

*Pacific
Journal of
Mathematics*

Volume 262 No. 1

March 2013

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

EDITORS

V. S. Varadarajan (Managing Editor)
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
pacific@math.ucla.edu

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

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

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

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

Robert Finn
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
finn@math.stanford.edu

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

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

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

Jie Qing
Department of Mathematics
University of California
Santa Cruz, CA 95064
qing@cats.ucsc.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 2013 is US \$400/year for the electronic version, and \$485/year for print and electronic. Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by [Mathematical Reviews](#), [Zentralblatt MATH](#), [PASCAL CNRS Index](#), [Referativnyi Zhurnal](#), [Current Mathematical Publications](#) and the [Science Citation Index](#).

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

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

PUBLISHED BY

 **mathematical sciences publishers**
nonprofit scientific publishing

<http://msp.org/>

© 2013 Mathematical Sciences Publishers

ON THE SECOND K -GROUP OF A RATIONAL FUNCTION FIELD

KARIM JOHANNES BECHER AND MÉLANIE RACZEK

We give an optimal bound on the minimal length of a sum of symbols in the second Milnor K -group of a rational function field in terms of the degree of the ramification.

1. Introduction

Let E be an arbitrary field and F the function field of the projective line \mathbb{P}_E^1 over E . For $m \in \mathbb{N}$, there is a well-known exact sequence

$$(1.1) \quad 0 \longrightarrow K_2^{(m)} E \longrightarrow K_2^{(m)} F \xrightarrow{\partial} \bigoplus_{x \in \mathbb{P}_E^{1(1)}} K_1^{(m)} E(x) \longrightarrow K_1^{(m)} E \longrightarrow 0,$$

due to Milnor and Tate; see [Milnor 1970, (2.3)]. Here, $K_1^{(m)}$ and $K_2^{(m)}$ are the functors that associate to a field its first and second K -groups modulo m , respectively, and $\mathbb{P}_E^{1(1)}$ is the set of closed points of \mathbb{P}_E^1 . The map ∂ is called the *ramification map*. By [Gille and Szamuely 2006, (7.5.4)], for m prime to the characteristic of E , the sequence (1.1) translates into a sequence in Galois cohomology, and the proof of its exactness essentially goes back to [Faddeev 1951].

In this article we study how for a given element ρ in the image of ∂ one finds a good $\xi \in K_2^{(m)} F$ with $\partial(\xi) = \rho$. Our main result [Theorem 3.10](#) states that there is such a ξ that is a sum of r symbols (canonical generators of $K_2^{(m)} F$) where r is bounded by half the degree of the support of ρ . This generalizes results from [Kunyavskii et al. 2006; Rowen et al. 2005; Sivatski 2007], where the problem has been studied in terms of Brauer groups in the presence of a primitive m -th root of unity in E for $m > 0$. Developing further an idea in [Sivatski 2007, Proposition 2], we provide examples ([Example 4.3](#)) where the bound on r cannot be improved.

This work was done while Becher was a Fellow of the Zukunftskolleg and Raczek was a Postdoctoral Fellow of the Fonds de la Recherche Scientifique – FNRS. The project was further supported by the Deutsche Forschungsgemeinschaft (project “Quadratic Forms and Invariants”, BE 2614/3).

MSC2010: 12Y05, 12E30, 12G05, 19D45.

Keywords: Milnor K -theory, field extension, valuation, ramification.

2. Milnor K -theory of a rational function field

We recall the basic terminology of K -theory for fields as introduced in [Milnor 1970], with slightly different notation. Let F be a field. For $m, n \in \mathbb{N}$, let $K_n^{(m)} F$ denote the abelian group generated by elements called *symbols*, which are of the form $\{a_1, \dots, a_n\}$ with $a_1, \dots, a_n \in F^\times$, subject to the defining relations that $\{\cdot, \dots, \cdot\} : (F^\times)^n \rightarrow K_n^{(m)} F$ is a multilinear map, that $\{a_1, \dots, a_n\} = 0$ whenever $a_i + a_{i+1} = 1$ in F for some $i < n$, and that $m \cdot \{a_1, \dots, a_n\} = 0$. For $a, b \in F^\times$ we have $\{ab\} = \{a\} + \{b\}$ in $K_1^{(m)} F$. The second relation above is void when $n = 1$, hence $K_1^{(m)} F$ is the same as $F^\times / F^{\times m}$, only with different notation for the elements and the group operation. As shown in [Milnor 1970, (1.1) and (1.3)], it follows from the defining relations that, for $a_1, \dots, a_n \in F^\times$, we have $\{a_{\sigma(1)}, \dots, a_{\sigma(n)}\} = \varepsilon \{a_1, \dots, a_n\}$ for any permutation σ of the numbers $1, \dots, n$ with signature $\varepsilon = \pm 1$, and furthermore $\{a_1, \dots, a_n\} = 0$ whenever $a_i + a_{i+1} = 0$ for some $i < n$.

With this notation, $K_n^{(0)} F$ is the full Milnor K -group $K_n F$ introduced in [Milnor 1970], and $K_n^{(m)} F$ is its quotient modulo m for $m \geq 1$.

By a \mathbb{Z} -valuation we mean a valuation with value group \mathbb{Z} . Given a \mathbb{Z} -valuation v on F we denote by \mathbb{O}_v its valuation ring and by κ_v its residue field. For $a \in \mathbb{O}_v$ let \bar{a} denote the natural image of a in κ_v . By [ibid., (2.1)], for $n \geq 2$ and a \mathbb{Z} -valuation v on F , there is a unique homomorphism $\partial_v : K_n^{(m)} F \rightarrow K_{n-1}^{(m)} \kappa_v$ such that

$$\partial_v(\{f, g_2, \dots, g_n\}) = v(f) \cdot \{\bar{g}_2, \dots, \bar{g}_n\} \quad \text{for } f \in F^\times \text{ and } g_2, \dots, g_n \in \mathbb{O}_v^\times.$$

When $n = 2$, for $f, g \in F^\times$ we have $f^{-v(g)} g^{v(f)} \in \mathbb{O}_v^\times$ and

$$\partial_v(\{f, g\}) = \{(-1)^{v(f)v(g)} \overline{f^{-v(g)} g^{v(f)}}\} \quad \text{in } K_1^{(m)} \kappa_v.$$

We turn to the situation where F is the function field of \mathbb{P}^1 over E . By the choice of a generator, we identify F with the rational function field $E(t)$ in the variable t over E . Let \mathcal{P} denote the set of monic irreducible polynomials in $E[t]$. Any $p \in \mathcal{P}$ determines a \mathbb{Z} -valuation v_p on $E(t)$ that is trivial on E and such that $v_p(p) = 1$. There is further a unique \mathbb{Z} -valuation v_∞ on $E(t)$ such that $v_\infty(f) = -\deg(f)$ for any $f \in E[t] \setminus \{0\}$. We set $\mathcal{P}' = \mathcal{P} \cup \{\infty\}$. For $p \in \mathcal{P}'$ we write ∂_p for ∂_{v_p} and we denote by E_p the residue field of v_p . Note that E_p is naturally isomorphic to $E[t]/(p)$ for $p \in \mathcal{P}$, and E_∞ is naturally isomorphic to E .

It follows from [ibid., Section 2] that the sequence

$$(2.1) \quad 0 \longrightarrow K_n^{(m)} E \longrightarrow K_n^{(m)} E(t) \xrightarrow{\bigoplus \partial_p} \bigoplus_{p \in \mathcal{P}} K_{n-1}^{(m)} E_p \longrightarrow 0$$

is split exact. We are going to reformulate this fact for $n = 2$ and to relate the sequences (2.1) and (1.1). We set

$$\mathfrak{R}'_m(E) = \bigoplus_{p \in \mathcal{P}'} K_1^{(m)} E_p.$$

For $p \in \mathcal{P}'$, the norm map of the finite extension E_p/E yields a group homomorphism $K_1^{(m)} E_p \rightarrow K_1^{(m)} E$. Summation over these maps for all $p \in \mathcal{P}'$ yields a homomorphism $N: \mathfrak{R}'_m(E) \rightarrow K_1^{(m)} E$. Let $\mathfrak{R}_m(E)$ denote the kernel of N . We set $\partial = \bigoplus_{p \in \mathcal{P}'} \partial_p$. By [Gille and Szamuely 2006, (7.2.4) and (7.2.5)] we obtain an exact sequence

$$(2.2) \quad 0 \longrightarrow K_2^{(m)} E \longrightarrow K_2^{(m)} E(t) \xrightarrow{\partial} \mathfrak{R}'_m(E) \xrightarrow{N} K_1^{(m)} E \longrightarrow 0.$$

In particular, $\mathfrak{R}_m(E)$ is equal to the image of $\partial: K_2^{(m)} E(t) \rightarrow \mathfrak{R}'_m(E)$.

The choice of the generator of F over E fixes a bijection $\phi: \mathbb{P}_E^{1(1)} \rightarrow \mathcal{P}'$ and for any $x \in \mathbb{P}_E^{1(1)}$ a natural isomorphism between $E(x)$ and $E_{\phi(x)}$. This identifies $\bigoplus_{x \in \mathbb{P}_E^{1(1)}} K_1^{(m)} E(x)$ with $\mathfrak{R}'_m(E)$, and further the sequence (1.1) with (2.2). We will work with (2.2) in the sequel.

For $\rho = (\rho_p)_{p \in \mathcal{P}'} \in \mathfrak{R}'_m(E)$ we denote $\text{Supp}(\rho) = \{p \in \mathcal{P}' \mid \rho_p \neq 0\}$ and $\deg(\rho) = \sum_{p \in \text{Supp}(\rho)} [E_p : E]$, and call this the *support* and the *degree* of ρ . The degree of an element of $\mathfrak{R}'_m(E)$ is invariant under automorphisms of $E(t)/E$.

3. Bound by representation by symbols in terms of the degree

In this section we study the relation between the degree of $\rho \in \mathfrak{R}_m(E)$ and the properties of elements $\xi \in K_2^{(m)} E(t)$ with $\partial(\xi) = \rho$. In Theorem 3.10 we will show that there always exists such ξ that is a sum of r symbols where r is the integral part of $\deg(\rho)/2$. In particular, any ramification of degree at most three is realized by a symbol. This settles a question in [Kunyavskiĭ et al. 2006, (2.5)]. In some of the following statements, we consider elements of $\mathfrak{R}'_m(E)$, rather than only of $\mathfrak{R}_m(E)$.

Proposition 3.1. *If $\rho \in \mathfrak{R}_m(E)$ then $\deg(\rho) \neq 1$.*

Proof. Consider an element $\rho \in \mathfrak{R}'_m(E)$ with $\deg(\rho) = 1$. The support of ρ consists of one rational point $p \in \mathcal{P}'$. Hence $N(\rho) = \rho_p \neq 0$ in $K_1^{(m)} E$, whereby $\rho \notin \mathfrak{R}_m(E)$. \square

We say that $p \in \mathcal{P}'$ is *rational* if $[E_p : E] = 1$. We call a subset of \mathcal{P}' *rational* if all its elements are rational. We give two examples showing how to realize a given ramification of small degree and with rational support by one symbol.

Examples 3.2. (1) Let $a, c \in E^\times$ and $c \notin E^{\times m}$. The symbol $\sigma = \{t - a, c\}$ in $K_2^{(m)} E(t)$ satisfies $\text{Supp}(\sigma) = \{t - a, \infty\}$, $\partial_{t-a}(\sigma) = \{c\}$ and $\partial_\infty(\sigma) = \{c^{-1}\}$.

(2) For $a_1, a_2, c_1, c_2 \in E^\times$ with $a_1 \neq a_2$, we compute the ramification of the symbol

$$\sigma = \left\{ \frac{t - a_1}{c_2(a_2 - a_1)}, \frac{c_1(t - a_2)}{a_1 - a_2} \right\}$$

in $K_2^{(m)}E(t)$. It has $\text{Supp}(\sigma) \subseteq \{t - a_1, t - a_2, \infty\}$, $\partial_{t-a_i}(\sigma) = \{c_i\}$ for $i = 1, 2$, and $\partial_\infty(\sigma) = \{(c_1c_2)^{-1}\}$.

A ramification of degree two can, under some extra conditions, be realized by a symbol one of whose entries is a constant:

Proposition 3.3. *Let $\rho \in \mathfrak{R}_m(E)$ be such that $\deg(\rho) = 2$. If $\text{Supp}(\rho)$ is rational or $\text{char}(E) \neq m = 2$, there exist $e \in E^\times$ and $f \in E(t)^\times$ such that $\rho = \partial(\{e, f\})$.*

Proof. Suppose first that the support of ρ is rational. We choose $a, e \in E^\times$ such that $t - a \in \text{Supp}(\rho)$ and $\rho_{t-a} = \{e\}$ in $K_1^{(m)}E$. Then $\text{Supp}(\rho) = \{t - a, p\}$ where $p \in \mathcal{P}'$ is rational. As $N(\rho) = 0$ we obtain that $\rho_p = \{e^{-1}\}$ in $K_1^{(m)}E_p$. If $p = \infty$, we set $f = 1/(t - a)$. Otherwise $p = t - b$ for some $b \in E$, and we set $f = (t - b)/(t - a)$. In either case we obtain $\rho = \partial(\{e, f\})$.

It remains to consider the case where $\text{char}(E) \neq m = 2$ and $\text{Supp}(\rho) = \{p\}$ for a quadratic polynomial $p \in \mathcal{P}$. Then E_p/E is a separable quadratic extension. Let $x \in E_p^\times$ be such that $\rho_p = \{x\}$. As $\text{Supp}(\rho) = \{p\}$ and $N(\rho) = 0$, we obtain that the norm of x with respect to the extension E_p/E lies in $E^{\times 2}$, and therefore $x E_p^{\times 2} = e E_p^{\times 2}$ for some $e \in E^\times$; see [Lam 2005, Chapter VII, (3.9)]. Hence, $\rho_p = \{x\} = \{e\}$ in $K_1^{(2)}E_p$, and we obtain $\rho = \partial(\{e, p\})$. \square

In Proposition 3.3 the rationality of the support when $m \neq 2$ is not a superfluous condition; the following example was pointed out to us by J.-P. Tignol.

Example 3.4. Let k be a field. We consider the rational function field in two variables u and v over k . Let τ denote the k -automorphism of $k(u, v)$ satisfying $\tau(u) = v$ and $\tau(v) = u$. Then τ^2 is the identity map on $k(u, v)$, and $E = \{x \in k(u, v) \mid \tau(x) = x\}$ is a subfield of $k(u, v)$ such that $[k(u, v) : E] = 2$. Consider the element $y = v/u \in k(u, v)$. Since $y \notin E$, the quadratic polynomial

$$p = (t - y)(t - \tau(y)) = t^2 - \frac{u^2 + v^2}{uv}t + 1$$

is irreducible over E .

Let m be an odd positive integer. We consider the symbol $\sigma = \{p, t\}$ in $K_2^{(m)}E(t)$. Note that the support of $\partial(\sigma)$ is contained in $\{p\}$ and $\partial_p(\sigma) = \{\bar{t}\}$. Moreover, mapping t to y induces an E -isomorphism $E_p \rightarrow k(u, v)$. Since y is not an m -th power in $k(u, v)$, it follows that $\partial_p(\sigma) \neq 0$. Hence, $\text{Supp}(\partial(\sigma)) = \{p\}$ and $\deg(\partial(\sigma)) = 2$.

We claim that $\partial_p(\sigma) \neq \partial_p(\{e, f\})$ for any $e \in E^\times$ and $f \in E(t)^\times$. Suppose on the contrary that there exist $e \in E^\times$ and $f \in E(t)^\times$ such that $\partial_p(\sigma) = \partial_p(\{e, f\})$. Then we obtain that $e^{v_p(f)}y$ is an m -th power in $k(u, v)$, and taking norms with respect to the extension $k(u, v)/E$ yields that $e^{2v_p(f)} \in E^{\times m}$. Since m is odd, it follows that $e^{v_p(f)} \in E^{\times m}$, and thus $\partial_p(\{e, f\}) = 0$, a contradiction.

The remainder of this section builds up to our main result, [Theorem 3.10](#).

Lemma 3.5. *Let $\rho \in \mathfrak{R}'_m(E)$ with $\deg(\rho) \geq 2$. There exists a symbol σ in $K_2^{(m)} E(t)$ such that $\deg(\rho - \partial(\sigma)) \leq \deg(\rho) - 1$ and where this inequality is strict if $\deg(\rho) \geq 3$ and $\rho_\infty \neq 0$. More precisely, one may choose $\sigma = \{fh, g\}$ where f is the product of the polynomials in $\text{Supp}(\rho)$ and where $g, h \in E[t] \setminus \{0\}$ are such that $\deg(g) < \deg(f)$ and, either $\deg(h) < \deg(g)$, or $gh \in E^\times$.*

Proof. Let f be the product of the polynomials in $\text{Supp}(\rho)$. By the Chinese Remainder Theorem, we may choose $g \in E[t]$ prime to f with $\deg(g) < \deg(f)$ such that $\partial_p(\{f, g\}) = \rho_p$ for all monic irreducible polynomials $p \in \text{Supp}(\rho)$. If g is constant, let $h = 1$. If g is not square-free, let h be the product of the different monic irreducible factors of g . If g is square-free and not constant, then using the Chinese remainder theorem we choose $h \in E[t]$ prime to g with $\deg(h) < \deg(g)$ such that

$$\partial_p(\{f, g\}) - \rho_p = \{\bar{h}\}$$

in $K_1^{(m)} E_p$ for every monic irreducible factor p of g . For $\sigma = \{fh, g\}$ we obtain that $\text{Supp}(\rho - \partial(\sigma)) \setminus \{\infty\}$ is contained in the set of monic irreducible factors of h , whereby g, h , and σ have the desired properties. \square

Lemma 3.6. *Let $d \in \mathbb{N} \setminus \{0\}$ and $f \in E[t]$ nonconstant and square-free such that $\deg(p) \geq d$ for every irreducible factor p of f . Let $F = E[t]/(f)$ and let ϑ denote the class of t in F . For any $a \in F^\times$ there exist nonzero polynomials $g, h \in E[t]$ with $\deg(h) \leq d - 1$ and $\deg(g) \leq \deg(f) - d$ such that $a = g(\vartheta)/h(\vartheta)$.*

Proof. Let

$$V = \bigoplus_{i=0}^{d-1} E \vartheta^i \quad \text{and} \quad W = \bigoplus_{i=0}^{e-d} E \vartheta^i,$$

where $e = \deg(f)$. By the choice of d and the Chinese Remainder Theorem, we have $V \setminus \{0\} \subseteq F^\times$, where F^\times denotes the group of invertible elements of F . As $a \in F^\times$ we have $\dim_E(Va) = \dim_E(V) = d$ and $\dim_E(Va) + \dim_E(W) = e + 1 > e = [F : E]$, so $Va \cap W \neq 0$. Therefore $h(\vartheta)a = g(\vartheta)$ for certain $h, g \in E[t] \setminus \{0\}$ with $\deg(h) \leq d - 1$ and $\deg(g) \leq e - d$. Thus $h(\vartheta) \in V \setminus \{0\} \subseteq F^\times$ and $a = g(\vartheta)/h(\vartheta)$. \square

Lemma 3.7. *Let $\rho \in \mathfrak{R}'_m(E)$ and $q \in \text{Supp}(\rho)$ such that $\deg(q) = 2n + 1$ with $n \geq 1$. There exists a symbol σ in $K_2^{(m)} E(t)$ such that $\deg(\rho - \partial(\sigma)) \leq \deg(\rho) - 2$. More precisely, one may choose $\sigma = \{qhf^{-2}g^{-2}, g^{-1}f\}$ with $f, g, h \in E[t] \setminus \{0\}$ such that $\deg(f), \deg(g) \leq n$ and $\deg(h) \leq 2n - 1$.*

Proof. Applying Lemma 3.6 for $d = n + 1$ we find $f, g \in E[t] \setminus \{0\}$ with $\deg(f), \deg(g) \leq n$ such that $\partial_q(\{q, g^{-1}f\}) = \rho_q$. Then q is prime to fg . If fg is constant, let $h = 1$. If fg is not square-free, let h be the product of the different monic irreducible factors of fg . If fg is square-free and not constant, we choose $h \in E[t]$ prime to fg and with $\deg(h) < \deg(fg)$ such that

$\partial_p(\{h, g^{-1}f\}) = \partial_p(\{q^{-1}f^2g^2, g^{-1}f\})$ for every monic irreducible factor p of fg . In any case $\deg(h) \leq 2n - 1 = \deg(q) - 2$.

Let $\sigma = \{qhf^{-2}g^{-2}, g^{-1}f\}$. Then we have $\partial_q(\sigma) = \rho_q$ and $\partial_p(\sigma) = 0$ for every monic irreducible polynomial $p \in E[t]$ prime to h and not contained in $\text{Supp}(\rho)$. It follows that $q \in \text{Supp}(\rho) \setminus \text{Supp}(\rho - \partial(\sigma))$ and that every polynomial in $\text{Supp}(\rho - \partial(\sigma)) \setminus \text{Supp}(\rho)$ divides h . Furthermore, if $\deg(h) = 2n - 1$, then $\deg(f) = \deg(g) = n$, so that $\deg(qh) = 4n = 2 \deg(fg)$ and thus $\partial_\infty(\sigma) = 0$. We conclude that $\deg(\rho - \partial(\sigma)) \leq \deg(\rho) - 2$ in any case. \square

Proposition 3.8. *Let $\rho \in \mathfrak{X}'_m(E)$ with $\deg(\rho) \geq 2$. There exists a symbol σ in $K_2^{(m)}E(t)$ such that $\deg(\rho - \partial(\sigma)) \leq \deg(\rho) - 1$. Moreover, if $\deg(\rho) \geq 3$ and $\text{Supp}(\rho)$ contains an element of odd degree, then there exists a symbol σ in $K_2^{(m)}E(t)$ such that $\deg(\rho - \partial(\sigma)) \leq \deg(\rho) - 2$.*

Proof. In view of [Lemma 3.5](#) only the second part of the statement remains to be proven. If $\text{Supp}(\rho)$ contains a nonrational point of odd degree, the statement follows from [Lemma 3.7](#). Suppose now that $\text{Supp}(\rho)$ contains a rational point. Note that the statement is invariant under E -automorphisms of $E(t)$. Hence, we may assume that $\infty \in \text{Supp}(\rho)$, in which case the statement follows from [Lemma 3.5](#). \square

Question 3.9. Given $\rho \in \mathfrak{X}_m(E)$ with $\deg(\rho) \geq 3$, does there always exist a symbol σ in $K_2^{(m)}E(t)$ such that $\deg(\rho - \partial(\sigma)) \leq \deg(\rho) - 2$?

For $x \in \mathbb{R}$, the unique $z \in \mathbb{Z}$ such that $z \leq x < z + 1$ is denoted by $\lfloor x \rfloor$.

Theorem 3.10. *For $\rho \in \mathfrak{X}_m(E)$ and $n = \lfloor \deg(\rho)/2 \rfloor$, there exist symbols $\sigma_1, \dots, \sigma_n$ in $K_2^{(m)}E(t)$ such that $\rho = \partial(\sigma_1 + \dots + \sigma_n)$.*

Proof. We proceed by induction on n . If $n = 0$ then $\rho = 0$ by [Proposition 3.1](#) and the statement is trivial. Assume that $n > 0$. We have either $\deg(\rho) = 2n + 1$, in which case ρ contains a point of odd degree, or $\deg(\rho) = 2n$. Hence, by [Proposition 3.8](#) there exists a symbol σ in $K_2^{(m)}E(t)$ with $\deg(\rho - \partial(\sigma)) \leq 2n - 1$. By the induction hypothesis there exist symbols $\sigma_1, \dots, \sigma_{n-1}$ in $K_2^{(m)}E(t)$ with $\rho - \partial(\sigma) = \partial(\sigma_1 + \dots + \sigma_{n-1})$. Then $\rho = \partial(\sigma_1 + \dots + \sigma_{n-1} + \sigma)$. \square

If we knew that for $m \geq 1$ every element of $\mathfrak{X}_m(E)$ had a lift to $\mathfrak{X}_0(E)$ of the same degree, it would be sufficient to formulate and prove [Theorem 3.10](#) for $m = 0$.

4. Example showing that the bound is sharp

In this section we show that the bound in [Theorem 3.10](#) is sharp for all m and in all degrees. In order to obtain an example in [Example 4.3](#) where the bound of [Theorem 3.10](#) is an equality, we adapt Sivatski's argument [[2007](#), Proposition 2].

For any $a \in E$, there is a unique homomorphism $s_a : K_n^{(m)}E(t) \rightarrow K_n^{(m)}E$ such that $s_a(\{f_1, \dots, f_n\}) = \{f_1(a), \dots, f_n(a)\}$ for any $f_1, \dots, f_n \in E[t]$ prime to $t - a$ and such that $s_a(\{t - a, \cdot, \dots, \cdot\}) = 0$; see [[Gille and Szamuely 2006](#), (7.1.4)].

Lemma 4.1. *The homomorphism $s = s_0 - s_1 : K_n^{(m)} E(t) \rightarrow K_n^{(m)} E$ has the following properties:*

- (a) $s(K_n^{(m)} E) = 0$.
- (b) $s(\{(1-a)t + a, b_2, \dots, b_n\}) = \{a, b_2, \dots, b_n\}$ for any $a, b_2, \dots, b_n \in E^\times$.
- (c) Any symbol in $K_n^{(m)} E(t)$ is mapped under s to a sum of two symbols in $K_n^{(m)} E$.

Proof. Since s_0 and s_1 both restrict to the identity on $K_n^{(m)} E$, part (a) is clear. For $a, b_2, \dots, b_n \in E^\times$ and $\sigma = \{(1-a)t + a, b_2, \dots, b_n\}$, we have $s_1(\sigma) = 0$ and thus $s(\sigma) = s_0(\sigma) = \{a, b_2, \dots, b_n\}$. This shows (b). Part (c) follows from the observation that both s_0 and s_1 map symbols to symbols. \square

Proposition 4.2. *Let $d \in \mathbb{N}$, $a_1, \dots, a_d \in E^\times$, and $\sigma_1, \dots, \sigma_d$ symbols in $K_{n-1}^{(m)} E$. Assume that $\sum_{i=1}^d \{a_i\} \cdot \sigma_i \in K_n^{(m)} E$ is not equal to a sum of less than d symbols and let*

$$\xi = \sum_{i=1}^d \{(1-a_i)t + a_i\} \cdot \sigma_i \in K_n^{(m)} E(t).$$

Then $\deg(\partial(\xi)) = d + 1$, and if $r \in \mathbb{N}$ is such that $\partial(\xi) = \partial(\tau_1 + \dots + \tau_r)$ for symbols τ_1, \dots, τ_r in $K_n^{(m)} E(t)$, then $r \geq \lfloor (d+1)/2 \rfloor$.

Proof. The hypothesis that $\sum_{i=1}^d \{a_i\} \cdot \sigma_i \in K_n^{(m)} E$ cannot be written as a sum of less than d symbols has a few consequences. For $i = 1, \dots, d$, it follows that $\{a_i\} \cdot \sigma_i \neq 0$, so in particular $a_i \neq 1$, and with $p = t + a_i/(1-a_i)$ we get that $\partial_p(\xi) = \sigma_i \neq 0$ in $K_{n-1}^{(m)} E$. Furthermore, since

$$\sum_{i=1}^d \{a_i\} \cdot \sigma_i \neq \sum_{i=1}^{d-1} \{a_i a_d^{-1}\} \cdot \sigma_i,$$

we have $\partial_\infty(\xi) = -\sum_{i=1}^d \sigma_i \neq 0$ in $K_{n-1}^{(m)} E$. Therefore we obtain

$$\text{Supp}(\partial(\xi)) = \left\{ t + \frac{a_i}{1-a_i} \mid 1 \leq i \leq d \right\} \cup \{\infty\}$$

and thus $\deg(\partial(\xi)) = d + 1$.

Assume now that $r \in \mathbb{N}$ and $\partial(\xi) = \partial(\tau_1 + \dots + \tau_r)$ for symbols τ_1, \dots, τ_r in $K_n^{(m)} E(t)$. Then $\tau_1 + \dots + \tau_r - \xi$ is defined over E . Let s be the map from Lemma 4.1. By Lemma 4.1 we obtain that $s(\tau_1 + \dots + \tau_r - \xi) = 0$ and thus

$$\sum_{i=1}^d \{a_i\} \cdot \sigma_i = s(\xi) = s(\tau_1) + \dots + s(\tau_r) \in K_n^{(m)} E,$$

which is a sum of $2r$ symbols. Hence $2r \geq d$, by the hypothesis on d . \square

Example 4.3. Let p be a prime dividing m . Let k be a field containing a primitive p -th root of unity ω and $a_1, \dots, a_d \in k^\times$ such that the Kummer extension $k(\sqrt[p]{a_1}, \dots, \sqrt[p]{a_d})$ of k has degree p^d . Let b_1, \dots, b_d be indeterminates over k and set $E = k(b_1, \dots, b_d)$. Using [Tignol 1987, (2.10)] and [Becher and Hoffmann 2004, (2.1)], it follows that $\sum_{i=1}^d \{a_i, b_i\}$ is not equal to a sum of less than d symbols in $K_2^{(p)}E$. Since p divides m , it follows immediately that $\sum_{i=1}^d \{a_i, b_i\} \in K_2^{(m)}E$ is not a sum of less than d symbols in $K_2^{(m)}E$. Consider

$$\xi = \sum_{i=1}^d \{(1 - a_i)t + a_i, b_i\}$$

in $K_2^{(m)}E(t)$. By Proposition 4.2, for $\rho = \partial(\xi)$ we have that $\deg(\rho) = d + 1$ and $\rho \neq \partial(\xi')$ for any $\xi' \in K_2^{(m)}E(t)$ that is a sum of less than $r = \lfloor \deg(\rho)/2 \rfloor$ symbols.

Acknowledgements

We wish to express our gratitude to Jean-Pierre Tignol for his interest in our work and all his support in its course. We further would like to thank the referee for several very valuable remarks.

References

- [Becher and Hoffmann 2004] K. J. Becher and D. W. Hoffmann, “Symbol lengths in Milnor K -theory”, *Homology Homotopy Appl.* **6**:1 (2004), 17–31. MR 2005b:19001 Zbl 1069.19004
- [Faddeev 1951] D. K. Faddeev, “Simple algebras over a field of algebraic functions of one variable”, *Trudy Mat. Inst. Steklov.* **38** (1951), 321–344. In Russian; translated in *AMS Transl. Ser. 2* **3** (1956), 15–38. MR 13,905c Zbl 0053.35602
- [Gille and Szamuely 2006] P. Gille and T. Szamuely, *Central simple algebras and Galois cohomology*, Cambridge Studies in Advanced Mathematics **101**, Cambridge University Press, Cambridge, 2006. MR 2007k:16033 Zbl 1137.12001
- [Kunyavskii et al. 2006] B. È. Kunyavskii, L. H. Rowen, S. V. Tikhonov, and V. I. Yanchevskii, “Bicyclic algebras of prime exponent over function fields”, *Trans. Amer. Math. Soc.* **358**:6 (2006), 2579–2610. MR 2007d:16034 Zbl 1101.16013
- [Lam 2005] T. Y. Lam, *Introduction to quadratic forms over fields*, Graduate Studies in Mathematics **67**, American Mathematical Society, Providence, RI, 2005. MR 2005h:11075 Zbl 1068.11023
- [Milnor 1970] J. Milnor, “Algebraic K -theory and quadratic forms”, *Invent. Math.* **9** (1970), 318–344. MR 41 #5465 Zbl 0199.55501
- [Rowen et al. 2005] L. H. Rowen, A. S. Sivatski, and J.-P. Tignol, “Division algebras over rational function fields in one variable”, pp. 158–180 in *Algebra and number theory*, edited by R. Tandon, Hindustan Book Agency, Delhi, 2005. MR 2006i:16029 Zbl 1089.16015
- [Sivatski 2007] A. S. Sivatski, “On the Faddeev index of an algebra over the function field of a curve”, preprint 255, Universität Bielefeld, 2007, <http://www.math.uni-bielefeld.de/lag/man/255>.
- [Tignol 1987] J.-P. Tignol, “Algèbres indécomposables d’exposant premier”, *Adv. Math.* **65**:3 (1987), 205–228. MR 88h:16028 Zbl 0642.16015

Received February 8, 2012. Revised June 1, 2012.

KARIM JOHANNES BECHER
UNIVERSITEIT ANTWERPEN
DEPARTMENT MATHEMATICS AND COMPUTER SCIENCE
MIDDELHEIMLAAN 1
B-2020 ANTWERPEN
BELGIUM

and

UNIVERSITÄT KONSTANZ
ZUKUNFTSKOLLEG / FB MATHEMATIK UND STATISTIK
D-78457 KONSTANZ
GERMANY

becher@maths.ucd.ie

MÉLANIE RACZEK
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
ICTEAM
CHEMIN DU CYCLOTRON 2
1348 LOUVAIN-LA-NEUVE
BELGIUM

melanie.raczek@uclouvain.be

ON EXISTENCE OF A CLASSICAL SOLUTION TO A GENERALIZED KELVIN–VOIGT MODEL

MIROSLAV BULÍČEK, PETR KAPLICKÝ AND MARK STEINHAEUER

We consider a two-dimensional generalized Kelvin–Voigt model describing a motion of a compressible viscoelastic body. We establish the existence of a unique classical solution to such a model in the spatially periodic setting. The proof is based on Meyers’ higher integrability estimates that guarantee the Hölder continuity of the gradient of velocity and displacement.

1. Introduction

In this paper we focus on qualitative properties of a solution to a generalized Kelvin–Voigt model that describes the motion of a two-dimensional compressible viscoelastic body. Hence, assuming that the body occupies a domain $\Omega := (0, 1)^2$ and that $T > 0$ is the length of time interest, such a model is described by the system of equations

$$(1-1) \quad \begin{aligned} \rho_0 \mathbf{u}_{tt} - \operatorname{div} \mathbf{T} &= \rho_0 \mathbf{f} && \text{in } Q, \\ \mathbf{u}(0, \cdot) &= \mathbf{u}_0(\cdot) && \text{in } \Omega, \\ \mathbf{u}_t(0, \cdot) &= \mathbf{v}_0(\cdot) && \text{in } \Omega, \end{aligned}$$

where $Q := (0, T) \times \Omega$. Here, $\rho_0 : \Omega \rightarrow \mathbb{R}_+$ is a given density of the body, assumed to be time-independent, $\mathbf{f} : Q \rightarrow \mathbb{R}^2$ is a given density of external body forces, $\mathbf{u} : Q \rightarrow \mathbb{R}^2$ denotes an unknown displacement field and $\mathbf{T} : Q \rightarrow \mathbb{R}^{2 \times 2}$ stands for the Cauchy stress tensor. The initial displacement is denoted by $\mathbf{u}_0 : \Omega \rightarrow \mathbb{R}^2$ and the initial velocity of the body is $\mathbf{v}_0 : \Omega \rightarrow \mathbb{R}^2$.

We assume that

$$(1-2) \quad \mathbf{T} = \mathbf{T}^T \quad \text{in } Q$$

Bulíček is a researcher of the University Center for Mathematical Modeling, Applied Analysis and Computational Mathematics (Math MAC). Kaplický is supported by grant 201/09/0917 of CSF, and also partially by the research project MSM 0021620839 financed by MEYS. Steinhauer acknowledges the support of the Nečas Center for Mathematical Modeling, project LC06052, and thanks the Center for its hospitality.

MSC2000: 35B65, 35Q74, 74D10.

Keywords: Kelvin–Voigt model, regularity, classical solution, large-data and long-time.

and that \mathbf{T} is given as a sum of a viscous and an elastic part,

$$(1-3) \quad \mathbf{T} = \mathbf{T}_v + \mathbf{T}_e,$$

$$(1-4) \quad \mathbf{T}_e = \mathbf{H}(\mathbf{D}(\mathbf{u})),$$

$$(1-5) \quad \mathbf{T}_v = \mathbf{G}(\mathbf{D}(\mathbf{u}_t)),$$

where $\mathbf{H}, \mathbf{G} : \mathbb{R}_{\text{sym}}^{2 \times 2} \rightarrow \mathbb{R}_{\text{sym}}^{2 \times 2}$ are continuous mappings and $\mathbf{D} = (\nabla + \nabla^T)/2$ is the symmetric part of the gradient.

In the context of continuum mechanics, (1-1)₁ represents the balance of linear momentum written in Lagrangian coordinates. The decomposition (1-3) of the Cauchy stress tensor corresponds to the fact that the material under consideration is compressible. The initial density of the body is the given function ρ_0 , while the density at time $t > 0$ can be reconstructed from a formula $\rho(t, \cdot)(1 + \text{div } \mathbf{u}(t, \cdot)) = \rho_0$; see [Bulíček et al. 2012, (26)]. Note that using the balance of angular momentum, the natural requirement for nonpolar materials is (1-2).

In general, most materials can be understood as viscoelastic and one can try to investigate their properties in full generality. Unfortunately, the resulting system is highly nonlinear and may be even hyperbolic and up to our best knowledge there is no satisfactory existence theory for such problems. Therefore it seems to be reasonable (and also necessary) to simplify the model in such a way that it still captures all essential phenomena but it is easier to handle from the mathematical (and even computational) point of view. One such possible procedure, which is also used here, is the assumption that the strains are small. Then, following the fundamental works of Kelvin [Thomson 1865] and Voigt [1892] and taking \mathbf{G} and \mathbf{H} to be linear operators, one obtains the standard Kelvin–Voigt model for a viscoelastic body. However, doing such simplification, and recalling that at the beginning we assumed that the strains were small, we directly obtained a model, where also stresses must be small. On the other hand, it is not true in the original model that even under the assumption that strains are small the Cauchy stress cannot be large, which is the main drawback of the linear Kelvin–Voigt model. Therefore, recently Rajagopal [2009] has reconsidered generalizations of the classical Kelvin–Voigt model wherein he allowed for both the elastic solid and viscous fluid to be described through implicit constitutive relations. These models were also obtained by considering small strains, but the essential assumption was that the strain is a function of the stress. Then using a linearization procedure, one can still end up with small deformations but keeps the essential nonlinearity in stresses. For a more sophisticated discussion, we refer the interested reader to [Rajagopal 2009] and [Bulíček et al. 2012], where the elastic and viscous part of the Cauchy stress are given by the general formula (1-4)–(1-5). This is also the model we are interested in here and one can think of the bodies described by these models as of mixtures

of a material that can store energy and a viscous fluid that can dissipate energy. Moreover, such models are used in practice and for this we refer to [Fung 1993], where the author proposed such models to describe the response of biological matter which exhibits viscoelastic response, and to [Ramberg and Osgood 1943], where the authors deal with the inelastic response of bodies wherein a linearized measure of strain is related nonlinearly to stress.

Let us now formulate precise assumptions on \mathbf{G} and \mathbf{H} . We assume that $\mathbf{H}, \mathbf{G} : \mathbb{R}_{\text{sym}}^{2 \times 2} \rightarrow \mathbb{R}_{\text{sym}}^{2 \times 2}$, $\mathbf{H}, \mathbf{G} \in \mathcal{C}^{0,1}(\mathbb{R}_{\text{sym}}^{2 \times 2})^{2 \times 2}$, $\mathbf{G}(\mathbf{0}) = \mathbf{0}$, $\mathbf{H}(\mathbf{0}) = \mathbf{0}$, and that there exists a function $F : [0, +\infty) \rightarrow [0, +\infty)$ such that the potential $\Phi(\mathbf{D}) := F(|\mathbf{D}|)$ for $\mathbf{D} \in \mathbb{R}_{\text{sym}}^{2 \times 2}$ satisfies $\mathbf{G} = \partial\Phi/\partial\mathbf{D}$. Moreover we assume the existence of $r \in [2, \infty)$ and positive constants ν_0, ν_1 and ν_2 such that

$$(1-6) \quad \nu_0(1 + |\mathbf{D}|^2)^{(r-2)/2} |\mathbf{B}|^2 \leq \frac{\partial\mathbf{G}(\mathbf{D})}{\partial\mathbf{D}} : \mathbf{B} \otimes \mathbf{B} \leq \nu_1(1 + |\mathbf{D}|^2)^{(r-2)/2} |\mathbf{B}|^2,$$

$$(1-7) \quad \left| \frac{\partial\mathbf{H}(\mathbf{D})}{\partial\mathbf{D}} \right| \leq \nu_2$$

for all $\mathbf{B} \in \mathbb{R}_{\text{sym}}^{2 \times 2}$ and almost all $\mathbf{D} \in \mathbb{R}_{\text{sym}}^{2 \times 2}$. The prototypical example of the model we are interested in is given by

$$(1-8) \quad \mathbf{G}(\mathbf{D}) = (1 + |\mathbf{D}|^2)^{(r-2)/2} \mathbf{D}, \quad \mathbf{H}(\mathbf{D}) = (1 + |\mathbf{D}|^2)^{(q-2)/2} \mathbf{D}$$

with some $r \geq 2$, $q \in (1, 2]$. For (1-8) it is easy to verify (1-6) and (1-7). Note that (1-7) allows one to consider more general examples than that introduced in (1-8). It is worth noticing that (1-7) says only that \mathbf{H} is uniformly Lipschitz continuous but does not require any additional structure assumption as potentiality or monotonicity.

Concerning the boundary condition, we restrict ourselves to periodic (with respect to Ω) boundary conditions that require some normalization condition (in order to guarantee uniqueness of a solution). For simplicity we choose the simplest one:

$$(1-9) \quad \int_{\Omega} \rho_0(x) \mathbf{u}(t, x) dx = \int_{\Omega} \rho_0(x) \mathbf{u}_t(t, x) dx = \mathbf{0} \quad \text{for all } t \in (0, T).$$

A direct consequence of (1-9) is that we need to assume a compatibility condition on the data, namely, for all $t \in (0, T)$ we need that

$$(1-10) \quad \int_{\Omega} \rho_0(x) \mathbf{f}(t, x) dx = \int_{\Omega} \rho_0(x) \mathbf{v}_0(x) dx = \int_{\Omega} \rho_0(x) \mathbf{u}_0(x) dx = \mathbf{0}.$$

Although we use the simplest possible boundary condition we believe that our result can be adopted to a more general setting with more reasonable physical boundary data.

Next, we introduce the assumption put on the data of (1-1). For the density ρ_0 , we assume that there are $0 < \rho_* \leq \rho^* < \infty$ such that

$$(1-11) \quad \rho_0 \in L^\infty, \quad \rho_* \leq \rho_0(x) \leq \rho^* \quad \text{for almost all } x \in \Omega.$$

Concerning the density of the external body forces we prescribe

$$(1-12) \quad \rho_0 \mathbf{f} \in (L^r(0, T; (W_{\text{per}}^{1,r})^2))^*.$$

Finally, for the initial displacement and the initial velocity we assume that in addition to (1-10) they also satisfy

$$(1-13) \quad \mathbf{u}_0 \in (W_{\text{per}}^{1,2})^2, \quad \mathbf{v}_0 \in (L^2)^2.$$

The existence of a weak solution for the problem (1-1) with nonlinear \mathbf{T}_v satisfying (1-6) with $r = 2$ can be found in [Friedman and Nečas 1988], [Demoulini 2000] and [Tvedt 2008] under certain structural assumptions on \mathbf{T} that are more general than (1-3), (1-4) and (1-5). Next, the existence theory was extended for $r \geq 2$ in [Bulíček et al. 2012], where the authors assumed that the Cauchy stress satisfies (1-3)–(1-7). In addition they showed the uniqueness of a solution

$$(1-14) \quad \mathbf{u} \in W^{1,\infty}(0, T; (L^2)^2) \cap W^{1,r}(0, T; (W^{1,r})^2)$$

to (1-1). Although all results in [Bulíček et al. 2012] treat the case of mixed boundary conditions, the method presented there works also in the easier periodic case in which we are interested in here. Moreover, assuming that the data are smooth, one can prove by the method introduced there that the unique solution \mathbf{u} to (1-1) is more regular. We state the result in the next theorem.

Theorem 1.1 [Bulíček et al. 2012]. *Let $r \geq 2$, $T > 0$ be arbitrary. Assume that \mathbf{T} satisfies (1-3)–(1-7) and the data $(\rho_0, \mathbf{f}, \mathbf{u}_0, \mathbf{v}_0)$ satisfy (1-10), (1-11)–(1-13). Then there exists a unique weak solution $\mathbf{u} \in W^{1,\infty}(0, T; (L^2)^2) \cap W^{1,r}(0, T; (W_{\text{per}}^{1,r})^2)$ of (1-1).*

In addition, assume that there is $p > 2$ such that the data fulfill

$$(1-15) \quad (\rho_0, \mathbf{u}_0, \mathbf{v}_0) \in W_{\text{per}}^{1,p} \times (W_{\text{per}}^{2,p})^2 \times (W_{\text{per}}^{2,p})^2, \\ \mathbf{f} \in W^{1,2}(0, T; (L^p)^2).$$

Then the weak solution satisfies

$$(1-16) \quad (1 + |\mathbf{D}(\mathbf{u}_t)|)^{(r-2)/2} \mathbf{D}(\nabla \mathbf{u}_t) \in L^2(0, T; (L^2)^{2 \times 2 \times 2}), \\ (1 + |\mathbf{D}(\mathbf{u}_t)|)^{(r-2)/2} \mathbf{D}(\mathbf{u}_{tt}) \in L^2(0, T; (L^2)^{2 \times 2}), \\ \mathbf{u}_{tt} \in L^{r'}(0, T; (L^{r'})^2).$$

The main result of our paper is that we improve (1-16) and get the Hölder continuity of the velocity gradient. Consequently, we use such information to obtain that the unique weak solution is in fact a classical one provided that the data are sufficiently smooth. The first improvement is this:

Theorem 1.2. *Let $r \geq 2$, $T > 0$ be arbitrary. Assume that \mathbf{T} satisfies (1-3)–(1-7) and the data $(\rho_0, \mathbf{f}, \mathbf{u}_0, \mathbf{v}_0)$ satisfy (1-10), (1-11)–(1-13). In addition, assume that there is $p > 2$ such that (1-15) holds. Then there exists some $s \in (2, p)$ such that the unique solution of (1-1) satisfies*

$$(1-17) \quad \mathbf{u}_t \in W^{1,\infty}(0, T; (L^s)^2) \cap L^\infty(0, T; (W_{\text{per}}^{2,s})^2).$$

Consequently, for all $\alpha \in (0, (1 - 2/s)/3)$,

$$(1-18) \quad \nabla \mathbf{u}_t \in (\mathcal{C}^{0,\alpha}(Q))^{2 \times 2}.$$

As a consequence of [Theorem 1.2](#) we obtain:

Theorem 1.3. *Let all assumptions of [Theorem 1.2](#) hold. Then the unique solution from [Theorem 1.2](#) satisfies*

$$(1-19) \quad \mathbf{u}_t \in W^{1,p}(0, T; (L^p)^2) \cap L^p(0, T; (W_{\text{per}}^{2,p})^2).$$

If we in addition assume that

$$(1-20) \quad \begin{aligned} (\rho_0, \mathbf{u}_0, \mathbf{v}_0) &\in W^{1,\infty} \times (W_{\text{per}}^{3,p})^2 \times (W_{\text{per}}^{3,p})^2, \\ \mathbf{f} &\in L^p(0, T; (W_{\text{per}}^{1,p})^2), \quad \mathbf{G}, \mathbf{H} \in \mathcal{C}_{\text{loc}}^{1,1}(\mathbb{R}_{\text{sym}}^{2 \times 2})^{2 \times 2}, \end{aligned}$$

then the unique solution from [Theorem 1.2](#) satisfies

$$(1-21) \quad \nabla \mathbf{u}_t \in W^{1,p}(0, T; (L^p)^{2 \times 2}) \cap L^p(0, T; (W_{\text{per}}^{2,p})^{2 \times 2}).$$

As an immediate consequence of [Theorem 1.3](#) and an interpolation [Lemma A.1](#) we get:

Corollary 1.1. *Let all assumptions of [Theorem 1.3](#) hold with some $p > 4$ and $\mathbf{f} \in \mathcal{C}(Q)$. Then the unique weak solution \mathbf{u} is a classical one.*

For general systems of partial differential equations Hölder continuity of weak solutions is a rare phenomenon, that can be obtained only under special circumstances. One of them is that if $\Omega \subset \mathbb{R}^2$ is as in [Theorem 1.2](#). As far as we know the only former result in this direction for the problem (1-1) is the one from [[Friedman and Nečas 1988](#)] where [Theorem 1.2](#) is proved in the case $r = 2$. Another special condition when regularity (1-18) can be obtained is a special structure of the elliptic term \mathbf{T}_v . If it is assumed that the system (1-1) is a linear system of equations, i.e., classical Kelvin–Voigt model, one can establish the existence of a unique smooth solution (provided that data are smooth) by standard results for linear systems. In [[DiBenedetto and Friedman 1984; 1985](#)] a nonlinear function \mathbf{G} is treated with the structure very similar to the one suggested in (1-8) but the symmetric gradient is replaced with the full gradient, i.e., $\mathbf{T}_v = \mathbf{G}(\nabla \mathbf{u}_t)$. It is a remarkable fact that the method from [[DiBenedetto and Friedman 1984; 1985](#)] cannot be applied in the situation of (1-5), i.e., if the elliptic term depends only on $\mathbf{D}(\mathbf{u}_t)$. According to

our best knowledge no results about the Hölder continuity of gradients of weak solutions are known if $\Omega \subset \mathbb{R}^d$, $d > 2$ and the elliptic part of the equation depends only on the symmetric part of $\nabla \mathbf{u}_t$.

The method of the proof of [Theorem 1.2](#) is based on the fact that a small improvement of regularity in (1-16) gives Hölder continuity of $\nabla \mathbf{u}_t$. This was first observed in [\[Boyarskiĭ 1957\]](#) and [\[Meyers 1963\]](#) in the stationary case and extended to parabolic systems in [\[Nečas and Šverák 1991\]](#) and [\[Frehse and Seregin 1999\]](#). This method was used in [\[Friedman and Nečas 1988\]](#) to prove [Theorem 1.2](#) for $r = 2$. First the integrability of \mathbf{u}_{tt} was improved and then the system was treated as an elliptic one on time levels. This method must be modified if $r > 2$ as in this case we do not know how to get separately only information about \mathbf{u}_{tt} . Regularity of \mathbf{u}_{tt} and $\nabla^2 \mathbf{u}_t$ must be dealt with simultaneously as it was suggested for generalized Navier–Stokes system in [\[Kaplický et al. 2002\]](#). This is also the approach that we adopt here to prove [Theorem 1.2](#).

The paper has the following structure. In the next section we introduce some auxiliary lemmas about linear stationary and parabolic systems with bounded measurable coefficients. The proofs can be found in the [Appendix](#). In [Section 3](#) we provide the proof of [Theorem 1.2](#) if $r = 2$. This result is not new, but it is a basis for the analysis in [Section 4](#) where [Theorem 1.2](#) is proved for $r > 2$. Finally, we present a sketch of the proof of [Theorem 1.3](#) in [Section 5](#).

In the paper we use standard notation for Lebesgue and Sobolev spaces and their norms. If the domain on which the functions are considered is $\Omega = (0, 1)^2$, we shorten the notation and write only $W^{1,q}$, L^q or $\|\cdot\|_q$, $\|\cdot\|_{1,q}$. The subscript per denotes periodicity with respect to Ω . Particularly, $W_{\text{per}}^{1,q}$ are spaces of functions from $W_{\text{loc}}^{1,q}(\mathbb{R}^2)$ for which there is a representative that is periodic with respect to Ω . Moreover, scalar-, vector- and tensor-valued functions are denoted by small letters, small bold letters and bold capital letters in what follows. Also in order to distinguish between scalars, vectors and tensors we use the abbreviations X^d and $X^{d \times d}$ for vector- and tensor-valued function in a Banach space X . The symbol $\mathbb{R}_{\text{sym}}^{2 \times 2}$ denotes the space of all symmetric 2×2 matrices and for $\boldsymbol{\xi} \in \mathbb{R}^{2 \times 2}$, $\boldsymbol{\xi}_{\text{sym}}$ is its symmetric part. For a function $\mathbf{G} : \mathbb{R}_{\text{sym}}^{2 \times 2} \rightarrow \mathbb{R}^{2 \times 2}$ we denote its gradient by $\partial_{\mathbf{D}} \mathbf{G}$. Then for any $\mathbf{B}, \mathbf{D} \in \mathbb{R}_{\text{sym}}^{2 \times 2}$ we denote by $\partial_{\mathbf{D}} \mathbf{G}(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B}$ a scalar product of the matrices $\partial_{\mathbf{D}} \mathbf{G}(\mathbf{D})$ and $\mathbf{B} \otimes \mathbf{B}$. Symbols \mathbf{u}_t and $\partial_t \mathbf{u}$ denote derivative of \mathbf{u} with respect to $t \in (0, T)$.

2. Auxiliary results

In this section we recall some results for a linear system similar to (1-1), the proof of these results can be found in the [Appendix](#). This linear system will play a crucial role in the proof of [Theorem 1.2](#), where it will be used as the comparison problem.

Lemma 2.1. *Let $T > 0$ be given and assume that $\mathbf{A} : (0, T) \times \Omega \rightarrow \mathbb{R}^{2 \times 2 \times 2 \times 2}$ is a measurable tensor-valued function satisfying for some $0 < \lambda_1 \leq \lambda_2 < \infty$ and for almost all $(t, x) \in (0, T) \times \Omega$ the following symmetry and ellipticity conditions:*

$$(2-1) \quad \mathbf{A}_{ij}^{kl}(t, x) = \mathbf{A}_{kl}^{ij}(t, x) = \mathbf{A}_{kl}^{ji}(t, x) \quad \text{for all } i, j, k, l = 1, \dots, 2,$$

$$(2-2) \quad \lambda_1 |\mathbf{D}|^2 \leq \sum_{i,j,k,l=1}^2 \mathbf{A}_{kl}^{ij}(t, x) \mathbf{D}_{ij} \mathbf{D}_{kl} \leq \lambda_2 |\mathbf{D}|^2 \quad \text{for all } \mathbf{D} \in \mathbb{R}_{\text{sym}}^{2 \times 2}.$$

Then for any $\mathbf{F} \in L^2(0, T; (L^2)^{2 \times 2})$ and any Ω -periodic $\mathbf{w}_0 \in (L^2)^2$ having zero mean value, a unique Ω -periodic weak solution

$$\mathbf{w} \in \mathcal{C}(0, T; (L^2)^2) \cap L^2(0, T; (W^{1,2})^2), \quad \int_{\Omega} \mathbf{w}(t, x) \, dx = \mathbf{0}$$

exists to the problem

$$(2-3) \quad \begin{aligned} \mathbf{w}_t - \operatorname{div}(\mathbf{A}\mathbf{D}(\mathbf{w})) &= -\operatorname{div} \mathbf{F} && \text{in } (0, T) \times \Omega, \\ \mathbf{w}(0, \cdot) &= \mathbf{w}_0(\cdot) && \text{in } \Omega. \end{aligned}$$

Moreover, there exist positive constants $K, L > 0$ that are independent of T, \mathbf{A} and \mathbf{F} such that, for all s satisfying

$$(2-4) \quad 2 \leq s \leq 2 + \frac{L\lambda_1}{\lambda_2},$$

the following estimate holds:

$$(2-5) \quad \sup_{t \in (0, T)} \|\mathbf{w}(t)\|_s^2 \leq K \left(\frac{1}{\lambda_1} \|\mathbf{F}\|_{L^2(0, T; L^s)}^2 + \|\mathbf{w}_0\|_s^2 \right).$$

In case one replaces $\mathbf{D}(\mathbf{w})$ by $\nabla \mathbf{w}$ in [Lemma 2.1](#), the statement was proved in [\[Nečas and Šverák 1991\]](#). However, following the procedure in that paper almost step by step one can prove [Lemma 2.1](#) in full generality; see the [Appendix](#) for a detailed proof.

Note that in the previous lemma we did not improve the estimate for the gradient of the solution. As it is usual in parabolic equations the information on the spatial gradient of the solution will be deduced by comparing the equation with its steady form. Therefore, we recall the following lemma; see for example [\[Nečas 1967\]](#).

Lemma 2.2. *Let $\mathbf{A} : \Omega \rightarrow \mathbb{R}^{2 \times 2 \times 2 \times 2}$ be a measurable tensor-valued function satisfying, for some $0 < \lambda_1 \leq \lambda_2 < \infty$ and almost all $x \in \Omega$,*

$$(2-6) \quad \mathbf{A}_{ij}^{kl}(x) = \mathbf{A}_{kl}^{ij}(x) = \mathbf{A}_{kl}^{ji}(x) \quad \text{for all } i, j, k, l = 1, \dots, 2,$$

$$(2-7) \quad \lambda_1 |\mathbf{D}|^2 \leq \sum_{i,j,k,l=1}^2 \mathbf{A}_{kl}^{ij}(x) \mathbf{D}_{ij} \mathbf{D}_{kl} \leq \lambda_2 |\mathbf{D}|^2 \quad \text{for all } \mathbf{D} \in \mathbb{R}_{\text{sym}}^{2 \times 2}.$$

Then, for any $\mathbf{F} \in (L^2)^{2 \times 2}$, there exists a unique Ω -periodic weak solution $\mathbf{w} \in (W^{1,2})^2$ such that $\int_{\Omega} \mathbf{w} \, dx = \mathbf{0}$ solving the problem

$$(2-8) \quad -\operatorname{div}(\mathbf{A}\mathbf{D}(\mathbf{w})) = -\operatorname{div} \mathbf{F} \quad \text{in } \Omega.$$

Moreover, there exist $K, L > 0$ independent of T, \mathbf{A} and \mathbf{F} such that, for all

$$2 \leq s \leq 2 + \frac{L\lambda_1}{\lambda_2},$$

we have

$$(2-9) \quad \|\mathbf{D}(\mathbf{w})\|_s \leq \frac{K}{\lambda_1} \|\mathbf{F}\|_s.$$

In general the constants K, L from [Lemma 2.1](#) and [Lemma 2.2](#) may be different but without loss of generality we assume in what follows that they are the same.

3. Proof of [Theorem 1.2](#) in the case $r = 2$

This section is devoted to the proof of [Theorem 1.2](#) for $r = 2$. First, we introduce an ε -approximation to the problem (1-1), but we still write \mathbf{u} instead of \mathbf{u}^ε for its solution:

$$(3-1) \quad \rho_0 \mathbf{u}_{tt} - \operatorname{div}(\mathbf{G}(\mathbf{D}(\mathbf{u}_t))) = \rho_0 \mathbf{f} + \operatorname{div}(\mathbf{H}(\mathbf{D}(\mathbf{u} \star \omega^\varepsilon))) \quad \text{in } (0, T) \times \Omega,$$

with periodic boundary condition and initial data $(\mathbf{u}_0, \mathbf{v}_0)$. Here $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a standard regularizing kernel, i.e., $\omega \in C_0^\infty(U(0, 1))$ is nonnegative, radially symmetric, $\int_{\mathbb{R}^2} \omega \, dx = 1$, and we define

$$\omega^\varepsilon(x) = \varepsilon^{-2} \omega\left(\frac{x}{\varepsilon}\right).$$

Note that the convolution in the last term of (3-1) is taken only in space direction. Next, we formulate the existence result for (3-1), that is the starting point of our analysis.

Lemma 3.1. *Let \mathbf{H} and \mathbf{G} satisfy (1-6)–(1-7) with $r = 2$. Assume that $\rho_0, \mathbf{f}, \mathbf{u}_0$ and \mathbf{v}_0 satisfy (1-10) and (1-11)–(1-13). In addition, assume that $\mathbf{f} \in W^{1,2}(0, T; (L^2)^2)$, $\rho_0 \in W_{\text{per}}^{1,2+\delta}$ for a certain $\delta > 0$ and $\mathbf{u}_0, \mathbf{v}_0 \in (W_{\text{per}}^{2,2})^2$. Then for any $\varepsilon > 0$ there exists a unique Ω -periodic weak solution \mathbf{u} to (3-1), (1-1)₂–(1-1)₃ that obeys the a priori estimate*

$$(3-2) \quad \begin{aligned} \|\nabla^2 \mathbf{u}\|_{L^\infty(0, T; L^2)} + \|\nabla \mathbf{u}_t\|_{L^\infty(I, L^2)} + \|\nabla \mathbf{u}_t\|_{L^2(I, W^{1,2})} + \|\mathbf{u}_{tt}\|_{L^\infty(I, L^2)} + \|\mathbf{u}_{tt}\|_{L^2(I, W^{1,2})} \\ \leq C_1(1 + \|\mathbf{G}(\mathbf{D}(\mathbf{v}_0))\|_{1,2} + \|\mathbf{H}(\mathbf{D}(\mathbf{u}_0))\|_{1,2}), \end{aligned}$$

where $C_1 > 0$ is independent of v_1 and ε . Moreover, this solution converges to the unique solution of (1-1) as $\varepsilon \rightarrow 0_+$.

Proof. The proof is presented in [Bulíček et al. 2012] for the system (1-1) with mixed (Dirichlet and Neumann) boundary conditions. In our situation, smoothing of the term with \mathbf{H} in (3-1) simplifies the situation and also the periodic boundary conditions are simpler to deal with. Since the proof of Lemma 3.1 follows [Bulíček et al. 2012, Theorem 4.1, p. 9] line by line we do not present it here. \square

Lemma 3.2. *Let all the assumptions of Lemma 3.1 hold. Let $\varepsilon > 0$ be arbitrary and \mathbf{u} the unique weak solution to (3-1). Assume that for some $\delta > 0$ and s satisfying*

$$(3-3) \quad 2 \leq s \leq 2 + \min\left(\frac{Lv_0\rho_*}{v_1\rho_*}, \frac{\delta}{2}\right),$$

the data fulfill

$$(3-4) \quad (\mathbf{u}_0, \mathbf{v}_0, \rho_0) \in (W_{\text{per}}^{2,s})^2 \times (W_{\text{per}}^{2,s})^2 \times W_{\text{per}}^{1,2+\delta}(\Omega), \\ \mathbf{f} \in W^{1,2}(0, T; (L^s)^2).$$

Then the following estimate holds:

$$(3-5) \quad \sup_{t \in (0, T)} \|\mathbf{u}_{tt}\|_s \leq (1 + v_1)C(\mathbf{u}_0, \mathbf{v}_0, \delta, \mathbf{f}, v_0, v_2).$$

Proof. First, we construct an Ω -periodic \mathbf{F} having zero mean value over Ω such that

$$\operatorname{div} \mathbf{F} = \rho_0 \mathbf{f} \quad \text{in } (0, T) \times \Omega.$$

Such a construction is possible due to the compatibility condition (1-10). Moreover, using the theory for the divergence equation (see for example [Feireisl and Novotný 2009; Novotný and Straškraba 2004]) and (1-11) we have

$$(3-6) \quad \|\mathbf{F}\|_{W^{1,2}(0, T; W^{1,s})} \leq C\|\rho_0 \mathbf{f}\|_{W^{1,2}(0, T; L^s)} \leq C,$$

where the last inequality follows from (3-4). Next, we set $\mathbf{w} := \rho_0 \mathbf{u}_{tt}$ and applying ∂_t to (3-1) (in view of (3-2), this procedure is rigorous) we see that \mathbf{w} is a weak solution of the system

$$(3-7) \quad \mathbf{w}_t - \operatorname{div}(\mathbf{A}\mathbf{D}(\mathbf{w})) = \operatorname{div} \tilde{\mathbf{F}} \quad \text{in } (0, T) \times \Omega,$$

where

$$\mathbf{A} := \frac{1}{\rho_0} \frac{\partial \mathbf{G}(\mathbf{D}(\mathbf{u}_t))}{\partial \mathbf{D}}, \\ \tilde{\mathbf{F}} := \mathbf{F}_t + \frac{\partial \mathbf{H}(\mathbf{D}(\mathbf{u} \star \omega^\varepsilon))}{\partial \mathbf{D}} \mathbf{D}(\mathbf{u}_t \star \omega^\varepsilon) - \left[\frac{\partial \mathbf{G}(\mathbf{D}(\mathbf{u}_t))}{\partial \mathbf{D}} \right] \left(\frac{\nabla \rho_0}{\rho_0} \otimes \mathbf{u}_{tt} \right).$$

Since \mathbf{G} is assumed to satisfy (1-6) with $r = 2$, and ρ_0 satisfies (1-11), we see that the matrix \mathbf{A} fulfills (2-1)–(2-2) with

$$(3-8) \quad \lambda_1 := \frac{v_0}{\rho_*} \quad \text{and} \quad \lambda_2 := \frac{v_1}{\rho_*}.$$

Hence, assuming that s satisfies (3-3), it also satisfies $s \in [2, 2 + L\lambda_1/\lambda_2]$ and we can use Lemma 2.1 to deduce that

$$(3-9) \quad \sup_{t \in (0, T)} \|\mathbf{w}(t)\|_s^2 \leq K \left(\|\mathbf{w}(0)\|_s^2 + \frac{1}{\lambda_1} \int_0^T \|\tilde{\mathbf{F}}\|_s^2 \right).$$

We check that the right side is finite. To see this, we first evaluate the initial value $\mathbf{w}(0)$. Using (3-1) we see that

$$\mathbf{w}(0) := \rho_0 \mathbf{u}_{tt}(0) = \operatorname{div}(\mathbf{G}(\mathbf{D}(\mathbf{v}_0)) + \mathbf{H}(\mathbf{D}(\mathbf{u}_0 \star \omega^\varepsilon))) + \rho_0 \mathbf{f}(0)$$

and by using (1-6)–(1-7) and (3-4) we obtain that (for estimating \mathbf{f} we use the embedding $W^{1,2}(0, T) \hookrightarrow \mathcal{C}^{0,1/2}([0, T]) \hookrightarrow \mathcal{C}([0, T])$ in dimension one)

$$(3-10) \quad \begin{aligned} \|\mathbf{w}(0)\|_s &\leq \|\operatorname{div} \mathbf{G}(\mathbf{D}(\mathbf{v}_0))\|_s + \|\operatorname{div} \mathbf{H}(\mathbf{D}(\mathbf{u}_0 \star \omega^\varepsilon))\|_s + \|\rho_0 \mathbf{f}(0)\|_s \\ &\leq \nu_1 \|\mathbf{v}_0\|_{2,s} + \nu_2 \|\mathbf{u}_0\|_{2,s} + \rho^* \|\mathbf{f}(0)\|_s \leq C(1 + \nu_1). \end{aligned}$$

It remains to estimate the norm of $\tilde{\mathbf{F}}$ appearing on the right side of (3-9). Using (3-6) and (1-6)–(1-7) we obtain that

$$(3-11) \quad \begin{aligned} \int_0^T \|\tilde{\mathbf{F}}\|_s^2 &\leq \int_0^T (\|\mathbf{F}_t\|_s^2 + \nu_2 \|\mathbf{D}(\mathbf{u}_t)\|_s^2 + \|\rho_0^{-1} \partial_{\mathbf{D}} \mathbf{G}(\mathbf{D}(\mathbf{u}_t)) \nabla \rho_0 \otimes \mathbf{u}_{tt}\|_s^2) \\ &\leq \int_0^T \left(\|\mathbf{F}_t\|_{1,s}^2 + \nu_2 \|\mathbf{D}(\mathbf{u}_t)\|_s^2 + \frac{\nu_1^2}{\rho_*^2} \|\nabla \rho_0\|_{2+\delta}^2 \|\mathbf{u}_{tt}\|_{s(2+\delta)/(2+\delta-s)}^2 \right) \\ &\leq C(\mathbf{v}_0, \mathbf{u}_0, \mathbf{f}, \rho_0, \nu_2) + C(\rho_0, \delta) \nu_1^2 \int_0^T \|\mathbf{u}_{tt}\|_{1,2}^2. \end{aligned}$$

Consequently, using the uniform estimate (3-2) we can bound the last term on the right side of (3-11) and inserting this and (3-10) into (3-9) we deduce (3-5). \square

Since, we already know that \mathbf{u}_{tt} belongs to a better space than L^2 uniformly in time, we can improve the spatial regularity of \mathbf{u} with help of Lemma 2.2.

Lemma 3.3. *Let all assumptions of Lemma 3.1 hold. Then for any $\varepsilon > 0$, $\delta > 0$ and $s > 0$ fulfilling (3-3) and any data satisfying (3-4), the unique solution \mathbf{u} to the problem (3-1) satisfies for almost all $t \in (0, T)$ the estimate:¹*

$$(3-12) \quad \|\nabla^2 \mathbf{u}_t(t)\|_s \leq C(1 + \|\mathbf{u}_{tt}\|_{L^\infty(0, T; L^s)} + \|\nabla^2(\mathbf{u}(t) \star \omega^\varepsilon)\|_s),$$

with C depending only on $(\rho_0, \mathbf{f}, \mathbf{v}_0, \mathbf{u}_0, \nu_0, \nu_2, T)$.

Proof. Since we know from Lemma 3.1 that (3-1) holds pointwise at almost all time levels $t \in (0, T)$. We fix such an arbitrary $t \in (0, T)$ and rewrite the problem

¹The right side of (3-12) is finite, since for the time derivative we have an estimate due to Lemma 3.2 and the last term in (3-12) is finite due to regularization.

(3-1) as

$$(3-13) \quad -\operatorname{div}(\mathbf{G}(\mathbf{D}(\mathbf{u}_t))) = \operatorname{div}(\mathbf{F} + \mathbf{H}(\mathbf{D}(\mathbf{u} \star \omega^\varepsilon)) - \mathbf{F}_0) \quad \text{in } \Omega,$$

where \mathbf{F}_0 is found such that (note that t is fixed in what follows)

$$(3-14) \quad \operatorname{div} \mathbf{F}_0 = \rho_0 \mathbf{u}_{tt} \quad \text{in } \Omega, \quad \|\mathbf{F}_0(t)\|_{1,s} \leq C \rho^* \|\mathbf{u}_{tt}(t)\|_s$$

and \mathbf{F} satisfies

$$(3-15) \quad \operatorname{div} \mathbf{F} = \rho_0 \mathbf{f} \quad \text{in } \Omega, \quad \|\mathbf{F}(t)\|_{1,s} \leq C \rho^* \|\mathbf{f}(t)\|_s.$$

Next, we fix $k \in \{1, 2\}$, denote $\mathbf{w} := \partial_k \mathbf{u}_t$ and

$$\mathbf{A} := \frac{\partial \mathbf{G}(\mathbf{D}(\mathbf{u}_t))}{\partial \mathbf{D}}$$

and differentiate (3-13) in the weak sense with respect to x_k . We obtain the system of equations

$$(3-16) \quad -\operatorname{div}(\mathbf{A}\mathbf{D}(\mathbf{w})) = \operatorname{div}(\partial_k \mathbf{F} + \partial_k \mathbf{H}(\mathbf{D}(\mathbf{u} \star \omega^\varepsilon)) - \partial_k \mathbf{F}_0) \quad \text{in } \Omega,$$

equipped with periodic boundary conditions and requiring zero mean value for \mathbf{w} . Similarly as in the proof of Lemma 3.2, \mathbf{A} satisfies the assumption of Lemma 2.2 with $\lambda_1 := \nu_0$ and $\lambda_2 := \nu_1$. Hence for any $s^* \in [2, 2 + L\nu_0/\nu_1]$ we have the estimate

$$(3-17) \quad \|\mathbf{D}(\mathbf{w})\|_{s^*} \leq \frac{K}{\nu_0} (\|\mathbf{F}\|_{1,s^*} + \|\mathbf{F}_0\|_{1,s^*} + \|\mathbf{H}(\mathbf{D}(\mathbf{u} \star \omega^\varepsilon))\|_{1,s^*}).$$

Since, we know that $s \leq 2 + L(\nu_0 \rho^*)/(\nu_1 \rho^*) \leq 2 + L\nu_0/\nu_1$, we see that (3-17) also holds for $s^* := s$. Moreover, since it holds for any $k = 1, 2$ we can deduce from (3-17) by using the definition of \mathbf{F} and \mathbf{F}_0 that

$$(3-18) \quad \|\nabla^2 \mathbf{u}_t(t)\|_s \leq \frac{C(\rho_0)}{\nu_0} (\|\mathbf{f}(t)\|_s + \|\mathbf{u}_{tt}(t)\|_s + \nu_2 \|\mathbf{D}(\mathbf{u} \star \omega^\varepsilon(t))\|_{1,s}).$$

Consequently, using (3-4), (3-3) and the a priori uniform estimates (3-2), we deduce (3-12). \square

Having all previous estimates, we are ready to prove Theorem 1.2 for $r = 2$. Since, the case $r = 2$ will be used in the proof of Theorem 1.2 for $r > 2$ we formulate it as a special theorem where we trace the important constant ν_1 .

Theorem 3.1. *Let $T > 0$ be arbitrary. Assume that \mathbf{T} satisfies (1-3)–(1-7) with $r = 2$ and that data $(\rho_0, \mathbf{f}, \mathbf{u}_0, \mathbf{v}_0)$ satisfy (1-10), (1-11)–(1-13). In addition, assume that there is $p > 2$ such that the data fulfill*

$$(3-19) \quad (\mathbf{u}_0, \mathbf{v}_0, \rho_0) \in (W_{\text{per}}^{2,p})^2 \times (W_{\text{per}}^{2,p})^2 \times W_{\text{per}}^{1,p}(\Omega), \\ \mathbf{f} \in W^{1,2}(0, T; (L^p)^2).$$

Then there exists a constant C depending only on $(\rho_0, \mathbf{v}_0, \mathbf{u}_0, \mathbf{f}, T, \nu_0, \nu_2, p)$ such that for any s fulfilling

$$(3-20) \quad 2 \leq s \leq 2 + \min\left(\frac{L\nu_0\rho_*}{\nu_1\rho_*}, \frac{p-2}{2}\right),$$

the unique weak solution (1-1) satisfies the estimate

$$(3-21) \quad \|\mathbf{u}_{tt}\|_{L^\infty(0,T;L^s)} + \|\nabla^2 \mathbf{u}_t\|_{L^\infty(0,T;L^s)} \leq C(1 + \nu_1).$$

Proof. To prove the theorem it is enough to show estimate (3-21) for the unique solutions of the approximating problem (3-1). Indeed, having uniform (ε -independent) estimate (3-21) for the solution of the approximate problem it is easy to let $\varepsilon \rightarrow 0+$ and to obtain a solution of the original problem (1-1). The estimate (3-21) is valid for this solution due to the weak*-lower semicontinuity of the norm in $L^\infty(0, T; L^s)$. Uniqueness of the solution follows by the method of [Bulíček et al. 2012]; compare Lemma 3.1. Due to our assumption on the data and s we see that also all assumptions of Lemmas 3.1–3.3 are satisfied. We can use (3-12) to prove (3-21). To do so, we need to estimate the last two terms on the right side of (3-12). Note that both of them are finite, so we directly have an estimate of the form (3-21) but with right side depending on ε . To avoid this dependence we estimate both terms as follows. We start with the time derivative for which we obtain by direct use of Lemma 3.2 that

$$(3-22) \quad \|\mathbf{u}_{tt}\|_{L^\infty(0,T;L^s)} \leq C(1 + \nu_1).$$

Next, for the second term, we get, by (3-19),

$$(3-23) \quad \begin{aligned} \|\nabla^2(\mathbf{u}(t) \star \omega^\varepsilon)\|_s &\leq C\|\nabla^2 \mathbf{u}(t)\|_s = C\left\|\int_0^t \nabla^2 \mathbf{u}_t(\tau) d\tau + \nabla^2 \mathbf{u}_0\right\|_s \\ &\leq C\left(1 + \int_0^t \|\nabla^2 \mathbf{u}_t(\tau)\|_s d\tau\right). \end{aligned}$$

Using (3-22) and (3-23), we see that (3-12) reduces to

$$\|\nabla^2 \mathbf{u}_t(t)\|_s \leq C\left(1 + \nu_1 + \int_0^t \|\nabla^2 \mathbf{u}_t(\tau)\|_s d\tau\right).$$

Applying Gronwall's lemma in its integral form, we deduce (3-21). \square

4. Proof of Theorem 1.2 in the case $r > 2$

This section is devoted to the proof of Theorem 1.2 for $r > 2$. It is based on a direct application of the result from the previous section onto a suitable approximating problem.

First, we introduce a quadratic approximation of the problem (1-1). For any $\lambda > 1$ we define Lipschitz continuous functions η_λ and μ_λ as follows:

$$(4-1) \quad \eta_\lambda(s) := \begin{cases} 1 & \text{for } s \in [0, 2\lambda^2], \\ -\frac{s-3\lambda^2}{\lambda^2} & \text{for } s \in (2\lambda^2, 3\lambda^2), \\ 0 & \text{for } s \geq 3\lambda^2, \end{cases}$$

$$(4-2) \quad \mu_\lambda(s) := \begin{cases} 0 & \text{for } s \in [0, \lambda^2], \\ \gamma_\lambda \frac{s-\lambda^2}{\lambda^2} & \text{for } s \in (\lambda^2, 2\lambda^2), \\ \gamma_\lambda & \text{for } s \geq 2\lambda^2, \end{cases}$$

with some constant $\gamma_\lambda \in \mathbb{R}_+$ to be specified later. We approximate \mathbf{G} by \mathbf{G}^λ as

$$(4-3) \quad \mathbf{G}^\lambda(\mathbf{D}) := \eta_\lambda(|\mathbf{D}|^2)\mathbf{G}(\mathbf{D}) + \mu_\lambda(|\mathbf{D}|^2)\mathbf{D}.$$

Note that for \mathbf{G}^λ a potential can be constructed. The most important properties of this approximation are introduced in the following lemma.

Lemma 4.1. *Let \mathbf{G} satisfy the assumption (1-6) with $r > 2$ and $\nu_0, \nu_1 > 0$. Let $\lambda > 1$ be arbitrary. We set in (4-1) and (4-2)*

$$(4-4) \quad \gamma_\lambda := 7\nu_1(1 + 3\lambda^2)^{(r-2)/2}.$$

Then for all $\mathbf{B} \in \mathbb{R}_{\text{sym}}^{2 \times 2}$ and almost all $\mathbf{D} \in \mathbb{R}_{\text{sym}}^{2 \times 2}$ it holds

$$(4-5) \quad \bar{\nu}_0|\mathbf{B}|^2 \leq \partial_{\mathbf{D}}\mathbf{G}^\lambda(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B} \leq \bar{\nu}_1|\mathbf{B}|^2$$

with $\bar{\nu}_0$ and $\bar{\nu}_1$ given as

$$(4-6) \quad \bar{\nu}_0 := \nu_0,$$

$$(4-7) \quad \bar{\nu}_1 := \bar{\nu}_1(\lambda) := 36\nu_1(1 + 3\lambda^2)^{(r-2)/2}.$$

Moreover, setting $\bar{\lambda}(\mathbf{D}) := \min(\lambda, |\mathbf{D}|)$, we get

$$(4-8) \quad \nu_0(1 + \bar{\lambda}(\mathbf{D})^2)^{(r-2)/2}|\mathbf{B}|^2 \leq \partial_{\mathbf{D}}\mathbf{G}^\lambda(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B} \leq 36\nu_1(1 + 3\bar{\lambda}(\mathbf{D})^2)^{(r-2)/2}|\mathbf{B}|^2.$$

Proof. To shorten the notation we write $\partial_{\mathbf{D}}\mathbf{G}^\lambda(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B}$ only as $\partial_{\mathbf{D}}\mathbf{G}(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B}$. Using the definition of \mathbf{G}^λ we get

$$I = \eta_\lambda(|\mathbf{D}|^2)\partial_{\mathbf{D}}\mathbf{G}(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B} + 2\eta'_\lambda(|\mathbf{D}|^2)(\mathbf{D} \cdot \mathbf{B})(\mathbf{G}(\mathbf{D}) \cdot \mathbf{B}) \\ + \mu_\lambda(|\mathbf{D}|^2)|\mathbf{B}|^2 + 2\mu'_\lambda(|\mathbf{D}|^2)(\mathbf{D} \cdot \mathbf{B})^2.$$

From this identity and the definition of η_λ and μ_λ we finally conclude that

$$I = \begin{cases} \partial_{\mathbf{D}} \mathbf{G}(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B} & \text{if } |\mathbf{D}|^2 < \lambda^2, \\ \partial_{\mathbf{D}} \mathbf{G}(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B} + \gamma_\lambda \frac{|\mathbf{D}|^2 - \lambda^2}{\lambda^2} |\mathbf{B}|^2 + \frac{2\gamma_\lambda}{\lambda^2} (\mathbf{D} \cdot \mathbf{B})^2 & \text{if } |\mathbf{D}|^2 \in (\lambda^2, 2\lambda^2), \\ -\frac{|\mathbf{D}|^2 - 3\lambda^2}{\lambda^2} \partial_{\mathbf{D}} \mathbf{G}(\mathbf{D}) : \mathbf{B} \otimes \mathbf{B} - \frac{2}{\lambda^2} (\mathbf{D} \cdot \mathbf{B})(\mathbf{G}(\mathbf{D}) \cdot \mathbf{B}) + \gamma_\lambda |\mathbf{B}|^2 & \text{if } |\mathbf{D}|^2 \in (2\lambda^2, 3\lambda^2), \\ \gamma_\lambda |\mathbf{B}|^2 & \text{if } |\mathbf{D}|^2 > 3\lambda^2. \end{cases}$$

Now we remark that by the assumption (1-6) on \mathbf{G} we get

$$(\mathbf{G}(\mathbf{D}) \cdot \mathbf{B}) \leq \nu_1 (1 + |\mathbf{D}|^2)^{(r-2)/2} |\mathbf{B}| |\mathbf{D}|.$$

Defining $Y := I/|\mathbf{B}|^2$ and noting that $\lambda > 1$, it follows that

$$\begin{aligned} \nu_0 (1 + |\mathbf{D}|^2)^{(r-2)/2} \leq Y \leq \nu_1 (1 + |\mathbf{D}|^2)^{(r-2)/2} & \quad \text{if } |\mathbf{D}|^2 < \lambda^2, \\ \nu_0 (1 + |\mathbf{D}|^2)^{(r-2)/2} \leq Y \leq \nu_1 (1 + |\mathbf{D}|^2)^{(r-2)/2} + 5\gamma_\lambda & \quad \text{if } |\mathbf{D}|^2 \in (\lambda^2, 2\lambda^2), \\ \gamma_\lambda - 6\nu_1 (1 + |\mathbf{D}|^2)^{(r-2)/2} \leq Y \leq \gamma_\lambda + 7\nu_1 (1 + |\mathbf{D}|^2)^{(r-2)/2} & \quad \text{if } |\mathbf{D}|^2 \in (2\lambda^2, 3\lambda^2), \\ \gamma_\lambda = Y & \quad \text{if } |\mathbf{D}|^2 > 3\lambda^2, \end{aligned}$$

and we see that (4-5)–(4-8) follows. \square

Next, we find $\lambda_0 > 1$ such that² $\min(L\bar{\nu}_0\rho_*/(\bar{\nu}_1\rho^*), (p-2)/2) = L\bar{\nu}_0\rho_*/(\bar{\nu}_1\rho^*)$ and $\bar{\nu}_1 \geq 1$ for all $\lambda > \lambda_0$.

Finally, for arbitrary fixed $\lambda > \lambda_0$, we consider an approximation of (1-1) of the form

$$(4-9) \quad \begin{aligned} \rho_0 \mathbf{u}_{tt} - \operatorname{div} \mathbf{G}^\lambda(\mathbf{D}(\mathbf{u}_t)) - \operatorname{div} \mathbf{H}(\mathbf{D}(\mathbf{u})) &= \rho_0 \mathbf{f} & \text{in } Q, \\ \mathbf{u}(0, \cdot) &= \mathbf{u}_0(\cdot) & \text{in } \Omega, \\ \mathbf{u}_t(0, \cdot) &= \mathbf{v}_0(\cdot) & \text{in } \Omega, \end{aligned}$$

$$\int_{\Omega} \rho_0(x) \mathbf{u}(t, x) dx = \int_{\Omega} \rho_0(x) \mathbf{u}_t(t, x) dx = \mathbf{0} \quad \text{for all } t \in (0, T),$$

equipped with periodic boundary conditions for \mathbf{u} .

According to Lemma 4.1, \mathbf{G}^λ satisfies all assumptions of Theorem 3.1 and we get that for all s satisfying

$$(4-10) \quad 2 \leq s \leq 2 + \frac{L\bar{\nu}_0\rho_*}{\bar{\nu}_1\rho^*},$$

²The constant $p > 2$ appears in Theorem 1.2 and it is assumed to be the same as in Theorem 3.1.

the unique weak solution $\overline{(4-9)}$ satisfies the estimate

$$(4-11) \quad \begin{aligned} \|\nabla^2 \mathbf{u}_t\|_{L^\infty(0,T;L^s)} &\leq C(\rho_0, \mathbf{f}, T, \bar{v}_0, \nu_2, \mathbf{v}_0, \mathbf{u}_0, p)(\bar{v}_1 + 1) \\ &\leq C(\rho_0, \mathbf{f}, T, \bar{v}_0, \nu_2, \mathbf{v}_0, \mathbf{u}_0, p)\bar{v}_1. \end{aligned}$$

Using the definition of \bar{v}_0 and \bar{v}_1 we can set

$$(4-12) \quad 2 \leq s = 2 + R\lambda^{2-r}$$

for a fixed $R \in \left(0, \frac{L\rho_*\nu_0}{36\rho_*\nu_1} \left(\frac{1}{2}\right)^{r-2}\right)$ and rewrite the estimate (4-11) as

$$(4-13) \quad \|\nabla^2 \mathbf{u}_t\|_{L^\infty(0,T;L^s)} \leq C(\rho_0, \mathbf{f}, T, \nu_0, \nu_2, \mathbf{v}_0, \mathbf{u}_0, p, r)\lambda^{r-2}.$$

Our main goal, based on the estimate (4-13), is to find a sufficiently large $\lambda > \lambda_0$ such that

$$(4-14) \quad M := \left\| 1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2 \right\|_{L^\infty(0,T;L^\infty)} \leq \lambda^2.$$

For such λ the equality $\mathbf{G}^\lambda(\mathbf{D}(\mathbf{u}_t)) = \mathbf{G}(\mathbf{D}(\mathbf{u}_t))$ holds a.e. in $(0, T) \times \Omega$; hence, \mathbf{u} solves the original problem (1-1).

We start with estimates uniform with respect to λ . In the following the positive constant C is always independent of λ but it can depend on the data $(\mathbf{f}, \rho_0, \mathbf{u}_0, \mathbf{v}_0, p, r, \nu_0, \nu_1, \nu_2)$. From Lemma 3.1 we know that

$$(4-15) \quad \|\nabla^2 \mathbf{u}\|_{L^\infty(0,T;L^2)} + \|\mathbf{u}_{tt}\|_{L^\infty(0,T;L^2)} \leq C(1 + \|\mathbf{G}^\lambda(\mathbf{D}(\mathbf{v}_0))\|_{1,2}).$$

This estimate is still λ -dependent. However, using the definition of \mathbf{G}^λ and the assumptions on the data (1-15), we see that

$$\|\mathbf{G}^\lambda(\mathbf{D}(\mathbf{v}_0))\|_{1,2} \leq C(1 + \|\mathbf{D}(\mathbf{v}_0)\|^{r-1}\|_2 + \|\mathbf{D}(\mathbf{v}_0)\|^{r-2}\|\nabla^2 \mathbf{v}_0\|_2) \leq C(1 + \|\mathbf{v}_0\|_{2,p}^{r-1}),$$

where for the last inequality we used the Hölder inequality and the embedding $W^{2,p} \hookrightarrow W^{1,\infty}$ (valid for $p > 2$). Consequently, we see that (4-15) can be rewritten as

$$(4-16) \quad \|\nabla^2 \mathbf{u}\|_{L^\infty(0,T;L^2)} + \|\mathbf{u}_{tt}\|_{L^\infty(0,T;L^2)} \leq C.$$

Uniform estimates on $\nabla^2 \mathbf{u}_t$ are obtained by the same method as in the proof of Lemma 3.3. We rewrite (4-9) for a.e. $t \in (0, T)$ as

$$-\operatorname{div} \mathbf{G}^\lambda(\mathbf{D}(\mathbf{u}_t(t))) = \rho_0 \mathbf{f}(t) + \operatorname{div} \mathbf{H}(\mathbf{D}(\mathbf{u}(t))) - \rho_0 \mathbf{u}_{tt}(t).$$

This equation holds pointwise in Ω due to (4-11) and it is allowed to test it with $\mathbf{u}_t(t)$ and $-\Delta \mathbf{u}_t(t)$. Doing so, one gets with help of (4-16) and (4-8) that

$$\int_{\Omega} (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t(t)))^2)^{(r-2)/2} |\nabla^2 \mathbf{u}_t(t)|^2 \leq C \quad \text{for any } t \in (0, T),$$

and by a simple algebraic manipulation we deduce that

$$(4-17) \quad \left\| (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;W^{1,2})} \leq C.$$

Finally, we combine the nonuniform estimate (4-13) with the uniform ones (4-16) and (4-17) to deduce (4-14). First, we consider $\bar{s} \in (2, s)$ and $\alpha \in (0, 2)$ such that $1 = \alpha/2 + (\bar{s} - \alpha)/s$, i.e.,

$$(4-18) \quad \bar{s} - 2 = (2 - \alpha)(s - 2)/2.$$

We use the Hölder inequality to get

$$\begin{aligned} & \left\| (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;W^{1,\bar{s}})}^{\bar{s}} \\ & \leq C \left\| (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;W^{1,2})}^\alpha \left\| (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;W^{1,s})}^{\bar{s}-\alpha}. \end{aligned}$$

To estimate the term on the right, we use the definition of $\bar{\lambda}$, the uniform estimate (4-17) and the nonuniform estimate (4-13) to conclude that

$$(4-19) \quad \begin{aligned} \left\| (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;W^{1,\bar{s}})}^{\bar{s}} & \\ & \leq C \left(1 + \left\| \nabla (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;L^s)}^{\bar{s}-\alpha} \right) \\ & \leq C \left(1 + \left\| \nabla^2 \mathbf{u}_t \right\|_{L^\infty(0,T;L^s)}^{\bar{s}-\alpha} \lambda^{(r-2)(\bar{s}-\alpha)/2} \right) \\ & \leq C \lambda^{(r-2)(\bar{s}-\alpha)(3/2)}. \end{aligned}$$

Finally, we focus on finding such $\lambda > \lambda_0$ so that (4-14) holds. Using the embedding theorem $W^{1,\bar{s}} \hookrightarrow L^\infty$ with the precise embedding constant (see [Zierner 1989, proof of Theorem 2.4.1]), the definition (4-14) of M and the estimate (4-19), we get

$$(4-20) \quad \begin{aligned} M^{r/4} & \leq \left(\frac{C}{\bar{s}-2} \right)^{1-1/\bar{s}} \left\| (1 + \bar{\lambda}(\mathbf{D}(\mathbf{u}_t))^2)^{r/4} \right\|_{L^\infty(0,T;W^{1,\bar{s}})} \\ & \leq \left(\frac{C}{\bar{s}-2} \right)^{1-1/\bar{s}} \lambda^{\frac{3(r-2)(\bar{s}-\alpha)}{2\bar{s}}}. \end{aligned}$$

Hence, to show (4-14) and consequently to finish the proof, it is enough to find $\lambda > \lambda_0$, $\bar{s} \in (2, s)$ and $\alpha \in (0, 2)$ fulfilling (4-18) such that

$$(4-21) \quad \left(\frac{C}{\bar{s}-2} \right)^{1-1/\bar{s}} \lambda^{\frac{3(r-2)(\bar{s}-\alpha)}{2\bar{s}}} \leq \lambda^{r/2}.$$

Next, using (4-12) and (4-18) it is not difficult to deduce the identities

$$\begin{aligned} \left(\frac{C}{\bar{s}-2} \right)^{1-1/\bar{s}} \lambda^{(r-2)(3/2)\frac{\bar{s}-\alpha}{s}} & = \left(\frac{2C}{(2-\alpha)(s-2)} \right)^{1-1/\bar{s}} \lambda^{(r-2)(3/2)\frac{\bar{s}-\alpha}{s}} \\ & = \left(\frac{2C}{(2-\alpha)R} \right)^{1-1/\bar{s}} \lambda^{(r-2)\left(1-\frac{1}{s}+(3/2)\frac{\bar{s}-\alpha}{s}\right)} \end{aligned}$$

and we see that (4-21) is equivalent to

$$(4-22) \quad \left(\frac{2C}{(2-\alpha)R} \right)^{1-1/\bar{s}} \lambda^{(r-2)(1-\frac{1}{\bar{s}}+\frac{\bar{s}-\alpha}{\bar{s}}\frac{3}{2})} \leq \lambda^{r/2}.$$

Since $\lim_{\alpha \rightarrow 2-} \bar{s} = 2$ we have

$$\lim_{\alpha \rightarrow 2-} (r-2) \left(1 - \frac{1}{\bar{s}} + \frac{\bar{s}-\alpha}{\bar{s}} \frac{3}{2} \right) = \frac{r-2}{2} < \frac{r}{2}$$

and therefore it is always possible to find $\alpha \in (0, 2)$ (and consequently \bar{s}) and $\varepsilon > 0$ such that

$$\frac{r}{2} - (r-2) \left(1 - \frac{1}{\bar{s}} + \frac{\bar{s}-\alpha}{\bar{s}} \frac{3}{2} \right) > \varepsilon.$$

Thus, we fix such α and \bar{s} and we see that to fulfill (4-22) it is enough to find $\lambda > \lambda_0$ such that

$$\left(\frac{2C}{(2-\alpha)R} \right)^{1-1/\bar{s}} \leq \lambda^\varepsilon,$$

which is clearly possible and therefore the proof of (1-17) for the case $r > 2$ is complete. The regularity statement in (1-18) follows from Lemma A.1, part 2. Theorem 1.2 is proved.

5. Proof of Theorem 1.3

We start this section by formulating a result on L^p regularity for certain parabolic systems with Hölder continuous coefficients.

Theorem 5.1. *Let $d \in \mathbb{N}$, $\alpha \in (0, 1]$, $p > 1$, $\Omega = (0, 1)^d$ and $Q = (0, T) \times \Omega$. Assume that $\mathbf{A}_{ij}^{kl} : (0, T) \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfy the symmetry condition (2-1) and in addition for all $i, j, k, l \in \{1, \dots, d\}$ there hold*

$$(5-1) \quad \mathbf{A}_{ij}^{kl} \in C^{0,\alpha}([0, T] \times \mathbb{R}^d), \quad \mathbf{A}_{ij}^{kl} \text{ is periodic with respect to } \Omega,$$

$$(5-2) \quad \exists \gamma > 0, \forall \boldsymbol{\xi} \in \mathbb{R}^{d \times d}, t > 0, x \in \mathbb{R}^d : (\mathbf{A}(t, x) : \boldsymbol{\xi} \otimes \boldsymbol{\xi}) \geq \gamma |\boldsymbol{\xi}_{\text{sym}}|^2.$$

Let $1 < q < p$ and $\mathbf{w} \in L^q(0, T, W^{2,q}(\mathbb{R}^d))$ with $\partial_t \mathbf{w} \in L^q(0, T, L^q(\mathbb{R}^d))$ be a strong solution of the problem

$$(5-3) \quad \partial_t \mathbf{w}_j - \mathbf{A}_{ij}^{kl} \partial_i \partial_k \mathbf{w}_l = \mathbf{F}_j \quad \text{in } (0, T) \times \mathbb{R}^d$$

such that \mathbf{w} is periodic with respect to Ω and $\mathbf{w}(0, \cdot) = \mathbf{w}_0$, where $\mathbf{w}_0 \in W_{\text{per}}^{2,p}(\Omega)$ and $\mathbf{F} \in L^p(Q)^d$. Then this solution satisfies $\nabla^2 \mathbf{w} \in L^p(Q)^{d \times d \times d}$, $\partial_t \mathbf{w} \in L^p(Q)^d$ with the following uniform estimate:

$$\exists C > 0, \forall t \in (0, T) : \|\nabla^2 \mathbf{w}\|_{p, Q_t} + \|\partial_t \mathbf{w}\|_{p, Q_t} \leq C(\|\mathbf{F}\|_{p, Q_t} + \|\mathbf{w}_0\|_{2,p}),$$

where $Q_t = (0, t) \times \Omega$ and $C > 0$ may depend on T but is independent of t .

Proof. First one finds a smooth approximation of the solution \mathbf{w} by convolutions. Then it is possible to apply slightly modified L^p theory from [Schlag 1996] provided $q > 2$. If $q < 2$ first one needs to develop by a duality argument an L^p theory for $p < 2$ based on the results from the same paper. (The result also follows from [Ladyzhenskaja et al. 1968, Theorem VII.10.4].) \square

In the rest of this section we provide only formal a priori estimates. However, they can be made rigorous by the method of Section 3, see the approximation (3-1) and the proof of Theorem 3.1.

Let \mathbf{u} be the unique solution constructed in Theorem 1.2 and assume that it is sufficiently smooth. We suppose that all assumptions of Theorem 1.2 hold and show an estimate leading to (1-19). We denote $\mathbf{w} = \mathbf{u}_t$. It follows from Theorem 1.2 that \mathbf{w} is a strong solution of the problem (5-3) with $\mathbf{w}_0 = \mathbf{v}_0$ and

$$\mathbf{A}_{ij}^{kl} = \frac{1}{\rho_0} \frac{\partial \mathbf{G}_{ij}(\mathbf{D}(\mathbf{u}_t))}{\partial \mathbf{D}_{kl}}, \quad \mathbf{F}_j = \mathbf{f}_j + \frac{1}{\rho_0} \sum_{i=1}^2 \partial_{\mathbf{D}} \mathbf{H}_{ij}(\mathbf{D}\mathbf{u}) : \mathbf{D} \partial_i \mathbf{u}.$$

Here, the symmetry of \mathbf{A} was used. Since we already have (1-18) we know that \mathbf{A} satisfies (5-1) and (5-2), $\partial_{\mathbf{D}} \mathbf{H}(\mathbf{D}\mathbf{u})$ is bounded and we can apply Theorem 5.1 to get

$$(5-4) \quad \|\nabla^2 \mathbf{u}_t\|_{p, Q_t}^p + \|\mathbf{u}_{tt}\|_{p, Q_t}^p \leq C(\|\mathbf{f}\|_{p, Q_t}^p + \|\nabla^2 \mathbf{u}\|_{p, Q_t}^p + \|\mathbf{v}_0\|_{2, p}^p).$$

Using the inequality $\|\nabla^2 \mathbf{u}(t)\|_p^p \leq C(\|\nabla^2 \mathbf{u}_0\|_p^p + \|\nabla^2 \mathbf{u}_t\|_{p, Q_t}^p + \|\nabla^2 \mathbf{u}\|_{p, Q_t}^p)$ we conclude that

$$\|\nabla^2 \mathbf{u}(t)\|_p^p \leq C(\|\mathbf{u}_0\|_{2, p}^p + \|\mathbf{v}_0\|_{2, p}^p + \|\mathbf{f}\|_{p, Q_T}^p + \|\nabla^2 \mathbf{u}\|_{p, Q_t}^p).$$

Gronwall's lemma with (5-4) then gives (1-19).

To get estimates for (1-21) we proceed similarly as in the first step. We a priori assume sufficient smoothness of \mathbf{u} and define $\mathbf{w} = \partial_k \mathbf{u}_t$ for fixed $k \in \{1, 2\}$. We differentiate (1-1) with respect to x_k and find that \mathbf{w} solves the problem (5-3) with $\mathbf{w}_0 = \partial_k \mathbf{v}_0$ and

$$\begin{aligned} \mathbf{A}_{ij}^{kl} &= \frac{1}{\rho_0} \frac{\mathbf{G}_{ij}(\mathbf{D}(\mathbf{u}_t))}{\partial \mathbf{D}_{kl}}, \\ \mathbf{F}_j &= \frac{1}{\rho_0} \left(\sum_{i=1}^2 [\partial_{\mathbf{D}}^2 \mathbf{G}_{ij}(\mathbf{D}(\mathbf{u}_t)) : (\mathbf{D}(\partial_i \mathbf{u}_t) \otimes \mathbf{D}(\partial_k \mathbf{u}_t)) + \partial_{\mathbf{D}} \mathbf{H}_{ij}(\mathbf{D}(\mathbf{u})) : \mathbf{D}(\partial_i \partial_k \mathbf{u}) \right. \\ &\quad \left. + \partial_{\mathbf{D}}^2 \mathbf{H}_{ij}(\mathbf{D}(\mathbf{u})) : (\mathbf{D}(\partial_i \mathbf{u}) \otimes \mathbf{D}(\partial_k \mathbf{u}))] - \partial_k \rho_0 \partial_t^2 \mathbf{u}_j + \partial_k (\rho_0 \mathbf{f}_j) \right). \end{aligned}$$

For an arbitrary $\sigma \in (1, p]$, $t \in (0, T)$ we obtain from the properties of \mathbf{G} , \mathbf{H} and (1-18) that

$$\|\mathbf{F}\|_{\sigma, Q_t} \leq C(1 + \|\nabla^2 \mathbf{u}_t\|_{2\sigma, Q_t}^2 + \|\nabla^3 \mathbf{u}\|_{\sigma, Q_t} + \|\nabla \mathbf{u}\|_{2\sigma, Q_t}^2 + \|\mathbf{u}_{tt}\|_{\sigma, Q_t}).$$

The constant $C > 0$ may depend on $\mathbf{H}, \mathbf{G}, T, \mathbf{u}_0, \rho_0$.

