Pacific Journal of Mathematics

Volume 343 No. 2 August 2026

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2026 is US $710/year for the electronic version, and $965/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber address should be sent to Mathematical Sciences Publishers, 2000 Allston Way # 59, Berkeley, CA 94701-4004, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 1945-5844 electronic, 0030-8730 printed) at Mathematical Sciences Publishers, 2000 Allston Way # 59, Berkeley, CA 94701-4004, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, 2000 Allston Way # 59, Berkeley, CA 94701-4004.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.261

INFINITELY DIVISIBLE MODIFIED BESSEL DISTRIBUTIONS

ÁRPÁD BARICZ, DHIVYA PRABHU K,
SANJEEV SINGH AND ANTONY VIJESH V

Á. Baricz dedicates this paper to Mourad E. H. Ismail on the occasion of his 80th birthday

We study certain continuous univariate probability distributions supported on $[0, \infty)$ — the McKay distribution and its generalizations, the generalized inverse Gaussian distribution and the K-distribution — , all of which are related to modified Bessel functions of the first and second kinds. In most cases we show that they belong to the class of infinitely divisible distributions, self-decomposable distributions, generalized gamma convolutions and hyperbolically completely monotone densities. Some of the results are known, but new proofs are provided using special functions techniques: Integral representations of quotients of Tricomi hypergeometric functions, Gaussian hypergeometric functions, and modified Bessel functions of the second kind, play an important role in our study. In addition, by using a different approach based on asymptotic properties of modified Bessel functions, we rediscover a Stieltjes transform representation due to Hermann Hankel for the product of modified Bessel functions of the first and second kinds and we deduce a series of new Stieltjes transform representations for products, quotients and their reciprocals concerning modified Bessel functions of the first and second kinds. By using these results we obtain new infinitely divisible modified Bessel distributions with Laplace transforms related to modified Bessel functions of the first and second kind. We show that the new Stieltjes transform representations have some interesting applications and we list some open problems that may be of interest for further research. In addition, we present a new proof, using the Pick function characterization theorem, for the infinite divisibility of the ratio of two gamma random variables and some new Stieltjes transform representations of quotients of Tricomi hypergeometric functions.

MSC2020: 60E07, 44A15, 33C10.
Keywords: modified Bessel functions, Tricomi hypergeometric function, infinitely divisible distribution, self-decomposable distribution, generalized gamma convolution, hyperbolically completely monotone distribution.

1. Introduction

1.1. ***Preliminaries on infinite divisibility.*** A probability distribution is *infinitely divisible* (see Steutel and Van Harn [SH03]) if it can be expressed as the probability distribution of the sum of an arbitrary number of independent and identically distributed random variables. The concept of infinite divisibility of probability distributions was introduced in 1929 by Bruno de Finetti and the most basic results were discovered by Andrey Kolmogorov, Paul Lévy and Aleksandr Khinchin. This type of decomposition of a distribution is used in probability and statistics to find families of probability distributions that might be natural choices for certain models or applications. Infinitely divisible distributions also play an important role in the context of limit theorems. The characteristic function of any infinitely divisible distribution is called an infinitely divisible characteristic function and such a function may be represented, for any value of n, as the n-th power of some other characteristic function. More precisely, a probability distribution ν on the half-line $[0, \infty)$ is infinitely divisible if for any $n \in \mathbb{N}$ there exists a probability distribution ν_n on $(0, \infty)$ such that

$$\int_0^\infty e^{-xt} d\nu = \Big(\int_0^\infty e^{-xt} d\nu_n\Big)^n.$$

A function $f : (0, \infty) \to \mathbb{R}$ is *completely monotone* (or completely monotonic) if it has derivatives of all orders and $(-1)^n f^{(n)}(x) > 0$ for all $x > 0$ and $n \in \{0, 1, 2, \dots\}$. The classes of completely monotonic functions and infinitely divisible distributions are related by the following well-known result (see [Fe66, p. 425]).

Lemma 1. *The function $\omega : (0, \infty) \to (0, \infty)$ is the Laplace transform of an infinitely divisible distribution if and only if $\omega(x) = e^{-\varphi(x)}$, where $\varphi(0^+) = 0$ and $\varphi : (0, \infty) \to (0, \infty)$ is a Bernstein function, that is, φ' is completely monotonic.*

Every continuous-time Lévy process has infinitely divisible distributions, and conversely every infinitely divisible distribution generates uniquely a Lévy process (see for example Steutel and Van Harn [SH03]).

In various real life situations — for instance in biology, physics, economics and actuarial science — certain models postulate a random effect that is the sum of several independent random components with the same distribution. A convenient way to enforce this condition is to suppose the infinite divisibility of the distribution of these random effects.

The concept of *self-decomposability of probability measures* is due to Paul Lévy and goes back to 1937. A random variable X, distributed according to a law, is called *self-decomposable* if for every $c \in (0, 1)$ there exists a random variable X_c, independent of X, such that X and $X_c + cX$ are equal in law. A distribution is self-decomposable if and only if it is a weak limit of partial normed centered sums of a simple sequence of independent random variables. More precisely, a probability distribution is self-decomposable if it is the limit of

$$(X_1 + \cdots + X_n - b_n)/a_n,$$

where the X_i are independent random variables and $\{a_n\}$ and $\{b_n\}$ are sequences of constants with $a_n \to \infty$ and $a_{n+1}/a_n \to 1$. Every self-decomposable distribution is infinitely divisible, and the class of self-decomposable distributions is closed under the convolution and the weak convergence. This class contains stable distributions and generalized gamma convolutions.

The next auxiliary result (see [Fe66, p. 589]) characterizes self-decomposable distributions with support $[0, \infty)$.

Lemma 2. *A random variable X with support* $[0, \infty)$ *having the Laplace transform* $\omega : (0, \infty) \to (0, \infty)$ *is self-decomposable if and only if for every* α, *where* $\alpha \in (0, 1)$, *the function* $\omega(x)/\omega(\alpha x)$ *is a Laplace transform.*

Now, let us focus on two other subclasses of infinitely distributions. A function $f : (0, \infty) \to (0, \infty)$ is *hyperbolically completely monotone* if for each $u > 0$ the function $f(uv) f(u/v)$ is completely monotone as a function of $w = v + 1/v$, where $v > 0$. A distribution is said to be hyperbolically completely monotone if its probability density function is hyperbolically completely monotone. This class of lifetime distributions was discovered by Lennart Bondesson (see [Bo92]) and it is strongly connected to the class of *generalized gamma convolutions*, which was introduced by Olof Thorin [Th77] and constitutes the smallest class of distributions on $(0, \infty)$ that contains all gamma distributions and is closed under convolution and weak convergence. A positive continuous random variable X belongs to the class of generalized gamma convolutions if its Laplace transform is of the form

$$L(s) = \exp\Bigl(-as + \int_0^\infty \ln \frac{t}{s+t}\, d\mu(t)\Bigr),$$

where $s \geq 0$, $a \geq 0$ and $d\mu(t)$ is a nonnegative measure. Since the gamma distribution is infinitely divisible and self-decomposable, so is every generalized gamma convolution. An interesting result of Bondesson (see [Bo92]) states that a probability density that is hyperbolically completely monotone is the density of a generalized gamma convolution.

The next result is a *Pick function characterization theorem* of generalized gamma convolutions.

Lemma 3 [Bo92, Theorem 3.1.2]. *The probability distribution of a continuous random variable X on* $[0, \infty)$ *is a generalized gamma convolution if and only if its moment generating function* ψ *is analytic and zero-free in* $\mathbb{C} \setminus [0, \infty)$, *and satisfies* $\operatorname{Im}\bigl(\psi'(s)/\psi(s)\bigr) \geq 0$ *for* $\operatorname{Im} s > 0$.

Another simple characterization of generalized gamma convolutions is due to Bondesson:

Lemma 4 [Bo92, Theorem 6.1.1]. *A function ϕ on* $[0, \infty)$ *is the Laplace transform of a generalized gamma convolution if and only if* $\phi(0) = 1$ *and ϕ is hyperbolically completely monotone.*

Thorin's class of generalized gamma convolutions is closed with respect to rescaling, weak limits, and addition of independent random variables. Bondesson [Bo15, Theorem 1] has shown that the generalized gamma convolution class also has the remarkable property of being closed with respect to multiplication of independent random variables.

Proving or disproving infinite divisibility of a certain distribution is sometimes a complicated task and it may need a specialized approach; see for example the papers of John Kent [Ke78] and those of Pierre Bosch and Thomas Simon [BS15; BS16]. The Laplace transforms of probability measures are usually transcendental special functions, which has led authors to study the complete monotonicity of various quotients of special functions (such as modified Bessel functions, Tricomi hypergeometric functions, parabolic cylinder functions) and of the logarithmic derivatives of solutions of differential and difference equations; see for example the papers of Philip Hartman [Ha78; Ha79].

The present study is motivated by the *special function technique* approach of Mourad Ismail and coauthors (see [Is77a; Is77b; IK79; IM79; IM82; Is90]); we extend and complement the results from these papers by deducing a series of new Stieltjes transform representations for products and quotients of modified Bessel functions of the first and second kinds and by obtaining new infinitely divisible modified Bessel distributions.

Some of our main results are proved using ideas from the works just cited, but we also work with the theory of generalized gamma convolutions and hyperbolically completely monotone densities.

Outline. In the remainder of this section we recall some basic lemmas on Stieltjes transforms. In Section 2 we present a series of results on infinite divisibility of McKay distributions and its generalizations, the K-distribution and generalized inverse Gaussian distribution. In Section 3 we obtain a series of new Stieltjes transform representations for products and quotients of modified Bessel functions of the first and second kinds and using these results we obtain new infinitely divisible modified Bessel distributions. These new distributions have Laplace transforms related to modified Bessel functions of the first and second kinds. Section 4 is devoted to remarks and open problems concerning some distributions related to modified Bessel functions and Tricomi hypergeometric functions, while Section 5 contains the proofs of all the main results of this paper.

1.2. ***Preliminaries on Stieltjes transforms.*** Before presenting our results on infinitely divisible distributions whose probability density function or Laplace transform involves modified Bessel functions, we recall some basic facts concerning Stieltjes transforms. The first two lemmas of this subsection are variants of the *representation theorem for Stieltjes transforms*. In some applications we use Lemma 5 and in others Lemma 6. The ensuing Lemma 7 is called the *inversion theorem for Stieltjes transforms* or *Perron–Stieltjes inversion formula* and it is also a key ingredient in our proofs.

For proofs of Lemmas 5 and 6 we refer to [HW55] (pp. 235 and 210, respectively), while Lemma 7 can be found in [St32]. (Alternatively, Lemmas 5–7 can be found in [Is77a, Lemma 2.1], [IK79, Theorem 1.2] and [Is77a, Lemma 2.2], respectively.)

Lemma 5 (first representation theorem for Stieltjes transforms). *A complex function $F(z)$ admits a Stieltjes transform representation*

$$F(z)=\int_0^\infty \frac{d\mu(t)}{z+t},\quad \text{with } \int_0^\infty |d\mu(t)|<\infty, \tag{1-1}$$

if and only if the following conditions hold true:

a. *$F(z)$ is analytic for $|\arg z|<\pi$.*
b. *$F(z)=o(1)$ as $|z|\to\infty$ and $F(z)=o(|z|^{-1})$ as $|z|\to 0$, uniformly in every sector $|\arg z|\le \pi-\varepsilon$ for $\varepsilon>0$.*

Lemma 6 (second representation theorem for Stieltjes transforms). *If*

a. *$F(z)$ is analytic for $|\arg z|<\pi/\theta$ for some θ such that $0<\theta<1$, and*
b. *$F(z)=o(1)$ as $z\to\infty$ and $F(z)=o(|z|^{-1})$ as $z\to 0$, uniformly in every sector $|\arg z|\le\pi/\eta$ with $\theta<\eta<1$,*

then the Stieltjes transform representation

$$F(x)=\frac{1}{\pi}\int_0^\infty \frac{dt}{x+t}\,\frac{1}{2\pi\mathrm{i}}\int_C \frac{ze^{\frac{z}{2}}F(te^z)}{z^2+\pi^2}dz \tag{1-2}$$

is valid for all $x>0$, where C is a rectifiable closed curve going around $[-\mathrm{i}\pi,\mathrm{i}\pi]$ in the positive direction and lying in the strip $|\operatorname{Im} z|<\pi/\theta$.

Lemma 7 (inversion theorem for Stieltjes transforms). *If F has the representation* (1-1), *then*

$$\mu(t_2)-\mu(t_1)=\lim_{\eta\to 0^+}\frac{1}{2\pi\mathrm{i}}\int_{t_1}^{t_2}[F(-t-\mathrm{i}\eta)-F(-t+\mathrm{i}\eta)]\,dt, \tag{1-3}$$

where $\mu(t)$ is normalized by $\mu(0)=\mu(0^+)=0$ and $2\mu(t)=\mu(t^+)+\mu(t^-)$ for $t>0$.

2. Distributions whose probability density function involves modified Bessel functions

In this section we consider some known distributions whose probability density function involves the modified Bessel function of the first or second kind and study whether these distributions belong to the class of infinitely divisible distributions or to one of its subclasses: self-decomposable distributions, generalized gamma convolutions and hyperbolically completely monotone densities.

2.1. ***The McKay distribution of type I.*** The McKay distribution of type I involves the modified Bessel function of the first kind, I_μ. Its probability density function is given by

$$\varphi_{\mu,a,b}(x)=\frac{\sqrt{\pi}\,(b^2-a^2)^{\mu+\frac{1}{2}}}{(2a)^\mu\Gamma\big(\mu+\frac{1}{2}\big)}x^\mu e^{-bx}I_\mu(ax), \tag{2-1}$$

where $b>a>0$, $\mu>-\frac{1}{2}$, and its support is $(0,\infty)$. In view of the asymptotic relation

$$I_\mu(x)\sim\frac{x^\mu}{2^\mu\Gamma(\mu+1)}$$

as $x\to 0$ and by using the Legendre duplication formula for the Euler gamma function

$$\Gamma(2x)=\frac{1}{\sqrt{\pi}}2^{2x-1}\Gamma(x)\Gamma\big(x+\tfrac{1}{2}\big)$$

we obtain

$$\lim_{a\to 0}\varphi_{\mu,a,b}(x)=\frac{b^{2\mu+1}}{\Gamma(2\mu+1)}e^{-bx}x^{2\mu},$$

which shows that the McKay distribution of type I for $a\to 0$ reduces to a gamma distribution with shape parameter $2\mu+1$ and inverse scale parameter b (also known as the rate parameter). Since the gamma distribution is infinitely divisible and self-decomposable, and obviously belongs to the class of generalized gamma convolutions, it is natural to ask whether the McKay distribution of type I belongs to those same classes of distributions. It is known (see [Ba10, p. 580]) that $x\mapsto e^{-x}x^\mu I_\mu(x)$ is completely monotonic on $(0,\infty)$ for all $\mu\in\big[-\frac{1}{2},0\big]$, which in turn implies that $x\mapsto e^{-ax}(ax)^\mu I_\mu(ax)$ is completely monotonic on $(0,\infty)$ for all $\mu\in\big[-\frac{1}{2},0\big]$ and $a>0$. On the other hand, the function $x\mapsto e^{(a-b)x}$ is also completely monotonic on $(0,\infty)$ for all $b>a$, and since the product of two completely monotonic functions is also completely monotonic, we conclude:

Fact. *The probability density function $\varphi_{\mu,a,b}$ is completely monotonic on $(0,\infty)$ for all $\mu\in\big(-\frac{1}{2},0\big]$ and $b>a>0$, and according to the Goldie–Steutel law the McKay distribution of type I is infinitely divisible for all $\mu\in\big(-\frac{1}{2},0\big]$ and $b>a>0$.*

With a more sophisticated analysis it is possible to show the next theorem, to the effect that the McKay distribution of type I is infinitely divisible for all $\mu > -\frac{1}{2}$ and $b > a > 0$. The proof may be found in Section 5.

Theorem 1. *If* $\mu > -\frac{1}{2}$ *and* $b > a > 0$, *the McKay distribution, whose probability density function is defined by* (2-1), *belongs to the class of infinitely divisible distributions, self-decomposable distributions and generalized gamma convolutions.*

2.2. ***Another McKay-type distribution.*** Another distribution similar to the McKay distribution of type I also involves the modified Bessel function of the first kind I_μ. Its support is $[0, \infty)$ and its probability density function is given by

$$\psi_{\mu,a,b}(x) = \frac{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{3}{2}}}{2b(2a)^\mu\Gamma\left(\mu+\frac{3}{2}\right)} x^{\mu+1}e^{-bx}I_\mu(ax), \tag{2-2}$$

where $b > a > 0$, $\mu > -1$.

Theorem 2. *If* $\mu > -1$ *and* $b > a > 0$, *the distribution with probability density function* (2-2) *belongs to the class of infinitely divisible distributions.*

2.3. ***Generalization of the McKay distribution of type I.*** We turn to a generalization of McKay-type distributions. Its support is also $(0, \infty)$ and its probability density function is given by

$$\varphi_{\mu,\nu,a,b}(x) = \frac{1}{c_{\mu,\nu,a,b}} x^{\nu-1}e^{-bx}I_\mu(ax), \tag{2-3}$$

where $\mu+1 > 0$, $\mu+\nu > 0$, $b > a > 0$, and

$$c_{\mu,\nu,a,b} = \frac{(a/2)^\mu}{b^{\mu+\nu}}\frac{\Gamma(\mu+\nu)}{\Gamma(\mu+1)} \cdot {}_2F_1\left(\tfrac{\mu+\nu}{2}, \tfrac{\mu+\nu+1}{2}, \mu+1, a^2/b^2\right)$$

(here ${}_2F_1(a, b, c, x)$ is the Gaussian hypergeometric function). Observe that

$$c_{\mu,\mu+1,a,b} = \frac{(2a)^\mu\Gamma\left(\mu+\frac{1}{2}\right)}{\sqrt{\pi}\, b^{2\mu+1}}\, {}_1F_0\left(\mu+\tfrac{1}{2}, a^2/b^2\right) = \frac{(2a)^\mu\Gamma\left(\mu+\frac{1}{2}\right)}{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{1}{2}}}$$

and

$$c_{\mu,\mu+2,a,b} = \frac{2(2a)^\mu\Gamma\left(\mu+\frac{3}{2}\right)}{\sqrt{\pi}\, b^{2\mu+2}}\, {}_1F_0\left(\mu+\tfrac{3}{2}, a^2/b^2\right) = \frac{(2b)(2a)^\mu\Gamma\left(\mu+\frac{3}{2}\right)}{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{3}{2}}},$$

and consequently for all $x > 0$, $\mu > -\frac{1}{2}$ and $b > a > 0$ we have

$$\varphi_{\mu,\mu+1,a,b}(x) = \varphi_{\mu,a,b}(x) = \frac{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{1}{2}}}{(2a)^\mu\Gamma\left(\mu+\frac{1}{2}\right)} x^\mu e^{-bx}I_\mu(ax)$$

and for all $x > 0$, $\mu > -1$ and $b > a > 0$ we have

$$\varphi_{\mu,\mu+2,a,b}(x) = \psi_{\mu,a,b}(x) = \frac{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{3}{2}}}{2b(2a)^{\mu}\Gamma\left(\mu+\frac{3}{2}\right)} x^{\mu+1} e^{-bx} I_{\mu}(ax).$$

From [Ba10, p. 578] we know that the function $x \mapsto e^{-x}x^{-\mu}I_{\mu}(x)$ is completely monotonic on $(0, \infty)$ for all $\mu \geq -\frac{1}{2}$, which in turn implies that $x \mapsto e^{-ax}(ax)^{-\mu}I_{\mu}(ax)$ is completely monotonic on $(0, \infty)$ for all $\mu \geq -\frac{1}{2}$ and $a > 0$. On the other hand, the function $x \mapsto e^{(a-b)x}$ is also completely monotonic on $(0, \infty)$ for all $b > a$. As the product of completely monotonic functions, therefore, the probability density function $\varphi_{\mu,\nu,a,b}$ is completely monotonic on $(0, \infty)$ for all $\mu \geq -\frac{1}{2}$, $\mu + \nu \leq 1$ and $b > a > 0$; thus, according to the Goldie–Steutel law, *the generalization of the McKay distribution of type I is infinitely divisible for such values of* μ, ν, b, a.

Theorem 3. *If* $\mu+\nu > 0$, $\mu+1 \geq \nu$ *and* $b > a > 0$, *the distribution with probability density function defined by* (2-3) *is infinitely divisible.*

With the previous paragraph, this gives two domains in the (μ, ν)-plane where our distribution is infinitely divisible (with $b > a > 0$). Neither domain contains the other. It would of interest to find the largest domain of μ and ν with this property.

2.4. ***One more McKay-type distribution.*** We turn to another distribution similar to the above generalization of the McKay distribution of type I. Its support is also $(0, \infty)$ and its probability density function is given by

$$\xi_{\mu,a,b}(x) = \frac{1}{c_{\mu,a,b}} x^{2\mu} e^{-bx} \left(I_{\mu}(ax)\right)^2, \tag{2-4}$$

where $\mu > -\frac{1}{4}$, $b > 2a > 0$ and

$$c_{\mu,a,b} = \frac{2^{4\mu}a^{2\mu}}{\pi b^{4\mu+1}} \frac{\Gamma\left(\mu+\frac{1}{2}\right)\Gamma\left(2\mu+\frac{1}{2}\right)}{\Gamma(\mu+1)} \cdot {}_2F_1\left(\mu+\tfrac{1}{2}, 2\mu+\tfrac{1}{2}, \mu+1, 4a^2/b^2\right).$$

From [Ba10, p. 580] we know that the function $x \mapsto e^{-x}x^{\mu}I_{\mu}(x)$ is completely monotonic on $(0, \infty)$ for all $\mu \in \left[-\frac{1}{2}, 0\right]$, and thus so are $x \mapsto e^{-ax}(ax)^{\mu}I_{\mu}(ax)$ and $x \mapsto e^{-2ax}(ax)^{2\mu}(I_{\mu}(ax))^2$ for all $a > 0$. Taking the product with $x \mapsto e^{(2a-b)x}$, which is completely monotonic on $(0, \infty)$ for $b > 2a$, we see that $\xi_{\mu,a,b}$ is completely monotonic on $(0, \infty)$ for all $\mu \in \left(-\frac{1}{4}, 0\right]$ and $b > 2a > 0$. Again by Goldie–Steutel, then, *the above McKay type distribution is infinitely divisible for all* $\mu \in \left(-\frac{1}{4}, 0\right]$ *and* $b > 2a > 0$. With a finer analysis the range of μ can be extended:

Theorem 4. *If* $\mu \in \left(-\frac{1}{4}, \frac{1}{2}\right]$ *and* $b > 2a > 0$, *the distribution with probability density function* (2-4) *is infinitely divisible.*

2.5. ***The K-distribution or gamma-gamma distribution.*** Suppose that a random variable X has gamma distribution with mean σ and shape parameter α, where σ is a random variable also having a gamma distribution, with mean μ and shape parameter β, where $\alpha, \beta, \mu > 0$. The result (see [JP78] and [Re99]) is that X has the following probability density function, supported on $(0, \infty)$:

$$\omega_{\alpha,\beta,\mu}(x) = \frac{2}{\Gamma(\alpha)\Gamma(\beta)}(\alpha\beta/\mu)^{\frac{\alpha+\beta}{2}} x^{\frac{\alpha+\beta}{2}-1} K_{\alpha-\beta}(2\sqrt{\alpha\beta x/\mu}). \tag{2-5}$$

The corresponding probability distribution, known as the K-distribution, besides being a compound distribution, it is also a product distribution: namely, the distribution of the product of two independent gamma-distributed random variables, one with mean 1 and shape parameter $\alpha > 0$, the other with mean $\mu > 0$ and shape parameter $\beta > 0$.

Since gamma densities are hyperbolically completely monotone and the product of hyperbolically completely monotone functions belongs to the same class (see [Bo92] or [Bo15, Proposition 4]), we conclude that the K-distribution is a hyperbolically completely monotone distribution. This proves the next theorem, but we will later give an alternative proof of it using special function techniques.

Theorem 5. *If* $\alpha, \beta, \mu > 0$, *the* K*-distribution, with probability density function defined by* (2-5), *belongs to the class of infinitely divisible distributions, self-decomposable distributions, generalized gamma convolutions and hyperbolically completely monotone distributions.*

The K-distribution (also known as the gamma-gamma distribution, since it arises from the product of two independent gamma random variables) is well-known in engineering, having been used for example in modeling land and sea radar clutter, as well as the combined effects of fading and shadowing encountered in mobile communications channels. It has been applied also to optical wireless systems, involving the transmission of optical signals through the atmosphere (see [BSMKR06; CK11]), and in modeling microwave sea echo (see [JP78]).

In [BPV11, Theorem 5] the authors proved that the probability density function of the gamma-gamma distribution $x \mapsto \omega_{\alpha,\beta,\mu}(x)$ is *geometrically concave* on $(0, \infty)$ for all $\alpha, \beta, \mu > 0$, that is, the function $x \mapsto x\omega'_{\alpha,\beta,\mu}(x)/\omega_{\alpha,\beta,\mu}(x)$ is decreasing on $(0, \infty)$ for all $\alpha, \beta, \mu > 0$. It can be shown easily (see for example [Bo92, p. 102]) that this is equivalent to the probability density function of the K-distribution being hyperbolically monotone (of order 1), that is, the function $w \mapsto \omega_{\alpha,\beta,\mu}(uv)\omega_{\alpha,\beta,\mu}(u/v)$ is decreasing on $(0, \infty)$ for all $u, v, \alpha, \beta, \mu > 0$, where $w = v + 1/v$. Clearly Theorem 5 states much more than this: the function $w \mapsto \omega_{\alpha,\beta,\mu}(uv)\omega_{\alpha,\beta,\mu}(u/v)$ is not only decreasing on $(0, \infty)$ for all $u, v, \alpha, \beta, \mu > 0$, it is even completely monotonic.

2.6. ***The generalized inverse Gaussian distribution.*** The generalized inverse Gaussian distribution is a three-parameter family of continuous probability distributions with probability density function

$$\pi_{\mu,a,b}(x) = \frac{(a/b)^{\mu/2}}{2K_\mu(\sqrt{ab})} x^{\mu-1} e^{-\frac{1}{2}(ax+b/x)} \tag{2-6}$$

and support $(0, \infty)$, where $a, b > 0$ and μ is a real parameter. Barndorff-Nielsen and Halgreen [BH77] have proved that the generalized Gaussian distribution is infinitely divisible. Barndorff-Nielsen et al. [BBH78] have shown that the generalized inverse Gaussian distribution is a first hitting time for certain time-homogeneous diffusion processes provided the parameter μ is negative, and in this case infinite divisibility follows from the general central limit theorem. Halgreen [Ha79] and Bondesson [Bo79] have shown that the generalized inverse Gaussian distribution is a generalized gamma convolution, and according to [Bo92, p. 74] the generalized inverse Gaussian distribution belongs to the class of hyperbolically completely monotone densities and hence to the class of generalized gamma convolutions and self-decomposable distributions. Thus the next theorem is previously known, but we will provide an alternative proof, based on the special function approach, for the fact that the generalized inverse Gaussian distribution is a generalized gamma convolution.

