarrow \text{Mod}(S[(KK')^{-1}]_0)$ sends a graded S -module, X , viewed as an object in $\text{QGr}(S)$ to $X[(KK')^{-1}]_0 \in \text{Mod}(S[(KK')^{-1}]_0)$. The action of j^* on a morphism $f : M \rightarrow N$ in $\text{QGr}(S)$ is defined by choosing a lift of f to a morphism ϕ in $\text{Gr}(S)$, then applying the localization functor $X \mapsto X[(KK')^{-1}]_0$ to ϕ .

3. Point and line modules for D , a Zhang twist of S

In this section, we replace S by a Zhang twist of itself [Zhang 1996]. The appropriate Zhang twist is an algebra D that has a central element $z \in D_1$ such that $D/(z)$ is a 3-dimensional Artin–Schelter regular algebra. In the terminology of [Le Bruyn et al. 1996], this makes D a *central extension* of $D/(z)$. We use the results in that paper to determine the point and line modules for D . The point and line modules for $D/(z)$ are already understood due to [Artin et al. 1990; 1991].

In Section 4 we use Zhang’s fundamental equivalence $\text{Gr}(D) \cong \text{Gr}(S)$ [Zhang 1996] to transfer the results about the point and line modules for D to S .

3A. The Zhang twist. Let S be a graded \mathbb{k} -algebra and ϕ a degree-preserving \mathbb{k} -algebra automorphism of S . Define D to be the \mathbb{k} -algebra that is equal to S as a graded \mathbb{k} -vector space, but endowed with the new multiplication

$$c * d := \phi^n(c)d$$

for $c \in D = S$ and $d \in D_n = S_n$. We call D a *Zhang twist* of S . Zhang [1996] showed that there is an equivalence of categories $\Phi : \text{Gr}(S) \rightarrow \text{Gr}(D)$ defined as

follows: if M is a graded left S -module, then ΦM is M as a graded \mathbb{k} -vector space, but endowed with the D -action

$$c * m := \phi^n(c)m$$

for $c \in D = S$ and $m \in (\Phi M)_n = M_n$.

On morphisms Φ is the “identity”: if $f \in \text{Hom}_{\text{Gr}(S)}(M, M')$, then $\Phi(f) = f$ considered now as a morphism $\Phi M \rightarrow \Phi M'$. Note that f is a morphism of graded D -modules because if $c \in D$ and $m \in M_n$, then $f(c * m) = f(\phi^n(c)m) = \phi^n(c)f(m) = c * f(m)$.

We use the graded algebra automorphism $\phi : S \rightarrow S$ defined by

$$(3-1) \quad \phi(s) := K's(K')^{-1}.$$

This is a homomorphism because $K'S = SK'$, and is an automorphism because S is a 4-dimensional Artin–Schelter regular algebra and therefore a domain [Artin et al. 1991, Theorem 3.9].

Proposition 3.1. *Let D be the Zhang twist of S with respect to ϕ (3-1). Then D is isomorphic to $\mathbb{C}\langle E, F, K, K' \rangle$ modulo the relations*

$$\begin{aligned} [K', E] &= [K', F] = [K', K] = 0, \\ KE &= q^2 EK, \quad KF = q^{-2} FK, \quad qEF - q^{-1} FE = \frac{K^2 - K'^2}{q - q^{-1}}. \end{aligned}$$

In particular, K' belongs to the center of D .

Proof. Since $\phi(E) = q^{-1}E$, $\phi(F) = qF$, $\phi(K) = K$, and $\phi(K') = K'$,

$$K' * E = \phi(K')E = K'E = q^{-1}EK' = \phi(E)K' = E * K',$$

$$K' * F = \phi(K')F = K'F = qFK' = \phi(F)K' = F * K',$$

$$K * E = \phi(K)E = KE = qEK = q^2\phi(E)K = q^2E * K,$$

$$K * F = \phi(K)F = KF = q^{-1}FK = q^{-2}\phi(F)K = q^{-2}F * K,$$

and

$$qE * F - q^{-1}F * E = EF - FE = \frac{K^2 - K'^2}{q - q^{-1}}.$$

By the very definition of ϕ , K' belongs to the center of D . □

Corollary 3.2. *Let A be $\mathbb{C}\langle E, F, K \rangle$ modulo the relations*

$$KE = q^2 EK, \quad KF = q^{-2} FK, \quad qEF - q^{-1} FE = \frac{K^2}{q - q^{-1}}.$$

Then $A \cong D/(K')$ and D is a central extension of A in the sense of [Le Bruyn et al. 1996, Definition 3.1.1].

3B. Applying the results in [Le Bruyn et al. 1996]. In the notation of [Le Bruyn et al. 1996], our (E, F, K, K') is their (x_1, x_2, x_3, z) . Following the notation of equation (3.1) and Section 4.2 in that paper, if $A = \mathbb{C}\langle x_1, x_2, x_3 \rangle / (f_1, f_2, f_3)$, then the defining relations for a central extension D of A can be written as²

$$\begin{aligned} zx_i - x_i z &= 0, \quad j = 1, 2, 3, \\ g_j &:= f_j + z l_j + \alpha_j z^2 = 0, \quad j = 1, 2, 3, \end{aligned}$$

for some $l_j \in A_1$ and $\alpha_j \in \mathbb{C}$. For our D ,

$$(3-2) \quad \begin{aligned} \mathbf{f} &:= \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} -q^3 K F + q F K \\ q^{-3} K E - q^{-1} E K \\ q E F - q^{-1} F E + \kappa K^2 \end{pmatrix}, \quad \mathbf{l} := \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \\ \boldsymbol{\alpha} &:= \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\kappa \end{pmatrix}. \end{aligned}$$

Thus

$$(3-3) \quad \mathbf{g} := \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} -q^3 K F + q F K \\ q^{-3} K E - q^{-1} E K \\ q E F - q^{-1} F E + \kappa K^2 - \kappa K'^2 \end{pmatrix}.$$

The relations of A are said to be in *standard form* [Artin et al. 1990, p. 34] if, in the notation of [Le Bruyn et al. 1996, p. 181], there is a 3×3 matrix M , and a matrix $Q \in \text{GL}(3)$, such that $\mathbf{f} = M\mathbf{x}$ and $\mathbf{x}^\top M = (Q\mathbf{f})^\top$, where $\mathbf{f}^\top = (f_1, f_2, f_3)$ and A is generated as an algebra by the entries of the column vector \mathbf{x} .

Proposition 3.3. *The relations \mathbf{f} for A in (3-2) are in standard form, where*

$$(3-4) \quad \begin{aligned} \mathbf{x} &= (E, F, K)^\top, \\ Q &= \text{diag}(q^{-4}, q^4, 1), \end{aligned}$$

and

$$(3-5) \quad M = \begin{pmatrix} 0 & -q^3 K & q F \\ q^{-3} K & 0 & -q^{-1} E \\ -q^{-1} F & q E & \kappa K \end{pmatrix}.$$

Proof. It is easy to check that $\mathbf{f} = M\mathbf{x}$. On the other hand,

$$\mathbf{x}^\top M = (E, F, K)M = (q^{-3} F K - q^{-1} K F, -q^3 E K + q K E, q E F - q^{-1} F E + \kappa K^2),$$

²In the notation of [Le Bruyn et al. 1996, Theorem 3.1.3], $\gamma_j = 0$ for all j .

so

$$(\mathbf{x}^\top M)^\top = \begin{pmatrix} q^{-3}FK - q^{-1}KF \\ -q^3EK + qKE \\ qEF - q^{-1}FE + \kappa K^2 \end{pmatrix} = \begin{pmatrix} q^{-4} & 0 & 0 \\ 0 & q^4 & 0 \\ 0 & 0 & 1 \end{pmatrix} f.$$

Thus $\mathbf{x}^\top M = (Qf)^\top$ as claimed. \square

We use (E, F, K) as homogeneous coordinates on the plane $\mathbb{P}(A_1^*) \cong \mathbb{P}^2$ and identify this plane with the hyperplane $\{K' = 0\}$ in $\mathbb{P}(D_1^*)$.

Proposition 3.4. *The point scheme $(\mathcal{P}_A, \sigma_A)$ for A is the cubic divisor on $\{K' = 0\} = \mathbb{P}(A_1^*)$ consisting of the line $K = 0$ and the conic $\kappa^2 K^2 + EF = 0$. The line meets the conic at the points $(1, 0, 0)$ and $(0, 1, 0)$.*

(1) *If (ξ_1, ξ_2, ξ_3) lies on the conic $\kappa^2 K^2 + EF = 0$, then*

$$\sigma_A(\xi_1, \xi_2, \xi_3) = (q^2 \xi_1, q^{-2} \xi_2, \xi_3).$$

(2) *If (ξ_1, ξ_2, ξ_3) lies on the line $K = 0$; i.e., $\xi_3 = 0$, then*

$$\sigma_A(\xi_1, \xi_2, 0) = (q \xi_1, q^{-1} \xi_2, 0).$$

Proof. By [Artin et al. 1990], the subscheme of $\mathbb{P}(A_1^*)$ parametrizing the left point modules for A is given by the equation

$$\det \begin{pmatrix} 0 & -q^2 K & F \\ q^{-2} K & 0 & -E \\ -q^{-1} F & qE & \kappa K \end{pmatrix} = 0.$$

The vanishing locus of this determinant is the union of the line $K = 0$ and the smooth conic $\kappa^2 K^2 + EF = 0$. The line meets the conic at the points $(1, 0, 0)$ and $(0, 1, 0)$.

We denote this cubic curve by \mathcal{P}_A . The point module corresponding to a point $p \in \mathcal{P}_A$ is $M_p := A/Ap^\perp$, where p^\perp is the subspace of A_1 consisting of the linear forms that vanish at p .

If M_p is a point module for A so is $(M_p)_{\geq 1}(1)$. In keeping with the notation in [Le Bruyn et al. 1996], we write σ_A (in this proof we will use σ for brevity) for the automorphism of \mathcal{P}_A such that

$$(3-6) \quad M_{\sigma^{-1}(p)} \cong (M_p)_{\geq 1}(1).$$

To determine σ explicitly, let $p \in \mathcal{P}_A$ and suppose that $p = (\xi'_1, \xi'_2, \xi'_3)$ and $\sigma^{-1}(p) = (\xi_1, \xi_2, \xi_3)$ with respect to the homogeneous coordinates (E, F, K) . Then M_p has a homogeneous basis e_0, e_1, \dots , where $\deg(e_n) = n$, and

$$Ee_0 = \xi'_1 e_1, \quad Fe_0 = \xi'_2 e_1, \quad Ke_0 = \xi'_3 e_1,$$

and

$$Ee_1 = \xi_1 e_2, \quad Fe_1 = \xi_2 e_2, \quad Ke_1 = \xi_3 e_2.$$

Since $KE - q^2EK = 0$ in A , $(KE - q^2EK)e_0 = 0$; i.e., $\xi_3\xi'_1 - q^2\xi_1\xi'_3 = 0$. The other two relations for A in Corollary 3.2 imply $\xi_3\xi'_2 - q^{-2}\xi_2\xi'_3 = 0$ and $q\xi_1\xi'_2 - q^{-1}\xi_2\xi'_1 + \kappa\xi_3\xi'_3 = 0$. These equalities can be expressed as the single equality

$$\begin{pmatrix} 0 & \xi_3 & -q^{-2}\xi_2 \\ \xi_3 & 0 & -q^2\xi_1 \\ -q^{-1}\xi_2 & q\xi_1 & \kappa\xi_3 \end{pmatrix} \begin{pmatrix} \xi'_1 \\ \xi'_2 \\ \xi'_3 \end{pmatrix} = 0.$$

Since $\sigma(\xi_1, \xi_2, \xi_3) = (\xi'_1, \xi'_2, \xi'_3)$, we can now determine σ .

If (ξ_1, ξ_2, ξ_3) lies on the line $K = 0$; i.e., $\xi_3 = 0$, then

$$\begin{pmatrix} 0 & 0 & -q^{-2}\xi_2 \\ 0 & 0 & -q^2\xi_1 \\ -q^{-1}\xi_2 & q\xi_1 & 0 \end{pmatrix} \begin{pmatrix} \xi'_1 \\ \xi'_2 \\ \xi'_3 \end{pmatrix} = 0,$$

so $\sigma(\xi_1, \xi_2, 0) = (q\xi_1, q^{-1}\xi_2, 0)$. If (ξ_1, ξ_2, ξ_3) lies on the conic $\kappa^2 K^2 + EF = 0$, then

$$\begin{pmatrix} 0 & \xi_3 & -q^{-2}\xi_2 \\ \xi_3 & 0 & -q^2\xi_1 \\ -q^{-1}\xi_2 & q\xi_1 & \kappa\xi_3 \end{pmatrix} \begin{pmatrix} q^2\xi_1 \\ q^{-2}\xi_2 \\ \xi_3 \end{pmatrix} = 0,$$

so $\sigma(\xi_1, \xi_2, \xi_3) = (q^2\xi_1, q^{-2}\xi_2, \xi_3)$. □

The algebra A is of Type S'_1 in the terminology of [Artin et al. 1990, Proposition 4.13, p. 54]. See also [Le Bruyn et al. 1996, p. 187], where it is stated that D is the unique central extension of A that is not a polynomial extension, up to the notion of equivalence at [Le Bruyn et al. 1996, §3.1, p. 180].

In the next result, which is similar to Proposition 3.4, we use (E, F, K') as homogeneous coordinate functions on the plane $K = 0$ in $\mathbb{P}(D_1^*)$.

Proposition 3.5. *Let $A' = D/(K)$. The point scheme $(\mathcal{P}_{A'}, \sigma_{A'})$ for A' is the cubic divisor on the plane $K = 0$ consisting of the line $K' = 0$ and the smooth conic $EF + \kappa^2 K'^2 = 0$. The line meets the conic at the points $(1, 0, 0)$ and $(0, 1, 0)$.*

(1) *If (ξ_1, ξ_2, ξ_4) lies on the conic $\kappa^2 K'^2 + EF = 0$, then*

$$\sigma_{A'}(\xi_1, \xi_2, \xi_4) = (\xi_1, \xi_2, \xi_4).$$

(2) *If (ξ_1, ξ_2, ξ_4) lies on the line $K = 0$; i.e., $\xi_4 = 0$, then*

$$\sigma_{A'}(\xi_1, \xi_2, 0) = (q\xi_2, q^{-1}\xi_1, 0).$$

Proof. Since $[E, K'] = [F, K'] = qEF - q^{-1}FE - \kappa K'^2 = 0$ are defining relations for $A' = \mathbb{C}[E, F, K']$, the left point modules for A' are naturally parametrized by the scheme-theoretic zero locus of

$$\det \begin{pmatrix} 0 & K' & -F \\ K' & 0 & -E \\ -q^{-1}F & qE & -\kappa K' \end{pmatrix}$$

in $\mathbb{P}(A_1'^*)$, namely the union of the line $K' = 0$ and the smooth conic $\kappa^2 K'^2 + EF = 0$.

We denote this cubic curve by $\mathcal{P}_{A'}$ and define $\sigma_{A'} : \mathcal{P}_{A'} \rightarrow \mathcal{P}_{A'}$ (in this proof we will use σ for brevity) by the requirement that $M_{\sigma^{-1}(p)} \cong (M_p)_{\geq 1}(1)$ for all $p \in \mathcal{P}_{A'}$. Calculations like those in Proposition 3.4 show that σ is the identity on the conic and is given by $\sigma(\xi_1, \xi_2, 0) = (q\xi_2, q^{-1}\xi_1, 0)$ on the line $K' = 0$. \square

3C. The point scheme for D . By [Le Bruyn et al. 1996, Theorem 4.2.2] the point scheme $(\mathcal{P}_D, \sigma_D)$ for D exists. That result also gives an explicit description of \mathcal{P}_D . It is also pointed out there that the restriction of σ_D to $\mathcal{P}_D - \mathcal{P}_A$ is the identity.

Warning: The g_1, g_2, g_3 in (3-3) belong to the tensor algebra $T(D_1)$. The g_1, g_2, g_3 in the next result are the images of the g_1, g_2, g_3 in (3-3) in the polynomial ring generated by the indeterminates E, F, K, K' .

Proposition 3.6 [Le Bruyn et al. 1996, Lemma 4.2.1 and Theorem 4.2.2]. *Let $x_1 = E, x_2 = F$, and $x_3 = K$. The equations for \mathcal{P}_D are*

(1) $g_1 = g_2 = g_3 = 0$ on $\mathcal{P}_D \cap \{K' \neq 0\}$, where

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} (q - q^3)FK \\ (q^{-3} - q^{-1})EK \\ -\kappa^{-1}EF + \kappa K^2 - \kappa K'^2 \end{pmatrix} = \kappa^{-1} \begin{pmatrix} q^2 FK \\ q^{-2} EK \\ -EF + \kappa^2(K^2 - K'^2) \end{pmatrix},$$

(2) $K'g_1 = K'g_2 = K'g_3 = h_i = 0$ on $\mathcal{P}_D \cap \{x_i \neq 0\}$, where

$$\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = \kappa^{-1} K \begin{pmatrix} E(EF + \kappa^2 K^2 - \kappa^2 q^2 K'^2) \\ F(EF + \kappa^2 K^2 - \kappa^2 q^{-2} K'^2) \\ K(EF + \kappa^2 K^2 - \kappa^2 K'^2) \end{pmatrix}.$$

Proof. The polynomials h_1, h_2 , and h_3 are defined in [Le Bruyn et al. 1996, Lemma 4.2.1].

Denote the columns of M by M_1, M_2, M_3 , so that $M = [M_1 \ M_2 \ M_3]$, and note that

$$\det(M) = (\kappa K^2 + \kappa^{-1} EF)K.$$

In this case, since $\mathbf{l} = 0$, the definitions of h_1 , h_2 , and h_3 , in [Le Bruyn et al. 1996, Lemma 4.2.1] become

$$\begin{aligned} h_1 &= E \det(M) + z^2 \det[\alpha M_2 M_3], \\ h_2 &= F \det(M) + z^2 \det[M_1 \alpha M_3], \\ h_3 &= K \det(M) + z^2 \det[M_1 M_2 \alpha]. \end{aligned}$$

Since $\alpha = (0, 0, -\kappa)^\top$,

$$\begin{aligned} h_1 &= EK(\kappa K^2 + \kappa^{-1}EF) - \kappa q^2 K'^2 EK, \\ h_2 &= FK(\kappa K^2 + \kappa^{-1}EF) - \kappa q^{-2} K'^2 FK, \\ h_3 &= K^2(\kappa K^2 + \kappa^{-1}EF) - \kappa K'^2 K^2. \end{aligned}$$

Hence the result. □

Theorem 3.7. *The point scheme \mathcal{P}_D is reduced and is the union of*

- (1) *the conics $EF + \kappa^2 K^2 = K' = 0$ and $EF + \kappa^2 K'^2 = K = 0$,*
- (2) *the line $K = K' = 0$, and*
- (3) *the points $(0, 0, 1, \pm 1)$.*

Let $p \in \mathcal{P}_D$.

- (4) *If $p = (\xi_1, \xi_2, \xi_3, 0)$ is on the conic $EF + \kappa^2 K^2 = K' = 0$, then $\sigma_D(p) = (q^2 \xi_1, q^{-2} \xi_2, \xi_3, 0)$.*
- (5) *If $p = (\xi_1, \xi_2, 0, \xi_4)$ is on the conic $EF + \kappa^2 K'^2 = K = 0$, then $\sigma_D(p) = p$.*
- (6) *If $p = (\xi_1, \xi_2, 0, 0)$ is on the line $K = K' = 0$, then $\sigma_D(p) = (q \xi_1, q^{-1} \xi_2, 0, 0)$.*
- (7) *If $p = (0, 0, 1, \pm 1)$, then $\sigma_D(p) = p$.*

Proof. By [Le Bruyn et al. 1996, Theorem 4.2.2], $(\mathcal{P}_D)_{\text{red}} = (\mathcal{P}_A)_{\text{red}} \cup V(g_1, g_2, g_3)_{\text{red}}$, where $V(g_1, g_2, g_3)$ is the scheme-theoretic zero locus of the ideal $(EK, FK, EF - \kappa^2(K'^2 - K^2))$. Certainly \mathcal{P}_A is reduced. Straightforward computations on the open affine pieces $E \neq 0$, $F \neq 0$, $K \neq 0$, and $K' \neq 0$, show that $V(g_1, g_2, g_3)$ is reduced. Hence \mathcal{P}_D is reduced.

If $p = (0, 0, 1, \pm 1)$, then $M_p = D/Dp^\perp = D/DE + DF + D(K \mp K')$. But $DE + DF + D(K \mp K')$ is a two-sided ideal and the quotient by it is the polynomial ring in one variable. Hence $\sigma_D(p) = p$. □

3D. The line modules for D . We now use the results in [Le Bruyn et al. 1996, §5] to characterize the line modules for D . Recall from Section 2C2 that

$$\mathcal{L}_D = \{\text{lines } \ell \text{ in } \mathbb{P}(D_1^*) \mid D/D\ell^\perp \text{ is a line module}\}.$$

For each point $p \in \mathcal{P}_A$, let

$$\mathcal{L}_p := \{\ell \in \mathcal{L}_D \mid p \in \ell\}.$$

Then

$$\mathcal{L}_D = \{\text{lines on the plane } K' = 0\} \cup \bigcup_{p \in \mathcal{P}_A} \mathcal{L}_p.$$

Proposition 3.8. *Let M be a line module for D .*

- (1) *There is a unique line ℓ in $\mathbb{P}(D_1^*)$ such that $M \cong D/D\ell^\perp$.*
- (2) *If $K'M = 0$, then $\ell \subseteq \{K' = 0\}$ and M is a line module for $A = D/(K')$.*
- (3) *The line modules for A are, up to isomorphism, $A/A\ell^\perp$, where $\ell \subseteq \{K' = 0\}$.*
- (4) *If $K'M \neq 0$, then $M/K'M$ is a point module for A and isomorphic to A/Ap^\perp , where $\{p\} = \ell \cap \{K' = 0\}$.*

Proof. (1) This is a consequence of [Levasseur and Smith 1993, Proposition 2.8], which says that if A is a noetherian, Auslander regular, graded \mathbb{k} -algebra having Hilbert series $(1-t)^{-4}$, and is generated by A_1 , and satisfies the Cohen–Macaulay property, then there is a bijection

$$\{\text{lines } u = v = 0 \text{ in } \mathbb{P}(A_1^*) \mid \text{there is a rank 2 relation } a \otimes u - b \otimes v\} \longleftrightarrow \frac{A}{Au + Av}$$

between certain lines in $\mathbb{P}(A_1^*)$ and the set of isomorphism classes of line modules for A .

(2) This is obvious.

(3) Since A is a 3-dimensional Artin–Schelter regular algebra, by [Artin et al. 1990], the isoclasses of the line modules for A are the modules $A/A\ell^\perp$, where ℓ ranges over all lines in $\mathbb{P}(A_1^*) = \{K' = 0\}$.

(4) See the discussion at [Le Bruyn et al. 1996, p. 204]. \square

3D1. *The quadrics Q_p .* By [Le Bruyn et al. 1996, Theorem 5.1.6], if $p \in \mathcal{P}_A$ there is a quadric Q_p containing p such that

$$\mathcal{L}_p = \{\text{lines } \ell \subseteq \{K' = 0\} \text{ such that } p \in \ell\} \cup \{\text{lines } \ell \subseteq Q_p \text{ such that } p \in \ell\}.$$

By [Le Bruyn et al. 1996, Proposition 5.1.7], if $\sigma_D(p) = (\zeta_1, \zeta_2, \zeta_3, 0)$, then Q_p is given by the equation $\zeta^\top Q \mathbf{g} = 0$, where $\zeta = (\zeta_1, \zeta_2, \zeta_3)^\top$, Q is the matrix in Proposition 3.3, $\mathbf{g} = (g_1, g_2, g_3)^\top = \mathbf{f} + \alpha K'^2$, \mathbf{f} is the image in the polynomial ring $\mathbb{C}[E, F, K]$ of the column vector \mathbf{f} introduced in the proof of Proposition 3.3, and $\alpha = (0, 0, -\kappa)^\top$. Thus, Q_p is given by the equation

$$(\zeta_1, \zeta_2, \zeta_3) \begin{pmatrix} q^{-4} & 0 & 0 \\ 0 & q^4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \kappa^{-1} \begin{pmatrix} q^2 FK \\ q^{-2} EK \\ -EF + \kappa^2 K^2 - \kappa^2 K'^2 \end{pmatrix} = 0$$

or, equivalently, by

$$\kappa^{-1}(\zeta_1, \zeta_2, \zeta_3) \begin{pmatrix} q^{-2}FK \\ q^2EK \\ -EF + \kappa^2K^2 - \kappa^2K'^2 \end{pmatrix} = 0.$$

The next result determines \mathcal{L}_p for each $p \in \mathcal{P}_A$. We take coordinates with respect to the coordinate functions (E, F, K, K') .

Proposition 3.9. *Suppose $p = (\xi_1, \xi_2, \xi_3, 0) \in \mathcal{P}_A$.*

(1) *If $p = (1, 0, 0, 0)$, then*

$$\mathcal{L}_p = \{\text{lines } \ell \subseteq \{F = 0\} \cup \{K = 0\} \cup \{K' = 0\} \text{ such that } p \in \ell\}.$$

(2) *If $p = (0, 1, 0, 0)$, then*

$$\mathcal{L}_p = \{\text{lines } \ell \subseteq \{E = 0\} \cup \{K = 0\} \cup \{K' = 0\} \text{ such that } p \in \ell\}.$$

(3) *If $\xi_3 = 0$ and $\xi_1\xi_2 \neq 0$, then*

$$\mathcal{L}_p = \{\text{lines } \ell \subseteq \{K = 0\} \cup \{K' = 0\} \text{ such that } p \in \ell\}.$$

(4) *If $\xi_3 \neq 0$, then Q_p is the cone with vertex p given by the equation*

$$\xi_1 FK + \xi_2 EK + \xi_3(-EF + \kappa^2 K^2 - \kappa^2 K'^2) = 0,$$

and

$$\mathcal{L}_p = \{\text{lines } \ell \subseteq Q_p \cup \{K' = 0\} \text{ such that } p \in \ell\}.$$

Proof. (1)–(3) Suppose $\xi_3 = 0$. Then p is on the line $\{K = K' = 0\}$ whence $\sigma_D(p) = (\xi_1, \xi_2, 0, 0)$ and Q_p is given by the equation

$$\kappa^{-1}(\xi_1 q^{-2}F + \xi_2 q^2 E)K = 0.$$

Thus, ℓ either lies on the pair of planes $\{KK' = 0\}$ or the plane $\{\xi_1 q^{-2}F + \xi_2 q^2 E = 0\}$. Suppose $\ell \not\subseteq \{KK' = 0\}$. We are assuming that $q^4 + 1 \neq 0$ so, if p is neither $(1, 0, 0, 0)$ nor $(0, 1, 0, 0)$, then $\xi_1 q^{-2}F + \xi_2 q^2 E$ does not vanish at p whence there are no lines through p that lie on the plane $\{\xi_1 q^{-2}F + \xi_2 q^2 E = 0\}$.

Suppose $p = (1, 0, 0, 0)$. Then $Q_p = \{FK = 0\}$ and \mathcal{L}_p consists of the lines through p that are contained in $\{F = 0\} \cup \{K = 0\}$. Similarly, if $p = (0, 1, 0, 0)$, then \mathcal{L}_p consists of the lines through p that are contained in $\{E = 0\} \cup \{K = 0\}$.

(4) Suppose $\xi_3 \neq 0$. Then p lies on the conic $\kappa^2 K^2 + EF = K' = 0$ so $\sigma_A(p) = (q^2 \xi_1, q^{-2} \xi_2, \xi_3, 0)$. Thus, Q_p is given by the equation

$$\kappa^{-1}(q^2 \xi_1, q^{-2} \xi_2, \xi_3) \begin{pmatrix} q^{-2}FK \\ q^2EK \\ -EF + \kappa^2 K^2 - \kappa^2 K'^2 \end{pmatrix} = 0;$$

i.e., by the equation $\xi_1 FK + \xi_2 EK + \xi_3(-EF + \kappa^2 K^2 - \kappa^2 K'^2) = 0$. The quadric Q_p is singular at p (and is therefore a cone) because the partial derivatives at the point $p = (\xi_1, \xi_2, \xi_3, 0)$ are

$$\begin{aligned}\partial_E : (\xi_2 K - \xi_3 F)|_p &= 0, \\ \partial_F : (\xi_1 K - \xi_3 E)|_p &= 0, \\ \partial_K : (\xi_1 F + \xi_2 E + 2\xi_3 \kappa^2 K)|_p &= 0, \\ \partial_{K'} : 2\xi_3 \kappa^2 K'|_p &= 0.\end{aligned}$$

It follows that every line contained in Q_p passes through p , and hence all such lines correspond to line modules by [Le Bruyn et al. 1996, Theorem 5.1.6], as recalled above in Section 3D1. \square

Let C denote the conic $K' = \kappa^2 K^2 + EF = 0$ and C' the conic $K = \kappa^2 K'^2 + EF = 0$. The two isolated points in \mathcal{P}_D , namely $(0, 0, 1, \pm 1)$, lie on Q_p for every $p \in \{K' = \kappa^2 K^2 + EF = 0\}$ so there is a pencil of lines (or two pencils) through each of these points that correspond to line modules.

Since Q_p is singular at p , [Le Bruyn et al. 1996, Lemma 5.1.10] implies that $p = p^\vee$ in the notation defined on page 204 of *loc. cit.* Thus, according to [Le Bruyn et al. 1996, Definition 5.1.9], p is of the third kind. Thus, we are in the last case of [Le Bruyn et al. 1996, Table 1].

4. Points, lines, and quadrics in $\text{Proj}_{\text{nc}}(S)$

We now transfer the results in Section 3 from D to S . Recall that the automorphism $\phi : S \rightarrow S$ defined in (3-1) induces an equivalence of categories $\Phi : \text{Gr}(S) \rightarrow \text{Gr}(D)$. We first note how ϕ and Φ act on linear modules.

4A. Left ideals and linear modules over S and D . If W is a graded subspace of $S = D$, then $D_m * W_j = \phi^j(S_m)W_j = S_m W_j$ so, dropping the $*$, $DW = SW$, i.e., the left ideal of D generated by W is equal to the left ideal of S generated by W . In particular, if I is a graded left ideal of S , then $I = SI = DI$ so I is also a left ideal of D . Likewise, if J is graded left ideal of D , then $J = DJ = SJ$ so J is also a left ideal of S .

In summary, D and S have exactly the same left ideals.

Let I be a graded left ideal of S . The equality $S/I = D/I$ is an equality in the category of graded vector spaces. In fact, more is true: $\Phi(S/I) = D/I$ in the category $\text{Gr}(D)$. To see this, observe, first, that the result of applying Φ to the exact sequence $0 \rightarrow I \rightarrow S \rightarrow S/I \rightarrow 0$ in $\text{Gr}(S)$ is the exact sequence $0 \rightarrow I \rightarrow D \rightarrow \Phi(S/I) \rightarrow 0$ in $\text{Gr}(D)$, where $I \rightarrow S$ and $I \rightarrow D$ are the inclusion maps, then use the fact that $\Phi(S/I) = S/I = D/I$ as graded vector spaces.

Now let M be a d -linear S -module. Then $M \cong S/I$ for a unique graded left ideal I in S . Hence $\Phi M \cong \Phi(S/I) = D/I$. In particular, $\Phi(M)$ is a d -linear D -module. Hence

$$\{\text{left ideals } I \text{ in } S \mid S/I \text{ is } d\text{-linear}\} = \{\text{left ideals } I \text{ in } D \mid D/I \text{ is } d\text{-linear}\}.$$

Similarly, if $S_0 = k$, then

$$\{\text{subspaces } W \subset S_1 \mid S/SW \text{ is } d\text{-linear}\} = \{\text{subspaces } W \subset D_1 \mid D/DW \text{ is } d\text{-linear}\}.$$

Lemma 4.1. (1) $\mathcal{P}_S = \mathcal{P}_D$ and $\mathcal{L}_S = \mathcal{L}_D$.

(2) If $\sigma_S : \mathcal{P}_S \rightarrow \mathcal{P}_S$ is a bijection such that $(M_p)_{\geq 1}(1) \cong M_{\sigma_S^{-1}p}$ for all $p \in \mathcal{P}_S$, then there is a bijection $\sigma_D : \mathcal{P}_D \rightarrow \mathcal{P}_D$ such that $(M_p)_{\geq 1}(1) \cong M_{\sigma_D^{-1}p}$ for all $p \in \mathcal{P}_D$, namely $\sigma_D = \sigma_S \phi$.

Proof. (1) This follows from the discussion prior to the lemma.

(2) Let $p \in \mathcal{P}_S$. Suppose that $p = (\xi'_0, \dots, \xi'_n)$ and $\sigma_S^{-1}(p) = (\xi_0, \dots, \xi_n)$ with respect to an ordered basis x_1, \dots, x_n for S_1 . There is a homogeneous basis e_0, e_1, \dots , where $\deg(e_n) = n$, for the S -module M_p such that $x_i e_0 = \xi'_i e_1$ and $x_i e_1 = \xi_i e_2$ for all i . Hence $(\xi_i x_j - \xi_j x_i) e_1 = 0$ for all $1 \leq i, j \leq n$.

Since

$$\begin{aligned} \sigma_D^{-1}(p)^\perp &= \{x \in D_1 \mid x * e_1 = 0\} \\ &= \{x \in D_1 \mid \phi(x) e_1 = 0\} \\ &= \{\phi^{-1}(x) \in D_1 \mid x e_1 = 0\} \\ &= \phi^{-1}(\{x \in D_1 = S_1 \mid x e_1 = 0\}) \\ &= \phi^{-1}(\sigma_S^{-1}(p)^\perp) \\ &= \phi^{-1}(\sigma_S^{-1}(p))^\perp, \end{aligned}$$

$$\sigma_D^{-1} = \phi^{-1} \sigma_S^{-1} \text{ and } \sigma_D = \sigma_S \phi. \quad \square$$

4B. Points in $\text{Proj}_{\text{nc}}(S)$, the point scheme of S , and point modules. We restate Lemma 4.1(1) explicitly in the following theorem.

Theorem 4.2. *The point scheme \mathcal{P}_S is reduced. It is the union of*

- (1) *the conics $EF + \kappa^2 K^2 = K' = 0$ and $EF + \kappa^2 K'^2 = K = 0$,*
- (2) *the line $K = K' = 0$, and*
- (3) *the points $(0, 0, 1, \pm 1)$.*

Furthermore,

- (4) *if $p = (\xi_1, \xi_2, \xi_3, 0)$ is on the conic $EF + \kappa^2 K^2 = K' = 0$, then $\sigma_S(p) = (q\xi_1, q^{-1}\xi_2, \xi_3, 0)$;*

- (5) if $p = (\xi_1, \xi_2, 0, \xi_4)$ is on the conic $EF + \kappa^2 K'^2 = K = 0$, then $\sigma_S(p) = (q^{-1}\xi_1, q\xi_2, 0, \xi_4)$;
- (6) if $p = (\xi_1, \xi_2, 0, 0)$ is on the line $K = K' = 0$, then $\sigma_S(p) = p$;
- (7) if $p = (0, 0, 1, \pm 1)$, then $\sigma_S(p) = p$.

Proof. By Lemma 4.1, $\mathcal{P}_S = \mathcal{P}_D$. However, \mathcal{P}_D is reduced so \mathcal{P}_S is reduced. By Theorem 3.7, the irreducible components of \mathcal{P}_D are the varieties in parts (1), (2) and (3) of this theorem. Hence the same is true of \mathcal{P}_S .

Now suppose there is an ordered basis x_1, \dots, x_n for S_1 and scalars $\lambda_1, \dots, \lambda_n$ such that $\phi(x_i) = \lambda_i x_i$ for all i . Let $p = (\xi_1, \dots, \xi_n) \in \mathbb{P}(S_1^*)$, where the coordinates are written with respect to the ordered basis x_1, \dots, x_n . Let ξ be the point in S_1^* with coordinates (ξ_1, \dots, ξ_n) ; i.e., $x_i(\xi) = \xi_i$ or, equivalently, $\xi(x_i) = \xi_i$. Since $\phi(\xi)(x_i) = \xi(\phi^{-1}x_i) = \xi(\lambda_i^{-1}x_i)$, $\phi(\xi) = (\lambda_1^{-1}\xi_1, \dots, \lambda_n^{-1}\xi_n)$. Hence

$$\phi(p) = (\lambda_1^{-1}\xi_1, \dots, \lambda_n^{-1}\xi_n).$$

Note that E, F, K , and K' in S are eigenvectors for ϕ with eigenvalues $q^{-1}, q, 1$, and 1 , respectively. Thus if $p = (\xi_1, \xi_2, \xi_3, \xi_4) \in \mathbb{P}(S_1^*)$ with respect to the ordered basis E, F, K, K' , then $\phi(p) = (q\xi_1, q^{-1}\xi_2, \xi_3, \xi_4)$ and $\phi^{-1}(p) = (q^{-1}\xi_1, q\xi_2, \xi_3, \xi_4)$.

We now use the description of σ_D in Theorem 3.7 to obtain:

- (1) If $p = (\xi_1, \xi_2, \xi_3, 0) \in \{EF + \kappa^2 K^2 = K' = 0\}$, then $\sigma_D(p) = (q^2\xi_1, q^{-2}\xi_2, \xi_3, 0)$ so $\sigma_S(p) = \sigma_D\phi^{-1}(p) = (q\xi_1, q^{-1}\xi_2, \xi_3, 0)$.
- (2) If $p = (\xi_1, \xi_2, 0, \xi_4) \in \{EF + \kappa^2 K'^2 = K = 0\}$, then $\sigma_D(p) = p$ so $\sigma_S(p) = \sigma_D\phi^{-1}(p) = (q^{-1}\xi_1, q\xi_2, \xi_3, 0)$.
- (3) If $p = (\xi_1, \xi_2, 0, 0) \in \{K = K' = 0\}$, then $\sigma_D(p) = (q\xi_1, q^{-1}\xi_2, 0, 0)$ so $\sigma_S(p) = \phi^{-1}(p) = (\xi_1, \xi_2, 0, 0)$.
- (4) If $p = (0, 0, 1, \pm 1)$, then $\sigma_D(p) = p$ so $\sigma_S(p) = \phi^{-1}(p) = p$. \square

The algebra D is less symmetric than S : the fact that σ_D is the identity on one of the conics but not on the other indicates a certain asymmetry about D . The asymmetry is a result of the fact that we favored K' over K when we formed the Zhang twist of S which made K' , but not K , a central element. Theorem 4.2 shows that the symmetry is restored when the results for \mathcal{P}_D are transferred to \mathcal{P}_S .

4C. Lines and quadrics in $\text{Proj}_{\text{nc}}(S)$. Proposition 3.9 classified the line modules for D , and therefore the line modules for S . Theorem 4.5 below gives a new description of the line modules for S : it says that the line modules correspond to the lines lying on a certain pencil of quadrics. This is analogous to the description in [Le Bruyn and Smith 1993, Theorem 2] of the line modules for the homogenized enveloping algebra of \mathfrak{sl}_2 and the description in [Levasseur and Smith

1993, Theorem 4.5] of the line modules for the 4-dimensional Sklyanin algebra $S(\alpha, \beta, \gamma)$.

The new description provides a unifying picture. The pencil of quadrics becomes more degenerate as one passes from the $S(\alpha, \beta, \gamma)$'s to the homogenizations of the various $U_q(\mathfrak{sl}_2)$ and more degenerate still for $H(\mathfrak{sl}_2)$. The pencil for $H(\mathfrak{sl}_2)$ contains a double plane $t^2 = 0$, that for S contains the pair of planes $\{KK' = 0\}$, and that for $S(\alpha, \beta, \gamma)$ contains 4 cones and the other quadrics in the pencil are smooth.

The vertices of the cones in each pencil play a special role: for $H(\mathfrak{sl}_2)$ there is only one cone and its vertex corresponds to the trivial representation of $U(\mathfrak{sl}_2)$; for S there are two cones and their vertices correspond to the two 1-dimensional representations of $U_q(\mathfrak{sl}_2)$; for $S(\alpha, \beta, \gamma)$ the vertices of the four cones “correspond” to the four special 1-dimensional representations.

4C1. The line through two points $(\xi_1, \xi_2, \xi_3, 0)$ and $(\eta_1, \eta_2, 0, \eta_4)$ in $\mathbb{P}(S_1^*)$ is given by the equations

$$(4-1) \quad \xi_3\eta_4E - \xi_1\eta_4K - \xi_3\eta_1K' = \xi_3\eta_4F - \xi_2\eta_4K - \xi_3\eta_2K' = 0.$$

4C2. The pencil of quadrics $Q(\lambda) \subseteq \mathbb{P}(S_1^*)$. For each $\lambda \in \mathbb{P}^1$, let $Q(\lambda) \subseteq \mathbb{P}(D_1^*)$ be the quadric where

$$g_\lambda := \kappa^{-2}EF + K^2 - (\lambda + \lambda^{-1})KK' + K'^2$$

vanishes. The points on the conics

$$C' : \kappa^2 K'^2 + EF = K = 0,$$

$$C : \kappa^2 K^2 + EF = K' = 0$$

correspond to point modules for S . If $\lambda \neq 0, \infty$, then

$$C' = Q(\lambda) \cap \{K = 0\} \quad \text{and} \quad C = Q(\lambda) \cap \{K' = 0\}.$$

Proposition 4.3. (1) The base locus of the pencil $Q(\lambda)$ is $C \cup C'$.

(2) The $Q(\lambda)$'s are the only quadrics that contain $C \cup C'$.

(3) The singular quadrics in the pencil are the cones $Q(\pm 1)$ with vertices at $(0, 0, 1, \pm 1)$ respectively, and $Q(0) = Q(\infty) = \{KK' = 0\}$.

(4) The lines on $Q(1)$ are $\kappa^{-1}E - s(K - K') = s\kappa^{-1}F + (K - K') = 0$, $s \in \mathbb{P}^1$, and the lines on $Q(-1)$ are $\kappa^{-1}E - s(K + K') = s\kappa^{-1}F + (K + K') = 0$, $s \in \mathbb{P}^1$.

(5) Suppose $\lambda \notin \{0, \pm 1, \infty\}$. The two rulings on $Q(\lambda)$ are

$$(4-2) \quad \kappa^{-1}E - s(K - \lambda K') = s\kappa^{-1}F + (K - \lambda^{-1}K') = 0, \quad s \in \mathbb{P}^1,$$

$$(4-3) \quad \kappa^{-1}E - s(K - \lambda^{-1}K') = s\kappa^{-1}F + (K - \lambda K') = 0, \quad s \in \mathbb{P}^1.$$

Proof. (1) The base locus is, by definition, the intersection of all $Q(\mu)$ so is given by the equations $KK' = \kappa^{-2}EF + K^2 + K'^2 = 0$ so is $\{K' = \kappa^{-2}EF + K^2 = 0\} \cup \{K = \kappa^{-2}EF + K'^2 = 0\}$.

(2), (3) The proofs are straightforward. To prove (3) observe that the determinant of the symmetric matrix representing the g_λ has zeroes at $\lambda = \pm 1$ and a zero at $\lambda = 0, \infty$.

(4), (5) Let ℓ be the line defined by (4-2). Suppose $s \notin \{0, \infty\}$. Then $\kappa^{-1}E = s(K - \lambda K')$ and $s\kappa^{-1}F = -(K - \lambda^{-1}K')$ on ℓ , so $s\kappa^{-2}EF = -s(K - \lambda K')(K - \lambda^{-1}K')$ on ℓ . Canceling s , this says that $\kappa^{-2}EF + (K - \lambda K')(K - \lambda^{-1}K')$ vanishes on ℓ . Since the equation for $Q(\lambda)$ can be written as $\kappa^{-2}EF + (K - \lambda K')(K - \lambda^{-1}K') = 0$, $\ell \subseteq Q(\lambda)$. If $s = 0$, then ℓ is the line $E = K - \lambda^{-1}K' = 0$ which is on $Q(\lambda)$. If $s = \infty$, then ℓ is the line $K + \lambda^{-1}K' = F = 0$ which is on $Q(\lambda)$. The other case, (4-3), is similar. \square

4C3. There are exactly four singular quadrics in a generic pencil of quadrics in \mathbb{P}^3 . The point modules for the 4-dimensional Sklyanin algebras $S(\alpha, \beta, \gamma)$ are parametrized by a quartic elliptic curve $E \subseteq \mathbb{P}^3$ and 4 isolated points that are the vertices of the singular quadrics that contain E . The point modules corresponding to those isolated points correspond to the four 1-parameter families of 1-dimensional representations of a 4-dimensional Sklyanin algebra.

4C4. The vertices of the cones $Q(\pm 1)$ are the points $(0, 0, 1, \pm 1)$. These are the isolated points in the point scheme \mathcal{P}_S (see Theorem 4.2). Later, we will see that the points $(0, 0, 1, \pm 1)$ correspond to the two 1-dimensional $U_q(\mathfrak{sl}_2)$ -modules. More precisely, if p is one of those points, then $M_p[(KK')^{-1}]_0$ is a 1-dimensional $U_q(\mathfrak{sl}_2)$ -module.

4C5. The lines on $Q(1)$ meet C' and C at points of the form $(\xi_1, \xi_2, \xi_3, 0)$ and $(\xi_1, \xi_2, 0, -\xi_3)$ respectively. The lines on $Q(-1)$ meet C' and C at points of the form $(\xi_1, \xi_2, \xi_3, 0)$ and $(\xi_1, \xi_2, 0, \xi_3)$ respectively.

Theorem 4.4. Let ℓ be a line in $\mathbb{P}(S_1^*)$. Then $S/S\ell^\perp$ is a line module if and only if $\ell \subseteq Q(\lambda)$ for some $\lambda \in \mathbb{P}^1$.

Proof. (\Rightarrow) Suppose $S/S\ell^\perp$ is a line module. By Section 4A, $D/D\ell^\perp$ is a line module for D .

The result is true if $\ell \subseteq \{KK' = 0\} = Q(\infty)$ so, from now on, suppose $\ell \not\subseteq \{KK' = 0\}$.

Let $p = (\xi_1, \xi_2, \xi_3, 0)$ be the point where ℓ meets $\{K' = 0\}$. By the discussion at the beginning of Section 3D, A/Ap^\perp is a point module for $A = D/(K')$ so $p \in C \cup \{K = K' = 0\}$.

Suppose $p = (1, 0, 0, 0)$. By Proposition 3.9(1), ℓ is on the plane $\{F = 0\}$. Since $p \in \ell$ it follows that $\ell = \{F = K - \lambda K' = 0\}$ for some $\lambda \in \mathbb{P}^1$. Since $\ell \not\subseteq \{KK' = 0\}$, $\lambda \neq 0, \infty$. Thus ℓ is the line in (4-2) corresponding to the point $s = \infty \in \mathbb{P}^1$. Hence $\ell \subseteq Q(\lambda)$.

If $p = (0, 1, 0, 0)$, a similar argument shows that ℓ lies on some $Q(\lambda)$.

Now suppose that $p \notin \{(1, 0, 0, 0), (0, 1, 0, 0)\}$. Since $\ell \not\subseteq \{KK' = 0\}$, it follows from Proposition 3.9(3) that $\xi_3 \neq 0$. Hence by Proposition 3.9(4), ℓ lies on the quadric

$$\xi_1 FK + \xi_2 EK + \xi_3(-EF + \kappa^2 K - \kappa^2 K'^2) = 0.$$

The conic $C' = \{K = EF + \kappa^2 K'^2 = 0\}$ also lies on this quadric so $C' \cap \ell \neq \emptyset$. by a result analogous to Proposition 3.9(3), $\eta_4 \neq 0$. Let $(\eta_1, \eta_2, 0, \eta_4) \in C' \cap \ell$. Then ℓ is given by the equations in (4-1) so lies on the surface cut out by the equation

$$\begin{aligned} \xi_3^2 \eta_4^2 EF &= (\xi_1 \eta_4 K + \xi_3 \eta_1 K')(\xi_2 \eta_4 K + \xi_3 \eta_2 K') \\ &= \xi_1 \xi_2 \eta_4^2 K^2 + \xi_3 \eta_4 (\xi_1 \eta_2 + \xi_2 \eta_1) K K' + \xi_3^2 \eta_1 \eta_2 K'^2 \\ &= -\kappa^2 \xi_3^2 \eta_4^2 K^2 + \xi_3 \eta_4 (\xi_1 \eta_2 + \xi_2 \eta_1) K K' - \kappa^2 \xi_3^2 \eta_4^2 K'^2, \end{aligned}$$

which can be rewritten as $\xi_3^2 \eta_4^2 (\kappa^{-2} EF + K^2 + K'^2) - \xi_3 \eta_4 (\xi_1 \eta_2 + \xi_2 \eta_1) K K' = 0$. Thus, ℓ lies on some $Q(\lambda)$.

(\Leftarrow) Let ℓ be a line on $Q(\lambda)$. If $\ell \subseteq \{K' = 0\}$, then $D/D\ell^\perp = A/A\ell^\perp$ is a line module. From now on suppose that $\ell \not\subseteq \{K' = 0\}$.

To show that $D/D\ell^\perp$, and hence $S/S\ell^\perp$, is a line module we must show there is a point, p say, in $\ell \cap C$ such that $\ell \subseteq Q_p$, where Q_p is the quadric in Section 3D1.

Suppose $\ell \subseteq Q(1)$. By Proposition 4.3(4), ℓ is given by

$$\kappa^{-1} E - s(K - K') = s\kappa^{-1} F + (K - K') = 0$$

for some $s \in \mathbb{P}^1$. Since the point $p = (-s^2, 1, -s\kappa^{-1}, 0)$ belongs to $\ell \cap C$, $S/S\ell^\perp$ is a line module if and only if $\ell \subseteq Q_p$. Since Q_p is given by the equation

$$-s^2 FK + EK - s\kappa^{-1}(-EF + \kappa^2 K^2 - \kappa^2 K'^2) = 0,$$

the point $(-s^2, 1, 0, s\kappa^{-1})$ is in $\ell \cap Q_p$. Thus, ℓ passes through the vertex of the cone Q_p and through a second point on Q_p , whence $\ell \subseteq Q_p$. Therefore $S/S\ell^\perp$ is a line module.

The case $\ell \subseteq Q(-1)$ is similar.

Suppose $\lambda \notin \{0, \pm 1, \infty\}$. Since $\ell \subseteq Q(\lambda)$ we suppose, without loss of generality, that ℓ belongs to the ruling (4-3) on $Q(\lambda)$. Thus $\ell = \{\kappa^{-1} E - s(K - \lambda K') =$

$s\kappa^{-1}F + (K - \lambda^{-1}K') = 0\}$ for some $s \in \mathbb{P}^1$. The point $p = (-\kappa s^2, \kappa, -s, 0)$, which is the vertex of the cone Q_p , is in $\ell \cap C$. Thus ℓ passes through the vertex of Q_p and the point $(-\kappa s^2 v^2, \kappa, 0, sv)$ which is also on Q_p , $\ell \subseteq Q_p$. Hence $S/S\ell^\perp$ is a line module. \square

Theorem 4.5. *Let ℓ be a line in $\mathbb{P}(S_1^*)$. Then $S/S\ell^\perp$ is a line module if and only if ℓ meets $C \cup C'$ with multiplicity 2; i.e., if and only if ℓ is a secant line to $C \cup C'$.*

Proof. (\Rightarrow) Since $S/S\ell^\perp$ is a line module for S , $D/D\ell^\perp$ is a line module for D . Let λ be such that $\ell \subseteq Q(\lambda)$.

Suppose $Q(\lambda)$ is smooth. The Picard group of $Q(\lambda)$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}$ and equal to $\mathbb{Z}[L] \oplus \mathbb{Z}[L']$, where $[L]$ is the class of the line in (4-2) corresponding to $s = 0$ and $[L']$ is the class of the line in (4-3) corresponding to $s = 0$. Since $L = \{E = K - \lambda^{-1}K' = 0\}$, the scheme-theoretic intersection $L \cap (C \cup C')$ is the zero locus of the ideal

$$\begin{aligned} (E, K - \lambda^{-1}K') + (KK', g_\lambda) \\ = (E, K - \lambda^{-1}K', KK', \kappa^{-2}EF + (K - \lambda K')(K - \lambda^{-1}K')) \\ = (E, K - \lambda^{-1}K', KK'). \end{aligned}$$

Hence $L \cap (C \cup C')$ is a finite scheme of length 2. Therefore $[L] \cdot [C \cup C'] = 2$. A similar calculation shows that $[L'] \cdot [C \cup C'] = 2$. Hence $[C \cup C'] = 2[L] + 2[L']$. It follows that $[\ell] \cdot [C \cup C'] = 2$.

Suppose ℓ lies on the cone $Q(1)$. Then ℓ is the line $\kappa^{-1}E - s(K + K') = s\kappa^{-1}F + (K + K') = 0$ for some $s \in \mathbb{P}^1$. Therefore the scheme-theoretic intersection $\ell \cap (C \cup C')$ is the zero locus of the ideal

$$(4-4) \quad (\kappa^{-1}E - s(K - K'), s\kappa^{-1}F + (K - K')) + (KK', g_1).$$

Since $\ell \subset Q(1)$, g_1 belongs to the ideal vanishing on ℓ . The ideal in (4-4) is therefore equal to $(\kappa^{-1}E - s(K - K'), s\kappa^{-1}F + (K - K'), KK')$. Thus $\ell \cap (C \cup C')$ is a finite scheme of length 2.

If $\ell \subseteq Q(-1)$, a similar argument shows that $\ell \cap (C \cup C')$ is a finite scheme of length 2.

Suppose $\ell \subseteq Q(\infty) = \{KK' = 0\}$. Without loss of generality we can, and do, assume that $\ell \subseteq \{K' = 0\}$. By Bézout's theorem, ℓ meets C with multiplicity two. Thus, if $\ell \cap C' = \emptyset$, then ℓ meets $C \cup C'$ with multiplicity two. Now suppose that ℓ meets C with multiplicity two and C' with multiplicity ≥ 1 . If ℓ meets C at two distinct points, then ℓ is transversal to some $Q(\lambda')$ so meets $Q(\lambda')$, and hence $C \cup C'$, with multiplicity two. It remains to deal with the case where ℓ is tangent to C and meets C' . We now assume that is the case. Since $C \cap C' = \{(1, 0, 0, 0), (0, 1, 0, 0)\}$ it follows that ℓ is tangent to C at $(1, 0, 0, 0)$ or at $(0, 1, 0, 0)$. Since the two cases are similar we assume that ℓ is tangent to C at $(1, 0, 0, 0)$. It follows that

$\ell = \{K' = F = 0\}$. The scheme-theoretic intersection $\ell \cap (C \cup C')$ is the zero locus of the ideal

$$(K, F) + (KK', g_1) = (K, F, KK', \kappa^{-2}EF + K^2 + K'^2) = (K, F, K'^2).$$

Hence $\ell \cap (C \cup C')$ is a finite scheme of length 2.

(\Leftarrow) Suppose ℓ is a line that meets $C \cup C'$ with multiplicity 2.

If ℓ lies on the plane $\{K' = 0\}$, then $K' \in \ell^\perp$ so the ideal (K') of D is contained in $D\ell^\perp$ and $D/D\ell^\perp$ is a module over $A = D/(K')$. However, the dual of the map $D_1 \rightarrow A_1$ embeds $\mathbb{P}(A_1^*)$ in $\mathbb{P}(D_1^*)$ and the image of this embedding is $\{K' = 0\}$. In short, ℓ is a line in $\mathbb{P}(A_1^*)$. This implies that $A/A\ell^\perp$ is a line module for A and hence a line module for D . But $A/A\ell^\perp = D/D\ell^\perp$ so $D/D\ell^\perp$ is a line module for D . Therefore $S/S\ell^\perp$ is a line module for S . A similar argument shows that if ℓ lies on the plane $\{K = 0\}$, then $S/S\ell^\perp$ is a line module.

For the remainder of the proof we assume that $\ell \not\subseteq \{KK' = 0\}$.

A line that meets C with multiplicity two lies in the plane $K' = 0$, so ℓ meets C and C' with multiplicity one. Let $p = (\xi_1, \xi_2, \xi_3, 0)$ be the point where ℓ meets C and $p = (\eta_1, \eta_2, 0, \eta_4)$ be the point where ℓ meets C' . Since p is the vertex of Q_p and $C' \subseteq Q_p$, ℓ passes through the vertex of Q_p and another point on Q_p . Hence $\ell \subseteq Q_p$. \square

4C6. Lemmas 4.4 and 4.5 are analogous to results for the 4-dimensional Sklyanin algebras: the line modules correspond to the lines in \mathbb{P}^3 that lie on the quadrics that contain the quartic elliptic curve E , and those are exactly the lines in \mathbb{P}^3 that meet E with multiplicity two, i.e., the secant lines to E . Similar results hold for the homogenization of \mathfrak{sl}_2 [Le Bruyn and Smith 1993].

4C7. Notation for line modules. By Theorem 4.5, the lines that correspond to line modules for S are the secant lines to $C \cup C'$. If $(p) + (p')$ is a degree-two divisor on $C \cup C'$ we write $M_{p,p'}$ for the line module $M_\ell = S/S\ell^\perp$, where ℓ is the unique line that meets $C \cup C'$ at $(p) + (p')$. Thus, up to isomorphism, the line modules for S are

$$\{M_{p,p'} \mid p, p' \in C \cup C'\}.$$

4D. Incidence relations between lines and points in $\text{Proj}_{\text{nc}}(S)$. Let $(p) + (p')$ be a degree-two divisor on $C \cup C'$. There is a surjective map $M_{p,p'} \twoheadrightarrow M_p$ in $\text{Gr}(S)$ and, by [Levasseur and Smith 1993, Lemma 5.3], the kernel of that homomorphism is isomorphic to $M_{\ell'}(-1)$ for some ℓ' . Our next goal is to determine ℓ' . We do that in Proposition 4.8 below.

First we need the rather nice observation in the next lemma.

We call a degree-three divisor on a plane cubic curve *linear* if it is the scheme-theoretic intersection of that curve and a line.

Lemma 4.6. *Let C be a nondegenerate conic in \mathbb{P}^2 , σ an automorphism of C that fixes two points, and L the line through those two points. Let $p, p' \in C$ and $p'' \in L$. The divisor $(p) + (p') + (p'') \in \text{Div}(C \cup L)$ is linear if and only if $(\sigma p) + (\sigma^{-1} p') + (p'')$ is.*

Proof. By symmetry, it suffices to show that if $(p) + (p') + (p'')$ is linear so is $(\sigma p) + (\sigma^{-1} p') + (p'')$. That is what we will prove. So, assume $(p) + (p') + (p'')$ is linear.

If τ is an automorphism of \mathbb{P}^1 that fixes two points, there are nonzero scalars λ and μ and a choice of coordinates such that $\tau(s, t) = (\lambda s, \mu t)$ for all $(s, t) \in \mathbb{P}^1$. We assume, without loss of generality, that (C, σ) is the image of (\mathbb{P}^1, τ) under the 2-Veronese embedding. Thus, we can assume that C is the curve $xy - z^2 = 0$ and $\sigma(\alpha, \beta, \gamma) = (\lambda^2 \alpha, \mu^2 \beta, \lambda \mu \gamma)$. The line L is the line through $(1, 0, 0)$ and $(0, 1, 0)$.

Let $p = (\alpha, \beta, \gamma)$, $p' = (\alpha', \beta', \gamma')$, and $p'' = (a, b, 0)$. By hypothesis, these three points are collinear. Therefore

$$\det \begin{pmatrix} a & b & 0 \\ \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \end{pmatrix} = 0.$$

I.e., $a(\beta\gamma' - \beta'\gamma) - b(\alpha\gamma' - \alpha'\gamma) = 0$.

To show that $(\sigma p) + (\sigma^{-1} p') + (p'')$ is linear we must show that the points $\sigma p = (\lambda^2 \alpha, \mu^2 \beta, \lambda \mu \gamma)$, $\sigma^{-1} p' = (\lambda^{-2} \alpha', \mu^{-2} \beta', \lambda^{-1} \mu^{-1} \gamma')$, and p'' , are collinear. This is the case if and only if

$$\det \begin{pmatrix} a & b & 0 \\ \lambda^2 \alpha & \mu^2 \beta & \lambda \mu \gamma \\ \lambda^{-2} \alpha' & \mu^{-2} \beta' & \lambda^{-1} \mu^{-1} \gamma' \end{pmatrix} = 0.$$

This determinant is $a(\lambda^{-1} \mu \beta \gamma' - \lambda \mu^{-1} \beta' \gamma) - b(\lambda \mu^{-1} \alpha \gamma' - \lambda^{-1} \mu \alpha' \gamma)$. It is zero if and only if

$$(\alpha \gamma' - \alpha' \gamma)(\lambda^{-1} \mu \beta \gamma' - \lambda \mu^{-1} \beta' \gamma) - (\beta \gamma' - \beta' \gamma)(\lambda \mu^{-1} \alpha \gamma' - \lambda^{-1} \mu \alpha' \gamma) = 0.$$

This expression is equal to

$$\lambda^{-1} \mu (\alpha \beta \gamma'^2 - \alpha' \beta' \gamma^2) + \lambda \mu^{-1} (\alpha' \beta' \gamma^2 - \alpha \beta \gamma'^2).$$

But $\alpha \beta - \gamma^2 = \alpha' \beta' - \gamma'^2 = 0$ so $\alpha \beta \gamma'^2 - \alpha' \beta' \gamma^2 = 0$. Thus, the determinant is zero and we conclude that $\sigma^{-1} p'$, σp , and p'' , are collinear. \square

Remark 4.7. For an alternative approach to Lemma 4.6, note first that the statement can be recast as the claim that if η is the involution of C obtained by “reflection

across p''' meaning that

$$\eta(p) = \text{the second intersection of the line } pp'' \text{ with } C,$$

then $\pi = \eta \circ \sigma$ is an involution.

In turn, the involutivity of π follows from the fact that it interchanges the points p and p' , and any automorphism of \mathbb{P}^1 that interchanges two points is, after a coordinate change identifying said points with $0, \infty$, of the form $z \mapsto tz^{-1}$ for some constant t . \blacklozenge

The next result is analogous to [Levasseur and Smith 1993, Theorem 5.5], which shows for the 4-dimensional Sklyanin algebras that if $(p) + (p')$ is a degree-two divisor on the quartic elliptic curve E , then there is an exact sequence

$$0 \rightarrow M_{p+\tau, p'-\tau}(-1) \rightarrow M_{p, p'} \rightarrow M_p \rightarrow 0,$$

where τ is the point on E such that $\sigma p = p + \tau$ for all $p \in E$.

Proposition 4.8. *If $(p) + (p')$ is a degree-two divisor on $C \cup C'$, there is an exact sequence*

$$0 \rightarrow M_{\sigma p, \sigma^{-1} p'}(-1) \rightarrow M_{p, p'} \rightarrow M_p \rightarrow 0.$$

Proof. Let ℓ be the unique line in $\mathbb{P}^3 = \mathbb{P}(S_1^*)$ such that $\ell \cap C = (p) + (p')$; i.e., $M_{p, p'} = M_\ell$. By [Levasseur and Smith 1993, Lemma 5.3], there is an exact sequence $0 \rightarrow M_{\ell'}(-1) \rightarrow M_\ell \rightarrow M_p \rightarrow 0$ for some line module $M_{\ell'}$. We complete the proof by showing that $\ell' \cap C = (\sigma p) + (\sigma^{-1} p')$; i.e., $M_{\ell'} = M_{\sigma p, \sigma^{-1} p'}$. There are several cases depending on the location of p and p' .

Case 0. Suppose $\ell = L$. Then $K M_\ell = K' M_\ell = 0$ so M_ℓ , and consequently $M_{\ell'}$, is a module over $S/(K, K')$. Since $S/(K, K')$ is a commutative polynomial ring on two indeterminates it has a unique line module up to isomorphism, itself. In particular, $\ell' = \ell$. Hence there is an exact sequence $0 \rightarrow M_{p, p'}(-1) \rightarrow M_{p, p'} \rightarrow M_p \rightarrow 0$. But σ is the identity on L by Theorem 4.2, so $M_{p, p'} = M_{\sigma p, \sigma^{-1} p'}$. Thus, the previous exact sequence is exactly the sequence in the statement of this proposition.

Case 1. Suppose $p, p' \in C$. Then ℓ meets C with multiplicity two and therefore the plane $\{K' = 0\}$ with multiplicity ≥ 2 . Hence $\ell \subseteq \{K' = 0\}$. It follows that M_ℓ and $M_{\ell'}$ are modules over $S/(K')$. Given Case 0 treated above, for the remainder of Case 1 we can, and do, assume that $\ell \neq L$.

Since $\ell \neq L$, $\ell \cap (C + L) = (p) + (p') + (p'')$, where p'' is the point where ℓ and L meet. Since $S/(K')$ is a 3-dimensional Artin–Schelter regular algebra, [Artin et al. 1991, Proposition 6.24] tells us that ℓ' is the unique line such that $\ell' \cap (C + L)$ contains the divisor $(\sigma^{-1} p') + (\sigma^{-1} p'')$.³ By Theorem 4.2, $\sigma p'' = p''$ so ℓ' is the unique

³Since [Artin et al. 1991, Proposition 6.24] is for right modules and we are working with left modules we replaced σ by σ^{-1} in the conclusion of that result.

line in $\{K' = 0\}$ such that $\ell' \cap (C + L)$ contains $(\sigma^{-1}p') + (p'')$. By Lemma 4.6, σp , $\sigma^{-1}p'$, and p'' , are collinear. Therefore $\ell' \cap (C + L) = (\sigma p) + (\sigma^{-1}p') + (p'')$.

Case 2. If $p, p' \in C'$, the “same” argument as in Case 1 proves the proposition.

Case 3. Suppose $p \in C - C'$ and $p' \in C' - C$. Let $p = (\xi_1, \xi_2, \xi_3, 0)$ and $p' = (\eta_1, \eta_2, 0, \eta_4)$. Since $p \notin C'$, $\xi_3 \neq 0$. Since $p' \notin C$, $\eta_4 \neq 0$.

By (4-1), ℓ is given by the equations

$$X := \xi_3\eta_4E - \xi_1\eta_4K - \xi_3\eta_1K' = 0,$$

$$Y := \xi_3\eta_4F - \xi_2\eta_4K - \xi_3\eta_2K' = 0.$$

The corresponding linear modules are

$$M_\ell = M_{p,p'} = \frac{S}{SX + SY},$$

$$M_p = \frac{S}{SK' + SX + SY},$$

$$M_{p'} = \frac{S}{SK + SX + SY}.$$

By (4-1), the line through the points $\sigma p = (q\xi_1, q^{-1}\xi_2, \xi_3, 0)$ and $\sigma^{-1}p' = (q\eta_1, q^{-1}\eta_2, 0, \eta_4)$ is $\{X' = Y' = 0\}$, where

$$X' := \xi_3\eta_4E - q\xi_1\eta_4K - q\xi_3\eta_1K',$$

$$Y' := \xi_3\eta_4F - q^{-1}\xi_2\eta_4K - q^{-1}\xi_3\eta_2K'.$$

The corresponding line module is $M_{\sigma p, \sigma^{-1}p'} = S/SX' + SY'$.

The image of K' in M_ℓ generates the kernel of $M_\ell \rightarrow M_p$. Since $X'K' = qK'X$ and $Y'K' = q^{-1}K'Y$, X' and Y' annihilate the image of K' in M_ℓ . It follows that there is a map from $M_{\sigma p, \sigma^{-1}p'}(-1)$ onto the kernel of $M_\ell \rightarrow M_p$. Thus, the kernel of $M_\ell \rightarrow M_p$ is isomorphic to a quotient of $M_{\sigma p, \sigma^{-1}p'}$. But every nonzero submodule of a line modules has GK dimension 2, and every proper quotient of a line module has GK dimension 1, so every nonzero homomorphism map $M_{\sigma p, \sigma^{-1}p'}(-1) \rightarrow M_\ell$ is injective. This shows that the kernel of $M_\ell \rightarrow M_p$ is isomorphic to $M_{\sigma p, \sigma^{-1}p'}(-1)$.

Case 4. If $p' \in C - C'$ and $p \in C' - C$, the “same” argument as in Case 3 proves the proposition. \square

We continue to write L for the line $\{K = K' = 0\}$.

Proposition 4.9. *Let ℓ be a line in $\text{Proj}_{\text{nc}}(S)^4$ and suppose $p'' \in \ell \cap L$.*

- (1) *There are points $p, p' \in C \cup C'$ such that the scheme-theoretic intersection $\ell \cap (C \cup L)$ contains the divisor $(p) + (p') + (p'')$.*

⁴This means that $M_\ell := S/S\ell^\perp$ is a line module.

(2) *There is an exact sequence*

$$0 \rightarrow M_{\sigma^{-1}p, \sigma^{-1}p'}(-1) \rightarrow M_{p, p'} \rightarrow M_{p''} \rightarrow 0.$$

Proof. Since M_ℓ is a line module, ℓ meets $C \cup C'$ with multiplicity two. It therefore meets either $C \cup L$ or $C' \cup L$ with multiplicity ≥ 2 . We can, and do, assume without loss of generality that ℓ meets $C \cup L$ with multiplicity ≥ 2 . Hence ℓ meets the plane $\{K' = 0\}$ with multiplicity ≥ 2 . Therefore $\ell \subseteq \{K' = 0\}$. By Bézout's theorem, ℓ is either equal to L or meets $C \cup L$ with multiplicity 3.

Suppose $\ell = L$. Then M_ℓ is a module over the commutative polynomial ring $S/(K, K')$ and there is an exact sequence $0 \rightarrow M_\ell(-1) \rightarrow M_\ell \rightarrow M_{p''} \rightarrow 0$. Let p and p' be the points where L meets $C \cup C'$. Then $M_\ell = M_L = M_{p, p'}$ and, since σ is the identity on L , $M_\ell = M_{\sigma p, \sigma^{-1}p'}$. Thus, (1) and (2) hold when $\ell = L$.

Suppose $\ell \neq L$. Let p and p' be the points in $\ell \cap C$; i.e., $\ell \cap C = (p) + (p')$. By [Artin et al. 1991, Proposition 6.24], there is an exact sequence $0 \rightarrow M_{\ell'}(-1) \rightarrow M_\ell = M_{p, p'} \rightarrow M_{p''} \rightarrow 0$, where ℓ' is the unique line whose scheme-theoretic intersection with $C \cup L$ is $\geq (\sigma^{-1}p) + (\sigma^{-1}p')$. Hence $M_{\ell'} = M_{\sigma^{-1}p, \sigma^{-1}p'}$. \square

5. Relation to $U_q(\mathfrak{sl}_2)$ -modules

In this section, we relate our results about fat point and line modules for S to classical results about the finite-dimensional irreducible representations and Verma modules of $U_q(\mathfrak{sl}_2)$. Briefly, fat points in $\text{Proj}_{\text{nc}}(S)$ correspond to finite-dimensional irreducible $U_q(\mathfrak{sl}_2)$ -modules and lines in $\text{Proj}_{\text{nc}}(S)$ correspond to Verma modules.

5A. Facts about $U_q(\mathfrak{sl}_2)$. First, we recall a few facts about $U_q(\mathfrak{sl}_2)$ that can be found in [Jantzen 1996, Chapter 2].

5A1. Verma modules. For each $\lambda \in \mathbb{C}$, we call

$$M(\lambda) := \frac{U_q(\mathfrak{sl}_2)}{U_q(\mathfrak{sl}_2)e + U_q(\mathfrak{sl}_2)(k - \lambda)}$$

a Verma module for $U_q(\mathfrak{sl}_2)$, and λ its *highest weight*.

5A2. Casimir element. The Casimir element

$$(5-1) \quad C := ef + \frac{q^{-1}k + qk^{-1}}{(q - q^{-1})^2} = fe + \frac{qk + q^{-1}k^{-1}}{(q - q^{-1})^2}$$

is in the center of $U_q(\mathfrak{sl}_2)$ and acts on $M(\lambda)$ as multiplication by

$$\frac{q\lambda + q^{-1}\lambda^{-1}}{(q - q^{-1})^2}.$$

5A3. Finite-dimensional simple modules. For each $n \geq 1$, there are exactly two simple $U_q(\mathfrak{sl}_2)$ -modules of dimension $n + 1$. They can be labeled $L(n, +)$ and $L(n, -)$ in such a way that there are exact sequences

$$(5-2) \quad 0 \rightarrow M(\pm q^{-n-2}) \rightarrow M(\pm q^n) \rightarrow L(n, \pm) \rightarrow 0.$$

The module $L(n, \pm)$ has basis m_0, \dots, m_n with action

$$(5-3) \quad \begin{aligned} km_i &= \pm q^{n-2i} m_i, \\ fm_i &= \begin{cases} m_{i+1} & \text{if } i < n, \\ 0 & \text{if } i = n, \end{cases} \quad em_i = \begin{cases} \pm[i][n+1-i]m_{i-1} & \text{if } i > 0, \\ 0 & \text{if } i = 0, \end{cases} \end{aligned}$$

where we have made use of the *quantum integers*

$$[m] := \frac{q^m - q^{-m}}{q - q^{-1}}.$$

5B. Lines in $\text{Proj}_{\text{nc}}(S) \longleftrightarrow$ Verma modules for $U_q(\mathfrak{sl}_2)$. First, we show that Verma modules are “affine pieces” of line modules.

Proposition 5.1. *Let $\lambda \in \mathbb{C} \cup \{\infty\} = \mathbb{P}^1$ and let ℓ be the line $E = K - \lambda K' = 0$.*

- (1) ℓ lies on the quadric $Q(\lambda)$.
- (2) $S/S\ell^\perp$ is a line module.
- (3) If $\lambda \notin \{0, \infty\}$, then $(S/S\ell^\perp)[(KK')^{-1}]_0 \cong M(\lambda)$.

Proof. A simple calculation proves (1), and then (2) follows from Theorem 4.4.