Using [Theorem 5.1](#) we obtain, since $k = 1, 2$ was arbitrary,

$$\begin{aligned} & \|\nabla \mathbf{u}_{tt}\|_{\sigma, Q_t} + \|\nabla^3 \mathbf{u}_t\|_{\sigma, Q_t} \\ & \leq C(1 + \|\nabla^2 \mathbf{u}_t\|_{2\sigma, Q_t}^2 + \|\nabla^3 \mathbf{u}\|_{\sigma, Q_t} + \|\nabla \mathbf{u}\|_{2\sigma, Q_t}^2 + \|\mathbf{u}_{tt}\|_{\sigma, Q_t} + \|\mathbf{v}_0\|_{3, \sigma}). \end{aligned}$$

Now we use the inequality $\|\nabla^3 \mathbf{u}(t)\|_{\sigma}^{\sigma} \leq C(\|\nabla^3 \mathbf{u}\|_{\sigma, Q_t}^{\sigma} + \|\nabla^3 \mathbf{u}_t\|_{\sigma, Q_t}^{\sigma} + \|\nabla^3 \mathbf{u}_0\|_{\sigma}^{\sigma})$ to get

$$\begin{aligned} & \|\nabla^3 \mathbf{u}(t)\|_{\sigma, \Omega}^{\sigma} + \|\nabla \mathbf{u}_{tt}\|_{\sigma, Q_t}^{\sigma} + \|\nabla^3 \mathbf{u}_t\|_{\sigma, Q_t}^{\sigma} \\ & \leq C(1 + \|\nabla^2 \mathbf{u}_t\|_{2\sigma, Q_t}^2 + \|\nabla^3 \mathbf{u}\|_{\sigma, Q_t}^{\sigma} + \|\nabla \mathbf{u}\|_{2\sigma, Q_t}^2 + \|\mathbf{u}_{tt}\|_{\sigma, Q_t}^{\sigma} + \|\mathbf{u}_0\|_{3, \sigma}^{\sigma} + \|\mathbf{v}_0\|_{3, \sigma}^{\sigma}). \end{aligned}$$

Due to the assumption on \mathbf{u}_0 and \mathbf{v}_0 we know that $\|\mathbf{u}_0\|_{3, \sigma}^{\sigma} + \|\mathbf{v}_0\|_{3, \sigma}^{\sigma} < +\infty$ for all $\sigma \leq p$. If $\|\nabla^2 \mathbf{u}_t\|_{2\sigma, Q_t}^2 + \|\nabla \mathbf{u}\|_{2\sigma, Q_t}^2 + \|\mathbf{u}_{tt}\|_{\sigma, Q_t}^{\sigma}$ is bounded we get by Gronwall's inequality

$$(5-5) \quad \|\nabla \mathbf{u}_{tt}\|_{\sigma, Q_t}^{\sigma} + \|\nabla^3 \mathbf{u}_t\|_{\sigma, Q_t}^{\sigma} < +\infty.$$

This is always true for $\sigma \in (1, s/2]$, where $s > 2$ is taken from [Theorem 1.2](#). By the following multiplicative inequality [[Ladyzhenskaja et al. 1968](#), Theorem II.2.2] we get that, for any measurable function z ,

$$\|z\|_{\sigma(s+2)/2, Q}^{\sigma(s+2)/2} \leq C \|z\|_{L^{\infty}(0, T; L^s)}^{\sigma s/2} \|\nabla z\|_{\sigma, Q}^{\sigma}.$$

If we take into account (5-5) and (1-17) we find that, for $\sigma \in (1, s/2]$,

$$\|\nabla^2 \mathbf{u}\|_{\sigma(s+2)/2} + \|\nabla^2 \mathbf{u}_t\|_{\sigma(s+2)/2} + \|\mathbf{u}_{tt}\|_{\sigma(s+2)/2, Q} < +\infty.$$

Since $\sigma \frac{s+2}{2} > 2\sigma$ for all $\sigma > 1$ we get the statement (1-21) of [Theorem 1.3](#) after finite number of iterations.

Appendix: Proof of [Lemma 2.1](#)

First, we sketch the proof of [Lemma 2.1](#). We focus on main differences compared to [[Nečas and Šverák 1991](#)], where the full gradient case is treated.

Proof of [Lemma 2.1](#). The existence and uniqueness of a weak solution to (2-3) is standard. Therefore, we focus here only on the proof of (2-5). We provide only the formal proof but everything can be done rigorously by mollifying \mathbf{A} and \mathbf{w}_0 , applying standard results for the heat equation, deriving uniform bounds of the type (2-5) and then passing to the limit.

The main idea of the proof is to use $\mathbf{w}|\mathbf{w}|^{s-2}$ as a test function in the weak formulation of (2-3). Due to the presence of a nonlinearity in the test function we

obtain a pollution term coming from the elliptic term. To handle it, we use a simple inequality that can be deduced by an integration by parts formula. Hence, for any $s \geq 2$ and any smooth periodic \mathbf{u} we have

$$\begin{aligned}
& \int_{\Omega} |\mathbf{u}|^{s-2} |\nabla \mathbf{u}|^2 \\
&= - \int_{\Omega} |\mathbf{u}|^{s-2} \Delta \mathbf{u} \cdot \mathbf{u} - (s-2) \int_{\Omega} |\mathbf{u}|^{s-2} |\nabla |\mathbf{u}||^2 \\
&= -(s-2) \int_{\Omega} |\mathbf{u}|^{s-2} |\nabla |\mathbf{u}||^2 - 2 \int_{\Omega} |\mathbf{u}|^{s-2} \operatorname{div}(\mathbf{D}(\mathbf{u})) \cdot \mathbf{u} + \int_{\Omega} |\mathbf{u}|^{s-2} \nabla(\operatorname{div} \mathbf{u}) \cdot \mathbf{u} \\
&= -(s-2) \int_{\Omega} |\mathbf{u}|^{s-2} |\nabla |\mathbf{u}||^2 + 2 \int_{\Omega} |\mathbf{u}|^{s-2} |\mathbf{D}(\mathbf{u})|^2 + 2(s-2) \int_{\Omega} |\mathbf{u}|^{s-3} \mathbf{D}(\mathbf{u}) \mathbf{u} \cdot \nabla |\mathbf{u}| \\
&\quad - \int_{\Omega} |\mathbf{u}|^{s-2} |\operatorname{div} \mathbf{u}|^2 - (s-2) \int_{\Omega} |\mathbf{u}|^{s-3} \operatorname{div} \mathbf{u} \nabla |\mathbf{u}| \cdot \mathbf{u}.
\end{aligned}$$

Consequently, moving the term with the good sign to the left side we obtain the inequality

$$\begin{aligned}
& \int_{\Omega} (|\mathbf{u}|^{s-2} |\operatorname{div} \mathbf{u}|^2 + (s-2) |\mathbf{u}|^{s-2} |\nabla |\mathbf{u}||^2 + |\mathbf{u}|^{s-2} |\nabla \mathbf{u}|^2) \\
&\leq 2 \int_{\Omega} |\mathbf{u}|^{s-2} |\mathbf{D}(\mathbf{u})|^2 + 2(s-2) \int_{\Omega} |\mathbf{u}|^{s-2} |\mathbf{D}(\mathbf{u})| |\nabla |\mathbf{u}|| \\
&\quad + (s-2) \int_{\Omega} |\mathbf{u}|^{s-2} |\operatorname{div} \mathbf{u}| |\nabla |\mathbf{u}||.
\end{aligned}$$

Finally, using the pointwise estimate $|\operatorname{div} \mathbf{u}| \leq C |\mathbf{D}(\mathbf{u})|$ and applying Young's inequality we deduce that for any $s > 2$ there exists $C_s > 0$ such that

$$(A-1) \quad \int_{\Omega} |\mathbf{u}|^{s-2} |\nabla \mathbf{u}|^2 \leq C_s \int_{\Omega} |\mathbf{u}|^{s-2} |\mathbf{D}(\mathbf{u})|^2.$$

If we restrict to $s \in [2, 10]$ we can find $C^* > 0$ such that for all $s \in [2, 10]$, $C_s < C^*$.

With estimate (A-1) we can easily continue by using the standard procedure; see [Nečas and Šverák 1991; Frehse and Seregin 1999]. We test (2-3) by $|\mathbf{w}|^{s-2} \mathbf{w}$ with arbitrary $s \in [2, 4]$ to get

$$\begin{aligned}
(A-2) \quad & \frac{1}{s} \frac{d}{dt} \|\mathbf{w}\|_s^s + \int_{\Omega} |\mathbf{w}|^{s-2} \mathbf{A} \mathbf{D}(\mathbf{w}) \cdot \mathbf{D}(\mathbf{w}) \\
&= - \int_{\Omega} \mathbf{F} \cdot \nabla (|\mathbf{w}|^{s-2} \mathbf{w}) - \int_{\Omega} \mathbf{A} \mathbf{D}(\mathbf{w}) \cdot (\nabla |\mathbf{w}|^{s-2} \otimes \mathbf{w}).
\end{aligned}$$

Consequently, using (2-2) and (A-1) we observe that (we use Young's inequality to

get the second estimate)

$$\begin{aligned}
\text{(A-3)} \quad & \frac{1}{s} \frac{d}{dt} \|\mathbf{w}\|_s^s + \lambda_1 \int_{\Omega} |\mathbf{w}|^{s-2} |\mathbf{D}(\mathbf{w})|^2 \\
& \leq C \left(\int_{\Omega} |\mathbf{F}| |\mathbf{w}|^{s-2} |\nabla \mathbf{w}| + \lambda_2 (s-2) \int_{\Omega} |\mathbf{w}|^{s-2} |\nabla \mathbf{w}|^2 \right) \\
& \leq \left(C \lambda_2 (s-2) + \frac{\lambda_1}{2C^*} \right) \int_{\Omega} |\mathbf{w}|^{s-2} |\nabla \mathbf{w}|^2 + \frac{C^2 C^*}{2\lambda_1} \int_{\Omega} |\mathbf{F}|^2 |\mathbf{w}|^{s-2} \\
& \leq \lambda_1 \int_{\Omega} |\mathbf{w}|^{s-2} |\mathbf{D}(\mathbf{w})|^2 + \frac{C^2 C^*}{2\lambda_1} \int_{\Omega} |\mathbf{F}|^2 |\mathbf{w}|^{s-2},
\end{aligned}$$

provided that

$$\text{(A-4)} \quad CC^* \lambda_2 (s-2) + \lambda_1 / 2 < \lambda_1.$$

Thus, defining $L := 1/(2CC^*)$, we see that for all $2 \leq s \leq 2 + L\lambda_1/\lambda_2$ the condition (A-4) is automatically met and the inequality (A-3) implies that

$$\frac{d}{dt} \|\mathbf{w}\|_s^s \leq \frac{C}{\lambda_1} \int_{\Omega} |\mathbf{F}|^2 |\mathbf{w}|^{s-2} \leq \frac{C}{\lambda_1} \|\mathbf{F}\|_s^2 \|\mathbf{w}\|_s^{s-2}$$

and we finally obtain

$$\frac{d}{dt} \|\mathbf{w}\|_s^2 \leq \frac{C}{\lambda_1} \|\mathbf{F}\|_s^2,$$

which leads to (2-5) after integration with respect to $t \in (0, T)$. \square

Lemma A.1 (See also [Ladyzhenskaja et al. 1968]). *Let $T > 0$ and $\Omega \subset \mathbb{R}^2$. Assume that $p > 4$; then the embedding*

$$\text{(A-5)} \quad W^{1,p}(0, T; L^p) \cap L^p(0, T; W_{\text{per}}^{2,p}) \hookrightarrow \mathcal{C}([0, T], \mathcal{C}_{\text{per}}^1(\overline{\Omega})).$$

holds. In addition for any $s > 2$ and $\alpha \in (0, (1 - 2/s)/3)$ we have

$$\text{(A-6)} \quad W^{1,\infty}(0, T; L^s) \cap L^\infty(0, T; W^{2,s}) \hookrightarrow \mathcal{C}^{0,\alpha}([0, T], \mathcal{C}_{\text{per}}^{1,\alpha}(\overline{\Omega})).$$

Proof. By using an interpolation theorem (see [Amann 2000, proof of Corollary 4.5(ii)]) we find that, for any $\alpha \in [0, 1]$, $p_1, p_2 \in (1, \infty)$,

$$\text{(A-7)} \quad W^{1,p_1}(0, T; L^{p_2}) \cap L^{p_1}(0, T; W_{\text{per}}^{2,p_2}) \hookrightarrow W^{\alpha,p_1}(0, T; W_{\text{per}}^{2(1-\alpha),p_2}).$$

Consequently, setting $\alpha := \frac{1}{4}$, $p_1 = p_2 := p$ and using the standard Sobolev embedding we get (A-5), provided that $p > 4$.

To prove the second part of the lemma, we first note that for any $p \in [1, \infty]$ we have $W^{1,\infty}(0, T; L^s) \cap L^\infty(0, T; W^{2,s}) \hookrightarrow W^{1,p}(0, T; L^s) \cap L^p(0, T; W^{2,s})$.

Consequently, setting $p_2 := s$, $p_1 := p$ in (A-7), we deduce that

$$\text{(A-8)} \quad W^{1,\infty}(0, T; L^s) \cap L^\infty(0, T; W_{\text{per}}^{2,s}) \hookrightarrow W^{\alpha,p}(0, T; W_{\text{per}}^{2(1-\alpha),s})$$

for any $p \in (1, \infty)$ and any $\alpha \in [0, 1]$. Finally, assuming that $p > \frac{2s}{s-2}$ and setting

$$\alpha := \frac{ps - 2p + s}{3ps}$$

in (A-8), we get after using the standard embedding theorem that

$$W^{1,\infty}(0, T; L^s) \cap L^\infty(0, T; W_{\text{per}}^{2,s}) \hookrightarrow \mathcal{C}^{0,\beta}(0, T; \mathcal{C}_{\text{per}}^{1,\beta}(\bar{\Omega}))$$

with $\beta := \frac{1}{3} - \frac{2}{3s} - \frac{2}{3p}$. Since p is arbitrarily large the embedding (A-6) follows.

References

- [Amann 2000] H. Amann, “Compact embeddings of vector-valued Sobolev and Besov spaces”, *Glas. Mat. Ser. III* **35**:1 (2000), 161–177. [MR 2001h:46056](#) [Zbl 0997.46029](#)
- [Boyarskiĭ 1957] B. V. Boyarskiĭ, “Generalized solutions of a system of differential equations of first order and of elliptic type with discontinuous coefficients”, *Mat. Sb. N.S.* **43(85)** (1957), 451–503. In Russian. [MR 21 #5058](#)
- [Buliček et al. 2012] M. Buliček, J. Málek, and K. R. Rajagopal, “On Kelvin-Voigt model and its generalizations”, *Evolution Equations and Control Theory* **1**:1 (2012), 17–42.
- [Demoulini 2000] S. Demoulini, “Weak solutions for a class of nonlinear systems of viscoelasticity”, *Arch. Ration. Mech. Anal.* **155**:4 (2000), 299–334. [MR 2002d:74027](#) [Zbl 0991.74021](#)
- [DiBenedetto and Friedman 1984] E. DiBenedetto and A. Friedman, “Regularity of solutions of nonlinear degenerate parabolic systems”, *J. Reine Angew. Math.* **349** (1984), 83–128. [MR 85j:35089](#) [Zbl 0527.35038](#)
- [DiBenedetto and Friedman 1985] E. DiBenedetto and A. Friedman, “Hölder estimates for nonlinear degenerate parabolic systems”, *J. Reine Angew. Math.* **357** (1985), 1–22. [MR 87f:35134a](#) [Zbl 0549.35061](#)
- [Feireisl and Novotný 2009] E. Feireisl and A. Novotný, *Singular limits in thermodynamics of viscous fluids*, Birkhäuser, Basel, 2009. [MR 2011b:35001](#) [Zbl 1176.35126](#)
- [Frehse and Seregin 1999] J. Frehse and G. A. Seregin, “Full regularity for a class of degenerated parabolic systems in two spatial variables”, *Manuscripta Math.* **99**:4 (1999), 517–539. [MR 2000e:35117](#) [Zbl 0931.35029](#)
- [Friedman and Nečas 1988] A. Friedman and J. Nečas, “Systems of nonlinear wave equations with nonlinear viscosity”, *Pacific J. Math.* **135**:1 (1988), 29–55. [MR 90b:35152](#) [Zbl 0685.35070](#)
- [Fung 1993] Y.-C. Fung, *Biomechanics: Mechanical properties of living tissues*, 2nd ed., Springer, New York, 1993.
- [Kaplický et al. 2002] P. Kaplický, J. Málek, and J. Stará, “Global-in-time Hölder continuity of the velocity gradients for fluids with shear-dependent viscosities”, *NoDEA Nonlinear Differential Equations Appl.* **9**:2 (2002), 175–195. [MR 2003i:35233](#) [Zbl 0991.35066](#)
- [Ladyzhenskaja et al. 1968] O. A. Ladyzhenskaja, V. A. Solonnikov, and N. N. Ural’ceva, *Linear and quasilinear equations of parabolic type*, vol. 23, Translations of Mathematical Monographs, American Mathematical Society, Providence, 1968. [MR 39 #3159b](#)
- [Meyers 1963] N. G. Meyers, “An L^p -estimate for the gradient of solutions of second order elliptic divergence equations”, *Ann. Scuola Norm. Sup. Pisa* (3) **17** (1963), 189–206. [MR 28 #2328](#) [Zbl 0127.31904](#)

- [Nečas 1967] J. Nečas, “Sur la régularité des solutions variationnelles des équations elliptiques non-linéaires d’ordre $2k$ en deux dimensions”, *Ann. Scuola Norm. Sup. Pisa* (3) **21** (1967), 427–457. [MR 37 #2057](#) [Zbl 0171.09401](#)
- [Nečas and Šverák 1991] J. Nečas and V. Šverák, “On regularity of solutions of nonlinear parabolic systems”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **18**:1 (1991), 1–11. [MR 92d:35058](#) [Zbl 0735.35035](#)
- [Novotný and Straškraba 2004] A. Novotný and I. Straškraba, *Introduction to the mathematical theory of compressible flow*, Oxford Lecture Series in Mathematics and its Applications **27**, Oxford University Press, 2004. [MR 2005i:35220](#) [Zbl 1088.35051](#)
- [Rajagopal 2009] K. R. Rajagopal, “A note on a reappraisal and generalization of the Kelvin–Voigt model”, *Mech. Res. Comm.* **36**:2 (2009), 232–235.
- [Ramberg and Osgood 1943] W. Ramberg and W. R. Osgood, “Description of stress-strain curves by three parameters”, *Tech. Notes Nat. Adv. Comm. Aeronaut.* **1943**:902 (1943), 1–13. [MR 7,229h](#)
- [Schlag 1996] W. Schlag, “Schauder and L^p estimates for parabolic systems via Campanato spaces”, *Comm. Partial Differential Equations* **21**:7-8 (1996), 1141–1175. [MR 97k:35108](#) [Zbl 0864.35023](#)
- [Thomson 1865] W. Thomson, “On the elasticity and viscosity of metals”, *Proc. Roy. Soc. London* **14** (1865), 289–297.
- [Tvedt 2008] B. Tvedt, “Quasilinear equations for viscoelasticity of strain-rate type”, *Arch. Ration. Mech. Anal.* **189**:2 (2008), 237–281. [MR 2009m:74028](#) [Zbl 1147.74008](#)
- [Voigt 1892] W. Voigt, “Ueber innere Reibung fester Körper, insbesondere der Metalle”, *Annalen der Physik* **283**:12 (1892), 671–693. [JFM 24.0932.01](#)
- [Ziemer 1989] W. P. Ziemer, *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, Graduate Texts in Mathematics **120**, Springer, New York, 1989. [MR 91e:46046](#) [Zbl 0692.46022](#)

Received September 19, 2011. Revised November 8, 2012.

MIROSLAV BULÍČEK
MATHEMATICAL INSTITUTE, FACULTY OF MATHEMATICS AND PHYSICS
CHARLES UNIVERSITY IN PRAGUE
186 75 PRAHA 8
CZECH REPUBLIC
mbul8060@karlin.mff.cuni.cz

PETR KAPLICKÝ
DEPARTMENT OF MATHEMATICAL ANALYSIS, FACULTY OF MATHEMATICS AND PHYSICS
CHARLES UNIVERSITY IN PRAGUE
186 75 PRAHA 8
CZECH REPUBLIC
kaplicky@karlin.mff.cuni.cz

MARK STEINHAUER
MATHEMATICAL INSTITUTE
UNIVERSITY OF KOBLENZ-LANDAU
CAMPUS KOBLENZ
56070 KOBLENZ
GERMANY
steinhauerm@uni-koblenz.de

A LOWER BOUND FOR EIGENVALUES OF THE POLY-LAPLACIAN WITH ARBITRARY ORDER

QING-MING CHENG, XUERONG QI AND GUOXIN WEI

We study eigenvalues of the poly-Laplacian of arbitrary order on a bounded domain in an n -dimensional Euclidean space. We obtain a lower bound for these eigenvalues, significantly improving on that of Levine and Protter. In particular, the result of Melas (2003) is subsumed.

1. Introduction

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with piecewise smooth boundary $\partial\Omega$ in an n -dimensional Euclidean space \mathbb{R}^n . Let λ_i be the i -th eigenvalue of the Dirichlet eigenvalue problem of the poly-Laplacian with arbitrary order:

$$(1-1) \quad \begin{cases} (-\Delta)^l u = \lambda u & \text{in } \Omega, \\ u = \frac{\partial u}{\partial \nu} = \cdots = \frac{\partial^{l-1} u}{\partial \nu^{l-1}} = 0 & \text{on } \partial\Omega, \end{cases}$$

where Δ is the Laplacian in \mathbb{R}^n and ν denotes the outward unit normal vector field of the boundary $\partial\Omega$. It is well known that the spectrum of this eigenvalue problem is real and discrete:

$$0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \rightarrow +\infty,$$

where each λ_i has finite multiplicity and is repeated according to its multiplicity.

Let $V(\Omega)$ denote the volume of Ω and let B_n denote the volume of the unit ball in \mathbb{R}^n . When $l = 1$, the eigenvalue problem (1-1) is called a fixed membrane problem. In this case, one has Weyl's asymptotic formula

$$(1-2) \quad \lambda_k \sim \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}}, \quad k \rightarrow +\infty.$$

Cheng was partially supported by JSPS Grant-in-Aid for Scientific Research (B) no. 24340013. Qi was partially supported by NSFC grant no. 11171091. Wei was partially supported by NSFC grant no. 11001087 and the project of Pear River New Star of Guangzhou (grant no. 2012J2200028). Wei is the corresponding author.

MSC2010: 35P15.

Keywords: eigenvalue problem, lower bound for eigenvalues, poly-Laplacian with arbitrary order.

From the above asymptotic formula, one can obtain

$$(1-3) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \sim \frac{n}{n+2} \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}}, \quad k \rightarrow +\infty.$$

Pólya [1961] proved that

$$(1-4) \quad \lambda_k \geq \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}}, \quad \text{for } k = 1, 2, \dots,$$

if Ω is a tiling domain in \mathbb{R}^n . Moreover, he proposed the following:

Conjecture of Pólya. If Ω is a bounded domain in \mathbb{R}^n , then the k -th eigenvalue λ_k of the fixed membrane problem satisfies

$$(1-5) \quad \lambda_k \geq \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}}, \quad \text{for } k = 1, 2, \dots$$

Berezin [1972] and Lieb [1980] gave a partial solution to this conjecture. Li and Yau [1983] proved that

$$(1-6) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \geq \frac{n}{n+2} \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}}, \quad \text{for } k = 1, 2, \dots$$

Formula (1-3) shows that (1-6) is sharp in the sense of averages. From (1-6), one can derive

$$(1-7) \quad \lambda_k \geq \frac{n}{n+2} \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}}, \quad \text{for } k = 1, 2, \dots,$$

which gives a partial solution for the conjecture of Pólya with a factor $\frac{n}{n+2}$. Melas [2003] has improved the estimate (1-6) to

$$(1-8) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \geq \frac{n}{n+2} \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} k^{\frac{2}{n}} + \frac{1}{24(n+2)} \frac{V(\Omega)}{I(\Omega)}, \quad \text{for } k = 1, 2, \dots,$$

where

$$I(\Omega) = \min_{a \in \mathbb{R}^n} \int_{\Omega} |x - a|^2 dx$$

is called *the moment of inertia* of Ω .

When $l = 2$, the eigenvalue problem (1-1) is called the clamped plate problem. For the eigenvalues of the clamped plate problem, it follows from [Agmon 1965] and [Pleijel 1950] that

$$(1-9) \quad \lambda_k \sim \frac{16\pi^4}{(B_n V(\Omega))^{\frac{4}{n}}} k^{\frac{4}{n}}, \quad k \rightarrow +\infty.$$

This implies that

$$(1-10) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \sim \frac{n}{n+4} \frac{16\pi^4}{(B_n V(\Omega))^{\frac{4}{n}}} k^{\frac{4}{n}}, \quad k \rightarrow +\infty.$$

Furthermore, Levine and Protter [1985] proved that the eigenvalues of the clamped plate problem satisfy

$$(1-11) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \geq \frac{n}{n+4} \frac{16\pi^4}{(B_n V(\Omega))^{\frac{4}{n}}} k^{\frac{4}{n}}.$$

Formula (1-10) shows that the coefficient of $k^{\frac{4}{n}}$ is the best possible constant. Very recently, Cheng and Wei [2011] obtained the following improvement of (1-11):

$$(1-12) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \geq \frac{n}{n+4} \frac{16\pi^4}{(B_n V(\Omega))^{\frac{4}{n}}} k^{\frac{4}{n}} + c_n \frac{n}{n+2} \frac{4\pi^2}{(B_n V(\Omega))^{\frac{2}{n}}} \frac{V(\Omega)}{I(\Omega)} k^{\frac{2}{n}} + d_n \left(\frac{V(\Omega)}{I(\Omega)} \right)^2,$$

where c_n and d_n are constants depending only on the dimension n .

When $l \geq 3$, Levine and Protter [1985] proved that

$$(1-13) \quad \frac{1}{k} \sum_{i=1}^k \lambda_i \geq \frac{n}{n+2l} \frac{(2\pi)^{2l}}{(B_n V(\Omega))^{\frac{2l}{n}}} k^{\frac{2l}{n}}, \quad \text{for } k = 1, 2, \dots$$

From the above formula, one can obtain

$$(1-14) \quad \lambda_k \geq \frac{n}{n+2l} \frac{(2\pi)^{2l}}{(B_n V(\Omega))^{\frac{2l}{n}}} k^{\frac{2l}{n}}, \quad \text{for } k = 1, 2, \dots$$

In this paper we investigate eigenvalues of the Dirichlet eigenvalue problem (1-1) for the Laplacian with any order. We give a significant improvement of (1-13) by adding l lower-order terms than $k^{2l/n}$ to its right-hand side. In fact, we prove:

Theorem. *Let Ω be a bounded domain in an n -dimensional Euclidean space \mathbb{R}^n . Let $\lambda_i, i = 1, 2, \dots$, be the i -th eigenvalue of the eigenvalue problem (1-1). Then*

$$\begin{aligned} \frac{1}{k} \sum_{j=1}^k \lambda_j &\geq \frac{n}{n+2l} \frac{(2\pi)^{2l}}{(B_n V(\Omega))^{\frac{2l}{n}}} k^{\frac{2l}{n}} \\ &+ \frac{n}{n+2l} \sum_{p=1}^l \frac{(l+1-p)}{(24)^p n \dots (n+2p-2)} \frac{(2\pi)^{2(l-p)}}{(B_n V(\Omega))^{\frac{2(l-p)}{n}}} \left(\frac{V(\Omega)}{I(\Omega)} \right)^p k^{\frac{2(l-p)}{n}}. \end{aligned}$$

Remark. If we take $l = 1$, we obtain the inequality (1-8).

2. Proof of the Theorem

Before giving the proof, we introduce some definitions and basic facts about symmetric decreasing rearrangements.

For a bounded domain $\Omega \subset \mathbb{R}^n$, the *moment of inertia* of Ω is defined by

$$I(\Omega) = \min_{a \in \mathbb{R}^n} \int_{\Omega} |x - a|^2 dx.$$

By translating the origin, we may assume that

$$I(\Omega) = \int_{\Omega} |x|^2 dx.$$

Let Ω^* be the symmetric rearrangement of Ω , that is, Ω^* is the open ball centered at the origin with the same volume as Ω . Then

$$\Omega^* = \left\{ x \in \mathbb{R}^n; |x| < \left(\frac{V(\Omega)}{B_n} \right)^{\frac{1}{n}} \right\}.$$

By using the symmetric rearrangement Ω^* of Ω , we have

$$(2-1) \quad I(\Omega) = \int_{\Omega} |x|^2 dx \geq \int_{\Omega^*} |x|^2 dx = \frac{n}{n+2} V(\Omega) \left(\frac{V(\Omega)}{B_n} \right)^{\frac{2}{n}}.$$

Let f be a nonnegative continuous function on Ω . We consider its *distribution function* $\mu_f(t)$ defined by

$$\mu_f(t) = \text{Vol}(\{x \in \Omega; f(x) > t\}).$$

The distribution function can be viewed as a function from $[0, +\infty)$ to $[0, V(\Omega)]$. The *symmetric decreasing rearrangement* f^* of f is defined by

$$f^*(x) = \inf \{ t \geq 0; \mu_f(t) < B_n |x|^n \}, \quad \text{for } x \in \Omega^*.$$

By definition, we know that $f^*(x)$ is a radially symmetric function and

$$\text{Vol}(\{x \in \Omega; f(x) > t\}) = \text{Vol}(\{x \in \Omega^*; f^*(x) > t\}) \quad \text{for all } t > 0.$$

Let $f^*(x) = \phi(|x|)$. Then one gets that $\phi : [0, +\infty) \rightarrow [0, \sup f]$ is a decreasing function of $|x|$. We may assume that ϕ is absolutely continuous. It is well known that

$$(2-2) \quad \int_{\Omega} f(x) dx = \int_{\Omega^*} f^*(x) dx = n B_n \int_0^{+\infty} s^{n-1} \phi(s) ds$$

and

$$(2-3) \quad \int_{\Omega} |x|^{2l} f(x) dx \geq \int_{\Omega^*} |x|^{2l} f^*(x) dx = n B_n \int_0^{+\infty} s^{n+2l-1} \phi(s) ds.$$

Good sources of further information on rearrangements are [Bandle 1980; Pólya and Szegő 1951].

One gets from the coarea formula that

$$\mu_f(t) = \int_t^{\sup f} \int_{\{f=s\}} |\nabla f|^{-1} d\sigma_s ds.$$

Since f^* is radial, we have

$$\begin{aligned} \mu_f(\phi(s)) &= \text{Vol} \{x \in \Omega; f(x) > \phi(s)\} = \text{Vol} \{x \in \Omega^*; f^*(x) > \phi(s)\} \\ &= \text{Vol} \{x \in \Omega^*; \phi(|x|) > \phi(s)\} = B_n s^n. \end{aligned}$$

It follows that

$$n B_n s^{n-1} = \mu'_f(\phi(s)) \phi'(s)$$

for almost every s . Putting $\tau := \sup |\nabla f|$, we obtain from the above equations and the isoperimetric inequality that

$$-\mu'_f(\phi(s)) = \int_{\{f=\phi(s)\}} |\nabla f|^{-1} d\sigma_{\phi(s)} \geq \tau^{-1} \text{Vol}_{n-1}(\{f=\phi(s)\}) \geq \tau^{-1} n B_n s^{n-1}.$$

Therefore, one obtains, for almost every s ,

$$(2-4) \quad -\tau \leq \phi'(s) \leq 0.$$

In order to prove our theorem, we need the following lemma.

Lemma. *Let $b \geq 1$ and $\eta, A > 0$, and let $\psi : [0, +\infty) \rightarrow [0, +\infty)$ be a decreasing, absolutely continuous function such that*

$$-\eta \leq \psi'(s) \leq 0, \quad A = \int_0^{+\infty} s^{b-1} \psi(s) ds.$$

For any positive integer l , let

$$A_l := \int_0^{+\infty} s^{b+2l-1} \psi(s) ds.$$

Then, we have

$$A_l \geq \frac{1}{b+2l} \left[(bA)^{\frac{b+2l}{b}} \psi(0)^{-\frac{2l}{b}} + \sum_{p=1}^l \frac{(l+1-p)(bA)^{\frac{b+2(l-p)}{b}} \psi(0)^{\frac{2pb-2(l-p)}{b}}}{6pb \cdots (b+2p-2)\eta^{2p}} \right].$$

Proof. The proof is by induction. Firstly, one can get from the lemma of [Melas 2003] that

$$(2-5) \quad A_1 = \int_0^{+\infty} s^{b+1} \psi(s) ds \geq \frac{1}{b+2} \left[(bA)^{\frac{b+2}{b}} \psi(0)^{-\frac{2}{b}} + \frac{A\psi(0)^2}{6\eta^2} \right].$$

To prove the induction step, we assume the statement holds for $l = r$, that is,

$$A_r \geq \frac{1}{b+2r} \left[(bA)^{\frac{b+2r}{b}} \psi(0)^{-\frac{2r}{b}} + \sum_{p=1}^r \frac{(r+1-p)(bA)^{\frac{b+2(r-p)}{b}} \psi(0)^{\frac{2pb-2(r-p)}{b}}}{6^p b \cdots (b+2p-2)\eta^{2p}} \right].$$

Since the formula (2-5) holds for any $b \geq 1$, we have

$$\begin{aligned} & A_{r+1} \\ &= \int_0^{+\infty} s^{b+2r+1} \psi(s) ds \\ &\geq \frac{1}{b+2r+2} \left\{ [(b+2r)A_r]^{\frac{b+2r+2}{b+2r}} \psi(0)^{-\frac{2}{b+2r}} + \frac{A_r \psi(0)^2}{6\eta^2} \right\} \\ &\geq \frac{\psi(0)^{-\frac{2}{b+2r}}}{b+2r+2} \\ &\quad \times \left[(bA)^{\frac{b+2r}{b}} \psi(0)^{-\frac{2r}{b}} + \sum_{p=1}^r \frac{(r+1-p)(bA)^{\frac{b+2(r-p)}{b}} \psi(0)^{\frac{2pb-2(r-p)}{b}}}{6^p b \cdots (b+2p-2)\eta^{2p}} \right]^{\frac{b+2r+2}{b+2r}} \\ &\quad + \frac{1}{(b+2r)(b+2r+2)} \sum_{p=1}^r \frac{(r+1-p)(bA)^{\frac{b+2(r-p)}{b}} \psi(0)^{\frac{2(p+1)b-2(r-p)}{b}}}{6^{p+1} b \cdots (b+2p-2)\eta^{2p+2}} \\ &\quad + \frac{(bA)^{\frac{b+2r}{b}} \psi(0)^{\frac{2b-2r}{b}}}{6(b+2r)(b+2r+2)\eta^2} \\ &= \frac{\psi(0)^{-\frac{2}{b+2r}}}{b+2r+2} \left[(bA)^{\frac{b+2r}{b}} \psi(0)^{-\frac{2r}{b}} \right]^{\frac{b+2r+2}{b+2r}} \\ &\quad \times \left[1 + \sum_{p=1}^r \frac{(r+1-p)(bA)^{\frac{-2p}{b}} \psi(0)^{\frac{2pb+2p}{b}}}{6^p b \cdots (b+2p-2)\eta^{2p}} \right]^{\frac{b+2r+2}{b+2r}} + \frac{(bA)^{\frac{b+2r}{b}} \psi(0)^{\frac{2b-2r}{b}}}{6(b+2r)(b+2r+2)\eta^2} \\ &\quad + \frac{1}{(b+2r)(b+2r+2)} \sum_{p=2}^{r+1} \frac{(r+2-p)(bA)^{\frac{b+2r-2p+2}{b}} \psi(0)^{\frac{2pb-2r+2p-2}{b}}}{6^p b \cdots (b+2p-4)\eta^{2p}} \\ &= \frac{(bA)^{\frac{b+2r+2}{b}} \psi(0)^{-\frac{2r+2}{b}}}{b+2r+2} \left[1 + \sum_{p=1}^r \frac{(r+1-p)(bA)^{\frac{-2p}{b}} \psi(0)^{\frac{2pb+2p}{b}}}{6^p b \cdots (b+2p-2)\eta^{2p}} \right]^{\frac{b+2r+2}{b+2r}} \\ &\quad + \frac{(bA)^{\frac{b+2r}{b}} \psi(0)^{\frac{2b-2r}{b}}}{6(b+2r)(b+2r+2)\eta^2} + \frac{(bA)\psi(0)^{2(r+1)}}{6^{r+1} b \cdots (b+2r+2)\eta^{2(r+1)}} \\ &\quad + \frac{1}{(b+2r)(b+2r+2)} \sum_{p=2}^r \frac{(r+2-p)(bA)^{\frac{b+2(r+1-p)}{b}} \psi(0)^{\frac{2pb-2(r+1-p)}{b}}}{6^p b \cdots (b+2p-4)\eta^{2p}}. \end{aligned}$$

It follows from the Taylor formula that

$$\begin{aligned}
& A_{r+1} \\
& \geq \frac{1}{b+2r+2} (bA)^{\frac{b+2r+2}{b}} \psi(0)^{-\frac{2r+2}{b}} \\
& \quad \times \left[1 + \frac{b+2r+2}{b+2r} \sum_{p=1}^r \frac{(r+1-p)(bA)^{-\frac{2p}{b}} \psi(0)^{\frac{2pb+2p}{b}}}{6^p b \cdots (b+2p-2) \eta^{2p}} \right] \\
& \quad + \frac{(bA)^{\frac{b+2r}{b}} \psi(0)^{\frac{2b-2r}{b}}}{6(b+2r)(b+2r+2) \eta^2} + \frac{(bA) \psi(0)^{2(r+1)}}{6^{r+1} b \cdots (b+2r+2) \eta^{2(r+1)}} \\
& \quad + \frac{1}{(b+2r)(b+2r+2)} \sum_{p=2}^r \frac{(r+2-p)(bA)^{\frac{b+2(r+1-p)}{b}} \psi(0)^{\frac{2pb-2(r+1-p)}{b}}}{6^p b \cdots (b+2p-4) \eta^{2p}} \\
& = \frac{1}{b+2r+2} (bA)^{\frac{b+2r+2}{b}} \psi(0)^{-\frac{2r+2}{b}} \\
& \quad + \frac{1}{b+2r} \sum_{p=1}^r \frac{(r+1-p)(bA)^{\frac{b+2(r+1-p)}{b}} \psi(0)^{\frac{2pb-2(r+1-p)}{b}}}{6^p b \cdots (b+2p-2) \eta^{2p}} \\
& \quad + \frac{(bA)^{\frac{b+2r}{b}} \psi(0)^{\frac{2b-2r}{b}}}{6(b+2r)(b+2r+2) \eta^2} + \frac{(bA) \psi(0)^{2(r+1)}}{6^{r+1} b \cdots (b+2r+2) \eta^{2(r+1)}} \\
& \quad + \frac{1}{(b+2r)(b+2r+2)} \sum_{p=2}^r \frac{(r+2-p)(bA)^{\frac{b+2(r+1-p)}{b}} \psi(0)^{\frac{2pb-2(r+1-p)}{b}}}{6^p b \cdots (b+2p-4) \eta^{2p}} \\
& = \frac{1}{b+2r+2} (bA)^{\frac{b+2r+2}{b}} \psi(0)^{-\frac{2r+2}{b}} \\
& \quad + \left[\frac{r}{b(b+2r)} + \frac{1}{(b+2r)(b+2r+2)} \right] \frac{1}{6 \eta^2} (bA)^{\frac{b+2r}{b}} \psi(0)^{\frac{2b-2r}{b}} \\
& \quad + \sum_{p=2}^r \left[\frac{r+1-p}{b+2r} + \frac{(r+2-p)(b+2p-2)}{(b+2r)(b+2r+2)} \right] \frac{(bA)^{\frac{b+2(r+1-p)}{b}} \psi(0)^{\frac{2pb-2(r+1-p)}{b}}}{6^p b \cdots (b+2p-2) \eta^{2p}} \\
& \quad + \frac{(bA) \psi(0)^{2(r+1)}}{6^{r+1} b \cdots (b+2r+2) \eta^{2(r+1)}} \\
& \geq \frac{1}{b+2(r+1)} (bA)^{\frac{b+2(r+1)}{b}} \psi(0)^{-\frac{2(r+1)}{b}} \\
& \quad + \frac{1}{b+2(r+1)} \sum_{p=1}^{r+1} \frac{(r+2-p)(bA)^{\frac{b+2(r+1-p)}{b}} \psi(0)^{\frac{2pb-2(r+1-p)}{b}}}{6^p b \cdots (b+2p-2) \eta^{2p}}.
\end{aligned}$$

This completes the proof of the lemma. \square

Proof of the Theorem. Let u_j be an orthonormal eigenfunction corresponding to the eigenvalue λ_j , that is, u_j satisfies

$$(2-6) \quad \begin{cases} (-\Delta)^l u_j = \lambda_j u_j, & \text{in } \Omega, \\ u_j = \frac{\partial u_j}{\partial \nu} = \dots = \frac{\partial^{l-1} u_j}{\partial \nu^{l-1}} = 0, & \text{on } \partial\Omega, \\ \int_{\Omega} u_i u_j = \delta_{ij}, & \text{for any } i, j. \end{cases}$$

Thus, $\{u_j\}_{j=1}^{\infty}$ forms an orthonormal basis of $L^2(\Omega)$. We define a function φ_j by

$$(2-7) \quad \varphi_j(x) = \begin{cases} u_j(x), & x \in \Omega, \\ 0, & x \in \mathbb{R}^n \setminus \Omega. \end{cases}$$

The Fourier transform $\widehat{\varphi}_j(z)$ of $\varphi_j(x)$ is then given by

$$(2-8) \quad \widehat{\varphi}_j(z) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} \varphi_j(x) e^{i\langle x, z \rangle} dx = (2\pi)^{-n/2} \int_{\Omega} u_j(x) e^{i\langle x, z \rangle} dx.$$

We fix $k \geq 1$ and set

$$f(z) = \sum_{j=1}^k |\widehat{\varphi}_j(z)|^2, \quad \text{for } z \in \mathbb{R}^n.$$

From Bessel's inequality, it follows that

$$(2-9) \quad \begin{aligned} 0 \leq f(z) &= \sum_{j=1}^k |\widehat{\varphi}_j(z)|^2 = (2\pi)^{-n} \sum_{j=1}^k \left| \int_{\Omega} u_j(x) e^{i\langle x, z \rangle} dx \right|^2 \\ &\leq (2\pi)^{-n} \int_{\Omega} |e^{i\langle x, z \rangle}|^2 dx = (2\pi)^{-n} V(\Omega). \end{aligned}$$

By Parseval's identity, we have

$$(2-10) \quad \begin{aligned} \int_{\mathbb{R}^n} f(z) dz &= \sum_{j=1}^k \int_{\mathbb{R}^n} |\widehat{\varphi}_j(z)|^2 dz = \sum_{j=1}^k \int_{\mathbb{R}^n} \varphi_j^2(x) dx \\ &= \sum_{j=1}^k \int_{\Omega} u_j^2(x) dx = k. \end{aligned}$$

Furthermore, we deduce from integration by parts and Parseval's identity that

$$\begin{aligned}
(2-11) \quad & \int_{\mathbb{R}^n} |z|^{2l} f(z) dz \\
&= \sum_{j=1}^k \int_{\mathbb{R}^n} |z|^{2l} |\widehat{\varphi}_j(z)|^2 dz \\
&= \sum_{j=1}^k \int_{\mathbb{R}^n} |z|^{2l} \left| (2\pi)^{-n/2} \int_{\Omega} u_j(x) e^{i\langle x, z \rangle} dx \right|^2 dz \\
&= \sum_{j=1}^k \sum_{r_1, \dots, r_l=1}^n \int_{\mathbb{R}^n} \left| (2\pi)^{-n/2} \int_{\Omega} z_{r_1} \cdots z_{r_l} u_j(x) e^{i\langle x, z \rangle} dx \right|^2 dz \\
&= \sum_{j=1}^k \sum_{r_1, \dots, r_l=1}^n \int_{\mathbb{R}^n} \left| (2\pi)^{-n/2} \int_{\Omega} u_j(x) \frac{\partial^l e^{i\langle x, z \rangle}}{\partial x_{r_1} \cdots \partial x_{r_l}} dx \right|^2 dz \\
&= \sum_{j=1}^k \sum_{r_1, \dots, r_l=1}^n \int_{\mathbb{R}^n} \left| (2\pi)^{-n/2} \int_{\Omega} \frac{\partial^l u_j(x)}{\partial x_{r_1} \cdots \partial x_{r_l}} e^{i\langle x, z \rangle} dx \right|^2 dz \\
&= \sum_{j=1}^k \sum_{r_1, \dots, r_l=1}^n \int_{\mathbb{R}^n} \left| \frac{\partial^l u_j}{\partial x_{r_1} \cdots \partial x_{r_l}} \right|^2 dz \\
&= \sum_{j=1}^k \sum_{r_1, \dots, r_l=1}^n \int_{\mathbb{R}^n} \left(\frac{\partial^l u_j}{\partial x_{r_1} \cdots \partial x_{r_l}} \right)^2 dx \\
&= \sum_{j=1}^k \int_{\Omega} u_j (-\Delta)^l u_j dx = \sum_{j=1}^k \lambda_j.
\end{aligned}$$

Since

$$(2-12) \quad \nabla \widehat{\varphi}_j(z) = (2\pi)^{-n/2} \int_{\Omega} i x u_j(x) e^{i\langle x, z \rangle} dx,$$

we obtain from Bessel's inequality that

$$(2-13) \quad \sum_{j=1}^k |\nabla \widehat{\varphi}_j(z)|^2 \leq (2\pi)^{-n} \int_{\Omega} |i x e^{i\langle x, z \rangle}|^2 dx = (2\pi)^{-n} I(\Omega).$$

It follows from (2-9), (2-13) and the Cauchy–Schwarz inequality that, for every $z \in \mathbb{R}^n$,

$$\begin{aligned}
(2-14) \quad & |\nabla f(z)| \leq 2 \left(\sum_{j=1}^k |\widehat{\varphi}_j(z)|^2 \right)^{1/2} \left(\sum_{j=1}^k |\nabla \widehat{\varphi}_j(z)|^2 \right)^{1/2} \\
& \leq 2(2\pi)^{-n} \sqrt{V(\Omega) I(\Omega)}.
\end{aligned}$$

Using the symmetric decreasing rearrangement f^* of f and noting that

$$f^*(x) = \phi(|x|), \quad \tau = \sup |\nabla f| \leq 2(2\pi)^{-n} \sqrt{V(\Omega)I(\Omega)} := \eta,$$

we obtain, from (2-4),

$$(2-15) \quad -\eta \leq -\tau \leq \phi'(s) \leq 0$$

for almost every s . According to (2-2) and (2-10), we infer

$$(2-16) \quad k = \int_{\mathbb{R}^n} f(z) dz = \int_{\mathbb{R}^n} f^*(z) dz = nB_n \int_0^{+\infty} s^{n-1} \phi(s) ds.$$

From (2-3) and (2-11), we obtain

$$(2-17) \quad \begin{aligned} \sum_{j=1}^k \lambda_j &= \int_{\mathbb{R}^n} |z|^{2l} f(z) dz \geq \int_{\mathbb{R}^n} |z|^{2l} f^*(z) dz \\ &= nB_n \int_0^{+\infty} s^{n+2l-1} \phi(s) ds. \end{aligned}$$

Now, we can apply the [Lemma](#) to the function ϕ with

$$(2-18) \quad b = n, \quad A = \frac{k}{nB_n}, \quad \eta = 2(2\pi)^{-n} \sqrt{V(\Omega)I(\Omega)}.$$

We conclude that

$$(2-19) \quad \begin{aligned} \sum_{j=1}^k \lambda_j &\geq \frac{nB_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} \phi(0)^{-\frac{2l}{n}} \\ &\quad + \frac{nB_n}{n+2l} \sum_{p=1}^l \frac{(l+1-p)}{6^p n \cdots (n+2p-2) \eta^{2p}} \left(\frac{k}{B_n}\right)^{\frac{n+2l-2p}{n}} \phi(0)^{\frac{2pn+2p-2l}{n}}. \end{aligned}$$

Note that $0 < \phi(0) \leq \sup f \leq (2\pi)^{-n} V(\Omega)$. Hence we consider the function F defined by

$$(2-20) \quad \begin{aligned} F(t) &= \frac{nB_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} t^{-\frac{2l}{n}} \\ &\quad + \frac{nB_n}{n+2l} \sum_{p=1}^l \frac{(l+1-p)}{6^p n \cdots (n+2p-2) \eta^{2p}} \left(\frac{k}{B_n}\right)^{\frac{n+2l-2p}{n}} t^{\frac{2pn+2p-2l}{n}}, \end{aligned}$$

for $t \in (0, (2\pi)^{-n} V(\Omega)]$. From (2-1), we have

$$(2-21) \quad \eta \geq (2\pi)^{-n} B_n^{-\frac{1}{n}} V(\Omega)^{\frac{n+1}{n}}.$$

By a direct calculation, one gets from $B_n = \frac{2\pi^{n/2}}{n\Gamma(n/2)}$ that

$$(2-22) \quad \frac{B_n^{4/n}}{(2\pi)^2} < \frac{1}{2},$$

where $\Gamma(\frac{n}{2})$ is the gamma function. Thus, it follows from (2-21) and (2-22) that

$$\begin{aligned} F'(t) &= \frac{2B_n t^{-\frac{n+2l}{n}}}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} \left[-l + \sum_{p=1}^l \frac{(l+1-p)(pn+p-l)t^{\frac{2p(n+1)}{n}}}{6^p n \cdots (n+2p-2)\eta^{2p}} \left(\frac{k}{B_n}\right)^{-\frac{2p}{n}} \right] \\ &\leq \frac{2B_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} t^{-\frac{n+2l}{n}} \left[-l + \sum_{p>\frac{l}{n+1}}^l \frac{(l+1-p)(pn+p-l)}{6^p n \cdots (n+2p-2)} \left(\frac{B_n^{\frac{4}{n}}}{(2\pi)^2}\right)^p \right] \\ &< \frac{2B_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} t^{-\frac{n+2l}{n}} \left[-l + \sum_{p>\frac{l}{n+1}}^l \frac{(l+1-p)(pn+p-l)}{(12)^p n \cdots (n+2p-2)} \right] \\ &< \frac{2B_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} t^{-\frac{n+2l}{n}} \left[-l + \frac{l(n+1-l)}{12n} + \sum_{\substack{p>\frac{l}{n+1} \\ p \neq 1}}^l \frac{p^2 n(n+1)}{(12)^p n \cdots (n+2p-2)} \right] \\ &< \frac{2B_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} t^{-\frac{n+2l}{n}} \left[-l + \frac{l}{12} + \sum_{\substack{p>\frac{l}{n+1} \\ p \neq 1}}^l \frac{p^2}{(12)^p} \right] \\ &< \frac{2B_n}{n+2l} \left(\frac{k}{B_n}\right)^{\frac{n+2l}{n}} t^{-\frac{n+2l}{n}} \left[-l + \frac{l}{12} + \frac{1}{12} \right] < 0. \end{aligned}$$

We obtain that $F(t)$ is a decreasing function on $(0, (2\pi)^{-n}V(\Omega)]$. Then we can replace $\phi(0)$ by $(2\pi)^{-n}V(\Omega)$ in (2-19), namely,

$$\begin{aligned} \sum_{j=1}^k \lambda_j &\geq \frac{n}{n+2l} \frac{(2\pi)^{2l}}{(B_n V(\Omega))^{\frac{2l}{n}}} k^{\frac{n+2l}{n}} \\ &\quad + \frac{n}{n+2l} \sum_{p=1}^l \frac{(l+1-p)}{6^p n \cdots (n+2p-2)\eta^{2p}} \frac{(V(\Omega))^{\frac{2pn+2p-2l}{n}}}{(2\pi)^{2pn+2p-2n} B_n^{\frac{2l-2p}{n}}} k^{\frac{n+2l-2p}{n}} \\ &= \frac{n}{n+2l} \frac{(2\pi)^{2l}}{(B_n V(\Omega))^{\frac{2l}{n}}} k^{\frac{n+2l}{n}} \\ &\quad + \frac{n}{n+2l} \sum_{p=1}^l \frac{(l+1-p)}{24^p n \cdots (n+2p-2)} \frac{(2\pi)^{2(l-p)}}{(B_n V(\Omega))^{\frac{2(l-p)}{n}}} \left(\frac{V(\Omega)}{I(\Omega)}\right)^p k^{\frac{n+2(l-p)}{n}}. \end{aligned}$$

This completes the proof of the [Theorem](#). □

References

- [Agmon 1965] S. Agmon, “On kernels, eigenvalues, and eigenfunctions of operators related to elliptic problems”, *Comm. Pure Appl. Math.* **18** (1965), 627–663. [MR 33 #6446](#) [Zbl 0151.20203](#)
- [Bandle 1980] C. Bandle, *Isoperimetric inequalities and applications*, Monographs and Studies in Mathematics **7**, Pitman, Boston, 1980. [MR 81e:35095](#) [Zbl 0436.35063](#)
- [Berezin 1972] F. A. Berezin, “Ковариантные и контравариантные символы операторов”, *Izv. Akad. Nauk SSSR Ser. Mat.* **36:5** (1972), 1134–1167. Translated as “Covariant and contravariant symbols of operators” in *Math. USSR-Izv.* **6:5** (1972), 1117–1151. [MR 50 #2996](#) [Zbl 0259.47004](#)
- [Cheng and Wei 2011] Q.-M. Cheng and G. Wei, “A lower bound for eigenvalues of a clamped plate problem”, *Calc. Var. Partial Differential Equations* **42:3-4** (2011), 579–590. [MR 2012i:35070](#) [Zbl 1234.35158](#)
- [Levine and Protter 1985] H. A. Levine and M. H. Protter, “Unrestricted lower bounds for eigenvalues for classes of elliptic equations and systems of equations with applications to problems in elasticity”, *Math. Methods Appl. Sci.* **7:2** (1985), 210–222. [MR 87d:35101](#) [Zbl 0591.35050](#)
- [Li and Yau 1983] P. Li and S. T. Yau, “On the Schrödinger equation and the eigenvalue problem”, *Comm. Math. Phys.* **88:3** (1983), 309–318. [MR 84k:58225](#) [Zbl 0554.35029](#)
- [Lieb 1980] E. H. Lieb, “The number of bound states of one-body Schrödinger operators and the Weyl problem”, pp. 241–252 in *Geometry of the Laplace operator* (Honolulu, HI, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, American Mathematical Society, Providence, RI, 1980. Also in *The stability of matter: from atoms to stars, selecta of Elliott H. Lieb*, 4th ed., edited by Thirring, W., 245–256, Springer, Berlin, 2005. [MR 82i:35134](#) [Zbl 0445.58029](#)
- [Melas 2003] A. D. Melas, “A lower bound for sums of eigenvalues of the Laplacian”, *Proc. Amer. Math. Soc.* **131:2** (2003), 631–636. [MR 2003i:35218](#) [Zbl 1015.58011](#)
- [Pleijel 1950] Å. Pleijel, “On the eigenvalues and eigenfunctions of elastic plates”, *Comm. Pure Appl. Math.* **3** (1950), 1–10. [MR 12,265a](#) [Zbl 0040.05403](#)
- [Pólya 1961] G. Pólya, “On the eigenvalues of vibrating membranes”, *Proc. London Math. Soc.* (3) **11** (1961), 419–433. [MR 23 #B2256](#) [Zbl 0107.41805](#)
- [Pólya and Szegő 1951] G. Pólya and G. Szegő, *Isoperimetric inequalities in mathematical physics*, Annals of Mathematics Studies **27**, Princeton University Press, Princeton, NJ, 1951. [MR 13,270d](#) [Zbl 0044.38301](#)

Received December 14, 2010.

QING-MING CHENG
DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF SCIENCES
FUKUOKA UNIVERSITY
FUKUOKA 814-0180
JAPAN
cheng@fukuoka-u.ac.jp

XUERONG QI
DEPARTMENT OF MATHEMATICS
ZHENGZHOU UNIVERSITY
450001 ZHENGZHOU
CHINA
xrqi@zzu.edu.cn

GUOXIN WEI
SCHOOL OF MATHEMATICAL SCIENCES
SOUTH CHINA NORMAL UNIVERSITY
510631 GUANGZHOU
CHINA
weiguoxin@tsinghua.org.cn
weigx@scnu.edu.cn

QUIVER ALGEBRAS, PATH COALGEBRAS AND COREFLEXIVITY

SORIN DĂSCĂLESCU, MIODRAG C. IOVANOV
AND CONSTANTIN NĂSTĂSESCU

Dedicated to our friend Margaret Beattie on the occasion of her retirement

We study the connection between two combinatorial notions associated to a quiver: the quiver algebra and the path coalgebra. We show that the quiver coalgebra can be recovered from the quiver algebra as a certain type of finite dual, and we show precisely when the path coalgebra is the classical finite dual of the quiver algebra, and when all finite-dimensional quiver representations arise as comodules over the path coalgebra. We discuss when the quiver algebra can be recovered as the rational part of the dual of the path coalgebra. Similar results are obtained for incidence (co)algebras. We also study connections to the notion of coreflexive (co)algebras, and give a partial answer to an open problem concerning tensor products of coreflexive coalgebras.

1. Introduction and preliminaries

Let Γ be a quiver, and let K be an arbitrary ground field, which will be fixed throughout the paper. The associated quiver algebra $K[\Gamma]$ is an important object studied extensively in representation theory, and one theme in the field is to relate and understand combinatorial properties of the quiver via the properties of the category of representations of the quiver, and vice versa. Quiver algebras also play a role in general representation theory of algebras; for example, every finite-dimensional pointed algebra is a quiver algebra “with relations”. A closely related object is the path coalgebra $K\Gamma$, introduced in [Chin and Montgomery 1997], together with its comodules (quiver corepresentations). Comodules over path coalgebras turn out to form a special kind of representations of the quiver, called locally nilpotent representations in [Chin et al. 2002]. A natural question arises then: what is the precise connection between the two objects $K[\Gamma]$ and $K\Gamma$. We aim to provide such connections, by finding out when one of these objects can be recovered from the

MSC2010: 05C38, 06A11, 16T05, 16T15, 16T30.

Keywords: quiver algebra, incidence algebra, incidence coalgebra, path coalgebra, reflexive, coreflexive, coreflexive coalgebra.

other one. This is also important from the following viewpoint: one can ask when the finite-dimensional locally nilpotent representations of the quiver (i.e., quiver corepresentations), provide all the finite-dimensional quiver representations. This situation will be exactly the one in which the path coalgebra is recovered from the quiver algebra by a certain natural construction involving representative functions, which we recall below.

Given a coalgebra C , its dual C^* is always an algebra. Given an algebra A , one can associate a certain subspace A^0 of the dual A^* , which has a coalgebra structure. This is called the finite dual of A , and it plays an important role in the representation theory of A , since the category of locally finite left A -modules (i.e., modules which are sums of their finite-dimensional submodules) is isomorphic to the category of right A^0 -comodules (see, for example, [Green 1976]). A^0 is sometimes also called the coalgebra of representative functions, and consists of all $f : A \rightarrow K$ whose kernel contains a cofinite (i.e., having finite codimension) ideal. We show that the path coalgebra $K\Gamma$ can be reconstructed from the quiver algebra $K[\Gamma]$ as a certain type of “graded” finite dual, that is, $K\Gamma$ embeds in the dual space $K[\Gamma]^*$ as the subspace of linear functions $f : K[\Gamma] \rightarrow K$ whose kernel contains a cofinite monomial ideal. This is an “elementwise” answer to the recovery problem; its categorical analogue states that the comodules over the quiver coalgebra are precisely those quiver representations in which the annihilator of every element contains a cofinite monomial ideal. In order to connect these to the classical categorical duality, we first note that in general the quiver algebra does not have identity, but it has enough idempotents. Therefore, we first extend the construction of the finite dual to algebras with enough idempotents (Section 2). To such an algebra A we associate a coalgebra A^0 with counit, and we show that the category of right A^0 -comodules is isomorphic to the category of locally finite unital A -modules. In Section 3 we show that the path coalgebra $K\Gamma$ embeds in $K[\Gamma]^0$, and we prove that this embedding is an isomorphism, i.e., the path coalgebra can be recovered as the finite dual of the quiver algebra, if and only if the quiver has no oriented cycles and there are finitely many arrows between any two vertices. On the other hand, $K[\Gamma]$ embeds as an algebra without identity in the dual algebra $(K\Gamma)^*$ of the path coalgebra. We show that the image of this embedding is the rational (left or right) part of $(K\Gamma)^*$, i.e., the quiver algebra can be recovered as the rational part of the dual of the path coalgebra, if and only if for any vertex v of Γ there are finitely many paths starting at v and finitely many paths ending at v . This is also equivalent to the fact that $K\Gamma$ is a left and right semiperfect coalgebra.

In Section 4 we obtain similar results for another class of (co)algebras which are also objects of great combinatorial interest, namely for incidence (co)algebras. See [Joni and Rota 1979], for instance. We show that the incidence coalgebra of a partially ordered set X is always the finite dual of a subalgebra $FIA(X)$ of the

incidence algebra which consists of functions of finite support. In this setting, this algebra $FIA(X)$ is the natural analogue of the quiver algebra.

It is also interesting to know when can $K\Gamma$ be recovered from $(K\Gamma)^*$, and how this relates to the results of Section 3. This problem is related to an important notion in coalgebra theory, that of coreflexive coalgebra. A coalgebra C over K is coreflexive if the natural coalgebra embedding $C \rightarrow (C^*)^0$ is an isomorphism. In other words, C is coreflexive if it can be completely recovered from its dual. In Section 5 we aim to study this condition for path coalgebras and their subcoalgebras, and give the connection with the results of Section 3. We show that, in fact, a path coalgebra of a quiver with no loops and finitely many arrows between any two vertices is not necessarily coreflexive, and also, that the quivers of coreflexive path coalgebras can contain loops. We then prove a general result stating that under certain conditions a coalgebra C is coreflexive if and only if its coradical is coreflexive. In particular, this result holds for subcoalgebras of a path coalgebra $K\Gamma$ with the property that there are finitely many paths between any two vertices of Γ . The result applies in particular to incidence coalgebras. For both a path coalgebra and an incidence coalgebra the coradical is a grouplike coalgebra (over the set of vertices of the quiver for the first one, or the underlying ordered set for the second one). Thus the coreflexivity of such a coalgebra reduces to the coreflexivity of a grouplike coalgebra $K^{(X)}$. By [Heyneman and Radford 1974, Theorem 3.7.3], if K is an infinite field, then $K^{(X)}$ is coreflexive for most sets in X in set theory and any set of practical use (see Section 5).

We use our results to give a partial answer to a question of E. J. Taft and D. E. Radford asking whether the tensor product of two coreflexive coalgebras is coreflexive. In particular, we show that the tensor product of two coreflexive pointed coalgebras, which embed in path coalgebras of quivers with only finitely many paths between any two vertices, is coreflexive.

Throughout the paper $\Gamma = (\Gamma_0, \Gamma_1)$ will be a quiver. Γ_0 is the set of vertices, and Γ_1 is the set of arrows of Γ . If a is an arrow from the vertex u to the vertex v , we denote $s(a) = u$ and $t(a) = v$. A path in Γ is a finite sequence of arrows $p = a_1 a_2 \dots a_n$, where $n \geq 1$, such that $t(a_i) = s(a_{i+1})$ for any $1 \leq i \leq n - 1$. We will write $s(p) = s(a_1)$ and $t(p) = t(a_n)$. Also the length of such a p is n . Vertices v in Γ_0 are also considered as paths of length zero, and we write $s(v) = t(v) = v$. If p and q are two paths such that $t(p) = s(q)$, we consider the path pq by taking the arrows of p followed by the arrows of q . We denote by $K\Gamma$ the path coalgebra, which is the vector space with a basis consisting of all paths in Γ , comultiplication Δ defined by $\Delta(p) = \sum_{qr=p} q \otimes r$ for any path p , and counit ϵ defined by $\epsilon(v) = 1$ for any vertex v , and $\epsilon(p) = 0$ for any path of positive length. The underlying space of $K\Gamma$ can be also endowed with a structure of an algebra, not necessarily with identity, with the multiplication defined such that the product of two paths p and

q is pq if $t(p) = s(q)$, and 0 otherwise. We denote this algebra by $K[\Gamma]$; this is known in literature as the quiver algebra or the path algebra of Γ . It has identity if and only if Γ_0 is finite, and in this case the sum of all vertices is the identity.

Besides the above mentioned recovery connections between quiver algebras and path coalgebras, one can also ask whether there is any compatibility between them. More precisely, when do the two structures on the same vector space $K\Gamma$ give rise to a bialgebra structure. This turns out to be only the case for very special quivers. Specifically, consider $K[\Gamma]$ to be the vector space with basis the oriented paths of Γ , and with the quiver algebra and path coalgebra structures. Then $K[\Gamma]$ is a bialgebra (with enough idempotents in general) if and only if in Γ there are no (directed) paths of length ≥ 2 and no multiple edges between vertices (i.e., for any two vertices a, b of Γ there is at most one edge from a to b). Indeed, straightforward computations show that whenever multiple edges $\bullet \xrightarrow{x,y} \bullet$ or paths $\bullet \xrightarrow{x} \bullet \xrightarrow{y} \bullet$ of length at least 2 occur, then $\Delta(xy) \neq \Delta(x)\Delta(y)$. Conversely, a case by case computation for $\Delta(pq)$ with p, q paths of possible length 0 or 1 will show that $\Delta(pq) = \Delta(p)\Delta(q)$.

This shows that the relation between the path coalgebra and quiver algebra is more of a dual nature than an algebraic compatibility. For basic terminology and notation about coalgebras and comodules we refer to [Dăscălescu et al. 2001; Montgomery 1993; Sweedler 1969]. All (co)algebras and (co)modules considered here will be vector spaces over K , and duality $(-)^*$ represents the dual K -vector space.

2. The finite dual of an algebra with enough idempotents

In this section we extend the construction of the finite dual of an algebra with identity to the case where A does not necessarily have a unit, but it has enough idempotents. Throughout this section we consider a K -algebra A , not necessarily having a unit, but having a system $(e_\alpha)_{\alpha \in R}$ of pairwise orthogonal idempotents, such that $A = \bigoplus_{\alpha \in R} Ae_\alpha = \bigoplus_{\alpha \in R} e_\alpha A$. Such an algebra is said to have “enough idempotents”, and it is also called an algebra with a complete system of orthogonal idempotents in the literature. Let us note that A has local units, i.e., if $a_1, \dots, a_n \in A$, then there exists an idempotent $e \in A$ (which can be taken to be the sum of some e_α 's) such that $ea_i = a_i e = a_i$ for any $1 \leq i \leq n$. Our aim is to show that there exists a natural structure of a coalgebra (with counit) on the space

$$A^0 = \{f \in A^* \mid \text{Ker}(f) \text{ contains an ideal of } A \text{ of finite codimension}\}.$$

We will call A^0 the finite dual of the algebra A .

Lemma 2.1. *Let I be an ideal of A of finite codimension. Then the set $R' = \{\alpha \in R \mid e_\alpha \notin I\}$ is finite.*

Proof. Denote by \hat{a} the class of an element $a \in A$ in the factor space A/I . We have that $(\hat{e}_\alpha)_{\alpha \in R'}$ is linearly independent in A/I . Indeed, if $\sum_{\alpha \in R'} \lambda_\alpha \hat{e}_\alpha = 0$, then $\sum_{\alpha \in R'} \lambda_\alpha e_\alpha \in I$. Multiplying by some e_α with $\alpha \in R'$, we find that $\lambda_\alpha e_\alpha \in I$, so then necessarily $\lambda_\alpha = 0$. Since A/I is finite-dimensional, the set R' must be finite. \square

Assume now that B is another algebra with enough idempotents, say that $(f_\beta)_{\beta \in S}$ is a system of orthogonal idempotents in B such that $B = \bigoplus_{\beta \in S} Bf_\beta = \bigoplus_{\beta \in S} f_\beta B$.

Lemma 2.2. *Let H be an ideal of $A \otimes B$ of finite codimension. Let*

$$I = \{a \in A \mid a \otimes B \subseteq H\} \quad \text{and} \quad J = \{b \in B \mid A \otimes b \subseteq H\}.$$

Then I is an ideal of A of finite codimension, J is an ideal of B of finite codimension and $I \otimes B + A \otimes J \subseteq H$.

Proof. Let $a \in I$ and $a' \in A$. If $b \in B$ and f is an idempotent in B such that $fb = b$, we have that $a'a \otimes b = a'a \otimes fb = (a' \otimes f)(a \otimes b) \in H$. Thus $a'a \otimes B \subseteq H$, so $a'a \in I$. Similarly $aa' \in I$, showing that I is an ideal of A . Similarly J is an ideal of B .

It is clear that $(e_\alpha \otimes f_\beta)_{\alpha \in R, \beta \in S}$ is a set of orthogonal idempotents in $A \otimes B$ and

$$A \otimes B = \bigoplus_{\substack{\alpha \in R \\ \beta \in S}} (A \otimes B)(e_\alpha \otimes f_\beta) = \bigoplus_{\substack{\alpha \in R \\ \beta \in S}} (e_\alpha \otimes f_\beta)(A \otimes B).$$

By [Lemma 2.1](#), there are finitely many idempotents $e_{\alpha_1} \otimes f_{\beta_1}, \dots, e_{\alpha_n} \otimes f_{\beta_n}$ which lie outside H . If $\alpha \in R \setminus \{\alpha_1, \dots, \alpha_n\}$, then for any $\beta \in S$ we have that $e_\alpha \otimes f_\beta \in H$, so $e_\alpha \otimes Bf_\beta = (e_\alpha \otimes Bf_\beta)(e_\alpha \otimes f_\beta) \subseteq H$. Then $e_\alpha \otimes B \subseteq H$, so $e_\alpha \in I$. Similarly for any $\beta \in S \setminus \{\beta_1, \dots, \beta_n\}$ we have that $f_\beta \in J$.

For any $\beta \in S$ let $\phi_\beta : A \rightarrow A \otimes B$ be the linear map defined by $\phi_\beta(a) = a \otimes f_\beta$. If $a \in A$, then $a \in I$ if and only if for any $\beta \in S$ we have $a \otimes Bf_\beta \subseteq H$; because there is a local unit for a , this is further equivalent to $a \otimes f_\beta \in H$ for $\beta \in S$. This condition is obviously satisfied for $\beta \in S \setminus \{\beta_1, \dots, \beta_n\}$ since $f_\beta \in J$, so we obtain that

$$I = \bigcap_{1 \leq i \leq n} \phi_{\beta_i}^{-1}(H),$$

a finite intersection of finite codimensional subspaces of A , thus a finite codimensional subspace itself. Similarly J has finite codimension in B . The fact that $I \otimes B + A \otimes J \subseteq H$ is obvious. \square

Now we essentially proceed as in [\[Sweedler 1969, Chapter VI\]](#) or [\[Dăscălescu et al. 2001, Section 1.5\]](#), with some arguments adapted to the case of enough idempotents.

Lemma 2.3. *Let A and B be algebras with enough idempotents. The following assertions hold.*

- (i) If $f : A \rightarrow B$ is a morphism of algebras, then $f^*(B^0) \subseteq A^0$, where f^* is the dual map of f .
- (ii) If $\phi : A^* \otimes B^* \rightarrow (A \otimes B)^*$ is the natural linear injection, then $\phi(A^0 \otimes B^0) = (A \otimes B)^0$.
- (iii) If $M : A \otimes A \rightarrow A$ is the multiplication of A , and $\psi : A^* \otimes A^* \rightarrow (A \otimes A)^*$ is the natural injection, then $M^*(A^0) \subseteq \psi(A^0 \otimes A^0)$.

Proof. It goes as the proof of [Dăscălescu et al. 2001, Lemma 1.5.2], with part of the argument in (ii) replaced by using the construction and the result of Lemma 2.2. \square

Lemma 2.3 shows that by restriction and corestriction we can regard the natural linear injection ψ as an isomorphism $\psi : A^0 \otimes A^0 \rightarrow (A \otimes A)^0$. We consider the map $\Delta : A^0 \rightarrow A^0 \otimes A^0$, $\Delta = \psi^{-1}M^*$. Thus $\Delta(f) = \sum_i u_i \otimes v_i$ is equivalent to $f(xy) = \sum_i u_i(x)v_i(y)$ for any $x, y \in A$. On the other hand, we define a linear map $\varepsilon : A^0 \rightarrow K$ as follows. If $f \in A^0$, then $\text{Ker}(f)$ contains a finite codimensional ideal I . By Lemma 2.1, there are finitely many e_α 's outside I . Therefore only finitely many e_α 's lie outside $\text{Ker}(f)$, so it makes sense to define $\varepsilon(f) = \sum_{\alpha \in R} f(e_\alpha)$ (only finitely many terms are nonzero).

Proposition 2.4. *Let A be an algebra with enough idempotents. Then $(A^0, \Delta, \varepsilon)$ is a coalgebra with counit.*

Proof. The proof of the coassociativity works exactly as in the case where A has a unit; see [Dăscălescu et al. 2001, Proposition 1.5.3]. To check the property of the counit, let $f \in A^0$ and $\Delta(f) = \sum_i u_i \otimes v_i$. Let $a \in A$ and F a finite subset of R such that $a \in \sum_{\alpha \in F} e_\alpha A$. Then clearly $(\sum_{\alpha \in F} e_\alpha)a = a$. We have that

$$\begin{aligned} \left(\sum_i \varepsilon(u_i)v_i \right)(a) &= \sum_{i,\alpha} u_i(e_\alpha)v_i(a) = \sum_{\alpha} f(e_\alpha a) \\ &= \sum_{\alpha \in F} f(e_\alpha a) = f\left(\left(\sum_{\alpha \in F} e_\alpha\right)a\right) = f(a), \end{aligned}$$

so $\sum_i \varepsilon(u_i)v_i = f$. Similarly $\sum_i \varepsilon(v_i)u_i = f$, and this ends the proof. \square

Let us note that if $f : A \rightarrow B$ is a morphism of algebras with enough idempotents, then the map $f^0 : B^0 \rightarrow A^0$ induced by f^* is compatible with the comultiplications of A^0 and B^0 , but not necessarily with the counits (unless f is compatible in some way to the systems of orthogonal idempotents in A and B).

We denote by \rightarrow (respectively \leftarrow) the usual left (respectively right) actions of A on A^* . As in the unitary case, we have the following characterization of the elements of A^0 .

Proposition 2.5. *Let $f \in A^*$. With notation as above, the following assertions are equivalent.*

- (1) $f \in A^0$.
- (2) $M^*(f) \in \psi(A^0 \otimes A^0)$.
- (3) $M^*(f) \in \psi(A^* \otimes A^*)$.
- (4) $A \rightarrow f$ is finite-dimensional.
- (5) $f \leftarrow A$ is finite-dimensional.
- (6) $A \rightarrow f \leftarrow A$ is finite-dimensional.
- (7) $\text{Ker}(f)$ contains a left ideal of finite codimension.
- (8) $\text{Ker}(f)$ contains a right ideal of finite codimension.

Proof. (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) and (1) \Rightarrow (6) work exactly as in the case where A has identity; see [Dăscălescu et al. 2001, Proposition 1.5.6]. We adapt the proof of (4) \Rightarrow (1) to the case of enough idempotents. Since $A \rightarrow f$ is a left A -submodule of A^* , there is a morphism of algebras (without unit) $\pi : A \rightarrow \text{End}(A \rightarrow f)$ defined by $\pi(a)(m) = a \rightarrow m$ for any $a \in A, m \in A \rightarrow f$. Since $\text{End}(A \rightarrow f)$ has finite dimension, we have that $I = \text{Ker}(\pi)$ is an ideal of A of finite codimension. Let $a \in I$. Then $a \rightarrow (b \rightarrow f) = (ab) \rightarrow f = 0$ for any $b \in A$, so $f(xab) = 0$ for any $x, b \in A$. Let $e \in A$ such that $ea = ae = a$. Then $f(a) = f(eae) = 0$, so $a \in \text{Ker}(f)$. Thus $I \subseteq \text{Ker}(f)$, showing that $f \in A^0$. The equivalence (1) \Leftrightarrow (5) is proved similarly.

(6) \Rightarrow (1) can be adapted from the unital case; see [Montgomery 1993, Lemma 9.1.1], with a small change. Indeed, $R = (A \rightarrow f \leftarrow A)^\perp = \{x \in A \mid g(x) = 0 \text{ for any } g \in A \rightarrow f \leftarrow A\}$ is an ideal of A of finite codimension, and $R \subseteq \text{Ker}(f)$, since for any $r \in R$ there exists $e \in A$ such that $r = er = re$, so then $f(r) = f(ere) = (e \rightarrow f \leftarrow e)(r) = 0$.

(1) \Rightarrow (7) is obvious, while (7) \Rightarrow (1) follows from the fact that a left ideal I of finite codimension contains the finite codimensional ideal $J = \{r \in A \mid rA \subseteq I\}$.

(1) \Leftrightarrow (8) is similar. \square

We end this section with an interpretation of the connection between an algebra A with enough idempotents and its finite dual A^0 from the representation theory point of view. This extends the results presented in [Abe 1980, Chapter 3, Section 1.2] in the case where A has identity. Let M be a left A -module. Then M is called unital if $AM = M$. Also, M is called locally finite if the submodule generated by any element is finite-dimensional. Denote by $\text{LocFin}_A \mathcal{M}$ the full subcategory of the category of left A -modules consisting of all locally finite unital modules. We will also use the notations ${}_A \mathcal{M}, \mathcal{M}_A$ for the categories of left, or right modules over A ; similarly, for a coalgebra C , ${}^C \mathcal{M}$ and \mathcal{M}^C will be used to denote the categories of left and respectively right comodules.

Proposition 2.6. *Let A be an algebra with enough idempotents. Then the category \mathcal{M}^{A^0} of right A^0 -comodules is isomorphic to the category $\text{LocFin}_A \mathcal{M}$.*

Proof. Let M be a right A^0 -comodule with comodule structure $m \mapsto \sum m_0 \otimes m_1$. Then M is a left A -module with the action $am = \sum m_1(a)m_0$ for any $a \in A$ and $m \in M$. The counit property $m = \sum \varepsilon(m_1)m_0$, with all m_1 's in A^0 , shows that $m = \sum_{\alpha \in F} e_\alpha m$ for a finite set F , so M is unital. Since Am is contained in the subspace spanned by all m_0 's, we have that M is also locally finite.

Conversely, let $M \in \text{LocFin}_A \mathcal{M}$. Let $m \in M$, and let $(m_i)_{i=1,n}$ be a (finite) basis of Am . Define $a_1^*, \dots, a_n^* \in A^*$ such that $am = \sum_{i=1,n} a_i^*(a)m_i$ for any $a \in A$. Since $\bigcap_{i=1,n} \text{ann}_A(m_i) = \text{ann}_A(Am) \subseteq \text{ann}_A(m) = \bigcap_{i=1,n} \text{Ker } a_i^*$ and each $\text{ann}_A(m_i)$ has finite codimension, we get that $a_i^* \in A^0$ for any i . Now we define $\rho : M \rightarrow M \otimes A^0$ by $\rho(m) = \sum_{i=1,n} m_i \otimes a_i^*$. It is easy to see that the definition of $\rho(m)$ does not depend on the choice of the basis $(m_i)_i$, and that $(\rho \otimes I)\rho = (I \otimes \Delta)\rho$. To show that M is a right A^0 -comodule it remains to check the counit property, and this follows from the fact that M is unital.

It is clear that the above correspondences define an isomorphism of categories. \square

3. Quiver algebras and path coalgebras

We examine the connection between the quiver algebra $K[\Gamma]$ and the path coalgebra $K\Gamma$ associated to a quiver Γ . The algebra $K[\Gamma]$ has identity if and only if Γ has finitely many vertices. However, $K[\Gamma]$ always has enough idempotents (the set of all vertices). Thus by Section 2 we can consider the finite dual $K[\Gamma]^0$, which is a coalgebra with counit. One has that $K[\Gamma]^0 \supseteq K\Gamma$, i.e., the path coalgebra can be embedded in the finite dual of the quiver algebra. The embedding is given as follows: for each path $p \in \Gamma$, denote by $\theta(p) \in K[\Gamma]^*$ the function $\theta(p)(q) = \delta_{p,q}$. We have that $\theta(p) \in K[\Gamma]^0$ since if we denote by $S(p)$ the set of all subpaths of p , and by P the set of all paths in Γ , the span of $P \setminus S(p)$ is a finite codimensional ideal of $K[\Gamma]$ contained in $\text{Ker } \theta(p)$. It is easy to see that $\theta : K\Gamma \hookrightarrow K[\Gamma]^0$ is an embedding of coalgebras. In general, $K[\Gamma] \hookrightarrow (K\Gamma)^*$ is surjective if and only if the quiver Γ is finite. Also, in general, θ is not surjective. To see this, let A be the quiver algebra of a loop Γ , i.e., a quiver with one vertex and one arrow:



so $A = K[X]$, the polynomial algebra in one indeterminate. The finite dual of this algebra is

$$\lim_{\substack{\longrightarrow \\ n \in \mathbb{Z}_{\geq 0}}} (K[X]/(f^n))^* = \bigoplus_{f \text{ irreducible}} [\lim_{\substack{\longrightarrow \\ n \in \mathbb{Z}_{\geq 0}} (K[X]/(f^n))^*],$$

while the path coalgebra is precisely the divided power coalgebra, which can be written as $\varinjlim_{n \in \mathbb{Z}_{\geq 0}} (K[X]/(X^n))^*$. These two coalgebras are not isomorphic, so the map θ above is not a surjection. Indeed, $K\Gamma$ has just one grouplike element, the vertex of Γ , while the grouplike elements of A^0 , which are the algebra morphisms from $A = K[X]$ to K , are in bijection to K .

The embedding of coalgebras $\theta : K\Gamma \hookrightarrow K[\Gamma]^0$ also gives rise to a functor $F^\theta : {}^{K\Gamma}\mathcal{M} \rightarrow {}^{K[\Gamma]^0}\mathcal{M}$, associating to a left $K\Gamma$ -comodule the left $K[\Gamma]^0$ -comodule structure obtained by extension of coscalars via θ . We aim to provide a criterion for when the above map θ is bijective, that is, when the path coalgebra is recovered as the finite dual of the quiver algebra. Even though this is not always the case, we show that it is possible to interpret the quiver algebra as a certain kind of “graded” finite dual. We will think of $K\Gamma$ as embedded into $K[\Gamma]^0$ through θ , and sometimes write $K\Gamma$ instead of $\theta(K\Gamma)$.

Recall that in a quiver algebra $K[\Gamma]$, there is an important class of ideals, those which have a basis of paths; equivalently, the ideals generated by paths. Let us call such an ideal a monomial ideal. When I is a cofinite monomial right ideal, the quotient $K[\Gamma]/I$ produces an interesting type of representation often considered in the representation theory of quivers; we refer to [Villarreal 2001] for the theory monomial algebras and representations. In fact, such a representation also becomes a left $K\Gamma$ -comodule, i.e., it is in the “image” of the functor F^θ . To see this, let B be basis of paths for I and let E be the set of paths not belonging to I ; then E is finite, and because I is a right ideal, one sees that if $p \in E$ and $p = qr$, then $q \in E$. This shows that KE , the span of E , is a right $K\Gamma$ -subcomodule of $K\Gamma$, so it is a rational left $(K\Gamma)^*$ -module (for example, by [Dăscălescu et al. 2001, Theorem 2.2.5]). By [Dăscălescu et al. 2001, Lemma 2.2.12], the right $(K\Gamma)^*$ -module $(KE)^*$ is rational, and so it has a compatible left $K\Gamma$ -comodule structure. Hence $(KE)^*$ is a right $K[\Gamma]$ -module via the algebra map $K[\Gamma] \hookrightarrow (K\Gamma)^*$. Now, it is straightforward to see that $K[\Gamma]/I \cong (KE)^*$ as right $K[\Gamma]$ -modules, and this proves the claim. Thus, $F^\theta({}^{K\Gamma}((KE)^*)) = {}^{K[\Gamma]^0}(K[\Gamma]/I)$, since every finite-dimensional right $K[\Gamma]$ -module is a left $K[\Gamma]^0$ -comodule.

We can now state a characterization of the path coalgebra in terms of the quiver algebra, as a certain type of finite dual.

Proposition 3.1. *The coalgebra $\theta(K\Gamma)$ consists of all $f \in K[\Gamma]^*$ such that $\ker(f)$ contains a two-sided cofinite monomial ideal.*

Proof. Let P be the set of paths in Γ . If p is a path, and $S(p)$ is the set of subpaths of p , then the cofinite-dimensional vector space with basis $P \setminus S(p)$ is an ideal, and it is obviously contained in $\ker(\theta(p))$. Then clearly $\ker(\theta(z))$ contains a cofinite monomial ideal for any $z \in K\Gamma$.

Let now $f \in K[\Gamma]^*$ such that $\ker(f)$ contains the cofinite monomial ideal I . Let B be a basis of I consisting of paths, and let $E = P \setminus B$, which is finite, since I is cofinite. Then if $q \in B$, we have $f(q) = 0 = \sum_{p \in E} f(p)\theta(p)(q)$, while if $q \in E$, we have $(\sum_{p \in E} f(p)\theta(p))(q) = f(q)$. Therefore $f = \sum_{p \in E} f(p)\theta(p)$ lies in $\theta(K\Gamma)$. \square

The core of our characterization is the following easy combinatorial condition:

Proposition 3.2. *Let Γ be a quiver. The following conditions are equivalent:*

- (i) Γ has no oriented cycles and between any two vertices there are only finitely many arrows.
- (ii) For any finite set of vertices $E \subset \Gamma$, there are only finitely many paths passing only through vertices of E .

We recall that a representation of the quiver Γ is a pair $\mathcal{R} = ((V_u)_{u \in \Gamma_0}, (f_a)_{a \in \Gamma_1})$ consisting of a family of vector spaces and a family of linear maps, such that $f_a : V_u \rightarrow V_v$, where $u = s(a)$ and $v = t(a)$ for any $a \in \Gamma_1$. A morphism of representations is a family of linear maps (indexed by Γ_0) between the corresponding V_u 's, which are compatible with the corresponding linear morphisms in the two representations. The category $\text{Rep } \Gamma$ of representations of Γ is equivalent to the category $u.\mathcal{M}_{K[\Gamma]}$ of unital right $K[\Gamma]$ -modules. The equivalence $H : u.\mathcal{M}_{K[\Gamma]} \rightarrow \text{Rep } \Gamma$ works as follows. To a unital right $K[\Gamma]$ -module V we associate the representation $H(V) = ((V_u)_{u \in \Gamma_0}, (f_a)_{a \in \Gamma_1})$, where $V_u = Vu$ for any $u \in \Gamma_0$, and for an arrow a from u to v we define $f_a : V_u \rightarrow V_v$, $f_a(x) = xa$. An inverse equivalence functor associates to representation \mathcal{R} as above the space $\bigoplus_{u \in \Gamma_0} V_u$ endowed with a right $K[\Gamma]$ -action defined by $xp = f_p(x)$ for $p = a_1 \dots a_n$ and $x \in V_u$ such that $s(a_1) = u$. Here we denote $f_p = f_{a_n} \dots f_{a_1}$. If $s(a_1) \neq u$, the action is $xp = 0$.

A representation \mathcal{R} is locally finite if for any $u \in \Gamma_0$ and any $x \in V_u$ the subspace $\langle f_p(x) \mid p \text{ is a path with } s(p) = u \rangle$ of $\bigoplus_{u \in \Gamma_0} V_u$ is finite-dimensional. Denote the subcategory of locally finite representations by $\text{LocFinRep } \Gamma$. The equivalence H restricts to an equivalence $H_1 : \text{LocFin } \mathcal{M}_{K[\Gamma]} \rightarrow \text{LocFinRep } \Gamma$.

Recall from [Chin et al. 2002] that a representation \mathcal{R} is locally nilpotent if for any $u \in \Gamma_0$ and any $x \in V_u$, the set $\{p \mid p \text{ path with } f_p(x) \neq 0\}$ is finite. This is easily seen to be equivalent to each $x \in \bigoplus_{u \in \Gamma_0} V_u$ being annihilated by a monomial ideal of finite codimension. Denote by $\text{LocNilpRep } \Gamma$ the category of locally nilpotent representations, which is clearly a subcategory of $\text{LocFinRep } \Gamma$.

We have the following diagram:

$$\begin{array}{ccccc}
 K\Gamma\mathcal{M} & \xrightarrow{F^\theta} & K[\Gamma]^0\mathcal{M} & \xrightarrow{\sim G} & \text{LocFin } \mathcal{M}_{K[\Gamma]} & \xrightarrow{I_1} & u.\mathcal{M}_{K[\Gamma]} \\
 \downarrow H_2 \sim & & & & \downarrow H_1 \sim & & \downarrow H \sim \\
 \text{LocNilpRep } \Gamma & \xrightarrow{I_2} & & & \text{LocFinRep } \Gamma & \xrightarrow{I_3} & \text{Rep } \Gamma
 \end{array}$$

Here G is the equivalence of categories as in [Proposition 2.6](#) (the version for right modules), and the I_j 's are inclusion functors. It is easy to see that the image (on objects) of the functor $H_1 G F^\theta$ lies in $\text{LocNilpRep } \Gamma$, so we actually have a functor $H_2 : {}^{K\Gamma}\mathcal{M} \rightarrow \text{LocNilpRep } \Gamma$, and this is just the equivalence noticed in [[Chin et al. 2002](#), Proposition 6.1]. In this way, at the level of representations, the functor F^θ can be regarded as a functor (embedding) from the locally nilpotent quiver representations to the locally finite quiver representations.

We can now characterize precisely when the path coalgebra can be recovered from the quiver algebra, that is, when the above mentioned embedding θ is an isomorphism.

Theorem 3.3. *Let Γ be a quiver. The following assertions are equivalent:*

- (i) Γ has no oriented cycles and between every two vertices of Γ there are only finitely many arrows.
- (ii) $\theta(K\Gamma) = K[\Gamma]^0$.
- (iii) Every cofinite ideal of $K[\Gamma]$ contains a cofinite monomial ideal.
- (iv) The functor $F^\theta : {}^{K\Gamma}\mathcal{M} \rightarrow {}^{K[\Gamma]^0}\mathcal{M}$ is an equivalence.
- (v) Every locally finite quiver representation of Γ is locally nilpotent.

Proof. The equivalence of (ii) and (iv) is a general coalgebra fact: if $C \subseteq D$ is an inclusion of coalgebras, then the corestriction of scalars $F : {}^C\mathcal{M} \rightarrow {}^D\mathcal{M}$ is an equivalence if and only if $C = D$. Indeed, if F is an equivalence, pick an arbitrary $x \in D$ and let $N = xD^* \in {}^D\mathcal{M}$ be the finite-dimensional D -subcomodule of D generated by x . Then $N \simeq F(M)$, $M \in {}^C\mathcal{M}$, and considering the coalgebras of coefficients C_N and C_M of N and M , we see that $C_N = C_M \subseteq C$ by the definition of F . Since $x \in C_N$, this ends the argument.

The equivalence of (ii) and (iii) follows immediately from [Proposition 3.1](#).

(iv) \Leftrightarrow (v) The previous remarks on F^θ (and the diagram drawn there) show that F^θ is an equivalence functor if and only if so is I_2 . On the other hand, the inclusion functor I_2 is an equivalence if and only if every locally finite quiver representation of Γ is locally nilpotent.

(i) \Rightarrow (iii) Let I be an ideal of $K[\Gamma]$ of finite codimension. By [Lemma 2.1](#) applied for the algebra $K[\Gamma]$ and the complete set of orthogonal idempotents Γ_0 , we have that the set $S' = \{a \in \Gamma_0 \mid a \notin I\}$ must be finite. Let $S = \{a \in \Gamma_0 \mid a \in I\}$. Note that any path p starting or ending at a vertex in S belongs to I , since $p = s(p)p = pt(p) \in I$ if either $s(p) \in I$ or $t(p) \in I$. Furthermore, this shows that if p contains a vertex in S , then $p \in I$, since in that case $p = qr$ with $x = t(q) = s(r) \in S$. Denote the set of paths containing some vertex in S by M . Let H be the vector space spanned by M and let M' be the set of the rest of the paths in Γ . Obviously, M' consists of

the paths whose all vertices belong to S' . Since S' is finite, we see that M' is finite, by the conditions of (i) and [Proposition 3.2](#). Therefore H has finite codimension. Also, since H is spanned by paths passing through some vertex in S , we see that H is an ideal. We conclude that I contains the cofinite monomial ideal H .

(iii) \Rightarrow (i) We show first that there are no oriented cycles in Γ . Assume Γ has a cycle

$$C : v_0 \xrightarrow{x_0} v_1 \xrightarrow{x_1} \cdots \longrightarrow v_{s-1} \xrightarrow{x_{s-1}} v_s = v_0,$$

and consider such a cycle that does not self-intersect. We can consider the vertices v_0, \dots, v_{s-1} modulo s . Denote by $q_{n,k}$ the path starting at the vertex v_n ($0 \leq n \leq s-1$), winding around the cycle C and of length k . Denote again by P the set of all paths in Γ , and by $X = \{q_{n,k} \mid 0 \leq n \leq s-1, k \geq 0\}$. Since the set X is closed under subpaths, it is easy to see that the vector space H spanned by the set $P \setminus X$ is an ideal of $K[\Gamma]$. Let E be the subspace spanned by $S = \{q_{n,ks+i} - q_{n,i} \mid 0 \leq n \leq s-1, i \geq 0, k \geq 1\}$, and let $I = E + H$. We have

$$\begin{aligned} (q_{n,ks+i} - q_{n,i})q_{n+i,j} &= q_{n,ks+i+j} - q_{n,i+j} \in S, \\ (q_{n,ks+i} - q_{n,i})q_{m,j} &= 0 \text{ for } m \neq n+i, \\ q_{m,j}(q_{n,ks+i} - q_{n,i}) &= q_{m,ks+i+j} - q_{m,i+j} \in S \text{ if } m+j = n, \\ q_{m,j}(q_{n,ks+i} - q_{n,i}) &= 0 \text{ if } m+j \neq n. \end{aligned}$$

Here in the notation $q_{n,i}$ the first index is considered modulo s , while the second index is a nonnegative integer. The above equations show that if we multiply an element of S to the left (or right) by an element of X , we obtain either an element of S or 0. Combined with the fact that H is an ideal, this shows that I is an ideal.

It is clear that I has finite codimension, since $S \cup \{q_{n,i} \mid 0 \leq n \leq s-1, 0 \leq i \leq s-1\}$ spans $KC = \langle X \rangle$. Indeed, if $0 \leq n \leq s-1$ and j is a nonnegative integer, write $j = ks + i$ with $k \geq 0$ and $0 \leq i \leq s-1$, and we have that $q_{n,j} = q_{n,ks+i} = (q_{n,ks+i} - q_{n,i}) + q_{n,i}$.

On the other hand, I does not contain a cofinite monomial ideal. Indeed, it is easy to see that an element of the form $q_{m,j}$ cannot be in $\langle S \cup (P \setminus X) \rangle = I$, so any monomial ideal contained in I must have infinite codimension.

Thus, we have found a cofinite ideal I which does not contain a cofinite monomial ideal. This contradicts (iii), and we conclude that Γ cannot contain cycles.

We now show that in Γ there are no pair of vertices with infinitely many arrows between them. Assume such a situation exists between two vertices a, b : $a \xrightarrow{x_n} b$, $n \in \mathbb{Z}_{\geq 0}$. We let $X = \{x_n \mid n \in \mathbb{Z}_{\geq 0}\} \cup \{a, b\}$, H be the span of $P \setminus X$, which is an ideal since X is closed under taking subpaths. Let $S = \{x_n - x_0 \mid n \geq 1\}$ and I be the span of $S \cup (P \setminus X)$. As above, since $x_n - x_0$ multiplied by an element of X gives either $x_n - x_0$ or 0, we have that I is an ideal. I has finite codimension since

$\{a, b, x_0\} \cup S \cup (P \setminus X)$ spans $K\Gamma$. Also, I does not contain a monomial ideal of finite codimension since no x_n lies in I . Thus we contradict (iii). In conclusion there are finitely many arrows between any two vertices, and this ends the proof. \square

It is clear that a finite quiver Γ (i.e., Γ_0 and Γ_1 are finite) without oriented cycles satisfies condition (i) in Theorem 3.3. In this case $K\Gamma$ is finite-dimensional, and we obviously have $K[\Gamma] = (K\Gamma)^*$ (i.e., the map θ is bijective) and also $K\Gamma = K[\Gamma]^0$. This can also be thought as a trivial case of the above theorem.

We now present a few examples to further illustrate the above theorem.

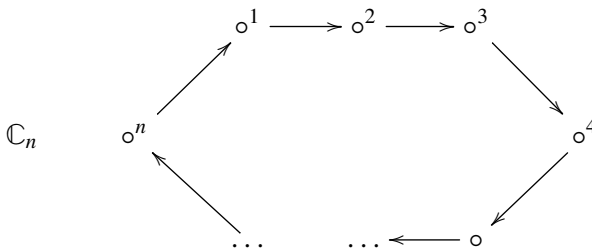
Example 3.4. Let ${}_{\infty}\mathbb{A}_{\infty}$ be the infinite line quiver

$$\dots \rightarrow \bullet \rightarrow \bullet \rightarrow \dots \rightarrow \bullet \rightarrow \dots$$

The quiver coalgebra $C = K_{\infty}\mathbb{A}_{\infty}$ of this quiver is serial, that is, the injective indecomposable left and right comodules are uniserial, i.e., they have a unique composition series; see [Gómez-Torrecillas and Navarro 2008]. For such a coalgebra, the finite dimensional comodules are easily classified: they are all serial [ibid.]. Moreover, the indecomposable finite-dimensional comodules, i.e., the uniserial ones, correspond to finite paths in ${}_{\infty}\mathbb{A}_{\infty}$. Note that this quiver satisfies the conditions of Theorem 3.3, and so the locally nilpotent representations of ${}_{\infty}\mathbb{A}_{\infty}$ (i.e., the comodules over $K_{\infty}\mathbb{A}_{\infty}$) coincide with the locally finite representations of the quiver algebra $A = K[{}_{\infty}\mathbb{A}_{\infty}]$, and also, the finite-dimensional quiver representations of ${}_{\infty}\mathbb{A}_{\infty}$ are the comodules over $K_{\infty}\mathbb{A}_{\infty}$. Moreover, the coalgebra of representative functions on $K[{}_{\infty}\mathbb{A}_{\infty}]$ is isomorphic to $K_{\infty}\mathbb{A}_{\infty}$.

Note that in general, it is not easy to describe arbitrary comodules even for a serial coalgebra. By results in [Iovanov 2011], if an infinite dimensional indecomposable injective comodule exists, then there are comodules which do not decompose into indecomposable comodules (and, in particular, are not indecomposable). Moreover, for the left bounded infinite quiver $\mathbb{A}_{\infty} : \bullet \rightarrow \bullet \rightarrow \dots \rightarrow \bullet \rightarrow \dots$, it is shown in [Iovanov 2011] that all left comodules over $K\mathbb{A}_{\infty}$ are serial direct sums of indecomposable uniserial comodules corresponding to finite paths, while in the category of right comodules over $K\mathbb{A}_{\infty}$ there are objects which do not decompose into direct sums of indecomposables.

Example 3.5. Let \mathbb{C}_n be the following quiver of affine Dynkin type \tilde{A}_n :



The path coalgebra $K\mathbb{C}_n$ is again serial, and the finite dimensional (left and right) comodules are all direct sum of uniserial objects (corresponding to finite paths). These correspond to finite-dimensional locally nilpotent representations. This quiver does not satisfy the hypothesis of [Theorem 3.3](#). We give an example of a quiver representation which is locally finite (even finite-dimensional) but not locally nilpotent. Let x_1, \dots, x_n denote the arrows of the quiver, with $a_i = s(x_i)$. Let $M = K[\mathbb{C}_n]/I$ where I is the (two sided) ideal generated by elements $p - 1$, where p is a path of length n and with $1 = a_1 + \dots + a_n$ (M is actually an algebra). One can easily see that M is spanned as a vector space by paths of length less than n . A not too difficult computation shows that I does not contain any monomial ideal of finite codimension, and so M as a representation of $K[\mathbb{C}_n]$ is not locally nilpotent, but it is finite-dimensional. We again note that the infinite dimensional comodules over the coalgebra $K\mathbb{C}_n$ are hard to understand, as there are both left and right comodules which are not direct sum of indecomposable comodules.

An easy particular example of this can be obtained for $n = 1$; in this case, $K[\mathbb{C}_1] \cong K[X]$ — the polynomial algebra. As noted before, the finite dual of this algebra is not the path coalgebra of \mathbb{C}_1 . Also, the representation $K[X]/(X - 1)$ is not locally nilpotent.

Let $\psi : K[\Gamma] \rightarrow (K\Gamma)^*$ be the linear map defined by $\psi(p)(q) = \delta_{p,q}$ for any paths p and q . In fact ψ is just θ as a linear map, but we denote it differently since we regard it now as a morphism in the category of algebras not necessarily with identity. Indeed, it is easy to check that ψ is multiplicative. Thus the quiver algebra embeds in the dual of the path coalgebra. Our aim is to show that in certain situations $K[\Gamma]$ can be recovered from $(K\Gamma)^*$ as the rational part. Obviously, this is the case when $K[\Gamma]$ is finite-dimensional, which will also be seen as a consequence of the next result, which characterizes completely these situations. We recall that if C is a coalgebra, the rational part of the left C^* -module C^* , consisting of all elements $f \in C^*$ such that there exist finite families $(c_i)_i$ in C and $(f_i)_i$ in C^* with $c^* f = \sum_i c^*(c_i) f_i$ for any $c^* \in C^*$, is denoted by $C_l^{*\text{rat}}$. This is the largest C^* -submodule which is rational, i.e., whose C^* -module structure comes from a right C -comodule structure. Similarly, $C_r^{*\text{rat}}$ denotes the rational part of the right C^* -module C^* . A coalgebra C is called right (respectively left) semiperfect if the category of right (respectively left) C -comodules has enough projectives. This is equivalent to the fact that $C_l^{*\text{rat}}$ (respectively $C_r^{*\text{rat}}$) is dense in C^* in the finite topology, see [[Dăscălescu et al. 2001](#), Section 3.2].