Theorem 6. *If $\mu \in \mathbb{R}$ and $a, b > 0$, the generalized inverse Gaussian distribution, with probability density function* (2-6), *belongs to the class of generalized gamma convolutions and hence to the class of self-decomposable distributions and infinitely divisible distributions.*

3. Distributions whose Laplace transform involves modified Bessel functions

At the end of his book [Bo92] Bondesson wrote: "Since the class of infinitely divisible distributions is extremely large, it is not a very interesting class. In fact, as the research during the last two decades has shown, infinitely divisibility seems to be more a rule than an exception. This is not surprising if one considers that the class of (univariate) distributions which are infinitely divisible with respect to the maximum operation contains all distributions. On the other hand, to investigate whether or not infinitely divisibility holds for a particular distribution may lead to a deep insight into the structure of that and other distributions and also to a lot of by-products. (Cf. Riemann's hypothesis in mathematics, for example.) This work would certainly not have been written had not Steutel (1973) asked whether the lognormal distribution is infinitely divisible and had not Thorin (1977) attacked and solved this problem."

In this section our aim is to consider some new lifetime distributions whose Laplace transform contains the modified Bessel function of the first and/or second kind and to study whether these distributions belong to the class of infinitely divisible distributions or to one of its subclasses such as self-decomposable distributions or generalized gamma convolutions. It is an interesting problem to find out whether or not particular distributions are infinitely divisible; in this connection, the first part of this section was motivated by the following open conjecture of Ismail and Miller [IM82, p. 234]:

Conjecture. *If* $\nu > \mu \geq 0$ *and* $b > a > 0$, *is the function*

$$x \mapsto \left(\frac{b}{a}\right)^{\mu-\nu} \frac{I_\mu(a\sqrt{x})I_\nu(b\sqrt{x})}{I_\mu(b\sqrt{x})I_\nu(a\sqrt{x})}$$

the Laplace transform of an infinitely divisible probability distribution?

Theorems 8 and 9 show that it is possible to generate the Laplace transform of an infinitely divisible probability distribution from the above quotient of modified Bessel functions of the first kind, with some slight modifications.

3.1. ***Quotients of modified Bessel functions of the first kind.*** Due to Ismail and Kelker [IK79, Theorem 1.10], we know that the function

$$x \mapsto \rho_{\mu,a}(x) = \frac{1}{2^\mu \Gamma(\mu+1)} \frac{(a\sqrt{x})^\mu}{I_\mu(a\sqrt{x})} \tag{3-1}$$

is the Laplace transform of an infinitely divisible distribution on $[0, \infty)$, when $\mu > 0$ and $a > 0$. The same distribution is also self-decomposable; the proof follows naturally from Lemma 2.

The next result shows that in fact the above function is the Laplace transform of a generalized gamma convolution.

Theorem 7. *If* $\mu > -1$ *and* $a > 0$, *then* $x \mapsto \rho_{\mu,a}(x)$ *is the Laplace transform of a generalized gamma convolution and therefore it is also the Laplace transform of a self-decomposable distribution.*

Theorem 8. *If* $\mu > -1$, $\nu > \sigma > -1$ *and* $b > a > 0$, *then*

$$\Omega_{\mu,\nu,\sigma,a,b}(x) = \left(\frac{b}{a}\right)^{\mu-\nu} \frac{I_\mu(a\sqrt{x})I_\nu(b\sqrt{x})}{I_\mu(b\sqrt{x})I_\nu(a\sqrt{x})} \cdot \rho_{\sigma,b}(x)$$

is the Laplace transform of an infinitely divisible distribution with support $[0, \infty)$.

Theorem 9. *If* $\mu > -1$, $\nu > \frac{1}{2}$ *and* $b > a > 0$, *the function*

$$\Omega_{\mu,\nu,a,b}(x) = \left(\frac{b}{a}\right)^{\mu-\nu} \frac{I_\mu(a\sqrt{x})I_\nu(b\sqrt{x})}{I_\mu(b\sqrt{x})I_\nu(a\sqrt{x})} \cdot e^{-b\sqrt{x}}$$

is the Laplace transform of an infinitely divisible distribution with support $[0, \infty)$.

3.2. ***Stieltjes transform representations and infinite divisibility.*** We now list results related to Stieltjes transforms of modified Bessel functions of the first and second kinds, their products and quotients. We show that these new Stieltjes transform representations are closely related to some new infinitely divisible modified Bessel distributions. The proofs of the results are based on representation and inversion theorems for Stieltjes transforms. Some of the proofs related to infinitely divisibility were inspired from the papers of Mourad Ismail, but in each case we will also show the integral representation of the probability density functions in question.

Theorem 10. *If* $a>0$, $\mu>-\frac{1}{2}$ *and* $|\arg z|<\pi$, *the following Stieltjes transform representation is valid*:

$$e^{-a\sqrt{z}}z^{-\frac{\mu}{2}}I_\mu(a\sqrt{z})=\frac{1}{\pi}\int_0^\infty\frac{t^{-\frac{\mu}{2}}}{z+t}J_\mu(a\sqrt{t})\sin(a\sqrt{t})\,dt.$$

It can be rewritten as the two-fold Laplace transform

$$e^{-a\sqrt{z}}z^{-\frac{\mu}{2}}I_\mu(a\sqrt{z})=\frac{1}{\pi}\int_0^\infty e^{-zs}\left(\int_0^\infty e^{-st}t^{-\frac{\mu}{2}}J_\mu(a\sqrt{t})\sin(a\sqrt{t})\,dt\right)ds.$$

We know from [Is90, Theorem 2] that if $\mu>\frac{1}{2}$, the function

$$x\mapsto 2^\mu\Gamma(\mu+1)x^{-\mu/2}I_\mu(\sqrt{x})e^{-\sqrt{x}}$$

is the Laplace transform of an infinitely divisible distribution that is not a generalized gamma convolution. This implies that the function

$$x\mapsto\frac{2^\mu\Gamma(\mu+1)}{a^\mu}x^{-\frac{\mu}{2}}e^{-a\sqrt{x}}I_\mu(a\sqrt{x})$$

is also the Laplace transform of an infinitely divisible distribution that is not a generalized gamma convolution. In view of Theorem 10, the function

$$s\mapsto\frac{1}{\pi}\frac{2^\mu\Gamma(\mu+1)}{a^\mu}\int_0^\infty e^{-st}t^{-\frac{\mu}{2}}J_\mu(a\sqrt{t})\sin(a\sqrt{t})\,dt$$

is the corresponding probability density function for the above distribution when $\mu>\frac{1}{2}$. This complements [Is90, Theorem 2]. Also, from [EMOT54, eq. (19), p. 226] we know that

$$z^{-\frac{\mu}{2}}e^{-a\sqrt{z}}I_\mu(b\sqrt{z})=\frac{1}{\pi}\int_0^\infty\frac{t^{-\frac{\mu}{2}}}{z+t}J_\mu(b\sqrt{t})\sin(a\sqrt{t})\,dt,$$

where $\operatorname{Re}\mu>-\frac{1}{2}$, $a>b>0$.

Theorem 10 can be extended to the case when $a>0$, $\operatorname{Re}\mu>-\frac{1}{2}$ and $|\arg z|<\pi$, which clearly complements the above formula for the case $a=b$.

The next theorem is a similar result for the product of modified Bessel functions.

Theorem 11. *Let $a \le b$, $\mu, \nu > -1$, $\nu - \mu < 1$ and $|\arg z| < \pi$. Then*

$$z^{\frac{\nu-\mu}{2}} I_\mu(a\sqrt{z})K_\nu(b\sqrt{z}) = \frac{1}{2}\int_0^\infty \frac{t^{\frac{\nu-\mu}{2}}}{z+t} J_\mu(a\sqrt{t})J_\nu(b\sqrt{t})\,dt. \tag{3-2}$$

This Stieltjes transform representation can be rewritten as

$$z^{\frac{\nu-\mu}{2}} I_\mu(a\sqrt{z})K_\nu(b\sqrt{z}) = \frac{1}{2}\int_0^\infty e^{-zs}\Big(\int_0^\infty e^{-st}t^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t})J_\nu(b\sqrt{t})\,dt\Big)\,ds.$$

Theorem 11 also holds for modified Bessel functions whose orders μ and ν are complex numbers: specifically, under the conditions $a \le b$, $\operatorname{Re}\mu > -1$, $\operatorname{Re}\nu > -1$, $\operatorname{Re}(\nu-\mu) < 1$ and $|\arg z| < \pi$.

During our work we found out that the integral representation in (3-2) existed in the literature (see [MOS66, p. 96]) in the form

$$z^{\nu-\mu} I_\mu(az)K_\nu(bz) = \int_0^\infty \frac{t^{\nu-\mu+1}}{z^2+t^2} J_\mu(at)J_\nu(bt)\,dt \tag{3-3}$$

with the better conditions $a \le b$, $\operatorname{Re}\nu > -1$, $\operatorname{Re}(\nu-\mu) < 2$ and $\operatorname{Re} z > 0$. By replacing z by $\sqrt{z}$ and introducing a suitable transformation, equation (3-3) becomes (3-2) with the conditions $a \le b$, $\operatorname{Re}\nu > -1$, $\operatorname{Re}(\nu-\mu) < 2$ and $|\arg z| < \pi$. The representation (3-2) is also given in [EMOT54, eq. (24), p. 227] with conditions $0 < a < b$, $2 + \operatorname{Re}\mu > \operatorname{Re}\nu > -1$. By using the asymptotic relations 10.30.4 and 10.25.3 in [NIST10], we have, as $z \to \infty$,

$$z^{\frac{\nu-\mu}{2}} I_\mu(a\sqrt{z})K_\nu(b\sqrt{z}) \sim z^{\frac{\nu-\mu}{2}} \frac{e^{a\sqrt{z}}}{\sqrt{2\pi a\sqrt{z}}}\sqrt{\frac{\pi}{2b\sqrt{z}}}\,e^{-b\sqrt{z}} = \frac{1}{2\sqrt{ab}}\,e^{(a-b)\sqrt{z}}z^{\frac{\nu-\mu-1}{2}}.$$

The factor $e^{(a-b)\sqrt{z}}$ does not goes to zero as $|z| \to \infty$ uniformly in every sector $|\arg z| \le \pi - \epsilon$, for $\epsilon > 0$. Consequently, condition (b) in Lemma 5 is not valid when $\operatorname{Re}(\nu-\mu) \ge 1$. Thus, when $\operatorname{Re}(\nu-\mu) \ge 1$ the Stieltjes measure presented in [MOS66, p. 96] and [EMOT54, eq. (24), p. 227] is not absolutely integrable.

For integer $\mu = \nu = n$, equation (3-3) appears in [Mi36] with a reference to von Hermann Hankel [Ha75], while a more general form of (3-3) for integer orders has been considered in [Ha75]. Namely, the extension of (3-3) in the case when $k \in \mathbb{N}$ and $n + m + q$ is an even integer has the form

$$-\frac{1}{2\Gamma(k+1)}\Big(\frac{\partial}{\partial r^2}\Big)^k r^q J_n(ar)\big(Y_n(br) - \pi \mathrm{i} J_n(br)\big) = \int_0^\infty t^{q+1}\frac{J_n(at)J_m(bt)}{(t^2-r^2)^{k+1}}dt.$$

The following result provides an immediate application of (3-2).

Theorem 12. *If $\mu > -1$ and $|\arg z| < \pi$, the following representations hold*:

$$2I_\mu(\sqrt{z})K_\mu(\sqrt{z}) = \int_0^\infty \frac{J_\mu^2(\sqrt{t})}{z+t}dt = \int_0^\infty e^{-zs}\Big(\int_0^\infty e^{-st}J_\mu^2(\sqrt{t})\,dt\Big)\,ds. \tag{3-4}$$

and consequently for $\mu > 0$ *the function* $x \mapsto 2\mu I_\mu(\sqrt{x})K_\mu(\sqrt{x})$ *is the Laplace transform of an infinitely divisible probability distribution with support* $[0, \infty)$, *and its corresponding probability density function is*

$$\varsigma_\mu(x) = \mu \int_0^\infty e^{-tx} J_\mu^2(\sqrt{t})\, dt = \frac{\mu}{x} e^{-\frac{1}{2x}} I_\mu\Big(\frac{1}{2x}\Big).$$

A Stieltjes transform can be viewed, at least formally, as a two-fold Laplace transform:

$$\int_0^\infty \frac{d\mu(t)}{z+t} = \int_0^\infty e^{-zs} \int_0^\infty e^{-st} d\mu(t)\, ds.$$

This relation immediately implies the second equation in Theorems 10 and 11, as well as the second equality in (3-4). The formula (3-4) preexisted in the literature as [EMOT54, eq. (22), p. 226]) with condition $\mu \in \mathbb{C}$. But for the case when $\operatorname{Re}\mu \le -1$, $\mu \ne -1, -2, \ldots$, by using the asymptotic relations [NIST10, 10.30.1, 10.30.2 and 10.27.3], we observe that as $z \to 0$

$$I_\mu(\sqrt{z})K_\mu(\sqrt{z}) = I_\mu(\sqrt{z})K_{-\mu}(\sqrt{z}) \sim \frac{\Gamma(-\mu)}{2^{2\mu+1}\Gamma(\mu+1)} z^\mu.$$

It is clear that $|z|\, I_\mu(\sqrt{z})K_\mu(\sqrt{z})$ does not goes to zero as $z \to 0$. This violates one of the conditions in Lemma 5. Thus, when $\operatorname{Re}\mu \le -1$, $\mu \ne -1, -2, \ldots$ the Stieltjes measure presented in [EMOT54, eq. (22), p. 226] is not absolutely integrable.

The next corollary contains some immediate applications of the formula in (3-4) concerning the product of modified Bessel functions of the first and second kinds. The first part of the next corollary is well-known and was proved using nontrivial arguments by Penfold et al. [PVG07] for $\mu \ge 0$, by Baricz [Ba09] for $\mu \ge -\frac{1}{2}$ and by Segura [Se21] for $\mu \ge -1$.

Corollary 1.

a. *The function* $x \mapsto I_\mu(x)K_\mu(x)$ *is decreasing on* $(0, \infty)$ *for all* $\mu > -1$.

b. *The function* $x \mapsto I_\mu(\sqrt{x})K_\mu(\sqrt{x})$ *is completely monotonic on* $(0, \infty)$ *for all* $\mu > -1$.

c. *The function* $x \mapsto x I_\mu(\sqrt{x})K_\mu(\sqrt{x})$ *is a Bernstein function on* $(0, \infty)$ *for all* $\mu > -1$.

d. *The function* $x \mapsto \big(x I_\mu(\sqrt{x})K_\mu(\sqrt{x})\big)^{-1}$ *is completely monotonic on* $(0, \infty)$ *for all* $\mu \ge 0$.

e. *If* $\mu > 0$ *and* $x > 0$, *then* $I_\mu(x)K_\mu(x) \le \dfrac{\pi c_L^2}{\sqrt{3}x^{\frac{2}{3}}}$, *where* $c_L = \sup_{t\in\mathbb{R}_+} \sqrt[3]{t}\, J_0(t) \simeq 0.7857468704\ldots$.

In [BMPS16, Theorem 1] the authors obtained a more general bound for $\mu \ge \nu$

and $x > 0$:

$$I_\mu(x)K_\nu(x) \le \frac{2\pi^{\frac{3}{2}}c_L}{\sqrt{3}\Gamma\left(\frac{2}{3}\right)\Gamma\left(\frac{5}{6}\right)(2x)^{\frac{1}{3}}}.$$

Since

$$c_L < \frac{2^{\frac{2}{3}}\sqrt{\pi}}{\Gamma\left(\frac{2}{3}\right)\Gamma\left(\frac{5}{6}\right)} \simeq 1.8407427466\ldots,$$

the upper bound in Corollary 1**e** is sharper than the above bound from [BMPS16] in the case when $\mu = \nu$.

Another infinitely divisible distribution related to the product of $I_\mu(a\sqrt{x})K_\nu(b\sqrt{x})$ can be found in the next theorem.

Theorem 13. *If* $\mu, \nu > -1$, $\nu - \mu < 1$, $a, b > 0$ *and* $|\arg z| < \pi$, *then*

$$z^{\frac{\nu-\mu}{2}} e^{-a\sqrt{z}} I_\mu(a\sqrt{z})K_\nu(b\sqrt{z}) = \frac{1}{2}\int_0^\infty \frac{t^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t})}{z+t}\big(J_\nu(b\sqrt{t})\cos(a\sqrt{t}) - Y_\nu(b\sqrt{t})\sin(a\sqrt{t})\big)\,dt.$$

This Stieltjes transform representation can be rewritten as

$$\frac{I_\mu(a\sqrt{z})K_\nu(b\sqrt{z})}{z^{\frac{\mu-\nu}{2}}e^{a\sqrt{z}}} = \frac{1}{2}\int_0^\infty e^{-zs}\Big(\int_0^\infty e^{-st}t^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t})\big(J_\nu(b\sqrt{t})\cos(a\sqrt{t}) - Y_\nu(b\sqrt{t})\sin(a\sqrt{t})\big)\,dt\Big)\,ds.$$

The next theorem shows that the function

$$x \mapsto \frac{2^{\mu-\nu+1}\Gamma(\mu+1)b^\nu}{a^\mu\Gamma(\nu)} e^{-a\sqrt{x}} x^{\frac{\nu-\mu}{2}} I_\mu(a\sqrt{x})K_\nu(b\sqrt{x})$$

is the Laplace transform of an infinitely divisible distribution on $(0, \infty)$ with the conditions $a, b, \nu > 0$ and $\mu > \frac{1}{2}$. Based on Theorem 13 we can express the probability density function of the corresponding distribution as

$$s \mapsto \frac{2^{\mu-\nu}\Gamma(\mu+1)b^\nu}{a^\mu\Gamma(\nu)} \times \int_0^\infty e^{-st}t^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t})\big(J_\nu(b\sqrt{t})\cos(a\sqrt{t}) - Y_\nu(b\sqrt{t})\sin(a\sqrt{t})\big)\,dt$$

whenever $a, b, \nu > 0$, $\mu > \frac{1}{2}$ and $\nu - \mu < 1$.

Theorem 14. *If* $a, b, \nu > 0$ *and* $\mu > \frac{1}{2}$, *the function*

$$\chi_{\mu,\nu,a,b}(x) = \frac{2^{\mu-\nu+1}\Gamma(\mu+1)b^\nu}{a^\mu\Gamma(\nu)} e^{-a\sqrt{x}} x^{\frac{\nu-\mu}{2}} I_\mu(a\sqrt{x})K_\nu(b\sqrt{x})$$

is the Laplace transform of an infinitely divisible distribution on $(0, \infty)$.

The next result is also related to an infinitely divisible distribution, but involves only the modified Bessel functions of the second kind.

Theorem 15. *If* $a, b, \mu, \nu > 0$, $\nu + \mu < 1$ *and* $|\arg z| < \pi$, *then*

$$z^{\frac{\mu+\nu}{2}} K_\mu(a\sqrt{z}) K_\nu(b\sqrt{z}) = -\frac{\pi}{4}\int_0^\infty \frac{t^{\frac{\mu+\nu}{2}}}{z+t}\big(J_\mu(a\sqrt{t})Y_\nu(b\sqrt{t}) + J_\nu(b\sqrt{t})Y_\mu(a\sqrt{t})\big)\,dt. \tag{3-5}$$

This representation can be rewritten as

$$z^{\frac{\mu+\nu}{2}} K_\mu(a\sqrt{z}) K_\nu(b\sqrt{z}) = \int_0^\infty e^{-zs}\Big(-\frac{\pi}{4}\int_0^\infty e^{-st} t^{\frac{\mu+\nu}{2}}\big(J_\mu(a\sqrt{t})Y_\nu(b\sqrt{t}) + J_\nu(b\sqrt{t})Y_\mu(a\sqrt{t})\big)\,dt\Big)\,ds.$$

The following result shows that the above product of modified Bessel functions naturally generates a generalized gamma convolution. On the other hand, if we let a, b, μ, ν and s be strictly positive real numbers, and $\nu + \mu < 1$, then the function

$$s \mapsto \frac{-\pi a^\mu b^\nu}{2^{\mu+\nu}\Gamma(\mu)\Gamma(\nu)}\int_0^\infty e^{-st} t^{\frac{\mu+\nu}{2}}\big(J_\mu(a\sqrt{t})Y_\nu(b\sqrt{t}) + J_\nu(b\sqrt{t})Y_\mu(a\sqrt{t})\big)\,dt$$

is in fact a probability density function and

$$\int_0^\infty\Big(\int_0^\infty e^{-st} t^{\frac{\mu+\nu}{2}}\big(J_\mu(a\sqrt{t})Y_\nu(b\sqrt{t}) + J_\nu(b\sqrt{t})Y_\mu(a\sqrt{t})\big)\,dt\Big)\,ds = \frac{-2^{\mu+\nu}\Gamma(\mu)\Gamma(\nu)}{\pi a^\mu b^\nu}.$$

This follows from Theorem 15 and the fact that function in (3-6) is the Laplace transform of an infinitely divisible distribution. In the case when $a = b = 1$, in view of the integral representation

$$J_\mu(x)Y_\nu(x) + J_\nu(x)Y_\mu(x) = -\frac{4}{\pi}\int_0^\infty J_{\mu+\nu}(2x\cosh s)\cosh((\mu-\nu)s)\; Fds$$

(see [EMOT53b, eq. (65), p. 97]), we obtain that (3-5) reduces to

$$z^{\frac{\mu+\nu}{2}} K_\mu(\sqrt{z}) K_\nu(\sqrt{z}) = \int_0^\infty\Big(\int_0^\infty \frac{t^{\frac{\mu+\nu}{2}}}{z+t} J_{\mu+\nu}(2\sqrt{t}\cosh s)\cosh((\mu-\nu)s)\,ds\Big)\,dt,$$

where $\mu, \nu > 0$, $\nu + \mu < 1$ and $|\arg z| < \pi$, as before.

In Theorem 15 it is possible to assume that the orders μ and ν are complex numbers with the conditions such that $\operatorname{Re}\mu > 0$, $\operatorname{Re}\nu > 0$ and $\operatorname{Re}(\nu+\mu) < 1$. The integral representation in (3-5) preexisted in the literature (see [MOS66, p. 96]) in two forms:

$$(-1)^{l+1}\frac{2}{\pi} z^{\mu+\nu+2l} K_\nu(az) K_\mu(bz) = \int_0^\infty \frac{t^{\mu+\nu+2l+1}}{z^2+t^2}\big(J_\nu(at)Y_\mu(bt) + J_\mu(bt)Y_\nu(at)\big)\,dt,$$

where $l \in \{0, \pm 1, \pm 2, \dots\}$, $\mathrm{Re}(\nu+l) > -1$, $\mathrm{Re}(\mu+l) > -1$, $l-1 < \mathrm{Re}(\mu+\nu+2l) < l$, $\mathrm{Re}\, z > 0$, and

$$(-1)^{l+1}\frac{2}{\pi}z^{\mu+\nu+2l-1}K_\nu(az)K_\mu(bz) = \int_0^\infty \frac{t^{\mu+\nu+2l}}{z^2+t^2}\big(J_\nu(at)J_\mu(bt) - Y_\nu(at)Y_\mu(bt)\big)\,dt,$$

where $l \in \{0, \pm 1, \pm 2, \dots\}$, $\mathrm{Re}(\nu+l) > -\frac{1}{2}$, $\mathrm{Re}(\mu+l) > -\frac{1}{2}$, $l-\frac{1}{2} < \mathrm{Re}(\mu+\nu+2l) < l$, $\mathrm{Re}\, z > 0$. Clearly our product representation (3-5) corresponds to case when $z^{\mu+\nu+2l}$ reduces to $z^{\mu+\nu}$ and in this case the conditions will be $\mathrm{Re}\,\nu > -1$, $\mathrm{Re}\,\mu > -1$, $-1 < \mathrm{Re}(\nu+\mu) < 0$ and $\mathrm{Re}\, z > 0$, which complements our conditions $\mathrm{Re}\,\mu > 0$, $\mathrm{Re}\,\nu > 0$, $\mathrm{Re}(\nu+\mu) < 1$ and $|\arg z| < \pi$.

Theorem 16. *If a, b, μ, and ν are strictly positive real numbers, the function*

$$\vartheta_{\mu,\nu,a,b}(x) = \frac{a^\mu b^\nu}{2^{\mu+\nu-2}\Gamma(\mu)\Gamma(\nu)}x^{\frac{\mu+\nu}{2}}K_\mu(a\sqrt{x})K_\nu(b\sqrt{x}) \tag{3-6}$$

is the Laplace transform of an infinitely divisible distribution on $(0,\infty)$ that is self-decomposable and a generalized gamma convolution.

Ismail and Kelker [IK79, Theorem 1.8] proved that if $\mu > \nu > -1$, the function $x \mapsto (\sqrt{x})^{\nu-\mu}K_\nu(\sqrt{x})/K_\mu(\sqrt{x})$ is the Laplace transform of an infinitely divisible distribution with support $(0,\infty)$ and consequently taking into account that $K_\mu(x) \sim 2^{\mu-1}x^{-\mu}\Gamma(\mu)$ as $\mu \to \infty$ for $x > 0$ fixed, the function $x \mapsto (\sqrt{x})^\nu K_\nu(\sqrt{x})/\left[2^{\nu-1}\Gamma(\nu)\right]$ is also the Laplace transform of an infinitely divisible distribution with support $(0,\infty)$ whenever $\nu > -1$. However, after verifying the proof of [IK79, Theorem 1.8] we reached the conclusion that the above results are only true when $\mu > \nu > 0$, and $\nu > 0$, respectively.

The next result is analogous to Theorem 15.

Theorem 17. *If $a, b > 0$, $\mu, \nu > -1$, $\mu+\nu > -1$ and $|\arg z| < \pi$, then*

$$e^{-(a+b)\sqrt{z}}z^{-\frac{\mu+\nu}{2}}I_\mu(a\sqrt{z})I_\nu(b\sqrt{z}) = \frac{1}{\pi}\int_0^\infty \frac{t^{-\frac{\mu+\nu}{2}}}{z+t}\big(J_\mu(a\sqrt{t})J_\nu(b\sqrt{t})\sin((a+b)\sqrt{t})\big)\,dt.$$

This Stieltjes transform representation can be rewritten as

$$e^{-(a+b)\sqrt{z}}z^{-\frac{\mu+\nu}{2}}I_\mu(a\sqrt{z})I_\nu(b\sqrt{z}) = \frac{1}{\pi}\int_0^\infty e^{-zs}\Big(\int_0^\infty e^{-st}t^{-\frac{\mu+\nu}{2}}\big(J_\mu(a\sqrt{t})J_\nu(b\sqrt{t})\sin((a+b)\sqrt{t})\big)\,dt\Big)\,ds.$$

The product of two modified Bessel functions of the first kind also generates an infinitely divisible distribution, which however will not be a generalized gamma convolution:

Theorem 18. *If* $a, b > 0$ *and* $\mu, \nu > \frac{1}{2}$, *the function*

$$\zeta_{\mu,\nu,a,b}(x) = \frac{2^{\mu+\nu}\Gamma(\mu+1)\Gamma(\nu+1)}{a^\mu b^\nu} e^{-(a+b)\sqrt{x}} x^{-\frac{\mu+\nu}{2}} I_\mu(a\sqrt{x}) I_\nu(b\sqrt{x})$$

is a Laplace transform of an infinitely divisible distribution on $(0, \infty)$, *but is not a generalized gamma convolution.*

From Theorems 17 and 18 for $a, b > 0$ and $\mu, \nu > \frac{1}{2}$ we conclude that

$$s \mapsto \frac{2^{\mu+\nu}\Gamma(\mu+1)\Gamma(\nu+1)}{a^\mu b^\nu \pi} \int_0^\infty e^{-st} t^{-\frac{\mu+\nu}{2}} \left(J_\mu(a\sqrt{t}) J_\nu(b\sqrt{t}) \sin(a+b)\sqrt{t}\right) dt$$

is a probability density function and

$$\int_0^\infty \left(\int_0^\infty e^{-st} t^{-\frac{\mu+\nu}{2}} \left(J_\mu(a\sqrt{t}) J_\nu(b\sqrt{t}) \sin(a+b)\sqrt{t}\right) dt\right) ds = \frac{\pi a^\mu b^\nu}{2^{\mu+\nu}\Gamma(\mu+1)\Gamma(\nu+1)}.$$

The next result is the counterpart of Theorem 15.