(3) The functor $j^*\pi^* : \text{Gr}(S) \rightarrow \text{Mod}(U_q(\mathfrak{sl}_2))$ defined by $j^*\pi^*M = M[(KK')^{-1}]_0$ is exact, so $(S/S\ell^\perp)[(KK')^{-1}]_0$ is isomorphic to $S(KK')^{-1}]_0 / (S\ell^\perp)[(KK')^{-1}]_0$. Using the isomorphism given by (2-7), it is clear that $(S\ell^\perp)[(KK')^{-1}]_0$ is the left ideal of $U_q(\mathfrak{sl}_2)$ generated by e and $k - \lambda$. \square

5B1. “Heretical” Verma modules. Proposition 5.1 illustrates the importance of line modules for Artin–Schelter regular algebras with Hilbert series $(1 - t)^{-4}$. Line modules are just like Verma modules. Indeed, Verma modules for $U(\mathfrak{sl}_2)$ and $U_q(\mathfrak{sl}_2)$ are “affine pieces” of line modules.

From the point of view of noncommutative projective algebraic geometry, the line modules that correspond to Verma modules are no more special than other line modules. One is tempted to declare that if ℓ is any line on any $Q(\lambda)$, $\lambda \neq 0, \infty$, then $(S/S\ell^\perp)[(KK')^{-1}]_0$ should be considered as a Verma module.

Doing that would place $U_q(\mathfrak{sl}_2)$ on a more equal footing with $U(\mathfrak{sl}_2)$: if one varies both the Borel subalgebra and the highest weight, then $U(\mathfrak{sl}_2)$ has a 2-parameter family of Verma modules; if were to define Verma modules for $U_q(\mathfrak{sl}_2)$ as “affine pieces” of line modules, then $U_q(\mathfrak{sl}_2)$ would also have a 2-parameter family of Verma modules.

5B2. Central (Casimir) elements. We define $\Omega(0) = \Omega(\infty) = KK'$ and, for each $\lambda \in \mathbb{C} - \{0, \infty\}$, we define

$$\begin{aligned}\Omega(\lambda) &:= EF + \frac{q^{-1}K^2 + qK'^2}{(q - q^{-1})^2} - \frac{q\lambda + q^{-1}\lambda^{-1}}{(q - q^{-1})^2}KK' \\ &= EF + \kappa^2(q^{-1}K - q\lambda K')(K - \lambda^{-1}K') \\ &= FE + \kappa^2(qK - q^{-1}\lambda^{-1}K')(K - \lambda K').\end{aligned}$$

The elements $\Omega(\lambda)$, $\lambda \in \mathbb{P}^1$, belong to the center of S and span a 2-dimensional subspace of S_2 .

We take note that $\Omega(\lambda) = \Omega(q^{-2}\lambda^{-1})$ and $\Omega(\mu) \neq \Omega(\lambda)$ if $\mu \notin \{\lambda, q^{-2}\lambda^{-1}\}$.

Under the isomorphism $S[(KK')^{-1}]_0 \cong U_q(\mathfrak{sl}_2)$ given in (2-7), we have

$$\Omega(\lambda)(KK')^{-1} = C - \frac{q\lambda + q^{-1}\lambda^{-1}}{(q - q^{-1})^2},$$

where C is the Casimir element defined in (5-1).

The reader will notice similarities between the pencil of central subspaces $\mathbb{C}\Omega(\lambda) \subseteq S_2$ and the pencil of quadrics $Q(\lambda) \subseteq \mathbb{P}(S_1^*)$. For example, exactly one $\Omega(\lambda)$ is a product of two degree-1 elements, namely $\Omega(0) = \Omega(\infty) = KK'$, and exactly one $Q(\lambda)$ that is a union of two planes, namely $Q(0) = Q(\infty) = \{KK' = 0\}$. In a similar vein, we expect that $S/(\Omega(\lambda))$ is a prime ring if and only if λ is not 0 or ∞ . The precise relation between the $\Omega(\lambda)$'s and the $Q(\lambda)$'s is established in Proposition 5.3.

Lemma 5.2. *Let $\lambda \in \mathbb{C}^\times$. The central element $\Omega(\lambda)$ annihilates M_ℓ for all lines ℓ of the form*

$$E - \kappa s(K - \lambda K') = sF + \kappa(K - \lambda^{-1}K') = 0, \quad s \in \mathbb{P}^1.$$

Proof. Let s be any point on \mathbb{P}^1 . Since $\Omega(\lambda)$ equals

$$\begin{aligned}FE + \frac{1}{(q - q^{-1})^2}(qK - q^{-1}\lambda^{-1}K')(K - \lambda K') \\ = F(E - \kappa s(K - \lambda^{-1}K')) + \kappa(qK - q^{-1}\lambda^{-1}K')(sF + \kappa(K - \lambda K')), \end{aligned}$$

it belongs to the left ideal generated by $E - \kappa s(K - \lambda^{-1}K')$ and $sF + \kappa(K - \lambda K')$. That left ideal is $S\ell^\perp$ so, since $\Omega(\lambda)$ is in the center of S , it annihilates $S/S\ell^\perp$. \square

Proposition 5.3. *Let M_ℓ be a line module. If $\lambda \in \mathbb{C}^\times$, then $\Omega(\lambda)$ annihilates M_ℓ if and only if either*

- (1) $\ell \subseteq Q(\lambda)$ and is in the same ruling as the line $E = K - \lambda K' = 0$, or
- (2) $\ell \subseteq Q(q^{-2}\lambda^{-1})$ and is in the same ruling as the line $E = K - q^{-2}\lambda^{-1}K' = 0$.

Furthermore, $\Omega(0) = \Omega(\infty)$ annihilates M_ℓ if and only if $\ell \subseteq Q(0) = Q(\infty) = \{KK' = 0\}$.

Proof. It is easy to see that the last sentence in the statement of the proposition is true so we will assume that $\ell \not\subseteq \{KK' = 0\}$. Since M_ℓ is a line module, ℓ lies on $Q(\mu) = Q(\mu^{-1})$ for some $\mu \in \mathbb{C}^\times$. We fix such a λ . Since $\ell \not\subseteq \{KK' = 0\}$, $\mu \neq 0, \infty$.

(\Rightarrow) Fix $\lambda \in \mathbb{C}^\times$ and suppose that $\Omega(\lambda)$ annihilates M_ℓ . The lines on $Q(\mu)$ are given by (4-2) and (4-3). Replacing μ by μ^{-1} if necessary, we can assume that ℓ belongs to the same ruling on $Q(\mu)$ as $E = K - \mu K' = 0$. Hence M_ℓ is annihilated by $\Omega(\mu)$. If $\Omega(\mu) \neq \Omega(\lambda)$, then M_ℓ is annihilated by KK' . That is not the case, so $\Omega(\mu) = \Omega(\lambda)$. Hence $\mu \in \{\lambda, q^{-2}\lambda^{-1}\}$. Hence either (1) or (2) holds.

(\Leftarrow) This implication follows from Lemma 5.2. If $\ell \subseteq Q(\lambda)$ and is in the same ruling as the line $E = K - \lambda K' = 0$, then M_ℓ is annihilated by $\Omega(\lambda)$. If $\ell \subseteq Q(q^{-2}\lambda^{-1})$ and is in the same ruling as the line $E = K - q^{-2}\lambda^{-1}K' = 0$, then M_ℓ is annihilated by $\Omega(q^{-2}\lambda^{-1}) = \Omega(\lambda)$. \square

We only care about the ideal generated by $\Omega(\lambda)$ and the matter of which modules are annihilated by which $\Omega(\lambda)$'s. Thus, we only care about $\Omega(\lambda)$ up to nonzero scalar multiples. For this reason it is often better to think of $\Omega(\lambda)$ as an element in \mathbb{P}^1 .

5C. Fat points in $\text{Proj}_{\text{nc}}(S) \longleftrightarrow$ finite-dimensional simple $U_q(\mathfrak{sl}_2)$ -modules.

As the title suggests, this subsection establishes a connection between the finite-dimensional simple $U_q(\mathfrak{sl}_2)$ -modules $L(n, \pm)$ discussed in Section 5A3 and certain fat points $F(n, \pm)$ of the noncommutative scheme $\text{Proj}_{\text{nc}}(S)$ that are defined below. Proposition 5.6 makes this connection explicit. We have not addressed the question of whether the $F(n, \pm)$'s are all the fat points.

5C1. *Some finite-dimensional simple S -modules.* We fix a square root, \sqrt{q} , of q and adopt the convention that $q^{n/2-i} = (\sqrt{q})^{n-2i}$ and $q^{i-n/2} = (\sqrt{q})^{2i-n}$. Let $V(n, \pm)$ be the vector space with basis v_0, \dots, v_n and define

$$Kv_i = \sqrt{\pm 1} q^{n/2-i} v_i, \quad K'v_i = \pm \sqrt{\pm 1} q^{i-n/2} v_i,$$

$$Fv_i = \begin{cases} [n-i]v_{i+1} & \text{if } i < n, \\ 0 & \text{if } i = n, \end{cases} \quad Ev_i = \begin{cases} \pm[i]v_{i-1} & \text{if } i > 0, \\ 0 & \text{if } i = 0. \end{cases}$$

5C2. *Automorphisms of S and autoequivalences of $\text{Gr}(S)$.* Let $\theta : S \rightarrow S$ be the algebra automorphism defined by $\theta(K) = -K$, $\theta(K') = K'$, $\theta(E) = E$, $\theta(F) = F$.

If $\varepsilon \in \mathbb{C}^\times$ let $\phi_\varepsilon : S \rightarrow S$ be the algebra automorphism $\phi_\varepsilon(a) = \varepsilon^n a$ for all $a \in S_n$.

Let ϕ be a degree-preserving algebra automorphism of S . The functor $\phi^* : \text{Gr}(S) \rightarrow \text{Gr}(S)$ is defined as follows: if $M \in \text{Gr}(S)$, then $\phi^*(M)$ is M as a graded vector space and if $a \in S$ and $m \in M^*$, then $a \cdot m = \phi(a)m$. The functor ϕ^* is an autoequivalence.

Proposition 5.4. *Let $\varepsilon = -\sqrt{-1}$.*

- (1) $V(n, \pm)$ is a simple S -module of dimension $n + 1$.
- (2) $V(n, \pm)$ is a $S[(KK')^{-1}]$ -module with $(KK')^{-1}$ acting as the identity.
- (3) Identifying $U_q(\mathfrak{sl}_2)$ with $S[(KK')^{-1}]_0$ as in Proposition 2.4, $V(n, \pm) \cong L(n, \pm)$ as a $U_q(\mathfrak{sl}_2)$ -module.
- (4) $\Omega(\pm q^n)$ annihilates $V(n, \pm)$.
- (5) $V(n, -) \cong \phi_\varepsilon^* \theta^* V(n, +)$.

Proof. (1) First we check that the action makes $V(n, \pm)$ a left S -module. If $v_{-1} = 0$, then $EKv_i = \pm\sqrt{\pm 1} q^{n/2-i} [i]v_{i-1}$ and $KEv_i = \pm\sqrt{\pm 1} q^{n/2-i+1} [i]v_{i-1} = qEKv_i$. Hence $KE - qEK$ acts on $V(n, \pm)$ as 0. With the understanding that $v_{n+1} = 0$, $FKv_i = \sqrt{\pm 1} q^{n/2-i} [n-i]v_{i-1}$ and $KFv_i = \sqrt{\pm 1} q^{n/2-i-1} [n-i]v_{i+1} = q^{-1}FKv_i$, so $KF - q^{-1}FK$ acts on $V(n, \pm)$ as 0. Similar calculations show that $K'E - q^{-1}EK'$ and $K'F - qFK'$ act on $V(n, \pm)$ as 0 also. Furthermore,

$$\begin{aligned} [E, F]v_i &= \pm([n-i]Ev_{i+1} - [i]Fv_{i-1}) \\ &= \pm([n-i][i+1][i][n-i+1])v_i \\ &= \pm[n-2i]v_i \\ &= \frac{K^2 - K'^2}{q - q^{-1}}v_i, \end{aligned}$$

so $V(n, \pm)$ really is a left S -module.

To see it is simple, first observe that the v_i are eigenvectors for K with pairwise distinct eigenvalues. It follows that if $V(n, \pm)$ is not simple, then there it has a proper submodule that contains some v_i . However, looking at the actions of E and F on the v_j , a submodule that contains one v_i contains all v_i . Hence $V(n, \pm)$ is simple.

(2) Since KK' acts on $V(n, \pm)$ as multiplication by 1, the module-action of S on $V(n, \pm)$ extends to a module-action of $S[(KK')^{-1}]$.

(3) Since $e = \frac{1}{\sqrt{q}}EK^{-1}$, $f = \frac{1}{\sqrt{q}}F(K')^{-1}$, and $k = K(K')^{-1}$,

$$\begin{aligned} kv_i &= -q^{n-2i}v_i, \\ fv_i &= \begin{cases} \frac{1}{\sqrt{\pm q}}q^{n/2-i}[n-i]v_{i+1} & \text{if } i < n, \\ 0 & \text{if } i = n, \end{cases} \quad ev_i = \begin{cases} \frac{\pm 1}{\sqrt{\pm q}}q^{i-n/2}[i]v_{i-1} & \text{if } i > 0, \\ 0 & \text{if } i = 0. \end{cases} \end{aligned}$$

Choose nonzero scalars $\lambda_0, \dots, \lambda_n$ such that

$$\lambda_{i-1}/\lambda_i = \sqrt{\pm q} q^{n/2-i} [n+1-i].$$

The linear isomorphism $\phi : V(n, \pm) \rightarrow L(n, \pm)$ defined by $\phi(v_i) = \lambda_i m_i$ is a $U_q(\mathfrak{sl}_2)$ -module isomorphism because $\phi(kv_i) = k\phi(v_i)$,

$$\phi(ev_i) = \frac{\pm 1}{\sqrt{\pm q}} q^{i-n/2} [i] \lambda_{i-1} m_{i-1} = \pm [i] [n+1-i] \lambda_i m_{i-1} = e\phi(v_i),$$

and

$$\phi(fv_i) = \frac{1}{\sqrt{\pm q}} q^{n/2-i} [n-i] \lambda_{i+1} v_{i+1} = \lambda_i m_{i+1} = f\phi(v_i).$$

Hence $V(n, \pm) \cong L(n, \pm)$ as claimed.

(4) By Schur's lemma, $\Omega(\lambda)$ acts on $V(n, \pm)$ as multiplication by a scalar. Therefore, if $\Omega(\lambda)$ annihilates v_0 it annihilates $V(n, \pm)$. Since $Ev_0 = 0$, $\Omega(\lambda)v_0 = \kappa^2(qK - q^{-1}\lambda^{-1}K')(K - \lambda K')v_0$. The result follows from $(K \mp q^n K')v_0 = 0$.

(5) Let v_0, \dots, v_n be the basis for $V(n, +)$ in Section 5C1 and, to avoid confusion, write v'_i for the basis element v_i in $V(n, -)$. Thus, $Kv'_i = -\varepsilon q^{n/2-i} v'_i$.

Define $\psi : \phi_\varepsilon^* \theta^* V(n, +) \rightarrow V(n, -)$ by $\psi(v_i) := (-1)^i \varepsilon^i v'_i$. To show ψ is an S -module isomorphism it suffices to show it is an S -module homomorphism. To this end, consider v_i as an element in $\phi_\varepsilon^* \theta^* V(n, +)$. Because $\theta\phi_\varepsilon(K) = -\varepsilon K$, $Kv_i = -\varepsilon q^{n/2-i} v_i$. Hence

$$\psi(Kv_i) = \psi(-\varepsilon q^{n/2-i} v_i) = -\varepsilon q^{n/2-i} (-1)^i \varepsilon^i v'_i = (-1)^i \varepsilon^i K v'_i = K\psi(v_i).$$

Similarly, because $\theta\phi_\varepsilon(K') = \varepsilon K'$ and $K'v'_i = \varepsilon q^{i-n/2} v'_i$,

$$\psi(K'v_i) = \psi(\varepsilon q^{i-n/2} v_i) = \varepsilon q^{i-n/2} (-1)^i \varepsilon^i v'_i = (-1)^i \varepsilon^i K' v'_i = K'\psi(v_i).$$

We also have

$$\psi(Fv_i) = \psi(\varepsilon[n-i]v_{i+1}) = \varepsilon[n-i](-1)^{i+1} \varepsilon^{i+1} v'_{i+1} = (-1)^i \varepsilon^i Fv'_i = F\psi(v_i)$$

and

$$\psi(Ev_i) = \psi(\varepsilon[i]v_{i-1}) = \varepsilon[i](-1)^{i-1} \varepsilon^{i-1} v'_{i-1} = (-1)^i \varepsilon^i Ev'_i = E\psi(v_i). \quad \square$$

5C3. Fat points and fat point modules. For each $n \in \mathbb{N}$ we define

$$F(n, \pm) := V(n, \pm) \otimes \mathbb{C}[z]$$

and make this a graded left S -module according to the recipe in Lemma 2.1. It is a fat point module. Proposition 5.6 makes the statement that the fat point (module) $F(n, \pm)$ corresponds to the finite-dimensional simple $U_q(\mathfrak{sl}_2)$ -module $L(n, \pm)$ precise.

Lemma 5.5. *If θ is the automorphism in Section 5C2, then $\theta^* F(n, \pm) \cong F(n, \mp)$.*

Proof. If V is any left S -module and ϕ_ε the automorphism in Section 5C2 associated to $\varepsilon \in \mathbb{k}^\times$, then the map $\Phi : V \otimes \mathbb{k}[z] \rightarrow (\phi_\varepsilon^* V) \otimes \mathbb{k}[z]$, $\Phi(v \otimes z^i) = v \otimes (\varepsilon z)^i$, is an isomorphism in $\text{Gr}(S)$. Hence

$$\begin{aligned} F(n, -) &= \phi_\varepsilon^* \theta^* V(n, +) \otimes \mathbb{k}[z] \cong \theta^* V(n, +) \otimes \mathbb{k}[z] \\ &\cong \theta^*(V(n, +) \otimes \mathbb{k}[z]) = \theta^* F(n, +). \quad \square \end{aligned}$$

Proposition 5.6. *If $\pi^* : \text{Gr}(S) \rightarrow \text{QGr}(S)$ and $j^* : \text{QGr}(S) \rightarrow U_q(\mathfrak{sl}_2)$ are the functors in Section 1A2, then $j^* \pi^* F(n, \pm) \cong L(n, \pm)$; i.e., there is an isomorphism of $U_q(\mathfrak{sl}_2)$ -modules*

$$F(n, \pm)[(KK')^{-1}]_0 \cong L(n, \pm).$$

Proof. The functor $j^* \pi^*$ sends $M \in \text{Gr}(S)$ to $M[(KK')^{-1}]_0$, where the latter is made into a $U_q(\mathfrak{sl}_2)$ -module via the isomorphism $U_q(\mathfrak{sl}_2) \rightarrow S[(KK')^{-1}]_0$ in Proposition 2.4.

Since KK' acts on $V(n, \pm)$ as the identity, it acts on $F(n, \pm) = V(n, \pm) \otimes k[z]$ as multiplication by z^2 . Hence, $F(n, \pm)[(KK')^{-1}]_0 = V(n, \pm) \otimes k[z, z^{-2}]_0 = V(n, \pm) \otimes 1$.

Let $\hat{S} = S[(KK')^{-1}]$. Applying the functor $\hat{S} \otimes_S -$ to the surjective S -module homomorphism $F(n, \pm) \rightarrow V(n, \pm)$, $v \otimes z^i \mapsto v$, produces a surjective homomorphism

$$\psi : F[(KK')^{-1}] = \hat{S} \otimes_S F(n, \pm) \rightarrow \hat{S} \otimes_S V(n, \pm)$$

of \hat{S} -modules. Of course, ψ is a homomorphism of \hat{S}_0 -modules. Every homogeneous component of $F(n, \pm)[(KK')^{-1}]$ is an \hat{S}_0 -submodule of $F(n, \pm)[(KK')^{-1}]$ so ψ restricts to a homomorphism $F(n, \pm)[(KK')^{-1}]_0 \rightarrow \hat{S} \otimes_S V(n, \pm)$ of \hat{S}_0 -modules. But $\hat{S} \otimes_S V(n, \pm)$ is isomorphic to $L(n, \pm)$ as an \hat{S}_0 -module by Proposition 5.4(3) and, by the previous paragraph, $\dim(F(n, \pm)[(KK')^{-1}]_0) = \dim(V(n, \pm)) = n + 1 = \dim(L(n, \pm))$ so the restriction of ψ to $F(n, \pm)[(KK')^{-1}]_0$ is an isomorphism of \hat{S}_0 -modules. \square

Proposition 5.7. *Let $n \geq 0$. Let ℓ_\pm be any line on $Q(\pm q^n)$ that is in the same ruling as the line $E = K \mp q^n K' = 0$.*

- (1) *There is a surjective S -module homomorphism $M_{\ell_\pm} \rightarrow V(n, \pm)$.*
- (2) *There is a homomorphism $M_{\ell_\pm} \rightarrow F(n, \pm)$ in $\text{Gr}(S)$ that becomes an epimorphism in $\text{QGr}(S)$.*
- (3) *In $\text{Proj}_{\text{nc}}(S)$, the fat point $F(n, \pm)$ lies on the line ℓ_\pm .*

Proof. Let $s \in \mathbb{P}^1$ be such that ℓ_\pm is the line

$$\kappa(K \mp q^n K') - s^{-1} E = \kappa(K \mp q^{-n} K') + s F = 0.$$

Thus,

$$M_{\ell_{\pm}} \cong \frac{S}{SX_{\pm} + SY_{\pm}},$$

where $X_{\pm} = \kappa(K \mp q^{-n}K') - s^{-1}E$ and $Y_{\pm} = \kappa(K \mp q^n K') + sF$.

(1) Since $V(n, \pm)$ is a simple S -module it suffices to show there is a nonzero homomorphism $M_{\ell_{\pm}} \rightarrow V(n, \pm)$. For this, it suffices to show there is a nonzero element in $V(n, \pm)$ annihilated by both X_{\pm} and Y_{\pm} .

If $s = 0$, and $v_{\pm} = v_0 \in V(n, \pm)$, then $X_{\pm}v_{\pm} = Ev_0 = 0$ and $Y_{\pm}v_{\pm} = (K \mp q^n K')v_{\pm} = 0$. If $s = \infty$ and $v_{\pm} = v_n \in V(n, \pm)$, then $X_{\pm}v_{\pm} = (K \mp q^{-n}K')v_n = 0$ and $Y_{\pm}v_{\pm} = Fv_n = 0$. Thus, (1) is true if s equals 0 or ∞ .

From now on, assume that $s \neq 0, \infty$. Let $\lambda_0, \dots, \lambda_n \in \mathbb{k}^{\times}$ be such that

$$\lambda_{i+1}/\lambda_i = \pm\sqrt{\pm 1} \frac{[n-i]}{[i+1]} sq^{-n/2}$$

for all i . If

$$v_{\pm} = \sum_{i=0}^n \lambda_i v_i \in V(n, \pm),$$

then

$$\begin{aligned} X_{\pm}v_{\pm} &= \sum_{i=0}^n (\kappa\sqrt{\pm 1} (q^{n/2-i} \mp q^{-n}q^{i-n/2})\lambda_i v_i - s^{-1}(\pm 1)[i]\lambda_i v_{i-1}) \\ &= \sum_{i=0}^n (-q^{-n/2}\sqrt{\pm 1}[n-i]\lambda_i \mp s^{-1}[i+1]\lambda_{i+1})v_i \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} Y_{\pm}v_{\pm} &= \sum_{i=0}^n (\kappa\sqrt{\pm 1} (q^{n/2-i} \mp q^n q^{i-n/2})\lambda_i v_i + s[n-i]\lambda_i v_{i+1}) \\ &= \sum_{i=0}^n (-q^{n/2}\sqrt{\pm 1}[i]\lambda_i + s[n-i+1]\lambda_{i-1})v_i \\ &= 0. \end{aligned}$$

(2) By Lemma 2.1, the existence of a nonzero homomorphism $M_{\ell_{\pm}} \rightarrow V(n, \pm)$ implies the existence of a nonzero homomorphism $M_{\ell_{\pm}} \rightarrow F(n, \pm)$ in $\text{Gr}(S)$. However, as an object in $\text{QGr}(S)$, $F(n, \pm)$ is irreducible so (2) follows.

(3) This is just terminology. □

If one of the lines $\ell_{\pm} = \{X_{\pm} = Y_{\pm} = 0\}$ in Proposition 5.7 meets C at $(\xi_1, \xi_2, \xi_3, 0)$, then it meets C' at $(q^{-n}\xi_1, q^n\xi_2, 0, \pm\xi_3)$. Combining this with Theorem 4.2(5) gives the following result.

Corollary 5.8. *Let $p = (\xi_1, \xi_2, \xi_3, 0) \in C$ and define $p_{\pm} = (\xi_1, \xi_2, 0, \pm\xi_3)$. Let ℓ_{\pm} be the secant line to $C \cup C'$ passing through p and $\sigma_S^n(p_{\pm})$. There is a surjective homomorphism $M_{\ell_{\pm}} \rightarrow V(n, \pm)$ in $\text{Mod}(S)$ and an epimorphism $M_{\ell_{\pm}} \rightarrow F(n, \pm)$ in $\text{QGr}(S)$.*

The analogue of (5-2) requires results from the next section, and can be found in Theorem 6.2.

6. Relation to the nondegenerate Sklyanin algebras

We remind the reader that $S(\alpha, \beta, \gamma)$ denotes one of the nondegenerate Sklyanin algebras defined in (2-2).

In this section, we show that some of our results about S can be obtained as “degenerations” of results in [Smith and Stafford 1992; Chirvasitu and Smith 2017; Smith and Staniszkis 1993] about $S(\alpha, \beta, \gamma)$. We also complete the characterization of those line modules that surject onto fat point modules that we alluded to in the last section.

6A. The point scheme of a nondegenerate Sklyanin algebra. We follow [Smith and Stafford 1992]. The point scheme of $S(\alpha, \beta, \gamma)$ embedded in \mathbb{P}^3 with coordinates x_0, x_1, x_2, x_3 is

$$(6-1) \quad E' = E \cup \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)\},$$

where E is the elliptic curve defined by

$$(6-2) \quad x_0^2 + x_1^2 + x_2^2 + x_3^2 = 0 = x_3^2 + \frac{1-\gamma}{1+\alpha}x_1^2 + \frac{1+\gamma}{1-\beta}x_2^2.$$

Equivalently, E is the intersection of any two of the following quadrics:

$$(6-3) \quad \begin{aligned} x_0^2 + x_1^2 + x_2^2 + x_3^2 &= 0, \\ x_0^2 - \beta\gamma x_1^2 - \gamma x_2^2 + \beta x_3^2 &= 0, \\ x_0^2 + \gamma x_1^2 - \alpha\gamma x_2^2 - \alpha x_3^2 &= 0, \\ x_0^2 - \beta x_1^2 + \alpha x_2^2 - \alpha x_3^2 &= 0. \end{aligned}$$

There is an automorphism σ of E' that fixes the four isolated points and on E is given by the formula

$$(6-4) \quad \sigma : \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} -2\alpha\beta\gamma & x_1x_2x_3 & - & x_0(-x_0^2 + \beta\gamma x_1^2 + \alpha\gamma x_2^2 + \alpha\beta x_3^2) \\ 2\alpha & x_0x_2x_3 & + & x_1(x_0^2 - \beta\gamma x_1^2 + \alpha\gamma x_2^2 + \alpha\beta x_3^2) \\ 2\beta & x_0x_1x_3 & + & x_2(x_0^2 + \beta\gamma x_1^2 - \alpha\gamma x_2^2 + \alpha\beta x_3^2) \\ 2\gamma & x_0x_1x_2 & + & x_3(x_0^2 + \beta\gamma x_1^2 + \alpha\gamma x_2^2 - \alpha\beta x_3^2) \end{pmatrix}.$$

6B. Degenerate point scheme. In the degenerate case, substituting $(\alpha, \beta, \gamma) = (0, b^2, -b^2)$ into equations (6-1) through (6-4) yields the following results.

We will compare the point scheme of $S = S(0, b^2, -b^2)$ to

$$(6-5) \quad E'_{\text{deg}} := E_{\text{deg}} \cup \{(1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1)\},$$

where the curve E_{deg} is defined by

$$(6-6) \quad x_0^2 + x_1^2 + x_2^2 + x_3^2 = 0 = x_3^2 + (1 + b^2)x_1^2 + x_2^2,$$

or as the intersection of any two of the quadrics

$$(6-7) \quad \begin{aligned} x_0^2 + x_1^2 + x_2^2 + x_3^2 &= 0, \\ x_0^2 + b^4 x_1^2 + b^2 x_2^2 + b^2 x_3^2 &= 0, \\ x_0^2 - b^2 x_1^2 &= 0, \\ x_0^2 - b^2 x_1^2 &= 0. \end{aligned}$$

The automorphism on E'_{deg} fixes the four isolated points and is defined on E_{deg} by

$$(6-8) \quad \sigma_{\text{deg}} : \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_0(x_0^2 + b^4 x_1^2) \\ x_1(x_0^2 + b^4 x_1^2) \\ 2b^2 x_0 x_1 x_3 + x_2(x_0^2 - b^4 x_1^2) \\ -2b^2 x_0 x_1 x_2 + x_3(x_0^2 - b^4 x_1^2) \end{pmatrix}$$

6C. Comparison with our results. We now compare $(E'_{\text{deg}}, \sigma_{\text{deg}})$ with $(\mathcal{P}_S, \sigma_S)$ from Theorem 4.2. Recall our definitions of E, F, K, K' from (2-5):

$$(6-9) \quad \begin{aligned} E &= \frac{i}{2}(1 - ib)(x_2 + ix_3), & F &= \frac{i}{2}(1 + ib)(x_2 - ix_3), \\ K &= x_0 + bx_1, & K' &= x_0 - bx_1. \end{aligned}$$

With respect to the homogeneous coordinates E, F, K , and K' ,

$$\mathcal{P}_S = C \cup C' \cup L \cup \{(0, 0, 1, \pm 1)\},$$

where C, C' and L are given by

$$\begin{aligned} C' : EF + \kappa^2 K'^2 &= K = 0, \\ C : EF + \kappa^2 K^2 &= K' = 0, \\ L : K = K' &= 0. \end{aligned}$$

The conics C and C' lie on the planes $K' = 0$ and $K = 0$, respectively, and the line L is the intersection of those two planes. With respect to the homogeneous coordinates E, F, K , and K' , (6-5) becomes

$$E'_{\text{deg}} = E_{\text{deg}} \cup \{(0, 0, 1, 1), (0, 0, 1, -1), (q, 1, 0, 0), (-q, 1, 0, 0)\}.$$

The isolated points $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$ in (6-5) remain isolated after degeneration, but the points $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$ in (6-5), which are $(q, 1, 0, 0)$ and $(-q, 1, 0, 0)$ in the E, F, K, K' coordinates, become points on the line L in \mathcal{P}_S after degeneration.

Next, we compare E_{deg} with $C \cup C' \cup L$. The equation (6-6) yields

$$x_0^2 - b^2 x_1^2 = (x_0 - bx_1)(x_0 + bx_1) = KK' = 0.$$

Hence $E_{\text{deg}} \subseteq \{K = 0\} \cup \{K' = 0\}$.

On the plane $K' = 0$, $x_0 = bx_1$ so both sides of (6-6) for E_{deg} become

$$(1 + b^2)x_1^2 + x_2^2 + x_3^2 = 0.$$

On the other hand, C' is given by

$$\begin{aligned} 0 = EF + \kappa^2 K'^2 &= -\frac{1}{4}(1 + b^2)(x_2^2 + x_3^2) + 4\kappa^2 b^2 x_1^2 \\ &= -\frac{1}{4}(1 + b^2)(x_2^2 + x_3^2) - \frac{1}{4}(1 + b^2)^2 x_1^2 \\ &= -\frac{1}{4}(1 + b^2)(x_2^2 + x_3^2 + (1 + b^2)x_1^2). \end{aligned}$$

Hence $E_{\text{deg}} \cap \{K' = 0\} = C'$. A similar calculation yields the analogous result for the plane $K = 0$. We thus conclude that

$$E_{\text{deg}} = C \cup C'.$$

Finally, we compare σ_{deg} and σ_S . On the plane $K = 0$,

$$\sigma_{\text{deg}} : \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} (b^2 + b^4)x_1^2 x_0 \\ (b^2 + b^4)x_1^3 \\ 2b^4 x_1^2 x_3 + (b^2 - b^4)x_1^2 x_2 \\ -2b^4 x_1^2 x_2 + (b^2 - b^4)x_1^2 x_3 \end{pmatrix} = \begin{pmatrix} (1 + b^2)x_0 \\ (1 + b^2)x_1 \\ (1 - b^2)x_2 + 2b^2 x_3 \\ (1 - b^2)x_3 - 2b^2 x_2 \end{pmatrix}.$$

Changing coordinates,

$$\sigma_{\text{deg}} : \begin{pmatrix} x_2 + ix_3 \\ x_2 - ix_3 \\ x_0 + bx_1 \\ 0 \end{pmatrix} \mapsto \begin{pmatrix} (1 - ib)^2(x_2 + ix_3) \\ (1 + ib)^2(x_2 - ix_3) \\ (1 + b^2)(x_0 + bx_1) \\ 0 \end{pmatrix} = \begin{pmatrix} q(x_2 + ix_3) \\ q^{-1}(x_2 - ix_3) \\ (x_0 + bx_1) \\ 0 \end{pmatrix}.$$

Therefore, in the E, F, K, K' coordinates, $\sigma_{\text{deg}}(\xi_1, \xi_2, \xi_3, 0) = (q\xi_1, q^{-1}\xi_2, \xi_3, 0) = \sigma_S(\xi_1, \xi_2, \xi_3, 0)$. Similar calculations on the plane $K' = 0$ and on the isolated points yield $\sigma_{\text{deg}} = \sigma_S$.

6D. Degenerations of Heisenberg automorphisms. Recall (e.g., from [Chirvasitu and Smith 2017, Proposition 2.6]) that the Heisenberg group of order 4^3 acts on the Sklyanin algebra $S(\alpha, \beta, \gamma)$ as follows.

	x_0	x_1	x_2	x_3
ϕ_1	bcx_1	$-ix_0$	$-ibx_3$	$-cx_2$
ϕ_2	acx_2	$-ax_3$	$-ix_0$	$-icx_1$
ϕ_3	abx_3	$-iax_2$	$-bx_1$	$-ix_0$

Table 2. Automorphisms of $S(\alpha, \beta, \gamma)$.

	E	F	K	K'
ϕ_1	bqF	$-bq^{-1}E$	$-ibK'$	ibK
ϕ_2	$\frac{1}{2}(1-ib)K'$	$\frac{1}{2}(1+ib)K$	0	0
ϕ_3	$\frac{i}{2}(1-ib)K'$	$-\frac{i}{2}(1+ib)K$	0	0

Table 3. Endomorphisms of $S(0, \beta, -\beta)$.

First, fix square roots a , b and c of α , β and γ respectively. We define automorphisms ϕ_i of $S(\alpha, \beta, \gamma)$ via Table 2.

Now fix $v_1, v_2, v_3 \in k^\times$ such that $av_1^2 = bv_2^2 = cv_3^2 = -iabc$, and define $\varepsilon_1 = v_1^{-1}\phi_1$, $\varepsilon_2 = v_2^{-1}\phi_2$, $\varepsilon_3 = v_3^{-1}\phi_3$, and $\delta = i$. The subgroup $\langle \varepsilon_1, \varepsilon_2, \varepsilon_3, \delta \rangle \subseteq \text{Aut}(S)$ is isomorphic to the Heisenberg group of order 4^3 , defined by generators and relations as

$$H_4 := \langle \varepsilon_1, \varepsilon_2, \delta \mid \varepsilon_1^4 = \varepsilon_2^4 = \delta^4 = 1, \delta\varepsilon_1 = \varepsilon_1\delta, \varepsilon_2\delta = \delta\varepsilon_2, \varepsilon_1\varepsilon_2 = \delta\varepsilon_2\varepsilon_1 \rangle.$$

The algebras we are considering are of the form $S(0, \beta, -\beta)$. We define $c = ib$. The map ϕ_1 still extends to an algebra automorphism but ϕ_2 and ϕ_3 degenerate to endomorphisms. In terms of E , F , K , and K' , the endomorphisms ϕ_i act as in Table 3.

Although ϕ_2 and ϕ_3 are not isomorphisms, there are associated endofunctors ϕ_2^* and ϕ_3^* of $\text{Gr}(S)$. The application of ϕ_2^* and ϕ_3^* to the point modules $M_{(0,0,1,\pm 1)} \in \text{Gr}(S)$ produces point modules. Indeed,

$$\phi_2(S) = \phi_3(S) = \mathbb{C}[K, K'] \subseteq S,$$

and the two point modules referred to above are cyclic $\mathbb{C}[K, K']$ -modules. With this in hand, the next result describes how the ϕ_i act on the four $S(0, \beta, -\beta)$ -points obtained by degeneration from $S(\alpha, \beta, \gamma)$. The proof is a direct application of the formulas in Table 2 above.

Proposition 6.1. *The endomorphisms ϕ_i of S move the four special point modules of S as follows.*

- (1) ϕ_1^* interchanges $M_{(0,0,1,\pm 1)}$ and interchanges $M_{(\pm q, 1, 0, 0)}$;

$$(2) \phi_2^* M_{(0,0,1,\pm 1)} \cong M_{(\pm q,1,0,0)};$$

$$(3) \phi_3^* M_{(0,0,1,\pm 1)} \cong M_{(\mp q,1,0,0)}.$$

6E. Degenerations of fat point–line incidences. In this section we describe resolutions of fat points by line modules by degenerating the analogous statements in [Smith and Staniszkis 1993] for the algebras $S(\alpha, \beta, \gamma)$.

If $\ell \subset \mathbb{P}^3$ is the line passing through $p, p' \in C \cup C'$, we will sometimes denote the line module M_ℓ by $M_{p,p'}$ for clarity. We will also use the following notation: if $p = (\xi_1, \xi_2, \xi_3, 0) \in C$, then $p_\pm = (\xi_1, \xi_2, 0, \pm \xi_3) \in C'$ is the point for which M_{p,p_\pm} surjects onto $F(0, \pm)$. Similarly, in order to keep the notation symmetric, if $p \in C'$ then p_\pm is the point on C for which M_{p,p_\pm} surjects onto $F(0, \pm)$.

Finally, we denote by $\sigma = \sigma_S : (C \cup C')^2 \rightarrow (C \cup C')^2$ the diagonal action of $\sigma = \sigma_S$ on $(C \cup C')^2$, and by ψ the automorphism

$$\psi := (\text{id}, \sigma) : (C \cup C')^2 \rightarrow (C \cup C')^2.$$

By a slight abuse of notation, we use the same symbols to refer to the induced automorphisms on the variety of lines through pairs of points on $C \cup C'$.

Theorem 6.2. *Let n be a nonnegative integer, ℓ_\pm a line through $p, p_\pm \in C \cup C'$, and $\ell_{\pm n}$ the line $\psi^n(\ell_\pm)$. In $\text{QGr}(S)$, there is an exact sequence*

$$(6-10) \quad 0 \rightarrow M_{\sigma^{-(n+1)}(\ell_{\pm n})}(-n-1) \rightarrow M_{\ell_\pm} \rightarrow F(n, \pm) \rightarrow 0.$$

Proof. We will prove this for ℓ_+ . To that end, let ℓ be the line through p and p_+ .

The relation $\{(p, p_\pm)\}$ on $C \cup C'$ is the fiber over $(0, b^2, -b^2)$ of a family of relations over the space of parameters (α, β, γ) for the Sklyanin algebras. Specifically, let us write

$$(x_0, x_1, x_2, x_3) \mapsto (-x_0, x_1, x_2, x_3)$$

for the $-$ maps on the elliptic curves $E = E(\alpha, \beta, \gamma)$ and let

$$(x_0, x_1, x_2, x_3) \mapsto (x_0, x_1, -x_2, -x_3)$$

be addition by the 2-torsion point $\omega \in E$.

Claim. $\{(p, p_+)\}$ is the limit of the graphs of the maps $p \mapsto \omega - p$.

Proof of claim. In terms of the x_i coordinates, the map $p \mapsto \omega - p$ amounts to changing the sign of x_0 . On the other hand, the discussion at the beginning of Section 6E shows that in (E, F, K, K') -coordinates the map $p \mapsto p_+$ simply interchanges K and K' . Since $C \cup C'$ is the degeneration of the family (6-2) of elliptic curves, the truth of the claim follows from the coordinate change formulas (6-9). \square

The claim implies that the resolutions

$$0 \rightarrow M_{\sigma(p), \sigma(\omega-p)}(-1) \rightarrow M_{p, \omega-p} \rightarrow \bullet \rightarrow 0$$

of the point modules associated to $(x_0, x_1, x_2, x_3) = (1, 0, 0, 0)$ (e.g., from [Levasseur and Smith 1993, Theorem 5.7]) degenerate to (6-10) for $n = 0$ in the $+$ case.

Similarly, for larger n we have, in the nondegenerate case, resolutions

$$0 \rightarrow M_{\sigma^{-(n+1)}(p), \sigma^{-1}(\omega-p)}(-n-1) \rightarrow M_{p, \sigma^n(\omega-p)} \rightarrow \bullet \rightarrow 0$$

of 1-critical fat points of multiplicity $n+1$ as explained in [Smith and Staniszkis 1993, Proposition 4.4(b)]. These degenerate to a resolution of the form (6-10) of a certain fat S -point module of multiplicity $n+1$ (denoted momentarily by the same symbol \bullet):

$$(6-11) \quad 0 \rightarrow M_{\sigma^{-(n+1)}(\ell_n)}(-n-1) \rightarrow M_{\ell_n} \rightarrow \bullet \rightarrow 0,$$

where ℓ is the line through p and p_+ and $\ell_n = \psi^n(\ell)$; note that \bullet is the same fat point (up to isomorphism in $\text{QGr}(S)$) for all choices of p .

Finally, to argue that $\bullet \cong F(n, +)$ in the present case, simply specialize to the line ℓ for which (6-11) is the homogenized version of the standard BGG resolution (5-2) of the simple $U_q(\mathfrak{sl}_2)$ -module $L(n, +)$.

There is a similar argument for $F(n, -)$, or one can use the observation in Lemma 5.5 that $F(n, -) \cong \theta^* F(n, +)$. \square

Remark 6.3. Incidentally, one can give a proof of Proposition 4.8 in the same spirit as that of Theorem 6.2 by degenerating the exact sequences

$$0 \rightarrow M_{\sigma p, \sigma^{-1} p'}(-1) \rightarrow M_{p, p'} \rightarrow M_p \rightarrow 0$$

from [Levasseur and Smith 1993, Theorem 5.5] for the Sklyanin algebras $S(\alpha, \beta, \gamma)$, where p, p' belong to the elliptic curve component $E = E(\alpha, \beta, \gamma)$ of the point scheme of $S(\alpha, \beta, \gamma)$ and σ is the translation automorphism of E . The result then follows from the observation made above that $E(\alpha, \beta, \gamma)$, together with its translation automorphism, degenerates to $C \cup C'$ equipped with our automorphism (also denoted by σ throughout) when $\alpha \rightarrow 0$. \blacklozenge

The next result completes the description of the fat point–line incidences.

Proposition 6.4. *For $n \geq 0$ the line modules M_{ℓ_n} from Theorem 6.2 are the only ones having $F(n, \pm)$ as a quotient in $\text{QGr}(S)$.*

Proof. The only central element $\Omega(\lambda)$ annihilating $F(n, \pm)$ is $\Omega(\pm q^n)$. In turn, Proposition 5.3 tells us that the only line modules annihilated by $\Omega(\pm q^n)$ are the lines M_{ℓ_n} in question and the lines $M_{\sigma^{-(n+1)}(\ell_n)}$ appearing as the leftmost terms in (6-10). In conclusion, it suffices to show that there are no surjections

$$(6-12) \quad M_{\sigma^{-(n+1)}(p), \sigma^{-1}(p_{\pm})} \rightarrow F(n, \pm)$$

in $\text{QGr}(S)$.

Let us specialize to $F(n, +)$, to fix notation. Upon localizing to $S[(KK')^{-1}]_0 \cong U = U_q(\mathfrak{sl}(2))$, (6-12) becomes a surjection

$$(6-13) \quad \frac{U}{UX + UY} \rightarrow L(n, +),$$

where $X = \kappa(1 - q^{n+2}k^{-1}) - s^{-1}q^{-1/2}e$ and $Y = \kappa(k - q^{-(n+2)}) + sq^{-1/2}f$ for some $s \in \mathbb{P}^1$. If $s = 0$ or ∞ then the left-hand side of (6-13) is the simple Verma module of highest weight $q^{-(n+2)}$ (respectively lowest weight q^{n+2}), thus contradicting the existence of such a surjection. On the other hand, if $s \in \mathbb{C}^\times$, then we obtain surjections (6-13) for *all* $s \in \mathbb{C}^\times$ by applying the \mathbb{G}_m -action on U given by

$$k \mapsto k, \quad e \mapsto s^{-1}e, \quad f \mapsto sf \quad \text{for } s \in \mathbb{C}^\times.$$

By continuity in $s \in \mathbb{P}^1$, we then get such surjections for $s = 0, \infty$ as well, and the previous argument applies. \square

We end with the following remark on certain modules over $U = U_q(\mathfrak{sl}_2)$. In the proof of Proposition 6.4 we showed that the modules (6-13) of the form $U/(UX_\pm + UY_\pm)$ do not surject onto the simple modules $L(n, \pm)$ for

$$(6-14) \quad \begin{aligned} X_\pm &= \kappa(1 \mp q^{n+2}k^{-1}) - s^{-1}q^{-1/2}e, \\ Y_\pm &= \kappa(k \mp q^{-(n+2)}) + sq^{-1/2}f, \end{aligned}$$

where $s \in \mathbb{P}^1$. In fact, we can do somewhat more:

Proposition 6.5. *For X_\pm and Y_\pm as in (6-14) the module $U/(UX_\pm + UY_\pm)$ is simple.*

Proof. As in the proof of Proposition 6.4, we focus on $X = X_+$ and $Y = Y_+$ to fix notation.

Assume otherwise. Then, using the equivalence between the category of modules over $U \cong S[(KK')^{-1}]_0$ and a full subcategory of $\text{QGr}(S)$, this assumption implies that the line module

$$M = M_{\sigma^{-(n+1)(p)}, \sigma^{-1}(p_\pm)}$$

from (6-12) has a nonobvious subobject in $\text{QGr}(S)$. The criticality of line modules then implies that such a subobject would be a shifted line module, and hence there would be a surjection from M to a nonzero fat point. Localizing back to U this would give a surjection of $U/(UX + UY)$ onto a nonzero finite-dimensional U -module, which would be a contradiction as in the proof of Proposition 6.4. \square

The significance of Proposition 6.5 is that it fits the simple Verma modules of highest and lowest weights $q^{-(n+2)}$ and respectively q^{n+2} into “continuous” \mathbb{P}^1 -families of simple modules.

Acknowledgements

We thank Richard Chandler for sharing his results with us. We also thank him and Michaela Vancliff for useful conversations about their work and ours.

We are also grateful to an anonymous referee whose careful reading and comments improved an earlier version of this paper.

Chirvasitu was partially supported by NSF grant DMS-1565226.

References

- [Artin et al. 1990] M. Artin, J. Tate, and M. Van den Bergh, “Some algebras associated to automorphisms of elliptic curves”, pp. 33–85 in *The Grothendieck Festschrift, I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR Zbl
- [Artin et al. 1991] M. Artin, J. Tate, and M. Van den Bergh, “Modules over regular algebras of dimension 3”, *Invent. Math.* **106**:2 (1991), 335–388. MR Zbl
- [Brown and Goodearl 2002] K. A. Brown and K. R. Goodearl, *Lectures on algebraic quantum groups*, Birkhäuser, Basel, 2002. MR Zbl
- [Chandler 2016] R. G. Chandler, Jr., *On the quantum spaces of some quadratic regular algebras of global dimension four*, Ph.D. thesis, University of Texas Arlington, 2016, available at <https://search.proquest.com/docview/1859920925>.
- [Chirvasitu and Smith 2017] A. Chirvasitu and S. P. Smith, “Exotic elliptic algebras of dimension 4”, *Adv. Math.* **309** (2017), 558–623. MR Zbl
- [Jantzen 1996] J. C. Jantzen, *Lectures on quantum groups*, Graduate Studies in Mathematics **6**, American Mathematical Society, Providence, RI, 1996. MR Zbl
- [Jimbo 1985] M. Jimbo, “A q -difference analogue of $U(\mathfrak{g})$ and the Yang–Baxter equation”, *Lett. Math. Phys.* **10**:1 (1985), 63–69. MR Zbl
- [Kassel 1995] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, New York, 1995. MR Zbl
- [Klimyk and Schmüdgen 1997] A. Klimyk and K. Schmüdgen, *Quantum groups and their representations*, Springer, Berlin, 1997. MR Zbl
- [Le Bruyn and Smith 1993] L. Le Bruyn and S. P. Smith, “Homogenized $\mathfrak{sl}(2)$ ”, *Proc. Amer. Math. Soc.* **118**:3 (1993), 725–730. MR Zbl
- [Le Bruyn et al. 1996] L. Le Bruyn, S. P. Smith, and M. Van den Bergh, “Central extensions of three-dimensional Artin–Schelter regular algebras”, *Math. Z.* **222**:2 (1996), 171–212. MR Zbl
- [Levasseur and Smith 1993] T. Levasseur and S. P. Smith, “Modules over the 4-dimensional Sklyanin algebra”, *Bull. Soc. Math. France* **121**:1 (1993), 35–90. MR Zbl
- [Lusztig 1988] G. Lusztig, “Quantum deformations of certain simple modules over enveloping algebras”, *Adv. in Math.* **70**:2 (1988), 237–249. MR Zbl
- [Lusztig 1990] G. Lusztig, “On quantum groups”, *J. Algebra* **131**:2 (1990), 466–475. MR Zbl
- [Shelton and Vancliff 2002a] B. Shelton and M. Vancliff, “Schemes of line modules, I”, *J. London Math. Soc.* (2) **65**:3 (2002), 575–590. MR Zbl
- [Shelton and Vancliff 2002b] B. Shelton and M. Vancliff, “Schemes of line modules, II”, *Comm. Algebra* **30**:5 (2002), 2535–2552. MR Zbl
- [Smith 1994] S. P. Smith, “The four-dimensional Sklyanin algebras”, *K-Theory* **8**:1 (1994), 65–80. MR Zbl

- [Smith and Stafford 1992] S. P. Smith and J. T. Stafford, “Regularity of the four-dimensional Sklyanin algebra”, *Compositio Math.* **83**:3 (1992), 259–289. MR Zbl
- [Smith and Staniszkis 1993] S. P. Smith and J. M. Staniszkis, “Irreducible representations of the 4-dimensional Sklyanin algebra at points of infinite order”, *J. Algebra* **160**:1 (1993), 57–86. MR Zbl
- [Smith and Van den Bergh 2013] S. P. Smith and M. Van den Bergh, “Noncommutative quadric surfaces”, *J. Noncommut. Geom.* **7**:3 (2013), 817–856. MR Zbl
- [Van den Bergh 2001] M. Van den Bergh, *Blowing up of non-commutative smooth surfaces*, Mem. Amer. Math. Soc. **734**, 2001. MR Zbl
- [Zhang 1996] J. J. Zhang, “Twisted graded algebras and equivalences of graded categories”, *Proc. London Math. Soc.* (3) **72**:2 (1996), 281–311. MR Zbl

Received December 12, 2016. Revised July 14, 2017.

ALEX CHIRVASITU

Current address:

DEPARTMENT OF MATHEMATICS

UNIVERSITY AT BUFFALO

BUFFALO, NY

UNITED STATES

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF WASHINGTON

SEATTLE, WA

UNITED STATES

achirvas@buffalo.edu

S. PAUL SMITH

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF WASHINGTON

SEATTLE, WA

UNITED STATES

smith@math.washington.edu

LIANG ZE WONG

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF WASHINGTON

SEATTLE, WA

UNITED STATES

wonglz@uw.edu

A GENERALIZATION OF “EXISTENCE AND BEHAVIOR OF THE RADIAL LIMITS OF A BOUNDED CAPILLARY SURFACE AT A CORNER”

JULIE N. CRENSHAW, ALEXANDRA K. ECHART AND KIRK E. LANCASTER

The principal existence theorem (i.e., Theorem 1) of “Existence and behavior of the radial limits of a bounded capillary surface at a corner” (*Pacific J. Math.* 176:1 (1996), 165–194) is extended to the case of a contact angle γ which is not bounded away from 0 and π (and depends on position in a bounded domain $\Omega \in \mathbb{R}^2$ with a convex corner at $\mathcal{O} = (0, 0)$). The lower bound on the size of “side fans” (i.e., Theorem 2 in the above paper) is extended to the case of such contact angles for convex and nonconvex corners.

1. Introduction and theorems

Consider the capillary problem

$$(1) \quad Nf = \kappa f + \lambda \quad \text{in } \Omega,$$

$$(2) \quad Tf \cdot \mathbf{v} = \cos \gamma \quad \text{on } \partial\Omega,$$

where Ω is a region in \mathbb{R}^2 with a corner at \mathcal{O} , $\mathcal{O} \in \partial\Omega$, $Nf = \nabla \cdot Tf$, $Tf = \nabla f / \sqrt{1 + |\nabla f|^2}$, κ and λ are constants, \mathbf{v} is the exterior unit normal on $\partial\Omega$, and $\gamma = \gamma(s)$ is a function of position on $\partial\Omega$, $0 \leq \gamma(s) \leq \pi$. The surface $z = f(x, y)$ describes the shape of the static liquid–gas interface in a vertical cylindrical tube of cross-section Ω ; see [Finn 1986; Lancaster and Siegel 1996] for background.

We are interested in the behavior of solutions to (1) and (2) in a neighborhood of a corner point of the boundary. We take the corner point to be $\mathcal{O} = (0, 0)$. Let $\Omega^* = \Omega \cap B_{\delta^*}(\mathcal{O})$, where $B_{\delta^*}(\mathcal{O})$ is the ball of radius δ^* about \mathcal{O} . Polar coordinates relative to \mathcal{O} will be denoted by r and θ . We assume that $\partial\Omega$ is piecewise smooth and that $\partial\Omega \cap B_{\delta^*}(\mathcal{O})$ consists of two arcs $\partial^+\Omega^*$ and $\partial^-\Omega^*$, whose tangent lines approach the lines $L^+ : \theta = \alpha$ and $L^- : \theta = -\alpha$, respectively, as the point \mathcal{O} is approached. The points where $\partial B_{\delta^*}(\mathcal{O})$ intersect $\partial\Omega$ are labeled A and B ; also, $\Gamma^* = \partial B_{\delta^*}(\mathcal{O}) \cap \overline{\Omega}$. Set

$$\Omega_\infty = \{(r \cos(\theta), r \sin(\theta)) : r > 0, -\alpha < \theta < \alpha\}.$$

MSC2010: primary 35J93, 58E12; secondary 35J60, 53A10.

Keywords: radial limits, capillary surfaces, corners, discontinuities.

Let $(x^+(s), y^+(s))$ be an arclength parametrization of $\partial^+\Omega^*$ and $(x^-(s), y^-(s))$ be an arclength parametrization of $\partial^-\Omega^*$, each measured from the corner at \mathcal{O} , so that $(x^\pm(0), y^\pm(0)) = (0, 0)$. Let $(x_*^+(s), y_*^+(s))$ be an arclength parametrization of $\partial^+\Omega_\infty = \{(r \cos(\alpha), r \sin(\alpha)) : r \geq 0\}$ and $(x_*^-(s), y_*^-(s))$ be an arclength parametrization of $\partial^-\Omega_\infty = \{(r \cos(-\alpha), r \sin(-\alpha)) : r \geq 0\}$, each measured from the corner at \mathcal{O} . Define

$$\gamma^+(s) = \gamma(x^+(s), y^+(s)) \quad \text{and} \quad \gamma^-(s) = \gamma(x^-(s), y^-(s)).$$

For $0 \leq \alpha \leq \frac{\pi}{2}$, the corner will be said to be *convex* and for $\frac{\pi}{2} < \alpha \leq \pi$, the corner will be said to be *nonconvex*.

In [Lancaster and Siegel 1996], the existence of radial limits of a bounded solution f to (1) that satisfies (2) on the smooth portions of $\partial\Omega$ is proven provided that γ was bounded away from 0 and π , and for a convex corner an additional condition is satisfied coupling γ^+ and γ^- . In this paper, we eliminate the requirement that γ is bounded away from 0 and π ; an additional condition must still be satisfied at a convex corner. The radial limits of f will be denoted by

$$Rf(\theta) = \lim_{r \rightarrow 0^+} f(r \cos \theta, r \sin \theta), \quad -\alpha < \theta < \alpha,$$

and $Rf(\pm\alpha) = \lim_{\partial^\pm\Omega^* \ni \mathbf{x} \rightarrow \mathcal{O}} f(\mathbf{x})$, $\mathbf{x} = (x, y)$, which are the limits of the boundary values of f on the two sides of the corner if these exist.

Theorem 1. *Let f be a bounded solution to (1) satisfying (2) on $\partial^\pm\Omega^* \setminus \{\mathcal{O}\}$, which is discontinuous at \mathcal{O} .*

- (a) *If $\alpha > \frac{\pi}{2}$ then $Rf(\theta)$ exists for all $\theta \in (-\alpha, \alpha)$.*
- (b) *If $\alpha \leq \frac{\pi}{2}$ and there exist constants $\underline{\gamma}^\pm, \bar{\gamma}^\pm$, $0 \leq \underline{\gamma}^\pm \leq \bar{\gamma}^\pm \leq \pi$ satisfying*

$$\pi - 2\alpha < \underline{\gamma}^+ + \underline{\gamma}^- \leq \bar{\gamma}^+ + \bar{\gamma}^- < \pi + 2\alpha$$

such that $\underline{\gamma}^\pm \leq \gamma^\pm(s) \leq \bar{\gamma}^\pm$ for all $s \in (0, s_0)$, for some $s_0 > 0$, then $Rf(\theta)$ exists for all $\theta \in (-\alpha, \alpha)$.

Furthermore, in either case, $Rf(\theta)$ is a continuous function on $(-\alpha, \alpha)$ which behaves in one of the following ways:

- (i) *$Rf(\theta)$ is a constant function of θ and f has a nontangential limit at \mathcal{O} .*
- (ii) *There exist α_1 and α_2 so that $-\alpha \leq \alpha_1 < \alpha_2 \leq \alpha$ and Rf is constant on $(-\alpha, \alpha_1]$ and $[\alpha_2, \alpha)$ and strictly increasing or strictly decreasing on $[\alpha_1, \alpha_2] \cap (-\alpha, \alpha)$. Label these case (I) and case (D), respectively.*
- (iii) *There exist $\alpha_1, \alpha_L, \alpha_R, \alpha_2$ so that $-\alpha \leq \alpha_1 < \alpha_L < \alpha_R < \alpha_2 \leq \alpha$, $\alpha_R = \alpha_L + \pi$, and Rf is constant on $(-\alpha, \alpha_1]$, $[\alpha_L, \alpha_R]$, and $[\alpha_2, \alpha)$ and either increasing on $[\alpha_1, \alpha_L] \cap (-\alpha, \alpha)$ and decreasing on $[\alpha_R, \alpha_2] \cap (-\alpha, \alpha)$ or decreasing on*

$[\alpha_1, \alpha_L] \cap (-\alpha, \alpha)$ and increasing on $[\alpha_R, \alpha_2] \cap (-\alpha, \alpha)$. Label these case (ID) and case (DI), respectively.

In Theorem 1 of [Lancaster and Siegel 1996] and Theorem 1 above, the existence of two intervals $(-\alpha, \alpha_1]$ and $[\alpha_2, \alpha)$ on which $Rf(\cdot)$ is constant (i.e., “side fans”) is established but the relationship between the sizes of these side fans and the contact angle is unclear. Theorem 2 of [Lancaster and Siegel 1996] establishes lower bounds on these sizes when the

$$\lim_{\partial^+ \Omega \ni (x, y) \rightarrow \mathcal{O}} \gamma(x, y) = \gamma_0^+ \quad \text{and} \quad \lim_{\partial^- \Omega \ni (x, y) \rightarrow \mathcal{O}} \gamma(x, y) = \gamma_0^-$$

are assumed to exist. (In [Lancaster 2010; 2012], these lower bounds were shown to be the actual sizes of the side fans.) What happens if the limits of γ at \mathcal{O} do not exist? Theorem 2 and Corollary 3 provide lower bounds in this situation.

For $0 < b < 1$, define

$$A_I^\pm(b) = \liminf_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_0^{b\epsilon} \cos(\gamma^\pm(t)) dt \quad \text{and} \quad A_S^\pm(b) = \limsup_{\epsilon \downarrow 0} \frac{1}{\epsilon} \int_0^{b\epsilon} \cos(\gamma^\pm(t)) dt.$$

Notice that $b \cos(\limsup_{t \downarrow 0} \gamma^\pm(t)) \leq A_I^\pm(b) \leq A_S^\pm(b) \leq b \cos(\liminf_{t \downarrow 0} \gamma^\pm(t))$.

Theorem 2. *Let f be a bounded solution to (1) satisfying (2) on $\partial^\pm \Omega^* \setminus \{\mathcal{O}\}$, which is discontinuous at \mathcal{O} . Assume $Rf(\theta)$ exists for all $\theta \in (-\alpha, \alpha)$. Then:*

- (a) *$Rf(\theta)$ is a continuous function on $(-\alpha, \alpha)$ which behaves as described in (i), (ii) or (iii) of Theorem 1.*
- (b) *There exist fans of constant radial limits adjacent to each tangent direction at \mathcal{O} and lower bounds on the sizes of these side fans exist.*

In terms of the cases labeled in Theorem 1, the sizes of the side fans $\beta^- = \alpha_1 + \alpha$ and $\beta^+ = \alpha - \alpha_2$ satisfy the following conditions:

- (1) $A_I^+ \left(\frac{\sin(\lambda - \beta^+)}{\sin(\lambda)} \right) + \frac{\sin(\beta^+)}{\sin(\lambda)} \geq 1$ for all $\lambda \in (\beta^+, \pi)$ for (I) and (DI).
- (2) $A_I^- \left(\frac{\sin(\lambda - \beta^-)}{\sin(\lambda)} \right) + \frac{\sin(\beta^-)}{\sin(\lambda)} \geq 1$ for all $\lambda \in (\beta^-, \pi)$ for (D) and (DI).
- (3) $1 + A_S^- \left(\frac{\sin(\lambda - \beta^-)}{\sin(\lambda)} \right) \leq \frac{\sin(\beta^-)}{\sin(\lambda)}$ for all $\lambda \in (\beta^-, \pi)$ for (I) and (ID).
- (4) $1 + A_S^+ \left(\frac{\sin(\lambda - \beta^+)}{\sin(\lambda)} \right) \leq \frac{\sin(\beta^+)}{\sin(\lambda)}$ for all $\lambda \in (\beta^+, \pi)$ for (D) and (ID).

2. Proofs of Theorems 1 and 2

The proof of Theorem 1 follows that established in [Lancaster 1985] and [Elcrat and Lancaster 1986] in which (i) the graph of the solution in $\Omega \times \mathbb{R}$ is represented

in isothermal coordinates, (ii) comparison arguments are used to prove that the component functions of the isothermal parametrization of the graph are uniformly continuous and so extend to be continuous on the closure of the parameter domain, (iii) boundary regularity theory (e.g., [Heinz 1970]) is used to prove that radial limits exist for almost every direction, (iv) cusp solutions are excluded (e.g., [Echart and Lancaster 2017]) and (v) the behavior of the radial limit function is determined. The only step which does not follow from previous work is (ii) and so the proof of Theorem 1 comes down to establishing (ii). The proof of Theorem 2 follows from standard “blow up” arguments.

2.1. Proof of Theorem 1. When $\alpha > \frac{\pi}{2}$, Theorem 1 is a consequence of [Entekhabi and Lancaster 2016]. Suppose now that $\alpha \leq \frac{\pi}{2}$. Since f is bounded and the prescribed mean curvature is $H(x, y, z) = \kappa z + \lambda$, there exist $M_1 \in (0, \infty)$ and $M_2 \in [0, \infty)$ such that

$$(3) \quad \sup_{(x,y) \in \Omega} |f(x, y)| \leq M_1 \quad \text{and} \quad \sup_{(x,y) \in \Omega} |H(x, y, f(x, y))| \leq M_2.$$

In §2.1 of [Entekhabi and Lancaster 2017], a specific torus is constructed which depends solely on M_2 and which is used as a comparison surface; one should compare this with, for example, [Lancaster and Siegel 1996], where several types of comparison surfaces are used, or [Entekhabi and Lancaster 2016], where an unduloid is used as a comparison surface. We shall use this torus as our comparison surface here. We will denote by q the modulus of continuity of the function h^- whose graph is the set \mathcal{T} which is the inner half of a torus with axis of symmetry $\{(2, y, 0) : y \in \mathbb{R}^2\}$, major radius $R_0 = 2$, and minor radius r_0 ; here

$$(4) \quad r_0 = \begin{cases} 1 & \text{if } M_2 = 0, \\ \frac{1}{M_2} + 1 - \sqrt{\left(\frac{1}{M_2}\right)^2 + 1} & \text{if } M_2 > 0. \end{cases}$$

Then q is also the modulus of continuity of functions (i.e., h^+ , h_β^- , h_β^+) whose graphs are obtained by rotations and translations in the horizontal plane of \mathcal{T} (see [Entekhabi and Lancaster 2017, p. 59]).

Let $\mathcal{S}_0 = \text{gra}(f) = \{(x, y, f(x, y)) : (x, y) \in \Omega^*\}$ and allow \mathcal{S} to be the closure of \mathcal{S}_0 in \mathbb{R}^3 . As in §2.2 of [Entekhabi and Lancaster 2017], there exists an isothermal parametrization $Y : E \rightarrow \mathbb{R}^3$ given by

$$Y(u, v) = (a(u, v), b(u, v), c(u, v))$$

such that $Y(\bar{E}) = \mathcal{S}$, $Y(E) = \mathcal{S}_0$, and (a_1) – (a_5) of [Entekhabi and Lancaster 2017] hold, where $E = B_1(\mathcal{O}) = \{(u, v) : u^2 + v^2 < 1\}$. By (a_2) of that paper, if we let $G(u, v) = (a(u, v), b(u, v))$ for $(u, v) \in E$, then $G \in C^0(\bar{E})$. From (a_3) of that paper, there exists a connected arc $\sigma \subset \partial E$ that Y maps strictly monotonically onto

$\{(x, y, f(x, y)) : (x, y) \in \partial\Omega^* \setminus \{\mathcal{O}\}\}$. Let the endpoints of σ be denoted \mathbf{o}_1 and \mathbf{o}_2 . There exists points $\mathbf{a}, \mathbf{b} \in \sigma$ such that $G(\mathbf{a}) = A$, $G(\mathbf{b}) = B$, G maps the arc $\mathbf{o}_2\mathbf{a}$ onto $\partial^-\Omega$ and G maps the arc $\mathbf{o}_1\mathbf{b}$ onto $\partial^+\Omega$. We must consider the two cases:

(A) $\mathbf{o}_1 = \mathbf{o}_2$,

(B) $\mathbf{o}_1 \neq \mathbf{o}_2$.

Assume first that (A) holds. Set $\mathbf{o} = \mathbf{o}_1 = \mathbf{o}_2$. We wish to prove that c is uniformly continuous on E and hence c extends to be continuous on \bar{E} . If so, then the existence and behavior of the radial limits of f follows as in [Entekhabi and Lancaster 2017; Lancaster and Siegel 1996]. There are three possible cases:

(i) $\underline{\gamma}^- > 0$ and $\bar{\gamma}^- < \pi$,

(ii) $\underline{\gamma}^+ > 0$ and $\bar{\gamma}^+ < \pi$,

(iii) $(\underline{\gamma}^- = 0 \text{ or } \bar{\gamma}^- = \pi)$ and $(\underline{\gamma}^+ = 0 \text{ or } \bar{\gamma}^+ = \pi)$.

Case (i). Let $\lambda_1 = \underline{\gamma}^+$, $\lambda_2 = \bar{\gamma}^+$, $\gamma_2 = \underline{\gamma}^-$. We observe that $\lambda_2 = \bar{\gamma}^+ < \pi + 2\alpha - \bar{\gamma}^-$, $\lambda_1 = \underline{\gamma}^+ > \pi - 2\alpha - \underline{\gamma}^-$, and so $\lambda_2 - \lambda_1 < 4\alpha$. We wish to use the argument in the proof of Theorem 2 of [Entekhabi and Lancaster 2017]. Since $\pi - 2\alpha - \lambda_1 < \gamma_2 < \pi + 2\alpha - \lambda_2$, we can choose $\tau_1, \tau_2 \in (0, \pi)$ such that $\tau_1 \in (\pi - 2\alpha - \lambda_1, \gamma_2)$ and $\tau_2 \in (\gamma_2, \pi + 2\alpha - \lambda_2)$. Set $\beta_1 = \frac{\pi}{2} - \tau_1$ and $\beta_2 = \tau_2 - \frac{\pi}{2}$. With these choices of β_1 and β_2 , notice that

$$T(h^- \circ T_{\beta_1})(x_1, 0) \cdot (0, -1) = \cos(\tau_1) > \cos(\gamma_2) \quad \text{for } 0 < x_1 < 2 - r_0$$

and

$$T(h^+ \circ T_{\beta_2})(x_1, 0) \cdot (0, -1) = \cos(\tau_2) < \cos(\gamma_2) \quad \text{for } 0 < x_1 < 2 - r_0$$

(see [Entekhabi and Lancaster 2017, p. 59]). This implies that for $\delta_1 = \delta_1(\beta_1, \beta_2) > 0$ small enough and $\mathbf{x} \in \partial^-\Omega$ with $|\mathbf{x}| < \delta_1$, we have

$$(5) \quad T(h_{\beta_1}^-)(\mathbf{x}) \cdot \vec{v}(\mathbf{x}) > \cos(\gamma(\mathbf{x})) \quad \text{and} \quad T(h_{\beta_2}^+)(\mathbf{x}) \cdot \vec{v}(\mathbf{x}) < \cos(\gamma(\mathbf{x})).$$

Since $\beta_1, \beta_2 \neq \pm\frac{\pi}{2}$, there exists $R = R(\beta_1, \beta_2) > 0$ such that $B_R(\mathcal{O}) \cap \Omega^* \subset \Delta_{\beta_1} \cap \Delta_{\beta_2}$, where Δ_β is as in §2.1 of [Entekhabi and Lancaster 2017]. For each $\delta \in (0, 1)$, allow

$$(6) \quad p(\delta) = \sqrt{\frac{8\pi M_0}{\ln(1/\delta)}}$$

where M_0 is the area of S_0 .

Let $\epsilon > 0$. Choose $\delta > 0$ such that

$$\begin{aligned} \sqrt{\delta} &< \min\{\|\mathbf{o} - \mathbf{a}\|, \|\mathbf{o} - \mathbf{b}\|\}, \\ p(\delta) &< \delta_1(\beta_1, \beta_2), \quad p(\delta) < R(\beta_1, \beta_2), \quad p(\delta) + q(p(\delta)) < \frac{\epsilon}{2}. \end{aligned}$$

Let $\mathbf{w}_0 = (u_0, v_0) \in E$. From the Courant–Lebesgue lemma, there exists a $\rho(\delta) \in (\delta, \sqrt{\delta})$ such that the arclength $l_{\rho(\delta)}$ of $C'_{\rho(\delta)}$ is less than $p(\delta)$, where $C_\delta = \{\mathbf{w} \in E : \|\mathbf{w} - \mathbf{w}_0\| = \delta\}$ and $C'_\delta = Y(C_\delta)$. Set $B_\delta = \{\mathbf{w} \in E : \|\mathbf{w} - \mathbf{w}_0\| < \delta\}$ and $B'_\delta = Y(B_\delta)$. Then, for $\mathbf{w} \in C'_{\rho(\delta)}$, there exist functions

$$(7) \quad b^+(x, y) = f(\mathbf{w}) + p(\delta) + h_{\beta_1}^-(x, y) \quad \text{for } (x, y) \in \Delta_{\beta_1},$$

$$(8) \quad b^-(x, y) = f(\mathbf{w}) - p(\delta) - h_{\beta_2}^+(x, y) \quad \text{for } (x, y) \in \Delta_{\beta_2}$$

where $\beta_1 = \frac{\pi}{2} - \tau_1$ and $\beta_2 = \tau_2 - \frac{\pi}{2}$. From (10) of [Entekhabi and Lancaster 2017], we have that $\operatorname{div}(b^+) \leq -M_2$ in Δ_{β_1} and $\operatorname{div}(b^-) \geq M_2$ in Δ_{β_2} . So in $\Omega \cap \Delta_{\beta_1}$, $\operatorname{div}(Tb^+) \leq \operatorname{div}(Tf)$. On $\partial^-\Omega \cap B_{\delta_1}(\mathcal{O})$, $Tb^+ \cdot \nu \geq Tf \cdot \nu$. As in the proof of Theorem 2 of that paper,

$$(9) \quad f(x, y) < b^+(x, y) \quad \text{for } (x, y) \in \Delta_{\beta_1} \cap B'_{\rho(\delta)},$$

where $B'_{\rho(\delta)} = Y(B_{\rho(\delta)})$. This follows since $Tb^+ \cdot \nu \geq Tf \cdot \nu$ on $\partial^+\Omega \cap B_{\delta_2}(\mathcal{O})$ by (15) of that paper if $\tau_1 + 2\alpha \leq \pi$ and no boundary condition on $\partial^+\Omega$ is required if $\tau_1 + 2\alpha > \pi$.

Repeat the same argument with $\lambda_1 = \underline{\gamma}^+$, $\lambda_2 = \bar{\gamma}^+$ and $\gamma_2 = \bar{\gamma}^-$. In the same way as above, there exist functions

$$(10) \quad b_*^+(x, y) = f(\mathbf{w}) + p(\delta) + h_{\beta_1}^-(x, y) \quad \text{for } (x, y) \in \Delta_{\beta_1},$$

$$(11) \quad b_*^-(x, y) = f(\mathbf{w}) - p(\delta) - h_{\beta_2}^+(x, y) \quad \text{for } (x, y) \in \Delta_{\beta_2}$$

such that

$$(12) \quad b_*^-(x, y) < f(x, y)$$

for $(x, y) \in \Delta_{\beta_2} \cap B'_{\rho(\delta)}$ where $B'_{\rho(\delta)} = Y(B_{\rho(\delta)})$. Then combining (9) and (12) we get

$$(13) \quad b_*^-(x, y) < f(x, y) < b^+(x, y)$$

for $(x, y) \in \Delta_{\beta_1} \cap \Delta_{\beta_2} \cap B'_{\rho(\delta)}$. As in [Entekhabi and Lancaster 2017], it follows that $c(u, v)$ is uniformly continuous on E .