Theorem 3.6. *The following are equivalent.*

- (i) $\text{Im}(\psi) = (K\Gamma)_l^{*\text{rat}}$.
- (ii) $\text{Im}(\psi) = (K\Gamma)_r^{*\text{rat}}$.

- (iii) For any vertex v of Γ there are finitely many paths starting at v and finitely many paths ending at v .
- (iv) The path coalgebra $K\Gamma$ is left and right semiperfect.

Proof. (iii) \Rightarrow (i) Let p be a path. We show that $p^* = \psi(p) \in \text{Im}(\psi)$. If $c^* \in (K\Gamma)^*$ and q is a path, we have that

$$(c^* p^*)(q) = \sum_{rs=q} c^*(r)p^*(s) = \begin{cases} c^*(r) & \text{if } q = rp \text{ for some path } r, \\ 0 & \text{if } q \text{ does not end with } p. \end{cases}$$

Let $q_1 = r_1 p, \dots, q_n = r_n p$ be all the paths ending with p . By the formula above, $(c^* p^*)(q_i) = c^*(r_i)$ for any $1 \leq i \leq n$, and $(c^* p^*)(q) = 0$ for any path $q \neq q_1, \dots, q_n$. This shows that $c^* p^* = \sum_{1 \leq i \leq n} c^*(r_i) q_i^*$, thus $p^* \in (K\Gamma)_l^{\text{rat}}$, and we have that $\text{Im}(\psi) \subseteq (K\Gamma)_l^{\text{rat}}$.

Now let $c^* \in (K\Gamma)_l^{\text{rat}}$, so there exist $(c_i)_{1 \leq i \leq n}$ in $K\Gamma$ and $(c_i^*)_{1 \leq i \leq n}$ in $(K\Gamma)^*$ such that $d^* c^* = \sum_{1 \leq i \leq n} d^*(c_i) c_i^*$ for any $d^* \in (K\Gamma)^*$. Let p_1, \dots, p_m be all the paths that appear with nonzero coefficients in some of the c_i 's (represented as a linear combination of paths). Then for any $p \neq p_1, \dots, p_m$ we have that $p^*(c_i) = 0$, so then $p^* c^* = 0$. Let v be a vertex such that no one of p_1, \dots, p_m passes through v . Then for any path p starting at v we have that $0 = (v^* c^*)(p) = v^*(v) c^*(p) = c^*(p)$. Therefore c^* may be nonzero on a path p only if $s(p) \in \{p_1, \dots, p_m\}$. By condition (iii), there are only finitely many such paths p , denote them by q_1, \dots, q_e . Then $c^* = \sum_{1 \leq i \leq e} c^*(q_i) q_i^* \in \text{Im}(\psi)$, and we also have that $(K\Gamma)_l^{\text{rat}} \subseteq \text{Im}(\psi)$.

(i) \Rightarrow (iii) Let v be a vertex. Then $v^* = \psi(v) \in (K\Gamma)_l^{\text{rat}}$, so there exist finite families $(c_i) \subseteq K\Gamma$ and $(c_i^*)_i \subseteq (K\Gamma)^*$ such that $c^* v^* = \sum_i c^*(c_i) c_i^*$ for any $c^* \in (K\Gamma)^*$. Then for any path q ,

$$(1) \quad \sum_i c^*(c_i) c_i^*(q) = (c^* v^*)(q) = \begin{cases} c^*(q) & \text{if } q \text{ ends at } v, \\ 0 & \text{otherwise.} \end{cases}$$

If there exist infinitely many paths ending at v , we can find one such path q which does not appear in the representation of any c_i as a linear combination of paths. Then there exists $c^* \in (K\Gamma)^*$ with $c^*(q) \neq 0$ and $c^*(c_i) = 0$ for any i , in contradiction with (1). Thus only finitely many paths can end at v . In particular Γ does not have cycles.

On the other hand, if we assume that there are infinitely many paths p_1, p_2, \dots starting at v , let $c^* \in (K\Gamma)^*$ which is 1 on each p_i and 0 on any other path. Clearly $c^* \notin \text{Im}(\psi)$. We show that $c^* \in (K\Gamma)_l^{\text{rat}}$, and the obtained contradiction shows that only finitely many paths start at v . Indeed, we have

$$(2) \quad (d^* c^*)(q) = \begin{cases} d^*(r) & \text{if } q = r p_i \text{ for some } i \geq 1 \text{ and some path } r, \\ 0 & \text{otherwise.} \end{cases}$$

Let r_1, \dots, r_m be all the paths ending at v (they are finitely many as we proved above). For each $1 \leq j \leq m$ we consider the element $c_j^* \in (K\Gamma)^*$ which is 1 on every path of the form $r_j p_i$, and 0 on any other path. Using (2) and the fact that $r_j p_i \neq r_{j'} p_{i'}$ for $(i, j) \neq (i', j')$ (this follows because $r_j, r_{j'}$ end at v and $p_i, p_{i'}$ start at v , and there are no cycles containing v), we see that $d^* c^* = \sum_{1 \leq j \leq m} d^*(r_j) c_j^*$, and this will guarantee that c^* is a rational element.

(ii) \Leftrightarrow (iii) is similar to (i) \Leftrightarrow (iii).

(iii) \Leftrightarrow (iv) follows from [Chin et al. 2002, Corollary 6.3]. \square

4. Incidence coalgebras and incidence algebras

In this section we parallel the results in Section 3 in the framework of incidence (co)algebras. Let (X, \leq) be a partially ordered set which is locally finite, i.e., the set $\{z \mid x \leq z \leq y\}$ is finite for any $x \leq y$ in X . The incidence coalgebra of X , denoted by KX , is the vector space with basis $\{e_{x,y} \mid x, y \in X, x \leq y\}$, and comultiplication and counit defined by $\Delta(e_{x,y}) = \sum_{x \leq z \leq y} e_{x,z} \otimes e_{z,y}$, $\epsilon(e_{x,y}) = \delta_{x,y}$ for any $x, y \in X$ with $x \leq y$. For such a X , we can consider the quiver Γ with vertices the elements of X , and such that there is an arrow from x to y if and only if $x < y$ and there is no element z with $x < z < y$. It was proved in [Dăscălescu et al. \geq 2013] that the linear map $\phi : KX \rightarrow K\Gamma$, defined by

$$\phi(e_{x,y}) = \sum_{\substack{p \text{ path} \\ \text{from } x \text{ to } y}} p$$

for any $x, y \in X, x \leq y$, is an injective coalgebra morphism. We note that this map is surjective if and only if in Γ there is at most one path between any two vertices $x, y \in X$. To see this, let $P(x, y)$ denote the set of paths from x to y . Note that the incidence coalgebra KX is then $KX = \bigoplus_{x,y \in X} \langle P(x, y) \rangle$, and that $\phi(\langle e_{x,y} \rangle) \subset P(x, y)$ for $x \leq y$. Thus, ϕ is surjective if and only if $\dim(P(x, y)) = 1$ for all $x \leq y$, which is equivalent to the above stated condition. In fact, this is also a consequence of the following more general fact.

Proposition 4.1. *A coalgebra C is both an incidence coalgebra and a path coalgebra if and only if it is the path coalgebra of a quiver Γ for which there is at most one path between any two vertices.*

Proof. If the condition is satisfied for a quiver Γ , we can introduce an obvious order on the set X of vertices of Γ setting $x \leq y$ if and only if there is a path from x to y . It is easy to check that this is an ordering, and so the above map $\phi : KX \rightarrow K\Gamma$ is bijective. Conversely, let $C \cong KX \cong K\Gamma$ for a locally finite partially ordered set X and a quiver Γ . We note that the simple subcoalgebras (and simple left subcomodules, simple right subcomodules) of C are precisely the spaces Kx for $x \in X$ and Kv for v

vertex in Γ , and X , respectively Γ correspond to the group-like elements of C . Thus, X must be the set of vertices of Γ . Furthermore, we note that in either an incidence coalgebra or a path coalgebra, the injective hull of a simple left comodule Kx is uniquely determined as follows (note that in general, given an injective module M and a submodule N of M , there is an injective hull of N contained in M but it is not necessarily uniquely determined). For incidence/path coalgebras, the right (left) injective hull $E_r(Kx)$ of Kx (respectively, $E_l(Kx)$) of the right (respectively, left) comodule Kx is the span of all segments/paths starting (respectively, ending) at x (see the proof of [Simson 2009, Proposition 2.5] for incidence coalgebras and [Chin et al. 2002, Corollary 6.2(b)] for path coalgebras). Then, for $x \leq y$, from the incidence coalgebra results we get $E_r(Kx) \cap E_l(Ky) = \langle e_{x,y} \rangle$ and from the path coalgebra we get $E_r(Kx) \cap E_l(Ky) = \langle P(x, y) \rangle$. This shows that $\langle P(x, y) \rangle$ is one dimensional, and the proof is finished. \square

Apart from the incidence coalgebra KX , there is another associated algebraic object with a combinatorial relevance. This is the incidence algebra $IA(X)$, which is the space of all functions $f : \{(x, y) \mid x, y \in X, x \leq y\} \rightarrow K$ (functions on the set of closed intervals of X), with multiplication given by convolution:

$$(fg)(x, y) = \sum_{x \leq z \leq y} f(x, z)g(z, y)$$

for any $f, g \in IA(X)$ and any $x, y \in X, x \leq y$. See [Spiegel and O'Donnell 1997] for details on the combinatorial relevance of the incidence algebra. It is clear that $IA(X)$ is isomorphic to the dual algebra of KX , if we identify a map $f \in IA(X)$ with the element of $(KX)^*$ which takes $e_{x,y}$ to $f(x, y)$ for any $x \leq y$. For simplicity, we will identify $IA(X)$ with $(KX)^*$.

Comparing to path coalgebras and quiver algebras, the situation is different, since the incidence algebra always has identity. However, we can consider the subspace $FIA(X)$ of $IA(X)$ spanned by all the elements $E_{x,y}$ with $x \leq y$, where $E_{x,y}(e_{u,v}) = \delta_{x,u}\delta_{y,v}$ for any $u \leq v$. Equivalently, $FIA(X)$ consists of all functions on $\{(x, y) \mid x, y \in X, x \leq y\}$ that have finite support. Then $FIA(X)$ is a subalgebra of $IA(X)$ which does not have an identity when X is infinite, but it has enough idempotents, the set of all $E_{x,x}$. The algebra $FIA(X)$ plays the role of the quiver algebra in this new framework.

The subspace $FIA(X)$ is dense in $IA(X)$ in the finite topology, since it is easy to see that $FIA(X)^\perp = 0$ (see [Dăscălescu et al. 2001, Corollary 1.2.9]). We have a coalgebra morphism $\theta : KX \rightarrow FIA(X)^0$, defined by $\theta(c)(c^*) = c^*(c)$ for any $c \in KX$ and any $c^* \in FIA(X)$. We note that $\theta(c)$ indeed lies in $FIA(X)^0$, since $\text{Ker}(\theta(c)) = \langle c \rangle^\perp \cap FIA(X) \supseteq D^\perp \cap FIA(X)$, where D is the (finite dimensional) subcoalgebra generated by c in KX . Then D^\perp is an ideal of $IA(X)$ of finite codimension, and then $D^\perp \cap FIA(X)$ is an ideal of $FIA(X)$ of finite codimension.

Since $FIA(X)$ is dense in $IA(X)$, θ is injective. The next result shows that we can recover the incidence coalgebra KX as the finite dual of the algebra with enough idempotents $FIA(X)$. The result parallels [Theorem 3.3](#); note that the conditions analogous to the ones in (i) in [Theorem 3.3](#) are always satisfied in incidence algebras.

Theorem 4.2. *For any locally finite partially ordered set X , the map*

$$\theta : KX \rightarrow FIA(X)^0$$

is an isomorphism of coalgebras.

Proof. It is enough to show that θ is surjective. Let $F \in FIA(X)^0$, so F maps $FIA(X)$ to K and $\text{Ker}(F)$ contains an ideal I of $FIA(X)$ of finite codimension. Then the set $X_0 = \{x \in X \mid E_{x,x} \notin I\}$ is finite by [Lemma 2.1](#).

If $x \in X \setminus X_0$, then $E_{x,y} = E_{x,x}E_{x,y} \in I$ for any $x \leq y$. Similarly $E_{x,y} \in I$ for any $y \in X \setminus X_0$ and $x \leq y$. Thus in order to have $E_{x,y} \notin I$, both x and y must lie in X_0 . This shows that only finitely many $E_{x,y}$'s lie outside I . Let \mathcal{F} be the set of all pairs (x, y) such that $E_{x,y} \notin I$. Then we have that $F = \sum_{(x,y) \in \mathcal{F}} F(E_{x,y})\theta(e_{x,y})$. Indeed, when evaluated at $E_{u,v}$, both sides are 0 if $(u, v) \notin \mathcal{F}$, or $F(E_{u,v})$ if $(u, v) \in \mathcal{F}$. Thus $F \in \text{Im}(\theta)$. \square

The next result and its proof parallel [Theorem 3.6](#).

Theorem 4.3. *Let $C = KX$. The following assertions are equivalent.*

- (i) $FIA(X) = C_l^{*\text{rat}}$.
- (ii) $FIA(X) = C_r^{*\text{rat}}$.
- (iii) *For any $x \in X$ there are finitely many elements $u \in X$ such that $u \leq x$, and finitely many elements $y \in X$ such that $x \leq y$.*
- (iv) *KX is a left and right semiperfect coalgebra.*

Proof. (i) \Rightarrow (iii) Since $E_{x,x} \in C_l^{*\text{rat}}$, there exist finite families $(c_i)_i$ in C and $(c_i^*)_i$ in C^* such that $c^*E_{x,x} = \sum_i c^*(c_i)c_i^*$ for any $c^* \in C^*$. If there are infinitely many elements u in X such that $u \leq x$, then we can choose such an element u_0 for which $e_{u_0,x}$ does not show up in the representation of any c_i (as a linear combination of the standard basis of C). Since $E_{u_0,x}(e_{p,q}) = \delta_{u_0,p}\delta_{x,q}$, we get $E_{u_0,x}(c_i) = 0$ for any i , so $\sum_i E_{u_0,x}(c_i)c_i^* = 0$, while $(E_{u_0,x}E_{x,x})(e_{u_0,x}) = 1$, a contradiction.

Assume now that for some $x \in X$ the set of all elements y with $x \leq y$, say $(y_i)_i$, is infinite. Let $c^* \in C^*$ which is 1 on each e_{x,y_i} and 0 on any other $e_{p,q}$. Then it is easy to see that

$$(d^*c^*)(e_{u,v}) = \begin{cases} d^*(e_{u,x}) & \text{if } u \leq x \leq v, \text{ and } v \in \{y_i \mid i\}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $(u_j)_j$ be the family of all elements u with $u \leq x$. As we proved above, this family is finite. For each j , let $c_j^* \in C^*$ equal 1 on every e_{u_j,y_i} , and 0 on any

other $e_{p,q}$. We have that $d^*c^* = \sum_j d^*(e_{u_j,x})c_j^*$ for any $d^* \in C^*$. Indeed, using the formula above we see that both sides equal $d^*(e_{u_{j_0},x})$ when evaluated at some $e_{u_{j_0},y_i}$, and 0 when evaluated at any other $e_{p,q}$.

Therefore $c^* \in C_l^{\text{rat}}$, but obviously $c^* \notin FIA(X)$, since it is nonzero on infinitely many $e_{p,q}$'s.

(iii) \Rightarrow (i) Choose some x, y with $x \leq y$. Then for any $c^* \in C^*$ we have that

$$(c^*E_{x,y})(e_{u,v}) = \begin{cases} c^*(e_{u,x}) & \text{if } u \leq x \leq y = v, \\ 0 & \text{otherwise.} \end{cases}$$

This shows that if $(u_j)_j$ is the finite family of all elements u with $u \leq x$, then $c^*E_{x,y} = \sum_j c^*(e_{u_j,x})E_{u_j,y}$, so $E_{x,y}$ lies in C_l^{rat} .

Now let $c^* \in C_l^{\text{rat}}$, so

$$d^*c^* = \sum_i d^*(c_i)c_i^*$$

for some finite families $(c_i)_i$ in C and $(c_i^*)_i$ in C^* . If $x \in X$ such that $e_{x,x}$ does not appear in any c_i (with nonzero coefficient), then $E_{x,x}c^* = 0$. In particular $0 = (E_{x,x}c^*)(e_{x,y}) = c^*(e_{x,y})$ for any $x \leq y$. Since only finitely many $e_{u,u}$ appear in the representations of the c_i 's, and for any such u there are finitely many v with $u \leq v$, we obtain that $c^*(e_{u,v})$ is nonzero for only finitely many $e_{u,v}$. So c^* lies in the span of all $E_{x,y}$'s, which is $FIA(X)$.

(ii) \Leftrightarrow (iii) is similar.

(iii) \Leftrightarrow (iv) follows from [Simson 2009, Lemma 5.1]. \square

5. Coreflexivity for path subcoalgebras and subcoalgebras of incidence coalgebras

We recall from [Radford 1973; Taft 1972] that a coalgebra C is called coreflexive if any finite-dimensional left (or equivalently, any finite-dimensional right) C^* -module is rational. This is also equivalent to asking that the natural embedding of C into the finite dual of C^* , $C \rightarrow (C^*)^0$ is surjective (so an isomorphism), or that any left (equivalently, any right) cofinite ideal is closed in the finite topology. See [Radford 1974; 1973; Taft 1972; 1977] for further equivalent characterizations.

Given the definition of coreflexivity and the characterizations of the previous section, it is natural to ask what is the connection between the situation when the path coalgebra can be recovered as the finite dual of the quiver algebra, and the coreflexivity of the path coalgebra. We note that these two are closely related. We have an embedding $\iota : K\Gamma \hookrightarrow (K\Gamma)^{*0}$; at the same time, we note that the embedding of algebras (without identity) $\psi : K[\Gamma] \hookrightarrow (K\Gamma)^*$ which is dense in the finite topology of $(K\Gamma)^*$, produces a comultiplicative morphism $\varphi : (K\Gamma)^{*0} \rightarrow K[\Gamma]^0$.

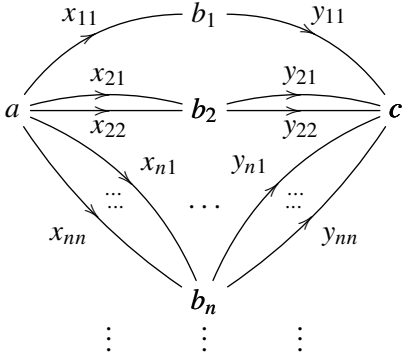
Note that φ is not necessarily a morphism of coalgebras, since it may not respect the counits. It is easy to see that these canonical morphisms are compatible with θ , i.e., they satisfy $\theta = \varphi \circ \iota$:

$$\begin{array}{ccc}
 K\Gamma & \xrightarrow{\iota} & (K\Gamma)^{*0} \\
 \searrow \theta & \searrow \hookrightarrow & \downarrow \varphi \\
 & & K[\Gamma]^0
 \end{array}$$

It is then natural to ask what is the connection between the bijectivity of θ , and coreflexivity of $K\Gamma$, i.e., bijectivity of ι . In fact, we notice that if C is coreflexive (equivalently, ι is surjective), then φ is necessarily injective.

The following two examples will show that, in fact, C can be coreflexive and θ not an isomorphism, and also that θ can be an isomorphism without C being coreflexive.

Example 5.1. Consider the path coalgebra of the following quiver Γ :



Here there are n arrows from vertex a to vertex b_n and n arrows from b_n to c for each positive integer n . We note that the one-dimensional vector space I spanned by $a - c$ is a coideal, since $a - c$ is an (a, c) -skew-primitive element. It is not difficult to observe that the quotient coalgebra C/I is isomorphic to the coalgebra from [Radford 1974, Example 3.4], and so C/I is not coreflexive, as shown in [Radford 1974]. By [Heyneman and Radford 1974, 3.1.4], we know that if I is a finite-dimensional coideal of a coalgebra C then C is coreflexive if and only if C/I is coreflexive. Therefore, C is not coreflexive. However, it is obvious that C satisfies the quiver conditions of Theorem 3.3, and therefore, $K\Gamma = K[\Gamma]^0$.

Hence, a path coalgebra of a quiver with no cycles and finitely many arrows between any two vertices is not necessarily coreflexive. Conversely, we note that in a coreflexive path coalgebra there are only finitely many arrows between any two vertices. This is true since a coreflexive coalgebra is locally finite by [Heyneman and

[Radford 1974, 3.2.4], which means that the wedge $X \wedge Y = \Delta^{-1}(X \otimes C + C \otimes Y)$ of any two finite-dimensional vector subspaces X, Y of C is finite-dimensional (one applies this for $X = Ka$ and $Y = Kb$). However, if a path coalgebra $K\Gamma$ is coreflexive, Γ may contain cycles: consider the path coalgebra C of a loop (a graph with one vertex and one arrow); C is then the divided power coalgebra, $C^* = K[[X]]$, the ring of formal power series, and its only ideals are (X^n) , which are closed in the finite topology of C^* . Thus, every finite dimensional C^* -module is rational and C is coreflexive.

We will prove coreflexivity of an interesting class of path coalgebras, whose quiver satisfy a slightly stronger condition than that required by Theorem 3.3 (so in particular, they will satisfy $K\Gamma = K[\Gamma]^0$). We first prove a general coreflexivity criterion.

Theorem 5.2. *Let C be a coalgebra with the property that for any finite dimensional subcoalgebra V there exists a finite-dimensional subcoalgebra W such that $V \subseteq W$ and $W^\perp W^\perp = W^\perp$. Then C is coreflexive if and only if its coradical C_0 is coreflexive.*

Proof. If C is coreflexive, then so is C_0 , since a subcoalgebra of a coreflexive coalgebra is coreflexive (see [Heyneman and Radford 1974, Proposition 3.1.4]). Conversely, let C_0 be coreflexive. We prove that any finite-dimensional left C^* -module M is rational, by induction on the length $l(M)$ of M . If $l(M) = 1$, i.e., M is simple, then M is also a left $C^*/J(C^*)$ -module, where $J(C^*)$ is the Jacobson radical of C^* . Since $C^*/J(C^*) \simeq C_0^*$ and C_0 is coreflexive, we have that M is a rational C_0^* -module, so then it is a rational C^* -module, too.

Assume now that the statement is true for length $< n$, where $n > 1$, and let M be a left C^* -module of length n . Let M' be a simple submodule of M , and consider the associated exact sequence

$$0 \rightarrow M' \rightarrow M \rightarrow M'' \rightarrow 0.$$

By the induction hypothesis M' and M'' are rational. By [Dăscălescu et al. 2001, Theorem 2.2.14] we have that $\text{ann}_{C^*}(M')$ and $\text{ann}_{C^*}(M'')$ are finite codimensional closed two-sided ideals in C^* . Using [Dăscălescu et al. 2001, Corollary 1.2.8 and Proposition 1.5.23], $\text{ann}_{C^*}(M') = U_1^\perp$ and $\text{ann}_{C^*}(M'') = U_2^\perp$ for some finite-dimensional subcoalgebras of C . Using the hypothesis for $V = U_1 + U_2$, there is a finite dimensional subcoalgebra W of C such that $U_1 \subseteq W$, $U_2 \subseteq W$ and $W^\perp W^\perp = W^\perp$. Then, by [Dăscălescu et al. 2001, Proposition 1.5.23],

$$W^\perp = W^\perp W^\perp \subseteq U_1^\perp U_2^\perp = \text{ann}_{C^*}(M') \text{ann}_{C^*}(M'') \subseteq \text{ann}_{C^*}(M)$$

is a two-sided closed ideal of C^* , of finite codimension. Therefore, M is a rational C^* -module by using again [Dăscălescu et al. 2001, Theorem 2.2.14]. \square

Proposition 5.3. *Let C be the path coalgebra $K\Gamma$, where Γ is a quiver such that there are finitely many paths between any two vertices. Then for any finite-dimensional subcoalgebra V of C there exists a finite-dimensional subcoalgebra W such that $V \subseteq W$ and $W^\perp W^\perp = W^\perp$. As a consequence, C is coreflexive if and only if the coradical C_0 (which is the grouplike coalgebra over the set of vertices of Γ) is coreflexive.*

Proof. Let V be a finite-dimensional subcoalgebra of $C = K\Gamma$. An element $c \in V$ is of the form

$$c = \sum_{i=1}^n \alpha_i p_i, \quad \alpha_i \neq 0,$$

a linear combination of paths p_1, \dots, p_n . Consider the set of all vertices at least one of these paths passes through, and let S_0 be the union of all these sets of vertices when c runs through the elements of V . Since V is finite-dimensional, we have that S_0 is finite (in fact, one can see that S_0 consists of all vertices in Γ which belong to V , so that KS_0 is the socle of V). Let P be the set of all paths p such that $s(p), t(p) \in S_0$. We consider the set S of all vertices at least one path of P passes through. It is clear that P is finite, and then so is S . We note that if $v_1, v_2 \in S$ and p is a path from v_1 to v_2 , then any vertex on p lies in S . Indeed, v_1 is on a path from u_1 to u'_1 (vertices in S_0), and let p_1 be its subpath from u_1 to v_1 . Similarly, v_2 is on a path from u_2 to u'_2 (in S_0), and let p_2 be the subpath from v_2 to u'_2 . Then $p_1 p p_2 \in P$, so any vertex of p is in S . Let W be the subspace spanned by all paths starting and ending at vertices in S . It is clear that any subpath of a path in W is also in W , so then W is a finite-dimensional subcoalgebra containing V (since S_0 is contained in S).

We show that $W^\perp W^\perp = W^\perp$. For this, given $\eta \in W^\perp$, we construct elements $f_1, f_2, g_1, g_2 \in W^\perp$ such that $\eta = f_1 g_1 + f_2 g_2$. We define $f_i(p)$ and $g_i(p)$, $i = 1, 2$, on all paths p by induction on the length of p . For paths p of length zero, i.e., if p is a vertex v , we define $f_i(v) = g_i(v) = 0$, $i = 1, 2$, for any $v \in S$, while for $v \notin S$, we set $f_1(v) = g_2(v) = 1$, and $f_2(v)$ and $g_1(v)$ are such that $g_1(v) + f_2(v) = \eta(v)$. Then clearly $\eta = f_1 g_1 + f_2 g_2$ on paths of length zero. For the induction step, assume that we have defined f_i and g_i , $i = 1, 2$, on all paths of length $< l$, and that $\eta = f_1 g_1 + f_2 g_2$ on any such path. Let now p be a path of length l , starting at u and ending at v . If $u, v \in S$, then we define $f_i(p) = g_i(p) = 0$, $i = 1, 2$, and clearly $\eta(p) = \sum_{i=1,2} \sum_{qr=p} f_i(q) g_i(r)$, since both sides are zero. If either $u \notin S$ or $v \notin S$, we need the following equality to hold:

$$\begin{aligned} (3) \quad & f_1(u)g_1(p) + f_1(p)g_1(v) + f_2(u)g_2(p) + f_2(p)g_2(v) \\ &= \eta(p) - \sum_{i=1,2} \sum_{\substack{qr=p \\ q \neq p, r \neq p}} f_i(q)g_i(r). \end{aligned}$$

We note that the terms of the right-hand side of the equality (3) have already been defined, because when $p = qr$ and $q \neq p$, $r \neq p$, the length of the paths q and r is strictly less than the length of p . We define $f_1(p)$ and $g_2(p)$ to be any elements of K , and then since either $f_1(u) = 1$ or $g_2(v) = 1$ (since either $u \notin S$ or $v \notin S$), we can choose suitable $g_1(p)$ and $f_2(p)$ such that (3) holds true.

The fact that C is coreflexive if and only if so is C_0 follows now follows directly from Theorem 5.2. \square

Moreover, we can extend the result in the previous proposition to subcoalgebras of path coalgebras.

Proposition 5.4. *Let C be a subcoalgebra of a path coalgebra $K\Gamma$, such that there are only finitely many paths between any two vertices in Γ . Then C is coreflexive if and only if C_0 is coreflexive.*

Proof. Let Γ' be the subquiver of Γ whose vertices are all the vertices v of Γ such that there is an element $c = \sum_i \alpha_i p_i \in C$, where the α_i 's are nonzero scalars and the p_i 's are pairwise distinct paths, and at least one p_i passes through v . The arrows of Γ' are all the arrows of Γ between vertices of Γ' . Clearly, there are only finitely many paths between any two vertices in Γ' . Then we have that C is a subcoalgebra of $K\Gamma'$ and $C_0 = (K\Gamma')_0$. Obviously, $C_0 \subset (K\Gamma')_0$; for the converse, let us consider a vertex u in Γ' , so there is $c \in C$ such that $c = \sum_i \alpha_i p_i$, with $\alpha_i \neq 0$ and distinct p_i 's, and some p_k passes through u . Let us write then $p_k = qr$ such that q ends at u and r begins at u . Since C is a subcoalgebra of $K\Gamma'$ it is also a sub-bicomodule, so then $r^*cq^* \in C$, where $q^*, r^* \in (K\Gamma')^*$ are equal to 1 on q, r respectively and 0 on all other paths of $K\Gamma'$. Now

$$r^*p_iq^* = \sum_{p_i=stw} q^*(s)tr^*(w)$$

and the only nonzero terms can occur if $p_i = qt_i r$, where t_i is a path starting and ending at u (loop at u). Let J be the set of these indices. In this situation $r^*p_iq^* = t_i$. Note that since the p_i 's are distinct, the t_j 's, $j \in J$ are distinct too. Also, since $p_k = qr$, there is at least such a j . We have

$$r^*cq^* = \sum_j \alpha_j t_j,$$

with all t_j beginning and ending at u , and $t_k = u$. Let $l \in J$ be an index such that t_l has maximum length among the t_j 's, $j \in J$. We note then that $t_l^*t_j = 0$ if $j \neq l$, since for any decomposition $t_j = st$, we have $t \neq t_l$ because of the maximality of t_l and of the fact that $t_j \neq t_l$. However, $t_l^*t_l = u$. Therefore, $t_l^*c = \alpha_l u \in C$, so $u \in C$ since $\alpha_l \neq 0$.

Thus if C_0 is coreflexive, we have that $(K\Gamma')_0$ is coreflexive, and then by [Proposition 5.3](#), we have that $K\Gamma'$ is coreflexive. Then C is coreflexive, as a subcoalgebra of $K\Gamma'$. Conversely, if C is coreflexive, clearly C_0 is coreflexive. \square

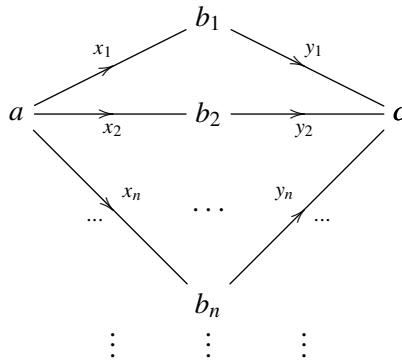
Corollary 5.5. *Let C be a subcoalgebra of an incidence coalgebra KX . Then C is coreflexive if and only if C_0 is coreflexive.*

Proof. As explained in [Section 4](#), KX can be embedded in a path coalgebra $K\Gamma$, where Γ is a quiver for which there are finitely many paths between any two vertices. Then C is isomorphic to a subcoalgebra of $K\Gamma$ and we apply [Proposition 5.4](#). \square

Recall that for a path coalgebra or incidence coalgebra C , $C_0 \sim K^{(X)}$, where X is the set of grouplike elements in C . At this point, we believe it is worth mentioning that by [[Heyneman and Radford 1974](#), Theorem 3.7.3], $K^{(X)}$ is coreflexive whenever X is a nonmeasurable cardinal. More precisely, an ultrafilter \mathcal{F} on a set X is called an *Ulam ultrafilter* if \mathcal{F} is closed under countable intersection. X is called *nonmeasurable* (or *reasonable* in the language of [[Heyneman and Radford 1974](#)]) if every Ulam ultrafilter is principal (i.e., it equals the collection of all subsets of X containing some fixed $x \in X$). The class of nonmeasurable sets contains the countable sets and is closed under usual set-theoretic constructions, such as the power set, subsets, products, and unions. If a nonreasonable (i.e., measurable) set exists, its cardinality has to be “very large” (inaccessible in the sense of set theory).

We now give an example to show that it is possible to have a coalgebra which is both coreflexive, and satisfies the path coalgebra “recovery” conditions of [Theorem 3.3](#); however, in its quiver, some vertices are joined by infinitely many paths. Thus, in general, the coreflexivity question for path coalgebras is more complicated.

Example 5.6. Consider the path coalgebra C of the following quiver Γ :



Here there are infinitely many vertices b_n , one for each positive integer n . Let W_n be the finite-dimensional subcoalgebra of C with basis

$$B = \{a, c, b_1, \dots, b_n, x_1, \dots, x_n, y_1, \dots, y_n, x_1y_1, \dots, x_ny_n\}.$$

We show that $W_n^\perp = W_n^\perp \cdot W_n^\perp$. Let $f \in W_n^\perp$. We show that we can find elements $g_1, g_2, h_1, h_2 \in W_n^\perp$ such that $f = g_1 h_1 + g_2 h_2$. This condition is already true on elements of B if we set g_1, g_2, h_1, h_2 to be zero on W_n . For $k > n$ we define g_1, g_2, h_1, h_2 on x_k, y_k and $x_k y_k$ such that

$$\begin{aligned} f(x_k y_k) &= \sum_{i=1,2} (g_i(a) h_i(x_k y_k) + g_i(x_k) h_i(y_k) + g_i(x_k y_k) h_i(c)), \\ f(x_k) &= \sum_{i=1,2} (g_i(a) h_i(x_k) + g_i(x_k) h_i(b_k)), \\ f(y_k) &= \sum_{i=1,2} (g_i(b_k) h_i(y_k) + g_i(y_k) h_i(c)), \\ f(b_k) &= \sum_{i=1,2} g_i(b_k) h_i(b_k). \end{aligned}$$

and since $g_i(a) = h_i(a) = g_i(c) = h_i(c) = 0$ this is equivalent to the matrix equality

$$\begin{pmatrix} f(b_k) & f(y_k) \\ f(x_k) & f(x_k y_k) \end{pmatrix} = \begin{pmatrix} g_1(b_k) \\ g_1(x_k) \end{pmatrix} \cdot (h_1(b_k) \ h_1(y_k)) + \begin{pmatrix} g_2(b_k) \\ g_2(x_k) \end{pmatrix} \cdot (h_2(b_k) \ h_2(y_k)).$$

But it is a standard linear algebra fact that any arbitrary 2×2 matrix can be written this way as a sum of two matrices of rank 1, and thus the claim is proved. Since every finite-dimensional subcoalgebra V of C is contained in some W_n with $W_n^\perp = W_n^\perp \cdot W_n^\perp$ and $C_0 \cong K^{\mathbb{Z}_{>0}}$ is coreflexive, by [Theorem 5.2](#) we obtain that C is coreflexive.

Reflexivity for quiver and incidence algebras. Recall from [\[Taft 1972\]](#) that an algebra is called reflexive if the natural (evaluation) map from A to A^{0*} is an isomorphism. Using our construction in Section 2, we can extend this to algebras with enough idempotents, and call such an algebra *reflexive* if the map $\Phi : a \mapsto (f \mapsto f(a)) \in A^{0*}$ is an isomorphism. We note that in general the coalgebra A^0 is a coalgebra with counit, and therefore, A^{0*} is an algebra with unit. Hence, a reflexive algebra must be unital. Parallel to algebras with unit we call an algebra *proper* if the map Φ is injective and we call A *weakly reflexive* if Φ is surjective. It is not difficult to see that an algebra is proper if and only if the intersection of all cofinite ideals is 0.

Theorem 5.7. *Let Γ be a quiver.*

- (i) *The quiver algebra $K[\Gamma]$ is proper.*
- (ii) *$K[\Gamma]$ is reflexive (equivalently, weakly reflexive) if and only if it is finite-dimensional, equivalently, Γ has finitely many vertices and arrows, and has no oriented cycles.*

Proof. (i) follows since $K[\Gamma]$ embeds in $(K\Gamma)^*$ which is proper by Proposition 3.1 of [Taft 1972], and one can easily see that Proposition 3.4 of the same reference, stating that a subalgebra of a proper algebra is proper can be extended to algebras with enough idempotents. Alternatively, one can see that the intersection of cofinite ideals of $K[\Gamma]$ is always 0.

(ii) Assume $K[\Gamma]$ is weakly reflexive, so $K[\Gamma] \rightarrow K[\Gamma]^{0*}$ is surjective. The inclusion $K\Gamma \subseteq K[\Gamma]^0$ yields a surjective morphism of algebras

$$K[\Gamma]^{0*} \rightarrow (K\Gamma)^*.$$

This shows that the natural map $\psi : K[\Gamma] \hookrightarrow (K\Gamma)^*$ is surjective (and, in fact, bijective). Consider the “gamma function” on $K[\Gamma]$, i.e., the function $\gamma \in K[\Gamma]$ equal to 1 on all paths. Then γ is in the image of ψ , and since every function in the image of ψ has finite support as a function on the set of paths of Γ , it follows that there are only finitely many paths in Γ . Therefore, $K[\Gamma]$ is finite-dimensional. The converse is obvious (as noticed before). \square

In the case of incidence algebras, using [Taft 1972, Proposition 6.1] which states that a coalgebra C is coreflexive if and only if C^* is reflexive, and using also Corollary 5.5, we immediately get this:

Theorem 5.8. *Let X be a locally finite partially ordered set. The following assertions are equivalent:*

- (i) *The incidence algebra $IA(X)$ of X over K is reflexive.*
- (ii) *The incidence coalgebra KX is coreflexive.*
- (iii) *The coalgebra $(KX)_0 = KX_0$ (the grouplike coalgebra on the elements of X) is coreflexive.*
- (iv) *The algebra K^X of functions on X is reflexive.*

These yield as a corollary the algebra analogue of Proposition 4.1.

Corollary 5.9. *Let A be an algebra of a nonmeasurable cardinality. Then A is isomorphic both to a quiver algebra and to an incidence algebra if and only if and only if it is the quiver algebra of a finite quiver with no oriented cycles, equivalently, it is elementary, finite dimensional and hereditary.*

Proof. If $A \cong K[\Gamma] \cong IA(X)$ for a quiver Γ and a locally finite partially ordered set X , then $K^{(X)}$ is coreflexive by [Heyneman and Radford 1974] since X is also nonmeasurable. Now $A \cong IA(X)$ is reflexive since $K^X \cong (K^{(X)})^*$ is reflexive by [Taft 1972, Proposition 6.1]. By Theorem 5.7, $A \cong K[\Gamma]$ must be finite dimensional since it is reflexive. The final statements follow from the well known characterizations of finite-dimensional quiver algebras. \square

An application. We give now an application of our considerations on coreflexive coalgebras. If Γ, Γ' are quivers, then we consider the quiver $\Gamma \times \Gamma'$ defined as follows. The vertices are all pairs (a, a') for a, a' vertices in Γ and Γ' respectively. The arrows are the pairs (a, x') , which is an arrow from (a, a'_1) to (a, a'_2) , where a is a vertex in Γ and x' is an arrow from a'_1 to a'_2 in Γ' , and the pairs (x, a') , which is an arrow from (a_1, a') to (a_2, a') , where x is an arrow from a_1 to a_2 in Γ , and a' is a vertex in Γ' . Let $p = x_1x_2 \dots x_n$ be a path in Γ going (in order) through the vertices a_0, a_1, \dots, a_n and $q = y_1y_2 \dots y_k$ be a path in Γ' going through vertices b_0, b_1, \dots, b_k (some vertices may repeat). We consider the 2 dimensional lattice $L = \{0, \dots, n\} \times \{0, \dots, k\}$. A lattice walk is a sequence of elements of L starting with $(0, 0)$ and ending with (n, k) , and always going either one step to the right or one step upwards in L , i.e., (i, j) is followed either by $(i + 1, j)$ or by $(i, j + 1)$. There are $\binom{n+k}{k}$ such walks.

To p, q and a lattice walk $(0, 0) = (i_0, j_0), (i_1, j_1), \dots, (i_{n+k}, j_{n+k}) = (n, k)$ in L we associate a path of length $n + k$ in $\Gamma \times \Gamma'$, starting at (a_0, b_0) and ending at (a_n, b_k) such that the r -th arrow of the path, from $(a_{i_{r-1}}, b_{j_{r-1}})$ to (a_{i_r}, b_{j_r}) is $(x_{r-1}, b_{j_{r-1}})$ if $i_r = i_{r-1} + 1$, and $(a_{i_{r-1}}, y_{r-1})$ if $j_r = j_{r-1} + 1$.

Conversely, if γ is a path in $\Gamma \times \Gamma'$, there are (uniquely determined) paths p in Γ and q in Γ' , and a lattice walk such that γ is associated to p, q and that lattice walk as above. Indeed, we take p to be the path in Γ formed by considering the arrows x such that there are arrows of the form (x, a') in γ , taken in the order they appear in γ . Similarly, q is formed by considering the arrows of the form (a, y) in γ . The lattice walk is defined according to the succession of arrows in γ .

For two such paths p, q let us denote $W(p, q)$ the set of all paths in $\Gamma \times \Gamma'$ associated to p and q via lattice walks.

Functoriality, (co)products of quivers and recovery problems. We note that if Γ and Γ' satisfy condition (i) in Theorem 3.3 (i.e., if their path coalgebras can be recovered as finite duals of the corresponding quiver algebras), then $\Gamma \times \Gamma'$ satisfies this condition, too. Indeed, the description of the arrows in $\Gamma \times \Gamma'$ shows that there are finitely many arrows between any two vertices. Also, if an oriented cycle existed in $\Gamma \times \Gamma'$, then it would produce an oriented cycle in each of Γ and Γ' .

Also, if Γ and Γ' satisfy condition (iii) in Theorem 3.6 (i.e., if their quiver algebras can be recovered as the rational part of the dual of the corresponding path coalgebras), then $\Gamma \times \Gamma'$ satisfies this condition, too. Indeed, a path in $\Gamma \times \Gamma'$ starting at the vertex (a, a') is determined by a path in Γ starting at a , a path in Γ' starting at a' (and there are finitely many such paths in both cases), and a lattice walk (chosen from a finite family). These can be extended to finite products of quivers in the obvious way.

Given a family of quivers $(\Gamma_i)_i$, one can consider the coproduct quiver $\Gamma = \coprod_i \Gamma_i$.

The path coalgebra functor commutes with coproducts and one has

$$K\Gamma = \bigoplus_i K\Gamma_i.$$

Also, the quiver algebra functor from the category of quivers to the category of algebras with enough idempotents has the property that

$$K\left[\coprod_i \Gamma_i\right] = \bigoplus_i K[\Gamma_i].$$

It is clear that $\coprod_i \Gamma_i$ satisfies the conditions of [Theorem 3.3](#) (i) if and only if each Γ_i satisfies the same condition, so each $K\Gamma_i$ can be recovered from $K[\Gamma_i]$ if and only if $K\Gamma$ is recoverable from $K[\Gamma]$. Also, each of the quivers $(\Gamma_i)_i$ satisfies condition (iii) in [Theorem 3.6](#), if and only if so does their disjoint union $\coprod_i \Gamma_i$. In coalgebra terms, this is justified by the fact that a direct sum $\bigoplus_i C_i$ of coalgebras is semiperfect if and only if each C_i is semiperfect.

Returning to coreflexivity problems, we need the following.

Lemma 5.10. *The linear map $\alpha : K\Gamma \otimes K\Gamma' \hookrightarrow K(\Gamma \times \Gamma')$ defined by $\alpha(p \otimes q) = \sum_{w \in W(p,q)} w$, where $p \in \Gamma$ and $q \in \Gamma'$ are paths, is an injective morphism of K -coalgebras.*

Proof. We keep the notations above. Denote δ and Δ the comultiplications of $K\Gamma \otimes K\Gamma'$ and $K(\Gamma \times \Gamma')$. We have

$$\begin{aligned} \delta\alpha(p \otimes q) &= \sum_{w \in W(p,q)} \sum_{w'w''=w} w' \otimes w'', \\ (\alpha \otimes \alpha)\Delta(p \otimes q) &= \sum_{p'p''=p} \sum_{q'q''=q} \sum_{\substack{u \in W(p',q') \\ v \in W(p'',q'')}} u \otimes v. \end{aligned}$$

On the one hand, if $p = p'p''$, $q = q'q''$, $u \in W(p', q')$ and $v \in W(p'', q'')$, we have $uv \in W(p, q)$. On the other hand, if $w \in W(p, q)$ and $w = w'w''$, then there exist $p'p''$ in Γ and q', q'' in Γ' such that $p = p'p''$, $q = q'q''$, $w' \in W(p', q')$ and $w'' \in W(p'', q'')$. These show that $\delta\alpha(p \otimes q) = (\alpha \otimes \alpha)\Delta(p \otimes q)$, i.e., α is a morphism of coalgebras (the compatibility with counits is easily verified).

To prove injectivity, if $p = x_1x_2 \dots x_n$ is a path in Γ starting at a_0 and ending at a_n , and $q = y_1y_2 \dots y_k$ is a path in Γ' starting at b_0 and ending at b_k , we denote by (p^*, q^*) the linear map on $K(\Gamma \times \Gamma')$ which equals 1 on the path $(x_1, b_0), \dots, (x_n, b_0), (a_n, y_1), \dots, (a_n, y_k)$ (for simplicity we also denote this path by $(p, b_0); (a_n, q)$) and 0 on the rest of the paths. Let $\sum_i \lambda_i p_i \otimes q_i \in \text{Ker}(\alpha)$. Then we have

$$(4) \quad \sum_i \sum_{w \in W(p_i, q_i)} \lambda_i w = 0.$$

Fix some j . Say that p_j ends at a_n and q_j starts at b_0 . We have that

$$(p_j^*, q_j^*)(w) = \begin{cases} 0 & \text{if } w \in W(p_i, q_i), i \neq j, \\ 0 & \text{if } w \in W(p_j, q_j) \text{ and } w \neq (p_j, b_0), (a_n, q_j), \\ 1 & \text{if } w = (p_j, b_0), (a_n, q_j). \end{cases}$$

Note that we used the fact that $W(p, q) \cap W(p', q') = \emptyset$ for $(p, q) \neq (p', q')$. Now applying (p_j^*, q_j^*) to (4) we see that $\lambda_j = 0$. We conclude that α is injective. \square

Combining the above, we derive a result about tensor products of certain coreflexive coalgebras. It is known that a tensor product of a coreflexive and a strongly coreflexive coalgebra is coreflexive (see [Radford 1973]; see also [Taft 1977]). It is not known whether the tensor product of coreflexive coalgebras is necessarily coreflexive. We have the following consequences.

Proposition 5.11. *Let C, D be coreflexive subcoalgebras of path coalgebras $K\Gamma$ and $K\Gamma'$ respectively such that between any two vertices in Γ and Γ' respectively there are only finitely many paths. Then $C \otimes D$ is coreflexive.*

Proof. Without any loss of generality we may assume that $C_0 = (K\Gamma)_0 = K^{(\Gamma_0)}$ and $D_0 = (K\Gamma')_0 = K^{(\Gamma'_0)}$ (otherwise we replace Γ and Γ' by appropriate subquivers), where $K^{(\Gamma_0)}$ denotes the grouplike coalgebra with basis the set Γ_0 of vertices of Γ . Now $C \otimes D$ is a subcoalgebra of $K\Gamma \otimes K\Gamma'$, so by Lemma 5.10, it also embeds in $K(\Gamma \times \Gamma')$. Since the coradical of $K(\Gamma \times \Gamma')$ is $K^{(\Gamma_0 \times \Gamma'_0)}$, and

$$K^{(\Gamma_0 \times \Gamma'_0)} \simeq K^{(\Gamma_0)} \otimes K^{(\Gamma'_0)} = C_0 \otimes D_0 \subseteq C \otimes D,$$

we must have that $(C \otimes D)_0 = K^{(\Gamma_0 \times \Gamma'_0)}$. We claim that $K^{(\Gamma_0 \times \Gamma'_0)}$ is coreflexive. Indeed, this is obvious if Γ_0 and Γ'_0 are both finite. Otherwise, $\text{card}(\Gamma_0 \times \Gamma'_0) = \max\{\text{card}(\Gamma_0), \text{card}(\Gamma'_0)\}$, hence $K^{(\Gamma_0 \times \Gamma'_0)}$ is isomorphic either to $K^{(\Gamma_0)}$ or to $K^{(\Gamma'_0)}$, both of which are coreflexive by Proposition 5.4. Since it is clear that in $\Gamma \times \Gamma'$ there are also finitely many paths between any two vertices, we can use Proposition 5.4 to show that $C \otimes D$ is coreflexive. \square

Corollary 5.12. *If C, D are coreflexive subcoalgebras of incidence coalgebras, then $C \otimes D$ is coreflexive.*

Proof. It follows immediately from the embedding of C and D in path coalgebras verifying the hypothesis of Proposition 5.11. \square

Acknowledgment

The authors would like to greatly acknowledge the very careful reading and detailed comments of the referee, which led to improvements in the mathematics and the exposition. In particular, we are in debt for questions and suggestions that prompted the expansion of Sections 3, 4 and 5 with several new results.

Acknowledgements

The research of this paper was partially supported by the UEFISCDI Grant PN-II-ID-PCE-2011-3-0635, contract no. 253/5.10.2011 of CNCSIS. For Iovanov, this work was supported by the strategic grant POSDRU/89/1.5/S/58852, Project “Postdoctoral program for training scientific researchers” cofinanced by the European Social Fund within the Sectorial Operational Program Human Resources Development 2007–2013.

References

- [Abe 1980] E. Abe, *Hopf algebras*, Cambridge Tracts in Mathematics **74**, Cambridge University Press, Cambridge, 1980. [MR 83a:16010](#) [Zbl 0476.16008](#)
- [Chin and Montgomery 1997] W. Chin and S. Montgomery, “Basic coalgebras”, pp. 41–47 in *Modular interfaces: modular Lie algebras, quantum groups, and Lie superalgebras* (Riverside, CA, 1995), edited by V. Chari and I. B. Penkov, AMS/IP Stud. Adv. Math. **4**, Amer. Math. Soc., Providence, RI, 1997. [MR 99c:16037](#) [Zbl 0920.16018](#)
- [Chin et al. 2002] W. Chin, M. Kleiner, and D. Quinn, “Almost split sequences for comodules”, *J. Algebra* **249**:1 (2002), 1–19. [MR 2003e:16045](#) [Zbl 1005.16034](#)
- [Dăscălescu et al. 2001] S. Dăscălescu, C. Năstăsescu, and Ş. Raianu, *Hopf algebras: an introduction*, Pure and Applied Mathematics **235**, Marcel Dekker, New York, 2001. [MR 2001j:16056](#) [Zbl 0962.16026](#)
- [Dăscălescu et al. \geq 2013] S. Dăscălescu, M. C. Iovanov, and C. Năstăsescu, “Path subcoalgebras, finiteness properties and quantum groups”, preprint. To appear in *J. Noncommut. Geom.* [arXiv 1012.4335](#)
- [Gómez-Torrecillas and Navarro 2008] J. Gómez-Torrecillas and G. Navarro, “Serial coalgebras and their valued Gabriel quivers”, *J. Algebra* **319**:12 (2008), 5039–5059. [MR 2010e:16054](#) [Zbl 1159.16029](#)
- [Green 1976] J. A. Green, “Locally finite representations”, *J. Algebra* **41**:1 (1976), 137–171. [MR 54 #348](#) [Zbl 0369.16008](#)
- [Heyneman and Radford 1974] R. G. Heyneman and D. E. Radford, “Reflexivity and coalgebras of finite type”, *J. Algebra* **28** (1974), 215–246. [MR 49 #10727](#) [Zbl 0291.16008](#)
- [Iovanov 2011] M. C. Iovanov, “Serial linear categories, quantum groups, and open questions in corepresentation theory”, preprint, 2011, Available at <http://dornsife.usc.edu/assets/sites/199/docs/Papers/ProbCoalg.pdf>.
- [Joni and Rota 1979] S. A. Joni and G.-C. Rota, “Coalgebras and bialgebras in combinatorics”, *Stud. Appl. Math.* **61**:2 (1979), 93–139. [MR 81c:05002](#) [Zbl 0471.05020](#)
- [Montgomery 1993] S. Montgomery, *Hopf algebras and their actions on rings*, CBMS Regional Conference Series in Mathematics **82**, Amer. Math. Soc., Providence, RI, 1993. [MR 94i:16019](#) [Zbl 0793.16029](#)
- [Radford 1973] D. E. Radford, “Coreflexive coalgebras”, *J. Algebra* **26** (1973), 512–535. [MR 48 #6160](#) [Zbl 0272.16012](#)
- [Radford 1974] D. E. Radford, “On the structure of ideals of the dual algebra of a coalgebra”, *Trans. Amer. Math. Soc.* **198** (1974), 123–137. [MR 49 #10728](#) [Zbl 0293.16012](#)

- [Simson 2009] D. Simson, “Incidence coalgebras of intervally finite posets, their integral quadratic forms and comodule categories”, *Colloq. Math.* **115**:2 (2009), 259–295. [MR 2010b:16076](#) [Zbl 1173.16009](#)
- [Spiegel and O’Donnell 1997] E. Spiegel and C. J. O’Donnell, *Incidence algebras*, Pure and Applied Mathematics **206**, Marcel Dekker, New York, 1997. [MR 98g:06001](#) [Zbl 0871.16001](#)
- [Sweedler 1969] M. E. Sweedler, *Hopf algebras*, W. A. Benjamin, New York, 1969. [MR 40 #5705](#) [Zbl 0194.32901](#)
- [Taft 1972] E. J. Taft, “Reflexivity of algebras and coalgebras”, *Amer. J. Math.* **94**:4 (1972), 1111–1130. [MR 46 #9095](#) [Zbl 0261.16013](#)
- [Taft 1977] E. J. Taft, “Reflexivity of algebras and coalgebras, II”, *Comm. Algebra* **5**:14 (1977), 1549–1560. [MR 58 #781](#) [Zbl 0369.16009](#)
- [Villarreal 2001] R. H. Villarreal, *Monomial algebras*, Pure and Applied Mathematics **238**, Marcel Dekker, New York, 2001. [MR 2002c:13001](#) [Zbl 1002.13010](#)

Received November 7, 2011. Revised July 20, 2012.

SORIN DĂSCĂLESCU
FACULTATEA DE MATEMATICA SI INFORMATICA
UNIVERSITY OF BUCHAREST
STR ACADEMIEI NR. 14, SECTOR 1
010014 BUCHAREST
ROMANIA
sdascal@fmi.unibuc.ro

MIODRAG C. IOVANOV
FACULTATEA DE MATEMATICA SI INFORMATICA
UNIVERSITY OF BUCHAREST
STR ACADEMIEI NR. 14, SECTOR 1
010014 BUCHAREST
ROMANIA

and

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
3620 S. VERMONT AVENUE, KAP 108
LOS ANGELES, CA 90089
UNITED STATES

iovanov@usc.edu

CONSTANTIN NĂSTĂSESCU
FACULTATEA DE MATEMATICA SI INFORMATICA
UNIVERSITY OF BUCHAREST
STR ACADEMIEI NR. 14, SECTOR 1
010014 BUCHAREST
ROMANIA

Constantin_nastasescu@yahoo.com

A POSITIVE DENSITY OF FUNDAMENTAL DISCRIMINANTS WITH LARGE REGULATOR

ÉTIENNE FOUVRY AND FLORENT JOUVE

We prove that there is a positive density of positive fundamental discriminants D such that the fundamental unit $\varepsilon(D)$ of the ring of integers of the field $\mathbb{Q}(\sqrt{D})$ is essentially greater than D^3 .

1. Introduction

Let $D > 1$ be a fundamental discriminant which means that D is the discriminant of the quadratic field $K := \mathbb{Q}(\sqrt{D})$. Let \mathbb{Z}_K be its ring of integers and let $\omega = (D + \sqrt{D})/2$. Then \mathbb{Z}_K is a \mathbb{Z} -module of rank 2

$$(1) \quad \mathbb{Z}_K = \mathbb{Z} \oplus \mathbb{Z} \omega.$$

Furthermore there exists a unique element $\varepsilon(D) > 1$ such that the group \mathbb{U}_K of invertible elements of \mathbb{Z}_K has the shape

$$\mathbb{U}_K = \{\pm \varepsilon(D)^n; n \in \mathbb{Z}\}.$$

The element $\varepsilon(D)$ is called the *fundamental unit* of \mathbb{Z}_K and its logarithm $R(D) := \log \varepsilon(D)$ is called the *regulator*. The regulator $R(D)$ is a central object of algebraic number theory. For instance $R(D)$ plays a role in the computation of the class number (see (34)). The study of the properties of the unruly function $D \mapsto R(D)$ is a fascinating problem in both theoretical and computational aspects (see [Cohen 1993], for instance).

A rather similar but not completely equivalent problem — see the discussion in Section 5 — is the study of the fundamental solution ε_d to the so-called *Pell equation*

$$\text{PE}(d) \quad T^2 - dU^2 = 1,$$

where the parameter d is a nonsquare positive integer and the unknown is the pair (T, U) of integers. It is convenient to write any given solution of $\text{PE}(d)$ under the form $T + U\sqrt{d}$. Let ε_d be the least of these solutions greater than 1. Then the set of solutions of $\text{PE}(d)$ is infinite and also has the shape $\{\pm \varepsilon_d^n; n \in \mathbb{Z}\}$.

MSC2010: primary 11D09; secondary 11R11.

Keywords: regulator of a real quadratic field, Pell equation.

It is known that there exists an absolute constant C such that the following inequalities hold

$$(2) \quad \sqrt{D} < \varepsilon(D) \leq \exp(C\sqrt{D} \log D) \quad \text{and} \quad 2\sqrt{d} < \varepsilon_d \leq \exp(C\sqrt{d} \log d).$$

It is widely believed that most of the time $\varepsilon(D)$ and ε_d are huge compared to the size of D or d , and this fact is confirmed by numerical evidence. One can find more precise conjectures ([Hooley 1984; Sarnak 1985], for instance) which would imply in particular that for all $\varepsilon > 0$ the inequality

$$(3) \quad \varepsilon_d \geq \exp d^{(1/2)-\varepsilon},$$

holds for almost all nonsquare d (and for almost all fundamental discriminants D , since these D form a subset of positive density). Recall that a subset \mathcal{A} of positive integers is said to have a *positive density* if its counting function satisfies the inequality

$$\liminf \frac{\#\{a \in \mathcal{A}; 1 \leq a \leq x\}}{x} > 0 \quad (x \rightarrow \infty).$$

The set \mathcal{A} is said to be *negligible* (or *with zero density*) if one has

$$\limsup \frac{\#\{a \in \mathcal{A}; 1 \leq a \leq x\}}{x} = 0 \quad (x \rightarrow \infty).$$

Since a proof of (3) still seems to be out of reach, it is a challenging problem to construct infinite sequences of fundamental discriminants D (resp. of nonsquare d) with a huge $\varepsilon(D)$ (resp. with a huge ε_d). In the case of fundamental discriminants D , it is now proved that there exists $c > 1$ such that the inequality $\varepsilon(D) > \exp(\log^c D)$ is true for infinitely many D 's; see, for example, [Yamamoto 1971; Reiter 1985; Halter-Koch 1989].

In the case of a nonsquare d the situation is better understood. Indeed we know that for some positive c there exists infinitely many d 's such that $\varepsilon_d > \exp(d^c)$. We refer the reader to the pioneering work of Dirichlet [1856] leading to the optimality of (2), and to more recent work on the subject, for instance [Zagier 1981, pp. 74, 85; Fouvry and Jouve 2012, Theorem 2]. See also [Golubeva 1987] for the study of the case $d = 5p^2$. However none of these works manages to produce an infinite family of squarefree d 's.

Besides, it is not known whether there exists a constant $c > 1$ such that the inequality $\varepsilon_d \geq \exp(\log^c d)$ holds for a positive density of d 's. So we may ask for the frequency of weaker inequalities, such as $\varepsilon_d > d^\theta$ or $\varepsilon(D) > D^\theta$, where $\theta > \frac{1}{2}$ is a fixed constant. In that direction, Hooley [1984, Corollary] proved that for almost every d one has $\varepsilon_d > d^{(3/2)-\varepsilon}$. This was improved to $\varepsilon_d > d^{(7/4)-\varepsilon}$ by Fouvry and Jouve [2013, Corollary 1] ($\varepsilon > 0$ arbitrary).

The same work of Hooley implies that there exists a positive density of d satisfying $\varepsilon_d > d^{3/2} / \log d$. By a complete different technique based on the theory of continued fractions, Golubeva [2002, Theorem] constructed a set of d 's of positive density such that $\varepsilon_d \geq d^{2-\varepsilon}$ ($\varepsilon > 0$ arbitrary). It does not seem to be an easy task to extend these two results to the case of a fundamental D because the condition for an integer to be squarefree seems hard to insert in the corresponding proofs of Hooley and Golubeva.

Our main result asserts that there is a positive density of positive fundamental discriminants D with fundamental unit of size essentially larger than D^3 . In fact we can say more: first we show it is enough to consider the contribution of positive fundamental discriminants with fundamental unit of positive norm to get our density estimate. Moreover we can further restrict our study to positive fundamental discriminants D that satisfy a very specific divisibility property. This property is of an algebraic nature. To explain precisely what it is, we state the following proposition the first version of which goes back (at least) to Dirichlet (see the beginning of Section 3 for historical background and references).

If $D > 1$ is a fundamental discriminant set

$$D' = \begin{cases} D & \text{if } D \text{ is odd,} \\ D/2 & \text{if } D = 4d, d \equiv 3 \pmod{4}, \\ D/4 & \text{if } 8 \mid D. \end{cases}$$

In other words D' is the kernel of D . Finally let Fund^+ denote the set of fundamental discriminants $D > 1$ such that $\varepsilon(D)$ has norm 1.

Proposition 1. *For every $D \in \text{Fund}^+$ there exists exactly two distinct positive divisors of D' , both different from 1 and $D/(4, D)$, among the set of norms of principal ideals of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$.*

Let Φ be the function on Fund^+ sending D to the minimum of the two distinct divisors of D' the existence of which is guaranteed by Proposition 1. With notation as above our main result can be stated as follows.

Theorem 2. *For every $\delta > 0$ there exists $x_0(\delta) > 0$ and $c_0(\delta) > 0$ such that*

$$(4) \quad \#\{D \in \text{Fund}^+ ; X < D \leq 2X, 2^2 \parallel D, \Phi(D) < D^\delta, \varepsilon(D) \geq D^{3-\delta}\} \geq c_0(\delta)X,$$

for every $X > x_0(\delta)$.

Similar statements are true when the condition $2^2 \parallel D$ in the set on the left-hand side is replaced by $8 \mid D$, or $D \equiv 1 \pmod{4}$.

We shall mainly concentrate on the case $2^2 \parallel D$ since the situation is simplified a lot thanks to an easy link between units of $\mathbb{Q}(\sqrt{D})$ and the equation PE($d/4$) via the equality

$$(5) \quad \varepsilon(D) = \varepsilon_{D/4}.$$

Proposition 1 can naturally be seen as a feature of the algebraic interpretation of the transformation of Legendre and Dirichlet we describe in [Section 2.1](#). We devote [Section 3](#) to the proof of this statement. The proof of (4) in [Theorem 2](#) is given in [Section 4](#). The cases $8 \mid D$ and D odd will be treated in [Section 5](#).

The last part of the paper explains another application of the ideas leading to [Theorem 2](#). It is well known that any information on the size of $\varepsilon(D)$ can be interpreted in terms of the ordinary class number $h(D)$ of the field $\mathbb{Q}(\sqrt{D})$. Among the various possible illustrations, we have selected the following one.

Theorem 3. *Let C_0 denote the converging Euler product:*

$$C_0 := \prod_{p \geq 3} \left(1 + \frac{p}{(p+1)^2(p-1)} \right).$$

There exists a constant $\delta > 0$ such that for every sufficiently large x one has the inequality

$$(6) \quad \sum_{\substack{D \leq x \\ 2^2 \parallel D}} h(D) \leq \left(\frac{8}{21\pi^2} C_0 - \delta \right) \frac{x^{3/2}}{\log x}.$$

The proof of this theorem is essentially based on [\[Fouvry and Jouve 2013\]](#) and [Proposition 7](#). It will be given in [Section 6](#) where we will explain why the inequality (6) is better than the trivial upper bound by some constant factor strictly larger than 3.5. We shall also use in a crucial way the fact that the set of D 's with a large $\varepsilon(D)$ exhibited in [Theorem 2](#) has some regularity. More precisely this set consists, up to a few exceptions, in integers of the form pm with p large (see (29) for the definition of $\mathcal{D}_m^\gamma(x)$). However the inequality (6) is certainly far from giving a crucial step towards the proof of the following expected asymptotic formula

$$\sum_{\substack{D \leq x \\ 2^2 \parallel D}} h(D) \sim c_0 x \log^2 x,$$

where x tends to infinity and c_0 is some absolute positive constant.

2. Preliminaries

2.1. Legendre and Dirichlet's transformation. In this subsection d denotes any positive integer, not necessarily a fundamental discriminant. We describe and use an easy transformation of the Pell equation [PE\(\$d\$ \)](#) which was initiated by Legendre [\[Legendre 1830, Chapter VII, pp. 61–74\]](#) and then extended by Dirichlet [\[Dirichlet 1834, Section 1\]](#). For the sake of completeness we give the detail of Legendre's argument. For a more detailed presentation together with historical background

and interpretations of this technique we refer to [Lemmermeyer 2003]. See also [Hooley 1984, p. 109; Cremona and Odoni 1989, pp. 18–19].

Let us write $\text{PE}(d)$ under the form

$$(7) \quad \frac{T^2 - 1}{d} = U^2.$$

Since $d \mid T^2 - 1$, we have $d = (T^2 - 1, d) = ((T + 1)(T - 1), d)$. Because the gcd of $T + 1$ and $T - 1$ can only take the values 1 or 2, we are led to consider the two corresponding cases:

- If $T + 1$ and $T - 1$ are coprime (i.e., T is even), we factorize

$$d = (T + 1, d)(T - 1, d) =: d_1 d_2,$$

in a unique way. Combining this splitting of d with (7) yields the four equations

$$T + 1 = d_1 U_1^2, \quad T - 1 = d_2 U_2^2, \quad d = d_1 d_2, \quad U = U_1 U_2,$$

which are equivalent to

$$(8) \quad d_1 U_1^2 - d_2 U_2^2 = 2, \quad T = -1 + d_1 U_1^2, \quad d = d_1 d_2, \quad U = U_1 U_2, \quad 2 \nmid d_1 U_1.$$

- If $2 = (T + 1, T - 1)$, two subcases are to be considered:

– either $4 \nmid d$, in which case U is even and the Equation (7) can be written

$$\frac{((T + 1)/2) \cdot ((T - 1)/2)}{d} = (U/2)^2.$$

Arguing as in the previous case we are reduced to considering the following set of equations:

$$(9) \quad d_1 U_1^2 - d_2 U_2^2 = 1, \quad T = -1 + 2d_1 U_1^2, \quad d = d_1 d_2, \quad U = 2U_1 U_2, \quad 4 \nmid d,$$

– or $4 \mid d$, in which case we can write (7) as follows:

$$\frac{((T + 1)/2) \cdot ((T - 1)/2)}{(d/4)} = U^2.$$

We factorize $d/4 = ((T + 1)/2, d/4)((T - 1)/2, d/4) =: d_1 d_2$ and get the set of equations

$$(10) \quad d_1 U_1^2 - d_2 U_2^2 = 1, \quad T = -1 + 2d_1 U_1^2, \quad d = 4d_1 d_2, \quad U = U_1 U_2.$$

The following statement summarizes the above decomposition in a more concise and applicable way.

Lemma 4 (Legendre and Dirichlet). *Let $d, U \in \mathbb{N}_{\geq 1}$ be fixed integers. Set*

$$\mathcal{A}(d, U) := \{T \geq 1; T^2 - dU^2 = 1\}$$

and

• if $2 \nmid dU$:

$$\mathcal{B}(d, U) := \{(d_1, d_2, U_1, U_2) \in \mathbb{N}_{\geq 1}^4; U_1 U_2 = U, d_1 d_2 = d, d_1 U_1^2 - d_2 U_2^2 = 2\},$$

• if $2 \mid dU$ and $4 \nmid d$:

$$\mathcal{B}(d, U) := \{(d_1, d_2, U_1, U_2) \in \mathbb{N}_{\geq 1}^4; 2U_1 U_2 = U, d_1 d_2 = d, d_1 U_1^2 - d_2 U_2^2 = 1\},$$

• if $4 \mid d$:

$$\mathcal{B}(d, U) := \{(d_1, d_2, U_1, U_2) \in \mathbb{N}_{\geq 1}^4; U_1 U_2 = U, 4d_1 d_2 = d, d_1 U_1^2 - d_2 U_2^2 = 1\}.$$

Then in each case, we have

$$\#\mathcal{A}(d, U) = \#\mathcal{B}(d, U) \in \{0, 1\}.$$

Proof. We start with the obvious observation that $\#\mathcal{A}(d, U) \in \{0, 1\}$. We give the rest of the argument in detail only in the first case, the other two cases being exactly similar.

Next, $\#\mathcal{B}(d, U) \in \{0, 1\}$. To see this we fix (d_1, d_2, U_1, U_2) a quadruple in $\mathcal{B}(d, U)$ and we show that the values of d_1, U_1 are prescribed by those of d, U . We compute the square of $d_1 U_1^2 - 1 = d_2 U_2^2 + 1$: it is $(d_1 U_1^2 - 1)(d_2 U_2^2 + 1) = dU^2 + 1$. Thus $d_1 U_1^2 - 1$ is determined by d, U and so is the gcd $(d_1 U_1^2, d)$. We claim this gcd is d_1 . Indeed $(d_1, d_2) = 1$ since these integers satisfy $d_1 U_1^2 - d_2 U_2^2 = 2$ and $2 \nmid dU$. Thus if $(d_1 U_1^2, d) \neq d_1$, there is a nontrivial common factor q to U_1 and d_2 . Again using the equation satisfied by (d_1, d_2, U_1, U_2) we deduce $q = 2$, contradicting the condition $2 \nmid dU$.

To conclude the proof we observe that both the implications

$$(\#\mathcal{A}(d, U) = 1) \Rightarrow (\#\mathcal{B}(d, U) \geq 1) \quad \text{and} \quad (\#\mathcal{B}(d, U) = 1) \Rightarrow (\#\mathcal{A}(d, U) \geq 1)$$

hold. The first implication is just a way of rephrasing the reduction step explained before the statement of the lemma. To prove the second implication we notice that a quadruple (d_1, d_2, U_1, U_2) gives rise to an element $T := d_2 U_2^2 + 1 = d_1 U_1^2 - 1$ belonging to $\mathcal{A}(d, U)$. \square

2.2. Remarks on Lemma 4. The first remark concerns the implicit decomposition $(d, T, U) \mapsto (d_1, d_2, U_1, U_2)$ of Lemma 4, which should really be seen as a square rooting process. This explains the efficiency of the method as a tool to study the size of the solutions to the Pell equation $\text{PE}(d)$. More precisely, a solution $T + U\sqrt{d}$ to $\text{PE}(d)$ produces via Lemma 4 the algebraic integer $U_1\sqrt{d_1} + U_2\sqrt{d_2}$ which has degree at most 4 (and at least 2 when d is not a square) over \mathbb{Q} and which satisfies

$$(U_1\sqrt{d_1} + U_2\sqrt{d_2})^2 = d_1 U_1^2 + d_2 U_2^2 + 2U_1 U_2 \sqrt{d_1 d_2}.$$

If T is odd this is precisely $T + U\sqrt{d}$. If T is even this number is $2(T + U\sqrt{d})$. Therefore [Lemma 4](#) enables us to significantly reduce the order of magnitude of the algebraic integers we work with.

The second remark concerns the special case where $d = p \equiv \pm 1 \pmod{4}$. In that case the integer d can only be factored in two ways under the form $d = d_1d_2$: either $(d_1, d_2) = (1, p)$ or $(d_1, d_2) = (p, 1)$. Hence the study of the equation $T^2 - pU^2 = 1$ is reduced to the four equations

$$U_1^2 - pU_2^2 = \begin{cases} \pm 2 & \text{if } 2 \nmid U, \\ \pm 1 & \text{if } 2 \mid U. \end{cases}$$

Since $U_2 \geq 1$ we deduce that $U_1 \geq \sqrt{p-2}$, and also that in every case one has the inequality $U \geq \sqrt{p-2}$. Hence any nontrivial solution $\Xi = T + U\sqrt{p}$ of the Pell equation $T^2 - pU^2 = 1$ satisfies the inequality

$$\Xi = \sqrt{pU^2 + 1} + U\sqrt{p} \geq \sqrt{p(p-2) + 1} + \sqrt{p(p-2)} \geq p.$$

This shows that the fundamental solution ε_p of [PE\(p\)](#) satisfies the inequality

$$(11) \quad \varepsilon_p > p.$$

For $p \equiv 3 \pmod{4}$ we deduce the lower bound

$$(12) \quad \varepsilon(4p) > p,$$

for the fundamental unit of $\mathbb{Q}(\sqrt{4p})$. In the general case of the equation $T^2 - dU^2 = 1$, the corresponding fundamental solution is greater than $2\sqrt{d}$ and this bound is essentially best possible, as the choice $d = n^2 - 1$ shows.

As E. P. Golubeva pointed out to us, the lower bound [\(11\)](#), which is certainly already in the literature, can be deduced from properties of the continued fraction expansion of \sqrt{p} . For instance, by Perron [\[1913, Satz 14, p. 94\]](#) we know that if the nonsquare integer d is such that the period k of the expansion of \sqrt{d} is even then it has the shape

$$\sqrt{d} = [b_0; \overline{b_1, \dots, b_{\nu-1}, b_\nu, b_{\nu-1}, \dots, b_1, 2b_0}],$$

where b_0 is the integral part of \sqrt{d} , the central coefficient b_ν of index $\nu := k/2$ either equals b_0 or $b_0 - 1$ or is less than $(2/3)b_0$, and where any b_ℓ , $1 \leq \ell < \nu$, is less than $(2/3)b_0$. If d is divisible by some prime congruent to $3 \pmod{4}$ it is well known that the associated integer k is even. In the particular case where $d = p \equiv 3 \pmod{4}$ we even know that $b_\nu = b_0$ or $b_0 - 1$ (see [\[Golubeva 1993, p. 1277\]](#)). Note that this last property is false if $d \equiv 3 \pmod{4}$ is not a prime. Consider for instance $\sqrt{15} = [3; \overline{1, 6}]$.

Classical properties of continued fraction expansions of quadratic integers imply that if \sqrt{d} has even period $k = 2\nu$, the fundamental solution $T_0 + U_0\sqrt{d}$ of [PE\(d\)](#)

satisfies

$$\frac{T_0}{U_0} = [b_0; b_1, \dots, b_{v-1}, b_v, b_{v-1}, \dots, b_1].$$

We deduce from the above discussion that in the case $d = p \equiv 3 \pmod{4}$ one has

$$U_0 \geq b_v \geq b_0 - 1 \geq \sqrt{p} - 2.$$

This gives (11).

3. Proof of Proposition 1

This result has been known for a long time. Dirichlet [1834, Section 5]) was the first to solve the question of the uniqueness of the decomposition $d = d_1 d_2$ (or $d = 4d_1 d_2$) appearing in (8), (9) and (10) but without, of course, using the language of modern algebraic number theory. We reprove this uniqueness result for squarefree d in passing in Section 3.1. For a statement using the language of binary quadratic forms see [Pall 1969], where the author notes that the result at issue essentially follows from a theorem due to Gauss (see the references in [Pall 1969]). For more on this subject we refer the reader to [Lemmermeyer 2003], in particular Theorem 3.3 there and the subsequent discussion. (The statement of that theorem contains a minor typo. One should allow the right-hand side of the equation to be negative since, e.g., the set of integral solutions (r, s) to each of the two equations $pr^2 - s^2 = 1$ and $pr^2 - s^2 = 2$ is empty if $p \equiv 7 \pmod{8}$.)

3.1. Applying Gauss's theorem on the 2-rank of C_D . Let $D \in \text{Fund}^+$. We denote by Cl_D (resp. C_D) the group of ideal classes of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$ in the ordinary (resp. narrow) sense. Let p_i , $1 \leq i \leq t$, be the pairwise distinct prime divisors of D . These primes are precisely the ones ramifying in $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$. For each $1 \leq i \leq t$, let \mathfrak{p}_i be the prime ideal of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$ above p_i . Let us define:

$$(13) \quad M = \{\mathfrak{p}_1^{\delta_1} \cdots \mathfrak{p}_t^{\delta_t}; \delta_i \in \{0, 1\} \text{ for all } i\}.$$

It is exactly the set of integral ideals of norm dividing D' .

Let \mathcal{S} be the subgroup of the group of fractional ideals of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$ generated by the prime ideals \mathfrak{p}_i , $1 \leq i \leq t$. Of course M is a subset of \mathcal{S} . Moreover a well known result of Gauss (see, e.g., [Fröhlich and Taylor 1993, Chapter V, Theorem 39]) asserts that the narrow class map

$$\nu : \mathcal{S} \rightarrow C_D$$

induces a surjection

$$\mathcal{S}/\mathcal{S}^2 \rightarrow C_{D,2} := \{g \in C_D : g^2 = 1\},$$

whose kernel has order 2 and where \mathcal{G}^2 denotes the subgroup of squares of the abelian group \mathcal{G} .

One deduces that each class in $C_{D,2}$ has exactly two representatives in M . In particular, the image under the narrow class map of

$$P_{\mathbb{Q}(\sqrt{D})}^+ :=$$

{fractional principal ideals of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$ generated by a totally positive element},

which is the trivial class of $C_{D,2}$, has two representatives in M . These representatives are (1) and a nontrivial ideal $I \in M$. By definition of M the norm of I divides D' . Besides it is easily seen that the norm of I is not $D/(4, D)$. Indeed if by contradiction the norm of I were $D/(4, D)$ then the ideal I would be principal and equal to $(\sqrt{D/(4, D)})$, since $I \in M$. However $D \in \text{Fund}^+$ and $(\sqrt{D/(4, D)})$ is generated by an element of negative norm. Thus $(\sqrt{D/(4, D)})$ cannot be a representative of the trivial class of $C_{D,2}$.

It turns out the ideal I can be described explicitly thanks to the Legendre–Dirichlet transformation. To see this let us analyze each case separately.

(i) *Assume first that $D = 4d$, $d \equiv 3 \pmod{4}$.* The fundamental unit of $\mathbb{Q}(\sqrt{D})$ may be written $\varepsilon(D) = T + U\sqrt{d}$. Applying the transformation described in [Section 2.1](#) to the norm equation $T^2 - dU^2 = 1$ leads either to [\(8\)](#) or [\(9\)](#) depending on whether T is even or odd.

- In case we are led to [\(8\)](#) (i.e., T is even) the integer $2d_1 > 1$ is a divisor of D' thus the ideal I is $(d_1U_1 + U_2\sqrt{d})$. Indeed the norm of the algebraic integer $d_1U_1 + U_2\sqrt{d}$ is $2d_1 > 0$ (note that $U_1\sqrt{d} + U_2d_2$ has norm $-2d_2 < 0$).
- Otherwise T is odd, hence U is even. Therefore, as explained in [Section 2.2](#), $\varepsilon(D) = T + U\sqrt{d}$ is the square of the algebraic integer $U_1\sqrt{d_1} + U_2\sqrt{d_2}$. We deduce $d_1 > 1$ since otherwise this algebraic integer would be a unit (it would have norm 1) of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$ contradicting the minimality of $\varepsilon(D)$. Thus one also has $I = (d_1U_1 + U_2\sqrt{d})$, the element $d_1U_1 + U_2\sqrt{d}$ having norm $d_1 > 0$.

(ii) *The second case we consider is $D \equiv 1 \pmod{4}$.* For convenience and to unify the notation we set in that case $d := D$. We may write $\varepsilon(D) = T/2 + (U/2)\sqrt{d}$, where $T \equiv U \pmod{2}$. If T and U are both even we argue as in the previous case (note that by reducing modulo 4 we see that $T/2$ has to be odd). Otherwise T and U are both odd and satisfy $T^2 - dU^2 = 4$. Mimicking the transformation of Legendre and Dirichlet described in [Section 2.1](#) (see also [Lemma 9](#)) one easily gets a set of equalities analogous to [\(8\)](#) and [\(9\)](#):

$$(14) \quad d_1U_1^2 - d_2U_2^2 = 4, \quad T = -2 + d_1U_1^2, \quad d = d_1d_2, \quad U = U_1U_2.$$

Therefore the integral principal ideal $(d_1U_1/2 + (U_2/2)\sqrt{d})$ (note that both d_1U_1

and U_2 are odd) is generated by an element of norm $d_1 > 0$. To see that this ideal is I it is enough to prove that $d_1 > 1$. Indeed if by contradiction $d_1 = 1$ then $(U_1/2)\sqrt{d_1} + (U_2/2)\sqrt{d_2}$ would be a unit of $\mathbb{Z}_{\mathbb{Q}(\sqrt{D})}$ the square of which equals $\varepsilon(D)$ contradicting the minimality of the fundamental unit.

(iii) Finally let us consider the case where $D = 4d$, $d \equiv 2 \pmod{4}$. As in the first case the fundamental unit may be written $\varepsilon(D) = T + U\sqrt{d}$. From the norm equation $T^2 - dU^2 = 1$ we deduce that T is odd and U is even; i.e., the transformation of Legendre and Dirichlet leads to (9). As in the first case one easily shows that $I = (d_1U_1 + U_2\sqrt{d})$.

However what we want to understand is how (the narrow classes of) the elements of $P_{\mathbb{Q}(\sqrt{D})} := \{ \text{fractional principal ideals of } \mathbb{Z}_{\mathbb{Q}(\sqrt{D})} \} \supset P_{\mathbb{Q}(\sqrt{D})}^+$ are represented in M . It turns out (see [Fouvry and Klüners 2010a, (6)], for instance) that one has a short exact sequence

$$1 \rightarrow F_\infty \rightarrow C_D \rightarrow Cl_D \rightarrow 1,$$

where F_∞ has order at most 2. It is straightforward from the definitions that $|F_\infty| = [P_{\mathbb{Q}(\sqrt{D})} : P_{\mathbb{Q}(\sqrt{D})}^+]$. Moreover one knows that $|F_\infty| = 2$ if and only if $\varepsilon(D)$ has norm 1 (see the discussion following [Fouvry and Klüners 2010a, (6)] and the references therein). Since we have assumed $D \in \text{Fund}^+$ we have $[P_{\mathbb{Q}(\sqrt{D})} : P_{\mathbb{Q}(\sqrt{D})}^+] = 2$ and the above discussion then implies that $P_{\mathbb{Q}(\sqrt{D})}$ has four representatives in M . We can even argue in a completely explicit way: $P_{\mathbb{Q}(\sqrt{D})}$ is the disjoint union of two left cosets with respect to the subgroup $P_{\mathbb{Q}(\sqrt{D})}^+$. We have exhibited two elements ((1) and $I = (a)$) in the coset $P_{\mathbb{Q}(\sqrt{D})}^+$. In the other coset obviously lies the ideal (\sqrt{d}) : the algebraic integer \sqrt{d} has norm $-d$ dividing D' . Using (a) and (\sqrt{d}) we easily deduce the construction of the fourth suitable ideal. Indeed in the decomposition of $(a\sqrt{d})$ as a product of prime ideals, the \mathfrak{p}_i 's are the only prime ideals that may appear. Reducing the exponent of each \mathfrak{p}_i appearing modulo 2 we get a principal ideal (recall that $\mathfrak{p}_j^2 = (p_j)$ for each j) the norm of which divides D' . Clearly this ideal is different from (1), (a) and (\sqrt{d}) . (We can deduce more: since both I and (\sqrt{d}) are elements of M and since d differs from D' by at most a factor 2 then either the norm \tilde{d} of $I = (a)$ divides d and therefore the norm of the “fourth” ideal is d/\tilde{d} or \tilde{d} is even and the norm of the fourth ideal is $4d/\tilde{d}$.)

In terms of the Legendre–Dirichlet transformation and besides (1) and $I = (a)$ the ideals (\sqrt{d}) and $(U_1\sqrt{d} + d_2U_2)$ (or $((U_1/2)\sqrt{d} + d_2U_2/2)$ in the case $d = D \equiv 1 \pmod{4}$) are representatives of $P_{\mathbb{Q}(\sqrt{D})}$ in M . Of these four integral principal ideals one has norm 1 and one has norm d . The norms of the other two are d_1 and d_2 (or $2d_1$ and $2d_2$ in the case where $D = 4d$, $d \equiv 3 \pmod{4}$, and the coordinate T of the fundamental unit $\varepsilon(D) = T + U\sqrt{d}$ is even) respectively. This concludes the proof of Proposition 1.

3.2. Remarks on Proposition 1 and its proof. Among the constraints defining the sets on the left-hand side of (4) one may object that there is some redundancy in imposing both the conditions $D \in \text{Fund}^+$ and $2^2 \parallel D$. However the norm of the fundamental unit is of course no longer automatically positive in the cases D odd or $8 \mid D$.

In view of the above proof of Proposition 1, we see that the integer $\Phi(D)$ can be given explicitly via the Legendre–Dirichlet transformation. Indeed we deduce from the proof the following explicit version of Proposition 1.

Proposition 5. *Let $D \in \text{Fund}^+$ and $d := D/(4, D)$. Let $d = d_1 d_2$ be the coprime factorization of d obtained by applying (8), (9) or (14) to the norm equation satisfied by the fundamental unit $\varepsilon(D)$. Then*

$$\Phi(D) = \begin{cases} \min(2d_1, 2d_2) & \text{if } D = 4d, d \equiv 3 \pmod{4}, T \equiv 0 \pmod{2}, \\ \min(d_1, d_2) & \text{otherwise,} \end{cases}$$

where in the first case $\varepsilon(D) = T + U\sqrt{d}$.

As a consequence one deduces $\Phi(D) < \sqrt{d}$ unless $D = 4d$, $d \equiv 3 \pmod{4}$ and the coordinate T of the fundamental unit $\varepsilon(D) = T + U\sqrt{d}$ is even. In the latter case one can only infer $\Phi(D) < \sqrt{D}$.

Example 6. Assuming $D \in \text{Fund}^+$ one might get the intuitive idea that among the four integral principal ideals of norm dividing D' , the ideal (\sqrt{d}) is the one with norm of maximal absolute value. Of course this is true if the norms of the four ideals in question divide d which is always the case unless $D = 4d$, $d \equiv 3 \pmod{4}$, and $\varepsilon(D) = T + U\sqrt{d}$ with T even. However this intuitive idea is not necessarily true in the latter situation. Consider the case $D = 12$. Thus $D' = 6$ and $d = 3$. If \mathcal{N} denotes the norm map relative to the extension $\mathbb{Q}(\sqrt{3})/\mathbb{Q}$, one easily checks that

$$\mathcal{N}(\sqrt{3}) = -3, \quad \mathcal{N}(1 + \sqrt{3}) = -2, \quad \mathcal{N}(3 + \sqrt{3}) = 6.$$

In the notation of the Legendre–Dirichlet transformation the maximum of the absolute values of the three norms above is $2d_1 = 6$. Moreover $\Phi(12) = 2$ and one notices as expected the identity among ideals:

$$(\sqrt{3}) \cdot (3 + \sqrt{3}) = (3) \cdot (1 + \sqrt{3}),$$

which is congruent to $(1 + \sqrt{3})$ modulo squares (i.e., modulo \mathcal{I}^2 in the notation of the proof of Proposition 1).

This example contains even more information. Not only does it show that d is not in general the maximum of the four divisors of D' among the norms of integral principal ideals, but also that at most one of the other three divisors is larger than d . Otherwise we would have $2d_1 \geq d$ and $2d_2 \geq d$, by virtue of Proposition 5. Since

$d = d_1 d_2 \equiv 3 \pmod{4}$ this implies $d = 3$. This corresponds to $D = 12$, in which case $d_2 = 1$, as shown above.

4. Proof of Theorem 2 when $2^2 \parallel D$

4.1. Notation. The letter p is reserved for prime numbers. The Möbius function is denoted by μ , the number of distinct prime divisors of the integer n is $\omega(n)$, the cardinality of the set of primes $p \leq x$ which are congruent to $a \pmod{q}$ is denoted by $\pi(x; q, a)$. The condition $n \sim N$ means that the variable n has to satisfy the inequalities $N < n \leq 2N$. As it shall not lead to confusion the symbol \sim will also be used in the usual sense: if f, g are two functions of the real variable x defined on a neighborhood of a on which g does not vanish, $f(x) \sim g(x)$ as $x \rightarrow a$ means that f/g approaches 1 as $x \rightarrow a$.

4.2. The basic splitting. Let D be a fundamental discriminant such that $2^2 \parallel D$. Hence $d := D/4$ is squarefree and congruent to $3 \pmod{4}$. In that particular case (1) simplifies into $\mathbb{Z}_K = \mathbb{Z} \oplus \mathbb{Z}\sqrt{d}$. As already mentioned both the facts that $D \in \text{Fund}^+$ and that D is divisible by some $p \equiv 3 \pmod{4}$ imply that there is no unit with norm -1 . Hence $T + U\sqrt{d}$ belongs to \cup_K if and only if $T^2 - dU^2 = 1$, i.e., (5) holds.

We construct a sequence of fundamental discriminants $D = 4d$ with a large $\varepsilon(D) = \varepsilon_d$ by starting from

$$d = pm,$$

where $p \equiv 3 \pmod{4}$ and $m \equiv 1 \pmod{4}$ is squarefree. We keep in mind that m is small compared to p , hence m is coprime with p .

For any squarefree integer m and any $x \geq 2$ let

$$(15) \quad \mathcal{D}_m(x) := \{pm; pm \sim x, p \geq 7, p \equiv 3 \pmod{4}\}.$$

Dirichlet's Theorem on primes in arithmetic progressions directly implies

$$(16) \quad \#\mathcal{D}_m(x) \sim \frac{x}{2m \log(x/m)},$$

as $x \rightarrow \infty$ uniformly for $m \leq \sqrt{x}$. We now introduce the following subset of $\mathcal{D}_m(x)$ consisting of elements pm with a small ε_{pm} : for $\delta = \delta(x) > 0$, we consider

$$\mathcal{D}_m(x, \delta) := \{pm; pm \in \mathcal{D}_m(x), \varepsilon_{pm} \leq (4pm)^{3-\delta}\}.$$

By counting solutions that may not be fundamental, we have the inequality

$$(17) \quad \#\mathcal{D}_m(x, \delta) \leq \#\{(p, T, U); T, U \geq 1, pm \in \mathcal{D}_m(x), T^2 - pmU^2 = 1, T + U\sqrt{pm} \leq (4pm)^{3-\delta}\}.$$

We now want to apply [Lemma 4](#) with the choice $d = pm$, where m satisfies

$$(18) \quad 2 \nmid m \text{ and } \mu^2(m) = 1.$$

Let $m_1 m_2 = m$ be a decomposition of m . For

$$(19) \quad \eta \in \{\pm 1, \pm 2\}.$$

we consider the equation

$$E(m_1, m_2, \eta) \quad m_1 U_1^2 - pm_2 U_2^2 = \eta.$$

By [\(17\)](#) and using the values of T appearing in [\(8\)](#) & [\(9\)](#) we get the inequality

$$(20) \quad \begin{aligned} & \# \mathcal{D}_m(x, \delta) \\ & \leq \sum_{m_1 m_2 = m} \sum_{\eta = \pm 1} \# \{ (p, U_1, U_2); pm \in \mathcal{D}_m(x), m_1 U_1^2 - pm_2 U_2^2 = \eta, \\ & \quad \quad \quad -1 + 2m_1 U_1^2 + 2U_1 U_2 \sqrt{pm} \leq (4pm)^{3-\delta} \} \\ & \quad + \sum_{m_1 m_2 = m} \sum_{\eta = \pm 2} \# \{ (p, U_1, U_2); pm \in \mathcal{D}_m(x), m_1 U_1^2 - pm_2 U_2^2 = \eta, \\ & \quad \quad \quad -1 + m_1 U_1^2 + U_1 U_2 \sqrt{pm} \leq (4pm)^{3-\delta} \}. \end{aligned}$$

We now want to simplify the above inequality by studying the orders of magnitude of the variables U_1 and U_2 . The equation $E(m_1, m_2, \eta)$ and the assumption $p \geq 7$ in [\(15\)](#) imply that we have

$$\frac{1}{2} m_1 U_1^2 \leq pm_2 U_2^2 \leq 2m_1 U_1^2.$$

Multiplying these inequalities by m_1 and using the assumption $pm \sim x$ we obtain:

$$(21) \quad \frac{1}{2} m_1 U_1 x^{-1/2} \leq U_2 \leq 2m_1 U_1 x^{-1/2}.$$

From the inequalities defining the sets in the right-hand side of [\(20\)](#) we deduce

$$U_1 U_2 \sqrt{pm} \leq 64 (pm)^{3-\delta},$$

which implies in turn

$$(22) \quad U_1 U_2 \leq 400 x^{(5/2)-\delta}.$$

Also note that [\(21\)](#) and [\(22\)](#) imply the inequalities

$$(23) \quad U_2 \leq 30 m_1^{1/2} x^{1-(\delta/2)} \quad \text{and} \quad U_1 \leq 2m_1^{-1} x^{1/2} U_2.$$

Now we drop the condition that p is prime in [\(20\)](#). We deduce the inequality

$$(24) \quad \# \mathcal{D}_m(x, \delta) \leq \sum_{m_1 m_2 = m} \sum_{\eta = \pm 1, \pm 2} F(m_1, m_2, \eta).$$

Here $F(m_1, m_2, \eta)$ is the number of solutions to the congruence

$$(25) \quad m_1 U_1^2 \equiv \eta \pmod{m_2 U_2^2},$$

where (U_1, U_2) is subject to (23). Let $\rho_{\eta, m_1}(t)$ be the number of solutions to the congruence

$$m_1 u^2 - \eta \equiv 0 \pmod{t},$$

where η satisfies (19) and m_1 is odd. The study of the function $\rho_{\eta, m_1}(t)$ is classically reduced to the study of $\rho_{\eta, m_1}(p^k)$. Since we always have $(m_1, \eta) = 1$ in every case one has $\rho_{\eta, m_1}(2^k) \leq 4$ and $\rho_{\eta, m_1}(p^k) \leq 2$ ($k \geq 1$ and $p \geq 3$). This leads to the inequality

$$(26) \quad \rho_{\eta, m_1}(t) \leq 2 \cdot 2^{\omega(t)} \quad \text{for any } t \geq 1.$$

Looking back at (24) we split the interval of variation of U_1 into intervals of length $m_2 U_2^2$ together with perhaps an incomplete one. Inserting (26) and noting that η can take four distinct values we obtain the inequality

$$(27) \quad \begin{aligned} \#\mathcal{D}_m(x, \delta) &\leq 8 \sum_{m_1 m_2 = m} \sum_{U_2 \leq 30 m_1^{1/2} x^{1-\delta/2}} 2^{\omega(m_2 U_2)} \left(2 \frac{x^{1/2}}{m_1 m_2 U_2} + 1 \right) \\ &\leq 16 \frac{x^{1/2}}{m} \Sigma_1 + 8 \Sigma_2, \end{aligned}$$

with

$$\Sigma_1 := \sum_{m_1 m_2 = m} 2^{\omega(m_2)} \sum_{U_2 \leq 30 m_1^{1/2} x^{1-\delta/2}} \frac{2^{\omega(U_2)}}{U_2},$$

and

$$\Sigma_2 := \sum_{m_1 m_2 = m} 2^{\omega(m_2)} \sum_{U_2 \leq 30 m_1^{1/2} x^{1-\delta/2}} 2^{\omega(U_2)}.$$

It remains to apply techniques for summing multiplicative functions (recall that m is squarefree). We obtain

$$\Sigma_1 \ll \sum_{m_1 m_2 = m} 2^{\omega(m_2)} \log^2 x \ll 3^{\omega(m)} \log^2 x,$$

and

$$\begin{aligned} \Sigma_2 &\ll x^{1-\delta/2} \log x \sum_{m_1 m_2 = m} 2^{\omega(m_2)} m_1^{1/2} = (x^{1-\delta/2} \log x) m^{1/2} \sum_{m_2 | m} \frac{2^{\omega(m_2)}}{\sqrt{m_2}}, \\ &\ll_{\kappa} \kappa^{\omega(m)} m^{1/2} x^{1-\delta/2} \log x, \end{aligned}$$

for any fixed $\kappa > 1$. Putting everything together via (27) we have finally proved:

Proposition 7. *For every $\kappa > 1$ there exists $c(\kappa) > 0$ such that the inequality*

$$(28) \quad \#\mathcal{D}_m(x, \delta) \leq c(\kappa) (3^{\omega(m)} m^{-1} x^{1/2} \log^2 x + \kappa^{\omega(m)} m^{1/2} x^{1-\delta/2} \log x),$$

holds for every $x \geq 2$, for every odd squarefree $m \leq \sqrt{x}$ and for every $\delta = \delta(x) \geq 0$.

Applying this proposition with $m = 1$ one instantly deduces:

Corollary 8. *Let $t \mapsto \psi(t)$ be any increasing function of the variable $t \geq 1$, approaching infinity as $t \rightarrow \infty$. Then as x tends to infinity one has*

$$\#\{p \leq x; p \equiv 3 \pmod{4}, \varepsilon(4p) \leq p^3/(\psi(p) \log^4 p)\} = o(x/(\log x)).$$

In other words, this corollary tells us that for almost every $p \equiv 3 \pmod{4}$, the regulator $R(4p)$ of the field $\mathbb{Q}(\sqrt{4p})$ is greater than $(3 - \varepsilon) \log p$ (where $\varepsilon > 0$ is arbitrary). However [Corollary 8](#) is not new: it is slightly weaker by a power of $\log p$ than [\[Golubeva 1993, Corollary 5\]](#) which was obtained by Golubeva via the theory of continued fractions. In the statement of [Corollary 8](#) it is possible to make the power of $\log p$ decrease. It requires a better control of the function $\rho_{\eta,1}(p)$ which can be achieved by appealing to oscillations of some Legendre symbol. One essentially deduces the fact that this ρ -function has mean value 1 as long as $\eta \neq 1$. Actually, requiring that $T + U\sqrt{p}$ be a fundamental solution to [PE\(p\)](#) is enough to reduce to this case.

4.3. End of the proof of the lower bound in [Theorem 2](#). Let γ be a constant satisfying $0 \leq \gamma \leq \frac{1}{2}$. Let

$$(29) \quad \mathcal{D}^\gamma(x) := \bigcup_m \mathcal{D}_m(x),$$

where the union is taken over the integers m satisfying

$$(30) \quad 1 \leq m \leq x^\gamma, \quad \mu^2(m) = 1 \text{ and } m \equiv 1 \pmod{4}.$$

Since the sets $\mathcal{D}_m(x)$ are pairwise disjoint (when m runs over the set of integers satisfying [\(30\)](#)) we have the equality

$$\#\mathcal{D}^\gamma(x) = \sum_{m \text{ satisfies (30)}} \#\mathcal{D}_m(x).$$

Inserting [\(16\)](#), summing over m , and using the formula

$$\sum_{\substack{m \leq y \\ m \equiv 1 \pmod{4}}} \mu^2(m) \sim \frac{2}{\pi^2} y \quad (y \rightarrow \infty),$$

we deduce that for every $\gamma_0 > 0$ and for $x \rightarrow \infty$, one has

$$(31) \quad \#\mathcal{D}^\gamma(x) \sim -\frac{\log(1-\gamma)}{\pi^2} x,$$

uniformly for $\gamma_0 \leq \gamma \leq \frac{1}{2}$.

Now we apply [Proposition 7](#) and [\(31\)](#) with the choice $\gamma = \delta/4$. Consider

$$\mathcal{E}(x, \delta) := \bigcup_m (\mathcal{D}_m(x) \setminus \mathcal{D}_m(x, \delta)),$$

where the union is taken over the indices m satisfying [\(30\)](#). Every element $pm \in \mathcal{E}(x, \delta)$ is squarefree and congruent to $3 \pmod{4}$. Hence $D := 4pm$ is a fundamental discriminant and it satisfies the inequality $\varepsilon_d = \varepsilon(D) \geq D^{3-\delta}$ and the inequality $D \leq 8x$. Furthermore, because the sets $\mathcal{D}_m(x)$ appearing in the definition of $\mathcal{E}(x, \delta)$ are pairwise disjoint, one trivially has:

$$\mathcal{E}(x, \delta) = \mathcal{D}^\gamma(x) \setminus \left(\bigcup_m \mathcal{D}_m(x, \delta) \right),$$

where the union appearing on the right-hand side is a disjoint union. Therefore,

$$\begin{aligned} \#\mathcal{E}(x, \delta) &\geq -\frac{(1-o(1)) \log(1-\delta/4)}{\pi^2} \cdot x - O\left(x^{1-\delta/2} \log x \sum_{m \leq x^{\delta/4}} (3/2)^{\omega(m)} m^{1/2}\right) \\ &\geq -\frac{(1-o(1)) \log(1-\delta/4)}{\pi^2} \cdot x. \end{aligned}$$

This gives the first case of [Theorem 2](#). Indeed the argument so far has only involved splittings of positive fundamental discriminants D of type $D/4 = d_1d_2$ with $d_1 = m_1$ and $d_2 = pm_2$ (see [\(20\)](#)). Since $m = m_1m_2$ is a divisor of D of very small size (see [\(30\)](#)) the condition on $\Phi(D)$ on the left-hand side of [\(4\)](#) is automatically fulfilled for the particular D 's under consideration in view of [Proposition 1](#) or rather its explicit version [Proposition 5](#).