Theorem 19. *If* $a, b > 0$, $\mu, \nu \in \mathbb{R}$, $\mu + \nu > 1$ *and* $|\arg z| < \pi$, *then*

$$\frac{e^{-(a+b)\sqrt{z}}}{z^{\frac{\mu+\nu}{2}} K_\mu(a\sqrt{z}) K_\nu(b\sqrt{z})} = \frac{4}{\pi^3} \int_0^\infty \frac{t^{-\frac{\mu+\nu}{2}}}{z+t} \Gamma_{\mu,\nu,a,b}(t)\, dt.$$

This representation is equivalent to

$$\frac{e^{-(a+b)\sqrt{z}}}{z^{\frac{\mu+\nu}{2}} K_\mu(a\sqrt{z}) K_\nu(b\sqrt{z})} = \frac{4}{\pi^3} \int_0^\infty e^{-zs} \left(\int_0^\infty e^{-st} t^{-\frac{\mu+\nu}{2}} \Gamma_{\mu,\nu,a,b}(t)\, dt\right) ds,$$

where

$$\Gamma_{\mu,\nu,a,b}(t) = \frac{{}_1T_{\mu,\nu,a,b}(t)\cos((a+b)\sqrt{t}) - {}_2T_{\mu,\nu,a,b}(t)\sin((a+b)\sqrt{t})}{\left(J_\mu^2(a\sqrt{t}) + Y_\mu^2(a\sqrt{t})\right)\left(J_\nu^2(b\sqrt{t}) + Y_\nu^2(b\sqrt{t})\right)}$$

with

$$\begin{aligned} {}_1T_{\mu,\nu,a,b}(t) &= J_\mu(a\sqrt{t}) Y_\nu(b\sqrt{t}) + J_\nu(b\sqrt{t}) Y_\mu(a\sqrt{t}), \\ {}_2T_{\mu,\nu,a,b}(t) &= J_\mu(a\sqrt{t}) J_\nu(b\sqrt{t}) - Y_\mu(a\sqrt{t}) Y_\nu(b\sqrt{t}). \end{aligned}$$

The next theorem shows that the reciprocal of the product of two modified Bessel functions of the second kind also generates an infinitely divisible distribution.

Theorem 20. *If* $a, b > 0$, $\mu, \nu > \frac{1}{2}$, *the function*

$$\kappa_{\mu,\nu,a,b}(x) = \frac{2^{\mu+\nu-2}\Gamma(\mu)\Gamma(\nu)}{a^\mu b^\nu} \cdot \frac{e^{-(a+b)\sqrt{x}}}{x^{\frac{\mu+\nu}{2}} K_\mu(a\sqrt{x}) K_\nu(b\sqrt{x})}$$

is the Laplace transform of an infinitely divisible distribution on $(0, \infty)$, *which is a generalized gamma convolution.*

In view of Theorems 19 and 20, the function

$$s \mapsto \frac{2^{\mu+\nu}\Gamma(\mu)\Gamma(\nu)}{a^\mu b^\nu \pi^3}\int_0^\infty e^{-st} t^{-\frac{\mu+\nu}{2}} \Gamma_{\mu,\nu,a,b}(t)\, dt$$

is a probability density function on $(0,\infty)$ whenever $a, b > 0$, $\mu, \nu > \frac{1}{2}$, and moreover

$$\int_0^\infty \Big(\int_0^\infty e^{-st} t^{-\frac{\mu+\nu}{2}} \Gamma_{\mu,\nu,a,b}(t)\, dt\Big)\, ds = \frac{a^\mu b^\nu \pi^3}{2^{\mu+\nu}\Gamma(\mu)\Gamma(\nu)}.$$

Now, we focus on the quotient of modified Bessel functions of the first and second kinds.

Theorem 21. *If $a, b > 0$, $\mu > -1$, $\mu + \nu > 0$ and $|\arg z| < \pi$, then*

$$\frac{z^{-\frac{\mu+\nu}{2}}}{e^{(a+b)\sqrt{z}}} \frac{I_\mu(a\sqrt{z})}{K_\nu(b\sqrt{z})} = -\frac{2}{\pi^2}\int_0^\infty \frac{t^{-\frac{\mu+\nu}{2}}}{z+t} J_\mu(a\sqrt{t})\gamma_{\nu,a,b}(t)\, dt.$$

This representation can be rewritten as

$$\frac{z^{-\frac{\mu+\nu}{2}}}{e^{(a+b)\sqrt{z}}} \frac{I_\mu(a\sqrt{z})}{K_\nu(b\sqrt{z})} = -\frac{2}{\pi^2}\int_0^\infty e^{-zs}\Big(\int_0^\infty \frac{e^{-st} J_\mu(a\sqrt{t})}{t^{\frac{\mu+\nu}{2}}} \gamma_{\nu,a,b}(t)\, dt\Big)\, ds,$$

where

$$\gamma_{\nu,a,b}(t) = \frac{J_\nu(b\sqrt{t})\cos((a+b)\sqrt{t}) + Y_\nu(b\sqrt{t})\sin((a+b)\sqrt{t})}{J_\nu^2(b\sqrt{t}) + Y_\nu^2(b\sqrt{t})}.$$

In particular, when $a \to 0$, we obtain the following result.

Corollary 2. *If $b > 0$, $\nu > \frac{1}{2}$ and $|\arg z| < \pi$, then*

$$z^{-\frac{\nu}{2}} e^{-b\sqrt{z}} \frac{1}{K_\nu(b\sqrt{z})} = -\frac{2}{\pi^2}\int_0^\infty \frac{t^{-\frac{\nu}{2}}}{z+t}\gamma_{\nu,0,b}(t)\, dt.$$

Finally, we show that the above quotient of modified Bessel functions of the first and second kinds also generates an infinitely divisible distribution.

Theorem 22. *If $a, b > 0$, $\nu, \mu > \frac{1}{2}$, then the function*

$$\varepsilon_{\mu,\nu,a,b}(x) = \frac{2^{\mu+\nu-1}}{a^\mu b^\nu}\Gamma(\nu)\Gamma(\mu+1) e^{-(a+b)\sqrt{x}} x^{-\frac{\mu+\nu}{2}} \frac{I_\mu(a\sqrt{x})}{K_\nu(b\sqrt{x})}$$

is the Laplace transform of an infinitely divisible distribution on $(0,\infty)$. *If $a, b > 0$, $\nu > 0$ and $\mu > -1$, then the reciprocal of $e^{(a+b)\sqrt{x}}\varepsilon_{\mu,\nu,a,b}(x)$, that is,*

$$\frac{e^{-(a+b)\sqrt{x}}}{\varepsilon_{\mu,\nu,a,b}(x)} = \frac{a^\mu b^\nu}{2^{\mu+\nu-1}\Gamma(\nu)\Gamma(\mu+1)} x^{\frac{\mu+\nu}{2}} \frac{K_\nu(b\sqrt{x})}{I_\mu(a\sqrt{x})},$$

is also a Laplace transform of an infinitely divisible distribution on $(0,\infty)$.

If $a, b > 0$ and $\mu, \nu > \frac{1}{2}$, Theorems 21 and 22 imply that the function

$$s \mapsto \frac{-2^{\mu+\nu}\Gamma(\nu)\Gamma(\mu+1)}{a^{\mu}b^{\nu}\pi^2}\int_0^\infty e^{-st}t^{-\frac{\mu+\nu}{2}}J_\mu(a\sqrt{t})\gamma_{\nu,a,b}(t)\,dt$$

is a probability density function and

$$\int_0^\infty\Big(\int_0^\infty e^{-st}t^{-\frac{\mu+\nu}{2}}J_\mu(a\sqrt{t})\gamma_{\nu,a,b}(t)\,dt\Big)\,ds = -\frac{a^{\mu}b^{\nu}\pi^2}{2^{\mu+\nu}\Gamma(\nu)\Gamma(\mu+1)}.$$

4. Remarks, open problems and challenges for future direction

4.1. ***Remarks on quotients of Tricomi hypergeometric functions.*** We now state some facts related to quotients of Tricomi hypergeometric functions. Goovaerts et al. [GHP78] proved that the distribution of the ratio of two independent gamma-distributed random variables is infinitely divisible; this solved a problem posed by Steutel [St73] in a survey of the theory of infinite divisibility. More precisely, based on a relation between the Laplace transform of the density of the quotient of two gamma random variables and the Tricomi confluent hypergeometric function, Goovaerts et al. proved that the distribution of the ratio of two independent gamma-distributed random variables is a generalized gamma convolution. In their proof the logarithmic derivative of the Laplace transform is obtained by the argument principle and contour integration. At the same time, Ismail and Kelker [IK79, Theorem 1.5] proved with the Stieltjes transform technique that the distribution of the two gamma random variables is self-decomposable, hence is infinitely divisible. In this subsection we provide another way to show that the distribution of the ratio of two gamma random variables is a generalized gamma convolution.

Theorem 23. *The distribution of the quotient of two gamma random variables is a generalized gamma convolution.*

Ismail and Kelker's proof of [IK79, Theorem 1.5] was based on the integral representation

$$\frac{\psi(a+1,c+1,z)}{\psi(a,c,z)} = \int_0^\infty \frac{t^{-c}e^{-t}|\psi(a,c,te^{\mathrm{i}\pi})|^{-2}}{(z+t)\Gamma(a+1)\Gamma(a-c+1)}dt \tag{4-1}$$

(see [IK79, p. 885]), where $a > 0$, $c < 1$ and $|\arg z| < \pi$. Our proof of Theorem 23 is also based on (4-1), bur our approach is based on the Pick function characterization theorem, Lemma 3. Moreover, we show that it is possible to obtain similar Stieltjes transform representations for quotients of Tricomi hypergeometric functions. These results complement [IK79, Theorem 1.4].

Theorem 24. *If* $a > 0$, $c < 1$ *and* $|\arg z| < \pi$, *then the following representations* *are valid*:

$$\frac{\psi(a,c-1,z)}{\psi(a,c,z)}=\frac{1-c}{a-c+1}+\int_0^\infty \frac{z}{z+t}\frac{t^{-c}e^{-t}|\psi(a,c,e^{\mathrm{i}\pi}t)|^{-2}}{\Gamma(a)\Gamma(a-c+2)}dt, \tag{4-2}$$

$$\frac{\psi(a+1,c,z)}{\psi(a,c,z)}=\frac{1}{a-c+1}-\int_0^\infty \frac{z}{z+t}\frac{(a-c+1)t^{-c}e^{-t}}{\Gamma(a+1)\Gamma(a-c+2)}|\psi(a,c,e^{\mathrm{i}\pi}t)|^{-2}dt,$$

$$\frac{\psi(a,c+1,z)}{\psi(a,c,z)}=1+\int_0^\infty \frac{t^{-c}e^{-t}}{z+t}\frac{|\psi(a,c,e^{\mathrm{i}\pi}t)|^{-2}}{\Gamma(a)\Gamma(a-c+1)}dt, \tag{4-3}$$

$$\frac{\psi(a-1,c,z)}{\psi(a,c,z)}=z-c+a+\int_0^\infty \frac{z}{z+t}\frac{t^{-c}e^{-t}|\psi(a,c,e^{\mathrm{i}\pi}t)|^{-2}}{\Gamma(a)\Gamma(a-c+1)}dt.$$

This follows naturally from (4-1), but for (4-3) we give a more detailed proof via the Stieltjes representation and inversion theorems. We mention that by using the Stieltjes transform technique Ismail and Kelker [IK79, eq. (1.5)] proved that

$$\frac{\psi(a,c-1,z)}{\psi(a,c,z)}=\int_0^\infty \frac{z}{z+t}\frac{t^{-c}e^{-t}|\psi(a,c,e^{\mathrm{i}\pi}t)|^{-2}}{\Gamma(a)\Gamma(a-c+2)}dt$$

for $a>0,\ 1<c<a+1$ and $|\arg z|<\pi$, and our result (4-2) complements this naturally.

Bondesson [Bo92, Example 4.3.1] remarked that the distribution of the quotient of two gamma random variables belongs also to the class of hyperbolically completely monotone densities. More precisely, if X and Y are independent gamma distributed random variables with parameters (α,β) and (α_0,β_0), then the probability density function of their quotient $Z=X/Y$ is given by

$$f(x)=\frac{\Gamma(\alpha+\alpha_0)}{\Gamma(\alpha)\Gamma(\alpha_0)}\Big(\frac{\beta_0}{\beta}\Big)^{\alpha}x^{\alpha-1}\Big(1+\frac{\beta_0}{\beta}x\Big)^{-(\alpha+\alpha_0)},$$

where $x>0$; see [IK79, p. 889]. Now, if $u,v>0$ and $w=v+1/v$, we have, by [Bo92, Example 4.3.1],

$$\begin{aligned}f(uv)f\Big(\frac{u}{v}\Big)&=\bigg(\frac{\Gamma(\alpha+\alpha_0)}{\Gamma(\alpha)\Gamma(\alpha_0)}\Big(\frac{\beta_0}{\beta}\Big)^{\alpha}\bigg)^2u^{2\alpha-2}\Big((1+\frac{\beta_0}{\beta}uv)\Big(1+\frac{\beta_0}{\beta}\frac{u}{v}\Big)\Big)^{-(\alpha+\alpha_0)}\\&=\bigg(\frac{\Gamma(\alpha+\alpha_0)}{\Gamma(\alpha)\Gamma(\alpha_0)}\Big(\frac{\beta_0}{\beta}\Big)^{\alpha}\bigg)^2u^{2\alpha-2}\Big(1+\Big(\frac{u\beta_0}{\beta}\Big)^2+\frac{\beta_0}{\beta}uw\Big)^{-(\alpha+\alpha_0)},\end{aligned}$$

and thus $w\mapsto f(uv)f(u/v)$ is completely monotonic on $(0,\infty)$ for all $\alpha,\alpha_0,\beta,\beta_0>0$.

As discussed in [AB23], the integral representation (4-1) was important in the study of the infinite divisibility of the Whittaker distribution. It is possible to obtain similar results on ratios of Tricomi hypergeometric functions where the difference between the parameters is not necessarily an integer; see [FS23, Section 2] for details.

4.2. ***The noncentral chi square distribution.*** The noncentral chi square distribution is a noncentral generalization of the chi-squared distribution and has the probability density function

$$\chi_{\mu,\lambda}(x) = \tfrac{1}{2}e^{-\frac{(x+\lambda)}{2}}(x/\lambda)^{\frac{\mu}{4}-\frac{1}{2}}I_{\frac{\mu}{2}-1}(\sqrt{\lambda x}), ,$$

where $\lambda > 0$ is the noncentral parameter and $\mu > 0$ is the degree of freedom. From [IK79, Theorem 1.6] we know that the noncentral chi square distribution is infinitely divisible for all degrees of freedom, including fractional ones. From [Bo92, Example 9.2.2] we also know that the noncentral chi square distribution belongs to the so-called class of generalized convolutions of mixtures of exponential distributions, introduced in [Bo81, p. 43], and which is in fact the smallest class of distributions that is closed under convolution and weak convergence and contains all mixtures of exponential distributions. We know that the class of generalized convolutions of mixtures of exponential distributions is a subclass of the infinitely divisible distributions, but contains all generalized gamma convolutions and hence hyperbolically completely monotone densities. Thus, it is very natural to ask whether the noncentral chi square distribution belongs to the class of generalized gamma convolutions or to the class of hyperbolically completely monotone densities. The next result suggests that under some conditions on the parameters μ and λ the noncentral chi square distribution belongs to the class of hyperbolically completely monotone densities, and hence to the class of generalized gamma convolutions. The ***first open problem*** is to find the optimal range for the parameters μ and λ such that the noncentral chi square distribution belongs to the class of hyperbolically completely monotone densities.

Theorem 25. *Let $\mu > 1$, $u, v > 0$ and $w = v + 1/v$. If $\lambda \leq \mu$, the function $w \mapsto \chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v)$ is strictly decreasing on $(2, \infty)$ and if $2\lambda \leq \mu$, then $w \mapsto \chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v)$ is strictly convex on $(2, \infty)$.*

4.3. ***Hyperbolically complete monotonicity of McKay distributions.*** In Theorems 1–4 of Section 2 we studied the infinite divisibility and self-decomposability of McKay-type distributions and checked their membership in the class of generalized gamma convolutions. We were not able to check whether these distributions belong to the class of hyperbolically completely monotone densities. (In Theorem 5, Macdonald's integral representation of the K-distribution was crucial; but to our knowledge there is no similar result for modified Bessel functions of the first kind.)

A function $f : (0, \infty) \to \mathbb{R}$ is *absolutely monotone* (or absolutely monotonic) if it has derivatives of all orders and $f^{(n)}(x) > 0$ for all $x > 0$ and $n \in \{0, 1, 2, \dots\}$. The next result shows that $w \mapsto I_\mu(uv)I_\mu(u/v)$ is absolutely monotonic on $(2, \infty)$; thus a more sophisticated analysis is needed to verify whether McKay type distributions belong to the class of hyperbolically completely monotone densities.

Theorem 26. *If* $\mu > -\frac{1}{2}$, $u, v > 0$ *and* $w = v + 1/v$, *then the function* $w \to I_\mu(uv)I_\mu(u/v)$ *is absolutely monotonic on* $(2, \infty)$.

The ***second open problem*** is to verify under what conditions the distributions considered in Theorems 1–4 belong to the class of hyperbolically completely monotone densities, and under what conditions the distributions in Theorems 2–4 belong to the class of self-decomposable distributions as well as of generalized gamma convolutions. One can also ask under what conditions the probability density function in Theorem 12 is hyperbolically completely monotonic.

4.4. ***Self-decomposability of modified Bessel distributions.*** Let $w(x)$ be the Laplace transform of a probability distribution on $[0, \infty)$. It is known that if $-dw(x)/dx$ is the Stieltjes transform of a positive measure, then the original distribution is self-decomposable, and hence is infinitely divisible. In other words, a nonnegative random variable is self-decomposable if its Laplace transform satisfies

$$\int_0^\infty e^{-xt}\,d\varpi(t) = e^{-h(x)}, \quad h'(x) = \int_0^\infty \frac{d\varpi(t)}{x+t}, \quad d\varpi(t) \ge 0. \tag{4-4}$$

By Lemma 1, a probability measure $d\omega$ supported on $[0, \infty)$ is infinitely divisible if and only if its Laplace transform satisfies

$$\int_0^\infty e^{-xt}\,d\omega(t) = e^{-h(x)}, \quad h(0) = 0, \quad \text{and } h'(x) \text{ is completely monotonic.} \tag{4-5}$$

Recall that self-decomposable functions are infinitely divisible and a probability distribution satisfying (4-4) and (4-5) is called a generalized gamma convolution. In Theorems 14, 18 and 22 we have infinitely divisible modified Bessel distributions whose Laplace transform can be written as Stieltjes transforms, but these Stieltjes transforms do not have positive kernels. Thus, the ***third open problem*** is to verify under which conditions the distributions considered in Theorems 14, 18 and 22 are self-decomposable.

5. Proofs of the main results

Proof of Theorem 1. From Prudnikov et al. [PBM88, eq. 2.15.3.2], we have

$$\int_0^\infty x^\mu e^{-bx} I_\mu(ax)\,dx = \frac{(2a)^\mu \Gamma\big(\mu + \frac{1}{2}\big)}{\sqrt{\pi}(b^2 - a^2)^{\mu+\frac{1}{2}}},$$

which in turn implies that the Laplace transform

$$L[\varphi_{\mu,a,b}(x)] = \int_0^\infty e^{-xt} \varphi_{\mu,a,b}(t)\,dt$$

of the probability density function $\varphi_{\mu,a,b}$ is given by

$$L[\varphi_{\mu,a,b}(x)]=\frac{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{1}{2}}}{(2a)^{\mu}\Gamma\big(\mu+\frac{1}{2}\big)}\int_0^\infty t^{\mu}e^{-(b+x)t}I_{\mu}(at)\,dt=\left(\frac{b^2-a^2}{(x+b)^2-a^2}\right)^{\mu+\frac{1}{2}}.$$

First we show that the McKay distribution with probability density function (2-1) is infinitely divisible. In view of Lemma 1, this is equivalent to

$$x\mapsto-\frac{d\ln L[\varphi_{\mu,a,b}(x)]}{dx}=\big(\mu+\tfrac{1}{2}\big)\Big(\frac{1}{x+b-a}+\frac{1}{x+b+a}\Big)$$

being complete monotone on $(0,\infty)$ when $\mu>-\frac{1}{2}$ and $b>a>0$, which is true.

We show that the same distribution is self-decomposable. Write $\omega_{\mu,a,b}(x)=L[\varphi_{\mu,a,b}(x)]/L[\varphi_{\mu,a,b}(\alpha x)]$. By Lemma 2, self-decomposability is equivalent to the complete monotonicity on $(0,\infty)$ of

$$x\mapsto-\frac{d\ln\omega_{\mu,a,b}(x)}{dx}=\big(\mu+\tfrac{1}{2}\big)\left(\frac{(1-\alpha)(b-a)}{(x+b-a)(\alpha x+b-a)}+\frac{(1-\alpha)(b+a)}{(x+b+a)(\alpha x+b+a)}\right)$$

when $\alpha\in(0,1)$, $\mu>-\frac{1}{2}$ and $b>a>0$, which is also true.

We show that our distribution belongs to the class of generalized gamma convolutions. The moment generating function of the distribution is

$$\psi(z):=M_X(z)=\int_0^\infty e^{tz}\varphi_{\mu,a,b}(t)\,dt=L[\varphi_{\mu,a,b}(-z)]=\left(\frac{b^2-a^2}{(z-b)^2-a^2}\right)^{\mu+\frac{1}{2}}.$$

This function is analytic and zero-free in $\mathbb{C}\setminus[0,\infty)$, so to apply Lemma 3 we just need to verify that $\operatorname{Im}\big(\psi'(s)/\psi(s)\big)\geq 0$ for $\operatorname{Im}s>0$. We have

$$\frac{\psi'(s)}{\psi(s)}=-\big(\mu+\tfrac{1}{2}\big)\Big(\frac{1}{s+a-b}+\frac{1}{s-a-b}\Big),$$

which implies that for $s=x+\mathrm{i}y$ and $y=\operatorname{Im}s>0$ we have, as needed,

$$\operatorname{Im}\frac{\psi'(s)}{\psi(s)}=\big(\mu+\tfrac{1}{2}\big)\left(\frac{y}{(x+a-b)^2+y^2}+\frac{y}{(x-a-b)^2+y^2}\right)>0.$$

(Another proof that the distribution belongs to the class of generalized gamma convolutions uses Lemma 4: we just observe that if $\phi(x)=L[\varphi_{\mu,a,b}(x)]$, then $\phi(uv)\phi(u/v)$ can be written as

$$\phi(uv)\phi(u/v)=\frac{(b^2-a^2)^{\mu+\frac{1}{2}}u^{-2\mu-1}}{\left(w+\dfrac{u^2+(b-a)^2}{(b-a)u}\right)^{\mu+\frac{1}{2}}\left(w+\dfrac{u^2+(b+a)^2}{(b+a)u}\right)^{\mu+\frac{1}{2}}},$$

which is completely monotonic in $w=v+1/v>0$ for all $\mu>-\frac{1}{2}$ and $b>a>0$.) □

Proof of Theorem 2. By [PBM88, equation 2.15.3.2] we have

$$\int_0^\infty x^{\mu+1} e^{-bx} I_\mu(ax)\,dx = \frac{2b(2a)^\mu \Gamma\big(\mu+\frac{3}{2}\big)}{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{3}{2}}},$$

which implies that the Laplace transform $L[\psi_{\mu,a,b}(x)] = \int_0^\infty e^{-xt}\psi_{\mu,a,b}(t)\,dt$ of the probability density function $\psi_{\mu,a,b}$ is given by

$$\frac{\sqrt{\pi}(b^2-a^2)^{\mu+\frac{3}{2}}}{2b(2a)^\mu \Gamma\big(\mu+\frac{3}{2}\big)} \int_0^\infty t^{\mu+1} e^{-(b+x)t} I_\mu(at)\,dt = \Big(1+\frac{x}{b}\Big)\Big(\frac{b^2-a^2}{(x+b)^2-a^2}\Big)^{\mu+\frac{3}{2}}.$$

By Lemma 1, showing that the distribution with probability density function (2-2) is infinitely divisible is equivalent to showing that

$$x \mapsto -\frac{d\ln L[\psi_{\mu,a,b}(x)]}{dx} = \frac{\mu+1}{x+b-a} + \frac{\mu+1}{x+b+a} + \frac{a^2}{(x+b)(x+b-a)(x+b+a)}$$

is completely monotone on $(0,\infty)$ when $\mu > -1$ and $b > a > 0$, which is true. □

Proof of Theorem 3. Equation 2.15.3.2 of [PBM88] reads $\int_0^\infty x^{\nu-1} e^{-bx} I_\mu(ax)\,dx = c_{\mu,\nu,a,b}$, so the Laplace transform

$$L[\varphi_{\mu,\nu,a,b}(x)] = \int_0^\infty e^{-xt}\varphi_{\mu,\nu,a,b}(t)\,dt$$

of the probability density function $\varphi_{\mu,\nu,a,b}$ is given by

$$L[\varphi_{\mu,\nu,a,b}(x)] = \Big(\frac{b}{x+b}\Big)^{\mu+\nu} \cdot \frac{{}_2F_1\Big(\frac{\mu+\nu}{2}, \frac{\mu+\nu+1}{2}, \mu+1, \frac{a^2}{(x+b)^2}\Big)}{{}_2F_1\Big(\frac{\mu+\nu}{2}, \frac{\mu+\nu+1}{2}, \mu+1, \frac{a^2}{b^2}\Big)}.$$