Case (ii). Case (ii) is simply case (i) reflected about the xz -plane and the proof follows as above.

Case (iii). Notice that

$$0 \leq \pi - 2\alpha < \underline{\gamma}^+ + \underline{\gamma}^- \leq \bar{\gamma}^+ + \bar{\gamma}^- < \pi + 2\alpha \leq \pi$$

and so $\underline{\gamma}^- = 0$ and $\underline{\gamma}^+ = 0$ cannot both occur and $\bar{\gamma}^+ = \pi$ and $\bar{\gamma}^- = \pi$ cannot both occur. The result follows from this, using the arguments in cases 1 and 2. In particular, if $\underline{\gamma}^- > 0$, then we obtain a supersolution b^+ as in case (i) (see Figure 1)

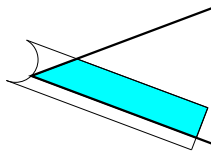


Figure 1. The domain of a supersolution in case (i).

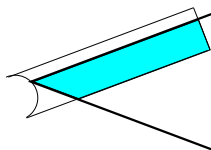


Figure 2. The domain of a supersolution in case (ii).

and if $\gamma^- = 0$, we obtain a supersolution b^+ as in case (ii) (see Figure 2); if $\bar{\gamma}^- < \pi$, we obtain a subsolution b_*^- as in case (i) and if $\bar{\gamma}^- = \pi$, we obtain a subsolution as in case (ii).

Now assume (B) holds. Let $B = \{(x, y) \in \mathbb{R}^2 : \sqrt{x^2 + y^2} < 1, y \geq 0\}$ and let \bar{B} be the closure of B in \mathbb{R}^2 . Let $g : \bar{B} \rightarrow \bar{E}$ be a conformal or anticonformal map taking $\{(u, 0) : -1 \leq u \leq 1\}$ onto $\partial E \setminus \sigma$ such that the map $X = Y \circ g : B \rightarrow \mathbb{R}^3$ has a downward orientation (i.e., the normal $X_u \times X_v$ to \mathcal{S}_0 gives a downward orientation). Writing $X(u, v) = (x(u, v), y(u, v), z(u, v))$ and $K(u, v) = (x(u, v), y(u, v))$, we have $K \in C^0(\bar{B})$ and $K(u, 0) = (0, 0)$ while $X(u, 0) = (0, 0, z(u, 0))$ for $u \in [-1, 1]$. Then the argument follows from [Lancaster and Siegel 1996] and the previous argument here, as explained in [Entekhabi and Lancaster 2017]. \square

2.2. Proof of Theorem 2. We first note that if $\delta_1, \delta_2 \in (-\alpha, \alpha)$ with $\delta_1 < \delta_2$ and $Rf(\delta_1)$ and $Rf(\delta_2)$ both exist, then it follows from [Elcrat and Lancaster 1986] that $Rf(\theta)$ exists for all $\theta \in [\delta_1, \delta_2]$ and $Rf(\theta)$ is a continuous function of θ on $[\delta_1, \delta_2]$ which behaves as described in (i), (ii) or (iii) of Theorem 1. The first part of Theorem 2 (i.e., (a)) follows from this.

Suppose $\{\epsilon_j\}$ is a decreasing sequence with $\lim_{j \rightarrow \infty} \epsilon_j = 0$. Let $I = (-1, 1)$ and set

$$\gamma_j(s) = \begin{cases} \gamma^+(\epsilon_j s) & \text{if } 0 < s < 1, \\ \gamma^-(-\epsilon_j s) & \text{if } -1 < s < 0 \end{cases}$$

for $j \in \mathbb{N}$; then $\{\cos(\gamma_j) : j \in \mathbb{N}\}$ is a subset of the unit ball in $L^\infty(I) = (L^1(I))^*$. By the Banach–Alaoglu theorem, there exist a subsequence $\{\epsilon_{j_k}\}$ of $\{\epsilon_j\}$ and a function $h = h_{\{\epsilon_{j_k}\}} \in L^\infty(I)$ such that $\cos(\gamma_{j_k})$ converges weak-star to h ; that is, for each $m \in L^1(I)$,

$$\lim_{k \rightarrow \infty} \int_{-1}^1 \cos(\gamma_{j_k}(s)) m(s) ds = \int_{-1}^1 h(s) m(s) ds.$$

Let us define $\gamma^* = \gamma_{\{\epsilon_{j_k}\}}^* = \cos^{-1}(h)$ (almost everywhere on $(-1, 1)$). For any $b \in (0, 1)$, by choosing m to be the characteristic function of the interval $(0, b)$ we see that

$$\int_0^b h(s) ds = \lim_{k \rightarrow \infty} \int_0^b \cos(\gamma_{j_k}(s)) ds = \lim_{k \rightarrow \infty} \frac{1}{\epsilon_{j_k}} \int_0^{b\epsilon_{j_k}} \cos(\gamma^+(t)) dt$$

and, by choosing m to be the characteristic function of the interval $(-b, 0)$,

$$\int_{-b}^0 h(s) ds = \lim_{k \rightarrow \infty} \int_{-b}^0 \cos(\gamma_{j_k}(s)) ds = \lim_{k \rightarrow \infty} \frac{1}{\epsilon_{j_k}} \int_0^{b\epsilon_{j_k}} \cos(\gamma^-(t)) dt;$$

hence

$$\int_0^b \cos(\gamma^*(s)) ds = \lim_{k \rightarrow \infty} \frac{1}{\epsilon_{j_k}} \int_0^{b\epsilon_{j_k}} \cos(\gamma^+(t)) dt \geq \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{\epsilon b} \cos(\gamma^+(t)) dt$$

and

$$\int_{-b}^0 \cos(\gamma^*(s)) ds = \lim_{k \rightarrow \infty} \frac{1}{\epsilon_{j_k}} \int_0^{b\epsilon_{j_k}} \cos(\gamma^-(t)) dt \geq \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{\epsilon b} \cos(\gamma^-(t)) dt.$$

Thus

$$(14) \quad \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{\epsilon b} \cos(\gamma^\pm(t)) dt \leq \liminf_{j \rightarrow \infty} \int_0^b \cos(\gamma^\pm(\epsilon_j s)) ds.$$

Choose a sequence $\{\epsilon_j\}$ with $\lim_{j \rightarrow \infty} \epsilon_j = 0$ such that

$$\lim_{j \rightarrow \infty} \frac{1}{\epsilon_j} \int_0^{b\epsilon_j} \cos(\gamma^+(t)) dt = \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{\epsilon b} \cos(\gamma^+(t)) dt;$$

as above, there exist a subsequence $\{\epsilon_{j_k}\}$ of $\{\epsilon_j\}$ and $\gamma_* \in L^\infty(I)$ such that $\cos(\gamma_{j_k})$ converges weak-star to $\cos(\gamma_*)$. Then

$$\begin{aligned} \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{\epsilon b} \cos(\gamma^+(t)) dt &= \lim_{k \rightarrow \infty} \frac{1}{\epsilon_{j_k}} \int_0^{b\epsilon_{j_k}} \cos(\gamma^+(t)) dt \\ &= \lim_{k \rightarrow \infty} \int_0^b \cos(\gamma_{j_k}(s)) ds = \int_0^b \cos(\gamma_*(s)) ds. \end{aligned}$$

Case 1. Suppose case (I) or (DI) of Theorem 1 holds and $\alpha_2 = \alpha - \beta^+$. Let us assume there exists $\lambda \in (\beta^+, \pi)$ such that

$$(15) \quad A_I^+ \left(\frac{\sin(\lambda - \beta^+)}{\sin(\lambda)} \right) + \frac{\sin(\beta^+)}{\sin(\lambda)} < 1;$$

we shall show that this leads to a contradiction. Set

$$b = \frac{\sin(\lambda - \beta^+)}{\sin(\lambda)}.$$

Choose a sequence $\{\epsilon_j\}$ with $\lim_{j \rightarrow \infty} \epsilon_j = 0$ such that

$$(16) \quad \lim_{j \rightarrow \infty} \frac{1}{\epsilon_j} \int_0^{b\epsilon_j} \cos(\gamma^+(t)) dt = \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{b\epsilon} \cos(\gamma^+(t)) dt;$$

as before, there exist a subsequence $\{\epsilon_{j_k}\}$ of $\{\epsilon_j\}$ and $\gamma_* \in L^\infty(I)$ such that $\cos(\gamma_{j_k})$ converges weak-star to $\cos(\gamma_*)$. Then

$$\begin{aligned} \liminf_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int_0^{b\epsilon} \cos(\gamma^+(t)) dt &= \lim_{k \rightarrow \infty} \frac{1}{\epsilon_{j_k}} \int_0^{b\epsilon_{j_k}} \cos(\gamma^+(t)) dt \\ &= \lim_{k \rightarrow \infty} \int_0^b \cos(\gamma_{j_k}(s)) ds = \int_0^b \cos(\gamma_*(s)) ds. \end{aligned}$$

Let $\theta_0 \in (\sigma, \alpha_2)$, where $\sigma = \alpha_1$ if case (I) holds and $\sigma = \alpha_R$ if case (DI) holds, and $z_0 = Rf(\theta_0)$. Set $\Omega_k = \{(x, y) \in \mathbb{R}^2 : (\epsilon_{j_k}x, \epsilon_{j_k}y) \in \Omega\}$ and define $f_k \in C^\infty(\Omega_k)$ by

$$f_k(x, y) = \frac{1}{\epsilon_{j_k}} (f(\epsilon_{j_k}x, \epsilon_{j_k}y) - z_0)$$

for $(x, y) \in \Omega_k$. Let γ_k be defined on $\partial\Omega_k \setminus \{\mathcal{O}\}$ by

$$\gamma_k(x, y) = \gamma(\epsilon_{j_k}x, \epsilon_{j_k}y)$$

and let $\nu_k = \nu_k(x, y)$ denote the outward unit normal to $\partial\Omega_k$. Then f_k satisfies the capillary problem

$$\begin{aligned} Nf_k(x, y) &= \epsilon_{j_k} \kappa f(\epsilon_{j_k}x, \epsilon_{j_k}y) + \epsilon_{j_k} \lambda, & (x, y) \in \Omega_k, \\ T f_k \cdot \nu_k &= \cos(\gamma_k) & \text{on } \partial\Omega_k \setminus \{\mathcal{O}\}. \end{aligned}$$

Since $Rf(\theta) < z_0$ if $\sigma < \theta < \theta_0$ and $Rf(\theta) > z_0$ if $\theta_0 < \theta < \alpha$, we see (e.g., [Lancaster 2010; 2012; Simon 1980]; also see [Tam 1984; 1986]) that $\{f_k\}$ converges locally to the generalized solution f_∞ (in the sense of Miranda [1977] and Giusti [1980; 1984]) of the functional

$$\mathcal{F}_\infty(g) = \iint_{\Omega_\infty} \sqrt{1 + |Dg|^2} dx - \int_{\partial\Omega_\infty} \cos(\gamma_*(s)) g ds,$$

where

$$f_\infty(r \cos(\theta), r \sin(\theta)) = \begin{cases} -\infty & \text{if } -\alpha < \theta < \theta_0, \\ \infty & \text{if } \theta_0 < \theta < \alpha \end{cases}$$

if case (I) holds or case (DI) holds and $z_0 > Rf(\theta)$ for all $\theta \in (-\alpha, \alpha_L)$ and

$$f_\infty(r \cos(\theta), r \sin(\theta)) = \begin{cases} \infty & \text{if } -\alpha < \theta < \theta_h, \\ -\infty & \text{if } \theta_h < \theta < \theta_0, \\ \infty & \text{if } \theta_0 < \theta < \alpha \end{cases}$$

with $Rf(\theta_h) = z_0$ and $\theta_h < \alpha_L$ otherwise.

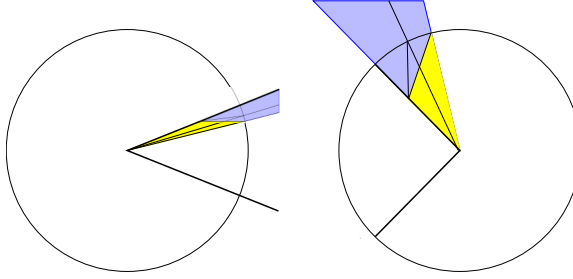


Figure 3. The yellow region represents Σ_{θ_0} .

Let us now define the sets

$$\mathcal{P} = \{(x, y) \in \Omega_\infty : f_\infty(x, y) = \infty\} \quad \text{and} \quad \mathcal{N} = \{(x, y) \in \Omega_\infty : f_\infty(x, y) = -\infty\}.$$

These sets have a special structure which follows from the fact that \mathcal{P} minimizes the functional

$$\Phi(A) = \iint_{\Omega_\infty} |D\chi_A| - \int_{\partial\Omega_\infty} \cos(\gamma_*) \chi_A dH^1$$

and \mathcal{N} minimizes the functional

$$\Psi(A) = \iint_{\Omega_\infty} |D\chi_A| + \int_{\partial\Omega_\infty} \cos(\gamma_*) \chi_A dH^1$$

in the appropriate sense (e.g., [Giusti 1980; Lancaster and Siegel 1996; Miranda 1977]). Let Σ_{θ_0} denote the (open) triangular region whose boundary is the triangle with vertices $(0, 0)$, $B = (b \cos(\alpha), b \sin(\alpha))$ and $C = (\cos(\theta_0), \sin(\theta_0))$ and set $A = \mathcal{P} \setminus \Sigma_{\theta_0}$ (see Figure 3). Simple trigonometric computations with $R > 2$ show that

$$(17) \quad \Phi(B(\mathcal{O}, R) \cap \mathcal{P}) - \Phi(B(\mathcal{O}, R) \cap \mathcal{P} \setminus \Sigma_{\theta_0}) = (1 - A_I^+(b)) - \left(\frac{\sin(\alpha - \theta_0)}{\sin(\omega)} \right),$$

where $\pi - \omega$ is the angle $\angle \mathcal{O}BC$. This holds for all $\theta_0 < \alpha_2 = \alpha - \beta^+$; taking the limit as $\theta_0 \uparrow \alpha - \beta^+$ and noticing that $\omega \rightarrow \lambda$ as $\theta_0 \uparrow \alpha - \beta^+$, we see that

$$\Phi(B(\mathcal{O}, R) \cap \mathcal{P}) - \Phi(B(\mathcal{O}, R) \cap \mathcal{P} \setminus \Sigma_{\alpha_2}) = (1 - A_I^+(b)) - \left(\frac{\sin(\beta^+)}{\sin(\lambda)} \right) > 0$$

or

$$\Phi(B(\mathcal{O}, R) \cap \mathcal{P}) > \Phi(B(\mathcal{O}, R) \cap \mathcal{P} \setminus \Sigma_{\alpha_2});$$

this contradicts the fact that \mathcal{P} (locally) minimizes Φ . Therefore (15) is false. This completes case 1.

Case 2. Suppose case (D) or (DI) of Theorem 1 holds and $\alpha_1 = -\alpha + \beta^-$. Let us assume there exists $\lambda \in (\beta^-, \pi)$ such that

$$(18) \quad A_I^- \left(\frac{\sin(\lambda - \beta^-)}{\sin(\lambda)} \right) + \frac{\sin(\beta^-)}{\sin(\lambda)} < 1.$$

Using an argument similar to that in case 1, we reach a contradiction.

Case 3. Suppose case (I) or (ID) of Theorem 1 holds and $\alpha_1 = -\alpha + \beta^-$. Let us assume there exists $\lambda \in (\beta^-, \pi)$ such that

$$(19) \quad 1 + A_S^- \left(\frac{\sin(\lambda - \beta^-)}{\sin(\lambda)} \right) > \frac{\sin(\beta^-)}{\sin(\lambda)}.$$

Set

$$b = \frac{\sin(\lambda - \beta^-)}{\sin(\lambda)}.$$

Arguing as in case 1, we see that the set $\mathcal{N} = \{(x, y) \in \Omega_\infty : f_\infty(x, y) = -\infty\}$ minimizes the functional

$$\Psi(A) = \iint_{\Omega_\infty} |D\chi_A| + \int_{\partial\Omega_\infty} \cos(\gamma_*) \chi_A dH^1$$

in the appropriate sense (e.g., [Giusti 1980; Lancaster and Siegel 1996; Miranda 1977]). Let Σ_{θ_0} denote the (open) triangular region whose boundary is the triangle with vertices $(0, 0)$, $B = (b \cos(-\alpha), b \sin(-\alpha))$ and $C = (\cos(\theta_0), \sin(\theta_0))$ and set $A = \mathcal{N} \setminus \Sigma_{\theta_0}$ (see Figure 3). Simple trigonometric computations with $R > 2$ show that

$$\Psi(B(\mathcal{O}, R) \cap \mathcal{N}) - \Psi(B(\mathcal{O}, R) \cap \mathcal{N} \setminus \Sigma_{\theta_0}) = (1 + A_S^-(b)) - \left(\frac{\sin(\alpha + \theta_0)}{\sin(\omega)} \right),$$

where ω is the angle $\angle OBC$. This holds for all $\theta_0 > \alpha_1 = -\alpha + \beta^-$; taking the limit as $\theta_0 \downarrow -\alpha + \beta^-$ and noticing that $\omega \rightarrow \lambda$ as $\theta_0 \downarrow -\alpha + \beta^-$, we see that

$$\Psi(B(\mathcal{O}, R) \cap \mathcal{N}) - \Psi(B(\mathcal{O}, R) \cap \mathcal{N} \setminus \Sigma_{\alpha_1}) = (1 + A_S^-(b)) - \left(\frac{\sin(\beta^-)}{\sin(\lambda)} \right) > 0$$

or

$$\Psi(B(\mathcal{O}, R) \cap \mathcal{N}) > \Psi(B(\mathcal{O}, R) \cap \mathcal{N} \setminus \Sigma_{\alpha_1});$$

this contradicts the fact that \mathcal{N} (locally) minimizes Ψ . Therefore (19) is false. This completes case 3.

Case 4. Suppose case (D) or (ID) of Theorem 1 holds and $\alpha_2 = \alpha - \beta^+$. Let us assume there exists $\lambda \in (\beta^+, \pi)$ such that

$$(20) \quad 1 + A_S^+ \left(\frac{\sin(\lambda - \beta^+)}{\sin(\lambda)} \right) > \frac{\sin(\beta^+)}{\sin(\lambda)}.$$

Using an argument similar to that in case 3, we reach a contradiction. The proof of Theorem 2 is then complete. \square

3. Corollaries and examples

Corollary 3. Suppose $m \in [-1, 1]$; set $\sigma = \cos^{-1}(m) \in [0, \pi]$.

- (a) If $A_I^+(b) \leq mb$ and case (I) or (DI) holds, then $\beta^+ \geq \sigma$.
- (b) If $A_I^-(b) \leq mb$ and case (D) or (DI) holds, then $\beta^- \geq \sigma$.
- (c) If $A_S^-(b) \geq mb$ and case (I) or (ID) holds, then $\beta^- \geq \pi - \sigma$.
- (d) If $A_S^+(b) \geq mb$ and case (D) or (ID) holds, then $\beta^+ \geq \pi - \sigma$.

Proof. (a) Suppose case (I) or (DI) of Theorem 1 holds, $\sigma \in [0, \pi]$, $\cos(\sigma) = m$, and $\beta^+ < \sigma$. By Theorem 2(a), we know that

$$\sin(\sigma) \left(\frac{\sin(\lambda - \beta^+)}{\sin(\lambda)} \right) + \frac{\sin(\beta^+)}{\sin(\lambda)} \geq A_I^+ \left(\frac{\sin(\lambda - \beta^+)}{\sin(\lambda)} \right) + \frac{\sin(\beta^+)}{\sin(\lambda)} \geq 1$$

or

$$\frac{\cos(\sigma) \sin(\lambda - \beta^+) + \sin(\beta^+)}{\sin(\lambda)} \geq 1$$

for all $\lambda \in (\beta^+, \pi)$. Since $\sigma > \beta^+$, we may set $\lambda = \sigma$ and obtain

$$\cos(\sigma - \beta^+) = \frac{\cos(\sigma) \sin(\sigma - \beta^+) + \sin(\beta^+)}{\sin(\sigma)} \geq 1,$$

which is a contradiction since $\sigma - \beta^+ \neq 0$. Thus $\beta^+ \geq \sigma$.

(b) This is essentially the same as (a).

(c) Suppose case (I) or (ID) of Theorem 1 holds, $\sigma \in [0, \pi]$, $\cos(\sigma) = m$, and $\beta^- < \pi - \sigma$. By Theorem 2(c), we know that

$$1 + \sin(\sigma) \left(\frac{\sin(\lambda - \beta^-)}{\sin(\lambda)} \right) \leq 1 + A_S^- \left(\frac{\sin(\lambda - \beta^-)}{\sin(\lambda)} \right) \leq \frac{\sin(\beta^-)}{\sin(\lambda)}$$

or

$$\frac{\sin(\lambda) + \cos(\sigma) \sin(\lambda - \beta^-) - \sin(\beta^-)}{\sin(\lambda)} \leq 0$$

for all $\lambda \in (\beta^-, \pi)$. Since $\beta^- < \pi - \sigma$, we may set $\lambda = \pi - \sigma$ and obtain

$$1 + \cos(\sigma + \beta^-) = \frac{\sin(\sigma) + \cos(\sigma) \sin(\sigma + \beta^-) - \sin(\beta^-)}{\sin(\sigma)} \leq 0,$$

which is a contradiction since $\sigma + \beta^- < \pi$. Thus $\beta^- \geq \pi - \sigma$.

(d) This is essentially the same as (c). \square

Example 4. Let $\alpha \in (0, \pi]$ and $\gamma_1^\pm, \gamma_2^\pm \in [0, \pi]$ with $\gamma_1^+ \leq \gamma_2^+$ and $\gamma_1^- \leq \gamma_2^-$. Set

$$\Omega = \{(r \cos(\theta), r \sin(\theta)) : 0 < r < 1, -\alpha < \theta < \alpha\}.$$

For each $n \in \mathbb{N}$, let $A_n = (2^{-n^2}, 2^{-n(n-1)}]$ and $B_n = (2^{-n(n+1)}, 2^{-n^2}]$. Define

$$\gamma(s) = \sum_{n=1}^{\infty} (\gamma_1^+ I_{A_n}(s) + \gamma_2^+ I_{B_n}(s) + \gamma_1^- I_{A_n}(-s) + \gamma_2^- I_{B_n}(-s)),$$

so that γ is defined on $\partial\Omega \cap B(\mathcal{O}, 1)$ by

$$\gamma(r \cos(\theta), r \sin(\theta)) = \begin{cases} \gamma_1^+ & \text{if } \theta = \alpha, 2^{-n^2} < r \leq 2^{-n(n-1)} \text{ for some } n \in \mathbb{N}, \\ \gamma_2^+ & \text{if } \theta = \alpha, 2^{-n(n+1)} < r \leq 2^{-n^2} \text{ for some } n \in \mathbb{N}, \\ \gamma_1^- & \text{if } \theta = -\alpha, 2^{-n^2} < r \leq 2^{-n(n-1)} \text{ for some } n \in \mathbb{N}, \\ \gamma_2^- & \text{if } \theta = -\alpha, 2^{-n(n+1)} < r \leq 2^{-n^2} \text{ for some } n \in \mathbb{N}. \end{cases}$$

Set

$$c_j = \begin{cases} 2^{-\frac{j}{2}(\frac{j}{2}+1)} & \text{if } j \text{ is even,} \\ 2^{-\left(\frac{j+1}{2}\right)^2} & \text{if } j \text{ is odd.} \end{cases}$$

Let $b \in (0, 1)$ be fixed for now. Set $\epsilon_j = c_{2j}/b$ ($j \in \mathbb{N}$); notice that $c_{2j+1}/c_{2j} = 2^{-(j+1)}$. Then

$$\begin{aligned} b \cos(\gamma_1^\pm) &\geq A_S^\pm(b) \\ &\geq \lim_{j \rightarrow \infty} \frac{1}{\epsilon_j} \int_0^{\epsilon_j b} \cos(\gamma^\pm(t)) dt \\ &= \lim_{j \rightarrow \infty} b \int_0^1 \cos(\gamma_j^\pm(sb)) ds \\ &= \lim_{j \rightarrow \infty} b \int_0^1 \cos(\gamma^\pm(c_{2j}s)) ds \\ &= \lim_{j \rightarrow \infty} b \left(\int_{\frac{c_{2j+1}}{c_{2j}}}^1 \cos(\gamma^\pm(c_{2j}s)) ds + \int_0^{\frac{c_{2j+1}}{c_{2j}}} \cos(\gamma^\pm(c_{2j}s)) ds \right) \\ &= \lim_{j \rightarrow \infty} b \left(\cos(\gamma_1^\pm)(1 - 2^{-(j+1)}) + \int_0^{2^{-(j+1)}} \cos(\gamma^\pm(c_{2j}s)) ds \right) \\ &= b \cos(\gamma_1^\pm). \end{aligned}$$

Using a similar argument for $A_I^\pm(b)$ with $\epsilon_j = c_{2j+1}/b$, $j \in \mathbb{N}$, we see that

$$(21) \quad A_I^\pm(b) = b \cos(\gamma_2^\pm) \quad \text{and} \quad A_S^\pm(b) = b \cos(\gamma_1^\pm).$$

Example 5. Let $\alpha \in (0, \pi]$ and $\gamma_1^\pm, \gamma_2^\pm \in [0, \pi]$ with $\gamma_1^+ \leq \gamma_2^+$ and $\gamma_1^- \leq \gamma_2^-$. Set

$$\Omega = \{(r \cos(\theta), r \sin(\theta)) : 0 < r < 1, -\alpha < \theta < \alpha\}.$$

For each $n \in \mathbb{N}$, let $A_n = \left(\frac{2}{4^n}, \frac{4}{4^n}\right)$, $B_n = \left(\frac{1}{4^n}, \frac{2}{4^n}\right)$, and $C_n = \left\{\frac{4}{4^n}\right\}$. Define

$$\gamma(s) = \sum_{n=1}^{\infty} (\gamma_1^+ I_{A_n}(s) + \gamma_2^+ I_{B_n}(s) + \pi I_{C_n}(s) + \gamma_1^- I_{A_n}(-s) + \gamma_2^- I_{B_n}(-s) + \pi I_{C_n}(-s)),$$

so that γ is defined on $\partial\Omega \cap B(\mathcal{O}, 1)$ by

$$\gamma(r \cos(\theta), r \sin(\theta)) = \begin{cases} \gamma_1^+ & \text{if } \theta = \alpha, 2/4^n < r < 4/4^n \text{ for some } n \in \mathbb{N}, \\ \gamma_2^+ & \text{if } \theta = \alpha, 1/4^n < r < 2/4^n \text{ for some } n \in \mathbb{N}, \\ \pi & \text{if } \theta = \alpha, r = 4/4^n \text{ for some } n \in \mathbb{N}, \\ 0 & \text{if } \theta = \alpha, r = 2/4^n \text{ for some } n \in \mathbb{N}, \\ \gamma_1^- & \text{if } \theta = -\alpha, 2/4^n < r < 4/4^n \text{ for some } n \in \mathbb{N}, \\ \gamma_2^- & \text{if } \theta = -\alpha, 1/4^n < r < 2/4^n \text{ for some } n \in \mathbb{N}, \\ \pi & \text{if } \theta = -\alpha, r = 4/4^n \text{ for some } n \in \mathbb{N}, \\ 0 & \text{if } \theta = -\alpha, r = 2/4^n \text{ for some } n \in \mathbb{N}. \end{cases}$$

Then

$$\begin{aligned} \liminf_{r \rightarrow 0} \gamma(r \cos(\pm\alpha), r \sin(\pm\alpha)) &= 0, \\ \limsup_{r \rightarrow 0} \gamma(r \cos(\pm\alpha), r \sin(\pm\alpha)) &= \pi, \\ \text{ess } \liminf_{r \rightarrow 0} \gamma(r \cos(\pm\alpha), r \sin(\pm\alpha)) &= \gamma_1^\pm, \\ \text{ess } \limsup_{r \rightarrow 0} \gamma(r \cos(\pm\alpha), r \sin(\pm\alpha)) &= \gamma_2^\pm. \end{aligned}$$

Thus $A_I^\pm(b) \geq b \cos(\gamma_2^\pm)$ and $A_S^\pm(b) \leq b \cos(\gamma_1^\pm)$.

Let $b \in (0, 1)$ be fixed for now. If we set $\epsilon_j = 1/(b4^j)$ ($j \in \mathbb{N}$), then $\gamma_j(s) = \gamma(s/b)$ and so

$$A_S^+(b) \geq \lim_{j \rightarrow \infty} \int_0^b \cos(\gamma_j(s)) ds = b \int_0^1 \cos(\gamma(s)) ds = b \left(\frac{2}{3} \cos(\gamma_1^+) + \frac{1}{3} \cos(\gamma_2^+) \right),$$

and, if we set $\epsilon_j = 2/(b4^j)$ ($j \in \mathbb{N}$), then $\gamma_j(s) = \gamma(s/(2b))$ and so

$$A_I^+(b) \leq \lim_{j \rightarrow \infty} \int_0^b \cos(\gamma_j(s)) ds = 2b \int_0^{\frac{1}{2}} \cos(\gamma(s)) ds = b \left(\frac{1}{3} \cos(\gamma_1^+) + \frac{2}{3} \cos(\gamma_2^+) \right);$$

similar estimates hold on $\partial^-\Omega$. Now suppose (η_j) is any decreasing sequence in $(0, 1)$ converging to zero. For each $j \in \mathbb{N}$, there exists a $k \in \mathbb{N}$ such that $\frac{1}{4} \leq 4^{k-1} \eta_j b < 1$ and, since γ is piecewise constant, a direct calculation shows

that $\int_0^b \cos(\gamma^\pm(\eta_j s)) ds = b \int_0^1 \cos(\gamma^\pm(\eta_j b s)) ds$ equals

$$\begin{cases} b \left(\frac{1}{6 \cdot 4^{k-1} \eta_j b} \cos \gamma_1^\pm + \left(1 - \frac{1}{6 \cdot 4^{k-1} \eta_j b} \right) \cos \gamma_2^\pm \right) & \text{if } \frac{1}{4} \leq 4^{k-1} \eta_j b < \frac{1}{2}, \\ b \left(\left(1 - \frac{1}{3 \cdot 4^{k-1} \eta_j b} \right) \cos \gamma_1^\pm + \frac{1}{3 \cdot 4^{k-1} \eta_j b} \cos \gamma_2^\pm \right) & \text{if } \frac{1}{2} \leq 4^{k-1} \eta_j b < 1. \end{cases}$$

The minimum occurs when $4^{k-1} \eta_j b = \frac{1}{2}$ and the minimum of $\int_0^b \cos(\gamma^\pm(\eta_j s)) ds$ is $b(\frac{1}{3} \cos(\gamma_1^\pm) + \frac{2}{3} \cos(\gamma_2^\pm))$. The maximum occurs when $4^{k-1} \eta_j b = \frac{1}{4}$ and the maximum of $\int_0^b \cos(\gamma^\pm(\eta_j s)) ds$ is $b(\frac{2}{3} \cos(\gamma_1^\pm) + \frac{1}{3} \cos(\gamma_2^\pm))$. Thus

$$(22) \quad A_I^\pm(b) = b\left(\frac{1}{3} \cos(\gamma_1^\pm) + \frac{2}{3} \cos(\gamma_2^\pm)\right)$$

and

$$(23) \quad A_S^\pm(b) = b\left(\frac{2}{3} \cos(\gamma_1^\pm) + \frac{1}{3} \cos(\gamma_2^\pm)\right).$$

In these examples, we have the same essential limits inferior and superior at \mathcal{O} and yet A_I^\pm and A_S^\pm behave differently. In Example 4, we have the “extreme values” (21); the “effective” contact angles in (a) and (b) of Corollary 3 are γ_2^\pm and in (c) and (d) of Corollary 3 are γ_1^\pm . On the other hand, in Example 5, we have the “intermediate values” (22) and (23). For Example 5, the “effective” contact angles in (a) and (b) of Corollary 3 are σ_2^\pm and in (c) and (d) of Corollary 3 are σ_1^\pm , where $\sigma_1^\pm, \sigma_2^\pm \in [0, \pi]$ satisfy

$$\cos \sigma_1^\pm = \frac{2}{3} \cos \gamma_1^\pm + \frac{1}{3} \cos \gamma_2^\pm \quad \text{and} \quad \cos \sigma_2^\pm = \frac{1}{3} \cos \gamma_1^\pm + \frac{2}{3} \cos \gamma_2^\pm.$$

If f is a bounded solution of (1) satisfying (2) on $\partial^\pm \Omega^* \setminus \{\mathcal{O}\}$ which is discontinuous at \mathcal{O} and $Rf(\theta)$ exists for all $\theta \in (-\alpha, \alpha)$, then bounds on the sizes β^+ and β^- of side fans can be computed using Corollary 3; the lower bounds on the sizes of these side fans differ between these two examples.

4. Comments and extensions

The last section of [Lancaster and Siegel 1996] dealt with extensions of (1) to equations of prescribed mean curvature. Consider the prescribed mean curvature contact angle problem

$$(24) \quad Nf = 2H(\cdot, f) \quad \text{in } \Omega,$$

$$(25) \quad Tf \cdot \nu = \cos \gamma \quad \text{a.e. on } \partial \Omega.$$

Suppose $f \in C^2(\Omega)$ satisfies (24) and (25) and also suppose the following conditions hold:

$$(i) \quad \sup_{x \in \Omega} |f(x)| < \infty \quad \text{and} \quad \sup_{x \in \Omega} |H(x, f(x))| < \infty.$$

$$(ii) \quad H(x, y, t) \text{ is weakly increasing in } t \text{ for each } (x, y) \in \Omega.$$

Using [Echart and Lancaster 2017], we see that Theorems 1 and 2 continue to hold for solutions f as above; the argument is the same as that in [Lancaster and Siegel 1996].

One might ask if the case considered in Theorem 2 is of “physical interest.” Is it possible for the contact angle to fail to have a limit at the corner \mathcal{O} ? In a sense this is a silly question since, at a small enough scale, the macroscopic description of a capillary surface becomes meaningless. On the other hand, one sometimes uses devices (e.g., homogenization) to obtain useful macroscopic information from knowledge of “small scale” properties. An experiment which might be of some interest would be to form a vertical wedge consisting of two planes of glass which have been coated in increasing narrow vertical strips with a nonwetting substance (e.g., paraffin) as the edge at which the two planes meet is approached; this would approximate the situation considered in Theorem 2 and one wonders if there is a “effective” contact angle at the corner which is larger than that for glass and smaller than that for paraffin.

References

- [Echart and Lancaster 2017] A. Echart and K. Lancaster, “On cusp solutions to a prescribed mean curvature equation”, *Pacific J. Math.* **288**:1 (2017), 47–54. MR Zbl
- [Elcrat and Lancaster 1986] A. R. Elcrat and K. E. Lancaster, “Boundary behavior of a nonparametric surface of prescribed mean curvature near a reentrant corner”, *Trans. Amer. Math. Soc.* **297**:2 (1986), 645–650. MR Zbl
- [Entekhabi and Lancaster 2016] M. Entekhabi and K. Lancaster, “Radial limits of bounded nonparametric prescribed mean curvature surfaces”, *Pacific J. Math.* **283**:2 (2016), 341–351. MR Zbl
- [Entekhabi and Lancaster 2017] M. Entekhabi and K. Lancaster, “Radial limits of capillary surfaces at corners”, *Pacific J. Math.* **288**:1 (2017), 55–67. MR Zbl
- [Finn 1986] R. Finn, *Equilibrium capillary surfaces*, Grundlehren Math. Wissenschaften **284**, Springer, New York, 1986. MR Zbl
- [Giusti 1980] E. Giusti, “Generalized solutions for the mean curvature equation”, *Pacific J. Math.* **88**:2 (1980), 297–321. MR Zbl
- [Giusti 1984] E. Giusti, *Minimal surfaces and functions of bounded variation*, Monographs in Mathematics **80**, Birkhäuser, Boston, 1984. MR Zbl
- [Heinz 1970] E. Heinz, “Über das Randverhalten quasilinearer elliptischer Systeme mit isothermen Parametern”, *Math. Z.* **113** (1970), 99–105. MR Zbl
- [Lancaster 1985] K. E. Lancaster, “Boundary behavior of a nonparametric minimal surface in \mathbb{R}^3 at a nonconvex point”, *Analysis* **5**:1-2 (1985), 61–69. MR Zbl
- [Lancaster 2010] K. E. Lancaster, “A proof of the Concus–Finn conjecture”, *Pacific J. Math.* **247**:1 (2010), 75–108. MR Zbl
- [Lancaster 2012] K. E. Lancaster, “Remarks on the behavior of nonparametric capillary surfaces at corners”, *Pacific J. Math.* **258**:2 (2012), 369–392. MR Zbl

- [Lancaster and Siegel 1996] K. E. Lancaster and D. Siegel, “Existence and behavior of the radial limits of a bounded capillary surface at a corner”, *Pacific J. Math.* **176**:1 (1996), 165–194. Correction in **179**:2 (1997), 397–402. MR Zbl
- [Miranda 1977] M. Miranda, “Superficie minime illimitate”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **4**:2 (1977), 313–322. MR Zbl
- [Simon 1980] L. Simon, “Regularity of capillary surfaces over domains with corners”, *Pacific J. Math.* **88**:2 (1980), 363–377. MR Zbl
- [Tam 1984] L.-F. Tam, *The behavior of capillary surfaces as gravity tends to zero*, Ph.D. thesis, Stanford University, 1984. Zbl
- [Tam 1986] L.-F. Tam, “Regularity of capillary surfaces over domains with corners: borderline case”, *Pacific J. Math.* **124**:2 (1986), 469–482. MR Zbl

Received January 17, 2017. Revised June 19, 2017.

JULIE N. CRENSHAW
 DEPARTMENT OF MATHEMATICS, STATISTICS AND PHYSICS
 WICHITA STATE UNIVERSITY
 WICHITA, KS
 UNITED STATES
 crenshaw@math.wichita.edu

ALEXANDRA K. ECHART
 DEPARTMENT OF MATHEMATICS, STATISTICS AND PHYSICS
 WICHITA STATE UNIVERSITY
 WICHITA, KS
 UNITED STATES
 echart@math.wichita.edu

KIRK E. LANCASTER
 WICHITA, KANSAS
 WICHITA, KS
 UNITED STATES
 redwoodsrunner@gmail.com

NORMS IN CENTRAL SIMPLE ALGEBRAS

DANIEL GOLDSTEIN AND MURRAY SCHACHER

To Robert Steinberg, a cherished teacher, colleague, and friend

Let A be a central simple algebra over a number field K . We study the question of which integers of K are reduced norms of integers of A . We prove that if K contains an integer that is the reduced norm of an element of A but not the reduced norm of an integer of A , then A is a totally definite quaternion algebra over a totally real field (i.e., A fails the *Eichler condition*).

1. Introduction

Let A be a central simple algebra over a number field K . Write $\text{Norm}(\cdot)$ for the reduced norm from A to K . If x is an integer in A , then clearly $\text{Norm}(x)$ lies in R , the ring of integers of K . It is also clear that x must be positive at the real primes of K at which A is ramified. Suppose that $m \in R$ satisfies this property, and so m is a norm from A (see Theorem 2). If m is not the reduced norm of an integer of A , we call m an *outlier for A* (this terminology is not standard).

The main result of this paper (combining Theorem A and Lemmas 6 and 7) is that if K contains an outlier for A , then K is totally real, A is a quaternion algebra over K , and A is totally definite. (One says in this case that A fails the Eichler condition).

We also prove a theorem of Deligne in Section 8 (because we couldn't find a proof in the literature), which states that if $n \geq 2$ is an integer, and E_1, \dots, E_n and F_1, \dots, F_n are supersingular elliptic curves defined over an algebraic closure of the finite field $GF(p)$, the field of p elements, then

$$E_1 \times \cdots \times E_n \cong F_1 \times \cdots \times F_n.$$

The main ingredient is Eichler's theorem on the uniqueness of a maximal order in a central simple algebra in which Eichler's condition holds. We also exploit the known fact that the endomorphism algebra of such an E_i is a maximal order in the quaternion algebra A_p over the rational field \mathbb{Q} ramified at p and ∞ and unramified everywhere else (and every maximal order arises in this context). Using this connection also allows one to interpret outliers in \mathbb{Q} for A_p as positive integers

MSC2010: primary 11S45; secondary 11S15.

Keywords: central simple algebras, reduced norms, super singular elliptic curves.

m for which no supersingular elliptic curve defined over the algebraic closure of $GF(p)$ has an endomorphism of degree m .

2. Notation and terminology

Throughout this paper, K is a number field, R its ring of integers, and A is a *central simple algebra* over K . By definition, A is a finite-dimensional algebra over K , the center of A is equal to K , and A has no nonzero 2-sided ideals. Equivalently, $A \otimes_K \bar{K}$ is isomorphic to the matrix algebra $M_n(\bar{K})$, where \bar{K} denotes an algebraic closure of K . For basic facts about central simple algebras, see [Pierce 1982].

The positive integer n is the *degree* of A . A central division algebra D is a central simple algebra, as is $M_k(D)$ for any k , and conversely every central simple algebra over K is of this form by Wedderburn's Theorem [Weil 1967, Chapter IX, §1, Proposition 2].

A division algebra of degree $n = 2$ is called a *quaternion algebra*.

If L is a field extension of K , then $A \otimes_K L$ is a central simple algebra over L . If $A \otimes_K L$ is isomorphic to $M_n(L)$ then L is said to *split* A .

Let M denote the set of places of the number field K . For each place $v \in M$, $A_v := A \otimes_K K_v$ is a central simple algebra over the completion K_v . By Wedderburn's theorem, it is a ring of matrices over a local division ring D_v central over K_v . We set $n_v = \text{degree}(D_v)$; n_v is called the *local degree*. A is said to be *split* at v if K_v splits A ($n_v = 1$); otherwise it is *ramified* at v ($n_v > 1$). A key fact is that a central simple algebra over K splits at all but finitely many places v of K .

We have the following splitting criterion:

Lemma 1 [Reiner 1975]. *Let A be a central simple algebra over the number field K . A finite extension L of K splits A if and only if, for each place v of K and for each extension w of v to L , the local dimension $[L_w : K_v]$ is a multiple of the local degree n_v .*

Note that to determine, using Lemma 1, whether a given finite extension L over K splits A , it is enough to check the stated condition at the finite set of places v of K where A is ramified.

The notion of reduced norms in a central simple algebra A is bound up with the two notions of subfields and splitting fields. A field extension L of K is a subfield of A if L embeds in A ; a maximal subfield of A is a maximal such. All maximal subfields of A have dimension $n = \text{degree}(A)$ over K . A maximal subfield of A is a splitting field for A , and conversely every n -dimensional splitting field for A embeds in A as a maximal subfield [Reiner 1975, Chapter 1, Section 7]. When A is a quaternion algebra, this translates as: maximal subfields of A are quadratic over K , and quadratic splitting fields of A embed in A . We will use this association later.

If A is a central simple algebra over K , L a maximal subfield of A , and $x \in L$, then the *reduced norm* $\text{Norm}(x)$ is the ordinary field norm from L to K . This notion is independent of the choice of L , or of the embedding of L into A . By norm we will always mean reduced norm, and the notation will be $\text{Norm}(x)$. In particular, for $a \in K$, $\text{Norm}(a) = a^n$. The usual property holds: $\text{Norm}(xy) = \text{Norm}(x) \text{Norm}(y)$ for $x, y \in A$, whether or not x and y commute. It follows that $\text{Norm}(ax) = a^n \text{Norm}(x)$ for $a \in K$.

An element $a \in A$ is an *integer* if the monic irreducible polynomial of a over K has coefficients in R . Sums and products of commuting integers are integers, but, as we shall see later, products of integers need not be integers.

Suppose A is central simple over K of degree n . Which elements of K are reduced norms of elements of A ? The answer is given by the theorem of Hasse, Maass and Schilling; see [Reiner 1975, p. 289]:

Theorem 2 (Hasse–Maass–Schilling). *An element m of K is a reduced norm of an element of A if and only if m is positive at every real place of K at which A is ramified.*

For convenience we will call this the HMS theorem. Note that there is no condition at the complex places of K , at the finite places of K , or at the real places of K where A does not ramify.

Suppose $m \in R$ and m is a norm in A . It need not happen that m is the norm of an integer of A . We will call $m \in R$ an *outlier* if m is a norm in A but not the norm of an integer. Equivalently, m is not the norm of an element of any maximal order. We will be concerned with the existence of, and properties of, outliers.

If K is a number field, we say K is *totally real* if K_v is real at all the infinite places v of K . If K is totally real, m in K is *totally positive* if the real number m_v is > 0 at all the infinite places v of K . The m in K for which $m_v > 0$ at the real places of A that ramify are, by Theorem 2, the reduced norms of elements of A , and conversely.

We recast the identification of outliers in terms of Lemma 1. Suppose A is central simple over K of degree n , R the ring of integers of K .

Lemma 3. *Suppose $m \in R$ is a norm in A . Then m is **not** an outlier if and only if there is a monic irreducible polynomial $f(t) \in R[t]$ such that*

- (1) *We have $f(0) = (-1)^n m$.*
- (2) *For each place v of K , let $f(t) = \prod f_i(t)$ be the factorization of $f(t)$ into irreducible monic factors in $K_v[t]$. Then each $d_i = \text{degree}(f_i)$ is a multiple of the local degree $n_v(A)$.*

Proof. Let $L = K(\alpha)$ be the root field of f . Then $[L : K] = n$ since f is irreducible, and α is an integer since $f \in R[t]$ is monic. The first condition says that the norm of

α is m . The second condition, by Lemma 1, says that L splits A , and so L embeds in A since its dimension is n . Then the reduced norm of α is m . The other direction of Lemma 3 is clear. \square

Corollary 4. *Suppose A and B are central simple algebras over K of the same degree n , and that the local degree $n_v(A)$ divides the local degree $n_v(B)$ for all places v of K . Then any outlier m of A is a priori also an outlier of B . In particular, if B has no outliers then A has no outliers.*

Proof. The polynomial requirements of Lemma 3 for B are more restrictive than those for A . \square

In Section 3 we review maximal orders in the central simple algebra A and recall Eichler's condition. In Section 4 we prove that when Eichler's condition is satisfied, then there are no outliers. In other words, if A has outliers then A is a quaternion algebra over a totally real number field K , and all real places of K are ramified in A . However, this condition is sufficient but not necessary: there are definite quaternion algebras over totally real number fields that have no outliers. We remark that there is no logical relation between having outliers and having a unique (up to conjugacy) maximal order; neither condition implies the other. In Section 5 we study quaternion algebras over the field of rational numbers \mathbb{Q} . We particularly study definite quaternion algebras ramified at a single finite prime. We write A_r for the definite quaternion algebra over \mathbb{Q} unramified away from the places ∞ and r . We show, for example, that if A_r has an outlier then it has an outlier less than an explicit bound $(r^2/16)$.

We give heuristic evidence that for infinitely many r , A_r has no outliers, as well as examples when chosen square-free integers are outliers.

We are grateful to Joel Rosenberg for many discussions about the contents of this paper, and for posing the questions which started us on this research.

3. Maximal orders

Let K be a number field, R its ring of integers, and A a central simple algebra over K . A subring O of R that contains 1, is finitely generated as an R module, and that contains a basis of A over K is called an *order* of A . Any order O of A is a projective R -module of rank equal to n , the degree of A over K . A *maximal order* of A is an order which is maximal with respect to containment. Maximal orders are isomorphic if and only if they are conjugate, so we will speak of conjugacy classes of maximal orders. All elements of a maximal order are *integral* over R , and every integral element of A is contained in some maximal order.

It is known that the number of maximal orders of A , up to conjugacy by an element of A , is finite. Let $\{O_1, \dots, O_t\}$ be a set of representatives.

For any given maximal order O of A , let $I(O)$ be the group of two-sided fractional ideals of O modulo principal two-sided fractional ideals. Set $i(O) = |O(I)|$. It is known that each $i(O)$ is finite although the cardinalities $i(O_1), \dots, i(O_t)$ may be distinct. Their sum $c := i(O_1) + \dots + i(O_t)$ turns out to be equal to the number of fractional left ideals of O modulo principal fractional left ideals for any maximal order O .

The terminology is that t is called the type number and c is called the class number. We have just seen that the type number is at most the class number.

Let A be a CSA over a number field K . Consider the following three conditions:

- (1) A is a quaternion algebra.
- (2) The field K is totally real.
- (3) A is ramified at every infinite place of K .

It is customary to say that A *fails* the Eichler condition when all three conditions hold. For example, the quaternion algebra A_p over \mathbb{Q} ramified at p and ∞ , and unramified away from those places, fails the Eichler condition.

This description can be refined if the Eichler condition holds. Assume now that A satisfies the Eichler condition. Then the $i(O_1), \dots, i(O_t)$ are all equal. In fact each $I(O_i)$ can be identified with an abelian group $I = I(A)$, as do the types T and the classes C . These three abelian groups fit into an exact sequence

$$0 \rightarrow T \rightarrow C \rightarrow I \rightarrow 0.$$

These three groups are related, via the reduced norm map to certain generalizations of the class group of the center K of A , by results of Eichler.

The group C is isomorphic to the group C' of fractional ideals of K modulo principal fractional ideals that can be generated by an invertible element $a \in K$ that is positive at all infinite places of K that ramify in A .

Let n be as usual the square root of $\dim_K(A)$. If \mathfrak{p} is a prime ideal of K that is ramified in A , then at the corresponding finite place v of K , $A \otimes K_v = M_r D'$ for some division algebra D' over K_v and for some r dividing n . The group T is isomorphic to the subgroup T' of C generated by nC and the class of \mathfrak{p}^r for each finite prime \mathfrak{p} (and note that this gives nothing new for the unramified primes since $r = n$).

I is isomorphic to the (abelian, finite) quotient group C'/T' .

4. Higher degree central simple algebras

Let A be a central simple algebra of degree n over the number field K . The main result of this section is:

Theorem A. *If $n > 2$ then A has no outliers.*

We need first a review of the proof of the HMS theorem in order to build a variant that works for integers. A first ingredient is:

Krasner's Lemma. *Let v be a place of K , and $f(t) = t^n + a_1 t^{n-1} + \cdots + a_n$ a separable irreducible polynomial in $K_v[t]$. If $g(t) \in K_v[t]$ is close enough to $f(t)$, then g is separable irreducible and $K_v[a] = K_v[b]$, where a is a root of $f(t)$ and b is a root of $g(t)$.*

Eichler's proof of the HMS theorem goes as follows. Let R be the integers of K , and $m \in R$ satisfy the required condition: m is positive at all places v of K which are real and ramified in A . Let S be the set of infinite places of K at which A ramifies. Let S' be a finite set of finite primes of K , including those that ramify in A . We insist that S' be nonempty; if necessary, we include an irrelevant extra prime where A is unramified but where the polynomial constructed below is irreducible. We construct a polynomial

$$(5) \quad f(t) = t^n + c_1 t^{n-1} + \cdots + (-1)^n m \in K[t]$$

so that:

- For each $v \in S'$, c_i is close enough to an irreducible polynomial $f_v(t) = t^n + a_1 t^{n-1} + \cdots + (-1)^n m \in R_v[t]$ to guarantee f is irreducible in $K_v[t]$. There is such a polynomial [Weil 1967, XI, §3, Lemma 2] but we don't show that here.
- For each $v \in S$, f is close to $f_v(t) = t^n + (-1)^n m$, i.e., each c_i is positive and close to 0. (Note that if any such v exists, then n is necessarily even). This guarantees f_v has no real roots. If A is not ramified at any infinite place of K , then this condition is vacuous.

Since S' is nonempty, f is irreducible in $K[t]$. Let $L = K(\alpha)$ where α is a root of f ; $[L : K] = n$. The first condition on f says that L splits A at the finite primes, and the second condition guarantees that L splits A at the ramified infinite places, since the root field of f must be complex. The sign $(-1)^n$ guarantees that the norm from L to K of α is m . Finally, since L is a splitting field of degree n , then L embeds in A as a maximal subfield, and the reduced norm of α is m .

This is the proof rendered by Eichler, and is the one presented in [Reiner 1975], [Vignéras 1980], and [Weil 1967]. Note that it made crucial use of the weak approximation theorem.

To go further, we use the strong approximation theorem [Weil 1967, Corollary 2, page 70], which better suits our purposes. Let w be a place of K at which A is unramified. Then we can insist that the c_i are in R_v for all $v \neq w$. We call this the strong proof of the HMS theorem. We conclude any $m \in R$ which is positive at all real places of K that ramify in A is the reduced norm of an element α of A that is

integral at all places v of K not equal to w . So if K has a complex place, or a real place that is not ramified in A , then, taking this for w shows that A has no outliers.

Lemma 6. *If A has an outlier then, K is totally real, A is totally definite, i.e., A is ramified at all the real infinite places of K .*

Proof. Let $m \in R$ be a norm in A . If the conditions are not satisfied, then A must have an infinite place w at which A is unramified. We use this extra place in the strong proof of the HMS theorem. Then the polynomial f is in $R[t]$, α is an integer, and m is the norm of an integer. \square

Lemma 7. *If A has an outlier then there is a finite place of K that ramifies in A .*

Proof. Suppose A is unramified at all finite places. By Lemma 6, we may assume n is even. Let $m \in R$ be totally positive. The polynomial $t^n + m$ does the trick in the strong proof of HMS. \square

We finish the proof of Theorem A. By Lemma 6 we may assume that K is totally real, A is ramified at all real places, $n > 2$ is even, and m is positive at all infinite places. First we treat the finite places. By [Weil 1967, Ch. XI, §3, Lemma 2], for each finite place v , for any n , and for any nonzero m in K there exists a monic degree- n irreducible polynomial $f(t) \in K_v[t]$ with coefficients in R_v such that $f(0) = (-1)^n m$.

Let M_f denote the set of finite places of K that ramify in A . Note that M_f is finite, and for each v in M_f we have $f_v(t)$ as required by Lemma 3, but we have not yet treated the infinite places.

For each $1 \leq k \leq n$, apply the Chinese remainder theorem to the coefficient of t^k in f_v to get a monic polynomial $g(t) \in K[t]$ with $g(0) = m$ and integral coefficients so that each localization $g_v(t)$ at each K_v is close enough to f_v to be irreducible by Krasner's Lemma. We have lifted the required polynomials at the finite primes, but the infinite places are still at bay; there is yet no reason why $g(t)$ has only complex embeddings.

Each $v \in M_f$ lies over some rational prime p_v . Let $N = \prod_{v \in M_f} p_v$ be their product.

Let M_{inf} be the set of real places of K that ramify in A . For any $v \in M_{\text{inf}}$ we have a real polynomial g_v which is positive at $-\infty$, ∞ and 0 by construction. Therefore, there is some integer multiple M_v of N so that $g_v(t) + M_v t^2$ is positive everywhere. Let M be the largest of the M_v . Furthermore by replacing M by $N^k M$ for a sufficiently large k , we can insure by Krasner's Lemma again that $g_v(t) + N^k M t^2$ is irreducible at each $v \in M_f$.

The polynomial $f(t) = g(t) + N^k M t^2 \in K[t]$ does the trick: it is monic of degree n , has no real roots, and for each place of K that ramifies in A , each irreducible factor of f_v has degree a multiple of $n_v(A)$. This finishes the proof of Theorem A. \square

Note that the coefficient of t^2 was available for modification only because $n > 2$. For quaternion algebras, the coefficient of t^2 is constant equal to 1. We get to that case next.

5. Quaternion algebras

We write \mathbb{Q} for the field of rational numbers and \mathbb{Z} for the ring of integers.

We consider definite quaternion algebras over \mathbb{Q} with special attention to A_r = the definite quaternion algebra ramified at the prime r and unramified at all other finite primes. Of course A_r is also ramified at ∞ , and so at all infinite places. The simplification here is that the integers which are norms in A_r are exactly the set of positive integers, and so the only issue is whether they are norms of integers. We now investigate how this could happen.

Let m be a positive integer. Let $f(t) = t^2 + bt + m$ with $b \in \mathbb{Z}$. Let $L = \mathbb{Q}(\alpha)$ with $f(\alpha) = 0$. Then L splits A_r if and only if:

- f is r -adically irreducible,
- f has degree 2 at ∞ , i.e., $d = b^2 - 4m < 0$.

When either of the conditions above hold, then f is irreducible and $[L : \mathbb{Q}] = 2$. When they both hold, L embeds in A_r by Lemma 1, $\text{Norm}(\alpha) = m$, and so m is the norm of an integer in A_r . Moreover, m is the norm of an integer if and only if this search succeeds for some $b \in \mathbb{Z}$. There are a finite number of eligible b by the last condition; $|b| < \sqrt{4m}$. Furthermore, b can be assumed to be positive; if α is a root of $t^2 + bt + m$ then $-\alpha$ is a root of $t^2 - bt + m$. Of course $b = 0$ is legitimate as a possibility. We record this in:

Lemma 8. *The positive integer m is the norm of an integer in A_r if and only if there is a polynomial $f(t) = t^2 + bt + m$ satisfying the two conditions above for some $b \in \mathbb{Z}$. It is sufficient to search only in the range $0 \leq b < \sqrt{4m}$.*

For polynomials of the right shape, they are irreducible r -adically if and only if they are irreducible mod r . So when is $m = 2$ an outlier in A_r ? We illustrate the search below, where we assume $r > 2$:

$$\begin{array}{ll}
 b = 0 & d = -8 \\
 (9) \quad b = 1 & d = -7 \\
 b = 2 & d = -4
 \end{array}$$

Of course -8 is an r -adic square if and only if -2 is, and this happens if and only if the Legendre symbol $\left(\frac{-2}{r}\right) = 1$. Similarly, -4 is a square if and only if -1 is. For each of the three conditions in (9), a random prime r satisfies it with probability $\frac{1}{2}$. We conclude:

Theorem 10. *The integer 2 is an outlier in A_r if and only if*

$$(11) \quad \left(\frac{-2}{r}\right) = \left(\frac{-7}{r}\right) = \left(\frac{-1}{r}\right) = 1.$$

By considering the value of $r \bmod 56$, it follows from the Dirichlet density theorem [Serre 1973, Chapter VI, §4, Theorem 2] that the set of primes r for which this holds has density $\frac{1}{8}$. In particular it is infinite.

We do this once more to determine when 3 is an outlier. The data gives the following list:

$$(12) \quad \begin{array}{ll} b = 0 & d = -12 \\ b = 1 & d = -11 \\ b = 2 & d = -8 \\ b = 3 & d = -3 \end{array}$$

There is a redundancy; -12 is a square if and only if -3 is. We conclude, for $r > 3$:

Theorem 13. *The integer 3 is an outlier in A_r if and only if*

$$(14) \quad \left(\frac{-3}{r}\right) = \left(\frac{-11}{r}\right) = \left(\frac{-2}{r}\right) = 1.$$

The set of primes r for which this holds is infinite and has density $\frac{1}{8}$.

By similar analysis we get, for $r > 6$:

Theorem 15. *The integer 6 is an outlier in A_r if and only if*

$$(16) \quad \left(\frac{-2}{r}\right) = \left(\frac{-3}{r}\right) = \left(\frac{-5}{r}\right) = \left(\frac{-23}{r}\right) = 1$$

The set of primes r for which this holds is infinite and has density $\frac{1}{16}$

Suppose 6 is an outlier for A_r . It does not follow that 2 and 3 are outliers. There may be integral α and β with $\text{Norm}(\alpha) = 2$ and $\text{Norm}(\beta) = 3$, and then $\text{Norm}(\alpha \cdot \beta) = 6$. It might happen that for all such occurrences α and β are in different maximal orders, and $\alpha \cdot \beta$ is not integral. When 6 is minimal as an outlier, this is what had to happen. This can be quantified; we state without proof:

Theorem 17. *A_r has the property that 2 and 3 are not outliers and 6 is an outlier if and only if $-2, -3, -5, -23$ are squares mod r and either*

$$(18) \quad \left(\frac{-1}{r}\right) = -1 \quad \text{or} \quad -1 \text{ is a square mod } r \text{ and } 11, 7 \text{ are nonsquares mod } r$$

The set of primes r for which this holds is infinite and has density $\frac{5}{128} = \left(\frac{1}{16}\right)\left(\frac{1}{2} + \frac{1}{8}\right)$.

We have not yet determined all outliers in A_r , nor have we answered whether they are infinite when nonempty. We need two results to prepare for this. We take on the second issue first. Since the next result holds more generally than for the A_r , we state it in full generality. In all of the following, the symbol (a, b) stands for the quaternion algebra over some ground field generated by i and j where $i^2 = a$, $j^2 = b$, $ij = -ji$.

Theorem 19. *Let A be a definite quaternion algebra over \mathbb{Q} ramified at the finite prime r . If m is a positive integer, then m is an outlier for A if and only if mr^2 is also.*

Proof. For the easy direction: if $\text{Norm}(\alpha) = m$ with α an integral element of A , then $\text{Norm}(r \cdot \alpha) = mr^2$. We need to show conversely that when mr^2 is the norm of an integer, so is m .

Let O be any maximal order of A . It is enough to show that whenever mr^2 is a norm of an element α of O , then $\alpha/r \in O$. The completion of O at r is the norm form of the unique quaternion algebra D over \mathbb{Q}_r . By [Serre 1979], D has the form (a, r) where a is an appropriate nonresidue mod r . When r is odd, any nonresidue will do, whereas when $r = 2$, $a = -3$ will do (in all cases, \sqrt{a} determines the unique unramified quadratic extension). The norm form for this algebra is:

$$(20) \quad F = x^2 - ay^2 - r(z^2 - aw^2)$$

Assume that $F(x, y, z, w) = mr^2$. It follows that $x^2 - ay^2 \equiv 0 \pmod{r}$. As a is a nonresidue, this forces x and y to be $\equiv 0 \pmod{r}$. But then $x^2 - ay^2 \equiv 0 \pmod{r^2}$, and so $r(z^2 - aw^2)$ is $0 \pmod{r^2}$. It follows that $z^2 - aw^2 \equiv 0 \pmod{r}$, so that z and w are $0 \pmod{r}$. Now all four coefficients x, y, z, w of α are divisible by r . Thus α/r is in O and has norm m . We conclude that whenever $\text{Norm}(\alpha)$ is mr^2 with α in O , then α/r is in O and has norm m . Since this holds for all maximal orders, the lemma is established. \square

Corollary 21. *With A as in Theorem 19, if the set of outliers for A is nonempty, then it is infinite; if m is an outlier for A , then so is mr^{2n} for any positive integer n .*

Remark 22. Corollary 21 allows division by r^2 , but not by r . In fact, if m is an outlier for A_r and relatively prime to r , then mr is not an outlier. The polynomial $t^2 + mr$ is irreducible at r by Eisenstein's criterion, and also irreducible at infinity; it satisfies the requirements of Lemma 3.

We need a bound up to which we can check for outliers not governed by Theorem 19. We do this for A_r ; the generalizations to definite quaternion algebras will be clear. One more preliminary is necessary.

Lemma 23. *Let $p > 2$ be a prime and m in $GF(p)$ nonzero. Then there exists b in $GF(p)$ such that $b^2 - 4m$ is a nonsquare mod p .*

Proof. Suppose not. Then, for every b , $b^2 - 4m$ is a square. But then $(b^2 - 4m) - 4m$ is a square and by induction $b^2 - 4mj$ is a square for all j . By our hypotheses $4m$ is invertible in $GF(p)$ so all elements of $GF(p)$ are squares, contradiction. \square

We can now establish a bound for A_r .

Theorem 24. *Suppose $r > 2$ is prime and m is a positive integer coprime to r . Set $C(r) = r^2/16$. If $m > C(r)$, then m is not an outlier for A_r .*

Proof. By Lemma 23, we choose an integer b such that $b^2 - 4m$ is a nonsquare mod r . We are free to assume of course that $b < r/2$. Set $f = t^2 + bt + m$, and $d = b^2 - 4m$. One checks that the bounds on b and m say that $d < 0$, so f is irreducible at infinity. Since d is a nonresidue at r , f is also irreducible in $\mathbb{Q}_r[t]$. Then f satisfies the requirements of Lemma 3, and so m is the norm of an integer. \square

Remark 25. Theorem 24 gives an effective strategy for finding all outliers in A_r . One checks all m in the interval $[0, C(r)]$ using Lemma 3. For $m > C(r)$: m is not an outlier if m is not divisible by r . If m is divisible by r to the first power, then m is not an outlier by Remark 22. If m is divisible by higher powers of r , then successive uses of Theorem 19 gets us to the case of first power or the range $[0, C(r)]$.

Here is one case where all outliers can be determined:

Corollary 26. *If $r = 67$, then the only outliers for A_r are of form $3 \cdot r^{2n}$, $n = 1, 2, 3, \dots$*

Proof. One checks in the range $[0, C(r)]$ that the only outlier is 3 using Lemma 3 for each possible m . Then Remark 25 does the rest. \square

Remark 27. Note that this corollary says that division by r^2 is not always possible when r is not a ramified prime. In A_{67} , $12 = 3 \cdot 2^2$ is the norm of an integer, but 3 is not, so division by the square of the unramified prime 2 is not possible.

An effective bound for more general definite quaternion algebras is not difficult. Suppose A is a quaternion algebra central over \mathbb{Q} ramified at infinity and the finite primes comprising a set S . Let C be the product of the finite ramified primes of A , and $M = C^2/16$. Then

Theorem 28. *M is an effective bound for determining all the outliers for A .*

Proof. The proof is exactly as in Theorem 24 and Remark 25. \square

The symbol $B = (-58, -17)$ over \mathbb{Q} is ramified at infinity and the finite primes $S = \{2, 17, 29\}$. Using Theorem 28 one can show:

Corollary 29. *The outliers for B are the set*

$$\{10r^{2n} : n = 1, 2, 3, \dots; r \text{ a product of elements of } S\}.$$

The minimal outlier of B is 10. Therefore, there are integers α and β in B with $\text{Norm}(\alpha) = 2$ and $\text{Norm}(\beta) = 5$. Whenever this happens, the product $\alpha\beta$ is not integral.

The appearance of 6 and 10 in this context is general, as seen in the next theorem.

Theorem 30. *Let m be a positive integer that is not a square. Then there are infinitely many primes r such that $\{m \cdot r^{2n} : n \geq 1\}$ are outliers for A_r .*

Proof. For b in the range $[0, C]$, $C = \sqrt{4m}$, and $d = b^2 - 4m$, we must have the Legendre symbol $\left(\frac{d}{r}\right)$ equal to 1; the Chebotarev density theorem says there are infinitely many such primes r . In fact their density is $1/2^s$ for some appropriate integer s . \square

6. Open questions

We begin with:

Are there infinitely many rational primes r such that A_r has no outliers?

Heuristically, the answer is yes. Computer searches for small bounds show that A_r has no outliers a little more than half the time.

We have seen that the set of primes r for which $m = 2$ is an outlier for A_r has density $\frac{1}{8}$. Similarly, the set of primes for which $m = 3$ is an outlier for A_r has density $\frac{1}{8}$. Adding together these probabilities for small m appears to give something like density 0.7; this is roughly the probability that neither 2 nor 3 is an outlier. However for large m , the density of primes r for which m is an outlier in A_r should be something like $2^{-c\sqrt{4m}}$, since we are asking that the floor of $\sqrt{4m} + 1$ numbers are all squares r -adically and some constant c is required because these numbers may not be linearly independent in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$. However the sum $\sum_{m>0} 2^{-c\sqrt{4m}}$ converges. Therefore we cannot distinguish whether our set is finite or infinite.

Another interesting question concerns totally definite quaternion algebras over totally real number fields. Do they have outliers? Sometimes? Often?

We have worked out only one example. Let $B = (-1, -7)_K$ where K is the real subfield of seventh roots of unity. Then B has no outliers. The argument is technical, so we will not reproduce it here; it requires a detailed study of units, totally positive units, class number, and the establishment of a bound as in Theorem 24; the bound is 1792. However, when $K = \mathbb{Q}(\sqrt{2})$, the same algebra tensored up to K does have outliers. Thus, restriction maps may or may not preserve the property of having no outliers.

On the other hand, let A be the algebra $(-1, -67)$ over \mathbb{Q} ; by Corollary 26, 3 is an outlier for A . If $K = \mathbb{Q}(\sqrt{67})$, then $A \otimes_{\mathbb{Q}} K$ is ramified at only the infinite places of K , and so by Lemma 7 has no outliers. Thus the restriction map may also fail to preserve the property of having outliers.

The last remark can be generalized. From Lemma 7, if A is a quaternion algebra over \mathbb{Q} , then there is a real quadratic field K such that:

- $A \otimes_{\mathbb{Q}} K$ is a division ring,
- $A \otimes_{\mathbb{Q}} K$ has no outliers.

7. Application to supersingular elliptic curves and surfaces

We review the connection between supersingular elliptic curves in characteristic r and maximal orders in A_r , where A_r is the definite quaternion algebra ramified at ∞ and r and unramified away from these places.

Let E be a supersingular elliptic curve defined over an algebraic closure Ω of $GF(r)$, and write $\text{End}(E)$ for its endomorphism ring. Then $\text{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ is isomorphic to A_r . Under this isomorphism, $\text{End}(E)$ is a maximal order in A_r , and, conversely, any maximal order M of A_r is isomorphic to $\text{End}(E)$ for some E .

Furthermore the norm of an endomorphism $\phi : E \rightarrow E$ is, under this isomorphism, equal to the reduced norm of the corresponding $m \in M$.

The statement “ m is an outlier for A_r ” translates to: no supersingular elliptic curve defined over Ω has an endomorphism of degree m .

So we see, for example, that for every integer $m > 1$ there are infinitely many primes p such that no supersingular elliptic curve defined over Ω has an endomorphism of degree m .

Next we turn to products of supersingular elliptic curves.

Corollary 31. *Let E be a supersingular elliptic curve defined over an algebraic closure of $GF(p)$, and set $A = E^g$ for $g \geq 2$ an integer. Then the abelian variety A has an endomorphism of degree m for every positive integer m .*

Remark 32. Here we are considering all endomorphisms of A , not just those that preserve the obvious principal polarization.

Proof 1. $\text{End}(A)$ (which happens to equal $S = \text{Mat}_g(M)$) is a maximal order in the central simple algebra $\text{Mat}_g(A_r)$ of dimension $4g^2$ over \mathbb{Q} . Eichler’s methods, as outlined in Section 3, imply that S is the unique maximal order up to conjugacy since $g > 1$. Therefore, by Theorem A, m is a reduced norm of an element $\alpha \in \text{End}(S)$. However, reduced norm is in this case equal to the degree of the map α . \square

Our second proof of Corollary 31 uses a well-known theorem of Deligne, which we state below. As we have not found an adequate proof in the literature, for the reader’s convenience we include one in the next section.

Theorem 33 (Deligne). *Let p be a prime and let $n \geq 2$ be an integer. If E_1, \dots, E_n and F_1, \dots, F_n are supersingular elliptic curves defined over an algebraic closure of $GF(p)$, then*

$$E_1 \times \cdots \times E_n \cong F_1 \times \cdots \times F_n$$

Proof 2 of Corollary 31. It turns out it is enough, using Deligne's result, to show that for each rational prime ℓ that $A = E^n$ has an endomorphism of degree ℓ .

There does exist for each ℓ an isogeny of supersingular elliptic curves $\phi : E \rightarrow E'$ of degree ℓ (for $\ell = p$, the Frobenius has degree p and for $\ell \neq p$ mod out by any subgroup H of order ℓ). However, Deligne's theorem gives an isomorphism

$$\psi : E^n \cong E' \times E^{n-1}.$$

Thus the composite $\psi^{-1} \circ (\phi \times \text{id}^{n-1})$ furnishes the desired endomorphism of degree ℓ . \square

We are grateful to Bruce Jordan for suggesting the second proof of Corollary 31.

8. Proof of Deligne's theorem

In this section we prove Theorem 33. For p a prime, let Ω denote an algebraic closure of $GF(p)$.

Remark 34. (1) It would suffice by induction to prove the theorem for $n = 2$ (although we will not use this remark).

(2) It will suffice to show (by transitivity of isomorphism) that $F_1 \times \cdots \times F_n \cong E^n$ for some particular supersingular elliptic curve E defined over Ω .

The remainder of this section is devoted to the proof of Theorem 33.

Note that $\Delta = \text{End}(E)$ is a maximal order in the quaternion algebra $A_p = \Delta \otimes_{\mathbb{Z}} \mathbb{Q}$. The left Δ -module $\text{Hom}(F_1 \times \cdots \times F_n, E)$ being a projective module of rank $n \geq 2$ is free by [Reiner 1975][Corollary 35.11 (iv)] (by the results of Section 3 since the Eichler condition holds for $M_n(\Delta)$, and since visibly any ray class field over \mathbb{Q} is trivial). This is the key point in the proof.

Let ϕ_1, \dots, ϕ_n be a basis. The freeness means that any homomorphism ψ from $F_1 \times \cdots \times F_n$ to E is uniquely a sum

$$\psi = \delta_1 \circ \phi_1 + \cdots + \delta_n \circ \phi_n$$

for some $\delta_1, \dots, \delta_n$ in Δ , noting that the Δ -action on $\text{Hom}(A, E)$, is composition of functions. Setting $\Phi = (\phi_1, \dots, \phi_n)$, we have constructed a homomorphism

$$\Phi : F_1 \times \cdots \times F_n \rightarrow E^n,$$

and to finish the proof of the theorem it will suffice to prove that Φ is an isomorphism.

Let K be the kernel of Φ . If Φ is not an isomorphism, then K is nontrivial, and therefore some projection $\pi_i(K)$ is nontrivial in F_i . Let $\rho : F_i \rightarrow E$ be a homomorphism, and set $\psi : F_1 \times \cdots \times F_n \rightarrow E$ to be $\psi(x_1, \dots, x_n) = \rho(x_i)$. It follows that ρ and therefore any homomorphism from F_i to E must kill $\pi_i(K)$.