4.4. Comments on the proof of [Proposition 7](#). To obtain the inequality [\(24\)](#) we have dropped the condition p prime. By sieve techniques it is possible to handle this constraint. The upshot of this would consist in saving a power of $\log x$ in the first term of the right-hand side of [\(28\)](#). This improvement does not seem to affect the exponent $3 - \delta$ in the statement of [\(4\)](#).

A more promising way to improve this exponent is to apply a better treatment of the congruence [\(25\)](#) in small intervals. After a classical expansion via Fourier techniques we would be led to bound the general exponential sum

$$\sum_{m_1 m_2 = m} \sum_{U_2 \leq 30 m^{1/2} x^{1-\delta/2}} \sum_{\substack{U_1 \pmod{m_2 U_2^2} \\ m_1 U_1^2 \equiv \eta \pmod{m_2 U_2^2}}} \sum_{\substack{1 \leq |h| \leq \\ m_1 m_2 x^{-1/2} U_2^{1+\varepsilon}}} \exp\left(2\pi i h \frac{U_1}{m_2 U_2^2}\right).$$

5. Proof of the remaining cases

5.1. The case D divisible by 8. In that case set $d := D/4$. We still have

$$K := \mathbb{Q}(\sqrt{D}) = \mathbb{Q}(\sqrt{d}) \quad \text{and} \quad \mathbb{Z}_K = \mathbb{Z} \oplus \mathbb{Z}\sqrt{d}.$$

However, contrary to the case $2^2 \parallel D$, the fact that $D \in \text{Fund}^+$ is no longer guaranteed which means that the negative Pell equation $T^2 - dU^2 = -1$ may be solvable.

Since we are only dealing with discriminants in Fund^+ we are led to modify (15):

$$\mathcal{D}_m(x) := \{2pm ; 2pm \sim x, p \equiv 3 \pmod{4}\},$$

hence $d \in \mathcal{D}_m(x)$ implies $D \in \text{Fund}^+$. We shall consider these sets for m squarefree and congruent to 1 mod 4. The proof of [Theorem 2](#) is essentially the same in this case.

5.2. The case D odd. In that case D is squarefree and congruent to 1 mod 4, write $d := D$. Then $K = \mathbb{Q}(\sqrt{D})$ and $\mathbb{Z}_K = \{(a + b\sqrt{d})/2 ; a, b \in \mathbb{Z}, a \equiv b \pmod{2}\}$. Hence the study of the fundamental unit of K is reduced to the question of finding the smallest nontrivial solution to the equation

$$T^2 - dU^2 = \pm 4.$$

As above we can ensure the equation $T^2 - dU^2 = -4$ has no integral solution (thus $D \in \text{Fund}^+$) by imposing d to be divisible by some $p \equiv 3 \pmod{4}$. To deal with the equation $T^2 - dU^2 = 4$ we appeal to a variant of [Lemma 4](#) that we state without proof.

Lemma 9. *Let d and U be positive integers such that $2 \nmid d$. Define $\mathcal{A}(d, U)$ as in [Lemma 4](#). Set*

$$\tilde{\mathcal{A}}(d, U) := \{T \geq 1 ; T^2 - dU^2 = 4\},$$

$$\tilde{\mathcal{B}}(d, U) := \{(d_1, d_2, U_1, U_2) \in \mathbb{N}_{\geq 1}^4 ; U_1U_2 = U, d_1d_2 = d, d_1U_1^2 - d_2U_2^2 = 4\}.$$

Then we have

$$(32) \quad \tilde{\mathcal{A}}(d, U) = 2 \cdot \mathcal{A}(d, U/2) \quad \text{if } 2 \mid U,$$

$$(33) \quad \#\tilde{\mathcal{A}}(d, U) = \#\tilde{\mathcal{B}}(d, U) \in \{0, 1\} \quad \text{if } 2 \nmid U.$$

We are led to modify (15) in the following way:

$$\mathcal{D}_m(x) := \{pm ; pm \sim x, p \equiv 3 \pmod{4}\}.$$

We shall consider these sets for m squarefree and congruent to 3 mod 4. Thanks to [Lemma 9](#) the proof of [Theorem 2](#) in this last case is once more essentially the same.

The proof of [Theorem 2](#) is now complete.

Remark 10. The ‘‘algebraic interpretation’’ provided by [Proposition 1](#) and translated by the condition on the function Φ in (4) relies heavily on the assumption that for the D ’s under consideration the fundamental unit $\varepsilon(D)$ has norm 1 (see [Section 3](#)). If $\varepsilon(D)$ has norm -1 then $d = D/(4, D)$ is the norm of the algebraic integer $\varepsilon(D)\sqrt{d}$. Gauss’s theorem on the 2-rank of C_D still applies and shows that the only two

divisors of D' among norms of integral principal ideals generated by totally positive elements are 1 and d . Recall that if $\varepsilon(D)$ has norm -1 then the groups $P_{\mathbb{Q}(\sqrt{D})}$ and $P_{\mathbb{Q}(\sqrt{D})}^+$ coincide.

Remark 11. One may wonder why neglecting the contribution of positive fundamental discriminants with fundamental unit of negative norm has such little influence on the difficulty of showing the lower bound (4). This comes from the fact that the set of fundamental discriminants with fundamental unit of norm -1 is negligible. More precisely the number of *special discriminants* (i.e., positive fundamental discriminants only divisible by 2 or primes congruent to 1 modulo 4) up to X is asymptotic to $c \cdot X (\log X)^{-1/2}$, where c is an absolute constant (see [Fouvry and Klüners 2010a, Section 1] and the references therein).

6. Proof of Theorem 3

Our starting point is the following well known *class number formula* (see [Cohen 1993, Proposition 5.6.9, p. 262], for instance)

$$(34) \quad h(D) = \frac{L(1, \chi_D)}{2 R(D)} \sqrt{D},$$

where D is a positive fundamental discriminant and $L(s, \chi_D)$ is the Dirichlet L -function associated to the Kronecker symbol $\chi_D = (D/\cdot)$

$$L(s, \chi_D) := \sum_{n=1}^{\infty} \chi_D(n) n^{-s} \quad (\Re s > 1).$$

Recall the classical upper bound

$$(35) \quad L(1, \chi) \ll \log(q+1),$$

which holds for any nonprincipal Dirichlet character χ modulo $q > 1$. To prove Theorem 3 we have to study the sum

$$\Sigma(x) := \sum_{\substack{D \leq x \\ 2^2 \parallel D}} h(D),$$

and prove the inequality

$$(36) \quad \Sigma(x) \leq \left(\frac{8}{21\pi^2} C_0 - \delta \right) \frac{x^{3/2}}{\log x},$$

for sufficiently large x . Define the two positive valued functions

$$\kappa(D) := R(D)/\log D, \quad \xi(D) := L(1, \chi_D) \sqrt{D}$$

and

$$(37) \quad \tilde{\Sigma}(x) := \sum_{\substack{D \leq x \\ 2^2 \parallel D}} \frac{\xi(D)}{\kappa(D)}.$$

By (34) and by partial summation, we see that (36) can be deduced from the inequality

$$(38) \quad \tilde{\Sigma}(x) \leq 2 \left(\frac{8}{21\pi^2} C_0 - 2\delta \right) x^{3/2},$$

for sufficiently large x .

Let γ , η and η' be small positive numbers and let $\mathcal{E}(x)$ be the set of indices over which the summation (37) is performed. We write any $D \in \mathcal{E}(x)$ under the form $D = 4d$. Hence $D \in \mathcal{E}(x)$ if and only if $d \in \mathcal{F}(x)$ where

$$(39) \quad \mathcal{F}(x) := \{d; \mu^2(d) = 1, d \equiv 3 \pmod{4} \text{ and } d \leq x/4\}.$$

We now consider two disjoint subsets of $\mathcal{F}(x)$ defined as follows:

$$\mathcal{F}_1(x) := \{d \in \mathcal{F}(x); \kappa(4d) \leq \frac{7}{4} - \eta'\},$$

$$\mathcal{F}_2(x) := \{d \in \mathcal{F}(x); \kappa(4d) > \frac{7}{4} - \eta',$$

$$d = pm, pm \sim x/8, p \equiv 3 \pmod{4}, m \equiv 1 \pmod{4}, m \leq x^\gamma \}.$$

We denote by $\mathcal{G}(x)$ the complement of $\mathcal{F}_1(x) \cup \mathcal{F}_2(x)$ in $\mathcal{F}(x)$. Let us then use the condition $\kappa(4d) \leq (7/4) + \eta$ to split further $\mathcal{F}_2(x)$ into the partition $\mathcal{F}_2^+(x) \cup \mathcal{F}_2^-(x)$ where:

$$\mathcal{F}_2^-(x) := \{d \in \mathcal{F}_2(x); \kappa(4d) \leq \frac{7}{4} + \eta\},$$

$$\mathcal{F}_2^+(x) := \{d \in \mathcal{F}_2(x); \kappa(4d) > \frac{7}{4} + \eta\}.$$

Using this decomposition we split the sum $\tilde{\Sigma}(x)$ accordingly:

$$(40) \quad \tilde{\Sigma}(x) = \sigma_{\mathcal{F}_1}(x) + \sigma_{\mathcal{F}_2^-}(x) + \sigma_{\mathcal{F}_2^+}(x) + \sigma_{\mathcal{G}}(x),$$

where each term on the right-hand side is a sum over the corresponding obvious subset of $\mathcal{F}(x)$ we have just defined. To upper bound $\sigma_{\mathcal{F}_1}(x)$ we use [Fouvry and Jouve 2013, Theorem 1] which asserts that for any $\varepsilon > 0$ one has

$$\#\{(D, \varepsilon_D); D \text{ nonsquare}, 2 \leq D \leq x, \varepsilon_D \leq D^{(1/2)+\alpha}\} = O_\varepsilon(x^{(\alpha/3)+(7/12)+\varepsilon}),$$

uniformly for $\alpha \geq 0$ and $x \geq 2$. Together with (5) the above formula (with the choices $\varepsilon = \eta'/12$ and $\alpha = 5/4 - \eta'$) implies:

$$\#\mathcal{F}_1(x) \ll_\gamma x^{1-\eta'/4}.$$

Hence by the inequality $\kappa(4d) \geq 1/2$ (see (2)) and by (35), we deduce the inequality

$$(41) \quad \sigma_{\mathcal{F}_1}(x) \ll x^{(3/2)-(\eta'/4)} \log x.$$

By (28), we also know that

$$\#\mathcal{F}_2^-(x) \ll x^{1-\eta/10},$$

with the choice $\gamma = \eta/10$. As for the proof of (41) we deduce that

$$(42) \quad \sigma_{\mathcal{F}_2^-}(x) \ll x^{(3/2)-(\eta/10)} \log x.$$

Next note the following easy inequality, consequence of the definitions of the sets $\mathcal{F}_2^+(x)$, $\mathcal{F}_2^-(x)$ and $\mathcal{G}(x)$:

$$\sigma_{\mathcal{F}_2^+}(x) + \sigma_{\mathcal{G}}(x) \leq \frac{1}{7/4+\eta} \sum_{d \in \mathcal{F}_2^+(x)} \xi(4d) + \frac{1}{7/4-\eta'} \sum_{d \in \mathcal{G}(x)} \xi(4d).$$

Set

$$(43) \quad \tilde{\mathcal{F}}_2(x) := \{d; d = pm, \mu^2(d) = 1, pm \sim x/8, m \leq x^\gamma, \\ p \equiv 3 \pmod{4}, m \equiv 1 \pmod{4}\}.$$

From the inclusion $\mathcal{F}_1(x) \cup \mathcal{F}_2(x) \supset \tilde{\mathcal{F}}_2(x)$ one deduces

$$\sum_{d \in \mathcal{F}_1(x) \cup \mathcal{F}_2(x)} \xi(4d) \geq \sum_{d \in \tilde{\mathcal{F}}_2(x)} \xi(4d).$$

Combining the last two inequalities with the following obvious facts:

$$\sum_{d \in \mathcal{G}(x)} \xi(4d) = \sum_{d \in \mathcal{F}(x)} \xi(4d) - \sum_{d \in \mathcal{F}_1(x) \cup \mathcal{F}_2(x)} \xi(4d), \quad \sum_{d \in \mathcal{F}_2^+(x)} \xi(4d) \leq \sum_{d \in \tilde{\mathcal{F}}_2(x)} \xi(4d)$$

we deduce the inequality

$$(44) \quad \sigma_{\mathcal{F}_2^+}(x) + \sigma_{\mathcal{G}}(x) \leq \frac{1}{7/4-\eta'} \sum_{d \in \mathcal{F}(x)} \xi(4d) - \frac{\eta + \eta'}{(7/4+\eta)(7/4-\eta')} \sum_{d \in \tilde{\mathcal{F}}_2(x)} \xi(4d).$$

It remains to evaluate each of the two sums in (44). To that end we state and prove two lemmas, the most classical of which is the following:

Lemma 12. *As $y \rightarrow \infty$, one has*

$$\sum_{\substack{d \leq y \\ d \equiv 3 \pmod{4}}} \mu^2(d) L(1, \chi_{4d}) \sqrt{d} \sim \frac{4C_0}{3\pi^2} y^{3/2}.$$

Proof. Let $A_1(y)$ be the sum we want to evaluate. By the properties of the Kronecker symbol we have the equality

$$A_1(y) = \sum_{\substack{d \leq y \\ d \equiv 3 \pmod{4}}} \mu^2(d) \sqrt{d} \sum_{n \geq 1, 2 \nmid n} \frac{(d/n)}{n},$$

that now involves a Jacobi symbol. By the fact that the sum over n varying in any interval of length $4d$ of the symbols $(4d/n)$ equals zero, we can express using partial summation the above infinite series as a finite sum with a small enough error term:

$$\sum_{\substack{n \geq 1 \\ 2 \nmid n}} \frac{(d/n)}{n} = \sum_{\substack{1 \leq n \leq y^2 \\ 2 \nmid n}} \frac{(d/n)}{n} + O(y^{-1}),$$

uniformly for $d \leq y$. Inserting this equality in the definition of $A_1(y)$ and splitting the sum according to whether n is a square or not, we get the equality

$$(45) \quad A_1(y) = MT_1(y) + Err_1(y) + O(y^{1/2}).$$

In the above equality the sum $MT_1(y)$ which will appear as the main term is the following

$$(46) \quad MT_1(y) := \sum_{\substack{d \leq y \\ d \equiv 3 \pmod{4}}} \sum_{\substack{1 \leq t \leq y \\ (t, 2d)=1}} \mu^2(d) \frac{\sqrt{d}}{t^2},$$

whereas $Err_1(y)$ is defined by

$$(47) \quad Err_1(y) := \sum_{\substack{d \leq y \\ d \equiv 3 \pmod{4}}} \sum_{\substack{1 \leq n \leq y^2 \\ 2 \nmid n, n \neq \square}} \mu^2(d) \frac{\sqrt{d}}{n} \left(\frac{d}{n} \right).$$

We first consider $Err_1(y)$. We want to prove that it behaves as an error term. More precisely we want to show:

$$(48) \quad Err_1(y) = o(y^{3/2}) \quad (y \rightarrow \infty).$$

To do so we split the double sum in (47) in $O(\log^2 y)$ subsums $Err_1(D, N)$ where the sizes of d and n are controlled:

$$(49) \quad Err_1(D, N) := \sum_{\substack{d \sim D \\ d \equiv 3 \pmod{4}}} \sum_{\substack{n \sim N \\ 2 \nmid n, n \neq \square}} \mu^2(d) \frac{\sqrt{d}}{n} \left(\frac{d}{n} \right),$$

with $D \leq y/2$ and $N \leq y^2/2$. Our purpose is to prove that in all these cases we have

$$(50) \quad \text{Err}_1(D, N) = O(y^{3/2} \log^{-3} y).$$

Of course the trivial bound is $\text{Err}_1(D, N) \ll D^{3/2}$. Hence (50) is proved for any (D, N) such that $D \leq y \log^{-2} y$. Thus for the rest of the proof we suppose that

$$(51) \quad D > y \log^{-2} y.$$

The sum $\text{Err}_1(y)$ is a particular case of a double sum of Jacobi or Kronecker symbols, which is nowadays quite common in analytic number theory. For instance we have (see [Fouvry and Klüners 2010b, Proposition 10]):

Lemma 13. *For every $A > 0$, there exists $c(A) > 0$, such that for every bounded complex sequences (α_m) and (β_n) and for every M and N satisfying the inequalities $M, N \geq \max(2, \log^A(MN))$, one has the inequality*

$$\left| \sum_{m \sim M} \sum_{n \sim N} \alpha_m \beta_n \mu^2(2m) \mu^2(2n) \left(\frac{m}{n}\right) \right| \leq c(A) \|\alpha\|_\infty \|\beta\|_\infty MN \log^{-A/2}(MN).$$

However in the definition (49) of $\text{Err}_1(D, N)$ the variable n is not squarefree. To circumvent this difficulty we decompose $n = \ell^2 n'$ where now n' is squarefree and we consider two cases. Either $\ell \leq N^{1/4}$ and we apply Lemma 13 where the parameters M and N respectively have the values D and $N\ell^{-2}$. Or $\ell > N^{1/4}$ and we apply the trivial bound. Summing over ℓ , choosing a big enough A in Lemma 13 and appealing to (51), we finally deduce the inequality

$$\text{Err}_1(D, N) \ll D^{3/2} \log^{-10}(DN) \ll y^{3/2} \log^{-3} y,$$

which holds uniformly for $N \geq \log^{100} y$. Hence we have also proved (50) in that case. Combining with (51) it remains to prove (50) in the case where D is large and N is small:

$$(52) \quad D \geq y \log^{-2} y \text{ and } N \leq \log^{100} y.$$

We shall now benefit from the oscillations of the character $d \mapsto (d/n)$ when d runs over squarefree integers $d \equiv 3 \pmod{4}$ as follows. Our argument uses the following rather standard lemma which can be found in [Prachar 1958, formula (1)].

Lemma 14. *The following equality*

$$\sum_{\substack{n \leq x \\ n \equiv \ell \pmod{k}}} \mu^2(n) = \frac{6}{\pi^2} \prod_{p|k} \left(1 - \frac{1}{p^2}\right)^{-1} \frac{x}{k} + O(x^{1/2}),$$

holds uniformly for $x \geq 2$, $k \geq 1$ and ℓ coprime with k .

Applying [Lemma 14](#) to each of the reduced classes ℓ modulo $4n$ such that $\ell \equiv 3 \pmod{4}$ and summing over these ℓ , we obtain the equality

$$(53) \quad \sum_{\substack{d \leq t \\ d \equiv 3 \pmod{4}}} \mu^2(d) \left(\frac{d}{n} \right) = O(nt^{1/2}),$$

uniformly for $t \geq 1$ and for $n \geq 1$ odd and nonsquare.

Integrating by part and summing over $n \sim N$, we easily see that [\(50\)](#) also holds under the condition [\(52\)](#). As a conclusion the proof of [\(48\)](#) is now complete.

We now deal with $\text{MT}_1(y)$. From [Lemma 14](#) we deduce that for any given $A > 0$ the formula

$$\sum_{\substack{d \leq z \\ (d,t)=1 \\ d \equiv 3 \pmod{4}}} \mu^2(d) \sim \frac{2}{\pi^2} \prod_{p|t} \left(1 + \frac{1}{p} \right)^{-1} z,$$

holds as $z \rightarrow \infty$ uniformly for t odd satisfying $t \leq z^A$. By a partial summation and by comparison with an integral we have

$$\sum_{\substack{d \leq z \\ (d,t)=1 \\ d \equiv 3 \pmod{4}}} \mu^2(d) \sqrt{d} \sim \frac{4}{3\pi^2} \cdot \prod_{p|t} \left(1 + \frac{1}{p} \right)^{-1} z^{3/2}.$$

Inserting this formula in the definition [\(46\)](#) and summing over every odd $t \leq y$ yields:

$$\text{MT}_1(y) \sim_{y \rightarrow \infty} \frac{4}{3\pi^2} y^{3/2} \sum_{2 \nmid t} t^{-2} \prod_{p|t} \left(1 + \frac{1}{p} \right)^{-1}.$$

The infinite series above admits an expansion as an Euler product

$$(54) \quad \text{MT}_1(y) \sim_{y \rightarrow \infty} \frac{4}{3\pi^2} \prod_{p \geq 3} \left(1 + \frac{p}{(p+1)^2(p-1)} \right) y^{3/2} = \frac{4C_0}{3\pi^2} y^{3/2}.$$

Putting together [\(45\)](#), [\(48\)](#) and [\(54\)](#) we complete the proof of [Lemma 12](#). \square

The second lemma we need in order to evaluate the sums in [\(44\)](#) is the following.

Lemma 15. *Let $0 < \gamma < \frac{1}{2}$ and, for any $x \geq 0$, let $\widetilde{\mathcal{F}}_2(x)$ be defined as in [\(43\)](#). Then there exists $c(\gamma) > 0$, such that as $x \rightarrow \infty$ one has*

$$\sum_{d \in \widetilde{\mathcal{F}}_2(x)} L(1, \chi_{4d}) \sqrt{d} \sim c(\gamma) x^{3/2}.$$

The asymptotics is uniform for $\gamma_0 \leq \gamma \leq \frac{1}{2} - \gamma_0$, whenever $0 < \gamma_0 < \frac{1}{4}$.

Proof. The proof is very similar to the proof of [Lemma 12](#). The main difference being that [\(53\)](#) is replaced by the following consequence of the classical Siegel–Walfisz theorem

$$(55) \quad \sum_{\substack{m \equiv 1 \pmod{4} \\ m \leq x^\gamma}} \mu^2(d) \sum_{\substack{p \equiv 3 \pmod{4} \\ p \sim D/m}} \left(\frac{pm}{n} \right) = O_A(\sqrt{n} D \log^{-A} D),$$

which holds for any constant $A > 0$. Note that the upper bound contained in [\(55\)](#) is only interesting if $n \leq \log^{2A} D$. This exactly fits the constraint we have on the summation over n (see [\(52\)](#)).

The corresponding main term will have the shape (see [\(46\)](#))

$$\sum_{\substack{m \leq x^\gamma \\ m \equiv 1 \pmod{4}}} \mu^2(m) \sqrt{m} \sum_{\substack{p \sim x/(8m) \\ p \equiv 3 \pmod{4}}} \sqrt{p} \sum_{t, (t, 2pm)=1} \frac{1}{t^2}.$$

Inverting summations we first sum over p (where we use a variant of [\(16\)](#)), then over m and finally over t , as in the proof of [\(54\)](#). We note in passing that $c(\gamma)$ could be given an explicit value. \square

6.1. End of the proof of [Theorem 3](#) and remarks. Putting together the definition [\(40\)](#), [Lemma 12](#), [Lemma 15](#) (with the choice $\gamma = \eta/10$), and the equalities [\(41\)](#), [\(42\)](#) and [\(44\)](#), we get the inequality

$$\begin{aligned} & \tilde{\Sigma}(x) \\ & \leq \left\{ \frac{4C_0}{3\pi^2(7/4 - \eta')} (1 + o(1)) - \frac{(\eta + \eta')c(\eta/10)}{(7/4 + \eta)(7/4 - \eta')} (1 - o_\eta(1)) \right\} x^{3/2} + o_{\eta, \eta'}(x^{3/2}). \end{aligned}$$

Now fix $\eta = \frac{1}{10}$. Then by fixing a very small $\eta' > 0$ the above upper bound can be written

$$\tilde{\Sigma}(x) \leq K_0 x^{3/2},$$

for sufficiently large x and for some fixed K_0 satisfying the inequality

$$K_0 < \frac{16 C_0}{21\pi^2}.$$

This proves [\(38\)](#) hence [\(36\)](#) and completes the proof of [Theorem 3](#).

We now discuss the influence of the different results about the size of $\varepsilon(D)$ we have used on the sum we have studied. If our only input is the trivial lower bound $\varepsilon(D) \geq 2\sqrt{D}$ (see [\(2\)](#)), we cannot get anything better than

$$(56) \quad \sum_{\substack{D \leq x \\ 2^2 \parallel D}} h(D) \leq \left(\frac{4 C_0}{3\pi^2} + \delta \right) \frac{x^{3/2}}{\log x},$$

for every positive δ and every sufficiently large x .

Using [Fouvry and Jouve 2013, Theorem 1] has enabled us to improve the multiplicative coefficient in the above upper bound by the factor 3.5. Finally the purpose of our Proposition 7 has been to improve the inequality (56) by some factor slightly larger than 3.5.

6.2. A consequence of Corollary 8. A natural question is to ask for some upper bound on average for the class number $h(D)$ when D is essentially prime. So we consider the sum

$$S(x) := \sum_{\substack{p \leq x \\ p \equiv 3 \pmod{4}}} h(4p).$$

By techniques very similar to those presented in the beginning of Section 6 and the trivial bound $\varepsilon(4p) \geq 2\sqrt{p}$, we can prove that we have the trivial asymptotic inequality

$$S(x) \leq \left(\frac{1}{2} + o(1)\right) \frac{x^{3/2}}{\log^2 x}.$$

When appealing instead to (12), we improve this upper bound by a factor 2. Finally, Corollary 8 improves by a factor 6 the trivial asymptotic inequality. More precisely we get the following result the proof of which easily follows from Corollary 8 and is left to the reader.

Corollary 16. *As $x \rightarrow \infty$, one has the inequality*

$$S(x) \leq \left(\frac{1}{12} + o(1)\right) \frac{x^{3/2}}{\log^2 x}.$$

Acknowledgements

The authors thank E. P. Golubeva, J. Klüners, F. Lemmermeyer and D. Milovic for discussions and comments concerning a previous version of this work.

References

- [Cohen 1993] H. Cohen, *A course in computational algebraic number theory*, Graduate Texts in Mathematics **138**, Springer, Berlin, 1993. [MR 94i:11105](#) [Zbl 0786.11071](#)
- [Cremona and Odoni 1989] J. E. Cremona and R. W. K. Odoni, “Some density results for negative Pell equations: An application of graph theory”, *J. London Math. Soc.* (2) **39**:1 (1989), 16–28. [MR 90b:11019](#) [Zbl 0678.10015](#)
- [Dirichlet 1834] P. G. Lejeune-Dirichlet, “Einige neue Sätze über unbestimmte Gleichungen”, pp. 649–604 [i.e., 664], 1834. Reprinted as pp. 219–236 in his *Werke*, vol. I, G. Reimer, Berlin, 1889; available at <http://archive.org/stream/abhandlungenderk1834deut#page/n734>. [JFM 21.0016.01](#)
- [Dirichlet 1856] P. G. Lejeune-Dirichlet, “Sur une propriété des formes quadratiques à déterminant positif”, *J. Math. Pure Appl.* (2) **1** (1856), 76–79. Reprinted as pp. 191–194 in his *Werke*, vol. II, G.

- Reimer, Berlin, 1897; available at http://portail.mathdoc.fr/JMPA/PDF/JMPA_1856_2_1_A7_0.pdf. [JFM 28.0014.01](#)
- [Fouvry and Jouve 2012] É. Fouvry and F. Jouve, “Fundamental solutions to Pell equation with prescribed size”, *Proc. Steklov Inst. Math.* **276**:1 (2012), 40–50.
- [Fouvry and Jouve 2013] É. Fouvry and F. Jouve, “Size of regulators and consecutive square-free numbers”, *Math. Z.* **273**:3–4 (2013), 869–882. [MR 3030681](#)
- [Fouvry and Klüners 2010a] É. Fouvry and J. Klüners, “On the negative Pell equation”, *Ann. of Math.* (2) **172**:3 (2010), 2035–2104. [MR 2011h:11122](#) [Zbl 1230.11136](#)
- [Fouvry and Klüners 2010b] É. Fouvry and J. Klüners, “The parity of the period of the continued fraction of \sqrt{d} ”, *Proc. Lond. Math. Soc.* (3) **101**:2 (2010), 337–391. [MR 2011g:11213](#) [Zbl 1244.11092](#)
- [Fröhlich and Taylor 1993] A. Fröhlich and M. J. Taylor, *Algebraic number theory*, Cambridge Studies in Advanced Mathematics **27**, Cambridge University Press, 1993. [MR 94d:11078](#) [Zbl 0744.11001](#)
- [Golubeva 1987] E. P. Golubeva, “On the lengths of the periods of a continued fraction expansion of quadratic irrationalities and on the class numbers of real quadratic fields”, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **160**:Anal. Teor. Chisel i Teor. Funktsii. 8 (1987), 72–81. In Russian; translated in *J. Soviet Math.* **52**:3 (1990), 3049–3056. [MR 88j:11055](#) [Zbl 0900.11015](#)
- [Golubeva 1993] E. P. Golubeva, “Class numbers of real quadratic fields of discriminant $4p$ ”, *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **204**:Anal. Teor. Chisel i Teor. Funktsii. 11 (1993), 11–36. In Russian; translated in *J. Math. Sc.* **79**:5 (1996), 1277–1292. [MR 94f:11115](#) [Zbl 0814.11023](#)
- [Golubeva 2002] E. P. Golubeva, “On the Pell equation”, *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **286**:Anal. Teor. Chisel i Teor. Funkts. 18 (2002), 36–39. In Russian; translated in *J. Math. Sc.* **12**:6 (2004), 3600–3602. [MR 2003j:11027](#) [Zbl 1077.11021](#)
- [Halter-Koch 1989] F. Halter-Koch, “Reell-quadratische Zahlkörper mit großer Grundeinheit”, *Abh. Math. Sem. Univ. Hamburg* **59** (1989), 171–181. [MR 91h:11116](#) [Zbl 0718.11054](#)
- [Hooley 1984] C. Hooley, “On the Pellian equation and the class number of indefinite binary quadratic forms”, *J. Reine Angew. Math.* **353** (1984), 98–131. [MR 86d:11032](#) [Zbl 0539.10019](#)
- [Legendre 1830] A.-M. Legendre, *Théorie des nombres, I*, 3rd ed., Didot, Paris, 1830.
- [Lemmermeyer 2003] F. Lemmermeyer, “Higher descent on Pell conics, I: From Legendre to Selmer”, preprint, 2003. [arXiv math/0311309v1](#)
- [Pall 1969] G. Pall, “Discriminantal divisors of binary quadratic forms”, *J. Number Theory* **1** (1969), 525–533. [MR 40 #1335](#) [Zbl 0186.08602](#)
- [Perron 1913] O. Perron, *Die Lehre von den Kettenbrüchen*, Teubner, Berlin, 1913. [JFM 43.0283.04](#)
- [Prachar 1958] K. Prachar, “Über die kleinste quadratfreie Zahl einer arithmetischen Reihe”, *Monatsh. Math.* **62** (1958), 173–176. [MR 19,1160g](#) [Zbl 0083.03704](#)
- [Reiter 1985] C. Reiter, “Effective lower bounds on large fundamental units of real quadratic fields”, *Osaka J. Math.* **22**:4 (1985), 755–765. [MR 87h:11107](#) [Zbl 0586.12005](#)
- [Sarnak 1985] P. C. Sarnak, “Class numbers of indefinite binary quadratic forms, II”, *J. Number Theory* **21**:3 (1985), 333–346. [MR 87h:11027](#) [Zbl 0571.10022](#)
- [Yamamoto 1971] Y. Yamamoto, “Real quadratic number fields with large fundamental units”, *Osaka J. Math.* **8** (1971), 261–270. [MR 45 #5107](#) [Zbl 0243.12001](#)
- [Zagier 1981] D. B. Zagier, *Zetafunktionen und quadratische Körper: Eine Einführung in die höhere Zahlentheorie*, Springer, Berlin, 1981. [MR 82m:10002](#) [Zbl 0459.10001](#)

Received November 23, 2011. Revised November 21, 2012.

ÉTIENNE FOUVRY
LABORATOIRE DE MATHÉMATIQUE
CAMPUS D'ORSAY, UNIVERSITÉ DE PARIS-SUD
BATIMENT 425 UMR 8628
91405 ORSAY CEDEX
FRANCE
etienne.fouvry@math.u-psud.fr

FLORENT JOUVE
LABORATOIRE DE MATHÉMATIQUE
CAMPUS D'ORSAY, UNIVERSITÉ DE PARIS-SUD
BATIMENT 425 UMR 8628
91405 ORSAY CEDEX
FRANCE
florent.jouve@math.u-psud.fr

ON THE ISENTROPIC COMPRESSIBLE EULER EQUATION WITH ADIABATIC INDEX $\gamma = 1$

DONG LI, CHANGXING MIAO AND XIAOYI ZHANG

We consider the isentropic compressible Euler equations with polytropic gamma law $P(\rho) = \rho^\gamma$ in dimensions $d \leq 3$. We address the borderline case when adiabatic index $\gamma = 1$ and establish local theory in the Sobolev space $C_t^0 L_x^p \cap C_t^0 \dot{H}_x^k$ for $d < p \leq 4$. This covers a class of physical solutions which can decay to vacuum at spatial infinity and are not compact perturbations of steady states. We construct a blowup scenario where initially the fluid is quiet in a neighborhood of the origin but is supersonic near the spatial infinity. For this special class of noncompact initial data, we prove the formation of singularities in finite time.

1. Introduction and main results

We consider the Cauchy problem for the d -dimensional, $d \leq 3$, isentropic compressible Euler equation

$$(1-1) \quad \begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0, \\ \rho(\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v}) + \nabla P = 0, \\ (\rho, \mathbf{v})(0, x) = (\rho_0, \mathbf{v}_0)(x). \end{cases} \quad (t, x) \in \mathbb{R}^+ \times \mathbb{R}^d.$$

Here, $\rho = \rho(t, x)$ is a scalar function representing density, $\mathbf{v} = \mathbf{v}(t, x)$ is a vector-valued function representing velocity. P is the pressure, satisfying the polytropic gamma law

$$P(\rho) = A\rho^\gamma, \quad \gamma \geq 1,$$

where $A > 0$ is a constant and γ is so-called adiabatic index. In this paper, we will mainly consider the borderline case $\gamma = 1$. For simplicity we shall set $A = 1$.

There is an extensive one-dimensional theory on the singularity formation of solutions to the compressible Euler equation and related equations (see [John 1974; Klainerman and Majda 1980; Lax 1964; Liu 1979]). The proofs are usually

Li is supported in part by an NSERC Discovery grant and by the NSF under agreement DMS-1128155. Miao is supported by the NSF of China (11171033 and 11231006). Zhang is supported by an Alfred P. Sloan fellowship.

MSC2010: primary 35Q35; secondary 76N10.

Keywords: compressible Euler equation, blowup solutions.

based on method of characteristics which is not robust enough to treat dimensions $d \geq 2$ (see, however, [Chae and Ha 2009] for a blowup result in 3D using method of characteristics). Sideris [1985] considered the following three-dimensional compressible Euler system:

$$(1-2) \quad \begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0, \\ \rho(\partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v}) + \nabla P = 0, \\ \partial_t S + \mathbf{v} \cdot \nabla S = 0, \\ (\rho, \mathbf{v}, S)(0, x) = (\rho_0, \mathbf{v}_0, S_0)(x). \end{cases}$$

Here $S = S(t, x)$ denotes the specific entropy and the pressure law is given by

$$(1-3) \quad P(\rho, S) = A\rho^\gamma e^S, \quad A > 0, \quad \gamma > 1.$$

If we set $S(t, x) \equiv \bar{S} = \text{const}$, the system (3-1) reduces to (1-1) (hence the name “isentropic”). The following set of initial data was considered in [Sideris 1985], where $R > 0$ is fixed:

$$(1-4) \quad \begin{aligned} \rho_0(x) &> 0 && \text{for all } x, \\ \rho_0(x) &= \bar{\rho}, \quad \mathbf{v}_0(x) = 0, \quad S_0(x) = \bar{S} && \text{if } |x| \geq R. \end{aligned}$$

Such initial data can be viewed as compact perturbations of the steady state $(\rho, \mathbf{v}, S) \equiv (\bar{\rho}, 0, \bar{S})$. By a change of variable $c = \text{const} \cdot \rho^{(\gamma-1)/2}$, one can rewrite (3-1) as a symmetric positive hyperbolic system in terms of (c, \mathbf{v}, S) . For initial data of the form (1-4), local wellposedness of (c, \mathbf{v}, S) in $C_t^0 H_x^{(S/2)^+}$ then follows easily; see [Kato 1975]. The speed of sound σ is defined by

$$(1-5) \quad \sigma = \left(\frac{\partial P(\rho, S)}{\partial \rho} \right)^{1/2} \Big|_{(\rho, S) = (\bar{\rho}, \bar{S})} = (A\gamma \bar{\rho}^{\gamma-1} e^{\bar{S}})^{1/2}.$$

A result of [Sideris 1985], roughly speaking, is that for a set of initial data (1-4) which is supersonic in a neighborhood of the origin, the corresponding C^1 solution to (1-2)–(1-3) must have finite lifespan. This was extended to the two-dimensional case by Rammaha [1989]. There are also more precise results on the estimate of lifespan of blowup solutions which are small perturbations of steady states. For the 3D compressible Euler equation (1-1) with irrotational (i.e., $\nabla \times \mathbf{v} = 0$) initial data $(\rho_0, \mathbf{v}_0) = (\epsilon \tilde{\rho}_0 + \bar{\rho}, \epsilon \tilde{\mathbf{v}}_0)$, where $\tilde{\rho}_0 \in \mathcal{S}(\mathbb{R}^3)$, $\tilde{\mathbf{v}}_0 \in \mathcal{S}(\mathbb{R}^3)^3$ ($\mathcal{S}(\mathbb{R}^3)$ is the usual Schwartz space), Sideris [1991] proved that the lifespan of the classical solution $T_\epsilon > \exp(C/\epsilon)$. For the upper bound it follows from [Sideris 1985] that $T_\epsilon < \exp(C/\epsilon^2)$ under some mild conditions on the initial data. For initial data which is spherically symmetric and is smooth compact ϵ -perturbation of the constant state, Godin [2005] obtained by using a suitable approximation solution the precise

asymptotic of the lifespan T_ϵ as

$$\lim_{\epsilon \rightarrow 0} \epsilon \log T_\epsilon = T^*,$$

where T^* is a constant. These results rely crucially on the observation that after some simple manipulations, the compressible Euler equation in rescaled variables is given by a vectorial nonlinear wave equation with pure quadratic nonlinearities. This fact together with the positivity of fundamental solutions of the wave operator were also exploited in [Sideris 1985; Rammaha 1989] to establish a different set of blowup results which are analogs of corresponding results on nonlinear wave equations.

In this paper we will be concerned with the d -dimensional isentropic compressible Euler system (1-1) with adiabatic index $\gamma = 1$. This is the borderline case, since previous results in the literature are mainly for the case $\gamma > 1$. We discuss first the local theory. In the case $\gamma > 1$, all the results mentioned before essentially deal with initial data which contain no vacuum states and are compact perturbations of steady states, cf. (1-4). Local wellposedness to (1-1) in $C_t^0 H_x^s$ for some regularity index $s > d/2 + 1$ then follow easily from [Kato 1975] after some change of variables transforming to a symmetric positive hyperbolic system. In principle one can essentially repeat this kind of analysis in the case $\gamma = 1$ and obtain local wellposedness for initial data which are compact perturbations of steady states. However we shall not discuss this simple case and will focus instead on the more interesting case where the initial data can be essentially *noncompact*. A useful example is where the initial density $\rho_0(x)$ decays as $(1 + |x|^2)^{-\beta}$ for some large exponent β as $|x| \rightarrow \infty$; in other words, we allow the density to decay to vacuum at spatial infinity. As it turns out, even the local theory for such initial data requires a bit of work, since the standard H_x^k spaces which fit so well with the usual symmetric hyperbolic systems are not suitable for closing the estimates due to problems at low frequencies. Instead, we will establish the local existence in $L^p(\mathbb{R}^d) \cap \dot{H}^k(\mathbb{R}^d)$:

Theorem 1.1 (local existence). *Let the dimension be $d \leq 3$. Let $k \geq 10d$ be a large integer and take p such that $d < p \leq 4$. Assume the initial data satisfy*

$$(1-6) \quad \begin{aligned} \rho_0 &> 0, & \rho_0^{-1} \nabla \rho_0 &\in L^p(\mathbb{R}^d) \cap \dot{H}^{k-1}(\mathbb{R}^d), \\ \rho_0 &\in L^p(\mathbb{R}^d) \cap \dot{H}^{k-1}(\mathbb{R}^d), & \mathbf{v}_0 &\in L^p(\mathbb{R}^d) \cap \dot{H}^k(\mathbb{R}^d). \end{aligned}$$

Then there exists $T > 0$ such that the Cauchy problem (1-1) admits a unique solution

$$\rho \in C([0, T]; L^p(\mathbb{R}^d) \cap \dot{H}^{k-1}(\mathbb{R}^d)), \quad \mathbf{v} \in C([0, T]; L^p(\mathbb{R}^d) \cap \dot{H}^k(\mathbb{R}^d)),$$

with $\rho > 0$. Moreover, $\rho^{-1} \nabla \rho \in C([0, T], L^p(\mathbb{R}^d) \cap \dot{H}^{k-1}(\mathbb{R}^d))$. If in addition $\rho_0 \in L^1(\mathbb{R}^d)$, we have mass conservation:

$$\int \rho(t, x) dx = \int \rho_0(x) dx.$$

Remark 1.2. In [Theorem 1.1](#), the restriction $p > d$ comes from the physical assumptions we put on the initial data ρ_0 . Since we allow ρ_0 to be essentially noncompact, in particular we can take $\rho_0 \sim (1 + |x|^2)^{-\beta}$ for $|x| \gg 1$. It is not difficult to check that in this case $\rho_0^{-1} \nabla \rho_0 \in L^p(\mathbb{R}^d)$ *only for* $p > d$. On the other hand the upper bound $p \leq 4$ comes from bounding certain L^2 -norm of products in the nonlinear estimates. For example (see also [\(2-2\)](#)), if we have two functions f, g with frequencies supported on the ball $|\xi| \leq 1$, that is, $f \sim P_{\leq 1} f, g \sim P_{\leq 1} g$ (here $P_{\leq 1}$ is the usual Littlewood–Paley projector, see [Section 2](#)), and we only know that f and g are bounded in L^p , then

$$\|fg\|_{L_x^2(\mathbb{R}^d)} \lesssim \|f\|_{L_x^{2p/(p-2)}} \|g\|_p \lesssim \|f\|_p \|g\|_p,$$

where in the last inequality we have to use the Bernstein inequality for which the constraint $2p/(p-2) \geq p$ or $p \leq 4$ is deduced. By the constraint $d < p \leq 4$ we deduce $d \leq 3$ and this is the main reason for the restriction of the dimension.

The next result is on the formation of singularities in finite time. We will show that the local solutions constructed in [Theorem 1.1](#) have finite life spans. As was mentioned before, the class of data that leads to blowups is a not a compact perturbation of the constant state. More precisely we have the following

Theorem 1.3 (blowup from spatial infinity). *Let ρ_0, \mathbf{v}_0 satisfy the conditions in [\(1-6\)](#) and $\rho_0 \in L^1(\mathbb{R}^d)$. For $d = 2, 3$, we also assume \mathbf{v}_0 is irrotational: $\text{curl}(\mathbf{v}_0) = 0$. Let $\rho_0(x) = 1, \mathbf{v}_0(x) = 0$, for all $|x| \leq 10$. Let $\phi(x)$ be a Schwartz function such that $\nabla^2 \phi(x)$ is positive definite on $|x| > 1$. Set*

$$(1-7) \quad N := \int \rho_0 \mathbf{v}_0 \cdot \nabla \phi(x) dx.$$

Then there exist a constant $C = C(\|\rho_0\|_1) > 0$ such that whenever $N > C$, the corresponding solution constructed in [Theorem 1.1](#) blows up at some time $T^ < 1$.*

Remark 1.4. The blowup constructed in [Theorem 1.3](#) is different from the usual case where the initial data is concentrated near the origin. In our scenario, the bulk of the initial data is concentrated near spatial infinity and the quantity N defined in [\(1-7\)](#) measures this concentration. The intuitive picture is that initially the fluid is quiet in an $O(1)$ -neighborhood of the origin but is supersonic near the spatial infinity. After an $O(1)$ -finite time the fluid develops singularities in the transient region away from the origin.

2. Preliminaries

We will often use the notation $X \lesssim Y$ whenever there exists some constant C such that $X \leq CY$. For any two operators A, B , we use the notation $[A, B] := AB - BA$ to denote the commutator.

We will also need to use the Littlewood–Paley theory. Let $\varphi(\xi)$ be a smooth bump function supported in the ball $|\xi| \leq 2$ and equal to one on the ball $|\xi| \leq 1$. For each dyadic number $N \in 2^{\mathbb{Z}}$ we define the Littlewood–Paley operators

$$\begin{aligned}\widehat{P_{\leq N} f}(\xi) &:= \varphi(\xi/N) \widehat{f}(\xi), & \widehat{P_{> N} f}(\xi) &:= [1 - \varphi(\xi/N)] \widehat{f}(\xi), \\ \widehat{P_N f}(\xi) &:= [\varphi(\xi/N) - \varphi(2\xi/N)] \widehat{f}(\xi).\end{aligned}$$

Similarly we can define $P_{< N}$, $P_{\geq N}$, and $P_{M < \cdot \leq N} := P_{\leq N} - P_{\leq M}$, whenever M and N are dyadic numbers. We will frequently write $f_{\leq N}$ for $P_{\leq N} f$ and similarly for the other operators. We recall the following standard Bernstein- and Sobolev-type inequalities:

Lemma 2.1. *For any $1 \leq p \leq q \leq \infty$ and $s > 0$, we have*

$$\begin{aligned}\|P_{\geq N} f\|_{L_x^p} &\lesssim N^{-s} \| |\nabla|^s P_{\geq N} f \|_{L_x^p}, \\ \| |\nabla|^s P_{\leq N} f \|_{L_x^p} &\lesssim N^s \| P_{\leq N} f \|_{L_x^p}, \\ \| |\nabla|^{\pm s} P_N f \|_{L_x^p} &\sim N^{\pm s} \| P_N f \|_{L_x^p}, \\ \| P_{\leq N} f \|_{L_x^q} &\lesssim N^{d/p-d/q} \| P_{\leq N} f \|_{L_x^p}, \\ \| P_N f \|_{L_x^q} &\lesssim N^{d/p-d/q} \| P_N f \|_{L_x^p}.\end{aligned}$$

We will use the following simple estimate frequently:

$$(2-1) \quad \|f\|_{\infty} \lesssim \|P_{\leq 1} f\|_p + \sum_{\substack{N>1 \\ N \in 2^{\mathbb{Z}}}} N^{d/2} \|P_N f\|_2 \lesssim \|P_{\leq 1} f\|_p + \|P_{> 1} f\|_{\dot{H}^{d/2+1}}.$$

We prove below some commutator estimates which will be useful in controlling the nonlinear terms. To simple notations we shall assume that the functions are scalar-valued. The extension to vector-valued functions is rather trivial. In order not to be burdened with notations, we will sometimes use the same notations for vector-valued functions as in the scalar-valued case. For example if $\mathbf{v} = (v_1, \dots, v_d)$ and $v_j \in L_x^2(\mathbb{R}^d)$, we shall simply write $\mathbf{v} \in L_x^2(\mathbb{R}^d)$ in place of $\mathbf{v} \in L_x^2(\mathbb{R}^d)^d$.

Lemma 2.2. *Let $f, g \in \mathcal{S}(\mathbb{R}^d)$. Let ∂ denote any partial derivative. Let $2 \leq p \leq 4$ and $k > d + 2$. Then*

$$\begin{aligned}\|[\partial^k, f \partial] g\|_2 &\lesssim \|f\|_{\dot{H}^k \cap L^p} \|g\|_{\dot{H}^k \cap L^p}, & \|[\partial^k, f] g\|_2 &\lesssim \|f\|_{\dot{H}^k \cap L^p} \|g\|_{\dot{H}^{k-1} \cap L^p}, \\ \|[\partial^{k-1}, f] \partial g\|_2 &\lesssim \|f\|_{\dot{H}^k \cap L^p} \|g\|_{\dot{H}^{k-1} \cap L^p}.\end{aligned}$$

Proof. We only prove the first one. By the chain rule and the triangle inequality, we have the bound

$$\|[\partial^k, f \partial] g\|_2 \lesssim \sum_{1 \leq j \leq k} \|\partial^j f \partial^{k+1-j} g\|_2.$$

In the case $1 \leq j \leq k/2$, we split g into low and high frequencies. For the low-frequency piece, we use the fact $p \leq 4$ and Bernstein to get

$$\begin{aligned} \|\partial^j f \partial^{k+1-j} P_{\leq 1} g\|_2 &\leq \|\partial^j f\|_{2p/(p-2)} \|\partial^{k+1-j} P_{\leq 1} g\|_p \\ &\lesssim (\|\partial^j P_{\leq 1} f\|_{2p/(p-2)} + \|\partial^j P_{> 1} f\|_{2p/(p-2)}) \|g\|_p \\ &\lesssim \|f\|_{\dot{H}^k \cap L^p} \|g\|_p. \end{aligned}$$

In the last estimate, we used a similar estimate as in (2-1). For the high-frequency piece, we use Sobolev embedding and Bernstein to get

$$(2-2) \quad \begin{aligned} \|\partial^j f \partial^{k+1-j} P_{> 1} g\|_2 &\leq \|\partial^j f\|_{\infty} \|\partial^{k+1-j} P_{> 1} g\|_2 \\ &\lesssim \|f\|_{\dot{H}^k \cap L^p} \|g\|_{\dot{H}^k}. \end{aligned}$$

Again we invoke (2-1) in the last step. In the case $k/2 < j \leq k$, we can instead split f into low and high frequencies. Then the estimate just follows by symmetry. \square

We need to use the following space which will be useful for proving some contraction estimates in Section 3. For any positive integer k , define

$$(2-3) \quad X_k = \{f, \|f\|_{X_k} := \|f\|_p + \|P_{> 1} f\|_{\dot{H}^k} < \infty\}.$$

It is not difficult to check that for $k > d/2$ the space X_k forms an algebra. This fact together with some useful commutator estimates and product estimates are stated in the next

Lemma 2.3. *Under the same conditions as in Lemma 2.2, we have:*

$$\begin{aligned} \|[\partial^{k-1}, f \partial] P_{> 1} g\|_2 &\lesssim \|f\|_{X_{k-1}} \|P_{> 1} g\|_{\dot{H}^{k-1}}, \\ \|\partial^{k-1} (f P_{\leq 1} g)\|_2 &\lesssim \|f\|_{X_{k-1}} \|g\|_p, \\ \|[\partial^{k-1}, f] P_{> 1} g\|_2 &\lesssim \|f\|_{X_{k-1}} \|g\|_{X_{k-2}}, \\ \|\partial^{k-1} (f \partial g)\|_2 &\lesssim \|f\|_{X_{k-1}} \|g\|_{X_k}, \\ \|\partial^{k-1} (fg)\|_2 &\lesssim \|f\|_{X_{k-1}} \|g\|_{X_{k-1}}. \end{aligned}$$

Proof. The proof proceeds in a similar way as in Lemma 2.2. One has to split both f and g into high- and low-frequency pieces and discuss several cases. We omit the details. \square

3. Proof of Theorem 1.1

To construct the local solution, we will use the usual Picard iteration but in a slightly nonstandard space and exploiting in an essential way the structure of the system. Due to the singular nature of the problem, we need both the hyperbolic formulation of the equation and the original formulation. The tricky part of the analysis is to define a good iteration scheme.

To this end, we define

$$f = \log \rho,$$

and rewrite the Cauchy problem (1-1) in terms of (f, \mathbf{v}) as

$$(3-1) \quad \begin{cases} \partial_t f + \mathbf{v} \cdot \nabla f + \nabla \cdot \mathbf{v} = 0, \\ \partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} + \nabla f = 0, \\ f(0, x) = \log \rho_0(x), \quad \mathbf{v}(0, x) = \mathbf{v}_0(x). \end{cases}$$

By bringing in the f -function, we have obtained the hyperbolic formulation (3-1) for the original system.

Remark 3.1. It is tempting to invoke the usual wellposedness theory in $H^k, k > d/2$ spaces and conclude that the system (3-1) admits local solutions in $C_t^0 H_x^k$. However there is a serious problem with this due to the physical assumptions we put on the initial data. Namely $f = \log \rho$ does not lie in L_x^2 in general. To see it one can consider the sample case $\rho(x) = \langle x \rangle^{-C}$ which immediately yields $f \sim \log \langle x \rangle \notin L_x^2$. In fact it is not difficult to check $f \notin \dot{H}_x^k(\mathbb{R}^d)$ for any $k \leq d/2$.

By Remark 3.1, we shall proceed differently from the usual fashion and work with an enlarged (and redundant) system of equations which includes both the hyperbolic formulation and the original system. The advantage is that with a little bit of work we can obtain regularity of all functions at one stroke.

We start with the zeroth iterate, defined as

$$\rho^{(0)}(t, x) = \rho_0(x), \quad \mathbf{v}^{(0)}(t, x) = \mathbf{v}_0(x), \quad f^{(0)}(t, x) = \log \rho_0(x).$$

For any integer $n \geq 0$, we inductively define $(\rho^{(n+1)}, \mathbf{v}^{(n+1)}, f^{(n+1)})$ as solutions to the linear system

$$(3-2) \quad \begin{cases} \partial_t \rho^{(n+1)} + \nabla \cdot (\rho^{(n+1)} \mathbf{v}^{(n)}) = 0, \\ \partial_t f^{(n+1)} + \mathbf{v}^{(n)} \cdot \nabla f^{(n+1)} + \nabla \cdot \mathbf{v}^{(n+1)} = 0, \\ \partial_t \mathbf{v}^{(n+1)} + (\mathbf{v}^{(n)} \cdot \nabla) \mathbf{v}^{(n+1)} + \nabla f^{(n+1)} = 0, \\ \rho^{(n+1)}(0, x) = \rho_0(x), \quad f^{(n+1)}(0, x) = \log \rho_0(x), \quad \mathbf{v}^{(n+1)}(0, x) = \mathbf{v}_0(x). \end{cases}$$

Remark 3.2. Strictly speaking, instead of $f^{(n+1)}$, we should be working with $\mathbf{g}^{(n+1)} = \nabla f^{(n+1)}$ and write the second equation in (3-2) as

$$\partial_t \mathbf{g}^{(n+1)} + \nabla(\mathbf{v}^{(n)} \cdot \mathbf{g}^{(n+1)}) + \nabla(\nabla \cdot \mathbf{v}^{(n+1)}) = 0,$$

with initial data $\mathbf{g}^{(n+1)} = \nabla \rho_0 / \rho_0$. Correspondingly in the third equation of (3-2) we should replace $\nabla f^{(n+1)}$ by $\mathbf{g}^{(n+1)}$. In this way we do not need to prove any regularity or solvability estimates of $f^{(n+1)}$ themselves in the iteration system.

We first show that the sequence of functions $\mathbf{v}^{(n)}$ are uniformly bounded in the space $L_t^\infty([0, T]; \dot{H}^k \cap L^p)$, $(\rho^{(n)}, \nabla f^{(n)})$ are uniformly bounded in the space $L_t^\infty([0, T]; \dot{H}^{k-1} \cap L^p)$ for some suitably small T .

Step 1: The L^p boundedness of the iterates $(\rho^{(n+1)}, \mathbf{v}^{(n+1)}, \nabla f^{(n+1)})$. Multiplying the first equation in (3-2) by $|\rho^{(n+1)}|^{p-2}\rho^{(n+1)}$ and integrating by parts, we get

$$\frac{1}{p} \frac{d}{dt} \|\rho^{(n+1)}(t)\|_p^p + \frac{p-1}{p} \int (\rho^{(n+1)})^p \nabla \cdot \mathbf{v}^{(n)} dx = 0.$$

Therefore

$$(3-3) \quad \begin{aligned} \frac{d}{dt} \|\rho^{(n+1)}(t)\|_p &\leq \|\nabla \cdot \mathbf{v}^{(n)}(t)\|_\infty \|\rho^{(n+1)}(t)\|_p \\ &\lesssim \|\mathbf{v}^{(n)}(t)\|_{\dot{H}^k \cap L^p} \|\rho^{(n+1)}(t)\|_p. \end{aligned}$$

Next we take the inner product with $|\mathbf{v}^{(n+1)}|^{p-2}\mathbf{v}^{(n+1)}$ on both sides of the third equation in (3-2). After integrating on \mathbb{R}^d , we get

$$\begin{aligned} \frac{1}{p} \frac{d}{dt} \|\mathbf{v}^{(n+1)}(t)\|_p^p - \frac{1}{p} \int \nabla \cdot \mathbf{v}^{(n)} |\mathbf{v}^{(n+1)}|^p dx \\ + \int |\mathbf{v}^{(n+1)}|^{p-2} \nabla f^{(n+1)} \cdot \mathbf{v}^{(n+1)} dx = 0. \end{aligned}$$

Hölder's inequality yields

$$(3-4) \quad \begin{aligned} \frac{d}{dt} \|\mathbf{v}^{(n+1)}(t)\|_p &\lesssim \|\nabla \cdot \mathbf{v}^{(n)}(t)\|_\infty \|\mathbf{v}^{(n+1)}(t)\|_p + \|\nabla f^{(n+1)}(t)\|_p \\ &\lesssim \|\mathbf{v}^{(n)}(t)\|_{\dot{H}^k \cap L^p} \|\mathbf{v}^{(n+1)}(t)\|_p + \|\nabla f^{(n+1)}(t)\|_p. \end{aligned}$$

To close the estimate, we need to estimate $\|\nabla f^{(n+1)}\|_p$. Differentiating the second equation in (3-2) once, we have the equation for $\partial_i f^{(n+1)}$:

$$\partial_i \partial_i f^{(n+1)} + \partial_i (\mathbf{v}^{(n)} \cdot \nabla f^{(n+1)}) + \nabla \cdot \partial_i \mathbf{v}^{(n+1)} = 0.$$

Multiplying both sides by $|\partial_i f^{(n+1)}|^{p-2} \partial_i f^{(n+1)}$ and integrating by parts, we get

$$\begin{aligned} \frac{1}{p} \frac{d}{dt} \|\partial_i f^{(n+1)}(t)\|_p^p + \int \partial_i \mathbf{v}^{(n)} \cdot \nabla f^{(n+1)} |\partial_i f^{(n+1)}|^{p-2} \partial_i f^{(n+1)} dx \\ - \frac{1}{p} \int \nabla \cdot \mathbf{v}^{(n)} |\partial_i f^{(n+1)}|^p dx + \int \nabla \cdot \partial_i \mathbf{v}^{(n+1)} |\partial_i f^{(n+1)}|^{p-2} \partial_i f^{(n+1)} dx = 0. \end{aligned}$$

By Hölder's inequality,

$$\begin{aligned} \frac{1}{p} \frac{d}{dt} \|\partial_i f^{(n+1)}(t)\|_p^p &\leq \|\partial_i \mathbf{v}^{(n)}(t)\|_\infty \|\nabla f^{(n+1)}(t)\|_p \|\partial_i f^{(n+1)}(t)\|_p^{p-1} \\ &\quad + \frac{1}{p} \|\nabla \cdot \mathbf{v}^{(n)}(t)\|_\infty \|\partial_i f^{(n+1)}(t)\|_p^p + \|\partial_i \nabla \cdot \mathbf{v}^{(n+1)}(t)\|_p \|\partial_i f^{(n+1)}(t)\|_p^{p-1}. \end{aligned}$$

Summing in $i = 1, \dots, d$ gives

$$\begin{aligned}
 (3-5) \quad & \frac{d}{dt} \|\nabla f^{(n+1)}(t)\|_p \\
 & \lesssim \sum_{i=1}^d \|\partial_i \mathbf{v}^{(n)}(t)\|_\infty \|\nabla f^{(n+1)}(t)\|_p + \sum_{j,i=1}^d \|\partial_j \partial_i \mathbf{v}^{(n+1)}(t)\|_p \\
 & \lesssim \|\mathbf{v}^{(n)}(t)\|_{\dot{H}^k \cap L^p} \|\nabla f^{(n+1)}(t)\|_p + \|\mathbf{v}^{(n+1)}(t)\|_{\dot{H}^k \cap L^p}.
 \end{aligned}$$

This ends the L^p -estimate. Next we turn to high-order energy estimates.

Step 2: \dot{H}^k -estimates. Let ∂^k denote a differential operator of order k , we compute

$$\begin{aligned}
 (3-6) \quad & \frac{d}{dt} \int |\partial^k \mathbf{v}^{(n+1)}|^2 dx \\
 & = 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot \partial^k \partial_t \mathbf{v}^{(n+1)} dx \\
 & = -2 \int \partial^k \mathbf{v}^{(n+1)} \cdot \partial^k [(\mathbf{v}^{(n)} \cdot \nabla) \mathbf{v}^{(n+1)}] dx - 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot \partial^k \nabla f^{(n+1)} dx \\
 & = -2 \int \partial^k \mathbf{v}^{(n+1)} \cdot [(\mathbf{v}^{(n)} \cdot \nabla) \partial^k \mathbf{v}^{(n+1)}] - 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot [\partial^k, (\mathbf{v}^{(n)} \cdot \nabla)] \mathbf{v}^{(n+1)} dx \\
 & \quad - 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot \partial^k \nabla f^{(n+1)} dx \\
 & = \int \nabla \cdot \mathbf{v}^{(n)} |\partial^k \mathbf{v}^{(n+1)}|^2 dx - 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot [\partial^k, (\mathbf{v}^{(n)} \cdot \nabla)] \mathbf{v}^{(n+1)} dx \\
 & \quad - 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot \nabla \partial^k f^{(n+1)} dx.
 \end{aligned}$$

Similarly for $f^{(n+1)}$ we have

$$\begin{aligned}
 (3-7) \quad & \frac{d}{dt} \int |\partial^k f^{(n+1)}|^2 dx = \int \nabla \cdot \mathbf{v}^{(n)} |\partial^k f^{(n+1)}|^2 \\
 & \quad - 2 \int \partial^k f^{(n+1)} [\partial^k, \mathbf{v}^{(n)}] \cdot \nabla f^{(n+1)} - 2 \int \partial^k f^{(n+1)} \partial^k \nabla \cdot \mathbf{v}^{(n+1)} dx.
 \end{aligned}$$

Adding (3-6), (3-7) together, we have

$$\begin{aligned}
 & \frac{d}{dt} (\|\partial^k \mathbf{v}^{(n+1)}(t)\|_2^2 + \|\partial^k f^{(n+1)}(t)\|_2^2) \\
 & = \int \nabla \cdot \mathbf{v}^{(n)} |\partial^k \mathbf{v}^{(n+1)}|^2 dx - 2 \int \partial^k \mathbf{v}^{(n+1)} \cdot [\partial^k, (\mathbf{v}^{(n)} \cdot \nabla)] \mathbf{v}^{(n+1)} dx \\
 & \quad + \int \nabla \cdot \mathbf{v}^{(n)} |\partial^k f^{(n+1)}|^2 dx - 2 \int \partial^k f^{(n+1)} [\partial^k, \mathbf{v}^{(n)}] \cdot \nabla f^{(n+1)} dx.
 \end{aligned}$$

By Hölder's inequality and [Lemma 2.2](#), we have

$$\begin{aligned} & \frac{d}{dt} (\|\partial^k \mathbf{v}^{(n+1)}(t)\|_2^2 + \|\partial^k f^{(n+1)}(t)\|_2^2) \\ & \lesssim \|\nabla \cdot \mathbf{v}^{(n)}\|_\infty (\|\partial^k \mathbf{v}^{(n+1)}\|_2^2 + \|\partial^k f^{(n+1)}\|_2^2) \\ & \quad + \|\mathbf{v}^{(n)}\|_{\dot{H}^k \cap L^p} (\|\partial^k \mathbf{v}^{(n+1)}\|_2 \|\mathbf{v}^{(n+1)}\|_{\dot{H}^k \cap L^p} + \|\partial^k f^{(n+1)}\|_2 \|\nabla f^{(n+1)}\|_{\dot{H}^{k-1} \cap L^p}). \end{aligned}$$

Since $\|\nabla \cdot \mathbf{v}^{(n)}\|_\infty \lesssim \|\mathbf{v}^{(n)}\|_{\dot{H}^k \cap L^p}$, we then obtain

$$(3-8) \quad \begin{aligned} & \frac{d}{dt} (\|\mathbf{v}^{(n+1)}(t)\|_{\dot{H}^k}^2 + \|f^{(n+1)}(t)\|_{\dot{H}^k}^2) \\ & \lesssim \|\mathbf{v}^{(n)}(t)\|_{\dot{H}^k \cap L^p} (\|\mathbf{v}^{(n+1)}(t)\|_{\dot{H}^k \cap L^p}^2 + \|\nabla f^{(n+1)}(t)\|_{\dot{H}^{k-1} \cap L^p}^2). \end{aligned}$$

The estimates are now complete. However, to prove the contraction estimates, we still need the high-order energy estimate of $\rho^{(n+1)}$: the \dot{H}^{k-1} -norm. By using integration by parts, we compute

$$\begin{aligned} \frac{d}{dt} \int |\partial^{k-1} \rho^{(n+1)}|^2 dx &= - \int \partial^{k-1} \rho^{(n+1)} \nabla \partial^{k-1} \rho^{(n+1)} \cdot \mathbf{v}^{(n)} dx \\ & \quad - \int \partial^{k-1} \rho^{(n+1)} [\partial^{k-1}, \mathbf{v}^{(n)}] \cdot \nabla \rho^{(n+1)} dx \\ & \quad - \int \partial^{k-1} \rho^{(n+1)} \partial^{k-1} (\rho^{(n+1)} \nabla \cdot \mathbf{v}^{(n)}) dx. \end{aligned}$$

By Hölder and using again [Lemma 2.2](#), we obtain

$$(3-9) \quad \frac{d}{dt} \|\rho^{(n+1)}(t)\|_{\dot{H}^{k-1}} \lesssim \|\mathbf{v}^{(n)}(t)\|_{\dot{H}^k \cap L^p} \|\rho^{(n+1)}(t)\|_{\dot{H}^{k-1} \cap L^p}.$$

Set

$$M^{(n+1)}(t) := \|\rho^{(n+1)}(t)\|_{\dot{H}^{k-1} \cap L^p}^2 + \|\mathbf{v}^{(n+1)}(t)\|_{\dot{H}^k \cap L^p}^2 + \|\nabla f^{(n+1)}(t)\|_{\dot{H}^{k-1} \cap L^p}^2.$$

Collecting the estimates [\(3-3\)](#), [\(3-4\)](#), [\(3-5\)](#), [\(3-8\)](#) and [\(3-9\)](#), we have

$$\begin{cases} \frac{d}{dt} M^{(n+1)}(t) \leq C M^{(n+1)}(t) (1 + M^{(n)}(t)), \\ M^{(n+1)}(0) = \|\rho_0\|_{\dot{H}^{k-1} \cap L^p}^2 + \|\mathbf{v}_0\|_{\dot{H}^k \cap L^p}^2 + \|\nabla f_0\|_{\dot{H}^{k-1} \cap L^p}^2 := M_0. \end{cases}$$

Here the constant depends only on p, d . Applying Gronwall's inequality, we obtain

$$(3-10) \quad M^{(n+1)}(t) \leq M_0 \exp \left\{ C \int_0^t (1 + M^{(n)}(s)) ds \right\}.$$

It suffices to take T small enough such that

$$(3-11) \quad 8CT(1 + M_0) \leq \frac{1}{100}.$$

Then the sequence $M^{(n)}(t)$ are uniformly bounded as

$$(3-12) \quad \|M^{(n)}\|_{L_t^\infty([0, T])} \leq 2M_0.$$

Therefore, for the chosen T , the sequence $\{\rho^{(n)}, \nabla f^{(n)}\}$ are bounded in

$$L_t^\infty([0, T]; (\dot{H}^{k-1} \cap L^p)),$$

and $\{\mathbf{v}^{(n)}\}$ are bounded in $L_t^\infty([0, T]; (\dot{H}^k \cap L^p))$. In the next step, we shall show that they are Cauchy in an intermediate topology.

Step 3: Contraction estimates. It is easy to check that the differences $\rho^{(n+1)} - \rho^{(n)}$, $\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}$, and $f^{(n+1)} - f^{(n)}$ satisfy the system of equations

$$\begin{aligned} \partial_t(\rho^{(n+1)} - \rho^{(n)}) + \nabla \cdot ((\rho^{(n+1)} - \rho^{(n)})\mathbf{v}^{(n)}) + \nabla \cdot (\rho^{(n)}(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})) &= 0, \\ \partial_t(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) + (\mathbf{v}^{(n)} \cdot \nabla)(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \\ &\quad + [(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}) \cdot \nabla]\mathbf{v}^{(n)} + \nabla(f^{(n+1)} - f^{(n)}) = 0, \\ \partial_t(f^{(n+1)} - f^{(n)}) + \mathbf{v}^{(n)} \cdot (\nabla f^{(n+1)} - \nabla f^{(n)}) \\ &\quad + (\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}) \cdot \nabla f^{(n)} + \nabla \cdot (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) = 0. \end{aligned}$$

We shall prove that the sequence $\mathbf{v}^{(n)}$ is Cauchy in X_{k-1} and $(\rho^{(n)}, f^{(n)})$ is Cauchy in X_{k-2} . Here the space X_j is defined in (2-3). We first estimate the L^p norm as

$$\begin{aligned} \frac{d}{dt} \|\rho^{(n+1)} - \rho^{(n)}(t)\|_p^p &\lesssim \|\nabla \cdot \mathbf{v}^{(n)}\|_\infty \|\rho^{(n+1)} - \rho^{(n)}\|_p^p \\ &\quad + \|\nabla \rho^{(n)}\|_\infty \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_p \|\rho^{(n+1)} - \rho^{(n)}\|_p^{p-1} \\ &\quad + \|\rho^{(n)}\|_\infty \|\nabla \cdot (\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})\|_p \|\rho^{(n+1)} - \rho^{(n)}\|_p^{p-1}. \end{aligned}$$

Note that

$$\begin{aligned} \|\rho^{(n)}\|_\infty + \|\nabla \rho^{(n)}\|_\infty &\lesssim \|\rho^{(n)}\|_{X_{k-1}}, \\ \|\nabla \cdot \mathbf{v}^{(n-1)}\|_\infty &\lesssim \|\mathbf{v}^{(n-1)}\|_{X_k}, \\ \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_p &\lesssim \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}, \\ \|\nabla \cdot (\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})\|_p &\lesssim \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}. \end{aligned}$$

Therefore

$$(3-13) \quad \begin{aligned} \frac{d}{dt} \|\rho^{(n+1)} - \rho^{(n)}\|_p \\ \lesssim \|\mathbf{v}^{(n)}\|_{X_k} \|\rho^{(n+1)} - \rho^{(n)}\|_p + \|\rho^{(n)}\|_{X_{k-1}} \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}. \end{aligned}$$

Similarly we also have

$$\begin{aligned}
& \frac{d}{dt} \|\mathbf{v}^{(n+1)}(t) - \mathbf{v}^{(n)}(t)\|_p^p \\
& \lesssim \|\nabla \cdot \mathbf{v}^{(n)}(t)\|_\infty \|(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})(t)\|_p^p \\
& \quad + \|\nabla \mathbf{v}^{(n)}(t)\|_\infty \|(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})(t)\|_p \|(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})(t)\|_p^{p-1} \\
& \quad + \|(\nabla f^{(n+1)} - \nabla f^{(n)})(t)\|_p \|(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})(t)\|_p^{p-1}.
\end{aligned}$$

Using the fact $\|\nabla f^{(n+1)} - \nabla f^{(n)}\|_p \lesssim \|\nabla f^{(n+1)} - \nabla f^{(n)}\|_{X_{k-2}}$, we arrive at

$$\begin{aligned}
(3-14) \quad \frac{d}{dt} \|(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})(t)\|_p & \lesssim \|\mathbf{v}^{(n)}(t)\|_{X_k} \|(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})(t)\|_{X_{k-1}} \\
& \quad + \|(\nabla f^{(n+1)} - \nabla f^{(n)})(t)\|_{X_{k-2}}.
\end{aligned}$$

For the L^p estimate of $\nabla f^{(n+1)} - \nabla f^{(n)}$, we have

$$\begin{aligned}
& \frac{d}{dt} \|\partial f^{(n+1)}(t) - \partial f^{(n)}(t)\|_p^p \\
& \lesssim \|\nabla \mathbf{v}^{(n)}\|_\infty \|\nabla(f^{(n+1)} - f^{(n)})\|_p \|\partial f^{(n+1)} - \partial f^{(n)}\|_p^{p-1} \\
& \quad + \|\nabla \cdot \mathbf{v}^{(n)}\|_\infty \|\partial f^{(n+1)} - \partial f^{(n)}\|_p^p \\
& \quad + \|\nabla f^{(n)}\|_\infty \|(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})\|_p \|\partial f^{(n+1)} - \partial f^{(n)}\|_p^{p-1} \\
& \quad + \|\partial \nabla f^{(n)}\|_\infty \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_p \|\partial f^{(n+1)} - \partial f^{(n)}\|_p^{p-1} \\
& \quad + \|\partial \nabla \cdot (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_p \|\partial f^{(n+1)} - \partial f^{(n)}\|_p^{p-1}.
\end{aligned}$$

Or, simplifying a bit,

$$\begin{aligned}
(3-15) \quad \frac{d}{dt} \|(\nabla f^{(n+1)} - \nabla f^{(n)})(t)\|_p \\
& \lesssim \|\mathbf{v}^{(n)}(t)\|_{X_k} \|(\nabla f^{(n+1)} - \nabla f^{(n)})(t)\|_{X_{k-2}} \\
& \quad + \|\nabla f^{(n)}(t)\|_{X_{k-1}} \|(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)})(t)\|_{X_{k-1}} + \|(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})(t)\|_{X_{k-1}}.
\end{aligned}$$

We now turn to the \dot{H}^{k-1} estimates of the high-frequency part of the iterate differences. From direct computation, we have

$$(3-16) \quad \frac{d}{dt} \int |P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})|^2 dx = I_1 + I_2 + I_3,$$

where we have set

$$I_1 = -2 \int P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot P_{>1} \partial^{k-1} [(\mathbf{v}^{(n)} \cdot \nabla)(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})] dx,$$

$$I_2 = -2 \int P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot P_{>1} \partial^{k-1} [(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}) \cdot \nabla \mathbf{v}^{(n)}] dx,$$

$$I_3 = -2 \int P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot P_{>1} \partial^{k-1} \nabla (f^{(n+1)} - f^{(n)}) dx.$$

We can write

$$\begin{aligned}
I_1 &= -2 \int P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot P_{>1} \partial^{k-1}[(\mathbf{v}^{(n)} \cdot \nabla) P_{>1}^2(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})] dx \\
&\quad - 2 \int P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot P_{>1} \partial^{k-1}[(\mathbf{v}^{(n)} \cdot \nabla)(I - P_{>1}^2)(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})] dx \\
&= -2 \int P_{>1}^2 \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot (\mathbf{v}^{(n)} \cdot \nabla) \partial^{k-1} P_{>1}^2(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) dx \\
&\quad - 2 \int P_{>1}^2 \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot [\partial^{k-1}, (\mathbf{v}^{(n)} \cdot \nabla)] P_{>1}^2(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) dx \\
&\quad - 2 \int P_{>1}^2 \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot \partial^{k-1}[(\mathbf{v}^{(n)} \cdot \nabla)(I - P_{>1}^2)(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})] dx.
\end{aligned}$$

Integrating by parts and using Hölder's inequality together with [Lemma 2.3](#) (the first two), we obtain the estimate

$$\begin{aligned}
I_1 &\lesssim \|\nabla \cdot \mathbf{v}^{(n)}\|_\infty \|\partial^{k-1} P_{>1}^2(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2^2 \\
&\quad + \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \|[\partial^{k-1}, (\mathbf{v}^{(n)} \cdot \nabla)] P_{>1}^2(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \\
&\quad + \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \|\partial^{k-1}((\mathbf{v}^{(n)} \cdot \nabla)(I - P_{>1}^2)(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}))\|_2 \\
&\lesssim \|\mathbf{v}^{(n)}\|_{X_k} \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2^2 \\
&\quad + \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \|\mathbf{v}^{(n)}\|_{X_k} \|\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}\|_{X_{k-1}} \\
&\lesssim \|\mathbf{v}^{(n)}\|_{X_k} \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \|\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}\|_{X_{k-1}}.
\end{aligned}$$

For the next term, we use [Lemma 2.3](#) to write

$$\begin{aligned}
(3-17) \quad I_2 &\lesssim \|\partial^{k-1} P_{>1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \|P_{>1} \partial^{k-1}((\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}) \cdot \nabla) \mathbf{v}^{(n)}\|_2 \\
&\lesssim \|\mathbf{v}^{(n)}\|_{X_k} \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}.
\end{aligned}$$

Collecting the estimates above, we have

$$\begin{aligned}
(3-18) \quad \frac{d}{dt} &\|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2^2 \\
&\leq C \|\mathbf{v}^{(n)}\|_{X_k} \|P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2 \\
&\quad \times [\|\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}\|_{X_{k-1}} + \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}] \\
&\quad - 2 \int P_{>1} \partial^{k-1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) \cdot P_{>1} \partial^{k-1} \nabla(f^{(n+1)} - f^{(n)}) dx.
\end{aligned}$$

The estimate for $f^{(n+1)} - f^{(n)}$ follows similarly. We compute

$$\begin{aligned}
& \frac{d}{dt} \int |P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})|^2 dx \\
&= -2 \int P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)}) P_{>1} \partial^{k-1} [\mathbf{v}^{(n)} \cdot (\nabla f^{(n+1)} - \nabla f^{(n)})] dx \\
&\quad - 2 \int P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)}) P_{>1} \partial^{k-1} [(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}) \cdot \nabla f^{(n)}] dx \\
&\quad - 2 \int P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)}) \nabla \cdot P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) dx \\
&= -2 \int P_{>1}^2 \partial^{k-1} (f^{(n+1)} - f^{(n)}) \mathbf{v}^{(n)} \cdot \partial^{k-1} P_{>1}^2 (\nabla f^{(n+1)} - \nabla f^{(n)}) dx \\
&\quad - 2 \int P_{>1}^2 \partial^{k-1} (f^{(n+1)} - f^{(n)}) [\partial^{k-1}, \mathbf{v}^{(n)}] \cdot P_{>1}^2 (\nabla f^{(n+1)} - \nabla f^{(n)}) dx \\
&\quad - 2 \int P_{>1}^2 \partial^{k-1} (f^{(n+1)} - f^{(n)}) \partial^{k-1} (\mathbf{v}^{(n)}) \cdot (I - P_{>1}^2) (\nabla f^{(n+1)} - \nabla f^{(n)}) dx \\
&\quad - 2 \int P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)}) P_{>1} \partial^{k-1} [(\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}) \cdot \nabla f^{(n)}] dx \\
&\quad - 2 \int P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)}) \nabla \cdot P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) dx.
\end{aligned}$$

Applying [Lemma 2.3](#) (the last two estimates), we get

$$\begin{aligned}
(3-19) \quad & \frac{d}{dt} \|P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})\|_2^2 \\
& \leq C \|\mathbf{v}^{(n)}\|_{X_k} \|\nabla f^{(n+1)} - \nabla f^{(n)}\|_{X_{k-2}} \|P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})\|_2 \\
& \quad + C \|\nabla f^{(n)}\|_{X_{k-1}} \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}} \|P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})\|_2 \\
& \quad - 2 \int P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)}) \nabla \cdot P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}) dx.
\end{aligned}$$

Adding together [\(3-18\)](#) and [\(3-19\)](#), we have

$$\begin{aligned}
& \frac{d}{dt} \|P_{>1} \partial^{k-1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_2^2 + \frac{d}{dt} \|P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})\|_2^2 \\
& \leq C \|\mathbf{v}^{(n)}\|_{X_k} \|P_{>1} (\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_{\dot{H}^{k-1}} \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}} \\
& \quad + C \|\mathbf{v}^{(n)}\|_{X_k} \|\nabla f^{(n+1)} - \nabla f^{(n)}\|_{X_{k-2}} \|P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})\|_2 \\
& \quad + C \|\nabla f^{(n)}\|_{X_{k-1}} \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}} \|P_{>1} \partial^{k-1} (f^{(n+1)} - f^{(n)})\|_2.
\end{aligned}$$

Summing over all the partial derivatives and using Cauchy–Schwartz, we have

$$\begin{aligned}
 (3-20) \quad & \frac{d}{dt} (\|P_{>1}(\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)})\|_{\dot{H}^{k-1}}^2 + \|P_{>1}\nabla(f^{(n+1)} - f^{(n)})\|_{\dot{H}^{k-2}}^2) \\
 & \lesssim \|\mathbf{v}^{(n)}\|_{X_k} \cdot (\|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}^2 + \|\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}\|_{X_{k-1}}^2) \\
 & \quad + \|\mathbf{v}^{(n)}\|_{X_k} \cdot \|\nabla(f^{(n+1)} - f^{(n)})\|_{X_{k-2}}^2 \\
 & \quad + \|\nabla f^{(n)}\|_{X_{k-1}} \cdot (\|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}}^2 + \|\nabla(f^{(n+1)} - f^{(n)})\|_{X_{k-2}}^2).
 \end{aligned}$$

Similarly, we get the estimate for ρ as follows:

$$\begin{aligned}
 (3-21) \quad & \frac{d}{dt} \|P_{>1}(\rho^{(n+1)} - \rho^{(n)})\|_{\dot{H}^{k-2}} \lesssim \|\rho^{(n)}\|_{X_{k-1}} \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}} \\
 & \quad + \|\mathbf{v}^{(n)}\|_{X_k} \|\rho^{(n+1)} - \rho^{(n)}\|_{X_{k-2}}.
 \end{aligned}$$

Let

$$N^n(t) = \|\mathbf{v}^{(n+1)} - \mathbf{v}^{(n)}\|_{X_{k-1}}^2 + \|\nabla f^{(n+1)} - \nabla f^{(n)}\|_{X_{k-2}}^2 + \|\rho^{(n+1)} - \rho^{(n)}\|_{X_{k-2}}^2.$$

Collecting estimates (3-13)–(3-15), (3-20), (3-21), and integrating in t , we have

$$N^{(n+1)}(t) \leq C \int_0^t (1 + M^{(n)}(\tau)) N^{(n)}(\tau) d\tau + C \int_0^t (1 + M^{(n)}(\tau)) N^{(n+1)}(\tau) d\tau.$$

Using Gronwall's inequality we get

$$\|N^{(n+1)}\|_{L_t^\infty([0, T])} \leq CT \|M^{(n)} N^{(n)}\|_{L_t^\infty([0, T])} \exp\{CT(1 + \|M^{(n)}\|_{L_t^\infty([0, T])})\}.$$

From (3-12) and the choice of T (3-11), we have

$$\begin{aligned}
 (3-22) \quad & \|N^{(n+1)}\|_{L_t^\infty([0, T])} \leq 2CM_0T \|N^{(n)}\|_{L_t^\infty([0, T])} \exp\{2CT(1 + M_0)\} \\
 & \leq \frac{1}{2} \|N^{(n)}\|_{L_t^\infty([0, T])}.
 \end{aligned}$$

Step 4: Limiting system and regularity of solutions. The estimate (3-22) easily implies that

$$\begin{aligned}
 & \{\rho^{(n)}\}_{n=1}^\infty \text{ is Cauchy in } L_t^\infty([0, T], X_{k-2}), \\
 & \{\mathbf{v}^{(n)}\}_{n=1}^\infty \text{ is Cauchy in } L_t^\infty([0, T], X_{k-1}), \\
 & \{\nabla f^{(n)}\}_{n=1}^\infty \text{ is Cauchy in } L_t^\infty([0, T], X_{k-2}).
 \end{aligned}$$

From the condition $k \geq 10d$, and using the embedding $X_{k-2} \subset W^{[k/2], p}$, we know that all sequences $\{\rho^{(n)}, \mathbf{v}^{(n)}, \nabla f^{(n)}\}_{n=1}^\infty$ are Cauchy in $L_t^\infty([0, T]; W^{[k/2], p})$. Using the iteration system (3-2) and noting $W^{[k/5], p}$ is an algebra, we can upgrade the regularity in time and obtain that $\{\rho^{(n)}, \mathbf{v}^{(n)}, \nabla f^{(n)}\}_{n=1}^\infty$ are Cauchy in

$W_t^{[k/5],\infty}([0, T]; W^{[k/5],p})$. Therefore there exist

$$\begin{aligned}\rho &\in L_t^\infty([0, T], \dot{H}^{k-1} \cap L^p) \cap W_t^{[k/5],\infty}([0, T]; W^{[k/5],p}), \\ \mathbf{g} &\in L_t^\infty([0, T], \dot{H}^{k-1} \cap L^p) \cap W_t^{[k/5],\infty}([0, T]; W^{[k/5],p}), \\ v &\in L_t^\infty([0, T], \dot{H}^k \cap L^p) \cap W_t^{[k/5],\infty}([0, T]; W^{[k/5],p}).\end{aligned}$$

such that the following equations hold true in the classical sense:

$$(3-23) \quad \begin{cases} \partial_t \rho + \nabla \cdot (\rho \mathbf{v}) = 0, \\ \partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} + \mathbf{g} = 0, \\ \partial_t \mathbf{g} + \nabla (\mathbf{v} \cdot \mathbf{g}) + \nabla (\nabla \cdot \mathbf{v}) = 0. \end{cases}$$

Step 5: Continuity in highest norm. Since $(\rho, \mathbf{v}, \mathbf{g}) \in C([0, T], L_x^p)$, we only need to show $(\rho, \mathbf{g}) \in C([0, T], \dot{H}_x^{k-1})$, $\mathbf{v} \in C([0, T], \dot{H}^k)$. We shall only prove it for ρ as the others are similar. Fix any $t_0 \in [0, T]$, we compute

$$(3-24) \quad \begin{aligned} &\|\partial^{k-1}(\rho(t) - \rho(t_0))\|_2^2 \\ &= \|\partial^{k-1}\rho(t)\|_2^2 - \|\partial^{k-1}\rho(t_0)\|_2^2 + 2\langle \partial^{k-1}\rho(t_0) - \partial^{k-1}\rho(t), \partial^{k-1}\rho(t_0) \rangle, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the usual L^2 -pairing. By a simple density argument and the fact that $\rho \in C([0, T], L_x^p)$, we have¹

$$\lim_{t \rightarrow t_0} |(3-24)| = 0.$$

Therefore we only need to check the norm continuity, that is:

$$\lim_{t \rightarrow t_0} \|\partial^{k-1}\rho(t)\|_2^2 = \|\partial^{k-1}\rho(t_0)\|_2^2.$$

But this follows from a simple Gronwall estimate, which was essentially done in [Step 1](#). We omit the details.

Finally to recover the equation in (1-1) we still need to show $\rho > 0$ and $\mathbf{g} = \nabla \rho / \rho$. Since the initial data ρ_0 is positive, the positivity of ρ follows easily from the method of characteristics and the fact that $\mathbf{v} \in C^2$. We leave the proof that $\mathbf{g} = \nabla \rho / \rho$ to the next step.

Step 6: Identification of \mathbf{g} with $\nabla \rho / \rho$. We first show that

$$(3-25) \quad \frac{\nabla \rho}{\rho} \in C([0, T], L_x^p).$$

¹If $t_0 = 0$, then the left continuity can be obtained by the simple fact that our solution actually belongs to $C([-T_1, T], L_x^p)$ for some small T_1 since our system is inviscid.

From [Step 4](#) and using the positivity of $\rho^{(n)}$ and ρ , it is not difficult to check that up to a subsequence,

$$\frac{\nabla \rho^{(n)}}{\rho^{(n)}}(t, x) \rightarrow \frac{\nabla \rho}{\rho}(t, x), \quad a.e. (t, x) \in [0, T] \times \mathbb{R}^d.$$

Thus [\(3-25\)](#) can be proved if the sequence $\nabla \rho^{(n)}/\rho^{(n)}$ is Cauchy in $C_t^0 L_x^p$. To this end, we set $\mathbf{g}_1^{(n+1)} = \nabla \rho^{(n+1)}/\rho^{(n+1)}$. By the ρ -equation in [\(3-2\)](#) we have

$$\partial_t \mathbf{g}_1^{(n+1)} + \nabla(\nabla \cdot \mathbf{v}^{(n)}) + \nabla(\mathbf{v}^{(n)} \cdot \mathbf{g}_1^{(n+1)}) = 0.$$

Using integration by parts (note that $\mathbf{g}_1^{(n+1)}$ is gradient-like), we obtain

$$\frac{d}{dt} \|\mathbf{g}_1^{(n+1)}(t)\|_p \lesssim \|\mathbf{v}^{(n)}(t)\|_{X_k} (1 + \|\mathbf{g}_1^{(n+1)}(t)\|_p).$$

From Gronwall's inequality and the choice of T (shrinking T if necessary), we obtain

$$\|\mathbf{g}_1^{(n+1)}\|_{L_t^\infty([0, T]; L_x^p)} \leq 2M_0.$$

Similarly, we have

$$\|\partial \mathbf{g}_1^{(n+1)}\|_{L_t^\infty([0, T]; L_x^p)} \leq 2M_0.$$

Summing over all partial derivatives we see $\nabla \mathbf{g}_1^{(n+1)}$ is bounded in L^p .

For the L^p norm of the difference, we have

$$\frac{d}{dt} \|\mathbf{g}_1^{(n+1)} - \mathbf{g}_1^{(n)}\|_p \lesssim \|\mathbf{v}^{(n)}\|_{X_k} \|\mathbf{g}_1^{(n+1)} - \mathbf{g}_1^{(n)}\|_p + \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{X_{k-1}} \|\mathbf{g}_1^{(n)}\|_{W^{1,p}}.$$

Using the boundedness of $\mathbf{g}_1^{(n)}$ in $W^{1,p}$ and Gronwall, we have

$$\|\mathbf{g}_1^{(n+1)} - \mathbf{g}_1^{(n)}\|_{L_t^\infty([0, T]; L^p)} \leq C \|\mathbf{v}^{(n)} - \mathbf{v}^{(n-1)}\|_{L_t^\infty([0, T]; X_{k-1})}.$$

Therefore $\mathbf{g}_1^{(n)}$ is Cauchy in $C_t^0([0, T]; L_x^p)$. This completes the proof of [\(3-25\)](#).

We are now ready to show $\mathbf{g} = \nabla \rho / \rho$. Indeed, from the first equation in [\(3-23\)](#), we see $\nabla \rho / \rho$ satisfies in the classical sense

$$\partial_t \left(\frac{\nabla \rho}{\rho} \right) + \nabla \left(\mathbf{v} \cdot \frac{\nabla \rho}{\rho} \right) + \nabla(\nabla \cdot \mathbf{v}) = 0.$$

This equation has exactly the same form as the g -equation in [\(3-23\)](#). The identification of \mathbf{g} with $\nabla \rho / \rho$ then follows from the uniqueness of the solutions in the L^p class, to the following vector equation

$$\partial_t \mathbf{h} + \nabla(\mathbf{v} \cdot \mathbf{h}) = 0, \quad \mathbf{h}(0) \in L^p.$$

The uniqueness in L^p follows from a simple energy estimate which is omitted here. We note that if $\rho_0 \in L^1$, then the mass conservation follows from a standard truncation argument. We omit the details.

4. Proof of Theorem 1.3

By Theorem 1.1, for any chosen ρ_0, \mathbf{v}_0 , there exists a time $T > 0$ such that (1-1) admits a unique solution $(\rho(t, x), \mathbf{v}(t, x))$ on $[0, T]$. In particular, the local solution is at least C^2 and satisfies the equation in the classical sense. Since $\text{curl}(\mathbf{v}_0) = 0$, it is easy to check that $\text{curl}(\mathbf{v}(t)) = 0$ for any t . We first observe the property of finite propagation speed. Indeed, set

$$f = \log \rho.$$

Then the Euler equation (1-1) can be written as

$$\begin{cases} \partial_t f + \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla f = 0, \\ \partial_t \mathbf{v} + (\mathbf{v} \cdot \nabla) \mathbf{v} + \nabla f = 0. \end{cases}$$

Taking one more derivative in t for both equations and using the irrotational condition $\text{curl} \mathbf{v} = 0$, we have

$$\begin{cases} \partial_{tt} f - \Delta f = \frac{1}{2} \Delta(|\mathbf{v}|^2) + \frac{1}{2} \nabla f \cdot \nabla(|\mathbf{v}|^2) + |\nabla f|^2 + \mathbf{v} \cdot \nabla(\nabla \cdot \mathbf{v}) + \mathbf{v} \cdot \nabla(\mathbf{v} \cdot \nabla f), \\ \partial_{tt} \mathbf{v} - \Delta \mathbf{v} = \nabla(\mathbf{v} \cdot (\mathbf{v} \cdot \nabla) \mathbf{v}) + 2\nabla(\mathbf{v} \cdot \nabla f). \end{cases}$$

This is a standard quasilinear wave equation. The standard arguments (compare [Sogge 1995]), yields the finite propagation speed. In particular, we have $\rho(t, x) = 1, \mathbf{v}(t, x) = 0$ for all t, x such that $|x| \leq 1 - t$ and $t \leq T$.

We claim that the corresponding local solution $\rho(t, x), \mathbf{v}(t, x)$ must blow up before $t = 1$. We argue by contradiction. Suppose ρ, \mathbf{v} exist on $[0, 1]$, then we have

$$\frac{d}{dt} \int \rho \phi \, dx = \int \rho \mathbf{v} \cdot \nabla \phi \, dx.$$

Taking one more derivative in t , we get

$$\begin{aligned} \frac{d^2}{dt^2} \int \rho \phi \, dx &= \frac{d}{dt} \int \rho \mathbf{v} \cdot \nabla \phi \, dx \\ &= \int \rho \partial_t \mathbf{v} \cdot \nabla \phi + \int \partial_t \rho \mathbf{v} \cdot \nabla \phi \, dx \\ &= - \int (\rho(\mathbf{v} \cdot \nabla) \mathbf{v}) \cdot \nabla \phi \, dx - \int \nabla \cdot (\rho \mathbf{v}) \mathbf{v} \cdot \nabla \phi \, dx - \int \nabla \rho \cdot \nabla \phi \, dx \\ &= \int \rho \mathbf{v}_j \mathbf{v}_k \partial_{jk} \phi(x) \, dx + \int \rho \Delta \phi \, dx. \end{aligned}$$

Note $\mathbf{v}(t, x)$ vanishes on $|x| \leq 1$ for all $t \in [0, 1]$. For $|x| > 1$, we use the fact that $\nabla^2 \phi$ is positive definite and the boundedness of $\Delta \phi$ to get

$$\frac{d^2}{dt^2} \int \rho \phi \, dx > -C,$$

for some C depending on $\|\rho_0\|_1$. Therefore from the condition (1-7), we have

$$\frac{d}{dt} \int \rho \phi \, dx \geq N - C \quad \text{for } t \in [0, 1].$$

This implies

$$\int \rho(1, x) \phi(x) \, dx \geq \int \rho_0(x) \phi(x) \, dx + N - C,$$

which, for N large enough, contradicts the fact that

$$\int \rho(1, x) \phi(x) \, dx \leq \|\rho(1)\|_1 \|\phi\|_\infty.$$

This completes the proof of [Theorem 1.3](#).

References

- [Chae and Ha 2009] D. Chae and S.-Y. Ha, “On the formation of shocks to the compressible Euler equations”, *Commun. Math. Sci.* **7**:3 (2009), 627–634. [MR 2011a:35331](#) [Zbl 1183.35225](#)
- [Godin 2005] P. Godin, “The lifespan of a class of smooth spherically symmetric solutions of the compressible Euler equations with variable entropy in three space dimensions”, *Arch. Ration. Mech. Anal.* **177**:3 (2005), 479–511. [MR 2006i:76092](#) [Zbl 1075.76052](#)
- [John 1974] F. John, “Formation of singularities in one-dimensional nonlinear wave propagation”, *Comm. Pure Appl. Math.* **27** (1974), 377–405. [MR 51 #6163](#) [Zbl 0302.35064](#)
- [Kato 1975] T. Kato, “The Cauchy problem for quasi-linear symmetric hyperbolic systems”, *Arch. Rational Mech. Anal.* **58**:3 (1975), 181–205. [MR 52 #11341](#) [Zbl 0343.35056](#)
- [Klainerman and Majda 1980] S. Klainerman and A. Majda, “Formation of singularities for wave equations including the nonlinear vibrating string”, *Comm. Pure Appl. Math.* **33**:3 (1980), 241–263. [MR 81f:35080](#) [Zbl 0443.35040](#)
- [Lax 1964] P. D. Lax, “Development of singularities of solutions of nonlinear hyperbolic partial differential equations”, *J. Mathematical Phys.* **5** (1964), 611–613. [MR 29 #2532](#) [Zbl 0135.15101](#)
- [Liu 1979] T. P. Liu, “Development of singularities in the nonlinear waves for quasilinear hyperbolic partial differential equations”, *J. Differential Equations* **33**:1 (1979), 92–111. [MR 80g:35075](#) [Zbl 0379.35048](#)
- [Rammaha 1989] M. A. Rammaha, “Formation of singularities in compressible fluids in two-space dimensions”, *Proc. Amer. Math. Soc.* **107**:3 (1989), 705–714. [MR 90f:35127](#) [Zbl 0692.35015](#)
- [Sideris 1985] T. C. Sideris, “Formation of singularities in three-dimensional compressible fluids”, *Comm. Math. Phys.* **101**:4 (1985), 475–485. [MR 87d:35127](#) [Zbl 0606.76088](#)
- [Sideris 1991] T. C. Sideris, “The lifespan of smooth solutions to the three-dimensional compressible Euler equations and the incompressible limit”, *Indiana Univ. Math. J.* **40**:2 (1991), 535–550. [MR 92f:35119](#) [Zbl 0736.35087](#)
- [Sogge 1995] C. D. Sogge, *Lectures on nonlinear wave equations*, International Press, Boston, MA, 1995. [MR 2000g:35153](#) [Zbl 1089.35500](#)

Received January 8, 2012. Revised August 1, 2012.

DONG LI
INSTITUTE FOR ADVANCED STUDY
PRINCETON, NJ 08540
UNITED STATES

and

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BC V6T 1Z2
CANADA

mpdongli@gmail.com

CHANGXING MIAO
INSTITUTE OF APPLIED PHYSICS AND COMPUTATIONAL MATHEMATICS
BEIJING 100088
CHINA

miao_changxing@iapcm.ac.cn

XIAOYI ZHANG
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF IOWA
14 MACLEAN HALL
IOWA CITY, IA 52242
UNITED STATES

and

ACADEMY OF MATHEMATICS AND SYSTEMS SCIENCES
BEIJING 100080
CHINA

zh.xiaoyi@gmail.com

SYMMETRIC REGULARIZATION, REDUCTION AND BLOW-UP OF THE PLANAR THREE-BODY PROBLEM

RICHARD MOECKEL AND RICHARD MONTGOMERY

We carry out a sequence of coordinate changes for the planar three-body problem, which successively eliminate the translation and rotation symmetries, regularize all three double collision singularities and blow-up the triple collision. Parametrizing the configurations by the three relative position vectors maintains the symmetry among the masses and simplifies the regularization of binary collisions. Using size and shape coordinates facilitates the reduction by rotations and the blow-up of triple collision while emphasizing the role of the shape sphere. By using homogeneous coordinates to describe Hamiltonian systems whose configuration spaces are spheres or projective spaces, we are able to take a modern, global approach to these familiar problems. We also show how to obtain the reduced and regularized differential equations in several convenient local coordinate systems.

1. Introduction and history

The three-body problem of Newton has symmetries and singularities. The reduction process eliminates symmetries thereby reducing the number of degrees of freedom. The Levi-Civita regularization eliminates binary collision singularities by a noninvertible coordinate change together with a time reparametrization. The McGehee blow-up eliminates the triple collision singularity by an ingenious polar coordinate change and another time reparametrization. Each process has been applied individually and in various combinations to the three-body problem, many times.

In this paper we apply all three processes globally and systematically, with no one body singled out in the various transformations. The end result is a complete flow on a five-dimensional manifold with boundary. We focus attention on the geometry of the various spaces and maps appearing along the way. At the heart of this paper is a beautiful degree-4 octahedral covering map of the shape sphere, branched over the binary collision points (see [Figure 4](#) on page 179). This map

Research supported by NSF grant DMS-1208908.

MSC2010: primary 37N05, 70F07, 70G45; secondary 53A20, 53CXX.

Keywords: celestial mechanics, three-body problem, regularization.

first appears in the work of Lemaître [1954; 1964]. One of our goals is to give a modern, geometrical approach to this regularizing map.

The reduction procedure for the three body problem dates back to Lagrange [1772] who found elegant differential equations for 10 translation and rotation invariant variables, including the squares of the lengths of the three sides of the triangle formed by the bodies. These equations are valid for the three-body problem in any dimension. The variables of Lagrange also have the advantage of maintaining the symmetry among the masses. On the other hand, for the planar problem they are subject to 3 nonlinear constraints in addition to the energy and angular momentum integrals. Moreover, we do not know a way to regularize the binary collision singularities in Lagrange's equations. For a modern introduction to Lagrange's equations; see [Albouy and Chenciner 1998; Albouy 2004; Chenciner 2011].