Recall that the Gaussian hypergeometric function satisfies

$$\frac{d}{dx}\left({}_2F_1(a,b,c,x)\right) = \frac{ab}{c}\cdot {}_2F_1(a+1,b+1,c+1,x)$$

and for a, b and c such that $-1 \le a \le c$ and $0 < b \le c$ we have the Küstner integral representation [Kü02]

$$\frac{z\cdot {}_2F_1(a+1,b+1,c+1,z)}{{}_2F_1(a,b,c,z)} = \int_0^1 \frac{z}{1-tz} dq_{a,b,c}(t), \tag{5-1}$$

where $z \in \mathbb{C}\setminus[1,\infty)$, $q_{a,b,c}(1) - q_{a,b,c}(0) = 1$ and $q_{a,b,c}$ is a nondecreasing self-mapping of $[0, 1]$. To prove infinite divisibility, in view of Lemma 1, we just need to observe that the function

$$x \mapsto -\frac{d\ln L[\varphi_{\mu,\nu,a,b}(x)]}{dx}$$

$$= \frac{\mu+\nu}{x+b} + \frac{a^2(\mu+\nu)(\mu+\nu+1)}{2(\mu+1)(x+b)^3} \cdot \frac{{}_2F_1\left(\frac{\mu+\nu+2}{2}, \frac{\mu+\nu+3}{2}, \mu+2, \frac{a^2}{(x+b)^2}\right)}{{}_2F_1\left(\frac{\mu+\nu}{2}, \frac{\mu+\nu+1}{2}, \mu+1, \frac{a^2}{(x+b)^2}\right)}$$
$$= \frac{\mu+\nu}{x+b}\left(1 + \frac{a^2(\mu+\nu+1)}{2(\mu+1)} \int_0^1 \frac{dq_{\mu,\nu}(t)}{(x+b-a\sqrt{t})(x+b+a\sqrt{t})}\right)$$

is completely monotonic on $(0, \infty)$ for all $\mu+\nu>0$, $\nu \le \mu+1$ and $b>a>0$, where in view of (5-1) we have $q_{\mu,\nu}(1) - q_{\mu,\nu}(0) = 1$ and $q_{\mu,\nu}$ is a nondecreasing self-mapping of [0, 1]. □

Proof of Theorem 4. Equation 2.15.20.5 of [PBM88] reads $\int_0^\infty x^{2\mu}e^{-bx}(I_\mu(ax))^2dx = c_{\mu,a,b}$, so the Laplace transform

$$L[\xi_{\mu,a,b}(x)] = \int_0^\infty e^{-xt}\xi_{\mu,a,b}(t)\,dt$$

of the probability density function $\xi_{\mu,a,b}$ is given by

$$L[\xi_{\mu,a,b}(x)] = \left(\frac{b}{x+b}\right)^{4\mu+1} \cdot \frac{{}_2F_1\left(\mu+\frac{1}{2}, 2\mu+\frac{1}{2}, \mu+1, 4a^2/(x+b)^2\right)}{{}_2F_1\left(\mu+\frac{1}{2}, 2\mu+\frac{1}{2}, \mu+1, 4a^2/b^2\right)}.$$

To prove infinite divisibility, in view of Lemma 1, we just need to observe that

$$x \mapsto -\frac{d\ln L[\xi_{\mu,a,b}(x)]}{dx}$$
$$= \frac{4\mu+1}{x+b} + \frac{8a^2(\mu+\frac{1}{2})(2\mu+\frac{1}{2})}{(\mu+1)(x+b)^3} \frac{{}_2F_1\left(\mu+\frac{3}{2}, 2\mu+\frac{3}{2}, \mu+2, 4a^2/(x+b)^2\right)}{{}_2F_1\left(\mu+\frac{1}{2}, 2\mu+\frac{1}{2}, \mu+1, 4a^2/(x+b)^2\right)}$$
$$= \frac{4\mu+1}{x+b}\left(1 + \frac{2a^2(2\mu+1)}{\mu+1} \int_0^1 \frac{dq_\mu(t)}{(x+b-2a\sqrt{t})(x+b+2a\sqrt{t})}\right)$$

is completely monotonic on $(0, \infty)$ for all $-1 \le \mu+\frac{1}{2} \le \mu+1$, $0 < 2\mu+\frac{1}{2} \le \mu+1$ and $b > 2a > 0$, where in view of (5-1) we have $q_\mu(1) - q_\mu(0) = 1$ and q_μ is a nondecreasing self-mapping of [0, 1]. □

Proof of Theorem 5. We will consider only the case when $\alpha \ne \beta$ for the proof of the infinite divisibility, self-decomposability as well as when we prove that the K-distribution belongs to the class of generalized gamma convolutions. However, in the case of hyperbolically complete monotonicity we will consider also the case when $\alpha = \beta$.

By using Lemma 1, first we show that the K-distribution is infinitely divisible. Equation 2.16.8.4 of [PBM88] reads

$$\int_0^\infty t^{q-1/2}e^{-rt}K_{2\nu}(2s\sqrt{t})\,dt = \frac{\Gamma\left(q+\nu+\frac{1}{2}\right)\Gamma\left(q-\nu+\frac{1}{2}\right)}{2sr^qe^{-s^2/2r}} W_{-q,\nu}\left(\frac{s^2}{r}\right),$$

where $W_{\kappa,\mu}$ stands for the Whittaker function of the second kind. Hence

$$L[\omega_{\alpha,\beta,\mu}(x)] = \int_0^\infty e^{-xt}\omega_{\alpha,\beta,\mu}(t)\,dt = \left(\frac{\alpha\beta}{\mu x}\right)^{\frac{\alpha+\beta-1}{2}} e^{\frac{\alpha\beta}{2\mu x}} W_{-\frac{\alpha+\beta-1}{2},\frac{\alpha-\beta}{2}}\left(\frac{\alpha\beta}{\mu x}\right).$$

$W_{\kappa,\mu}$ is related to the Tricomi hypergeometric function by

$$W_{\kappa,\mu}(x) = e^{-\frac{x}{2}}x^{\mu+\frac{1}{2}}\cdot\psi\left(\mu-\kappa+\tfrac{1}{2},\,1+2\mu,\,x\right),$$

so the Laplace transform of the K-distribution is

$$L[\omega_{\alpha,\beta,\mu}(x)] = \left(\frac{\alpha\beta}{\mu x}\right)^{\alpha}\psi\left(\alpha,\,1+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right).$$

Now, recall the recurrence relation [NIST10, eq. 13.3.22]

$$\psi'(a,c,x) = -a\psi(a+1,c+1,x)$$

and the integral representation (4-1), which is valid for $|\arg z| < \pi$, $a > 0$ and $c < 1$. In view of Lemma 1 and the above relations, to show that the K-distribution is infinitely divisible we just need to show that for $\theta(x) = L[\omega_{\alpha,\beta,\mu}(x)]$ we have

$$-\frac{d}{dx}\ln\theta(x) = \frac{\alpha}{x}+\frac{\alpha\beta}{\mu x^2}\frac{\psi'\left(\alpha,\,1+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right)}{\psi\left(\alpha,\,1+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right)} = \frac{\alpha}{x}-\frac{\alpha^2\beta}{\mu x^2}\frac{\psi\left(\alpha+1,\,2+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right)}{\psi\left(\alpha,\,1+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right)},$$

that is,

$$-\frac{d}{dx}\ln\theta(x) = \frac{\alpha}{x}\left(1-\frac{\alpha\beta}{\mu x}\int_0^\infty \frac{\omega_{\alpha,\beta}(t)}{\frac{\alpha\beta}{\mu x}+t}\,dt\right),$$

where

$$\omega_{\alpha,\beta}(t) = \frac{t^{\beta-\alpha-1}e^{-t}}{\Gamma(\alpha+1)\Gamma(\beta)}\left|\psi(\alpha,\,1+\alpha-\beta,\,te^{\mathrm{i}\pi})\right|^{-2},$$

is completely monotonic on $(0,\infty)$. To show this, observe that if in

$$\text{(5-2)}\qquad \frac{\alpha\beta}{\mu x}\frac{\psi\left(\alpha+1,\,2+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right)}{\psi\left(\alpha,\,1+\alpha-\beta,\,\frac{\alpha\beta}{\mu x}\right)} = \int_0^\infty \frac{\frac{\alpha\beta}{\mu x}\omega_{\alpha,\beta}(t)}{\frac{\alpha\beta}{\mu x}+t}\,dt$$

we make the change of variable $\frac{\alpha\beta}{\mu x} = s$ and take s to infinity, then in view of the asymptotic expansion [NIST10, eq. 13.7.3]

$$\psi(a,c,x) \sim x^{-a}\left(1+a(c-a-1)\frac{1}{x}+\frac{1}{2}a(a+1)(a+1-c)(a+2-c)\frac{1}{x^2}+\dots\right),$$

which is valid for large real x and fixed a and c, we obtain

$$\text{(5-3)}\qquad \int_0^\infty \omega_{\alpha,\beta}(t)\,dt = 1.$$

This implies that

$$-\frac{d}{dx}\ln\theta(x)=\frac{\alpha}{x}\Big(\int_0^\infty \omega_{\alpha,\beta}(t)\,dt-\frac{\alpha\beta}{\mu x}\int_0^\infty \frac{\omega_{\alpha,\beta}(t)}{\frac{\alpha\beta}{\mu x}+t}dt\Big)=\int_0^\infty \frac{\alpha\omega_{\alpha,\beta}(t)}{x+\frac{\alpha\beta}{\mu t}}dt,$$

and this is completely monotonic in x on $(0,\infty)$ for all $\alpha,\beta,\mu>0$ such that $\alpha<\beta$. All we need is to observe that the Whittaker function of the first kind is symmetric in the second parameter, that is, $W_{\kappa,\mu}=W_{\kappa,-\mu}$, and this shows that for every $\alpha,\beta,\mu>0$ the Laplace transform of the K-distribution $L[\omega_{\alpha,\beta,\mu}(x)]$ is the Laplace transform of an infinitely divisible distribution. Another way to show the complete monotonicity of $-d\ln\theta(x)/dx$ is to use the Kummer transformation [NIST10, eq. 13.2.40]

$$\psi(a,c,x)=x^{1-c}\psi(a-c+1,2-c,x)$$

and then apply the previous approach to the Laplace transform

$$L[\omega_{\alpha,\beta,\mu}(x)]=\Big(\frac{\alpha\beta}{\mu x}\Big)^{\beta}\psi\Big(\beta,1-\alpha+\beta,\frac{\alpha\beta}{\mu x}\Big), \tag{5-4}$$

to check that it is the Laplace transform of an infinite divisible distribution for all $\alpha,\beta,\mu>0$ such that $\beta<\alpha$.

Now set $\theta(x)=L[\omega_{\alpha,\beta,\mu}(x)]$. By Lemma 2, the K-distribution will be self-decomposable if, for every $a\in(0,1)$, the function $\eta(x)=\theta(x)/\theta(ax)$ is a Laplace transform. Lemma 1 reduces the proof of the last condition to checking that $-d\ln\eta(x)/dx$ is completely monotonic on $(0,\infty)$. Thus, by (5-2) we have

$$\begin{aligned}-\frac{d}{dx}[\ln\eta(x)]&=-\frac{\alpha\beta}{a\mu x^2}\frac{\psi'\big(\alpha,1+\alpha-\beta,\frac{\alpha\beta}{a\mu x}\big)}{\psi\big(\alpha,1+\alpha-\beta,\frac{\alpha\beta}{a\mu x}\big)}+\frac{\alpha\beta}{\mu x^2}\frac{\psi'\big(\alpha,1+\alpha-\beta,\frac{\alpha\beta}{\mu x}\big)}{\psi\big(\alpha,1+\alpha-\beta,\frac{\alpha\beta}{\mu x}\big)}\\
&=\frac{\alpha^2\beta}{a\mu x^2}\frac{\psi\big(\alpha+1,2+\alpha-\beta,\frac{\alpha\beta}{a\mu x}\big)}{\psi\big(\alpha,1+\alpha-\beta,\frac{\alpha\beta}{a\mu x}\big)}-\frac{\alpha^2\beta}{\mu x^2}\frac{\psi\big(\alpha+1,2+\alpha-\beta,\frac{\alpha\beta}{\mu x}\big)}{\psi\big(\alpha,1+\alpha-\beta,\frac{\alpha\beta}{\mu x}\big)}\\
&=\frac{\alpha}{x}\Bigg(\int_0^\infty\frac{\frac{\alpha\beta}{a\mu x}\omega_{\alpha,\beta}(t)}{\frac{\alpha\beta}{a\mu x}+t}\,dt-\int_0^\infty\frac{\frac{\alpha\beta}{\mu x}\omega_{\alpha,\beta}(t)}{\frac{\alpha\beta}{\mu x}+t}\,dt\Bigg)\\
&=\int_0^\infty\frac{\alpha^2\beta(1-a)\omega_{\alpha,\beta}(t)}{a\mu t\big(x+\frac{\alpha\beta}{t\mu}\big)\big(x+\frac{\alpha\beta}{at\mu}\big)}\,dt,\end{aligned}$$

which is completely monotonic as a function of x on $(0,\infty)$, $a\in(0,1)$, for all $\alpha,\beta,\mu>0$ and $\alpha<\beta$. Thus the K-distribution is self-decomposable if $\alpha<\beta$. The case $\beta<\alpha$ is handled by repeating the process for the other version of the Laplace transform, (5-4).

Lemma 3 is used to prove that the K-distribution belongs to the class of generalized gamma convolutions. The moment generating function

$$M_X(s)=L[\omega_{\alpha,\beta,\mu}(-s)]=\Big(-\frac{\alpha\beta}{\mu s}\Big)^{\alpha}\psi\Big(\alpha,\,1+\alpha-\beta,\,-\frac{\alpha\beta}{\mu s}\Big)$$

of the K-distribution is analytic and zero-free in $\mathbb{C}\setminus[0,\infty)$; we show that it satisfies $\operatorname{Im}\big(\psi'(s)/\psi(s)\big)\ge 0$ for $\operatorname{Im} s>0$. In view of (5-2) and (5-3) for $\psi(s)=M_X(s)$ and $\alpha<\beta$ we obtain

$$\frac{\psi'(s)}{\psi(s)}=-\frac{\alpha}{s}-\frac{\alpha^2\beta}{\mu s^2}\,\frac{\psi\big(\alpha+1,\,2+\alpha-\beta,\,-\frac{\alpha\beta}{\mu s}\big)}{\psi\big(\alpha,\,1+\alpha-\beta,\,-\frac{\alpha\beta}{\mu s}\big)}=-\frac{\alpha}{s}+\frac{\alpha}{s}\int_0^\infty\frac{-\frac{\alpha\beta}{\mu s}\omega_{\alpha,\beta}(t)}{-\frac{\alpha\beta}{\mu s}+t}\,dt$$
$$=-\alpha\int_0^\infty\frac{\omega_{\alpha,\beta}(t)}{\big(s-\frac{\alpha\beta}{\mu t}\big)}\,dt,$$

so for $s=x+\mathrm{i}y$ such that $y>0$ we have

$$\operatorname{Im}\frac{\psi'(s)}{\psi(s)}=\int_0^\infty\frac{\alpha y\,\omega_{\alpha,\beta}(t)}{\big(\big(x-\frac{\alpha\beta}{\mu t}\big)^2+y^2\big)}\,dt>0.$$

This concludes the proof in the case $\alpha<\beta$. The case $\alpha>\beta$ is handled by applying the same process to the other version of the Laplace transform, (5-4).

Finally, we show that the probability density function of the K-distribution is hyperbolically completely monotone. By using Macdonald's integral representation

$$K_\mu(x)K_\mu(y)=\frac{1}{2}\int_0^\infty\exp\Big(-\frac{t}{2}-\frac{x^2+y^2}{2t}\Big)K_\mu\Big(\frac{xy}{t}\Big)\frac{dt}{t}\tag{5-5}$$

(see [Wa44, p. 439]), where $x,y>0$ and $\mu\in\mathbb{R}$, we obtain

$$\omega_{\alpha,\beta,\mu}(uv)\omega_{\alpha,\beta,\mu}(u/v)$$
$$=\frac{2u^{\alpha+\beta-2}}{\Gamma^2(\alpha)\Gamma^2(\beta)}\Big(\frac{\alpha\beta}{\mu}\Big)^{\alpha+\beta}\int_0^\infty\exp\Big(-\frac{t}{2}-\frac{w}{\mu t}(2\alpha\beta u)\Big)K_{\alpha-\beta}\Big(\frac{4\alpha\beta u}{\mu t}\Big)\frac{dt}{t},$$

which is completely monotonic on $(0,\infty)$ as a function of $w=v+1/v$ for all $\alpha,\beta,\mu,u>0$. □

Proof of Theorem 6. We apply Lemma 4. Observe that the Laplace transform of the generalized inverse Gaussian distribution is given by

$$\phi(x):=L[\pi_{\mu,a,b}(x)]=\Big(\frac{a}{2x+a}\Big)^{\frac{\mu}{2}}\cdot\frac{K_\mu\sqrt{b(2x+a)}}{K_\mu(\sqrt{ab})}.$$

Macdonald's integral representation (5-5) then gives

$$\phi(uv)\phi(u/v)=\frac{1}{2}\,\frac{a^\mu\big(K_\mu(\sqrt{ab})\big)^{-2}}{(\alpha w+\beta)^{\mu/2}}\int_0^\infty e^{-\frac{t}{2}-\frac{b}{t}(uw+a)}K_\mu(b\sqrt{\alpha w+\beta})\frac{dt}{t},$$

where $\alpha=2au>0$ and $\beta=4u^2+a^2>0$. On the other hand, by [Ba10, p. 589], the function $x\mapsto x^{\mu/2}K_\mu(\sqrt{x})$ is completely monotonic on $(0,\infty)$ for all $\mu\in\mathbb{R}$. This implies that the function $w \mapsto (b\sqrt{\alpha w+\beta})^{\mu} K_{\mu}(b\sqrt{\alpha w+\beta})$ is also completely monotonic on $(0,\infty)$ for all $\mu \in \mathbb{R}$ and $b, \alpha, \beta > 0$. Since $w \mapsto (\alpha w+\beta)^{-\mu}$ is completely monotonic on $(0,\infty)$ for all $\mu \geq 0$ and $\alpha, \beta > 0$, and $w \mapsto e^{-\frac{b}{t}(uw+a)}$ is completely monotonic on $(0,\infty)$ for all $a, b, u, t > 0$, we conclude that indeed the function $w \mapsto \phi(uv)\phi(u/v)$ is completely monotonic on $(0,\infty)$ for all $\mu \geq 0$ and $a, b > 0$. For the case when $\mu < 0$ we just need to observe that $K_{\mu}(x) = K_{-\mu}(x)$ and in view of the above mentioned results $x \mapsto x^{-\mu/2} K_{\mu}(\sqrt{x})$ is completely monotonic on $(0,\infty)$ for all $\mu < 0$, which implies that the function $w \mapsto (b\sqrt{\alpha w+\beta})^{-\mu} K_{\mu}(b\sqrt{\alpha w+\beta})$ is also completely monotonic on $(0,\infty)$ for all $\mu < 0$ and $b, \alpha, \beta > 0$. Thus, we conclude that the function $w \mapsto \phi(uv)\phi(u/v)$ is completely monotonic on $(0,\infty)$ also for all $\mu < 0$ and $a, b > 0$. Now, applying Lemma 4, the proof is complete. □

Proof of Theorem 7. We will use Lemma 3. From [IK79, Theorem 1.9] we know that $x \mapsto \rho_{\mu,a}(x)$ is the Laplace transform of a probability distribution, and so the moment generating function $s \mapsto \phi_{\mu,a}(s)$ is

$$\phi_{\mu,a}(s) = \rho_{\mu,a}(-s) = \left(\frac{a}{2}\right)^{\mu} \frac{(-s)^{\frac{\mu}{2}}}{\Gamma(\mu+1)} \frac{1}{I_{\mu}(a\sqrt{-s})} = \left(\frac{a}{2}\right)^{\mu} \frac{s^{\frac{\mu}{2}}}{\Gamma(\mu+1)} \frac{1}{J_{\mu}(a\sqrt{s})},$$

in view of the relation $I_{\mu}(\mathrm{i}x) = \mathrm{i}^{\mu} J_{\mu}(x)$. Therefore $s \mapsto \phi_{\mu,a}(s)$ is analytic and zero-free in $\mathbb{C} \setminus [0,\infty)$. Taking the logarithmic derivative of both sides of the preceding display, we obtain

$$\frac{\phi'_{\mu,a}(s)}{\phi_{\mu,a}(s)} = \frac{\mu}{2s} - \frac{a}{2\sqrt{s}} \frac{J'_{\mu}(a\sqrt{s})}{J_{\mu}(a\sqrt{s})}. \tag{5-6}$$

Taking the logarithmic derivative of both sides of the infinite product representation

$$J_{\mu}(x) = \frac{\left(\frac{1}{2}x\right)^{\mu}}{\Gamma(\mu+1)} \prod_{n\geq 1}\left(1 - \frac{x^2}{j_{\mu,n}^2}\right),$$

where $j_{\mu,n}$ stands for the n-th positive zero of $x \mapsto J_{\mu}(x)$, we obtain the classical Mittag-Leffler expansion

$$\frac{J'_{\mu}(x)}{J_{\mu}(x)} = \frac{\mu}{x} - \sum_{n\geq 1} \frac{2x}{j_{\mu,n}^2 - x^2},$$

which implies that

$$\operatorname{Im} \frac{\phi'_{\mu,a}(s)}{\phi_{\mu,a}(s)} = \operatorname{Im}\left(\sum_{n\geq 1} \frac{a^2}{j_{\mu,n}^2 - a^2 s}\right) = \sum_{n\geq 1} \frac{a^4 y}{(j_{\mu,n}^2 - a^2 x)^2 + (a^2 y)^2} > 0,$$

whenever $x = \operatorname{Re} s \in \mathbb{R}$ and $y = \operatorname{Im} s > 0$. Consequently, the conditions in Lemma 3 are satisfied and the proof is complete. □

Proof of Theorem 8. Since $2^{\mu}\Gamma(\mu+1)x^{-\mu}I_{\mu}(x)\to 1$ as $x\to 0$, it follows that $\Omega_{\mu,\nu,\sigma,a,b}(x)\to 1$ as $x\to 0$. Using the recurrence relation $I'_{\mu}(x)=I_{\mu+1}(x)+(\mu/x)I_{\mu}(x)$ (see [Wa44, p. 79]) and the Mittag-Leffler expansion

$$\frac{I_{\mu+1}(x)}{I_{\mu}(x)}=\sum_{n\geq 1}\frac{2x}{x^2+j_{\mu,n}^2},$$

we obtain for $-d\ln\Omega_{\mu,\nu,\sigma,a,b}(x)/dx$ the value

$$\begin{aligned}
&-\frac{a}{2\sqrt{x}}\frac{I'_{\mu}(a\sqrt{x})}{I_{\mu}(a\sqrt{x})}-\frac{b}{2\sqrt{x}}\frac{I'_{\nu}(b\sqrt{x})}{I_{\nu}(b\sqrt{x})}+\frac{b}{2\sqrt{x}}\frac{I'_{\mu}(b\sqrt{x})}{I_{\mu}(b\sqrt{x})}+\frac{a}{2\sqrt{x}}\frac{I'_{\nu}(a\sqrt{x})}{I_{\nu}(a\sqrt{x})}+\sum_{n\geq 1}\Big(\frac{b^2}{b^2x+j_{\sigma,n}^2}\Big)\\
&=\sum_{n\geq 1}\Big(-\frac{1}{x+j_{\mu,n}^2a^{-2}}-\frac{1}{x+j_{\nu,n}^2b^{-2}}+\frac{1}{x+j_{\mu,n}^2b^{-2}}+\frac{1}{x+j_{\nu,n}^2a^{-2}}+\frac{1}{x+j_{\sigma,n}^2b^{-2}}\Big)\\
&=\int_0^{\infty}e^{-xt}\Big(\sum_{n\geq 1}\big(-e^{-j_{\mu,n}^2a^{-2}t}-e^{-j_{\nu,n}^2b^{-2}t}+e^{-j_{\mu,n}^2b^{-2}t}+e^{-j_{\nu,n}^2a^{-2}t}+e^{-j_{\sigma,n}^2b^{-2}t}\big)\Big)\,dt\\
&=\int_0^{\infty}e^{-xt}\Big(\sum_{n\geq 1}\big(\big(-e^{-j_{\mu,n}^2a^{-2}t}+e^{-j_{\mu,n}^2b^{-2}t}\big)+\big(-e^{-j_{\nu,n}^2b^{-2}t}+e^{-j_{\sigma,n}^2b^{-2}t}\big)+e^{-j_{\nu,n}^2a^{-2}t}\big)\Big)\,dt.
\end{aligned}$$

Since $a<b$ and $-1<\sigma<\nu$, we have $e^{-j_{\mu,n}^2a^{-2}t}<e^{-j_{\mu,n}^2b^{-2}t}$ and $e^{-j_{\nu,n}^2b^{-2}t}<e^{-j_{\sigma,n}^2b^{-2}t}$ for every $n\in\mathbb{N}$ and $\mu>-1$, where we used the fact that $\mu\mapsto j_{\mu,n}$ is increasing on $(-1,\infty)$ for every $n\in\mathbb{N}$. Consequently, the last expression is positive, and this implies that $-d\ln\Omega_{\mu,\nu,\sigma,a,b}(x)/dx$ is completely monotonic in x on $(0,\infty)$. In view of Lemma 1 we conclude that $\Omega_{\mu,\nu,\sigma,a,b}(x)$ is indeed the Laplace transform of an infinitely divisible distribution. □

Proof of Theorem 9. The result will be established by verifying the conditions in Lemma 1 and by using the main idea of the proof of [Is90, Theorem 1]. Clearly $\Omega_{\mu,\nu,a,b}(x)\to 1$ as $x\to 0$, and by using the same approach as in the proof of Theorem 8, we obtain

$$\begin{aligned}
&-\frac{d\ln\Omega_{\mu,\nu,a,b}(x)}{dx}\\
&\quad=\frac{b}{2\sqrt{x}}+\sum_{n\geq 1}\Big(-\frac{1}{x+j_{\mu,n}^2a^{-2}}-\frac{1}{x+j_{\nu,n}^2b^{-2}}+\frac{1}{x+j_{\mu,n}^2b^{-2}}+\frac{1}{x+j_{\nu,n}^2a^{-2}}\Big)\\
&\quad=\frac{b}{2\sqrt{x}}-\sum_{n\geq 1}\frac{1}{x+j_{\nu,n}^2b^{-2}}+\int_0^{\infty}e^{-xt}\sum_{n\geq 1}\big(-e^{-j_{\mu,n}^2a^{-2}t}+e^{-j_{\mu,n}^2b^{-2}t}+e^{-j_{\nu,n}^2a^{-2}t}\big)\,dt.
\end{aligned}$$

Note that $e^{-j_{\mu,n}^2a^{-2}t}<e^{-j_{\mu,n}^2b^{-2}t}$ for all $\mu>-1$, $b>a$, $t>0$ and $n\in\mathbb{N}$. Consequently,

$$x\mapsto\int_0^{\infty}e^{-xt}\sum_{n\geq 1}\big(-e^{-j_{\mu,n}^2a^{-2}t}+e^{-j_{\mu,n}^2b^{-2}t}+e^{-j_{\nu,n}^2a^{-2}t}\big)\,dt$$

is a completely monotonic function on $(0,\infty)$ for all $\mu>-1$, $\nu>-1$, and $b>a$. Now, define

$$\eta_{\mu,b}(x)=\frac{b}{2\sqrt{x}}-\sum_{n\geq 1}\frac{1}{x+j_{\nu,n}^2b^{-2}}.$$