Lemma 35. *There is a supersingular elliptic curve E_0 defined over $GF(p)$.*

Proof. Let E be a supersingular elliptic curve defined over Ω . Then there is only one isogeny of order p from E to another elliptic curve, namely the Frobenius isogeny $Fr : E \rightarrow E^{(p)}$. It follows that E has an endomorphism of degree p if and only if E is defined over $GF(p)$.

Consider now the element $\sqrt{-p}$ in the quadratic number field $L = \mathbb{Q}(\sqrt{-p})$. It has norm p and is integral. As L splits A_p , it embeds in A_p . So $\mathbb{Z}[\sqrt{-p}]$ a fortiori embeds and thus is contained in a maximal order O of A_p . The usual norm on $\mathbb{Z}[\sqrt{-p}]$ is equal to the restriction of the reduced norm under the embedding. In the correspondence between maximal orders of A_p and supersingular elliptic curves over Ω , the elliptic curve corresponding to O is thus defined over $GF(p)$. \square

The proof of Deligne's theorem will be completed by the following lemma.

Lemma 36. *Let E and F be supersingular elliptic curves over Ω . Then the intersection, as subgroup schemes of $\ker(\phi)$ as ϕ ranges over $\text{Hom}(E, F)$, is trivial.*

Remark 37. (1) It will suffice to find a collection of isogenies from E to F whose degrees are coprime.

(2) It will suffice to prove the lemma for a fixed elliptic curve E_0 (and F varying), then precomposing with the dual isogenies from E_0 to E coming from (1).

Proof.

First of all we know that $\text{Hom}(E, F)$ is nonzero. If $O = \text{End}(E)$ is the maximal order of $\text{End}(E) \otimes_{\mathbb{Z}} \mathbb{Q}$ corresponding to E , then O has an ideal whose right order is equal to the maximal order corresponding to F and this furnishes a nonzero isogeny. So $\text{Hom}(E, F)$ is a finitely generated projective left module over O (and not the zero module). Let K denote the intersection (as subgroup schemes) of $\ker(\phi)$ as ϕ ranges over $\text{Hom}(E, F)$. We have just showed that K is finite. Among all isogenies from E to F , let ϕ be one of least degree. Let ℓ be a prime dividing $\deg(\phi)$, hence also the order of K . We first treat the somewhat easier case $\ell \neq p$. If $\phi(E[\ell]) = 0$ then $(1/\ell)\phi$ is a nonzero isogeny from E to F of smaller degree, a contradiction. Thus $W = \ker(\phi) \cap E[\ell]$ is one-dimensional. However, $\text{End}(E)$ acts transitively on the one-dimensional subspaces of $E[\ell]$. Thus there is a σ in $\text{End}(E)$ that does not fix W . Then $\phi + \phi \circ \sigma$ is an isogeny from E to F of order prime to ℓ .

We finish the proof of the lemma in the case $\ell = p$.

It is enough by transitivity to assume (by Lemma 35) that $E = E_0$ is defined over $GF(p)$. Assume that every isogeny from E_0 to E has degree divisible by p . Let $\phi : E_0 \rightarrow E$ be the nonzero isogeny of least degree. If ϕ has degree divisible by p then ϕ factors through the Frobenius isogeny: $\phi = \psi \circ Fr$ for some $\psi : E_0^{(p)} \rightarrow E$. But since $E_0^{(p)} = E_0$, $\psi : E_0 \rightarrow E$ is an isogeny of degree smaller than $\deg(\phi)$. \square

This finishes the proof of Lemma 36 and of Theorem 33.

References

- [Pierce 1982] R. S. Pierce, *Associative algebras*, Graduate Texts in Mathematics **88**, Springer, Berlin, 1982. MR Zbl
- [Reiner 1975] I. Reiner, *Maximal orders*, London Mathematical Society Monographs **5**, Academic Press, London, 1975. MR Zbl
- [Serre 1973] J.-P. Serre, *A course in arithmetic*, Graduate Texts in Mathematics **7**, Springer, New York, 1973. MR Zbl
- [Serre 1979] J.-P. Serre, *Local fields*, Graduate Texts in Mathematics **67**, Springer, New York, 1979. MR Zbl
- [Vignéras 1980] M.-F. Vignéras, *Arithmétique des algèbres de quaternions*, Lecture Notes in Mathematics **800**, Springer, Berlin, 1980. MR Zbl
- [Weil 1967] A. Weil, *Basic number theory*, Grundlehren der Math. Wissenschaften **144**, Springer, New York, 1967. MR Zbl

Received January 22, 2017. Revised July 7, 2017.

DANIEL GOLDSTEIN
CENTER FOR COMMUNICATIONS RESEARCH
SAN DIEGO, CA
UNITED STATES
danielgolds@gmail.com

MURRAY SCHACHER
CENTER FOR COMMUNICATIONS RESEARCH
SAN DIEGO, CA
UNITED STATES
mms@math.ucla.edu

GLOBAL EXISTENCE AND BLOWUP OF SMOOTH SOLUTIONS OF 3-D POTENTIAL EQUATIONS WITH TIME-DEPENDENT DAMPING

FEI HOU, INGO WITT AND HUICHENG YIN

In this paper, we are concerned with the global existence and blowup of smooth solutions of the 3-D irrotational compressible Euler equation with time-dependent damping

$$\begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u + pI_3) = -\alpha(t)\rho u, \\ \rho(0, x) = \bar{\rho} + \varepsilon \rho_0(x), \quad u(0, x) = \varepsilon u_0(x), \end{cases}$$

where $x \in \mathbb{R}^3$, the frictional coefficient $\alpha(t) = \mu/(1+t)^\lambda$ with $\mu > 0$ and $\lambda \geq 0$, $\bar{\rho} > 0$ is a constant, $\rho_0, u_0 \in C_0^\infty(\mathbb{R}^3)$, $(\rho_0, u_0) \neq 0$, $\rho(0, x) > 0$, $\operatorname{curl} u_0 \equiv 0$, and $\varepsilon > 0$ is sufficiently small. For $0 \leq \lambda \leq 1$, we show that there exists a global $C^\infty([0, \infty) \times \mathbb{R}^3)$ -smooth solution (ρ, u) by introducing and establishing some uniform time-weighted energy estimates of (ρ, u) , while for $\lambda > 1$, in general, the smooth solution (ρ, u) blows up in finite time. Therefore, $\lambda = 1$ appears to be the critical value for the global existence of small amplitude smooth solution (ρ, u) .

1. Introduction

In this paper, we are concerned with the global existence and blowup of smooth solutions of the three-dimensional irrotational compressible Euler equations with time-dependent damping

$$(1-1) \quad \begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u + pI_3) = -\alpha(t)\rho u, \\ \rho(0, x) = \bar{\rho} + \varepsilon \rho_0(x), \quad u(0, x) = \varepsilon u_0(x), \end{cases}$$

Hou and Yin were supported by the NSFC (No. 11571177) and the Priority Academic Program Development of Jiangsu Higher Education Institutions. Ingo Witt was partly supported by the DFG via the Sino-German project Analysis of Partial Differential Equations and Applications.

MSC2010: primary 35L70; secondary 35L65, 35L67, 76N15.

Keywords: compressible Euler equations, damping, time-weighted energy inequality,

Klainerman–Sobolev inequality, blowup, hypergeometric function.

where $x = (x_1, x_2, x_3)$, $\rho, u = (u_1, u_2, u_3)$, and p stand for the density, velocity, and pressure, respectively, I_3 is the 3×3 identity matrix, the frictional coefficient $\alpha(t) = \mu/(1+t)^\lambda$ with $\mu > 0$ and $\lambda \geq 0$, and $u_0 = (u_{1,0}, u_{2,0}, u_{3,0})$,

$$\operatorname{curl} u_0 = (\partial_2 u_{3,0} - \partial_3 u_{2,0}, \partial_3 u_{1,0} - \partial_1 u_{3,0}, \partial_1 u_{2,0} - \partial_2 u_{1,0}) \equiv 0.$$

The equation of state of the gases is assumed to be $p(\rho) = A\rho^\gamma$, where $A > 0$ and $\gamma > 1$ are constants. Furthermore, $\bar{\rho} > 0$ is a constant, $\rho_0, u_0 \in C_0^\infty(\mathbb{R}^3)$, $(\rho_0, u_0) \not\equiv 0$, $\rho(0, x) > 0$, and $\varepsilon > 0$ is sufficiently small. With respect to the physical background of (1-1), it can be found in [Dafermos 1995].

For $\mu = 0$ in $\alpha(t)$, (1-1) is the standard compressible Euler equation. It is well known that C^∞ -smooth solution (ρ, u) of (1-1) will in general blow up in finite time. For the extensive literature on blowup results and the blowup mechanism for (ρ, u) , see [Alinhac 1999a; 1999b; 1993; Christodoulou 2007; Christodoulou and Miao 2014; Christodoulou and Lisibach 2016; Ding et al. 2016; Hörmander 1997; Sideris 1997; 1985; Speck 2016; Yin and Qiu 1999; Yin 2004] and so on.

For $\lambda = 0$ in $\alpha(t)$, it has been shown that (1-1) admits a global C^∞ -smooth solution (ρ, u) and the large time behavior of (ρ, u) is governed by a parabolic equation derived by using Darcy's law; see [Dafermos 1995; Hsiao and Serre 1996; Hsiao and Liu 1992; Kawashima and Yong 2004; Nishihara 1997; Pan and Zhao 2009; Sideris et al. 2003; Tan and Guochun 2012; Wang and Yang 2001].

For $\mu > 0$ and $\lambda > 0$ in $\alpha(t)$, an interesting problem arises: does the C^∞ -smooth solution (ρ, u) of (1-1) blow up in finite time or does it exist globally? In this paper, we will systematically study this problem under the assumption of $\operatorname{curl} u_0 \equiv 0$. In this case it is not hard to see that $\operatorname{curl} u(t, \cdot) \equiv 0$ for all $t \geq 0$ as long as the smooth solution (ρ, u) of (1-1) exists. Then one can introduce a potential function $\varphi = \varphi(t, x)$ such that $u = \nabla \varphi$ (here and below, $\nabla = \nabla_x$), where the C^∞ scalar function φ has a compact support in x (as $u(t, \cdot)$ has a compact support for any fixed $t \geq 0$ in view of $u_0 \in C_0^\infty(\mathbb{R}^3)$ and admits a finite propagation speed which holds for hyperbolic systems). Substituting $u = \nabla \varphi$ into the second equation of (1-1), we obtain

$$(1-2) \quad \partial_t \varphi + \frac{1}{2} |\nabla \varphi|^2 + h(\rho) + \frac{\mu}{(1+t)^\lambda} \varphi = 0,$$

where $h'(\rho) = c^2(\rho)/\rho$ with $c(\rho) = \sqrt{p'(\rho)}$ and $h(\bar{\rho}) = 0$.

From $h'(\rho) > 0$ for $\rho > 0$ we have that

$$(1-3) \quad \rho = h^{-1} \left(- \left(\partial_t \varphi + \frac{1}{2} |\nabla \varphi|^2 + \frac{\mu}{(1+t)^\lambda} \varphi \right) \right),$$

where $\bar{\rho} = h^{-1}(0)$ and h^{-1} is the inverse function of $h = h(\rho)$.

Substituting (1-3) into the first equation of (1-1) yields

$$(1-4) \quad \partial_t^2 \varphi - c^2(\rho) \Delta \varphi + 2 \sum_{k=1}^3 (\partial_k \varphi) \partial_{tk}^2 \varphi + \sum_{i,k=1}^3 (\partial_i \varphi) (\partial_k \varphi) \partial_{ik}^2 \varphi \\ + \frac{\mu}{(1+t)^\lambda} |\nabla \varphi|^2 + \partial_t \left(\frac{\mu}{(1+t)^\lambda} \varphi \right) = 0.$$

As for the initial data $\varphi(0, x)$ and $\partial_t \varphi(0, x)$ for (1-4): Obviously, $\varphi(0, x) = \varepsilon \varphi_0(x)$, where

$$\varphi_0(x) = \int_{-\infty}^{x_1} u_{1,0}(s, x_2, x_3) ds.$$

Note that $\varphi_0 \in C_0^\infty(\mathbb{R}^3)$ in view of $\text{curl } u_0 \equiv 0$ and $u_0 \in C_0^\infty(\mathbb{R}^3)$. Furthermore, from (1-2) we infer that $\partial_t \varphi(0, x) = \varepsilon \varphi_1(x) + \varepsilon^2 g(x, \varepsilon)$, where

$$\varphi_1 = - \left(\mu \varphi_0 + \frac{c^2(\bar{\rho})}{\bar{\rho}} \rho_0 \right)$$

and

$$g(x, \varepsilon) = -\rho_0^2(x) \int_0^1 \left(\frac{c^2(\rho)}{\rho} \right)' \Big|_{\rho=\bar{\rho}+\theta \varepsilon \rho_0(x)} d\theta - \frac{1}{2} \sum_{i=1}^3 u_{i,0}^2(x).$$

Notice that $g(x, \varepsilon)$ is smooth in (x, ε) and has compact support in x . Consequently, studying problem (1-1) under the assumption $\text{curl } u_0 \equiv 0$ is equivalent to investigating the problem

$$(1-5) \quad \begin{cases} \partial_t^2 \varphi - c^2(\rho) \Delta \varphi + 2 \sum_{k=1}^3 (\partial_k \varphi) \partial_{tk}^2 \varphi + \sum_{i,k=1}^3 (\partial_i \varphi) (\partial_k \varphi) \partial_{ik}^2 \varphi \\ \quad + \frac{\mu}{(1+t)^\lambda} |\nabla \varphi|^2 + \partial_t \left(\frac{\mu}{(1+t)^\lambda} \varphi \right) = 0, \\ \varphi(0, x) = \varepsilon \varphi_0(x), \quad \partial_t \varphi(0, x) = \varepsilon \varphi_1(x) + \varepsilon^2 g(x, \varepsilon). \end{cases}$$

Here we mention that

$$c^2(\rho) = c^2(\bar{\rho}) - (\gamma - 1) \left(\partial_t \varphi + \frac{1}{2} |\nabla \varphi|^2 + \frac{\mu}{(1+t)^\lambda} \varphi \right)$$

which follows by direct computation.

We now state the first main result of this paper.

Theorem 1.1 (global existence for $0 \leq \lambda \leq 1$). *Suppose that $\text{curl } u_0 \equiv 0$. If $\mu > 0$ and $0 \leq \lambda \leq 1$, then, for $\varepsilon > 0$ small enough, (1-5) admits a global C^∞ -smooth solution φ . As a consequence, (1-1) has a global C^∞ -smooth solution (ρ, u) which fulfills $\rho > 0$ and which is uniformly bounded for $t \geq 0$ together with all its derivatives.*

Remark. The principal part of the linearization of the equation in (1-5) about $(\rho, \varphi) = (\bar{\rho}, 0)$ is

$$(1-6) \quad \mathcal{L}(\dot{\varphi}) \equiv \partial_t^2 \dot{\varphi} - c^2(\bar{\rho}) \Delta \dot{\varphi} + \frac{\mu}{(1+t)^\lambda} \partial_t \dot{\varphi} - \frac{\mu\lambda}{(1+t)^{\lambda+1}} \dot{\varphi}.$$

For the linear operator \mathcal{L}_0 with

$$\mathcal{L}_0(\dot{\varphi}) \equiv \partial_t^2 \dot{\varphi} - c^2(\bar{\rho}) \Delta \dot{\varphi} + \frac{\mu}{(1+t)^\lambda} \partial_t \dot{\varphi},$$

which appears as part of (1-6), it is shown in [Wirth 2006; 2007] that the large-term behavior of solutions $\dot{\varphi}$ of $\mathcal{L}_0(\dot{\varphi}) = 0$ depends on the value of λ . For $0 \leq \lambda < 1$ it is the same as the large-term behavior of solutions of the linear heat equation $\partial_t \dot{\varphi} - c^2(\bar{\rho}) \Delta \dot{\varphi} = 0$, while for $\lambda > 1$ it is the same as the large-term behavior of solutions of the linear wave equation $\partial_t^2 \dot{\varphi} - c^2(\bar{\rho}) \Delta \dot{\varphi} = 0$. In addition, precise microlocal large-term decay properties of solutions $\dot{\varphi}$ of $\mathcal{L}(\dot{\varphi}) = 0$ have been established in [do Nascimento and Wirth 2015] for a special range of values of λ and μ . It seems to be difficult, however, to apply these microlocal estimates to attack the quasilinear problem (1-5). (In general, those microlocal estimates are useful when treating semilinear damped wave equations; see [D’Abbicco and Reissig 2014; D’Abbicco et al. 2015].)

Remark. For the 1-D Burgers equation with time-dependent damping term

$$(1-7) \quad \begin{cases} \partial_t w + w \partial_x w = -\frac{\mu}{(1+t)^\lambda} w, & (t, x) \in \mathbb{R}_+ \times \mathbb{R}, \\ w(0, x) = \varepsilon w_0(x), \end{cases}$$

where $\mu > 0$ and $\lambda \geq 0$ are constants, $w_0 \in C_0^\infty(\mathbb{R})$, $w_0 \not\equiv 0$, and $\varepsilon > 0$ is sufficiently small, one concludes by the method of characteristics that

$$\begin{cases} T_\varepsilon = \infty & \text{if } 0 \leq \lambda < 1 \text{ or } \lambda = 1, \mu > 1, \\ T_\varepsilon < \infty & \text{if } \lambda > 1 \text{ or } \lambda = 1, 0 \leq \mu \leq 1, \end{cases}$$

where T_ε is the lifespan of the C^∞ -smooth solution w of (1-7). Therefore, $\lambda = 1$ again appears to be the critical value for the global existence of smooth solutions w of (1-7) in the presence of the damping term

$$\frac{\mu}{(1+t)^\lambda} w.$$

Remark. The smallness of $\varepsilon > 0$ in Theorem 1.1 is necessary in order to guarantee the global existence of smooth solution (ρ, u) . Indeed, as in [Sideris et al. 2003], large amplitude smooth solution of (1-1) may blow up in finite time even for $0 \leq \lambda \leq 1$. See also Theorem 4.1.

Next we concentrate on the case of $\lambda > 1$. As in [Sideris 1985], introduce the two functions

$$q_0(l) = \int_{|x|>l} \frac{(|x| - l)^2}{|x|} (\rho(0, x) - \bar{\rho}) dx,$$

$$q_1(l) = \int_{|x|>l} \frac{|x|^2 - l^2}{|x|^3} x \cdot (\rho u)(0, x) dx.$$

Before stating our blowup result for problem (1-1) with $\lambda > 1$, we require to introduce a special hypergeometric function $\Psi(a, b, c; z)$, where the constants a and b satisfy $a + b = 1$ and $ab = \frac{1}{2}\mu\lambda$, $c \in \mathbb{R}^+$, the variable $z \in \mathbb{R}$, and

$$\Psi(a, b, c; z) = \sum_{n=0}^{+\infty} \frac{(a)_n (b)_n}{n! (c)_n} z^n$$

with $(a)_n = a(a+1) \cdots (a+n-1)$ and $(a)_0 = 1$. It is known from [Erdélyi et al. 1953] that $\Psi(a, b, c; z)$ is an analytic function of z in $(-1, 1)$ and $\Psi(a, b, c; 0) = \Psi(a+1, b+1, c; 0) = 1$. Therefore, there exists a small constant $\delta_0 > 0$ depending on a and b (i.e., μ and λ) such that for $-\frac{1}{2}\delta_0 \leq z \leq 0$,

$$(1-8) \quad \frac{1}{2} \leq \Psi(a, b, 1; z), \Psi(a+1, b+1, 2; z) \leq \frac{3}{2}.$$

Theorem 1.2 (blowup for $\lambda > 1$). *Suppose $\text{supp } \rho_0, \text{supp } u_0 \subseteq \{x : |x| \leq M\}$ and let*

$$(1-9) \quad q_0(l) > 0,$$

$$(1-10) \quad q_1(l) \geq 0$$

hold for all $l \in (\tilde{M}, M)$, where \tilde{M} is some fixed constant satisfying $0 \leq \tilde{M} < M$. Moreover, we assume that there exist two constants $M_0 \geq \tilde{M}$ and $\Lambda \geq \frac{3}{2}\mu\lambda$ such that

$$(1-11) \quad q_1(l) \geq \Lambda q_0(l),$$

holds for all $l \in (M_0, M)$, where $M - M_0 < \delta_0$ and δ_0 is given in (1-8). If $\mu > 0$ and $\lambda > 1$, then there exists an $\varepsilon_0 > 0$ such that, for $0 < \varepsilon \leq \varepsilon_0$, the lifespan T_ε of C^∞ -smooth solution (ρ, u) of (1-1) is finite.

Remark. It is not hard to find a large number of initial data $(\rho, u)(0, x)$ such that (1-9)–(1-11) are satisfied. For instance, choosing $\rho_0(x) > 0$ and $u_0(x) = x\rho_0(x)\Lambda/\bar{\rho}$, then we get (1-9)–(1-11).

Remark. Sideris [1985] showed the formation of singularities in three-dimensional compressible equations under the assumptions of (1-9)–(1-10). However, in order to prove the blowup result of smooth solution (ρ, u) to problem (1.1) and overcome the difficulty arisen by the time-dependent frictional coefficient $\mu/(1+t)^\lambda$ with $\mu > 0$

and $\lambda > 1$, we pose an extra assumption (1-11) except (1-9)–(1-10), which leads to the nonnegativity of $P(t, l)$ in (3-7) so that an ordinary typed blowup inequalities (3-23)–(3-24) can be established. One can see more details in Section 3.

Let us indicate the proofs of Theorems 1.1 and 1.2. To prove Theorem 1.1, we first introduce the function $\psi = \varphi/(1+t)^\lambda$ which fulfills the second-order quasilinear wave equation

$$\partial_t^2 \psi - \Delta \psi + \frac{\mu}{(1+t)^\lambda} \partial_t \psi + \frac{2\lambda}{1+t} \partial_t \psi - \frac{\lambda(1-\lambda)}{(1+t)^2} \psi = Q(\psi, \partial \psi, \partial^2 \psi),$$

where $Q(\psi, \partial \psi, \partial^2 \psi)$ stands for an error term which is of the second order in $(\psi, \partial \psi, \partial^2 \psi)$; $\partial = (\partial_t, \nabla)$. Then, in order to establish the global existence of ψ , we introduce the time-weighted energy

$$E_N(\psi)(t) = \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \Gamma^a \psi|^2 + |\Gamma^a \psi|^2) dx,$$

where $N \geq 8$ is a fixed number, $\Gamma = (\Gamma_0, \Gamma_1, \dots, \Gamma_7) = (\partial, \Omega, S)$ with $\Omega = (\Omega_1, \Omega_2, \Omega_3) = x \wedge \nabla$, $S = t \partial_t + \sum_{k=1}^3 x_k \partial_k$, and $\Gamma^a = \Gamma_0^{a_0} \Gamma_1^{a_1} \dots \Gamma_7^{a_7}$. Note that the vector fields Γ which appear in the definition of the energy $E_N(\psi)(t)$ only comprise part of the standard Klainerman vector fields $\{\partial, \Omega, S, H\}$, where $H = (H_1, H_2, H_3) = (x_1 \partial_t + t \partial_1, x_2 \partial_t + t \partial_2, x_3 \partial_t + t \partial_3)$. This is due to the fact that the equation in (1-5) is not invariant under the Lorentz transformations H in view of the presence of the time-dependent damping. By a rather technical and involved analysis of the resulting equation for ψ , we eventually show that $E_N(\psi)(t) \leq \frac{1}{2} K^2 \varepsilon^2$ when $E_N(\psi)(t) \leq K^2 \varepsilon^2$ is assumed for some suitably large constant $K > 0$ and small $\varepsilon > 0$. Here we emphasize that the condition of $0 \leq \lambda \leq 1$ plays an essential role in the process of deriving the uniform boundedness of $E_N(\psi)(t)$ (see Lemmas 2.3–2.5). This, together with the continuous induction argument, yields the global existence of ψ and further completes the proof of Theorem 1.1 for $0 \leq \lambda \leq 1$. To prove the blowup result of Theorem 1.2 for $\lambda > 1$, as in [Sideris 1985], we derive a related second-order ordinary differential inequality. From this and assumptions (1-9)–(1-11), an upper bound of the lifespan T_ε is derived by making essential use of $\lambda > 1$. In this way the proof of Theorem 1.2 is completed. In Theorem 4.1, we show that for large data smooth solution (ρ, u) of (1-1), even in case $0 \leq \lambda \leq 1$, (ρ, u) will in general blow up in finite time. In addition, the proof on the nonnegativity of $P(t, l)$, which is introduced in (3-1), is given in the Appendix.

Throughout, we shall use the following notation and conventions:

- ∇ stands for ∇_x ;
- $r = |x| = \sqrt{x_1^2 + x_2^2 + x_3^2}$;
- $\langle r - t \rangle = (1 + (r - t)^2)^{1/2}$;

- $\|u(t, x)\| = \left(\int_{\mathbb{R}^3} |u(t, x)|^2 dx\right)^{1/2}$ and $\|u(t, x)\|_{L^\infty} = \sup_{x \in \mathbb{R}^3} |u(t, x)|$;
- Γ denotes one of the vector fields $\{\partial, S, \Omega\}$ on $\mathbb{R}_+ \times \mathbb{R}^3$, where $\partial = (\partial_t, \nabla)$, $S = t\partial_t + \sum_{k=1}^3 x_k \partial_k$, $\Omega = (\Omega_1, \Omega_2, \Omega_3) = x \wedge \nabla$;
- β is the solution of $\beta'(t) = \frac{\mu}{(1+t)^\lambda} \beta(t)$ for $t \geq 0$, $\beta(0) = 1$, i.e.,

$$(1-12) \quad \beta(t) \equiv \begin{cases} e^{\frac{\mu}{1-\lambda}[(1+t)^{1-\lambda}-1]}, & \lambda \geq 0, \lambda \neq 1, \\ (1+t)^\mu, & \lambda = 1; \end{cases}$$

- $c(\bar{\rho}) = 1$ will be assumed throughout (introduce $X = x/c(\bar{\rho})$ as a new space coordinate if necessary).

2. Global existence for small amplitude in case $0 \leq \lambda \leq 1$

Throughout this section, $C > 0$ stands for a generic constant which is independent of K , ε , and t .

We start by recalling a Sobolev-type inequality (see [Klainerman 1987]).

Lemma 2.1. *Let $u = u(t, x)$ be a smooth function of $(t, x) \in [0, \infty) \times \mathbb{R}^3$. Then*

$$(2-1) \quad |u(t, x)| \leq C(1+r)^{-1} \sum_{|a| \leq 2} \|\Gamma^a u(t, x)\|.$$

Moreover, we shall make use of the following inequalities (see [Klainerman and Sideris 1996, Lemma 3.1 and Theorem 5.1]).

Lemma 2.2. *For $u \in C^2([0, \infty) \times \mathbb{R}^3)$,*

$$(2-2) \quad \|\langle r-t \rangle \nabla \partial u(t, x)\| \leq C \left(\sum_{|b| \leq 1} \|\partial \Gamma^b u(t, x)\| + t \|\square u(t, x)\| \right),$$

$$(2-3) \quad (1+r) \langle r-t \rangle |\nabla \partial u(t, x)| \leq C \left(\sum_{|b| \leq 3} \|\partial \Gamma^b u(t, x)\| + t \|\square u(t, x)\| \right),$$

where $\square = \partial_t^2 - \Delta = \partial_t^2 - \sum_{k=1}^3 \partial_k^2$.

We now reformulate problem (1-5). Let $\psi = \varphi/(1+t)^\lambda$. From (1-5) and $c(\bar{\rho}) = 1$ we then have

$$(2-4) \quad \square \psi + \frac{\mu}{(1+t)^\lambda} \partial_t \psi + \frac{2\lambda}{1+t} \partial_t \psi - \frac{\lambda(1-\lambda)}{(1+t)^2} \psi = Q(\psi, \partial \psi, \partial^2 \psi),$$

where

$$\begin{aligned} Q(\psi, \partial \psi, \partial^2 \psi) = & (c^2(\rho) - 1) \Delta \psi - 2(1+t)^\lambda \partial_t \nabla \psi \cdot \nabla \psi - 2\lambda(1+t)^{\lambda-1} |\nabla \psi|^2 \\ & - \mu |\nabla \psi|^2 - (1+t)^{2\lambda} \sum_{1 \leq i, j \leq 3} (\partial_i \psi)(\partial_j \psi) \partial_{ij}^2 \psi. \end{aligned}$$

We define a time-weighted energy for (2-4),

$$E_N(\psi(t)) = \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \Gamma^a \psi|^2 + |\Gamma^a \psi|^2) dx,$$

where $N \geq 8$ is a fixed number. Moreover, we assume that for any $t \geq 0$,

$$(2-5) \quad E_N(\psi(t)) \leq K^2 \varepsilon^2,$$

where $K > 0$ is a suitably large constant. It follows from (2-1) and (2-5) that, for all $|a| \leq N-2$,

$$(2-6) \quad \begin{aligned} |\partial \Gamma^a \psi| &\leq C(1+r)^{-1} \sum_{|b| \leq 2} \|\Gamma^b \partial \Gamma^a \psi(t, x)\| \\ &\leq C(1+r)^{-1} \sum_{|b| \leq N} \|\partial \Gamma^b \psi(t, x)\| \\ &\leq C(1+r)^{-1} (1+t)^{-\lambda} \sqrt{E_N(\psi(t))} \\ &\leq CK\varepsilon(1+r)^{-1} (1+t)^{-\lambda} \end{aligned}$$

and

$$(2-7) \quad |\Gamma^a \psi| \leq C(1+r)^{-1} \sum_{|b| \leq N} \|\Gamma^b \psi(t, x)\| \leq CK\varepsilon(1+r)^{-1}.$$

In view of Lemma 2.2 and (2-5), we have

Lemma 2.3. *Let ψ be a solution of (2-4). Then, for all $|a| \leq N-3$ and $0 \leq \lambda \leq 1$, we have the pointwise estimate*

$$(2-8) \quad \|\nabla \partial \Gamma^a \psi\|_{L^\infty} \leq CK\varepsilon(1+t)^{-2\lambda}.$$

Moreover, for $0 \leq l \leq N-1$, the weighted L^2 estimate

$$(2-9) \quad \begin{aligned} \sum_{|b| \leq l} \|\langle r-t \rangle \nabla \partial \Gamma^b \psi(t, x)\| \\ \leq C \sum_{|c| \leq l+1} \|\partial \Gamma^c \psi(t, x)\| + C(1+t)^{1-\lambda} \sum_{|c| \leq l} \|\nabla \Gamma^c \psi(t, x)\| \\ + C(1-\lambda)(1+t)^{-1} \sum_{|c| \leq l} \|\Gamma^c \psi(t, x)\| \end{aligned}$$

holds.

Proof. It follows from (2-3)–(2-4) and (2-6)–(2-7) that

$$\begin{aligned}
& (1+t) \sum_{|a| \leq N-3} |\nabla \partial \Gamma^a \psi| \\
& \leq C \sum_{|a| \leq N-3} (1+r) \langle r-t \rangle |\nabla \partial \Gamma^a \psi| \\
& \leq C \sum_{|c| \leq N} \|\partial \Gamma^c \psi\| + Ct \sum_{|a| \leq N-3} \|\square \Gamma^a \psi\| \\
& \leq CK\varepsilon(1+t)^{-\lambda} + C(1+t)^{1-\lambda} \sum_{|a| \leq N-3} \|\partial_t \Gamma^a \psi\| + C(1+t)^{-1} \sum_{|a| \leq N-3} \|\Gamma^a \psi\| \\
& \quad + C(1+t) \sum_{|b|+|c| \leq N-3} \|\nabla \partial \Gamma^b \psi \Gamma^c \psi\| + C(1+t)^{1+\lambda} \sum_{|a| \leq N-3} \|\Gamma^a (\partial_t \nabla \psi \cdot \nabla \psi)\| \\
& \leq CK\varepsilon(1+t)^{1-2\lambda} + CK\varepsilon(1+t) \sum_{|a| \leq N-3} \|\nabla \partial \Gamma^a \psi\|_{L^\infty},
\end{aligned}$$

which derives (2-7) in view of the smallness of $\varepsilon > 0$.

By (2-2), (2-6)–(2-8) and (2-4), we have that, for $l \leq N-1$,

(2-10)

$$\begin{aligned}
& \sum_{|b| \leq l} \|\langle r-t \rangle \nabla \partial \Gamma^b \psi\| \\
& \leq C \sum_{|c| \leq l+1} \|\partial \Gamma^c \psi\| + Ct \sum_{|b| \leq l} \|\Gamma^b \square \psi\| \\
& \leq C \sum_{|c| \leq l+1} \|\partial \Gamma^c \psi\| + C(1+t)^{1-\lambda} \sum_{|c| \leq l} \|\nabla \Gamma^c \psi\| + C(1-\lambda)(1+t)^{-1} \sum_{|c| \leq l} \|\Gamma^c \psi\| \\
& \quad + C(1+t)^{1+\lambda} \sum_{|b| \leq l} \|\Gamma^b (\partial_t \nabla \psi \cdot \nabla \psi)\| \\
& \quad + C(1+t) \sum_{\substack{|c| \leq N-3, \\ |b| \leq l-|c|}} \|\langle r-t \rangle^{-1} \Gamma^c \psi\|_{L^\infty} \|\langle r-t \rangle \nabla \partial \Gamma^b \psi\| \\
& \quad + C(1+t) \sum_{\substack{2-N \leq |c| \leq l, \\ |b| \leq l+2-N}} \|(1+r) \nabla \partial \Gamma^b \psi\|_{L^\infty} \|(1+r)^{-1} \Gamma^c \psi\| \\
& \leq C \sum_{|c| \leq l+1} \|\partial \Gamma^c \psi\| + C(1+t)^{1-\lambda} \sum_{|c| \leq l} \|\nabla \Gamma^c \psi\| + C(1-\lambda)(1+t)^{-1} \sum_{|c| \leq l} \|\Gamma^c \psi\| \\
& \quad + CK\varepsilon \sum_{|b| \leq l} \|\langle r-t \rangle \nabla \partial \Gamma^b \psi\| + CK\varepsilon(1+t)^{1-\lambda} \sum_{2-N \leq |c| \leq l} \|(1+r)^{-1} \Gamma^c \psi\|.
\end{aligned}$$

Note that $\Gamma^c \psi(t, x)$ is supported in $\{x : |x| \leq t + M\}$. Then it follows from Hardy inequality that

$$(2-11) \quad \|(1+r)^{-1} \Gamma^c \psi\| \leq C \|\nabla \Gamma^c \psi\|.$$

Substituting (2-11) into (2-10) and applying the smallness of ε , we derive (2-9). \square

Next we derive the time-weighted energy estimate for the solution ψ of (2-4).

Lemma 2.4. *Let $\mu > 0$ and $\lambda \in (0, 1]$. Under assumption (2-5), for all $t > 0$ and $N \geq 8$, it holds that*

$$(2-12) \quad \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \partial^a \psi|^2 + \psi^2) dx + C \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \partial^a \psi|^2 dx d\tau \\ \leq C\varepsilon^2 + C(1+K\varepsilon) \int_0^t A(\tau) \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+\tau)^{2\lambda} |\partial \partial^a \psi|^2 + \psi^2) dx d\tau,$$

where $A(\cdot)$ stands for a generic nonnegative function such that $A \in L^1((0, \infty))$, and $\|A\|_{L^1}$ is independent of K but dependent on μ and λ .

Proof. First we show (2-12) in case $|a| = 0$. Multiplying (2-4) by $m(1+t)^{2\lambda} \partial_t \psi + (1+t)^{2\lambda-1} \psi$ yields by a direct computation

$$(2-13) \quad \frac{1}{2} \partial_t [m(1+t)^{2\lambda} |\partial \psi|^2 + 2(1+t)^{2\lambda-1} \psi \partial_t \psi + (\mu(1+t)^{\lambda-1} + 2\lambda(1+t)^{2\lambda-2}) \psi^2] \\ + \operatorname{div}(\cdots) + (\mu m(1+t)^\lambda + (\lambda m - 1)(1+t)^{2\lambda-1}) (\partial_t \psi)^2 \\ + (1 - \lambda m)(1+t)^{2\lambda-1} |\nabla \psi|^2 + \frac{\mu}{2} (1 - \lambda)(1+t)^{\lambda-2} \psi^2 \\ + C_1(\lambda - 1)(1+t)^{2\lambda-2} \psi \partial_t \psi + C_2(\lambda - 1)(1+t)^{2\lambda-3} \psi^2 \\ = (m(1+t)^{2\lambda} \partial_t \psi + (1+t)^{2\lambda-1} \psi) \mathcal{Q}(\psi, \partial \psi, \partial^2 \psi),$$

where the constant $m > 0$ will be determined later and C_i ($i = 1, 2$) are suitable constants. Note that in the square bracket of the first line in (2-13),

$$(2-14) \quad m(1+t)^{2\lambda} |\partial \psi|^2 + 2(1+t)^{2\lambda-1} \psi \partial_t \psi + (\mu(1+t)^{\lambda-1} + 2\lambda(1+t)^{2\lambda-2}) \psi^2 \\ = m(1+t)^{2\lambda} \left(\frac{1}{3} |\partial_t \psi|^2 + |\nabla \psi|^2 \right) + \left(\mu(1+t)^{\lambda-1} + \left(2\lambda - \frac{3}{2m} \right) (1+t)^{2\lambda-2} \right) \psi^2 \\ + \left(\sqrt{\frac{2m}{3}} (1+t)^\lambda \partial_t \psi + \sqrt{\frac{3}{2m}} (1+t)^{\lambda-1} \psi \right)^2.$$

We choose $m > 0$ to fulfill

$$\lambda < \frac{1}{m} < \min\{\mu + \lambda, 2\lambda\};$$

together with $\lambda \leq 1$ (i.e., $2\lambda - 2 \leq \lambda - 1 \leq 0$), this yields that (2-14) is equivalent to

$$(1+t)^{2\lambda} |\partial \psi|^2 + (1+t)^{\lambda-1} \psi^2.$$

On the other hand, the coefficients

$$\mu m(1+t)^\lambda + (\lambda m - 1)(1+t)^{2\lambda-1}$$

and

$$(1 - \lambda m)(1+t)^{2\lambda-1}$$

of $(\partial_t \psi)^2$ and $|\nabla \psi|^2$ in the left-hand side of (2-13) are both positive.

Then integrating (2-13) over $[0, t] \times \mathbb{R}^3$ yields

$$\begin{aligned} (2-15) \quad & \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \psi|^2 + (1+t)^{\lambda-1} \psi^2) dx \\ & + C \int_0^t \int_{\mathbb{R}^3} ((1+\tau)^\lambda (\partial_t \psi)^2 + (1+\tau)^{2\lambda-1} |\nabla \psi|^2 + (1+\tau)^{\lambda-2} \psi^2) dx d\tau \\ & \leq C\varepsilon^2 + \int_0^t A(\tau) \int_{\mathbb{R}^3} (1+\tau)^{\lambda-1} \psi^2 dx d\tau \\ & + C \left| \int_0^t \int_{\mathbb{R}^3} (m(1+\tau)^{2\lambda} \partial_t \psi + (1+\tau)^{2\lambda-1} \psi) Q(\psi, \partial \psi, \partial^2 \psi) dx d\tau \right|. \end{aligned}$$

Next we improve the time-weighted estimate of ψ in the left-hand side of (2-15). Multiplying both sides of (2-4) by $(1+t)^\lambda \psi$ yields by direct computation

$$\begin{aligned} & \partial_t \left((1+t)^\lambda \psi \partial_t \psi + \frac{\mu}{2} \psi^2 \right) + \operatorname{div}(\cdots) - (1+t)^\lambda (\partial_t \psi)^2 - \lambda(1+t)^{\lambda-1} \psi \partial_t \psi \\ & + (1+t)^\lambda |\nabla \psi|^2 + 2\lambda(1+t)^{\lambda-1} \psi \partial_t \psi + \lambda(\lambda-1)(1+t)^{\lambda-2} \psi^2 \\ & = (1+t)^\lambda \psi Q(\psi, \partial \psi, \partial^2 \psi). \end{aligned}$$

From this and (2-15), we can choose the multiplier

$$m(1+t)^{2\lambda} \partial_t \psi + (1+t)^{2\lambda-1} \psi + \kappa(1+t)^\lambda \psi$$

for (2-4) with a small $\kappa > 0$ and then obtain

$$\begin{aligned} (2-16) \quad & \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \psi|^2 + \psi^2) dx + C \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \psi|^2 dx d\tau \\ & \leq C\varepsilon^2 + \int_0^t A(\tau) \int_{\mathbb{R}^3} \psi^2 dx d\tau \\ & + C \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \psi) Q(\psi, \partial \psi, \partial^2 \psi) dx d\tau \right| \\ & + C \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda \psi Q(\psi, \partial \psi, \partial^2 \psi) dx d\tau \right|. \end{aligned}$$

Next we derive the time-weighted estimates of $\partial^a \psi$ with $1 \leq |a| \leq N$. Taking ∂^a on both sides of (2-4) yields

$$\begin{aligned} \square \partial^a \psi &+ \frac{\mu}{(1+t)^\lambda} \partial_t \partial^a \psi + \frac{2\lambda}{1+t} \partial_t \partial^a \psi \\ &= \partial^a Q(\psi, \partial \psi, \partial^2 \psi) + \sum_{1 \leq |b| \leq |a|} \frac{1}{(1+t)^\lambda} (1 + O((1+t)^{\lambda-1})) \partial^b \psi \\ &\quad - \lambda(\lambda-1) \partial^a \left(\frac{1}{(1+t)^2} \right) \psi. \end{aligned}$$

Exactly as for (2-16), multiplying this by

$$m(1+t)^{2\lambda} \partial_t \partial^a \psi + (1+t)^{2\lambda-1} \partial^a \psi + \kappa(1+t)^\lambda \partial^a \psi,$$

we obtain

$$\begin{aligned} (2-17) \quad &\sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \partial^a \psi|^2 + \psi^2) dx + C \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \partial^a \psi|^2 dx d\tau \\ &\leq C\varepsilon^2 + \int_0^t A(\tau) \int_{\mathbb{R}^3} \psi^2 dx d\tau \\ &\quad + C \sum_{0 \leq |a| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \partial^a \psi) \partial^a Q(\psi, \partial \psi, \partial^2 \psi) dx d\tau \right| \\ &\quad + C \sum_{0 \leq |a| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\partial^a \psi) \partial^a Q(\psi, \partial \psi, \partial^2 \psi) dx d\tau \right|. \end{aligned}$$

We now deal with the last two terms in the right-hand side of (2-17). We first analyze the integrand $(1+t)^{2\lambda} (\partial_t \partial^a \psi) \partial^a Q(\psi, \partial \psi, \partial^2 \psi)$ of the penultimate term. Direct computation yields

$$\begin{aligned} \partial^a Q(\psi, \partial \psi, \partial^2 \psi) \\ = (c^2(\rho) - 1) \Delta \partial^a \psi - 2(1+t)^\lambda \nabla \partial_t \partial^a \psi \cdot \nabla \psi - (1+t)^{2\lambda} (\partial_i \psi) (\partial_j \psi) \partial_{ij}^2 \partial^a \psi + \text{l.o.t.} \end{aligned}$$

and

$$\begin{aligned} (2-18) \quad &(1+t)^{2\lambda} (\partial_t \partial^a \psi) \partial^a Q(\psi, \partial \psi, \partial^2 \psi) \\ &= \text{div}((1+t)^{2\lambda} (c^2(\rho) - 1) (\partial_t \partial^a \psi) \nabla \partial^a \psi) - \text{div}((1+t)^{3\lambda} |\partial_t \partial^a \psi|^2 \nabla \psi) \\ &\quad - \frac{1}{2} \partial_t ((1+t)^{2\lambda} (c^2(\rho) - 1) |\nabla \partial^a \psi|^2) \\ &\quad + (1+t)^{3\lambda} |\partial_t \partial^a \psi|^2 \Delta \psi + \lambda(1+t)^{2\lambda-1} (c^2(\rho) - 1) |\nabla \partial^a \psi|^2 \\ &\quad + \frac{1}{2} (1+t)^{2\lambda} (c^2(\rho))' \partial_t \rho |\nabla \partial^a \psi|^2 \\ &\quad - (1+t)^{4\lambda} (\partial_i \psi) (\partial_j \psi) (\partial_{ij}^2 \partial^a \psi) \partial_t \partial^a \psi + \text{l.o.t.}, \end{aligned}$$

where here and below l.o.t. designates lower-order terms which are of the form

$$(\partial^{b_1} \psi)(\partial^{b_2} \psi) \dots (\partial^{b_l} \psi)$$

(multiplied by $\partial \partial^a \psi$ or $\partial^a \psi$) with $l \geq 2$ and $1 \leq |b_1| + \dots + |b_l| \leq |a| + 1$. Here we are concerned with the top-order derivatives only. Note that the term $(1+t)^{4\lambda}(\partial_i \psi)(\partial_j \psi)(\partial_{ij}^2 \partial^a \psi) \partial_t \partial^a \psi$ in (2-18) can be expressed as

$$\begin{aligned} (2-19) \quad & (1+t)^{4\lambda}(\partial_i \psi)(\partial_j \psi)(\partial_{ij}^2 \partial^a \psi) \partial_t \partial^a \psi \\ &= \frac{1}{2} \{ \partial_i ((1+t)^{4\lambda}(\partial_i \psi)(\partial_j \psi)(\partial_j \partial^a \psi) \partial_t \partial^a \psi) \\ & \quad + \partial_j ((1+t)^{4\lambda}(\partial_i \psi)(\partial_j \psi)(\partial_i \partial^a \psi) \partial_t \partial^a \psi) \\ & \quad - \partial_t ((1+t)^{4\lambda}(\partial_i \psi)(\partial_j \psi)(\partial_i \partial^a \psi) \partial_j \partial^a \psi) \\ & \quad + \partial_t ((1+t)^{4\lambda}(\partial_i \psi) \partial_j \psi)(\partial_i \partial^a \psi) \partial_j \partial^a \psi + \text{l.o.t.} \}. \end{aligned}$$

Similarly, for the integrand of

$$\left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\partial^a \psi) \partial^a Q(\psi, \partial \psi, \partial^2 \psi) dx d\tau \right|,$$

one has

$$\begin{aligned} (2-20) \quad & (1+t)^\lambda \partial^a \psi \partial^a Q(\psi, \partial \psi, \partial^2 \psi) \\ &= \text{div}((1+t)^\lambda (c^2(\rho) - 1) \nabla(\partial^a \psi) \partial^a \psi) - \frac{1}{2} \partial_i ((1+t)^{3\lambda} (\partial_i \psi) \partial^a (|\nabla \psi|^2) \partial^a \psi) \\ & \quad - \partial_t ((1+t)^\lambda (c^2(\rho) - 1) |\nabla \psi|^2) \partial^a \psi - (1+t)^\lambda (c^2(\rho) - 1) |\nabla \partial^a \psi|^2 \\ & \quad - (1+t)^\lambda (c^2(\rho))' \nabla \rho \cdot \nabla(\partial^a \psi) \partial^a \psi + \lambda(1+t)^{\lambda-1} \partial^a (|\nabla \psi|^2) \partial^a \psi \\ & \quad + (1+t)^\lambda \partial^a (|\nabla \psi|^2) \partial_t \partial^a \psi + \frac{1}{2} (1+t)^{3\lambda} (\Delta \psi) \partial^a (|\nabla \psi|^2) \partial^a \psi \\ & \quad + \frac{1}{2} (1+t)^{3\lambda} \nabla \psi \cdot \nabla(\partial^a \psi) \partial^a (|\nabla \psi|^2) + \text{l.o.t.} \end{aligned}$$

From the expression $(\partial^{b_1} \psi)(\partial^{b_2} \psi) \dots (\partial^{b_l} \psi)$ ($l \geq 2$, $1 \leq |b_1| + \dots + |b_l| \leq N+1$) of the lower-order terms one readily obtains that there exists at most one b_j ($1 \leq j \leq l$) such that

$$\left\lceil \frac{N+3}{2} \right\rceil < |b_j| \leq N+1.$$

Moreover, $\left\lceil \frac{N+3}{2} \right\rceil \leq N-2$ by $N \geq 8$. Thus, applying (2-5)–(2-7) and subsequently substituting (2-18)–(2-20) into (2-17) completes the proof of Lemma 2.4. \square

Next we focus on the general time-weighted energy estimate of $\partial \Gamma^a \psi$ with $0 \leq |a| \leq N$ and $N \geq 8$.

Lemma 2.5 (time-weighted energy estimate of $\partial\Gamma^a\psi$ for $|a| \leq N$). *Let $\mu > 0$ and $\lambda \in (0, 1]$. Under assumption (2-5), we have that, for $t > 0$,*

$$(2-21) \quad \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial\Gamma^a\psi|^2 + |\Gamma^a\psi|^2) dx \\ + C \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial\Gamma^a\psi|^2 dx d\tau \\ \leq C\varepsilon^2 + C(1+K\varepsilon) \int_0^t A(\tau) \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} ((1+\tau)^{2\lambda} |\partial\Gamma^a\psi|^2 + \psi^2) dx d\tau,$$

where the function A has been defined in Lemma 2.4.

Proof. Writing $\Gamma^a = \tilde{\Gamma}^b \partial^c$ with $\tilde{\Gamma} \in \{\Omega, S\}$, we will use induction on $|b|$ to prove (2-21). In view of Lemma 2.4, it is enough to assume that $|c| = 0$.

Suppose that (2-21) holds for $|b| \leq l-1$, where $1 \leq l \leq N$. We then intend to establish (2-21) for $|b| = l$.

Acting with $\tilde{\Gamma}^a$ (where $a = b$ and $|b| = l$) on both sides of (2-4) yields

$$(2-22) \quad \square \tilde{\Gamma}^a \psi + \frac{\mu}{(1+t)^\lambda} \partial_t \tilde{\Gamma}^a \psi + \frac{2\lambda}{1+t} \partial_t \tilde{\Gamma}^a \psi \\ = \sum_{|b_1| < |b|} \tilde{\Gamma}^{b_1} \partial^c \square \psi + \tilde{\Gamma}^a Q(\psi, \partial\psi, \partial^2\psi) \\ - \left[\tilde{\Gamma}^a, \frac{\mu}{(1+t)^\lambda} \partial_t \right] \psi - \left[\tilde{\Gamma}^a, \frac{2\lambda}{1+t} \partial_t \right] \psi + \tilde{\Gamma}^a ((\lambda-1)(1+t)^{-2} \psi).$$

Starting from (2-22), as in the proof of Lemma 2.4, we can choose the multiplier

$$m(1+t)^{2\lambda} \partial_t \tilde{\Gamma}^a \psi + (1+t)^{2\lambda-1} \tilde{\Gamma}^a \psi + \kappa(1+t)^\lambda \tilde{\Gamma}^a \psi$$

to derive (2-21). For the commutators, we see from (2-4) that

$$(2-23) \quad \left| \int_0^t \int_{\mathbb{R}^3} \left[\tilde{\Gamma}^a, \frac{\mu}{(1+t)^\lambda} \partial_t \right] \psi (1+t)^\lambda \tilde{\Gamma}^a \psi dx d\tau \right| \\ \leq C \sum_{|a_1| < |a|} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda \square \tilde{\Gamma}^{a_1} \psi \tilde{\Gamma}^a \psi dx d\tau \right| \\ + C \sum_{|a_1| < |a|} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda \tilde{\Gamma}^{a_1} Q(\psi, \partial\psi, \partial^2\psi) \tilde{\Gamma}^a \psi dx d\tau \right| \\ + C \sum_{|a_1| < |a|} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{\lambda-1} \tilde{\Gamma}^a \psi (\partial_t \tilde{\Gamma}^{a_1} \psi + (1-\lambda)(1+\tau)^{-1} \tilde{\Gamma}^{a_1} \psi) dx d\tau \right|$$

$$\begin{aligned}
&\leq C\varepsilon^2 + C \sum_{|a_1| < |a|} \left| \int_{\mathbb{R}^3} (1+t)^\lambda \partial_t \tilde{\Gamma}^{a_1} \psi \tilde{\Gamma}^a \psi \, dx \right| \\
&\quad + C \sum_{|a_1| < |a|} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda \tilde{\Gamma}^{a_1} Q(\psi, \partial\psi, \partial^2\psi) \tilde{\Gamma}^a \psi \, dx \, d\tau \right| \\
&\quad + C \sum_{|a_1| < |a|} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{\lambda-1} \tilde{\Gamma}^a \psi (\partial_t \tilde{\Gamma}^{a_1} \psi + (1-\lambda)(1+\tau)^{-1} \tilde{\Gamma}^{a_1} \psi) \, dx \, d\tau \right| \\
&\quad + C \sum_{|a_1| < |a|} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda \partial \tilde{\Gamma}^{a_1} \psi \partial \tilde{\Gamma}^a \psi \, dx \, d\tau \right|.
\end{aligned}$$

By the finite propagation speed, we have for $a > 0$

$$(2-24) \quad |\tilde{\Gamma}^a \psi| \leq C(1+t) \sum_{|a_1| < |a|} |\partial \tilde{\Gamma}^{a_1} \psi|.$$

It follows from (2-23)–(2-24) and a direct computation that

$$\begin{aligned}
(2-25) \quad &\sum_{\substack{|b|=l, \\ |c| \leq N-l}} \int_{\mathbb{R}^3} ((1+t)^{2\lambda} |\partial \tilde{\Gamma}^b \partial^c \psi|^2 + |\tilde{\Gamma}^b \partial^c \psi|^2) \, dx \\
&\quad + C \sum_{\substack{|b|=l, \\ |c| \leq N-l}} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^b \partial^c \psi|^2 \, dx \, d\tau \\
&\leq C\varepsilon^2 + CE_{l-1}(\psi(t)) + C \sum_{\substack{|b_1| < l, \\ |c_1| \leq N-|b_1|}} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^{b_1} \partial^{c_1} \psi|^2 \, dx \, d\tau \\
&\quad + C(1+K\varepsilon) \int_0^t A(\tau) \sum_{\substack{|b_1| \leq l, \\ |c_1| \leq N-|b_1|}} \int_{\mathbb{R}^3} ((1+\tau)^{2\lambda} |\partial \tilde{\Gamma}^{b_1} \partial^{c_1} \psi|^2 + |\tilde{\Gamma}^{b_1} \partial^{c_1} \psi|^2) \, dx \, d\tau \\
&\quad + C \sum_{\substack{|b_1| \leq l, \\ |c_1| \leq N-|b_1|}} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} \partial^{c_1} Q(\psi, \partial\psi, \partial^2\psi) \, dx \, d\tau \right| \\
&\quad + C \sum_{\substack{|b_1| \leq l, \\ |c_1| \leq N-|b_1|}} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} \partial^{c_1} Q(\psi, \partial\psi, \partial^2\psi) \, dx \, d\tau \right|.
\end{aligned}$$

Next we deal with the last two terms in the right-hand side of (2-25). Note that

$$c^2(\rho) - 1 = -G(\psi, \partial\psi) \int_0^1 (c^2)'(-sG(\psi, \partial\psi)) \, ds,$$

where $G(\psi, \partial\psi) = (1+t)^\lambda \partial_t \psi + (1+t)^{\lambda-1} \psi + (1+t)^{2\lambda} |\nabla \psi|^2 / 2 + \mu \psi$. From this, it is readily seen that the typical terms in $Q(\psi, \partial\psi, \partial^2\psi)$ are of the form $\psi \Delta \psi$, $(1+t)^\lambda \partial_t \nabla \psi \cdot \nabla \psi$, and $(1+t)^{2\lambda} (\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi$. We analyze them separately. Without loss of generality, we assume $|c_1| = 0$ in the last two terms of (2-25); the treatment of the other cases is easier.

Part A: *Estimates of*

$$\sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} (\psi \Delta \psi) dx d\tau \right|$$

and

$$\sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} (\psi \Delta \psi) dx d\tau \right|.$$

Note that

$$\tilde{\Gamma}^{b_1} (\psi \Delta \psi) = I_1 + I_2 + I_3,$$

where

$$\begin{aligned} I_1 &= \psi \Delta \tilde{\Gamma}^{b_1} \psi, \\ I_2 &= \sum_{\substack{|b_1|=|b_2|+|b_3|, \\ 1 \leq |b_2| \leq N-2}} (\tilde{\Gamma}^{b_2} \psi) \Delta \tilde{\Gamma}^{b_3} \psi, \\ I_3 &= \sum_{\substack{|b_1|=|b_2|+|b_3|, \\ N-1 \leq |b_2| \leq l}} (\tilde{\Gamma}^{b_2} \psi) \Delta \tilde{\Gamma}^{b_3} \psi. \end{aligned}$$

In view of $b_1 = a$ and

$$\begin{aligned} & (1+t)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \psi \Delta \tilde{\Gamma}^a \psi \\ &= \operatorname{div}((1+t)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \psi \nabla \tilde{\Gamma}^a \psi) + \frac{1}{2} \partial_t ((1+t)^{2\lambda} |\nabla \tilde{\Gamma}^a \psi|^2 \psi) \\ & \quad - (1+t)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \nabla \psi \cdot \nabla \tilde{\Gamma}^a \psi - \lambda (1+t)^{\lambda-1} |\nabla \tilde{\Gamma}^a \psi|^2 \psi - \frac{1}{2} (1+t)^{2\lambda} |\nabla \tilde{\Gamma}^a \psi|^2 \partial_t \psi, \end{aligned}$$

we have by an integration by parts and (2-6)–(2-7)

$$\begin{aligned} (2-26) \quad & \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) I_1 dx d\tau \right| \\ & \leq C\varepsilon^2 + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} (1+t)^{2\lambda} |\partial \tilde{\Gamma}^a \psi|^2 dx \\ & \quad + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau. \end{aligned}$$

Moreover, it follows from (2-7) and (2-9) that

$$\begin{aligned}
 (2-27) \quad & \int_{\mathbb{R}^3} |(1+t)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \mathbf{I}_2| dx \\
 & \leq (1+t)^{2\lambda} \|\langle r-t \rangle^{-1} \tilde{\Gamma}^{b_2} \psi\|_{L^\infty} \cdot \|\partial_t \tilde{\Gamma}^a \psi\| \cdot \|\langle r-t \rangle \Delta \tilde{\Gamma}^{b_3} \psi\| \\
 & \leq CK\varepsilon(1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\| \sum_{|b_4| \leq |b_3|+1} (\|\nabla \tilde{\Gamma}^{b_4} \psi\| + (1-\lambda)(1+t)^{-1} \|\tilde{\Gamma}^{b_4} \psi\|) \\
 & \leq CK\varepsilon(1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\| \sum_{|b_4| \leq |b_3|+1} \|\nabla \tilde{\Gamma}^{b_4} \psi\| + CK\varepsilon(1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\|^2 \\
 & \quad + CK\varepsilon(1-\lambda)(1+t)^{\lambda-2} \sum_{|b_4| \leq |b_3|+1} \|\tilde{\Gamma}^{b_4} \psi\|^2.
 \end{aligned}$$

On the other hand, we have that by (2-6) and Hardy's inequality

$$\begin{aligned}
 (2-28) \quad & \int_{\mathbb{R}^3} |(1+t)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \mathbf{I}_3| dx \\
 & \leq (1+t)^{2\lambda} \|(1+r) \Delta \tilde{\Gamma}^{b_3} \psi\|_{L^\infty} \cdot \|\partial_t \tilde{\Gamma}^a \psi\| \|(1+r)^{-1} \tilde{\Gamma}^{b_2} \psi\| \\
 & \leq CK\varepsilon(1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\| \sum_{|b_4| \leq |b_2|} \|\nabla \tilde{\Gamma}^{b_4} \psi\|.
 \end{aligned}$$

Combining (2-26)–(2-28) together with $0 \leq \lambda \leq 1$ (this means that the coefficient $CK\varepsilon(1-\lambda)(1+t)^{\lambda-2}$ of $\sum_{|b_4| \leq |b_3|+1} \|\tilde{\Gamma}^{b_4} \psi\|^2$ in the last line of (2-27) is nonnegative and in $L^1(0, \infty)$) yields

$$\begin{aligned}
 (2-29) \quad & \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \Gamma^a \psi) \Gamma^{b_1} (\psi \Delta \psi) dx d\tau \right| \\
 & \leq C\varepsilon^2 + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} (1+t)^{2\lambda} |\partial \tilde{\Gamma}^a \psi|^2 dx \\
 & \quad + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau \\
 & \quad + CK\varepsilon \sum_{|b_4| \leq N} \int_0^t A(\tau) \int_{\mathbb{R}^3} |\tilde{\Gamma}^{b_4} \psi|^2 dx d\tau.
 \end{aligned}$$

Note that

$$\begin{aligned}
 (1+t)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} (\psi \Delta \psi) &= \sum_{|b_2|+|b_3|=|b_1|} (1+t)^\lambda (\tilde{\Gamma}^a \psi) (\tilde{\Gamma}^{b_2} \psi) \Delta \tilde{\Gamma}^{b_3} \psi \\
 &= \operatorname{div} \left(\sum_{|b_2|+|b_3|=|b_1|} (1+t)^\lambda (\tilde{\Gamma}^a \psi) (\tilde{\Gamma}^{b_2} \psi) \nabla \tilde{\Gamma}^{b_3} \psi \right) + \sum_{i=4}^5 \mathbf{I}_i,
 \end{aligned}$$

where

$$\begin{aligned}
 I_4 &= - \sum_{\substack{|b_2| \leq N-2, \\ |b_2|+|b_3|=|b_1|}} (1+t)^\lambda (\tilde{\Gamma}^{b_2} \psi) (\nabla \tilde{\Gamma}^a \psi) \cdot (\nabla \tilde{\Gamma}^{b_3} \psi), \\
 I_5 &= - \sum_{\substack{N-1 \leq |b_2| \leq l-1, \\ |b_2|+|b_3|=|b_1|}} (1+t)^\lambda (\tilde{\Gamma}^{b_2} \psi) (\nabla \tilde{\Gamma}^a \psi) \cdot (\nabla \tilde{\Gamma}^{b_3} \psi) \\
 &\quad - \sum_{|b_2|+|b_3|=|b_1|} (1+t)^\lambda (\tilde{\Gamma}^a \psi) (\nabla \tilde{\Gamma}^{b_2} \psi) \cdot (\nabla \tilde{\Gamma}^{b_3} \psi).
 \end{aligned}$$

Therefore, by (2-7) and Hardy's inequality, we have

$$\int_{\mathbb{R}^3} |I_4| dx \leq CK\varepsilon (1+t)^\lambda \|\nabla \tilde{\Gamma}^a \psi\| \sum_{|b_1|+3-N \leq |b_3| \leq N} \|\nabla \tilde{\Gamma}^{b_3} \psi\|$$

and

$$\int_{\mathbb{R}^3} |I_5| dx \leq CK\varepsilon \|(1+r)^{-1} \tilde{\Gamma}^{b_2} \psi \nabla \tilde{\Gamma}^a \psi\|_{L^1} \leq CK\varepsilon \|\nabla \tilde{\Gamma}^{b_2} \psi\| \|\nabla \tilde{\Gamma}^a \psi\|.$$

This yields

$$\begin{aligned}
 (2-30) \quad \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} (\psi \Delta \psi) dx d\tau \right| \\
 \leq CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau.
 \end{aligned}$$

Part B: *Estimates of*

$$\sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} ((1+\tau)^\lambda \partial_t \nabla \psi \cdot \nabla \psi) dx d\tau \right|$$

and

$$\sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} ((1+\tau)^\lambda \partial_t \nabla \psi \cdot \nabla \psi) dx d\tau \right|.$$

One has

$$\begin{aligned}
 &\tilde{\Gamma}^{b_1} ((1+t)^\lambda \partial_t \nabla \psi \cdot \nabla \psi) \\
 &= (1+t)^\lambda \partial_t \nabla \tilde{\Gamma}^{b_1} \psi \cdot \nabla \psi + \sum_{N-2 \leq |b_2| \leq l-1} (1+t)^\lambda (\partial_t \nabla \tilde{\Gamma}^{b_2} \psi) \nabla \tilde{\Gamma}^{b_3} \psi \\
 &\quad + \sum_{|b_2| \leq N-3} (1+t)^\lambda (\partial_t \nabla \tilde{\Gamma}^{b_2} \psi) \nabla \tilde{\Gamma}^{b_3} \psi \\
 &= \Pi_1 + \Pi_2 + \Pi_3.
 \end{aligned}$$

By (2-8), we have

$$(2-31) \quad \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \Pi_1 dx d\tau \right| \leq CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau.$$

In addition, it follows from (2-6), (2-9) and a direct computation that

$$(2-32) \quad \begin{aligned} & (1+t)^{2\lambda} \|(\partial_t \Gamma^a \psi) \Pi_2\|_{L^1} \\ & \leq (1+t)^{3\lambda} \sum_{|b_2| \leq N-4} \|\langle r-t \rangle^{-1} \nabla \Gamma^{b_3} \psi\|_{L^\infty} \cdot \|\partial_t \Gamma^a \psi\| \cdot \|\langle r-t \rangle \partial_t \nabla \Gamma^{b_2} \psi\| \\ & \leq CK\varepsilon (1+t)^\lambda \|\partial_t \Gamma^a \psi\| \sum_{|c| \leq |b_2|+1} (\|\nabla \Gamma^c \psi\| + (1-\lambda)(1+t)^{-1} \|\Gamma^c \psi\|) \\ & \leq CK\varepsilon (1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\| \sum_{|b_4| \leq |b_3|+1} \|\nabla \tilde{\Gamma}^{b_4} \psi\| + CK\varepsilon (1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\|^2 \\ & \quad + CK\varepsilon (1-\lambda)(1+t)^{\lambda-2} \sum_{|b_4| \leq |b_3|+1} \|\tilde{\Gamma}^{b_4} \psi\|^2. \end{aligned}$$

Treating Π_3 , we obtain by (2-8)

$$(2-33) \quad \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \Pi_3 dx d\tau \right| \leq CK\varepsilon \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau.$$

Collecting (2-31)–(2-33) together with $0 \leq \lambda \leq 1$ (this means that the coefficient $CK\varepsilon(1-\lambda)(1+t)^{\lambda-2}$ of $\sum_{|b_4| \leq |b_3|+1} \|\tilde{\Gamma}^{b_4} \psi\|^2$ in the last line of (2-32) is nonnegative and in $L^1(0, \infty)$) yields

$$(2-34) \quad \begin{aligned} & \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \Gamma^a \psi) \Gamma^{b_1} ((1+t)^\lambda \partial_t \nabla \psi \cdot \nabla \psi) dx d\tau \right| \\ & \leq CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau \\ & \quad + CK\varepsilon \sum_{|b_4| \leq N} \int_0^t A(\tau) \int_{\mathbb{R}^3} |\tilde{\Gamma}^{b_4} \psi|^2 dx d\tau. \end{aligned}$$

In addition, one notes that

$$\begin{aligned} & 2(1+t)^{2\lambda} (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^a (\partial_t \nabla \psi \cdot \nabla \psi) \\ & = \sum_{|c| \leq |a|} \partial_t ((1+t)^{2\lambda} \tilde{\Gamma}^a \psi \Gamma^c (|\nabla \psi|^2)) \\ & \quad - 2\lambda(1+t)^{2\lambda-1} (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^c (|\nabla \psi|^2) - (1+t)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \tilde{\Gamma}^c (|\nabla \psi|^2). \end{aligned}$$

From this, (2-6) and Hardy's inequality, we have

$$(2-35) \quad \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} ((1+\tau)^\lambda \partial_t \nabla \psi \cdot \nabla \psi) dx d\tau \right| \\ \leq C\varepsilon^2 + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_{\mathbb{R}^3} (1+t)^{2\lambda} |\partial \tilde{\Gamma}^a \psi|^2 dx \\ + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau.$$

Part C: *Estimates of*

$$\sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} ((1+\tau)^{2\lambda} (\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi) dx d\tau \right|$$

and

$$\sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} ((1+\tau)^{2\lambda} (\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi) dx d\tau \right|.$$

A direct computation yields

$$\begin{aligned} & \tilde{\Gamma}^{b_1} ((\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi) \\ &= \partial_i \psi \partial_j \psi \partial_{ij} \tilde{\Gamma}^{b_1} \psi + \sum_{N-2 \leq |b_2| \leq |b_1|-1} (\nabla^2 \tilde{\Gamma}^{b_2} \psi)(\nabla \tilde{\Gamma}^{b_3} \psi) \nabla \tilde{\Gamma}^{b_4} \psi \\ & \quad + \sum_{|b_2| \leq N-3} (\nabla^2 \tilde{\Gamma}^{b_2} \psi)(\nabla \tilde{\Gamma}^{b_3} \psi) \nabla \tilde{\Gamma}^{b_4} \psi \\ &= \text{III}_1 + \text{III}_2 + \text{III}_3. \end{aligned}$$

As in the treatment of II_1 in Part B, we have

$$(2-36) \quad \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \text{III}_1 dx d\tau \right| \\ \leq C\varepsilon^2 + CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau.$$

By (2-6) and (2-9), for the term III_2 , we have

$$(2-37) \quad (1+t)^{4\lambda} \|(\partial_t \tilde{\Gamma}^a \psi)(\nabla^2 \tilde{\Gamma}^{b_2} \psi)(\nabla \tilde{\Gamma}^{b_3} \psi) \nabla \tilde{\Gamma}^{b_4} \psi\|_{L^1} \\ \leq (1+t)^{4\lambda} \|\langle r-t \rangle^{-1} (\nabla \tilde{\Gamma}^{b_3} \psi) \nabla \tilde{\Gamma}^{b_4} \psi\|_{L^\infty} \cdot \|\partial_t \tilde{\Gamma}^a \psi\| \cdot \|\langle r-t \rangle \nabla^2 \tilde{\Gamma}^{b_2} \psi\| \\ \leq CK\varepsilon (1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\| \sum_{|b_4| \leq |b_3|+1} \|\nabla \tilde{\Gamma}^{b_4} \psi\| + CK\varepsilon (1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\|^2 \\ + CK\varepsilon (1-\lambda)(1+t)^{\lambda-2} \sum_{|b_4| \leq |b_3|+1} \|\tilde{\Gamma}^{b_4} \psi\|^2.$$

By (2-6) and (2-8), for the term III_3 , one has

$$(2-38) \quad (1+t)^{4\lambda} \|(\partial_t \tilde{\Gamma}^a \psi)(\nabla^2 \tilde{\Gamma}^{b_2} \psi)(\nabla \tilde{\Gamma}^{b_3} \psi) \nabla \tilde{\Gamma}^{b_4} \psi\|_{L^1} \\ \leq CK\varepsilon(1+t)^\lambda \|\partial_t \tilde{\Gamma}^a \psi\| \sum_{|c| \leq |b_1|} \|\nabla \tilde{\Gamma}^c \psi\|.$$

Collecting (2-36)–(2-38) together with $0 \leq \lambda \leq 1$ (this means that the coefficient $CK\varepsilon(1-\lambda)(1+t)^{\lambda-2}$ of $\sum_{|b_4| \leq |b_3|+1} \|\tilde{\Gamma}^{b_4} \psi\|^2$ in the last line of (2-37) is nonnegative and in $L^1(0, \infty)$) yields

$$(2-39) \quad \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^{2\lambda} (\partial_t \tilde{\Gamma}^a \psi) \tilde{\Gamma}^{b_1} ((1+\tau)^{2\lambda} (\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi) dx d\tau \right| \\ \leq CK\varepsilon \sum_{0 \leq |a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \tilde{\Gamma}^a \psi|^2 dx d\tau \\ + CK\varepsilon \sum_{|b_4| \leq N} \int_0^t A(\tau) \int_{\mathbb{R}^3} |\tilde{\Gamma}^{b_4} \psi|^2 dx d\tau.$$

In addition,

$$2(1+t)^{3\lambda} (\Gamma^a \psi) \Gamma^{b_1} ((\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi) \\ = \operatorname{div}((1+t)^{3\lambda} (\Gamma^a \psi)(\nabla \psi) \Gamma^{b_1} (|\nabla \psi|^2)) - (1+t)^{3\lambda} (\nabla \Gamma^a \psi)(\nabla \psi) \Gamma^{b_1} (|\nabla \psi|^2) \\ - (1+t)^{3\lambda} (\Gamma^a \psi)(\Delta \psi) \Gamma^{b_1} (|\nabla \psi|^2) \\ + \sum_{|b_2| \leq |b_1|-1} (1+t)^{3\lambda} (\Gamma^a \psi)(\nabla^2 \Gamma^{b_2} \psi)(\nabla \Gamma^{b_3} \psi) \nabla \Gamma^{b_4} (|\psi|^2).$$

Together with (2-6) and Hardy's inequality this yields

$$(2-40) \quad \sum_{|b_1| \leq N} \left| \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda (\Gamma^a \psi) \Gamma^{b_1} ((1+\tau)^{2\lambda} (\partial_i \psi)(\partial_j \psi) \partial_{ij} \psi) dx d\tau \right| \\ \leq CK\varepsilon \sum_{|a| \leq N} \int_0^t \int_{\mathbb{R}^3} (1+\tau)^\lambda |\partial \Gamma^a \psi|^2 dx d\tau.$$

Therefore, substituting (2-29)–(2-30), (2-34)–(2-35), and (2-39)–(2-40) into (2-25) and utilizing the smallness of $\varepsilon > 0$ gives (2-21). \square

Based on Lemmas 2.4 and 2.5, we now prove Theorem 1.1.

Proof of Theorem 1.1. By Lemmas 2.4 and 2.5, one has that, for fixed $N \geq 8$,

$$E_N(t) \leq C\varepsilon^2 + C(1+K\varepsilon) \int_0^t A(t') E_N(t') dt'.$$

Choosing the constants $K > 0$ large and $\varepsilon > 0$ small, by Gronwall's inequality one gets that, for any $t \geq 0$,

$$E_N(t) \leq e^{C(1+K\varepsilon)\|A(t)\|_{L^1}} E_N(0) \leq \frac{1}{2} K^2 \varepsilon^2.$$

Thus, Theorem 1.1 is proved by the assumption that $E_N(t) \leq K^2 \varepsilon^2$ and a continuous induction argument. \square

3. Blowup for small data in case $\lambda > 1$

In this section, we shall prove the blowup result of Theorem 1.2 which is valid in case $\lambda > 1$.