Jacobi eliminates the translation symmetry by the familiar device of fixing the center of mass at the origin and introducing Jacobi coordinates [1843]. The elimination of rotations is achieved by introducing some angular variable (or variables in the spatial case) to describe the overall rotation of the triangle together with some complementary, rotation-invariant variables. This method, which is the basis for much of the later work on the three-body problem, has some disadvantages. First, the Jacobi coordinates break the symmetry among the masses, making it much more difficult to regularize all three binary collisions at once. Second, for topological reasons, there is no way to choose an angular variable suitable for a global reduction that includes the binary collision configurations, namely, the map from the normalized configuration space to the shape sphere is a Hopf fibration, a nontrivial circle bundle. If we delete the binary collision points, the bundle becomes trivial but this deletion is not conducive to subsequent regularization.

Murnaghan [1936] derived a symmetrical Hamiltonian for the planar three-body problem in terms of the lengths of the sides and an angular variable representing the overall rotation of the triangle with respect to an inertial coordinate system. Then he obtains a reduced system by ignoring the angular variable. Van Kampen and Wintner [1937] carry out a similar reduction for the spatial three-body problem. While these reductions avoid breaking the symmetry, they are still subject to the problem about the use of angular variables in a nontrivial bundle. In addition, using the side lengths as variables leads to differential equations that are not smooth at the collinear configurations (a problem seemingly avoided somehow by Lagrange).

Lemaître [1954] introduced a symmetrical approach to reduction and regularization of binary collisions leading to the octahedral branched covering map of the sphere mentioned above. After using Euler angles to reduce by rotations, he introduces a size variable and two shape variables, which can be viewed as spherical coordinates on the shape sphere which we use below. The regularization of binary collisions is done in the shape variables by means of the octahedral covering map.

The use of Euler angles limits the validity of the reduction step of Lemaître's work and the derivations are based on rather heavy trigonometric computations. But much of this paper can be viewed as a modern, global way to arrive at his covering map.

In this endeavor we have the advantage of the modern theory of reduction of Hamiltonian systems with symmetry. Smale [1970] describes the reduction process for the three-body problem as the formation of a quotient manifold with a reduced Hamiltonian flow. Meyer [1973] and Marsden and Weinstein [1974] formalized the reduction procedure into what is now called "symplectic reduction theory". Fixing the integrals of motion determines invariant manifolds in phase space. The quotient spaces of these invariant manifolds are the reduced phase spaces and the flows induced on them are again Hamiltonian with respect to an appropriate symplectic structure and a reduced Hamiltonian function.

The regularization procedure goes back to Levi-Civita [1920], who showed how to regularize binary collisions in perturbed planar Kepler problems by using the complex squaring map (a branched double covering of the complex plane). It is easy to adapt his method to regularize one of the binary collisions in the three-body problem, but regularizing all three requires more ingenuity. Lemaître's regularizing map behaves like the complex squaring map at each of the binary collision points on the shape sphere. Another approach to simultaneous regularization (without reduction) was introduced by Waldvogel [1972], who used a quadratic mapping of the translation-reduced configuration space \mathbb{C}^2 . We use a similar quadratic mapping applied to certain homogeneous shape variables below. Heggie [1974] found an elegant, symmetrical way to regularize all of the binary collisions for the N -body problem. In the planar case, his method is to apply separate Levi-Civita transformations to each of the difference vectors $q_i - q_j$. We apply this same idea below, but to the homogeneous shape variables, where it is found to induce Lemaître's octahedral covering.

Triple collision acts like an essential singularity in the three-body problem. McGehee [1974] showed how an extension of spherical coordinates, together with a time reparametrization, yields a flow with no singularities at triple collision. This "McGehee blow-up" has the effect of replacing the triple collision point by a manifold called the collision manifold. Relative to the new parametrization, it takes forever to reach triple collision, whereas the Newtonian time to triple collision is finite. The flow on the triple collision manifold governs the behavior of near-triple collision solutions. One aspect of the blow-up procedure is the use of separate size and shape coordinates to describe the configuration of the bodies. As shown below, such a splitting also facilitates the global reduction by rotations.

Several authors have combined blow-up of triple collision with reduction and/or regularization of binary collision. Waldvogel [1982] reduced and regularized the flow on the zero-angular-momentum triple collision manifold. The first part of his

paper combines Murnaghan's reduction procedure with some formulas of Lemaître to obtain a reduced and regularized Hamiltonian for the zero-angular momentum three-body problem. Binary collisions are not regularized on the nonzero angular momentum levels. However, it is known that triple collisions can only occur when the angular momentum is zero. After restricting to the zero angular momentum manifold, Waldvogel blows up the triple collision to get reduced, regularized and blown-up differential equations. Simó and Susín [1991] used these coordinates in their study of the dynamics on the collision manifold. These coordinates are very much in the spirit of this paper but do not achieve a full reduction, regularization and blow-up due to the restriction to zero angular momentum.

The present paper draws on all these sources. We begin with some symplectic reduction theory. Turning to the three-body problem, we eliminate translation symmetry by introducing the three difference vectors $Q_{ij} = q_i - q_j$ as coordinates. Since these are linearly dependent, some effort is needed to justify the change of coordinates. Next we introduce a size variable and associated spherical coordinates X_{ij} . One novelty of our approach is that we use homogeneous coordinates to describe points on spheres. Instead of constraining the spherical coordinates to have a fixed norm, we only ask them to avoid the origin and then we find differential equations for them that are invariant under scaling.

Once this point of view is adopted, it is relatively easy to carry out a global reduction by rotations. Using complex coordinates, the combined action of scaling and rotation is just scaling by a complex number. Quotienting by complex scaling, we end up with a complex projective space, in fact with $\mathbb{C}\mathbb{P}^1$. Of course, as real manifolds, $\mathbb{C}\mathbb{P}^1 \simeq S^2$, and this is our version of the shape sphere. We finally obtain a global reduction of the planar three-body problem with a six-dimensional reduced phase space, the cotangent bundle of $\mathbb{R}^+ \times S^2$.

Turning to regularization, we use simultaneous Levi-Civita transformations of the homogeneous variables X_{ij} to regularize all three binary collisions. This regularizing map is applied to both the rotation-reduced and unreduced problems. In the reduced case we get a reduced and regularized system on the cotangent bundle of $\mathbb{R}^+ \times S^2$, which is related to the unregularized version by Lemaître's map.

Finally we show how McGehee's blow-up procedure can be applied to the various Hamiltonians we have found.

2. Symplectic reduction

In this section we recall some results about the reduction of a Hamiltonian system with symmetry. In addition we show how to tell when two symmetric Hamiltonian systems lead to equivalent reduced systems.

First we describe the basic symplectic reduction theory of Meyer [1973] and Marsden and Weinstein [1974] in the case of a system with symmetry. Suppose (M, ω) is a symplectic manifold and G is a Lie group which acts on M as a group of symplectic diffeomorphisms. Let $J : M \rightarrow \mathfrak{g}^*$ be the momentum map, where \mathfrak{g}^* is the dual of the Lie algebra of G . If we fix a momentum value $\mu \in \mathfrak{g}^*$ and suppose that the action of G maps the level set $J^{-1}(\mu)$ into itself, the quotient space

$$P_\mu = J^{-1}(\mu)/G$$

is called the *reduced phase space*.

If the group action is free and proper, then this space is a smooth manifold. There is an induced symplectic form ω_μ on P_μ , which is obtained as follows. First, for $x \in M$, restrict $\omega(x)$ to the tangent spaces $T_x J^{-1}(\mu)$. The resulting two-form has a kernel, which is precisely the tangent space to the group orbit through x . This implies that there is an induced two-form on the quotient vector space that is the tangent space to the quotient manifold.

Now if $H : M \rightarrow \mathbb{R}$ is a G -invariant Hamiltonian then the corresponding Hamiltonian flow has $J^{-1}(\mu)$ as an invariant set and G -orbits map to G -orbits under the flow. Hence there is a well-defined quotient flow on $J^{-1}(\mu)/G$. There is also a reduced Hamiltonian $H_\mu : P_\mu \rightarrow \mathbb{R}$ and the reduction theorem states that the quotient flow on (P_μ, ω_μ) is the Hamiltonian flow of the reduced Hamiltonian.

Now suppose we have two such Hamiltonian systems with symmetry. For $i = 1, 2$, there will be symplectic manifolds (M_i, ω_i) , symmetry groups G_i and momentum maps J_i . If we fix momentum values μ_i , we get reduced phase spaces $P_i = J_i^{-1}(\mu_i)/G_i$ with symplectic forms ω_{μ_i} . Suppose $H_i : M_i \rightarrow \mathbb{R}$ are G_i -invariant Hamiltonians and let $H_{\mu_i} : P_i \rightarrow \mathbb{R}$ be the corresponding reduced Hamiltonians. We want to give a concrete way to check that the two reduced Hamiltonian flows are equivalent.

Suppose we have a smooth map $F : J_1^{-1}(\mu_1) \rightarrow J_2^{-1}(\mu_2)$ that maps G_1 -orbits into G_2 -orbits; that is, F is equivariant. Then F induces a smooth map of quotient manifolds $\hat{F} : P_1 \rightarrow P_2$. We will call F *partially symplectic* if it preserves the restrictions of the symplectic forms, that is,

$$F^*(\omega_2|_{J_2^{-1}(\mu_2)}) = \omega_1|_{J_1^{-1}(\mu_1)}.$$

It follows that $\hat{F} : (P_1, \omega_{\mu_1}) \rightarrow (P_2, \omega_{\mu_2})$ is symplectic. Hence \hat{F} is a local diffeomorphism, even if F itself is locally neither injective nor surjective. Then the usual theory of symplectic maps applied to \hat{F} gives:

Theorem 1. *Suppose $F : J_1^{-1}(\mu_1) \rightarrow J_2^{-1}(\mu_2)$ is a partially symplectic, equivariant map and that the restrictions of the Hamiltonians are related by $H_1 = H_2 \circ F$. Then*

$\hat{F} : P_1 \rightarrow P_2$ is a symplectic, local diffeomorphism of the reduced phase spaces, which takes orbits of the reduced Hamiltonian flow of H_{μ_1} to those of H_{μ_2} .

Definition 2. A partially symplectic, equivariant map $G : J_2^{-1}(\mu_2) \rightarrow J_1^{-1}(\mu_1)$ such that $F \circ G = \text{id} \pmod{G_2}$ and $G \circ F = \text{id} \pmod{G_1}$ (so that these maps take group orbits into group orbits) will be called a *pseudoinverse* for F .

A partial inverse G for F induces a bona fide inverse \hat{G} for \hat{F} , which exhibits an equivalence between the two reduced Hamiltonian flows.

As a special case, suppose the two Hamiltonians are both defined on the same space and have the same symmetry group. If their restrictions to $J^{-1}(\mu)$ agree then they will lead to the same reduced system. The identity map will provide the required partially symplectic map. We will call two such Hamiltonians *equivalent*. Equivalent Hamiltonians may produce different flows on $J^{-1}(\mu)$ but the quotient flows will agree.

The following theorems about the symplectic reduction of a cotangent bundle $M = T^*X$ will be used later. (See [Abraham and Marsden 1978, Theorem 4.3.3] for a version of these theorems.) Suppose G acts freely on the configuration space X and that the G -action on M is the canonical lift of this base action. Suppose that the orbit space B for the G action on X is a manifold and the projection $\pi : X \rightarrow B$ a submersion.

Theorem 3. *Under the above assumptions, the reduced space P_0 of T^*X at $\mu = 0$ is isomorphic to T^*B with its canonical symplectic structure ω_B .*

The theorem can be proved as a special case of [Theorem 1](#). Because π is onto, $d\pi_x : T_x X \rightarrow T_{\pi(x)} B$ is an onto linear map for each $x \in X$. Consequently the dual map $d\pi_x^* : T_{\pi(x)}^* B \rightarrow T_x^* X$ is injective. In the next paragraph we will show that the image of this dual is $J^{-1}(0)_x$:

$$(1) \quad \text{im}(d\pi_x^*) = J^{-1}(0)_x := J^{-1}(0) \cap T_x^* X.$$

It follows that we can invert $d\pi_x^*$ on the fiber $J^{-1}(0)_x \subset T_x^* X$. Define

$$F : J^{-1}(0) \rightarrow T^*B ; F(x, p) = (\pi(x), d\pi_x^{*-1}(p)).$$

One verifies that F is a partially symplectic map relative to G acting on $J^{-1}(0)$, and the trivial group acting on T^*B . A particularly easy way to see the partially symplectic nature of F is to introduce local bundle coordinates $X \supset \pi^{-1}(U) \cong U \times G$. (X is covered by sets of this nature.) In bundle coordinates $\pi(x, g) = x$, and so $T_U^* X \cong T^*U \times G \times \mathfrak{g}^*$. We write elements of T^*X over U as $(b, P; g, \mu)$, $b \in U$, $P \in T_b^*U$, $g \in G$, $\mu \in \mathfrak{g}^*$. In these coordinates $J(b, P; g, \mu) = \mu$, so that the general element of $J^{-1}(0)_U$ can be written $(b, P_b, g, 0)$ and $F(b, P_b, g, 0) = (b, P_b)$. We have $\omega_X = dx \wedge dP + dg \wedge d\mu$ and, $\omega_B = dx \wedge dP$, where we hope the meaning of these

symbolic expressions is obvious. It follows immediately that $F^*\omega_B = \omega_X|_{J^{-1}(0)}$, which is the claimed partially symplectic nature of F . [Theorem 3](#) follows.

We explain why [\(1\)](#) holds, and in the process gain some understanding of the momentum map. The group action is a map $G \times X \rightarrow X$ which, when differentiated with respect to $g \in G$ at the identity, yields the “infinitesimal action” $\sigma : \mathfrak{g} \times X \rightarrow TX$. For each frozen x , the map $\sigma_x : \mathfrak{g} \rightarrow T_x X$ is linear and, because G acts freely, injective. As we vary x , σ forms a vector bundle map, part of an exact sequence of vector bundle maps over X :

$$0 \rightarrow \mathfrak{g} \times X \xrightarrow{\sigma} TX \xrightarrow{d\pi} \pi^*TB$$

where $\pi^*TB = \{(x, V); x \in X, V \in T_{\pi(x)}B\}$ is the pull-back of TB over B by the map $\pi : X \rightarrow B$. (Exactness of the sequence follows by differentiating the statement that the fibers of π are the G -orbits.) Dualizing, we get

$$0 \leftarrow \mathfrak{g}^* \times X \xleftarrow{\sigma^*} T^*X \xleftarrow{d\pi^*} \pi^*T^*B.$$

The momentum map for the G -action on T^*X is $\pi_1 \circ \sigma^*$, where $\pi_1 : \mathfrak{g}^* \times X \rightarrow \mathfrak{g}^*$ is the projection onto the first factor. In other words,

$$J(x, p) = \sigma_x^* p.$$

It follows from the exactness of the dual sequence that $\text{im}(d\pi_x^*) = \ker(\sigma_x^*)$, which is precisely [\(1\)](#).

In order to identify the reduction of $M = T^*X$ at a nonzero value, $\mu \neq 0$, we introduce a connection Γ for the bundle $G \rightarrow X \rightarrow B$. The curvature of the connection Γ is a \mathfrak{g} -valued two-form Ω on B , which we may pull-back to T^*B via the canonical projection $\tau_B : T^*B \rightarrow B$. Then $\mu \cdot \Omega$ is a scalar-valued two-form on B .

Theorem 4. *Under the same assumptions as above on G , the reduced space P_μ of T^*X at μ is isomorphic to T^*B with the twisted symplectic structure $\omega_B - \tau_B^* \mu \cdot \Omega$.*

We only present the proof in the case $G = S^1$, whose Lie algebra we identify with \mathbb{R} in the usual way. Then a connection is a G -invariant one-form on T^*X that satisfies the normalization property $J(x, \Gamma(x)) = 1$. Its curvature Ω is defined by $d\Gamma = \pi^*\Omega$. We define the momentum shift map

$$\Phi_\mu : J^{-1}(0) \rightarrow J^{-1}(\mu), \quad \Phi_\mu(x, p) = (x, p + \mu\Gamma(x)),$$

which adds $\mu\Gamma$ pointwise to each covector. The fiber-linearity of J shows that Φ_μ does indeed map $J^{-1}(0)$ onto $J^{-1}(\mu)$. (The inverse of Φ_μ subtracts $\mu\Gamma$.) The map is G -equivariant since Γ is G -invariant. Thus Φ_μ induces a G -equivariant diffeomorphism $J^{-1}(0)/G \rightarrow J^{-1}(\mu)/G$. We have already identified $J^{-1}(0)/G$ with T^*B . However, Φ_μ is not partially symplectic, so we cannot directly apply

Theorem 1. To understand and quantify this failure, let $\Theta = P dQ$ denote the canonical one-form on T^*X . Compute $\Phi_\mu^* \Theta = \Theta + \mu \tau_X^* \Gamma$. Taking the exterior derivative, using $\omega_X = -d\Theta$, we find that $\Phi_\mu^* \omega_X = \omega_X - \mu \tau_X^* \pi^* \Omega$. This equation implies that if we shift the canonical two-form on $J^{-1}(0)$ by subtracting $\mu \tau_X^* \pi^* \Omega$ then Φ_μ is a partially symplectic map between $J^{-1}(0)$ and $J^{-1}(\mu)$. **Theorem 4** follows.

3. Reduction by translations

To formulate the Newtonian planar three-body problem, it is convenient to use the complex plane, where we identify $(x, y) \in \mathbb{R}^2$ with $x + iy \in \mathbb{C}$.

Let $q_1, q_2, q_3 \in \mathbb{C}$ be the positions of the three bodies and let $q = (q_1, q_2, q_3) \in \mathbb{C}^3$. We will adopt the Hamiltonian point of view, where the conjugate momentum variables p_i are covectors rather than vectors. If we identify a covector $(a, b) \in \mathbb{R}^{2*}$ with $a + ib \in \mathbb{C}$, then we have momentum variables

$$p_i \in \mathbb{C}^* \simeq \mathbb{C} \quad \text{and} \quad p = (p_1, p_2, p_3) \in \mathbb{C}^{3*}.$$

The planar three-body problem is the Hamiltonian system on the phase space $(\mathbb{C}^3 \setminus \Delta) \times \mathbb{C}^{3*}$ with Hamiltonian

$$(2) \quad \begin{aligned} H(q, p) &= K_0(p) - U(q), \\ K_0(p) &= \frac{|p_1|^2}{2m_1} + \frac{|p_2|^2}{2m_2} + \frac{|p_3|^2}{2m_3}, \\ U(q) &= \frac{m_1 m_2}{|q_1 - q_2|} + \frac{m_3 m_1}{|q_3 - q_1|} + \frac{m_2 m_3}{|q_2 - q_3|}, \end{aligned}$$

where $\Delta = \{q : q_i = q_j \text{ for some } i \neq j\}$, the singular set. From now on, we will not explicitly mention that the singular set must be deleted from the domains of the various Hamiltonians we construct.

The Newtonian potential is invariant under the group $G = \mathbb{C}$ acting by translation on the position vectors and leaving the momenta fixed. The momentum map is given by

$$p_{\text{tot}} = p_1 + p_2 + p_3 \in \mathbb{C}^*.$$

By fixing a value of this integral and passing to the quotient space, one obtains a reduced Hamiltonian system. A simple and familiar way to accomplish this reduction is to assume $p_{\text{tot}} = 0$ and then fix the center of mass at the origin: $m_1 q_1 + m_2 q_2 + m_3 q_3 = 0$.

However, we will now describe an alternative method for eliminating the translation symmetry, which will make it easier to regularize double collisions later on. This approach is a variation on the one used in [Heggie 1974]. We will view it as an application of **Theorem 1**.

3.1. Relative coordinates. Introduce relative position variables $Q_{12}, Q_{31}, Q_{23} \in \mathbb{C}$ and corresponding momentum variables $P_{12}, P_{31}, P_{23} \in \mathbb{C}^*$. The relative coordinates are related to the positions variables q_i by a linear map $Q = Lq$

$$(3) \quad L : \mathbb{C}^3 \rightarrow \mathbb{C}^3, \quad Q_{12} = q_1 - q_2, \quad Q_{31} = q_3 - q_1, \quad Q_{23} = q_2 - q_3.$$

The dual map, which describes the pull-back of the relative momenta P_{ij} to p space, is given by

$$(4) \quad L^* : \mathbb{C}^{3*} \rightarrow \mathbb{C}^{3*}, \quad p_1 = P_{12} - P_{31}, \quad p_2 = P_{23} - P_{12}, \quad p_3 = P_{31} - P_{23}.$$

We naturally have $Q_{ji} = -Q_{ij}$ and consequently $P_{ji} = -P_{ij}$ so that (4) can be written $p_i = \sum_j P_{ij}$, a form which extends to the N -body problem.

The linear map L is neither one-to-one nor onto. Its kernel,

$$\ker L = \{q : q = (c, c, c) \text{ for some } c \in \mathbb{R}^2 = \mathbb{C}\},$$

is the subspace of translation symmetries in q -space. So its image

$$\mathfrak{W} = \text{im } L = \{Q : Q_{12} + Q_{31} + Q_{23} = 0\}$$

is isomorphic to the quotient space of \mathbb{C}^3 by translations. \mathfrak{W} is a complex subspace of \mathbb{C}^3 with complex dimension two, or real dimension 4. We can define a map in the other direction, $q = L^\dagger(Q)$:

$$(5) \quad L^\dagger : q_1 = \frac{m_2 Q_{12} - m_3 Q_{31}}{m}, \quad q_2 = \frac{m_3 Q_{23} - m_1 Q_{12}}{m}, \quad q_3 = \frac{m_1 Q_{31} - m_2 Q_{23}}{m},$$

where $m = m_1 + m_2 + m_3$. L^\dagger maps \mathbb{C}^3 onto

$$\mathfrak{W}' = \text{im } L^\dagger = \{q : m_1 q_1 + m_2 q_2 + m_3 q_3 = 0\},$$

the zero-center of mass subspace, and it is easy to check that the restrictions $L|_{\mathfrak{W}'}$ and $L^\dagger|_{\mathfrak{W}}$ are inverses.

For the dual map, we find that the kernel is generated by translations in P -momentum space

$$\ker L^* = \{P : P = (c, c, c) \text{ for some } c \in \mathbb{C}^*\}$$

while the image is the zero-momentum subspace

$$\mathfrak{V} = \text{im } L^* = \{p : p_1 + p_2 + p_n = 0\}.$$

The map $L^{\dagger*} : \mathbb{C}^{3*} \rightarrow \mathbb{C}^{3*}$

$$(6) \quad L^{\dagger*} : \quad P_{12} = \frac{m_2 p_1 - m_1 p_2}{m}, \quad P_{31} = \frac{m_1 p_3 - m_3 p_1}{m}, \quad P_{23} = \frac{m_3 p_2 - m_2 p_3}{m}$$

maps \mathbb{C}^{3*} onto

$$\mathcal{V}' = \text{im } L^{\dagger*} = \{P : m_3 P_{12} + m_2 P_{31} + m_1 P_{23} = 0\},$$

and the restrictions $L^*|_{\mathcal{V}'}$ and $L^{\dagger*}|_{\mathcal{V}'}$ are inverses.

Define a relative coordinate Hamiltonian on the (Q, P) phase space $\mathbb{C}^3 \times \mathbb{C}^{3*}$ by

$$(7) \quad \begin{aligned} H_{\text{rel}}(Q, P) &= K(P) - U(Q), \\ K(P) &= K_0(L^*P) = \frac{|P_{12} - P_{31}|^2}{2m_1} + \frac{|P_{23} - P_{12}|^2}{2m_2} + \frac{|P_{31} - P_{23}|^2}{2m_3}, \\ U(Q) &= \frac{m_1 m_2}{|Q_{12}|} + \frac{m_3 m_1}{|Q_{31}|} + \frac{m_2 m_3}{|Q_{23}|}, \end{aligned}$$

so that

$$(8) \quad H(q, L^*P) = H_{\text{rel}}(Lq, P).$$

The kinetic energy can be written

$$(9) \quad K(P) = \frac{1}{2} P^T B P, \quad \text{with } B = \begin{bmatrix} \left(\frac{1}{m_1} + \frac{1}{m_2}\right)I & -\frac{1}{m_1}I & -\frac{1}{m_2}I \\ -\frac{1}{m_1}I & \left(\frac{1}{m_3} + \frac{1}{m_1}\right)I & -\frac{1}{m_3}I \\ -\frac{1}{m_2}I & -\frac{1}{m_3}I & \left(\frac{1}{m_2} + \frac{1}{m_3}\right)I \end{bmatrix},$$

where I denotes the 2×2 identity matrix.

3.2. Equivalence to the translation-reduced three-body problem. We will now show that the reduction of the Hamiltonian system with Hamiltonian $H_{\text{rel}}(Q, P)$ by translations in momentum space is equivalent to the reduction of the three-body Hamiltonian H by translations in configuration space.

Theorem 5. $\mathcal{W} \times \mathbb{C}^{3*}$ is invariant under the Hamiltonian flow of $H_{\text{rel}}(Q, P)$. The restricted flow is invariant under translations in momentum space and it induces a quotient flow, which is conjugate to the zero total momentum flow of the three-body problem reduced by translations.

The proof will be an application of [Theorem 1](#). First we describe how the relevant symplectic structures look in complex coordinates. If $Q \in \mathbb{C}^3$ and $P \in \mathbb{C}^{3*}$ it is convenient to define a Hermitian variant of the natural evaluation pairing:

$$(10) \quad \langle P, Q \rangle = \bar{P}_{12} Q_{12} + \bar{P}_{31} Q_{31} + \bar{P}_{23} Q_{23}.$$

As a result, if $Q_{jk} = x_{jk} + i y_{jk}$ and $P_{jk} = a_{jk} + i b_{jk}$, we get

$$(11) \quad \begin{aligned} \text{re}\langle P, Q \rangle &= a_{12}x_{12} + b_{12}y_{12} + \cdots, \\ \text{im}\langle P, Q \rangle &= a_{12}y_{12} - b_{12}x_{12} + \cdots. \end{aligned}$$

Thus the real part of the complex pairing agrees with the usual real pairing and, as a bonus, the imaginary part is $-\mu$, where μ is the angular momentum. With this convention, the canonical one-forms on (q, p) -space and (Q, P) -space can be written

$$(12) \quad \begin{aligned} \theta &= \operatorname{re}\langle p, dq \rangle = \operatorname{re}(\bar{p}_1 dq_1 + \bar{p}_2 dq_2 + \bar{p}_3 dq_3) \\ \Theta &= \operatorname{re}\langle P, dQ \rangle = \operatorname{re}(\bar{P}_{12} dQ_{12} + \bar{P}_{31} dQ_{31} + \bar{P}_{23} dQ_{23}). \end{aligned}$$

Proof of Theorem 5. For the three-body problem we have the phase space

$$M_1 = \mathbb{C}^6 \times \mathbb{C}^{6*} = \{(q, p)\},$$

with the standard symplectic structure. The Hamiltonian $H(q, p)$ is invariant under the action of the group $G_1 = \mathbb{C}$ acting by

$$c \cdot (q, p) = (q_1 + c, q_2 + c, q_3 + c, p_1, p_2, p_3), \quad c \in \mathbb{C}.$$

We fix the momentum level $p_{\text{tot}} = 0$ and obtain a quotient Hamiltonian flow.

For the Hamiltonian H_{rel} , the phase space is $M_2 = \mathbb{C}^3 \times \mathbb{C}^{3*} = \{(Q, P)\}$ with the standard symplectic structure. $H_{\text{rel}}(Q, P)$ is invariant under the action of the group $G_2 = \mathbb{C}^*$ acting on by $c \cdot (Q, P) = (Q_{12}, Q_{31}, Q_{23}, P_{12} + c, P_{31} + c, P_{23} + c)$, $c \in \mathbb{C}^*$. The momentum map is $Q_{\text{tot}} = Q_{12} + Q_{31} + Q_{23}$ and we fix the momentum level $Q_{\text{tot}} = 0$ giving a second quotient Hamiltonian flow.

To see that these two quotient flows are equivalent we apply [Theorem 1](#). Define

$$F(q, p) = (Lq, L^{\dagger*}p), \quad G(Q, P) = (L^{\dagger}Q, L^*P).$$

Then, $F : \{p_{\text{tot}} = 0\} \rightarrow \{Q_{\text{tot}} = 0\}$ and $G : \{Q_{\text{tot}} = 0\} \rightarrow \{p_{\text{tot}} = 0\}$. Moreover, $G \circ F(q, p) = c \cdot (q, p)$, where $-c = \frac{1}{m}(m_1q_1 + m_2q_2 + m_3q_3) \in \mathbb{C}$ is the center of mass. Similarly, $F \circ G(Q, P) = c \cdot (Q, P)$, where

$$-c = \frac{1}{m}(m_3P_{12} + m_2P_{31} + m_1P_{23}) \in \mathbb{C}^*.$$

In other words $G \circ F = \text{id} \pmod{G_1}$ and $F \circ G = \text{id} \pmod{G_2}$.

It remains to verify that F and G are partially symplectic. Consider the canonical one-forms (12). From (3) and (6). We find, for example $F^*\bar{P}_{12} = (m_2\bar{p}_1 - m_1\bar{p}_2)/m$ and $F^*dQ_{12} = dq_1 - dq_2$. After a bit of algebra we get

$$F^*\Theta = \theta - \operatorname{re}\left(\frac{\bar{p}_{\text{tot}}(m_1dq_1 + m_2dq_2 + m_3dq_3)}{m}\right).$$

Restricting to $\{p_{\text{tot}} = 0\}$ shows that F is partially symplectic. Similarly,

$$G^*\theta = \Theta - \frac{\operatorname{re}((m_3\bar{P}_{12} + m_2\bar{P}_{31} + m_1\bar{P}_{23})(dQ_{12} + dQ_{31} + dQ_{23}))}{m},$$

which we restrict to $\{Q_{\text{tot}} = 0\}$ to see that G is also partially symplectic. We have shown that F and G are pseudoinverses in the sense of [Definition 2](#). According to [\(8\)](#) these pseudoinverses intertwine H and H_{rel} . The hypotheses of [Theorem 1](#) have been verified, completing the proof. \square

Hamilton's equations for the Hamiltonian $H_{\text{rel}}(Q, P)$ are simply

$$(13) \quad \begin{aligned} \dot{Q} &= BP, \\ \dot{P} = U_Q &= -\left(\frac{m_1 m_2 Q_{12}}{r_{12}^3}, \frac{m_3 m_1 Q_{31}}{r_{31}^3}, \frac{m_2 m_3 Q_{23}}{r_{23}^3} \right), \end{aligned}$$

where $r_{ij} = |Q_{ij}|$. (Note that here and in all of the differential equations below, partial derivatives like U_Q are calculated by simply calculating the corresponding real partial derivatives and converting the resulting real vector or covector to complex notation; no complex differentiations are involved.) Differential equations for the three-body problem reduced by translations are obtained by restricting Q to \mathfrak{W} . Then Q remains in \mathfrak{W} under the flow. Moreover, covectors P, P' , which are initially equivalent under translation remain so.

Since the symmetry group \mathbb{C}^* acts only on the momenta P_{ij} , the reduced phase space is the eight-dimensional space $\mathfrak{W} \times (\mathbb{C}^{3*}/\mathbb{C}^*) \simeq \mathfrak{W} \times \text{im } L^* = \mathfrak{W} \times \mathfrak{V}$. This can be identified with the cotangent bundle $T^*\mathfrak{W} = \mathfrak{W} \times \mathfrak{W}^*$ as follows. Let $P \in \mathbb{C}^{3*}$. Then $P|_{\mathfrak{W}} \in \mathfrak{W}^*$ and two covectors $P, P' \in \mathbb{C}^{3*}$ have the same restriction to \mathfrak{W} if they differ by an element of $\ker L^*$; that is, if they are equivalent under the symmetry group.

So far we have not really accomplished any ‘‘reduction’’ since there are still twelve (Q, P) variables. Essentially, we have traded the constraint

$$p_{\text{tot}} = p_1 + p_2 + p_3 = 0$$

and the translation symmetry in q for the constraint $Q_{\text{tot}} = Q_{12} + Q_{31} + Q_{23} = 0$ and translation symmetry in P . We will see below that the use of the Q_{ij} is advantageous for regularizing double collisions. A genuine reduction of dimension can be easily achieved by introducing a basis for \mathfrak{W} . Moreover, this can be accomplished in several ways as we will see in [Section 3.4](#) below. But one virtue of [\(7\)](#) is that it avoids making a choice of parametrization and thereby preserves the symmetry of the problem under permutations of the masses.

3.3. Mass metrics and the kinetic energy. The potential energy $U(Q)$ of [\(7\)](#) is particularly simple, but the kinetic energy $K(P)$ seems less natural. In this section we will see that it is related by duality to a Hermitian metric which will play an important role later on.

Define a Hermitian *mass metric* on \mathbb{C}^3 by

$$(14) \quad \langle V, W \rangle = \frac{1}{m} (m_1 m_2 \bar{V}_{12}^T W_{12} + m_3 m_1 \bar{V}_{31}^T W_{31} + m_2 m_3 \bar{V}_{23}^T W_{23}).$$

The corresponding norm is given by

$$(15) \quad |Q|^2 = \frac{1}{m} (m_1 m_2 |Q_{12}|^2 + m_3 m_1 |Q_{31}|^2 + m_2 m_3 |Q_{23}|^2).$$

The mass norm

$$r = |Q| = \sqrt{\langle Q, Q \rangle}$$

provides a natural measure of the size of a configuration $Q = (Q_{12}, Q_{31}, Q_{23}) \in \mathbb{C}^3$. In particular, $r = 0$ represent triple collision. There is a *dual mass metric* on \mathbb{C}^{3*} given by

$$(16) \quad \langle P, R \rangle = m \left(\frac{\bar{P}_{12}^T R_{12}}{m_1 m_2} + \frac{\bar{P}_{31}^T R_{31}}{m_3 m_1} + \frac{\bar{P}_{23}^T R_{23}}{m_2 m_3} \right),$$

with squared norm

$$(17) \quad |P|^2 = m \left(\frac{|P_{12}|^2}{m_1 m_2} + \frac{|P_{31}|^2}{m_3 m_1} + \frac{|P_{23}|^2}{m_2 m_3} \right).$$

Note: Altogether we have three interpretations of $\langle \cdot, \cdot \rangle$ depending on whether the arguments are two vectors (14), two covectors (16), or a vector and a covector, (10). All three pairings are Hermitian, being complex-linear in the second argument and antilinear in the first.

Introduce the notation $\mathcal{W}_0 = \mathcal{W} \setminus 0$ (and a similar notation for any vector space). If $Q \in \mathcal{W}_0$ then it is easy to check that the vectors Q, N, T form a Hermitian-orthogonal complex basis for $T_Q \mathbb{C}^3$ with respect to the Hermitian mass metric, where

$$(18) \quad \begin{aligned} Q &= (Q_{12}, Q_{31}, Q_{23}), & N &= (m_3, m_2, m_1), \\ T &= \left(\frac{\bar{Q}_{31}}{m_2} - \frac{\bar{Q}_{23}}{m_1}, \frac{\bar{Q}_{23}}{m_1} - \frac{\bar{Q}_{12}}{m_3}, \frac{\bar{Q}_{12}}{m_3} - \frac{\bar{Q}_{31}}{m_2} \right). \end{aligned}$$

Q is a radial vector and N, T are, respectively, normal and tangent to \mathcal{W} . Clearly $\{Q, T\}$ is a basis for \mathcal{W} .

The next lemma shows the relationship between the kinetic energy and the dual of the mass metric.

Remark on terminology. A nondegenerate quadratic form on a vector space, or on the fibers of a vector bundle, determines uniquely a quadratic form on the dual vector space, or on the fibers of the dual vector bundle. We refer to this dual quadratic form as either the “cometric” or the “dual norm”.

Lemma 6. *The kinetic energy satisfies*

$$(19) \quad K(P) = \frac{1}{2} \frac{|\langle P, Q \rangle|^2}{|Q|^2} + \frac{1}{2} \frac{|\langle P, T \rangle|^2}{|T|^2} = \frac{1}{2} |P|^2 - \frac{1}{2} \frac{|\langle P, N \rangle|^2}{|N|^2} = \frac{1}{2} |\pi_{\mathcal{W}}^* P|^2,$$

where $|P|$ is the dual mass norm and where $\pi_{\mathcal{W}} : \mathbb{C}^3 \rightarrow \mathbb{C}^3$ is orthogonal projection onto \mathcal{W} with respect to the mass metric.

Moreover, $K(P)$ can be characterized as one-half of the unique translation-invariant quadratic form on $T_Q^* \mathbb{C}^3$ representing the dual of the restriction of the mass norm to $T_Q \mathcal{W}$.

Proof. A direct computation shows that

$$|P|^2 - \frac{|\langle P, N \rangle|^2}{|N|^2} = \frac{|P_{12} - P_{31}|^2}{2m_1} + \frac{|P_{23} - P_{12}|^2}{2m_2} + \frac{|P_{31} - P_{23}|^2}{2m_3} = 2K(P).$$

On the other hand, dual norms, or cometrics, can be characterized by the property that for any orthogonal basis $\{Q, N, T\}$,

$$|P|^2 = \frac{|\langle P, Q \rangle|^2}{|Q|^2} + \frac{|\langle P, N \rangle|^2}{|N|^2} + \frac{|\langle P, T \rangle|^2}{|T|^2}.$$

Hence

$$2K(P) = |P|^2 - \frac{|\langle P, N \rangle|^2}{|N|^2} = \frac{|\langle P, Q \rangle|^2}{|Q|^2} + \frac{|\langle P, T \rangle|^2}{|T|^2},$$

and this is also the formula for $|P \circ \pi_{\mathcal{W}}|^2$.

If we view $T_Q^* \mathcal{W}$ as the quotient space of $T_Q^* \mathbb{C}^3$ under momentum translations, then any norm on $T_Q^* \mathcal{W}$ is represented by a unique translation-invariant quadratic form on $T_Q^* \mathbb{C}^3$. In particular, this applies to the dual norm of the restriction of the mass norm to $T_Q \mathcal{W}$. Since $\{Q, T\}$ is an orthogonal basis for $T_Q \mathcal{W}$ with respect to the mass metric, this “lift” of the dual norm will be given by

$$\frac{|\langle P, Q \rangle|^2}{|Q|^2} + \frac{|\langle P, T \rangle|^2}{|T|^2} = 2K(P). \quad \square$$

3.4. Parametrizing \mathcal{W} . Let $e_1 = (a_{12}, a_{31}, a_{23})$, $e_2 = (b_{12}, b_{31}, b_{23}) \in \mathcal{W}$ be any complex basis for \mathcal{W} . The corresponding coordinate map is

$$f : \mathbb{C}^2 \rightarrow \mathcal{W} \subset \mathbb{C}^3, \quad f(\xi_1, \xi_2) = \xi_1 e_1 + \xi_2 e_2 \quad \text{or} \quad Q_{ij} = \xi_1 a_{ij} + \xi_2 b_{ij},$$

where $\xi = (\xi_1, \xi_2) \in \mathbb{C}^2$ are the new coordinates.

Extend f to a map $F : T^* \mathbb{C}^2 \rightarrow \mathcal{W} \times \mathbb{C}^{3*}$ by letting $P \in \mathbb{C}^{3*}$ be any solution to the equations $\langle P, e_1 \rangle = \bar{\eta}_1$, $\langle P, e_2 \rangle = \bar{\eta}_2$, where $\eta = (\eta_1, \eta_2) \in \mathbb{C}^{2*}$ is the dual momentum to ξ and N is the normal vector to \mathcal{W} from (18). Any two solutions will differ by a momentum translation, which will not affect the computations below.

This definition makes F partially symplectic, where the symplectic structure on $T^*\mathbb{C}^2$ derives from the canonical one-form

$$\theta = \text{re}\langle \eta, \xi \rangle = \text{re}(\bar{\eta}_1 \xi_1 + \bar{\eta}_2 \xi_2).$$

To find the new Hamiltonian, note that the pull-back of the Hermitian mass metric is

$$\langle \xi, \xi' \rangle = \bar{\xi}^T G \xi', \quad \text{with } G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}, \quad g_{ij} = \langle e_i, e_j \rangle.$$

Clearly this can be viewed as the pull-back of the restriction of the mass metric to \mathcal{W} . The dual of this metric is

$$\langle \eta, \eta' \rangle = \bar{\xi}^T G \xi', \quad \text{with } G^{-1} = \frac{1}{g} \begin{bmatrix} g_{22} & -g_{21} \\ -g_{12} & g_{11} \end{bmatrix}, \quad g = \det G.$$

It follows from [Lemma 6](#) and the fact that the momenta also transform as pull-backs that the kinetic energy will be one-half of the dual norm.

The Hamiltonian becomes

$$(20) \quad H(\xi, \eta) = \frac{1}{2} \bar{\eta}^T G^{-1} \eta - U(\xi),$$

where

$$U(\xi) = \frac{m_1 m_2}{\rho_{12}} + \frac{m_1 m_3}{\rho_{31}} + \frac{m_2 m_3}{\rho_{23}}, \quad \rho_{ij} = |Q_{ij}| = |a_{ij} \xi_1 + b_{ij} \xi_2|.$$

Example 7 (heliocentric coordinates). One can form such a parametrization of \mathcal{W} by choosing one of the masses, say m_1 , to play the role of the origin. Set $Q_{12} = -\xi_1$, $Q_{31} = \xi_2$, $Q_{23} = \xi_1 - \xi_2$ so that $\xi_1, \xi_2 \in \mathbb{C}$ are the coordinates of m_2, m_3 relative to m_1 . The corresponding basis for \mathcal{W} $e_1 = (-1, 0, 1)$, $e_2 = (0, 1, -1)$, and the momenta $\bar{\eta}_i = \langle P, e_i \rangle$ satisfy $\eta_1 = P_{23} - P_{12}$, $\eta_2 = P_{31} - P_{23}$. For example, we can choose $P_{12} = -\eta_1$, $P_{31} = \eta_2$, $P_{23} = 0$. Substituting into H_{red} gives the familiar Hamiltonian

$$H(\xi, \eta) = \frac{|\eta_1 + \eta_2|^2}{2m_1} + \frac{|\eta_1|^2}{2m_2} + \frac{|\eta_2|^2}{2m_3} - \frac{m_1 m_2}{|\xi_1|} - \frac{m_1 m_3}{|\xi_2|} - \frac{m_2 m_3}{|\xi_1 - \xi_2|}.$$

Example 8 (Jacobi coordinates). Alternatively one can introduce Jacobi coordinates ξ_1, ξ_2 by setting

$$Q_{12} = -\xi_1, \quad Q_{31} = \xi_2 + v_2 \xi_1, \quad Q_{23} = -\xi_2 + v_1 \xi_1, \quad v_i = \frac{m_i}{m_1 + m_2}.$$

This corresponds to the orthogonal basis $e_1 = (-1, v_2, v_1)$, $e_2 = (0, 1, -1)$, and we have mass metric

$$G = \begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix}, \quad \text{with } \mu_1 = \frac{m_1 m_2}{m_1 + m_2}, \quad \mu_2 = \frac{(m_1 + m_2) m_3}{m}.$$

The momenta satisfy $\eta_1 = -P_{12} + \nu_2 P_{31} + \nu_1 P_{23}$, $\eta_2 = P_{31} - P_{23}$, and for an inverse we could choose $P_{12} = 0$, $P_{31} = \eta_1 + \nu_1 \eta_2$, $P_{23} = \eta_1 - \nu_2 \eta_2$. From (20) we get the equally familiar Hamiltonian

$$H(\xi, \eta) = \frac{|\eta_1|^2}{2\mu_1} + \frac{|\eta_2|^2}{2\mu_2} - \frac{m_1 m_2}{|\xi_1|} - \frac{m_1 m_3}{|\xi_2 + \nu_2 \xi_1|} - \frac{m_2 m_3}{|\xi_2 - \nu_1 \xi_1|}.$$

4. Spherical-homogeneous coordinates

The Hamiltonian $H_{\text{rel}}(Q, P)$ of (7), representing the translation-reduced planar three-body problem, has further symmetries. The potential function $U(Q)$ is symmetric under simultaneous rotation of the Q_{ij} in \mathbb{C} and is also homogeneous of degree -1 with respect to scaling. In this section we exploit the scaling symmetry by converting the system to spherical coordinates. This will be useful later when we blow-up the triple collision singularity.

We use the mass norm $r = |Q|$ as a measure of the size of a configuration $Q = (Q_{12}, Q_{31}, Q_{23}) \in \mathbb{C}^3$. In particular, $r = 0$ represent triple collision. For $Q \in \mathbb{C}_0^3$ we want to define a spherical variable $X \in \mathcal{S}^5$ to describe the normalized configuration. However, instead of using the unit sphere $\mathcal{S}^5 = \{X \in \mathbb{C}^3 : |X| = 1\}$ we will view the sphere as the quotient space of \mathbb{C}_0^3 under scaling by positive real numbers. This gives a convenient way to work globally on \mathcal{S}^5 . We will take a similar approach when working with the complex projective space $\mathbb{C}\mathbb{P}^2$ in the next section.

Let $M = T^*\mathbb{C}_0^3 \simeq \mathbb{C}_0^3 \times \mathbb{C}^{3*}$ with the standard symplectic structure. Let $G = \mathbb{R}^+$ be the group of positive real numbers and let G act on M by $k \cdot (X, Y) = (kX, Y/k)$, where $X \in \mathbb{C}_0^3$, $Y \in \mathbb{C}^{3*}$, $k > 0$. We will use the notation $[X]$, $[X, Y]$ to denote equivalence classes under scaling. In other words, two vectors $X, X' \in \mathbb{C}_0^3$ are equivalent, denoted $X' \sim X$, if $X' = kX$ for some $k > 0$. Similarly $(X', Y') \sim (X, Y)$ if $X' = kX$, $Y' = Y/k$ for some $k > 0$.

The momentum map for this group action is given by $S(X, Y) = \text{re}\langle Y, X \rangle$, where the angle bracket denotes the Hermitian evaluation pairing (10). Fixing this scaling-momentum to be $\text{re}\langle Y, X \rangle = 0$ and passing to the quotient space we get a reduced symplectic manifold, which can be identified with the cotangent bundle $T^*\mathcal{S}^5$. This is a special case of cotangent bundle reduction at zero momentum, as described in Theorem 3. Introduce the notation

$$T_{\text{sph}}^*\mathbb{C}^3 = S^{-1}(0) = \{(X, Y) \in T^*\mathbb{C}_0^3 : \text{re}\langle Y, X \rangle = 0\}.$$

Then we have $T_{\text{sph}}^*\mathbb{C}^3/\mathbb{R}^+ \simeq T^*\mathcal{S}^5$.

We are going to pass from the relative configuration variable $Q \in \mathbb{C}_0^3$ to a size variable r and a homogeneous variable $X \in \mathbb{C}_0^3$.

Definition 9. If $r = |Q|$ and $[X] = [Q]$, we say that $(r, X) \in \mathbb{R}^+ \times \mathbb{C}_0^3$ are spherical-homogeneous coordinates for the configuration $Q \in \mathbb{C}_0^3$

X will be defined only up to a positive real factor and will be viewed as representing a point of S^5 . We can use Q itself as a homogeneous representative of the corresponding point in S^5 . Hence we define a spherical-homogeneous coordinate map

$$f : \mathbb{C}_0^3 \rightarrow \mathbb{R}^+ \times \mathbb{C}_0^3 \quad r = |Q|, \quad X = Q.$$

Extend $f(Q)$ to a map $F(Q, P)$, $F : T^*\mathbb{C}_0^3 \rightarrow T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}^3$ by setting

$$F : \quad p_r = \frac{\text{re}\langle P, Q \rangle}{|Q|}, \quad Y = P - \frac{\text{re}\langle P, Q \rangle}{|Q|^2} Q^*.$$

Here $p_r \in \mathbb{R}^*$, $Y \in \mathbb{C}^{3*}$ are the conjugate momentum variables to r , X and Q^* is the dual covector to Q with respect to the mass metric. By definition, this means the unique covector in \mathbb{C}^{3*} such that $\langle Q^*, V \rangle = \langle Q, V \rangle$, where the first angle bracket is the evaluation pairing and the second is the mass metric. We find

$$(21) \quad Q^* = \frac{1}{m} (m_1 m_2 Q_{12}, m_1 m_3 Q_{31}, m_2 m_3 Q_{23}) \in \mathbb{C}^{3*}.$$

A pseudoinverse $G(r, p_r, X, Y)$, $G : T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}^3 \rightarrow T^*\mathbb{C}_0^3$ to F is given by

$$(22) \quad G : \quad Q = \frac{rX}{|X|}, \quad P = \frac{Pr}{|X|} X^* + \frac{|X|}{r} Y.$$

We have $G \circ F = \text{id}$ and

$$F \circ G(r, p_r, X, Y) = (r, p_r, kX, Y/k), \quad \text{where } k = \frac{r}{|X|}.$$

Hence $f \circ G = \text{id} \text{ mod } \mathbb{R}^+$.

To check that F, G are partially symplectic, compute the pull-backs of the canonical one-forms

$$(23) \quad \theta = p_r dr + \text{re}(\bar{Y}_{12} dX_{12} + \bar{Y}_{31} dX_{31} + \bar{Y}_{23} dX_{23})$$

and Θ from (12). We find $G^*\theta = \Theta$ while $F^*\Theta = \theta + \dots$, where the omitted terms are divisible by $\text{re}\langle Y, X \rangle$. Hence the maps preserve the restricted symplectic forms as required.

The spherical-homogeneous Hamiltonian is $H_{\text{sph}} = H_{\text{rel}} \circ G$. Using the formula for Q in (22), the potential $U(Q)$ becomes $U_{\text{sph}}(r, X) = (1/r)V(X)$, where

$$(24) \quad V(X) = |X| U(X) = |X| \left(\frac{m_1 m_2}{|X_{12}|} + \frac{m_3 m_1}{|X_{31}|} + \frac{m_2 m_3}{|X_{23}|} \right).$$

Note that V is invariant with respect to scaling of X so it determines a well-defined function, $V : S^5 \rightarrow \mathbb{R}$, which we will sometimes write as $V([X])$.

The kinetic energy is $K_{\text{sph}} = K(P)$, where P is given by (22). It follows from Lemma 6 that the two terms in (22) are orthogonal with respect to the quadratic form K . To see this, note that they are orthogonal with respect to the dual mass metric since $\langle Y, X^* \rangle = \langle Y, X \rangle = 0$. Since $X \in \mathfrak{W}$ we have

$$\langle Y \circ \pi_{\mathfrak{W}}, X^* \circ \pi_{\mathfrak{W}} \rangle = \langle Y \circ \pi_{\mathfrak{W}}, \pi_{\mathfrak{W}} X \rangle = \langle Y, X \rangle = 0,$$

so $X^* \circ \pi_{\mathfrak{W}}$ and $Y \circ \pi_{\mathfrak{W}}$ are still orthogonal. Evaluating K separately on the two terms of (22), we find

$$(25) \quad K_{\text{sph}} = \frac{1}{2} p_r^2 + \frac{|X|^2}{r^2} K(Y),$$

and so the spherical-homogeneous Hamiltonian is

$$(26) \quad H_{\text{sph}}(r, p_r, X, Y) = \frac{1}{2} p_r^2 + \frac{|X|^2}{r^2} K(Y) - \frac{1}{r} V(|X|).$$

Theorem 10. *The Hamiltonian flow of H_{sph} on $T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ has invariant submanifold $\{\text{re}\langle Y, X \rangle = 0\}$ and the quotient of the restricted flow by the scaling symmetry is equivalent to the Hamiltonian flow of H_{rel} on $T^*\mathbb{C}_0^3$. This submanifold contains a codimension 2 invariant submanifold $\{\text{re}\langle Y, X \rangle = 0, X_{12} + X_{31} + X_{23} = 0\}$ for which the quotient of the restricted flow by the symmetry of scaling and translations of the Y_{ij} is conjugate to the flow of the zero total momentum three-body problem reduced by translations.*

Proof. For the first part we apply Theorem 1 with $M_1 = T^*\mathbb{C}_0^3$, $M_2 = T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ and symmetry groups $G_1 = \{\text{id}\}$ and $G_2 = \mathbb{R}^+$. The momentum level is

$$S(X, Y) = \text{re}\langle Y, X \rangle = 0.$$

It was shown above that the maps F, G between $T^*\mathbb{C}_0^3$ and $S^{-1}(0)$ are partially symplectic pseudoinverses.

For the second part we change the groups to be $G_1 = \mathbb{C}^*$ and G_2 is a semidirect product of the scaling group \mathbb{R}^+ and the momentum translation group \mathbb{C}^* with group multiplication $(k_2, c_2) \cdot (k_1, c_1) = (k_2 k_1, c_1/k_2 + c_2)$, where $(k_i, c_i) \in \mathbb{R}^+ \times \mathbb{C}^*$. The momentum levels are $\{Q_{\text{tot}} = 0\}$ and $\{X_{\text{tot}} = 0, \text{re}\langle Y, X \rangle = 0\}$, respectively, and these are fixed by the actions of the groups. The maps F, G restrict to maps between these level sets and the restrictions are partially symplectic pseudoinverses. \square

If we use the formula $K(Y) = \frac{1}{2} \bar{Y}^T B Y$, with B from (9), we find that Hamilton's equations for H_{sph} are

$$(27) \quad \begin{aligned} \dot{r} &= p_r, & \dot{p}_r &= \frac{2|X|^2 K(Y)}{r^3} - \frac{1}{r^2} V(X), \\ \dot{X} &= \frac{|X|^2}{r^2} B Y, & \dot{Y} &= \frac{1}{r} D V(X) - \frac{2K(Y)}{r^2} X. \end{aligned}$$

The quotient space of $T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}_0^3$ mentioned in [Theorem 10](#) is diffeomorphic to $T^*\mathbb{R}^+ \times T^*\mathcal{S}^5$ (by simply thinking of X, Y as homogeneous coordinates for $[X, Y] \in T^*\mathcal{S}^5$). The quotient space of $T^*\mathbb{R}^+ \times T_{\text{sph}, \mathcal{W}}^*\mathbb{C}_0^3$ is diffeomorphic to $T^*\mathbb{R}^+ \times T^*S(\mathcal{W})$, where $S(\mathcal{W}) = \mathcal{W} \cap \mathcal{S}^5$ is diffeomorphic to \mathcal{S}^3 . Hence the reduced space is eight-dimensional as before. The reduced flow is just the translation-reduced three-body problem in spherical coordinates.

At this point, instead of reducing the number of dimensions, we have actually increased it from twelve to fourteen. The value of the present formulation lies in the fact that it has been put in a form where double collisions can be easily regularized and the triple collision easily blown-up without destroying the symmetry among the masses. As in the previous section, one could explicitly realize the reduction to eight dimensions by parametrizing the subspace \mathcal{W} . However we will not do this here.

5. Reduction by rotations: the shape sphere

Next we form the quotient by rotations. Since we are using complex coordinates, the combined action of scaling Q by a real factor $r > 0$ and rotating Q by an angle θ is represented by $Q \mapsto kQ$, where $k = re^{i\theta} \in \mathbb{C}_0 = \mathbb{C} \setminus 0$, the space of nonzero complex numbers. A point in the resulting quotient space represents the size and shape of a configuration.

5.1. Projective-homogeneous coordinates. As before we will measure the size by $r = |Q|$. To represent the shape, we project $Q \in \mathbb{C}_0^3$ to the quotient of \mathbb{C}_0^3 by the action of \mathbb{C}_0 . This quotient space is the complex projective plane $\mathbb{P}(\mathbb{C}^3) = \mathbb{C}\mathbb{P}^2$. Homogeneous coordinates will provide a way to work globally on the projective plane, just as they did for the sphere \mathcal{S}^5 in the last section. For $X \in \mathbb{C}_0^3$ let $[X] \in \mathbb{C}\mathbb{P}^2$ denote the corresponding element of the projective plane, that is, the equivalence class of X under the relation that $X \sim Q$ if $X = kQ$ for some $k \in \mathbb{C}, k \neq 0$. (Thus the square bracket will now mean a projective point rather than a spherical one.)

Definition 11. (r, X) are a pair of projective-homogeneous coordinates for $Q \in \mathbb{C}_0^3$ if $r = |Q|$ and $[X] = [Q] \in \mathbb{C}\mathbb{P}^2$.

X is defined only up to a nonzero complex factor. We can take $X = Q$ itself to define the projective-homogeneous coordinate map

$$f : \mathbb{C}_0^3 \rightarrow \mathbb{R}^+ \times \mathbb{C}_0^3, \quad r = |Q|, \quad X = Q.$$

Remark. Despite the fact that spherical-homogeneous coordinates and projective-homogeneous coordinates are both denoted (r, X) , there are differences between the two coordinate systems. Spherical-homogeneous coordinates represent points

in $\mathbb{C}_0^3 \simeq \mathbb{R}^+ \times \mathcal{S}^5$, whereas projective-homogeneous coordinates represent points in the quotient space $(\mathbb{C}_0^3)/\mathcal{S}^1 \simeq \mathbb{R}^+ \times \mathbb{C}\mathbb{P}^2$.

If we include the origin and form the quotient space under rotations we have $\mathbb{C}^3/\mathcal{S}^1 = \text{Cone}(\mathbb{C}\mathbb{P}^2)$, the cone over $\mathbb{C}\mathbb{P}^2$, where the cone point corresponds to total collision $0 \in \mathbb{C}^3$. For any topological space X , we can form the space $\text{Cone}(X)$ which has a distinguished cone point $*$ and $\text{Cone}(X) \setminus * = \mathbb{R}^+ \times X$. In this case, the cone is not a smooth manifold.

The equivalence class $[X] = [Q] \in \mathbb{C}\mathbb{P}^2$ represents the shape of a three-body configuration only if $Q \in \mathcal{W}$. Restricting to such Q we get $[Q] \in \mathbb{P}(\mathcal{W})$, where $\mathbb{P}(\mathcal{W})$ is the projective space of the subspace $\mathcal{W} \subset \mathbb{C}^3$. Since \mathcal{W} is a two-dimensional complex subspace, $\mathbb{P}(\mathcal{W})$ is a projective line, that is, $\mathbb{P}(\mathcal{W}) \simeq \mathbb{C}\mathbb{P}^1 \simeq \mathcal{S}^2$. $\mathbb{P}(\mathcal{W})$ will be called the *shape sphere*.

Any function on our original configuration space that is invariant under translation, rotation, and scaling induces a function on the shape sphere, the most important example being our homogenized potential

$$V(X) = |X|U(X) : \mathbb{P}\mathcal{W} \rightarrow \mathbb{R}.$$

We will also use homogeneous momentum variables. A pair

$$(X, Y) \in T^*\mathbb{C}_0^3 \simeq \mathbb{C}_0^3 \times \mathbb{C}^{3*}$$

will represent a point of $T^*\mathbb{C}\mathbb{P}^2$. Let $G = \mathbb{C}_0$ be the group of nonzero complex numbers and let G act on $T^*\mathbb{C}_0^3$ by $k \cdot (X, Y) = (kX, Y/\bar{k})$. We will use the notation $[X, Y]$ to denote equivalence classes under scaling. In other words, $(X', Y') \sim (X, Y)$ if $X' = kX$, $Y' = Y/\bar{k}$ for some nonzero $k \in \mathbb{C}$. The momentum map for this group action is given by the Hermitian evaluation pairing $\sigma(X, Y) = \langle Y, X \rangle \in \mathbb{C}$. The real part of the complex number $\sigma(X, Y)$ is the real scaling-momentum $S(X, Y)$ (which we want to be zero as in the last section). On the other hand, from (11) we see that $\text{im } \sigma(X, Y) = -i\mu$, where μ is the angular momentum.

If we fix the complex scaling-momentum to be $\langle Y, X \rangle = 0$ and pass to the quotient space, then as in [Theorem 3](#) we get a reduced symplectic manifold, which is naturally identified with the cotangent bundle $T^*\mathbb{C}\mathbb{P}^2$ with its natural symplectic structure. Introduce the notation

$$T_{\text{pr}}^*\mathbb{C}^3 = \sigma^{-1}(0) = \{(X, Y) \in T^*\mathbb{C}_0^3 : \langle Y, X \rangle = 0\}.$$

Then we have

$$T_{\text{pr}}^*\mathbb{C}^3/\mathbb{C}_0 \simeq T^*\mathbb{C}\mathbb{P}^2.$$

If, on the other hand, we fix the complex scaling-momentum to be $\langle Y, X \rangle = -i\mu$ and pass to the quotient space we still get a reduced symplectic manifold, which

can be identified with the cotangent bundle $T^*\mathbb{C}\mathbb{P}^2$ but with a twisted symplectic structure, as described in [Theorem 4](#). More about this below.

To get a system equivalent to the reduced three-body problem we will also need to include the radial variables. Restrict X to \mathcal{W} and quotient by the action of the group \mathbb{C} of translations in Y -momentum space. Let $M = T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ with coordinates (r, p_r, X, Y) and let $G = \mathbb{C}_0 \times \mathbb{C}$ act by

$$(k, c) \cdot (r, p_r, X, Y) = (r, p_r, kX, c \cdot (Y/\bar{k})), \quad \text{with } c \cdot Y = (Y_{12} + c, Y_{31} + c, Y_{23} + c).$$

Fixing the momentum level $J(X, Y) = (\sigma(X, Y), X_{\text{tot}}) = (-i\mu, 0) \in \mathbb{C}^2$ and passing to the quotient space gives the reduced phase space

$$P = \{(r, p_r, X, Y) : \langle Y, X \rangle = -i\mu, X_{12} + X_{31} + X_{23} = 0\} / G$$

of real dimension $\dim P = 14 - 4 - 4 = 6$ as expected. In fact we have

$$P \simeq T^*\mathbb{R}^+ \times T^*\mathbb{P}(\mathcal{W}) \simeq T^*\mathbb{R}^+ \times T^*\mathcal{S}^2.$$

We still need to find the reduced Hamiltonian and show that the reduced Hamiltonian system is equivalent to the reduced three-body problem. This is easy to do starting from the spherical Hamiltonian in the last section. Indeed, the passage from the spherical-homogeneous variables $(r, p_r, X, Y) \in T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ to the corresponding projective-homogeneous ones is just given by the identity map. The new feature here is that the symmetry group is enlarged from $\mathbb{R}^+ \times \mathbb{C}^* \simeq \mathbb{R}^+ \times \mathbb{C}$ to $\mathbb{C}_0 \times \mathbb{C}$. Then we have the following extension of [Theorem 10](#):

Theorem 12. *The Hamiltonian flow of H_{sph} on $T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ has an invariant set where $\langle Y, X \rangle = -i\mu$. The quotient of the restricted flow by the complex scaling symmetry is equivalent to the Hamiltonian flow of H_{rel} on $T^*\mathbb{C}_0^3/\mathcal{S}^1$. There is another invariant set where $\langle Y, X \rangle = -i\mu$ and $X_{12} + X_{31} + X_{23} = 0$ and the quotient of the restricted flow by the complex scaling symmetry and by translations of the Y_{ij} is conjugate to the flow of the three-body problem with zero total momentum and angular momentum μ , reduced by translations and rotations.*

Proof. The maps F and G as in the proof of [Theorem 10](#) restrict to maps of the μ angular momentum levels. They are still partially symplectic pseudoinverses. \square

The next step is to use a momentum shift map to pull-back the problem to the zero-angular-momentum level. This expresses all of the reduced problems on the same phase space and makes the role of the angular momentum constant explicit. Let

$$(28) \quad \Phi_\mu(r, p_r, X, Z) = (r, p_r, X, Y), \quad Y = Z + \mu\Gamma(X), \quad \Gamma(X) = \frac{iX^*}{|X|^2},$$

where

$$X^* = \frac{1}{m}(m_1m_2X_{12}, m_3m_1X_{31}, m_2m_3X_{23}) \in \mathbb{C}^{3*}.$$

Note that $\Phi_\mu : J^{-1}(0, 0) \rightarrow J^{-1}(-i\mu, 0)$, since if $\langle Z, X \rangle = 0$ we have

$$\operatorname{im}\langle Y, X \rangle = \operatorname{im}\left\langle i\mu \frac{X^*}{|X|^2}, X \right\rangle = -\mu \operatorname{re}\left\langle \frac{X^*}{|X|^2}, X \right\rangle = -\mu.$$

Composing H_{sph} with Φ_μ we get a Hamiltonian

$$(29) \quad H_\mu(r, p_r, X, Z) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} \right) + \frac{|X|^2}{r^2} K(Z) - \frac{1}{r} V(|X|).$$

To verify this we need to show that the kinetic energy can be written

$$(30) \quad K_\mu = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} \right) + \frac{|X|^2}{r^2} K(Z).$$

This decomposition follows from an orthogonality argument based on [Lemma 6](#). Namely, the vectors iX and Z are orthogonal with respect to the mass metric and the first one lies in ${}^{\mathcal{W}}$. Then, as in the last section, [Lemma 6](#) shows that they are orthogonal with respect to the quadratic form K and so $K(Y) = K(\mu\Gamma(X)) + K(Z)$. $K(\mu\Gamma(X))$ gives μ^2 -term in K_μ .

[Equation \(30\)](#) gives a decomposition of the kinetic energy into radial and angular parts and a third term which can be viewed as the kinetic energy due to changes in the shape of the configuration. Some authors call this decomposition of kinetic energy, or the consequent orthogonal decomposition of velocities the ‘‘Saari decomposition’’. (See [\[Saari 1984\]](#).) In the next subsection we show how this last shape term can be understood in terms of the Fubini–Study metric on the shape sphere.

5.2. Fubini–Study metrics and the shape kinetic energy. Using a complex orthogonal basis, we give a simple decomposition of the dual mass metric, which leads to deeper insights into the kinetic energy decomposition [\(30\)](#). Since the shape sphere has complex dimension one, there are some very simple formulas for the shape term of this decomposition.

To describe the Fubini–Study metric (also called the Kähler metric), let \mathcal{V} denote any complex vector space and let $\langle V, W \rangle$ be any Hermitian metric on \mathcal{V} . If $X \in \mathcal{V}_0 = \mathcal{V} \setminus 0$ then the corresponding *Fubini–Study* metric on $T_X\mathcal{V}$ is

$$(31) \quad \langle V, W \rangle_{\text{FS}} = \frac{\langle V, W \rangle \langle X, X \rangle - \langle V, X \rangle \langle X, W \rangle}{\langle X, X \rangle^2}.$$

As a bilinear form on $T_X\mathcal{V}$, the Fubini–Study ‘‘metric’’ is degenerate with kernel the complex line spanned by the vector X . But it induces a bona fide Hermitian metric on the projective space $\mathbb{P}(\mathcal{V})$.

To see this, let $\pi : \mathcal{V}_0 \rightarrow \mathbb{P}(\mathcal{V})$ denote the projection map: $\pi(X) = [X]$. The tangent map $T\pi : T\mathcal{V}_0 \rightarrow T\mathbb{P}(\mathcal{V})$, $T\pi(X, V) = ([X], D\pi(X)V)$ has the property that $T\pi(X, V) = T\pi(X', V')$ if and only if $X' = kX$ and $V' = kV + lX$ for some

complex numbers $k \neq 0, l$. So it is natural to view the tangent bundle $T\mathbb{P}(\mathcal{V})$ as the set of equivalence classes $[X, V]$ of pairs $(X, V) \in \mathcal{V}_0 \times \mathcal{V}$ under this equivalence relation. It is easy to check that the formula for $\langle \cdot, \cdot \rangle_{\text{FS}}$ is invariant under this equivalence relation and so it gives a well-defined Hermitian metric on $\mathbb{P}(\mathcal{V})$. The real part $\text{re}\langle V, W \rangle_{\text{FS}}$ gives a Riemannian metric on $\mathbb{P}(\mathcal{V})$ and the imaginary part gives a two-form called the *Fubini–Study form*, which will be important later

$$\Omega_{\text{FS}}(V, W) = \text{im}\langle V, W \rangle_{\text{FS}}.$$

Starting with the mass metric on $\mathcal{V} = \mathbb{C}^3$, we get a Fubini–Study metric on $\mathbb{C}\mathbb{P}^2$. However, because of [Lemma 6](#), we will be interested in its restriction to the two-dimensional complex subspace $\mathcal{W} \subset \mathbb{C}^3$, which we denote by $\langle \cdot, \cdot \rangle_{\text{FS}, \mathcal{W}}$, which induces a Hermitian metric on the shape sphere $\mathbb{P}(\mathcal{W})$.

Our goal is to show that the shape kinetic energy is the cometric dual to this Fubini–Study metric on $\mathbb{P}(\mathcal{W})$. (By a “cometric” on a manifold X we mean the fiberwise quadratic form on T^*X that is dual to a Riemannian metric on X .) To this end we will need to describe cometrics on projective space in homogeneous coordinates. We continue to identify $T^*\mathbb{C}\mathbb{P}^2$ with the quotient space of $T^*_{\text{pr}}\mathbb{C}^3 = \{(X, Z) \in \mathbb{C}_0^3 \times \mathbb{C}^{3*} : \langle Z, X \rangle = 0\}$ under the complex scaling symmetry. In the same spirit, the cotangent bundle $T^*\mathbb{P}(\mathcal{W})$ is the quotient space (a symplectic reduced space)

$$T^*\mathbb{P}(\mathcal{W}) \simeq (T^*_{\text{pr}, \mathcal{W}}\mathbb{C}_0^3) / \mathbb{C}_0 \times \mathbb{C},$$

where

$$T^*_{\text{pr}, \mathcal{W}}\mathbb{C}_0^3 = \{(X, Z) \in \mathcal{W} \times \mathbb{C}^{3*} : \langle Z, X \rangle = 0, X \neq 0\}$$

and where the group $\mathbb{C}_0 \times \mathbb{C}$ represents the scaling symmetry and the momentum translation in Z -space. We refer to (X, Z) as homogeneous coordinates on $\mathbb{P}(\mathcal{W})$. The restriction of $Z \in \mathbb{C}^{3*}$ to \mathcal{W} representing a covector in $T^*_{[X]} \mathbb{P}(\mathcal{W})$. Expressed in homogeneous coordinates a cometric on $\mathbb{P}(\mathcal{W})$ is a function of the form $Q(X, Z)$ which is quadratic in Z and invariant under the $\mathbb{C}_0 \times \mathbb{C}$ action.

Theorem 13. *The Fubini–Study cometric $|Z|_{\text{FS}, \mathcal{W}}^2$ at $[X] \in \mathbb{P}(\mathcal{W})$ is related to the kinetic energy (formula (19)) by*

$$\frac{1}{2}|Z|_{\text{FS}, \mathcal{W}}^2 = |X|^2 K(Z).$$

Proof. Substitute (X, Z) for (Q, P) in formula (19). Use $\langle Z, X \rangle = 0$ to get $K(Z) = (1/2|T|^2)\langle Z, T \rangle$. The vector field $T(X)$ appearing in that formula is tangent to \mathcal{W} and orthogonal to X , hence fits the hypothesis of [Lemma 14](#) immediately below. The lemma asserts that we have

$$|Z|_{\text{FS}, \mathcal{W}}^2 = |\langle Z, e(X) \rangle|^2, \quad \text{with } e(X) = \frac{|X|}{|T(X)|} T(X). \quad \square$$

Lemma 14. *Let $T(X)$, $X \in \mathfrak{W}_0$ be a nonzero complex vector field tangent to \mathfrak{W}_0 and normal to X with respect to the Hermitian metric mass metric. Then*

$$e(X) = \frac{|X|}{|T(X)|} T(X)$$

is a unit tangent vector field on \mathfrak{W}_0 with respect to the pulled back Fubini–Study metric $\langle \cdot, \cdot \rangle_{\text{FS}, \mathfrak{W}}$. Moreover

$$(32) \quad \langle V, W \rangle_{\text{FS}, \mathfrak{W}} = \frac{\langle V, e(X) \rangle \langle e(X), W \rangle}{|X|^4}, \quad \text{with } V, W \in \mathfrak{W}/(\mathbb{C}X) \cong T_{[X]} \mathbb{P}(\mathfrak{W}),$$

and the pulled-back cometric is given by the quadratic form

$$(33) \quad |Z|_{\text{FS}, \mathfrak{W}}^2 = |\langle Z, e(X) \rangle|^2, \quad \text{with } Z \in T_{X, \text{pr}}^* \mathbb{C}^3.$$

Proof. Since $T(X)$ is orthogonal to X , (31) gives $|T|_{\text{FS}}^2 = |T|^2/|X|^2$ and so $e(X)$ is a Fubini–Study unit vector at X .

The tangent space $T_X \mathfrak{W}$ has complex dimension two and $\{X, e(X)\}$ is a basis. If we expand $V \in T_X \mathfrak{W}$ as

$$V = \frac{\langle V, X \rangle}{|X|^2} X + \frac{\langle V, T(X) \rangle}{|T(X)|^2} T(X)$$

and similarly for W , then since X is in the kernel of $\langle \cdot, \cdot \rangle_{\text{FS}}$ we get

$$\langle V, W \rangle_{\text{FS}, \mathfrak{W}} = \langle V, W \rangle_{\text{FS}} = \frac{\langle V, T(X) \rangle \langle T(X), W \rangle}{|X|^2 |T(X)|^2} = \frac{\langle V, e(X) \rangle \langle e(X), W \rangle}{|X|^4},$$

as claimed.

Observe that if \mathbb{E} , $\langle \cdot, \cdot \rangle$ is a one-dimensional complex Hermitian vector space with unit vector e then the cometric on \mathbb{E}^* is given by the quadratic form

$$Z \in \mathbb{E}^* \mapsto |\langle Z, e \rangle|^2.$$

From this observation the last formula of the lemma follows. □

Remark. The manifold $\mathbb{P}(\mathfrak{W})$, being a two-sphere, admits no nonvanishing vector field. So how did we just construct a unit vector field $e(X)$ to this two-sphere? We did not! The gadget $e(X)$ is a unit section of the pull-back $f^* T \mathbb{P}(\mathfrak{W})$ of this tangent bundle by the homogenization map $f : \mathfrak{W}_0 \rightarrow \mathbb{P}(\mathfrak{W})$ that sends $X \rightarrow [X]$. This pull-back bundle can be viewed as a subbundle of $T \mathfrak{W}_0$, and hence $e(X)$ is a vector field on \mathfrak{W}_0 .

Using the vector field $T(X)$ of formula (19) (with X substituted for Q), we obtain the Fubini–Study unit tangent vector

$$e(X) = \sqrt{\frac{m_1 m_2 m_3}{m}} \left(\frac{\bar{X}_{31}}{m_2} - \frac{\bar{X}_{23}}{m_1}, \frac{\bar{X}_{23}}{m_1} - \frac{\bar{X}_{12}}{m_3}, \frac{\bar{X}_{12}}{m_3} - \frac{\bar{X}_{31}}{m_2} \right).$$

From this expression we get simple formulas for the Fubini–Study metric and two-form on \mathcal{W} :

$$(34) \quad \langle \cdot, \cdot \rangle_{\text{FS}, \mathcal{W}} = \frac{m_1 m_2 m_3}{m |X|^4} \bar{\sigma} \otimes \sigma, \quad \Omega_{\text{FS}, \mathcal{W}} = \frac{m_1 m_2 m_3}{m |X|^4} \text{im } \bar{\sigma} \otimes \sigma,$$

where the complex-valued one-form σ is given by any of the following formulas

$$(35) \quad \begin{aligned} \sigma &= \langle e, dX \rangle = X_{31} dX_{12} - X_{12} dX_{31} \\ &= X_{12} dX_{23} - X_{23} dX_{12} = X_{23} dX_{31} - X_{31} dX_{23}. \end{aligned}$$

For example, the second formula for σ is obtained by eliminating X_{23} , dX_{23} from $\langle e, dX \rangle$ using the equations $X_{23} = -X_{12} - X_{31}$ and $dX_{23} = -dX_{12} - dX_{31}$. Note that the formulas for σ are independent of the masses. This implies that the Fubini–Study metrics for different masses are all conformal to one another.

Similarly we get a formula for the dual norm and the shape kinetic energy:

$$(36) \quad |X|^2 K(Z) = \frac{1}{2} |Z|_{\text{FS}, \mathcal{W}}^2 = \frac{m |\alpha(Z)|^2}{2m_1 m_2 m_3},$$

where $\alpha(Z)$ is given by any of the following formulas:

$$(37) \quad \begin{aligned} \alpha &= \frac{1}{m} (m_1 m_2 X_{12} (Z_{23} - Z_{31}) + m_3 m_1 X_{31} (Z_{12} - Z_{23}) + m_2 m_3 X_{23} (Z_{31} - Z_{12})) \\ &= \frac{|X|^2 (Z_{31} - Z_{12})}{\bar{X}_{23}} = \frac{|X|^2 (Z_{12} - Z_{23})}{\bar{X}_{31}} = \frac{|X|^2 (Z_{23} - Z_{31})}{\bar{X}_{12}}. \end{aligned}$$

Our identification of the shape kinetic energy with the Fubini–Study cometric gives an alternative formula for the reduced Hamiltonian on $T_{\text{pr}}^* \mathbb{C}^3$

$$(38) \quad H_\mu(r, p_r, X, Z) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} \right) + \frac{1}{2r^2} |Z|_{\text{FS}, \mathcal{W}}^2 - \frac{1}{r} V(X),$$

where $|Z|_{\text{FS}, \mathcal{W}}^2$ is the Fubini–Study cometric on \mathcal{W} .

5.3. Induced symplectic structure and the reduced differential equations. Using the momentum shift map, we have pulled back the Hamiltonian to the reduced Hamiltonian H_μ defined on the zero-angular momentum level $T^* \mathbb{R}^+ \times T_{\text{pr}}^* \mathbb{C}^3$, where

$$T_{\text{pr}}^* \mathbb{C}^3 = \{(X, Z) \in T^* \mathbb{C}^3 : \langle Z, X \rangle = 0\}.$$

However, as described in [Theorem 4](#), there is also an induced symplectic structure on this set which different from the restriction of the standard one. The pull-back of the canonical one-form θ under the momentum shift map [\(28\)](#) is

$$\Phi_\mu^* \theta = p_r dr + \text{re} \langle Z, dX \rangle + \frac{\mu}{|X|^2} \text{im} \langle X^*, dX \rangle = \Theta + \mu \Theta_1$$

with

$$\Theta_1 = \text{im} \frac{\langle X^*, dX \rangle}{|X|^2} = \text{im} \frac{\langle X, dX \rangle}{|X|^2},$$

where we changed the evaluation pairing to the mass metric in the second equation. The modified symplectic form will be $\Omega_\mu = \Omega - \mu d\Theta_1$, where we find

$$(39) \quad d\Theta_1 = 2 \text{im} \frac{\langle dX, dX \rangle |X|^2 - \langle dX, X \rangle \langle X, dX \rangle}{|X|^4} = 2\Omega'_{\text{FS}},$$

where Ω'_{FS} is the Fubini–Study two-form determined by the mass metric on \mathbb{C}^3 (as opposed to its restriction to ${}^{\circ}\mathcal{W}$ as in [Section 5.2](#)). Geometrically, Ω'_{FS} represents the curvature of the circle bundle $S^5 \rightarrow \mathbb{C}\mathbb{P}^2$.

Once we have Ω_μ we calculate Hamilton’s differential equations using the defining equation for Hamiltonian vector fields:

$$(40) \quad (\dot{r}, \dot{p}_r, \dot{X}, \dot{Z}) \lrcorner \Omega_\mu = dH_\mu.$$

The interior product with the standard form gives the usual result:

$$(\dot{r}, \dot{p}_r, \dot{X}, \dot{Z}) \lrcorner \Omega = -\dot{p}_r dr + \dot{r} dp_r - \text{re}\langle \dot{Z}, dX \rangle + \text{re}\langle \dot{X}, dZ \rangle.$$

Since Ω'_{FS} involves only dX , it can be viewed as a two-form on C^3 instead of on phase space. Moreover, it only affects the differential equations for \dot{Z} . Hamilton’s equations read:

$$(41) \quad \dot{r} = H_{\mu, p_r}, \quad \dot{p}_r = -H_{\mu, r}, \quad \dot{X} = H_{\mu, Z}, \quad \dot{Z} = -H_{\mu, X} - 2\mu H_{\mu, Z} \lrcorner \Omega'_{\text{FS}},$$

where H_μ is given by [\(29\)](#). The term involving the Fubini–Study metric will be called the *curvature term*, $T'_{\text{curv}} = -2\mu H_{\mu, Z} \lrcorner \Omega'_{\text{FS}}$.

Lemma 15. *If $X \in {}^{\circ}\mathcal{W}$ and $\langle Z, X \rangle = 0$, then the vector $H_{\mu, Z}$ is in ${}^{\circ}\mathcal{W}$ and $\langle X, H_{\mu, Z} \rangle = 0$. In fact*

$$(42) \quad H_{\mu, Z} = \frac{\overline{\langle Z, e \rangle}}{r^2} e \in {}^{\circ}\mathcal{W},$$

where $e(X)$ is as in [Lemma 14](#).

The curvature term T'_{curv} is equivalent under the translation symmetry in \mathbb{C}^{3*} to

$$(43) \quad T_{\text{curv}} = -\frac{2\mu}{r^2} iZ.$$

Proof. From [\(29\)](#) we have $H_{\mu, Z} = (|X|^2/r^2)DK(Z)$. Note that since $Z \in \mathbb{C}^{3*}$, we have $DK(Z) : \mathbb{C}^{3*} \rightarrow \mathbb{R}$. By duality we can view $DK(Z)$ as a vector in \mathbb{C}^3 . Let $X \in {}^{\circ}\mathcal{W}$. Since $\dot{X} = H_{\mu, Z}$ and ${}^{\circ}\mathcal{W}$ is invariant, we must have $H_{\mu, Z} \in {}^{\circ}\mathcal{W}$. If $\langle Z, X \rangle = 0$ then an orthogonality argument as above shows $K(Z + X^*) = K(Z) + K(X^*)$,

which implies, since K is a quadratic form, that $DK(Z)(X^*) = \langle DK(Z), X \rangle = 0$, as required.

In Section 5.2 we showed that in the subspace $\{Z : \langle Z, X \rangle = 0\}$ we have $|X|^2 K(Z) = \frac{1}{2} |\langle Z, e \rangle|^2$. In fact, we will see that the Z -derivatives of these two functions also agree:

$$(44) \quad |X|^2 DK(Z) = \overline{\langle Z, e \rangle} e.$$

To see that (44) indeed holds, note that differentiation along the subspace shows that they must agree when evaluated on any δZ with $\langle \delta Z, X \rangle = 0$. On the other hand, both sides vanish on the complementary covector $Z' = X^*$. Note that the right hand side was calculated, as always, by converting to real variables, finding the real derivative and then converting back to a complex vector. Equivalently, we expand

$$\frac{1}{2} |\langle Z + \delta Z, e \rangle|^2 = \frac{1}{2} |\langle Z, e \rangle|^2 + \text{re} \langle \delta Z, \overline{\langle Z, e \rangle} e \rangle + \dots$$

for all δZ , showing that the vector in question is the complex representative of the real vector derivative.

To show the equivalence of T'_{curv} and T_{curv} we will show that they agree when restricted to \mathcal{W} . The argument can be based on a kind of Fubini–Study duality. Namely, if $V \in \mathcal{W}$ we will show that

$$(45) \quad \langle H_{\mu,Z}, V \rangle_{\text{FS}} = \frac{1}{r^2} \langle Z, V \rangle,$$

which means that $r^2 H_{\mu,Z}$ is a dual vector to Z with respect to the Fubini–Study metric on \mathcal{W} . To see this note that (44) gives

$$\langle H_{\mu,Z}, V \rangle_{\text{FS}} = \frac{1}{r^2} \frac{\langle \overline{\langle Z, e \rangle} e, V \rangle}{|X|^2} = \frac{\langle Z, e \rangle \langle e, V \rangle}{r^2 |X|^2}.$$

On the other hand any $V \in \mathcal{W}$ is a linear combination

$$V = \frac{\langle X, V \rangle}{|X|^2} X + \frac{\langle e, V \rangle}{|e|^2} e.$$

Since e is a Fubini–Study unit vector, we have $|e| = |X|$ and so

$$\frac{1}{r^2} \langle Z, V \rangle = \frac{\langle Z, e \rangle \langle e, V \rangle}{r^2 |e|^2} = \frac{\langle Z, e \rangle \langle e, V \rangle}{r^2 |X|^2}$$

and (45) holds. From this we can calculate that for any $V \in \mathcal{W}$

$$T'_{\text{curv}}(V) = -2\mu \text{im} \langle H_{\mu,Z}, V \rangle_{\text{FS}} = -\frac{2\mu}{r^2} \text{im} \langle Z, V \rangle = -\frac{2\mu}{r^2} \text{re} \langle iZ, V \rangle.$$

Thus T'_{curv} and T_{curv} agree as real-valued one-forms on \mathcal{W} as claimed. Replacing T'_{curv} by T_{curv} introduces only an irrelevant translation of the momentum Z . \square

Taking this lemma into account we finally get Hamilton's equations for the reduced Hamiltonian in the form

$$(46) \quad \begin{aligned} \dot{r} &= p_r, & \dot{p}_r &= \frac{\mu^2 + |X|^2 2K(Z)}{r^3} - \frac{1}{r^2} V(X), \\ \dot{X} &= \frac{|X|^2}{r^2} DK(Z), & \dot{Z} &= \frac{1}{r} DV(X) - \frac{2K(Z)}{r^2} X - \frac{2\mu}{r^2} iZ. \end{aligned}$$

Applying [Theorem 1](#) to the momentum shift map and remembering [Theorem 12](#), we have:

Theorem 16. *The Hamiltonian flow of H_μ on $T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ has an invariant set $T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^3$, where $\langle Z, X \rangle = 0$ with symplectic structure given by the restriction of the standard form minus $2\mu\Omega_{\text{FS}}$. The quotient of the restricted flow by the complex scaling symmetry is equivalent to the Hamiltonian flow of H on $T^*\mathbb{C}_0^3/\mathcal{S}^1$. There is another invariant set $T^*\mathbb{R}^+ \times T_{\text{pr},\mathcal{W}}^*\mathbb{C}^3$, where $\langle Z, X \rangle = 0$ and $X_{12} + X_{31} + X_{23} = 0$ and the quotient of the restricted flow by the complex scaling symmetry and by translations of the Z_{ij} is conjugate to the flow of the three-body problem with zero total momentum and angular momentum μ , reduced by translations and rotations.*

This Hamiltonian system represents the reduced three-body problem in a way which is convenient for regularization of binary collisions and blow-up of triple collision. However, the phase space is still fourteen-dimensional. Next we describe how to find lower-dimensional representations of the reduced three-body problem by parametrizing the shape sphere in various ways.

5.4. Parametrizing the shape sphere. The shape sphere is the projective space $\mathbb{P}(\mathcal{W})$. As in [Section 3.4](#), choosing a complex basis $\{e_1, e_2\}$ for \mathcal{W} gives a map $f: \mathbb{C}^2 \rightarrow \mathcal{W}$, $X = f(\xi)$. By viewing $X \in \mathcal{W}$ and $\xi \in \mathbb{C}^2$ as homogeneous coordinates we get an induced parametrization of the shape sphere $f_{\text{pr}}: \mathbb{C}\mathbb{P}^1 \rightarrow \mathbb{P}(\mathcal{W})$.

The formulas of [Section 3.4](#) (with (Q, P) replaced by (X, Z)) allow us to find the reduced Hamiltonian for any such basis. If

$$e_1 = (a_{12}, a_{31}, a_{23}), \quad e_2 = (b_{12}, b_{31}, b_{23}) \in \mathcal{W},$$

then we have, as before, $X_{ij} = \xi_1 a_{ij} + \xi_2 b_{ij}$ and $\bar{\eta}_1 = \langle Y, e_1 \rangle$, $\bar{\eta}_2 = \langle Y, e_2 \rangle$. We define a Hermitian mass metric and dual mass metric for ξ, η to be the pull-backs of the metrics for X, Y . The squared norms are

$$|\xi|^2 = \bar{\xi}^T G \xi, \quad |\eta|^2 = \bar{\eta}^T G^{-1} \eta,$$

where G is the matrix with entries $G_{ij} = \langle e_i, e_j \rangle$, and these squared norms represent the mass metric and cometric on \mathcal{W} .

The relation between the cometric and kinetic energy yields the Hamiltonian (see (29), (30) and Theorem 13):

$$(47) \quad H_\mu(r, p_r, \xi, \eta) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} + \frac{|\xi|^2 |\eta|^2}{r^2} \right) - \frac{1}{r} V(\xi),$$

where the shape potential is

$$V(\xi) = |\xi| \left(\frac{m_1 m_2}{\rho_{12}} + \frac{m_1 m_3}{\rho_{31}} + \frac{m_2 m_3}{\rho_{23}} \right), \quad \rho_{ij} = |X_{ij}| = |a_{ij} \xi_1 + b_{ij} \xi_2|.$$

To make the map F of Section 3.4 be partially symplectic we need to alter the standard symplectic form in (ξ, η) -space by subtracting $2\mu F^* \Omega'_{\text{FS}}$. Pulling back the Fubini–Study metric $\langle \cdot, \cdot \rangle_{\text{FS}}$ by f gives the Fubini–Study metric in ξ space

$$\langle \cdot, \cdot \rangle_{\text{FS}} = \frac{\langle d\xi, d\xi \rangle \langle \xi, \xi \rangle - \langle d\xi, \xi \rangle \langle \xi, d\xi \rangle}{\langle \xi, \xi \rangle^2}.$$

With the help of (34) one can show

$$\langle \cdot, \cdot \rangle_{\text{FS}} = \frac{g}{|\xi|^4} \bar{\sigma}_0 \otimes \sigma_0, \quad \text{where } \sigma_0 = \xi_1 d\xi_2 - \xi_2 d\xi_1, \quad g = \det G.$$

The Fubini–Study two-form is the imaginary part.

Since σ_0 is independent of the choice of basis, the Fubini–Study metrics for various choices of basis are all conformal to one another. If we choose an orthonormal basis the metrics are Euclidean. The Fubini–Study metric for a general basis is related to the Euclidean one by

$$\langle \cdot, \cdot \rangle_{\text{FS}} = \kappa(\xi) \langle \cdot, \cdot \rangle_{\text{FS, euc}},$$

where the conformal factor is

$$(48) \quad \kappa(\xi) = \frac{g |\xi|_{\text{euc}}^4}{|\xi|^4},$$

where $|\xi|_{\text{euc}}^2 = |\xi_1|^2 + |\xi_2|^2$.

The curvature term can be calculated directly from the definition $H_{\mu, \eta} \lrcorner \Omega_{\text{FS}}$ and we find

$$T_{\text{curv}} = -\frac{2\mu}{r^2} i\eta.$$

Hamilton's equations in $T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^2$ are

$$(49) \quad \begin{aligned} \dot{r} &= p_r, & \dot{p}_r &= \frac{\mu^2 + |\xi|^2 |\eta|^2}{r^3} - \frac{1}{r^2} V(\xi), \\ \dot{\xi} &= \frac{|\xi|^2}{r^2} G^{-1} \eta, & \dot{\eta} &= \frac{1}{r} DV(\xi) - \frac{|\eta|^2}{r^2} G \xi - \frac{2\mu}{r^2} i\eta. \end{aligned}$$

There are still 10 variables but the invariant set $T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^2$ with $\langle \eta, \xi \rangle = 0$ is eight-dimensional and we have a complex scaling symmetry. The introduction of an affine coordinate on the projective line yields a full *local* reduction to 6 variables. For example, consider those points $[\xi] = [\xi_1, \xi_2] \in \mathbb{C}\mathbb{P}^1$ with $\xi_1 \neq 0$. If ρ is any nonzero constant complex number then every such point has a unique representative of the form $[\xi_1, \xi_2] = [\rho, z]$, $z = x + iy \in \mathbb{C}$, thus parametrizing almost all of the shape sphere by a single complex variable z , the *affine coordinate*. Of course the roles of ξ_1, ξ_2 could be reversed to parametrize the subset with $\xi_2 \neq 0$.

If $\zeta = \alpha + i\beta \in \mathbb{C}^*$ denotes the momentum vector dual to z then the unique extension of $f(z) = (\rho, z)$ to a partially symplectic map $T^*\mathbb{C} \rightarrow T_{\text{pr}}^*\mathbb{C}^2 = \{\langle \eta, \xi \rangle = 0\}$ is defined by $\xi_1 = \rho$, $\xi_2 = z$, $\eta_1 = -\bar{z}\zeta/\rho$, $\eta_2 = \zeta$. One computes the mass metric is

$$|\xi(z)|^2 = g_{11}|\rho|^2 + g_{22}|z|^2 + 2\operatorname{re}(\bar{\rho}g_{12}z)$$

and the cometric is

$$|\zeta|^2 = \frac{|\xi(z)|^2|\zeta|^2}{g|\rho|^2}, \quad \text{with } g = \det(G_{ij}).$$

This gives a Hamiltonian system with 3 degrees of freedom:

$$(50) \quad H_\mu(r, p_r, x, y, \alpha, \beta) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} + \frac{|\xi(z)|^4|\zeta|^2}{g|\rho|^2 r^2} \right) - \frac{1}{r} V(x, y),$$

where

$$V(z) = |\xi(z)| \left(\frac{m_1 m_2}{\rho_{12}} + \frac{m_1 m_3}{\rho_{31}} + \frac{m_2 m_3}{\rho_{23}} \right), \quad \text{with } \rho_{ij} = |a_{ij} + b_{ij}z|.$$

The Fubini–Study form is

$$\Omega_{\text{FS}} = \frac{g}{|\rho|^2|\xi(z)|^2} \operatorname{im} d\bar{z} \otimes dz = \frac{g dx \wedge dy}{|\rho|^2|\xi(z)|^2}.$$

The curvature term is just $T_{\text{curv}} = -\frac{2\mu}{r^2} i\zeta$, as usual.

Example 17 (projective Jacobi coordinates). As a first example, consider using Jacobi coordinates as in [Section 3.4](#), only this time applied to the homogeneous variables X, Z . As before, the basis which defines the Jacobi coordinates is the orthogonal basis $e_1 = (-1, v_2, v_1)$, $e_2 = (0, 1, -1)$. We have

$$\begin{aligned} X &= (-\xi_1, \xi_2 + v_2\xi_1, -\xi_2 + v_1\xi_1), & \xi &= (-X_{12}, v_1X_{31} - v_2X_{23}), \\ Z &= (0, \eta_1 + v_1\eta_2, \eta_1 - v_2\eta_2), & \eta &= (-Z_{12} + v_2Z_{31} + v_1Z_{23}, Z_{31} - Z_{23}), \end{aligned}$$

where, as usual, Z is nonunique.

The Hamiltonian is (47), where the shape potential is

$$V(\xi) = |\xi| \left(\frac{m_1 m_2}{|\xi_1|} + \frac{m_1 m_3}{|\xi_2 + \nu_2 \xi_1|} + \frac{m_2 m_3}{|\xi_2 - \nu_1 \xi_1|} \right).$$

The mass matrix $G = \text{diag}(\mu_1, \mu_2)$ has determinant $g = \mu_1 \mu_2 = m_1 m_2 m_3 / m$ and associated norm and conorm

$$|\xi|^2 = \mu_1 |\xi_1|^2 + \mu_2 |\xi_2|^2 \quad \text{and} \quad |\eta|^2 = \frac{|\eta_1|^2}{\mu_1} + \frac{|\eta_2|^2}{\mu_2}.$$

Hamilton’s equations with the curvature term are given by (49).

If we introduce affine variables by setting $\xi_1 = \rho$, $\xi_2 = z$ as above and if we choose $\rho = \sqrt{\mu_2 / \mu_1}$ the mass norm reduces to $|\xi|^2 = \mu_2(1 + x^2 + y^2)$ and we get the affine Jacobi Hamiltonian

$$H_\mu(r, p_r, x, y, \alpha, \beta) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} + \frac{(1 + x^2 + y^2)^2 |\zeta|^2}{r^2} \right) - \frac{1}{r} V(x, y).$$

Hamilton’s equations with the curvature term are

$$\begin{aligned} \dot{r} &= p_r, & \dot{p}_r &= \frac{1}{r^3} [\mu^2 + (1 + x^2 + y^2)^2 (\alpha^2 + \beta^2)] - \frac{1}{r^2} V(\xi), \\ \dot{x} &= \frac{(1 + x^2 + y^2)^2}{r^2} \alpha, & \dot{y} &= \frac{(1 + x^2 + y^2)^2}{r^2} \beta, \\ \dot{\alpha} &= \frac{1}{r} V_x(x, y) - \frac{2}{r^2} (1 + x^2 + y^2) (\alpha^2 + \beta^2) x + \frac{2\mu}{r^2} \beta, \\ \dot{\beta} &= \frac{1}{r} V_y(x, y) - \frac{2}{r^2} (1 + x^2 + y^2) (\alpha^2 + \beta^2) y - \frac{2\mu}{r^2} \alpha. \end{aligned} \tag{51}$$

Example 18 (equilateral coordinates). In projective Jacobi coordinates (ξ_1, ξ_2) , the binary collision points b_{12}, b_{13}, b_{23} are located at the projective points

$$[1, 0], [1, -\nu_2], [1, \nu_1] \in \mathbb{C}\mathbb{P}^1$$

while the equilateral triangle configurations (the Lagrange points) are at

$$[1, \ell_\pm] \in \mathbb{C}\mathbb{P}^1, \quad \text{where } \ell_\pm = \frac{m_1 - m_2}{2(m_1 + m_2)} \pm \frac{\sqrt{3}}{2} i = \frac{\nu_1 - \nu_2}{2} \pm \frac{\sqrt{3}}{2} i.$$

Using a Möbius transformation, we can put three points anywhere we like on the shape sphere, $\mathbb{C}\mathbb{P}^1$. Remarkably, it turns out that if we put the binary collisions at the third roots of unity

$$[1, \xi_2] = [1, 1], [1, \omega], [1, \bar{\omega}] \in \mathbb{C}\mathbb{P}^1 \quad \text{with } \omega = \frac{1}{2}(-1 + i\sqrt{3}), \tag{52}$$

then the equilateral points are automatically moved to the north and south poles $[1, 0], [0, 1]$. These coordinates were introduced in [Moeckel et al. 2012].

These coordinates are obtained by choosing the basis

$$e_1 = (1, \omega, \bar{\omega}), \quad e_2 = -\bar{e}_1 = (-1, -\bar{\omega}, -\omega)$$

for \mathcal{W} . The coordinate change map is $X = \xi_1 e_1 + \xi_2 e_2$ or

$$X_{12} = \xi_1 - \xi_2, \quad X_{31} = \omega\xi_1 - \bar{\omega}\xi_2, \quad X_{23} = \bar{\omega}\xi_1 - \omega\xi_2,$$

and indeed takes the roots of unity (52) to the binary collisions. Setting $\xi_2 = 0$, we see that $|X_{12}| = |X_{32}| = |X_{23}|$ corresponding to an equilateral triangle, with the same result if $\xi_1 = 0$. Thus the coordinate change map sends the poles $\xi = [1, 0]$, $[0, 1]$ to the equilateral triangles.

The mutual distances (of the homogeneous variables) $\rho_{ij} = |X_{ij}|$ that appear in the shape potential are very simple:

$$\rho_{12} = |\xi_1 - \xi_2|, \quad \rho_{31} = |\xi_1 - \omega\xi_2|, \quad \rho_{23} = |\xi_1 - \bar{\omega}\xi_2|.$$

The mass metric can also be written in terms of these

$$|\xi|^2 = \frac{1}{m}(m_1 m_2 \rho_{12}^2 + m_3 m_1 \rho_{31}^2 + m_2 m_3 \rho_{23}^2).$$

It is represented by the matrix G with entries $g_{ij} = \langle e_i, e_j \rangle$:

$$g_{11} = g_{22} = \frac{m_1 m_2 + m_3 m_1 + m_2 m_3}{m}, \quad g_{12} = \bar{g}_{21} = -\frac{m_1 m_2 + m_3 m_1 \omega + m_2 m_3 \bar{\omega}}{m},$$

and determinant $g = \det G = 3m_1 m_2 m_3 / m$.

The inverse transformation is given by

$$\xi_1 = \frac{1}{3}(X_{12} + \bar{\omega}X_{31} + \omega X_{23}), \quad \xi_2 = -\frac{1}{3}(X_{12} + \omega X_{31} + \bar{\omega}X_{23}),$$

and the momenta satisfy $\eta_1 = Z_{12} + \bar{\omega}Z_{31} + \omega Z_{23}$, $\eta_2 = -Z_{12} - \omega Z_{31} - \bar{\omega}Z_{23}$.

Choosing affine variables by setting $\xi_1 = z$, $\xi_2 = 1$, we get the Hamiltonian (50) with

$$|\xi(z)|^2 = \frac{1}{m}(m_1 m_2 |z - 1|^2 + m_3 m_1 |z - \omega|^2 + m_2 m_3 |z - \bar{\omega}|^2).$$

The complexity of mass norm is perhaps outweighed by the fact that the potential is given by the wonderful expression

$$V(z) = |\xi(z)| \left(\frac{m_1 m_2}{|z - 1|} + \frac{m_1 m_3}{|z - \omega|} + \frac{m_2 m_3}{|z - \bar{\omega}|} \right).$$

The advantage of these coordinates is that they provide the homogenized potential V with “radial monotonicity”. Let $E = x(\partial/\partial x) + y(\partial/\partial y)$ be the radial vector field in the z plane, where $z = x + iy$. Then $E[V] > 0$ for $0 < |z| < 1$, $E[V] < 0$ for $|z| > 1$, and $E[V] = 0$ if and only if $|z| = 1$ or $z = 0$. (See Proposition 4 of [Moeckel

et al. 2012].) This monotonicity was the key ingredient to the main theorem of [Montgomery 2002].

5.5. Making the shape sphere round. Instead of using projective or local affine coordinates, one can map the shape sphere to the unit sphere in \mathbb{R}^3 . First we do this homogeneously, then restrict to the unit sphere to get another version with 6 degrees of freedom. Let $\xi = (\xi_1, \xi_2) \in \mathbb{C}^2$ be coordinates associated with some choice of basis e_1, e_2 for \mathcal{W} .