Then $\frac{b}{2\sqrt{x}}=\frac{1}{\pi}\int_0^\infty\frac{dt}{x+b^{-2}t^2}$ and thus

$$\eta_{\mu,b}(x)=\frac{1}{\pi}\int_0^\infty\frac{dt}{x+b^{-2}t^2}-\sum_{n\geq 1}\frac{1}{x+j_{\nu,n}^2b^{-2}}.$$

On the other hand, for every s such that $j_{\nu,m}\leq s<j_{\nu,m+1}$, we have

$$\begin{aligned}\int_0^s\frac{dt}{x+b^{-2}t^2}&\geq\int_{j_{\nu,0}}^{j_{\nu,m}}\frac{dt}{x+b^{-2}t^2}=\sum_{n=1}^m\int_{j_{\nu,n-1}}^{j_{\nu,n}}\frac{dt}{x+b^{-2}t^2}\\&\geq\sum_{n=1}^m\int_{j_{\nu,n-1}}^{j_{\nu,n}}\frac{dt}{x+b^{-2}j_{\nu,n}^2}=\sum_{n=1}^m\frac{j_{\nu,n}-j_{\nu,n-1}}{x+b^{-2}j_{\nu,n}^2},\end{aligned}$$

where we used the fact that $t\mapsto 1/(x+b^{-2}t^2)$ is a decreasing function on $(0,\infty)$ for all $b>0$ and $x>0$. We thus arrive at

$$\eta_{\mu,b}(x)\geq\frac{1}{\pi}\sum_{n\geq 1}\frac{j_{\nu,n}-j_{\nu,n-1}}{x+b^{-2}j_{\nu,n}^2}-\sum_{n\geq 1}\frac{1}{x+j_{\nu,n}^2b^{-2}}.$$

Since $j_{\nu,n}-j_{\nu,n-1}>\pi$ for $\nu>\frac{1}{2}$ (see [Se01, Theorem 1.6], for example), we conclude that $\eta_{\mu,b}(x)>0$ for all $x>0$ and $b>0$. Moreover, from the previous results we obtain that

$$\begin{aligned}\frac{(-1)^k}{k!}\eta_{\mu,b}^{(k)}(x)&=\frac{1}{\pi}\int_0^\infty\frac{dt}{(x+b^{-2}t^2)^{k+1}}-\sum_{n\geq 1}\frac{1}{(x+j_{\nu,n}^2b^{-2})^{k+1}}\\&\geq\frac{1}{\pi}\sum_{n\geq 1}\frac{j_{\nu,n}-j_{\nu,n-1}}{(x+j_{\nu,n}^2b^{-2})^{k+1}}-\sum_{n\geq 1}\frac{1}{(x+j_{\nu,n}^2b^{-2})^{k+1}}\end{aligned}$$

for all $x>0$, $b>0$ and $k\in\mathbb{N}$, where we have used that $t\mapsto 1/(x+b^{-2}t^2)^{k+1}$ is a decreasing function on $(0,\infty)$ for all $x>0$, $b>0$ and $k\in\mathbb{N}$. Consequently, $(-1)^k\eta_{\mu,b}^{(k)}(x)\geq 0$ for all $k\in\mathbb{N}$ and $x>0$, $b>0$. Therefore $\eta_{\mu,b}$ is also a completely monotonic function on $(0,\infty)$ and this implies that $x\mapsto -d\ln\Omega_{\mu,\nu,a,b}(x)/dx$ is the sum of two completely monotonic functions on $(0,\infty)$. □

Proof of Theorem 10. Let $F(z)=e^{-a\sqrt{z}}z^{-\frac{\mu}{2}}I_\mu(a\sqrt{z})$. Using eq. 10.30.1 of [NIST10], we have

$$F(z)\sim\frac{e^{-a\sqrt{z}}a^\mu}{2^\mu\Gamma(\mu+1)}\quad\text{as } z\to 0,$$

leading to $F(z) = o(|z|^{-1})$ as $|z| \to 0$. Using eq. 10.30.4 of [NIST10], we obtain

$$F(z) \sim \frac{1}{\sqrt{2\pi a}} z^{-\frac{(\mu+1/2)}{2}} \quad \text{as } |z| \to \infty.$$

These relations hold uniformly in every sector $|\arg z| \le \pi - \epsilon$, $\epsilon > 0$. Hence conditions **a** and **b** of Lemma 5 hold, and by using (5-8) we arrive at

$$F(-t-\mathrm{i}\eta) = F(e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta)) = e^{a\mathrm{i}\sqrt{t+\mathrm{i}\eta}}(t+\mathrm{i}\eta)^{-\frac{\mu}{2}} J_\mu(a\sqrt{t+\mathrm{i}\eta}),$$
$$F(-t+\mathrm{i}\eta) = F(e^{\mathrm{i}\pi}(t-\mathrm{i}\eta)) = e^{-a\mathrm{i}\sqrt{t-\mathrm{i}\eta}}(t-\mathrm{i}\eta)^{-\frac{\mu}{2}} J_\mu(a\sqrt{t-\mathrm{i}\eta}).$$

By using the asymptotic relation

(5-7) $$J_\nu(z) \sim \frac{(z/2)^\nu}{\Gamma(\nu+1)}, \quad \nu \ne -1, -2, \dots, \quad \text{as } z \to 0$$

(see [NIST10, eq. 10.7.3]) we conclude that the function $F(-t-\mathrm{i}\eta) - F(-t+\mathrm{i}\eta)$ is continuous in every rectangle $[t_1, t_2] \times [0, \eta]$, where t_1, t_2, $\eta > 0$. Thus, the limit and the integral in (1-3) can be interchanged and the fundamental theorem of calculus yields

$$\alpha'(t) = \frac{1}{\pi} t^{-\frac{\mu}{2}} J_\mu(a\sqrt{t}) \sin(a\sqrt{t}),$$

so that $\alpha(t)$ is continuous and $\lim_{t\to 0^+} \alpha(t)$ exists. Letting $\tilde{\alpha}(t) = \alpha(t) - \alpha(0)$, we get the normalized measure $\tilde{\alpha}(t)$ with $\tilde{\alpha}'(t) = \alpha'(t)$. □

Proof of Theorem 11. Let us consider the function

$$F(z) = z^{\frac{\nu-\mu}{2}} I_\mu(a\sqrt{z}) K_\nu(b\sqrt{z}).$$

Using eqs. 10.30.1 and 10.30.2 of [NIST10] we get, for $\nu > 0$ and $\mu > -1$,

$$F(z) \sim \left(\tfrac{1}{2}\right)^{\mu-\nu+1} \frac{\Gamma(\nu)}{\Gamma(\mu+1)} \frac{a^\mu}{b^\nu} \quad \text{as } z \to 0.$$

And using eqs. 10.30.1 and 10.30.3 of [NIST10] we get, for $\nu = 0$ and $\mu > -1$,

$$F(z) \sim -\frac{a^\mu}{2^\mu \Gamma(\mu+1)} \ln(b\sqrt{z}) \quad \text{as } z \to 0.$$

Similarly, by using the relations $K_\nu(z) = K_{-\nu}(z)$ (see eqs. 10.27.3, 10.30.2 and 10.30.1 of [NIST10]) we have, for $-1 < \nu < 0$ and $\mu > -1$,

$$F(z) \sim \frac{a^\mu b^\nu}{2^{\mu+\nu+1}} \frac{\Gamma(-\nu)}{\Gamma(\mu+1)} z^{\nu+1} \quad \text{as } z \to 0.$$

We thus obtain $F(z) = o(|z|^{-1})$ as $z \to 0$ for all $\nu > -1$ and $\mu > -1$. Further using eqs. 10.30.4 and 10.25.3 of [NIST10], we see that

$$F(z) \sim z^{\frac{\nu-\mu-1}{2}} \frac{e^{-(b-a)\sqrt{z}}}{2} \quad \text{as } z \to \infty,$$

uniformly in every sector $|\arg z| \le \pi - \varepsilon$, $\varepsilon > 0$. This leads to $F(z) = o(1)$ as $|z| \to \infty$, whenever $b \ge a$ and $\nu - \mu < 1$. Thus, the conditions in Lemma 5 have been verified because $F(z)$ is analytic in $|\arg z| < \pi$.

Now, in order to apply Lemma 7, we observe that

$$\begin{aligned}
F(-t-\mathrm{i}\eta) &= F(e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta)) \\
&= (e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta))^{\frac{\nu-\mu}{2}} I_\mu(ae^{-\frac{\mathrm{i}\pi}{2}}\sqrt{t+\mathrm{i}\eta}) K_\nu(be^{-\frac{\mathrm{i}\pi}{2}}\sqrt{t+\mathrm{i}\eta}) \\
&= \tfrac{1}{2}\mathrm{i}\pi(t+\mathrm{i}\eta)^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t+\mathrm{i}\eta}) H_\nu^{(1)}(b\sqrt{t+\mathrm{i}\eta}), \\
F(-t+\mathrm{i}\eta) &= F(e^{\mathrm{i}\pi}(t-\mathrm{i}\eta)) \\
&= (e^{\mathrm{i}\pi}(t-\mathrm{i}\eta))^{\frac{\nu-\mu}{2}} I_\mu(ae^{\frac{\mathrm{i}\pi}{2}}\sqrt{t-\mathrm{i}\eta}) K_\nu(be^{\frac{\mathrm{i}\pi}{2}}\sqrt{t-\mathrm{i}\eta}) \\
&= -\tfrac{1}{2}\mathrm{i}\pi(t-\mathrm{i}\eta)^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t-\mathrm{i}\eta}) H_\nu^{(2)}(b\sqrt{t-\mathrm{i}\eta}),
\end{aligned}$$

where we have used the following relations (see [IM79, p. 459; NIST10, 10.27]):

$$\text{(5-8)} \quad \begin{cases}
H_\mu^{(1)}(z) = J_\mu(z) + \mathrm{i}Y_\mu(z), & H_\mu^{(2)}(z) = J_\mu(z) - \mathrm{i}Y_\mu(z), \\
K_\mu(ze^{-\frac{1}{2}\mathrm{i}\pi}) = \frac{1}{2}\mathrm{i}\pi e^{\frac{1}{2}\mathrm{i}\pi\mu} H_\mu^{(1)}(z), & K_\mu(ze^{\frac{1}{2}\mathrm{i}\pi}) = -\frac{1}{2}\mathrm{i}\pi e^{-\frac{1}{2}\mathrm{i}\pi\mu} H_\mu^{(2)}(z), \\
I_\mu(ze^{\pm\frac{1}{2}\mathrm{i}\pi}) = e^{\pm\frac{1}{2}\mathrm{i}\mu\pi} J_\mu(z), & I_\mu(z) = e^{-\frac{1}{2}\mathrm{i}\mu\pi} J_\mu(ze^{\frac{1}{2}\mathrm{i}\pi}).
\end{cases}$$

We next justify interchanging the limit and the integral in (1-3), in three cases:

(i) $\nu > 0$ and $\mu > -1$: By using the asymptotic relation

$$\text{(5-9)} \qquad H_\nu^{(1)}(z) \sim -H_\nu^{(2)}(z) \sim -\frac{\mathrm{i}}{\pi}\Gamma(\nu)(z/2)^{-\nu}, \quad z \to 0, \quad \operatorname{Re}\nu > 0$$

(see [NIST10, eq. 10.7.7]) and (5-7), we observe that the function $F(-t-\mathrm{i}\eta) - F(-t+\mathrm{i}\eta)$ is continuous in every rectangle $[t_1, t_2] \times [0, \eta]$, where $t_1,\ t_2,\ \eta > 0$. Thus, in this case, we can interchange the limit and the integral in (1-3).

(ii) $\nu = 0$ and $\mu > -1$: By using the asymptotic relation

$$\text{(5-10)} \qquad H_0^{(1)}(z) \sim -H_0^{(2)}(z) \sim \frac{2\mathrm{i}}{\pi}\ln(z), \quad z \to 0$$

(see [NIST10, eq. 10.7.2]) and (5-7), we obtain

$$\begin{aligned}
F(-t-\mathrm{i}\eta) &\sim -\frac{a^\mu}{2^\mu\Gamma(\mu+1)}\ln(b\sqrt{t+\mathrm{i}\eta}), \quad t \to 0,\ \eta \to 0, \\
F(-t+\mathrm{i}\eta) &\sim -\frac{a^\mu}{2^\mu\Gamma(\mu+1)}\ln(b\sqrt{t-\mathrm{i}\eta}), \quad t \to 0,\ \eta \to 0.
\end{aligned}$$

By the definition of asymptotic convergence, this means there are constants c, α, $\beta > 0$ such that

$$\left|F(-t\mp\mathrm{i}\eta)\right| \le \frac{ca^\mu}{2^\mu\Gamma(\mu+1)}\left|\ln(b\sqrt{t\pm\mathrm{i}\eta})\right|, \quad \text{for } 0 < t < \alpha,\ 0 < \eta < \beta.$$

Because the right side of this (double) inequality is integrable over $[t_1, t_2]$, where $0 < t_1 < t_2 < \alpha$, we are allowed to interchange the limit and the integral in (1-3).

(iii) $-1 < \nu < 0$ and $\mu > -1$: Using the relations

$$H^{(1)}_{-\nu}(z) = e^{\nu\pi \mathrm{i}} H^{(1)}_{\nu}(z), \quad H^{(2)}_{-\nu}(z) = e^{-\nu\pi \mathrm{i}} H^{(2)}_{\nu}(z) \tag{5-11}$$

(see [NIST10, eq. 10.4.6]) and the asymptotic relations (5-7) and (5-9), we have

$$F(-t-\mathrm{i}\eta) \sim \frac{e^{-\nu\pi\mathrm{i}}\Gamma(-\nu)a^{\mu}b^{\nu}}{2^{\mu+\nu+1}\Gamma(\mu+1)}(t+\mathrm{i}\eta)^{\nu},$$
$$F(-t+\mathrm{i}\eta) \sim \frac{e^{\nu\pi\mathrm{i}}\Gamma(-\nu)a^{\mu}b^{\nu}}{2^{\mu+\nu+1}\Gamma(\mu+1)}(t-\mathrm{i}\eta)^{\nu},$$

as $t \to 0$ and $\eta \to 0$. Then the same reasoning as in the previous case justifies interchanging the limit and the integral in (1-3).

Altogether we get

$$\alpha'(t) = \frac{1}{2\pi\mathrm{i}} \lim_{\eta\to 0^+} \big(F(-t-\mathrm{i}\eta) - F(-t+\mathrm{i}\eta)\big) = \tfrac{1}{2} t^{\frac{\nu-\mu}{2}} J_{\mu}(a\sqrt{t}) J_{\nu}(b\sqrt{t}),$$

showing that $\alpha(t)$ is continuous and $\lim_{t\to 0}\alpha(t)$ exists. Then $\tilde{\alpha}(t) = \alpha(t) - \alpha(0)$ is a normalized measure with $\tilde{\alpha}'(t) = \alpha'(t)$. □

Proof of Theorem 12. A function $x \mapsto \phi(x)$ is the Laplace transform of a probability distribution on $(0, \infty)$ if and only if $\phi(0) = 1$ and $\phi(x)$ has the form

$$\phi(x) = \int_0^{\infty} e^{-xt}\, d\mu(t)$$

and is completely monotonic on $(0, \infty)$ (see [Bo92, p. 8]). In view of (3-4) this implies that $x \mapsto 2\mu I_{\mu}(\sqrt{x})K_{\mu}(\sqrt{x})$ is the Laplace transform of a probability distribution with support $(0, \infty)$. Combining this with the second equality in (3-4) we see that this distribution's probability density function,

$$\varsigma_{\mu}(x) = \mu \int_0^{\infty} e^{-tx} J^2_{\mu}(\sqrt{t})\, dt,$$

is a completely monotonic function on $(0, \infty)$ for all $\mu > 0$, and by using the Goldie–Steutel law we obtain that indeed the function $x \mapsto 2\mu I_{\mu}(\sqrt{x})K_{\mu}(\sqrt{x})$ is the Laplace transform of an infinitely divisible probability distribution with support $(0, \infty)$. Now using the formula $\int_0^{\infty} e^{-st} J^2_{\mu}(\sqrt{t})\, dt = \frac{1}{s} e^{-\frac{1}{2s}} I_{\mu}\big(\frac{1}{2s}\big)$, where $\mu > 0$ (see [OB12, p. 139]), we get

$$\varsigma_{\mu}(x) = \frac{\mu}{x} e^{-\frac{1}{2x}} I_{\mu}\Big(\frac{1}{2x}\Big),$$

which completes the derivation of the probability density function. □

Proof of Corollary 1. The assertions in parts **a**, **b** and **c** follow immediately from the integral representation (3-4). We prove part **d**. From [Se21, p. 8], we have

$$[xI_\mu(x)K_\mu(x)]^{-1} = \frac{I_{\mu-1}(x)}{I_\mu(x)} + \frac{K_{\mu-1}(x)}{K_\mu(x)}$$

and in view of the three-term recurrence relation

$$I_{\mu-1}(x) = I_{\mu+1}(x) + \frac{2\mu}{x} I_\mu(x),$$

we arrive at

$$\big(xI_\mu(x)K_\mu(x)\big)^{-1} = \frac{2\mu}{x} + \frac{I_{\mu+1}(x)}{I_\mu(x)} + \frac{K_{\mu-1}(x)}{K_\mu(x)}.$$

Now, replacing x by $\sqrt{x}$ and dividing both sides by $\sqrt{x}$, we obtain

$$\big(xI_\mu(\sqrt{x})K_\mu(\sqrt{x})\big)^{-1} = \frac{2\mu}{x} + \frac{1}{\sqrt{x}}\frac{I_{\mu+1}(\sqrt{x})}{I_\mu(\sqrt{x})} + \frac{1}{\sqrt{x}}\frac{K_{\mu-1}(\sqrt{x})}{K_\mu(\sqrt{x})}.$$

Using the integral representation

$$\frac{K_{\mu-1}(\sqrt{x})}{\sqrt{x}K_\mu(\sqrt{x})} = \frac{2}{\pi^2}\int_0^\infty \frac{t^{-1}}{x+t}\big(J_\mu^2(\sqrt{t}) + Y_\mu^2(\sqrt{t})\big)^{-1} dt, \tag{5-12}$$

where $x > 0$, $\mu \geq 0$ (see for example [Is77a, eq. (1.3)] or [IM82, eq. (2.4)]) and the series and integral representations for the ratios of modified Bessel functions of the first and second kinds, we obtain

$$\big(xI_\mu(\sqrt{x})K_\mu(\sqrt{x})\big)^{-1} = \frac{2\mu}{x} + \sum_{n\geq 1}\frac{2}{x+j_{\mu,n}^2} + \frac{2}{\pi^2}\int_0^\infty \frac{t^{-1}}{x+t}\big(J_\mu^2(\sqrt{t})+Y_\mu^2(\sqrt{t})\big)^{-1} dt,$$

which is the sum of three completely monotonic functions on $(0,\infty)$ for all $\mu > 0$.

For **e** we use the integral representation of $I_\mu(\sqrt{x})K_\mu(\sqrt{x})$ and the Landau bound

$$|J_\mu(t)| \leq c_L |t|^{-\frac{1}{3}}, \quad \mu > 0, \quad t \in \mathbb{R}, \quad c_L = 0.7857468704\ldots$$

(see [La00] or [BMPS16]). We obtain, as desired,

$$I_\mu(\sqrt{x})K_\mu(\sqrt{x}) \leq \frac{c_L^2}{2}\int_0^\infty \frac{t^{-\frac{2}{3}}}{x+t} dt = \frac{\pi c_L^2}{\sqrt{3}x^{\frac{2}{3}}}. \qquad \square$$

Proof of Theorem 13. Let

$$F(z) = z^{\frac{\nu-\mu}{2}} e^{-a\sqrt{z}} I_\mu(a\sqrt{z}) K_\nu(b\sqrt{z}).$$

The steps on the first half-page of the proof of Theorem 11 apply unchanged until the use of eq. 10.30.4 of [NIST10], which is replaced by 10.30.5. We arrive at

$$F(z) \sim \frac{z^{\frac{\nu-\mu-1}{2}} e^{-b\sqrt{z}}}{2\sqrt{ab}} \quad \text{as } |z| \to \infty,$$

uniformly in every sector $|\arg z| \le \pi - \epsilon$, $\epsilon > 0$, and thus $F(z) = o(1)$ as $|z| \to \infty$. Hence conditions **a** and **b** in Lemma 5 have been verified, and this implies that the corresponding Stieltjes transform representation exists. The use of the inversion theorem to obtain the corresponding measure is also almost the same: from Lemma 7 and equations (5-8) we obtain

$$F(-t-\mathrm{i}\eta) = F(e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta)) = \frac{\mathrm{i}\pi}{2}(t+\mathrm{i}\eta)^{\frac{\nu-\mu}{2}} e^{a\mathrm{i}\sqrt{t+\mathrm{i}\eta}} J_\mu(a\sqrt{t+\mathrm{i}\eta}) H_\nu^{(1)}(b\sqrt{t+\mathrm{i}\eta}),$$

$$F(-t+\mathrm{i}\eta) = F(e^{\mathrm{i}\pi}(t-\mathrm{i}\eta)) = -\frac{\mathrm{i}\pi}{2}(t-\mathrm{i}\eta)^{\frac{\nu-\mu}{2}} e^{-a\mathrm{i}\sqrt{t-\mathrm{i}\eta}} J_\mu(a\sqrt{t-\mathrm{i}\eta}) H_\nu^{(2)}(b\sqrt{t-\mathrm{i}\eta}).$$

The justification for interchanging the limit and the integral in (1-3) also goes as in the proof of Theorem 11, and we omit it. Finally, from the relations (5-8), we get

$$\alpha'(t) = \tfrac{1}{2} t^{\frac{\nu-\mu}{2}} J_\mu(a\sqrt{t})\big(J_\nu(b\sqrt{t})\cos(a\sqrt{t}) - Y_\nu(b\sqrt{t})\sin(a\sqrt{t})\big). \qquad \square$$

Proof of Theorem 14. From the recurrence relation $K'_\mu(x) = -K_{\mu-1}(x) - \frac{\mu}{x}K_\mu(x)$ (see [Wa44, p. 79]) and the equivalent form of the integral representation (5-12), namely,

$$\frac{K_{\mu-1}(\sqrt{x})}{\sqrt{x}K_\mu(\sqrt{x})} = \frac{4}{\pi^2}\int_0^\infty \frac{t^{-1}}{x+t^2}\big(J_\mu^2(t) + Y_\mu^2(t)\big)^{-1} dt,$$

where $x > 0$, $\mu \ge 0$, we obtain

$$\begin{aligned}
-\frac{d \ln \chi_{\mu,\nu,a,b}(x)}{dx} &= \frac{a}{2\sqrt{x}} + \frac{\mu-\nu}{2x} - \frac{a}{2\sqrt{x}}\frac{I'_\mu(a\sqrt{x})}{I_\mu(a\sqrt{x})} - \frac{b}{2\sqrt{x}}\frac{K'_\nu(b\sqrt{x})}{K_\nu(b\sqrt{x})} \\
&= \frac{a}{2\sqrt{x}} - \frac{a}{2\sqrt{x}}\frac{I_{\mu+1}(a\sqrt{x})}{I_\mu(a\sqrt{x})} + \frac{b}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})} \\
&= \Theta(x) + \frac{2}{\pi^2}\int_0^\infty \frac{t^{-1}}{x+b^{-2}t^2}\big(J_\nu^2(t) + Y_\nu^2(t)\big)^{-1} dt,
\end{aligned}$$

where

$$\begin{aligned}
\Theta(x) &= \frac{1}{\pi}\int_0^\infty \frac{dt}{x+a^{-2}t^2} - \sum_{n\ge1}\frac{1}{x+a^{-2}j_{\mu,n}^2} \\
&\ge \frac{1}{\pi}\sum_{n\ge1}\frac{j_{\mu,n}-j_{\mu,n-1}}{x+a^{-2}j_{\mu,n}^2} - \sum_{n\ge1}\frac{1}{x+a^{-2}j_{\mu,n}^2} \ge 0.
\end{aligned}$$

Further,

$$\frac{(-1)^m}{m!}\frac{d^m\Theta(x)}{dx^m} \ge \frac{1}{\pi}\sum_{n\ge1}\frac{j_{\mu,n}-j_{\mu,n-1}}{(x+a^{-2}j_{\mu,n}^2)^{m+1}} - \sum_{n\ge1}\frac{1}{(x+a^{-2}j_{\mu,n}^2)^{m+1}} \ge 0,$$

since $j_{\mu,n} - j_{\mu,n-1} > \pi$ whenever $\mu > \frac{1}{2}$ and $n \in \mathbb{N}$. Thus, $x \mapsto -d \ln \chi_{\mu,\nu,a,b}(x)/dx$ is completely monotonic on $(0, \infty)$ for all $a, b, \nu > 0$ and $\mu > \frac{1}{2}$. In view of Lemma 1 the proof is complete. $\square$

Proof of Theorem 15. Let $F(z)=z^{\frac{\mu+\nu}{2}}K_\mu(a\sqrt{z})K_\nu(b\sqrt{z})$. Using the asymptotic eq. 10.30.2 from [NIST10] for a, b, μ, $\nu>0$, we have

$$z^{\frac{\mu+\nu}{2}}K_\mu(a\sqrt{z})K_\nu(b\sqrt{z})\sim \tfrac{1}{4}\Gamma(\mu)\Gamma(\nu)\frac{2^{\mu+\nu}}{a^\mu b^\nu}$$

as $z\to 0$. Hence, $f(z)=o(|z|^{-1})$ as $|z|\to 0$. Using [NIST10, eq. 10.25.3], we have

$$z^{\frac{\mu+\nu}{2}}K_\mu(a\sqrt{z})K_\nu(b\sqrt{z})\sim \frac{\pi}{2\sqrt{ab}}\sqrt{z}^{\mu+\nu-1}e^{-(a+b)\sqrt{z}}$$

as $|z|\to\infty$ and $|\arg z|<\frac{3}{2}\pi$. From this we see that $F(z)=o(1)$ as $|z|\to\infty$, uniformly in every sector $|\arg z|\le\pi-\varepsilon$, $\varepsilon>0$. Therefore, the conditions of Lemma 5 hold true. At the same time, using (5-8) we obtain

$$F(-t-\mathrm{i}\eta)=F(e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta))=-\tfrac{1}{4}\pi^2(t+\mathrm{i}\eta)^{\frac{\mu+\nu}{2}}H_\mu^{(1)}(a\sqrt{t+\mathrm{i}\eta})H_\nu^{(1)}(b\sqrt{t+\mathrm{i}\eta}),$$
$$F(-t+\mathrm{i}\eta)=F(e^{\mathrm{i}\pi}(t-\mathrm{i}\eta))=-\tfrac{1}{4}\pi^2(t-\mathrm{i}\eta)^{\frac{\mu+\nu}{2}}H_\mu^{(2)}(a\sqrt{t-\mathrm{i}\eta})H_\nu^{(2)}(b\sqrt{t-\mathrm{i}\eta}),$$

By using the asymptotic relation (5-9), we conclude that $F(-t-\mathrm{i}\eta)-F(-t+\mathrm{i}\eta)$ is continuous in every rectangle $[t_1,t_2]\times[0,\eta]$, $t_1,t_2,\eta>0$. Consequently, we can interchange the limit and the integral in (1-3). In view of Lemma 7 we arrive at

$$\alpha'(t)=-14\pi t^{\frac{\mu+\nu}{2}}\big(J_\mu(a\sqrt{t})Y_\nu(b\sqrt{t})+J_\nu(b\sqrt{t})Y_\mu(a\sqrt{t})\big).$$

Thus $\alpha(t)$ is continuous and $\lim_{t\to 0^+}\alpha(t)$ exists. Then $\tilde{\alpha}(t)=\alpha(t)-\alpha(0)$ is the normalized measure with $\tilde{\alpha}'(t)=\alpha'(t)$. □

Proof of Theorem 16. Infinite divisibility results from [IK79, Theorem 1.8]. Namely, we know that the function $x\mapsto(\sqrt{x})^\nu K_\nu(\sqrt{x})/(2^{\nu-1}\Gamma(\nu))$ is the Laplace transform of an infinitely divisible distribution with support $(0,\infty)$ whenever $\nu>0$. Thus the Laplace transform $\vartheta_{\mu,\nu,a,b}(x)$ is the product of Laplace transforms of two infinitely divisible distributions, and hence the Laplace transform of an infinitely divisible distribution, by [SH03, Proposition 2.1]. More precisely, (5-12) gives

$$\begin{aligned}-\frac{d\ln\vartheta_{\mu,\nu,a,b}(x)}{dx}&=-\frac{\mu+\nu}{2x}-\frac{a}{2\sqrt{x}}\frac{K_\mu'(a\sqrt{x})}{K_\mu(a\sqrt{x})}-\frac{b}{2\sqrt{x}}\frac{K_\nu'(b\sqrt{x})}{K_\nu(b\sqrt{x})}\\&=\frac{a}{2\sqrt{x}}\frac{K_{\mu-1}(a\sqrt{x})}{K_\mu(a\sqrt{x})}+\frac{b}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})}\\&=\frac{1}{\pi^2}\bigg(\int_0^\infty\frac{t^{-1}}{x+ta^{-2}}\big(J_\mu^2(\sqrt{t})+Y_\mu^2(\sqrt{t})\big)^{-1}dt\\&\qquad+\int_0^\infty\frac{t^{-1}}{x+tb^{-2}}\big(J_\nu^2(\sqrt{t})+Y_\nu^2(\sqrt{t})\big)^{-1}dt\bigg),\end{aligned}$$

which as a function of x is completely monotonic on $(0,\infty)$ for all strictly positive real numbers a, b, μ and ν. Applying Lemma 1, this shows that $x \mapsto \vartheta_{\mu,\nu,a,b}(x)$ is the Laplace transform of an infinitely divisible distribution.