Proof of Theorem 1.2. We divide the proof into two cases.

Case 1: $\gamma = 2$. Let (ρ, u) be a smooth solution of (1-1). For $l > 0$, we define

$$(3-1) \quad P(t, l) = \int_{|x|>l} \eta(x, l)(\rho(t, x) - \bar{\rho}) dx,$$

where

$$\eta(x, l) = |x|^{-1}(|x| - l)^2.$$

Employing the first equation in (1-1) and an integration by parts, we see that

$$\begin{aligned} \partial_t P(t, l) &= \int_{|x|>l} \eta(x, l) \partial_t (\rho(t, x) - \bar{\rho}) dx = - \int_{|x|>l} \eta(x, l) \operatorname{div}(\rho u)(t, x) dx \\ &= \int_{|x|>l} (\nabla_x \eta)(x, l) \cdot (\rho u)(t, x) dx, \end{aligned}$$

where we have used the fact that $\eta(x, l) = 0$ on $|x| = l$ and that $u(t, x) = 0$ for $|x| \geq t + M$.

By differentiating $\partial_t P(t, l)$ again and using the second equation in (1-1), we find that

$$\begin{aligned} (3-2) \quad \partial_t^2 P(t, l) &= \int_{|x|>l} (\nabla_x \eta)(x, l) \cdot \partial_t (\rho u)(t, x) dx \\ &= - \sum_{i,j} \int_{|x|>l} (\partial_{x_i} \eta) \partial_{x_j} (\rho u_i u_j) dx - \int_{|x|>l} (\nabla_x \eta)(x, l) \cdot \nabla(p - \bar{p}) dx \\ &\quad - \frac{\mu}{(1+t)^\lambda} \int_{|x|>l} (\nabla_x \eta)(x, l) \cdot (\rho u)(t, x) dx, \end{aligned}$$

where $\nabla_x \eta(x, l) = |x|^{-3}(|x|^2 - l^2)x$ vanishes on $|x| = l$ and $\bar{p} = p(\bar{\rho})$. Integration by parts implies that

$$(3-3) \quad \begin{aligned} \partial_t^2 P(t, l) + \frac{\mu}{(1+t)^\lambda} \partial_t P(t, l) \\ = \sum_{i,j} \int_{|x|>l} (\partial_{x_i x_j}^2 \eta) \rho u_i u_j dx + \int_{|x|>l} (\Delta \eta)(p - \bar{p}) dx \\ \equiv J_1(t, l) + J_2(t, l), \end{aligned}$$

where we have used that $p - \bar{p}$ vanishes for $|x| \geq t + M$. A direct computation of $\partial_{x_i x_j}^2 \eta$ shows that

$$(3-4) \quad \begin{aligned} J_1(t, l) = \int_{|x|>l} \frac{2l^2}{|x|^3} \rho \left(\frac{x}{|x|} \cdot u \right)^2 dx \\ - \int_{|x|>l} \frac{|x|^2 - l^2}{|x|^3} \rho \left(\frac{x}{|x|} \cdot u \right)^2 dx + \int_{|x|>l} \frac{|x|^2 - l^2}{|x|^3} \rho |u|^2 dx \geq 0. \end{aligned}$$

On the other hand, notice that

$$(3-5) \quad \partial_l^2 \eta(x, l) = 2|x|^{-1} = \Delta_x \eta(x, l).$$

Then

$$(3-6) \quad J_2(t, l) = \int_{|x|>l} \partial_l^2 \eta(x, l) (p(t, x) - \bar{p}) dx = \partial_l^2 \int_{|x|>l} \eta(x, l) (p(t, x) - \bar{p}) dx,$$

where we have used the fact that η and $\partial_l \eta$ vanish on $|x| = l$. Combining (3-3)–(3-6), we arrive at

$$(3-7) \quad \partial_t^2 P(t, l) - \partial_l^2 P(t, l) + \frac{\mu}{(1+t)^\lambda} \partial_t P(t, l) = f(t, l) \equiv J_1(t, l) + G(t, l) \geq G(t, l),$$

where

$$(3-8) \quad G(t, l) = \partial_l^2 \int_{|x|>l} \eta(x, l) (p - \bar{p} - (\rho - \bar{\rho})) dx = \int_{|x|>l} 2|x|^{-1} (p - \bar{p} - (\rho - \bar{\rho})) dx.$$

Thanks to $\gamma = 2$ and the sound speed $\bar{c} = \sqrt{2A\bar{\rho}} = 1$, we have

$$(3-9) \quad p - \bar{p} - (\rho - \bar{\rho}) = A(\rho^2 - \bar{\rho}^2 - 2\bar{\rho}(\rho - \bar{\rho})) = A(\rho - \bar{\rho})^2.$$

Substituting (3-9) into (3-8) gives

$$G(t, l) \geq 0.$$

For M_0 satisfying the condition (1-11), let $\Sigma \equiv \{(t, l) : t \geq 0, t + M_0 \leq l \leq t + M\}$ be the strip domain. By applying Riemann's representation (see [Courant and Hilbert

1962, §5.5]) with the assumptions (1-9)–(1-11), we see that the solution $P(t, l)$ to (3-7) is nonnegative in Σ . We put its proof in the Appendix. Rewrite (3-7) as

$$\partial_t^2 P(t, l) - \partial_l^2 P(t, l) + \frac{\mu}{(1+t)^\lambda} (\partial_t P(t, l) - \partial_l P(t, l)) = f(t, l) - \frac{\mu}{(1+t)^\lambda} \partial_l P(t, l).$$

By the method of characteristics we have

$$\begin{aligned} P(t, l) = & \frac{1}{2} P(0, l+t) + \frac{1}{2\beta(t)} P(0, l-t) + \frac{1}{2} \int_0^t \frac{1}{\beta(\tau)} \frac{\mu}{(1+\tau)^\lambda} P(0, l+t-2\tau) d\tau \\ & + \int_0^t \frac{1}{\beta(\tau)} \partial_t P(0, l+t-2\tau) d\tau + \frac{1}{2} \int_0^t \int_{l-t+\tau}^{l+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} f(\tau, y) dy d\tau \\ & + \frac{1}{2} \int_0^t \frac{\beta(\tau)}{\beta(t)} \frac{\mu}{(1+\tau)^\lambda} P(\tau, l-t+\tau) d\tau \\ & + \frac{1}{2} \int_0^t \int_\tau^t \frac{\beta(\tau)}{\beta(s)} \frac{\mu^2}{(1+\tau)^\lambda (1+s)^\lambda} P(\tau, l+t-2s+\tau) ds d\tau \\ & - \frac{1}{2} \int_0^t \frac{\mu}{(1+\tau)^\lambda} P(\tau, l+t-\tau) d\tau; \end{aligned}$$

see (1-12). Together with assumptions (1-9)–(1-10) and $P(t, l) \geq 0$ in Σ this yields, for $l \geq t + M_0$,

$$\begin{aligned} (3-10) \quad P(t, l) \geq & \frac{1}{2\beta(t)} q_0(l-t) + \frac{1}{2} \int_0^t \int_{l-t+\tau}^{l+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} G(\tau, y) dy d\tau \\ & - \frac{1}{2} \int_0^t \frac{\mu}{(1+\tau)^\lambda} P(\tau, l+t-\tau) d\tau. \end{aligned}$$

Define the function

$$(3-11) \quad F(t) \equiv \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} P(\tau, l) \frac{dl}{l} d\tau.$$

Then, by (3-10), we have that

$$\begin{aligned} (3-12) \quad F''(t) = & \int_{t+M_0}^{t+M} P(t, l) \frac{dl}{l} \\ \geq & \frac{1}{2\beta(t)} \int_{t+M_0}^{t+M} q_0(l-t) \frac{dl}{l} + \frac{1}{2} \int_{t+M_0}^{t+M} \int_0^t \int_{l-t+\tau}^{l+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} G(\tau, y) dy d\tau \frac{dl}{l} \\ & - \frac{1}{2} \int_{t+M_0}^{t+M} \int_0^t \frac{\mu}{(1+\tau)^\lambda} P(\tau, l+t-\tau) d\tau \frac{dl}{l} \\ \equiv & J_3 + J_4 - J_5. \end{aligned}$$

From $\lambda > 1$ and assumption (1-9), we see that

$$(3-13) \quad J_3 \geq \frac{c_1}{t+M} \int_{t+M_0}^{t+M} q_0(l-t) dl = \frac{c_1}{t+M} \int_{M_0}^M q_0(l) dl = \frac{c_2 \varepsilon}{t+M},$$

where $c_1, c_2 > 0$ are constants independent of ε . Note that $P(\tau, y)$ is supported in $\{y : y \leq \tau + M\}$ and nonnegative in Σ . Hence, there exists a constant $C_1 > 0$ such that

$$(3-14) \quad J_5 \leq \frac{C_1}{(1+t)^\lambda} \int_0^t \int_{\tau+M_0}^{\tau+M} P(\tau, y) \frac{dy}{y} d\tau = \frac{C_1}{(1+t)^\lambda} F'(t).$$

Substituting (3-14) into (3-12) gives

$$(3-15) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq J_3 + J_4.$$

To bound J_4 from below, we write

$$(3-16) \quad \begin{aligned} J_4 &= \frac{1}{2} \int_0^{t-M_1} \int_{\tau+M_0}^{\tau+M} G(\tau, y) \int_{t+M_0}^{y+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} \frac{dl}{l} dy d\tau \\ &\quad + \frac{1}{2} \int_{t-M_1}^t \int_{\tau+M_0}^{2t-\tau+M_0} G(\tau, y) \int_{t+M_0}^{y+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} \frac{dl}{l} dy d\tau \\ &\quad + \frac{1}{2} \int_{t-M_1}^t \int_{2t-\tau+M_0}^{\tau+M} G(\tau, y) \int_{y-t+\tau}^{y+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} \frac{dl}{l} dy d\tau \\ &\equiv J_{4,1} + J_{4,2} + J_{4,3}, \end{aligned}$$

where $M_1 = (M - M_0)/2$. For $t < M_1$, $t - M_1$ in the limits of integration is replaced by 0. By $\lambda > 1$, for the integrand in $J_{4,1}$ we have that

$$(3-17) \quad \int_{t+M_0}^{y+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} \frac{dl}{l} \geq c \frac{y-\tau-M_0}{t+M} \geq c \frac{(t-\tau)(y-\tau-M_0)^2}{(t+M)^2}.$$

Analogously, for the integrands in $J_{4,2}$ and $J_{4,3}$ we have that

$$(3-18) \quad \int_{t+M_0}^{y+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} \frac{dl}{l} \geq c \frac{(t-\tau)(y-\tau-M_0)^2}{(t+M)^2}$$

and

$$(3-19) \quad \int_{y-t+\tau}^{y+t-\tau} \frac{\beta(\tau)}{\beta\left(\frac{l+t+\tau-y}{2}\right)} \frac{dl}{l} \geq c \frac{t-\tau}{t+M} \geq c \frac{(t-\tau)(y-\tau-M_0)^2}{(t+M)^2},$$

where $c > 0$ is a constant. Substituting (3-17)–(3-19) into (3-16) yields

$$J_4 \geq \frac{c}{(t+M)^2} \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} (y-\tau-M_0)^2 \partial_y^2 \tilde{G}(\tau, y) dy d\tau,$$

where $\tilde{G}(t, l) = \int_{|x|>l} \eta(x, l)(p - \bar{p} - (\rho - \bar{\rho})) dx$. Note that $\tilde{G}(\tau, y) = \partial_y \tilde{G}(\tau, y) = 0$ for $y = \tau + M$. Thus, it follows from the integration by parts together with (3-8)–(3-9) that

$$\begin{aligned}
 (3-20) \quad J_4 &\geq \frac{c}{(t+M)^2} \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \tilde{G}(\tau, y) dy d\tau \\
 &\geq \frac{c}{(t+M)^2} \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \int_{|x|>y} \eta(x, y)(\rho(\tau, x) - \bar{\rho})^2 dx dy d\tau \\
 &\equiv \frac{c}{(t+M)^2} J_6.
 \end{aligned}$$

By applying the Cauchy–Schwartz inequality to $F(t)$ defined by (3-11), we arrive at

$$(3-21) \quad F^2(t) \leq J_6 \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \int_{y<|x|<\tau+M} \eta(x, y) dx \frac{dy}{y^2} d\tau \equiv J_6 J_7.$$

We estimate J_7 as

$$\begin{aligned}
 (3-22) \quad J_7 &= \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \int_{y<|x|<\tau+M} \frac{(|x|-y)^2}{|x|} dx \frac{dy}{y^2} d\tau \\
 &= \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \int_y^{\tau+M} 4\pi l(l-y)^2 dl \frac{dy}{y^2} d\tau \\
 &\leq C \int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} (\tau+M)(\tau+M-y)^3 \frac{dy}{y^2} d\tau \\
 &\leq C \int_0^t (t-\tau)(\tau+M) \int_{\tau+M_0}^{\tau+M} \frac{dy}{y^2} d\tau \\
 &\leq C \int_0^t \frac{t-\tau}{\tau+M} d\tau \leq C(t+M) \log(t/M+1).
 \end{aligned}$$

Combining (3-13), (3-15) and (3-20)–(3-22) gives the ordinary differential inequalities

$$(3-23) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq \frac{c_2 \varepsilon}{t+M}, \quad t \geq 0,$$

$$(3-24) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C[(t+M)^3 \log(t/M+1)]^{-1} F^2(t), \quad t \geq 0.$$

Next, we apply (3-23)–(3-24) to prove that the lifespan T_ε of smooth solution $F(t)$ is finite for all $0 < \varepsilon \leq \varepsilon_0$. The fact that $F(0) = F'(0) = 0$, together with (3-23), yields

$$(3-25) \quad F'(t) \geq C\varepsilon \log(t/M+1), \quad t \geq 0,$$

$$(3-26) \quad F(t) \geq C\varepsilon(t+M) \log(t/M+1), \quad t \geq t_1 \equiv Me^2,$$

where the constant $C > 0$ is independent of ε . Substituting (3-26) into (3-24) derives

$$F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C\varepsilon^2(t+M)^{-1} \log(t/M+1), \quad t \geq t_1,$$

which leads to the improvement

$$(3-27) \quad F(t) \geq C\varepsilon^2(t+M) \log^2(t/M+1), \quad t \geq t_2 \equiv Me^3 > t_1.$$

Substituting this into (3-24) derives

$$(3-28) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C\varepsilon^2(t+M)^{-2} \log(t/M+1) F(t), \quad t \geq t_2.$$

It follows from (3-25) that $F'(t) \geq 0$ for $t \geq 0$. Then multiplying (3-28) by $F'(t)$ and integrating from t_3 (which will be chosen later) to t yield

$$F'(t)^2 \geq C_2 F'(t_3)^2 + C_3 \varepsilon^2 \int_{t_3}^t (s+M)^{-2} \log(s/M+1) [F(s)^2]' ds.$$

Integrating by parts yields

$$(3-29) \quad \begin{aligned} F'(t)^2 &\geq C_2 F'(t_3)^2 \\ &\quad + C_3 \varepsilon^2 \left((t+M)^{-2} \log(t/M+1) F(t)^2 - (t_3+M)^{-2} \log(t_3/M+1) F(t_3)^2 \right) \\ &\quad - \int_{t_3}^t \left(\frac{\log(s/M+1)}{(s+M)^2} \right)' F(s)^2 ds, \quad t \geq t_3, \end{aligned}$$

where

$$\left(\frac{\log(s/M+1)}{(s+M)^2} \right)' \leq 0$$

for $t \geq t_3 \geq t_2$. On the other hand, (3-23) implies

$$\left(e^{-\frac{C_1}{\lambda-1}[(1+t)^{1-\lambda}-1]} F'(t) \right)' \geq 0, \quad t \geq 0,$$

which yields for $0 \leq t \leq \tau$

$$(3-30) \quad F'(t) \leq e^{\frac{C_1}{\lambda-1}[(1+t)^{1-\lambda}-(1+\tau)^{1-\lambda}]} F'(\tau).$$

Together with $F(0) = 0$, this yields

$$(3-31) \quad F(t) = \int_0^t F'(s) ds \leq C_4 t F'(t), \quad t > 0.$$

Choose

$$(3-32) \quad t_3 = M \left(e^{\frac{C_2}{2C_3 C_4 \varepsilon^2}} - 1 \right)$$

which satisfies $2C_3C_4\log(t_3/M+1)\varepsilon^2 = C_2$. Together with (3-29) and (3-31), this yields

$$(3-33) \quad F'(t) \geq \sqrt{C_3}\varepsilon(t+M)^{-1}\log^{\frac{1}{2}}(t/M+1)F(t), \quad t \geq t_3.$$

By integrating (3-33) from t_3 to t , we arrive at

$$\log \frac{F(t)}{F(t_3)} \geq C\varepsilon \log^{\frac{3}{2}}\left(\frac{t+M}{t_3+M}\right), \quad t \geq t_3.$$

If $t \geq t_4 \equiv Ct_3^2$, we then have

$$\log \frac{F(t)}{F(t_3)} \geq 8 \log(t/M+1).$$

Together with (3-27) for $F(t_3)$, this yields

$$(3-34) \quad F(t) \geq C\varepsilon^2(t+M)^8, \quad t \geq t_4.$$

Substituting this into (3-24) derives

$$F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C\varepsilon F(t)^{\frac{3}{2}}, \quad t \geq t_4.$$

Multiplying this differential inequality by $F'(t)$ and integrating from t_4 to t yields

$$F'(t)^2 \geq C\varepsilon \left(F(t)^{\frac{5}{2}} - F(t_4)^{\frac{5}{2}} \right).$$

On the other hand, (3-30) and (3-31) imply that, for $t \geq t_4$,

$$F(t) = F'(\xi)(t-t_4) + F(t_4) \geq CF'(t_4)(t-t_4) \geq CF(t_4) \frac{t-t_4}{t_4},$$

where $t_4 \leq \xi \leq t$. If $t \geq t_5 \equiv Ct_4$, then we have

$$F(t)^{\frac{5}{2}} - F(t_4)^{\frac{5}{2}} \geq \frac{1}{2} F(t)^{\frac{5}{2}}.$$

Thus

$$(3-35) \quad F'(t) \geq C\sqrt{\varepsilon}F(t)^{\frac{5}{4}}, \quad t \geq t_5.$$

If $T_\varepsilon > 2t_5$, then integrating (3-35) from t_5 to T_ε derives

$$F(t_5)^{-\frac{1}{4}} - F(T_\varepsilon)^{-\frac{1}{4}} \geq C\sqrt{\varepsilon}T_\varepsilon.$$

We see from (3-34) and $t_5 = Ct_3^2$ that

$$F(t_5) \geq C\varepsilon^2 e^{C/\varepsilon^2},$$

which together with $F(T_\varepsilon) > 0$ is a contradiction. Thus, $T_\varepsilon \leq 2t_5 = Ct_3^2$. From the choice of t_3 in (3-32), we see that $T_\varepsilon \leq e^{C/\varepsilon^2}$.

Case 2: $\gamma > 1$ and $\gamma \neq 2$. Recall that the sound speed is $\bar{c} = \sqrt{\gamma A \bar{\rho}^{\gamma-1}} = 1$. Instead of (3-9) we have

$$p - \bar{p} - (\rho - \bar{\rho}) = A(\rho^\gamma - \bar{\rho}^\gamma - \gamma \bar{\rho}^{\gamma-1}(\rho - \bar{\rho})) \equiv A\psi(\rho, \bar{\rho}).$$

The convexity of ρ^γ for $\gamma > 1$ implies that $\psi(\rho, \bar{\rho})$ is positive for $\rho \neq \bar{\rho}$. Applying Taylor's theorem, we have

$$\psi(\rho, \bar{\rho}) \geq C(\gamma, \bar{\rho}) \Phi_\gamma(\rho, \bar{\rho}),$$

where $C(\gamma, \bar{\rho})$ is a positive constant and Φ_γ is given by

$$\Phi_\gamma(\rho, \bar{\rho}) = \begin{cases} (\bar{\rho} - \rho)^\gamma, & \rho < \frac{1}{2}\bar{\rho}, \\ (\rho - \bar{\rho})^2, & \frac{1}{2}\bar{\rho} \leq \rho \leq 2\bar{\rho}, \\ (\rho - \bar{\rho})^\gamma, & \rho > 2\bar{\rho}. \end{cases}$$

For $\gamma > 2$, we have that $(\bar{\rho} - \rho)^\gamma = (\bar{\rho} - \rho)^2(\bar{\rho} - \rho)^{\gamma-2} \geq C(\gamma, \bar{\rho})(\rho - \bar{\rho})^2$ for $2\rho < \bar{\rho}$ and $(\rho - \bar{\rho})^\gamma = (\rho - \bar{\rho})^2(\rho - \bar{\rho})^{\gamma-2} \geq C(\gamma, \bar{\rho})(\rho - \bar{\rho})^2$ for $\rho > 2\bar{\rho}$. Thus, $\psi(\rho, \bar{\rho}) \geq C(\gamma, \bar{\rho})(\rho - \bar{\rho})^2$. In this case, Theorem 1.2 can be shown completely analogously to Case 1.

Next we treat the case $1 < \gamma < 2$. We define $F(t)$ as in (3-11),

$$F(t) = \int_0^t \int_{\tau+M_0}^{\tau+M} \frac{1}{l} \int_{|x|>l} \frac{(|x| - l)^2}{|x|} (\rho(\tau, x) - \bar{\rho}) dx dl d\tau.$$

Similarly to the case of $\gamma = 2$, we have

$$(3-36) \quad F''(t) \geq J_3 + J_4 - J_5,$$

where

$$J_3 \geq \frac{C\varepsilon}{t+M},$$

$$J_4 \geq C(t+M)^{-2} \tilde{J}_6,$$

$$J_5 \leq \frac{C_1}{(1+t)^\lambda} F'(t),$$

and

$$\tilde{J}_6 = \int_0^t (t - \tau) \int_{\tau+M_0}^{\tau+M} \int_{|x|>y} \frac{(|x| - y)^2}{|x|} \Phi_\gamma(\rho(\tau, x) - \bar{\rho}) dx dy d\tau.$$

Denote $\Omega_1 = \{(\tau, x) : \bar{\rho} \leq \rho(\tau, x) \leq 2\bar{\rho}\}$, $\Omega_2 = \{(\tau, x) : \rho(\tau, x) > 2\bar{\rho}\}$, and $\Omega_3 = \{(\tau, x) : \rho(\tau, x) < \bar{\rho}\}$. Divide $F(t)$ into a sum of the three integrals over the domains Ω_i ($1 \leq i \leq 3$)

$$F(t) = F_1(t) + F_2(t) + F_3(t) \equiv \int_{\Omega_1} \cdots + \int_{\Omega_2} \cdots + \int_{\Omega_3} \cdots.$$

Corresponding to the three parts of $F(t)$, we define $\tilde{J}_6 \equiv \tilde{J}_{6,1} + \tilde{J}_{6,2} + \tilde{J}_{6,3}$. In view of $F(t) \geq 0$ and $F_3(t) \leq 0$, we have

$$F(t) \leq F_1(t) + F_2(t).$$

Applying Hölder's inequality for the domains Ω_1 and Ω_2 , we obtain that

$$\begin{aligned} F(t) &\leq \tilde{J}_{6,1}^{\frac{1}{2}} \left(\int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \frac{1}{y^2} \int_{y<|x|\leq\tau+M} \frac{(|x|-y)^2}{|x|} dx dy d\tau \right)^{\frac{1}{2}} \\ &\quad + \tilde{J}_{6,2}^{\frac{1}{\gamma}} \left(\int_0^t (t-\tau) \int_{\tau+M_0}^{\tau+M} \frac{1}{y^{\frac{\gamma}{\gamma-1}}} \int_{y<|x|\leq\tau+M} \frac{(|x|-y)^2}{|x|} dx dy d\tau \right)^{\frac{\gamma-1}{\gamma}} \\ &\leq \tilde{J}_6^{\frac{1}{2}} (t+M)^{\frac{1}{2}} \log^{\frac{1}{2}}(t/M+1) + \tilde{J}_6^{\frac{1}{\gamma}} (t+M)^{\frac{\gamma-1}{\gamma}} \\ &= (\tilde{J}_6(t+M)^{-1})^{\frac{1}{2}} (t+M) \log^{\frac{1}{2}}(t/M+1) + (\tilde{J}_6(t+M)^{-1})^{\frac{1}{\gamma}} (t+M). \end{aligned}$$

In view of $1 < \gamma < 2$, we have $\frac{1}{2\gamma} < \frac{1}{2} < \frac{1}{\gamma}$. Applying Young's inequality yields

$$F(t) \leq \left((\tilde{J}_6(t+M)^{-1})^{\frac{1}{2\gamma}} + (\tilde{J}_6(t+M)^{-1})^{\frac{1}{\gamma}} \right) (t+M) \log^{\frac{1}{2}}(t/M+1), \quad t \geq \tilde{t}_1 \equiv Me.$$

Together with the fact that $F(t) \geq C\varepsilon(t+M) \log(t/M+1)$, this yields

$$\tilde{J}_6 \geq CF(t)^\gamma (t+M)^{1-\gamma} \log^{-\frac{\gamma}{2}}(t/M+1), \quad t \geq \tilde{t}_1.$$

Substituting this into (3-36) yields

$$(3-37) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq \frac{C\varepsilon}{t+M}, \quad t \geq 0,$$

$$(3-38) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq CF(t)^\gamma (t+M)^{-1-\gamma} \log^{-\frac{\gamma}{2}}(t/M+1), \quad t \geq \tilde{t}_1.$$

Substituting $F(t) \geq C\varepsilon(t+M) \log(t/M+1)$ into (3-38) yields

$$F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C\varepsilon^\gamma (t+M)^{-1} \log^{\frac{\gamma}{2}}(t/M+1).$$

Integrating this yields

$$F(t) \geq C\varepsilon^\gamma (t+M) \log^{\frac{\gamma+2}{2}}(t/M+1).$$

Substituting this into (3-38) again gives

$$\begin{aligned} F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \\ \geq C\varepsilon^{\gamma^2} (t+M)^{-1} \log^{\frac{\gamma(\gamma+1)}{2}}(t/M+1) = C\varepsilon^{\gamma^2} (t+M)^{-1} \log^{\frac{\gamma(\gamma^2-1)}{2(\gamma-1)}}(t/M+1). \end{aligned}$$

Repeating this process n times, we see that

$$(3-39) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C\varepsilon^{\gamma^n} (t+M)^{-1} \log^{\frac{\gamma(\gamma^n-1)}{2(\gamma-1)}} (t/M+1),$$

where $n = [\log_\gamma 2]$. Solving (3-39) yields

$$F(t) \geq C\varepsilon^{\gamma^n} (t+M) \log^{\frac{\gamma(\gamma^n-1)}{2(\gamma-1)}+1} (t/M+1), \quad t \geq \tilde{t}_2,$$

where $\tilde{t}_2 > 0$ is a constant only depending on γ . Substituting this into (3-38) derives

$$(3-40) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq CF(t)\varepsilon^{\gamma^n(\gamma-1)} (t+M)^{-2} \log^{\frac{\gamma^{n+1}-2}{2}} (t/M+1), \quad t \geq \tilde{t}_2,$$

where $\frac{1}{2}(\gamma^{n+1}-2) > 0$ by the choice of $n = [\log_\gamma 2]$. Since (3-40) is analogous to (3-28), as in Case 1, we can choose

$$\tilde{t}_3 = O\left(e^{C\varepsilon^{-\frac{2\gamma^n(\gamma-1)}{\gamma^{n+1}-2}}}\right)$$

such that

$$F'(t) \geq C\varepsilon^{\frac{\gamma^n(\gamma-1)}{2}} (t+M)^{-1} \log^{\frac{\gamma^{n+1}-2}{4}} (t/M+1) F(t), \quad t \geq \tilde{t}_3,$$

which is similar to (3-33) and yields

$$(3-41) \quad F(t) \geq C\varepsilon^{C_\gamma} (t+M)^{\frac{2(\gamma+2)}{\gamma-1}}, \quad t \geq \tilde{t}_4 \equiv C\tilde{t}_3^2,$$

where $C_\gamma > 0$ is a constant depending on γ . Substituting (3-41) into (3-38) yields

$$(3-42) \quad F''(t) + \frac{C_1}{(1+t)^\lambda} F'(t) \geq C\varepsilon^{C_\gamma} F(t)^{\frac{\gamma+1}{2}}, \quad t \geq \tilde{t}_4.$$

Multiplying (3-42) by $F'(t)$ and integrating over the variable t as in Case 1, we have

$$F'(t) \geq C\varepsilon^{C_\gamma} F(t)^{\frac{\gamma+3}{4}}, \quad t \geq \tilde{t}_5 \equiv C\tilde{t}_4.$$

Together with $\gamma > 1$ and the choice of \tilde{t}_3 , this yields $T_\varepsilon < \infty$.

Both Case 1 and Case 2 complete the proof of Theorem 1.2. \square

4. Blowup for large data

In this section, we establish a blowup result for large amplitude smooth solutions of (1-1) which is valid for all $\lambda \geq 0$. More precisely, instead of (1-1) we consider

the Cauchy problem

$$(4-1) \quad \begin{cases} \partial_t \rho + \operatorname{div}(\rho u) = 0, \\ \partial_t(\rho u) + \operatorname{div}(\rho u \otimes u + p I_3) = -\frac{\mu}{(1+t)^\lambda} \rho u, \\ \rho(0, x) = \bar{\rho} + \tilde{\rho}_0(x), \quad u(0, x) = \tilde{u}_0(x), \end{cases}$$

where $\tilde{\rho}_0, \tilde{u}_0 \in C_0^\infty(\mathbb{R}^3)$, $\operatorname{supp} \tilde{\rho}_0, \operatorname{supp} \tilde{u}_0 \subseteq B(0, M) \equiv \{x : |x| \leq M\}$, and $\rho(0, \cdot) > 0$. Motivated by the treatment of the special case of $\lambda = 0$ in [Sideris et al. 2003], we introduce the functions

$$\begin{aligned} H(t) &\equiv \int_{\mathbb{R}^3} x \cdot (\rho u)(t, x) dx, & L(t) &\equiv \int_{\mathbb{R}^3} (\rho(t, x) - \bar{\rho}) dx, \\ \gamma(t) &\equiv (t + M)^2 \left(L(0) + \frac{4\pi^2 \bar{\rho}}{3} (t + M)^3 \right), \end{aligned}$$

and also remind the reader of the definition of the function β in (1-12).

Then we have the following result:

Theorem 4.1. *Suppose that $L(0) \geq 0$ and*

$$(4-2) \quad H(0) \int_0^{T^*} \frac{d\tau}{\gamma(\tau)\beta(\tau)} > 1.$$

for some $T^ > 0$. Then $T < T^*$ holds for any solution $(\rho, u) \in C^1([0, T] \times \mathbb{R}^3)$ of (4-1).*

Proof. From the first equation of (4-1), we see that

$$L'(t) = - \int_{\mathbb{R}^3} \operatorname{div}(\rho u) dx = 0,$$

which implies $L(t) = L(0)$. Applying the second equation of (4-1), we find that

$$H'(t) = \int_{\mathbb{R}^3} x \cdot \partial_t(\rho u)(t, x) dx = \int_{\mathbb{R}^3} x \cdot \left[-\operatorname{div}(\rho u \otimes u) - \nabla p - \frac{\mu}{(1+t)^\lambda} \rho u \right] dx.$$

An integration by parts gives

$$(4-3) \quad H'(t) + \frac{\mu}{(1+t)^\lambda} H(t) = \int_{\mathbb{R}^3} (\rho |u|^2 + 3(p(\rho) - p(\bar{\rho}))) dx.$$

Note that the convexity of $p = A\rho^\gamma$ for $\gamma > 1$ and $c(\bar{\rho}) = 1$ imply that

$$(4-4) \quad \int_{\mathbb{R}^3} (p(\rho) - p(\bar{\rho})) dx \geq \int_{\mathbb{R}^3} A\gamma \bar{\rho}^{\gamma-1} (\rho - \bar{\rho}) dx = L(0).$$

Furthermore, by applying the Cauchy–Schwartz inequality to $H(t)$ and taking into account $\text{supp } u(t, \cdot) \subseteq B(0, M+t)$ for any fixed $t \geq 0$, we have

$$\begin{aligned}
 (4-5) \quad H(t)^2 &\leq \left(\int_{\mathbb{R}^3} \rho |u|^2 dx \right) \left(\int_{|x| \leq t+M} \rho |x|^2 dx \right) \\
 &\leq (t+M)^2 \left(L(0) + \frac{4\pi^2 \bar{\rho}}{3} (t+M)^3 \right) \int_{\mathbb{R}^3} \rho |u|^2 dx \\
 &= \gamma(t) \int_{\mathbb{R}^3} \rho |u|^2 dx.
 \end{aligned}$$

Substituting (4-4)–(4-5) into (4-3) yields

$$(4-6) \quad H'(t) + \frac{\mu}{(1+t)^\lambda} H(t) \geq \frac{H(t)^2}{\gamma(t)} + 3L(0).$$

Together with $L(0) \geq 0$ and $H(0) > 0$ due to (4-2), this shows that $H(t) > 0$ for all $t \in [0, T]$. Denoting $G(t) \equiv \beta(t)H(t)$, from (1-12) and (4-6) we then get that

$$(4-7) \quad G'(t) \geq \frac{G^2(t)}{\gamma(t)\beta(t)}.$$

Now suppose that $T \geq T^*$. Then integrating (4-7) from 0 to T yields

$$\frac{1}{H(0)} - \frac{1}{G(T)} \geq \int_0^T \frac{d\tau}{\gamma(\tau)\beta(\tau)} \geq \int_0^{T^*} \frac{d\tau}{\gamma(\tau)\beta(\tau)},$$

which is a contradiction in view of $G(T) > 0$ and (4-2). \square

Appendix: Proof of the nonnegativity of $P(t, l)$ in $\Sigma \equiv \{(t, l) : t \geq 0, t + M_0 \leq l \leq t + M\}$

We fixed a point $A = (t_A, l_A) \in \Sigma$. In the characteristic coordinates $\xi = 1 + t - l$ and $\zeta = 1 + t + l$, (3-7) can be written as

$$(A-1) \quad \mathcal{L}\bar{P} \equiv \partial_{\xi\zeta}^2 \bar{P} + \frac{2^{\lambda-2}\mu}{(\xi+\zeta)^\lambda} (\partial_\xi \bar{P} + \partial_\zeta \bar{P}) = \frac{\bar{f}}{4},$$

where $\bar{P}(\xi, \zeta) \equiv P\left(\frac{\zeta+\xi}{2} - 1, \frac{\zeta-\xi}{2}\right)$. The adjoint operator \mathcal{L}^* of \mathcal{L} has the form

$$(A-2) \quad \mathcal{L}^*\mathcal{R} \equiv \partial_{\xi\zeta}^2 \mathcal{R} - \frac{2^{\lambda-2}\mu}{(\xi+\zeta)^\lambda} (\partial_\xi \mathcal{R} + \partial_\zeta \mathcal{R}) + \frac{2^{\lambda-1}\mu\lambda}{(\xi+\zeta)^{\lambda+1}} \mathcal{R}.$$

For the point $A = (\xi_A, \zeta_A)$ with $\xi_A + \zeta_A = 2(1 + t_A) \geq 2$, write $B = (2 - \zeta_A, \zeta_A)$ and $C = (\xi_A, 2 - \xi_A)$, and let \mathcal{D} the domain surrounded by the triangle ABC (see Figure 1 below).

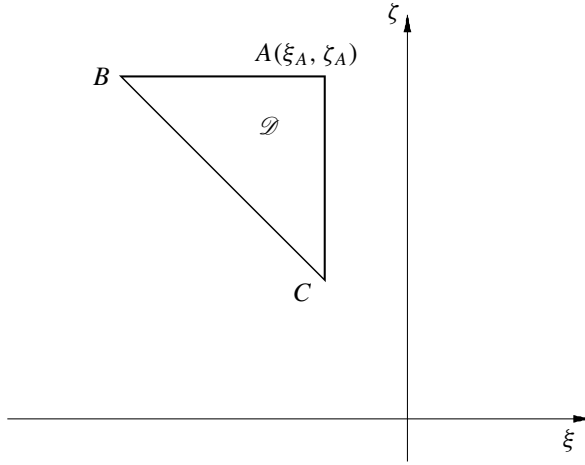


Figure 1. (ξ, ζ) -plane.

Let the numbers a and b satisfy $a + b = 1$ and $ab = \frac{1}{2}\mu\lambda$. We define

$$(A-3) \quad z \equiv -\frac{(\xi_A - \xi)(\zeta_A - \zeta)}{(\xi_A + \zeta_A)(\xi + \zeta)}$$

and

$$(A-4) \quad \mathcal{R}(\xi, \zeta; \xi_A, \zeta_A) \equiv \left[\frac{\beta(\xi + \zeta - 1)}{\beta(\xi_A + \zeta_A - 1)} \right]^{2^{\lambda-2}} \Psi(a, b, 1; z);$$

here the definition of function β is given in (1-12) and Ψ is the hypergeometric function. From this and direct calculation, we infer

$$(A-5) \quad \mathcal{L}^* \mathcal{R} = \left[\frac{2^{\lambda-2}\mu\lambda}{(\xi + \zeta)^{\lambda+1}} - \frac{\mu\lambda}{2(\xi + \zeta)^2} - \frac{4^{\lambda-2}\mu^2}{(\xi + \zeta)^{2\lambda}} \right] \mathcal{R}.$$

On the other hand, from (A-1)–(A-2) we arrive at

$$\mathcal{R} \mathcal{L} \bar{P} - \bar{P} \mathcal{L}^* \mathcal{R} = \partial_\zeta \left(\mathcal{R} \partial_\xi \bar{P} + \frac{2^{\lambda-2}\mu}{(\xi + \zeta)^\lambda} \mathcal{R} \bar{P} \right) - \partial_\xi \left(\bar{P} \partial_\zeta \mathcal{R} - \frac{2^{\lambda-2}\mu}{(\xi + \zeta)^\lambda} \mathcal{R} \bar{P} \right).$$

Integrating this over \mathcal{D} yields

$$(A-6) \quad \begin{aligned} \bar{P}(A) &= \frac{1}{2} \mathcal{R}(C; A) \bar{P}(C) + \frac{1}{2} \mathcal{R}(B; A) \bar{P}(B) \\ &+ \iint_{\mathcal{D}} (\mathcal{R} \mathcal{L} \bar{P} - \bar{P} \mathcal{L}^* \mathcal{R}) d\xi d\zeta + \int_{BC} \left(\frac{1}{2} \mathcal{R} \partial_\xi \bar{P} - \frac{1}{2} \bar{P} \partial_\xi \mathcal{R} + \frac{\mu}{4} \mathcal{R} \bar{P} \right) d\xi \\ &+ \left(\frac{1}{2} \bar{P} \partial_\zeta \mathcal{R} - \frac{1}{2} \mathcal{R} \partial_\zeta \bar{P} - \frac{\mu}{4} \mathcal{R} \bar{P} \right) d\zeta. \end{aligned}$$

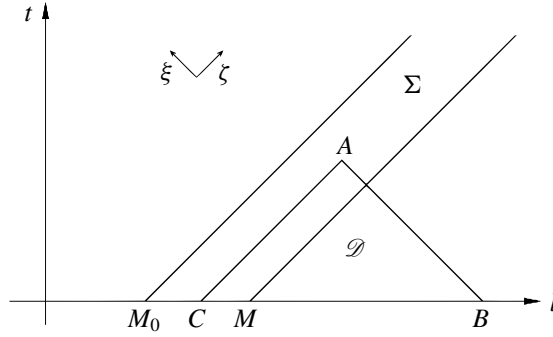


Figure 2. (t, l) -plane.

Returning to the variable (t, l) (see Figure 2), we find in the second line of (A-6) that

$$\begin{aligned}
 \text{(A-7)} \quad \int_{BC} \dots &= \int_B^C \left[\frac{1}{4} \mathcal{R}(\partial_t - \partial_l) P - \frac{1}{4} P(\partial_t - \partial_l) \mathcal{R} + \frac{\mu}{4} \mathcal{R} P \right] (-dl) \\
 &\quad + \left[\frac{1}{4} P(\partial_t + \partial_l) \mathcal{R} - \frac{1}{4} \mathcal{R}(\partial_t + \partial_l) P - \frac{\mu}{4} \mathcal{R} P \right] dl \\
 &= \int_{l_A - t_A}^{l_A + t_A} \left[\frac{\mu}{2} \mathcal{R} P + \frac{1}{2} \mathcal{R} \partial_t P - \frac{1}{2} P \partial_t \mathcal{R} \right] \Big|_{t=0} dl \\
 &= \int_{l_A - t_A}^{l_A + t_A} \beta(t_A)^{-\frac{1}{2}} \left[\Psi(a, b, 1; z|_{t=0}) \left(\frac{\mu}{4} q_0(l) + \frac{1}{2} q_1(l) \right) \right. \\
 &\quad \left. - \frac{\mu\lambda}{4} \Psi(a+1, b+1, 2; z|_{t=0}) q_0(l) z_t|_{t=0} \right] dl,
 \end{aligned}$$

where we have used the formula $\Psi'(a, b, c; z) = \frac{ab}{c} \Psi(a+1, b+1, c+1; z)$ (see [Erdélyi et al. 1953, page 58]). From the definition (A-3), we arrive at

$$z = -\frac{(t_A - l_A - t + l)(t_A + l_A - t - l)}{4(1 + t_A)(1 + t)}$$

and

$$\text{(A-8)} \quad z_t|_{t=0} = \frac{t_A}{2(1 + t_A)} - z|_{t=0}.$$

If $(t, l) \in \Sigma \cap \overline{\mathcal{D}}$, we infer

$$\text{(A-9)} \quad 0 \geq z \geq -\frac{1}{2}(M - M_0) \geq -\frac{1}{2}\delta_0,$$

which implies that (1-8) holds. This, together with (A-7)–(A-9) and the assumption (1-11) of $\Lambda \geq \frac{3}{2}\mu\lambda$, yields that the integral in the second line of (A-6) is nonnegative.

Next we prove that $P(t, l) \geq 0$ for all $(t, l) \in \Sigma$. Define

$$\bar{t} \equiv \inf\{t : \exists l \in (t + M_0, t + M) \text{ such that } P(t, l) < 0\}.$$

From assumption (1-9), we get $\bar{t} > 0$. If $\bar{t} < +\infty$, we see that there exists $\bar{l} \in (\bar{t} + M_0, \bar{t} + M)$ such that $P(\bar{t}, \bar{l}) = 0$. Moreover, we have $P(t, l) \geq 0$ for $t < \bar{t}$. Choose $A = (t_A, l_A) = (\bar{t}, \bar{l})$ in (A-6). From (A-4)–(A-5) and (1-8) we infer $\mathcal{L}^*\mathcal{R} \leq 0$ for $\lambda > 1$ and $(t, l) \in \Sigma \cap \mathcal{D}$. It follows from $f(t, l) \geq 0$ in (3-7), (1-8), (1-9), and (A-6) that

$$P(\bar{t}, \bar{l}) \geq \frac{1}{2}\mathcal{R}(C; A)P(0, \bar{l} - \bar{t}) + \iint_{\Sigma \cap \mathcal{D}} (\mathcal{R}\mathcal{L}\bar{P} - \bar{P}\mathcal{L}^*\mathcal{R}) d\xi d\zeta \geq \frac{1}{4}q_0(\bar{l} - \bar{t}) > 0,$$

which is a contradiction with $P(\bar{t}, \bar{l}) = 0$. Consequently, we conclude that $\bar{t} = +\infty$ and $P(t, l) \geq 0$ for all $(t, l) \in \Sigma$.

Acknowledgment

Yin Huicheng wishes to express his gratitude to Professor Michael Reissig, Technical University Bergakademie Freiberg, Germany, for his interest in this problem and some fruitful discussions in the past.

References

- [Alinhac 1993] S. Alinhac, “Temps de vie des solutions régulières des équations d’Euler compressibles axisymétriques en dimension deux”, *Invent. Math.* **111**:3 (1993), 627–670. MR Zbl
- [Alinhac 1999a] S. Alinhac, “Blowup of small data solutions for a quasilinear wave equation in two space dimensions”, *Ann. of Math.* (2) **149**:1 (1999), 97–127. MR Zbl
- [Alinhac 1999b] S. Alinhac, “Blowup of small data solutions for a class of quasilinear wave equations in two space dimensions, II”, *Acta Math.* **182**:1 (1999), 1–23. MR Zbl
- [Christodoulou 2007] D. Christodoulou, *The formation of shocks in 3-dimensional fluids*, European Mathematical Society, Zürich, 2007. MR Zbl
- [Christodoulou and Lisibach 2016] D. Christodoulou and A. Lisibach, “Shock development in spherical symmetry”, *Ann. PDE* **2**:1 (2016), art. id. 3. MR
- [Christodoulou and Miao 2014] D. Christodoulou and S. Miao, *Compressible flow and Euler’s equations*, Surveys of Modern Mathematics **9**, International Press, Somerville, MA, 2014. MR Zbl
- [Courant and Hilbert 1962] R. Courant and D. Hilbert, *Methods of mathematical physics, II: Partial differential equations*, Interscience, New York, 1962. MR Zbl
- [D’Abbico and Reissig 2014] M. D’Abbico and M. Reissig, “Semilinear structural damped waves”, *Math. Methods Appl. Sci.* **37**:11 (2014), 1570–1592. MR Zbl
- [D’Abbico et al. 2015] M. D’Abbico, S. Lucente, and M. Reissig, “A shift in the Strauss exponent for semilinear wave equations with a not effective damping”, *J. Differential Equations* **259**:10 (2015), 5040–5073. MR Zbl
- [Dafermos 1995] C. M. Dafermos, “A system of hyperbolic conservation laws with frictional damping”, *Z. Angew. Math. Phys.* **46**:special issue (1995), 294–307. MR Zbl

- [Ding et al. 2016] B. Ding, I. Witt, and H. Yin, “The small data solutions of general 3-D quasilinear wave equations, II”, *J. Differential Equations* **261**:2 (2016), 1429–1471. MR Zbl
- [Erdélyi et al. 1953] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions, I*, McGraw-Hill, New York, 1953. MR Zbl
- [Hörmander 1997] L. Hörmander, *Lectures on nonlinear hyperbolic differential equations*, Math. Appl. **26**, Springer, Berlin, 1997. MR Zbl
- [Hsiao and Liu 1992] L. Hsiao and T.-P. Liu, “Convergence to nonlinear diffusion waves for solutions of a system of hyperbolic conservation laws with damping”, *Comm. Math. Phys.* **143**:3 (1992), 599–605. MR Zbl
- [Hsiao and Serre 1996] L. Hsiao and D. Serre, “Global existence of solutions for the system of compressible adiabatic flow through porous media”, *SIAM J. Math. Anal.* **27**:1 (1996), 70–77. MR Zbl
- [Kawashima and Yong 2004] S. Kawashima and W.-A. Yong, “Dissipative structure and entropy for hyperbolic systems of balance laws”, *Arch. Ration. Mech. Anal.* **174**:3 (2004), 345–364. MR Zbl
- [Klainerman 1987] S. Klainerman, “Remarks on the global Sobolev inequalities in the Minkowski space \mathbb{R}^{n+1} ”, *Comm. Pure Appl. Math.* **40**:1 (1987), 111–117. MR Zbl
- [Klainerman and Sideris 1996] S. Klainerman and T. C. Sideris, “On almost global existence for nonrelativistic wave equations in 3D”, *Comm. Pure Appl. Math.* **49**:3 (1996), 307–321. MR Zbl
- [do Nascimento and Wirth 2015] W. N. do Nascimento and J. Wirth, “Wave equations with mass and dissipation”, *Adv. Differential Equations* **20**:7-8 (2015), 661–696. MR Zbl
- [Nishihara 1997] K. Nishihara, “Asymptotic behavior of solutions of quasilinear hyperbolic equations with linear damping”, *J. Differential Equations* **137**:2 (1997), 384–395. MR Zbl
- [Pan and Zhao 2009] R. Pan and K. Zhao, “The 3D compressible Euler equations with damping in a bounded domain”, *J. Differential Equations* **246**:2 (2009), 581–596. MR Zbl
- [Sideris 1985] T. C. Sideris, “Formation of singularities in three-dimensional compressible fluids”, *Comm. Math. Phys.* **101**:4 (1985), 475–485. MR Zbl
- [Sideris 1997] T. C. Sideris, “Delayed singularity formation in 2D compressible flow”, *Amer. J. Math.* **119**:2 (1997), 371–422. MR Zbl
- [Sideris et al. 2003] T. C. Sideris, B. Thomases, and D. Wang, “Long time behavior of solutions to the 3D compressible Euler equations with damping”, *Comm. Partial Differential Equations* **28**:3-4 (2003), 795–816. MR Zbl
- [Speck 2016] J. Speck, *Shock formation in small-data solutions to 3D quasilinear wave equations*, Mathematical Surveys and Monographs **214**, American Mathematical Society, Providence, RI, 2016. MR Zbl
- [Tan and Guochun 2012] Z. Tan and W. Guochun, “Large time behavior of solutions for compressible Euler equations with damping in \mathbb{R}^3 ”, *J. Differential Equations* **252**:2 (2012), 1546–1561. MR Zbl
- [Wang and Yang 2001] W. Wang and T. Yang, “The pointwise estimates of solutions for Euler equations with damping in multi-dimensions”, *J. Differential Equations* **173**:2 (2001), 410–450. MR Zbl
- [Wirth 2006] J. Wirth, “Wave equations with time-dependent dissipation, I: Non-effective dissipation”, *J. Differential Equations* **222**:2 (2006), 487–514. MR Zbl
- [Wirth 2007] J. Wirth, “Wave equations with time-dependent dissipation, II: Effective dissipation”, *J. Differential Equations* **232**:1 (2007), 74–103. MR Zbl
- [Yin 2004] H. Yin, “Formation and construction of a shock wave for 3-D compressible Euler equations with the spherical initial data”, *Nagoya Math. J.* **175** (2004), 125–164. MR Zbl

[Yin and Qiu 1999] H. Yin and Q. Qiu, “The blowup of solutions for 3-D axisymmetric compressible Euler equations”, *Nagoya Math. J.* **154** (1999), 157–169. MR Zbl

Received December 31, 2016.

FEI HOU
DEPARTMENT OF MATHEMATICS AND IMS
NANJING UNIVERSITY
NANJING
CHINA
houfeimath@gmail.com

INGO WITT
MATHEMATICAL INSTITUTE
UNIVERSITY OF GÖTTINGEN
GÖTTINGEN
GERMANY
iwitt@uni-math.gwdg.de

HUICHENG YIN
SCHOOL OF MATHEMATICAL SCIENCES
JIANGSU PROVINCIAL KEY LABORATORY FOR NUMERICAL SIMULATION OF LARGE SCALE
COMPLEX SYSTEMS
NANJING NORMAL UNIVERSITY
NANJING
CHINA
huicheng@nju.edu.cn
05407@njnu.edu.cn

FORMAL CONFLUENCE OF QUANTUM DIFFERENTIAL OPERATORS

BERNARD LE STUM AND ADOLFO QUIRÓS

We prove that a differential operator in the usual sense is formally the limit of quantum differential operators. For this purpose, we introduce the notion of a twisted differential operator of infinite level and prove that, formally, such an object is independent of the choice of the twist. Our method provides explicit formulas.

Introduction	427
1. Twisted principal parts	431
2. Twisted differential forms	439
3. Twisted binomial coefficient theorem for principal parts	445
4. Twisted differential operators of infinite level	447
5. Twisted Taylor series	453
6. Quantum differential operators	457
7. Formal deformations of twisted differential operators	461
8. Formal confluence	468
9. Formal confluence in positive quantum characteristic	472
References	477

Introduction

Classically, the finite difference operator and the q -difference operator are given by the formulas

$$\Delta_h(f)(x) = \frac{f(x+h) - f(x)}{h} \quad \text{and} \quad \delta_q(f)(x) = \frac{f(qx) - f(x)}{qx - x},$$

and we can obtain the differentiation operator by passing to the limit

$$\partial(f) = \lim_{h \rightarrow 0} \Delta_h(f) = \lim_{q \rightarrow 1} \delta_q(f).$$

By using any of these operators, we may consider the notions of finite difference systems, q -difference systems and differential systems. The confluence process

Quirós was supported by grant MTM2015-68524-P (MINECO/FEDER, Spain/EU).

MSC2010: 12H10.

Keywords: q -difference, differential operator, confluence.

consists in using the solutions of a difference (finite difference or q -difference) system in order to approximate the solutions of a differential system.

In the late 19th century, C. Guichard [1887] solved the problem of integration along a finite difference operator. In the early 20th century, in a couple of articles in the Transactions of the AMS, Georges D. Birkhoff [1911] and R. D. Carmichael [1911] gave reasonable conditions for solving linear finite difference systems. Shortly after, Carmichael [1912] applied the same methods to linear q -difference systems. This was later improved by C. Raymond Adams [1928/29]. There exists an abundant literature on these questions and we would like to jump to the present and mention the recent paper by Anne Duval and Julien Roques [2008] where they study Fuchsian difference systems and their confluence. More generally, the school of Jean-Pierre Ramis produced a lot of interesting material during the last two decades (see [Di Vizio et al. 2003] for a survey). Among them, we mention in particular the work of Jacques Sauloy on resolution, classification and confluence of Fuchsian differential equations (see [Sauloy 2000] for example) and on the Galois theory of these equations [Sauloy 2003].

While these works look at equations over the complex numbers, difference systems have also been studied over p -adic fields by Lucia di Vizio [2004], for example, and Andrea Pulita [2008; 2017]. Pulita showed that, over a p -adic field, confluence is a very rigid process in the sense that we can approximate a linear differential system by a q -difference system that has exactly the same (formal) solutions. Finally, we want to mention the work of Charlotte Hardouin [2010] who studies q -difference Galois theory over a field of positive characteristic in the “exotic” case in which q is a root of unity (which is also taken into account by Pulita).

Our approach here is very general in the sense that we do not make any assumption about the base field or the value of the parameter q for example. Actually, differential systems and difference systems may be seen as particular instances of what we called *twisted* differential systems in [Le Stum and Quirós 2015b] (see also [André 2001] for a similar approach). Let us be more precise. If R is a commutative ring, then a twisted R -algebra (or a difference R -algebra, if we name them in the spirit of [Cohn 1965]) is a commutative R -algebra A endowed with an R -linear ring endomorphism σ . Actually, in [Le Stum and Quirós 2015b] we allowed more generally families of endomorphisms satisfying some relations but we will concentrate here on the one-dimensional case and we will use either one endomorphism or a system of roots of this endomorphism. One may then consider the σ -derivations D of A , or more generally, the σ -derivations D of an A -module M : R -linear maps that satisfy the twisted Leibnitz rule

$$D(xs) = D(x)s + \sigma(x)D(s).$$

A σ -differential module is then defined as an A -module M endowed with an action by σ -derivations of all the σ -derivations of A . In the classical situation above, we had $\sigma(x) = x$, $\sigma(x) = qx$ or $\sigma(x) = x + h$, giving rise to an action of ∂ , δ_q or Δ_h respectively. In our more general setting, the confluence question amounts to comparing twisted differential systems for different twists.

Our method is original in the sense that it relies on a generalization of the classical notion of a ring of differential operators. In order to do that, we may first fix what we call a σ -coordinate $x \in A$, so that there exists a basis $\partial_{\sigma,A}$ for the σ -derivations of A such that $\partial_{\sigma,A}(x) = 1$. We may then define the twisted Weyl algebra D_σ as the Ore extension of A by σ and $\partial_{\sigma,A}$ (as in [Bourbaki 2012], Proposition 1.4). This is the free A -module on the generators ∂_σ^k for $k \in \mathbb{N}$ that satisfies

$$\partial_\sigma \circ f = \partial_{\sigma,A}(f) + \sigma(f)\partial_\sigma$$

when $f \in A$. A σ -differential module is then the same thing as a D_σ -module. So now, what we want to understand is the relation between D_σ and D_τ when τ is another R -linear endomorphism of A with the same twisted coordinate x . In the end, we will be essentially interested in $\tau(x) = x$ (the usual case where we simply write D instead of D_τ) and $\sigma(x) = qx + h$ with $q, h \in R$ (the quantum case). Note that the condition of q being a root of unity in the quantum case below is analogous to having a positive characteristic in the usual case and will require some care.

One may usually see a ring of differential operators as some dual of a ring of formal functions (see the first section of [Grothendieck 1967] for example). Doing this directly for D_σ would require understanding the notion of σ -divided powers on the function side. This is a very interesting question that we postpone to a forthcoming article. Here, we will actually replace the twisted Weyl algebra D_σ with a Grothendieck ring of differential operators $D_\sigma^{(\infty)}$ (so that the σ -divided powers live naturally on the differential operator side). It so happens that the classical construction works incredibly well for this particular generalization. It is actually sufficient to replace the usual powers of an ideal with the twisted powers introduced in [Le Stum and Quirós 2015a]. But now we are faced with two questions: the comparison between $D_\sigma^{(\infty)}$ and $D_\tau^{(\infty)}$ on one hand and the comparison of D_σ with $D_\sigma^{(\infty)}$ on the other.

We will show that, if we denote by $\widehat{D}_\sigma^{(\infty)}$ the completion along the divided powers of ∂_σ , then there always exists a natural isomorphism $\widehat{D}_\sigma^{(\infty)} \simeq \widehat{D}_\tau^{(\infty)}$ (the formal deformation of Proposition 7.4) and we will be able to give very explicit formulas. For example, in the case $\tau(x) = x$ and $\sigma(x) = qx$, we can write (over a field of characteristic zero when q is not a root of unity):

$$\partial_\sigma = \sum_{k=1}^{\infty} \frac{1}{k!} (q-1)^{k-1} x^{k-1} \partial^k \quad \text{and} \quad \partial = \sum_{k=1}^{\infty} \frac{(1-q)^k}{1-q^k} x^{k-1} \partial_\sigma^k.$$

In the quantum situation, there exists a canonical map $D_\sigma \rightarrow D_\sigma^{(\infty)}$ whose image is exactly the ring \bar{D}_σ of small (or naive) differential operators. Under some conditions on q (R q -divisible and q -char(R) = 0, see [Le Stum and Quirós 2015b]), then all three rings D_σ , \bar{D}_σ and $D_\sigma^{(\infty)}$ are actually equal. This happens for example in the usual case when R is a \mathbb{Q} -algebra and in the quantum case when q belongs to a subfield K of R and q is not a root of unity. When all these conditions are satisfied, we easily derive from the formal deformation our first confluence Theorem 8.3: there exists a map $D_\sigma \rightarrow \hat{D}$ with dense image.

We will also consider the case where q is a primitive p -th root of unity (but still R a \mathbb{Q} -algebra). It is then necessary to use a complete family \underline{q} of p^n -th roots of q . We can define what we call a rooted Weyl algebra D_σ by taking the limit on all D_{σ_n} and build a map $D_\sigma \rightarrow \hat{D}$ with dense image. This is the second confluence Theorem 9.13.

It is interesting to notice that this last theorem puts together the ring D that supports the conjecture of Jacques Dixmier [1968] (whose higher version is equivalent to the Jacobian conjecture [Belov-Kanel and Kontsevich 2007]) and an Azumaya algebra where the Morita equivalence could be applied. Recall that the Jacobian conjecture (or global inversion theorem) states that an endomorphism of the complex affine space whose Jacobian is invertible is necessarily an isomorphism. Recall also that Dixmier's conjecture states that any endomorphism of a complex Weyl algebra such as D is always an automorphism. This conjecture is not valid for D_σ . However, this last ring is a lot easier to study because it is essentially a matrix algebra.

In a forthcoming paper, we will prove an ultrametric version of the results presented here. More precisely, we will introduce the notion of twisted differential operator of given radius on an affinoid algebra and show that this notion is essentially independent of the choice of the twist. This will lead to an explicit equivalence between differential systems and difference systems, generalizing a theorem of Pulita [2008]. Recall that the first results in this direction were already obtained by Yves André and Lucia Di Vizio [2004].

In a forthcoming joint paper with Michel Gros, we will introduce the notion of quantum divided power and apply our methods in order to obtain a ring of quantum differential operators of level zero. It happens to be isomorphic to the quantum Weyl algebra. However, this new approach also provides the notion of quantum p -curvature. We will then use some Frobenius action and obtain an Azumaya splitting of the quantum Weyl algebra as well as a quantum Simpson's correspondence much as in [Gros and Le Stum 2014].

Both authors thank Michel Gros for all the fruitful conversations that we had all together.

For us, a *ring* has an associative multiplication (not commutative in general) and a two-sided unit. Morphisms of rings are always assumed to preserve the unit. A module always means a *left* module. We will essentially consider 1-twisted

rings from [Le Stum and Quirós 2015b] and simply call them twisted rings. More precisely, throughout the paper, R will denote a *commutative ring* and A a *twisted commutative R -algebra*: a commutative R -algebra endowed with an R -linear ring endomorphism σ_A .

1. Twisted principal parts

We will introduce here the notion of twisted principal part (functions on twisted infinitesimal neighborhoods of the diagonal).

We will begin by ignoring the endomorphism σ_A and concentrating on the commutative R -algebra A . The tensor product $P_{A/R} := A \otimes_R A$ has two A -algebra structures, one coming from the action on the left and the other one coming from the action on the right. Unless otherwise specified, we will use the *left* structure when we see $P_{A/R}$ as an A -module. However, when M is an A -module, the notation $P \otimes_A M$ will implicitly mean that we use the action of A on the *right* to build the tensor product and that the resulting object will be seen as an A -module using the action of A on the *left*.

In practice, we will write $x := x \otimes 1 \in P_{A/R}$ and $\tilde{x} := 1 \otimes x \in P_{A/R}$. In other words, with these notations, the action on the left is multiplication by x and the action on the right is multiplication by \tilde{x} . Any element of $P_{A/R}$ can be written as a finite sum $\sum x_i \tilde{y}_i$. At some point, we will call the embedding on the right the *Taylor map* and denote it by

$$\theta_{A/R} : A \rightarrow P_{A/R}, \quad x \mapsto \tilde{x}.$$

We will then call $\tilde{x} = \theta_{A/R}(x)$ the *Taylor expansion* of $x \in A$ (more on this vocabulary later on).

We will also consider the canonical map corresponding to the projection that forgets the middle term:

$$A \otimes_R A \xrightarrow{\delta_{A/R}} A \otimes_R A \otimes_R A, \quad x \otimes y \longmapsto x \otimes 1 \otimes y.$$

It is a morphism of R -algebras that may also be seen as a map

$$P_{A/R} \xrightarrow{\delta_{A/R}} P_{A/R} \otimes_A P_{A/R} \quad x \tilde{y} \longmapsto x \otimes \tilde{y},$$

where A acts on the *right* on the first factor and on the *left* on the second one in the tensor product.

Let $I_{A/R}$ be the kernel of the multiplication morphism

$$P_{A/R} \rightarrow A, \quad x \tilde{y} \mapsto xy$$

(that corresponds to the diagonal embedding). The ideal $I_{A/R}$ is generated by the image of the linear map

$$A \xrightarrow{d_{A/R}} P_{A/R}, \quad x \longmapsto \xi := \tilde{x} - x$$

(that corresponds to the difference between the projections). In practice, we will usually drop the index A/R and simply write P , I , θ , δ and d .

Since we will make regular use of linearization, we introduce it formally now:

Definition 1.1. Let M, N be two A -modules and $u : M \rightarrow N$ an R -linear map. Then the A -linearization of u is the A -linear map

$$\begin{array}{ccccc} & & \tilde{u} & & \\ & \nearrow \simeq & & \searrow & \\ P \otimes_A M & \longrightarrow & A \otimes_R M & \longrightarrow & N \\ & & x \otimes s \longmapsto & & xu(s). \end{array}$$

Recall that, in the tensor product, we use the action on the *right* on P . Therefore, we have for all $x, y \in A$, for all $s \in M$,

$$\tilde{u}(x\tilde{y} \otimes s) = xu(ys).$$

As an example, we see that the multiplication map $P \rightarrow A$ is the A -linearization of the identity of A , seen as an R -linear map.

For further use, we also mention the following result that follows immediately from the definitions:

Lemma 1.2. *If we are given two R -linear maps $\varphi : M \rightarrow N$ and $\psi : L \rightarrow M$, then the linearization of $\varphi \circ \psi$ factors as*

$$P \otimes_A L \xrightarrow{\delta \otimes \text{Id}} P \otimes_A P \otimes_A L \xrightarrow{\text{Id} \otimes_A \tilde{\psi}} P \otimes_A M \xrightarrow{\tilde{\varphi}} N.$$

Now we make the endomorphism σ_A enter the game. We will consider $P := A \otimes_R A$ as a twisted R -algebra by using σ_A on the left and the identity the right: in other words, we set $\sigma_P := \sigma_A \otimes_R \text{Id}_A$. Alternatively, this is the unique structure of twisted R -algebra on P such that for all $x \in A$,

$$\sigma_P(x) = \sigma_A(x) \quad \text{and} \quad \sigma_P(\tilde{x}) = \tilde{x}.$$

In particular, we may also consider P as a twisted A -algebra, in the sense that P is endowed with a σ_A -linear ring endomorphism σ_P . We will often drop the indexes A and P and simply write σ for both maps (so that $\sigma(\tilde{x}) = \tilde{x}$) when there is no ambiguity.

Before we do anything else, let us prove the following result, which is quite elementary, but very useful:

Lemma 1.3. *If $x \in A$ and $\xi := \tilde{x} - x \in P_{A/R}$, then*

$$\sigma_P(\xi) = \xi + y \quad \text{with} \quad y := x - \sigma_A(x).$$

Proof. We have

$$\sigma(\xi) = \sigma(\tilde{x} - x) = \sigma(\tilde{x}) - \sigma(x) = \tilde{x} - \sigma(x) = \tilde{x} - x + x - \sigma(x) = \xi + y. \quad \square$$

Lemma 1.4. *The kernel of the A -linearization*

$$P_{A/R} \xrightarrow{\tilde{\sigma}_A} A, \quad x\tilde{y} \longmapsto x\sigma_A(y)$$

of σ_A is $\sigma_P(I_{A/R})$.

Proof. Recall that $\sigma(I)$ denotes the *ideal* generated by the image of I . As a consequence, since I is generated by the image of d , we see that $\ker \tilde{\sigma}$ will be generated by the image of

$$A \xrightarrow{\sigma \circ d} P, \quad x \longmapsto \tilde{x} - \sigma(x).$$

Now, we have for all $x \in A$, $\tilde{\sigma}(\tilde{x} - \sigma(x)) = \sigma(x) - \sigma(x) = 0$, and it follows that $\sigma(I) \subset \ker \tilde{\sigma}$. Conversely, by definition, we have for all $x \in A$,

$$\sigma(x) \equiv \tilde{x} \pmod{\sigma(I)}.$$

Therefore, if $f := \sum x_i \tilde{y}_i \in \ker \tilde{\sigma}$, we have

$$f \equiv \sum x_i \sigma(y_i) = \tilde{\sigma}(f) = 0 \pmod{\sigma(I)},$$

and we see that $\ker \tilde{\sigma} \subset \sigma(I)$. \square

Remarks. (1) It is sometimes convenient to use the bimodule language. An A -sesquimodule M is an A -bimodule such that for all $x \in A$, $s \in M$,

$$\sigma_A(x) \cdot s = s \cdot x.$$

Note that we are using the *reverse* convention from André [2001] so that forgetting the right action induces an equivalence (an isomorphism) between A -sesquimodules and left A -modules (we will use this identification).

- (2) The R -algebra P has a canonical A -bimodule structure which is completely independent of the choice of σ_A . If we endow A with its sesquimodule structure, then the linearization $\tilde{\sigma}$ of σ_A is a morphism of A -bimodules: we always have for all $x, y \in A$, $f \in P$,

$$\tilde{\sigma}(x \cdot f \cdot y) = \tilde{\sigma}(xf\tilde{y}) = x\tilde{\sigma}(f)\sigma(y) = x \cdot \tilde{\sigma}(f) \cdot y.$$

It follows that $\ker \tilde{\sigma}$ has a natural A -bimodule structure. Actually, since $\tilde{\sigma}$ is a ring homomorphism, then $\ker \tilde{\sigma}$ is an ideal and therefore automatically an A -bimodule.

Recall from [Le Stum and Quirós 2015a] that the *twisted powers* of I are

$$I^{(0)} = P, \quad I^{(1)} = I, \quad I^{(2)} := I\sigma(I), \quad \dots, \quad I^{(n)} := I\sigma(I) \cdots \sigma^{n-1}(I),$$

where images and products of ideals are meant as *ideals*. We will write $I_{A/R}^{(n)\sigma}$ when we want to make clear the dependence on σ and A/R .

Definition 1.5. The A -module of *twisted principal parts of order $n \in \mathbb{N}$ (and infinite level)* of A is

$$P_{A/R, (n)\sigma} := P_{A/R} / I_{A/R}^{(n+1)\sigma}.$$

The A -module of *twisted principal parts of infinite order (and infinite level)* of A is the *twisted completion*:

$$\widehat{P}_{A/R, \sigma} := \varprojlim P_{A/R, (n)\sigma}.$$

Note that these A -modules all have natural R -algebra structures and that the definition also makes sense for $n = -1$ so that $P_{A/R, (-1)\sigma} = 0$. Again, we will often drop the indexes A/R when we believe that there is no risk of confusion and simply write $P_{(n)\sigma}$ and \widehat{P}_σ . We will also drop the index σ when $\sigma_A = \text{Id}_A$.

Examples. (1) When A is trivially twisted, which means that $\sigma_A := \text{Id}_A$, this notion of principal parts coincides with the usual one (definition 16.3.1 of [Grothendieck 1967]), and therefore many of the basic objects we will construct are twisted versions of those in that work.

(2) When $A = R[x]$, we have $P = R[x, \xi]$ with $\xi = \tilde{x} - x$ and $I = (\xi)$. Moreover, $\sigma(\xi) = \xi + y$ with $y = x - \sigma(x)$. It follows that

$$\sigma^n(\xi) = \xi + (n)_\sigma(y) = \xi + x - \sigma^n(x),$$

with $(n)_\sigma := 1 + \cdots + \sigma^{n-1}$. Therefore, we have

$$P_{(n)} = R[x, \xi] / \prod_{i=0}^n (\xi + (i)_\sigma(y)) = R[x, \xi] / \prod_{i=0}^n (\xi + x - \sigma^i(x)).$$

(a) In the case $\sigma(x) = x$, we get $P_{(n)} = R[x, \xi] / \xi^{n+1}$ as expected.

(b) More generally, if we assume that $\sigma(x) = x + h$ with $h \in R$, we obtain

$$P_{(n)} = R[x, \xi] / \prod_{i=0}^n (\xi - ih).$$

(c) On the other hand, if we let $\sigma(x) = qx$ with $q \in R$, we find

$$P_{(n)} = R[x, \xi] / \prod_{i=0}^n (\xi + (1 - q^i)x).$$

When $R \rightarrow R'$ is a homomorphism of commutative rings, we endow $A' := R' \otimes_R A$ with $\sigma_{A'} := \text{Id}_{R'} \otimes_R \sigma_A$.

Proposition 1.6. *Let $R \rightarrow R'$ be a homomorphism of commutative rings and $A' := R' \otimes_R A$. Then we have for all $n \in \mathbb{N}$,*

$$A' \otimes_A P_{A/R, (n)_\sigma} \simeq P_{A'/R', (n)_\sigma}.$$

Proof. If we let $P' := P_{A'/R'}$, then there exists a canonical isomorphism $A' \otimes_A P \simeq R' \otimes_R P \simeq P'$. Moreover, if we denote by I' the kernel of the multiplication morphism on P' , we have $A' \otimes_A I \simeq I'$. And finally, $\text{Id}_A \otimes_A \sigma_P$ corresponds to $\sigma_{P'}$ under this isomorphism. Our assertion is therefore an immediate consequence of right exactness of tensor product. \square

Recall from [Le Stum and Quirós 2015a] that a twisted A -algebra is an A -algebra B endowed with a σ_A -linear ring endomorphism σ_B (which simply means that for all $f \in A$, $\sigma_B(f 1_B) = \sigma_A(f)$).

Proposition 1.7. *If B is a twisted commutative A -algebra, then there exists a canonical morphism of B -algebras*

$$B \otimes_A P_{A/R, (n)_\sigma} \rightarrow P_{B/R, (n)_\sigma}.$$

When B is a quotient (resp. a localization) of A , this map is surjective (resp. bijective).

Recall from Definition 1.7 of [Le Stum and Quirós 2015b] that we call such a B a *twisted quotient* (resp. *twisted localization*) of A .

Proof. The morphism of twisted R -algebras $A \rightarrow B$ extends naturally to a morphism of twisted R -algebras $P_A \rightarrow P_B$. Since I_A is sent into I_B , we see that, for all $n \in \mathbb{N}$, $\sigma^n(I_A)$ is sent into $\sigma^n(I_B)$ and the first assertion formally follows. In the case of a quotient map, all the maps involved are surjective.