Consider the Hopf map $h : \mathbb{C}^2 \rightarrow \mathbb{R}^3$ given by $w_1 = 2 \operatorname{re} \bar{\xi}_1 \xi_2$, $w_2 = 2 \operatorname{im} \bar{\xi}_1 \xi_2$, $w_3 = |\xi_1|^2 - |\xi_2|^2$. Using the Euclidean metric for w we get

$$|w|^2 = w_1^2 + w_2^2 + w_3^2 = |\xi|_{\text{euc}}^4 = (|\xi_1|^2 + |\xi_2|^2)^2.$$

It follows that $2|\xi_1|^2 = |w| + w_3$, $2|\xi_2|^2 = |w| - w_3$, $2\bar{\xi}_1 \xi_2 = w_1 + i w_2$.

We will need formulas for $\rho_{ij} = |X_{ij}| = |a_{ij}\xi_1 + b_{ij}\xi_2|$ in the variables w_i . We have

$$(53) \quad \begin{aligned} \rho_{ij}^2 &= |a_{ij}|^2 |\xi_1|^2 + |b_{ij}|^2 |\xi_2|^2 + 2 \operatorname{re}(\bar{\xi}_1 \xi_2 \bar{a}_{ij} b_{ij}) \\ &= \frac{1}{2}(|a_{ij}|^2 + |b_{ij}|^2)|w| + \frac{1}{2}(|a_{ij}|^2 - |b_{ij}|^2)w_3 + \operatorname{re}(\bar{a}_{ij} b_{ij})w_1 - \operatorname{im}(\bar{a}_{ij} b_{ij})w_2. \end{aligned}$$

Then the mass metric will be given by

$$(54) \quad |\xi|^2 = \frac{1}{m}(m_1 m_2 \rho_{12}^2 + m_3 m_1 \rho_{31}^2 + m_2 m_3 \rho_{23}^2).$$

If we let $\alpha_1, \alpha_2, \alpha_3$ be dual momentum variables, we can extend the Hopf map h to a partially symplectic map $F : T_{\text{pr}}^* \mathbb{C}^2 \rightarrow T_{\text{sph}}^* \mathbb{R}^3$ by defining its (pseudo) inverse:

$$\eta = \alpha \circ Dh := Dh^t \alpha.$$

To find the reduced Hamiltonian in w coordinates we will exploit the fact that the Euclidean metric transforms nicely. Recall that the shape kinetic energy is the dual of the Fubini–Study metric and that the latter is related conformally to the Euclidean metric with conformal factor κ^{-1} , where κ is given by (48). In other words, since we are restricting to $\langle \eta, \xi \rangle = 0$ we have

$$|\xi|^2 |\eta|^2 = \kappa^{-1} |\xi|_{\text{euc}}^2 |\eta|_{\text{euc}}^2.$$

One can verify that the Euclidean norms transform under the Hopf map in such a way that

$$|\xi|_{\text{euc}}^2 |\eta|_{\text{euc}}^2 = 4|w|^2 |\alpha|^2,$$

where we are using the Euclidean norm on $\mathbb{R}^3, \mathbb{R}^{3*}$. Hence the reduced Hamiltonian on the sphere is given by

$$H_\mu(r, p_r, w, \alpha) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} + \frac{4|w|^2 |\alpha|^2}{\kappa(w)r^2} \right) - \frac{1}{r} V(w),$$

where $|w|^2 = w_1^2 + w_2^2 + w_3^2$ and $|\alpha|^2 = \alpha_1^2 + \alpha_2^2 + \alpha_3^2$ and where the shape potential is given by

$$V(w) = |\xi(w)| \left(\frac{m_1 m_2}{\rho_{12}} + \frac{m_3 m_1}{\rho_{31}} + \frac{m_2 m_3}{\rho_{23}} \right)$$

with the ρ_{ij} and $|\xi|$ as in (53) and (54).

The Fubini–Study form becomes a multiple $\kappa/4$ of the Euclidean solid angle form

$$\Omega_{\text{FS}} = \frac{\kappa}{4|w|^3} (w_1 dw_2 \wedge dw_3 + w_2 dw_3 \wedge dw_1 + w_3 dw_1 \wedge dw_2).$$

This leads to the curvature term

$$T_{\text{curv}} = \frac{2\mu}{|w|r^2} \alpha \times w,$$

where $w \times \alpha$ denotes the cross product in \mathbb{R}^3 .

The differential equations are

$$(55) \quad \begin{aligned} \dot{r} &= p_r, & \dot{p}_r &= \frac{1}{r^3} \left(\mu^2 + \frac{4|w|^2|\alpha|^2}{\kappa} \right) - \frac{1}{r^2} V(\xi), & \dot{w} &= \frac{4|w|^2}{\kappa r^2} \alpha, \\ \dot{\alpha} &= \frac{1}{r} DV(w) - \frac{4|\alpha|^2}{\kappa r^2} w + \frac{4|w|^2|\alpha|^2}{\kappa^2 r^2} \kappa_w + \frac{2\mu}{|w|r^2} \alpha \times w. \end{aligned}$$

From [Theorem 1](#), if we restrict to $T^*\mathbb{R}^+ \times T^*_{\text{sph}}\mathbb{R}^3 = \{(\alpha, w)_{\text{euc}} = 0\}$ and quotient by the scaling action of \mathbb{R}^+ , we get a reduced system equivalent to the reduced three-body problem. But $(\alpha, w)_{\text{euc}} = 0$ implies that $|w|$ is constant under the flow. Hence we have a six-dimensional invariant submanifold given by $|w| = 1$, $(\alpha, w)_{\text{euc}} = 0$ representing the reduced three-body problem. The reduced phase space is $T^*\mathbb{R}^+ \times T^*\mathcal{S}^2$ and the shape sphere is represented by the standard unit sphere.

To get to six dimensions with no constraints one could parametrize the sphere with two variables. If this is done with stereographic projection, the result is similar to the affine coordinate reduction of [Section 5.4](#). On the other hand one could also use spherical coordinates θ, ϕ . However, both of these are just local coordinates while the system above is global, albeit constrained.

Example 19 (Jacobi coordinates on \mathcal{S}^2). If we choose an orthonormal basis for ${}^{\circ}\mathcal{W}$ then we get the conformal factor $\kappa = 1$ and the resulting Hamiltonian will have a simpler shape kinetic energy. For example, we could normalize the Jacobi basis of [Example 17](#) to

$$e'_1 = \frac{1}{\sqrt{\mu_1}} (-1, v_2, v_1), \quad e'_2 = \frac{1}{\sqrt{\mu_2}} (0, 1, -1).$$

The coordinates ξ_i are replaced by $\sqrt{\mu_i} \xi_i$ in all of the formulas. We get rather complicated homogeneous mutual distances

$$\begin{aligned} 2\mu_1\mu_2\rho_{12}^2 &= \mu_2(|w| + w_3), \\ 2\mu_1\mu_2\rho_{31}^2 &= (\mu_2v_2^2 + \mu_1)|w| + (\mu_2v_2^2 - \mu_1)w_3 + 2v_2\sqrt{\mu_1\mu_2} w_1, \\ 2\mu_1\mu_2\rho_{23}^2 &= (\mu_2v_1^2 + \mu_1)|w| + (\mu_2v_1^2 - \mu_1)w_3 - 2v_1\sqrt{\mu_1\mu_2} w_1. \end{aligned}$$

In the equal mass case with $m_i = 1$ and $|w| = 1$, however, we get

$$\rho_{12}^2 = |w| + w_3, \quad \rho_{31}^2 = |w| + \frac{\sqrt{3}}{2}w_1 - \frac{1}{2}w_3, \quad \rho_{23}^2 = |w| - \frac{\sqrt{3}}{2}w_1 - \frac{1}{2}w_3.$$

On the other hand the Hamiltonian is

$$H_\mu(r, p_r, w, \alpha) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} + \frac{4|w|^2|\alpha|^2}{r^2} \right) - \frac{1}{r} V(w),$$

where the norms are Euclidean.

Example 20 (equilateral coordinates on S^2). If we use the basis of [Example 18](#) $e_1 = (1, \omega, \bar{\omega})$, $e_2 = -\bar{e}_1 = (-1, -\bar{\omega}, -\omega)$, we get simple mutual distances

$$\rho_{12}^2 = |w| - w_1 \quad \rho_{31}^2 = |w| + \frac{1}{2}w_1 - \frac{\sqrt{3}}{2}w_2 \quad \rho_{23}^2 = |w| + \frac{1}{2}w_1 + \frac{\sqrt{3}}{2}w_2.$$

Collinear shapes form the equator $w_3 = 0$ with the binary collisions placed at the roots of unity.

On the other hand we have a formidable conformal factor

$$\kappa = \frac{3m_1m_2m_3m(w_1^2 + w_2^2 + w_3^2)}{(m_1m_2\rho_{12}^2 + m_3m_1\rho_{31}^2 + m_2m_3\rho_{23}^2)^2}.$$

In the equal mass case ($m_i = 1$) we see $\kappa = 1$.

5.6. Visualizing the shape sphere. Having reduced the planar three-body problem by using size and shape coordinates, we will pause to have a closer look at the shape sphere and the shape potential.

Using the spherical variables $w = (w_1, w_2, w_3)$ we can visualize the shape sphere as the round unit sphere in \mathbb{R}^3 . The equilateral basis of [Example 20](#) puts the binary collisions at the third roots of unity on the equator and the Lagrange equilateral configurations at the poles. [Figure 1](#) shows some of the level curves of V for two choices of the masses. In addition to the binary collisions shapes where $V \rightarrow \infty$, there are three saddle points at the Eulerian central configurations. The Lagrange points are always minima of V .

If we use stereographic projection to map the sphere to the complex plane, we get the affine coordinate representation of [Example 18](#). [Figure 2](#) shows affine contour

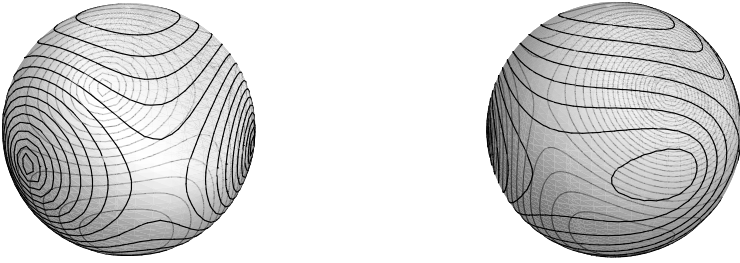


Figure 1. Contour plot of the shape potential on the unit sphere $w_1^2 + w_2^2 + w_3^2 = 1$ in the equal mass case (left) and for masses $m_1 = 1, m_2 = 2, m_3 = 10$ (right).

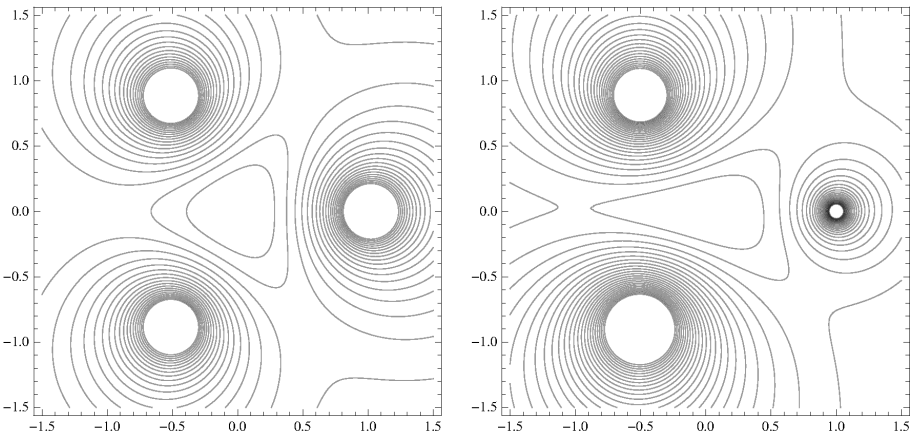


Figure 2. Contour plot of the shape potential on the complex plane in the equal mass case (left) and for masses $m_1 = 1, m_2 = 2, m_3 = 10$ (right). These plots can be viewed as stereographic projections of those in [Figure 1](#).

plots for the same two choices of the masses. Now the collinear shapes are on the real axis.

6. Levi-Civita regularization

In this section, we describe a way to simultaneously regularize all 3 binary collision using 3 separate Levi-Civita transformations. This approach to simultaneous regularization was introduced by Heggie [1974]. There are two versions depending on whether the variables Q_{ij} or the homogeneous variables X_{ij} are used. The former approach was used by Heggie; we will take the latter. We begin with a review of Levi-Civita regularization for the Kepler problem.

Levi-Civita showed how to regularize the two-body problem, which is to say, the Kepler problem. Let $q \in \mathbb{C}$ denote the position of a planet going around an infinitely massive sun placed at the origin. After a normalization, the Kepler Hamiltonian is $\frac{1}{2}|p|^2 - \alpha/|q|$. Levi-Civita's transformation is the map

$$z \mapsto z^2 = q$$

together with the induced map on momenta

$$\eta \mapsto \frac{1}{2\bar{z}}\eta = p$$

and the time rescaling

$$\frac{d}{d\tau} = r \frac{d}{dt}.$$

To understand the map on momenta, make the substitution $q = z^2$ in the expression $\langle p, dz \rangle$ for the canonical one-form. We have $\langle p, dq \rangle = \langle p, 2zdz \rangle = \langle 2\bar{z}p, dz \rangle$, which shows that if $\eta = 2\bar{z}p$ then $\langle \eta, dz \rangle = \langle p, dq \rangle$. This computation shows that the map $(\eta, z) \rightarrow (p, q)$ with $p = (1/(2\bar{z}))\eta, q = z^2$ is a 2:1 canonical transformation away from the origin. Observe that $r = |z|^2$. Thus in terms of the new variables

$$H = \frac{1}{2r} \left(|\eta|^2 - \frac{\alpha}{|z|^2} \right).$$

Time rescaling is equivalent to rescaling the Hamiltonian vector field. This rescaling can be implemented using the following ‘‘Poincaré trick’’. If X_H is the Hamiltonian vector field for H , and if h is a value of H , then fX_H is the Hamiltonian vector field for the Hamiltonian $\tilde{H} = f(H - h)$ provided we restrict ourselves to the level set $\{H = h\}$. We take $f = r = |z|^2$ and compute that

$$\tilde{H} = \frac{1}{2}(|\eta|^2 - h|z|^2 - \alpha),$$

which is the Hamiltonian for a harmonic oscillator when $h < 0$.

6.1. Simultaneous regularization. Let (r, X) denote either the spherical-homogeneous or projective-homogeneous coordinates. To simultaneously regularize all three double collisions we perform a Levi-Civita transformation on each of the homogeneous complex variables X_{ij} . Thus, we introduce three new complex variables $z_{ij} = -z_{ji}$ and set $X_{ij} = z_{ij}^2$. Define a regularizing map $f : \mathbb{C}_0^3 \rightarrow \mathbb{C}_0^3$ by

$$X = f(z_{12}, z_{31}, z_{23}) = (z_{12}^2, z_{31}^2, z_{23}^2).$$

The preimage of the subspace \mathcal{W} is the quadratic cone

$$\mathcal{C} : z_{12}^2 + z_{31}^2 + z_{23}^2 = 0.$$

We have $f : \mathcal{C}_0 \rightarrow \mathcal{W}_0$. Every $X \in \mathcal{W}_0$ has 8 preimages under f , except for the three binary collision points ($X_{ij} = 0$ some ij), which each have 4 preimages. (Since $X \neq 0$, at most one of the X_{ij} or z_{ij} can vanish at a time on \mathcal{W}_0 or \mathcal{C}_0 .)

Since f is homogeneous, it induces maps $f_{\text{sph}} : \mathbf{S}^5 \rightarrow \mathbf{S}^5$ and $f_{\text{pr}} : \mathbb{C}\mathbb{P}^2 \rightarrow \mathbb{C}\mathbb{P}^2$. In this case we also view z_{ij} as homogenous spherical or projective coordinates. These restrict to regularizing maps $f_{\text{sph}} : \mathbf{S}(\mathcal{C}) \rightarrow \mathbf{S}(\mathcal{W})$ and $f_{\text{pr}} : \mathbb{P}(\mathcal{C}) \rightarrow \mathbb{P}(\mathcal{W})$, where, as above, $\mathbf{S}(\cdot)$ and $\mathbb{P}(\cdot)$ denote quotient spaces under real and complex scaling, respectively.

The mutual distances become

$$(56) \quad \rho_{ij} = |X_{ij}| = |z_{ij}|^2$$

and the mass norm is

$$(57) \quad |X(z)|^2 = |f(z)|^2 = \frac{m_1 m_2 \rho_{12}^2 + m_1 m_3 \rho_{31}^2 + m_2 m_3 \rho_{23}^2}{m_1 + m_2 + m_3}.$$

We will use the standard Hermitian inner product, denoted $\langle \cdot, \cdot \rangle$, on z -space so

$$(58) \quad \|z\|^2 = |z_{12}|^2 + |z_{31}|^2 + |z_{23}|^2 = \rho_{12} + \rho_{31} + \rho_{23}.$$

Let η_{ij} be the conjugate momenta to z_{ij} and let Y_{ij} the homogenous momenta conjugate to X_{ij} . We extend f to a map $(r, p_r, X, Y) = F(r, p_r, z, \eta)$ by setting

$$Y_{ij} = \frac{1}{2\bar{z}_{ij}} \eta_{ij}.$$

Then F restricts to maps

$$T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}^3 \rightarrow T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}^3 \quad \text{and} \quad T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^3 \rightarrow T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^3,$$

where in (z, η) -variables we have the constraints $\text{re}\langle \eta, z \rangle = 0$ for the sphere and $\langle \eta, z \rangle = 0$ for the projective plane. We continue to denote these restricted maps by the letter F .

The action of $c \in \mathbb{C}$ by translation of the momenta Y_{ij} to $Y_{ij} + c$ pulls-back under F to translation of η_{ij} by $2c\bar{z}_{ij}$, that is, to the action

$$c \cdot (r, p_r, z, \eta) = (r, p_r, z, \eta + 2c\bar{z}).$$

The momentum map for this pulled back action is $\gamma = z_{12}^2 + z_{31}^2 + z_{23}^2$. Of course we will be interested in the level set $\gamma = 0$. We will call this the z -translation symmetry of η .

6.1.1. Geometry of \mathcal{C} and the regularized shape sphere. It is interesting to investigate the algebraic surface \mathcal{C} in more detail. If we write the complex vector $z \in \mathbb{C}^3$ as $z = a + ib$, where $a = \text{re } z$ and $b = \text{im } z \in \mathbb{R}^3$, then

$$z_{12}^2 + z_{31}^2 + z_{23}^2 = 0 \quad \text{if and only if} \quad |a|^2 = |b|^2, \quad a \cdot b = 0.$$

This means a, b are real, orthogonal vectors of equal length

$$s^2 = |a|^2 = |b|^2 = \frac{|z|^2}{2}.$$

If we define a third vector $c = a \times b$ we get an orthogonal frame in \mathbb{R}^3 and the matrix

$$(59) \quad A(z) = \frac{1}{s} \begin{bmatrix} a_{12} & b_{12} & c_{12}/s \\ a_{31} & b_{31} & c_{31}/s \\ a_{23} & b_{23} & c_{23}/s \end{bmatrix} \in \text{SO}(3).$$

The mapping $A(z)$ induces a diffeomorphism between the quotient space $\mathcal{S}(\mathcal{C})$ of \mathcal{C}_0 by positive real scalings to $\text{SO}(3)$ and hence, as is well-known, to the real projective space $\mathbb{R}\mathbb{P}(3)$ (and to the unit tangent bundle to S^2).

The projective curve $\mathbb{P}(\mathcal{C})$ turns out to be diffeomorphic to the two-sphere S^2 and, accordingly, we will call it the *regularized shape sphere*. One way to see this is to note that $\mathbb{P}(\mathcal{C}) \simeq \mathcal{S}(\mathcal{C})/S^1$ is the quotient of $\mathcal{S}(\mathcal{C})$ under rotations. It is easy to see that action the rotation group on z rotates the vectors $a, b \in \mathbb{R}^3$ above in their own plane and leaves $c = a \times b$ invariant. It follows that the map $z \mapsto c/|c|$ induces a diffeomorphism $\mathbb{P}(\mathcal{C}) \simeq S^2$.

In the sections below, we will apply the regularizing map to obtain several regularized Hamiltonians for the three-body problem. Starting with spherical-homogenous variables leads to a regularized system not reduced by rotations while the projective-homogenous variables lead to a Hamiltonian system which is both regularized and reduced. In addition we will consider several ways to parametrize the cone \mathcal{C} to obtain lower-dimensional systems. [Theorem 1](#) can be applied to show the equivalence of the Hamiltonian systems below, but we will omit most of the details.

6.2. Spherical regularization. First we will find the regularized Hamiltonian in spherical-homogeneous coordinates. This gives a regularization of binary collisions without reducing by the rotational symmetry. Let (r, X) be the spherical-homogeneous coordinates of [Section 4](#). The spherical Hamiltonian is

$$H_{\text{sph}}(r, p_r, X, Y) = \frac{1}{2} p_r^2 + \frac{|X|^2}{r^2} K(Y) - \frac{1}{r} V(X).$$

Using the formula analogous to the one in [\(7\)](#) for $K(Y)$ and applying the regularizing map gives

$$(60) \quad H_{\text{sph}}(r, p_r, z, \eta) = \frac{1}{2} p_r^2 + \frac{|X(z)|^2}{r^2} \left(\frac{|\pi_1|^2}{8m_1\rho_{12}\rho_{31}} + \frac{|\pi_2|^2}{8m_2\rho_{12}\rho_{23}} + \frac{|\pi_3|^2}{8m_3\rho_{31}\rho_{23}} \right) - \frac{1}{r} \left(\frac{m_1 m_2}{\rho_{12}} + \frac{m_3 m_1}{\rho_{31}} + \frac{m_2 m_3}{\rho_{23}} \right),$$

where

$$(61) \quad \pi_1 = \eta_{12}\bar{z}_{31} - \eta_{31}\bar{z}_{12}, \quad \pi_2 = \eta_{23}\bar{z}_{12} - \eta_{12}\bar{z}_{23}, \quad \pi_3 = \eta_{31}\bar{z}_{23} - \eta_{23}\bar{z}_{31}.$$

Next we rescale time using the Poincaré trick. One choice of time-rescaling factor is $|z_{12}z_{31}z_{23}|^2 = \rho_{12}\rho_{31}\rho_{23}$. But since X, z are homogeneous coordinates, a degree-zero homogeneous function such as

$$(62) \quad \tau = \frac{\rho_{12}\rho_{31}\rho_{23}}{(\rho_{12} + \rho_{31} + \rho_{23})^3} = \frac{\rho_{12}\rho_{31}\rho_{23}}{\|z\|^6}$$

seems more appropriate. Note that by the arithmetic-geometric mean inequality we have $0 \leq \tau \leq \frac{1}{27}$. In Section 6.3 we will choose a different time rescaling function λ .

The rescaled solution with energy $H_{\text{sph}} = h$ become the zero-energy solutions for the Hamiltonian $\tilde{H}_{\text{sph}}(r, p_r, z, \eta) = \tau(H_{\text{sph}} - h)$:

$$(63) \quad \tilde{H}_{\text{sph}} = \frac{\tau p_r^2}{2} + \frac{|X(z)|^2}{r^2\|z\|^6} \left(\frac{|\pi_1|^2\rho_{23}}{8m_1} + \frac{|\pi_2|^2\rho_{31}}{8m_2} + \frac{|\pi_3|^2\rho_{12}}{8m_3} \right) - \frac{1}{r}W(z) - h\tau,$$

where the *regularized shape potential* W is

$$(64) \quad W(z) = \frac{|X(z)|}{\|z\|^6} (m_1m_2\rho_{31}\rho_{23} + m_1m_3\rho_{12}\rho_{23} + m_2m_3\rho_{12}\rho_{31}).$$

Note that since z is a homogeneous variable representing $[z] \in \mathbf{S}^5$, we have $z \neq 0$. For a homogeneous coordinate representing a binary collision we will have exactly one of the variables $z_{ij} = 0$ and $\|z\| > 0$. Thus \tilde{H} is nonsingular at these points and the binary collisions are regularized.

Theorem 21. *The Hamiltonian flow of \tilde{H}_{sph} on $T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ has an invariant submanifold $T^*\mathbb{R}^+ \times T_{\text{sph}, \mathbb{C}}^*\mathbb{C}_0^3$ defined by $\text{re}\langle \eta, z \rangle = 0$ and $z_{12}^2 + z_{31}^2 + z_{23}^2 = 0$. The quotient of the restricted flow by scaling and by translation of η by \bar{z} represents the zero total momentum three-body problem with regularized binary collisions, reduced by translations (but not by rotations).*

The quotient space of $T_{\text{sph}, \mathbb{C}}^*\mathbb{C}_0^3$ by these symmetries can be identified with $T^*\mathcal{S}(\mathbb{C}) \simeq T^*\mathbb{RP}(3)$. The regularizing map induces an 8-to-1 branched covering map $f_{\text{sph}} : \mathcal{S}(\mathbb{C}) \rightarrow \mathcal{S}(\mathcal{W})$, that is, an 8-to-1 branched covering $\mathbb{RP}^3 \mapsto \mathcal{S}^3$. The map is a diffeomorphism except where (exactly) one of the $z_{ij} = 0$ and $X_{ij} = 0$. To describe the branching behavior, note that in the two-dimensional complex subspace \mathcal{W} , the set where $X_{12} = 0$ is a complex line which corresponds to a circle S^1 in the sphere $\mathcal{S}(\mathcal{W})$. The preimage of this circle will be 2 circles in the projective space $\mathcal{S}(\mathbb{C})$. Altogether, the map is branched over 3 circles, each circle having preimage 2 circles in the projective space \mathbb{RP}^3 .

6.2.1. Quadratic parametrization of \mathcal{C} . Instead of writing Hamilton’s equations for \tilde{H}_{sph} , we will describe a parametrization of the cone \mathcal{C} that leads to a lower-dimensional system of equations. There is a nice 2-to-1 parametrization by quadratic polynomials which is related to the double covers of $\mathbb{R}P^3$ by S^3 , of $SO(3)$ by the unit quaternions, and of $SO(3)$ by $SU(2)$.

Define a 2-to-1 mapping $g : \mathbb{C}^2 \rightarrow \mathcal{C} \subset \mathbb{C}^3$ by

$$(65) \quad g : \quad z_{12} = 2i x_1 x_2, \quad z_{31} = x_1^2 + x_2^2, \quad z_{23} = i(x_1^2 - x_2^2),$$

where $x_1, x_2 \in \mathbb{C}$. This can be seen as a variant of a map used by Waldvogel [1972] in his regularization of the planar problem. But here we are applying the idea to the homogeneous variables X , which makes it easier to blow-up triple collision later on.

By homogeneity, there is an induced map $g_{\text{sph}} : S^3 \rightarrow S(\mathcal{C})$. The induced map is given by the same formula except that x, z now denote homogenous coordinates for the points of S^3, S^5 . (This double covering map gives another way to see that $S(\mathcal{C})$ is diffeomorphic to the real projective space $\mathbb{R}P^3$.) The map g_{sph} can be motivated in several ways. First, after omitting the factors of i , it resembles the formulas for parametrizing Pythagorean triples. Next, write $x_1 = u_1 - i u_2, x_2 = u_3 + i u_4$ and define the unit quaternion $u = u_1 + i u_2 + j u_3 + k u_4$. Then the familiar conjugation map $v \mapsto uv\bar{u}$, where v is an imaginary quaternion, defines a rotation $R(x)$ on the three-dimensional space of v ’s. Up to a permutation of the columns, $R(x) = A(z)$, the matrix of (59), and hence the conjugation map defines a map $x \mapsto z$. As a variation on this construction, define the unitary x -dependent matrix

$$U = \begin{bmatrix} \bar{x}_1 & x_2 \\ -\bar{x}_2 & x_1 \end{bmatrix} \in SU(2).$$

Then the adjoint representation $v \mapsto U(x)vU(x)^{-1}$ on $\mathfrak{su}(2) \simeq \mathbb{R}^3$ produces the same rotation $R(x)$.

The composition $f \circ g_{\text{sph}}$ of the regularizing map and the quadratic parametrization gives a 16-to-1 branched cover $S^3 \mapsto S^3$, which becomes 8-to-1 over the binary collisions. Each binary collision is represented by a circle in the range which has 2 preimage circles for a total of 6 branching circles in the domain. Using stereographic projection, it is possible to get some idea of the behavior of this remarkable, regularizing map. Figure 3 shows the projection of the three-sphere. The three transparent surfaces are tori representing the collinear configurations with a given ordering of the bodies along the line. These intersect in 6 circles representing the binary collisions. The figure shows thin tubes around each of these circles.

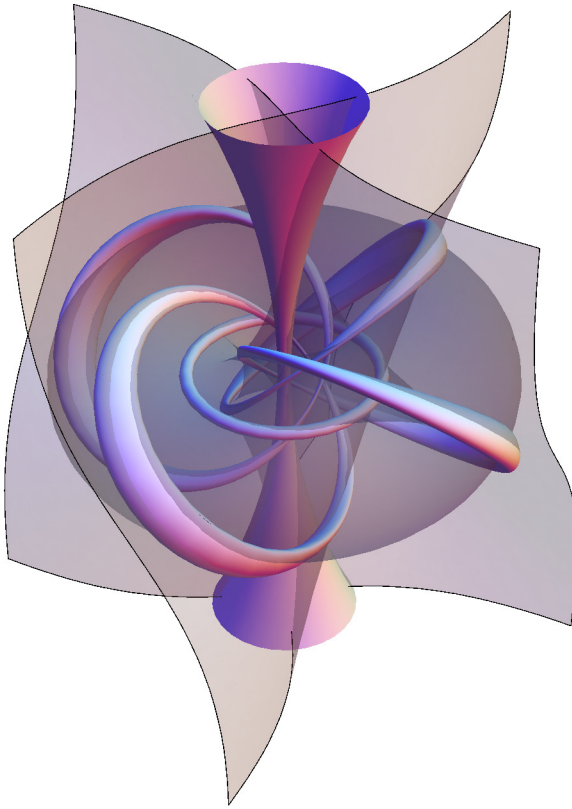


Figure 3. Stereographic projection of S^3 showing the preimage under the regularizing map of the collinear configurations and small tubes around the binary collision circles.

To extend g to a partially symplectic map $G : T^*\mathbb{R}^+ \times T^*\mathbb{C}^2 \rightarrow T^*\mathbb{R}^+ \times \mathcal{C} \times \mathbb{C}^{3*}$ we transform the momenta η, y so that $y = \eta \overline{Df(z)}$ or

$$\begin{bmatrix} y_1 & y_2 \end{bmatrix} = \begin{bmatrix} \eta_{12} & \eta_{31} & \eta_{23} \end{bmatrix} \begin{bmatrix} -2i\bar{x}_2 & -2i\bar{x}_1 \\ 2\bar{x}_1 & 2\bar{x}_2 \\ -2i\bar{x}_1 & 2i\bar{x}_2 \end{bmatrix}.$$

The value of η is not uniquely determined but any two solutions will yield equivalent covectors and the same transformed Hamiltonian. For example, we could take

$$\eta_{12} = 0, \quad \eta_{31} = \frac{1}{4} \left(\frac{y_1}{\bar{x}_1} + \frac{y_2}{\bar{x}_2} \right), \quad \eta_{23} = \frac{i}{4} \left(\frac{y_1}{\bar{x}_1} - \frac{y_2}{\bar{x}_2} \right).$$

G restricts to $G : T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}^2 \rightarrow T^*\mathbb{R}^+ \times T_{\text{sph}, \mathcal{C}}^*\mathbb{C}^{3*}$, where

$$T_{\text{sph}}^*\mathbb{C}^2 = \{(x, y) : \text{re}\langle y, x \rangle = 0\} \quad \text{and} \quad T_{\text{sph}, \mathcal{C}}^*\mathbb{C}^{3*} = \{(z, \eta) : z \in \mathcal{C}, \text{re}\langle \eta, z \rangle = 0\}.$$

The regularized spherical Hamiltonian becomes

$$(66) \quad \begin{aligned} \tilde{H}_{\text{sph}} &= \frac{\tau p_r^2}{2} + \frac{|X(x)|^2}{r^2 \|x\|^{12}} \left(\frac{|\pi'_1|^2 \rho_{23}}{256m_1} + \frac{|\pi'_2|^2 \rho_{31}}{256m_2} + \frac{|\pi'_3|^2 \rho_{12}}{256m_3} \right) - \frac{1}{r} W(x) - h\tau, \\ \pi'_1 &= y_1 \bar{x}_2 + y_2 \bar{x}_1, & \pi'_2 &= y_1 \bar{x}_2 - y_2 \bar{x}_1, & \pi'_3 &= y_1 \bar{x}_1 - y_2 \bar{x}_2, \\ \rho_{12} &= |2x_1 x_2|^2, & \rho_{31} &= |x_1^2 + x_2^2|^2, & \rho_{23} &= |x_1^2 - x_2^2|^2, \\ \|z\|^2 &= 2\|x\|^4 = \rho_{12} + \rho_{31} + \rho_{23}, \\ |X(x)|^2 &= \frac{m_1 m_2 \rho_{12}^2 + m_1 m_3 \rho_{31}^2 + m_2 m_3 \rho_{23}^2}{m_1 + m_2 + m_3}. \end{aligned}$$

Note that \tilde{H} is invariant under the scaling symmetry $(x, y) \rightarrow (kx, k^{-1}y)$, $k > 0$. The corresponding Hamiltonian system on the ten-dimensional space $T^*(\mathbb{R}^+ \times \mathbb{C}^2)$ can be reduced to the expected eight dimensions by restricting to the invariant set $T^*\mathbb{R}^+ \times T_{\text{sph}}^*\mathbb{C}^2$ and then passing to the quotient space under scaling.

6.3. Projective regularization. Next we will get a regularized version of the reduced three-body problem. Let (r, X) be the projective-homogeneous coordinates of [Section 5](#). For a fixed angular momentum, we have the reduced Hamiltonian on $T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^3$

$$H_\mu(r, p_r, X, Z) = \frac{1}{2} \left(p_r^2 + \frac{\mu^2}{r^2} \right) + \frac{|X|^2}{r^2} K(Z) - \frac{1}{r} V(|X|).$$

After making the Levi-Civita transformations, fixing an energy and changing time-scale by the factor τ from [\(62\)](#) we obtain a regularized reduced Hamiltonian

$$(67) \quad \tilde{H}_\mu = \frac{\tau p_r^2}{2} + \frac{\tau \mu^2}{2r^2} + \frac{|X(z)|^2}{r^2 \|z\|^6} \left(\frac{|\pi_1|^2 \rho_{23}}{8m_1} + \frac{|\pi_2|^2 \rho_{31}}{8m_2} + \frac{|\pi_3|^2 \rho_{12}}{8m_3} \right) - \frac{1}{r} W(\xi) - h\tau,$$

where the various quantities appearing in the formula are given by [\(56\)](#), [\(57\)](#), [\(58\)](#), [\(61\)](#) and [\(64\)](#). The only difference between the spherical and projective Hamiltonians is the term involving μ^2 . We also impose the extra constraint $\text{im}\langle \eta, z \rangle = 0$ and there will be extra curvature terms in the differential equations.

To find the curvature terms we need to pull-back the Fubini–Study form under the regularizing map $X = f(z)$, $X_{ij} = z_i^2 z_j$. The Fubini–Study metric on z -space is derived from the standard Hermitian metric on \mathbb{C}^3 by a formula analogous to [\(31\)](#). We can express its restriction to \mathcal{C} in terms of a tangent vector field as we did in [Lemma 14](#). The analogous formula to [\(32\)](#) is

$$(68) \quad \langle\langle V, W \rangle\rangle_{\text{FS}, \mathcal{C}} = \frac{\langle\langle V, e \rangle\rangle \langle\langle e, W \rangle\rangle}{\|z\|^4}, \quad V, W \in T_X \mathcal{G},$$

where $e(z)$ is a Fubini–Study unit vector field tangent to \mathcal{C} and normal to z . For example, observe that if $z \in \mathcal{C}_0 = \mathcal{C} \setminus 0$ then the vectors z, \bar{z}, T form a Hermitian-orthogonal complex basis for $T_z\mathbb{C}^3$, where

$$(69) \quad T = z \times \bar{z} = (z_{31}\bar{z}_{23} - z_{23}\bar{z}_{31}, z_{23}\bar{z}_{12} - z_{12}\bar{z}_{23}, z_{12}\bar{z}_{31} - z_{31}\bar{z}_{12}).$$

Hence we can take

$$e = \frac{\|z\|}{\|T\|} T = (z \times \bar{z}) / \|z\|.$$

This gives

$$(70) \quad \langle\langle \cdot, \cdot \rangle\rangle_{\text{FS}, \mathcal{C}} = \frac{\bar{\Sigma} \otimes \Sigma}{\|z\|^4},$$

where Σ is given by any of the formulas

$$(71) \quad \begin{aligned} \Sigma &= \frac{\langle\langle z \times \bar{z}, dz \rangle\rangle}{\|z\|} = \frac{\|z\| (z_{12} dz_{31} - z_{31} dz_{12})}{z_{23}} \\ &= \frac{\|z\| (z_{23} dz_{12} - z_{12} dz_{23})}{z_{31}} = \frac{\|z\| (z_{31} dz_{23} - z_{23} dz_{31})}{z_{12}}. \end{aligned}$$

For example, the first version is just $\Sigma = \langle\langle e, dz \rangle\rangle$ and the second is obtained by eliminating z_{23}, dz_{23} using the equations

$$z_{23}^2 = -z_{12}^2 - z_{31}^2 \quad \text{and} \quad z_{23} dz_{23} = -z_{12} dz_{12} - z_{31} dz_{31}.$$

Using these formulas, we find that the pull-back of the Fubini–Study metric on ${}^{\mathfrak{W}}$ is a conformal multiple of the Fubini–Study metric on \mathcal{C} .

Lemma 22. *The pull-back of the Fubini–Study metric on ${}^{\mathfrak{W}}$ is given by*

$$f^* \langle\langle \cdot, \cdot \rangle\rangle_{\text{FS}, \mathfrak{W}} = \lambda(z) \langle\langle \cdot, \cdot \rangle\rangle_{\text{FS}, \mathcal{C}},$$

where the conformal factor is

$$(72) \quad \lambda = \frac{4m_1 m_2 m_3 \rho_{12} \rho_{31} \rho_{23} \|z\|^2}{m |X(z)|^4} = \frac{4m_1 m_2 m_3 (\rho_{12} + \rho_{31} + \rho_{23}) \rho_{12} \rho_{31} \rho_{23}}{(m_1 m_2 \rho_{12}^2 + m_1 m_3 \rho_{31}^2 + m_2 m_3 \rho_{23}^2)^2}$$

and where $\rho_{ij} = |z_{ij}|^2$.

Proof. Equation (34) shows that we need to compute the pullback $f^*\sigma$, where σ is given by (35). Using the first formula for σ gives

$$f^*\sigma = 2z_{12}^2 z_{31} dz_{31} - 2z_{31}^2 z_{12} dz_{31} = 2z_{12} z_{31} z_{23} \Sigma.$$

Hence

$$f^* \langle\langle \cdot, \cdot \rangle\rangle_{\text{FS}, \mathfrak{W}} = \frac{m_1 m_2 m_3}{m |X(z)|^4} f^* \bar{\sigma} \otimes f^* \sigma = \frac{4m_1 m_2 m_3}{m |X(z)|^4} |z_{12}|^2 |z_{31}|^2 |z_{23}|^2 \bar{\Sigma} \otimes \Sigma.$$

Now use (57), (58) and (70) to get the formula in the proposition. \square

Similarly we can pull-back the Fubini–Study cometric on \mathcal{W} and compare it with the Fubini–Study cometric on \mathcal{C} . The formula analogous to (33) is

$$(73) \quad \|\eta\|_{\text{FS},\mathcal{C}}^2 = |\langle \eta, e \rangle|^2 = \frac{|\langle \eta, z \times \bar{z} \rangle|^2}{\|z\|^2}, \quad \eta \in T_{z,\text{pr}}^* \mathbb{C}^3.$$

This is a degenerate quadratic form, invariant under z -translation of η , which represents the Fubini–Study cometric on \mathcal{C} .

The next lemma relates this to the pull-back of the Fubini–Study cometric on \mathcal{W} and hence, to the shape kinetic energy.

Lemma 23. *The pull-back of the Fubini–Study cometric on \mathcal{W} is*

$$F^* \|\cdot\|_{\text{FS},\mathcal{W}}^2 = \lambda^{-1} \|\cdot\|_{\text{FS},\mathcal{C}}^2,$$

where λ is given by (72). Hence the shape kinetic energy in regularized coordinates is

$$\frac{1}{2} \lambda^{-1} \|\eta\|_{\text{FS},\mathcal{C}}^2 = \frac{1}{2} \frac{|\langle \eta, z \times \bar{z} \rangle|^2}{\lambda \|z\|^2}.$$

Proof. Equation (36) shows that we need to compute the pullback $F^* \alpha$, where α is given by (37). Using the second formula for α gives

$$\frac{|z_{23}|^2}{|X|^2} F^* \alpha = \frac{(\eta_{31} \bar{z}_{12} - \eta_{12} \bar{z}_{31}) z_{23}}{2 \bar{z}_{12} \bar{z}_{31} \bar{z}_{23}}$$

and there are two similar equations from the third and fourth formulas. Adding these gives

$$F^* \alpha = \frac{|X(z)|^2}{\|z\|^2} \langle \bar{\eta}, z \times \bar{z} \rangle.$$

Therefore,

$$F^* \|\eta\|_{\text{FS},\mathcal{W}}^2 = \frac{m|X(z)|^4 |\langle \eta, z \times \bar{z} \rangle|^2}{4m_1 m_2 m_3 \rho_{12} \rho_{31} \rho_{23} \|z\|^4} = \frac{m|X(z)|^4}{4m_1 m_2 m_3 \rho_{12} \rho_{31} \rho_{23} \|z\|^2} \|\eta\|_{\text{FS},\mathcal{C}}^2.$$

Comparing with the formula for λ completes the proof. \square

It follows from the lemma that we have an equivalent reduced, regularized Hamiltonian

$$\tilde{H}_\mu = \frac{\tau p_r^2}{2} + \frac{\tau \mu^2}{2r^2} + \frac{\tau \|\eta\|_{\text{FS},\mathcal{C}}^2}{2\lambda(z)r^2} - \frac{1}{r} W(\xi) - h\tau.$$

Some simplification is obtained by choosing the degree-zero homogeneous function λ as our time rescaling function instead of the function τ of (62), that is, by setting

$\tau = \lambda$. This gives the reduced, regularized Hamiltonian

$$(74) \quad \begin{aligned} \tilde{H}_\mu &= \frac{\lambda p_r^2}{2} + \frac{\lambda \mu^2}{2r^2} + \frac{\|\eta\|_{\text{FS}, \mathcal{C}}^2}{2r^2} - \frac{1}{r} W(\xi) - h\lambda \\ &= \frac{\lambda p_r^2}{2} + \frac{\lambda \mu^2}{2r^2} + \frac{|\langle \eta, z \times \bar{z} \rangle|^2}{2r^2 \|z\|^2} - \frac{1}{r} W(\xi) - h\lambda, \end{aligned}$$

where the new regularized shape potential is

$$(75) \quad W = \frac{4\sqrt{m} m_1 m_2 m_3 (\rho_{12} + \rho_{31} + \rho_{23}) (m_1 m_2 \rho_{31} \rho_{23} + m_1 m_3 \rho_{12} \rho_{23} + m_2 m_3 \rho_{12} \rho_{31})}{(m_1 m_2 \rho_{12}^2 + m_1 m_3 \rho_{31}^2 + m_2 m_3 \rho_{23}^2)^{3/2}}.$$

The factor of λ in the Fubini–Study two-form and the factor of λ^{-1} in the shape kinetic energy cancel out in the interior product defining the curvature term. Remembering the timescale factor λ we find that the curvature term is

$$(76) \quad T_{\text{curv}} = -\frac{2\mu\lambda}{r^2} i\eta,$$

which is added to the right hand side (that is to $-\partial H/\partial z$) of the Hamilton’s equation for $\dot{\eta}$.

Theorem 24. *The Hamiltonian flow of \tilde{H}_μ on $T^*\mathbb{R}^+ \times T^*\mathbb{C}_0^3$ has an invariant set $T^*\mathbb{R}^+ \times T_{\text{pr}, \mathcal{C}}^*\mathbb{C}^3$, where $\langle \eta, z \rangle = 0$ and $z_{12}^2 + z_{31}^2 + z_{23}^2 = 0$ with symplectic structure given by the restriction of the standard form minus $2\mu\lambda\Omega_{\text{FS}}$. The quotient of the restricted flow by the complex scaling symmetry and by \bar{z} -translations of η represents the three-body problem with zero total momentum and angular momentum μ , with regularized binary collisions, reduced by translations and rotations.*

The regularized, reduced Hamiltonian \tilde{H}_μ , together with the curvature term gives a system of differential equations on the fourteen-dimensional space $T^*(\mathbb{R}^+ \times \mathbb{C}^3)$ with variables (r, p_r, z, η) . The six-dimensional quotient space of $T^*\mathbb{R}^+ \times T_{\text{pr}, \mathcal{C}}^*\mathbb{C}^3$ is diffeomorphic to $T^*\mathbb{R}^+ \times T^*\mathbb{P}(\mathcal{C})$. Instead of writing these fourteen-dimensional differential equations, we will describe several ways to parametrize the regularized shape sphere $P(\mathcal{C})$ to arrive at lower-dimensional systems of equations.

6.3.1. Quadratic parametrization of the regularized shape sphere. We can parametrize \mathcal{C} using the same quadratic map $g : \mathbb{C}^2 \rightarrow \mathcal{C} \subset \mathbb{C}^3$ as in [Section 6.2.1](#):

$$z_{12} = 2ix_1x_2, \quad z_{31} = x_1^2 + x_2^2, \quad z_{23} = i(x_1^2 - x_2^2).$$

Since g is homogeneous with respect to complex scaling, it induces a map $g_{\text{pr}} : \mathbb{C}\mathbb{P}^1 \rightarrow P(\mathcal{C})$ from the projective line onto $P(\mathcal{C})$. Although g and the induced map g_{sph} of \mathcal{S}^3 in [Section 6.2.1](#) are both 2-to-1, the extra quotienting makes g_{pr} a

diffeomorphism. This shows again that $P(\mathcal{C})$ is diffeomorphic to the two-sphere. The same partially symplectic extension

$$G : T^*\mathbb{R}^+ \times T^*\mathbb{C}^2 \rightarrow T^*\mathbb{R}^+ \times \mathcal{C} \times \mathbb{C}^{3*}$$

restricts to a map $G : T^*\mathbb{R}^+ \times T_{\text{pr}}^*\mathbb{C}^2 \rightarrow T^*\mathbb{R}^+ \times T_{\text{pr},\mathcal{C}}^*\mathbb{C}^3$, where

$$T_{\text{pr}}^*\mathbb{C}^2 = \{(x, y) : \langle y, x \rangle = 0\} \quad \text{and} \quad T_{\text{pr},\mathcal{C}}^*\mathbb{C}^3 = \{(z, \eta) : z \in \mathcal{C}, \langle \eta, z \rangle = 0\}.$$

If we use (74) together with the formula (73) for the dual Fubini–Study metric, we obtain, after some simplification, the reduced, regularized Hamiltonian

$$(77) \quad \begin{aligned} \tilde{H}_\mu &= \frac{\lambda p_r^2}{2} + \frac{\lambda \mu^2}{2r^2} + \frac{|y_1 x_2 - x_1 y_2|^2}{4r^2} - \frac{1}{r} W(x) - h\lambda, \\ \rho_{12} &= |2x_1 x_2|^2, \quad \rho_{31} = |x_1^2 + x_2^2|^2, \quad \rho_{23} = |x_1^2 - x_2^2|^2, \end{aligned}$$

where $W(x)$ is still given by (75) and $\lambda(x)$ by (72) but with the ρ_{ij} replaced by the given expressions in terms of x .

We have the complex constraint $\langle y, x \rangle = 0$ and the system is invariant under complex scaling symmetry $(x, y) \rightarrow (kx, y/\bar{k})$, $k \in \mathbb{C}_0$. Applying the constraint and passing to the quotient space reduces the dimension from 10 to 6. As usual, Hamilton’s differential equations will have a curvature term

$$T_{\text{curv}} = -\frac{2\mu\lambda}{r^2} iy$$

added to the \dot{y} equation.

6.3.2. Dynamics in regularized affine coordinates. As in Section 5.4 we can use affine local coordinates on $\mathbb{C}\mathbb{P}^1$. Every projective point $[x_1, x_2] \in \mathbb{C}\mathbb{P}^1$ with $x_1 \neq 0$ has a representative of the form $[x_1, x_2] = [1, z] = [1, x + iy]$, where $x, y \in \mathbb{R}$. The appropriate momentum substitution is $y_1 = -\bar{z}\zeta$, $y_2 = \zeta$, where $\zeta = \alpha + i\beta \in \mathbb{C}^*$ is a momentum vector dual to z .

We get a Hamiltonian system with 6 degrees of freedom:

$$(78) \quad \begin{aligned} \tilde{H}_\mu &= \frac{\lambda p_r^2}{2} + \frac{\lambda \mu^2}{2r^2} + \frac{(1+x^2+y^2)^2(\alpha^2+\beta^2)}{4r^2} - \frac{1}{r} W(x, y) - h\lambda, \\ \rho_{12} &= 4(x^2+y^2), \quad \rho_{31} = (1+x^2-y^2)^2+4x^2y^2, \quad \rho_{23} = (1-x^2+y^2)^2+4x^2y^2. \end{aligned}$$

The Fubini–Study form becomes

$$\Omega_{\text{FS}} = \frac{dx \wedge dy}{(1+x^2+y^2)^2}.$$

Hamilton's equations with the curvature term are

$$\begin{aligned}
 \dot{r} &= \lambda p_r, & \dot{p}_r &= \frac{1}{r^3} \left[(1+x^2+y^2)^2 (\alpha^2 + \beta^2) + \lambda \mu^2 \right] - \frac{1}{r^2} W(x, y), \\
 \dot{x} &= \frac{(1+x^2+y^2)^2}{2r^2} \alpha, & \dot{y} &= \frac{(1+x^2+y^2)^2}{2r^2} \beta, \\
 (79) \quad \dot{\alpha} &= \frac{1}{r} W_x - \lambda_x \left[\frac{p_r^2}{2} + \frac{\mu^2}{2r^2} - h \right] - \frac{(1+x^2+y^2)(\alpha^2 + \beta^2)x}{r^2} + \frac{2\lambda\mu\beta}{r^2}, \\
 \dot{\beta} &= \frac{1}{r} W_y - \lambda_y \left[\frac{p_r^2}{2} + \frac{\mu^2}{2r^2} - h \right] - \frac{(1+x^2+y^2)(\alpha^2 + \beta^2)y}{r^2} - \frac{2\lambda\mu\alpha}{r^2}.
 \end{aligned}$$

6.3.3. Dynamics in regularized spherical coordinates. Instead of using projective or local affine coordinates, one can map the regularized shape sphere to the unit sphere in \mathbb{R}^3 . A particularly elegant way to do this is to use the diffeomorphism between \mathcal{C} and $\text{SO}(3)$ described in [Section 6.1.1](#).

Given $z \in \mathcal{C}$ we write $z = a + ib$, where $a, b \in \mathbb{R}^3$, and define $c = a \times b \in \mathbb{R}^3$. We saw that the matrix

$$A(z) = \frac{1}{s} \begin{bmatrix} a_{12} & b_{12} & c_{12}/s \\ a_{31} & b_{31} & c_{31}/s \\ a_{23} & b_{23} & c_{23}/s \end{bmatrix}$$

is in $\text{SO}(3)$, where $s^2 = |z|^2/2 = |a|^2 = |b|^2 = |c|$.

We will work homogeneously and define a map $g : \mathcal{C} \rightarrow \mathbb{R}^3$,

$$g(z) = c = \text{re}(z) \times \text{im}(z).$$

By homogeneity, there is an induced map $g_{\text{pr}} : \mathbb{P}(\mathcal{C}) \rightarrow \mathcal{S}(\mathbb{R}^3) \simeq \mathcal{S}^2$, where we view z and c as homogeneous coordinates with respect to complex and positive real scaling respectively.

The orthogonality of the matrix $A(z)$ can be used to derive some useful formulas. Since the rows as well as the columns are unit vectors, we find

$$\rho_{ij} = |z_{ij}|^2 = a_{ij}^2 + b_{ij}^2 = \frac{|c|^2 - c_{ij}^2}{|c|},$$

which gives the beautiful formulas

$$(80) \quad \rho_{12} = \frac{c_{31}^2 + c_{23}^2}{|c|}, \quad \rho_{31} = \frac{c_{12}^2 + c_{23}^2}{|c|}, \quad \rho_{23} = \frac{c_{12}^2 + c_{31}^2}{|c|},$$

for the homogeneous mutual distances. Similar formulas were given in [\[Lemaître 1964\]](#).

Next, consider the quantity

$$\bar{z}_{12}z_{31} = a_{12}a_{31} + b_{12}b_{31} + i(a_{12}b_{31} - a_{31}b_{12}) = (a_{12}, b_{12}) \cdot (a_{31}, b_{31}) + ic_{23}.$$

Using the orthogonality of the rows we can express this entirely in terms of c . We find

$$\bar{z}_{12}z_{31} = -\frac{c_{12}c_{31}}{|c|} + ic_{23}, \quad \bar{z}_{23}z_{12} = -\frac{c_{23}c_{12}}{|c|} + ic_{31}, \quad \bar{z}_{31}z_{23} = -\frac{c_{31}c_{23}}{|c|} + ic_{12}.$$

These last formulas allow us to write down local inverses for g_{pr} . Namely, consider the map $h_{12} : \mathbb{R}^3 \rightarrow \mathbb{C}^3$,

$$\begin{aligned} h_{12}(c) &= |c|\bar{z}_{12}(z_{12}, z_{31}, z_{23}) = |c|(\bar{z}_{12}z_{12}, \bar{z}_{12}z_{31}, \bar{z}_{12}z_{23}) \\ &= (c_{31}^2 + c_{23}^2, -c_{12}c_{31} + i|c|c_{23}, -c_{12}c_{23} - i|c|c_{31}). \end{aligned}$$

If $z_{12} \neq 0$, then $h_{12}(c)$ represents the same projective point in $\mathbb{P}(\mathcal{C})$ as z does so $h_{12}(c)$ give a local inverse for the projective map g_{pr} . There are similar partial inverses h_{31}, h_{23} .

To find the regularized, reduced Hamiltonian system, we need to convert the Fubini–Study metric and its dual norm (that is, cometric) to c -coordinates. The spherical analogue of the Fubini–Study metric is the spherical metric

$$\langle \cdot, \cdot \rangle_{\text{sph}} = \frac{|c|^2 \langle dc, dc \rangle - \langle dc, c \rangle \langle c, dc \rangle}{|c|^4} = \frac{|c \times dc|^2}{|c|^4},$$

where we are using the Euclidean inner product on \mathbb{R}^3 . We will see that

$$g^* \langle \cdot, \cdot \rangle_{\text{sph}} = 2 \langle \langle \cdot, \cdot \rangle \rangle_{\text{FS}, \mathcal{C}} = \frac{2|\langle z \times \bar{z}, dz \rangle|^2}{\|z\|^6}.$$

To see this, note that $z \times \bar{z} = -2ia \times b = -2ic$. Hence

$$dc = \frac{i}{2}(dz \times \bar{z} + z \times d\bar{z}).$$

This, together with the fact that $\langle z, \bar{z} \rangle = 0$ on \mathcal{C} leads, after some algebra, to the pull-back formula. Correspondingly, the Euclidean solid angle form pulls back to twice the Fubini–Study form, hence

$$\lambda \Omega_{\text{FS}, \mathbb{C}} = g^* \frac{\lambda}{2|c|^3} (c_1 dc_2 \wedge dc_3 + c_2 dc_3 \wedge dc_1 + c_3 dc_1 \wedge dc_2).$$

Let $\gamma \in \mathbb{R}^{3*}$ be a dual momentum vector to $c \in \mathbb{R}^3$. From the spherical scaling, we will have $\gamma \cdot c = 0$. If we split the momentum vector η into real and imaginary parts, $\eta = u + iv$, then the momenta transform via

$$u = b \times \gamma, \quad v = -a \times \gamma, \quad \text{with } \gamma = -\frac{u \cdot c}{|c|^2} a - \frac{v \cdot c}{|c|^2} b.$$

From this we find that the dual spherical norm

$$|\gamma|_{\text{sph}}^2 = |\gamma \times c|^2 = |c|^2 |\gamma|^2$$

corresponds to $\frac{1}{2} \|\cdot\|_{\mathbb{F}\mathbb{S}, \mathcal{C}}$. So we get the reduced, regularized Hamiltonian

$$(81) \quad \begin{aligned} \tilde{H}_\mu &= \frac{\lambda p_r^2}{2} + \frac{\lambda \mu^2}{2r^2} + \frac{|c|^2 |\gamma|^2}{r^2} - \frac{1}{r} W(c) - h\lambda, \\ \rho_{12} &= c_{31}^2 + c_{23}^2, \quad \rho_{31} = c_{12}^2 + c_{23}^2, \quad \rho_{23} = c_{12}^2 + c_{31}^2. \end{aligned}$$

Here we have used the homogeneity of the formulas to redefine ρ_{ij} to eliminate the factors of $|c|$. The curvature term is

$$(82) \quad T_{\text{curv}} = \frac{2\mu\lambda}{|c|r^2} \gamma \times c.$$

6.4. Visualizing the regularized shape sphere — Lemaître’s conformal map. The map of projective curves $f_{\text{pr}} : \mathbb{P}(\mathcal{C}) \rightarrow \mathbb{P}(\mathcal{W})$, induced by the squaring map, can be visualized as a map of the two-sphere into itself. Indeed this is the point of view taken by Lemaître [1964], but he arrived at it in a rather different way.

The map is a four-to-one branched covering map with octahedral symmetry (see Figure 4). The map is generically four-to-one except at the binary collision points, where it is two-to-one. In the figure, each octant of the regularized sphere maps to one or the other hemisphere of the unregularized sphere. Thus, for example, the north pole of the unregularized sphere (representing a Lagrangian, equilateral central configuration) has four preimages, which lie in alternate octants. Each binary collision point on the equator of the unregularized shape sphere has two preimages, which lie on a coordinate axes of the regularized sphere.

Using affine coordinates, it is possible to express the regularizing map as a map of the complex plane. For example, let $u = x_2/x_1$, where (x_1, x_2) are the parameters of Section 6.3.1. Choose a basis for \mathcal{W} so that the coordinates (ξ_1, ξ_2) satisfy $\xi_1 = X_{12}$, $\xi_2 = X_{23} - X_{31}$ and let $v = \xi_2/\xi_1$. Then it is easy to check that the regularizing map $X_{ij} = z_{ij}^2$ is given by the degree-four rational map

$$v = \frac{1}{2}(u^2 + u^{-2}).$$

The three-dimensional sphere of Figure 3 is just the preimage of the regularized two-sphere sphere in Figure 4 under a Hopf-map. Each point of the two-sphere determines a circle in the three-sphere. The three large tori in Figure 3 are the preimages of the collinear circles in the two-sphere (where the coordinate planes cut the sphere). The six tubes in Figure 3 are the preimages of small circles around the binary collision points (where the coordinate axes cut the sphere).

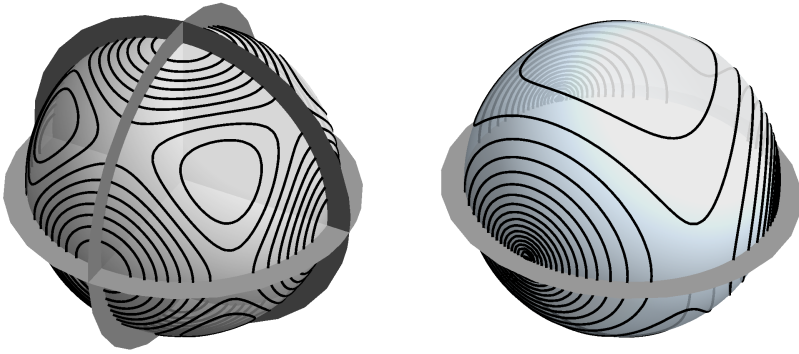


Figure 4. The regularizing map is a four-to-one branched cover of the two-sphere with octahedral symmetry. Each octant of the regularized shape sphere (left) maps onto a hemisphere of the unregularized shape sphere (right). The planes represent collinear configurations. The figure also shows level curves of the unregularized shape potential and their preimages in the equal mass case.

7. Blowing up triple collision

Our systematic use of the radial coordinate r together with the homogeneous coordinates used to describe the shape make it easy to implement McGehee’s method for blowing-up total collision. We need only rescale momenta and change the timescale. The changes can be made before or after reduction. The changes are noncanonical, so destroy the Hamiltonian character of the equations. We will describe the general method for the rotation-reduced and unreduced cases and then make some comments on the results of applying it to some of the Hamiltonians described above.

7.1. Before reduction. Consider a Hamiltonian of the general form

$$(83) \quad H(r, p_r, X, Y) = \frac{1}{2r^2} B(X)(Y, Y) - \frac{1}{r} V(X) + \left[\frac{1}{2} A(X) p_r^2 - C(X) \right]$$

when expanded in powers of r . This covers the rotation-unreduced Hamiltonian H_{sph} of Section 4 and the corresponding regularized Hamiltonians $\tilde{H}_{\text{sph}}(r, p_r, z, \eta)$ and $\tilde{H}_{\text{sph}}(r, p_r, x, y)$ of Section 6.2 (after changing the names of the variables). For the unregularized Hamiltonian H_{sph} we have $A(X) = 1$, $C(X) = 0$, while for the regularized Hamiltonians \tilde{H}_{sph} we have $A(X) = \tau(X)$, $C(X) = h \tau(X)$. The quantity $B(X)(Y, Y)$ represents the nonradial part of the kinetic energy. It is a quadratic form in Y , which we represent by a symmetric matrix $B(X)$ depending on X . The dependence of B on X must also be quadratic since H must be homogeneous of degree 0 with respect to the scaling $(X, Y) \mapsto (kX, (1/k)Y)$.

Let $f(r)$ be a positive, real-valued function. We will introduce a new timescale such that $' = f(r) \cdot$. The usual choice is McGehee's scaling factor $f_1(r) = r^{3/2}$ but we will also consider $f_2(r) = (r/(r+1))^{3/2}$, which has better behavior for large r . (With the first choice, solutions can reach $r = \infty$ in finite time.) For any such $f(r)$, we replace (p_r, Y) by rescaled momentum variables

$$(84) \quad v = \frac{f(r)p_r}{r}, \quad \alpha = \frac{f(r)Y}{r^2}.$$

The shape variable X remains the same. When we make these substitutions of independent and dependent variables in the Hamilton's differential equations resulting from (83), we get

$$(85) \quad \begin{aligned} r' &= A(X)vr, \\ v' &= \frac{1}{2}(1+r(\ln v)_r)A(X)v^2 + B(X)(\alpha, \alpha) - v(r)V(X) \\ X' &= B(X)\alpha, \\ \alpha' &= -\frac{1}{2}\left((1-r(\ln v)_r)A(X)v\alpha + A_X v^2 + B_X(\alpha, \alpha)\right) + v(r)V_X + rv(r)C_X, \end{aligned}$$

where $v(r) = f(r)^2/r^3$ and the subscripts denote differentiation. For McGehee's scaling $f(r) = f_1(r) = r^{3/2}$ we have $v(r) = 1$, $(\ln v)_r = 0$ and the equations simplify considerably. For $f_2(r)$ we have $v(r) = (1+r)^{-3}$ and both v and $(\ln v)_r$ are still smooth all the way down to $r = 0$.

Writing the energy equations $H_{\text{sph}} = h$ or $\tilde{H}_{\text{sph}} = 0$ in terms of the rescaled momenta gives

$$(86) \quad \frac{1}{2}A(X)v^2 + \frac{1}{2}B(X)(\alpha, \alpha) - v(r)V(X) = rv(r)C(X).$$

For example if we use the $r^{3/2}$ rescaling with H_{sph} , we have

$$A = 1, \quad B(X) = |X|^2 B_0, \quad C = 0, \quad V(X) = |X| \sum_{i < j} \frac{m_i, m_j}{|X_{ij}|},$$

where B_0 is the constant symmetric matrix (9). We get the blown-up differential equations

$$\begin{aligned} r' &= vr, & v' &= \frac{1}{2}v^2 - |X|^2 B_0(\alpha, \alpha) + V(X), \\ X' &= |X|^2 B_0 \alpha, & \alpha' &= -\frac{1}{2}v\alpha - B_0(\alpha, \alpha)X + V_X, \end{aligned}$$

with the energy relation $\frac{1}{2}v^2 + \frac{1}{2}B_0(X)(\alpha, \alpha) - V(X) = rh$.

The regularized equations arising from \tilde{H}_{sph} are considerably more complicated due to the $B(X)$ terms (or rather the $B(z)$ or $B(x)$ terms). Instead of writing them explicitly, we will just make some observations about them. Consider, for example,

$\tilde{H}_{\text{sph}}(r, p_r, x, y)$ from (66). $B(x)$ will be a complicated, 4×4 real matrix arising from the second term in (66). The phase space before blow-up is

$$T^*\mathbb{R}^+ \times T^*\mathbb{C}^2 \simeq (0, \infty) \times \mathbb{R} \times \mathbb{C}^2 \times \mathbb{C}^2.$$

In addition to the energy relation $\tilde{H}_{\text{sph}} = 0$, we have $\text{re}\langle y, x \rangle = 0$ and the scaling symmetry by positive real numbers so there is an induced flow on an quotient manifold of real dimension 7. After blow-up we have variables

$$(r, v, x, \alpha) \in [0, \infty) \times \mathbb{R} \times \mathbb{C}^2 \times \mathbb{C}^2,$$

where we have extended the flow to the *collision manifold* where $r = 0$, which is an invariant set for the differential equations. We have a real-analytic vector field on this manifold-with-boundary. Imposing the constraints and passing to the quotient under scaling gives a real-analytic vector field on a seven-dimensional manifold-with-boundary representing the planar three-body problem on a fixed energy manifold, with all binary collisions regularized and with triple collision blown-up. Note in particular that the regularization of binary collisions passes smoothly to the boundary.

We claim that if the timescale factor $f(r) = f_2(r) = (r/(r + 1))^{3/2}$ is used, then the differential equations define a complete flow on $[0, \infty) \times \mathbb{R} \times \mathbb{C}^2 \times \mathbb{C}^2$ and hence the induced seven-dimensional flow is complete. Since the differential equations are smooth, the only obstruction to completeness would be orbits that become unbounded in finite time. It is well-known that, with the usual timescale, such orbits do not exist for the three-body problem. It follows that if we use only bounded time-rescaling factors, the same will hold for the modified differential equations. McGehee’s factor $r^{3/2}$ is unbounded and it is possible for orbits to escape in finite time. Indeed, there are solutions of the three body problem for which $r(t) = O(t)$ as $t \rightarrow \infty$ with respect to the usual time-scale and these will reach infinity in finite rescaled time. The factor f_2 , while producing less elegant differential equations, eliminates this problem.

7.2. After reduction. The rotation-reduced Hamiltonians H_μ and their many regularized forms \tilde{H}_μ have the general form

$$(87) \quad H_\mu(r, p_r, X, Z) = \frac{1}{2r^2}[B(X)(Z, Z) + A(X)\mu^2] - \frac{1}{r}V(X) + [\frac{1}{2}A(X)p_r^2 - C(X)]$$

(after changing the names of the variables). The only new term here, when compared to the Hamiltonian of Section 7.1, is the quadratic term in the angular momentum μ . We have a momentum constraint $\langle Z, X \rangle = 0$ and there will be a curvature term, T_{curv} , added to the \dot{Z} equation. As in Section 7.1, for the unregularized Hamiltonians H_μ ,

we have

$$A(X) = 1, \quad C(X) = 0, \quad T_{\text{curv}} = -\frac{2\mu}{r^2}iZ,$$

while for the regularized Hamiltonians \tilde{H}_μ , we have

$$A(X) = \lambda(X), \quad C(X) = h\lambda(X), \quad T_{\text{curv}} = -\frac{2\mu\lambda}{r^2}iZ.$$

As in the last section, the variables X, Z can denote either homogeneous coordinates on the cotangent bundle of projective space, before or after Levi-Civita transformation, or they can be local holomorphic coordinates on the cotangent bundle of the shape sphere or of the regularized shape sphere $\mathbb{P}(\mathcal{C})$ (see the examples below). Our computations immediately below hold for all these cases.

We rescale time and the momenta as in (84) with Z replacing Y . We must also rescale angular momentum according to

$$(88) \quad \tilde{\mu} = \frac{f(r)\mu}{r^2}.$$

Then energy equations $H_\mu = h$ or $\tilde{H}_\mu = 0$ become

$$(89) \quad \frac{1}{2}A(X)(v^2 + \tilde{\mu}^2) + \frac{1}{2}B(X)(\alpha, \alpha) - v(r)V(X) = rv(r)C(X),$$

where

$$(90) \quad v = \frac{f^2}{r^3},$$

so that $v = 1$ for $f = r^{3/2}$ and $v = (1+r)^{-3}$ for $f = f_2$.

In order to express the differential equations succinctly, let

$$\tilde{K} = \frac{1}{2}A(X)(v^2 + \tilde{\mu}^2) + \frac{1}{2}B(X)(\alpha, \alpha)$$

denote the blown-up kinetic energy and let

$$(91) \quad \phi(r) = -\frac{1}{2}(1 - r(\ln v)_r).$$

Then the equations of motion are

$$(92) \quad \begin{aligned} r' &= A(X)vr, & v' &= \phi(r)A(X)v^2 + 2\tilde{K} - v(r)V, \\ \tilde{\mu}' &= \phi(r)A(X)v\tilde{\mu}, & X' &= B(X)\alpha, \\ \alpha' &= \phi(r)A(X)v\alpha - \tilde{K}_{X,+} + v(r)V_X + rv(r)C_X + T_{\text{curv}} \end{aligned}$$

where

$$T_{\text{curv}} = -2i\tilde{\mu}\alpha \quad \text{or} \quad -2i\tilde{\mu}\tau(X)\alpha$$

for the unregularized and regularized cases, respectively. We remark that the v' equation can also be written

$$v' = (\phi + 1)A(X)v^2 + B(X)(\alpha, \alpha) + A(X)\tilde{\mu}^2 - v(r)V(X).$$

In these equations, we are regarding $\tilde{\mu}$ as a new variable subject, by definition, to the constraint

$$(93) \quad \sqrt{r} \tilde{\mu} = \sqrt{v(r)} \mu,$$

where μ is the old angular momentum constant. This point of view is necessary to make the curvature term smooth at $r = 0$.

As in [Section 7.1](#), all functions of r extend smoothly to $r = 0$. If we start with one of the regularized Hamiltonians \tilde{H}_μ , then for the resulting differential equations, all binary collisions have been regularized and the triple collision blown-up. We obtain a flow on a manifold-with-boundary of dimension 5 after fixing μ , setting $\tilde{H}_\mu = 0$, imposing the constraint on $\tilde{\mu}$, the constraints that $X \in \mathcal{C}$ and $\langle Z, X \rangle = 0$ and passing to the quotient under complex scaling. Binary collisions are regularized for all values of μ and if the time rescaling is done using $f_2(r)$, the flows on these manifolds will be complete.

It is well-known that triple collisions are possible in the three-body problem only when $\mu = 0$. In this case, (93) shows that either $\tilde{\mu} = 0$ or $r = 0$. Both of these submanifolds are invariant sets for the dynamical system. The five-dimensional manifold-with-boundary with the above constraints and with $\tilde{\mu} = 0$ represents the closure of zero-angular-momentum three-body problem. The four-dimensional manifold where $\tilde{\mu} = r = 0$ forms the boundary. Even though orbit with $\mu \neq 0$ cannot have $r \rightarrow 0$, the part of the collision manifold $\{r = 0\}$ where $\tilde{\mu} \neq 0$ is relevant for studying low-angular-momentum orbits passing close to triple collision [[Moeckel 1984](#); [1989](#)].

We will now present a couple of versions of the regularized, reduced and blown-up differential equations for the three-body problem.

Example 25 (the blown-up regularized affine equations). In [Section 6.3.2](#), we used affine local coordinates on the regularized shape sphere to obtain a regularized Hamiltonian $\tilde{H}(z, \zeta)$ with 6 degrees of freedom. (We wrote $z = x + iy$, $\zeta = \alpha + i\beta$ in [Section 6.3.2](#).) Comparing with the general form (87) we have

$$\begin{aligned} A(X) &= \lambda(z), & B(X)(Z, Z) &= \frac{1}{2}(1 + |z|^2)^2 |\zeta|^2, \\ C(X) &= h\lambda(z), & V(X) &= W(z). \end{aligned}$$

Recall that λ and W are given by the formulas (72) and (75) with $\rho_{12} = 4|z|^2$, $\rho_{31} = |1 + z^2|^2$, $\rho_{23} = |1 - z^2|^2$. As per the preceding subsection, we continue to write the rescaled momentum variable as α (thus $\alpha = (f/r^2)\zeta$), trusting that

there will be no confusing with the previous use of α . The rescaled kinetic energy satisfies

$$2\tilde{K} = \lambda v^2 + \lambda \tilde{\mu}^2 + \frac{1}{2}(1 + |z|^2)^2 |\zeta|^2.$$

Then the regularized, blown-up equations read:

$$(94) \quad \begin{aligned} r' &= \lambda(z)vr, & v' &= \phi(r)\lambda(z)v^2 + 2\tilde{K} - v(r)W(z), \\ \tilde{\mu}' &= \phi(r)\lambda(z)v\tilde{\mu}, & z' &= \frac{1}{2}(1 + |z|^2)^2\alpha, \\ \alpha' &= \phi(r)\lambda(z)v\alpha - \tilde{K}_z + v(r)W_z + rv(r)h\tau_z(z) - 2i\tilde{\mu}\lambda(z)\alpha. \end{aligned}$$

The possibilities for $v(r)$, $\phi(r)$ are described in the previous subsection, in equations (90), (91).

We have 7 variables, $(r, v, \tilde{\mu}, z, \alpha) \in [0, \infty) \times \mathbb{R} \times \mathbb{R} \times \mathbb{C} \times \mathbb{C}$. The constraints are $\frac{1}{2}\lambda(z)(v^2 + \tilde{\mu}^2) + \frac{1}{4}(1 + |z|^2)^2 |\alpha|^2 - v(r)W(z) = rv(r)\lambda(z)h$ and $\sqrt{r}\tilde{\mu} = \sqrt{v(r)}\mu$.

Example 26 (the blown-up regularized spherical equations). In [Section 6.3.3](#), we used spherical-homogeneous variables $c = (c_1, c_2, c_3)$ to give a global representation of the regularized shape sphere. We found a regularized Hamiltonian

$$\tilde{H}_\mu(r, c, p_r, \gamma).$$

Comparing with the general form (87), we have

$$A(X) = \lambda(c), \quad B(X)(Z, Z) = 2|c|^2|\gamma|^2, \quad C(X) = h\lambda(c), \quad V(X) = W(c).$$

λ and W are given by the usual formulas with

$$\rho_{12} = c_{31}^2 + c_{23}^2, \quad \rho_{31} = c_{12}^2 + c_{23}^2, \quad \rho_{23} = c_{12}^2 + c_{31}^2.$$

With $\alpha = (f/r^2)\gamma$, the rescaled kinetic energy satisfies $2\tilde{K} = \lambda v^2 + \lambda \tilde{\mu}^2 + 2|c|^2 |\alpha|^2$.

Then the regularized, blown-up equations read:

$$(95) \quad \begin{aligned} r' &= \lambda(c)vr, & v' &= \phi(r)\lambda(c)v^2 + 2\tilde{K} - v(r)W(c), \\ \tilde{\mu}' &= \phi(r)\lambda(c)v\tilde{\mu}, & c' &= 2|c|^2\alpha, \\ \alpha' &= \phi(r)\lambda(c)v\alpha - \tilde{K}_c + vW_c + rv(r)h\lambda_c(c) + \frac{2\tilde{\mu}\lambda(c)}{|c|}\alpha \times c. \end{aligned}$$

We have 9 variables, $(r, v, \tilde{\mu}, c, \alpha) \in [0, \infty) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}_0^3 \times \mathbb{R}^3$. However, (c, α) are homogeneous variables. They satisfy $\langle \alpha, c \rangle = 0$ and the equations are invariant under the real scaling $(c, \alpha) \rightarrow (kc, (1/k)\alpha)$. Taking this into account, we have an induced system on the seven-dimensional quotient space $[0, \infty) \times \mathbb{R} \times \mathbb{R} \times T^*\mathcal{S}^2$. The energy and angular momentum constraints are

$$(96) \quad \frac{1}{2}\lambda(c)(v^2 + \tilde{\mu}^2) + |c|^2 |\alpha|^2 - v(r)W(c) = rv(r)\lambda(c)h$$

and $\sqrt{r}\tilde{\mu} = \sqrt{v(r)}\mu$, giving a subvariety of dimension 5.

A nice alternative to the quotient construction is just to observe that $\langle \alpha, c \rangle = 0$ implies that $|c|$ is invariant under the differential equations (95). Instead of quotienting by the scaling symmetry, we can simply restrict c to the unit sphere. Let

$$\mathcal{M}(h, \mu) = \{(r, v, \tilde{\mu}, c, \alpha) : |c| = 1, \langle \alpha, c \rangle = 0, \sqrt{r}\tilde{\mu} = \sqrt{v(r)}\mu, (96) \text{ holds}\}.$$

Then $\mathcal{M}(h, \mu)$ is a five-dimensional submanifold (or subvariety when $\mu = 0$) of $[0, \infty) \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}^3$, which is invariant under (95). The flow on $\mathcal{M}(h, \mu)$ globally represents the planar three-body problem reduced by translations and rotations, with all binary collisions regularized and with triple collision blown-up.

8. Summary

In Section 2 we recall the theory of symplectic reduction by an Abelian group G of a cotangent bundle T^*X of some configuration space X . The theory asserts that the reduced space is the manifold $T^*(X/G)$ — the cotangent bundle of the quotient space X/G . There is a twist: the symplectic structure of this cotangent bundle is typically not the standard one. Reduction depends on selecting a value μ of the “angular momentum” and the symplectic structure on $T^*(X/G)$ depends linearly on μ , becoming the standard one only when $\mu = 0$. In Sections 3, 4, and 5 we apply this reduction theory to the non-Abelian group G of orientation-preserving similarities acting on the phase space $T^*\mathbb{C}^3$ of the configuration space \mathbb{C}^3 of the planar three-body problem. In order to apply the theory we break the group up into its three Abelian parts: translations, scalings, and rotations. Reduction by these three subgroups make up the next three sections: Section 3 (translations), Section 4 (scalings), and Section 5 (rotations).

In Section 3 we use the linear map

$$L : \mathbb{C}^3 \rightarrow \mathbb{C}^3, \quad L(q_1, q_2, q_3) = (q_1 - q_2, q_2 - q_3, q_3 - q_1) = (Q_{12}, Q_{23}, Q_{31})$$

to form the quotient of \mathbb{C}^3 by translations. The image of L realizes the quotient of \mathbb{C}^3 by translations. This image is the two-dimensional complex subspace $\mathcal{W} \subset \mathbb{C}^3$ consisting of those Q ’s that satisfy the “triangle closure” relation

$$Q_{12} + Q_{23} + Q_{31} = 0.$$

In Sections 4 and 5, we form the quotient of the $\mathcal{W} = \text{im}(L)$ from Section 3 by the group of scalings (Section 4) and the group of rotations (Section 5). These two groups combine to form the Abelian group \mathbb{C}^* of nonzero complex numbers acting by scalar multiplication on the \mathbb{C}^3 of Q_{ij} ’s, and hence on its subspace $\text{im}(L)$. To form the quotient we must subtract out the triple collision point $0 \in \mathcal{W} \subset \mathbb{C}^3$ obtaining

$\mathcal{W}_0 := \mathcal{W} \setminus \{0\}$. We then implement the well-known fact that $\mathcal{W}_0 \simeq \mathbb{C}_0^2/\mathbb{C}^* = \mathbb{C}\mathbb{P}^1 = S^2 = \text{shape sphere}$.

Three-body dynamics does depend on overall size so we cannot possibly get a reduced dynamics on $T^*\mathbb{C}\mathbb{P}^1$. Instead we use the reduction by scale in Section 4 as a tool for coherently separating the size variable r from the shape variables X_{ij} . Together the r, X_{ij} form the “projective-homogeneous” coordinates of Section 5.

In Section 6.1 we introduce the Levi-Civita regularizing map $f : \mathbb{C}^3 \rightarrow \mathbb{C}^3$ to regularize all three binary collisions. The map sends z_{ij} to $X_{ij} = z_{ij}^2$. The map is \mathbb{C}^* -equivariant and so induces the following commutative diagram, which summarizes the paper:

$$\begin{array}{ccc}
 \mathbb{C}^3 & \xrightarrow{\text{Levi-Civita } f \text{ (Section 6.1)}} & \mathbb{C}^3 \\
 \uparrow & & \uparrow \text{Section 3} \\
 \mathcal{C} \setminus \{0\} & \xrightarrow{f \text{ restricted}} & \mathcal{W} \setminus \{0\} \\
 \downarrow \mathbb{C}^* & & \downarrow \mathbb{C}^*, \text{ Sections 4, 5} \\
 \mathbb{P}(\mathcal{C}) & \xrightarrow{\text{Lemaître}} & \mathbb{P}(\mathcal{W}) = \mathbb{C}\mathbb{P}^1 \\
 \parallel & & \parallel \\
 \text{regularized shape sphere} & & \text{shape sphere}
 \end{array}
 \tag{97}$$

The space $\mathcal{C} = \{z_{12}^2 + z_{23}^2 + z_{31}^2 = 0\}$ is an affine cone and is the pullback of $\mathcal{W} = \{Q_{12} + Q_{23} + Q_{31} = 0\}$ by the regularizing map f . The downward arrows are the standard projections used in defining projective space.

To obtain the phase spaces of the paper, take the cotangent bundles T^*X of each space X in the diagram (97), and cross with the space $T^*(0, \infty) = (0, \infty) \times \mathbb{R}$, which encodes the radial variable r and its momentum p_r . For angular momentum μ nonzero, the twist referred to in the first paragraph of this summary arises as the pull-back of the Fubini–Study form on $\mathbb{C}\mathbb{P}^1$, or of its Levi-Civita pull-back.

The separation into radial and shape variables begun in Section 4 allows us to make the final McGehee blow-up rescalings of time and momenta in Section 7. We end with a dynamical system, which is regular through all binary collisions and whose flow is complete.

We will close the paper with some pictures illustrating how the size and shape variables can help to visualize the behavior of orbits of the planar three-body problem. The figure-eight orbit of [Chenciner and Montgomery 2000] features three equal masses moving on a single curve in the plane, as shown in the top image of Figure 5. The other two images show how the size and shape of the triangle formed by the bodies varies using unregularized and regularized shape variables.

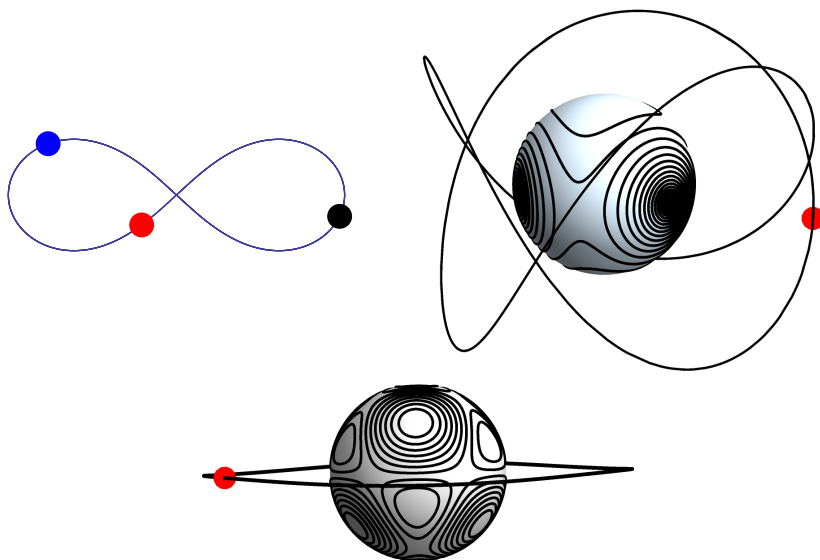


Figure 5. The famous figure-eight orbit of three equal masses. As the three bodies chase one another on the figure-eight curve in the plane, the size and shape vary as shown in the top right picture. The behavior seems much simpler in the regularized covering space (bottom).

The shape spheres are represented by the unit sphere in \mathbb{R}^3 . The size and shape are treated as spherical coordinates with the radial variable in \mathbb{R}^3 representing size $r + 1$ (so the unit spheres represent triple collision). For the figure-eight orbit, the size is nearly constant while the shape almost follows a level curve of the shape potential. The behavior of the regularized shape is surprisingly simple with the orbit close to a great circle on the sphere.

References

- [Abraham and Marsden 1978] R. Abraham and J. E. Marsden, *Foundations of mechanics*, 2nd ed., Benjamin/Cummings Publishing Co. Advanced Book Program, Reading, MA, 1978. [MR 81e:58025](#)
- [Albouy 2004] A. Albouy, “[Mutual distances in celestial mechanics](#)”, lecture notes, 2004, http://www.imcce.fr/fr/presentation/equipes/ASD/preprints/rep.2004/Albouy_%20Nankai09_2004.pdf.
- [Albouy and Chenciner 1998] A. Albouy and A. Chenciner, “[Le problème des \$n\$ corps et les distances mutuelles](#)”, *Invent. Math.* **131**:1 (1998), 151–184. [MR 98m:70017](#) [Zbl 0919.70005](#)
- [Chenciner 2011] A. Chenciner, “The Lagrange reduction of the N -body problem”, preprint, 2011.
- [Chenciner and Montgomery 2000] A. Chenciner and R. Montgomery, “[A remarkable periodic solution of the three-body problem in the case of equal masses](#)”, *Ann. of Math. (2)* **152**:3 (2000), 881–901. [MR 2001k:70010](#) [Zbl 0987.70009](#)

- [Heggie 1974] D. Heggie, “A global regularisation of the gravitational N -body problem”, *Celest. Mech.* **10** (1974), 217–241. [Zbl 0312.70015](#)
- [Jacobi 1843] C. Jacobi, “Sur l’elimination des noeuds dans le problème des trois corps”, *J. Reine Angew. Math.* **26** (1843), 115–131.
- [Kampen and Wintner 1937] E. R. V. Kampen and A. Wintner, “On a symmetrical canonical reduction of the problem of three bodies”, *Amer. J. Math.* **59**:1 (1937), 153–166. [MR 1507227](#) [Zbl 0015.42101](#)
- [Lagrange 1772] J.-L. Lagrange, “Essai sur le problème des trois corps”, in *Prix de l’académie royale des sciences de Paris*, tome IX, 1772. Reprinted as pages 229–331 in his *Œuvres complètes*.
- [Lemaître 1954] G. Lemaître, “Régularisation dans le problème des trois corps”, *Acad. Roy. Belgique. Bull. Cl. Sci.* (5) **40**. (1954), 759–767. [MR 16,964l](#) [Zbl 0057.16103](#)
- [Lemaître 1964] G. Lemaître, “The three body problem”, technical report CR-110, NASA, 1964.
- [Levi-Civita 1920] T. Levi-Civita, “Sur la régularisation du problème des trois corps”, *Acta Math.* **42**:1 (1920), 99–144. [MR 1555161](#) [JFM 47.0837.01](#)
- [Marsden and Weinstein 1974] J. Marsden and A. Weinstein, “Reduction of symplectic manifolds with symmetry”, *Rep. Mathematical Phys.* **5**:1 (1974), 121–130. [MR 53 #6633](#) [Zbl 0327.58005](#)
- [McGehee 1974] R. McGehee, “Triple collision in the collinear three-body problem”, *Invent. Math.* **27** (1974), 191–227. [MR 50 #11912](#) [Zbl 0297.70011](#)
- [Meyer 1973] K. R. Meyer, “Symmetries and integrals in mechanics”, pp. 259–272 in *Dynamical systems* (Salvador, Brazil, 1971), edited by M. M. Peixoto, Academic Press, New York, 1973. [MR 48 #9760](#) [Zbl 0293.58009](#)
- [Moeckel 1984] R. Moeckel, “Heteroclinic phenomena in the isosceles three-body problem”, *SIAM J. Math. Anal.* **15**:5 (1984), 857–876. [MR 86j:58047](#) [Zbl 0593.70009](#)
- [Moeckel 1989] R. Moeckel, “Chaotic dynamics near triple collision”, *Arch. Rational Mech. Anal.* **107**:1 (1989), 37–69. [MR 90i:58167](#) [Zbl 0697.70021](#)
- [Moeckel et al. 2012] R. Moeckel, R. Montgomery, and A. Venturelli, “From brake to syzygy”, *Arch. Ration. Mech. Anal.* **204**:3 (2012), 1009–1060. [MR 2917128](#) [Zbl 06102023](#)
- [Montgomery 2002] R. Montgomery, “Infinitely many syzygies”, *Arch. Ration. Mech. Anal.* **164**:4 (2002), 311–340. [MR 2004c:70018](#) [Zbl 1024.70005](#)
- [Murnaghan 1936] F. D. Murnaghan, “A symmetric reduction of the planar three-body problem”, *Amer. J. Math.* **58**:4 (1936), 829–832. [MR 1507204](#) [Zbl 0015.32406](#)
- [Saari 1984] D. G. Saari, “From rotations and inclinations to zero configurational velocity surfaces, I: A natural rotating coordinate system”, *Celestial Mech.* **33**:4 (1984), 299–318. [MR 86g:70003](#) [Zbl 0549.70003](#)
- [Simó and Susín 1991] C. Simó and A. Susín, “Connections between critical points in the collision manifold of the planar 3-body problem”, pp. 497–518 in *The geometry of Hamiltonian systems* (Berkeley, CA, 1989), edited by T. Ratiu, Math. Sci. Res. Inst. Publ. **22**, Springer, New York, 1991. [MR 92f:70007](#) [Zbl 0732.70006](#)
- [Smale 1970] S. Smale, “Topology and mechanics, I”, *Invent. Math.* **10** (1970), 305–331. [MR 46 #8263](#) [Zbl 0202.23201](#)
- [Waldvogel 1972] J. Waldvogel, “A new regularization of the planar problem of three bodies”, *Celest. Mech.* **6** (1972), 221–231. [Zbl 0242.70012](#)
- [Waldvogel 1982] J. Waldvogel, “Symmetric and regularized coordinates on the plane triple collision manifold”, *Celestial Mech.* **28**:1-2 (1982), 69–82. [MR 83m:70021](#) [Zbl 0551.70007](#)

RICHARD MOECKEL
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455
UNITED STATES

rick@math.umn.edu

RICHARD MONTGOMERY
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SANTA CRUZ
SANTA CRUZ, CA 95064
UNITED STATES

rmont@ucsc.edu

CANONICAL CLASSES AND THE GEOGRAPHY OF NONMINIMAL LEFSCHETZ FIBRATIONS OVER S^2

YOSHIHISA SATO

The Stipsicz conjecture on the fiber-sum decomposability of Lefschetz fibrations states that nonminimal Lefschetz fibrations over S^2 are irreducible with respect to fiber-sum decompositions. We can conclude that such Lefschetz fibrations are prime and fundamental. In this paper, we determine the canonical classes of nonminimal Lefschetz fibrations admitting spheres of square -1 whose total intersection number with generic fiber is big. As a consequence, we consider the Kodaira dimension and the geography problem of such Lefschetz fibrations.

1. Introduction

If a 4-dimensional manifold M admits some fibration structure, then we can understand its topology in detail. Elliptic surfaces, which are complex surfaces admitting elliptic fibrations whose generic fibers are smooth elliptic curves, were deeply studied by Kodaira, Kas, Moishezon and so on. Much is known about not only the topology of elliptic surfaces but also the differentiable structures on elliptic surfaces [Matsumoto 1986; Ue 1986; Donaldson 1987; Kametani and Sato 1994].

After that, symplectic structures are often studied as well as differentiable structures in 4-dimensional topology. In particular, *Lefschetz fibrations* have been studied in 4-dimensional symplectic topology since the latter half in the 1990's. A Lefschetz fibration is a smooth fibration of a smooth 4-manifold over a surface with finitely many critical points as complex analogues of Morse functions. Elliptic fibrations can be regarded as genus-1 Lefschetz fibrations. The importance of Lefschetz fibrations from the viewpoint of topology was reverified by Matsumoto [1996]. Lefschetz pencils and Lefschetz fibrations have played a major role in 4-dimensional symplectic topology by the support of the remarkable works of Donaldson [1998] and Gompf [1999], which imply that Lefschetz fibrations provide a topological

This research is supported by Grant-in-Aid for Scientific Research (C) (No. 23540096), Japan Society for the Promotion of Science.

MSC2010: primary 14J80, 57R17; secondary 14D06, 32Q65.

Keywords: Lefschetz fibrations, canonical class, geography, 4-manifolds, pseudoholomorphic curves, Gromov invariants.

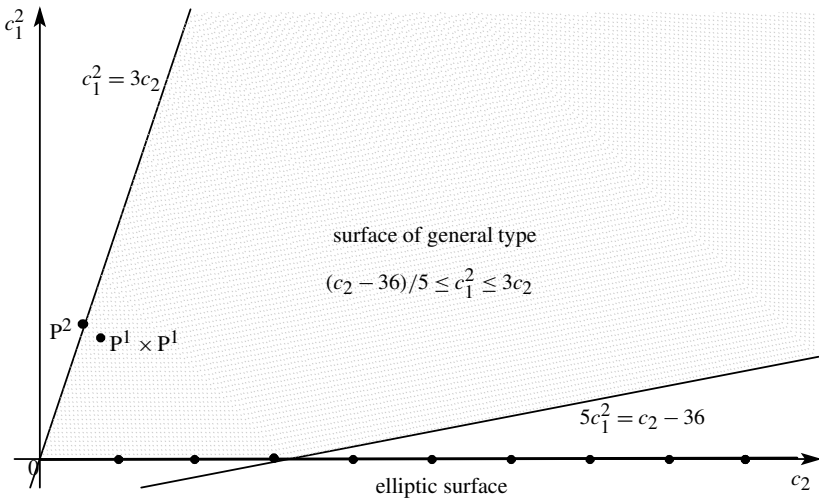


Figure 1. Geography of simply connected, minimal complex surfaces.

characterization of symplectic 4-manifolds and that most of symplectic 4-manifolds correspond to 4-manifolds with Lefschetz fibrations.

The geography problem in complex surfaces is the characterization of pairs of integers that are realized as the pairs (c_1^2, c_2) of Chern numbers of complex surfaces, and it is well studied in algebraic geometry. By the classification of complex surfaces due to Kodaira, a simply connected complex surface is rational, elliptic or of general type. We know completely the range that rational surfaces and elliptic surfaces cover in the (c_1^2, c_2) -plane. Minimal surfaces of general type must satisfy $c_1^2, c_2 > 0$ (Noether inequality) and $(c_2 - 36)/5 \leq c_1^2 \leq 3c_2$ (Bogomolov–Miyaoka–Yau inequality). In fact, the range of (c_1^2, c_2) of simply connected, minimal complex surfaces is as shown in [Figure 1](#).

A complex surface with even first Betti number b_1 is Kähler and so symplectic. Since any symplectic 4-manifold M with symplectic structure ω admits an ω -compatible almost-complex structure J , we can define Chern classes $c_1(TM, J), c_2(TM, J)$ for a symplectic 4-manifold M . Therefore, the geography for symplectic 4-manifolds comes into our mind. These problems are raised by McCarthy and Wolfson [\[1994\]](#):

- (1) Which pairs of integers are realized as (c_1^2, c_2) of a symplectic 4-manifold?
- (2) If there is a symplectic 4-manifold corresponding to a given lattice point (m, n) , how many symplectic 4-manifolds realize (m, n) as the pair (c_1^2, c_2) of Chern numbers?

Questions (1) and (2) are called the *geography problem* and the *botany problem*, respectively. Since simply connected complex surfaces are symplectic, the range of

(c_1^2, c_2) for symplectic 4-manifolds contains the range for simply connected complex surfaces. The remarkable works of Donaldson and Gompf suggest that the geography of symplectic 4-manifolds is nearly the same as one of Lefschetz fibrations. Every lattice point (c_1^2, c_2) , except finitely many lying in $(c_2 - 36)/5 \leq c_1^2 \leq 2c_2$, is realized as the total space of a Lefschetz fibration [Persson 1987]. On the other hand, Fintushel and Stern [1998] showed that there exists a minimal Lefschetz fibration that does not satisfy the Noether inequality. Stipsicz [1998] addressed the Bogomolov–Miyaoka–Yau inequality for Lefschetz fibrations. Therefore, there exists a difference between the range in the complex case and one in the symplectic case. See also [Gompf 1995; Stipsicz 1996].

Instead of investigating all of the objects, we restrict them to *prime* (or *irreducible*) things and examine these. Topologists often construct new manifolds by the cut-and-paste method. As one can make a new manifold from given manifolds by taking the connected sum, we can make a new Lefschetz fibration from given Lefschetz fibrations by taking the *fiber sum*. In the category of Lefschetz fibrations, Lefschetz fibrations that cannot be decomposed as any nontrivial fiber sum are prime (or irreducible) with respect to fiber sum decompositions. Therefore, it is natural and enough to investigate the geography of *irreducible* Lefschetz fibrations, that is, Lefschetz fibrations that cannot be decomposed as any nontrivial fiber sum. Lefschetz fibrations with smooth spheres of square -1 have the following properties:

- (1) Every projective complex surface admits the structure of a Lefschetz pencil and the notion of a Lefschetz pencil is important in the 4-dimensional topology. The blow-up of a Lefschetz pencil along the base locus yields a Lefschetz fibration with sections of square -1 . Conversely, by blowing down, we can obtain a Lefschetz pencil from a Lefschetz fibration with sections of square -1 .
- (2) Any nonminimal Lefschetz fibration over S^2 that has smooth spheres of square -1 cannot be decomposed as a nontrivial fiber sum.

Fact (2) was conjectured by Stipsicz [2001]. The Stipsicz conjecture asserting the minimality of Lefschetz fibrations with fiber sum decomposability was proved by Usher [2006] affirmatively. In [Sato 2006] the author gave an independent and easier proof of the Stipsicz conjecture in the case of fiber genus 2. Thus, nonminimal Lefschetz fibrations S^2 are irreducible with respect to the fiber sum decompositions and we can conclude that such Lefschetz fibrations are fundamental.

The canonical class K_M of a symplectic 4-manifold (M, ω, J) is defined by $K_M = -c_1(TM, J)$. Thus, if we determine the canonical class K_M , then we can calculate $c_1^2(TM, J) = K_M^2$. In this paper, we determine the canonical classes for nonminimal Lefschetz fibrations over S^2 . By using K_X^2 of nonminimal Lefschetz fibrations $X \rightarrow S^2$, we can calculate the symplectic Kodaira dimension κ^s and solve

the geography problem for nonminimal Lefschetz fibrations. By using the symplectic Kodaira dimension κ^s , we can answer a question of Endo [2008, Problem 4.13] on the diffeomorphism type of three symplectic 4-manifolds admitting nonminimal Lefschetz fibrations.

Smith [2001b] showed the finiteness of the geography of genus-2 Lefschetz pencils; that is, there are only finitely many possible Chern pairs (c_1^2, c_2) of genus-2 Lefschetz pencils. This implies that there is an upper bound on the number of singular fibers of a genus-2 Lefschetz pencil. In fact, the number of singular fibers of such a pencil is less than or equal to 40. From this, the following question comes to our mind:

Question 1-1 (Smith [Auroux 2006b]). Is there an upper bound (in terms of the genus only) on the number of singular fibers of a Lefschetz fibration admitting a section of square -1 ?

In [Sato 2008], the author generalized Smith's result on genus-2 Lefschetz pencils to the geography on nonminimal genus-2 Lefschetz fibrations over S^2 . In this paper, we consider the geography problem of nonminimal genus- g (≥ 3) Lefschetz fibrations over S^2 and show the finiteness of the geography of certain classes of nonminimal genus- g Lefschetz fibrations, which gives us a partial answer for Question 1-1. For example, in the case where nonminimal Lefschetz fibrations are hyperelliptic and have only (-1) -sections as smooth spheres of square -1 , we have:

Theorem 1-2. *For $g \geq 3$, there are only finitely many possible Chern pairs (c_1^2, c_2) of hyperelliptic genus- g Lefschetz fibrations with $2g - 2$ or $2g - 3$ sections of square -1 whose total spaces are neither the blow-up of a rational surface nor the blow-up of a ruled surface. As a consequence, there is an upper bound on the number of singular fibers of such a Lefschetz fibration. In fact, for any such hyperelliptic genus- g Lefschetz fibration $f : X \rightarrow S^2$, the number $\mu(f)$ of singular fibers of f satisfies*

$$\mu(f) \leq \frac{(8g - 9)(2g + 1)}{g - 1} + \sum_{h=1}^{\lfloor g/2 \rfloor} \frac{16g^2 - 11g - 8}{12h(g - h) - (2g + 1)}.$$

We can answer Question 1-1 in a generic situation; see Section 6. On the other hand, considering the fiber sum construction, we see that minimal Lefschetz fibrations can have arbitrarily many singular fibers.