Set $\xi(x) = \vartheta_{\mu,\nu,a,b}(x)/\vartheta_{\mu,\nu,a,b}(\alpha x)$. Then

$$\begin{aligned}
&-\frac{d\ln\xi(x)}{dx}\\
&= -\frac{a}{2\sqrt{x}}\frac{K'_\mu(a\sqrt{x})}{K_\mu(a\sqrt{x})} - \frac{b}{2\sqrt{x}}\frac{K'_\nu(b\sqrt{x})}{K_\nu(b\sqrt{x})} + \frac{a\sqrt{\alpha}}{2\sqrt{x}}\frac{K'_\mu(a\sqrt{\alpha x})}{K_\mu(a\sqrt{\alpha x})} + \frac{b\sqrt{\alpha}}{2\sqrt{x}}\frac{K'_\nu(b\sqrt{\alpha x})}{K_\nu(b\sqrt{\alpha x})}\\
&= \frac{a}{2\sqrt{x}}\frac{K_{\mu-1}(a\sqrt{x})}{K_\mu(a\sqrt{x})} + \frac{b}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})} - \frac{a\sqrt{\alpha}}{2\sqrt{x}}\frac{K_{\mu-1}(a\sqrt{\alpha x})}{K_\mu(a\sqrt{\alpha x})} - \frac{b\sqrt{\alpha}}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{\alpha x})}{K_\nu(b\sqrt{\alpha x})}\\
&= \frac{a^2}{\pi^2}\int_0^\infty \frac{(1-\alpha)}{(a^2\alpha x+t)(a^2x+t)}\big(J_\mu^2(\sqrt{t})+Y_\mu^2(\sqrt{t})\big)^{-1}dt\\
&\qquad\qquad + \frac{b^2}{\pi^2}\int_0^\infty \frac{(1-\alpha)}{(b^2\alpha x+t)(b^2x+t)}\big(J_\nu^2(\sqrt{t})+Y_\nu^2(\sqrt{t})\big)^{-1}dt
\end{aligned}$$

is completely monotonic with respect to x on $(0, \infty)$ for all $\alpha \in (0, 1)$ and strictly positive real numbers a, b, μ and ν. Applying Lemmas 1 and 2 we conclude that ξ is the Laplace transform of an infinitely divisible distribution, and so $x \mapsto \vartheta_{\mu,\nu,a,b}(x)$ is the Laplace transform of a self-decomposable distribution.

Finally, let $w = v + 1/v$. For a, b, μ, ν arbitrary positive real numbers set

$$\alpha_{\mu,\nu,a,b} = \frac{a^\mu b^\nu}{2^{\mu+\nu-2}\Gamma(\mu)\Gamma(\nu)}.$$

Using Macdonald's integral representation (5-5) for modified Bessel functions of the second kind we can write

$$\begin{aligned}
&\vartheta_{\mu,\nu,a,b}(uv)\vartheta_{\mu,\nu,a,b}(u/v)\\
&\quad = \alpha^2_{\mu,\nu,a,b}\cdot u^{\mu+\nu}K_\mu(a\sqrt{uv})K_\mu(a\sqrt{u/v})K_\nu(b\sqrt{uv})K_\nu(b\sqrt{u/v})\\
&\quad = \tfrac{1}{4}\alpha^2_{\mu,\nu,a,b}\,u^{\mu+\nu}\\
&\qquad \times \int_0^\infty\!\!\int_0^\infty \exp\Big(-wu\Big(\frac{b^2t+a^2s}{2ts}\Big)\Big)\exp\Big(-\frac{t+s}{2}\Big)K_\mu\Big(\frac{au^2}{t}\Big)K_\nu\Big(\frac{bu^2}{s}\Big)\,\frac{dt}{t}\,\frac{ds}{s}.
\end{aligned}$$

The factor containing w makes the function $w \mapsto \vartheta_{\mu,\nu,a,b}(uv)\vartheta_{\mu,\nu,a,b}\,(u/v)$ completely monotonic with respect to w on $(0, \infty)$. Then Lemma 4 gives that $x \mapsto \vartheta_{\mu,\nu,a,b}(x)$ is the Laplace transform of a generalized gamma convolution. □

Proof of Theorem 17. Let $F(z) = e^{-(a+b)\sqrt{z}}z^{-\frac{\mu+\nu}{2}}I_\mu(a\sqrt{z})I_\nu(b\sqrt{z})$. As $z \to 0$,

$$z^{-\frac{\mu+\nu}{2}}I_\mu(a\sqrt{z})I_\nu(b\sqrt{z}) \sim \left(\tfrac{1}{2}\right)^{\mu+\nu}\frac{a^\mu b^\nu}{\Gamma(\mu+1)\Gamma(\nu+1)}$$

and hence $F(z) = o(|z|^{-1})$ as $|z| \to 0$. Using [NIST10, eq. 10.25.3] we obtain

$$F(z) \sim \frac{1}{2\pi\sqrt{ab}} z^{-\frac{\mu+\nu+1}{2}}$$

as $|z| \to \infty$, uniformly in every sector $|\arg z| \le \pi - \varepsilon$, $\varepsilon > 0$. Thus, conditions **a** and **b** of Lemma 5 are met. Now we use (5-8) to evaluate

$$F(-t-\mathrm{i}\eta) = F(e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta)) = e^{(a+b)\mathrm{i}\sqrt{t+\mathrm{i}\eta}}(t+\mathrm{i}\eta)^{-\frac{\mu+\nu}{2}} J_\mu(a\sqrt{t+\mathrm{i}\eta}) J_\nu(b\sqrt{t+\mathrm{i}\eta}),$$

$$F(-t+\mathrm{i}\eta) = F(e^{\mathrm{i}\pi}(t-\mathrm{i}\eta)) = e^{-(a+b)\mathrm{i}\sqrt{t-\mathrm{i}\eta}}(t-\mathrm{i}\eta)^{-\frac{\mu+\nu}{2}} J_\mu(a\sqrt{t-\mathrm{i}\eta}) J_\nu(b\sqrt{t-\mathrm{i}\eta}).$$

By using the asymptotic relation (5-7), we can validate the interchange of the integral and limit in (1-3). Then Lemma 7 gives

$$\alpha'(t) = \frac{1}{\pi} t^{-\frac{\mu+\nu}{2}} J_\mu(a\sqrt{t}) J_\nu(b\sqrt{t}) \sin((a+b)\sqrt{t}),$$

and thus $\alpha(t)$ is continuous and $\lim_{t\to 0^+} \alpha(t)$ exists. Letting $\tilde{\alpha}(t) = \alpha(t) - \alpha(0)$ we get the normalized measure with $\tilde{\alpha}'(t) = \alpha'(t)$. □

Proof of Theorem 18. From [Is90, Theorem 2] we know that if $\nu > \frac{1}{2}$, the function $x \mapsto 2^\nu \Gamma(\nu+1) x^{-\nu/2} I_\nu(\sqrt{x}) e^{-\sqrt{x}}$ is the Laplace transform of an infinitely divisible distribution that is not a generalized gamma convolution. Hence $\zeta_{\mu,\nu,a,b}(x)$ is the Laplace transform of an infinitely divisible distribution that is not a generalized gamma convolution, being a product of two such.

To prove infinite divisibility we start with the observation that $-\ln \zeta_{\mu,\nu,a,b}(x) \to 0$ as $x \to 0$ (see [NIST10, eq. 10.30.1]). Using the same idea as in the proof of Theorem 9 we obtain

$$\begin{aligned}
&-\frac{\ln \zeta_{\mu,\nu,a,b}(x)}{dx}(x) \\
&= \frac{a+b}{2\sqrt{x}} + \frac{\mu+\nu}{2x} - \frac{a}{2\sqrt{x}} \frac{I'_\mu(a\sqrt{x})}{I_\mu(a\sqrt{x})} - \frac{b}{2\sqrt{x}} \frac{I'_\nu(b\sqrt{x})}{I_\nu(b\sqrt{x})} \\
&= \frac{a}{2\sqrt{x}} - \sum_{n\ge 1} \frac{1}{x + j^2_{\mu,n} a^{-2}} + \frac{b}{2\sqrt{x}} - \sum_{n\ge 1} \frac{1}{x + j^2_{\nu,n} b^{-2}} \\
&= \left(\frac{1}{\pi} \int_0^\infty \frac{dt}{x+t^2 a^{-2}} - \sum_{n\ge 1} \frac{1}{x+j^2_{\mu,n} a^{-2}} \right) + \left(\frac{1}{\pi} \int_0^\infty \frac{dt}{x+t^2 b^{-2}} - \sum_{n\ge 1} \frac{1}{x+j^2_{\nu,n} b^{-2}} \right);
\end{aligned}$$

this is a sum of two completely monotonic functions on $(0, \infty)$ if $a, b > 0$ and $\mu, \nu > \frac{1}{2}$.

To show that $\zeta_{\mu,\nu,a,b}(x)$ is not a Laplace transform of a generalized gamma convolution we use Pick's characterization (Lemma 3). If $\psi(s) = \zeta_{\mu,\nu,a,b}(-s)$ is the moment generating function, we have $\operatorname{Im}\big(\psi'(s)/\psi(s)\big) < 0$ whenever $\operatorname{Im} s > 0$, as follows: With $s = re^{\mathrm{i}\theta} = x + \mathrm{i}y$, we can write

$$\begin{aligned}\frac{\psi'(s)}{\psi(s)} &= -\frac{(a+b)\mathrm{i}}{2\sqrt{s}} - \frac{\mu+\nu}{2s} + \frac{a}{2\sqrt{s}}\frac{J_\mu'(a\sqrt{s})}{J_\mu(a\sqrt{s})} + \frac{b}{2\sqrt{s}}\frac{J_\nu'(b\sqrt{s})}{J_\nu(b\sqrt{s})} \\ &= -\frac{(a+b)\mathrm{i}}{2\sqrt{s}} + \sum_{n\ge 1}\frac{a^2}{s-j_{\mu,n}^2} + \sum_{n\ge 1}\frac{b^2}{s-j_{\nu,n}^2} \\ &= \frac{\big(-\sin\frac{\theta}{2} - \mathrm{i}\cos\frac{\theta}{2}\big)(a+b)}{2\sqrt{r}} + \sum_{n\ge 1}\frac{a^2(x-j_{\mu,n}^2-\mathrm{i}y)}{(x-j_{\mu,n}^2)^2+y^2} + \sum_{n\ge 1}\frac{b^2(a-j_{\nu,n}^2-\mathrm{i}y)}{(x-j_{\nu,n}^2)^2+y^2},\end{aligned}$$

which is negative when $\operatorname{Im} s > 0$, as desired. □

Proof of Theorem 19. Define

$$F(z) = \frac{e^{-(a+b)\sqrt{z}}}{z^{\frac{\mu+\nu}{2}}K_\mu(a\sqrt{z})K_\nu(b\sqrt{z})}.$$

First take $\mu > 0$ and $\nu > 0$ and use [NIST10, eq. 10.30.2] to get

$$F(z) \sim \frac{a^\mu b^\nu e^{-(a+b)\sqrt{z}}}{2^{\mu+\nu-2}\Gamma(\mu)\Gamma(\nu)} \quad \text{as } z \to 0.$$

Then take $\mu > 0$ and $\nu = 0$ and use [NIST10, eq. 10.30.3] to get

$$F(z) \sim -\frac{a^\mu e^{-(a+b)\sqrt{z}}}{2^{\mu-1}\Gamma(\mu)\ln(b\sqrt{z})} \quad \text{as } z \to 0.$$

When $\mu > 0$ and $\nu < 0$, we use [NIST10, eq. 10.30.2] and $K_\nu(z) = K_{-\nu}(z)$ to get

$$F(z) \sim \frac{e^{-(a+b)\sqrt{z}}a^\mu b^{-\nu}z^{-\nu}}{2^{\mu-\nu-2}\Gamma(\mu)\Gamma(-\nu)} \quad \text{as } z \to 0.$$

To sum up, $F(z) = o(|z|^{-1})$ as $|z| \to 0$ whenever $\mu > 0$ and $\nu \in \mathbb{R}$, and by interchanging ν and μ we can conclude that in fact $F(z) = o(|z|^{-1})$ as $|z| \to 0$ whenever $\mu \in \mathbb{R}$, $\nu \in \mathbb{R}$ and $\mu + \nu > 1$. Using [NIST10, eq. 10.25.3], we also have

$$F(z) \sim \frac{2\sqrt{ab}z^{\frac{1-(\mu+\nu)}{2}}}{\sqrt{\pi}} \quad \text{as } |z| \to \infty$$

for $\mu + \nu > 1$. Hence, $F(z) = o(1)$ as $|z| \to \infty$ in the region $|\arg z| < \frac{3}{2}\pi$, showing that condition **b** in Lemma 6 is met.

$K_\mu(z)$ has no zeros in $|\arg z| \le \frac{\pi}{2}$ and has finitely many zeros in $\mathbb{C} \setminus \Omega$, where $\Omega = \big\{z : |\arg z| < \frac{\pi}{2}\big\}$. Consequently, $K_\mu(a\sqrt{z})$ has no zeros in $|\arg z| \le \pi$ and has finitely many zeros in $\pi < |\arg z| < 2\pi$. Hence, we can find a $\theta \in \big(\frac{2}{3}, 1\big)$ such that the function $F(z)$ is analytic in the region $|\arg| < \pi/\theta$, showing that condition **a** of Lemma 6 is also met.

Using the residue theorem, we evaluate the contour integral in (1-2) to

$$\varrho(t) = \frac{1}{2\pi \mathrm{i}} \int_C \frac{z e^{\frac{z}{2}}}{z^2+\pi^2} F(e^z t)\, dz = \frac{\mathrm{i}}{2}\big(F(te^{\mathrm{i}\pi}) - F(te^{-\mathrm{i}\pi})\big),$$

where C is a rectifiable closed curve going around $[-\mathrm{i}\pi, \mathrm{i}\pi]$ in the positive direction and lying in the strip $|\mathrm{Im}\, z| < \pi/\theta$. On the other hand, in view of (5-8), we have

$$F(te^{\mathrm{i}\pi}) = \frac{-4e^{-(a+b)\mathrm{i}\sqrt{t}}}{\pi^2 t^{\frac{\mu+\nu}{2}} H_\mu^{(2)}(a\sqrt{t}) H_\nu^{(2)}(b\sqrt{t})},$$
$$F(te^{-\mathrm{i}\pi}) = \frac{-4e^{(a+b)\mathrm{i}\sqrt{t}}}{\pi^2 t^{\frac{\mu+\nu}{2}} H_\mu^{(1)}(a\sqrt{t}) H_\nu^{(1)}(b\sqrt{t})}.$$

All this leads to

$$\varrho(t) = \frac{4}{t^{\frac{\mu+\nu}{2}}\pi^2} \cdot \frac{{}_1T_{\mu,\nu,a,b}(t)\cos((a+b)\sqrt{t}) - {}_2T_{\mu,\nu,a,b}(t)\sin((a+b)\sqrt{t})}{\big(J_\mu^2(a\sqrt{t}) + Y_\mu^2(a\sqrt{t})\big)\big(J_\nu^2(b\sqrt{t}) + Y_\nu^2(b\sqrt{t})\big)}. \qquad \square$$

Proof of Theorem 20. From [Is90, Theorem 1] we know that if $\nu > \frac{1}{2}$, the function $x \mapsto 2^{\nu-1}\Gamma(\nu)x^{-\nu/2}e^{-\sqrt{x}}/K_\nu(\sqrt{x})$ is the Laplace transform of an infinitely divisible distribution that is a generalized gamma convolution. This implies that $\kappa_{\mu,\nu,a,b}(x)$ is the Laplace transform of an infinitely divisible distribution that is also a generalized gamma convolution, being a product of two such. (See [Bo15, Theorem 1] or [BB17, Proposition 7].)

For infinite divisibility we calculate

$$\begin{aligned}
-&\frac{d\ln\kappa_{\mu,\nu,a,b}(x)}{dx}\\
&= \frac{a+b}{2\sqrt{x}} + \frac{\mu+\nu}{2x} + \frac{a}{2\sqrt{x}}\frac{K_\mu'(a\sqrt{x})}{K_\mu(a\sqrt{x})} + \frac{b}{2\sqrt{x}}\frac{K_\nu'(b\sqrt{x})}{K_\nu(b\sqrt{x})}\\
&= \frac{a}{2\sqrt{x}} - \frac{a}{2\sqrt{x}}\frac{K_{\mu-1}(a\sqrt{x})}{K_\mu(a\sqrt{x})} + \frac{b}{2\sqrt{x}} - \frac{b}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})}\\
&= \frac{1}{\pi}\int_0^\infty \frac{1}{x+t^2a^{-2}}\left(1 - \frac{2(\pi t)^{-1}}{J_\mu^2(t)+Y_\mu^2(t)}\right)dt\\
&\qquad + \frac{1}{\pi}\int_0^\infty \frac{1}{x+t^2b^{-2}}\left(1 - \frac{2(\pi t)^{-1}}{J_\nu^2(t)+Y_\nu^2(t)}\right)dt.
\end{aligned}$$

which for $\mu > \frac{1}{2}$ implies, in view of the inequality $J_\mu^2(t) + Y_\mu^2(t) > 2/(\pi t)$ (see [Wa44, p. 447]), that $x \mapsto -d\ln\kappa_{\mu,\nu,a,b}(x)/dx$ is indeed completely monotonic on

$(0, \infty)$ for all $\mu, \nu > \frac{1}{2}$ and $a, b > 0$. Then Lemma 1 shows that $\kappa_{\mu,\nu,a,b}(x)$ is the Laplace transform of an infinitely divisible distribution. □

Proof of Theorem 21. Letting

$$F(z) = z^{-\frac{\mu+\nu}{2}} e^{-(a+b)\sqrt{z}} \frac{I_\mu(a\sqrt{z})}{K_\nu(b\sqrt{z})},$$

we go through the same manipulations as in the proof of Theorem 19 to obtain the asymptotics of F as $z \to 0$: when $\nu > 0$, use eqs. 10.30.1 and 10.30.2 of [NIST10]; when $\nu = 0$, use 10.30.1 and 10.30.3; when $\nu < 0$, use 10.30.1, 10.27.3 and 10.30.2. Finally, when $a, b > 0$ and $\mu + \nu > 0$ use 10.30.4 and 10.25.3 to obtain

$$F(z) \sim \sqrt{b/a}\,\frac{z^{-\frac{\mu+\nu}{2}}}{\pi} \quad \text{as } |z| \to \infty.$$

Thus, condition **b** of Lemma 5 holds true. $K_\nu(z)$ has no zeros in $|\arg z| \le \frac{\pi}{2}$, so neither does $K_\nu(b\sqrt{z})$ have zeros in $|\arg z| \le \pi$. Thus, condition **a** of Lemma 5 is also met. From (5-8) we then obtain

$$F(-t - \mathrm{i}\eta) = F(e^{-\mathrm{i}\pi}(t + \mathrm{i}\eta)) = -\frac{2\mathrm{i}(t+\mathrm{i}\eta)^{-\frac{\mu+\nu}{2}} e^{(a+b)\mathrm{i}\sqrt{t+\mathrm{i}\eta}}}{\pi} \frac{J_\mu(a\sqrt{t+\mathrm{i}\eta})}{H_\nu^{(1)}(b\sqrt{t+\mathrm{i}\eta})},$$

$$F(-t + \mathrm{i}\eta) = F(e^{\mathrm{i}\pi}(t - \mathrm{i}\eta)) = \frac{2\mathrm{i}(t-\mathrm{i}\eta)^{-\frac{\mu+\nu}{2}} e^{-(a+b)\mathrm{i}\sqrt{t-\mathrm{i}\eta}}}{\pi} \frac{J_\mu(a\sqrt{t-\mathrm{i}\eta})}{H_\nu^{(2)}(b\sqrt{t-\mathrm{i}\eta})}.$$

To validate the interchange of limit and integral in (1-3) we consider three cases as in the proof of Theorem 11:

<u>(i) $\nu > 0$ and $\mu > -1$</u>: From (5-7) and (5-9) we see that $F(-t - \mathrm{i}\eta) - F(-t + \mathrm{i}\eta)$ is continuous in every rectangle $[t_1, t_2] \times [0, \eta]$.

<u>(ii) $\nu = 0$ and $\mu > -1$</u>: From (5-7) and (5-10) we get, as $t \to 0$ and $\eta \to 0$,

$$F(-t - \mathrm{i}\eta) \sim -\frac{a^\mu}{2^\mu \Gamma(\mu+1)} \frac{1}{\ln(b\sqrt{t+\mathrm{i}\eta})},$$

$$F(-t + \mathrm{i}\eta) \sim -\frac{a^\mu}{2^\mu \Gamma(\mu+1)} \frac{1}{\ln(b\sqrt{t-\mathrm{i}\eta})}.$$

Thus, both functions are bounded in every rectangle $[t_1, t_2] \times [0, \eta]$.

<u>(iii) $\nu < 0$ and $\mu > -1$</u>: From (5-7), (5-9) and (5-11) we get, as $t \to 0$ and $\eta \to 0$,

$$F(-t - \mathrm{i}\eta) \sim \frac{e^{\nu\pi\mathrm{i}} a^\mu}{2^{\mu-\nu-1} b^\nu \Gamma(\mu+1)\Gamma(-\nu)} (t + \mathrm{i}\eta)^{-\nu},$$

$$F(-t + \mathrm{i}\eta) \sim \frac{e^{-\nu\pi\mathrm{i}} a^\mu}{2^{\mu-\nu-1} b^\nu \Gamma(\mu+1)\Gamma(-\nu)} (t - \mathrm{i}\eta)^{-\nu};$$

that is, both functions are bounded in every rectangle $[t_1, t_2] \times [0, \eta]$.

Finally, in view of Lemma 7, we obtain

$$\frac{d\alpha(t)}{dt} = -\frac{2t^{-\frac{\mu+\nu}{2}}}{\pi^2} J_\mu(a\sqrt{t})\, \frac{\cos((a+b)\sqrt{t})J_\nu(b\sqrt{t}) + \sin((a+b)\sqrt{t})Y_\nu(b\sqrt{t})}{J_\nu^2(b\sqrt{t}) + Y_\nu^2(b\sqrt{t})}.$$

□

Proof of Corollary 2. We can simply take $a \to 0$ in Theorem 21. However, we give another proof using the Stieltjes transform representation in Lemma 6. Let

$$F(z) = (\sqrt{z})^{-\nu} e^{-b\sqrt{z}} \frac{1}{K_\nu(b\sqrt{z})}.$$

Equation 10.30.2 in [NIST10] gives $F(z) \sim o(|z|^{-1})$ as $z \to 0$, while 10.25.3 gives $F(z) \sim (\sqrt{z})^{-\nu+\frac{1}{2}}\sqrt{2b/\pi}$ as $|z| \to \infty$. Hence condition **b** in Lemma 6 holds true.