Now, if $B := S^{-1}A$ is a twisted localization of A , then P_B is the localization of P_A with respect to the monoid S' generated by S and \tilde{S} , and we have $I_B = P_B \otimes_{P_A} I_A$. It immediately follows that for all $n \in \mathbb{N}$, we have $\sigma^n(I_B) = P_B \otimes_{P_A} \sigma^n(I_A)$, and therefore $I_B^{(n)} = P_B \otimes_{P_A} I_A^{(n)}$. Thus we see that

$$P_{B, (n)_\sigma} = P_B / (P_B \otimes_{P_A} I_A^{(n+1)}) = P_B \otimes_{P_A} P_{A, (n)} = B \otimes_A P_{A, (n)} \otimes_A B.$$

We need to remove the B on the right-hand side and it is sufficient to show that \tilde{x} is invertible in $B \otimes_A P_{A, (n)}$ whenever $x \in S$. But we have

$$\prod_{i=0}^n (\tilde{x} - \sigma^i(x)) = \prod_{i=0}^n \sigma^i(\tilde{x} - x) \in I_A^{(n+1)},$$

from which we derive that there exists $f \in P_A$ such that

$$f\tilde{x} \equiv \prod_{i=0}^n \sigma^i(x) \pmod{I_A^{(n+1)}}.$$

Since $B = S^{-1}A$, we must have $\sigma(S) \subset B^\times$ and it follows that $\prod_{i=0}^n \sigma^i(x) \in B^\times$. Thus, we see $f\tilde{x}$ is invertible in $B \otimes_A P_{A,(n)}$ and it follows \tilde{x} is invertible too. \square

As an illustration, we can give explicit formulas in the quantum situation. Recall that we introduced in [Le Stum and Quirós 2015a] the notion of *twisted powers* of an element in a twisted ring. In particular, for $f \in P$, we will have

$$f^{(0)} := 1, \quad f^{(1)} := f, \quad f^{(2)} = f\sigma(f), \quad \dots, \quad f^{(n+1)} = f\sigma(f) \cdots \sigma^n(f).$$

Recall also that the *quantum binomial coefficients* are defined by induction (see [Le Stum and Quirós 2015a] for example) as

$$\binom{n}{k}_q := \binom{n-1}{k-1}_q + q^k \binom{n-1}{k}_q.$$

Proposition 1.8. *Assume $\sigma(x) = qx$ with $q \in R$ and let $\xi = \tilde{x} - x$. Then we have for all $n \in \mathbb{N}$,*

$$(1) \quad \xi^{(n)} = \sum_{j=0}^n (-1)^j \binom{n}{j}_q q^{\frac{j(j-1)}{2}} x^j \tilde{x}^{n-j}$$

and for all $n \in \mathbb{N}$,

$$\tilde{x}^n = \sum_{i=0}^n \binom{n}{i}_q x^i \xi^{(n-i)}.$$

Proof. The first equality is essentially the quantum binomial formula (see proposition 2.14 in [Le Stum and Quirós 2015a]):

$$(\tilde{x} - x)^{(n)} = \sum_{j=0}^n \binom{n}{j}_q (-x)^{(j)} \tilde{x}^{(n-j)}.$$

For the second one, we compute the right-hand side with the help of formula (1):

$$\begin{aligned} \sum_{i=0}^n \binom{n}{i}_q x^i \xi^{(n-i)} &= \sum_{i=0}^n \binom{n}{i}_q x^i \left(\sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j}_q q^{\frac{j(j-1)}{2}} x^j \tilde{x}^{n-i-j} \right) \\ &= \sum_{k=0}^n \left(\sum_{i=0}^k \binom{n}{i}_q \binom{n-i}{k-i}_q (-1)^{k-i} q^{\frac{(k-i)(k-i-1)}{2}} \right) x^k \tilde{x}^{n-k} \end{aligned}$$

after rewriting $i + j = k$. Now, we have (using corollaries 2.7 and 2.8 in [Le Stum and Quirós 2015a], for example)

$$\sum_{i=0}^k \binom{n}{i}_q \binom{n-i}{k-i}_q (-1)^{k-i} q^{\frac{(k-i)(k-i-1)}{2}} = \binom{n}{k}_q \sum_{i=0}^k \binom{k}{i}_q (-1)^{k-i} q^{\frac{(k-i)(k-i-1)}{2}}.$$

But the quantum binomial formula again implies that for $k > 0$, we have

$$\sum_{i=0}^k \binom{k}{i}_q (-1)^{k-i} q^{\frac{(k-i)(k-i-1)}{2}} = \prod_{i=0}^{k-1} (1 - q^i) = 0.$$

And it follows that

$$\sum_{i=0}^n \binom{n}{i}_q x^i \xi^{(n-i)} = \tilde{x}^n,$$

as asserted. \square

We will need a slightly stronger notion of coordinate than the one used in [Le Stum and Quirós 2015b]:

Definition 1.9. Let $x \in A$ and $\xi := \tilde{x} - x \in P_{A/R}$.

- (1) Then x is a *twisted coordinate* for the twisted R -algebra A or a σ -*coordinate* for the R -algebra A if for all $n \in \mathbb{N}$, $P_{A/R, (n)_\sigma}$ is freely generated as an A -module by the images of $1, \xi, \xi^2, \dots, \xi^n$.
- (2) If x is a twisted coordinate such that $\sigma(x) = qx + h$ with $q, h \in R$, then we will call it a *quantum coordinate* or q -*coordinate* and call A a *quantum R -algebra* or q - *R -algebra*.

Examples. (1) When $\sigma_A = \text{Id}_A$, a twisted coordinate will be called a *usual coordinate*. In the case A/R is smooth (of pure relative dimension one), then a usual coordinate is nothing but an *étale coordinate*: it means that the map

$$R[T] \rightarrow A, \quad T \mapsto x$$

is étale.

- (2) If $A = R[x]$, then x is always a twisted coordinate, whatever σ is.

Proposition 1.10. (1) If $R \rightarrow R'$ is a homomorphism of commutative rings and x is a twisted coordinate on A , then x becomes a twisted coordinate on $A' := R' \otimes_R A$ (relatively to R').

- (2) If B is a twisted localization of A and x is a twisted coordinate on A , then x becomes a twisted coordinate on B .

Proof. Follows from propositions 1.6 and 1.7. \square

Remark. Assume A is a quantum R -algebra so that there exists a twisted coordinate x on A and $q, h \in R$ such that $\sigma(x) = qx + h$. Let us still denote by the same letter x an indeterminate over R and by σ again the endomorphism of $R[x]$ given by the same formula. Then A becomes an $R[x]$ -twisted algebra and we have a canonical isomorphism (free A -modules of the same rank with the same generators):

$$A \otimes_{R[x]} P_{R[x]/R, (n)_\sigma} \simeq P_{A/R, (n)_\sigma}.$$

In the next statement, we use the letter ξ as an indeterminate over A so that $A[\xi]$ denotes the polynomial ring and $A[\xi]_{\leq n}$ the submodule of polynomial of degree at most n . Ultimately, this should not create any confusion due to Corollary 1.12 below.

Proposition 1.11. *Let $x \in A$ and $y := x - \sigma(x)$. We endow $A[\xi]$ with the unique σ_A -linear endomorphism such that $\sigma(\xi) = \xi + y$. Then x is a twisted coordinate on A if and only if the morphism of twisted algebras*

$$\phi : A[\xi] \rightarrow P_{A/R}, \quad \xi \mapsto \tilde{x} - x$$

induces for all $n \in \mathbb{N}$ an isomorphism of A -algebras $A[\xi]/\xi^{(n+1)} \simeq P_{A/R, (n)_\sigma}$.

Proof. First of all, it follows from Lemma 1.3 that there exists such a morphism for all $n \in \mathbb{N}$. On the other hand,

$$(2) \quad \xi^{(n+1)} = \prod_{i=0}^n (\xi + (x - \sigma^i(x)))$$

is a monic polynomial of degree $n + 1$. Then euclidean division tells us that the composite map

$$A[\xi]_{\leq n} \rightarrow A[\xi] \rightarrow A[\xi]/\xi^{(n+1)}$$

is an isomorphism of A -modules. Therefore the condition on ϕ is equivalent to the fact that the map

$$A[\xi]_{\leq n} \rightarrow P_{(n)_\sigma}, \quad \xi \mapsto \overline{\tilde{x} - x}$$

is bijective. And this exactly means that $P_{(n)_\sigma}$ is freely generated by the $n + 1$ first powers of the images of $\tilde{x} - x$. \square

When the polynomial ring $A[\xi]$ is endowed with the structure of a σ_A -algebra, we will set

$$A[\![\xi]\!]_\sigma := \varprojlim A[\xi]/\xi^{(n+1)}.$$

Corollary 1.12. *With the same hypothesis, x is a twisted coordinate on A if and only if there exists an isomorphism of A -algebras*

$$A[\![\xi]\!]_\sigma \xrightarrow{\sim} \widehat{P}_{A/R, \sigma}, \quad \xi \longmapsto \tilde{x} - x.$$

Corollary 1.13. *Let $x \in A$ and $\xi := \tilde{x} - x \in P$. Then the following conditions are equivalent:*

- (1) x is a twisted coordinate on A ;
- (2) for all $n \in \mathbb{N}$, the A -module $P_{(n)\sigma}$ is freely generated by the images of $1, \xi, \xi^{(2)}, \dots, \xi^{(n)}$;
- (3) for all $n \in \mathbb{N}$, the A -module $I^{(n)}/I^{(n+1)}$ is free of rank one on the image of $\xi^{(n)}$.

Corollary 1.14. *If A is a twisted localization of $R[x]$, then x is a twisted coordinate on A .*

Proof. Using the second part of Proposition 1.10, we may assume that $A = R[x]$, in which case this is a trivial consequence of Proposition 1.11. \square

2. Twisted differential forms

In this section, we study the module of twisted differential forms (of degree one) and make the link with twisted derivations. We use the same notations as before.

Definition 2.1. The A -module of *twisted differential forms* on A/R is

$$\Omega_{A/R, \sigma}^1 := I_{A/R}/I_{A/R}^{(2)\sigma}.$$

Again, we will often drop the index A/R . Since we implicitly endow P with the action of A on the left, we will also always see Ω_σ^1 as an A -module through the action on the left.

Examples. (1) When A is trivially twisted, then $\Omega_\sigma^1 = I/I^2$ is the usual module of differential forms of A over R .

- (2) If $A = R[x]$ is endowed with any R -algebra endomorphism σ , then Ω_σ^1 is free of rank 1: with the notations of Lemma 1.3 (so that $y = x - \sigma(x)$), we have

$$\Omega_\sigma^1 \simeq \xi R[x, \xi]/\xi(\xi + y) \simeq R[x, \xi]/(\xi + y) \simeq R[x].$$

Remarks. (1) Clearly, Ω_σ^1 has a natural A -bimodule structure as a quotient of two ideals of P . It happens that this is identical to its A -sesquimodule structure: by definition, if $x \in A$, then $\tilde{x} \equiv \sigma(x) \pmod{\sigma(I)}$ and it follows that for all $f \in I$,

$$(3) \quad f\tilde{x} \equiv f\sigma(x) \pmod{I\sigma(I)}.$$

- (2) The Ω_σ^1 that appears in proposition 1.4.2.1 of [André 2001] is exactly the same as ours (André calls k what we call R).
- (3) Formula (3) is exactly the first step of the braiding described by Max Karoubi and Mariano Suárez-Álvarez [2003].

- (4) One can define more generally the *twisted de Rham complex* $\Omega_{\sigma}^{\bullet}$ of A as the quotient of the noncommutative tensor algebra of I by the graded differential ideal generated by $I\sigma(I)$. We will not consider this complex here.

Proposition 2.2. *There exists a split exact sequence*

$$0 \rightarrow \Omega_{A/R,\sigma}^1 \rightarrow P_{A/R,(1)\sigma} \xrightarrow{\tilde{\sigma}_A} A \rightarrow 0.$$

Proof. There exists such an exact sequence by definition of $P_{(1)\sigma}$ and Ω_{σ}^1 . The A -module structure of P provides a section of $\tilde{\sigma}$. \square

Proposition 2.3. (1) *If $R \rightarrow R'$ is a homomorphism of commutative rings and $A' := R' \otimes_R A$, then there exists an isomorphism*

$$A' \otimes_A \Omega_{A/R,\sigma}^1 \simeq \Omega_{A'/R',\sigma}^1.$$

- (2) *If B is a twisted commutative A -algebra, then there exists a canonical B -linear map*

$$(4) \quad B \otimes_A \Omega_{A/R,\sigma}^1 \rightarrow \Omega_{B/R,\sigma}^1.$$

When B is a quotient (resp. a localization) of A this map is surjective (resp. bijective).

Proof. Using Proposition 2.2, this follows from propositions 1.6 and 1.7. \square

Remark. This last result does *not* hold however if we only require $A \rightarrow B$ to be an étale map (and not a localization map) as the following example shows. Let R be any field of characteristic different from 2, $A := R[x]$ with $\sigma_A := \text{Id}_A$ and $B := R[x, x^{-1}]$ with $\sigma_B(x) := -x$. Then the morphism $x \mapsto x^2$ is an étale twisted morphism but the morphism (4) is the zero map. More precisely, if $\xi = \tilde{x} - x$, we have

$$B \otimes_A \Omega_{A,\sigma}^1 = (\xi)/(\xi^2) \quad \text{and} \quad \Omega_{B,\sigma}^1 = (\xi)/(\xi^2 + 2x\xi),$$

where the ideals are taken inside $R[x, x^{-1}, \xi]$, and the morphism (4) is induced by $\xi \mapsto \xi^2 + 2x\xi$.

Recall from [Le Stum and Quirós 2015b] that a *twisted derivation* of A is an R -linear map into an A -module M that satisfies the *twisted Leibnitz rule*: for all $x, y \in A$,

$$D(xy) = xD(y) + \sigma(y)D(x).$$

They form an A -module $\text{Der}_{R,\sigma}(A, M)$.

Proposition 2.4. *The canonical map $A \rightarrow \Omega_{A/R,\sigma}^1$ induced by d is a twisted derivation. It provides us with a natural isomorphism*

$$(5) \quad \text{Hom}_A(\Omega_{A/R,\sigma}^1, M) \xrightarrow{\sim} \text{Der}_{R,\sigma}(A, M), \quad u \longmapsto D := u \circ d$$

whenever M is an A -module.

In the future, we will also denote by $d : A \rightarrow \Omega_\sigma^1$ this universal twisted derivation when there is no risk of confusion.

Proof. Using formula (3), we see that, inside P , we have

$$\begin{aligned} yd(x) + \sigma(x)d(y) &= y(\tilde{x} - x) + \sigma(x)(\tilde{y} - y) \\ &\equiv y(\tilde{x} - x) + \tilde{x}(\tilde{y} - y) \\ &= \tilde{x}\tilde{y} - xy \\ &= d(xy) \pmod{I\sigma(I)}. \end{aligned}$$

It follows that the induced map $d : A \rightarrow \Omega_\sigma^1$ is indeed a twisted derivation. This also implies that the map in (5) is well defined. And it is clearly injective because I is generated by the image of d .

We now show that it is surjective. If D is a twisted derivation of M , we can consider its linearization

$$P \xrightarrow{\tilde{D}} M, \quad x\tilde{y} \longmapsto xD(y).$$

By definition, the ideal $I\sigma(I)$ is generated by elements of the form

$$f = (\tilde{x} - x)(\tilde{y} - \sigma(y)) = \tilde{x}\tilde{y} - x\tilde{y} - \sigma(y)\tilde{x} + x\sigma(y),$$

and we have

$$\tilde{D}(f) = D(xy) - xD(y) - \sigma(y)D(x) + x\sigma(y)D(1) = 0$$

because D is a twisted derivation (and in particular $D(1) = 0$). It follows that \tilde{D} factors through $P/I\sigma(I)$ and we may consider the induced map $u : \Omega_\sigma^1 \rightarrow M$. It only remains to notice that we have for all $x \in A$,

$$u(d(x)) = \tilde{D}(d(x)) = \tilde{D}(\tilde{x} - x) = D(x) - xD(1) = D(x). \quad \square$$

Remark. There exists a very elegant proof of this last result through the theory of bimodules. It is based on the fact (see proposition 17 in [Bourbaki 1970], Chapter III, section 10) that I is universal for bimodule derivations: there exists a natural isomorphism

$$\mathrm{Hom}_{A\text{-Bim}}(I, M) \xrightarrow{\sim} \mathrm{Der}_R(A, M), \quad u \longmapsto D := u \circ d,$$

where the right-hand side stands for bimodule derivations (see proposition 1.4.2.1 of [André 2001]).

As an immediate consequence of the proposition, writing $T_{A/R, \sigma} := \mathrm{Der}_{R, \sigma}(A, A)$, which we will often abbreviate to T_σ , we obtain the following:

Corollary 2.5. *The A -module $T_{A/R, \sigma}$ is the dual of $\Omega_{A/R, \sigma}^1$.*

Proposition 2.6. *Assume that $\Omega_{A/R,\sigma}^1$ is projective of finite rank. Then if M is an A -module we have the following:*

- (1) *If $R \rightarrow R'$ is a base extension and $A' = R' \otimes_R A$, then there exists a canonical isomorphism*

$$R' \otimes_R \operatorname{Der}_{R,\sigma}(A, M) \simeq \operatorname{Der}_{R',\sigma}(A', A' \otimes_A M).$$

- (2) *If B is a twisted A -algebra, there exists a canonical map*

$$B \otimes_A \operatorname{Der}_{R,\sigma}(A, M) \leftarrow \operatorname{Der}_{R,\sigma}(B, B \otimes_A M).$$

It is injective (resp. bijective) when B is a quotient (resp. a localization) of A .

Proof. Both assertions follow from Proposition 2.3. More precisely, in the first case,

$$R' \otimes_R \operatorname{Der}_{R,\sigma}(A, M) \simeq R' \otimes_R \operatorname{Hom}_A(\Omega_{A/R,\sigma}^1, M) \simeq \operatorname{Hom}_A(\Omega_{A/R,\sigma}^1, A' \otimes_A M)$$

because $\Omega_{A/R,\sigma}^1$ is projective of finite rank, and then

$$\begin{aligned} \operatorname{Hom}_A(\Omega_{A/R,\sigma}^1, A' \otimes_A M) &\simeq \operatorname{Hom}_{A'}(A' \otimes_A \Omega_{A/R,\sigma}^1, A' \otimes_A M) \\ &\simeq \operatorname{Hom}_{A'}(\Omega_{A'/R',\sigma}^1, A' \otimes_A M) \simeq \operatorname{Der}_{R',\sigma}(A', A' \otimes_A M). \end{aligned}$$

The proof of the second assertion follows exactly the same lines with the same arguments. We have

$$\begin{aligned} B \otimes_A \operatorname{Der}_{R,\sigma}(A, M) &\simeq B \otimes_A \operatorname{Hom}_A(\Omega_{A/R,\sigma}^1, M) \\ &\simeq \operatorname{Hom}_A(\Omega_{A/R,\sigma}^1, B \otimes_A M) \\ &\simeq \operatorname{Hom}_B(B \otimes_A \Omega_{A/R,\sigma}^1, B \otimes_A M) \leftarrow \operatorname{Hom}_A(\Omega_{B/R,\sigma}^1, B \otimes_A M) \\ &\simeq \operatorname{Der}_{R,\sigma}(B', B \otimes_A M), \end{aligned}$$

and the only map which is not always an isomorphism will be injective (resp. bijective) when B is a quotient (resp. a localization) of A . \square

The following immediate consequence is worth stating:

Corollary 2.7. *If $\Omega_{A/R,\sigma}^1$ is projective of finite rank, then $T_{A/R,\sigma}$ and $\Omega_{A/R,\sigma}^1$ are dual to each other and we have*

- (1) *If $R \rightarrow R'$ is a base extension and $A' = R' \otimes_R A$, then there exists a canonical isomorphism*

$$R' \otimes_R T_{A/R,\sigma} \simeq T_{A'/R',\sigma'}.$$

- (2) *If B is a twisted localization of A , then there exists a canonical isomorphism*

$$B \otimes_A T_{A/R,\sigma} \simeq T_{B/R,\sigma}.$$

Definition 2.8. A *twisted connection* on an A -module M is an R -linear map

$$\nabla : M \mapsto M \otimes_A \Omega_{A/R, \sigma}^1$$

such that for all $s \in M$, for all $x \in A$,

$$\nabla(xs) = s \otimes d(x) + \sigma(x)\nabla(s).$$

An A -linear map between two A -modules with twisted connections is said to be *horizontal* if it is compatible with the connections.

Clearly, A -modules endowed with a connection and horizontal maps form a category $\nabla_{\sigma_A}\text{-Mod}$.

Remark. This definition is compatible with definition 2.2.1 by André [2001]. In particular, all the tannakian formalism applies but this is not what we are interested in.

Recall from [Le Stum and Quirós 2015b] that if D is a twisted derivation of A , then a twisted D -derivation of an A -module M is an R -linear endomorphism D_M of M that satisfies the twisted Leibnitz rule: for all $x \in A$, for all $s \in M$,

$$D_M(xs) = D(x)s + \sigma_A(x)D_M(s).$$

One may then consider the notion of *action by twisted derivations* of $T_{A/R, \sigma}$ on M : it is an R -linear action such that whenever $D \in T_{A/R, \sigma}$, the map $D_M : s \mapsto D.s$ is a D -derivation.

Proposition 2.9. *There exists a functor from the category of A -modules endowed with a twisted connection to the category of A -modules endowed with a linear action of $T_{A/R, \sigma}$ by twisted derivations. It is an equivalence (an isomorphism) when $\Omega_{A/R, \sigma}^1$ is free of finite rank.*

Proof. If M is endowed with a twisted connection $\nabla : M \mapsto M \otimes_A \Omega_{\sigma}^1$ and D is a twisted derivation of A , we may write uniquely $D = u \circ d$ with $u : \Omega_{\sigma}^1 \rightarrow A$ and consider the composite map $D_M := (\text{Id}_M \otimes u) \circ \nabla : M \rightarrow M$. Then we will have

$$\begin{aligned} D_M(xs) &= (\text{Id}_M \otimes u)(\nabla(xs)) \\ &= (\text{Id}_M \otimes u)(s \otimes d(x) + \sigma(x)\nabla(s)) \\ &= (u \circ d)(x)s + \sigma(x)(\text{Id}_M \otimes u)(\nabla(s)) = D(x)s + \sigma(x)D_M(s). \end{aligned}$$

Conversely, assume that M is endowed with an action of T_{σ} by twisted derivations. Let D_1, \dots, D_n be a basis of T_{σ} and $\omega_1, \dots, \omega_n$ be the dual basis in Ω_{σ}^1 . Then we

can define $\nabla(s) = \sum D_{i,M}(s) \otimes \omega_i$ and check that

$$\begin{aligned} \nabla(xs) &= \sum D_{i,M}(xs) \otimes \omega_i \\ &= \left(\sum D_i(x)s + \sigma(x)D_{i,M}(s) \right) \otimes \omega_i \\ &= s \otimes \sum D_i(x)\omega_i + \sigma(x) \sum D_{i,M}(s) \otimes \omega_i = s \otimes d(x) + \sigma(x)\nabla(s). \end{aligned}$$

Clearly, this is an inverse to the previous functor. \square

Proposition 2.10. *If $x \in A$ is a **twisted coordinate** on A , then $\Omega_{A/R,\sigma}^1$ is free of rank 1, generated by dx . Moreover, there exists a unique twisted derivation $\partial_{x,\sigma}$ of A such that $\partial_{x,\sigma}(x) = 1$ and we have for all $D \in T_{A/R,\sigma}$,*

$$D = D(x)\partial_{x,\sigma}.$$

Proof. The first assertion is a particular case of Corollary 1.13. The second one then follows from Corollary 2.5. \square

In particular, we see that a twisted coordinate is also a coordinate in the sense of [Le Stum and Quirós 2015b]. In order to lighten the notations, we will usually drop the index x but we must not forget that ∂_σ depends on the choice of x . Also, we would rather write $\partial_{A,\sigma}$ than ∂_{σ_A} when we want to make clear the dependence on A .

If x is a twisted coordinate on A , one may consider the *twisted Weyl algebra* $D_{A/R,\sigma,\partial}$ (see [Le Stum and Quirós 2015b] for example), that we will usually denote by $D_{A/R,\sigma}$ and sometimes simply by D_σ . This is the noncommutative polynomial ring in one variable ∂_σ over A with the commutation rule for all $z \in A$,

$$\partial_\sigma z = \partial_{A,\sigma}(z) + \sigma_A(z)\partial_\sigma.$$

Moreover, there exists an equivalence (an isomorphism) of categories

$$D_{A/R,\sigma}\text{-Mod} \simeq \partial_{A,\sigma}\text{-Mod},$$

where the latter denotes the category of A -modules M endowed with a $\partial_{A,\sigma}$ -derivation.

Proposition 2.11. *Assume that x is a twisted coordinate on A . Then there exists an equivalence (an isomorphism) of categories*

$$\nabla_{\sigma_A}\text{-Mod} \simeq D_{A/R,\sigma}\text{-Mod}.$$

Proof. Follows from Proposition 2.9. \square

3. Twisted binomial coefficient theorem for principal parts

We prove here the main theorem that will allow us to recover twisted differential operators from principal parts. We use the same notation as before.

In Section 1 we introduced the canonical map (it is a morphism of R -algebras)

$$P_{A/R} \xrightarrow{\delta_{A/R}} P_{A/R} \otimes_A P_{A/R}, \quad x\tilde{y} \longmapsto x \otimes \tilde{y}.$$

We want to investigate the interaction between σ_A and $\delta_{A/R}$.

Recall that we also considered in Section 1 the maps

$$P_{A/R} \xrightarrow{\sigma_P} P_{A/R}, \quad x\tilde{y} \longmapsto \sigma_A(x)\tilde{y},$$

which is an R -linear ring homomorphism, and

$$A \xrightarrow{d_{A/R}} P_{A/R}, \quad x \longmapsto \tilde{x} - x,$$

which is only R -linear. As usual, we will drop the subscripts in order to lighten the notation, hoping that the meaning will always be clear from the context.

Lemma 3.1. *For all $i = 0, \dots, n$, we have in $P \otimes_A P$ for all $x \in A$,*

$$\delta(\sigma^n(d(x))) = 1 \otimes \sigma^i(d(x)) + \sigma^{n-i}(d(\sigma^i(x))) \otimes 1.$$

Proof. We do the computations in $A \otimes_R A \otimes_R A$. The right-hand side is

$$\begin{aligned} 1 \otimes (1 \otimes x - \sigma^i(x) \otimes 1) + \sigma^{n-i}(1 \otimes \sigma^i(x) - \sigma^i(x) \otimes 1) \otimes 1 \\ = 1 \otimes 1 \otimes x - 1 \otimes \sigma^i(x) \otimes 1 + 1 \otimes \sigma^i(x) \otimes 1 - \sigma^n(x) \otimes 1 \otimes 1 \\ = 1 \otimes 1 \otimes x - \sigma^n(x) \otimes 1 \otimes 1. \end{aligned}$$

If we develop the left-hand side, we obtain exactly the same thing:

$$\delta(\sigma^n(d(x))) = \delta(1 \otimes x - \sigma^n(x) \otimes 1) = 1 \otimes 1 \otimes x - \sigma^n(x) \otimes 1 \otimes 1. \quad \square$$

We endow $P \otimes_A P$ with the endomorphism $\sigma_P \otimes_A \text{Id}_P$ (which is the same thing as $\sigma_A \otimes_R \text{Id}_A \otimes_R \text{Id}_A$ on $A \otimes_R A \otimes_R A$).

Proposition 3.2. *The map $\delta : P \rightarrow P \otimes_A P$ is a morphism of twisted R -algebras.*

Proof. This is the case $n = 1$ and $i = 0$ of Lemma 3.1. More precisely, if $x \in A$ and $\xi = \tilde{x} - x$, we have

$$\delta(\sigma(\xi)) = 1 \otimes \xi + \sigma(\xi) \otimes 1 = \sigma(\delta(\xi)). \quad \square$$

Proposition 3.3. *We have in $P \otimes_A P$ for all $n \in \mathbb{N}$,*

$$\delta(I^{(n)}) \subset \sum_{i=0}^n I^{(i)} \otimes I^{(n-i)}.$$

Proof. First of all, since I is generated by the image of d , it follows from Lemma 3.1 that for all $i = 0, \dots, n$, we have

$$\delta(\sigma^n(I)) \subset P \otimes \sigma^i(I) + \sigma^{n-i}(I) \otimes P.$$

Using induction, we obtain

$$\begin{aligned} \delta(I^{(n+1)}) &= \delta(I^{(n)})\delta(\sigma^n(I)) \subset \sum_{i=0}^n (I^{(i)} \otimes I^{(n-i)})(P \otimes \sigma^{n-i}(I) + \sigma^i(I) \otimes P) \\ &\subset \sum_{i=0}^n (I^{(i)} \otimes I^{(n-i+1)}) + \sum_{i=0}^n (I^{(i+1)} \otimes I^{(n-i)}) \\ &\subset \sum_{i=0}^{n+1} I^{(i)} \otimes I^{(n+1-i)}. \end{aligned} \quad \square$$

Corollary 3.4. *For all $m, n \in \mathbb{N}$, we have in $P \otimes_A P$:*

$$\delta(I^{(n+m+1)}) \subset P \otimes_A I^{(m+1)} + I^{(n+1)} \otimes_A P.$$

In other words, δ induces a map

$$P_{(n+m)\sigma} \xrightarrow{\delta_{n,m}} P_{(n)\sigma} \otimes_A P_{(m)\sigma}.$$

Proof. If $0 \leq i \leq m+n+1$, we have either $i > n$ and then $I^{(i)} \subset I^{(n+1)}$ or else $i \leq n$ so that $m+n+1-i > m$ and then $I^{(m+n+1-i)} \subset I^{(m+1)}$.

Going to the limit, we obtain a canonical homomorphism of R -algebras

$$\widehat{P}_\sigma \xrightarrow{\widehat{\delta}} \widehat{P}_\sigma \widehat{\otimes}_A \widehat{P}_\sigma,$$

where the right-hand side is, *by definition*, the inverse limit of all the $P_{(n)\sigma} \otimes_A P_{(m)\sigma}$. In other words, we obtain a comultiplication on \widehat{P}_σ that will allow us to turn its “dual” into a ring (more on this later). \square

We finish this section with the *quantum binomial theorem for principal parts*:

Theorem 3.5. *let A be a twisted commutative R -algebra and $x \in A$ such that $\sigma(x) = qx + h$ with $q, h \in R$. If we set $\xi = \tilde{x} - x$, then we have*

$$\delta(\xi^{(n)}) := \sum_{i=0}^n \binom{n}{i}_q \xi^{(n-i)} \otimes \xi^{(i)}.$$

Proof. The formula is proved to be correct by induction on n . First of all, since δ is a ring homomorphism, we have

$$\delta(\xi^{(n+1)}) = \delta(\xi^{(n)}\sigma^n(\xi)) = \delta(\xi^{(n)})\delta(\sigma^n(\xi)).$$

Using induction and Lemma 3.6 below, we get

$$\begin{aligned}
 \delta(\xi^{(n+1)}) &= \sum_{i=0}^n \binom{n}{i}_q (\xi^{(n-i)} \otimes \xi^{(i)}) (1 \otimes \sigma^i(\xi) + q^i \sigma^{n-i}(\xi) \otimes 1) \\
 &= \sum_{i=0}^n \binom{n}{i}_q \xi^{(n-i)} \otimes \xi^{(i)} \sigma^i(\xi) + \sum_{i=0}^n \binom{n}{i}_q q^i \xi^{(n-i)} \sigma^{n-i}(\xi) \otimes \xi^{(i)} \\
 &= \sum_{i=0}^n \binom{n}{i}_q \xi^{(n-i)} \otimes \xi^{(i+1)} + \sum_{i=0}^n \binom{n}{i}_q q^i \xi^{(n-i+1)} \otimes \xi^{(i)} \\
 &= \sum_{i=1}^{n+1} \binom{n}{i-1}_q \xi^{(n-i+1)} \otimes \xi^{(i)} + \sum_{i=0}^n \binom{n}{i}_q q^i \xi^{(n-i+1)} \otimes \xi^{(i)} \\
 &= \sum_{i=0}^{n+1} \left(\binom{n}{i-1}_q + q^i \binom{n}{i}_q \right) \xi^{(n+1-i)} \otimes \xi^{(i)} \\
 &= \sum_{i=0}^{n+1} \binom{n+1}{i}_q \xi^{(n+1-i)} \otimes \xi^{(i)}. \quad \square
 \end{aligned}$$

Lemma 3.6. *Under the hypothesis of the proposition, we have for all $i = 0, \dots, n$,*

$$\delta(\sigma^n(\xi)) = 1 \otimes \sigma^i(\xi) + q^i \sigma^{n-i}(\xi) \otimes 1.$$

Proof. We have

$$d(\sigma(x)) = \widetilde{\sigma(x)} - \sigma(x) = \widetilde{qx + h} - (qx + h) = q(\tilde{x} - x) = q\xi.$$

The analogous result holds for the endomorphism σ^i . It follows that $d(\sigma^i(x)) = q^i \xi$ and we finish with Lemma 3.1. \square

4. Twisted differential operators of infinite level

We are now able to define the ring of twisted differential operators (of infinite level). We keep the previous notation.

Definition 4.1. If M and N are two A -modules, then a *twisted differential operator* $\varphi : M \rightarrow N$ of order at most n (and infinite level) is an R -linear map whose A -linearization

$$\begin{array}{ccc}
 & \xrightarrow{\quad \tilde{\varphi} \quad} & \\
 P \otimes_A M & \xrightarrow{\quad \simeq \quad} & A \otimes_R M \longrightarrow N \\
 & x \otimes s \longmapsto & x\varphi(s)
 \end{array}$$

factors through $P_{(n)_\sigma} \otimes_A M$.

Note that the condition means that the restriction of $\tilde{\varphi}$ to $I^{(n+1)} \otimes_A M$ is zero. We might still write $\tilde{\varphi}$ for the map induced on $P_{(n)\sigma} \otimes_A M$ when there is no risk of confusion.

We denote by $\text{Diff}_{n,\sigma}(M, N)$ the set of all twisted differential operators of order at most n . Thus, we have a canonical isomorphism

$$\text{Diff}_{n,\sigma}(M, N) \simeq \text{Hom}_A(P_{(n)\sigma} \otimes_A M, N),$$

where $P_{(n)\sigma}$ is seen as an A -module for the action on the *right* with respect to \otimes_A and for the action on the *left* with respect to Hom_A . In particular, $\text{Diff}_{n,\sigma}(M, N)$ has the natural structure of a $P_{(n)\sigma}$ -module given by $x\tilde{y} \cdot \varphi := x \circ \varphi \circ y$, where multiplication by y takes place in M while we multiply by x in N .

We will also denote by $\text{Diff}_\sigma(M, N)$ the set of all twisted differential operators of any order so that

$$\text{Diff}_\sigma(M, N) \simeq \varinjlim \text{Hom}_A(P_{(n)\sigma} \otimes_A M, N).$$

In particular, we see that $\text{Diff}_\sigma(M, N)$ has the natural structure of a \widehat{P}_σ -module.

In the case $N = M$, we will write $\text{Diff}_{n,\sigma}(M)$ and $\text{Diff}_\sigma(M)$. We also set $D_{A/R,\sigma}^{(\infty)} := \text{Diff}_\sigma(A)$ and we will often drop the index A/R and simply write $D_\sigma^{(\infty)}$.

Definition 4.2. Let x be a twisted coordinate on A and $\xi = \tilde{x} - x$. Then the *standard basis* of $\text{Diff}_{n,\sigma}(A)$ is the basis $\partial_\sigma^{[k]}$ dual to the images of the $\xi^{(k)}$ in $P_{(n)\sigma}$. We call $\partial_\sigma^{[k]}$ the *standard twisted divided differential operator of order k associated to x* .

In other words (we will come back to this in Section 7), the canonical basis is characterized by the property that for all $k, n \in \mathbb{N}$,

$$\widetilde{\partial_\sigma^{[k]}}(\xi^{(n)}) = \begin{cases} 1 & \text{if } k = n, \\ 0 & \text{otherwise,} \end{cases}$$

where $\widetilde{\partial_\sigma^{[k]}}$ denotes the A -linearization of the R -linear endomorphism $\partial_\sigma^{[k]}$ of A .

Thus, when x is a twisted coordinate, any $\varphi \in D_\sigma^{(\infty)}$ can be uniquely written as a finite sum $\sum z_k \partial_\sigma^{[k]}$ with $z_k \in A$ (and conversely, any such sum is in $D_\sigma^{(\infty)}$).

The next proposition shows that the A -module of twisted differential operators could have also been defined by induction on the order n (this is sometimes more convenient and does not require one to work out the theory of principal parts). Instead, it uses the notion of twisted bracket for all $\varphi \in \text{Hom}_R(M, N)$, for all $x \in A$,

$$[\varphi, x]_\sigma = \varphi \circ x - \sigma(x) \circ \varphi,$$

already used in [Le Stum and Quirós 2015b]. We will need this intermediate result:

Lemma 4.3. *Let M, N be two A -modules, $\varphi \in \text{Hom}_R(M, N)$ and $x \in A$. If we set $\varphi_x := [\varphi, x]_\sigma$, then*

$$\tilde{\varphi}_x = \tilde{\varphi} \circ (\sigma(\xi) \otimes_A \text{Id}_M) : P \otimes_A M \rightarrow N.$$

Proof. We do the computations in $A \otimes_R M$. Let $y \in A$ and $s \in M$. Then

$$\begin{aligned} \tilde{\varphi}(\sigma(\xi)(y \otimes s)) &= \tilde{\varphi}((1 \otimes x - \sigma(x) \otimes 1)(y \otimes s)) \\ &= \tilde{\varphi}(y \otimes xs) - \tilde{\varphi}(\sigma(x)y \otimes s) \\ &= y\varphi(xs) - \sigma(x)y\varphi(s) \\ &= y\varphi_x(s) = \tilde{\varphi}_x(y \otimes s). \end{aligned} \quad \square$$

Proposition 4.4. *Let M, N be two A -modules and $\varphi \in \text{Hom}_R(M, N)$. Then for all $n \in \mathbb{N}$, we have*

$$\varphi \in \text{Diff}_{n,\sigma}(M, N) \Leftrightarrow \text{for all } x \in A, [\varphi, x]_{\sigma^n} \in \text{Diff}_{n-1,\sigma}(M, N).$$

Note that we can start the induction process with $\text{Diff}_{0,\sigma}(M, N) = \text{Hom}_A(M, N)$ or with $\text{Diff}_{-1,\sigma}(M, N) = 0$ if we prefer.

Proof. For $x \in A$, we set $\varphi_x := [\varphi, x]_{\sigma^n}$. Then we consider the linearizations

$$\tilde{\varphi}, \tilde{\varphi}_x : P \otimes_A M \rightarrow N$$

of φ and φ_x , respectively, and apply Lemma 4.3 to σ^n so that

$$\tilde{\varphi}_x = \tilde{\varphi} \circ (\sigma^n(\xi) \otimes_A \text{Id}_M).$$

Thus we see that $\tilde{\varphi} = 0$ on $I^{(n+1)} \otimes_A M = I^{(n)} \sigma^n(I) \otimes_A M$ if and only if $\tilde{\varphi}_x = 0$ on $I^{(n)} \otimes_A M$ for all $x \in A$. In other words, $\tilde{\varphi}$ factors through $P_{(n)\sigma} \otimes_A M$ if and only if all $\tilde{\varphi}_x$ factor through $P_{(n-1)\sigma} \otimes_A M$. \square

Corollary 4.5. *A twisted differential operator of order at most n (and infinite level) from M to N is an R -linear map $\varphi : M \rightarrow N$ such that for all $x_0, \dots, x_n \in A$,*

$$[[\cdot \cdots [[\varphi, x_n]_{\sigma^n}, x_{n-1}]_{\sigma^{n-1}} \cdots]_\sigma, x_0] = 0.$$

Remarks. (1) Be careful that, with the notation of Valery Lunts and Alexander L. Rosenberg [1997], our $\text{Diff}_\sigma(M, N)$ is different from their $\text{Diff}(M, N^\sigma)$ which is defined by the condition for all $x_0, \dots, x_n \in A$,

$$[[\cdot \cdots [[\varphi, x_n]_\sigma, x_{n-1}]_\sigma \cdots]_\sigma, x_0] = 0.$$

They only coincide when $n = 0, 1$.

(2) Our $\text{Diff}_\sigma(M, N)$ should however coincide with some flavor of the $\text{Diff}_\beta(M, N)$ of Lunts and Rosenberg. More precisely, in order to define this module, they need a G -grading on A and a bilinear map $\beta : G \times G \rightarrow R^\times$ (they use k and R for our R

and A). In the simplest nontrivial case $A = R[x]$ and $\beta(m, n) = q^{-mn}$, we believe that their $D_\beta(A)$ coincides with our $D_{A/R, \sigma}^{(\infty)}$ but their $D_q(A)$ is bigger (see [Iyer and McCune 2002] for example).

(3) Charlotte Hardouin [2010, Definition 2.4] introduces what she calls an *iterative q -difference ring* or ID_q -ring for short. She chooses some nonzero $q \in K$ where K is a fixed algebraically closed field and endows the field $A := K(x)$ of rational functions on K with the automorphism $\sigma(x) = qx$. Then if we look carefully at her conditions, we see that an ID_q -ring is a finitely generated A -algebra B with a structure of $D_{A/K, \sigma}^{(\infty)}$ -module, denoted by $(\varphi, y) \mapsto \varphi(y)$, such that the map $y \mapsto \sigma(y)$ is an automorphism of the ring B and for all $k \in \mathbb{N}$, for all $y, z \in B$,

$$(6) \quad \partial_\sigma^{[k]}(yz) = \sum_{i+j=k} (\sigma^i \partial_\sigma^{[j]})(y) \partial_\sigma^{[i]}(z).$$

Note that B becomes a twisted A -algebra, that is in fact *inversive* (which simply means that σ is bijective on B , a condition that is built into Hardouin's definition), and that condition (6) is automatic if the q -characteristic of B is zero.

Proposition 4.6. *Composition of twisted differential operators gives a twisted differential operator. Moreover, its order is at most the sum of the order of the components.*

Proof. We let $\varphi : M \rightarrow N$ be a twisted differential operator of order n and $\psi : L \rightarrow M$ a twisted differential operator of order m and consider the factorization

$$P \otimes_A L \xrightarrow{\delta \otimes \text{Id}} P \otimes_A P \otimes_A L \xrightarrow{\text{Id} \otimes_A \tilde{\psi}} P \otimes_A M \xrightarrow{\tilde{\varphi}} N$$

of Lemma 1.2. The map $\tilde{\varphi}$ factors through $P_{(n)_\sigma} \otimes_A M$ and $\text{Id} \otimes \tilde{\psi}$ factors through $P \otimes_A P_{(m)_\sigma} \otimes_A L$. Therefore, their composite factors through $P_{(n)_\sigma} \otimes_A P_{(m)_\sigma} \otimes_A L$ and it follows from Corollary 3.4 that the whole thing will factor through $P_{(n+m)_\sigma} \otimes L$. Thus, $\varphi \circ \psi$ is a twisted differential operator of order at most $m + n$. \square

It is sometimes useful to have a general formula for the commutation of twisted differential operators with the twisted coordinate x :

Proposition 4.7. *We have for all $k \in \mathbb{N} \setminus \{0\}$,*

$$\partial_\sigma^{[k]} \circ x = \sigma^k(x) \partial_\sigma^{[k]} + \partial_\sigma^{[k-1]}.$$

Proof. In order to make the proof easier to understand, we will still write $\tilde{x} = 1 \otimes x \in P$, but we will denote the multiplication maps by

$$A \xrightarrow{m_x} A, \quad y \mapsto xy \quad \text{and} \quad P \xrightarrow{m_{\tilde{x}}} P \quad \varphi \mapsto \tilde{x}\varphi.$$

Then it is easy to see that $m_{\tilde{x}}$ splits as

$$P \xrightarrow{\delta} P \otimes_A P \xrightarrow{\text{Id}_P \otimes_A \tilde{m}_x} P,$$

where \widetilde{m}_x denotes the linearization of m_x . It then follows from Lemma 1.2 that

$$\widetilde{\partial_\sigma^{[k]} \circ m_x} = \widetilde{\partial_\sigma^{[k]} \circ m_{\tilde{x}}}.$$

On the other hand, we have

$$\xi^{(n+1)} = \xi^{(n)} \sigma^n(\xi) = \xi^{(n)}(\tilde{x} - \sigma^n(x)) = \tilde{x} \xi^{(n)} - \sigma^n(x) \xi^{(n)},$$

from which we derive $\tilde{x} \xi^{(n)} = \xi^{(n+1)} + \sigma^n(x) \xi^{(n)}$. Putting all these together, we get

$$\begin{aligned} (\widetilde{\partial_\sigma^{[k]} \circ m_x})(\xi^{(n)}) &= (\widetilde{\partial_\sigma^{[k]} \circ m_{\tilde{x}}})(\xi^{(n)}) \\ &= \widetilde{\partial_\sigma^{[k]}(\tilde{x} \xi^{(n)})} \\ &= \widetilde{\partial_\sigma^{[k]}(\xi^{(n+1)} + \sigma^n(x) \xi^{(n)})} = \begin{cases} 1 & \text{if } n = k-1, \\ \sigma^k(x) & \text{if } n = k, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

In other words, we have

$$\partial_\sigma^{[k]} \circ m_x = \sigma^k(x) \partial_\sigma^{[k]} + \partial_\sigma^{[k-1]}.$$

□

Corollary 4.8. *If $x \in A^\times$, then for all $k \in \mathbb{N} \setminus \{0\}$,*

$$\partial_\sigma^{[k]} \circ x^{-1} = \sum_{i=0}^k \frac{(-1)^i}{\prod_{j=0}^i \sigma^{k-j}(x)} \partial_\sigma^{[k-i]}.$$

Proof. Using for a moment the convention $\partial_\sigma^{[-1]} = 0$, which makes Proposition 4.7 formally valid for $k = 0$, we compute

$$\begin{aligned} \sum_{i=0}^k \frac{(-1)^i}{\prod_{j=0}^i \sigma^{k-j}(x)} \partial_\sigma^{[k-i]} \circ x &= \sum_{i=0}^k \frac{(-1)^i}{\prod_{j=0}^i \sigma^{k-j}(x)} (\sigma^{k-i}(x) \partial_\sigma^{[k-i]} + \partial_\sigma^{[k-i-1]}) \\ &= \sum_{i=0}^k \frac{(-1)^i}{\prod_{j=0}^{i-1} \sigma^{k-j}(x)} \partial_\sigma^{[k-i]} + \sum_{i=0}^{k-1} \frac{(-1)^i}{\prod_{j=0}^i \sigma^{k-j}(x)} \partial_\sigma^{[k-i-1]} = \partial_\sigma^{[k]}. \quad \square \end{aligned}$$

Proposition 4.9. *Let M, N be two A -modules.*

- (1) *Let $R \rightarrow R'$ be any morphism of commutative rings and $A' := R' \otimes_R A$, endowed with $\text{Id}_{R'} \otimes_R \sigma_A$. Then we have*

$$\text{Diff}_{\sigma_{A'}}(A' \otimes_A M, A' \otimes_A N) \simeq A' \otimes_A \text{Diff}_{\sigma_A}(M, N).$$

- (2) *If $A \rightarrow B$ is a twisted localization and M is finitely presented, then we have*

$$\text{Diff}_{\sigma_B}(B \otimes_A M, B \otimes_A N) \simeq B \otimes_A \text{Diff}_{\sigma_A}(M, N).$$

Proof. Follows from propositions 1.6 and 1.7 and the fact that direct limits commute with tensor product (and localization is flat). \square

Remark. (1) As a particular case, we will have

$$D_{A'/R',\sigma'}^{(\infty)} \simeq A' \otimes_A D_{A/R,\sigma}^{(\infty)} \quad \text{and} \quad D_{B/R,\sigma}^{(\infty)} \simeq B \otimes_A D_{A/R,\sigma}^{(\infty)}.$$

(2) When A is a quantum R -algebra, we will always have (see the remark following Proposition 1.10) a natural isomorphism of A -modules

$$D_{A/R,\sigma}^{(\infty)} \simeq A \otimes_{R[x]} D_{R[x]/R,\sigma}^{(\infty)}.$$

Note however that the ring structure (or equivalently the action on A) plays a fundamental role.

Proposition 4.10. *We have $\text{Diff}_{0,\sigma}(A) = A$ and $\text{Diff}_{1,\sigma}(A) = A \oplus T_{A/R,\sigma}$.*

Proof. The first assertion follows from the fact that $P_{(0)\sigma} = A$ and the second one from Proposition 2.2. \square

Recall that we introduced in [Le Stum and Quirós 2015b] the ring $\bar{D}_{A/R,\sigma}$ of *small* twisted differential operators of A/R as the smallest subring of $\text{End}_R(A)$ containing both A and $T_{A/R,\sigma}$. Again, we will simply write \bar{D}_σ when we believe that there is no risk of confusion. Then we have the following:

Corollary 4.11. *The ring of small twisted differential operators is contained inside the ring of twisted differential operators of infinite level: we have*

$$\bar{D}_{A/R,\sigma} \subset D_{A/R,\sigma}^{(\infty)}.$$

When there exists a twisted coordinate, we can make the twisted Weyl algebra enter the picture and we have:

Corollary 4.12. *If x is a twisted coordinate on A , there exists an epi-mono factorization*

$$D_{A/R,\sigma} \twoheadrightarrow \bar{D}_{A/R,\sigma} \hookrightarrow D_{A/R,\sigma}^{(\infty)}.$$

Proof. There exists a natural map $D_\sigma \rightarrow \bar{D}_\sigma$ that sends the parameter ∂_σ of $D_{\sigma,\partial}$ to the corresponding endomorphism $\partial_{A,\sigma}$ of A . And it is surjective since $\partial_{A,\sigma}$ is a generator of T_σ . \square

None of the maps are in general bijective, even in characteristic zero. However, as we will see in Theorem 6.3 below, there are some important cases where both maps are bijective.

At some point, we will need to be able to compare twisted differential operators with respect to σ and twisted differential operators with respect to the powers (or roots) of σ .

Proposition 4.13. *For all $m > 0$, if M, N are two A -modules, we have*

$$\text{Diff}_{\sigma^m}(M, N) \subset \text{Diff}_{\sigma}(M, N).$$

Proof. Since

$$I^{(mn)\sigma} = I\sigma(I) \cdots \sigma^{mn}(I) \subset I\sigma^m(I) \cdots \sigma^{mn}(I) = I^{(n)\sigma^m},$$

there exists an natural surjective map

$$P_{(nm)\sigma} \rightarrow P_{(n)\sigma^m},$$

from which we derive an inclusion $\text{Diff}_{n,\sigma^m}(M, N) \subset \text{Diff}_{mn,\sigma}(M, N)$. \square

Recall from [Le Stum and Quirós 2015a] that a system of roots of σ is a family $\underline{\sigma} := \{\sigma_n\}_{n \in S}$, with $S \subset \mathbb{N}$, of R -linear ring endomorphisms of A such that $\sigma_n^m = \sigma_{n'}^{m'}$ whenever $m/n = m'/n'$ and $\sigma_n^n = \sigma$. We will always assume that S is filtering for division in the sense that if $m, m' \in S$, then there always exists $m'' \in S$ such that $m \mid m''$ and $m' \mid m''$. We will call the pair $(A, \underline{\sigma})$ an S -twisted R -algebra.

Definition 4.14. Let $\underline{\sigma} := \{\sigma_n\}_{n \in S}$ be a system of roots of σ . Then the *ring of twisted differential operators (of infinite level)* $D_{A/R, \underline{\sigma}}^{(\infty)}$ is the R -subalgebra of $\text{End}_R(A)$ generated by all σ_n -differential operators (of infinite level) for all $n \in S$.

Proposition 4.13 then has the following consequence:

Corollary 4.15. *If $\underline{\sigma}$ is a system of roots of σ , we have*

$$D_{A/R, \underline{\sigma}}^{(\infty)} = \bigcup D_{A/R, \sigma_n}^{(\infty)}.$$

Note that in section 3 of [Le Stum and Quirós 2015b] we defined in exactly the same way the ring of small twisted differential operators $\bar{D}_{A/R, \underline{\sigma}}$ for any family $\underline{\sigma}$ (not necessarily a root system), but we showed that the analogous statement is *not* true in general.

5. Twisted Taylor series

We will develop here the formalism of twisted Taylor maps which describes the formal solutions of twisted differential modules. Notations are as before.

Lemma 5.1. *If M is a $D_{A/R, \sigma}^{(\infty)}$ -module, then the canonical map $D_{A/R, \sigma}^{(\infty)} \rightarrow \text{End}_R(M)$ induces, for all $n \in \mathbb{N}$, a $P_{A/R, (n)\sigma}$ -linear map*

$$\text{Diff}_{n, \sigma}(A) \rightarrow \text{Diff}_{n, \sigma}(M).$$

Hence, there exists a canonical \widehat{P}_{σ} -linear map

$$D_{A/R}^{(\infty)} \rightarrow \text{Diff}_{\sigma}(M).$$

Proof. Since the canonical map $\lambda : D_\sigma^{(\infty)} \rightarrow \text{End}_R(M)$ is a morphism of A -algebras, it will commute with the action of P . More precisely, for all $x, y \in A$ and $\varphi \in D_\sigma^{(\infty)}$, we have (see Section 4 for the definition of the action of P on $D_\sigma^{(\infty)}$)

$$\lambda(x\tilde{y} \cdot \varphi) = \lambda(x \circ \varphi \circ y) = x \circ \lambda(\varphi) \circ y = x\tilde{y} \cdot \lambda(\varphi).$$

In particular, if $\tilde{\varphi}$ is zero on $I^{(n+1)}$, then $\widetilde{\lambda(\varphi)}$ will be zero on $I^{(n+1)} \otimes_A M$. It means that the image of $\text{Diff}_{n,\sigma}(A)$ falls inside $\text{Diff}_{n,\sigma}(M)$. \square

We will usually denote by $\varphi_M \in \text{Diff}_\sigma(M)$ the image of $\varphi \in D_\sigma^{(\infty)}$. In other words, for $\varphi \in D_\sigma^{(\infty)}$ and $s \in M$, we will have $\varphi_M(s) = \varphi s$.

Definition 5.2. A *twisted Taylor structure (of infinite level)* on an A -module M is a compatible family of A -linear maps $\theta_n : M \rightarrow M \otimes_A P_{(n)\sigma}$ (called *twisted Taylor maps*) with $\theta_0 = \text{Id}$, making commutative all the diagrams

$$\begin{array}{ccc} M & \xrightarrow{\theta_m} & M \otimes_A P_{(m)\sigma} \\ \downarrow \theta_{n+m} & & \downarrow \theta_n \otimes \text{Id} \\ M \otimes_A P_{(n+m)\sigma} & \xrightarrow{\text{Id} \otimes \delta_{n,m}} & M \otimes_A P_{(n)\sigma} \otimes_A P_{(m)\sigma} \end{array}$$

For example, the canonical twisted Taylor structure on A is defined by the family of composite maps

$$\begin{array}{ccc} A & \xrightarrow{\theta} & P \\ & \searrow \theta_n & \downarrow \\ & & P_{(n)\sigma} \end{array}$$

where the upper map is the Taylor map $x \mapsto \tilde{x}$ given by the action on the right (see Section 1).

There exists an obvious notion of morphism of A -modules endowed with a twisted Taylor structure and they form a category.

Proposition 5.3. *Let M be an A -module endowed with a twisted Taylor structure $(\theta_n)_{n \in \mathbb{N}}$. Then there exists a unique structure of a $D_{A/R,\sigma}^{(\infty)}$ -module on M such that, for all $n \in \mathbb{N}$, we have*

$$(7) \quad \theta_n(s) = \sum s_k \otimes f_k \Rightarrow \text{for all } \varphi \in \text{Diff}_{n,\sigma}(A), \varphi_M(s) = \sum \tilde{\varphi}(f_k) s_k.$$

This is functorial in M in the sense that any morphism of A -modules $M \rightarrow N$ which is compatible with some twisted Taylor structures on M and N will automatically be $D_{A/R,\sigma}^{(\infty)}$ -linear. Moreover, this is an equivalence (an isomorphism) of categories if all $P_{(n)\sigma}$ are finite projective (for the left A -module structure).

Note that the last condition is satisfied if there exists a twisted coordinate on A .

Proof. First of all, there exists for all $n \in \mathbb{N}$ a canonical morphism of A -modules

$$(8) \quad M \otimes_A P_{(n)\sigma} \rightarrow \text{Hom}_A(\text{Hom}_A(P_{(n)\sigma}, A), M),$$

which is automatically $P_{(n)\sigma}$ -linear. Now, A -linear maps

$$(9) \quad \theta_n : M \rightarrow M \otimes_A P_{(n)\sigma}$$

correspond bijectively to $P_{(n)\sigma}$ -linear maps

$$\epsilon_n : P_{(n)\sigma} \otimes_A M \rightarrow M \otimes_A P_{(n)\sigma},$$

and we can compose with the map (8) in order to get

$$P_{(n)\sigma} \otimes_A M \rightarrow \text{Hom}_A(\text{Hom}_A(P_{(n)\sigma}, A), M),$$

or equivalently,

$$\text{Hom}_A(P_{(n)\sigma}, A) \rightarrow \text{Hom}_A(P_{(n)\sigma} \otimes_A M, M).$$

In other words, we obtain $P_{(n)\sigma}$ -linear maps

$$(10) \quad \text{Diff}_{n,\sigma}(A) \rightarrow \text{Diff}_{n,\sigma}(M), \quad \varphi \mapsto \varphi_M.$$

Formula (7) follows directly from the construction. Compatibility for various n in (10) follows from compatibility for various n in (9). We need to show that the corresponding map $D_\sigma^{(\infty)} \rightarrow \text{End}(M)$ is a morphism of rings. To do that, one can use the description of composition of twisted differential operators given in Proposition 4.6: we need to verify that the maps (9) are compatible with δ which is exactly the condition in the definition of Taylor structure.

This construction is clearly functorial. Moreover, if $P_{(n)\sigma}$ is finite projective (for the left A -module structure), then the map (8) is actually an isomorphism. And it follows from Lemma 5.1 that a $D_\sigma^{(\infty)}$ -module structure on M will provide us with a compatible family of maps as in (10). \square

Definition 5.4. Let M be an A -module endowed with a twisted Taylor structure. Then the *twisted Taylor map* of M is the map

$$\hat{\theta} = \varprojlim \theta_n : M \mapsto M \widehat{\otimes}_A \widehat{P}_\sigma := \varprojlim M \otimes_A P_{(n)\sigma}.$$

The *twisted Taylor series* of $s \in M$ is $\hat{\theta}(s) \in M \widehat{\otimes}_A \widehat{P}_\sigma$.

There exists a commutative diagram

$$\begin{array}{ccc} M & \xrightarrow{\hat{\theta}} & M \widehat{\otimes}_A \widehat{P}_\sigma \\ \downarrow \hat{\theta} & & \downarrow \hat{\theta} \widehat{\otimes} \text{Id} \\ M \widehat{\otimes}_A \widehat{P}_\sigma & \xrightarrow{\text{Id} \widehat{\otimes} \hat{\delta}} & M \widehat{\otimes}_A \widehat{P}_\sigma \widehat{\otimes}_A \widehat{P}_\sigma \end{array}$$

and the action of A on \widehat{P}_σ on the right is given by the Taylor map of A .

In practice, we will only consider the case of finitely presented A -modules M , and then the completed tensor product is the usual tensor product $M \otimes_A \widehat{P}_\sigma$.

We just showed in Proposition 5.3 that any $D_\sigma^{(\infty)}$ -module comes with a canonical twisted Taylor structure. When there exists a twisted coordinate on A , we can describe it explicitly as follows:

Proposition 5.5 (twisted Taylor formula). *Assume that x is a twisted coordinate on A and let $\xi = \tilde{x} - x$. If M is a $D_\sigma^{(\infty)}$ -module, we have for all $n \in \mathbb{N}$,*

$$\theta_n(s) = \sum_{k=0}^n \partial_\sigma^{[k]}(s) \otimes \xi^{(k)} \in M \otimes_A P_{(n)_\sigma}$$

and

$$\hat{\theta}(s) = \sum_{k=0}^{\infty} \partial_\sigma^{[k]}(s) \otimes \xi^{(k)} \in M \widehat{\otimes}_A \widehat{P}_\sigma.$$

Proof. This follows from equation (7): if we write

$$\theta_n(s) = \sum_{k=0}^n s_k \otimes \xi^{(k)},$$

we will have for all $l \in \mathbb{N}$,

$$\partial_\sigma^{[l]}(s) = \sum_{k=0}^n \tilde{\partial}_\sigma^{[l]}(\xi^{(k)}) s_k = s_l. \quad \square$$

In particular, we see that, if $z \in A$, then the image in \widehat{P}_σ of $\tilde{z} = 1 \otimes z \in P$ is the twisted Taylor series

$$\hat{\theta}(z) = \sum_k \partial_\sigma^{[k]}(z) \xi^{(k)}.$$

This explains why it is legitimate to call the map $z \mapsto \tilde{z}$ the Taylor map.

Examples. (1) If x is a twisted coordinate on A , we have $\hat{\theta}(x) = x + \xi$ and

$$\hat{\theta}(x^2) = x^2 + (x + \sigma(x))\xi + \xi^2.$$

(2) Assume that $x \in A^\times$ is an invertible twisted coordinate on A and that $\sigma(x) = qx$ with $q \in R^\times$. Then we have

$$\hat{\theta}\left(\frac{1}{x}\right) = \sum_{k=0}^{\infty} (-1)^k \frac{\xi^{(k)}}{q^{\frac{k(k+1)}{2}} x^{k+1}} = \frac{1}{x} - \frac{\xi}{qx^2} + \frac{\xi^{(2)}}{q^3 x^3} - \cdots.$$

Remark. If A is a twisted localization of $R[x]$, then there exists at most one R -algebra homomorphism $\hat{\theta} : A \rightarrow \widehat{P}_\sigma$ such that $\hat{\theta}(x) = \tilde{x}$. It means that the twisted

Taylor map

$$A \xrightarrow{\hat{\theta}} \widehat{P}_\sigma, \quad z \mapsto \sum_{k=0}^{\infty} \partial_\sigma^{[k]}(z) \xi^{(k)}$$

is the *unique* such map.

6. Quantum differential operators

In the quantum situation, we can be a lot more explicit as we shall see now. Thus, we assume in this section that A is a quantum R -algebra: we are given a twisted coordinate x such that $\sigma(x) = qx + h$ with $q, h \in R$.

Proposition 6.1. *We have for all $k, l \in \mathbb{N}$,*

$$\partial_\sigma^{[k]} \circ \partial_\sigma^{[l]} = \binom{k+l}{l}_q \partial_\sigma^{[k+l]}.$$

Proof. It follows from Lemma 1.2 that

$$\widetilde{\partial_\sigma^{[k]} \circ \partial_\sigma^{[l]}} = \widetilde{\partial_\sigma^{[k]}} \circ (\text{Id} \otimes \widetilde{\partial_\sigma^{[l]}}) \circ \delta.$$

Thus, using Theorem 3.5, we see that

$$\begin{aligned} (\widetilde{\partial_\sigma^{[k]} \circ \partial_\sigma^{[l]}})(\xi^{(n)}) &= \widetilde{\partial_\sigma^{[k]}} \left((\text{Id} \otimes \widetilde{\partial_\sigma^{[l]}}) \left(\sum_{i=0}^n \binom{n}{i}_q \xi^{(n-i)} \otimes \xi^{(i)} \right) \right) \\ &= \sum_{i=0}^n \binom{n}{i}_q \widetilde{\partial_\sigma^{[k]}}(\xi^{(n-i)} \otimes \widetilde{\partial_\sigma^{[l]}}(\xi^{(i)})) \\ &= \binom{n}{l}_q \widetilde{\partial_\sigma^{[k]}}(\xi^{(n-l)}) = \begin{cases} \binom{k+l}{l}_q & \text{if } n = k+l, \\ 0 & \text{otherwise.} \end{cases} \quad \square \end{aligned}$$

Recall from [Le Stum and Quirós 2015a] that if $m \in \mathbb{N}$, we write $(m)_q := \binom{m}{1}_q = 1 + \dots + q^{m-1}$ and we define by induction $(m)_q! := (m)_q(m-1)_q!$.

Corollary 6.2. *We have for all $k \in \mathbb{N}$, for all $z \in A$,*

$$\partial_\sigma^k(z) = (k)_q! \partial_\sigma^{[k]}(z).$$

Proof. We proceed by induction on k and obtain

$$\begin{aligned} \partial_\sigma^{k+1}(z) &= \partial_\sigma^k(\partial_\sigma(z)) = (k)_q! \partial_\sigma^{[k]}(\partial_\sigma(z)) \\ &= (k)_q! (\partial_\sigma^{[k]} \partial_\sigma)(z) \\ &= (k)_q! \binom{k+1}{1}_q \partial_\sigma^{[k+1]}(z) = (k+1)_q! \partial_\sigma^{[k+1]}(z). \quad \square \end{aligned}$$

The next result is important because it describes explicitly the relations between the different rings of twisted differential operators introduced so far. Recall from Corollary 4.12 that there exists an epi-mono factorization

$$D_{A/R,\sigma} \twoheadrightarrow \bar{D}_{A/R,\sigma} \hookrightarrow D_{A/R,\sigma}^{(\infty)}.$$

Recall from [Le Stum and Quirós 2015a] that the ring R is said to be q -flat (resp. q -divisible) if $(m)_q$ is never a zero-divisor (resp. is always invertible) unless $(m)_q = 0$ and also that the q -characteristic of R is the smallest positive integer p such that $(p)_q = 0$, if it exists, and 0 otherwise.

Theorem 6.3. *Assume R is q -divisible and let A be a q - R -algebra. Then:*

(1) *If q -char(A) = 0, we have*

$$D_{A/R,\sigma} = \bar{D}_{A/R,\sigma} = D_{A/R,\sigma}^{(\infty)}.$$

(2) *If q -char(A) = $p > 0$, we have*

$$D_{A/R,\sigma} / \partial_\sigma^p \simeq \bar{D}_{A/R,\sigma} \simeq D_{A/R,\sigma}^{(\infty)} / K_\sigma^{[p]},$$

where $K_\sigma^{[p]}$ is the free A -module generated by all $\partial_\sigma^{[k]}$ for $k \geq p$.

Note that with some extra conventions, assertions (1) and (2) could be combined.

Proof. We use the formula from Corollary 6.2.

In situation (1), all q -integers are invertible in A and the composite map sends the canonical basis $\{\partial_\sigma^k\}_{k \in \mathbb{N}}$ of D_σ bijectively onto a basis of $D_\sigma^{(\infty)}$. The first assertion follows.

In situation (2), we have $(p)_q = 0$ in A and the element $\partial_\sigma^p \in D_\sigma$ is therefore sent to 0. But since R is q -divisible, we have $(m)_q \in R^\times$ for $m < p$, and the composite map sends the family $\{\partial_\sigma^k\}_{k < p}$ of D_σ bijectively onto a basis of $D_\sigma^{(\infty)} / K_\sigma^{[p]}$. \square

Remarks. (1) The hypothesis in (1) is satisfied in the classical cases of differential equations, finite difference equations and q -difference equations.

More precisely, it is satisfied for example when

- (a) $q = 1$ and R is a \mathbb{Q} -algebra, or
- (b) q is not equal to zero, not a root of 1 and belongs to a subfield K of R .

The hypothesis in (2) is satisfied for differential equations and finite difference equations in positive characteristic as well as in the classical quantum case. More precisely, they are satisfied for example when

- (a) $q = 1$ and R is an \mathbb{F}_p -algebra, or
- (b) q is a nontrivial p -th root of 1 (p not necessarily prime) and belongs to a subfield K of R , or
- (c) q is a nontrivial p -th root of 1 with p prime (but q does not necessarily belong to a subfield of R).

- (2) Since both $D_{A/R,\sigma}$ and $D_{A/R,\sigma}^{(\infty)}$ commute with extensions of R (although $\bar{D}_{A/R,\sigma}$ does not), and $D_{A/R,\sigma}$ always commutes with extensions of A , we can sometimes (see the remark following Proposition 1.10) reduce questions to the generic case

$$(11) \quad R = \mathbb{Q}(t)[s], \quad A = R[x], \quad q = t \quad \text{and} \quad h = s,$$

and work as well over the latter. In this case, thanks to the theorem, we can identify $D_{A/R,\sigma}$ with $D_{A/R,\sigma}^{(\infty)}$.

- (3) The same proof shows that if A is a twisted R -algebra which is only q -flat (but not necessarily q -divisible), then
- (a) if q -char(A) = 0, then $D_{A/R,\sigma} = \bar{D}_{A/R,\sigma}$;
 - (b) if q -char(A) = $p > 0$, then $D_{A/R,\sigma}/\partial_\sigma^p \simeq \bar{D}_{A/R,\sigma}$.

The end of this section will be devoted to giving explicit formulas. They are usually quite formal to prove in the ring of twisted differential operators of infinite level and their analog in the twisted Weyl algebra is then easily obtained thanks to Theorem 6.3.

Proposition 6.4. *In $D_{A/R,\sigma}^{(\infty)}$, for all $k > 0$,*

$$\partial_\sigma^{[k]} \circ x = q^k x \partial_\sigma^{[k]} + (k)_q h \partial_\sigma^{[k]} + \partial_\sigma^{[k-1]}.$$

In $D_{A/R,\sigma}$, for all $k > 0$,

$$\partial_\sigma^k \circ x = q^k x \partial_\sigma^k + (k)_q (h \partial_\sigma^k + \partial_\sigma^{k-1}).$$

Proof. The first assertion is simply a reformulation of Proposition 4.7. For the second one, after a base change, we may reduce to the generic case (11) and thus assume that R is q -divisible and q -char(R) = 0. And we may then replace D_σ with $D_\sigma^{(\infty)}$ in which case we fall back onto the first equality. \square

Note that in order to prove the second formula, we cannot use Corollary 6.2 directly: our equality takes place in the Weyl algebra and is not just an assertion about endomorphisms of A .

We concentrate now on the case $\sigma(x) = qx$.

Proposition 6.5. *Assume $h = 0$, $q \in R^\times$ and $x \in A^\times$. Then in $D_{A/R,\sigma}^{(\infty)}$, for all $k \in \mathbb{N}$,*

$$\partial_\sigma^{[k]} \circ x^{-1} = \sum_{i=0}^k (-1)^i q^{-\frac{(2k-i)(i+1)}{2}} x^{-i-1} \partial_\sigma^{[k-i]}.$$

In $D_{A/R,\sigma}$, for all $k \in \mathbb{N}$,

$$\partial_\sigma^k \circ x^{-1} = \sum_{i=0}^k (-1)^i q^{-\frac{(2k-i)(i+1)}{2}} (k)_q \cdots (k-i+1)_q x^{-i-1} \partial_\sigma^{k-i}.$$

Proof. The first assertion is a particular case of Corollary 4.8 and the second one follows by the standard generic argument. \square

Proposition 6.6. *Assume that $h = 0$. Then in $D_{A/R, \sigma}^{(\infty)}$, for all $k, n \in \mathbb{N}$,*

$$\partial_{\sigma}^{[k]} \circ x^n = \sum_{i=0}^k q^{(n-i)(k-i)} \binom{n}{i}_q x^{n-i} \partial_{\sigma}^{[k-i]}.$$

In $D_{A/R, \sigma}$, for all $k, n \in \mathbb{N}$,

$$\partial_{\sigma}^k \circ x^n = \sum_{i=0}^k q^{(n-i)(k-i)} (i)_q! \binom{k}{i}_q \binom{n}{i}_q x^{n-i} \partial_{\sigma}^{k-i}.$$

Proof. As usual, the second formula will follow from the first, which we prove directly by induction on n . We have

$$\begin{aligned} \partial_{\sigma}^{[k]} \circ x^n &= \sum_{i=0}^k q^{(n-1-i)(k-i)} \binom{n-1}{i}_q x^{n-1-i} \partial_{\sigma}^{[k-i]} \circ x \\ &= \sum_{i=0}^k q^{(n-1-i)(k-i)} \binom{n-1}{i}_q x^{n-1-i} (q^{k-i} x \partial_{\sigma}^{[k-i]} + \partial_{\sigma}^{[k-i-1]}) \\ &= \sum_{i=0}^k q^{(n-1-i)(k-i)} \binom{n-1}{i}_q x^{n-1-i} q^{k-i} x \partial_{\sigma}^{[k-i]} \\ &\quad + \sum_{i=0}^k q^{(n-1-i)(k-i)} \binom{n-1}{i}_q x^{n-1-i} \partial_{\sigma}^{[k-i-1]} \\ &= \sum_{i=0}^k q^{(n-i)(k-i)} \binom{n-1}{i}_q x^{n-i} \partial_{\sigma}^{[k-i]} \\ &\quad + \sum_{i=0}^{k-1} q^{(n-i)(k-i+1)} \binom{n-1}{i-1}_q x^{n-i} \partial_{\sigma}^{[k-i]} \\ &= \sum_{i=0}^k q^{(n-i)(k-i)} \left(\binom{n-1}{i}_q + q^{n-i} \binom{n-1}{i-1}_q \right) x^{n-i} \partial_{\sigma}^{[k-i]} \\ &= \sum_{i=0}^k q^{(n-i)(k-i)} \binom{n}{i}_q x^{n-i} \partial_{\sigma}^{[k-i]}. \end{aligned}$$

\square

Corollary 6.7. *When $h = 0$, we have for all $n \in \mathbb{N}$,*

$$\partial_{\sigma}^{[k]}(x^n) = \begin{cases} \binom{n}{k}_q x^{n-k} & \text{if } n \geq k, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\partial_\sigma^k(x^n) = \begin{cases} (n)_q \cdots (n-k+1)_q x^{n-k} & \text{if } n \geq k, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The first assertion follows from the proposition and the second one uses the inclusion $\bar{D}_{A/R,\sigma} \subset D_{A/R,\sigma}^{(\infty)}$. \square

Corollary 6.8. *Assume that $h = 0$. Then we have for all $n \in \mathbb{N} \setminus \{0\}$,*

$$\partial_\sigma(x^n) = (n)_q x^{n-1}.$$

If $x \in A^\times$ and $q \in R^\times$, then the formula actually holds for all $n \in \mathbb{Z}$.

Proof. Only the second part needs a proof: if x is invertible, we have for $n > 0$

$$\begin{aligned} 0 &= \partial_\sigma(1) = \partial_\sigma(x^n x^{-n}) \\ &= \partial_\sigma(x^n) x^{-n} + \sigma(x^n) \partial_\sigma(x^{-n}) \\ &= (n)_q x^{n-1} x^{-n} + q^n x^n \partial_\sigma(x^{-n}) = (n)_q x^{-1} + q^n x^n \partial_\sigma(x^{-n}), \end{aligned}$$

from which we derive, when $q \in R^\times$,

$$\partial_\sigma(x^{-n}) = -q^{-n} (n)_q x^{-n-1} = (-n)_q x^{-n-1}. \quad \square$$

7. Formal deformations of twisted differential operators

In this section, we study the relation between twisted differential operators relative to various endomorphisms of A . We are particularly interested in the comparison of our twisted differential operators with usual differential operators. We assume that there exists a twisted coordinate x , that we fix for the rest of the section, and we write $y_\sigma := x - \sigma(x)$.