The organization of this paper is as follows: In Sections 2–3, we recall the notion of Lefschetz fibrations over S^2 and give some examples of nonminimal Lefschetz fibrations. In Section 4, we consider the geography of symplectic 4-manifolds and Lefschetz fibrations. In Section 5, we determine the canonical classes of nonminimal Lefschetz fibrations and answer Endo's question. In Section 6, we show

the finiteness of the geography of nonminimal hyperelliptic Lefschetz fibrations. Furthermore, we consider the geography of nonminimal, nonhyperelliptic genus-3 Lefschetz fibrations.

2. Lefschetz fibrations over S^2

The definition of Lefschetz fibrations. A smooth map $f : X \rightarrow \Sigma$ from a closed, connected, oriented smooth 4-manifold X onto a closed, connected, oriented smooth 2-manifold Σ is said to be a *Lefschetz fibration* if f admits finitely many critical points $C = \{p_1, p_2, \dots, p_k\}$ on which f is injective and around which there are orientation-preserving complex coordinate neighborhoods such that locally f can be expressed as $f(z_1, z_2) = z_1^2 + z_2^2$. It is a consequence of this definition that $f|_{X \setminus C} : X \setminus C \rightarrow \Sigma \setminus f(C)$ is a smooth fiber bundle with fiber a closed oriented 2-manifold.

If a generic fiber that is the inverse image of a regular value has genus g , or equivalently if $f|_{X \setminus C}$ is a surface bundle with fiber a closed orientable surface of genus g , we refer to f as a *genus- g Lefschetz fibration*. Moreover, we assume that f is relatively minimal, that is, there is no fiber containing a sphere of square -1 . Two Lefschetz fibrations $f : X \rightarrow \Sigma$ and $f' : X' \rightarrow \Sigma'$ are *isomorphic* if there are diffeomorphisms $\Phi : X \rightarrow X'$ and $\varphi : \Sigma \rightarrow \Sigma'$ such that $f' \circ \Phi = \varphi \circ f$. In this paper, we will assume that a base space Σ is a 2-sphere.

A fiber containing a critical point is called a *singular fiber*, which is obtained by collapsing a simple closed curve, called a *vanishing cycle*, on a nearby generic fiber to a point. A singular fiber is called *reducible* or *irreducible* according to whether the corresponding vanishing cycle separates or does not separate in the generic fiber. In particular, if a vanishing cycle α separates the closed surface Σ_g of genus g into two components with genera h and $g - h$ ($1 \leq h \leq [g/2]$), then the reducible singular fiber corresponding to α is said to be of *type II_h* .

Let Γ_g be the mapping class group of genus g , namely the group of all isotopy classes of orientation-preserving self-diffeomorphisms of Σ_g . The local monodromy around a singular fiber of a Lefschetz fibration $f : X \rightarrow S^2$ is a positive Dehn twist τ_a along the corresponding vanishing cycle a . See [Figure 2](#). Since the base space of f is a 2-sphere, the product of all the local monodromies of f is trivial in Γ_g . Such a relation in Γ_g

$$t_{a_1} t_{a_2} \cdots t_{a_\mu} = 1$$

is called a *positive relation*, where a_1, a_2, \dots, a_μ are vanishing cycles of f and each t_{a_i} is the isotopy class of τ_{a_i} in Γ_g .

Isomorphism classes of Lefschetz fibrations are determined by the monodromy representations as follows:

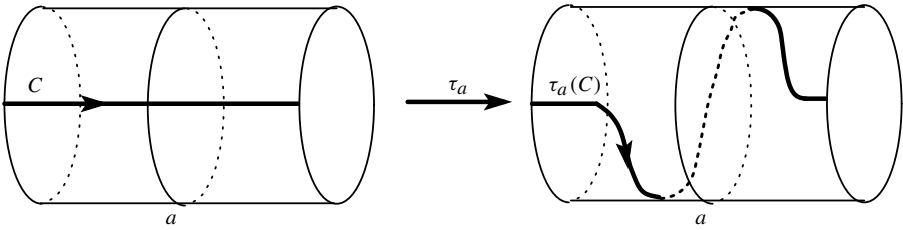


Figure 2. A positive Dehn twist.

Theorem 2-1 [Matsumoto 1996]. *Suppose that $g \geq 2$. Then, there is a one-to-one correspondence*

$$\left\{ \begin{array}{l} \text{isomorphism classes of Lefschetz} \\ \text{fibrations with } n \text{ singular fibers} \end{array} \right\} \longleftrightarrow \{\text{conjugacy classes of } \rho\},$$

where $\rho : \pi_1(S^2 \setminus \bigcup f(p_i), b_0) \rightarrow \Gamma_g$ is the monodromy representation.

From this theorem, it is well-known that a genus- g Lefschetz fibration is characterized by a positive relation $t_{a_1} t_{a_2} \cdots t_{a_\mu} = 1$ in Γ_g up to Hurwitz equivalence and simultaneous conjugation of all factors by a same element in Γ_g .

A *Lefschetz pencil* is a nonempty finite set $B = \{b_1, b_2, \dots, b_\ell\}$ of X , called the *base locus*, together with a smooth map $f : X \setminus B \rightarrow \mathbb{C}P^1$ such that each b_i has an orientation-preserving complex coordinate neighborhood in which locally f can be expressed as $f(z_1, z_2) = z_1/z_2$, and each critical point of f has a local coordinate neighborhood as a Lefschetz fibration. By the definitions of Lefschetz fibrations and pencils, the blow-up at the base locus points of a Lefschetz pencil yields a Lefschetz fibration over S^2 with sections of square -1 . It is well-known that every projective complex surface admits a Lefschetz pencil, which is generalized to symplectic 4-manifolds as follows:

Theorem 2-2 [Donaldson 1998]. *Every symplectic 4-manifold admits a Lefschetz pencil whose closed fibers are symplectic submanifolds.*

A result of Thurston [1976] on symplectic structures of surface-bundles over closed oriented surfaces can be generalized as follows to 4-manifolds admitting Lefschetz fibrations.

Theorem 2-3 [Gompf and Stipsicz 1999]. *Let $f : X \rightarrow S^2$ be a Lefschetz fibration and $[F]$ denote the homology class of the fiber. If $[F] \neq 0$ in $H_2(X; \mathbb{R})$, then X admits a symplectic structure such that fibers are symplectic submanifolds.*

If the fiber genus g is greater than 1, then the homology class of a generic fiber of f is not torsion in $H_2(X; \mathbb{Z})$, and so this theorem states that such an X admits a symplectic structure such that fibers of f are symplectic submanifolds.

From now on, we suppose that the fiber genus g is greater than 1 and we can use the symplectic topology. Then, combining the remarkable theorems of

Donaldson and Gompf gives the following topological characterization of symplectic 4-manifolds.

Corollary 2-4. *A 4-manifold X admits a symplectic structure if and only if it admits a Lefschetz pencil.*

Proof. By [Theorem 2-2](#), a symplectic 4-manifold X admits a Lefschetz pencil. If X admits a Lefschetz pencil, then the blow-up at the base locus points of a Lefschetz pencil yields a Lefschetz fibration $f : X \# n\overline{\mathbb{C}\mathbb{P}^2} \rightarrow S^2$ with sections of square -1 . Let F be a generic fiber of f and E a (-1) -section of f . Since $F \cdot E = 1$, the homology class of F is nontrivial in $H_2(X \# n\overline{\mathbb{C}\mathbb{P}^2}; \mathbb{R})$, and so it follows from [Theorem 2-3](#) that $X \# n\overline{\mathbb{C}\mathbb{P}^2}$ admits a symplectic structure ω . If X is rational or ruled, then X has a symplectic structure. Suppose that X is neither rational nor ruled. Then, by the (-1) -curve theorem [[Li and Liu 1995](#); [Taubes 1995](#); [1996](#)], we regard a smooth (-1) -section as an ω -symplectic sphere of square -1 . Hence, the symplectic blow-down of $X \# n\overline{\mathbb{C}\mathbb{P}^2}$ yields a symplectic structure on X . \square

Let $f_i : X_i \rightarrow S^2$ ($i = 1, 2$) be a genus- g Lefschetz fibration. Removing regular neighborhoods $N(F_1)$, $N(F_2)$ of generic fibers F_1 , F_2 in each, we glue these open remainders along their boundaries by using a fiber-preserving diffeomorphism $\varphi : \partial(X_1 - \text{Int } N(F_1)) \rightarrow \partial(X_2 - \text{Int } N(F_2))$ with $f_2 \circ \varphi = f_1$ on $\partial(X_1 - \text{Int } N(F_1))$. We denote the resulting 4-manifold by $X_1 \#_F X_2$, that is, $X_1 \#_F X_2 = (X_1 - \text{Int } N(F_1)) \cup_{\varphi} (X_2 - \text{Int } N(F_2))$. Then $X_1 \#_F X_2$ admits a genus- g Lefschetz fibration $f_1 \#_F f_2 : X_1 \#_F X_2 \rightarrow S^2$ associated to f_1 and f_2 . We call the genus- g Lefschetz fibration $f_1 \#_F f_2 : X_1 \#_F X_2 \rightarrow S^2$ the *fiber sum* of f_1 and f_2 . The diffeomorphism type of $X_1 \#_F X_2$ might depend on the choice of the gluing diffeomorphism φ . In fact, [Ozbagci and Stipsicz \[2000\]](#) constructed infinitely many Lefschetz fibrations as the fiber sums from the same building blocks by using various gluing diffeomorphisms. However, for the sake of brevity, we do not record those dependencies. By taking the fiber sums, we can obtain infinitely many genus- g Lefschetz fibrations. On the other hand, [Stipsicz \[2001\]](#) and [Smith \[2001a\]](#) showed that, if a Lefschetz fibration has a (-1) -section, then it cannot be decomposed as any nontrivial fiber sum. Furthermore, [Usher \[2006\]](#) showed that no nonminimal Lefschetz fibration can be decomposed as a nontrivial fiber sum (the Stipsicz conjecture).

Therefore, nonminimal Lefschetz fibrations are “irreducible” building blocks in the fiber sum construction. Thus, we consider nonminimal Lefschetz fibrations in this paper.

The signature of Lefschetz fibrations. The Hirzebruch signature theorem implies that the pair (c_1^2, c_2) of Chern numbers is determined by the signature and the Euler characteristic. So, when we consider the geography of Lefschetz fibrations later, it is important to calculate the signature and the Euler characteristic of a 4-manifold

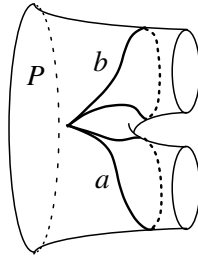


Figure 3. A pair of pants.

admitting a Lefschetz fibration. Every singular fiber of a genus- g Lefschetz fibration $f : X \rightarrow S^2$ contributes $+1$ to the Euler characteristic $e(X)$. If the fibration f has μ singular fibers, then we have $e(X) = 4(1 - g) + \mu$.

Compared with the calculation of the Euler characteristic, it is difficult to calculate the signature of X . Now we introduce two signature formulae. One is the Matsumoto–Endo formula for hyperelliptic Lefschetz fibrations and the other is the Smith formula for general (possibly nonhyperelliptic) Lefschetz fibrations. Let F_1, F_2, \dots, F_μ be singular fibers of $f : X \rightarrow S^2$. Let $N(F_i)$ denote the tubular neighborhood of F_i ($i = 1, 2, \dots, \mu$). We set $X_0 = X - \bigcup_{i=1}^\mu N(F_i)$. Then the restriction $f|_{X_0} : X_0 \rightarrow f(X_0)$ is the associated Σ_g -bundle over the punctured sphere. Since an irreducible singular fiber and a reducible singular fiber contribute 0 and -1 to the signature $\sigma(X)$, respectively, it follows from the Novikov additivity that we have

$$\sigma(X) = \sigma(X_0) - \sum_{h=1}^{\lfloor g/2 \rfloor} s_h,$$

where s_h denotes the number of singular fibers of type II_h . The signature $\sigma(X_0)$ of the bundle part X_0 can be calculated from the signature cocycle τ_g , which is a 2-cocycle of the Siegel modular group $\mathrm{Sp}(2g; \mathbb{Z})$ [Meyer 1973]. Let $P = S^2 - \bigsqcup_{i=1}^3 \mathrm{Int} D_i^2$ be a pair of pants and $E(\alpha, \beta) \rightarrow P$ the Σ_g -bundle defined by monodromies $\alpha, \beta \in \Gamma_g$.

Then, Meyer [1973] showed that for the signature of $E(\alpha, \beta)$ we have

$$\sigma(E(\alpha, \beta)) = -\tau_g(\alpha, \beta).$$

Since f has μ singular fibers, we can decompose the μ -punctured sphere $f(X_0)$ into $\mu - 2$ pairs $P_1, P_2, \dots, P_{\mu-2}$ of pants as in Figure 4.

Then it follows from Novikov additivity and Meyer’s theorem that we have

$$\sigma(X_0) = \sum_{i=1}^{\mu-2} \sigma(f^{-1}(P_i)) = - \sum_{i=1}^{\mu} \tau_g(t_{a_{i-1}} \cdots t_{a_2} t_{a_1}, t_{a_i}).$$

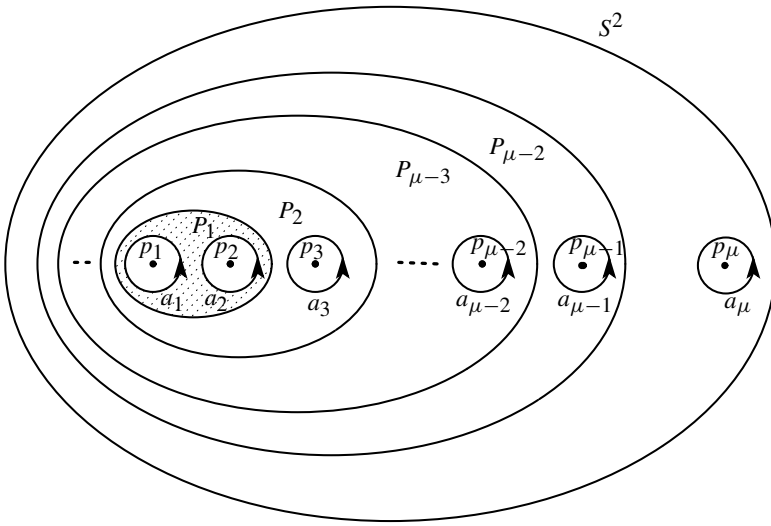


Figure 4. Decomposition of $f(X_0)$ into $\mu - 2$ pairs of pants.

Here a_1, a_2, \dots, a_μ are vanishing cycles of f and t_{a_0} denotes the identity map. Thus, in terms of the signature cocycle τ_g and monodromies $t_{a_1}, t_{a_2}, \dots, t_{a_\mu}$, the signature of X is given by

$$\sigma(X) = - \sum_{i=1}^{\mu} \tau_g(t_{a_{i-1}} \cdots t_{a_2} t_{a_1}, t_{a_i}) - \sum_{h=1}^{[g/2]} s_h.$$

The Matsumoto–Endo signature formula. A hyperelliptic Lefschetz fibration is a Lefschetz fibration whose monodromy representation ρ is equivalent to one taking isotopy classes commuting with the hyperelliptic involution $\iota : \Sigma_g \rightarrow \Sigma_g$. Since the hyperelliptic mapping class group Γ_2^{hyp} of genus 2 agrees with Γ_2 , every genus-2 Lefschetz fibration is hyperelliptic.

When we restrict the signature cocycle τ_g to the hyperelliptic mapping class group Γ_g^{hyp} , its cohomology class $[\tau_g^H] \in H^2(\Gamma_g^{\text{hyp}}; \mathbb{Z})$ is of finite order. So we can calculate the terms of signature cocycles by the coboundary maps called Meyer’s functions. Matsumoto [1996] and Endo [2000] calculated Meyer’s functions and obtained the signature formula for hyperelliptic Lefschetz fibrations.

Theorem 2-5 [Matsumoto 1996; Endo 2000]. *Suppose that $f : X \rightarrow S^2$ is a genus- g hyperelliptic Lefschetz fibration with n_0 irreducible singular fibers and s_h singular fibers of type II_h ($h = 1, 2, \dots, [g/2]$). Then, we have*

$$\sigma(X) = - \frac{g+1}{2g+1} n_0 + \sum_{h=1}^{[g/2]} \left(\frac{4h(g-h)}{2g+1} - 1 \right) s_h.$$

Smith’s signature formula. Smith obtained the signature formula for (possibly nonhyperelliptic) Lefschetz fibrations by using the geometry of the moduli space of stable curves. We denote the Deligne–Mumford compactified moduli space of stable curves of genus g by $\overline{\mathcal{M}}_g$. Let $f : X \rightarrow S^2$ be a genus- g Lefschetz fibration. Then, we can have a symplectic structure on X such that each fiber $f^{-1}(x)$ is a pseudoholomorphic curve. Since a 2-dimensional almost-complex structure is integrable, each fiber $f^{-1}(x)$ determines a point in the Deligne–Mumford compactified moduli space $\overline{\mathcal{M}}_g$.

Thus we can define the moduli map $\phi_f : S^2 \rightarrow \overline{\mathcal{M}}_g$ of f by

$$\phi_f(x) := [f^{-1}(x)] \in \overline{\mathcal{M}}_g \quad \text{for all } x \in S^2.$$

In particular, if $f : X \rightarrow \mathbb{C}P^1$ is holomorphic, then the image $\phi_f(\mathbb{C}P^1)$ is a rational curve in $\overline{\mathcal{M}}_g$.

Theorem 2-6 [Smith 1999]. *For any genus- g Lefschetz fibration $f : X \rightarrow S^2$ with μ singular fibers, namely $\mu = n_0 + \sum_{h=1}^{\lfloor g/2 \rfloor} s_h$, the signature of X is given by*

$$\sigma(X) = 4\langle c_1(\lambda), [\phi_f(S^2)] \rangle - \mu,$$

where $\lambda \rightarrow \overline{\mathcal{M}}_g$ denotes the Hodge bundle with fiber $\wedge^g H^0(C; K_C)$, the determinant line above $[C]$.

For a projective fibration $f : X \rightarrow \mathbb{C}P^1$, this theorem follows from Mumford’s formula. Smith’s formula is a generalization of Atiyah’s formula for smooth fibrations, and related work by Meyer.

3. Examples of Lefschetz fibrations

Let Γ_g be the mapping class group of Σ_g . For elements $\varphi, \psi \in \Gamma_g$, the product $\psi \cdot \varphi$ (or $\psi\varphi$) stands for applying φ first and then ψ .

Let $c_1, c_2, \dots, c_{2g+1}$ be the curves on Σ_g illustrated in Figure 5. The isotopy classes of the positive Dehn twists $\tau_{c_1}, \tau_{c_2}, \dots, \tau_{c_{2g+1}}$ along $c_1, c_2, \dots, c_{2g+1}$ are

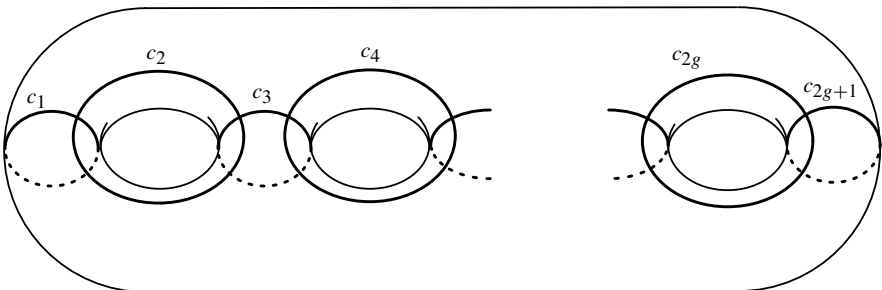


Figure 5. Lickorish generators.

Lickorish generators of the mapping class group Γ_g of genus g . For the sake of brevity, we denote the isotopy class of the positive Dehn twist τ_a along the curve a by the same symbol a .

It is well-known that Γ_g has the following positive relations:

$$\begin{aligned} W_1 &: (c_1 \cdot c_2 \cdots c_{2g} \cdot c_{2g+1} \cdot c_{2g+1} \cdot c_{2g} \cdots c_2 \cdot c_1)^2 = 1, \\ W_2 &: (c_1 \cdot c_2 \cdots c_{2g} \cdot c_{2g+1})^{2g+2} = 1, \\ W_3 &: (c_1 \cdot c_2 \cdots c_{2g})^{4g+2} = 1. \end{aligned}$$

From these positive relations, we can construct hyperelliptic genus- g Lefschetz fibrations with only irreducible singular fibers and with sections of square -1 . Furthermore, these Lefschetz fibrations are double branched covers of the Hirzebruch surfaces and so holomorphic. The total space of the Lefschetz fibration corresponding to W_1 is diffeomorphic to $\mathbb{C}P^2 \# (4g + 5)\overline{\mathbb{C}P^2}$.

Examples of nonminimal genus-2 Lefschetz fibrations. The Hirzebruch surface $\mathbb{F}_n = \mathbb{P}(\mathbb{C}_{\mathbb{C}P^1} \oplus \mathbb{C}_{\mathbb{C}P^1}(n))$ has two disjoint holomorphic sections Δ_n and Δ_{-n} of square $\pm n$.

(1) $M_1 = \mathbb{C}P^2 \# 13\overline{\mathbb{C}P^2}$: The positive relation $W_1 : (c_1 \cdot c_2 \cdot c_3 \cdot c_4 \cdot c_5^2 \cdot c_4 \cdot c_3 \cdot c_2 \cdot c_1)^2 = 1$ describes the genus-2 Lefschetz fibration on the rational surface M_1 obtained as a double covering of \mathbb{F}_0 branched along a smooth algebraic curve in the linear system $|6\Delta + 2F|$. This fibration is obtained from the composition of the covering projection with the bundle projection $\mathbb{F}_0 \rightarrow S^2$ and has 20 irreducible singular fibers and sections of square -1 .

(2) $M_2 = K3 \# 2\overline{\mathbb{C}P^2}$: The positive relation $W_2 : (c_1 \cdot c_2 \cdot c_3 \cdot c_4 \cdot c_5)^6 = 1$ describes the genus-2 Lefschetz fibration on M_2 obtained as a double covering of $\mathbb{F}_1 = \mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$ branched along a smooth algebraic curve in the linear system $|6L|$, where L is a line in $\mathbb{C}P^2$ avoiding the blown-up point. This fibration has 30 irreducible singular fibers and sections of square -1 .

(3) $M_3 = H'(1)$ (Horikawa surface): The positive relation $W_3 : (c_1 \cdot c_2 \cdot c_3 \cdot c_4)^{10} = 1$ describes the genus-2 Lefschetz fibration on M_3 obtained as a double covering of \mathbb{F}_2 branched along the disjoint union of a smooth curve in the linear system $|5\Delta_2|$ and Δ_{-2} . This fibration has 40 irreducible singular fibers and a section of square -1 . This section is a lift of the component of the branched set coming from Δ_{-2} . On the other hand, a fiber sum of two copies of the rational genus-2 Lefschetz fibration $\mathbb{C}P^2 \# 13\overline{\mathbb{C}P^2} \rightarrow S^2$ is a genus-2 Lefschetz fibration, which has 40 irreducible singular fibers and the total space is homeomorphic to $H'(1)$ but not diffeomorphic.

(4) $S^2 \times T^2 \# 4\overline{\mathbb{C}P^2}$: Matsumoto showed that $S^2 \times T^2 \# 4\overline{\mathbb{C}P^2}$ has a genus-2 Lefschetz fibration with 6 irreducible singular fibers and 2 reducible singular fibers. This also has a section of square -1 . The positive relation describing this fibration is $(\alpha_1 \cdot \sigma \cdot \alpha_2 \cdot \alpha_3)^2 = 1$, where $\alpha_1, \alpha_2, \alpha_3$ and σ are given by positive Dehn twists along the curves indicated in Figure 6.

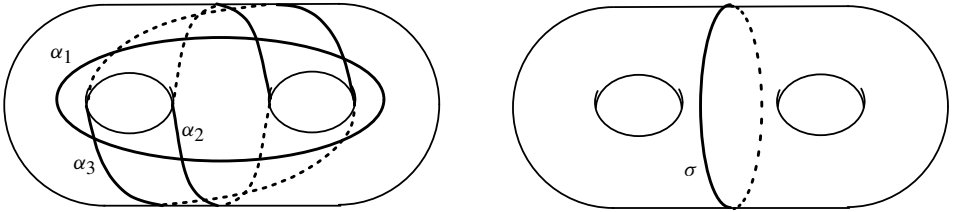


Figure 6

(5) Auroux’s genus-2 Lefschetz fibration : Auroux [2003] constructed the interesting genus-2 Lefschetz fibration $f : X \rightarrow \mathbb{C}P^1$ with 28 irreducible singular fibers and one reducible singular fiber. This fibration is nonminimal but does not admit section of square (-1) . See [Sato 2008]. The positive relation corresponding to it is given as follows:

$$\sigma \cdot (c_3 \cdot c_4 \cdot c_5 \cdot c_2 \cdot c_3 \cdot c_4 \cdot c_1 \cdot c_2 \cdot c_3)^2 \cdot (c_1 \cdot c_2 \cdot c_3 \cdot c_4 \cdot c_5 \cdot c_5 \cdot c_4 \cdot c_3 \cdot c_2 \cdot c_1) = 1.$$

For other examples of nonminimal genus-2 Lefschetz fibrations, see [Sato 2008].

Examples of nonminimal genus-3 Lefschetz fibrations.

(1) M_1, M_2 and M_3 corresponding to positive relations W_1, W_2 and W_3 for $g = 3$ have nonminimal, hyperelliptic and holomorphic genus-3 Lefschetz fibrations.

(2) $S^2 \times T^2 \# 8\overline{\mathbb{C}P^2}$: This has a nonhyperelliptic genus-3 Lefschetz fibration with positive relation $(\alpha_1 \cdot \alpha_2 \cdot \alpha_3 \cdot \alpha_4 \cdot \beta_1^2 \cdot \beta_2^2)^2$ indicated in Figure 7. This fibration also has a section of square -1 .

(3) Fuller’s example : Fuller constructed a nonhyperelliptic and nonholomorphic genus-3 Lefschetz fibration with positive relation

$$(\beta_1 \cdot \beta_2 \cdot c_4 \cdot c_3 \cdot c_2 \cdot c_1 \cdot c_5 \cdot c_4 \cdot c_3 \cdot c_2 \cdot c_6 \cdot c_5 \cdot c_4 \cdot c_3 \cdot (c_1 \cdots c_6)^{10}) = 1,$$

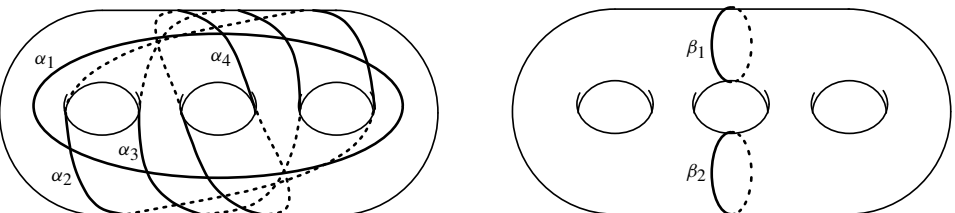


Figure 7

where β_1 and β_2 are given by positive Dehn twists along curves indicated in Figure 7. This fibration also has a section of square -1 .

(4) $T^4 \# 4\overline{\mathbb{C}P^2}$: Smith [2001c] showed that $T^4 \# 4\overline{\mathbb{C}P^2}$ has a hyperelliptic and holomorphic genus-3 Lefschetz fibration with 12 irreducible singular fibers and 4 sections of square -1 . This fibration is obtained by using the inverse of the usual Kummer construction of a $K3$ surface which is elliptically fibered over S^2 with 16 disjoint (-2) -spheres containing 4 sections and 12 singular fibers.

(5) Fermat surface of degree 4 : The Fermat surface of degree 4 is the smooth hypersurface in $\mathbb{C}P^3$ defined by the equation $z_0^4 + z_1^4 + z_2^4 + z_3^4 = 0$. Kuno [2010] proved that this surface admits a genus-3 Lefschetz pencil with 4 base locus points. The blow-up of this surface at the base locus points yields a nonminimal, nonhyperelliptic, holomorphic genus-3 Lefschetz fibration with only 36 irreducible singular fibers and 4 sections of square -1 . See [Kuno 2010] for its monodromies.

Examples of nonminimal genus- g Lefschetz fibrations. Endo [2008] generalized some parts of Chakiris' construction of holomorphic genus-2 Lefschetz fibrations topologically to give many examples of nonminimal hyperelliptic Lefschetz fibrations of arbitrary genus. Their examples are given in terms of positive relations in mapping class groups. Now, from Endo's list, we introduce some examples that we investigate in Section 5.

For simple closed curves c, a_1, a_2, \dots, a_r on Σ_g and

$$W = a_r^{\varepsilon_r} \cdots a_2^{\varepsilon_2} a_1^{\varepsilon_1} \quad (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_r \in \{\pm 1\}),$$

we put $w(c) := \tau_{a_r}^{\varepsilon_r} \cdots \tau_{a_2}^{\varepsilon_2} \tau_{a_1}^{\varepsilon_1}(c)$. Furthermore, for a factorization $V = c_{i_1} c_{i_2} \cdots c_{i_r}$, we put $wV := w(c_{i_1}) w(c_{i_2}) \cdots w(c_{i_r})$.

We define words I, J, C_I, C_{II}, P, Q and R in Γ_g^{hyp} as follows. Here, d denotes the boundary curve of a regular neighborhood of $c_1 \cup c_2 \cup \cdots \cup c_{2\lfloor g/2 \rfloor}$.

$$I := c_1 \cdot c_2 \cdots c_{2g} \cdot c_{2g+1}^2 \cdot c_{2g} \cdots c_2 \cdot c_1,$$

$$J := (c_1 \cdot c_2 \cdots c_{2g})^{2g+1},$$

$$C_I := (c_1 \cdot c_2 \cdots c_{2g+1})^{2g+2},$$

$$C_{II} := (c_1 \cdot c_2 \cdots c_{2g})^{4g+2}, \quad \text{namely } C_{II} = J^2,$$

$$P := d \cdot w(c_{g+1} \cdots c_3 \cdot c_2) \cdots w(c_{2g} \cdots c_{g+2} \cdot c_{g+1}) \\ \cdot (c_{g+1} \cdots c_3 \cdot c_2) \cdots (c_{2g} \cdots c_{g+2} \cdot c_{g+1}), \quad (g \text{ is even})$$

where $W := (c_1 \cdot c_2 \cdots c_g)^{-(g+1)}$. When g is even, the words Q and R are defined by

$$Q := (c_1 \cdot c_2 \cdots c_{2g+1})^{g+1} \cdot d \cdot w_1(c_{g+1}) \cdot w_2(c_{g+2}) \cdots w_{g+1}(c_{2g+1}),$$

$$R := d \cdot w_1(c_{g+1}) \cdot w_2(c_{g+2}) \cdots w_{g+1}(c_{2g+1}) \cdot (c_{2g+1} \cdots c_2 \cdot c_1)^{g+1},$$

where $W_i := (c_{i+g-1} \cdots c_{i+1} c_i)^{-1}$ for each $i \in \{1, 2, \dots, g+1\}$. When g is odd, the words Q and R are defined by

$$\begin{aligned} Q &:= (c_1 \cdot c_2 \cdots c_{2g+1})^{g+2} \cdot d \cdot (c_{g-1} \cdot c_2 \cdot c_1)^2 \cdot w_1(c_g) \cdot w_2(c_{g+1}) \cdots w_{g+2}(c_{2g+1}), \\ R &:= c_1 \cdot c_2 \cdots c_{2g+1} \cdot d \cdot (c_{g-1} \cdots c_2 \cdot c_1)^2 \cdot w_1(c_g) \cdot w_2(c_{g+1}) \cdots w_{g+2}(c_{2g+1}) \\ &\quad \cdot (c_{2g+1} \cdots c_2 \cdot c_1)^{g+1}, \end{aligned}$$

where $W_i := c_i^{-1} c_{i+1}^{-1} \cdots c_{i+g-2}^{-1}$ for each $i \in \{1, 2, \dots, g+2\}$.

Endo [2008] proved that the words I, J, C_I, C_{II}, P, Q and R , as products of positive Dehn twists, satisfy the following positive relations:

$$\begin{aligned} I^2 &= 1, & C_I &= 1, & C_{II} &= 1, \\ P^2 &= 1, & Q &= 1, & PI &= 1, & PJ &= 1, & RI &= 1 & (g \text{ is even}) \\ Q &= 1, & R &= 1 & (g \text{ is odd}). \end{aligned}$$

From these positive relations but for $RI = 1$, we can obtain nonminimal hyperelliptic Lefschetz fibrations admitting sections of square (-1) ; see [Endo 2008]. Of course, the three relations $I^2 = 1, C_I = 1$ and $C_{II} = 1$ are the same as the positive relations W_1, W_2, W_3 .

4. The geography of symplectic 4-manifolds

Smooth closed 4-manifolds can be endowed with different structures as complex structures, almost-complex structures and symplectic structures.

A *symplectic structure* on a smooth 4-manifold M is a closed 2-form ω that is nondegenerate as a bilinear form on each tangent space $T_x M$. A *symplectic 4-manifold* is a smooth 4-manifold M together with a symplectic structure ω .

An *almost-complex structure* on M is a bundle endomorphism $J : TM \rightarrow TM$ of the tangent bundle TM with $J^2 = -id_{TM}$. Since (TM, J) is regarded as a \mathbb{C}^2 -bundle over M , we can define *Chern classes*

$$c_1(M, J) := c_1(TM, J), \quad c_2(M, J) := c_2(TM, J).$$

Furthermore, it is well-known that any symplectic 4-manifold (M, ω) admits an ω -compatible almost-complex structure, which is an almost-complex structure J such that $g(u, v) := \omega(u, Jv)$ ($u, v \in TM$) is a Riemannian metric.

A smooth map $\varphi : \Sigma \rightarrow M$ from a possibly disconnected compact Riemann surface (Σ, j) to (M, J) is said to be *J-holomorphic* if the differential $d\varphi$ satisfies

$$d\varphi \circ j = J \circ d\varphi.$$

We call the image $\varphi(\Sigma)$ a *J-holomorphic curve* or a *pseudoholomorphic curve* with respect to J . If C is a pseudoholomorphic curve with respect to an ω -compatible

almost complex structure, then C is also ω -symplectic. A pseudoholomorphic curve on a symplectic 4-manifold is one of the most important tools in modern symplectic 4-dimensional topology and has a lot in common with holomorphic curves. For example, two distinct pseudoholomorphic curves intersect discretely and positively. Hence, the algebraic intersection number between two distinct pseudoholomorphic curves stands for the geometric intersection number.

Many typical examples of simply connected 4-manifolds are given by complex surfaces. Since they are simply connected, they are also *Kähler*, that is, symplectic 4-manifolds admitting symplectic structures whose compatible almost-complex structures are integrable. The geography of simply connected minimal complex surfaces, that is, the range of Chern pairs (c_1^2, c_2) of such complex surfaces, is as in [Figure 1](#). The boundary of the range is given by the Noether line ($5c_1^2 = c_2 - 36$) and the Bogomolov–Miyaoaka–Yau line ($c_1^2 = 3c_2$).

Symplectic 4-manifolds have a lot in common with complex surfaces: pseudoholomorphic curves play a role as holomorphic curves on complex surfaces. Donaldson’s theorem ([Theorem 2-2](#)) gives us a symplectic version of the ample divisor. By Taubes’ theorem [[1995](#); [1996](#)] on the Gromov–Witten invariants Gr_T and the canonical classes, we can regard a pseudoholomorphic representative of the canonical class K_M of a symplectic 4-manifold (M, ω) as a symplectic version of the canonical divisor. Thus, one would like to achieve a similar classification as in complex surfaces for symplectic 4-manifolds. We consider the geography problem for symplectic 4-manifolds: Which pairs (m, n) of integers are realized as the Chern pairs (c_1^2, c_2) of a symplectic 4-manifold?

We review Chern classes of symplectic 4-manifolds. We notice the following fundamental relations between c_1^2 and c_2 first:

$$\begin{aligned} c_1 &\equiv w_2 \pmod{2}, \\ c_1^2 &= 3\sigma + 2e \quad (\text{Hirzebruch's signature theorem}), \\ c_2 &= e. \end{aligned}$$

Thus, the pair (c_1^2, c_2) is determined uniquely by the pair (σ, e) . Conversely, the pair (σ, e) is determined uniquely by the pair (c_1^2, c_2) . Furthermore, since c_1 is characteristic, we have that $c_1^2 \equiv w_2 \pmod{8}$. Hence, the Noether formula holds also for symplectic 4-manifolds:

$$c_1^2 + c_2 \equiv 0 \pmod{12}.$$

By use of a pseudoholomorphic representative of the canonical class K_M , one can prove part (1) of the following theorem:

Theorem 4-1 [[Taubes 1996](#); [Liu 1996](#)]. (1) *If M is a minimal symplectic 4-manifold with $b_2^+(M) > 1$, then $c_1^2(M) \geq 0$.*

- (2) Let M be a minimal symplectic 4-manifold with $b_2^+(M) = 1$. If $c_1^2(M) < 0$, then M must be an irrational ruled surface.

It follows from an easy calculation in (co)homology that, if M is a symplectic 4-manifold with $b_1(M) \leq 1$, then $c_1^2(M) \leq 5c_2(M)$. As for nonminimal symplectic 4-manifolds, we have

$$c_1^2(M \# \overline{\mathbb{C}P^2}) = c_1^2(M) - 1 \quad \text{and} \quad c_2(M \# \overline{\mathbb{C}P^2}) = c_2(M) + 1.$$

Namely, the blow-up translates lattice points on the (c_2, c_1^2) -plane along the vector $(1, -1)$. Thus, the problem of the maximum for the slope c_1^2/c_2 comes to our mind. The author has no answer for this problem. However, it is expected that, if M is a symplectic 4-manifold which is not an irrational ruled surface, then $c_1^2(M) \leq 3c_2(M)$. See [Stipsicz 2000].

Since surfaces of general type are Kähler, it follows from the geography of complex surfaces that, for most of lattice points (m, n) with $\frac{1}{5}(n - 36) \leq m \leq 3n$ and $m \geq 0$, there are symplectic 4-manifolds with $(c_1^2, c_2) = (m, n)$. On the other hand, Gompf [1995], Stipsicz [1996], and Fintushel and Stern [1998] showed that the existential range of (c_1^2, c_2) of symplectic 4-manifolds is larger than that of complex surfaces.

As for the geography of Lefschetz fibrations, we can regard the geography of 4-manifolds admitting Lefschetz fibrations as one of symplectic 4-manifolds by the works of Donaldson and Gompf. Furthermore, we can also consider the geography of fibration structures of Lefschetz fibrations as follows:

Let $f : X \rightarrow S^2$ be a genus- g Lefschetz fibration with n_0 irreducible singular fibers and s_h reducible singular fibers of type II_h ($1 \leq h \leq [g/2]$). We denote the number of singular fibers of f by $\mu(f) := n_0 + \sum_{h=1}^{[g/2]} s_h$. Then, we have the following:

$$c_1^2(X) = 3\sigma(X) + 2e(X), \quad c_2(X) = e(X) = 4 - 4g + \mu(f).$$

Hence, by Theorem 2-5, we can calculate the Chern pairs $(c_1^2(X), c_2(X))$ from the $([g/2] + 1)$ -tuple $(n_0, s_1, \dots, s_{[g/2]})$ of the numbers of singular fibers for a hyperelliptic genus- g Lefschetz fibration $f : X \rightarrow S^2$. Thus, we regard the geography problem of (possibly nonhyperelliptic) genus- g Lefschetz fibrations as characterizing the $([g/2] + 1)$ -tuple $(n_0, s_1, \dots, s_{[g/2]})$ of the numbers of singular fibers.

We now recall some facts about the number of singular fibers. The following inequalities hold for the number of irreducible singular fibers and the number of reducible singular fibers:

Theorem 4-2 [Stipsicz 1999]. *Let $f : X \rightarrow S^2$ be a nontrivial genus- g Lefschetz fibration with n_0 irreducible singular fibers and s_h reducible singular fibers of type II_h ($1 \leq h \leq [g/2]$).*

$$(1) \quad 5n_0 \geq \sum_{h=1}^{\lfloor g/2 \rfloor} s_h.$$

(2) $n_0 > 0$, that is, there is no Lefschetz fibration with only reducible singular fibers.

Let $N(g)$ denote the minimal number of singular fibers in genus- g Lefschetz fibrations over S^2 , namely,

$$N(g) := \min\{\mu(f) \mid f : X \rightarrow S^2 \text{ is a genus-}g \text{ Lefschetz fibration}\}.$$

Theorem 4-3 [Korkmaz and Ozbagci 2001; Stipsicz 1999]. *We have estimates on $N(g)$ as follows:*

- (1) $N(2) = 7$, or 8.
- (2) $N(g) \geq \frac{1}{5}(4g + 2)$.

Proposition 4-4 [Sato 2010b]. *Let $f : X \rightarrow S^2$ be a genus-2 Lefschetz fibration with $\mu(f)$ singular fibers.*

- (1) *If $\mu(f) = 7$, then X is diffeomorphic to $S^2 \times T^2 \# 3\overline{\mathbb{C}P^2}$.*
- (2) *If $\mu(f) = 8$, then X is diffeomorphic to $S^2 \times T^2 \# 4\overline{\mathbb{C}P^2}$.*

By considering the abelianization of the global monodromy of a Lefschetz fibration, we can obtain the congruence on the number of singular fibers. The following proposition is proved by noting that the abelianization $H_1(\Gamma_2; \mathbb{Z})$ of Γ_2 is isomorphic to the cyclic group of order 10.

Proposition 4-5 [Persson 1992]. *Suppose that a genus-2 Lefschetz fibration over S^2 has n_0 irreducible singular fibers and s reducible singular fibers. Then, we have*

$$n_0 + 2s \equiv 0 \pmod{10}.$$

If $g \geq 3$, then $H_1(\Gamma_g; \mathbb{Z}) = 0$, and so we can get no information on the number of singular fibers. However, if we consider the hyperelliptic case, then we can get information for hyperelliptic Lefschetz fibrations. Since the abelianization $H_1(\Gamma_g^{\text{hyp}}; \mathbb{Z})$ of the hyperelliptic mapping class group Γ_g^{hyp} is isomorphic to $\mathbb{Z}/2(2g + 1)$ if g is even and $\mathbb{Z}/4(2g + 1)$ if g is odd, we obtain the congruence on the number of singular fibers of a hyperelliptic fibration.

Proposition 4-6 [Endo 2000]. *Suppose that $f : X \rightarrow S^2$ is a genus- g hyperelliptic Lefschetz fibration with n_0 irreducible singular fibers and s_h singular fibers of type II_h ($h = 1, 2, \dots, \lfloor g/2 \rfloor$). Then, we have*

$$n_0 + 4 \sum_{h=1}^{\lfloor g/2 \rfloor} h(2h + 1)s_h \equiv 0 \pmod{\left\{ \begin{array}{l} 2(2g + 1) \text{ (if } g \text{ is even)} \\ 4(2g + 1) \text{ (if } g \text{ is odd)} \end{array} \right\}}.$$

Remark 4-7. By taking the fiber sums, we can construct genus- g Lefschetz fibrations with arbitrarily large numbers of singular fibers. For example, for the genus- g hyperelliptic Lefschetz fibration $f : \mathbb{C}P^2 \# (4g + 5)\overline{\mathbb{C}P^2} \rightarrow \mathbb{C}P^1$ corresponding to the positive relation

$$W_1 : (c_1 \cdot c_2 \cdots c_{2g} \cdot c_{2g+1} \cdot c_{2g+1} \cdot c_{2g} \cdots c_2 \cdot c_1)^2 = 1$$

we consider the fiber sum $\#_{mF} f$ of m copies of f . Then, the total space of $\#_{mF} f$ is minimal and $\#_{mF} f$ has $4(2g + 1)m$ irreducible singular fibers. Hence, the set of all $(n_0, s_1, \dots, s_{[g/2]})$ of the numbers of singular fibers of genus- g Lefschetz fibrations over S^2 is not bounded.

We shall consider the geography problem of Lefschetz fibrations in [Section 6](#).

5. The canonical classes of nonminimal Lefschetz fibrations over S^2

2-spheres of square -1 in Lefschetz fibrations. Now we begin with two important theorems on smoothly embedded spheres in a symplectic 4-manifold with self-intersection number -1 .

Theorem 5-1 ((-1) -curve theorem, [\[Li and Liu 1995; Taubes 1996\]](#)). *Let (M, ω) be a closed symplectic 4-manifold. Suppose that M is neither the blow-up of a rational surface nor the blow-up of a ruled surface. Then, any smoothly embedded sphere of square -1 is \mathbb{Z} -homologous to a pseudoholomorphic rational curve of square -1 after the appropriate choice of an orientation of the sphere.*

Taubes showed this theorem for $b_2^+(M) > 1$, and Li and Liu showed this theorem for $b_2^+(M) = 1$.

Theorem 5-2 [\[Ohta and Ono 2005\]](#). *Let (M, ω) be a closed symplectic 4-manifold and F an irreducible pseudoholomorphic curve in M with respect to an ω -compatible almost-complex structure J_0 . Suppose that the genus of F is positive. Then, there exists an almost-complex structure J , which is arbitrarily close to J_0 , such that F and any symplectic sphere of square -1 are represented by J -holomorphic curves simultaneously.*

Next we consider spheres of square -1 in Lefschetz fibrations. Let $f : X \rightarrow S^2$ be a nonminimal genus- g Lefschetz fibration. Namely, we let X admit smoothly embedded spheres of square -1 . Since we suppose that $g \geq 2$, X has a symplectic structure ω with an ω -compatible almost complex structure J for which the fibers are pseudoholomorphic ([Theorem 2-3](#)). From now on, we assume that a Lefschetz fibration $f : X \rightarrow S^2$ is not minimal and X admits such structures ω and J .

Let $E \in H^2(X; \mathbb{Z})$ be the Poincaré dual of the homology class that is represented by a smoothly embedded sphere of square -1 in X . By changing the orientation of this sphere if necessary, we may assume that $E \cdot [\omega] > 0$, because we have

$E \cdot [\omega] \neq 0$ by the (-1) -curve theorem and the fact that $\omega|_\Sigma$ on a closed symplectic submanifold Σ is a volume form of Σ . We denote by \mathcal{E}_X the set of all the Poincaré duals of the homology classes E that can be represented by smoothly embedded spheres of square -1 and satisfy $E \cdot [\omega] > 0$. Moreover, let F denote the Poincaré dual of the homology class represented by a generic fiber. Then, we have the following theorem:

Theorem 5-3 [Sato 2008]. *Suppose that X is neither the blow-up of a rational surface nor the blow-up of a ruled surface. We set $\mathcal{E}_X = \{E_1, E_2, \dots, E_m\}$. Then:*

- (1) $E_i \cdot F \geq 1$ for any $E_i \in \mathcal{E}_X$,
- (2) $m \leq (\sum_{i=1}^m E_i) \cdot F \leq 2g - 2$.

Remark 5-4. Suppose that X is neither the blow-up of a rational surface nor the blow-up of a ruled surface. Then, by the (-1) -curve theorem, $E \in \mathcal{E}_X$ can be represented by an ω -symplectic sphere of square -1 . Hence, it follows from [Theorem 5-2](#) and the positivity of intersections of pseudoholomorphic curves that, if $E \in \mathcal{E}_X$ satisfies $E \cdot F = 1$, then E is represented by a (-1) -section of f .

Thus, by [Theorem 5-3](#), we can classify \mathcal{E}_X into several types. For example, \mathcal{E}_X in the cases of $g = 2$ and $g = 3$ are classified as follows:

The case $g = 2$: We consider a genus-2 Lefschetz fibration $f : X \rightarrow S^2$ with spheres of square -1 . If X is neither rational nor ruled, then [Theorem 5-3](#) states that \mathcal{E}_X is one of the following three:

Type (1, 1): $\mathcal{E}_X = \{E_1, E_2\}$, $E_1 \cdot F = E_2 \cdot F = 1$.

Type (1): $\mathcal{E}_X = \{E\}$, $E \cdot F = 1$.

Type (2): $\mathcal{E}_X = \{E\}$, $E \cdot F = 2$.

In the first and the second cases, spheres representing \mathcal{E}_X are (-1) -sections of $f : X \rightarrow S^2$. Note that $E_1 \cdot E_2 = 0$ for E_1 and E_2 in Type (1, 1), which follows from the proof of [Corollary 3](#) in [Li 1999].

The case $g = 3$: We consider a genus-3 Lefschetz fibration $f : X \rightarrow S^2$ with spheres of square -1 . If X is neither rational nor ruled, [Theorem 5-3](#) states that the set \mathcal{E}_X of spheres of square -1 is one of the following 11 types:

Type (1, 1, 1, 1): $\mathcal{E}_X = \{E_1, E_2, E_3, E_4\}$, $E_1 \cdot F = E_2 \cdot F = E_3 \cdot F = E_4 \cdot F = 1$.

Type (1, 1, 2): $\mathcal{E}_X = \{E_1, E_2, E\}$, $E_1 \cdot F = E_2 \cdot F = 1$, $E \cdot F = 2$.

Type (1, 3): $\mathcal{E}_X = \{E_1, E\}$, $E_1 \cdot F = 1$, $E \cdot F = 3$.

Type (2, 2): $\mathcal{E}_X = \{E_1, E_2\}$, $E_1 \cdot F = E_2 \cdot F = 2$.

Type (4): $\mathcal{E}_X = \{E\}$, $E \cdot F = 4$.

Type (1, 1, 1): $\mathcal{E}_X = \{E_1, E_2, E_3\}$, $E_1 \cdot F = E_2 \cdot F = E_3 \cdot F = 1$.

Type (1, 2): $\mathcal{C}_X = \{E_1, E\}$, $E_1 \cdot F = 1$, $E \cdot F = 2$.

Type (3): $\mathcal{C}_X = \{E\}$, $E \cdot F = 3$.

Type (1, 1): $\mathcal{C}_X = \{E_1, E_2\}$, $E_1 \cdot F = E_2 \cdot F = 1$.

Type (2): $\mathcal{C}_X = \{E\}$, $E \cdot F = 2$.

Type (1): $\mathcal{C}_X = \{E\}$, $E \cdot F = 1$.

Furthermore, if we set $\sum \mathcal{C}_X := \sum_{i=1}^m E_i$ for $\mathcal{C}_X = \{E_1, E_2, \dots, E_m\}$, then types of \mathcal{C}_X are shared as follows:

$(\sum \mathcal{C}_X) \cdot F = 4$: Type (1, 1, 1, 1), Type (1, 1, 2), Type (1, 3), Type (2, 2),
Type (4)

$(\sum \mathcal{C}_X) \cdot F = 3$: Type (1, 1, 1), Type (1, 2), Type (3)

$(\sum \mathcal{C}_X) \cdot F = 2$: Type (1, 1), Type (2)

$(\sum \mathcal{C}_X) \cdot F = 1$: Type (1)

In general, if the set $\mathcal{C}_X = \{E_1, E_2, \dots, E_m\}$ for a nonminimal genus- g Lefschetz fibration $f : X \rightarrow S^2$ satisfies the conditions

$$E_i \cdot F = j_i, \quad j_1 \leq j_2 \leq \dots \leq j_m,$$

then \mathcal{C}_X is said to be of Type (j_1, j_2, \dots, j_m) .

Now we can state the main theorem.

Theorem 5-5. *Let $f : X \rightarrow S^2$ be a nonminimal genus- g Lefschetz fibration. Let K_X be the canonical class of (X, ω) . Suppose that X is neither the blow-up of a rational surface nor the blow-up of a ruled surface. Then, the canonical class K_X can be determined according to the types of \mathcal{C}_X as follows:*

[1] If $g = 2$, we have:

(1) If \mathcal{C}_X is of Type (1, 1), then $K_X = E_1 + E_2$, where $\mathcal{C}_X = \{E_1, E_2\}$.

(2) If \mathcal{C}_X is of Type (2), then $K_X = E$, where $\mathcal{C}_X = \{E\}$.

(3) If \mathcal{C}_X is of Type (1), then $K_X = 2E + R$ or $K_X = 2E + F$. Here, $\mathcal{C}_X = \{E\}$ and R is a genus-1 irreducible component of a reducible singular fiber such that $E \cdot R = 1$. Moreover, in the case of $K_X = 2E + R$, the fibration f has only one reducible singular fiber. In the case of $K_X = 2E + F$, the fibration f has no reducible singular fiber.

[2] If $g \geq 3$, we have:

(1) If $(\sum \mathcal{C}_X) \cdot F = 2g - 2$, then $K_X = \sum \mathcal{C}_X$.

(2) If $(\sum \mathcal{C}_X) \cdot F = 2g - 3$, we have

$$K_X = 2E_1 + \sum_{\substack{E \in \mathcal{C}_X \\ E \neq E_1}} E + R.$$

Here, E_1 is a (-1) -section of f and R is a genus-1 irreducible component of a reducible singular fiber such that $E_1 \cdot R = 1$ and $E \cdot R = 0$ for any $E \in \mathcal{C}_X$ ($E \neq E_1$).

Proof. We can find out the proof in the case of $g = 2$ in the proof of [Theorem 5-1](#) of [\[Sato 2008\]](#). We suppose that the fiber genus of f is greater than two.

Equip X with an almost complex structure J such that fibers of f are J -holomorphic curves. Let $\mathcal{C}_X = \{E_1, E_2, \dots, E_m\}$ be the set of all cohomology classes represented by spheres of square -1 . Set $A = K_X - \sum \mathcal{C}_X = K_X - \sum_{i=1}^m E_i$. By the adjunction formula, we have $K_X \cdot F = 2g - 2$, $K_X \cdot E_i = -1$ for any i and so $A \cdot E_i = K_X \cdot E_i - E_i^2 = 0$ for any i . Furthermore, we have

$$A^2 = A \cdot (K_X - \sum \mathcal{C}_X) = A \cdot K_X - \sum_{i=1}^m A \cdot E_i = A \cdot K_X,$$

$$A \cdot F = K_X \cdot F - (\sum \mathcal{C}_X) \cdot F = (2g - 2) - (\sum \mathcal{C}_X) \cdot F.$$

Hence, if $(\sum \mathcal{C}_X) \cdot F$ is $2g - 2$ or $2g - 3$, then

$$A \cdot F = \begin{cases} 0 & \text{if } (\sum \mathcal{C}_X) \cdot F = 2g - 2, \\ 1 & \text{if } (\sum \mathcal{C}_X) \cdot F = 2g - 3. \end{cases}$$

Since each class E_i of \mathcal{C}_X is represented by a pseudoholomorphic curve and is a basic class of the Gromov–Taubes invariant Gr_T [\[Taubes 1995; 1996\]](#), it follows from the duality formula of the Gromov–Taubes invariant that A is also a basic class, that is, $\text{Gr}_T(A) \neq 0$. Hence, the class A has a J -holomorphic representative $\mathcal{C} = \{(C_j, m_j)\}_{1 \leq j \leq n}$ such that each C_j is a J -holomorphic curve and each m_j (≥ 1) is the multiplicity of C_j . The components C_j of \mathcal{C} are not always nonsingular.

(1) The case of $(\sum \mathcal{C}_X) \cdot F = 2g - 2$: The cohomology class A is represented by $\mathcal{C} = \{(C_j, m_j)\}_{1 \leq j \leq n}$ and we have $A = \sum_{j=1}^n m_j [C_j]$. Since $A \cdot F = 0$, we have $\sum_{j=1}^n m_j [C_j] \cdot F = 0$. Noting that pseudoholomorphic curves have locally positive intersections, this implies that each component C_j of \mathcal{C} is contained in a fiber. Hence, we have $[C_j]^2 = 0$ or -1 . If C_j is a generic fiber or an irreducible singular fiber, then $[C_j]^2 = 0$. If C_j is a component of a reducible singular fiber, then $[C_j]^2 = -1$. However, since $A \cdot E_i = 0$ and $E_i \cdot F \neq 0$ for any i , each component C_j is neither a generic fiber nor an irreducible singular fiber. Furthermore, since f is relatively minimal, every fiber contains no sphere-component. Therefore, C_j is a component of a reducible singular fiber with genus $(C_j) \geq 1$ and $[C_j]^2 = -1$. If distinct components C_j and C_k intersected, they would be components of a

reducible fiber and one of C_j and C_k would meet a section E_i . However, since $A \cdot E_i = 0$, any component C_j does not meet other components C_k .

Now we arrange the indices of components of \mathcal{C} . Let $C_{j,k}$ be a component of a reducible singular fiber such that the genus of $C_{j,k}$ is k . Let $m_{j,k}$ be the multiplicity of $C_{j,k}$. If \mathcal{C} does not contain a component of genus k , then we set up a virtual component $C_{1,k}$ and $m_{1,k} = 0$. Then, we have $A = \sum_{k=1}^{g-1} \sum_{j=1}^{n_j} m_{j,k} [C_{j,k}]$. Noting that $K_X \cdot [C_{j,k}] = 2k - 1$ and $[C_{i,\ell}] \cdot [C_{j,k}] = 0$ ($k \neq \ell$), we calculate A^2 and $A \cdot K_X$. We have

$$\begin{aligned} A^2 &= \sum_{j=1}^{n_1} m_{j,1}^2 [C_{j,1}]^2 + \sum_{j=1}^{n_2} m_{j,2}^2 [C_{j,2}]^2 + \cdots + \sum_{j=1}^{n_{g-1}} m_{j,g-1}^2 [C_{j,g-1}]^2 \\ &= - \left(\sum_{j=1}^{n_1} m_{j,1}^2 + \sum_{j=1}^{n_2} m_{j,2}^2 + \cdots + \sum_{j=1}^{n_{g-1}} m_{j,g-1}^2 \right), \\ A \cdot K_X &= \sum_{j=1}^{n_1} m_{j,1} [C_{j,1}] \cdot K_X + \sum_{j=1}^{n_2} m_{j,2} [C_{j,2}] \cdot K_X + \cdots \\ &\quad + \sum_{j=1}^{n_{g-1}} m_{j,g-1} [C_{j,g-1}] \cdot K_X \\ &= \sum_{j=1}^{n_1} m_{j,1} + 3 \sum_{j=1}^{n_2} m_{j,2} + \cdots + (2g - 3) \sum_{j=1}^{n_{g-1}} m_{j,g-1}. \end{aligned}$$

Hence, we have that $A^2 \leq 0$ and $A \cdot K_X \geq 0$. Since $A^2 = A \cdot K_X$, we have $A^2 = A \cdot K_X = 0$. Therefore, we have $m_{j,k} = 0$ for any j, k , in particular $A = 0$. Hence, $K_X = \sum \mathcal{C}_X$.

(2) The case of $(\sum \mathcal{C}_X) \cdot F = 2g - 3$: Since $A \cdot F = 1$, the pseudoholomorphic representative \mathcal{C} of A contains a section S as a component of \mathcal{C} . Then, we can see that S is smooth and the multiplicity of S is one. Suppose that S is singular and $x \in S$ is a singular point of S . The fiber $F_0 = f^{-1}(f(x))$ intersects S at the singular point x . This fact implies that $[S] \cdot [F_0] \geq 2$, because pseudoholomorphic curves have locally positive intersections. However, this contradicts the fact that $A \cdot [F_0] = A \cdot F = 1$. Hence, S is a smooth section. Moreover, since $A \cdot F = 1$, the multiplicity of S is one.

Let $\{C_j \mid j = 1, 2, \dots, n\}$ be the set of all components of \mathcal{C} except S . Then, we can see that each C_j contains in a fiber of f . Since $A = [S] + \sum_{j=1}^n m_j [C_j]$ and $A \cdot F = 1$, we have that $1 = A \cdot F = [S] \cdot F + \sum_{j=1}^n m_j [C_j] \cdot F = 1 + \sum_{j=1}^n m_j [C_j] \cdot F$ and $\sum_{j=1}^n m_j [C_j] \cdot F = 0$. Hence, we have $[C_j] \cdot F = 0$ for any j , and so each component C_j contains in a fiber.

Now we divide components of \mathcal{C} except S into generic/irreducible fibers and components of reducible fibers. Furthermore, we divide components of reducible fibers in \mathcal{C} according to genera. Namely, the class A is represented by the pseudo-holomorphic curve $\mathcal{C} = \{(S, 1), (F_i, k_i), (C_{j,\ell}, m_{j,\ell})\}$, where each F_i is a generic fiber or an irreducible singular fiber and each $C_{j,\ell}$ is a component of a reducible singular fiber such that the genus of $C_{j,\ell}$ is ℓ . Of course, we have that $[C_{j,\ell}]^2 = -1$, $[C_{i,k}] \cdot [C_{j,\ell}] = 0$ ($(i, k) \neq (j, \ell)$), $F \cdot [C_{j,\ell}] = 0$ and $[C_{j,\ell}] \cdot K_X = 2\ell - 1$. Since an irreducible singular fiber is homologous to the generic fiber F , components of \mathcal{C} which are generic fibers or irreducible singular fibers yield the homology class mF , and so the class A is given by

$$\begin{aligned} A &= [S] + mF + \sum_{j=1}^{n_1} m_{j,1}[C_{j,1}] + \sum_{j=1}^{n_2} m_{j,2}[C_{j,2}] + \cdots + \sum_{j=1}^{n_{g-1}} m_{j,g-1}[C_{j,g-1}] \\ &= [S] + mF + \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell}[C_{j,\ell}]. \end{aligned}$$

We compare A^2 with $A \cdot K_X$ in the same way as the case (1). We have

$$\begin{aligned} A^2 &= [S]^2 + m^2 F^2 + \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell}^2 [C_{j,\ell}]^2 + 2m[S] \cdot F \\ &\quad + 2 \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell} [S] \cdot [C_{j,\ell}] + 2 \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell} F \cdot [C_{j,\ell}] \\ &= [S]^2 - \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell}^2 + 2m + 2 \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell} [S] \cdot [C_{j,\ell}], \\ A \cdot K_X &= [S] \cdot K_X + mF \cdot K_X + \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell} [C_{j,\ell}] \cdot K_X \\ &= -2 - [S]^2 + 2m(g-1) + \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} (2\ell - 1)m_{j,\ell}. \end{aligned}$$

Since $A^2 - A \cdot K_X = 0$, we have

$$(5-6) \quad -2(1 + [S]^2) + 2m(g-1) + \sum_{\ell=1}^{g-1} \sum_{j=1}^{n_\ell} m_{j,\ell}(2\ell - 1 + m_{j,\ell} - 2[S] \cdot [C_{j,\ell}]) = 0.$$

By Lemma 2.1 of [Stipsicz 2001], the self-intersection number of S is negative, and so we have $-2(1 + [S]^2) \geq 0$. Furthermore, since S is a section, the intersection

number $[S] \cdot [C_{j,\ell}]$ is 0 or 1. Hence, we have that $m_{j,\ell}(2\ell - 1 + m_{j,\ell} - 2[S] \cdot [C_{j,\ell}]) \geq 0$. Hence, each term in the left side of (5-6) is nonnegative. Therefore, we obtain

$$\begin{cases} -2(1 + [S]^2) = 0, \\ 2m(g - 1) = 0, \\ m_{j,\ell}(2\ell - 1 + m_{j,\ell} - 2[S] \cdot [C_{j,\ell}]) = 0 \text{ for any } j, \ell. \end{cases}$$

If $\ell \geq 2$, then we have $2\ell - 1 + m_{j,\ell} - 2[S] \cdot [C_{j,\ell}] \geq 1$ since $0 \leq [S] \cdot [C_{j,\ell}] \leq 1$. Hence it follows from the third equation above that $m_{j,\ell} = 0$ for $\ell \geq 2$. Thus the pseudoholomorphic representative \mathcal{C} consists only of the section S and torus components of reducible singular fibers. Noting that $[S] \cdot [C_{j,1}]$ is 0 or 1, it follows that if $m_{j,1}(1 + m_{j,1} - 2[S] \cdot [C_{j,1}]) = 0$ and $m_{j,1} \neq 0$, then $m_{j,1} = 1$ and $[S] \cdot [C_{j,1}] = 1$. On the other hand, since $[S]^2 = -1$, the smooth section S is a sphere of square -1 and the class $[S]$ is a member of \mathcal{C}_X . Set $E_1 = [S]$.

Now we consider the case where there is a torus component in \mathcal{C} . Then, such a torus component meets the section S and its multiplicity is one. Thus, we can write A as $A = E_1 + \sum_{j=1}^{n_1} [C_{j,1}]$. Suppose that $n_1 \geq 2$. Since $A \cdot E_1 = 0$ and $[S] \cdot [C_{j,1}] = 1$, we have that $n_1 - 1 = \sum_{j \neq 1} [C_{j,1}] \cdot E_1 = A \cdot E_1 - (E_1^2 + [C_{1,1}] \cdot E_1) = 0$. This is a contradiction. Hence, the class A is written as $A = E_1 + [C_{1,1}]$. Therefore, we obtain

$$K_X = 2E_1 + \sum_{\substack{E \in \mathcal{C}_X \\ E \neq E_1}} E + R,$$

where $R = [C_{1,1}]$ is the class represented by a torus component of a reducible singular fiber and $E_1 \cdot R = 1$. Since $C_{1,1}$ is a torus, we have $K_X \cdot R = 1$ and so $\sum_{E \in \mathcal{C}_X, E \neq E_1} E \cdot R = K_X \cdot R - (2E_1 \cdot R + R^2) = 0$. Hence, it follows from the positiveness of local intersections of pseudoholomorphic curves that $E \cdot R = 0$ for any $E \in \mathcal{C}_X$ except E_1 .

Next we consider the case where there is no torus component in \mathcal{C} . Then, we can write A as $A = E_1$. Hence, we obtain

$$K_X = 2E_1 + \sum_{\substack{E \in \mathcal{C}_X \\ E \neq E_1}} E.$$

Then, the minimal model X_{\min} of X must satisfy that $K_{\min}^2 < 0$. Here, K_{\min} denotes the canonical class of X_{\min} . However, since X is not the blow-up of a ruled surface, it follows from Theorem 4-1 that K_{\min}^2 must be nonnegative. This is a contradiction. Hence, the case of $K_X = 2E_1 + \sum_{E \in \mathcal{C}_X, E \neq E_1} E$ cannot occur. This completes the proof. \square

Remark 5-7. (1) If a nonminimal genus- g Lefschetz fibration $f : X \rightarrow S^2$ satisfies $g \geq 3$ and \mathcal{C}_X is of class $(\sum \mathcal{C}_X) \cdot F = 2g - 3$, then f must have some reducible singular fibers.

(2) \mathcal{E}_X of class $(\sum \mathcal{E}_X) \cdot F = 2g - 3$ must contain some sections of square -1 . For example, the set \mathcal{E}_X of type (3) does not appear for nonminimal genus-3 Lefschetz fibrations.

(3) The square of the canonical class K_X for X in [Theorem 5-5](#) is bounded. In fact, when $(\sum \mathcal{E}_X) \cdot F = 2g - 2$, we have that $2 - 2g \leq K_X^2 \leq -1$. When $(\sum \mathcal{E}_X) \cdot F = 2g - 3$, we have that $3 - 2g \leq K_X^2 \leq -2$.

The symplectic Kodaira dimension of nonminimal Lefschetz fibrations. Given any smooth complex surface X , there are four possibilities of the behavior of the plurigenera $P_n(X)$. The Kodaira dimension $\kappa(X)$ of X is defined according to four possibilities: $\kappa(X) = -\infty, 0, 1$, or 2 . It is well-known that smooth compact complex surfaces are classified in terms of the Kodaira dimension.

The first notion of the symplectic version of the Kodaira dimension appeared in [\[McDuff and Salamon 1996\]](#) and the symplectic Kodaira dimension is defined and discussed in detail in [\[Li 2006\]](#).

Definition 5-8. Let (M, ω) be a minimal symplectic 4-manifold with symplectic canonical class K_ω . Then, the symplectic Kodaira dimension $\kappa^s(M, \omega)$ is defined as follows:

$$\kappa^s(M, \omega) = \begin{cases} -\infty & \text{if } K_\omega \cdot [\omega] < 0 \text{ or } K_\omega^2 < 0, \\ 0 & \text{if } K_\omega \cdot [\omega] = 0 \text{ and } K_\omega^2 = 0, \\ 1 & \text{if } K_\omega \cdot [\omega] > 0 \text{ and } K_\omega^2 = 0, \\ 2 & \text{if } K_\omega \cdot [\omega] > 0 \text{ and } K_\omega^2 > 0. \end{cases}$$

The symplectic Kodaira dimension of a nonminimal 4-manifold (M, ω) is defined to be that of any minimal model of (M, ω) .

Theorem 5-9 ([\[Li 2006\]](#); see also [\[Dorfmeister and Zhang 2009\]](#)). *Let (M, ω) be a closed symplectic 4-manifold with symplectic canonical class K_ω . If (M, ω) is minimal, then:*

(1) *With M given the orientation compatible with ω , the symplectic Kodaira dimension of (M, ω) only depends on the oriented diffeomorphism type of M , that is, if ω' is another symplectic form on M compatible with the orientation of M , then $\kappa^s(M, \omega) = \kappa^s(M, \omega')$.*

$$(2) \quad \kappa^s(M, \omega) = \begin{cases} -\infty & \text{if } M \text{ is rational or ruled,} \\ 0 & \text{if } K_\omega \text{ is torsion,} \\ 1 & \text{if } K_\omega \text{ is nontorsion but } K_\omega^2 = 0, \\ 2 & \text{if } K_\omega^2 > 0. \end{cases}$$

Hence, by [Theorem 5-9\(2\)](#), we can calculate the symplectic Kodaira dimension of (M, ω) in terms of the canonical class K_ω .

Now, we have two kinds of Kodaira dimension. In the case where (M, ω) admits a complex structure J , the equivalence of these Kodaira dimensions $\kappa(M, J)$ and $\kappa^s(M, \omega)$ was proved by Li:

Theorem 5-10 (Li; see [Dorfmeister and Zhang 2009, Theorem 3.1]). *Let M be a smooth 4-manifold that admits a symplectic structure ω as well as a complex structure J . Then, we have $\kappa^s(M, \omega) = \kappa(M, J)$.*

Remark 5-11. There are smooth 4-manifolds M that admits a symplectic structure ω and a complex structure J but whose (M, ω, J) is not Kähler. For example, Kodaira–Thurston manifolds are such manifolds. **Theorem 5-10** states that the equivalence of Kodaira dimensions $\kappa(M, J)$ and $\kappa^s(M, \omega)$ holds for not only the Kähler case but also the non-Kähler case.

For a nonminimal genus- g Lefschetz fibration $f : X \rightarrow S^2$ with

$$\left(\sum \mathcal{E}_X\right) \cdot F = 2g - 2 \quad \text{or} \quad \left(\sum \mathcal{E}_X\right) \cdot F = 2g - 3,$$

we calculate the square K_{\min}^2 of the canonical class K_{\min} of a minimal model of X by **Theorem 5-5**. Furthermore, we can calculate the symplectic Kodaira dimension from K_{\min}^2 .

Theorem 5-12. *Let $f : X \rightarrow S^2$ be a nonminimal genus- g Lefschetz fibration. Equip X with the natural symplectic structure ω given by **Theorem 2-3**. Let K_X be the canonical class of (X, ω) and K_{\min} the canonical class of a minimal model of X . Suppose that X is neither the blow-up of a rational surface nor the blow-up of a ruled surface.*

[1] *If $g = 2$, we have:*

- (1) *If \mathcal{E}_X is of Type (1, 1), then $K_{\min}^2 = 0$ and $\kappa^s(X, \omega) = 0$.*
- (2) *If \mathcal{E}_X is of Type (2), then $K_{\min}^2 = 0$ and $\kappa^s(X, \omega) = 0$.*
- (3) *If \mathcal{E}_X is of Type (1), then $K_{\min}^2 = 0$ and $\kappa^s(X, \omega) = 1$ when $K_X = 2E + R$. We have $K_{\min}^2 = 1$ and $\kappa^s(X, \omega) = 2$ when $K_X = 2E + F$.*

[2] *If $g \geq 3$, we have:*

- (1) *If $\left(\sum \mathcal{E}_X\right) \cdot F = 2g - 2$, then $K_{\min}^2 = 0$ and $\kappa^s(X, \omega) = 0$.*
- (2) *If $\left(\sum \mathcal{E}_X\right) \cdot F = 2g - 3$, then $K_{\min}^2 = 0$ and $\kappa^s(X, \omega) = 1$.*

For example, Tables 1 and 2 summarize the canonical class K_X and the symplectic Kodaira dimension $\kappa^s(X)$ for nonminimal Lefschetz fibrations of fiber genus 2 or 3.

The author has also investigated the Iitaka D -dimension of the adjoint divisor $K_X + F$ of F for holomorphic Lefschetz fibrations. See [Sato 2010a].

Now we can state about the relationship between the Kodaira dimension and the base loci of Lefschetz pencils.

$(\sum \mathcal{C}_X) \cdot F$	\mathcal{C}_X	K_X	κ^S
$(\sum \mathcal{C}_X) \cdot F = 2$	Type (1, 1)	$K_X = E_1 + E_2$	$\kappa^S = 0$
	Type (2)	$K_X = E$	
$(\sum \mathcal{C}_X) \cdot F = 1$	Type (1)	$K_X = 2E + R$	$\kappa^S = 1$
		$K_X = 2E + F$	$\kappa^S = 2$

Table 1. The canonical class and the Kodaira dimension of non-minimal genus-2 Lefschetz fibrations.

$(\sum \mathcal{C}_X) \cdot F$	\mathcal{C}_X	K_X	κ^S
$(\sum \mathcal{C}_X) \cdot F = 4$	Type (1,1,1,1)	$K_X = E_1 + E_2 + E_3 + E_4$	$\kappa^S = 0$
	Type (1,1,2)	$K_X = E_1 + E_2 + E$	$\kappa^S = 0$
	Type (1,3)	$K_X = E_1 + E$	$\kappa^S = 0$
	Type (2,2)	$K_X = E_1 + E_2$	$\kappa^S = 0$
	Type (4)	$K_X = E$	$\kappa^S = 0$
$(\sum \mathcal{C}_X) \cdot F = 3$	Type (1,1,1)	$K_X = 2E_1 + E_2 + E_3 + R$	$\kappa^S = 1$
	Type (1,2)	$K_X = 2E_1 + E + R$	$\kappa^S = 1$
	Type (3)	no existence	no existence

Table 2. The canonical class and the Kodaira dimension of non-minimal genus-3 Lefschetz fibrations.

Corollary 5-13. *Let (X, ω) be a minimal symplectic 4-manifold that is neither rational nor ruled. Suppose that X admits a Lefschetz pencil whose fiber genus g is more than 2.*

- (1) *If the base locus consists of $2g - 2$ base points, then the symplectic Kodaira dimension of X is 0.*
- (2) *If the base locus consists of $2g - 3$ base points, then the symplectic Kodaira dimension of X is 1.*
- (3) *If X is a minimal symplectic 4-manifold with $\kappa^S(X) = 2$ (for example, a minimal complex surface of general type), then the base locus consists of at most $2g - 4$ base points.*

Proof. Suppose that X admits a Lefschetz pencil whose base locus consists of k base points. By blowing up at these k base points, we obtain a nonminimal genus- g Lefschetz fibration $Y = X \# k\mathbb{C}P^2 \rightarrow S^2$ with k sections of square -1 . The set \mathcal{C}_Y is of Type $(1, 1, \dots, 1)$ with $(\sum \mathcal{C}_Y) \cdot F = k$. Therefore, the claim is proved by [Theorem 5-12](#). \square

Remark 5-14. (1) For any $g (\geq 1)$, a $K3$ surface admits a Lefschetz pencil whose

fiber genus is g and whose base locus consists of $2g - 2$ base points. See [Smith 2001b].

(2) Auroux [2006a] calculated the monodromies of the canonical Lefschetz pencils on a pair of homeomorphic Horikawa surfaces X_1 and X_2 . The Horikawa surface X_1 is a double cover of $\mathbb{C}P^1 \times \mathbb{C}P^1$ branched along a smooth algebraic curve of bidegree $(6, 12)$. The Horikawa surface X_2 is a double cover of the Hirzebruch surface \mathbb{F}_6 branched along $\Delta_{-6} \cup C$, where Δ_{-6} is the exceptional section of \mathbb{F}_6 and C is a smooth algebraic curve in the linear system $|5\Delta_6|$. By the construction of X_1 and X_2 , these Horikawa surfaces admit genus-2 Lefschetz fibrations with 120 singular fibers; the corresponding positive relations are

$$(c_1 \cdot c_2 \cdot c_3 \cdot c_4 \cdot c_5^2 \cdot c_4 \cdot c_3 \cdot c_2 \cdot c_1)^{12} = 1 \quad \text{and} \quad (c_1 \cdot c_2 \cdot c_3 \cdot c_4)^{30} = 1,$$

respectively. See also [Fuller 1998]. Since these have fiber-sum decompositions, X_1 and X_2 are minimal by Stipsicz’s conjecture. Furthermore, the Kodaira dimension of X_1 and X_2 is 2. Auroux showed that X_1 and X_2 admit genus-17 Lefschetz pencils with 16 base points.

(3) A surface S of general type embedded in a higher dimensional projective space $\mathbb{C}P^N$ has a Lefschetz pencil. The intersections of S with hyperplane sections $\{H_t \mid t \in \mathbb{C}P^1\}$ containing a generic linear subspace A of complex codimension 2 make the family of curves, which give S a Lefschetz pencil. If $[S] = r[\mathbb{C}P^2] \in H_4(\mathbb{C}P^N; \mathbb{Z})$, then the base locus $B = S \cap A$ of the pencil consists on r points. By noting the Enriques–Kodaira classification of complex surfaces, the genus g of the generic fiber must be at least 2. On the other hand, we have that $g \geq (r + 4)/2$ by Corollary 5-13.

Endo’s question. By generalizing Chakiris’ way to construct holomorphic genus-2 Lefschetz fibrations, Endo [2008] gave examples of hyperelliptic Lefschetz fibrations of arbitrary genus. We can find many examples of nonminimal hyperelliptic Lefschetz fibrations [ibid.].

For a positive relator W , we denote the Lefschetz fibration associated to the positive relation $W = 1$ by $M_W \rightarrow S^2$. If g is even, then each of M_{P^2} , M_Q , M_{PJ} and M_{PJ} in Section 3 is nonminimal. In the case of $g = 2$, the 4-manifolds M_{PJ} , M_{RI} and $K3 \# \mathbb{C}P^2$ are homeomorphic to $3\mathbb{C}P^2 \# 20\overline{\mathbb{C}P^2}$ by Freedman’s classification theorem. Furthermore, M_{PJ} for $g = 2$, $K3 \# \overline{\mathbb{C}P^2}$ and $3\mathbb{C}P^2 \# 20\overline{\mathbb{C}P^2}$ are mutually nondiffeomorphic. On the other hand, the manifold M_{RI} for $g = 2$ is not diffeomorphic to $3\mathbb{C}P^2 \# 20\overline{\mathbb{C}P^2}$. Hence, the following questions are natural.

Question 5-15 [Endo 2008, Problem 4.13]. Let PJ and RI be the positive relators introduced on page 204.

- (1) Determine whether M_{PJ} and M_{RI} are diffeomorphic or not when $g = 2$.
- (2) Is M_{RI} diffeomorphic to $K3 \# \overline{\mathbb{C}P^2}$?

Remark 5-16. The Lefschetz fibration $M_{RI} \rightarrow S^2$ for $g = 2$ is isomorphic to the genus-2 Lefschetz fibration constructed by Auroux [2003], whose set of spheres of square -1 is of Type (2). See also [Sato 2008]. Hence, $M_{RI} \rightarrow S^2$ for $g = 2$ admits no section of square -1 . Since $M_{PJ} \rightarrow S^2$ for $g = 2$ admits a section of square -1 , two genus-2 Lefschetz fibrations $M_{PJ} \rightarrow S^2$ and $M_{RI} \rightarrow S^2$ are not isomorphic. Question 5-15(1) is whether M_{PJ} and M_{RI} are diffeomorphic as manifolds.

In order to answer this question, we note the following proposition:

Proposition 5-17. *Let (M, ω) and (M', ω') be minimal symplectic 4-manifolds. Suppose that the Kodaira dimension $\kappa^s(M, \omega)$ of (M, ω) is equal to 0. If*

$$\kappa^s(M', \omega') \neq 0,$$

then M' is not orientation-preserving diffeomorphic to M .

Proof. Suppose there exists an orientation-preserving diffeomorphism $f : M \rightarrow M'$. Since the isomorphism $f^* : H^2(M'; \mathbb{Z}) \rightarrow H^2(M; \mathbb{Z})$ gives a one-to-one correspondence

$$\mathcal{B}as(M') \rightarrow \mathcal{B}as(M),$$

if $\mathcal{B}as(M')$ has a nontorsion class, then $\mathcal{B}as(M)$ has also a nontorsion class. Here, $\mathcal{B}as(M)$ and $\mathcal{B}as(M')$ denote the set of Seiberg–Witten basic classes of M and M' , respectively. By Theorem 5-9, the canonical class $K_\omega \in \mathcal{B}as(M)$ is a torsion class. Furthermore, it follows from a theorem of Taubes [1994] (Theorem 10.1.11 of [Gompf and Stipsicz 1999]) that

$$|K \cdot [\omega]| \leq |K_\omega \cdot [\omega]| = 0$$

for any $K \in \mathcal{B}as(M)$. Hence, $|K \cdot [\omega]| = |K_\omega \cdot [\omega]| = 0$ and $\mathcal{B}as(M) = \{\pm K_\omega\}$. In particular, $\mathcal{B}as(M)$ contains only torsion classes. However, since $\kappa^s(M', \omega') \neq 0$, the canonical class $K_{\omega'}$ is nontorsion, and so the set $\mathcal{B}as(M')$ contains a nontorsion class. This is a contradiction. \square

Answer to Question 5.1. (1) *The manifold M_{PJ} is not orientation-preserving diffeomorphic to the manifold M_{RI} .*

(2) *The manifold M_{RI} is diffeomorphic to $K3 \# \overline{\mathbb{C}P^2}$.*

Proof. (1) The genus-2 Lefschetz fibration $M_{PJ} \rightarrow S^2$ is of Type (1) and the genus-2 Lefschetz fibration $M_{RI} \rightarrow S^2$ is of Type (2). Let $M_{PJ\min}$ and $M_{RI\min}$ be minimal models of M_{PJ} and M_{RI} , respectively. By Theorem 5-12, we have that $\kappa^s(M_{PJ\min}) = 1$ and $\kappa^s(M_{RI\min}) = 0$. Therefore, it follows from Proposition 5-17 that $M_{PJ\min}$ and $M_{RI\min}$ are not mutually orientation-preserving diffeomorphic. Since M_{PJ} and M_{RI} are not the blow-ups of a rational surface nor the blow-ups of a ruled surface, two manifolds M_{PJ} and M_{RI} are not orientation-preserving diffeomorphic by the uniqueness of minimal models of symplectic 4-manifolds.

(2) Since the Lefschetz fibration $M_{RI} \rightarrow S^2$ for $g = 2$ is isomorphic to the genus-2 Lefschetz fibration constructed by Auroux, M_{RI} is a simply connected Kähler 4-manifold. Furthermore, we have that $\kappa(M_{RI\min}) = \kappa^s(M_{RI\min}) = 0$. Hence, $M_{RI\min}$ is a K3 surface. Noting the uniqueness of minimal models of symplectic 4-manifolds, M_{RI} is diffeomorphic to $K3 \# \overline{\mathbb{C}P^2}$. \square

6. The geography of nonminimal Lefschetz fibrations over S^2

The geography of nonminimal hyperelliptic Lefschetz fibrations. In this section, we consider the geography problem of genus- g Lefschetz fibrations as characterizing the $([g/2] + 1)$ -tuple $(n_0, s_1, \dots, s_{[g/2]})$ of the numbers of singular fibers.

Then, we have the finiteness theorem of the geography of nonminimal hyperelliptic Lefschetz fibrations:

Theorem 6-1. *Suppose that X is neither the blow-up of a rational surface nor the blow-up of a ruled surface. Then, only finitely many $(n_0, s_1, \dots, s_{[g/2]})$ can be realized as the $([g/2] + 1)$ -tuple of the numbers of singular fibers for nonminimal hyperelliptic genus- g (≥ 2) Lefschetz fibrations with $(\sum \mathbb{C}_X) \cdot F = 2g - 2, 2g - 3$.*

Proof. For a hyperelliptic genus- g Lefschetz fibration $f : X \rightarrow S^2$ with the $([g/2] + 1)$ -tuple $(n_0, s_1, \dots, s_{[g/2]})$ of the numbers of singular fibers, we can calculate K_X^2 from the number of singular fibers as follows:

$$\begin{aligned} K_X^2 &= 3\sigma(X) + 2e(X) \\ &= \frac{g-1}{2g+1}n_0 + \sum_{h=1}^{[g/2]} \left(\frac{12h(g-h) - (2g+1)}{2g+1} \right) s_h + 8(1-g). \end{aligned}$$

Since $12h(g-h) - (2g+1) = -12(h-g/2)^2 + (3g+1)(g-1) > 0$ for $1 \leq h \leq [g/2]$, we have

$$\frac{g-1}{2g+1} > 0 \quad \text{and} \quad \frac{12h(g-h) - (2g+1)}{2g+1} > 0.$$

Hence, every coefficient of $K_X^2 = K_X^2(n_0, s_1, \dots, s_{[g/2]})$ is positive. Therefore, since K_X^2 is bounded by Remark 5-7(3), the number of the $([g/2] + 1)$ -tuple $(n_0, s_1, \dots, s_{[g/2]})$ satisfying K_X^2 and the estimation in Section 4 is finite. In fact, since $K_X^2 \leq -1$, we have

$$\frac{g-1}{2g+1}n_0 + \sum_{h=1}^{[g/2]} \left(\frac{12h(g-h) - (2g+1)}{2g+1} \right) s_h + 8(1-g) \leq -1.$$

From the above inequality, we obtain

$$n_0 \leq \frac{(2g+1)(8g-9)}{g-1}.$$

Moreover, noting that the number n_0 of irreducible singular fibers is positive [Stipsicz 1999], we have

$$s_h \leq \frac{16g^2 - 11g - 8}{12h(g - h) - (2g + 1)}$$

for any h ($1 \leq h \leq [g/2]$). □

From Theorem 6-4, we can obtain an estimate for $\mu(f) = n_0 + \sum_{h=1}^{[g/2]} s_h$ and a partial answer for Smith’s question (Question 1-1).

Corollary 6-2. *There is an upper bound on the number of singular fibers of non-minimal hyperelliptic genus- g (≥ 3) Lefschetz fibrations with $(\sum \mathcal{C}_X) \cdot F = 2g - 2, 2g - 3$ whose total spaces are neither the blow-up of a rational surface nor the blow-up of a ruled surface. In fact, for such hyperelliptic genus- g Lefschetz fibration $f : X \rightarrow S^2$, the number $\mu(f)$ of singular fibers of f satisfies*

$$\mu(f) \leq \frac{(8g - 9)(2g + 1)}{g - 1} + \sum_{h=1}^{[g/2]} \frac{16g^2 - 11g - 8}{12h(g - h) - (2g + 1)}.$$

Remark 6-3. (1) The estimation of $\mu(f)$ given in Corollary 6-2 is rough. By using linear programming, one can obtain a strict estimation of $\mu(f)$.

(2) There is no upper bound on the number of singular fibers of minimal Lefschetz fibrations. In fact, if a Lefschetz fibration $f : X \rightarrow S^2$ has μ singular fibers, then the fiber sum $m \#_F f$ of m copies of f is a minimal Lefschetz fibration with $m\mu$ singular fibers. Hence, there are minimal Lefschetz fibrations with arbitrarily many singular fibers.

Given g (≥ 2), we can present the list of possible $([g/2] + 1)$ -tuples

$$(n_0, s_1, \dots, s_{[g/2]})$$

for nonminimal hyperelliptic genus- g Lefschetz fibrations with $(\sum \mathcal{C}_X) \cdot F = 2g - 2, 2g - 3$. The lists in the cases of $g = 2$ and $g = 3$ are given in Tables 3 and 4.

$(\sum \mathcal{C}_X) \cdot F$	\mathcal{C}_X	K_X	(n_0, s)	κ^s
$(\sum \mathcal{C}_X) \cdot F = 2$	Type (1, 1)	$K_X = E_1 + E_2$	(16, 2), (30, 0)	$\kappa^s = 0$
	Type (2)	$K_X = E$	(14, 3), (28, 1)	
$(\sum \mathcal{C}_X) \cdot F = 1$	Type (1)	$K_X = 2E + R$	(28, 1)	$\kappa^s = 1$
		$K_X = 2E + F$	(40, 0)	$\kappa^s = 2$

Table 3. Possible pairs (n_0, s) as geography in the case of $g = 2$.

$(\sum \mathcal{E}_X) \cdot F$	\mathcal{E}_X	K_X	(n_0, s)	κ^s
$(\sum \mathcal{E}_X) \cdot F = 4$	Type (1,1,1,1)	$K_X = E_1 + E_2 + E_3 + E_4$	(8, 4)	$\kappa^s = 0$
	Type (1,1,2)	$K_X = E_1 + E_2 + E$	(20, 3)	
	Type (1,3)	$K_X = E_1 + E$	(32, 2)	
	Type (2,2)	$K_X = E_1 + E_2$	(32, 2)	
	Type (4)	$K_X = E$	(44, 1)	
$(\sum \mathcal{E}_X) \cdot F = 3$	Type (1,1,1)	$K_X = 2E_1 + E_2 + E_3 + R$	(20, 3)	$\kappa^s = 1$
	Type (1,2)	$K_X = 2E_1 + E_2 + R$	(32, 2)	$\kappa^s = 1$
	Type (3)	none	none	none

Table 4. Possible pairs (n_0, s) as geography in the hyperelliptic case of $g = 3$.

The geography of nonminimal, nonhyperelliptic genus-3 Lefschetz fibrations.

At present, we do not know whether the signature $\sigma(X)$ of X can be calculated from the number of singular fibers for a nonhyperelliptic Lefschetz fibration $f : X \rightarrow S^2$. Hence, we do not know whether the finiteness theorem of the geography holds for nonhyperelliptic case. However, in the case of nonhyperelliptic genus-3 Lefschetz fibrations, we can show the finiteness of the geography for nonminimal holomorphic Lefschetz fibrations by using Smith’s signature formula and the Deligne–Mumford compactified moduli space $\overline{\mathcal{M}}_3$ of stable curves of genus 3.

Let Δ_0 and Δ_1 be the divisors of irreducible and reducible nodal curves, respectively. Then, the Deligne–Mumford compactified moduli space $\overline{\mathcal{M}}_3$ is given by adjoining Δ_0 and Δ_1 to the moduli space \mathcal{M}_3 of stable curves of genus 3. Let $\overline{\mathcal{H}}_3$ denote the divisor of hyperelliptic curves of genus 3 in $\overline{\mathcal{M}}_3$. Then, a theorem of Harer [1983] states that the Hodge class $c_1(\lambda)$, $[\Delta_0]$ and $[\Delta_1]$ generate $H^2(\overline{\mathcal{M}}_3; \mathbb{Z})$ and the cohomology class $[\overline{\mathcal{H}}_3]$ is given, up to a positive rational multiple, by

$$[\overline{\mathcal{H}}_3] = 9c_1(\lambda) - [\Delta_0] - 3[\Delta_1].$$

See [Harris and Morrison 1998].

Theorem 6-4. *Suppose that X is neither the blow-up of a rational surface nor the blow-up of a ruled surface. Then, only finitely many (n_0, s) can be realized as pairs of the numbers of singular fibers for nonminimal, nonhyperelliptic and holomorphic genus-3 Lefschetz fibrations with $(\sum \mathcal{E}_X) \cdot F = 3, 4$.*

Proof. Suppose that $f : X \rightarrow \mathbb{C}P^1$ is nonhyperelliptic and holomorphic. A holomorphic fibration f gives rise to a rational curve $\phi_f(\mathbb{C}P^1)$ in $\overline{\mathcal{M}}_3$. The rational curve $\phi_f(\mathbb{C}P^1)$ has positive intersection with any effective divisors that are not contained

in $\phi_f(\mathbb{C}P^1)$. In particular, we have

$$\langle [\overline{\mathcal{H}}_3], [\phi_f(\mathbb{C}P^1)] \rangle \geq 0.$$

Since $\langle [\overline{\mathcal{H}}_3], [\phi_f(\mathbb{C}P^1)] \rangle$ is given, up to a positive multiple, by

$$\begin{aligned} \langle [\overline{\mathcal{H}}_3], [\phi_f(\mathbb{C}P^1)] \rangle &= \langle 9c_1(\lambda) - [\Delta_0] - 3[\Delta_1], [\phi_f(\mathbb{C}P^1)] \rangle \\ &= \frac{9}{4}(\sigma(X) + n_0 + s) - n_0 - 3s \\ &= \frac{9}{4}\sigma(X) + \frac{5}{4}n_0 - \frac{3}{4}s, \end{aligned}$$

we can obtain the following inequality:

$$\sigma(X) \geq -\frac{5}{9}n_0 + \frac{1}{3}s.$$

Thus, we get the relations

$$\begin{cases} K_X^2 = 3\sigma(X) + 2e(X), \\ -4 \leq K_X^2 \leq -1, \\ \sigma(X) \geq -\frac{5}{9}n_0 + \frac{1}{3}s, \\ e(X) = n_0 + s - 8, \\ 5n_0 \geq s, \end{cases}$$

hence

$$\begin{cases} n_0 + 9s - 45 \leq 0, \\ 5n_0 - s \geq 0, \\ n_0 > 0, s \geq 0. \end{cases}$$

The number of pairs (n_0, s) satisfying these inequalities is finite. □

Acknowledgments

The author would like to thank Professors Tadashi Ashikaga, Hisaaki Endo and Susumu Hirose for helpful discussions. The author is also very grateful to the referee for his comments and suggestions.

References

[Auroux 2003] D. Auroux, “Fiber sums of genus 2 Lefschetz fibrations”, *Turkish J. Math.* **27**:1 (2003), 1–10. [MR 2004b:57033](#) [Zbl 1075.53087](#)

[Auroux 2006a] D. Auroux, “The canonical pencils on Horikawa surfaces”, *Geom. Topol.* **10** (2006), 2173–2217. [MR 2007m:14065](#) [Zbl 1129.57030](#)

[Auroux 2006b] D. Auroux, “Mapping class group factorizations and symplectic 4-manifolds: some open problems”, pp. 123–132 in *Problems on mapping class groups and related topics*, edited by B. Farb, Proc. Sympos. Pure Math. **74**, Amer. Math. Soc., Providence, RI, 2006. [MR 2007h:53134](#) [Zbl 05124680](#)

- [Donaldson 1987] S. K. Donaldson, “Irrationality and the h -cobordism conjecture”, *J. Differential Geom.* **26**:1 (1987), 141–168. [MR 88j:57035](#) [Zbl 0631.57010](#)
- [Donaldson 1998] S. K. Donaldson, “Lefschetz fibrations in symplectic geometry”, *Doc. Math. Extra Vol. 2* (1998), 309–314. [MR 99i:57044](#) [Zbl 0909.53018](#)
- [Dorfmeister and Zhang 2009] J. Dorfmeister and W. Zhang, “The Kodaira dimension of Lefschetz fibrations”, *Asian J. Math.* **13**:3 (2009), 341–357. [MR 2010k:53147](#) [Zbl 1207.53086](#)
- [Endo 2000] H. Endo, “Meyer’s signature cocycle and hyperelliptic fibrations”, *Math. Ann.* **316**:2 (2000), 237–257. [MR 2001b:57047](#) [Zbl 0948.57013](#)
- [Endo 2008] H. Endo, “A generalization of Chakiris’ fibrations”, pp. 251–282 in *Groups of diffeomorphisms*, edited by R. Penner et al., Adv. Stud. Pure Math. **52**, Math. Soc. Japan, Tokyo, 2008. [MR 2011e:57044](#) [Zbl 1180.57028](#)
- [Fintushel and Stern 1998] R. Fintushel and R. J. Stern, “Constructions of smooth 4-manifolds”, *Doc. Math. Extra Vol. 2* (1998), 443–452. [MR 99g:57033](#) [Zbl 0899.57012](#)
- [Fuller 1998] T. Fuller, “Diffeomorphism types of genus 2 Lefschetz fibrations”, *Math. Ann.* **311**:1 (1998), 163–176. [MR 99f:57035](#) [Zbl 0905.57014](#)
- [Gompf 1995] R. E. Gompf, “A new construction of symplectic manifolds”, *Ann. of Math. (2)* **142**:3 (1995), 527–595. [MR 96j:57025](#) [Zbl 0849.53027](#)
- [Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, *4-manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. [MR 2000h:57038](#) [Zbl 0933.57020](#)
- [Harer 1983] J. Harer, “The second homology group of the mapping class group of an orientable surface”, *Invent. Math.* **72**:2 (1983), 221–239. [MR 84g:57006](#) [Zbl 0533.57003](#)
- [Harris and Morrison 1998] J. Harris and I. Morrison, *Moduli of curves*, Graduate Texts in Mathematics **187**, Springer, New York, 1998. [MR 99g:14031](#) [Zbl 0913.14005](#)
- [Kametani and Sato 1994] Y. Kametani and Y. Sato, “0-dimensional moduli space of stable rank 2 bundles and differentiable structures on regular elliptic surfaces”, *Tokyo J. Math.* **17**:1 (1994), 253–267. [MR 95g:14012](#) [Zbl 0826.14023](#)
- [Korkmaz and Ozbagci 2001] M. Korkmaz and B. Ozbagci, “Minimal number of singular fibers in a Lefschetz fibration”, *Proc. Amer. Math. Soc.* **129**:5 (2001), 1545–1549. [MR 2001h:57019](#) [Zbl 1058.57011](#)
- [Kuno 2010] Y. Kuno, “On the global monodromy of a Lefschetz fibration arising from the Fermat surface of degree 4”, *Kodai Math. J.* **33**:3 (2010), 457–472. [MR 2754332](#) [Zbl 1221.14011](#)
- [Li 1999] T.-J. Li, “Smoothly embedded spheres in symplectic 4-manifolds”, *Proc. Amer. Math. Soc.* **127**:2 (1999), 609–613. [MR 99c:57055](#) [Zbl 0911.57018](#)
- [Li 2006] T.-J. Li, “Symplectic 4-manifolds with Kodaira dimension zero”, *J. Differential Geom.* **74**:2 (2006), 321–352. [MR 2007e:53114](#) [Zbl 1105.53068](#)
- [Li and Liu 1995] T. J. Li and A. Liu, “Symplectic structure on ruled surfaces and a generalized adjunction formula”, *Math. Res. Lett.* **2**:4 (1995), 453–471. [MR 96m:57052](#) [Zbl 0855.53019](#)
- [Liu 1996] A.-K. Liu, “Some new applications of general wall crossing formula, Gompf’s conjecture and its applications”, *Math. Res. Lett.* **3**:5 (1996), 569–585. [MR 97k:57038](#) [Zbl 0872.57025](#)
- [Matsumoto 1986] Y. Matsumoto, “Diffeomorphism types of elliptic surfaces”, *Topology* **25**:4 (1986), 549–563. [MR 88b:32061](#) [Zbl 0615.14023](#)
- [Matsumoto 1996] Y. Matsumoto, “Lefschetz fibrations of genus two—a topological approach”, pp. 123–148 in *Topology and Teichmüller spaces* (Katinkulta, 1995), edited by S. Kojima et al., World Scientific, River Edge, NJ, 1996. [MR 2000h:14038](#) [Zbl 0921.57006](#)

- [McCarthy and Wolfson 1994] J. D. McCarthy and J. G. Wolfson, “Symplectic normal connect sum”, *Topology* **33**:4 (1994), 729–764. [MR 95h:57038](#) [Zbl 0812.53033](#)
- [McDuff and Salamon 1996] D. McDuff and D. Salamon, “A survey of symplectic 4-manifolds with $b^+ = 1$ ”, *Turkish J. Math.* **20**:1 (1996), 47–60. [MR 97e:57028](#) [Zbl 0870.57023](#)
- [Meyer 1973] W. Meyer, “Die Signatur von Flächenbündeln”, *Math. Ann.* **201** (1973), 239–264. [MR 48 #9715](#) [Zbl 0241.55019](#)
- [Ohta and Ono 2005] H. Ohta and K. Ono, “Symplectic 4-manifolds containing singular rational curves with $(2, 3)$ -cusp”, pp. 233–241 in *Singularités Franco-Japonaises*, edited by J.-P. Brasselet and T. Suwa, Sémin. Congr. **10**, Soc. Math. France, Paris, 2005. [MR 2006g:53140](#) [Zbl 1082.53080](#)
- [Ozbagci and Stipsicz 2000] B. Ozbagci and A. I. Stipsicz, “Noncomplex smooth 4-manifolds with genus-2 Lefschetz fibrations”, *Proc. Amer. Math. Soc.* **128**:10 (2000), 3125–3128. [MR 2000m:57036](#) [Zbl 0951.57015](#)
- [Persson 1987] U. Persson, “An introduction to the geography of surfaces of general type”, pp. 195–218 in *Algebraic geometry, Bowdoin, 1985* (Brunswick, Maine, 1985), edited by S. J. Bloch, Proc. Sympos. Pure Math. **46**, Amer. Math. Soc., Providence, RI, 1987. [MR 89a:14057](#) [Zbl 0656.14020](#)
- [Persson 1992] U. Persson, “Genus two fibrations revisited”, pp. 133–144 in *Complex algebraic varieties* (Bayreuth, 1990), edited by K. Hulek et al., Lecture Notes in Math. **1507**, Springer, Berlin, 1992. [MR 1178724](#) [Zbl 0792.14018](#)
- [Sato 2006] Y. Sato, “The Stipsicz’s conjecture for genus-2 Lefschetz fibrations”, preprint, 2006.
- [Sato 2008] Y. Sato, “2-spheres of square -1 and the geography of genus-2 Lefschetz fibrations”, *J. Math. Sci. Univ. Tokyo* **15**:4 (2008), 461–491. [MR 2011c:57059](#) [Zbl 1236.57036](#)
- [Sato 2010a] Y. Sato, “An attempt to introduce the notion of the Itaka–Kodaira dimension into Lefschetz fibrations”, preprint, 2010.
- [Sato 2010b] Y. Sato, “The necessary condition on the fiber-sum decomposability of genus-2 Lefschetz fibrations”, *Osaka J. Math.* **47**:4 (2010), 949–963. [MR 2012j:57052](#) [Zbl 1228.57011](#)
- [Smith 1999] I. Smith, “Lefschetz fibrations and the Hodge bundle”, *Geom. Topol.* **3** (1999), 211–233. [MR 2000j:57059](#) [Zbl 0929.53047](#)
- [Smith 2001a] I. Smith, “Geometric monodromy and the hyperbolic disc”, *Q. J. Math.* **52**:2 (2001), 217–228. [MR 2002c:57046](#) [Zbl 0981.57013](#)
- [Smith 2001b] I. Smith, “Lefschetz pencils and divisors in moduli space”, *Geom. Topol.* **5** (2001), 579–608. [MR 2002f:57056](#) [Zbl 1066.57030](#)
- [Smith 2001c] I. Smith, “Torus fibrations on symplectic four-manifolds”, *Turkish J. Math.* **25**:1 (2001), 69–95. [MR 2002c:57047](#) [Zbl 0996.53055](#)
- [Stipsicz 1996] A. Stipsicz, “A note on the geography of symplectic manifolds”, *Turkish J. Math.* **20**:1 (1996), 135–139. [MR 97m:57035](#) [Zbl 0876.57039](#)
- [Stipsicz 1998] A. I. Stipsicz, “Simply connected 4-manifolds near the Bogomolov–Miyaoaka–Yau line”, *Math. Res. Lett.* **5**:6 (1998), 723–730. [MR 2000h:57047](#) [Zbl 0947.57032](#)
- [Stipsicz 1999] A. I. Stipsicz, “On the number of vanishing cycles in Lefschetz fibrations”, *Math. Res. Lett.* **6**:3-4 (1999), 449–456. [MR 2000g:57046](#) [Zbl 0955.57026](#)
- [Stipsicz 2000] A. I. Stipsicz, “The geography problem of 4-manifolds with various structures”, *Acta Math. Hungar.* **87**:4 (2000), 267–278. [MR 2001g:57050](#) [Zbl 0958.53034](#)
- [Stipsicz 2001] A. I. Stipsicz, “Indecomposability of certain Lefschetz fibrations”, *Proc. Amer. Math. Soc.* **129**:5 (2001), 1499–1502. [MR 2001h:57029](#) [Zbl 0978.57022](#)
- [Taubes 1994] C. H. Taubes, “The Seiberg–Witten invariants and symplectic forms”, *Math. Res. Lett.* **1**:6 (1994), 809–822. [MR 95j:57039](#) [Zbl 0853.57019](#)

- [Taubes 1995] C. H. Taubes, “The Seiberg–Witten and Gromov invariants”, *Math. Res. Lett.* **2**:2 (1995), 221–238. [MR 96a:57076](#) [Zbl 0854.57020](#)
- [Taubes 1996] C. H. Taubes, “SW \Rightarrow Gr: From the Seiberg–Witten equations to pseudo-holomorphic curves”, *J. Amer. Math. Soc.* **9**:3 (1996), 845–918. [MR 97a:57033](#) [Zbl 0867.53025](#)
- [Thurston 1976] W. P. Thurston, “Some simple examples of symplectic manifolds”, *Proc. Amer. Math. Soc.* **55**:2 (1976), 467–468. [MR 53 #6578](#) [Zbl 0324.53031](#)
- [Ue 1986] M. Ue, “On the diffeomorphism types of elliptic surfaces with multiple fibers”, *Invent. Math.* **84**:3 (1986), 633–643. [MR 87j:57019](#) [Zbl 0595.14028](#)
- [Usher 2006] M. Usher, “Minimality and symplectic sums”, *Int. Math. Res. Not.* **2006**:16 (2006), 1–17. Art. ID 49857. [MR 2007h:53139](#) [Zbl 1110.57017](#)

Received August 3, 2011. Revised July 11, 2012.