Next, $K_\nu(z)$ has no zero in the region $|\arg z| \le \frac{\pi}{2}$ and it has finitely many zeros in the regions $\frac{\pi}{2} < \arg z < \pi$ and $-\pi < \arg z < -\frac{\pi}{2}$. Hence $F(z)$ is analytic in $|\arg z| \le \pi$. We can find $\epsilon \in \left(\frac{2}{3}, 1\right)$ such that $F(z)$ is analytic in the region $|\arg z| < \frac{\pi}{\epsilon}$, so condition **a** in Lemma 6 is also met. In view of Lemma 7 we obtain

$$\frac{d\alpha(t)}{dt} = \frac{1}{2\pi \mathrm{i}}\big(F(e^{-\mathrm{i}\pi}t) - F(e^{\mathrm{i}\pi}t)\big),$$

where

$$F(e^{-\mathrm{i}\pi}t) = e^{\mathrm{i}\frac{\pi}{2}\nu}(\sqrt{t})^{-\nu} e^{\mathrm{i}b\sqrt{t}} \frac{1}{K_\nu(e^{-\mathrm{i}\frac{\pi}{2}}b\sqrt{t})} = \frac{2(\sqrt{t})^{-\nu}}{\mathrm{i}\pi} e^{\mathrm{i}b\sqrt{t}} \frac{J_\nu(b\sqrt{t}) - \mathrm{i}Y_\nu(b\sqrt{t})}{J_\nu^2(b\sqrt{t}) + Y_\nu^2(b\sqrt{t})},$$

$$F(e^{\mathrm{i}\pi}t) = e^{-\mathrm{i}\frac{\pi}{2}\nu}(\sqrt{t})^{-\nu} e^{-\mathrm{i}b\sqrt{t}} \frac{1}{K_\nu(e^{\mathrm{i}\frac{\pi}{2}}b\sqrt{t})} = -\frac{2(\sqrt{t})^{-\nu}}{\mathrm{i}\pi} e^{-\mathrm{i}b\sqrt{t}} \frac{J_\nu(b\sqrt{t}) + \mathrm{i}Y_\nu(b\sqrt{t})}{J_\nu^2(b\sqrt{t} + Y_\nu^2(b\sqrt{t}))}.$$

This leads to

$$\frac{d\alpha(t)}{dt} = -\frac{2}{\pi^2}(\sqrt{t})^{-\nu} \frac{J_\nu(b\sqrt{t})\cos(b\sqrt{t}) + Y_\nu(b\sqrt{t})\sin(b\sqrt{t})}{J_\nu^2(b\sqrt{t}) + Y_\nu^2(b\sqrt{t})}$$

□

Proof of Theorem 22. Applying Theorems 1 and 2 of [Is90] in the same way as in the proof of Theorem 18 shows that $\varepsilon_{\mu,\nu,a,b}(x)$ is the Laplace transform of an infinitely divisible distribution. More precisely, we use that

$$\begin{aligned}
-\frac{d\ln\varepsilon_{\mu,\nu,a,b}(x)}{dx}(x) &= \frac{a+b}{2\sqrt{x}} + \frac{\mu+\nu}{2x} - \frac{a}{2\sqrt{x}}\frac{I'(a\sqrt{x})}{I_\mu(a\sqrt{x})} + \frac{b}{2\sqrt{x}}\frac{K_\nu'(b\sqrt{x})}{K_\nu(b\sqrt{x})} \\
&= \frac{a+b}{2\sqrt{x}} - \frac{a}{2\sqrt{x}}\frac{I_{\mu+1}(a\sqrt{x})}{I_\mu(a\sqrt{x})} - \frac{b}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})} \\
&= \frac{1}{\pi}\int_0^\infty \frac{1}{x+b^{-2}t^2}\left(1 - \frac{2(\pi t)^{-1}}{J_\nu^2(t) + Y_\nu^2(t)}\right) dt \\
&\qquad + \frac{1}{\pi}\int_0^\infty \frac{dt}{x+a^{-2}t^2} - \sum_{n\ge1}\frac{1}{x+a^{-2}j_{\mu,n}^2}
\end{aligned}$$

is completely monotonic on $(0, \infty)$ for all $\mu, \nu > \frac{1}{2}$ and $a, b > 0$, being a sum of two completely monotonic functions: the expression on the penultimate line is completely monotonic because of the inequality $J_\nu^2(t) + Y_\nu^2(t) > 2/(\pi t)$ where $\nu > \frac{1}{2}$ (see [Wa44, p. 447]), while the expression on the last line satisfies

$$\begin{aligned}\Theta(x) &= \frac{1}{\pi}\int_0^\infty \frac{dt}{x + a^{-2}t^2} - \sum_{n\geq 1} \frac{1}{x + a^{-2} j_{\mu,n}^2} \\ &\geq \frac{1}{\pi}\sum_{n\geq 1} \frac{j_{\mu,n} - j_{\mu,n-1}}{x + a^{-2} j_{\mu,n}^2} - \sum_{n\geq 1} \frac{1}{x + a^{-2} j_{\mu,n}^2} \geq 0\end{aligned}$$

and

$$\frac{(-1)^m}{m!}\frac{d^m\Theta(x)}{dx^m} \geq \frac{1}{\pi}\sum_{n\geq 1} \frac{j_{\mu,n} - j_{\mu,n-1}}{(x + a^{-2} j_{\mu,n}^2)^{m+1}} - \sum_{n\geq 1} \frac{1}{(x + a^{-2} j_{\mu,n}^2)^{m+1}} \geq 0,$$

since $j_{\mu,n} - j_{\mu,n-1} > \pi$ whenever $\mu > \frac{1}{2}$ and $n \in \mathbb{N}$. In view of Lemma 1, this proves the first part of Theorem 22.

For the second part we observe, using a similar approach as before, that the corresponding expression

$$\begin{aligned}-\frac{d}{dx}\ln\frac{e^{-(a+b)\sqrt{x}}}{\varepsilon_{\mu,\nu,a,b}(x)} &= -\frac{\nu+\mu}{2x} - \frac{b}{2\sqrt{x}}\frac{K_\nu'(b\sqrt{x})}{K_\nu(b\sqrt{x})} + \frac{a}{2\sqrt{x}}\frac{I_\mu'(a\sqrt{x})}{I_\mu(a\sqrt{x})} \\ &= \frac{b}{2\sqrt{x}}\frac{K_{\nu-1}(b\sqrt{x})}{K_\nu(b\sqrt{x})} + \frac{a}{2\sqrt{x}}\frac{I_{\mu+1}(a\sqrt{x})}{I_\mu(a\sqrt{x})} \\ &= \frac{2}{\pi^2}\int_0^\infty \frac{1}{x + b^{-2}t^2}[J_\nu^2(t) + Y_\nu^2(t)]^{-1}dt + \sum_{n\geq 1}\frac{1}{x + a^{-2} j_{\mu,n}^2}\end{aligned}$$

is completely monotonic on $(0, \infty)$, being a sum of two completely monotonic functions whenever $a, b, \nu > 0$ and $\mu > -1$. □

Proof of Theorem 23. Let X and Y be independent gamma variables with parameters (α, β) and (α_0, β_0). The probability density function of the quotient $Z = X/Y$ is

$$f(z) = \frac{\Gamma(\alpha+\alpha_0)}{\Gamma(\alpha)\Gamma(\alpha_0)}\Big(\frac{\beta_0}{\beta}\Big)^\alpha x^{\alpha-1}\Big(1 + \frac{\beta_0}{\beta}x\Big)^{-(\alpha+\alpha_0)},$$

where $x > 0$ (see for example [IK79, p. 889]). Let $a = \alpha$ and $c = 1 - \alpha_0$. The Laplace transform $L(s)$ of the quotient of two gamma random variables is

$$L(s) = \frac{\Gamma(a-c+1)}{\Gamma(1-c)}\psi(a, c, s)$$

(see [IK79, p.889]) and thus the moment generating function $\phi(s) = L(-s)$ is

$$\phi(s) = \frac{\Gamma(a-c+1)}{\Gamma(1-c)}\psi(a, c, -s). \tag{5-13}$$

Since $a>0$, by using [NIST10, eq. 13.9.13], we observe that $\psi(a,c,-s)$ has no zeros in $\mathbb{C}\setminus[0,\infty)$. Thus the first condition of the Pick function characterization theorem (Lemma 3) is verified, and by taking the logarithmic derivative of both sides of (5-13), we obtain

$$\frac{\phi'(s)}{\phi(s)}=-\frac{\psi'(a,c,-s)}{\psi(a,c,-s)}=\frac{a\psi(a+1,c+1,-s)}{\psi(a,c,-s)}.$$

For $s=x+\mathrm{i}y$, by using the integral representation (4-1), which is valid for $|\arg z|<\pi$, $a>0$ and $c<1$, we arrive at

$$\begin{aligned}\frac{\phi'(s)}{\phi(s)}&=\int_0^\infty\frac{at^{-c}e^{-t}|\psi(a,c,te^{\mathrm{i}\pi})|^{-2}\,dt}{(-s+t)\Gamma(a+1)\Gamma(a-c+1)}\\&=\int_0^\infty\frac{at^{-c}e^{-t}|\psi(a,c,te^{\mathrm{i}\pi})|^{-2}\,dt}{(-x-\mathrm{i}y+t)\Gamma(a+1)\Gamma(a-c+1)}\\&=\int_0^\infty\frac{(t-x+\mathrm{i}y)at^{-c}e^{-t}|\psi(a,c,te^{\mathrm{i}\pi})|^{-2}\,dt}{((t-x)^2+y^2)\Gamma(a+1)\Gamma(a-c+1)}.\end{aligned}$$

This shows that $\operatorname{Im}\phi'(s)/\phi(s)\geq 0$ whenever $\operatorname{Im}s>0$. □

Proof of Theorem 24. By using equations (12)–(15) of [EMOT53a, p. 258] and the equality $\psi'(a,c,z)=-a\psi(a+1,c+1,z)$, we obtain

$$\begin{aligned}\psi(a,c-1,z)&=\frac{1-c}{a-c+1}\psi(a,c,z)+\frac{az}{a-c+1}\psi(a+1,c+1,z),\\ \psi(a+1,c,z)&=\frac{1}{a-c+1}\psi(a,c,z)-\frac{z}{a-c+1}\psi(a+1,c+1,z),\\ \psi(a,c+1,z)&=\psi(a,c,z)+a\psi(a+1,c+1,z),\\ \psi(a-1,c,z)&=z\psi(a,c,z)-(c-a)\psi(a,c,z)+az\phi(a+1,c+1,z).\end{aligned}$$

Dividing both sides of these equations by $\psi(a,c,z)$ and using the integral representation (4-1) yields the required results.

Thus, the representations in Theorem 24 follow naturally from (4-1), but for (4-3) we give a more detailed proof via the Stieltjes representation and inversion theorems. For this let

$$F(z)=\frac{\psi(a,c+1,z)}{\psi(a,c,z)}-1.$$

Using eq. 13.7.3 of [NIST10], we have

$$F(z)\sim\frac{\sum_{s\geq 1}\dfrac{(a)_s((a-c)_s-(a-c+1)_s)(-z)^{-s}}{s!}}{\sum_{s\geq 0}\dfrac{(a)_s(a-c+1)_s(-z)^{-s}}{s!}}\quad\text{as } z\to\infty,$$

and $|\arg z| < \frac{3}{2}\pi$, showing that $f(z) \to 0$ as $|z| \to \infty$. Next, for $c < -1$, using eq. 13.2.22, we have

$$F(z) \sim \frac{\Gamma(-c)/\Gamma(a-c)}{\Gamma(1-c)/\Gamma(a-c+1)} - 1 \quad \text{as } z \to 0;$$

for $c = -1$, eqs. 13.2.21 and 13.2.22 give

$$F(z) \sim \frac{1/\Gamma(a+1)}{\Gamma(1-c)/\Gamma(a-c+1)} - 1 \quad \text{as } z \to 0;$$

for $-1 < c < 0$, eqs. 13.2.20 and 13.2.22 yield

$$F(z) \sim \frac{\Gamma(-c)/\Gamma(a-c)}{\Gamma(1-c)/\Gamma(a-c+1)} - 1 \quad \text{as } z \to 0;$$

for $c = 0$, eqs. 13.2.19 and 13.2.21 lead to

$$F(z) \sim \frac{-1/\Gamma(a)(\ln(z)+d)}{1/\Gamma(a+1)} - 1 \quad \text{as } z \to 0,$$

(where d is the constant in eq. 13.2.19), and for $0 < c < 1$, eqs. 13.2.18 and 13.2.20 give

$$F(z) \sim \frac{\Gamma(c)/\Gamma(a)z^{-c} + \Gamma(-c)/\Gamma(a-c)}{\Gamma(1-c)/\Gamma(a-c+1)} \quad \text{as } z \to 0.$$

Thus $F(z) = o(|z|^{-1})$ as $z \to 0$, for $a > 0$ and $c < 1$. The conditions in Lemma 6 have been verified, because $\psi(a, c, z)$ has no zeros in $|\arg z| < \pi/\alpha$, where $\alpha \in \left(\frac{2}{3}, 1\right)$. By using the residue calculus we obtain

$$\begin{aligned}
\frac{d}{dt}\alpha(t) &= \frac{1}{2\pi \mathrm{i}}\big(F(te^{-\mathrm{i}\pi}) - F(te^{\mathrm{i}\pi})\big) \\
&= \frac{1}{2\pi \mathrm{i}} \lim_{\eta\to 0^+}\big(F(-t-\mathrm{i}\eta) - F(-t+\mathrm{i}\eta)\big) \\
&= \frac{1}{2\pi \mathrm{i}} \lim_{\eta\to 0^+} \frac{\begin{array}{c}\psi(a,c,e^{\mathrm{i}\pi}(t-\mathrm{i}\eta))\psi(a,c+1,e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta)) \\ -\,\psi(a,c,e^{-\mathrm{i}\pi}(t+\mathrm{i}\eta))\psi(a,c+1,e^{\mathrm{i}\pi}(t-\mathrm{i}\eta))\end{array}}{\big|\psi(a,c,e^{\mathrm{i}\pi}(t-\mathrm{i}\eta))\big|^2}.
\end{aligned}$$

Now, by using [EMOT53a, eq. (14), p. 263], for $-a < \min\{0, 1-c\}$ we have

$$\lim_{\eta\to 0^+} \psi(a, c, e^{\pm \mathrm{i}\pi}(t \mp \mathrm{i}\eta)) = k_1 y_1(-t) - e^{\mp \mathrm{i}\pi c} k_2 y_2(-t), \tag{5-14}$$

where

$$y_1(x) = \Phi(a, c, x) \quad \text{and} \quad y_2(x) = x^{1-c}\Phi(a-c+1, 2-c, x),$$

$\Phi(a, c, x)$ being the Kummer confluent hypergeometric function and

$$k_1 = \frac{\Gamma(1-c)}{\Gamma(a-c+1)}, \quad k_2 = (-1)^{1-c}\frac{\Gamma(c-1)}{\Gamma(a)}.$$

On the other hand, from [EMOT53a, eq. (9), p. 253] we have

$$W[y_1(x), y_2(x)] = y_1(x)y_2'(x) - y_2(x)y_1'(x) = (1-c)x^{-c}e^x \tag{5-15}$$

and by using the recurrence relation $\psi(a, c+1, x) = -\psi'(a, c, x) + \psi(a, c, x)$ (see [EMOT53a, eq. (14), p. 258]) and equation (5-14), we obtain

$$\begin{aligned}\lim_{\eta\to 0^+} &\psi(a, c+1, e^{\mp i\pi}(t \pm i\eta)) \\ &= k_1 y_1'(-t) - e^{\pm i\pi c} k_2 y_2'(-t) + k_1 y_1(-t) - e^{\pm i\pi c} k_2 y_2(-t).\end{aligned} \tag{5-16}$$

From (5-16), (5-15) and (5-14) it follows that

$$\frac{d\alpha(t)}{dt} = -\frac{\sin(\pi c)k_1k_2(y_1(-t)y_2'(-t) - y_2(-t)y_1'(-t))}{\pi|\psi(a, c, e^{i\pi}t)|^2}$$

and by using the formula $\sin(\pi c) = \dfrac{\pi}{\Gamma(c-1)\Gamma(2-c)}$ (see [IK79, p. 890]) we arrive at

$$\frac{d\alpha(t)}{dt} = \frac{t^{-c}e^{-t}\,|\psi(a, c, e^{i\pi}t)|^{-2}}{\Gamma(a)\Gamma(a-c+1)}.$$

This completes the proof of (4-3). □

Proof of Theorem 25. From [MOS66, p. 90] we have this integral representation for the product of modified Bessel functions of the first kind, where $\mu > -\frac{1}{2}$:

$$\begin{aligned}&I_\mu(a)I_\mu(b) = \\ &\frac{\left(\frac{1}{2}ab\right)^\mu}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}+\mu\right)}\int_0^\pi (a^2+b^2-2ab\cos t)^{-\frac{1}{2}\mu} I_\mu\left((a^2+b^2-2ab\cos t)^{1/2}\right)\sin^{2\mu} t\,dt,\end{aligned} \tag{5-17}$$

Combining this with

$$\chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v) = \tfrac{1}{4}e^{-\lambda}(u/\lambda)^{\frac{\mu}{2}-1}e^{-\frac{u}{2}w}I_{\frac{\mu}{2}-1}(\sqrt{\lambda uv})I_{\frac{\mu}{2}-1}(\sqrt{\lambda u/v}).$$

and using the notation $T = \sqrt{\lambda u(w - 2\cos t)}$, we obtain

$$\chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v) = c_{\mu,\lambda}(u)e^{-\frac{u}{2}w}\int_0^\pi \frac{I_{\frac{\mu}{2}-1}(T)}{T^{\frac{\mu}{2}-1}}\sin^{\mu-2} t\,dt, \tag{5-18}$$

with

$$c_{\mu,\lambda}(u) = \frac{e^{-\lambda}u^{(\mu-2)}}{2^{\frac{\mu}{2}+1}\sqrt{\pi}\,\Gamma\left(\frac{\mu-1}{2}\right)}.$$

Since $2dT/dw = \lambda u/T$ and by the recurrence relation $(z^{-\mu}I_\mu(z))' = z^{-\mu}I_{\mu+1}(z)$ (see for example [Wa44, p. 79]), after differentiating both sides of (5-18) with respect to $w = v + 1/v$, we arrive at

$$\frac{d}{dw}\left(\chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v)\right)$$

$$= c_{\mu,\lambda}(u)\bigg(-\frac{1}{2}ue^{-\frac{uw}{2}}\int_0^\pi \frac{I_{\frac{\mu}{2}-1}(T)}{T^{\frac{\mu}{2}-1}}\sin^{\mu-2}t\,dt + e^{-\frac{uw}{2}}\int_0^\pi \frac{\lambda u}{2}\frac{I_{\frac{\mu}{2}}(T)}{T^{\frac{\mu}{2}}}\sin^{\mu-2}t\,dt\bigg)$$

$$= c_{\mu,\lambda}(u)\bigg(\frac{u}{2}e^{-\frac{uw}{2}}\int_0^\pi \frac{\sin^{\mu-2}t}{T^{\frac{\mu}{2}-1}}\Big(-I_{\frac{\mu}{2}-1}(T)+\lambda\frac{I_{\frac{\mu}{2}}(T)}{T}\Big)\,dt\bigg).$$

The recurrence relation $(2\mu/z)I_\mu(z) = I_{\mu-1}(z) - I_{\mu+1}(z)$ (see [Wa44, p. 79]) yields

$$\frac{\lambda}{T}I_{\frac{\mu}{2}}(T) = \frac{\lambda}{\mu}I_{\frac{\mu}{2}-1}(T) - \frac{\lambda}{\mu}I_{\frac{\mu}{2}+1}(T),$$

which in turn gives

$$\frac{d}{dw}\big(\chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v)\big) = \frac{\frac{1}{2}uc_{\mu,\lambda}(u)}{e^{\frac{u}{2}w}}\int_0^\pi \frac{\sin^{\mu-2}t}{T^{\frac{\mu}{2}-1}}\bigg(\Big(\frac{\lambda}{\mu}-1\Big)I_{\frac{\mu}{2}-1}(T) - \frac{\lambda}{\mu}I_{\frac{\mu}{2}+1}(T)\bigg)\,dt < 0$$

whenever $0 < \lambda \le \mu$ and $u > 0$. By using a similar approach we obtain

$$\begin{aligned}\frac{d^2}{dw^2}\big(\chi_{\mu,\lambda}(uv)\chi_{\mu,\lambda}(u/v)\big) &= c_{\mu,\lambda}(u)\bigg[-\frac{1}{4}u^2e^{-\frac{u}{2}w}\int_0^\pi \sin^{\mu-2}t\Big(-\frac{I_{\frac{\mu}{2}-1}(T)}{T^{\frac{\mu}{2}-1}}+\lambda\frac{I_{\frac{\mu}{2}}(T)}{T^{\frac{\mu}{2}}}\Big)\,dt \\ &\qquad + \frac{u}{2}e^{-\frac{u}{2}w}\int_0^\pi \sin^{\mu-2}t\Big(-\frac{\lambda u}{2}\frac{I_{\frac{\mu}{2}}(T)}{T^{\frac{\mu}{2}}}+\frac{\lambda^2u}{2}\frac{I_{\frac{\mu}{2}+1}(T)}{T^{\frac{\mu}{2}+1}}\Big)\,dt\bigg] \\ &= c_{\mu,\lambda}(u)\cdot\frac{u^2}{4}e^{-\frac{u}{2}w}\int_0^\pi \frac{\sin^{\mu-2}t}{T^{\frac{\mu}{2}-1}}\Big(I_{\frac{\mu}{2}-1}(T)-\frac{2\lambda}{T}I_{\frac{\mu}{2}}(T)+\frac{\lambda^2}{T^2}I_{\frac{\mu}{2}+1}(T)\Big)\,dt,\end{aligned}$$

which can be rewritten as

$$\frac{u^2c_{\mu,\lambda}(u)}{4e^{\frac{u}{2}w}}\int_0^\pi \frac{\sin^{\mu-2}t}{T^{\frac{\mu}{2}-1}}\bigg(\Big(1-\frac{2\lambda}{\mu}\Big)I_{\frac{\mu}{2}-1}(T)+\frac{2\lambda}{\mu}I_{\frac{\mu}{2}+1}(T)+\frac{\lambda^2}{\mu^2}I_{\frac{\mu}{2}+1}(T)\bigg)\,dt$$

and this is strictly positive whenever $0 < \lambda \le \mu/2$ and $u > 0$. □

Proof of Theorem 26. If $t \in \big(0, \frac{\pi}{2}\big]$ and $a, b > 0$, then $a^2+b^2-2ab\cos t > (a-b)^2 > 0$; if $t \in \big(\frac{\pi}{2}, \pi\big)$ and $a, b > 0$, we likewise have $a^2+b^2-2ab\cos t > 0$. Inserting $a = uv$ and $b = u/v$ in the integral representation (5-17), we obtain

$$I_\mu(uv)I_\mu(u/v) = \frac{\big(\frac{1}{2}u^2\big)^\mu}{\sqrt{\pi}\,\Gamma\big(\frac{1}{2}+\mu\big)}\int_0^\pi f_\mu(S)\sin^{2\mu}t\,dt,$$

where $f_\mu(S) = S^{-\mu}I_\mu(S)$ and $S = \sqrt{u^2((w^2-2)-2\cos t)} > 0$. The recurrence relation $(z^{-\mu}I_\mu(z))' = z^{-\mu}I_{\mu+1}(z)$ (see [Wa44, p. 79]) implies that each of

$$\frac{d}{dw}f_\mu(S)=u^2wf_{\mu+1}(S),\qquad \frac{d^2}{dw^2}f_\mu(S)=u^2f_{\mu+1}(S)+(u^2w)^2f_{\mu+2}(S),$$
$$\frac{d^3}{dw^3}f_\mu(S)=3wu^4f_{\mu+2}(S)+(u^2w)^3f_{\mu+3}(S),$$
$$\frac{d^4}{dw^4}f_\mu(S)=3u^4f_{\mu+2}(S)+6u^6w^2f_{\mu+3}(S)+(u^2w)^4f_{\mu+4}(S)$$

is positive for all $a, b>0$, $w>2$ and $\mu>-\frac{1}{2}$. In view of these relations we make the induction hypothesis that the $(2n+1)$-st derivative is of the form

$$\frac{d^{2n+1}}{dw^{2n+1}}f_\mu(S)=\\ (u^2w)^{2n+1}f_{\mu+2n+1}(S)+\alpha_{2n}(u)w^{2n-1}f_{\mu+2n}(S)+\cdots+\alpha_2(u)wf_{\mu+k}(S)$$

and this expression is positive, where $k\le 2n$ and $\alpha_{2n}(u)$, $\alpha_{2n-2}(u)$, …, $\alpha_2(u)$ are nonnegative constants. We also make the induction hypothesis that the $2n$-th derivative is of the form

$$\frac{d^{2n}}{dw^{2n}}f_\mu(S)=(u^2w)^{2n}f_{\mu+2n}(S)+\beta_{2n}(u)w^{2n-1}f_{\mu+2n-1}(S)+\cdots+\beta_2(u)f_{\mu+k}(S)$$

and this expression is positive, where $k\le 2n-1$ and $\beta_{2n}(u)$, $\beta_{2n-2}(u)$, …, $\beta_2(u)$ are nonnegative constants. By using the recurrence relation $df_\mu(S)/dw=u^2wf_{\mu+1}(S)$ repeatedly we see that the derivatives $d^{2n+3}f_\mu(S)/dw^{2n+3}$ and $d^{2n+2}f_\mu(S)/dw^{2n+2}$ have a similar form as $d^{2n+1}f_\mu(S)/dw^{2n+1}$ and $d^{2n}f_\mu(S)/dw^{2n}$, respectively, and that each of these expressions is positive. Consequently, for all $\mu>-\frac{1}{2}$, $u, v>0$, $w>2$ and $n\in\mathbb{N}$ we have

$$\frac{d^n}{dw^n}\big(I_\mu(uv)I_\mu(u/v)\big)=\frac{\left(\frac{1}{2}u^2\right)^\mu}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}+\mu\right)}\int_0^\pi\frac{d^n}{dw^n}f_\mu(S)\sin^{2\mu}t\,dt\ge 0.\qquad\square$$

Acknowledgments. The research of Dhivya Prabhu K was supported in part by the Council of Scientific and Industrial Research India (Grant No. 09/1022(11054)/2021-EMR-I) and in part by a short-term Fellowship of the Indian Institute of Technology Indore, and was completed during his visit between March and May 2024 to Babeş-Bolyai University of Cluj-Napoca. This author is grateful to the Department of Business Administration, Extension Sfântu-Gheorghe, Faculty of Economics and Business Administration, Babeş-Bolyai University of Cluj-Napoca for hospitality.

References

[AB23] G. M. Assefa and A. Baricz, "Infinite divisibility of the Whittaker distribution", *Proc. Amer. Math. Soc.* **151**:12 (2023), 5429–5442. MR

[Ba09] A. Baricz, "On a product of modified Bessel functions", *Proc. Amer. Math. Soc.* **137**:1 (2009), 189–193. MR

[Ba10] A. Baricz, “Bounds for modified Bessel functions of the first and second kinds”, *Proc. Edinb. Math. Soc.* (2) **53**:3 (2010), 575–599. MR

[BB17] A. Behme and L. Bondesson, “A class of scale mixtures of Gamma(k)-distributions that are generalized gamma convolutions”, *Bernoulli* **23**:1 (2017), 773–787. MR

[BBH78] O. Barndorff-Nielsen, P. Blaesild, and C. Halgreen, “First hitting time models for the generalized inverse Gaussian distribution”, *Stochastic Process. Appl.* **7**:1 (1978), 49–54. MR

[BH77] O. Barndorff-Nielsen and C. Halgreen, “Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions”, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **38**:4 (1977), 309–311. MR

[BMPS16] A. Baricz, D. Jankov Maširević, S. Ponnusamy, and S. Singh, “Bounds for the product of modified Bessel functions”, *Aequationes Math.* **90**:4 (2016), 859–870. MR

[Bo15] L. Bondesson, “A class of probability distributions that is closed with respect to addition as well as multiplication of independent random variables”, *J. Theoret. Probab.* **28**:3 (2015), 1063–1081. MR

[Bo79] L. Bondesson, “A general result on infinite divisibility”, *Ann. Probab.* **7**:6 (1979), 965–979. MR

[Bo81] L. Bondesson, “Classes of infinitely divisible distributions and densities”, *Z. Wahrsch. Verw. Gebiete* **57**:1 (1981), 39–71. MR

[Bo92] L. Bondesson, *Generalized gamma convolutions and related classes of distributions and densities*, Lecture Notes in Statistics **76**, Springer, 1992. MR

[BPV11] A. Baricz, S. Ponnusamy, and M. Vuorinen, “Functional inequalities for modified Bessel functions”, *Expo. Math.* **29**:4 (2011), 349–414. MR

[BS15] P. Bosch and T. Simon, “On the infinite divisibility of inverse beta distributions”, *Bernoulli* **21**:4 (2015), 2552–2568. MR

[BS16] P. Bosch and T. Simon, “A proof of Bondesson’s conjecture on stable densities”, *Ark. Mat.* **54**:1 (2016), 31–38. MR

[BSMKR06] P. S. Bithas, N. C. Sagias, P. T. Mathiopoulos, G. K. Karagiannidis, and A. A. Rontogiannis, “On the performance analysis of digital communications over generalized-K fading channels”, *IEEE Commun. Lett.* **10**:5 (2006), 353–355.