Recall that the ring $D_\sigma^{(\infty)}$ of differential operators (of infinite level) comes with a natural increasing filtration by A -submodules

$$D_\sigma^{(\infty)} = \bigcup_{m=0}^{\infty} \text{Diff}_{m,\sigma}(A),$$

which is called the *order filtration*. The choice of the twisted coordinate determines a splitting of this filtration. To see this, we let for all $m \in \mathbb{N}$, $K_\sigma^{[m]} \subset D_\sigma^{(\infty)}$ be the free A -module generated by all $\partial_\sigma^{[k]}$ for $k \geq m$. Note that this is actually a filtration by *left* ideals.

Definition 7.1. The decreasing filtration by the $K_\sigma^{[m]}$ is called the *ideal filtration* on $D_\sigma^{(\infty)}$. The *module of twisted differential operators of infinite level and infinite order* on A is

$$\widehat{D}_{A/R,\sigma}^{(\infty)} = \varprojlim D_\sigma^{(\infty)} / K_\sigma^{[m+1]}.$$

We might again drop the index A/R and write $\widehat{D}_\sigma^{(\infty)}$. The decreasing filtration of $\widehat{D}_\sigma^{(\infty)}$ by the closures $\widehat{K}_\sigma^{[m]}$ of the $K_\sigma^{[m]}$ will also be called the *ideal filtration*.

Remarks. (1) The ideal filtration is *separated*, which means that $\cap_m K_\sigma^{[m+1]} = \{0\}$ and it follows that $D_\sigma^{(\infty)} \subset \widehat{D}_\sigma^{(\infty)}$. Actually, any $\varphi \in \widehat{D}_\sigma^{(\infty)}$ can be uniquely written as an infinite sum $\sum_0^\infty z_k \partial_\sigma^{[k]}$ with $z_k \in A$ (and conversely). In other words, we have the isomorphisms of A -modules

$$D_\sigma^{(\infty)} = \bigoplus_{k \in \mathbb{N}} A \partial_\sigma^{[k]} \quad \text{and} \quad \widehat{D}_\sigma^{(\infty)} = \prod_{k \in \mathbb{N}} A \partial_\sigma^{[k]}.$$

(2) The ideal filtration defines a splitting of the order filtration in the sense that

$$D_\sigma^{(\infty)} = \text{Diff}_{m,\sigma}(A) \oplus K_\sigma^{[m+1]} \quad \text{and} \quad \widehat{D}_\sigma^{(\infty)} = \text{Diff}_{m,\sigma}(A) \oplus \widehat{K}_\sigma^{[m+1]}$$

as A -modules.

(3) $\widehat{D}_\sigma^{(\infty)}$ is *not* a ring in general. More precisely, multiplication on $D_\sigma^{(\infty)}$ is not continuous for the ideal filtration: we always have $\partial_\sigma^{[m]} \rightarrow 0$ when $m \rightarrow \infty$ but, if $\sigma(x) = qx$ and $x \in A^\times$ for example, we can see that for all $m \in \mathbb{N}$,

$$\partial_\sigma^{[m]} \circ x^{-1} \equiv \partial_\sigma^{[m]}(x^{-1}) = x^{-m-1} \neq 0 \pmod{K_\sigma^{[1]}}.$$

If A is any ring and $A[\xi]$ denotes the polynomial ring in one variable ξ , then the *natural filtration* on $\text{Hom}_A(A[\xi], A)$ is the decreasing filtration by the kernels of the surjections

$$\text{Hom}_A(A[\xi], A) \rightarrow \text{Hom}_A(A[\xi]_{\leq m}, A),$$

where $A[\xi]_{\leq m}$ denotes as before the A -module of polynomials of degree at most m . The corresponding topology will be called the *natural topology* of $\text{Hom}_A(A[\xi], A)$. Note that $\text{Hom}_A(A[\xi], A)$ is separated and complete for the natural topology:

$$\varprojlim \text{Hom}_A(A[\xi]_{\leq m}, A) = \text{Hom}_A(\varinjlim A[\xi]_{\leq m}, A) = \text{Hom}_A(A[\xi], A).$$

Lemma 7.2 (formal density). *The map*

$$A[\xi] \rightarrow P_{A/R}, \quad \xi \mapsto \tilde{x} - x$$

induces by duality an isomorphism of topological A -modules

$$(12) \quad \widehat{D}_{A/R,\sigma}^{(\infty)} \xrightarrow{\sim} \text{Hom}_A(A[\xi], A).$$

More precisely, the ideal filtration corresponds to the natural filtration.

Proof. It follows from Corollary 1.13 that the cokernel of the map

$$A[\xi]_{\leq m} \hookrightarrow A[\xi] \rightarrow P \twoheadrightarrow P_{(n)\sigma}$$

is generated by $\xi^{(k)}$ for $m < k \leq n$. Moreover, this map is injective when $m \leq n$. Dually, it means that the corresponding map

$$\text{Diff}_{n,\sigma}(A) \simeq \text{Hom}_A(P_{(n)\sigma}, A) \rightarrow \text{Hom}_A(A[\xi]_{\leq m}, A)$$

is surjective for $m \leq n$ and that its kernel is exactly $K_\sigma^{[m+1]} \cap \text{Diff}_{n,\sigma}(A)$ (it is generated by $\partial_\sigma^{[k]}$ for $n \geq k > m$). Taking direct limits on the left, we see that the map

$$D_\sigma^{(\infty)} \rightarrow \text{Hom}_A(A[\xi]_{\leq m}, A)$$

is surjective with kernel exactly $K_\sigma^{[m+1]}$. In other words, we get a canonical isomorphism of A -modules

$$D_\sigma^{(\infty)} / K_\sigma^{[m+1]} \xrightarrow{\sim} \text{Hom}_A(A[\xi]_{\leq m}, A).$$

Thus, taking inverse limits on both sides gives the result. \square

Remarks. (1) By construction, there exists a commutative diagram

$$\begin{array}{ccc} D_{A/R,\sigma}^{(\infty)} & \hookrightarrow & \text{End}_R(A) \\ \downarrow & & \downarrow \simeq \\ & & \text{Hom}_A(P, A) \\ \downarrow & & \downarrow \\ \widehat{D}_{A/R,\sigma}^{(\infty)} & \xrightarrow{\simeq} & \text{Hom}_A(A[\xi], A) \end{array}$$

- (2) We will usually denote by $\tilde{\varphi} \in \text{Hom}_A(A[\xi], A)$ the image of $\varphi \in \widehat{D}_{A/R,\sigma}^{(\infty)}$. This is compatible with the previous notation for linearization. In particular, we have for all $k, n \in \mathbb{N}$,

$$\widetilde{\partial_\sigma^{[k]}(\xi^{(n)})} = \begin{cases} 1 & \text{if } k = n, \\ 0 & \text{otherwise.} \end{cases}$$

- (3) When $A = R[x]$, we actually get

$$D_{A/R,\sigma}^{(\infty)} \hookrightarrow \widehat{D}_{A/R,\sigma}^{(\infty)} \xrightarrow{\sim} \text{Hom}_A(A[\xi], A) \xrightarrow{\sim} \text{Hom}_A(P, A) \xrightarrow{\sim} \text{End}_R(A),$$

and $\widehat{D}_{A/R,\sigma}^{(\infty)}$ is a ring in this very special case.

- (4) There exists a complex analytic analog of the density lemma (for usual differential operators) as explained by Zoghman Mebkhout and Luis Narváez [1998]: the ring of algebraic differential operators is dense in the ring of continuous endomorphisms of the structural sheaf.

Although it is difficult to give an explicit description of the isomorphism (12), we can at least show the following:

Lemma 7.3. *We have for all $n \in \mathbb{N} \setminus \{0\}$,*

$$\tilde{\partial}_\sigma(\xi^n) = (\sigma(x) - x)^{n-1}.$$

More generally, if τ is another R -algebra endomorphism of A , we have for all $n \in \mathbb{N} \setminus \{0\}$,

$$\tilde{\partial}_\sigma(\xi^{(n)\tau}) = \prod_{i=1}^{n-1} (\sigma(x) - \tau^i(x)),$$

where $\xi^{(n)\tau} := \xi \tau(\xi) \cdots \tau^{n-1}(\xi)$ denotes the twisted power with respect to the endomorphism τ .

Proof. By definition, $\tilde{\partial}_\sigma$ is the unique A -linear function on $A[\xi]$ such that $\tilde{\partial}_\sigma(\xi^{(n)}) = 1$ for $n = 1$ and 0 otherwise. We consider now the unique A -linear function u on $A[\xi]$ such that $u(\xi^n) = (\sigma(x) - x)^{n-1}$ for $n > 0$ and $u(1) = 0$. We want to show that $u = \tilde{\partial}_\sigma$. The map u may be seen as the composition of division by ξ on $A[\xi]$ (after removing the constant term) and evaluation at $\sigma(x) - x$. But we have

$$\xi^{(n)} := \xi \sigma(\xi) \cdots \sigma^{n-1}(\xi),$$

and we know that $\sigma(\xi) = \xi + x - \sigma(x)$. Therefore, it is clear that for $n \geq 2$, we will get $u(\xi^{(n)}) = 0$. Of course, for $n = 0$ we have $\xi^{(n)} = 1$ and we also obtain 0. Finally, for $n = 1$, we have $\xi^{(n)} = \xi$ and we get 1.

The proof of the second formula follows the same lines. From the first part, we can interpret $\tilde{\partial}_\sigma$ as the composition of division by ξ on $A[\xi]$ and evaluation at $\sigma(x) - x$. We want to apply this to $\xi^{(n)\tau} = \prod_{i=0}^{n-1} \tau^i(\xi)$ and we know that $\tau^i(\xi) = \xi + x - \tau^i(x)$. Thus, we see that

$$\tilde{\partial}_\sigma(\xi^{(n)\tau}) = \prod_{i=1}^{n-1} (\sigma(x) - x + x - \tau^i(x)) = \prod_{i=1}^{n-1} (\sigma(x) - \tau^i(x)). \quad \square$$

We can actually derive a remarkable consequence of the density lemma (recall that when τ is an endomorphism of the R -algebra A , we call any twisted coordinate relative to τ a τ -coordinate):

Proposition 7.4 (formal deformation). *If x is also a τ -coordinate for another R -endomorphism τ of A , then there exists a canonical isomorphism of topological A -modules (that depends on x)*

$$\widehat{\mathbf{D}}_{A/R, \sigma}^{(\infty)} \xrightarrow{\sim} \widehat{\mathbf{D}}_{A/R, \tau}^{(\infty)}.$$

More precisely, it is compatible with the ideal filtrations.

Recall that the hypothesis is always satisfied when A is a τ -twisted localization of $R[x]$.

Proof. We may just compose the isomorphism of the formal density Lemma 7.2 with the inverse of the analogous isomorphism for τ . \square

When the coordinate x is fixed, we may (and will) *identify* these two topological (or even, filtered) A -modules. It is worth mentioning a particular case of this corollary that is of great interest (this is the case $\tau = \text{Id}_A$, in which case we say *usual coordinate* instead of τ -coordinate):

Corollary 7.5. *Assume that x is not only a σ -coordinate, but also a usual coordinate on A . Then there exists a canonical isomorphism*

$$\widehat{D}_{A/R, \sigma}^{(\infty)} \xrightarrow{\sim} \widehat{D}_{A/R}^{(\infty)}.$$

Note that the hypothesis is always satisfied when A is a localization of $R[x]$. More generally, when A is smooth over R , a usual coordinate is the same thing as an étale coordinate.

Remark. (1) In theorem 2 of the introduction to [Pulita 2017], Pulita shows that, in the p -adic world, some differential modules have the natural structure of a σ -module. The main idea is to realize a formal solution of the differential module as formal solution of some σ -module. This is analogous to the way we derive the formal deformation theorem from the formal density lemma. It seems that the first result in this direction was obtained by André and Di Vizio [2004].

(2) In [Gros and Le Stum 2014], Michel Gros and Bernard Le Stum were able to prove a quantum Simpson's correspondence. Assume $R = \mathbb{Z}[q]$, where q is a p -th root of unity with p prime, A is the polynomial ring and $\sigma(x) = qx$. Then the category of A -modules endowed with a quasiniipotent q -derivation is equivalent to the category of A -modules endowed with an A -linear quasiniipotent endomorphism. This is derived from an isomorphism quite analogous to (12) which reads

$$\widehat{D}_q \simeq \text{End}_{\widehat{\mathbb{Z}}_q}(\widehat{\mathbb{Z}}_q \widehat{A}).$$

(3) We want to insist on the fact that the isomorphism of Corollary 7.5 is not an isomorphism of rings. Actually, neither side of this isomorphism is a ring. In order to achieve this goal, it would be necessary to refine the topology. After the remarks following Corollary 7.7, we will give some elementary but nontrivial examples over the complex numbers, and refer the reader to forthcoming articles for a detailed study of the ultrametric and the q -characteristic zero cases.

We give now some explicit formulas in order to express the twisted derivation from one world as a twisted differential operator in another world:

Proposition 7.6. *Assume that τ is some other R -endomorphism of A and x is also a τ -coordinate. Then we have*

$$\partial_\sigma = \sum_{k=1}^{\infty} \left(\prod_{i=1}^{k-1} (\sigma(x) - \tau^i(x)) \right) \partial_\tau^{[k]}.$$

Proof. This follows directly by duality from Lemma 7.3. □

Corollary 7.7. *Assume that x is also a usual coordinate on A . Then*

$$\partial_\sigma = \sum_{k=1}^{\infty} (\sigma(x) - x)^{k-1} \partial^{[k]}$$

and

$$\sigma = \sum_{k=0}^{\infty} (\sigma(x) - x)^k \partial^{[k]}.$$

Proof. The first assertion follows directly from the proposition and the other one is deduced from the equality $\sigma = 1 - y_\sigma \partial_\sigma$. □

Remarks. (1) In the case $\sigma(x) = qx$, the formulas read

$$(13) \quad \partial_\sigma = \sum_{k=1}^{\infty} (q-1)^{k-1} x^{k-1} \partial^{[k]} \quad \text{and} \quad \sigma = \sum_{k=0}^{\infty} (q-1)^k x^k \partial^{[k]}.$$

(2) In the case $\sigma(x) = x + h$, the formulas read

$$(14) \quad \partial_\sigma = \sum_{k=1}^{\infty} h^{k-1} \partial^{[k]} \quad \text{and} \quad \sigma = \sum_{k=0}^{\infty} h^k \partial^{[k]}.$$

Examples. (1) Consider the differential equation

$$\frac{ds}{dz} = s$$

over the complex numbers. If we use the second formula of (13), we see that the corresponding q -difference equation is given by

$$s(qz) = e^{(q-1)z} s(z).$$

If we use (14) instead, we obtain the difference equation

$$s(z+h) = e^h s(z).$$

(2) Consider now the differential equation

$$\frac{ds}{dz} = \frac{a}{z} s$$

for some $a \in \mathbb{C}$. Then the q -difference equation will be given by

$$s(qz) = q^a s(z),$$

and the difference equation is

$$s(z+h) = \left(\frac{z+h}{z} \right)^a s(z).$$

We may also go the other way and express the usual derivation in term of twisted derivations. We only do the classical cases:

Corollary 7.8. *Assume that x is also a usual coordinate on A . Then*

(1) *If $\sigma(x) = qx$, we have*

$$\partial = \sum_{k=1}^{\infty} (1-q)^{k-1} (k-1)_q! x^{k-1} \partial_{\sigma}^{[k]}.$$

(2) *If $\sigma(x) = x + h$, we have*

$$\partial = \sum_{k=1}^{\infty} (-1)^{k-1} h^{k-1} (k-1)! \partial_{\sigma}^{[k]}.$$

Proof. In the formula of Proposition 7.6 we replace τ by σ and σ by Id respectively, in order to obtain

$$\begin{aligned} \partial &= \sum_{k=1}^{\infty} \prod_{i=1}^{k-1} (x - q^i x) \partial_{\sigma}^{[k]} = \sum_{k=1}^{\infty} \prod_{i=1}^{k-1} (1 - q^i) x^{k-1} \partial_{\sigma}^{[k]} \\ &= \sum_{k=1}^{\infty} (1-q)^{k-1} \prod_{i=1}^{k-1} (i)_q x^{k-1} \partial_{\sigma}^{[k]} = \sum_{k=1}^{\infty} (1-q)^{k-1} (k-1)_q! x^{k-1} \partial_{\sigma}^{[k]} \end{aligned}$$

in the first case, and

$$\partial = \sum_{k=1}^{\infty} \prod_{i=1}^{k-1} (-ih) \partial_{\sigma}^{[k]} = \sum_{k=1}^{\infty} (-h)^{k-1} (k-1)! \partial_{\sigma}^{[k]}$$

in the other one. \square

We may also apply Proposition 7.6 in order to make more precise the statement of Proposition 4.13.

Corollary 7.9. *Assume that x is also a σ^m -coordinate for A over R , then we have*

$$\partial_{\sigma^m} = \sum_{k=1}^m \left(\prod_{i=1}^{k-1} (\sigma^m(x) - \sigma^i(x)) \right) \partial_{\sigma}^{[k]}.$$

One can give a more concrete formula in the quantum situation:

Corollary 7.10. *Assume that x is also a σ^m -coordinate for A over R and that $\sigma(x) = qx$. Then we have*

$$\partial_{\sigma^m} = \sum_{k=1}^m q^{\frac{k(k-1)}{2}} (q-1)^{k-1} (m-1)_q \cdots (m-k+1)_q x^{k-1} \partial_{\sigma}^{[k]}.$$

Proof. We compute the coefficient of $\partial_\sigma^{[k]}$ for $1 \leq k \leq m$:

$$\begin{aligned} \prod_{i=1}^{k-1} (q^m x - q^i x) &= \left(\prod_{i=1}^{k-1} q^i \right) \left(\prod_{i=1}^{k-1} (q^{m-i} - 1) \right) x^{k-1} \\ &= q^{\frac{k(k-1)}{2}} (q-1)^{k-1} \left(\prod_{i=1}^{k-1} (m-i)_q \right) x^{k-1}. \quad \square \end{aligned}$$

Example. We have $\partial_{\sigma^2} = \partial_\sigma + q(q-1)x\partial_\sigma^{[2]}$.

8. Formal confluence

We explain here how one can use the quantum Weyl algebra in order to approximate a usual differential operator. We assume that A is a quantum R -algebra which means that we are given a twisted coordinate x on A with $\sigma(x) = qx + h$ with $q, h \in R$.

Recall from [Le Stum and Quirós 2015b] that the twisted Weyl R -algebra D_σ has an increasing filtration by free A -modules of finite rank

$$\text{Fil}^m D_\sigma = \bigoplus_{k \leq m} A \partial_\sigma^k$$

that is called the *order filtration*. But it also has a decreasing filtration by free A -modules (of infinite rank)

$$K_\sigma^m = \bigoplus_{k \geq m} A \partial_\sigma^k$$

that we called the *ideal filtration*. We will consider the completion

$$\widehat{D}_\sigma = \varprojlim D_\sigma / K_\sigma^{m+1}$$

that also comes with its *ideal filtration* and we have $D_\sigma \subset \widehat{D}_\sigma$. Note that \widehat{D}_σ is the set of all infinite sums $\sum_0^\infty z_k \partial_\sigma^k$ with $z_k \in A$. In other words, there exists an isomorphism of A -modules

$$\widehat{D}_\sigma = \prod_{k \in \mathbb{N}} A \partial_\sigma^k.$$

The A -module \widehat{D}_σ is not a ring in general. However, the result holds in the finite quantum characteristic case as we can check right now:

Proposition 8.1. *If A is q -flat and $q\text{-char}(A) = p > 0$, then multiplication is continuous for the ideal topology on $D_{A/R, \sigma}$ and turns $\widehat{D}_{A/R, \sigma}$ into an R -algebra.*

Proof. From the equality (see the remark following Theorem 6.3)

$$D_\sigma / \partial_\sigma^p \simeq \bar{D}_\sigma,$$

and the fact that \bar{D}_σ is a ring, we deduce that the (left) ideal generated by ∂_σ^p is a two-sided ideal. Since multiplication is automatically continuous for the ∂_σ^p -adic filtration, it follows that it is also continuous for the ideal filtration (which is the filtration by left ideals generated by the powers of ∂_σ). \square

Lemma 8.2. *The composite map*

$$D_{A/R,\sigma} \twoheadrightarrow \bar{D}_{A/R,\sigma} \hookrightarrow D_{A/R,\sigma}^{(\infty)}$$

is compatible with the ideal filtrations. Moreover, if R is q -divisible and if $q\text{-char}(R) = 0$, then $\widehat{D}_{A/R,\sigma} \simeq \widehat{D}_{A/R,\sigma}^{(\infty)}$.

This applies in particular when R is a \mathbb{Q} -algebra and $\sigma = \text{Id}$.

Proof. The first assertion follows from Corollary 6.2 and the second follows from Theorem 6.3. \square

We can now state our first confluence theorem (recall that a usual coordinate is the same thing as an étale coordinate when A is smooth over R):

Theorem 8.3 (formal quantum confluence 1). *Let R be a \mathbb{Q} -algebra, A a q - R -algebra for some $q \in R$. Assume that the quantum coordinate on A is also a usual coordinate. Then there exists a canonical map of filtered A -modules*

$$(15) \quad D_{A/R,\sigma} \rightarrow \widehat{D}_{A/R}.$$

Moreover, if R is q -divisible, we have

- (1) *If $q\text{-char}(A) = 0$, then the map (15) is injective and $D_{A/R,\sigma}$ is dense in $\widehat{D}_{A/R}$. Actually, the inclusion morphism $D_{A/R,\sigma} \subset \widehat{D}_{A/R}$ is strict for the ideal filtrations in the sense that $D_{A/R,\sigma} \cap \widehat{K}^m = K_\sigma^m$ for all $m \in \mathbb{N}$.*
- (2) *If $q\text{-char}(A) = p > 0$, then the map (15) induces an isomorphism*

$$(D_{A/R,\sigma}/\partial_\sigma^p \simeq) \quad \bar{D}_{A/R,\sigma} \simeq D_{A/R}/\partial^p.$$

Proof. The map (15) is simply the composite

$$D_\sigma \rightarrow D_\sigma^{(\infty)} \hookrightarrow \widehat{D}_\sigma^{(\infty)} \simeq \widehat{D}^{(\infty)} \simeq \widehat{D}_{A/R},$$

where the next to the last map is the formal deformation isomorphism of Corollary 7.5 and the last one comes from Lemma 8.2 applied to the case $\sigma = \text{Id}$ since R is a \mathbb{Q} -algebra.

If we assume that R is q -divisible and that $q\text{-char}(A) = 0$, then Lemma 8.2 tells us that $D_\sigma = D_\sigma^{(\infty)}$ as filtered rings (for the ideal filtrations) and we can use Corollary 7.5 again.

Finally, when R is q -divisible but $q\text{-char}(A) = p > 0$, then $(p)_q = 0$ in A and all $(m)_q \in R^\times$ for $m < p$. We can use the last assertion of Theorem 6.3 and the fact that the isomorphism of Corollary 7.5 is strictly compatible with the filtrations. \square

As in [Le Stum and Quirós 2015b], we denote by $A[T]_\sigma$ the twisted polynomial ring associated to σ : this is the noncommutative polynomial ring in T over A with the commutation rule $Tx = \sigma(x)T$. Recall also that the twisted coordinate x is said to be *strong* if $x - \sigma(x) \in A^\times$.

Corollary 8.4. *If R is a q -divisible \mathbb{Q} -algebra, $q\text{-char}(A) = 0$, x is strong and is also a σ^n -coordinate for all $n \in \mathbb{N}$, then the A -linear map*

$$(16) \quad A[T]_\sigma \rightarrow \widehat{D}_{A/R}, \quad T^n \mapsto \sum_{k=0}^{\infty} \frac{1}{k!} (\sigma^n(x) - x)^k \partial^k$$

has dense image.

Proof. Compose the isomorphism of theorem 6.13 of [Le Stum and Quirós 2015b] with the map of Theorem 8.3. The formula comes from Corollary 7.7. \square

Example. The theorem (and its corollary) applies in particular when q is not a root of unity in some field K of characteristic zero, R is an algebra containing K , $A = R[x, x^{-1}]$ and $\sigma(x) = qx$. In this situation, $D := D_{A/R}$ is the noncommutative ring $R[x, x^{-1}, \partial]$ with the commutation rule $\partial x = x\partial + 1$ and we may see \widehat{D} (which is *not* a ring) as the set of infinite sums $R[x, x^{-1}][[\partial]]$. Also, $D_\sigma := D_{A/R, \sigma}$ is the noncommutative ring $R[x, x^{-1}, \partial_\sigma]$ with the commutation rule $\partial_\sigma x = qx\partial_\sigma + 1$. The map of the theorem satisfies

$$\partial_\sigma \mapsto \sum_{k=1}^{\infty} \frac{1}{k!} (q-1)^{k-1} x^{k-1} \partial^k,$$

and the map of the corollary is given by

$$T^n \mapsto \sum_{k=0}^{\infty} \frac{1}{k!} (q^n - 1)^k x^k \partial^k.$$

In the next section, we will have to move between σ and powers of σ (or more precisely, the other way around, between σ and roots of σ).

Proposition 8.5. *Let $m \in \mathbb{N} \setminus \{0\}$. If $(m)_q! \in R^\times$, then there exists a unique A -linear ring homomorphism*

$$D_{A/R, \sigma^m} \xrightarrow{\iota_{\sigma, m}} D_{A/R, \sigma}, \quad \partial_{\sigma^m} \mapsto \sum_{k=1}^m \frac{1}{(k)_q!} \left(\prod_{i=1}^{k-1} (\sigma^m(x) - \sigma^i(x)) \right) \partial_\sigma^k$$

Moreover, the diagram

$$(17) \quad \begin{array}{ccccc} D_{A/R, \sigma^m} & \longrightarrow & D_{A/R, \sigma^m}^{(\infty)} & \hookrightarrow & \text{End}_R(A) \\ \downarrow \iota_{\sigma, m} & & \downarrow & & \parallel \\ D_{A/R, \sigma} & \longrightarrow & D_{A/R, \sigma}^{(\infty)} & \hookrightarrow & \text{End}_R(A) \end{array}$$

is commutative.

Proof. We may assume (see the second remark following Theorem 6.3) that we are in the generic situation

$$R = \mathbb{Q}(t)[s], \quad A = R[x], \quad q = t \quad \text{and} \quad h = s.$$

But then the left horizontal arrows in diagram (17) are bijective. In other words, we can identify the twisted Weyl R -algebras with the rings of twisted differential operators of infinite level and use Corollary 7.9 and Corollary 6.2. \square

Remarks. (1) In the case $h = 0$, we have thanks to Corollary 7.10 the more concrete formula:

$$\partial_{\sigma^m} \mapsto \sum_{k=1}^m q^{\frac{k(k-1)}{2}} \frac{(q-1)^{k-1}}{(k)_q} \binom{m-1}{k-1}_q x^{k-1} \partial_{\sigma}^k.$$

- (2) Assume that R is q -divisible. If q -char(R) = $p > 0$, then the hypothesis is satisfied if and only if $m < p$. Of course if q -char(R) = 0, then the hypothesis is satisfied for any $m \in \mathbb{N}$.
- (3) Just to give an idea of the geometric intuition, the geometric counterpart of the twisted localization hypothesis should be the requirement to work on a Zariski open subset of “the quantum line”.

Proposition 8.6. *Let $m, n \in \mathbb{N} \setminus \{0\}$. If $(mn)_q!$ is invertible, then*

$$\iota_{\sigma, mn} = \iota_{\sigma^m, n} \circ \iota_{\sigma, m}.$$

Proof. After a base change, we may assume that all q -integers are invertible and identify twisted Weyl R -algebras with rings of twisted differential operators of infinite level. The assertion then becomes a triviality. \square

Proposition 8.7. *If $(m)_q!$ is invertible in R , then the following diagram commutes:*

$$\begin{array}{ccc} T & A[T]_{\sigma^m} \longrightarrow D_{A/R, \sigma^m}, & T \mapsto 1 + (\sigma^m(x) - x) \partial_{\sigma^m} \\ \downarrow & \downarrow & \downarrow \iota_{\sigma, m} \\ T^m & A[T]_{\sigma} \longrightarrow D_{A/R, \sigma}, & T \mapsto 1 + (\sigma(x) - x) \partial_{\sigma} \end{array}$$

Proof. Again, we may assume that all q -integers are invertible in R and use the second remark following Theorem 6.13 of [Le Stum and Quirós 2015b]. \square

9. Formal confluence in positive quantum characteristic

In this section, we extend the formal confluence theorem to the case of positive quantum characteristic. In order to do that, it is necessary to use the S -twisted theory, where S is a filtering (for division) set of positive integers. Thus, we assume here that A is an S -twisted R -algebra in the sense of [Le Stum and Quirós 2015b] (it is endowed with a compatible system of n -th roots σ_n of σ for all $n \in S$).

We recall from [Le Stum and Quirós 2015a] that a *system of roots* of $q \in R$ is a compatible family $\underline{q} := \{q_n\}_{n \in S}$ of n -th roots of q . We call the system \underline{q} *admissible* if for all $n \in S$,

$$(n)_{q^n} \in R^\times.$$

This is a natural condition in order to define the q -rational number $(r)_q$ for any $r \in \mathbb{N}(1/S)$. More precisely, if $r = m/n$ with $m \in \mathbb{N}$ and $n \in S$, then

$$(r)_q := \frac{(m)_{q_n}}{(n)_{q_n}}$$

only depends on r and not on the choice of m and n . It is convenient to introduce the following terminology:

Definition 9.1. Let $\{q_n\}_{n \in S}$ be a system of roots of q in R . Then R is said to be \underline{q} -divisible if R is q_n -divisible for all $n \in S$.

When the system is admissible, there exists a nice equivalent definition:

Lemma 9.2. Let $\{q_n\}_{n \in S}$ be an admissible system of roots of q . Then R is \underline{q} -divisible if and only if for all $r \in \mathbb{N}(1/S)$,

$$(r)_q \in R^\times \cup \{0\}.$$

Proof. If $r = m/n$ with $m \in \mathbb{N}$ and $n \in S$, we have $(r)_q = (m)_{q_n}/(n)_{q_n}$. It follows that $(r)_q \in R^\times$ (resp. $= 0$) if and only if $(m)_{q_n} \in R^\times$ (resp. $= 0$). \square

Remarks. (1) If $1 - q \in R^\times$, then \underline{q} is admissible. In particular, if $q \in K$ where K is a subfield of R and $q \neq 0$, then \underline{q} is admissible.

(2) If $\underline{q} \subset K$ where K is a subfield of R , then R is \underline{q} -divisible.

Lemma 9.3. For a system $\underline{q} := \{q_n\}_{n \in S}$ of roots of q , the following are equivalent:

- (1) \underline{q} is admissible and for all $r \in \mathbb{N}(1/S) \cap [0, 1]$, we have $(r)_q \in R^\times$.
- (2) For all $n \in S$, we have $(n)_{q_n}! \in R^\times$.

Proof. First of all, the second condition also implies that \underline{q} is admissible and we may therefore make this assumption. If $r = m/n$ with $m \in \mathbb{N}$ and $n \in S$, we know that $(r)_q \in R^\times$ if and only if $(m)_{q_n} \in R^\times$. Also, we have $r \leq 1$ if and only if $m \leq n$. Thus, the second condition, which means that $(m)_{q_n} \in R^\times$ whenever $m \leq n$, is equivalent to the requirement that $(r)_q \in R^\times$ for $r \leq 1$. \square

Definition 9.4. Such a system of roots will be called *strongly admissible*.

Lemma 9.5. *If \underline{q} is an admissible (resp. a strongly admissible) system of roots, and we write for all $n \in S$, $p_n := q_n$ -char(R), then we have for all $n \in S$,*

$$p_n \nmid n \text{ (resp. } p_n > n) \quad \text{or} \quad p_n = 0.$$

When R is \underline{q} -divisible, the converse is true.

Proof. By definition, q_n is admissible (resp. strongly admissible) if and only if $(n)_{q_n} \in R^\times$ (resp. $(n)_{q_n}! \in R^\times$). By definition also, we have $p_n \neq 0$ and $p_n \mid n$ (resp. $p_n \leq n$) if and only if $(n)_{q_n} = 0$ (resp. $(n)_{q_n}! = 0$). We see that these two conditions are mutually exclusive in general and that they are exhaustive when R is q_n -divisible. \square

There usually exist strongly admissible systems as the next result shows:

Proposition 9.6. *Assume that R is \underline{q} -divisible and that q -char(R) = $p > 0$. Then if \underline{q} is a system of p^n -th root of q , it is strongly admissible.*

Proof. It follows from Proposition 1.16 of [Le Stum and Quirós 2015a] that for all $n \in S$, we have q_{p^n} -char(R) = $p^{n+1} > p^n$ and we can apply Lemma 9.5. \square

Definition 9.7. An $x \in A$ is called an *S-twisted coordinate*, or *rooted twisted coordinate*, (resp. an *S-quantum coordinate*, or *rooted quantum coordinate*), if x is a twisted (resp. quantum) coordinate for all (A, σ_n) . We call it *strong* if $x - \sigma_n(x) \in A^\times$ for all $n \in S$.

Thus, by definition, x is an *S-quantum coordinate* if and only if for all $n \in S$,

$$(18) \quad \sigma_n(x) = q_n x + h_n \quad \text{with} \quad q_n, h_n \in R.$$

When this is the case, we might also say \underline{q} -coordinate and call A a *rooted quantum R-algebra*, an *S-quantum R-algebra* or a \underline{q} -*R-algebra*. We call it *strong* when x is strong.

Lemma 9.8. *Assume that x is simultaneously a twisted coordinate and a rooted quantum coordinate on A so that (18) holds. Then x is a quantum coordinate on A and if we write $\sigma(x) = qx + h$ with $q, h \in R$, then $\underline{q} := \{q_n\}_{n \in S}$ is a system of roots of q . If this system is admissible, we have for all $n \in \mathbb{N}$,*

$$h_n = \left(\frac{1}{n}\right)_q h.$$

Proof. If we are given $m \in \mathbb{N}$ and $n \in S$, we know that σ_n^m only depends on $r := m/n$ and so does $\sigma_n^m(x) = q_n^m x + (m)_{q_n} h_n$. It immediately follows that \underline{q} is a system of roots. Moreover, the case $m = n$ implies that $\sigma(x) = qx + h$ with $q = q_n^n$ and $h = (n)_{q^n} h_n$ from which we derive the other assertions. \square

Examples. (1) If we are given a system \underline{q} of roots in R , we can endow $A := R[x]$ or $A = R[x, x^{-1}]$ with, for all $n \in S$,

$$\sigma_n(x) = q_n x.$$

In the later case, x is a strong \underline{q} -coordinate if and only if $1 - q \in R^\times$ (and then \underline{q} is also admissible).

(2) More generally, if we are given an *admissible* system \underline{q} of roots of $q \in R$ and some $h \in R$, we can endow $R[x]$ with, for all $n \in S$,

$$\sigma_n(x) = q_n x + \left(\frac{1}{n}\right)_q h.$$

Definition 9.9. Assume that x is a \underline{q} -coordinate on A with \underline{q} strongly admissible. A *rooted twisted differential A -module* is an A -module M endowed with a family of ∂_{A, σ_n} -derivations $\partial_{M, n}$ for all $n \in S$ such that whenever $n, n' \in S$ with $n \mid n'$, we have for all $s \in M$,

$$\partial_{M, n}(s) = \sum_{k=1}^{n'/n} \frac{1}{(k)_{q_{n'}}!} \left(\prod_{i=1}^{k-1} (\sigma_n(x) - \sigma_{n'}^i(x)) \right) \partial_{M, n'}^k(s).$$

They form a category that we will denote by $\partial_{A, \underline{\sigma}}\text{-Mod}^{\text{root}}$.

Definition 9.10. If A is a \underline{q} - R -algebra with \underline{q} strongly admissible, then the *rooted twisted Weyl algebra* of A is

$$D_{A/R, \underline{\sigma}} := \varinjlim D_{A/R, \sigma_n},$$

with transition maps (from Proposition 8.5) for $n, n' \in S$, with $n \mid n'$ given by

$$D_{A/R, \sigma_n} \xrightarrow{\iota_{\sigma_{n'}, n'/n}} D_{A/R, \sigma_{n'}}, \quad \partial_{\sigma_n} \longmapsto \sum_{k=1}^{n'/n} \frac{1}{(k)_{q_{n'}}!} \left(\prod_{i=1}^{k-1} (\sigma_n(x) - \sigma_{n'}^i(x)) \right) \partial_{\sigma_{n'}}^k.$$

Proposition 9.11. Assume that A is a \underline{q} - R -algebra with \underline{q} strongly admissible. Then the rooted twisted differential A -modules form an abelian category with sufficiently many injective and projective objects. Actually, if M is a $D_{A/R, \underline{\sigma}}$ -module, then the maps

$$M \xrightarrow{\partial_{M, n}} M, \quad s \longmapsto \partial_{M, n} s$$

turn M into a rooted twisted differential A -module and we obtain an equivalence (an isomorphism) of categories

$$D_{A/R,\underline{\sigma}}\text{-Mod} \simeq \partial_{A,\underline{\sigma}}\text{-Mod}^{\text{root}}.$$

Proof. This is a consequence of proposition 5.6 of [Le Stum and Quirós 2015b]. \square

Recall from the same work that $A[T^{1/S}]_{\underline{\sigma}}$ denotes the noncommutative Puiseux polynomial ring with the commutation rule for all $m \in \mathbb{N}$, for all $n \in S$,

$$T^{m/n}x = \sigma_n^m(x)T^{m/n}.$$

One can also consider the notion of $\underline{\sigma}_A$ -module: this is an A -module endowed with a compatible family of $\sigma_{A,n}$ -linear endomorphisms $\sigma_{M,n}$. There exists an equivalence (an isomorphism) between the category of $A[T^{1/S}]_{\underline{\sigma}}$ -modules and the category of $\underline{\sigma}_A$ -modules.

Proposition 9.12. *Assume that x is a \underline{q} -coordinate on A with \underline{q} strongly admissible. Then there exists a unique A -linear homomorphism of rings*

$$(19) \quad A[T^{1/S}]_{\underline{\sigma}} \rightarrow D_{A/R,\underline{\sigma}}, \quad T^{1/n} \mapsto 1 + (\sigma_n(x) - x)\partial_{\sigma_n}$$

inducing a functor

$$(20) \quad \partial_{A,\underline{\sigma}}\text{-Mod} \rightarrow \underline{\sigma}_A\text{-Mod}.$$

If x is a strong \underline{q} -coordinate, then the map (19) is an isomorphism and the functor (20) is an equivalence.

Proof. This will follow from Theorem 6.13 of [Le Stum and Quirós 2015b] once we know that the various maps for the different n are all compatible. More precisely, we have to check that the differential operators $1 + (\sigma_n(x) - x)\partial_{\sigma_n}$ form a system of roots in $D_{\underline{\sigma}}$. But this follows from Proposition 8.7. \square

Theorem 9.13 (formal quantum confluence 2). *Let R be a \mathbb{Q} -algebra, \underline{q} a strongly admissible system of roots in R and A a \underline{q} - R -algebra. Then there exists a canonical A -linear map*

$$D_{A/R,\underline{\sigma}} \rightarrow \widehat{D}_{A/R}.$$

If R is \underline{q} -divisible and \underline{q} is infinite, then the map (16) has dense image.

Proof. First of all, the map (16) is obtained by taking the direct limit of the maps (15) for all the σ_n . Now, we set $p_n := q_n\text{-char}(R)$ for all $n \in S$ and we apply the confluence Theorem 8.3 to σ_n . If $p_n = 0$, we are done. Otherwise, the theorem tells

us that the bottom map in the following commutative diagram is surjective:

$$\begin{array}{ccc} D_{\sigma} & \longrightarrow & \widehat{D} \\ \uparrow & & \downarrow \\ D_{\sigma_n} & \twoheadrightarrow & D/\partial^{p_n} \end{array}$$

And we proved in Lemma 9.5 that for all $n \in S$, we have $p_n > n$. Since \underline{q} is infinite, we have $p_n \rightarrow \infty$ and we see that the image of the upper map is dense. \square

Corollary 9.14. *Assume moreover that x is a strong \underline{q} -coordinate on A . Then the following A -linear map has dense image:*

$$A[T^{1/S}]_{\underline{\sigma}} \rightarrow \widehat{D}_{A/R}, \quad T^{m/n} \mapsto \sum_{k=0}^{\infty} \frac{1}{k!} (\sigma_n^m(x) - x)^k \partial^k$$

Proof. The formula comes from Corollary 7.7. \square

Example. Theorem 9.13 (and its corollary) applies in particular when q is a primitive p -th root of unity in some algebraically closed field K of characteristic zero, q is a system of p^n -th roots of q in K , R is an algebra containing K , $A = R[x, x^{-1}]$ and $\sigma_n(x) = q_n x$ (we should actually write σ_{p^n} and q_{p^n} but we will try to make the notation easier to use).

As we already saw, in this situation, $D := D_{A/R}$ is the noncommutative ring $R[x, x^{-1}, \partial]$ with the commutation rule $\partial x = x\partial + 1$ and we may see \widehat{D} (which is *not* a ring) as $R[x, x^{-1}][[\partial]]$. We also want to understand the left-hand side and see that $D_n := D_{A/R, \sigma_n}$ is the noncommutative ring $R[x, x^{-1}, \partial_n]$ with the commutation rule $\partial_n x = q_n x \partial_n + 1$. The transition maps are given by the rather tricky formulas

$$\partial_n \mapsto \sum_{k=1}^{p^{n'}-n} \frac{q_{n'}^{\frac{k(k-1)}{2}} (q_{n'} - 1)^{k-1}}{(k)_{q_{n'}}} \binom{p^{n'}-n-1}{k-1}_{q_{n'}} x^{k-1} \partial_{n'}^k,$$

and the map of the theorem satisfies

$$\partial_n \mapsto \sum_{k=1}^{\infty} \frac{1}{k!} (q_n - 1)^{k-1} x^{k-1} \partial^k.$$

Also, if we set $q^{m/n} := q_n^m$, the map of the corollary is given by

$$T^r \mapsto \sum_{k=1}^{\infty} \frac{1}{k!} (q^r - 1)^k x^k \partial^k.$$

Remark. Alexei Belov-Kanel and Maxim Kontsevich [2007] proved that the Jacobian conjecture is stably equivalent to the Dixmier conjecture [1968]. According to

the latter, any endomorphism of a Weyl algebra over \mathbb{C} is an automorphism. And this is still a conjecture even in dimension one. On the other hand, when q is a primitive p -th root of unity, then the quantum Weyl algebra D_σ does *not* satisfy the Dixmier conjecture but it is an Azumaya algebra. In particular, checking that an endomorphism is an automorphism can be done on the center. As explained by Backelin [2011], it is appealing to attack the Dixmier conjecture through quantum Weyl algebras and one can hope that Theorem 9.13 might provide a tool for this quest.

References

- [Adams 1928/29] C. R. Adams, “On the linear ordinary q -difference equation”, *Ann. of Math.* (2) **30**:1-4 (1928/29), 195–205. MR Zbl
- [André 2001] Y. André, “Différentielles non commutatives et théorie de Galois différentielle ou aux différences”, *Ann. Sci. École Norm. Sup.* (4) **34**:5 (2001), 685–739. MR Zbl
- [André and Di Vizio 2004] Y. André and L. Di Vizio, “ q -difference equations and p -adic local monodromy”, pp. 55–111 in *Analyse complexe, systèmes dynamiques, sommabilité des séries divergentes et théories galoisiennes, I*, edited by M. Loday-Richaud, Astérisque **296**, Société Mathématique de France, Paris, 2004. MR Zbl
- [Backelin 2011] E. Backelin, “Endomorphisms of quantized Weyl algebras”, *Lett. Math. Phys.* **97**:3 (2011), 317–338. MR Zbl
- [Belov-Kanel and Kontsevich 2007] A. Belov-Kanel and M. Kontsevich, “The Jacobian conjecture is stably equivalent to the Dixmier conjecture”, *Mosc. Math. J.* **7**:2 (2007), 209–218. MR Zbl
- [Birkhoff 1911] G. D. Birkhoff, “General theory of linear difference equations”, *Trans. Amer. Math. Soc.* **12**:2 (1911), 243–284. MR Zbl
- [Bourbaki 1970] N. Bourbaki, *Algèbre, chapitres 1 à 3*, Hermann, Paris, 1970. Translated as *Algebra, chapters 1-3*, Addison-Wesley, 1973. MR Zbl
- [Bourbaki 2012] N. Bourbaki, *Algèbre, chapitre 8: Modules et anneaux semisimples*, 2nd revised ed., Springer, Berlin, 2012. MR Zbl
- [Carmichael 1911] R. D. Carmichael, “Linear difference equations and their analytic solutions”, *Trans. Amer. Math. Soc.* **12**:1 (1911), 99–134. MR Zbl
- [Carmichael 1912] R. D. Carmichael, “The general theory of linear q -difference equations”, *Amer. J. Math.* **34**:2 (1912), 147–168. MR Zbl
- [Cohn 1965] R. M. Cohn, *Difference algebra*, Interscience, New York, 1965. MR Zbl
- [Di Vizio 2004] L. Di Vizio, “Introduction to p -adic q -difference equations (weak Frobenius structure and transfer theorems)”, pp. 615–675 in *Geometric aspects of Dwork theory, I, II*, edited by A. Adolphson et al., de Gruyter, Berlin, 2004. MR Zbl arXiv
- [Di Vizio et al. 2003] L. Di Vizio, J.-P. Ramis, J. Sauloy, and C. Zhang, “Équations aux q -différences”, *Gaz. Math.* **96** (2003), 20–49. MR Zbl
- [Dixmier 1968] J. Dixmier, “Sur les algèbres de Weyl”, *Bull. Soc. Math. France* **96** (1968), 209–242. MR Zbl
- [Duval and Roques 2008] A. Duval and J. Roques, “Familles fuchsienues d’équations aux (q -) différences et confluence”, *Bull. Soc. Math. France* **136**:1 (2008), 67–96. MR Zbl
- [Gros and Le Stum 2014] M. Gros and B. Le Stum, “Une neutralisation explicite de l’algèbre de Weyl quantique complétée”, *Comm. Algebra* **42**:5 (2014), 2163–2170. MR Zbl

- [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), 5–333. MR Zbl
- [Guichard 1887] C. Guichard, “Sur la résolution de l’équation aux différences fines $G(x+1) - G(x) = H(x)$ ”, *Ann. Sci. École Norm. Sup. (3)* **4** (1887), 361–380. MR Zbl
- [Hardouin 2010] C. Hardouin, “Iterative q -difference Galois theory”, *J. Reine Angew. Math.* **644** (2010), 101–144. MR Zbl
- [Iyer and McCune 2002] U. N. Iyer and T. C. McCune, “Quantum differential operators on $\mathbb{K}[x]$ ”, *Internat. J. Math.* **13**:4 (2002), 395–413. MR Zbl
- [Karoubi and Suárez-Álvarez 2003] M. Karoubi and M. Suárez-Álvarez, “Twisted Kähler differential forms”, *J. Pure Appl. Algebra* **181**:2-3 (2003), 279–289. MR Zbl
- [Le Stum and Quirós 2015a] B. Le Stum and A. Quirós, “On quantum integers and rationals”, pp. 107–130 in *Trends in number theory* (Seville, 2013), edited by F. Chamizo et al., *Contemp. Math.* **649**, American Mathematical Society, Providence, RI, 2015. MR Zbl
- [Le Stum and Quirós 2015b] B. Le Stum and A. Quirós, “Twisted calculus”, preprint, 2015. arXiv
- [Lunts and Rosenberg 1997] V. A. Lunts and A. L. Rosenberg, “Differential operators on noncommutative rings”, *Selecta Math. (N.S.)* **3**:3 (1997), 335–359. MR Zbl
- [Mebkhout and Narváez Macarro 1998] Z. Mebkhout and L. Narváez Macarro, “Le théorème de continuité de la division dans les anneaux d’opérateurs différentiels”, *J. Reine Angew. Math.* **503** (1998), 193–236. MR Zbl
- [Pulita 2008] A. Pulita, “ p -adic confluence of q -difference equations”, *Compos. Math.* **144**:4 (2008), 867–919. MR Zbl
- [Pulita 2017] A. Pulita, “Infinitesimal deformation of p -adic differential equations on Berkovich curves”, *Math. Ann.* **368**:1-2 (2017), 111–164. MR Zbl
- [Sauloy 2000] J. Sauloy, “Systèmes aux q -différences singuliers réguliers: classification, matrice de connexion et monodromie”, *Ann. Inst. Fourier (Grenoble)* **50**:4 (2000), 1021–1071. MR Zbl
- [Sauloy 2003] J. Sauloy, “Galois theory of Fuchsian q -difference equations”, *Ann. Sci. École Norm. Sup. (4)* **36**:6 (2003), 925–968. MR Zbl

Received April 7, 2017.

BERNARD LE STUM
 INSTITUT DE RECHERCHE MATHÉMATIQUE (IRMAR)
 UNIVERSITE DE RENNES I
 CAMPUS DE BEAULIEU
 35042 RENNES
 FRANCE
 bernard.le-stum@univ-rennes1.fr

ADOLFO QUIRÓS
 DEPARTAMENTO DE MATEMÁTICAS
 UNIVERSIDAD AUTÓNOMA DE MADRID
 CIUDAD UNIVERSITARIA DE CANTOBLANCO
 28049 MADRID
 SPAIN
 adolfo.quiros@uam.es

RIGIDITY OF HAWKING MASS FOR SURFACES IN THREE MANIFOLDS

JIACHENG SUN

It is well known that Hawking mass is nonnegative for a stable sphere of constant mean curvature (CMC) in a three-manifold of nonnegative scalar curvature. R. Bartnik proposed the rigidity problem of the Hawking mass of stable CMC spheres. We show partial rigidity results of Hawking mass for stable CMC spheres in asymptotic flat (AF) manifolds with nonnegative scalar curvature. If the Hawking mass of a nearly round stable CMC surface vanishes, then the surface must be the standard sphere in \mathbb{R}^3 and the interior of the surface is flat. Similar results also hold for asymptotic hyperbolic manifolds. A complete AF manifold having small or large isoperimetric surface with zero Hawking mass must be flat. We use the mean field equation and monotonicity of Hawking mass as well as rigidity results of Y. Shi in our proof.

1. Introduction

One of the most important tasks in general relativity is to understand the mass of spacetime. The first attempt on this topic is the positive mass theorem, which says that the mass of an asymptotic flat manifold is nonnegative if the scalar curvature is nonnegative, and the mass vanishes if and only if the manifold is isometric to standard Euclidean space. Another important attempt is the Penrose inequality, which tells us that the mass is no less than $\sqrt{A/16\pi}$ when there is a horizon, where A is the area of the outmost minimal surface, and the equality holds if and only if the manifold is isometric to Schwarzschild space. From the Penrose inequality we see the impact of boundary behavior is also remarkable. This motivates us to study quasilocal mass for a compact manifold with boundary.

Brown–York mass is a well defined quasilocal mass for a domain with convex boundary, which characterizes the deviation of mean curvature compared with a Euclidean metric, whose positivity and rigidity is proved by [Shi and Tam 2002]. Another important quasilocal mass is Hawking mass, which played a key role in proving the Penrose inequality in [Bray 1997] and [Huiskens and Ilmanen 2001].

MSC2010: 53C21, 53C24, 53C80, 58C40.

Keywords: rigidity, Hawking mass, stable CMC, mean field equation.

Because the Willmore functional of a surface can be arbitrarily large, we cannot expect positivity for an arbitrary surface. But for a stable CMC sphere in a nonnegative scalar curvature manifold, the Hawking mass is nonnegative [Christodoulou and Yau 1988].

Bartnik [2002, p. 235] proposed the rigidity problem of Hawking mass, i.e., what can we say about the ambient manifold when the Hawking mass vanishes for some surface. This paper is devoted to a partial result for the rigidity problem if the surface is nearly round. We study the eigenvalue and eigenfunctions of the Jacobi operator for a stable CMC surface with zero Hawking mass, then transfer the rigidity problem to a mean field type equation with respect to the second eigenvalue 6 of the standard S^2 under some restriction. If the equation has only the zero solution, then the rigidity of Hawking mass holds. We get the local uniqueness by studying the spherical harmonics on S^2 carefully and also iteration methods. If the solution is small in some sense, we can get the power decay of both the kernel part of $\Delta + 6$ and also the orthogonal part. But we believe that the equation has only the zero solution with the integral restriction.

The main term in Hawking mass is the Willmore functional. In \mathbb{R}^3 the Willmore functional is constant 4π if and only if the surface is round sphere. So we can detect the curvature of ambient space by the Willmore functional. For this reason, we expect that manifolds with zero Hawking mass surface may have some flatness properties.

Theorem 1. *Let (M, g) be a complete Riemannian three-manifold with scalar curvature $R(g) \geq 0$ and $\Omega \subset M$ be a domain with boundary $\Sigma = \partial\Omega$. If Σ is a nearly round stable CMC sphere in M with $m_H(\Sigma) = 0$, then Ω is isometric to a Euclidean ball in \mathbb{R}^3 . In particular, Σ is isometric to the standard S^2 in \mathbb{R}^3 . In this paper, nearly round is in the sense that Gauss curvature satisfies*

$$(1-1) \quad \left| \frac{|\Sigma|}{4\pi} K_\Sigma - 1 \right|_{C^0} < \epsilon_0$$

for some universal constant $\epsilon_0 \ll 1$.

The hyperbolic case of the above rigidity is the following:

Theorem 2. *Let (M, g) be a complete Riemannian three-manifold with scalar curvature $R(g) \geq -6$ and $\Omega \subset M$ be a domain with boundary $\Sigma = \partial\Omega$. If Σ is a nearly round stable CMC sphere in M with $m_H(\Sigma) = 0$, then Ω is isometric to a hyperbolic ball in \mathbb{H}^3 .*

By the examples of A. Carlotto and R. Schoen [2016] there are manifolds with nonnegative scalar curvature which are flat in a half space of \mathbb{R}^3 , so we can only expect flatness inside the surface with zero Hawking mass for stable CMC surfaces.

But we can get global flatness for isoperimetric surfaces of sphere type:

Theorem 3. *Let (M, g) be a complete AF three-manifold with scalar curvature $R(g) \geq 0$. If there exists a nearly round isoperimetric sphere Σ with $m_H(\Sigma) = 0$, then (M, g) is isometric to (\mathbb{R}^3, δ) .*

This theorem also has a hyperbolic version:

Theorem 4. *Let (M, g) be a complete AH three-manifold with scalar curvature $R(g) \geq -6$. If there exists a nearly round isoperimetric sphere Σ with $m_H(\Sigma) = 0$, then (M, g) is isometric to $(\mathbb{H}^3, g_{\mathbb{H}})$.*

We already know from [Chodosh et al. 2016] that large surfaces of the canonical stable CMC foliation in [Huisken and Yau 1996; Qing and Tian 2007] are isoperimetric and close to the coordinate spheres. So we can get the rigidity result for large isoperimetric surfaces. For rigidity of small isoperimetric surfaces, we use the monotonicity of Hawking mass and also a rigidity result of Y. Shi.

Theorem 5. *Let (M, g) be a complete AF three-manifold with scalar curvature $R(g) \geq 0$ which has no compact minimal surface. If there is an small enough isoperimetric surface Σ with $m_H^+(\Sigma) = 0$ (see definition in Section 5.3), then (M, g) is isometric to \mathbb{R}^3 .*

Structure of this paper. In Section 2, we give the basic definitions. In Section 3, we prove the rigidity of Hawking mass for nearly round stable CMC spheres. We transform the rigidity problem to a mean field-type equation, and prove the local uniqueness of the zero solution. By doing so, we get that a surface with zero Hawking mass must be the standard S^2 and then use the rigidity of [Shi and Tam 2002; 2007] to finish the proof. In Section 4, we prove the global properties of manifolds with nearly round isoperimetric surfaces having zero Hawking mass. This directly implies the rigidity for large isoperimetric surfaces in the canonical stable CMC foliation by Huisken and Yau [1996] and Qing and Tian [2007]. In Section 5, we prove the rigidity for small isoperimetric surfaces by using the monotonicity of Hawking mass. This relies on the fact that the topology of a small isoperimetric surface must be a sphere. In Appendix A.1 we give the spherical harmonics and computations for the square of second order spherical harmonics. In Appendix A.2 we sketch a proof of the existence of isoperimetric surfaces for all volumes in AF three-manifolds. In Appendix A.3 we sketch the proof of continuity of isoperimetric profile for AF manifolds which is important to prove the right continuity of I'_+ .

2. Preliminaries

We give some basic notations to present our result. Let $\Sigma \subset (M, g)$ be a surface with unit normal vector field n , second fundamental form A and mean curvature H .

Definition. The Willmore functional of Σ is defined by:

$$(2-1) \quad W(\Sigma) = \frac{1}{4} \int_{\Sigma} H^2,$$

when $R(g) \geq 0$ and

$$(2-2) \quad W(\Sigma) = \frac{1}{4} \int_{\Sigma} (H^2 - 4).$$

when $R(g) \geq -6$.

The Willmore functional appears in various areas, such as bending energy of elastic membranes. It appears naturally in general relativity in the form of the Hawking mass of a surface:

Definition. The Hawking mass of Σ is defined by

$$(2-3) \quad m_H(\Sigma) = \frac{|\Sigma|^{1/2}}{(16\pi)^{3/2}} \left(16\pi - \int_{\Sigma} H^2 \right),$$

when $R(g) \geq 0$ and

$$(2-4) \quad m_H(\Sigma) = \frac{|\Sigma|^{1/2}}{(16\pi)^{3/2}} \left(16\pi - \int_{\Sigma} (H^2 - 4) \right).$$

when $R(g) \geq -6$.

Definition. If H is constant along Σ , we say Σ is a CMC surface; the Jacobi operator of a CMC surface Σ is the second variation of area:

$$(2-5) \quad L_{\Sigma} = -\Delta_{\Sigma} - (|A|^2 + \text{Ric}(n, n)).$$

A CMC surface Σ is *stable* if the first eigenvalue of L_{Σ} on mean-zero functions is nonnegative:

$$(2-6) \quad \Lambda_1(L_{\Sigma}) = \inf \left\{ \int_{\Sigma} f L_{\Sigma} f : \int_{\Sigma} f = 0, \int_{\Sigma} f^2 = 1 \right\} \geq 0,$$

i.e., it satisfies the following stability condition:

$$(2-7) \quad \int_{\Sigma} (|A|^2 + \text{Ric}(n, n)) f^2 \leq \int_{\Sigma} |\nabla f|^2$$

for all $f \in C_c^{\infty}(\Sigma)$ and $\int_{\Sigma} f = 0$.

Remark. The above definition of eigenvalue in mean-zero functions is different from the eigenvalue defined in the ordinary way by min-max construction:

$$(2-8) \quad \lambda_1(L_{\Sigma}) = \inf \left\{ \int_{\Sigma} f L_{\Sigma} f : \int_{\Sigma} f u_0 = 0, \int_{\Sigma} f^2 = 1 \right\}$$

where u_0 is the zeroth eigenfunction of L_Σ . By definition we have

$$(2-9) \quad \Lambda_1(L_\Sigma) \leq \lambda_1(L_\Sigma).$$

We also want to study isoperimetric surfaces in AF (resp. AH) three-manifolds. We will always use the bracket to denote the asymptotic hyperbolic case after related asymptotic flat situations.

Definition. A complete connected three-manifold (M, g) is called AF (resp. AH), if there exists a constant $C > 0$ and a compact set K such that $M \setminus K$ is diffeomorphic to $\mathbb{R}^3 \setminus B_R(0)$ for some $R > 0$, and in standard coordinates the metric g has the following properties:

$$(2-10) \quad g = \delta + h \quad (\text{resp. } g = g_{\mathbb{H}} + h)$$

and

$$(2-11) \quad |h_{ij}| + r|\partial h_{ij}| + r^2|\partial^2 h_{ij}| \leq Cr^{-\tau}$$

$\tau \in (\frac{1}{2}, 1]$ (resp. $\tau = 3$), where r and ∂ denote the Euclidean distance and standard derivative operator on \mathbb{R}^3 respectively. The region $M \setminus K$ is called the end of M . The standard hyperbolic space $(\mathbb{H}^3, g_{\mathbb{H}})$ is

$$(2-12) \quad g_{\mathbb{H}} = \frac{1}{1+r^2} dr^2 + r^2 g_{S^2}.$$

We also need the following definition of isoperimetric surface:

Definition. Given a complete Riemannian 3-manifold (M, g) , its isoperimetric profile with volume V is defined as

$$(2-13) \quad I(V) = \inf \left\{ \mathcal{H}^2(\partial^* \Omega) : \begin{array}{l} \Omega \subset M \text{ is a Borel set with finite perimeter} \\ \text{and } \mathcal{H}_g^3(\Omega) = V \end{array} \right\},$$

where \mathcal{H}^2 is a 2-dimensional Hausdorff measure for the reduced boundary of Ω denoted by $\partial^* \Omega$. A Borel set $\Omega \subset M$ of finite perimeter such that $\mathcal{H}_g^3(\Omega) = V$ and $I(V) = \mathcal{H}^2(\partial^* \Omega)$ is called an isoperimetric region of (M, g) of volume V . The surface $\partial \Omega$ is called an isoperimetric surface.

3. Rigidity of Hawking mass for nearly round stable CMC surfaces

It was shown in [Christodoulou and Yau 1988] that the Hawking mass is nonnegative for a stable CMC sphere. It is proved by using a Hersch-type test function in the stability condition and the nonnegativity of scalar curvature. Since we need to study the equality case, we prove it here for completeness.

Lemma 6 [Christodoulou and Yau 1988]. *Let (M, g) be a Riemannian three-manifold with scalar curvature $R(g) \geq 0$, if Σ is a stable CMC sphere in M , then $m_H(\Sigma) \geq 0$.*

Proof. By [Li and Yau 1982] there exists a conformal $\varphi : \Sigma \rightarrow S^2 \subseteq \mathbb{R}^3$ with $\int_{\Sigma} \varphi = 0$. We can plug these test functions in stability condition, and using that

$$(3-1) \quad \int_{\Sigma} |\nabla \varphi_i|^2 \geq \int_{\Sigma} (|A|^2 + \text{Ric}(n, n)) \varphi_i^2$$

for a surface conformal to $S^2 \subseteq \mathbb{R}^3$,

$$(3-2) \quad \int_{\Sigma} |\nabla \varphi_i|^2 d\mu_{\Sigma} = \int_{S^2} |\nabla x_i|^2 d\mu_{S^2} = - \int_{S^2} x_i \Delta x_i d\mu_{S^2} = 2 \int_{S^2} x_i^2 d\mu_{S^2} = \frac{8}{3}\pi.$$

Thus we can get

$$(3-3) \quad 8\pi \geq \int_{\Sigma} |A|^2 + \text{Ric}(n, n).$$

By Gauss's equation

$$(3-4) \quad K_{\Sigma} = \frac{R}{2} - \text{Ric}(n, n) + \frac{1}{2}(H^2 - |A|^2).$$

So we have

$$(3-5) \quad |A|^2 + \text{Ric}(n, n) = \frac{R}{2} - K_{\Sigma} + \frac{1}{2}(H^2 + |A|^2) = \frac{1}{2}(R + |A^0|^2) + \frac{3}{4}H^2 - K_{\Sigma},$$

where we have used that $|A|^2 = |A^0|^2 + \frac{1}{2}H^2$. We get

$$(3-6) \quad 8\pi \geq \frac{1}{2} \int_{\Sigma} (R + |A^0|^2) + \frac{3}{4} \int_{\Sigma} H^2 - \int_{\Sigma} K_{\Sigma},$$

so we obtain

$$(3-7) \quad 16\pi - \int_{\Sigma} H^2 \geq \frac{2}{3} \int_{\Sigma} (R + |A^0|^2) \geq 0 \quad \square$$

We can get an analogous result for the hyperbolic case; see also [Chodosh 2016]:

Lemma 7. *Let (M, g) be a Riemannian three-manifold with scalar curvature $R(g) \geq -6$. If Σ is a stable CMC sphere in M , then $m_H(\Sigma) \geq 0$.*

Now we start to study stable CMC surfaces with zero Hawking mass. First we can get a spectral characterization of them. We need the following lemma in [El Soufi and Ilias 1992], which gives a optimal estimate of the second eigenvalue of the Schrödinger operator. It also gives part of the rigidity of the second eigenvalue which is the case for a Jacobi operator on a stable CMC sphere.

Lemma 8 [El Soufi and Ilias 1992]. *For any continuous function q on surface Σ ,*

$$(3-8) \quad \lambda_1(-\Delta_{\Sigma} + q)|\Sigma| \leq 2A_c(\Sigma) + \int_{\Sigma} q.$$

The equality holds if and only if Σ admits a conformal map into the standard S^2 whose components are the first eigenfunctions. If Σ is of genus zero, then the equality implies that Σ is conformal to the standard S^2 in \mathbb{R}^3 and q is given by the energy density of a Möbius transform, where λ_1 is the first eigenvalue of $-\Delta_\Sigma + q$ in the sense of (2-8), $A_c(\Sigma)$ is the conformal volume in [Li and Yau 1982] and for a sphere, $A_c(\Sigma) = 4\pi$.

By the above lemma we have the following characterization of zero-Hawking mass stable CMC spheres.

Proposition 9. *Let (M, g) be a complete Riemannian three-manifold with scalar curvature $R(g) \geq 0$ (resp. $R(g) \geq -6$). If Σ is a stable CMC sphere with $m_H(\Sigma) = 0$ and area $|\Sigma| = 4\pi$, then the second eigenvalue $\lambda_1(-\Delta_\Sigma + K_\Sigma) = 3$, with three eigenfunctions $\varphi_1, \varphi_2, \varphi_3$, $\int_\Sigma \varphi_i = 0$, and $\sum_{i=1}^3 \varphi_i^2 = 1$. In particular, $|\nabla \varphi|^2 = 3 - K_\Sigma$, which is independent of eigenfunctions.*

Proof. From the above proof of Lemma 6 we can see if $m_H(\Sigma) = 0$ on Σ , we have $\int_\Sigma H^2 = 16\pi$ (resp. $\int_\Sigma (H^2 - 4) = 16\pi$), $R = 0$ (resp. $R = -6$), $A^0 = 0$ on Σ . The area $|\Sigma| = 4\pi$, then $H = 2$ (resp. $H = 2\sqrt{2}$), the Jacobi operator becomes

$$(3-9) \quad L_\Sigma = -\Delta_\Sigma + K_\Sigma - 3.$$

By the stability of Σ and Lemma 8, we have

$$(3-10) \quad 0 \leq 4\pi \Lambda_1(L_\Sigma) \leq 4\pi \lambda_1(L_\Sigma) \leq 8\pi + \int_\Sigma (K_\Sigma - 3) = 0,$$

so all the equalities hold, in particular

$$(3-11) \quad \lambda_1(-\Delta_\Sigma + K_\Sigma) = 3,$$

with three eigenfunctions $\varphi_1, \varphi_2, \varphi_3$, $\int_\Sigma \varphi_i = 0$, and $\sum_{i=1}^3 \varphi_i^2 = 1$, so

$$(3-12) \quad -\Delta_\Sigma \varphi + K_\Sigma \varphi - 3\varphi = 0.$$

By $|\varphi|^2 = \sum_{i=1}^3 \varphi_i^2 = 1$, we have

$$(3-13) \quad 0 = \Delta_\Sigma |\varphi|^2 = 2\varphi \Delta_\Sigma \varphi + 2|\nabla \varphi|^2$$

Taking inner product of φ with (3-12), we get

$$(3-14) \quad |\nabla \varphi|^2 = 3 - K_\Sigma. \quad \square$$

Remark. We see from the above lemma that the first eigenvalue and eigenfunctions of the Schrödinger operator $-\Delta_\Sigma + K_\Sigma$ equal those of the standard S^2 . We expect that the metric is isometric to the standard metric on S^2 .

In the following, we will always use Σ to denote a stable CMC surface with zero Hawking mass. Let $\varphi : \Sigma \rightarrow S^2 \subseteq \mathbb{R}^3$ be the conformal map in Proposition 9 with $\int_{\Sigma} \varphi = 0$. Denote the metric on Σ by $g = e^u g_0$, with g_0 the standard metric on S^2 . By definition of the conformal map φ ,

$$(3-15) \quad e^{-u} = \frac{1}{2} |\nabla \varphi|^2.$$

The standard formula for Gauss curvature under a conformal change of metric gives

$$(3-16) \quad K_{\Sigma} = e^{-u} \left(1 - \frac{1}{2} \Delta_{g_0} u \right).$$

So (3-14) gives

$$(3-17) \quad \Delta_{g_0} u = 6 - 6e^u.$$

Also the volume-preserving variation implies

$$(3-18) \quad \int_{S^2} x_i e^u = 0.$$

So for this stable CMC surface with zero Hawking mass Σ ,

$$(3-19) \quad K_{\Sigma} - 1 = e^{-u} (1 - 3 + 3e^u) - 1 = 2(1 - e^{-u}).$$

This means that if u is C^0 close to 0, then K_{Σ} is C^0 close to 1, which implies Σ is nearly round. If we can prove (3-17), (3-18) admit only the zero solution, then the stable CMC surface with vanishing Hawking mass is isometric to the standard S^2 .

Equations of the same type as (3-17) have been studied in various aspects, such as prescribed Gaussian curvature [Kazdan and Warner 1974], the mean field model and the Chern–Simons–Higgs model. This kind of equation may have bifurcation when approaching the eigenvalues of S^2 , so it may lose compactness. Ding, Jost, Li and Wang [Ding et al. 1997; 1998] have studied the equation at the first eigenvalue. Li [1999] has initiated the study of the existence of solutions by computing the Leray–Schauder topological degree. Lin [2000] computed the degree on S^2 and surface of any genus [Chen and Lin 2003], but there is little work on the uniqueness of this kind of equation at second eigenvalue of S^2 . In fact, because the bifurcation occurs after the first eigenvalue, it is hard to guarantee the uniqueness globally, but we can get local uniqueness of the constant solution for (3-17). That's why we put the nearly round condition in our results. We use the Lyapunov–Schmidt decomposition as in [Neves and Tian 2009] to estimate the kernel of $\Delta_{g_0} + 6$ and the orthogonal part separately.

Lemma 10. *Let u satisfy*

$$(3-20) \quad \Delta_{g_0} u = 6(1 - e^u)$$

on standard S^2 . There exists a universal constant $\delta_0 > 0$ such that if $\sup |u| < \delta_0$, then $u \equiv 0$.

Proof. In the following, the constant C is universal, and may differ from line to line. Denote $E_2 = \ker\{\Delta_{g_0} + 6\}$, which is the second eigenspace of $-\Delta_{g_0}$ on the standard S^2 . It is well known that $E_2 = \text{span}\{Y_{2,-2}, Y_{2,-1}, Y_{2,0}, Y_{2,1}, Y_{2,2}\}$ (see appendix below). Let P_2 be the projection operator on E_2 . Consider the decomposition $u = u_1 + u_2$, where $u_1 \in E_2^\perp$, and $u_2 \in E_2$. Then

$$(3-21) \quad \Delta_{g_0} u_1 + 6u_1 = 6(1 + u - e^u),$$

$$(3-22) \quad \Delta_{g_0} u_2 + 6u_2 = 0.$$

As $(\Delta_{g_0} + 6)^{-1}$ is bounded from L^2 to $W^{2,2}$ on E_2^\perp , we have

$$(3-23) \quad |u_1|_{W^{2,2}} \leq C|1 + u - e^u|_{L^2}.$$

We can assume

$$(3-24) \quad \sup |u| \leq \delta < 1.$$

Then from (3-23) and the Sobolev embedding, we have

$$(3-25) \quad |u_1|_{L^\infty} \leq C|u^2|_{L^2} \leq C\delta^2.$$

Also from equation (3-17), we know

$$(3-26) \quad |\Delta_{g_0} u + 6u + 3u^2|_{L^2} = 6|1 + u + \frac{1}{2}u^2 - e^u|_{L^2} \leq C|u^3|_{L^2} \leq C\delta^3.$$

By (3-25), we can get

$$|u_1^2|_{L^2} \leq C\delta^4.$$

By the decomposition of $u_2 = u - u_1$ we have

$$(3-27) \quad |\Delta_{g_0} u + 6u + 3u_2^2|_{L^2} \leq 2|u^3|_{L^2} + 6|u_1 u|_{L^2} + 3|u_1^2|_{L^2} \leq C\delta^3.$$

In order to get the estimate of u_2 , we project the above equation to E_2 . Then

$$(3-28) \quad |P_2 u_2^2|_{L^2} \leq C\delta^3.$$

By Lemma 11 below and (3-28) we have

$$(3-29) \quad |u_2|_{L^\infty} \leq C|u_2|_{L^2} \leq C\delta^{3/2}$$

Combining (3-25) and (3-29), we improve the initial assumption (3-24):

$$(3-30) \quad \sup |u| \leq C|u|_{L^2} < C\delta^{3/2}.$$

Taking $\delta_0 = \frac{1}{2}C^{-2}$ and iterate the procedure, we get

$$(3-31) \quad \sup |u| \leq C_0|u|_{L^2} < C^{-2}(C^2\delta_0)^{(3/2)^k} = C^{-2}\left(\frac{1}{2}\right)^{(3/2)^k},$$

and let $k \rightarrow \infty$, we get the desired result. \square

Lemma 11. *For all $u_2 \in E_2$,*

$$(3-32) \quad |P_2 u_2|_{L^2}^2 = \frac{1}{7} \sqrt{\frac{5}{\pi}} |u_2|_{L^2}^2.$$

Proof. Let

$$(3-33) \quad u_2 = \sum_{i=-2}^2 \lambda_i Y_{2,i},$$

where $Y_{2,i}$ are the second-order spherical harmonics (see Appendix A.1). By computations and projecting u_2^2 to E_2 , we have

$$(3-34) \quad \begin{aligned} P_2 u_2^2 &= \frac{1}{14} \sqrt{\frac{5}{\pi}} [2(\lambda_0^2 - \lambda_{-2}^2 - \lambda_2^2) + \lambda_1^2 + \lambda_{-1}^2] Y_{2,0} \\ &\quad + \frac{1}{7} \sqrt{\frac{5}{\pi}} (\sqrt{3} \lambda_{-1} \lambda_1 - 2 \lambda_{-2} \lambda_0) Y_{2,-2} \\ &\quad + \frac{1}{14} \sqrt{\frac{5}{\pi}} [\sqrt{3}(\lambda_1^2 - \lambda_{-1}^2) - 4 \lambda_0 \lambda_2] Y_{2,2} \\ &\quad + \frac{1}{7} \sqrt{\frac{5}{\pi}} [\lambda_{-1} \lambda_0 + \sqrt{3}(\lambda_{-2} \lambda_1 - \lambda_{-1} \lambda_2)] Y_{2,-1} \\ &\quad + \frac{1}{7} \sqrt{\frac{5}{\pi}} [\lambda_0 \lambda_1 + \sqrt{3}(\lambda_{-2} \lambda_{-1} + \lambda_1 \lambda_2)] Y_{2,1}. \end{aligned}$$

Thus

$$(3-35) \quad \begin{aligned} |P_2 u_2^2|_{L^2}^2 &= \left(\frac{1}{7} \sqrt{\frac{5}{\pi}} \right)^2 \left\{ \frac{1}{4} (2\lambda_0^2 - 2\lambda_{-2}^2 - 2\lambda_2^2 + \lambda_{-1}^2 + \lambda_1^2)^2 \right. \\ &\quad + (\sqrt{3} \lambda_{-1} \lambda_1 - 2 \lambda_{-2} \lambda_0)^2 + \frac{1}{4} [\sqrt{3}(\lambda_1^2 - \lambda_{-1}^2) - 4 \lambda_0 \lambda_2]^2 \\ &\quad + [\lambda_{-1} \lambda_0 + \sqrt{3}(\lambda_{-2} \lambda_1 - \lambda_{-1} \lambda_2)]^2 \\ &\quad \left. + [\lambda_0 \lambda_1 + \sqrt{3}(\lambda_{-2} \lambda_{-1} + \lambda_1 \lambda_2)]^2 \right\} \\ &= \left(\frac{1}{7} \sqrt{\frac{5}{\pi}} \right)^2 \left(\sum_{i=-2}^2 \lambda_i^2 \right)^2 = \left(\frac{1}{7} \sqrt{\frac{5}{\pi}} \right)^2 |u_2|_{L^2}^2. \quad \square \end{aligned}$$

The following rigidity result is a kind of positive mass theorem in the compact case (see [Miao 2002], [Shi and Tam 2002], and [Hang and Wang 2006]):

Lemma 12. *Let (M, g) be a compact, orientable Riemannian 3-manifold with scalar curvature $R(g) \geq 0$ and ∂M isometric to a round S^2 with mean curvature $H = 2$. Then (M, g) is isometric to the unit ball in (\mathbb{R}^3, δ) .*

To prove Theorem 2 we need a rigidity result for the hyperbolic case of the sphere; see Theorem 3.8 in [Shi and Tam 2007].