YOSHIHISA SATO
DEPARTMENT OF SYSTEMS DESIGN AND INFORMATICS
KYUSHU INSTITUTE OF TECHNOLOGY
680-4 KAWAZU, IIZUKA
FUKUOKA 820-8502
JAPAN
ysato@ces.kyutech.ac.jp

HILBERT–KUNZ INVARIANTS AND EULER CHARACTERISTIC POLYNOMIALS

LARRY SMITH

We study the Hilbert–Kunz invariants of homogeneous ideals in graded polynomial algebras and develop a homological formula for the Hilbert–Kunz multiplicity resembling the formula of J.-P. Serre using Koszul homology for the ordinary multiplicity of an ideal. We apply this in the special case of maximal primary irreducible ideals to obtain several new results, among which is a reciprocity formula for the Hilbert–Kunz invariants of directly linked ideals in a graded polynomial algebra.

The Hilbert–Kunz invariants grew out of the paper of E. Kunz [1969] characterizing regular local rings in characteristic $p \neq 0$ and they were put into their present form by P. Monsky [1983]. These invariants are defined by analogy with the Hilbert function and its associated multiplicity, but instead of using the ordinary powers of an ideal to do so, one uses its Frobenius powers instead. Specifically, fix a field \mathbb{F} of characteristic $p \neq 0$ and a commutative graded connected¹ \mathbb{F} -algebra A . Recall that for an ideal $I \subset A$ the *Frobenius power* $I^{[p]}$ of I is the ideal generated by the p -th powers of elements of I . If A is Noetherian and M is a finitely generated A -module, so M is of finite type,² one defines the *Hilbert–Kunz function* $\mathbf{HK}_{(I,M)} : \mathbb{N}_0 \longrightarrow \mathbb{N}_0$ of a maximal primary ideal $I \subset A$ on M by $\mathbf{HK}_{(I,M)}(e) = \dim_{\mathbb{F}}(M/I^{[p^e]} \cdot M)$ for $e \in \mathbb{N}_0$. The *Hilbert–Kunz multiplicity* of I on M is defined to be the real number

$$e_{\mathbf{HK}}(I, M) = \lim_{e \rightarrow \infty} \left\{ \frac{\dim_{\mathbb{F}}(M/I^{[p^e]} \cdot M)}{p^{e \cdot \dim(A)}} \right\} = \lim_{e \rightarrow \infty} \left\{ \frac{\mathbf{HK}_{(I,M)}(e)}{p^{e \cdot \dim(A)}} \right\}.$$

The fact that the numbers $\{\dim_{\mathbb{F}}(M/I^{[p^e]} \cdot M)/p^{e \cdot \dim(A)}\}_{e \in \mathbb{N}_0}$ form a bounded Cauchy sequence, so that the preceding limit makes sense, was proved by P. Monsky

MSC2010: 13A15, 13D02, 18G00.

Keywords: Hilbert–Kunz invariants, linkage of ideals.

¹By a *connected algebra* over \mathbb{F} is meant a nonnegatively graded algebra R whose degree 0 component is $\mathbb{F} \cdot 1$ where $1 \in R$ is the unit of the algebra. The terminology comes from algebraic topology: a (nonpathological) topological space X is connected if and only if its cohomology algebra is connected.

²A graded vector space V is of *finite type* if all its homogeneous components V_i are finite-dimensional.

(see, e.g., [Monsky 1983]). If the ideal I is the maximal ideal³ then one speaks of the Hilbert–Kunz function of M , and writes it as $\mathbf{HK}_M(-)$, and the Hilbert–Kunz multiplicity of M , and denotes it by $e_{\mathbf{HK}}(M)$.

The colength formula of [Watanabe and Yoshida 2000] for $e_{\mathbf{HK}}(I)$ provided our starting point. Using it we obtain a homological formula, Proposition 3.1, for $e_{\mathbf{HK}}(I)$ based on work of W. Smoke [1972] going back to D. Hilbert [1890]. The colength formula yields a relation between the Hilbert–Kunz multiplicity of bundle and base ideals in the context of the projective bundle theorem (see [Smith and Stong 2011] and Section 2), as well as the Hilbert–Kunz multiplicity of Gorenstein ideals with socle degree 2 or 3 in polynomial algebras (see Section 2 and Section 4). Proposition 3.1 also leads to a reciprocity relation for the Hilbert–Kunz multiplicity of a pair of directly linked ideals (see Section 5) in polynomial algebras.

We pay particular attention here to setting things up in a graded context. Being of the J. C. Moore school,⁴ a \mathbb{Z} -graded object X is a collection $\{X_i \mid i \in \mathbb{Z}\}$, not a direct sum, and only homogeneous elements are considered. If the direct sum makes sense we write $\text{Tot}(X)$ for the direct sum $\bigoplus X_i$ to distinguish it from the graded object X . Unless specifically mentioned to the contrary all graded vector spaces are nonnegatively graded, i.e., $X_i = 0$ for all $i < 0$.

1. Background from homological algebra

In this section we collect results from homological algebra needed for the proofs in the later sections. These consist of a brief review of [Smoke 1972] which formulates some fundamental ideas of D. Hilbert, in particular the syzygy theorem and its application to computing Poincaré series (see, e.g., [Hilbert 1890]) in the language of homological algebra.

Fix a ground field \mathbb{F} which for the present may be arbitrary. Let R denote a commutative graded connected algebra over \mathbb{F} . Unless otherwise stated to the contrary the algebra R should be assumed Noetherian, i.e., finitely generated over \mathbb{F} .

³In the graded context there is only one maximal ideal in A ; to wit, the *augmentation ideal*, which sometimes is referred to as the *irrelevant ideal*, denoted by \bar{A} and consisting of all the homogeneous elements of strictly positive degree and a zero in all other degrees, i.e., \bar{A} is the kernel of the *augmentation map* $\eta : A \rightarrow \mathbb{F}$ defined by

$$\eta(a) = \begin{cases} a & \text{if } \deg(a) = 0, \\ 0 & \text{otherwise.} \end{cases}$$

If $A = \mathbb{F}[x_1, \dots, x_n]$ is a polynomial algebra, then, since the notation $\mathbb{F}[x_1, \dots, x_n]$ is quite ugly we prefer to write \mathfrak{m} for the maximal ideal in this case and also to use the expression *\mathfrak{m} -primary* for a maximal primary ideal in $\mathbb{F}[x_1, \dots, x_n]$. More generally, we write \mathfrak{m}_A for the maximal ideal of A if A is a complicated symbol such as $\mathbb{F}[V]^G$.

⁴Though I myself am a Massey product.

Definition. A function ξ from isomorphism classes of R -modules of finite type to an abelian group A is called an *Euler characteristic with values in A* , or said to have the *Euler characteristic property*, if for every short exact sequence

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

of R -modules of finite type one has

$$\xi(M') + \xi(M'') = \xi(M).$$

Example 1. The Poincaré series,⁵ to wit,

$$P(M, t) = \sum_{i \in \mathbb{N}_0} \dim_{\mathbb{F}}(M_i) t^i,$$

for a finite type R -module M , defines an Euler characteristic with values in the abelian group $\mathbb{Z}[[t]]$ of formal power series with integral coefficients.

The following general nonsense result should at least be recorded. A proof is not really necessary (but if you insist on seeing one, consult, e.g., [Fraser 1969]).

Lemma 1.1. *Let R be a commutative graded connected algebra over \mathbb{F} and denote by $K(R)$ the Grothendieck group of the category of finite type R -modules. Then an Euler characteristic ξ with values in the abelian group A is nothing but a homomorphism of abelian groups $K(R) \longrightarrow A$.*

In other words, the map $[-]$ that assigns to an R -module of finite type its equivalence class in the Grothendieck group $K(R)$ is a universal function with the Euler characteristic property. This means that given any Euler characteristic ξ with values in the abelian group A , there is a unique group homomorphism $\Xi : K(R) \longrightarrow A$ such that the diagram

$$\begin{array}{ccc} \mathbf{MOD}_R & \xrightarrow{[-]} & K(R) \\ \xi \downarrow & \swarrow \Xi & \\ & & A \end{array}$$

commutes, where \mathbf{MOD}_R denotes the set (sic!) of isomorphism classes of R -modules of finite type.

Definition. A resolution of an R -module M of finite type

$$\mathcal{F} \quad \cdots \longrightarrow F_i \longrightarrow F_{i-1} \longrightarrow \cdots \longrightarrow F_0 \longrightarrow M \longrightarrow 0$$

⁵We prefer to work with Poincaré series in place of the Hilbert function.

is said to be *weakly minimal* if

$$\cdots > \text{cx}(F_i) > \text{cx}(F_{i-1}) > \cdots > \text{cx}(F_0),$$

where $\text{cx}(-)$ denotes the connectivity⁶ of the module $-$. If in addition all the induced maps $Q(F_i) \rightarrow Q(F_{i-1})$ of the vector spaces of indecomposables⁷ are zero for $i \in \mathbb{N}$ then we say that \mathcal{F} is *minimal*.

The notion of a weakly minimal resolution is a bit ad hoc, but it is the condition that was employed in [Broer et al. 2011] to prove the following lemmas culminating in the formula of Proposition 1.5 below, which is the nonequivariant version of the starting point for [Broer et al. 2011]. Also, by working with weakly minimal resolutions we can choose one resolution with some special algebraic structure to be put on homology modules, and another resolution to prove finiteness results such as in the next lemma. This kind of strategy was used in [Broer et al. 2011], especially Section 2, to incorporate a group action and character series. For another example of this kind see the discussion following Proposition 3.1 to follow where such a special structure is imposed for example by choosing a resolution as in [Buchsbaum and Eisenbud 1977], though one could alternatively invoke [Avramov and Golod 1971].

Lemma 1.2. *If \mathcal{F} is a weakly minimal resolution of an R -module of finite type with each term F_i of \mathcal{F} of finite type for $i \in \mathbb{N}_0$ then the alternating sum of their Poincaré series*

$$\sum_{i \in \mathbb{N}_0} (-1)^i P(F_i, t) \in \mathbb{Z}[[t]]$$

makes sense as a formal power series.

Proof. For any integer j there are only finitely many F_i for $i \in \mathbb{N}_0$ with $(F_i)_j \neq 0$, so for any i and j there are only finitely many $P(F_i, t)$ in which t^j has a nonzero coefficient. □

The next lemma says that the formal power series in Lemma 1.2 does not depend on the choice of the weakly minimal free resolution and provides a value for it.

Lemma 1.3. *If \mathcal{F} is a weakly minimal free resolution of an R -module M of finite type and each term F_i ($i \in \mathbb{N}_0$) of \mathcal{F} is of finite type, then*

$$P(R, t) \cdot \sum_{i \in \mathbb{N}_0} (-1)^i P(V_i, t) = \sum_{i \in \mathbb{N}_0} (-1)^i P(F_i, t) = P(M, t) \in \mathbb{Z}[[t]],$$

where $V_i = \mathbb{F} \otimes_R F_i$ is the indecomposable module of F_i for $i \in \mathbb{N}_0$.

⁶The *connectivity* $\text{cx}(M)$ of a graded vector space M is the largest integer such that $M_i = 0$ for $i \leq \text{cx}(M)$.

⁷If M is an R -module its vector space of *indecomposable elements* is $\mathbb{F} \otimes_R M$. It is often denoted by $Q(M)$.

Proof. This follows from the Euler characteristic property of the function $P(-, t)$ and the fact that $P(F_i, t) = P(R, t) \cdot P(V_i, t)$. □

So [Lemma 1.3](#) tells us for an R -module of finite type that the alternating sum

$$(1-1) \quad \sum_{i \in \mathbb{N}_0} (-1)^i P(F_i, t) \in \mathbb{Z}[[t]]$$

associated with a weakly minimal resolution \mathcal{F} of finite type⁸ is independent of the resolution \mathcal{F} . To evaluate it we are free to pick \mathcal{F} in a particular way; for example to be a minimal resolution. For a minimal resolution \mathcal{F} of a finite type module M one has⁹

$$F_i \cong R \otimes \text{Tor}_i^R(M, \mathbb{F}) \quad \text{for } i \in \mathbb{N}_0,$$

so we obtain a second way to evaluate the alternating sum (1-1). To wit:

Lemma 1.4. *If \mathcal{F} is a weakly minimal resolution of an R -module of finite type with each term F_i of \mathcal{F} of finite type for $i \in \mathbb{N}_0$ then*

$$\sum_{i \in \mathbb{N}_0} (-1)^i P(F_i, t) = P(R, t) \cdot \sum_{i \in \mathbb{N}_0} (-1)^i P(\text{Tor}_i^R(M, \mathbb{F}), t).$$

To summarize this part of the discussion we have proven the following result going back in spirit to [\[Hilbert 1890\]](#).

Proposition 1.5. *Let M be an R -module of finite type. Then*

$$P(M, t) = P(R, t) \cdot \sum_{i \in \mathbb{N}_0} (-1)^i \cdot P(\text{Tor}_i^R(M, \mathbb{F}), t).$$

2. Background on Hilbert–Kunz invariants and first applications

The definition of the Hilbert–Kunz multiplicity in general requires the asymptotics introduced by P. Monsky to prove its existence. However in the special case of ideals in a polynomial algebra this is unnecessary. The existence is a more or less a direct consequence of the formula for the Hilbert–Kunz function of an ideal in terms of the Frobenius functor and the exactness of that functor for polynomial algebras (see, e.g., [\[Huneke and Yao 2002, Remark 2.4\]](#)). For the sake of simplicity we assume that all algebras in this section have a *standard grading*, i.e., are generated as algebras by their homogeneous component of degree 1.

⁸We say that a resolution \mathcal{F} is of *finite type* if the modules F_i in it are finitely generated for all $i \in \mathbb{N}_0$.

⁹Minimal resolutions are unique up to isomorphism, hence *the* minimal resolution: though one should not overdo it here: the isomorphism is not unique nor functorial. The isomorphism $F_i \cong R \otimes \text{Tor}_i^R(\mathbb{F}, M)$ is a result of the fact that per definition of minimal resolution the differentials in the complex $\mathbb{F} \otimes_R \mathcal{F}$ are identically zero and of course F_i is a free R -module for $i \in \mathbb{N}_0$.

Proposition 2.1. *Let $I \subset \mathbb{F}[x_1, \dots, x_n]$ be a maximal primary ideal in the standard graded polynomial algebra $S = \mathbb{F}[x_1, \dots, x_n]$ over the field \mathbb{F} of characteristic $p \neq 0$ and set $H = \mathbb{F}[x_1, \dots, x_n]/I$. Then the Hilbert–Kunz function $\mathbf{HK}_{(I,S)}(-)$ is given by*

$$\mathbf{HK}_{(I,S)}(e) = p^{e \cdot n} \cdot \dim_{\mathbb{F}}(H) \quad \text{for } e \in \mathbb{N}_0,$$

and the Hilbert–Kunz multiplicity by $e_{\mathbf{HK}}(I, S) = \dim_{\mathbb{F}}(H)$, i.e., the colength of I .

The following simple example illustrates this; additional examples may be found further on in this section as well as [Section 4](#). It is due to F. S. Macaulay [[1916](#), Section 71] and has served ever since to demonstrate that irreducible ideals need not be generated by a regular sequence.

Example 2 (F. S. Macaulay). Let \mathbb{F} be a field and consider the five quadratic forms

$$z^2 - x^2, z^2 - y^2, xy, xz, yz \in \mathbb{F}[x, y, z]$$

and the ideal $M \subset \mathbb{F}[x, y, z]$ that they generate. The quotient algebra is easily seen to be a Poincaré duality algebra, so by [[Meyer and Smith 2005](#), Proposition I.1.4] the ideal M is irreducible. In fact, the quotient algebra is the \mathbb{F} -cohomology with the grading halved of the complex surface¹⁰ $\mathbb{C}\mathbb{P}(2) \# \mathbb{C}\mathbb{P}(2) \# \mathbb{C}\mathbb{P}(2)$, which is the connected sum of three copies of the complex projective plane $\mathbb{C}\mathbb{P}(2)$. The preceding formula tells us that for any field of characteristic $p \neq 0$ the Hilbert–Kunz multiplicity is $e_{\mathbf{HK}}(M, \mathbb{F}[x, y, z]) = 5$.

In [[Smith and Stong 2011](#)] we studied the algebra associated with the projective bundle theorem of algebraic topology (see, e.g., [[Stong 1968](#), p. 62]). For such ideals [Proposition 2.1](#) yields a formula for the Hilbert–Kunz multiplicity of the bundle ideal in terms of the base ideal and the bundle dimension. Recall that for $I \subset \mathbb{F}[V, X]$ an \mathfrak{m} -primary ideal and $J = I \cap \mathbb{F}[V]$ we call I a *projective bundle ideal* with *base ideal* J if $\mathbb{F}[V, X]/I$ is a free $\mathbb{F}[V]/J$ -module with respect to the module structure defined by the canonical inclusion $\mathbb{F}[V]/J \hookrightarrow \mathbb{F}[V, X]/I$. For such an ideal there is a coexact sequence¹¹

$$(2-1) \quad \mathbb{F} \longleftarrow \mathbb{F}[X]/(X^{k+1}) \longleftarrow \mathbb{F}[V, X]/I \longleftarrow \mathbb{F}[V]/J \longleftarrow \mathbb{F}$$

¹⁰Or of $\mathbb{C}\mathbb{P}(2) \# (S^2 \times S^2)$, which is the connected sum of a projective plain and a torus (see, e.g., the proof of Lemma 1.3 in [[Smith and Stong 2010](#)]).

¹¹If $A'' \xleftarrow{f''} A \xleftarrow{f'} A'$ are maps between commutative graded connected algebras, the sequence is called *coexact* if $\ker f''$ is the ideal $f'(\bar{A}') \cdot A$ of A generated by the image of the augmentation ideal \bar{A}' of A' . Equivalently, $f''(A) \cong \mathbb{F} \otimes_{A'} A$. The category $\mathcal{CC}\mathcal{A}_*$ of commutative graded connected algebras over a field has categorical images and cokernels, the image of a map $f : A' \rightarrow A''$ being the monomorphism $\iota_f : f(A') \hookrightarrow A''$ and the cokernel the epimorphism $\eta_f : A'' \rightarrow \mathbb{F} \otimes_{A'} A''$. To say that the sequence is coexact is equivalent to requiring that the natural map of the categorical cokernel of f' to categorical image of f'' is an isomorphism. The categorical cokernel of a map $f : R \rightarrow S$ in $\mathcal{CC}\mathcal{A}_*$ is often denoted by $R//f$ or $R//S$. So coexact is the categorical concept dual to exact.

of algebras. This sequence is an analogue of the coexact sequence of cohomology algebras

$$\mathbb{F} \longleftarrow H^*(\mathbb{C}\mathbb{P}(k); \mathbb{F}) \longleftarrow H^*(\mathbb{P}(\xi \downarrow B); \mathbb{F}) \longleftarrow H^*(B; \mathbb{F}) \longleftarrow \mathbb{F}$$

associated to a complex vector bundle $\xi \downarrow B$ of dimension $k + 1$ over the base space B , where $\mathbb{P}(\xi \downarrow B)$ is the associated projective space bundle (see, e.g., [Stong 1968, loc. cit.]). For this reason we use the following terminology in connection with the coexact sequence (2-1) of a projective bundle ideal. The integer $k + 1$ is called the *bundle dimension*, $\mathbb{F}[V, X]/I$ the *bundle algebra*, $\mathbb{F}[V]/J$ the *base algebra*, and $\mathbb{F}[X]/(X^{k+1})$ the *fiber algebra*. A detailed example of a projective bundle ideal follows Proposition 2.2 which relates the Hilbert–Kunz multiplicity of the three algebras in the coexact sequence (2-1).

The proof of Lemma 2.2 in [Smith and Stong 2011] yields the formula

$$(2-2) \quad P(\mathbb{F}[V, X]/I, t) = P(\mathbb{F}[V]/J, t) \cdot P(\mathbb{F}[X]/(X^{k+1}), t)$$

relating the Poincaré series of the terms of the coexact sequence (2-1). Therefore one has the following relation for the Hilbert–Kunz multiplicities (compare [Huneke and Yao 2002, Lemma 2.1]).

Proposition 2.2. *Let $I \subset \mathbb{F}[V, X]$ be a projective bundle ideal with base ideal $J \subset \mathbb{F}[V]$ and bundle dimension $k + 1$. Then*

$$e_{\mathbf{HK}}(I, \mathbb{F}[V, X]) = (k + 1) \cdot e_{\mathbf{HK}}(J, \mathbb{F}[V]).$$

Proof. Evaluate both sides of formula (2-2) at $t = 1$ and use Proposition 2.1. \square

Here is an example to illustrate Proposition 2.2. The choice of \mathbb{F}_2 as ground field is merely a convenience in relating this example to its topological origins.

Example 3. Let $t, r \in \mathbb{N}_0$ be nonnegative integers and $V = \mathbb{F}_2^{2t+r}$. Denote by $x_1, \dots, x_t, y_1, \dots, y_t, u_1, \dots, u_r$ a basis for the space V^* of linear forms on V . Choose a linear form $w_1 \in \mathbb{F}[V]_1$ and a quadratic form $w_2 \in \mathbb{F}[V]_2$. In the polynomial algebra $\mathbb{F}_2[V, X]$ consider the ideal I generated by the following $t^2 + \binom{t}{2} + 2tr + 1$ forms:

$$\begin{aligned} &x_1^2, \dots, x_t^2, y_1^2, \dots, y_t^2, \\ &x_i \cdot y_j \quad \text{for } 1 \leq i \neq j \leq t, \\ &x_i \cdot u_j \quad \text{for } 1 \leq i \leq t \text{ and } 1 \leq j \leq r, \\ &y_i \cdot u_j \quad \text{for } 1 \leq i \leq t \text{ and } 1 \leq j \leq r, \\ &X^2 + w_1 \cdot X + w_2. \end{aligned}$$

This is a projective bundle ideal with bundle dimension 2 and base ideal $J \subset \mathbb{F}_2[V]$ generated by all the previous forms except for $X^2 + w_1 \cdot X + w_2$. The quotient of

$\mathbb{F}_2[V]$ by the base ideal is the cohomology with \mathbb{F}_2 coefficients of the closed surface

$$F = (S^1 \times S^1) \# \cdots \# (S^1 \times S^1) \# \mathbb{R}P(2) \# \cdots \# \mathbb{R}P(2),$$

$$\xleftarrow{t} \xrightarrow{\quad} \xleftarrow{r} \xrightarrow{\quad}$$

where $\#$ denotes the connected sum of closed manifolds. So

$$e_{\mathbf{HK}}(\mathbb{F}[V]/J) = \dim_{\mathbb{F}_2}(\text{Tot}(\mathbb{F}[V]/J)) = 4t + 2r,$$

and since the bundle dimension is 2,

$$e_{\mathbf{HK}}(\mathbb{F}[V, X]/I) = 2 \cdot (4t + 2r).$$

The corresponding Poincaré duality quotient algebra $\mathbb{F}_2[V, X]/I$ is isomorphic to the \mathbb{F}_2 -cohomology of the projective space bundle of a 2-plane bundle ξ over the closed surface F whose Stiefel–Whitney classes are w_1 and w_2 .

Ever since the publication of [Cartan and Eilenberg 1956], change of rings phenomena have played an important role in algebra. An essential such result for Hilbert–Kunz multiplicity was proven in [Watanabe and Yoshida 2000]. Here it is in the graded form.

Theorem 2.3 (K. Watanabe and K. Yoshida). *Let $A \hookrightarrow B$ be a finite extension of graded connected commutative Noetherian integral domains over the field \mathbb{F} of characteristic $p \neq 0$ and $I \subset A$ a maximal primary ideal. Then*

$$e_{\mathbf{HK}}(I \cdot B, B) = r \cdot e_{\mathbf{HK}}(I, A),$$

where r is equal to the degree $|\mathbb{L} : \mathbb{K}|$ of the field extension $\mathbb{K} \subseteq \mathbb{L}$, and we have written \mathbb{K} for the field of fractions¹² $\mathbf{FF}(A)$ of A and \mathbb{L} for the field of fractions $\mathbf{FF}(B)$ of B .

The following example shows that an analogous formula for the Hilbert–Kunz function does not hold.

Example 4. Let A be the subalgebra of $B = \mathbb{F}[x, y]$ that is generated by x^2, xy, y^2 . Since $x^2, y^2 \in A$ is a system of parameters for B the extension $A \hookrightarrow B$ is finite. If \mathbb{K} is the field of fractions of A and \mathbb{L} the field of fractions of B then the field extension $\mathbb{K} \subset \mathbb{L}$ has degree $r = |\mathbb{L} : \mathbb{K}| = 2$. One way¹³ to see this is to let \mathbb{E} be

¹²The terminology *quotient field* or *field of fractions* of A , where A is a domain, is unfortunately not so clear as it might be. What is meant is the field consisting of all the fractions of the form a/b where $a, b \in A$ and $b \neq 0$; not the quotient of A by its maximal ideal. If A is graded then only homogeneous elements would be allowed and the resulting graded object is \mathbb{Z} -graded and a graded field. We employ the notation $\mathbf{FF}(A)$ for the field of fractions of an integral domain A , graded or not. In the special case of the field of fractions of $\mathbb{F}[V]$, we denote it by $\mathbb{F}(V)$, or $\mathbb{F}(z_1, \dots, z_n)$ if z_1, \dots, z_n is a basis for the linear forms.

¹³Alternatively, for \mathbb{F} of characteristic different from 2 one has $A \cong \mathbb{F}[x, y]^{\mathbb{Z}/2}$ where $\mathbb{Z}/2 < \text{GL}(2, \mathbb{F})$ is generated by $-\text{Id} \in \text{GL}(2, \mathbb{F})$. Galois theory would then tell us the degree of the extension is 2.

the field of fractions of $\mathbb{F}[x^2, y^2]$ so $\mathbb{E} \subsetneq \mathbb{K} \subsetneq \mathbb{L}$. Since $|\mathbb{L} : \mathbb{E}|$ has degree 4 the only possible value for $|\mathbb{L} : \mathbb{K}|$ is 2 since it must be a proper nontrivial divisor of 4.

We consider the augmentation ideal \bar{A} of A and note that $\dim_{\mathbb{F}}(A/\bar{A}) = 1 = \mathbf{HK}_{(\bar{A}, A)}(0)$. Note that $\bar{A} \cdot B = (x^2, xy, y^2) \subset \mathbb{F}[x, y] = B$ so $\mathbf{HK}_{(\bar{A} \cdot B, B)}(0) = \dim(B/(\bar{A} \cdot B)) = 4$. Therefore

$$\mathbf{HK}_{(\bar{A} \cdot B, B)}(0) = 4 \neq 2 \cdot 1 = r \cdot \mathbf{HK}_{(\bar{A}, A)}(0).$$

A similar computation would apply to any $e \in \mathbb{N}$ by [Proposition 2.1](#).

As an illustration of [Theorem 2.3](#) let us return to [Example 2](#) and instead of considering the ideal of $\mathbb{F}[x, y, z]$ generated by the five quadratic forms listed there the subalgebra they generate.

Example 5. Let $A \subset \mathbb{F}[x, y, z]$ be the subalgebra generated by the five forms

$$z^2 - x^2, z^2 - y^2, xy, xz, yz \in \mathbb{F}[x, y, z].$$

Then [Theorem 2.3](#) tells us that we can compute the Hilbert–Kunz multiplicity of A over a field of nonzero characteristic from the ideal in $\mathbb{F}[x, y, z]$ the five forms generate with the formula

$$e_{\mathbf{HK}}(A) = r \cdot e_{\mathbf{HK}}(\bar{A} \cdot \mathbb{F}[x, y, z], \mathbb{F}[x, y, z]),$$

where r is the degree of the field extension $\mathbf{FF}(A) \subset \mathbb{F}(x, y, z)$. That the degree of this field extension is 4 may be seen by enlarging \mathbb{F} to contain a square root \mathbf{i} of -1 . This does not change the degree of the resulting field extension. Then apply the automorphism α of $\mathbb{F}[x, y, z]$ given by sending z to $\mathbf{i} \cdot z$ and leaving x and y fixed. The algebra A gets mapped to $\alpha(A)$ which is generated by $z^2 + x^2, z^2 + y^2, xy, xz, yz \in \mathbb{F}[x, y, z]$. The element $z \in \mathbb{F}(x, y, z)$ is integral over $\alpha(A)$ with minimal polynomial

$$t^4 + (x^2 + y^2)t^2 + (xy)^2 \in \alpha(A)[t].$$

If we adjoin z to $\mathbf{FF}(\alpha(A))$ then the resulting field extension also contains $y = yz/z$ and $x = xz/z$ so coincides with $\mathbb{F}(x, y, z)$. Hence $r = 4$ and therefore $e_{\mathbf{HK}}(A) = e_{\mathbf{HK}}(\alpha(A)) = 5/4$.

If A is a commutative graded connected Noetherian algebra of Krull dimension $n = \dim(A)$ over the field \mathbb{F} then its Poincaré series has integral coefficients and a pole of order n at $t = 1$. Therefore the rational number

$$(1 - t)^n \cdot P(A, t) \Big|_{t=1} = \deg(A)$$

is well defined; it is called the *degree* of A (see [\[Smith 1995, Section 5.5\]](#) for a discussion of this invariant and its occurrence in invariant theory in particular). For a finite extension $A \hookrightarrow B$ of commutative graded connected Noetherian integral

domains, the ratio of their degrees is the degree¹⁴ of the corresponding field extension of their respective fields of fractions. Specifically, if \mathbb{K} is the field of fractions of A and \mathbb{L} the field of fractions of B then $\deg(B) = |\mathbb{L} : \mathbb{K}| \cdot \deg(A)$ (see, e.g., [Smith 1995, Proposition 5.5.2]). So using the notion of the degree of an algebra allows us to rephrase the change of rings theorem for integral domains in a more symmetric form.

Corollary 2.4. *Let $A \hookrightarrow B$ be a finite extension of Noetherian integral domains over the field \mathbb{F} of characteristic $p \neq 0$ and $I \subset A$ a maximal primary ideal. Then*

$$e_{\mathbf{HK}}(I \cdot B, B) \cdot \deg(A) = e_{\mathbf{HK}}(I, A) \cdot \deg(B).$$

In this next example it is easier to compute the degree of the subalgebra of $\mathbb{F}[V]$ being investigated rather than the degree of the field extension.

Example 6. Consider the subalgebra A in the polynomial algebra $\mathbb{F}[x, y, z]$ generated by the four forms x^2, xy, y^2, z^4 . It is not hard to see that x^2, y^2, z^4 is a system of parameters for A and that A is Cohen–Macaulay with basis $1, xy$ as an $\mathbb{F}[x^2, y^2, z^4]$ -module. Hence the Poincaré series of A is

$$P(A, t) = \frac{1+t^2}{(1-t^2)^2} \cdot \frac{1}{1-t^4} = \frac{1}{(1-t^2)^3}$$

so for the degree we have $\deg(A) = 1/8$. Since the quotient of $\mathbb{F}[x, y, z]$ by the ideal I has dimension 12 we obtain from Corollary 2.4 that $e_{\mathbf{HK}}(A) = 12/8 = 3/2$. So, although the Poincaré series of A looks like that of a polynomial algebra on three elements of degree 2 the Hilbert–Kunz invariant tells it is not (see, e.g., [Kunz 1969]). This example is well known from invariant theory (see, e.g., [Stanley 1979]).

3. An Euler characteristic formula for the Hilbert–Kunz multiplicity

In a famous paper, D. Hilbert [1890] proved not only the finiteness of the number of generators of the ring of invariants of certain classical groups, but also of the number of relations between invariants, and relations between relations, etc. In modern terms (we follow the notations and terminology of [Smith 1995]), and formulated for finite groups, what he did was to choose a minimal resolution¹⁵

$$(3-1) \quad 0 \longrightarrow F_n \longrightarrow F_{n-1} \longrightarrow \cdots \longrightarrow F_1 \longrightarrow F_0 \longrightarrow \mathbb{F}[V]_G \longrightarrow 0$$

of the ring¹⁶ of coinvariants $\mathbb{F}[V]_G$ of a representation $\rho : G \hookrightarrow \mathrm{GL}(n, \mathbb{F})$, of a finite group G over the field \mathbb{F} , regarded as an $\mathbb{F}[V]$ -module. Then, by the Euler

¹⁴This multiple use of *degree* hopefully will cause no confusion.

¹⁵Which he first had to prove existed!

¹⁶By definition the coinvariant algebra $\mathbb{F}[V]_G$ is $\mathbb{F} \otimes_{\mathbb{F}[V]^G} \mathbb{F}[V]$.

characteristic property of the Poincaré series one has, as in [Section 1](#),

$$(3-2) \quad P(\mathbb{F}[V]_G, t) = \sum_{i=0}^n (-1)^i P(F_i, t).$$

From the definition of a minimal resolution one finds (loc. cit.)

$$(3-3) \quad F_i \cong \mathbb{F}[V] \otimes \text{Tor}_i^{\mathbb{F}[V]}(\mathbb{F}[V]_G, \mathbb{F}) \quad \text{for } i = 0, \dots, n,$$

so putting (3-2) and (3-3) together leads to the formula

$$(3-4) \quad \begin{aligned} P(\mathbb{F}[V]_G, t) &= P(\mathbb{F}[V], t) \cdot \sum_{i=0}^n (-1)^i P(\text{Tor}_i^{\mathbb{F}[V]}(\mathbb{F}[V]_G, \mathbb{F}), t) \\ &= \frac{1}{(1-t)^n} \cdot \sum_{i=0}^n (-1)^i P(\text{Tor}_i^{\mathbb{F}[V]}(\mathbb{F}[V]_G, \mathbb{F}), t). \end{aligned}$$

The discussion in [Section 1](#) allows us to reformulate this in the following more general terms for use in computing Hilbert–Kunz multiplicities. It is an analog for Hilbert–Kunz multiplicity of the formula of J.-P. Serre (see, e.g., [\[Serre 1965, Part V\]](#)) for the ordinary multiplicity.

Proposition 3.1. *Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a graded polynomial algebra on generators with degrees $\deg(x_i) = d_i$ for $i = 1, \dots, n$ over the field \mathbb{F} of characteristic $p \neq 0$, $I \subset S$ an \bar{S} -primary ideal,¹⁷ and $S/I = R$. Then*

$$e_{\text{HK}}(I, S) = \left[\frac{1}{(1-t^{d_1}) \cdots (1-t^{d_n})} \cdot \sum_{i=0}^n (-1)^i P(\text{Tor}_i^S(R, \mathbb{F}), t) \right] \Big|_{t=1}.$$

Proof. This follows [Proposition 1.5](#), after accounting for the degrees of the variables, and [Proposition 2.1](#). □

Although this formula seems pretty useless on the surface (after all, how is one to compute the Poincaré series of the various syzygy modules without really having so firm a grip on R that one knows its Poincaré series already?), there are several answers to this objection in the case of irreducible ideals $I \subset S = \mathbb{F}[x_1, \dots, x_n]$, because, in this case, the algebra $\text{Tor}_*^S(R, \mathbb{F})_*$ carries the additional structure of a Poincaré duality algebra (see, e.g., [\[Meyer and Smith 2005, Part I\]](#) for a discussion of the relation between Poincaré duality quotients of graded Gorenstein algebras and irreducible ideals). Specifically, the modules (see [\[Avramov and Golod 1971\]](#) for the local case) $\text{Tor}_*^S(R, \mathbb{F})_*$ form a bigraded algebra, which, if we regrade it by total degree, are, apart from the cosmetic difference¹⁸ of being graded commutative

¹⁷There is no loss of generality in assuming that I is \bar{S} -primary since $e_{\text{HK}}(I, S) = 0$ if it is not.

¹⁸For an algebraic topologist in fact this is not a difference: it is with these commutation rules that Poincaré duality algebras arise as the cohomology of manifolds.

instead of commutative, a Poincaré duality algebra. Moreover, R itself is a Poincaré duality algebra (loc. cit.), and if R has formal dimension d (which means the socle of R is in homogeneous degree d) then the formal dimension of the singly graded algebra $\text{Tor}^S(R, \mathbb{F})$ is $n + d$, where $n = \dim_{\mathbb{F}}(V)$. Therefore the ordinary Poincaré series of this singly graded torsion algebra, to wit, the formal series

$$\sum_{i=0}^n P(\text{Tor}_i^S(R, \mathbb{F}), t),$$

must be a palindromic polynomial of degree $n + d$, i.e., if

$$\sum_{i=0}^n P(\text{Tor}_i^S(R, \mathbb{F}), t) = a_0 + a_1t + \cdots + a_{n+d}t^{n+d}, \quad a_0, \dots, a_{n+d} \in \mathbb{N}_0,$$

then $a_i = a_{n+d-i}$ for all $i = 0, \dots, \lfloor (n + d)/2 \rfloor$. This means that in case $n = 3$ we can actually write down a closed formula for the Hilbert–Kunz multiplicity of a maximal primary irreducible ideal $I \subset \mathbb{F}[x, y, z]$ knowing only the degrees of the generators of I and the socle degree of the quotient $R = \mathbb{F}[x, y, z]/I$.

Proposition 3.2. *Let \mathbb{F} be a field of characteristic $p \neq 0$, $I \subset \mathbb{F}[x, y, z]$ be a maximal primary irreducible ideal in the polynomial algebra on three generators x, y, z of degrees a, b, c , and set d equal to the degree of the socle of the quotient algebra $R = \mathbb{F}[x, y, z]/I$. Let the degrees of a minimal set of ideal generators for I be d_1, \dots, d_k . Then $e_{\mathbf{HK}}(I, \mathbb{F}[x, y, z])$ is equal to*

$$\left[\frac{1}{(1-t^a) \cdot (1-t^b) \cdot (1-t^c)} \cdot \left[1 - (t^{d_1} + \cdots + t^{d_k}) + (t^{3+d-d_1} + \cdots + t^{3+d-d_k}) - t^{3+d} \right] \right]_{t=1}.$$

Proof. Write S for $\mathbb{F}[x, y, z]$. Let (\mathcal{K}, ∂) be the Koszul resolution for \mathbb{F} regarded as an S -module which has

$$\mathcal{K} = S \otimes E(u, v, w), \quad \begin{cases} \partial(f \otimes 0) = 0 & \text{for } f \in S \text{ and} \\ \partial(1 \otimes u) = x, \quad \partial(1 \otimes v) = y, \quad \partial(1 \otimes w) = z. \end{cases}$$

So there are no boundaries of homological degree 3 and $f \otimes u \cdot v \cdot w$ is a cycle if and only if

$$0 = \partial(f \otimes u \cdot v \cdot w) = f \cdot x \otimes v \cdot w + f \cdot y \otimes u \cdot w + f \cdot w \otimes u \cdot v.$$

Since the elements vw, uw, uv are linearly independent in $E(u, v, w)$ this is the case if and only if $f \cdot x = f \cdot y = f \cdot w = 0$, and therefore $f \in \text{soc}(R)$. Hence

$$\text{Tor}_3^S(R, \mathbb{F}) = \text{soc}(R) \otimes uvw$$

and is 1-dimensional concentrated in degree $3 + \text{deg}(\text{soc}(R))$ just as it should be.

There is a short exact sequence $0 \rightarrow I \rightarrow S \rightarrow R \rightarrow 0$ of S -modules which leads to the long exact sequence of torsion modules

$$0 = \text{Tor}_1^S(S, \mathbb{F}) \rightarrow \text{Tor}_1^S(R, \mathbb{F}) \xrightarrow{\partial} I \otimes_S \mathbb{F} \rightarrow S \otimes_S \mathbb{F} \xrightarrow{\pi} R \otimes_S \mathbb{F} \rightarrow 0.$$

Since $S \otimes_S \mathbb{F} \cong \mathbb{F} \cong R \otimes_S \mathbb{F}$ the map π is an isomorphism and hence so is ∂ . This tells us that

$$P(\text{Tor}_1^S(R, \mathbb{F}), t) = t^{d_1} + \dots + t^{d_k}$$

and therefore the Euler characteristic polynomial for the torsion product is

$$P(\text{Tor}^S(R, \mathbb{F}), t) = 1 - (t^{d_1} + \dots + t^{d_k}) + (t^{3+d-d_1} + \dots + t^{3+d-d_k}) - t^{3+d},$$

as follows from the preceding discussion. The final formula is then a consequence of [Proposition 3.1](#). □

A maximal primary irreducible ideal I in a polynomial algebra $\mathbb{F}[V]$ would more often than not be specified by giving its Macaulay dual μ_I in the sense of [[Macaulay 1916](#), Part IV] (see also [[Meyer and Smith 2005](#), Parts I and VI]). The element μ_I may be viewed in several different ways: first, as an element of the divided polynomial algebra $\Gamma(V)$ with degree s equal to the formal dimension of the quotient algebra $\mathbb{F}[V]/I$; equivalently, as a form of degree $-s$ in the inverse polynomial algebra associated with $\mathbb{F}[V]$ and a basis for the space of linear forms V^* ; or, as an element in the local cohomology module $H_m^n(\mathbb{F}[V])_{-s-n}$, where $n = \dim_{\mathbb{F}}(V)$ (see, e.g., [[Greenlees and Smith 2008](#); [Smith 2013](#)]). So the degree of the socle would be a priori known. In the case of $\mathbb{F}_2[x, y, z]$ and socle degree 3 for the quotient algebra there are up to automorphism twenty-one possible choices for the Macaulay dual, and the corresponding ideals and quotient algebras have been classified and listed in [[Smith and Stong 2010](#)]. For all of these [Proposition 3.2](#) gives the Hilbert–Kunz multiplicity.¹⁹ Here is an example.

Example 7 [[Smith and Stong 2010](#), Section 5, Orbit 10]. Consider the inverse ternary cubic form²⁰

$$\theta_{10} = x^{-3} + y^{-3} + x^{-1}y^{-2} + x^{-1}y^{-1}z^{-1} \in \mathbb{F}_2[x^{-1}, y^{-1}, z^{-1}],$$

which defines a maximal primary ideal $I(\theta_{10}) \subset \mathbb{F}[x, y, z]$. Using the method of catalecticant matrices due to J. J. Sylvester (see, e.g., [[Meyer and Smith 2005](#),

¹⁹It would be interesting to know how to express the Hilbert–Kunz multiplicity of these examples in terms of the invariants from [[Smith and Stong 2010](#), Section 6] used to separate them.

²⁰This is the classical terminology for a form in three variables (cubic) of degree three. Since we are dealing with variables of degree -1 (inverse) this means that θ_{10} is a form in three inverse variables, here x^{-1} , y^{-1} , and z^{-1} , and has degree -3 .

Section 6.2]), one finds that this ideal is generated by the three forms $x^2 + xz + y^2$, $x^2 + yz$, z^2 as an ideal of $S = \mathbb{F}[x, y, z]$. If one examines the catalecticant matrix

$cat_{\theta_{10}}(1, 2)$	x^2	y^2	z^2	xy	xz	yz
x	1	1	0	0	0	1
y	0	1	0	1	1	0
z	0	0	0	1	0	0

representing this orbit, one can see that $z^2 = 0$, and that the algebra $H(\theta_{10}) = \mathbb{F}[x, y, z]/I(\theta_{10})$ corresponding to this matrix can be visualized as pictured in [Diagram 1](#). As in [\[Meyer and Smith 2005\]](#) the entries on a given horizontal line in the diagram are a basis for the homogeneous component of H of degree equal to the number of lines above the unit 1 of the algebra. So one reads off that the dimension of H is 8 and hence $e_{\mathbf{HK}}(I(\theta_{10}), S) = 8$.

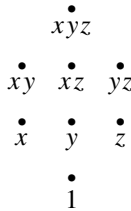


Diagram 1. The algebra $H(\theta_{10})$.

From [Diagram 1](#) one finds the relations

$$x^2 = yz, \quad y^2 = xz + yz.$$

This shows that $H(\theta_{10})$ is a free module over the subalgebra $\mathbb{F}[z]/(z^2) \subset H(\theta_{10})$ with basis the four elements $1, x, y, xy$. So $H(\theta_{10})$ looks like the \mathbb{F}_2 -cohomology of the total space M^3 of a fibering $S^1 \times S^1 \hookrightarrow M^3 \downarrow S^1$ which is totally nonhomologous to zero. Such a fibered manifold is constructed in [\[Smith and Stong 2010, Section 7\]](#).

The situation for surface algebras $H = \mathbb{F}[x, y, z]/I$, where I is a maximal primary irreducible ideal and the socle degree of H is 2 is somewhat simpler. Here is how this goes.

Example 8. Consider a nonzero inverse quadratic form μ in three inverse variables in $\mathbb{F}[x^{-1}, y^{-1}, z^{-1}]$ which defines a maximal primary ideal $I(\mu) \subset \mathbb{F}[x, y, z] = S$ with quotient algebra $R(\mu) = S/I(\mu)$ a Poincaré duality algebra of formal dimension 2; a surface algebra in the language of [\[Smith and Stong 2010\]](#). Making use of [Proposition 3.2](#) and [\[Eisenbud 1995, Exercise 21.6\]](#) allows us to construct the following table:

$\text{rank}(\mu)$	$P(\text{Tor}^S(R(\mu), t))$	$e_{\text{HK}}(I(\mu))$
1	$1 - 2t + t^2 - t^3 + 2t^4 - t^5$	3
2	$1 - t - t^2 + t^3 + t^4 - t^5$	4
3	$1 - 5t + 5t^2 - t^5$	5

Table 1. Hilbert–Kunz multiplicity of ternary surface algebras.

In [Smith and Stong 2010, Section 2] we showed that any surface algebra over \mathbb{F}_2 can be written as a connected sum of the two basic examples: $\mathbb{F}_2[x, y]/(x^2, y^2)$, with Macaulay dual form of rank 2, and $\mathbb{F}_2[z]/(z^3)$, with Macaulay dual form of rank 1, so with the aide of the above table and Proposition 4.1 one has a formula for the Hilbert–Kunz multiplicity of any surface algebra at least over \mathbb{F}_2 . See the discussion of connected sums in Section 4 and Examples 10, 11 there.

In fact, already the two-variable case of Proposition 3.1 is interesting, as we explain next. We use its proof to provide a short and simple proof of the result of F. S. Macaulay that an irreducible ideal in a polynomial algebra in two variables is generated by a regular sequence (for a different modern proof, see, e.g., [Vasconcelos 1967]).

Theorem 3.3 [Macaulay 1904]. *Let \mathbb{F} be a field and $I \subset \mathbb{F}[x, y] = S$ an ideal such that $R = S/I$ is a Poincaré duality algebra. Then I is generated by a regular sequence.*

Proof. To evaluate the formula in Proposition 3.1 in this case we recycle the proof of Proposition 3.2 to compute $\text{Tor}_i^S(R, \mathbb{F})$ for $i = 1$ and 2. This results in the formula

$$P(R, t) = \frac{1}{(1-t^a) \cdot (1-t^b)} \cdot [1 - (t_1^k + \dots + t^{k_r}) + t^{2+d}],$$

where $\deg(x) = a$, $\deg(y) = b$, $d = a + b$, k_1, \dots, k_r are the degrees of a minimal set of generators for I , and $d = \text{f-dim}(R)$, i.e., the socle degree of R . The left hand side of this equality is a polynomial so the right hand side must be one also. This says that

$$(1 - t^a) \cdot (1 - t^b) = (1 - t)^2 \cdot (1 + t + \dots + t^{a-1}) \cdot (1 + t + \dots + t^{b-1})$$

must divide

$$(3-5) \quad p(t) = 1 - (t^{k_1} + \dots + t^{k_r}) + t^{2+d}$$

so $p(1) = 0$. Evaluating $p(1)$ from the formula (3-5) and equating the result to zero gives $0 = p(1) = 2 - r$, so $r = 2$ and I is generated by two elements f, h which must then be a system of parameters since R is totally finite. Since S is Cohen–Macaulay it follows that $f, h \in S = \mathbb{F}[x, y]$ is a regular sequence. \square

We can change viewpoint and replace the ideal I in the statement of [Proposition 3.1](#) with a subalgebra A such as an algebra of invariants. The reformulated result takes the following form. An illustrative example is given in [Example 9](#).

Proposition 3.4. *Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a graded polynomial algebra on generators with degrees $\deg(x_i) = d_i$ for $i = 1, \dots, n$ over the field \mathbb{F} of characteristic $p \neq 0$, $A \subset S$ a subalgebra making S into a finitely generated A -module, and set²¹ $R = S // A \cong S \otimes_A \mathbb{F}$. Then*

$$e_{\text{HK}}(A) = \deg(A) \cdot \left[\frac{1}{(1-t^{d_1}) \cdots (1-t^{d_n})} \cdot \sum_{i=0}^n (-1)^i P(\text{Tor}_i^S(R, \mathbb{F}), t) \right] \Big|_{t=1}.$$

Proof. This follows from [Theorem 2.3](#), [Proposition 1.5](#) after accounting for the degrees of the variables, and [Proposition 2.1](#). □

4. Further applications and examples

In this section we collect some examples of computations of Hilbert–Kunz invariants to illustrate the behavior of these in special circumstances. We begin with the possibility that there is an integral form of the algebra being studied. Then one can ask if, and if so how, these invariants change with the characteristic. Rings of invariants of permutation groups are natural candidates in this context. The following example provides such a case where there seems to be a connection with \mathbb{F} -rationality (see, e.g., [[Glassbrenner 1995](#); [Singh 1998](#); [Smith 2004](#)]).

Example 9. Consider the ring of invariants $\mathbb{F}[z_1, \dots, z_n]^{A_n}$ of the alternating group A_n acting by means of its tautological permutation representation on the variables. Denote by $e_1, \dots, e_n \in \mathbb{F}[z_1, \dots, z_n]$ the elementary symmetric polynomials in z_1, \dots, z_n . These are invariants of the full symmetric group Σ_n and hence also of its alternating subgroup, so they belong to $\mathbb{F}[z_1, \dots, z_n]^{A_n}$. If the characteristic of \mathbb{F} is not 2 and we restrict the permutation representation of Σ_n to the alternating subgroup A_n , then, as is also well known, the ring of invariants $\mathbb{F}[z_1, \dots, z_n]^{A_n}$ is a complete intersection generated by e_1, \dots, e_n and the discriminant

$$\Delta_n = \prod_{1 \leq i < j \leq n} (z_i - z_j) = \sum_{\sigma \in \Sigma_n} \text{sgn}(\sigma) \cdot z_{\sigma(1)}^0 z_{\sigma(2)}^1 \cdots z_{\sigma(n)}^{n-1},$$

the square Δ_n^2 being a polynomial in e_1, \dots, e_n given by the resultant of φ_n and φ'_n (see, e.g., [[Smith 1995](#), Section 1.3; [Glassbrenner 1995](#), Section 12]), where

$$\varphi_n(t) = e_n + e_{n-1} \cdot t + \cdots + e_1 \cdot t^{n-1} + t^n = \prod_{i=1}^n (t + z_i) \in \mathbb{F}[z_1, \dots, z_n][t].$$

²¹Recall that $S // A$ is defined to be $S \otimes_A \mathbb{F}$ and is the categorical cokernel of the map including A into S in the category of commutative graded connected algebras over the field \mathbb{F} .

Less well known²² would appear to be the invariants in characteristic 2. If we set

$$\mathfrak{S}_n = \sum_{\sigma \in A_n} z_{\sigma(1)}^0 z_{\sigma(2)}^1 \cdots z_{\sigma(n)}^{n-1},$$

then *regardless of the characteristic*, $\mathbb{F}[z_1, \dots, z_n]^{A_n}$ is a hypersurface algebra (hence Gorenstein) generated by e_1, \dots, e_n and \mathfrak{S}_n , the square \mathfrak{S}_n^2 being a polynomial in e_1, \dots, e_n (see, e.g., [Smith 1995, Theorem 1.3.5]).

D. Glassbrenner [1995] discovered that the Hilbert ideals²³ $\mathfrak{h}(A_n)$ and $\mathfrak{h}(\Sigma_n)$ coincide if the characteristic p of the field \mathbb{F} divides $\binom{n}{2}$. This was extended to all odd $p \leq n$ in [Singh 1998] and all $p \leq n$ in [Smith 2004]. Specifically, one has

$$\mathfrak{h}(A_n) = (e_1, \dots, e_n) = \mathfrak{h}(\Sigma_n) \iff p \leq n.$$

This being the case, we get from Proposition 3.4 the following formulae for the Hilbert–Kunz multiplicity of the algebra $\mathbb{F}[z_1, \dots, z_n]^{A_n}$ as a function of the characteristic of the ground field \mathbb{F} (see also [Brenner 2010] for a more complete discussion of Hilbert–Kunz multiplicities of algebras of invariants):

$$e_{\text{HK}}(\mathbb{F}[z_1, \dots, z_n]^{A_n}) = \begin{cases} \frac{n!}{(1/2) \cdot n!} = 2 & \text{if } p \leq n, \\ \frac{n! - 1}{(1/2) \cdot n!} = 2 - \frac{2}{n!} & \text{otherwise.} \end{cases}$$

This follows from the discussion of this example in [Smith 2004], in particular the computation of a Macaulay dual for the Hilbert ideal $\mathfrak{h}(A_n)$, which shows that $\mathbb{F}[z_1, \dots, z_n]_{A_n}$ is the algebra $\mathbb{F}[z_1, \dots, z_n]_{\Sigma_n}$ with the socle removed if $p > n$, and that the degree of the algebra $\mathbb{F}[z_1, \dots, z_n]^{A_n}$ is $(1/2) \cdot n!$ independent of the characteristic of \mathbb{F} (see, e.g., [Smith 1995, Theorem 5.5]).

Remark. If one lets $n \uparrow \infty$ in these formulae one gets 2 in all cases, i.e., independent of p . Does this have any significance? Can it be explained by some integral analog for integral alternating invariants of the ring of integral symmetric polynomials in infinitely many variables?

A standard way to study ideals, or even to define special properties for them, is to examine the corresponding quotient algebra. In [Smith and Stong 2010; 2011] we studied a natural construction coming from algebraic topology on Poincaré duality algebras called the *connected sum*.²⁴ If R' and R'' are Poincaré duality algebras over the field \mathbb{F} of the same formal dimension d then their connected sum $R' \# R''$ is

²²This fact gets *rediscovered* every couple of years and published circa once a decade.

²³If $\rho : G \hookrightarrow \text{GL}(n, \mathbb{F})$ is a representation of a group over the field \mathbb{F} then the *Hilbert ideal* is the ideal in $\mathbb{F}[z_1, \dots, z_n]$ generated by all G -invariant forms of strictly positive degree.

²⁴This construction seems much more natural on the quotients than on the ideals defining them.

defined by identifying in their direct sum the two units and fundamental classes,²⁵ so by the requirements

$$(R' \# R'')_i = \begin{cases} \mathbb{F} \cdot 1 & \text{if } i = 0, \\ R'_i \oplus R''_i & \text{if } 0 < i < d, \text{ and} \\ \mathbb{F} \cdot [R' \# R''] & \text{if } i = d, \end{cases}$$

where $[R'] \in R'$ and $[R''] \in R''$ are chosen fundamental classes. Put another way, if $\mathbb{S}(d)$ denotes the Poincaré duality algebra²⁶ with $\mathbb{F} \cdot 1$ in degree 0 and $[\mathbb{S}] \cdot \mathbb{F}$ in degree d with all other homogeneous degrees being 0, then for any Poincaré duality algebra R of formal dimension d with fundamental class $[R]$ there is a natural map $\tau : \mathbb{S}(d) \rightarrow R$ sending unit to unit and fundamental class to fundamental class. The connected sum is defined by requiring that

$$\begin{array}{ccc} \mathbb{S}(d) & \xrightarrow{\tau} & R' \\ \downarrow & \square & \downarrow \\ R'' & \xrightarrow{\tau} & R' \# R'' \end{array}$$

be a pushout diagram.

If $R' = S'/I'$ and $R'' = S''/I''$ where S' and S'' are standard graded polynomial algebras so $I' \subset S'$ and $I'' \subset S''$ are maximal primary irreducible ideals, then $R' \# R''$ is of the form $(S' \otimes S'')/I$ for a maximal primary irreducible ideal $I' \# I'' = I \subset S = S' \otimes S''$ in the standard graded polynomial algebra S . From the colength formula in Proposition 2.1 we then get the following formula for the Hilbert–Kunz multiplicity.

Proposition 4.1. *Let $S' = \mathbb{F}[x'_1, \dots, x'_{n'}]$ and $S'' = \mathbb{F}[x''_1, \dots, x''_{n''}]$ be standard graded polynomial algebras over the field \mathbb{F} of characteristic $p \neq 0$, and let $I' \subset S'$ and $I'' \subset S''$ be maximal primary ideals with Poincaré duality quotients $R' = S'/I'$ and $R'' = S''/I''$ of the same formal dimension $d > 0$. If $I \subset S' \otimes S'' = \mathbb{F}[x_1, \dots, x_{n'}, x''_1, \dots, x''_{n''}]$ defines the Poincaré duality quotient algebra $R' \# R''$ as a quotient of S then $e_{\mathbf{HK}}(I, S) = e_{\mathbf{HK}}(I', S) + e_{\mathbf{HK}}(I'', S'') - 2$.*

The case $d = 0$ of the previous result is trivial because in this case $R' = \mathbb{F} = R''$ so I is the maximal ideal and $e_{\mathbf{HK}}(I, S) = 1$. The result for more than two parts in the connected sum is easily extended by induction to yield the formula

$$e_{\mathbf{HK}}(I(1) \# \dots \# I(k), S) = e_{\mathbf{HK}}(I(1), (S(1))) + \dots + e_{\mathbf{HK}}(I(k), (S(k))) - 2 \cdot (k - 1)$$

← k →

for the Hilbert–Kunz multiplicity of the ideal defining the connected sum of k irreducible ideals in k standard graded polynomial algebras.

²⁵A fundamental class is a nonzero element of the socle.

²⁶This is nothing but $H^*(S^d; \mathbb{F})$.

Corollary 4.2. *Let $I \subset \mathbb{F}_2[z_1, \dots, z_n] = S$ be a maximal primary irreducible ideal with quotient algebra $H = S/I$ a surface algebra, i.e., the socle degree of H is 2. If $q_I \in \mathbb{F}[z_1^{-1}, \dots, z_n^{-1}]$ is an inverse quadratic form that is a Macaulay dual for I then $e_{\mathbf{HK}}(I, S) = 2 + \text{rank}(q_I)$.*

Proof. By [Smith and Stong 2010, Corollary 2.6] any surface algebra over \mathbb{F}_2 is a connected sum of the algebras with ranks 1 or 2 listed in Table 1. The result then follows from Proposition 4.1 by induction on the number of terms in the connected sum. □

As an example of Proposition 4.1 we return to Example 2 from Section 2.

Example 10. The connected sum

$$(\mathbb{F}[x]/(x^3)) \# (\mathbb{F}[y]/(y^3)) \# (\mathbb{F}[z]/(z^3))$$

is a Poincaré duality quotient of $\mathbb{F}[x, y, z]$ with formal dimension 3. Its defining ideal is the ideal M of Example 2. Therefore we find from Proposition 4.1 for its Hilbert–Kunz multiplicity $e_{\mathbf{HK}}(M, \mathbb{F}[x, y, z]) = 3 + 3 + 3 - 2 \cdot 2 = 9 - 4 = 5$, just as computed previously.

In [Smith and Stong 2010] we provided several criteria to check if a Poincaré duality algebra is in fact a connected sum, one of which we use in the next example.

Example 11. Consider the ideal $I \subset \mathbb{F}[x, y, z] = S$ generated by the five quadratic forms

$$x^2, y^2, xz, yz, z^2 - xy.$$

It is not hard to see that Lemma 1.1 of [Smith and Stong 2010] applies to the quotient algebra $R = S/I$ with $H'_1 = \text{Span}_{\mathbb{F}}\{x, y\}$ and $H''_1 = \text{Span}_{\mathbb{F}}\{z\}$ so²⁷ $R = \mathbb{F}[x, y]/(x^2, y^2) \# \mathbb{F}[z]/(z^3)$. Therefore Proposition 4.1 tells us that $e_{\mathbf{HK}}(I, S) = 4 + 3 - 2 = 5$.

Remark. Let $S' = \mathbb{F}[x'_1, \dots, x'_{n'}]$ and $S'' = \mathbb{F}[x''_1, \dots, x''_{n''}]$ be standard graded polynomial algebras over the field \mathbb{F} of characteristic $p \neq 0$ and $I' \subset S'$, $I'' \subset S''$ maximal primary ideals with Poincaré duality quotients $R' = S'/I'$ and $R'' = S''/I''$ of the same formal dimension $d > 0$. If $I \subset S' \otimes S'' = \mathbb{F}[x_1, \dots, x_{n'}, x''_1, \dots, x''_{n''}]$ defines the Poincaré duality algebra $R = R' \# R''$ as a quotient of S , then the theorem of [Avramov and Golod 1971] tells us that the three torsion products

$$\text{Tor}^{S'}(R', \mathbb{F}), \quad \text{Tor}^{S''}(R'', \mathbb{F}), \quad \text{and} \quad \text{Tor}^S(R, \mathbb{F})$$

²⁷Topologists should recognize this as $H^*((S^2 \times S^2) \# \mathbb{C}\mathbb{P}(2); \mathbb{F})$ after halving the grading degrees; algebraists as the ideal with Macaulay dual $z^{-2} + x^{-1}y^{-1} \in \mathbb{F}[x^{-1}, y^{-1}, z^{-1}]$ (see, e.g., [Eisenbud 1995, Example 21.7]). In characteristic 2 the algebras in this and the previous example are isomorphic; see, e.g., [Smith and Stong 2010, Lemma 2.4].

are Poincaré duality algebras, the first of dimension $d + n'$, the second of dimension $d + n''$ and the third of dimension $d + n$, where $n = n' + n''$. So all three of the algebras

$$\text{Tor}^{S'}(R', \mathbb{F}) \otimes E'', \quad \text{Tor}^{S''}(R'', \mathbb{F}) \otimes E', \quad \text{and} \quad \text{Tor}^S(R, \mathbb{F})$$

are Poincaré duality algebras of formal dimension $d + n$, where E'', E' are exterior the algebras $E'' = \text{Tor}^{S''}(\mathbb{F}, \mathbb{F}) = E(u''_1, \dots, u''_{n''})$ and $E' = \text{Tor}^{S'}(\mathbb{F}, \mathbb{F}) = E(u'_1, \dots, u'_{n'})$. If we regard R' and R'' as quotients of $S = S' \otimes S''$ by means of the maps

$$R' \cong R' \otimes \mathbb{F} \xleftarrow{\pi' \otimes \epsilon''} S' \otimes S'' \xrightarrow{\epsilon' \otimes \pi''} \mathbb{F} \otimes R'' \cong R'',$$

where ϵ', ϵ'' are the augmentation maps of S' and S'' respectively, and π' and π'' the quotient maps from S' and S'' onto R' and R'' respectively, then

$$\begin{aligned} \text{Tor}^{S'}(R', \mathbb{F}) \otimes E'' &\cong \text{Tor}^S(R', \mathbb{F}), \\ \text{Tor}^{S''}(R'', \mathbb{F}) \otimes E' &\cong \text{Tor}^S(R'', \mathbb{F}), \end{aligned}$$

so all three of the torsion products

$$\text{Tor}^S(R', \mathbb{F}), \quad \text{Tor}^S(R'', \mathbb{F}), \quad \text{and} \quad \text{Tor}^S(R, \mathbb{F})$$

become Poincaré duality algebras of formal dimension $d + n$. Moreover there is a map

$$\eta : \text{Tor}^S(R', \mathbb{F}) \# \text{Tor}^S(R'', \mathbb{F}) \longrightarrow \text{Tor}^S(R' \# R'', \mathbb{F})$$

of degree one basically induced by forming the connected sum of the two maps

$$\text{Tor}^S(R', \mathbb{F}) \longrightarrow \text{Tor}^S(R' \# R'', \mathbb{F}) \longleftarrow \text{Tor}^S(R'', \mathbb{F}).$$

The map η being of degree one must be a monomorphism (see, e.g., the proof, not the statement, of Lemma I.3.1 in [Meyer and Smith 2005]). It does not seem to be an isomorphism for the case of the connected sum $\mathbb{F}[x]/(x^3) \# \mathbb{F}[y]/(y^3)$: so what can we say about it?

5. Reciprocity formulae for linked ideals

Recall that two ideals $I, J \subset A$ in a commutative graded connected algebra A over the field \mathbb{F} are said to be *directly linked* if there is a regular sequence $f_1, \dots, f_m \in \bar{A}$ such that

$$I = ((f_1, \dots, f_m) :_A J) \quad \text{and} \quad J = ((f_1, \dots, f_m) :_A I).$$

In this case one also says that I and J are linked over the complete intersection ideal $\mathfrak{f} = (f_1, \dots, f_m)$ in A . If A is a Gorenstein algebra, then an ideal generated by a regular sequence of maximal length is irreducible (see, e.g., [Meyer and Smith 2005,

Proposition I.1.4 and Lemma I.1.3]) and hence the Noether involution theorem [loc. cit., Theorem I.2.1] assures us that either one of these conditions implies the other.

The purpose of this section is to prove the following reciprocity formula for the Hilbert–Kunz multiplicity of a pair of directly linked maximal primary ideals in a polynomial algebra:

$$(5-1) \quad e_{\mathbf{HK}}(I, S) + e_{\mathbf{HK}}(J, S) = e_{\mathbf{HK}}(\mathfrak{f}, S).$$

Here $S = \mathbb{F}[x_1, \dots, x_n]$ is a polynomial algebra over the field \mathbb{F} , $\mathfrak{f} = (f_1, \dots, f_n) \subset S$ is an ideal generated by a regular sequence $f_1, \dots, f_n \in \bar{S}$ of maximal length, and $I \subset S$ is a maximal primary ideal containing \mathfrak{f} with $J = (\mathfrak{f} :_A I)$ the directly linked ideal. Note that the right hand side of (5-1) may be evaluated by means of the colength formula to yield

$$e_{\mathbf{HK}}(\mathfrak{f}, S) = \dim_{\mathbb{F}}(S/\mathfrak{f}) = \frac{\prod_{i=1}^n \deg(f_i)}{\prod_{i=1}^n \deg(x_i)},$$

since S is a free module over the subalgebra $\mathbb{F}[f_1, \dots, f_n]$.

The plan for the proof of formula (5-1) is to use Proposition 2.1 and first prove the reciprocity formula

$$(5-2) \quad \dim_{\mathbb{F}}(S/I) + \dim_{\mathbb{F}}(S/J) = \dim_{\mathbb{F}}(S/\mathfrak{f})$$

for the dimensions of the corresponding quotient algebras.²⁸ To do this we make use of some elementary homological *tic-toc-toe*. We begin with the following basic fact.²⁹

Lemma 5.1. *Let A be a commutative graded connected algebra over the field \mathbb{F} , $f_1, \dots, f_n \in \bar{A}$, and M an A -module. If f_1, \dots, f_n form a regular sequence on M , then on the category of $A/(f_1, \dots, f_n)$ -modules there are natural equivalences of functors*

$$\text{Ext}_A^i(-, M) \cong \begin{cases} \text{Hom}_A(-, M/(f_1, \dots, f_n) \cdot M) & \text{if } i = n, \\ 0 & \text{for } i < n. \end{cases}$$

The proof of this lemma rests on the following observation.

Lemma 5.2. *Let A be a commutative graded connected algebra over the field \mathbb{F} and M, N a pair of A -modules. If $\text{Ann}_A(N)$ contains a regular element on M then $\text{Hom}_A(N, M) = 0$.*

²⁸Here, and throughout this section, we abuse notation and write $\dim_{\mathbb{F}}(X)$, where X is a totally finite graded vector space for the more correct $\dim_{\mathbb{F}}(\text{Tot}(X))$.

²⁹Versions of these lemmas go back at least to [Serre 1965] and can be found in [Bass 1963, Proposition 2.9] as well as [Bruns and Herzog 1993, Lemmas 1.2.3 and 1.2.4].

Proof. Let $\varphi : N \rightarrow M$ be a homomorphism of A -modules and $u \in \text{Ann}_A(N)$ a regular element on M . If $w \in N$ then $u \cdot \varphi(w) = \varphi(u \cdot w) = \varphi(0) = 0$ implies that $\varphi(w) = 0$ since u is a regular element on M ; hence $\varphi = 0$ since w was arbitrary. \square

Proof of Lemma 5.1. By induction on n . For $n = 0$ there is nothing to prove, so suppose that $n > 0$ and the result is established for $n - 1$. Let N be an $A/(f_1, \dots, f_n)$ -module. By the induction hypothesis,

$$\text{Ext}_A^{n-1}(N, M) \cong \text{Hom}_A(N, M/(f_1, \dots, f_{n-1}) \cdot M).$$

Since $f_n \in \text{Ann}_A(N)$ is a regular element on the A -module $M/(f_1, \dots, f_{n-1}) \cdot M$, Lemma 5.2 tells us that $\text{Hom}_A(N, M/(f_1, \dots, f_{n-1}) \cdot M) = 0$. Therefore of course $\text{Ext}_A^{n-1}(N, M) = 0$ as well.

The element $f_n \in \bar{A}$ being regular on M means one has a short exact sequence of A -modules

$$0 \rightarrow M \xrightarrow{\cdot f_n} M \rightarrow M/f_n \cdot M \rightarrow 0.$$

The long exact sequence for $\text{Ext}^\bullet(N, -)$ associated to it yields³⁰

$$0 = \text{Ext}_A^{n-1}(N, M) \rightarrow \text{Ext}_A^{n-1}(N, M/f_n \cdot M) \xrightarrow{\delta} \text{Ext}_A^{n-1}(N, M) \xrightarrow{\cdot f_n} \text{Ext}_A^n(N, M) \rightarrow \dots$$

The map $\cdot f_n$ is induced by multiplication with f_n on M , but, $\text{Ext}_A^\bullet(-, -)$ is a balanced functor so it is equally well induced by multiplication with f_n on N which is the zero map. Therefore $\delta : \text{Ext}_A^{n-1}(N, M) \rightarrow \text{Ext}_A^n(N, M)$ is an isomorphism. The $n - 1$ elements f_1, \dots, f_{n-1} form a regular sequence on $M/f_n \cdot M$, so the induction hypothesis yields an isomorphism

$$\text{Ext}_A^{n-1}(N, M/f_n \cdot M) \cong \text{Hom}_A(N, M/(f_1, \dots, f_{n-1}, f_n) \cdot M),$$

completing the inductive proof that $\text{Ext}_A^n(N, M) \cong \text{Hom}_A(N, M/(f_1, \dots, f_n) \cdot M)$. To complete the inductive step note that for $k > 0$ we have $\text{Ext}_A^{n-k}(N, M) \cong \text{Hom}(N, M/(f_1, \dots, f_{n-k}) \cdot M)$ and f_n is a regular element on the quotient module $M/(f_1, \dots, f_{n-k}) \cdot M$. Since f_n annihilates N , Lemma 5.2 tells us that $\text{Hom}(N, M/(f_1, \dots, f_{n-k}) \cdot M) = 0$, and hence $\text{Ext}_A^{n-k}(N, M) = 0$ as well. \square

There are a number of special cases of these lemmas that are relevant to the notion of linkage. We need to record these, but before we do so, note that, if $f_1, \dots, f_n \in \bar{A}$ is a regular sequence in the commutative graded connected algebra A over the field \mathbb{F} and the ideal $\mathfrak{f} = (f_1, \dots, f_n)$ is irreducible, then the Noether involution theorem (see, e.g., [Meyer and Smith 2005, Theorem I.2.1]) implies that $J = (\mathfrak{f} :_A I)$ if and only if $I = (\mathfrak{f} :_A J)$. This is the case if $n = \dim(A)$ and A is Gorenstein. It will

³⁰We will use a \bullet to denote the indexing of derived functors rather than a $*$ to distinguish it from the internal grading index on these functors.

allow us under these circumstances to interchange the roles of I and J in the next result.

Lemma 5.3. *Let A be a commutative graded connected algebra over the field \mathbb{F} , $f_1, \dots, f_n \in \bar{A}$ a regular sequence, and $I \subset S$ a maximal primary ideal. Set $J = ((f_1, \dots, f_n) :_A I)$. Then $\text{Ext}_A^n(S/I, A) \cong J/(f_1, \dots, f_n)$.*

Proof. In Lemma 5.1, put $\text{---} = A/I$ and $M = A$. The result is an isomorphism

$$\text{Ext}_A^n(A/I, A) \cong \text{Hom}_A(A/I, A/(f_1, \dots, f_n)).$$

Any element $\varphi \in \text{Hom}_A(A/I, A/(f_1, \dots, f_n))$ is determined by $\varphi(1)$ from the requirement that it be an A -module homomorphism, viz., $\varphi(a) = \varphi(a \cdot 1) = a \cdot \varphi(1)$. In order that this formula define a map $A/I \rightarrow A/(f_1, \dots, f_n)$ it is necessary and sufficient that $\varphi(1)$ annihilate the image of I in $A/(f_1, \dots, f_n)$. Note that

$$\begin{aligned} \varphi(1) \in \text{Ann}_{A/(f_1, \dots, f_n)}(I/(f_1, \dots, f_n)) &= \left(0 \quad : \quad I/(f_1, \dots, f_n) \right)_{A/(f_1, \dots, f_n)} \\ &\cong ((f_1, \dots, f_n) :_A I)/(f_1, \dots, f_n) \\ &\cong J/(f_1, \dots, f_n). \end{aligned}$$

Hence the map $\text{Hom}_A(A/I, A/(f_1, \dots, f_n)) \rightarrow J/(f_1, \dots, f_n)$ defined by sending an element $\varphi \in \text{Hom}_A(A/I, A/(f_1, \dots, f_n))$ to $\varphi(1) \in J/(f_1, \dots, f_n)$ is an isomorphism, which combined with the isomorphism of Lemma 5.1, $\text{Ext}_A^n(A/I, A) \cong \text{Hom}_A(A/I, A/(f_1, \dots, f_n))$, yields the desired conclusion. \square

Remark. As a special case of Lemma 5.3 we can put $I = (f_1, \dots, f_n)$ and conclude

$$\text{Ext}_A^n(A/(f_1, \dots, f_n), A) \cong A/(f_1, \dots, f_n).$$

This will prove useful in the sequel.

In Lemma 5.3 the Noether involution theorem tells us that if the ideal $(f_1, \dots, f_n) \subset A$ is maximal primary and irreducible then we can interchange the roles of I and J . What is somewhat surprising is that we can also interchange the roles of A/I and $J/(f_1, \dots, f_n)$ if A is a polynomial algebra,³¹ to wit:

Lemma 5.4. *Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a graded polynomial algebra over the field \mathbb{F} and $f_1, \dots, f_n \in \bar{S}$ a regular sequence (so the ideal $(f_1, \dots, f_n) \subset S$ is maximal primary and irreducible). Let $I \subset S$ be an ideal containing f_1, \dots, f_n and $J = ((f_1, \dots, f_n) :_S I)$ the directly linked ideal. Then $\text{Ext}_S^n(J/(f_1, \dots, f_n), S) \cong S/I$.*

Proof. Consider the short exact sequence of S -modules

$$0 \rightarrow J/(f_1, \dots, f_n) \rightarrow S/(f_1, \dots, f_n) \rightarrow S/J \rightarrow 0.$$

³¹Careful study of the proof shows it would be enough to suppose that the ideal (f_1, \dots, f_n) is maximal primary and irreducible as well as $\text{Ext}_S^{n+1}(S/J, S) = 0$.

Apply the functor $\text{Ext}_S^n(-, S)$ to it. One gets a long exact sequence

$$(5-3) \quad \cdots \longleftarrow \text{Ext}_S^n(J/(f_1, \dots, f_n), S) \longleftarrow \text{Ext}_S^n(S/(f_1, \dots, f_n), S) \\ \longleftarrow \text{Ext}_S^n(S/J, S) \longleftarrow \cdots.$$

By [Lemma 5.3](#) and the Remark following it we find

$$\text{Ext}_S^n(S/J, S) \cong I/(f_1, \dots, f_n)$$

and

$$\text{Ext}_S^n(S/(f_1, \dots, f_n), S) \cong S/(f_1, \dots, f_n).$$

If we put this into [\(5-3\)](#) we obtain the exact sequence

$$\cdots \longleftarrow \text{Ext}_S^n(J/(f_1, \dots, f_n), S) \longleftarrow S/(f_1, \dots, f_n) \longleftarrow I/(f_1, \dots, f_n) \longleftarrow \cdots.$$

The map $I/(f_1, \dots, f_n) \rightarrow S/(f_1, \dots, f_n)$ is monic, and in addition the map $\text{Ext}_S^n(S/(f_1, \dots, f_n), S) \rightarrow \text{Ext}_S^n(J/(f_1, \dots, f_n), S)$ in the exact sequence [\(5-3\)](#) is epic since its cokernel lies in $\text{Ext}_S^{n+1}(S/J, S)$, which is zero because S has global dimension n . Therefore we have a short exact sequence

$$0 \longleftarrow \text{Ext}_S^n(J/(f_1, \dots, f_n), S) \longleftarrow S/(f_1, \dots, f_n) \longleftarrow I/(f_1, \dots, f_n) \longleftarrow 0$$

so $\text{Ext}_S^n(J/(f_1, \dots, f_n), S) \cong S/I$ as required. \square

Again, Noether's involution theorem tells us we can interchange the roles of I and J in this lemma. In the remainder of this section we will use the isomorphism

$$S/J \cong \text{Ext}_S^n(I/(f_1, \dots, f_n), S)$$

to prove the formula [\(5-2\)](#), from which the formula [\(5-1\)](#) follows by [Proposition 2.1](#). To do this we will construct a weakly minimal free resolution of $I/(f_1, \dots, f_n)$, use (see [Lemma 5.1](#)) that $\text{Ext}_S^i(I/(f_1, \dots, f_n), S) = 0$ for $i \neq n$, and the Euler characteristic of an exact sequence is zero, as well as [Lemma 1.3](#). We begin with a review of the mapping cone construction from homological algebra (see, e.g., [\[MacLane 1963, pp. 46–47\]](#)).

Recollection. If $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ is a map of chain complexes the *mapping cone* $\mathcal{C}(\varphi) = \mathcal{C}$ is the chain complex with chains $\mathcal{C}_i = \mathcal{B}_i \oplus \mathcal{A}_{i-1}$ for $i \in \mathbb{Z}$ and boundary ∂ maps defined by $\partial(b, a) = (\partial_B(b) + \varphi(a), \partial_A(a))$ where ∂_B, ∂_A are the boundary maps of the complexes \mathcal{B} and \mathcal{A} , respectively.

Note that the mapping cone \mathcal{C} of a chain map $\varphi : \mathcal{A} \rightarrow \mathcal{B}$ fits into a short exact sequence of complexes

$$0 \rightarrow \mathcal{B} \xrightarrow{\iota_\varphi} \mathcal{C} \xrightarrow{\pi} \Sigma(\mathcal{A}) \rightarrow 0,$$

where the map ι_φ is defined by $\iota_\varphi(b) = (b, 0)$. In the resulting long exact sequence in homology the boundary map $\partial : H_i(\Sigma(\mathcal{C})) \rightarrow H_{i-1}(\mathcal{B})$ may be identified up to sign with the induced map $\varphi_* : H_{i-1}(\mathcal{C}) \rightarrow H_{i-1}(\mathcal{B})$ (loc. cit.).

Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a polynomial algebra over the field \mathbb{F} , $f_1, \dots, f_n \in \bar{S}$ a regular sequence, and $I \subset S$ a maximal primary ideal containing f_1, \dots, f_n with directly linked ideal $J = ((f_1, \dots, f_n) :_S I)$. We next describe³² how to construct (see, e.g., [Peskin and Szpiro 1974, Proposition 2.6; Martsinkovsky and Strooker 2004, Proposition 10]) a (weakly minimal) free resolution of $I/(f_1, \dots, f_n)$ as S -module. Choose (weakly minimal) free resolutions of finite type, \mathcal{F} of S/I and \mathcal{K} of $S/(f_1, \dots, f_n)$ (e.g., \mathcal{K} could be the Koszul complex for $f_1, \dots, f_n \in \bar{S}$) as S -modules. Let \mathcal{C} be the mapping cone of a map of complexes $\varphi : \mathcal{K} \rightarrow \mathcal{F}$ lifting the natural quotient map $S/(f_1, \dots, f_n) \rightarrow S/I$. Then \mathcal{C} is a (weakly minimal) complex of free S -modules of finite type. We claim that apart from a degree shift it is a resolution of $I/(f_1, \dots, f_n)$. To see this we examine the long exact homology sequence associated with the exact sequence of complexes

$$0 \rightarrow \mathcal{F} \rightarrow \mathcal{C} \rightarrow \Sigma(\mathcal{K}) \rightarrow 0.$$

Since \mathcal{F} and \mathcal{K} are acyclic the only portion of this long exact sequence with nonzero terms looks as follows:

$$\begin{array}{ccccccc} 0 & \rightarrow & H_1(\mathcal{C}) & \rightarrow & H_1(\Sigma(\mathcal{K})) & \xrightarrow{\partial} & H_0(\mathcal{F}) \rightarrow H_0(\mathcal{C}) \rightarrow 0 \\ & & \uparrow \cong & & \uparrow \cong & & \\ & & S/(f_1, \dots, f_n) & \xrightarrow{\pi} & S/I, & & \end{array}$$

where π is the natural quotient map. Hence $H_0(\mathcal{C}) = 0$ and $H_1(\mathcal{C}) \cong I/(f_1, \dots, f_n)$. Therefore we have proven the following result (loc. cit.).

Lemma 5.5. *Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a polynomial algebra over the field \mathbb{F} , $f_1, \dots, f_n \in \bar{S}$ a regular sequence, and $I \subset S$ a maximal primary ideal containing f_1, \dots, f_n with directly linked ideal $J = ((f_1, \dots, f_n) :_S I)$. Let \mathcal{F} be a (weakly minimal) free resolution of S/I and \mathcal{K} of $S/(f_1, \dots, f_n)$. If \mathcal{C} is the mapping cone of a map of complexes lifting the natural quotient map $S/(f_1, \dots, f_n) \rightarrow S/I$ then $\Sigma^{-1}(\mathcal{C})$ is a (weakly minimal) free resolution of $I/(f_1, \dots, f_n)$, where $\Sigma^{-1}(\mathcal{C})$ denotes the shifted³³ complex $\Sigma^{-1}(\mathcal{C})_i = \mathcal{C}_{i+1}$ for $i \in \mathbb{Z}$.*

Continuing with the notations preceding Lemma 5.5 we note that the cocomplex $\mathcal{H} = \text{Hom}_S(\Sigma^{-1}(\mathcal{C}), S)$ has as cohomology $H^\bullet(\mathcal{H}) = \text{Ext}_S^\bullet(I/(f_1, \dots, f_n), S)$ and that by Lemma 5.1, $\text{Ext}_S^i(I/(f_1, \dots, f_n), S) = 0$ for $i \neq n$. We augment the

³²The geometric version of this construction would appear to be due to D. Ferrand (see [Peskin and Szpiro 1974, Section 2]).

³³A topologist would say *desuspended*.

cocomplex \mathcal{H} with $\text{Ext}_S^n(I/(f_1, \dots, f_n), S)$ to obtain an exact sequence, viz.,

$$0 \leftarrow \text{Ext}_S^n(I/(f_1, \dots, f_n), S) \leftarrow \mathcal{H}^n \leftarrow \dots \leftarrow \mathcal{H}^0 \leftarrow \mathcal{H}^{-1} \leftarrow 0.$$

The Euler characteristic of an exact sequence is zero, so after rearranging things we obtain the following equality for Euler characteristic polynomials:

$$P(\text{Ext}_S^n(I/(f_1, \dots, f_n), S), t) = \sum (-1)^i P(\mathcal{H}_i, t).$$

At this point we require an elementary, but necessary, lemma.

Lemma 5.6. *Let A be a commutative graded connected algebra over the field \mathbb{F} and L a finitely generated free A -module. Then as graded vector spaces $\text{Hom}_A(L, A) \cong Q(L)^* \otimes A$ where $Q(L)^* = \text{Hom}_{\mathbb{F}}(L \otimes_A \mathbb{F}, \mathbb{F})$, where $Q(L) = L \otimes_A \mathbb{F}$.*

Proof. Set $Q(L) = L \otimes_A \mathbb{F}$. We have isomorphisms of graded vector spaces

$$\begin{aligned} \text{Hom}_A(L, A) &\cong \text{Hom}_A(A \otimes Q(L), A) \cong \text{Hom}_{\mathbb{F}}(Q(L), \mathbb{F}) \\ &\cong \text{Hom}_{\mathbb{F}}(\text{Hom}_{\mathbb{F}}(\mathbb{F}, Q(L)), A) \cong \text{Hom}_{\mathbb{F}}(\mathbb{F}, Q(L)^* \otimes A) \cong Q(L)^* \otimes A, \end{aligned}$$

where the next to the last isomorphism results from the $\text{Hom} - \otimes$ interchange. \square

Returning to the discussion preceding the lemma, write $\mathcal{F} = S \otimes \mathcal{U}$ and $\mathcal{K} = S \otimes \mathcal{V}$ as bigraded vector spaces. Unravel the definition of the cocomplex \mathcal{H} and use [Lemma 5.6](#) to write

$$\begin{aligned} \mathcal{H} &= \text{Hom}_S(\Sigma^{-1}(\mathcal{C}), S) = \Sigma^{-1}(\text{Hom}_S(\mathcal{F} \oplus \Sigma(\mathcal{K}), S)) \\ &= \Sigma^{-1}(\text{Hom}_S(\mathcal{F}, S) \oplus \text{Hom}_S(\Sigma(\mathcal{K}), S)) \\ &= \Sigma^{-1}(\text{Hom}_S(S \otimes \mathcal{U}, S) \oplus \text{Hom}_S(S \otimes \Sigma(\mathcal{V}), S)) \\ &= [\Sigma^{-1}(S \otimes \text{Hom}_{\mathbb{F}}(\mathcal{U}, \mathbb{F}))] \oplus [S \otimes \text{Hom}_{\mathbb{F}}(\Sigma(\mathcal{V}), \mathbb{F})] \end{aligned}$$

as graded vector spaces. By taking Euler characteristic polynomials and applying [Lemma 1.3](#) we obtain

$$\begin{aligned} (5-4) \quad P(\text{Ext}_S^n(I/(f_1, \dots, f_n), S), t) &= P(S, t) \cdot \sum (-1)^i P(\mathcal{U}_i, t) - P(S, t) \cdot \sum (-1)^i P(\mathcal{V}_i, t) \\ &= P(S/(f_1, \dots, f_n), t) - P(S/I, t), \end{aligned}$$

and with this we can prove the formula [\(5-2\)](#); to wit:

Theorem 5.7. *Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a polynomial algebra over the field \mathbb{F} , $f_1, \dots, f_n \in \bar{S}$ a regular sequence, and $I \subset S$ a maximal primary ideal containing f_1, \dots, f_n with directly linked ideal $J = ((f_1, \dots, f_n) :_S I)$. Then*

$$\dim_{\mathbb{F}}(S/I) + \dim_{\mathbb{F}}(S/J) = \dim_{\mathbb{F}}(S/(f_1, \dots, f_n)).$$

Proof. By [Lemma 5.1](#) we have an isomorphism

$$(5-5) \quad S/J \cong \text{Ext}_S^n(I/(f_1, \dots, f_n), S).$$

By [\(5-4\)](#) we have an equality of Poincaré series

$$P(\text{Ext}_S^n(I/(f_1, \dots, f_n), S), t) = P(S/(f_1, \dots, f_n), t) - P(S/I, t),$$

so putting $t = 1$ into this equality yields an equality of dimensions

$$(5-6) \quad \dim_{\mathbb{F}}(\text{Ext}_S^n(I/(f_1, \dots, f_n), S)) = \dim_{\mathbb{F}}(S/(f_1, \dots, f_n)) - \dim_{\mathbb{F}}(S/I)$$

for the corresponding vector spaces. Combining the two equalities [\(5-5\)](#) and [\(5-6\)](#) completes the proof. \square

Corollary 5.8. *Let $S = \mathbb{F}[x_1, \dots, x_n]$ be a polynomial algebra over the field \mathbb{F} , $f_1, \dots, f_n \in \bar{S}$ a regular sequence, and $I \subset S$ a maximal primary ideal containing f_1, \dots, f_n with directly linked ideal $J = ((f_1, \dots, f_n) :_S I)$. Then*

$$e_{\text{HK}}(S/I) + e_{\text{HK}}(S/J) = e_{\text{HK}}(S/(f_1, \dots, f_n)).$$

Proof. This follows from [Theorem 5.7](#) and [Proposition 2.1](#). \square

References

- [Avramov and Golod 1971] L. L. Avramov and E. S. Golod, “The homology of algebra of the Koszul complex of a local Gorenstein ring”, *Mat. Zametki* **9** (1971), 53–58. In Russian, translated in *Math. Notes* **9** (1971), 30–32. [MR 43 #4883](#) [Zbl 0222.13014](#)
- [Bass 1963] H. Bass, “On the ubiquity of Gorenstein rings”, *Math. Z.* **82** (1963), 8–28. [MR 27 #3669](#) [Zbl 0112.26604](#)
- [Brenner 2010] H. Brenner, “A remark on the colength of the Hilbert Ideal”, preprint, Universität Osnabrück, 2010.
- [Broer et al. 2011] A. Broer, V. Reiner, L. Smith, and P. Webb, “Extending the coinvariant theorems of Chevalley, Shephard–Todd, Mitchell, and Springer”, *Proc. Lond. Math. Soc.* (3) **103**:5 (2011), 747–785. [MR 2012k:13017](#) [Zbl 1242.13008](#)
- [Bruns and Herzog 1993] W. Bruns and J. Herzog, *Cohen–Macaulay rings*, Cambridge Studies in Advanced Mathematics **39**, Cambridge University Press, 1993. [MR 95h:13020](#) [Zbl 0788.13005](#)
- [Buchsbaum and Eisenbud 1977] D. A. Buchsbaum and D. Eisenbud, “Algebra structures for finite free resolutions, and some structure theorems for ideals of codimension 3”, *Amer. J. Math.* **99**:3 (1977), 447–485. [MR 56 #11983](#) [Zbl 0373.13006](#)
- [Cartan and Eilenberg 1956] H. Cartan and S. Eilenberg, *Homological algebra*, Princeton University Press, 1956. [MR 17,1040e](#) [Zbl 0075.24305](#)
- [Eisenbud 1995] D. Eisenbud, *Commutative algebra: With a view toward algebraic geometry*, Graduate Texts in Mathematics **150**, Springer, New York, 1995. [MR 97a:13001](#) [Zbl 0819.13001](#)
- [Fraser 1969] M. Fraser, “Multiplicities and Grothendieck groups”, *Trans. Amer. Math. Soc.* **136** (1969), 77–92. [MR 38 #2138](#) [Zbl 0172.05103](#)
- [Glassbrenner 1995] D. Glassbrenner, “The Cohen–Macaulay property and F -rationality in certain rings of invariants”, *J. Algebra* **176**:3 (1995), 824–860. [MR 96h:13009](#) [Zbl 0847.14004](#)

- [Greenlees and Smith 2008] J. P. C. Greenlees and L. Smith, “Local cohomology and Macaulay’s dual systems”, preprint, 2008.
- [Hilbert 1890] D. Hilbert, “Ueber die Theorie der algebraischen Formen”, *Math. Ann.* **36**:4 (1890), 473–534. [MR 1510634](#) [JFM 22.0133.01](#)
- [Huneke and Yao 2002] C. Huneke and Y. Yao, “Unmixed local rings with minimal Hilbert–Kunz multiplicity are regular”, *Proc. Amer. Math. Soc.* **130**:3 (2002), 661–665. [MR 2002h:13026](#) [Zbl 0990.13011](#)
- [Kunz 1969] E. Kunz, “Characterizations of regular local rings for characteristic p ”, *Amer. J. Math.* **91** (1969), 772–784. [MR 40 #5609](#) [Zbl 0188.33702](#)
- [Macaulay 1904] F. S. Macaulay, “On a method of dealing with the intersections of plane curves”, *Trans. Amer. Math. Soc.* **5**:4 (1904), 385–410. [MR 1500679](#) [JFM 35.0587.01](#)
- [Macaulay 1916] F. S. Macaulay, *The algebraic theory of modular systems*, Cambridge University Press, 1916. [MR 95i:13001](#) [JFM 46.0167.01](#)
- [MacLane 1963] S. MacLane, *Homology*, Grundlehren der math. Wissenschaften **114**, Academic Press, New York, 1963. [MR 28 #122](#) [Zbl 0133.26502](#)
- [Martsinkovsky and Strooker 2004] A. Martsinkovsky and J. R. Strooker, “Linkage of modules”, *J. Algebra* **271**:2 (2004), 587–626. [MR 2005b:16002](#) [Zbl 1099.13026](#)
- [Meyer and Smith 2005] D. M. Meyer and L. Smith, *Poincaré duality algebras, Macaulay’s dual systems, and Steenrod operations*, Cambridge Tracts in Mathematics **167**, Cambridge University Press, 2005. [MR 2006h:13012](#) [Zbl 1083.13003](#)
- [Monsky 1983] P. Monsky, “The Hilbert–Kunz function”, *Math. Ann.* **263**:1 (1983), 43–49. [MR 84k:13012](#) [Zbl 0509.13023](#)
- [Peskine and Szpiro 1974] C. Peskine and L. Szpiro, “Liaison des variétés algébriques, I”, *Invent. Math.* **26** (1974), 271–302. [MR 51 #526](#) [Zbl 0298.14022](#)
- [Serre 1965] J.-P. Serre, *Algèbre locale: Multiplicités*, Lecture Notes in Mathematics **11**, Springer, Berlin, 1965. [MR 34 #1352](#) [Zbl 0142.28603](#)
- [Singh 1998] A. K. Singh, “Failure of F -purity and F -regularity in certain rings of invariants”, *Illinois J. Math.* **42**:3 (1998), 441–448. [MR 99g:13004](#) [Zbl 0915.13004](#)
- [Smith 1995] L. Smith, *Polynomial invariants of finite groups*, Research Notes in Mathematics **6**, A K Peters, Wellesley, MA, 1995. [MR 96f:13008](#) [Zbl 0864.13002](#)
- [Smith 2004] L. Smith, “On alternating invariants and Hilbert ideals”, *J. Algebra* **280**:2 (2004), 488–499. [MR 2005f:13006](#) [Zbl 1085.13002](#)
- [Smith 2013] L. Smith, “Ideals of generalized invariants, local cohomology, and Macaulay’s double duality theorem”, *Forum Math.* (2013).
- [Smith and Stong 2010] L. Smith and R. E. Stong, “Poincaré duality algebras mod two”, *Adv. Math.* **225**:4 (2010), 1929–1985. [MR 2011j:13020](#) [Zbl 1214.13002](#)
- [Smith and Stong 2011] L. Smith and R. E. Stong, “Projective bundle ideals and Poincaré duality algebras”, *J. Pure Appl. Algebra* **215**:4 (2011), 609–627. [MR 2011j:13006](#) [Zbl 1206.13004](#)
- [Smoke 1972] W. Smoke, “Dimension and multiplicity for graded algebras”, *J. Algebra* **21** (1972), 149–173. [MR 46 #9024](#) [Zbl 0231.13006](#)
- [Stanley 1979] R. P. Stanley, “Invariants of finite groups and their applications to combinatorics”, *Bull. Amer. Math. Soc. (N.S.)* **1**:3 (1979), 475–511. [MR 81a:20015](#) [Zbl 0497.20002](#)
- [Stong 1968] R. E. Stong, *Notes on cobordism theory*, Princeton University Press, 1968. [MR 40 #2108](#) [Zbl 0181.26604](#)

[Vasconcelos 1967] W. V. Vasconcelos, “Ideals generated by R -sequences”, *J. Algebra* **6** (1967), 309–316. [MR 35 #4209](#) [Zbl 0147.29301](#)

[Watanabe and Yoshida 2000] K.-i. Watanabe and K.-i. Yoshida, “Hilbert–Kunz multiplicity and an inequality between multiplicity and colength”, *J. Algebra* **230**:1 (2000), 295–317. [MR 2001h:13032](#) [Zbl 0964.13008](#)

Received December 1, 2011. Revised December 26, 2012.

LARRY SMITH
AG-INVARIANTENTHEORIE
MITTELWEG 3
D-37133 FRIEDLAND
GERMANY
larry@uni-math.gwdg.de

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use \LaTeX , but papers in other varieties of \TeX , and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as \LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of $\text{Bib}\TeX$ 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 262 No. 1 March 2013

On the second K -group of a rational function field	1
KARIM JOHANNES BECHER and MÉLANIE RACZEK	
On existence of a classical solution to a generalized Kelvin–Voigt model	11
MIROSLAV BULÍČEK, PETR KAPLICKÝ and MARK STEINHAUER	
A lower bound for eigenvalues of the poly-Laplacian with arbitrary order	35
QING-MING CHENG, XUERONG QI and GUOXIN WEI	
Quiver algebras, path coalgebras and coreflexivity	49
SORIN DĂSCĂLESCU, MIODRAG C. IOVANOV and CONSTANTIN NĂSTĂSESCU	
A positive density of fundamental discriminants with large regulator	81
ÉTIENNE FOUVRY and FLORENT JOUVE	
On the isentropic compressible Euler equation with adiabatic index $\gamma = 1$	109
DONG LI, CHANGXING MIAO and XIAOYI ZHANG	
Symmetric regularization, reduction and blow-up of the planar three-body problem	129
RICHARD MOECKEL and RICHARD MONTGOMERY	
Canonical classes and the geography of nonminimal Lefschetz fibrations over S^2	191
YOSHIHISA SATO	
Hilbert–Kunz invariants and Euler characteristic polynomials	227
LARRY SMITH	