Lemma 13. *Let (M, g) be a compact orientable Riemannian 3-manifold with scalar curvature $R(g) \geq -6$ and ∂M isometric to a round S^2 with mean curvature $H = 2\sqrt{2}$. Then (M, g) is isometric to the unit ball in hyperbolic space \mathbb{H}^3 .*

After Lemmas 10, 12 and 13, now we are in the position to prove Theorems 1 and 2.

Proof of Theorems 1 and 2. If $m_H(\Sigma) = 0$ on a nearly round stable CMC surface Σ , without loss of generality, assume $|\Sigma| = 4\pi$, and then $H = 2$ (resp. $H = 2\sqrt{2}$)

$$L_\Sigma = -\Delta_\Sigma + K - 3.$$

By Lemma 10 we get the nearly round stable CMC surface Σ is the standard S^2 in \mathbb{R}^3 . Then by Lemma 12 (resp. Lemma 13), we conclude that Ω isometric to a unit ball in \mathbb{R}^3 (\mathbb{H}^3_{-1}). \square

Theorem 1 (resp. Theorem 2) and Lemma 6 (resp. Lemma 7) can help us to understand the Willmore functional in manifolds with scalar curvature $R(g) \geq 0$ (resp. $R(g) \geq -6$).

Corollary 14. *Let (M, g) be a complete Riemannian three-manifold with scalar curvature $R(g) \geq 0$ (resp. $R(g) \geq -6$), $\Sigma = \partial\Omega$ a stable CMC sphere. Then $W(\Sigma) \leq 4\pi$. If Σ is nearly round, then equality holds if and only if Σ is the standard S^2 and Ω isometric to unit ball in \mathbb{R}^3 (resp. \mathbb{H}^3).*

4. Rigidity of Hawking mass for nearly round isoperimetric surfaces

Theorem 1 can be used to prove rigidity of isoperimetric surfaces in AF manifolds. By the manifold constructed by A. Carlotto and R. Schoen [2016]; see also [Chodosh et al. 2016]:

Example. There is an asymptotically flat Riemannian metric g on \mathbb{R}^3 with nonnegative scalar curvature and positive mass and such that $g = \delta$ on $\mathbb{R}^2 \times (0, +\infty)$.

We can only expect flatness inside the surface with zero Hawking mass for stable CMC surface. In order to prove Theorem 3 we need the following isoperimetric inequality of [Shi 2016], which also plays a key role in proving the existence of isoperimetric surfaces for all volumes in AF three-manifolds. It says that if there exists a Euclidean ball in an AF manifold with nonnegative scalar curvature, then the AF manifold must be \mathbb{R}^3 .

Lemma 15 [Shi 2016]. *Suppose (M, g) is an AF manifold with scalar curvature $R(g) \geq 0$. Then for any $V > 0$,*

$$(4-1) \quad I(V) \leq (36\pi)^{1/3} V^{2/3}.$$

There is a $V_0 > 0$ with

$$(4-2) \quad I(V_0) = (36\pi)^{1/3} V_0^{2/3}$$

if and only if (M, g) is isometric to \mathbb{R}^3 .

Also there is an analogous result for an isoperimetric profile on AH manifolds; see Proposition 3.3 in [Ji et al. 2016].

Lemma 16 [Ji et al. 2016]. *Suppose (M, g) is an AH manifold with scalar curvature $R(g) \geq -6$. Then for any $V > 0$,*

$$(4-3) \quad I(V) \leq I_{\mathbb{H}}(V).$$

There is a $V_0 > 0$ with

$$(4-4) \quad I(V_0) = I_{\mathbb{H}}(V_0)$$

if and only if (M, g) is isometric to $(\mathbb{H}^3, g_{\mathbb{H}})$.

Now we can prove the rigidity of nearly round isoperimetric surfaces:

Proof of Theorems 3 and 4. If there is a nearly round isoperimetric surface Σ with $m_H(\Sigma) = 0$, and we assume $|\Sigma| = 4\pi$, then $H = 2$. By Theorem 1, the isoperimetric region is a Euclidean ball of volume $\frac{4}{3}\pi$. So we have

$$(4-5) \quad 4\pi = I\left(\frac{4}{3}\pi\right) = (36\pi)^{1/3} \left(\frac{4}{3}\pi\right)^{2/3}.$$

By the rigidity part of Lemma 15, we conclude that (M, g) is isometric to \mathbb{R}^3 . Theorem 4 follows similarly from Theorem 2 and Lemma 16. \square

In fact, large surfaces of the canonical stable CMC foliation in [Huisken and Yau 1996; Qing and Tian 2007] are isoperimetric and close to the coordinate spheres.

Corollary 17. *Let (M, g) be an AF three-manifold with scalar curvature $R(g) \geq 0$. Then the Hawking mass of all the large enough surfaces in the canonical stable CMC foliation in [Huisken and Yau 1996; Qing and Tian 2007] are positive unless (M, g) is isometric to \mathbb{R}^3 .*

5. Rigidity of Hawking mass for small isoperimetric surfaces

For rigidity of small isoperimetric surfaces, we need to prove that such a surface is a sphere when the volume is small enough.

5.1. Topology of small isoperimetric surface. It is shown in [Ros 2005] that for a compact manifold without boundary, the isoperimetric surface is a topological sphere when the enclosing volume is small enough to be contained in a geodesic ball. For AF manifolds we still have this property; the proof follows that of the compact case in that work and relies on the behavior at infinity.

Lemma 18. *If (M, g) is a complete AF three-manifold without boundary, then there exists a $\delta_0 > 0$, such that for all volume $V \leq \delta_0$ the isoperimetric region is convex and contained in a small neighborhood of some point of M . In particular,*

$$(5-1) \quad I(V) \sim (36\pi)^{1/3} V^{2/3} \quad \text{when } V \rightarrow 0.$$

Proof. Let $\{\Sigma_n\}$ be a sequence of isoperimetric surfaces with second fundamental form A_n and volume $V_n \rightarrow 0$. There are two possibilities:

Case 1: $\{|A_n|\}$ is unbounded. Assume $r_n = \max |A_n| = |A_n|(x_n)$; by scaling Σ_n homothetically to $\Sigma'_n = r_n \Sigma_n$ with metric $g_n = r_n^2 g$, also $r_n \rightarrow \infty$, $x_n \in \Sigma'_n$, second fundamental form of Σ'_n satisfies $\max |A'_n| = |A'_n|(x_n)| = 1$. We have $(M, x_n, g_n) \rightarrow (\mathbb{R}^3, 0, \delta)$ smoothly; the limit manifold is standard \mathbb{R}^3 because the manifold is AF. Thus Σ'_n is a sequence of stable CMC surfaces with bounded curvature, and locally Σ'_n consists of a certain number of sheets as in Figure 13 of [Ros 2005]. Each one of the sheets is a graph over a bounded planar domain with bounded derivatives.

If two of the sheets become arbitrarily close near some point when $n \rightarrow \infty$, then we can modify the surface to get a new one with smaller area and the same volume. For details, see page 196 of [Ros 2005]. More precisely, if the two sheets of the surface become arbitrarily close near some point, we can reduce area without modifying the enclosed volume, which contradicts the minimizing property of Σ_n . There are three cases:

- (1) if there is a thin slab, then we can cut part of the volume to one end and reduce the area;
- (2) if there is a thin defect of the region, then we can fill part of the defect with volume gained from deforming a far-away portion of the boundary to reduce area;
- (3) if there are two close thin defects in the region, then we can reduce the area by moving the part between two defects to one of the defects.

Hence by compactness results [Pérez and Ros 2002], up to a subsequence, $\Sigma'_n \rightarrow \Sigma'$ smoothly with multiplicity one and $\Sigma' \subset \mathbb{R}^3$ is a surface of constant mean curvature $H_{\Sigma'}$ properly embedded in \mathbb{R}^3 endowed with a standard metric δ , $0 \in \Sigma'$, $|A'(0)|^2 = 1$. The fact that Σ_n are isoperimetric surfaces implies that Σ' is a stable CMC surface. By [da Silveira 1987] and the stability condition, we can conclude that Σ' is either a union of planes or a sphere. That the curvature at the origin is one implies Σ' is a unit sphere. Going back to Σ_n , for n large enough, the mean curvature H_{Σ_n} of Σ_n is large enough, such that

$$(5-2) \quad \frac{1}{2} H_{\Sigma_n}^2 + \text{Ric}(n, n) > 0.$$

If Σ_n is not connected, since the mean curvature of the isoperimetric surface Σ_n is the same (see Appendix A.4) for each component Σ_n^i , as $|A_n^i|^2 \geq \frac{1}{2} H_{\Sigma_n^i}^2$,

$$(5-3) \quad |A_n^i|^2 + \text{Ric}(n, n) > 0$$

on every component Σ_n^i . On the other hand, we can construct a variation f_i on Σ_n^i which is constant and $\sum_i \int_{\Sigma_n^i} f_i = 0$ in the stability condition of the isoperimetric inequality. This gives

$$(5-4) \quad 0 \geq \sum_i f_i^2 \int_{\Sigma_n^i} |A_n|^2 + \text{Ric}(n, n),$$

a contradiction. So for large n , we know Σ_n is connected and thus a sphere.

Case 2: $\{|A_n|\}$ is bounded. Scale Σ_n to enclose volume 1. By the above argument we get the limit consists of pairwise disjoint planes enclosing volume 1, a contradiction. So the lemma follows. \square

By the above lemma, the rigidity follows from Theorem 3. But it can also be proved by the monotonicity of Hawking mass with respect to the volume of the connected isoperimetric surface. This method relies on the connectedness of isoperimetric surface which was used by Bray [1997]. Bray needed the connectedness of isoperimetric surface when proving monotonicity of Hawking mass.

5.2. Properties of I . The isoperimetric profile I contains important geometric information of the manifold. It is nondecreasing in the outside of horizon. It is concave if the manifold has nonnegative Ricci curvature. The existence and regularity properties of isoperimetric regions for *all* volumes for AF is proved by combining [Shi 2016] with [Carlotto et al. 2016]; we sketch the proof in Appendix A.2 for completeness.

The continuity and differentiability of I for AF manifold is proved as in [Flores and Nardulli 2014] for manifolds with bounded geometry (Ricci curvature and volume of unit geodesic ball bounded below):

Lemma 19. *Given (M, g) is an AF manifold and $V \in (0, \infty)$, let $\Omega \subset M$ be an isoperimetric region with $\text{vol}(\Omega) = V$ and denote $\partial\Omega$ by Σ . The isoperimetric profile has the following regularity:*

- (a) *It is continuous and has left and right derivatives at V , $I'_+(V) \leq H_\Sigma \leq I'_-(V)$ and $I'_+(V)$ and $I'_-(V)$ are right and left continuous respectively.*
- (b) *The inequality $I''(V)I(V)^2 + \int_\Sigma (\text{Ric}(n, n) + |A_\Sigma|^2) \leq 0$ holds in the sense of comparison functions, i.e., for every $V_0 \geq 0$, there is a smooth function $I_{V_0}(V) \geq I(V)$, $I_{V_0}(V_0) = I(V_0)$ and $I''_{V_0}(V)I_{V_0}(V)^2 + \int_\Sigma (\text{Ric}(n, n) + |A_\Sigma|^2) \leq 0$.*

Proof. The continuity of I is proved in Appendix A.3 by adding and subtracting a small geodesic ball to the isoperimetric regions under the condition of bounded geometry. We only prove (b) which implies the differentiability of I . For every $V_0 > 0$, assume Ω_0 is the isoperimetric region with volume V_0 and $\Sigma_0 = \partial\Omega_0$ is the isoperimetric surface with unit outer normal n_0 , second fundamental form A_0 and mean curvature H_0 . In order to get an upper bound of I'' we do a unit normal

variation on Σ_0 . Let Σ_t denote the surface by flowing Σ_0 out with unit speed along the normal n_0 for time t . Since Σ_0 is a smooth embedded surface, there exists a $\delta > 0$ such that Σ_t exists for any $t \in (-\delta, \delta)$. Let $I_{V_0}(t) = \text{area}(\Sigma_t)$. By the first and second variational formula for area we have:

$$(5-5) \quad I'_{V_0}(t) = \int_{\Sigma_t} H \, d\mu,$$

$$(5-6) \quad V'(t) = I_{V_0}(t),$$

$$(5-7) \quad H'(t) = -|A|^2 - \text{Ric}(n, n).$$

We can also parametrize this isoperimetric surface by its volume as $\Sigma(V)$, and $I_{V_0}(V) = \text{area}(\Sigma(V))$. By definition of $\Sigma(V_0)$, $I_{V_0}(V) \geq I(V)$, $I_{V_0}(V_0) = I(V_0)$, so

$$(5-8) \quad I'_{V_0}(V) = \frac{\int_{\Sigma(V)} H \, d\mu}{I_{V_0}(V)} = H.$$

The second derivative of I_{V_0} is

$$(5-9) \quad \begin{aligned} I''_{V_0}(V) &= \frac{\int_{\Sigma(V)} (H^2 - |A|^2 - \text{Ric}(n, n)) \, d\mu}{I_{V_0}^2(V)} - \frac{H}{I_{V_0}^2(V)} \int_{\Sigma_t} H \, d\mu \\ &= -\frac{\int_{\Sigma(V)} |A|^2 + \text{Ric}(n, n) \, d\mu}{I_{V_0}^2(V)}. \end{aligned}$$

For an AF three-manifold, Ricci curvature is bounded below. Thus there exists $k \in \mathbb{R}$ such that $\text{Ric} \geq kg$, and it follows that

$$(5-10) \quad I''_{V_0}(V) \leq -\frac{k}{I_{V_0}(V)}.$$

If $k \geq 0$, then $I_{V_0}(V)$ is concave, and by Lemma 20 below we can get the concaveness of $I(V)$, and then the conclusion follows. In particular, I'_+ , I'_- are both nonincreasing functions, they are right and left continuous respectively and I'' exists almost everywhere.

If $k < 0$, let $\lambda = \lambda(k, a, b) := k/(2\delta(a, b))$, where $\delta(a, b) = \min\{I(V) : V \in [a, b]\}$ is strictly positive by continuity of I . For every $V_0 \in [a, b]$,

$$I_{V_0}(V) + \lambda V^2 \geq I(V) + \lambda V^2,$$

so we get $I_{V_0}(V) + \lambda V^2$ is concave. We can argue as above to get the same conclusion. \square

In the proof above, we used the following properties of concave functions:

Lemma 20. (a) [Morgan and Johnson 2000] *Let $f : (a, b) \rightarrow \mathbb{R}$ be a continuous function. Then f is concave if and only if for every $x_0 \in (a, b)$ there exists an*

open interval $I_{x_0} \subseteq (a, b)$ of x_0 and a concave smooth function $g_{x_0} : I_{x_0} \rightarrow \mathbb{R}$ such that $g_{x_0}(x_0) = f(x_0)$ and $g_{x_0}(x) \geq f(x)$ for every $x_0 \in I_{x_0}$.

- (b) If $f : (a, b) \rightarrow \mathbb{R}$ is a concave function, then f'_+ and f'_- are monotonic nonincreasing functions and also right and left continuous respectively. Moreover, f'' exists almost everywhere.

Proof. (a) If f is concave, just take g to be linear. If f is not concave, then there exists $\epsilon > 0$, such that $f_\epsilon(x) = f(x) - \epsilon x^2$ is not concave. So we can choose $x_1, x_3 \in (a, b)$, such that the graph of $f_\epsilon(x)$ lies below line $l(x)$ from $(x_1, f_\epsilon(x_1))$ to $(x_3, f_\epsilon(x_3))$. Assume $f_\epsilon(x) - l(x)$ attains its minimum at $x_2 \in (x_1, x_3)$.

By hypothesis, there is a concave smooth $g_{x_2}(x) \geq f(x)$, and $g_{x_2}(x_2) = f(x_2)$. Then $g_\epsilon(x) = g_{x_2}(x) - \epsilon x^2 \geq f_\epsilon(x)$, $g_\epsilon(x_2) = f_\epsilon(x_2)$, so we have that $g_\epsilon(x) - l(x)$ also attain its minimum at $x_2 \in (x_1, x_3)$, which implies $g''_\epsilon(x_2) \geq 0$, but $g''_\epsilon(x_2) = g''_{x_2}(x_2) - 2\epsilon \leq -2\epsilon$, a contradiction.

- (b) It is well known that f'_+ and f'_- are monotonic nonincreasing and f'' exists almost everywhere, so we just prove the right continuity of f'_+ , and left continuity of f'_- follows similarly. For any $x_0 \in (a, b)$, by monotonicity of f'_+ have

$$(5-11) \quad \lim_{x \rightarrow x_0^+} f'_+(x) \leq f'_+(x_0).$$

On the other hand,

$$(5-12) \quad f'_+(x_0) = \lim_{x \rightarrow x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x \rightarrow x_0^+} \frac{\int_{x_0}^x f'_+(t) dt}{x - x_0}$$

where we have used the stronger versions of the fundamental theorem of calculus [Walker 1977]

$$(5-13) \quad f(x) - f(x_0) = \int_{x_0}^x f'_+(t) dt$$

whenever f is continuous and $f'_+ \in L^1$. Again by the monotonicity we have

$$(5-14) \quad f'_+(t) \leq \lim_{x \rightarrow x_0^+} f'_+(x).$$

Combining with (5-12) and (5-14), we get

$$(5-15) \quad f'_+(x_0) \leq \lim_{x \rightarrow x_0^+} \frac{\int_{x_0}^x \lim_{x \rightarrow x_0^+} f'_+(x) dt}{x - x_0} = \lim_{x \rightarrow x_0^+} f'_+(x).$$

Then (5-11) and (5-15) give the right continuity of f'_+ . □

5.3. Monotonicity of m_H^+ . For differentiable points of I we have $H(V) = I'(V)$, so we can replace H with I' in Hawking mass in order to simplify Hawking mass to be a function of only volume. But I may not be differentiable for every volume, and there is a jump for H from I'_+ to I'_- at volumes which are not differentiable. By the compactness of isoperimetric surfaces, see [Meeks et al. 2014], there is a surface which achieves the minimal (maximal) mean curvature enclosing the same volume. So we can define the maximal Hawking mass as:

Definition. Let (M, g) be an AF three-manifold with nonnegative scalar curvature, $\Sigma \subset M$ be a isoperimetric surface of volume V . Then the maximal Hawking mass is $m_H^+(V) = \sqrt{I(V)}(16\pi - I(V)I'_+(V)^2)$.

When I is not differentiable, m_H^+ is the maximal Hawking mass, and it reduces to the ordinary Hawking mass at the differentiable points of I . We have the following result on the monotonicity of m_H^+ :

Lemma 21 [Bray 1997]. *Let (M, g) be an AF three-manifold with nonnegative scalar curvature. Assume for every $V > 0$ there is a connected isoperimetric surface enclosing volume V , and also $I(V)$ is increasing. Then $m_H^+(V)$ is nondecreasing.*

Proof. By Gauss's equation,

$$(5-16) \quad K = \frac{R}{2} - \text{Ric}(n, n) + \frac{1}{2}(H^2 - |A|^2).$$

So we have

$$(5-17) \quad |A|^2 + \text{Ric}(n, n) = \frac{R}{2} - K + \frac{1}{2}(H^2 + |A|^2)$$

by $|A|^2 = |A^0|^2 + \frac{1}{2}H^2$, and $R \geq 0$, so we have

$$(5-18) \quad I''_{V_0}(V) = -\frac{\int_{\Sigma(V)} |A|^2 + \text{Ric}(n, n) d\mu}{I_{V_0}^2(V)} \leq \frac{\int_{\Sigma(V)} K - \frac{3}{4}H^2 d\mu}{I_{V_0}^2(V)}.$$

By the connectedness of $\Sigma(V)$, we have

$$(5-19) \quad \int_{\Sigma(V)} K d\mu = 2\pi \chi(\Sigma(V)) \leq 4\pi.$$

Then

$$(5-20) \quad I''_{V_0}(V) \leq \frac{16\pi - 3I'_{V_0}(V)^2 I_{V_0}(V)}{4I_{V_0}^2(V)}.$$

As we have proved that $I'_+(V)$ is right continuous, so is maximal Hawking mass. Thus it is sufficient to prove $m_H^+(V)$ is weak nondecreasing, i.e., for any $[a, b] \in (0, \infty)$, $\int_a^b m_H^+(V)\phi'(V) dV \leq 0$ for all smooth nonnegative $\phi \in C_c^\infty(a, b)$, $\phi \geq 0$.

The reason to do so is that $m_H^+(V)$ has only countable jump points. Let the difference quotient be defined by

$$\Delta^h F(V) = \frac{1}{h}(F(V+h) - F(V)).$$

Then

$$\begin{aligned}
 (5-21) \quad & \int_a^b m_H^+(V) \phi'(V) dV \\
 &= \int_a^b \sqrt{I(V)} (16\pi - I(V) I_+'^2(V)) \phi'(V) dV \\
 &= \lim_{h \rightarrow 0^+} \int_a^b \sqrt{I(V)} (16\pi - I(V) \Delta^h I(V)^2) \Delta^h \phi(V) dV \\
 &= - \lim_{h \rightarrow 0^+} \int_a^b \Delta^{-h} \{ \sqrt{I(V)} (16\pi - I(V) \Delta^h I(V)^2) \} \phi(V) dV \\
 &= \lim_{h \rightarrow 0^+} \int_a^b \left\{ \phi I^{3/2} \left\{ \Delta^{-h} (\Delta^h I)^2 - I' \frac{16\pi - 3I'^2 I}{2I^2} \right\} \right\} dV,
 \end{aligned}$$

where we use the fact that $I_+' = I_-'$ almost everywhere.

Since $I_{V_0}(V_0) = I(V_0)$, and $I_{V_0}(V) \geq I(V)$, also $I(V)$ is increasing, we get $\Delta^{-h} (\Delta^h I)^2(V_0) \leq \Delta^{-h} (\Delta^h I_{V_0})^2(V_0)$, and $I'_{V_0} \geq 0$, so

$$\begin{aligned}
 (5-22) \quad & \int_a^b m_H^+(V) \phi'(V) dV \\
 &\leq \lim_{h \rightarrow 0^+} \int_a^b \left\{ \phi I_{V_0}^{3/2} \left\{ \Delta^{-h} (\Delta^h I_{V_0})^2 - I'_{V_0} \right\} \frac{16\pi - 3I_{V_0}'^2 I_{V_0}}{2I_{V_0}^2} \right\} dV_{V_0} \\
 &= \int_a^b 2\phi I_{V_0}^{3/2} I'_{V_0} \left\{ I''_{V_0} - \frac{16\pi - 3I_{V_0}'^2 I_{V_0}}{4I_{V_0}^2} \right\} dV_{V_0} \leq 0,
 \end{aligned}$$

where we used (5-20) and Fatou's lemma for the last equality. Hence, $m_H^+(V)$ is nondecreasing. \square

Remark. (1) Hawking mass is also monotonic along the stable CMC foliation as long as the area is nondecreasing; the proof is the same as above.

(2) We can see that the monotonicity of maximal Hawking mass relies heavily on the connectedness of the isoperimetric surface. If the isoperimetric surface has more than one components, Bray [1997] considers the sum of three halves of the area of the components

$$F(V) = \inf \left\{ \sum_i \text{area}(\Sigma_i)^{3/2} : \{\Sigma_i\} \text{ enclose volume } V \text{ outside the horizons} \right\}$$

under the condition the components are disjoint with each other. Then he proved the mass

$$m^+(V) = F(V)^{1/3}(36\pi - F'_+{}^2)/144\pi^{3/2}$$

is nondecreasing. In fact, for F he got the estimate

$$(5-23) \quad F''(V) \leq \frac{36\pi - F'(V)^2}{6F(V)},$$

and then the proof follows as above. The minimizing surfaces are CMC generally with different mean curvatures on each component. When the minimizer of F has only one component it must be an isoperimetric surface. We already know that for large enough volume in AF manifolds the isoperimetric surfaces are spheres close to coordinate spheres and $m^+(V) = m_H^+(V)$; their limits are the ADM mass of the manifold when volume goes to infinity.

Now we are in a position to prove the rigidity of small isoperimetric surfaces:

Proof of Theorem 5. First we claim that

$$(5-24) \quad \lim_{V \rightarrow 0} m_H^+(V) = 0.$$

In fact, by Lemma 18 we know the isoperimetric surface is of sphere type when the volume is small enough. Combined with Lemma 6, we get

$$(5-25) \quad \lim_{V \rightarrow 0} m_H^+(V) \geq 0.$$

By definition,

$$(5-26) \quad m_H^+(V) = \sqrt{I(V)}(16\pi - I(V)I'_+{}^2) \leq 16\pi\sqrt{I(V)},$$

which implies

$$(5-27) \quad \lim_{V \rightarrow 0} m_H^+(V) \leq 0.$$

Thus the claim follows by (5-25) and (5-27).

If there exists an isoperimetric surface Σ with volume $0 < V_0 \leq \delta_0$, such that $m_H^+(V_0) = 0$, then by monotonicity of Lemma 21 for m_H^+ and (5-24), we get

$$(5-28) \quad m_H^+(V) \equiv 0, \quad \text{for any } V \in [0, V_0].$$

Thus

$$(5-29) \quad I(V)I'_+{}^2(V) \equiv 16\pi \quad \text{on } [0, V_0].$$

Since I is continuous by Lemma 19, we get

$$(5-30) \quad I'_+(V) = I'(V) \quad \text{on } [0, V_0].$$

Since there are no compact minimal surfaces, I is increasing, and

$$(5-31) \quad I' = \sqrt{\frac{16\pi}{I}}.$$

Since $I(0) = 0$, we have

$$(5-32) \quad I(V) = (36\pi)^{1/3} V^{2/3} \quad \text{on } [0, V_0].$$

Then by Lemma 15 above we conclude that (M, g) is isometric to \mathbb{R}^3 . \square

Appendix

A.1. Spherical harmonics on S^2 . Write

$$\Delta_{S^2} = \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \varphi^2}.$$

The eigenvalues of $-\Delta_{S^2}$ are $\lambda = l(l+1)$, $l = 0, 1, 2, \dots$; the eigenfunctions are

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!}} \sin^{|m|} \theta P_l^{|m|}(\cos \theta) e^{im\varphi},$$

where $m = -l, \dots, l$, and the $P_l(x)$ are the Legendre polynomials, $P_0(x) = 1$, $P_1(x) = x$, $P_2(x) = \frac{1}{2}(3x^2 - 1)$. The reduction formula is

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

The real form of spherical harmonics are

$$l = 0$$

$$Y_{0,0} = \frac{1}{2} \sqrt{\frac{1}{\pi}},$$

$$l = 1$$

$$Y_{1,0} = \sqrt{\frac{3}{4\pi}} \cos \theta, \quad Y_{1,-1} = \sqrt{\frac{3}{4\pi}} \sin \theta \sin \varphi, \quad Y_{1,1} = \sqrt{\frac{3}{4\pi}} \sin \theta \cos \varphi,$$

$$l = 2$$

$$\begin{aligned} Y_{2,-2} &= \frac{1}{4} \sqrt{\frac{15}{\pi}} \sin^2 \theta \sin 2\varphi, & Y_{2,-1} &= \frac{1}{4} \sqrt{\frac{15}{\pi}} \sin 2\theta \sin \varphi, & Y_{2,0} &= \frac{1}{4} \sqrt{\frac{5}{\pi}} (3\cos^2 \theta - 1), \\ Y_{2,1} &= \frac{1}{4} \sqrt{\frac{15}{\pi}} \sin 2\theta \cos \varphi, & Y_{2,2} &= \frac{1}{4} \sqrt{\frac{15}{\pi}} \sin^2 \theta \cos 2\varphi, \end{aligned}$$

$$l = 4$$

$$\begin{aligned}
Y_{4,-4} &= \frac{3}{16} \sqrt{\frac{35}{\pi}} \sin^4 \theta \sin 4\varphi, & Y_{4,-3} &= \frac{3}{4} \sqrt{\frac{35}{2\pi}} \sin^3 \theta \cos \theta \sin 3\varphi, \\
Y_{4,-2} &= \frac{3}{8} \sqrt{\frac{5}{\pi}} \sin^2 \theta (7 \cos^2 \theta - 1) \sin 2\varphi, & Y_{4,-1} &= \frac{3}{8} \sqrt{\frac{10}{\pi}} \sin \theta \cos \theta (7 \cos^2 \theta - 3) \sin \varphi, \\
Y_{4,0} &= \frac{3}{16} \sqrt{\frac{1}{\pi}} (35 \cos^4 \theta - 30 \cos^2 \theta + 3), & Y_{4,1} &= \frac{3}{8} \sqrt{\frac{10}{\pi}} \sin \theta \cos \theta (7 \cos^2 \theta - 3) \cos \varphi, \\
Y_{4,2} &= \frac{3}{8} \sqrt{\frac{5}{\pi}} \sin^2 \theta (7 \cos^2 \theta - 1) \cos 2\varphi, & Y_{4,3} &= \frac{3}{4} \sqrt{\frac{35}{2\pi}} \sin^3 \theta \cos \theta \cos 3\varphi, \\
Y_{4,4} &= \frac{3}{16} \sqrt{\frac{35}{\pi}} \sin^4 \theta \cos 4\varphi.
\end{aligned}$$

To compute u_2^2 , we need to decompose the following terms into different order spherical harmonics:

$$\begin{aligned}
Y_{2,0}^2 &= \frac{3}{7} \sqrt{\frac{1}{\pi}} Y_{4,0} + \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{2,0} + \frac{1}{4\pi}, \\
Y_{2,-2}^2 &= -\frac{1}{2} \sqrt{\frac{5}{7\pi}} Y_{4,4} + \frac{1}{14} \sqrt{\frac{1}{\pi}} Y_{4,0} - \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{2,0} + \frac{1}{4\pi}, \\
Y_{2,2}^2 &= \frac{1}{2} \sqrt{\frac{5}{7\pi}} Y_{4,4} + \frac{1}{14} \sqrt{\frac{1}{\pi}} Y_{4,0} - \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{2,0} + \frac{1}{4\pi}, \\
Y_{2,-1}^2 &= -\frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{4,2} - \frac{2}{7} \sqrt{\frac{1}{\pi}} Y_{4,0} - \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,2} + \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{2,0} + \frac{1}{4\pi}, \\
Y_{2,1}^2 &= \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{4,2} - \frac{2}{7} \sqrt{\frac{1}{\pi}} Y_{4,0} + \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,2} + \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{2,0} + \frac{1}{4\pi}, \\
Y_{2,-2} Y_{2,2} &= \frac{1}{2} \sqrt{\frac{5}{7\pi}} Y_{4,-4}, & Y_{2,-2} Y_{2,0} &= \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{4,-2} - \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{2,-2}, \\
Y_{2,2} Y_{2,0} &= \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{4,2} - \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{2,2}, & Y_{2,-1} Y_{2,0} &= \frac{1}{7} \sqrt{\frac{15}{2\pi}} Y_{4,-1} + \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{2,-1}, \\
Y_{2,1} Y_{2,0} &= \frac{1}{7} \sqrt{\frac{15}{2\pi}} Y_{4,1} + \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{2,1}, & Y_{2,-1} Y_{2,1} &= \frac{1}{7} \sqrt{\frac{5}{\pi}} Y_{4,-2} + \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,-2}, \\
Y_{2,-2} Y_{2,-1} &= -\frac{1}{2} \sqrt{\frac{5}{14\pi}} Y_{4,3} - \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{4,1} + \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,1}, \\
Y_{2,2} Y_{2,1} &= \frac{1}{2} \sqrt{\frac{5}{14\pi}} Y_{4,3} - \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{4,1} + \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,1}, \\
Y_{2,-2} Y_{2,1} &= \frac{1}{2} \sqrt{\frac{5}{14\pi}} Y_{4,-3} - \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{4,-1} + \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,-1}, \\
Y_{2,2} Y_{2,-1} &= \frac{1}{2} \sqrt{\frac{5}{14\pi}} Y_{4,-3} + \frac{1}{14} \sqrt{\frac{5}{\pi}} Y_{4,-1} + \frac{1}{14} \sqrt{\frac{15}{\pi}} Y_{2,-1}.
\end{aligned}$$

A.2. Existence of isoperimetric surface for all volumes.

Lemma 22 [Carlotto et al. 2016]. *Let (M, g) be a three-manifold with nonnegative scalar curvature, maybe with horizon. Then the isoperimetric surface for all volumes exists.*

Proof. By Theorem 2.1 of [Ritoré and Rosales 2004], we can prove this in the same way as in [Eichmair and Metzger 2013]; for every $V > 0$, there exists an isoperimetric region Ω and a radius $r \geq 0$ such that

$$(A-1) \quad |\Omega|_g + \frac{4}{3}\pi r^3 = V, \quad |\partial\Omega|_g + 4\pi r^2 = I(V).$$

By the isoperimetric inequality of Shi [2016] on nonnegative scalar curvature manifolds, we get for every $r > 0$ that there is a bounded region Ω' with finite perimeter $|\partial\Omega'|_g$ lying arbitrary far out in the asymptotic flat region of (M, g) such that

$$(A-2) \quad |\partial\Omega'|_g = 4\pi r^2, \quad |\Omega'|_g > \frac{4}{3}\pi r^3.$$

If $r > 0$ in (A-1), then there is an Ω' that satisfies (A-2). We consider the region $\Omega \cup \Omega'$. Then

$$(A-3) \quad |\Omega|_g + |\Omega'|_g > V, \quad |\partial\Omega|_g + |\partial\Omega'|_g = I(V).$$

But by the definition of I and the above equality we get

$$(A-4) \quad I(|\Omega|_g + |\Omega'|_g) \leq |\partial\Omega|_g + |\partial\Omega'|_g = I(V).$$

By the fact that I is strictly increasing [Chodosh 2016], we have

$$(A-5) \quad I(|\Omega|_g + |\Omega'|_g) > I(V),$$

a contradiction. Thus $r = 0$, which implies that Ω is the isoperimetric region of volume V . \square

A.3. Continuity of I .

Lemma 23. *I is continuous on AF three-manifold.*

Proof. The proof is from [Flores and Nardulli 2014] for bounded geometry, where they don't have existence of isoperimetric surfaces. We need to prove the upper semicontinuity and lower semicontinuity for I , i.e., for any $V_0 > 0$,

$$(A-6) \quad \limsup_{V \rightarrow V_0^+} I(V) \leq I(V_0), \quad \limsup_{V \rightarrow V_0^-} I(V) \leq I(V_0),$$

$$(A-7) \quad I(V_0) \leq \liminf_{V \rightarrow V_0^+} I(V), \quad I(V_0) \leq \liminf_{V \rightarrow V_0^-} I(V).$$

Upper semicontinuity of I : Given $V_0 > 0$, there is isoperimetric region Ω_0 such that $\text{vol}(\Omega_0) = V_0$, $\text{area}(\partial\Omega_0) = I(V_0)$. For any $V \uparrow V_0$, we can subtract a small geodesic ball $B_r(p)$ such that $\text{vol}(B_r(p)) = V_0 - V$, $\text{vol}(\Omega_0 \setminus B_r(p)) = V$. Thus

$$(A-8) \quad I(V) \leq \text{area}(\partial\Omega_0) + \text{area}(\partial B_r(p)) = I(V_0) + \text{area}(\partial B_r(p)).$$

This implies

$$(A-9) \quad \limsup_{V \rightarrow V_0^+} I(V) \leq \text{area}(\partial\Omega_0) + \lim_{V \rightarrow V_0^+} \text{area}(\partial B_r(p)) = I(V_0).$$

For any $V \downarrow V_0$, we can add a small geodesic ball $B_r(p)$, such that $\text{vol}(B_r(p)) = V - V_0$, $\text{vol}(\Omega_0 \cup B_r(p)) = V$. Thus

$$(A-10) \quad I(V) \leq \text{area}(\partial\Omega_0) + \text{area}(\partial B_r(p)) = I(V_0) + \text{area}(\partial B_r(p)).$$

This implies

$$(A-11) \quad \limsup_{V \rightarrow V_0^-} I(V) \leq \text{area}(\partial\Omega_0) + \lim_{V \rightarrow V_0^-} \text{area}(\partial B_r(p)) = I(V_0).$$

So we get the upper semicontinuity of I from (A-9) and (A-11).

Lower semicontinuity of I : for $V \uparrow V_0$, there exists an isoperimetric region Ω such that $\text{vol}(\Omega) = V$. Adding a small geodesic ball $B_r(p)$ such that $\text{vol}(B_r(p)) = V_0 - V$,

$$(A-12) \quad I(V_0) \leq \text{area}(\partial\Omega) + \text{area}(\partial B_r(p)) = I(V) + \text{area}(\partial B_r(p)).$$

This implies

$$(A-13) \quad I(V_0) \leq \liminf_{V \rightarrow V_0^-} I(V) + \lim_{V \rightarrow V_0^-} \text{area}(\partial B_r(p)) \leq \liminf_{V \rightarrow V_0^-} I(V).$$

For $V \downarrow V_0$, subtract a small geodesic ball $B_r(p)$ such that $\text{vol}(B_r(p)) = V - V_0$, so that

$$(A-14) \quad I(V_0) \leq \text{area}(\partial\Omega) + \text{area}(\partial B_r(p)) = I(V) + \text{area}(\partial B_r(p)).$$

This implies

$$(A-15) \quad I(V_0) \leq \liminf_{V \rightarrow V_0^-} I(V) + \lim_{V \rightarrow V_0^-} \text{area}(\partial B_r(p)) \leq \liminf_{V \rightarrow V_0^-} I(V).$$

The lower semicontinuity follows from (A-13) and (A-15). \square

A.4. Mean curvature of isoperimetric surface.

Lemma 24. *The mean curvatures of all the components for an isoperimetric surface are the same.*

Proof. We know that an isoperimetric surface is stable CMC and the mean curvature is same on each component. This follows by the stability condition when choosing a piecewise constant variation function on each component. Assume $\Sigma = \Sigma_1 \cup \Sigma_2$ is an isoperimetric surface with disjoint components Σ_1 and Σ_2 . If the mean curvature of Σ_1 and Σ_2 are constants H_1 and H_2 , respectively, let

$$(A-16) \quad f = \begin{cases} -|\Sigma_2| & \text{on } \Sigma_1 \\ |\Sigma_1| & \text{on } \Sigma_2. \end{cases}$$

As Σ is an isoperimetric surface, so the first variation formula

$$\begin{aligned} 0 &= \int_{\Sigma} f H = \int_{\Sigma_1 \cup \Sigma_2} f H \\ (A-17) \qquad &= -|\Sigma_2|H_1|\Sigma_1| + |\Sigma_1|H_2|\Sigma_2| = |\Sigma_1||\Sigma_2|(H_2 - H_1). \end{aligned}$$

So $H_1 = H_2$, which implies mean curvature on each component is the same. \square

Acknowledgement

I would like to sincerely thank my advisor, Gang Tian, for his encouragement and constant support. I am very grateful to Yuguang Shi for suggesting this problem and for many stimulating discussions as well as pointing out the hyperbolic case for this paper. Many thanks to Prof. Ahmad El Soufi for an explanation of Theorem 1 in his paper. I also want to thank Jie Qing for pointing out the relation to eigenvalue problems, Jerry Kazdan and Yanyan Li for many discussions on mean field equation, and Qian Wang and Zhongmin Qian for getting me interested in this area. Also I want to thank Wenshuai Jiang and Dongyi Wei for many useful comments and discussions. Additional thanks go to Jianchun Chu, Shiguang Ma, Dong Zhang, Haobin Yu and Wenlong Wang for interesting discussions on the subject.

References

- [Bartnik 2002] R. Bartnik, “Mass and 3-metrics of non-negative scalar curvature”, pp. 231–240 in *Proceedings of the International Congress of Mathematicians, II* (Beijing, 2002), edited by T. Li, Higher Education Press, Beijing, 2002. MR Zbl arXiv
- [Bray 1997] H. L. Bray, *The Penrose inequality in general relativity and volume comparison theorems involving scalar curvature*, Ph.D. thesis, Stanford University, 1997, Available at <https://search.proquest.com/docview/304386501>.
- [Carlotto and Schoen 2016] A. Carlotto and R. Schoen, “Localizing solutions of the Einstein constraint equations”, *Invent. Math.* **205**:3 (2016), 559–615. MR Zbl
- [Carlotto et al. 2016] A. Carlotto, O. Chodosh, and M. Eichmair, “Effective versions of the positive mass theorem”, *Invent. Math.* **206**:3 (2016), 975–1016. MR Zbl
- [Chen and Lin 2003] C.-C. Chen and C.-S. Lin, “Topological degree for a mean field equation on Riemann surfaces”, *Comm. Pure Appl. Math.* **56**:12 (2003), 1667–1727. MR Zbl
- [Chodosh 2016] O. Chodosh, “Large isoperimetric regions in asymptotically hyperbolic manifolds”, *Comm. Math. Phys.* **343**:2 (2016), 393–443. MR Zbl
- [Chodosh et al. 2016] O. Chodosh, M. Eichmair, Y. Shi, and H. Yu, “Isoperimetry, scalar curvature, and mass in asymptotically flat Riemannian 3-manifolds”, preprint, 2016. arXiv
- [Christodoulou and Yau 1988] D. Christodoulou and S.-T. Yau, “Some remarks on the quasi-local mass”, pp. 9–14 in *Mathematics and general relativity* (Santa Cruz, CA, 1986), edited by J. A. Isenberg, Contemp. Math. **71**, American Mathematical Society, Providence, RI, 1988. MR Zbl
- [Ding et al. 1997] W. Ding, J. Jost, J. Li, and G. Wang, “The differential equation $\Delta u = 8\pi - 8\pi h e^u$ on a compact Riemann surface”, *Asian J. Math.* **1**:2 (1997), 230–248. MR Zbl

- [Ding et al. 1998] W. Ding, J. Jost, J. Li, and G. Wang, “An analysis of the two-vortex case in the Chern–Simons Higgs model”, *Calc. Var. Partial Differential Equations* **7**:1 (1998), 87–97. MR Zbl
- [Eichmair and Metzger 2013] M. Eichmair and J. Metzger, “Large isoperimetric surfaces in initial data sets”, *J. Differential Geom.* **94**:1 (2013), 159–186. MR Zbl
- [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 Zbl
- [Flores and Nardulli 2014] A. M. Flores and S. Nardulli, “Continuity and differentiability properties of the isoperimetric profile in complete noncompact Riemannian manifolds with bounded geometry”, preprint, 2014. arXiv
- [Hang and Wang 2006] F. Hang and X. Wang, “Rigidity and non-rigidity results on the sphere”, *Comm. Anal. Geom.* **14**:1 (2006), 91–106. MR Zbl
- [Huiskens and Ilmanen 2001] G. Huiskens and T. Ilmanen, “The inverse mean curvature flow and the Riemannian Penrose inequality”, *J. Differential Geom.* **59**:3 (2001), 353–437. MR Zbl
- [Huiskens and Yau 1996] G. Huiskens and S.-T. Yau, “Definition of center of mass for isolated physical systems and unique foliations by stable spheres with constant mean curvature”, *Invent. Math.* **124**:1-3 (1996), 281–311. MR Zbl
- [Ji et al. 2016] D. Ji, Y. Shi, and B. Zhu, “Exhaustion of isoperimetric regions in asymptotically hyperbolic manifolds with scalar curvature $R \geq -6$ ”, preprint, 2016. arXiv
- [Kazdan and Warner 1974] J. L. Kazdan and F. W. Warner, “Curvature functions for compact 2-manifolds”, *Ann. of Math. (2)* **99** (1974), 14–47. MR Zbl
- [Li 1999] Y. Y. Li, “Harnack type inequality: the method of moving planes”, *Comm. Math. Phys.* **200**:2 (1999), 421–444. MR Zbl
- [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 Zbl
- [Lin 2000] C.-S. Lin, “Topological degree for mean field equations on S^2 ”, *Duke Math. J.* **104**:3 (2000), 501–536. MR Zbl
- [Meeks et al. 2014] W. H. Meeks, III, P. Mira, J. Pérez, and A. Ros, “Isoperimetric domains of large volume in homogeneous three-manifolds”, *Adv. Math.* **264** (2014), 546–592. MR Zbl
- [Miao 2002] P. Miao, “Positive mass theorem on manifolds admitting corners along a hypersurface”, *Adv. Theor. Math. Phys.* **6**:6 (2002), 1163–1182. MR
- [Morgan and Johnson 2000] F. Morgan and D. L. Johnson, “Some sharp isoperimetric theorems for Riemannian manifolds”, *Indiana Univ. Math. J.* **49**:3 (2000), 1017–1041. MR Zbl
- [Neves and Tian 2009] A. Neves and G. Tian, “Existence and uniqueness of constant mean curvature foliation of asymptotically hyperbolic 3-manifolds”, *Geom. Funct. Anal.* **19**:3 (2009), 910–942. MR Zbl
- [Pérez and Ros 2002] J. Pérez and A. Ros, “Properly embedded minimal surfaces with finite total curvature”, pp. 15–66 in *The global theory of minimal surfaces in flat spaces* (Martina Franca, 1999), edited by G. P. Pirola, Lecture Notes in Math. **1775**, Springer, Berlin, 2002. MR Zbl
- [Qing and Tian 2007] J. Qing and G. Tian, “On the uniqueness of the foliation of spheres of constant mean curvature in asymptotically flat 3-manifolds”, *J. Amer. Math. Soc.* **20**:4 (2007), 1091–1110. MR Zbl
- [Ritoré and Rosales 2004] M. Ritoré and C. Rosales, “Existence and characterization of regions minimizing perimeter under a volume constraint inside Euclidean cones”, *Trans. Amer. Math. Soc.* **356**:11 (2004), 4601–4622. MR Zbl

- [Ros 2005] A. Ros, “The isoperimetric problem”, pp. 175–209 in *Global theory of minimal surfaces* (Berkeley, CA, 2001), edited by D. Hoffman, Clay Math. Proc. **2**, American Mathematical Society, Providence, RI, 2005. preliminary version at <http://www.ugr.es/~aros/isoper.pdf>. MR Zbl
- [Shi 2016] Y. Shi, “The isoperimetric inequality on asymptotically flat manifolds with nonnegative scalar curvature”, *Int. Math. Res. Not.* **2016**:22 (2016), 7038–7050. MR
- [Shi and Tam 2002] Y. Shi and L.-F. Tam, “Positive mass theorem and the boundary behaviors of compact manifolds with nonnegative scalar curvature”, *J. Differential Geom.* **62**:1 (2002), 79–125. MR Zbl
- [Shi and Tam 2007] Y. Shi and L.-F. Tam, “Rigidity of compact manifolds and positivity of quasi-local mass”, *Classical Quantum Gravity* **24**:9 (2007), 2357–2366. MR Zbl
- [da Silveira 1987] A. M. da Silveira, “Stability of complete noncompact surfaces with constant mean curvature”, *Math. Ann.* **277**:4 (1987), 629–638. MR Zbl
- [Walker 1977] P. L. Walker, “On Lebesgue integrable derivatives”, *Amer. Math. Monthly* **84**:4 (1977), 287–288. MR

Received April 13, 2017. Revised May 31, 2017.

JIACHENG SUN
BEIJING INTERNATIONAL CENTER FOR MATHEMATICAL RESEARCH
PEKING UNIVERSITY
QUANZHAI
BEIJING, 100871
CHINA
sunxujason@pku.edu.cn

ADDENDUM TO A STRONG MULTIPLICITY ONE THEOREM FOR SL_2

QING ZHANG

Volume **285**:2 (2016), 345–374

We remove the restrictions on the residue characteristic in the strong multiplicity one theorem proved by Chai and Zhang (2016). Similar considerations can remove the restrictions on the residue characteristic in the stability and local converse theorem for unitary groups considered by Zhang (2017).

Introduction

This is an addendum to the paper [Chai and Zhang 2016]. In this addendum, we remove the residue characteristic 2 restriction in the strong multiplicity one theorem for SL_2 [Chai and Zhang 2016, Theorem 4.8]. Following the proof of [Chai and Zhang 2016, Theorem 4.8], it suffices to remove the residue characteristic 2 restriction in the local converse theorem [Chai and Zhang 2016, Theorem 3.10]. The idea is to lift an irreducible representation of $SL_2(F)$ to $GL_2(F)$. Similar considerations can remove the residue characteristic 2 restrictions in the stability and local converse theorem for unitary groups considered by Zhang [2017a; 2017b; 2017c].

1. Strong multiplicity one theorem for SL_2

We start from the following stability and local converse theorem for SL_2 .

Theorem 1.1. *Let F be a p -adic field and ψ be a nontrivial additive character of F , which is also viewed as a character of $N(F)$. Let π and π' be two irreducible ψ -generic representations of $SL_2(F)$ with the same central character:*

- (1) *There exists an integer $l = l(\pi, \pi')$ such that if η is a quasicharacter of F^\times with $\text{cond}(\eta) > l$, then*

$$\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi).$$

- (2) *If $\gamma(s, \pi \times \eta, \psi) = \gamma(s, \pi' \times \eta, \psi)$ for all quasicharacters η , then $\pi \cong \pi'$.*

MSC2010: primary 11F70; secondary 22E50.

Keywords: strong multiplicity one, stability, local converse theorem.

Here the gamma factor $\gamma(s, \pi, \eta, \psi)$ is defined in [Chai and Zhang 2016, §2]. If the residue characteristic of F is not 2, Theorem 1.1 is [Chai and Zhang 2016, Theorem 3.10]. As noted in that paper, the difficulty in the proof when the residue characteristic is 2 comes from the fact that [Chai and Zhang 2016, Lemma 3.3] does not hold in that case. To remedy this, in the following, we lift a representation of $\mathrm{SL}_2(F)$ to $\mathrm{GL}_2(F)$ and give a proof of Theorem 1.1 uniformly. In the following F is p -adic field, \mathcal{O} is the ring of integers of F and \mathcal{P} is the maximal ideal of \mathcal{O} .

Before the proof, we recall some facts related to representations of SL_2 and GL_2 . All of these facts can be found in [Gelbart and Knapp 1982]. Let $\tilde{\pi}$ be an irreducible smooth representation of $\mathrm{GL}_2(F)$, then $\tilde{\pi}|_{\mathrm{SL}_2(F)}$ is a finite direct sum of irreducible representations of $\mathrm{SL}_2(F)$ with multiplicity one. Here the multiplicity one statement can be found in [Adler and Prasad 2006]. Moreover, for any irreducible smooth representation π of $\mathrm{SL}_2(F)$, there exists an irreducible smooth representation $\tilde{\pi}$ of $\mathrm{GL}_2(F)$ such that π is a direct summand of $\tilde{\pi}|_{\mathrm{SL}_2(F)}$.

For a representation π of $\mathrm{SL}_2(F)$ (or $\mathrm{GL}_2(F)$), denote its central character by ω_π .

Lemma 1.2. *Let π and π' be two irreducible representations of $\mathrm{SL}_2(F)$ with the same central character. Then there exist irreducible smooth representations $\tilde{\pi}$ and $\tilde{\pi}'$ of $\mathrm{GL}_2(F)$ with the same central character such that π and π' are direct summands of $\tilde{\pi}|_{\mathrm{SL}_2(F)}$ and $\tilde{\pi}'|_{\mathrm{SL}_2(F)}$, respectively.*

Proof. Let $\tilde{\pi}$ and $\tilde{\pi}'$ be two representations of $\mathrm{GL}_2(F)$ such that π and π' are direct summands of $\tilde{\pi}|_{\mathrm{SL}_2(F)}$ and $\tilde{\pi}'|_{\mathrm{SL}_2(F)}$, respectively. For any quasicharacter χ of F^\times , we have $\chi \otimes \tilde{\pi}|_{\mathrm{SL}_2(F)} = \tilde{\pi}|_{\mathrm{SL}_2(F)}$. Thus π is also a direct summand of $\chi \otimes \tilde{\pi}$. Since the central character of $\chi \otimes \tilde{\pi}$ is $\chi^2 \omega_\pi$, it suffices to find a quasicharacter χ of F^\times such that $\chi^2 = \omega_{\tilde{\pi}'} \omega_\pi^{-1}$. Denote $\eta = \omega_{\tilde{\pi}'} \omega_\pi^{-1}$. Since $\omega_\pi = \omega_{\pi'}$ by assumption, we get $\eta(\pm 1) = 1$. Thus there exists a character χ_1 of $F^{\times 2}$ such that $\eta = \chi_1^2$. Any extension χ of χ_1 to F^\times satisfies the desired property. \square

We now fix $\tilde{\pi}$ and $\tilde{\pi}'$ as in Lemma 1.2. We repeat part of the notations from [Chai and Zhang 2016, §3]. Let $\lambda \in \mathrm{Hom}_{N(F)}(\pi, \psi)$ be a nonzero ψ -Whittaker functional of π . Since π is a direct summand of $\tilde{\pi}$, λ can also be viewed as a ψ -Whittaker functional of $\tilde{\pi}$. Let v be a vector in the space V_π . We consider the Whittaker function

$$W_v(g) = \lambda(\tilde{\pi}(g)v), \quad g \in \mathrm{GL}_2(F).$$

We fix a vector $v \in V_\pi$ such that $W_v(1) = 1$. Similarly, we consider a Whittaker functional λ' of π' and fix a vector $v' \in V_{\pi'}$ such that $W_{v'}(1) = 1$. Let $C = C(v, v')$ be a positive integer such that v and v' are fixed by K_C under the action of $\tilde{\pi}$ and $\tilde{\pi}'$, respectively, where $K_C = I_2 + \mathrm{Mat}_{2 \times 2}(\mathcal{P}^C)$ is the standard congruence subgroup of $\mathrm{GL}_2(F)$ with level C . Recall that we have defined Howe vectors $v_m, v'_m, m \geq C$ associated with v and v' [Chai and Zhang 2016, (3-1)].

Denote by B the upper triangular Borel subgroup of $\mathrm{GL}_2(F)$.

Lemma 1.3. *For $m \geq C$, we have*

$$W_{v_m}(b) = W_{v'_m}(b), \quad \forall b \in B.$$

Proof. This is proved in [Baruch 1995, Corollary 6.2.11]. We give a sketch here. Since $\tilde{\pi}$ and $\tilde{\pi}'$ are both ψ -generic and have the same central character by Lemma 1.2, it suffices to show that $W_{v_m}(\mathrm{diag}(a, 1)) = W_{v'_m}(\mathrm{diag}(a, 1))$ for all $a \in F^\times$. As in the proof of [Chai and Zhang 2016, Corollary 3.4], we can check that $W_{v_m}(\mathrm{diag}(a, 1)) = 1$ if $a \in 1 + \mathcal{P}^m$ and $W_{v_m}(\mathrm{diag}(a, 1)) = 0$ if $a \notin 1 + \mathcal{P}^m$. The same is true for $W_{v'_m}$. The assertion follows. \square

After Lemma 1.3, all of the proof in [Chai and Zhang 2016, §3C] goes through. Part (1) of Theorem 1.1 is proved in this way. From the proof of [Chai and Zhang 2016, Theorem 3.10], if we assume the condition $\gamma(s, \pi, \eta, \psi) = \gamma(s, \pi', \eta, \psi)$ for all quasicharacter η of F^\times , we can obtain that

$$W_{v_m}(h) = W_{v'_m}(h), \quad \forall h \in \mathrm{SL}_2(F).$$

Since $\tilde{\pi}(h)v_m = \pi(h)v_m$ and $\tilde{\pi}'(h)v'_m = \pi'(h)v'_m$, we get

$$\lambda(\pi(h)v_m) = \lambda'(\pi'(h)v'_m), \quad \forall h \in \mathrm{SL}_2(F).$$

Then by the uniqueness of the Whittaker models, we get $\pi \cong \pi'$. This proves Theorem 1.1.

From the proof of the strong multiplicity one theorem given in [Chai and Zhang 2016], one can see that the restriction on the residue characteristic 2 in [Chai and Zhang 2016, Theorem 4.8] can be removed. We record the theorem for completeness.

Theorem 1.4. *Let F be a number field and \mathbb{A} be its ring of adeles. Let N be the upper triangular unipotent subgroup of SL_2 . Let ψ be a nontrivial additive character of $F \setminus \mathbb{A}_F$ which is also viewed as a character of $N(F) \setminus N(\mathbb{A})$. Let S be a finite set of finite places of F . Let $\pi = \otimes_v \pi_v$ and $\pi' = \otimes_v \pi'_v$ be two irreducible cuspidal automorphic representations of $\mathrm{SL}_2(\mathbb{A}_F)$ with the same central character. If π and π' are both generic with respect to ψ and $\pi_v \cong \pi'_v$ for all $v \notin S$, then $\pi = \pi'$.*

2. Stability and local converse theorem for unitary groups

Using the same trick as the $\mathrm{SL}(2)$ case in §1, one could remove the restrictions of the residue characteristic in the local converse theorem for $\mathrm{U}(1, 1)$ and $\mathrm{U}(2, 2)$ in [Zhang 2017a; 2017b] and the stability results for $\mathrm{U}(n, n)$ in [Zhang 2017c]. To be more precise, we introduce the following notations. Let F be a p -adic field and E/F be a quadratic field extension. Let $\mathrm{U}_{E/F}(n, n)$ be the unitary group defined

by the skew-Hermitian form $J_n = \begin{pmatrix} & I_n \\ -I_n & \end{pmatrix}$, where I_n is the $n \times n$ identity matrix. Let π be an irreducible smooth generic representation of $\mathrm{U}_{E/F}(n, n)$ and τ be an irreducible smooth generic representation of $\mathrm{GL}_k(E)$ with $k \leq n$. Then one can define a local gamma factor $\gamma(s, \pi \times \tau, \psi)$, where ψ is a fixed additive character of F . These local gamma factors come from the local functional equations of the local zeta integrals considered in [Ben-Artzi and Soudry 2009]. See [Zhang 2017a; 2017b] for some details when $n = 1, 2$ and [Zhang 2017c] for some details for the gamma factors for $\mathrm{U}_{E/F} \times \mathrm{GL}_1(E)$.

Theorem 2.1. (1) *Suppose that $n = 1, 2$. Let π and π' be two irreducible smooth generic representations of $\mathrm{U}_{E/F}(n, n)$ with the same central character and*

$$\gamma(s, \pi \times \tau, \psi) = \gamma(s, \pi' \times \tau, \psi)$$

for all irreducible smooth generic representations of $\mathrm{GL}_k(E)$ for all $k \leq n$. Then $\pi \cong \pi'$.

(2) *Suppose that n is an arbitrary positive integer. Let π and π' be two irreducible smooth generic representations of $\mathrm{U}_{E/F}(n, n)$ with the same central character. Then there exists a positive integer $l := l(\pi, \pi')$ such that for all quasicharacters η of E^\times with $\mathrm{cond}(\eta) > l$, one has*

$$\gamma(s, \pi \times \eta, \psi) = \gamma(s, \pi' \times \eta, \psi).$$

If E/F is unramified, or E/F is ramified but the residue characteristic of F is not 2, part (1) of the above theorem is proved in [Zhang 2017a; 2017b] and part (2) is proved in [Zhang 2017c]. To prove the general case, one can embed the representation π of $\mathrm{U}_{E/F}(n, n)$ into the similitude unitary group $\mathrm{GU}_{E/F}(n, n)$, where

$$\mathrm{GU}_{E/F}(n, n) = \{g \in \mathrm{GL}_{2n}(E) : {}^t \bar{g} J_n g = \lambda J_n, \lambda \in F^\times\}.$$

Here $x \mapsto \bar{x}$ is the nontrivial element in the Galois group of E/F . Note that the center of $\mathrm{GU}_{E/F}$ is E^\times and the center of $\mathrm{U}_{E/F}(n, n)$ is E^1 , where E^1 is the norm one element in E^\times .

The theory in [Gelbart and Knapp 1982] also works for the pair $\mathrm{GU}_{E/F}(n, n)$, $\mathrm{U}_{E/F}(n, n)$. In particular, the restriction of an irreducible smooth representation $\tilde{\pi}$ of $\mathrm{GU}_{E/F}(n, n)$ to $\mathrm{U}_{E/F}(n, n)$ is semisimple; on the other hand, for any irreducible smooth representation π of $\mathrm{U}_{E/F}(n, n)$, there exists an irreducible smooth representation $\tilde{\pi}$ of $\mathrm{GU}_{E/F}(n, n)$, such that π is a constituent of $\tilde{\pi}|_{\mathrm{U}_{E/F}(n, n)}$. As in Lemma 1.2, one has:

Lemma 2.2. *Let π and π' be two irreducible smooth representations of $\mathrm{U}_{E/F}(n, n)$ with the same central character. Then there exist irreducible smooth representations $\tilde{\pi}$ and $\tilde{\pi}'$ of $\mathrm{GU}_{E/F}(n, n)$ with the same central character such that π and π' are constituents of $\tilde{\pi}|_{\mathrm{U}_{E/F}(n, n)}$ and $\tilde{\pi}'|_{\mathrm{U}_{E/F}(n, n)}$, respectively.*

Proof. Let $\text{Sim} : \text{GU}_{E/F}(n, n) \rightarrow F^\times$ be the similitude character of the group $\text{GU}_{E/F}(n, n)$. Via the composition $\chi \circ \text{Sim}$, a quasicharacter χ of F^\times can be viewed as a character of $\text{GU}_{E/F}(n, n)$. Let $\tilde{\pi}$ be an irreducible smooth representation of $\text{GU}_{E/F}(n, n)$ with central character $\omega_{\tilde{\pi}}$. Then we can consider the representation $\chi \otimes \tilde{\pi}$. Note that $\chi \otimes \tilde{\pi}|_{\text{U}(n, n)} = \tilde{\pi}|_{\text{U}(n, n)}$ and the central character of $\chi \otimes \tilde{\pi}$ is $(\chi \circ \text{Nm}_{E/F})\omega_{\tilde{\pi}}$. Denote $\eta = \omega_{\tilde{\pi}}\omega_{\tilde{\pi}}^{-1}$. As in the proof of Lemma 1.2, it suffices to show that there is a character of χ of F^\times such that $\chi \circ \text{Nm}_{E/F} = \eta$. This follows from the fact that $\eta|_{E^\times} = 1$. \square

To proceed, we need the following multiplicity one property.

Proposition 2.3. *Let π be an irreducible generic representation of $\text{U}_{E/F}(n, n)$ and let $\tilde{\pi}$ be an irreducible representation of $\text{GU}_{E/F}(n, n)$. Then*

$$\dim \text{Hom}_{\text{U}_{E/F}(n, n)}(\tilde{\pi}, \pi) \leq 1.$$

Remark. For the pairs (GL, SL), (GSp, Sp) and (GO, O) (similitude orthogonal group and orthogonal group), similar results were proved in [Adler and Prasad 2006] without the genericity assumption.

Proof. Let U be the maximal unipotent subgroup of a fixed Borel subgroup of $\text{U}_{E/F}(n, n)$ and let ψ be a generic character of U . We assume that π is ψ -generic. Since U is also the maximal unipotent subgroup of a Borel subgroup of $\text{GU}_{E/F}(n, n)$, we have $\dim \text{Hom}_U(\tilde{\pi}, \psi) \leq 1$ by the uniqueness of Whittaker model for $\tilde{\pi}$. If π appears as a constituent of $\tilde{\pi}|_{\text{U}_{E/F}(n, n)}$, a Whittaker functional of π gives a Whittaker functional of $\tilde{\pi}$. Now the assertion follows from the uniqueness of Whittaker model of $\tilde{\pi}$. \square

After Lemma 2.2 and Proposition 2.3, the proof of Theorem 2.1 reduces the cases considered in [Zhang 2017a; 2017b; 2017c] as in the SL_2 case sketched in §1.

Finally, we remark that Theorem 2.1 also holds for symplectic groups with the same proof.

Acknowledgement

I would like to thank the anonymous referee for the careful reading of this draft and many useful suggestions.

References

- [Adler and Prasad 2006] J. D. Adler and D. Prasad, “On certain multiplicity one theorems”, *Israel J. Math.* **153** (2006), 221–245. MR Zbl
- [Baruch 1995] E. M. Baruch, *Local factors attached to representations of p -adic groups and strong multiplicity one*, Ph.D. thesis, Yale University, 1995, <http://tinyurl.com/baruchthesis>.

- [Ben-Artzi and Soudry 2009] A. Ben-Artzi and D. Soudry, “ L -functions for $U_m \times R_{E/F}GL_n$ ($n \leq [\frac{m}{2}]$)”, pp. 13–59 in *Automorphic forms and L -functions, I: Global aspects*, edited by D. Ginzburg et al., Contemp. Math. **488**, American Mathematical Society, Providence, RI, 2009. MR Zbl
- [Chai and Zhang 2016] J. Chai and Q. Zhang, “A strong multiplicity one theorem for SL_2 ”, *Pacific J. Math.* **285**:2 (2016), 345–374. MR Zbl
- [Gelbart and Knapp 1982] S. S. Gelbart and A. W. Knapp, “ L -indistinguishability and R groups for the special linear group”, *Adv. in Math.* **43**:2 (1982), 101–121. MR Zbl
- [Zhang 2017a] Q. Zhang, “A local converse theorem for $U(1, 1)$ ”, *Int. J. Number Theory* **13**:8 (2017), 1931–1981. MR Zbl
- [Zhang 2017b] Q. Zhang, “A local converse theorem for $U(2, 2)$ ”, *Forum Math.* (online publication January 2017).
- [Zhang 2017c] Q. Zhang, “Stability of Rankin–Selberg gamma factors for $Sp(2n)$, $\tilde{Sp}(2n)$ and $U(n, n)$ ”, *Int. J. Number Theory* **13**:9 (2017), 2393–2432.

Received December 21, 2016. Revised March 1, 2017.

QING ZHANG

qingzhang0@gmail.com

DEPARTMENT OF MATHEMATICS

THE OHIO STATE UNIVERSITY

COLUMBUS, OH 43210-1174

UNITED STATES

CONTENTS

Volume 292, no. 1 and no. 2

Luis J. Alías , Henrique F. de Lima and Fábio R. dos Santos: <i>New characterizations of linear Weingarten spacelike hypersurfaces in the de Sitter space</i>	1
Henning Haahr Andersen , Catharina Stroppel and Daniel Tubbenhauer: <i>Cellular structures using U_q-tilting modules</i>	21
David Bachman , Ryan Derby-Talbot and Eric Sedgwick: <i>Locally helical surfaces have bounded twisting</i>	257
Hari Bercovici , Jiun-Chau Wang and Ping Zhong: <i>Superconvergence to freely infinitely divisible distributions</i>	273
Michel Boileau , Yeonhee Jang and Richard Weidmann: <i>Meridional rank and bridge number for a class of links</i>	61
Akshay L. Chanillo , Sagun Chanillo and Ali Maalaoui: <i>Norm constants in cases of the Caffarelli–Kohn–Nirenberg inequality</i>	293
Sagun Chanillo with Akshay L. Chanillo and Ali Maalaoui	293
Alex Chirvasitu , S. Paul Smith and Liang Ze Wong: <i>Noncommutative geometry of homogenized quantum $\mathfrak{sl}(2, \mathbb{C})$</i>	305
Julie N. Crenshaw , Alexandra K. Echart and Kirk E. Lancaster: <i>A generalization of “Existence and behavior of the radial limits of a bounded capillary surface at a corner”</i>	355
Christophe Cuny and Michel Weber: <i>Pointwise convergence of almost periodic Fourier series and associated series of dilates</i>	81
Ryan Derby-Talbot with David Bachman and Eric Sedgwick	257
Alexandra K. Echart with Julie N. Crenshaw and Kirk E. Lancaster	355
Daniel Goldstein and Murray Schacher: <i>Norms in central simple algebras</i>	373
Joseph Gubeladze and Mateusz Michałek: <i>The poset of rational cones</i>	103
Qi Guo and Gabor Toth: <i>Dual mean Minkowski measures and the Grünbaum conjecture for affine diameters</i>	117
Fei Hou , Ingo Witt and Huicheng Yin: <i>Global existence and blowup of smooth solutions of 3-D potential equations with time-dependent damping</i>	389
Yeonhee Jang with Michel Boileau and Richard Weidmann	61
Kirk E. Lancaster with Julie N. Crenshaw and Alexandra K. Echart	355

Bernard Le Stum and Adolfo Quirós: <i>Formal confluence of quantum differential operators</i>	427
Jaepil Lee : <i>Bordered Floer homology of $(2, 2n)$-torus link complement</i>	139
Henrique F. de Lima with Luis J. Alías and Fábio R. dos Santos	1
Levi Lopes de Lima : <i>A Feynman–Kac formula for differential forms on manifolds with boundary and geometric applications</i>	177
Ali Maalaoui with Akshay L. Chanillo and Sagun Chanillo	293
Mateusz Michalek with Joseph Gubeladze	103
Sebastien Palcoux : <i>Ore’s theorem on cyclic subfactor planar algebras and beyond</i>	203
Adolfo Quirós with Bernard Le Stum	427
Fábio R. dos Santos with Luis J. Alías and Henrique F. de Lima	1
Murray Schacher with Daniel Goldstein	373
Eric Sedgwick with David Bachman and Ryan Derby-Talbot	257
John Shareshian and Russ Woodroffe: <i>Divisibility of binomial coefficients and generation of alternating groups</i>	223
Alexander S. Sivatski : <i>On rational points of certain affine hypersurfaces</i>	239
S. Paul Smith with Alex Chirvasitu and Liang Ze Wong	305
Catharina Stroppel with Henning Haahr Andersen and Daniel Tubbenhauer	21
Jiacheng Sun : <i>Rigidity of Hawking mass for surfaces in three manifolds</i>	479
Gabor Toth with Qi Guo	117
Daniel Tubbenhauer with Henning Haahr Andersen and Catharina Stroppel	21
Jun-Chau Wang with Hari Bercovici and Ping Zhong	273
Michel Weber with Christophe Cuny	81
Richard Weidmann with Michel Boileau and Yeonhee Jang	61
Ingo Witt with Fei Hou and Huicheng Yin	389
Liang Ze Wong with Alex Chirvasitu and S. Paul Smith	305
Russ Woodroffe with John Shareshian	223
Huicheng Yin with Fei Hou and Ingo Witt	389
Qing Zhang : <i>Addendum to “A strong multiplicity one theorem for SL_2”</i>	505
Ping Zhong with Hari Bercovici and Jun-Chau Wang	273

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX , but papers in other varieties of \TeX , and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of \BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, 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 292 No. 2 February 2018

Locally helical surfaces have bounded twisting	257
DAVID BACHMAN, RYAN DERBY-TALBOT and ERIC SEDGWICK	
Superconvergence to freely infinitely divisible distributions	273
HARI BERCOVICI, JIUN-CHAU WANG and PING ZHONG	
Norm constants in cases of the Caffarelli–Kohn–Nirenberg inequality	293
AKSHAY L. CHANILLO, SAGUN CHANILLO and ALI MAALAOUI	
Noncommutative geometry of homogenized quantum $\mathfrak{sl}(2, \mathbb{C})$	305
ALEX CHIRVASITU, S. PAUL SMITH and LIANG ZE WONG	
A generalization of “Existence and behavior of the radial limits of a bounded capillary surface at a corner”	355
JULIE N. CRENSHAW, ALEXANDRA K. ECHART and KIRK E. LANCASTER	
Norms in central simple algebras	373
DANIEL GOLDSTEIN and MURRAY SCHACHER	
Global existence and blowup of smooth solutions of 3-D potential equations with time-dependent damping	389
FEI HOU, INGO WITT and HUICHENG YIN	
Formal confluence of quantum differential operators	427
BERNARD LE STUM and ADOLFO QUIRÓS	
Rigidity of Hawking mass for surfaces in three manifolds	479
JIACHENG SUN	
Addendum to “A strong multiplicity one theorem for SL_2 ”	505
QING ZHANG	



0030-8730(201802)292:2;1-X