[CK11] N. D. Chatzidiamantis and G. Karagiannidis, “On the distribution of the sum of gamma-gamma variates and applications in RF and optical wireless communications”, *IEEE Trans. Comm.* **59** (2011), 1298–1308.

[EMOT53a] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions*, vol. 1, McGraw-Hill, New York, 1953. MR

[EMOT53b] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher transcendental functions*, vol. 1, McGraw-Hill, New York, 1953. MR

[EMOT54] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of integral transforms*, vol. 2, McGraw-Hill, New York, 1954. MR

[Fe66] W. Feller, *An introduction to probability theory and its applications*, vol. 2, Wiley, New York, 1966. MR

[FS23] R. A. C. Ferreira and T. Simon, “Convolution of beta prime distribution”, *Trans. Amer. Math. Soc.* **376**:2 (2023), 855–890. MR

[GHP78] M. J. Goovaerts, L. D’Hooge, and N. de Pril, “On the infinite divisibility of the ratio of two gamma-distributed variables”, *Stochastic Process. Appl.* **7**:3 (1978), 291–297. MR

[Ha75] H. Hankel, "Bestimmte Integrale mit Cylinderfunctionen", *Math. Ann.* **8** (1875), 453–470.

[Ha78] P. Hartman, "Difference equations: disconjugacy, principal solutions, Green's functions, complete monotonicity", *Trans. Amer. Math. Soc.* **246** (1978), 1–30. MR

[Ha79] C. Halgreen, "Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions", *Z. Wahrsch. Verw. Gebiete* **47**:1 (1979), 13–17. MR

[Ha79] P. Hartman, "Uniqueness of principal values, complete monotonicity of logarithmic derivatives of principal solutions, and oscillation theorems", *Math. Ann.* **241**:3 (1979), 257–281. MR

[HW55] I. I. Hirschman and D. V. Widder, *The convolution transform*, Princeton University Press, 1955. MR

[IK79] M. E. H. Ismail and D. H. Kelker, "Special functions, Stieltjes transforms and infinite divisibility", *SIAM J. Math. Anal.* **10**:5 (1979), 884–901. MR

[IM79] M. E. H. Ismail and C. P. May, "Special functions, infinite divisibility and transcendental equations", *Math. Proc. Cambridge Philos. Soc.* **85**:3 (1979), 453–464. MR

[IM82] M. E. H. Ismail and K. S. Miller, "An infinitely divisible distribution involving modified Bessel functions", *Proc. Amer. Math. Soc.* **85**:2 (1982), 233–238. MR

[Is77a] M. E. H. Ismail, "Bessel functions and the infinite divisibility of the Student t-distribution", *Ann. Probability* **5**:4 (1977), 582–585. MR

[Is77b] M. E. H. Ismail, "Integral representations and complete monotonicity of various quotients of Bessel functions", *Canadian J. Math.* **29**:6 (1977), 1198–1207. MR

[Is90] M. E. H. Ismail, "Complete monotonicity of modified Bessel functions", *Proc. Amer. Math. Soc.* **108**:2 (1990), 353–361. MR

[JP78] E. Jakeman and P. N. Pusey, "Significance of K-distributions in scattering experiments", *Phys. Rev. Lett.* **40**:9 (1978), 546–550.

[Ke78] J. Kent, "Some probabilistic properties of Bessel functions", *Ann. Probab.* **6**:5 (1978), 760–770. MR

[Kü02] R. Küstner, "Mapping properties of hypergeometric functions and convolutions of starlike or convex functions of order α", *Comput. Methods Funct. Theory* **2**:2 (2002), 597–610. MR

[La00] L. J. Landau, "Bessel functions: monotonicity and bounds", *J. London Math. Soc.* (2) **61**:1 (2000), 197–215. MR

[Mi36] S. C. Mitra, "On certain new connections between Legendre and Bessel Functions", *Math. Z.* **41**:1 (1936), 680–685. MR

[MOS66] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and theorems for the special functions of mathematical physics*, enlarged ed., Grundl. Math. Wissen. **52**, Springer, 1966. MR

[NIST10] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (editors), *NIST handbook of mathematical functions*, U.S. Department of Commerce (NIST), Washington, DC, 2010. MR

[OB12] F. Oberhettinger and L. Badii, *Tables of Laplace transforms*, Springer, 1973. MR

[PBM88] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and series*, *vol.* 2: *Special functions*, Second ed., Gordon and Breach, New York, 1988. MR

[PVG07] R. Penfold, J.-M. Vanden-Broeck, and S. Grandison, "Monotonicity of some modified Bessel function products", *Integral Transforms Spec. Funct.* **18**:1-2 (2007), 139–144. MR

[Re99] N. J. Redding, "Estimating the parameters of the K distribution in the intensity domain", technical report, Electronics Research Laboratory, Salisbury, SA, Australia, 1999.

[Se01] J. Segura, "Bounds on differences of adjacent zeros of Bessel functions and iterative relations between consecutive zeros", *Math. Comp.* **70**:235 (2001), 1205–1220. MR

[Se21] J. Segura, "Monotonicity properties for ratios and products of modified Bessel functions and sharp trigonometric bounds", *Results Math.* **76**:4 (2021), art. id. 221, 22 pp. MR

[SH03] F. W. Steutel and K. van Harn, *Infinite divisibility of probability distributions on the real line*, Monographs and Textbooks in Pure and Applied Mathematics **259**, Dekker, New York, 2004. MR

[St32] M. H. Stone, *Linear transformations in Hilbert space*, American Mathematical Society Colloquium Publications **15**, Amer. Math. Soc., 1932. MR

[St73] F. W. Steutel, "Some recent results in infinite divisibility", *Stochastic Process. Appl.* **1** (1973), 125–143. MR

[Th77] O. Thorin, "On the infinite divisibility of the Pareto distribution", *Scand. Actuar. J.* 1 (1977), 31–40. MR

[Wa44] G. N. Watson, *A treatise on the theory of Bessel functions*, Cambridge University Press, 1944. MR

Received October 8, 2025. Revised February 2, 2026.

ÁRPÁD BARICZ
DEPARTMENT OF ECONOMICS
BABEŞ-BOLYAI UNIVERSITY
CLUJ-NAPOCA 400591
ROMANIA

and

INSTITUTE OF APPLIED MATHEMATICS
ÓBUDA UNIVERSITY
BUDAPEST 1034
HUNGARY

bariczocsi@yahoo.com

DHIVYA PRABHU K
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY INDORE
INDORE 453552
INDIA

phd2001241004@iiti.ac.in

SANJEEV SINGH
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY INDORE
INDORE 453552
INDIA

snjvsngh@iiti.ac.in

ANTONY VIJESH V
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY INDORE
INDORE 453552
INDIA

vijesh@iiti.ac.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.315

ON $\mathbb{R}$-TREES, HOMOTOPIES, AND COVERING MAPS

JEREMY BRAZAS, GREGORY R. CONNER, PAUL FABEL AND CURTIS KENT

A map $p : E \to X$ has the *unique path lifting* property if every path in X, after a choice of an initial point, lifts uniquely to a path in E. We prove that if a group G acts on an $\mathbb{R}$-tree T in such a way that the quotient map $p : T \to T/G$ has the unique path lifting property, then the quotient space T/G does not contain a disc. As a consequence, we show that every map of manifolds with the unique path lifting property is a covering map. The proof requires a study of one-dimensional backtracking in paths. We show the surprising and counterintuitive result that the equivalence relation given by homotopies of paths rel. endpoints is generated by inserting and deleting one-dimensional backtracking.

1. Introduction

Actions of groups on $\mathbb{R}$-trees are a central tool in topology and geometric group theory. Paulin constructed essential actions of hyperbolic groups on $\mathbb{R}$-trees [18]. Culler and Vogtmann's outer space has a natural boundary consisting of actions of a free group on an $\mathbb{R}$-tree [3, Section 2]. Sela and Groves used isometric actions on $\mathbb{R}$-trees to study limit groups [13; 19]. Rips developed tools to understand finitely generated groups acting freely on $\mathbb{R}$-trees (see [4; 12]).

Our interest here is motivated by considering non-finitely generated groups acting on $\mathbb{R}$-trees. Berestovskii and Plaut in [1] proved that every length space X is the quotient of an $\mathbb{R}$-tree T obtained via an action by isometries of a locally free group G on T. They showed that the quotient map $p : T \to T/G = X$ has the property that every rectifiable path in X lifts uniquely to a path in T, after a choice of initial point. They constructed the $\mathbb{R}$-tree T by considering the set of rectifiable paths in X without backtracking and showed that these paths naturally formed an $\mathbb{R}$-tree.

We say that $p : E \to X$ has the *unique path lifting property* (or is a *UPL map*) if for every path $\alpha : [0, 1] \to X$ and every $e \in p^{-1}(\{\alpha(0)\})$, there is a unique path $\tilde{\alpha} : [0, 1] \to E$ with $\tilde{\alpha}(0) = e$ and $p \circ \tilde{\alpha} = \alpha$. The following is well-known:

($*$) If $p : E \to X$ is a covering map, then p has the unique path lifting property.

MSC2020: 54F50, 55R65, 57M10.

Keywords: $\mathbb{R}$-tree, geodesic $\mathbb{R}$-tree reduction, path homotopy, dendrite, unique path lifting, covering map.

If the quotient map $p : E \to X$ coming from a group action is a UPL map, one may understand the group action in terms of paths in the base space. Thus it is natural to ask when can a topological space X be the orbit space of a group acting on an $\mathbb{R}$-tree in such a way that the quotient map has the unique path lifting property. We prove the following.

Theorem 1.1. *If a group G acts on an $\mathbb{R}$-tree T in such a way that the quotient map $p : T \to T/G$ has the unique path lifting property, then the quotient space T/G does not contain a two-dimensional Euclidean disc.*

Theorem 1.1 follows from the following lemma.

Lemma 1.2. *If X is a first countable, locally path-connected, and simply connected space, E is a path-connected space, and $p : E \to X$ is a UPL map, then p is a homeomorphism.*

This gives the following surprising corollary for manifolds.

Corollary 1.3. *If $p : E \to X$ is a UPL map where X is first countable, locally path connected, and semilocally simply connected and E is locally path connected, then p is a covering map. In particular, every UPL map $p : M \to N$ of manifolds is a covering map.*

Lemma 1.2 is a positive answer to Dydak's unique lifting problem: Is every map with the unique path lifting property from a path-connected space E to the unit disc $\mathbb{D}^2$ in the Euclidean plane a homeomorphism? (See [9, Problem 2.3] and [5, Problem 4.6].) Additionally, Corollary 1.3 is a converse to (∗), which is surprisingly difficult to verify outside of one-dimensional cases.

The primary obstruction to establishing the converse to (∗) is showing that the unique lifting of paths rel. basepoint implies the lifting of path homotopies. If one assumes that path lifts vary continuously relative to their starting point, then standard techniques would apply; see, for example, [15, Section 12 of Chapter III]. However, the weaker lifting hypothesis of UPL maps need not imply continuous lifting in general. In fact, the aforementioned results of Berestovskii and Plaut may be considered as evidence that the converse of (∗) should not hold even when $X = \mathbb{D}^2$. Since arbitrary paths in $\mathbb{D}^2$ can be approximated by paths in one-dimensional subspaces and since $\mathbb{R}$-trees are closed under various limiting constructions, limiting methods seem a promising way to produce a counterexample extending that for rectifiable paths.

Our principal tool is understanding *one-dimensional backtracking*. We will make backtracking formal by considering maps into $\mathbb{R}$-trees. Suppose the nonconstant path $\alpha : [0, 1] \to X$ in a metric space factors through an $\mathbb{R}$-tree T as $\alpha = q \circ p$, where $p : [0, 1] \to T$ is a path and $q : T \to X$ is a map. Let $r : [0, 1] \to T$

parameterize the unique geodesic in T from $p(0)$ to $p(1)$. Then we say that the path $\beta = q \circ r$ is obtained from α by *geodesic* $\mathbb{R}$*-tree reduction*. Certainly, if β is a geodesic $\mathbb{R}$-tree reduction of α, then α and β are homotopic. The surprising result and main idea for proving Theorem 1.1 and Corollary 1.3 is that any two homotopic paths are both $\mathbb{R}$-tree reductions of a single common path. Hence, one may delete one-dimensional backtracking from this common path to obtain either of the two homotopic paths.

Theorem 1.4 (main theorem). *If X is a topological space and $\alpha, \beta : [0, 1] \to X$ are homotopic paths, then there exists a path $\gamma : [0, 1] \to X$ such that α and β are both geodesic $\mathbb{R}$-tree reductions of γ.*

Typically, the path γ will be very complicated, as it (1) will be space-filling in the image of some chosen homotopy between α and β and (2) must pass through all of the points of both α and β in the same order that each of these paths does. Despite its complexity, the construction of γ from the pair α, β is entirely explicit. Our main theorem implies that the equivalence relation on paths generated by geodesic $\mathbb{R}$-tree reduction coincides precisely with path homotopy (Corollary 5.7).

In [14, Corollary 2.4], it is shown that the following notion of "thin homotopy" is an equivalence relation on the set of paths in a Hausdorff space: *two paths are thinly homotopic if they form a loop that factors through a simplicial tree*. The key idea needed to verify transitivity is a technical result ensuring that certain pushouts of simplicial trees are simplicial trees [14, Proposition 5.5]. Theorem 1.4 implies that if one replaces "simplicial tree" with "$\mathbb{R}$-tree" in the definition of a thin homotopy, then the analogous relation is *not* transitive (Corollary 5.8). This occurs precisely because there exist extreme situations, namely those modeled by our main construction, where analogous pushouts of $\mathbb{R}$-trees are topological discs.

Outline of paper. Section 2 establishes notation for four equivalence relations on the path space that we will require throughout the paper and lays out some of their properties. Section 3 defines Cantor paths and staggered paths, which we will use in our construction of the path γ from Theorem 1.4 to control the distance between approximations of γ. In Section 4, we define CIP loops and LIP loops and state Lemmas 4.6 and 4.8. These two lemmas allow us to conclude that whenever a given path is modified by a recursive process of inserting out-and-back paths along parts of the domain where it is already constant, the uniform limit is $\mathbb{R}$-tree homotopic to the original path. In Section 5 we prove the technical result, Lemma 5.3, that constitutes the inductive step in the construction of γ for Theorem 1.4, and so complete the proof of that theorem. Section 6 contains the proofs of Theorem 1.1, Lemma 1.2, and Corollary 1.3. The proof of Lemma 4.6 is given in the Appendix.

2. $\mathbb{R}$-tree factorization of paths and loops

For paths α and β in a space X, $\alpha\beta$ will denote path concatenation and $\overline{\alpha}$ will denote the reverse path. If (X, d) is a metric space, $\rho(\alpha, \beta) = \sup\{d(\alpha(t), \beta(t)) \mid t \in [0, 1]\}$ defines the sup metric on the space of paths from $[0, 1]$ to X. Recall that when (X, d) is complete, so is its path space [11, 4.3.13]. We also require notation for a variety of relations on paths.

Definition 2.1. Let $\alpha : [a, b] \to X$, $\beta : [c, d] \to X$ be paths.

(1) We say α is *equivalent* to β and write $\alpha \equiv \beta$ if $\alpha = \beta \circ h$ for some increasing homeomorphism $h : [a, b] \to [c, d]$. If h is linear, we may say that α is a *linear reparameterization* of β.

(2) We say α and β are *Fréchet equivalent* and write $\alpha \approxeq \beta$ if $\alpha \circ f = \beta \circ g$ for nondecreasing continuous surjections $f : [0, 1] \to [a, b]$ and $g : [0, 1] \to [c, d]$.

Both $\equiv$ and $\approxeq$ are equivalence relations finer than the path homotopy relation, which we denote by $\simeq$. See [8] for a proof of the transitivity of $\approxeq$. Recall that an $\mathbb{R}$-*tree* is a uniquely arcwise connected, and locally arcwise connected geodesic metric space [2; 16] and a *Peano continuum* is a compact, connected, locally path-connected, metric space. A *dendrite* is a Peano continuum that does not contain a simple closed curve [17, Definition 10.1]. It is then an exercise to see that a dendrite is uniquely arcwise connected and locally arcwise connected. Mayer and Oversteegen showed that any uniquely arcwise connected and locally arcwise connected metric space admits a geodesic metric [16, Theorem 5.1]. Thus every dendrite admits a metric that makes it an $\mathbb{R}$-tree.

Definition 2.2. Let $\alpha, \beta : [0, 1] \to X$ be paths in a topological space X.

(1) We say that β is a *geodesic* $\mathbb{R}$*-tree reduction of* α, and we write $\alpha \geq_{\mathbb{R}} \beta$, if there is an $\mathbb{R}$-tree T, a path $p : [0, 1] \to T$, an injective path $r : [0, 1] \to T$ with $p(i) = r(i)$ for $i \in \{0, 1\}$, and a map $f : T \to X$ such that $f \circ p = \alpha$ and $f \circ r = \beta$.

(2) We say that α and β are $\mathbb{R}$*-tree homotopic*, and we write $\alpha \simeq_{\mathbb{R}} \beta$, if $\alpha\bar{\beta}$ is a loop that factors through an $\mathbb{R}$-tree, that is, if $\alpha(i) = \beta(i)$ for $i \in \{0, 1\}$ and if there exists an $\mathbb{R}$-tree T, a loop $\ell : [0, 1] \to T$ and a map $f : T \to X$ such that $f \circ \ell = \alpha\bar{\beta}$.

Note that

$$\alpha \equiv \beta \;\Rightarrow\; \alpha \geq_{\mathbb{R}} \beta \;\Rightarrow\; \alpha \simeq_{\mathbb{R}} \beta \;\Rightarrow\; \alpha \simeq \beta,$$

where the last implication holds since $\mathbb{R}$-trees are contractible. Geodesic $\mathbb{R}$-tree reduction is our formalization of removing backtracking, as mentioned in the introduction. Also, if $\alpha \simeq_{\mathbb{R}} \beta$, where β is injective, then $\alpha \geq_{\mathbb{R}} \beta$. Certainly, $\simeq_{\mathbb{R}}$ is reflexive and symmetric. However, the transitivity of $\simeq_{\mathbb{R}}$ does not hold in general; see Corollary 5.8. In a given space, transitivity of $\simeq_{\mathbb{R}}$ is equivalent to all paths admitting a unique (up to $\equiv$) "maximally reduced" geodesic ℝ-tree reduction. One must be wary of this temptation as our main result implies that there exist some paths in the plane which fail to have unique maximally reduced representatives.

We note some key properties of $\simeq_{\mathbb{R}}$ and $\geq_{\mathbb{R}}$ that *do* hold.

Remark 2.3. If T_1 and T_2 are ℝ-trees, $f_1 : [0, 1] \to T_1$ is an injective path, and $f_2 : [0, 1] \to T_2$ is any path, then the pushout of f_1 and f_2 is an ℝ-tree. Indeed, this pushout is obtained by attaching the closure of each connected component of $T_1 \setminus f_1([0, 1])$ to T_2 at a point along the image of f_2. As a consequence:

(1) $\geq_{\mathbb{R}}$ is transitive and is antisymmetric up to equivalence; that is, $\alpha \geq_{\mathbb{R}} \beta$ and $\beta \geq_{\mathbb{R}} \alpha \Rightarrow \alpha \equiv \beta$. Hence, $\geq_{\mathbb{R}}$ induces a partial order on path-equivalence classes.

(2) $\alpha \simeq_{\mathbb{R}} \beta$ if and only if α and β share a common geodesic ℝ-tree reduction, that is, if and only if there exists δ with $\alpha \geq_{\mathbb{R}} \delta$ and $\beta \geq_{\mathbb{R}} \delta$.

Lemma 2.4. *The following properties of* ℝ*-tree homotopy hold.*

(1) *If* $\alpha, \beta : [0, 1] \to X$ *are paths such that* $\alpha \simeq_{\mathbb{R}} \beta$ (*resp.* $\alpha \geq_{\mathbb{R}} \beta$) *and* $f : X \to Y$ *is a map, then* $f \circ \alpha \simeq_{\mathbb{R}} f \circ \beta$ (*resp.* $f \circ \alpha \geq_{\mathbb{R}} f \circ \beta$).

(2) *If* $\alpha_1 \simeq_{\mathbb{R}} \beta_1$, $\alpha_2 \simeq_{\mathbb{R}} \beta_2$, *and* $\alpha_1(1) = \alpha_2(0)$, *then* $\alpha_1\alpha_2 \simeq_{\mathbb{R}} \beta_1\beta_2$.

Assume moreover that X *is Hausdorff. Then*:

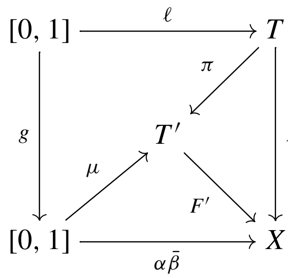
(3) *If* $\alpha, \beta : [0, 1] \to X$ *are paths and* $\alpha \circ g_1 \simeq_{\mathbb{R}} \beta \circ g_2$ *for nondecreasing continuous surjections* $g_1, g_2 : [0, 1] \to [0, 1]$, *then* $\alpha \simeq_{\mathbb{R}} \beta$.

(4) *If* $\alpha_1, \alpha_2 : [0, 1] \to X$ *are paths such that* $\alpha_1 \simeq_{\mathbb{R}} \alpha_2$ *and* $\beta_1, \beta_2 : [0, 1] \to X$ *are paths such that* $\alpha_1 \approx \beta_1$ *and* $\alpha_2 \approx \beta_2$, *then* $\beta_1 \simeq_{\mathbb{R}} \beta_2$.

Proof. (1) is clear. (2) holds since the one-point union of two ℝ-trees is an ℝ-tree. (3) Since $\alpha \circ g_1 \simeq_{\mathbb{R}} \beta \circ g_2$, there exists an ℝ-tree T, map $F : T \to X$, and loop $\ell : [0, 1] \to T$ such that $F \circ \ell = (\alpha \circ g_1)(\overline{\beta \circ g_2})$. We may replace T with the dendrite $\ell([0, 1])$ and assume ℓ is surjective. Since X is assumed to be Hausdorff, $F(T)$ is a compact metric space [17, 8.17] and we may replace X with $F(T)$. Define a nondecreasing surjection $g : [0, 1] \to [0, 1]$ by

$$g(t) = \begin{cases} \frac{1}{2} g_1(2t) & \text{if } t \in \left[0, \frac{1}{2}\right], \\ 1 - \frac{1}{2} g_2(2 - 2t) & \text{if } t \in \left[\frac{1}{2}, 1\right], \end{cases}$$

so that $F \circ \ell = (\alpha\bar{\beta}) \circ g$. Applying the monotone-light factorization theorem (see [10, Theorem 1] or [20, Theorem 2.3]) to the map $F : T \to X$ of compact metric spaces, we have $F = F' \circ \pi$ for a monotone map $\pi : T \to T'$ and a light map $F' : T' \to X$. Since π is a monotone quotient map on a dendrite, T' is a dendrite [17, Exercise 10.52]. Since g is monotone, F' is light, and $F' \circ \pi \circ \ell = (\alpha\bar{\beta}) \circ g$, the loop $\pi \circ \ell : [0, 1] \to T'$ is constant on the fibers of g. Thus, there is a unique loop $\mu : [0, 1] \to T'$ such that $\pi \circ \ell = \mu \circ g$:



Since $F' \circ \mu \circ g = (\alpha\bar{\beta}) \circ g$, where g is surjective, we have $F' \circ \mu = \alpha\bar{\beta}$, proving $\alpha \simeq_{\mathbb{R}} \beta$.

(4) Write $\alpha_1 \circ f_1 = \beta_1 \circ g_1$ and $\alpha_2 \circ f_2 = \beta_2 \circ g_2$ for nondecreasing continuous surjections $f_1, f_2, g_1, g_2 : [0, 1] \to [0, 1]$. Since $\alpha_1 \simeq_{\mathbb{R}} \alpha_2$, there exists an $\mathbb{R}$-tree T, a map $F : T \to X$, and loop $\ell : [0, 1] \to T$ such that $F \circ \ell = \alpha_1\overline{\alpha}_2$. Write $\ell = \mu_1\overline{\mu}_2$ so that $F \circ \mu_i = \alpha_i$, $i \in \{1, 2\}$. If $\ell' = (\mu_1 \circ f_1)(\overline{\mu_2 \circ f_2})$, then the factorization

$$F \circ \ell' = (\alpha_1 \circ f_1)(\overline{\alpha_2 \circ f_2}) = (\beta_1 \circ g_1)(\overline{\beta_2 \circ g_2})$$

shows that $\beta_1 \circ g_1 \simeq_{\mathbb{R}} \beta_2 \circ g_2$. Now, (3) implies $\beta_1 \simeq_{\mathbb{R}} \beta_2$. □

3. Cantor paths and their staggering

Definition 3.1. Recall that a *Cantor set* is any nonempty, compact, perfect, and totally disconnected metric space and any such space is homeomorphic to the standard ternary Cantor set. An open set $U \subseteq (0, 1)$ is a *Cantor complement* if $[0, 1] \setminus U$ is homeomorphic to a Cantor set.

An open set $U \subseteq (0, 1)$ is a Cantor complement if and only if U is dense in $[0, 1]$ and the set of connected components of U have the order type of $\mathbb{Q}$. It is necessarily the case that the connected components of a Cantor complement have pairwise disjoint closures. Given a path $\alpha : [0, 1] \to X$, let $\text{lc}(\alpha)$ denote the set of connected components of the open set $\mathcal{O}(\alpha) = \{t \in [0, 1] \mid \alpha \text{ is locally constant at } t\}$. Note that distinct elements of $\text{lc}(\alpha)$ have pairwise disjoint closures and that α is a light map if and only if $\mathcal{O}(\alpha) = \varnothing$.

Definition 3.2. A nonconstant path $\alpha : [0, 1] \to X$ is a *Cantor path* if $\mathcal{O}(\alpha)$ is a Cantor complement.

The standard ternary Cantor map $\tau : [0, 1] \to [0, 1]$ collapses the closure of each component of the complement of the ternary Cantor set to a point. Thus τ is a nondecreasing, surjective Cantor path. If α is a light path, then $\alpha \circ \tau$ is a Cantor path. A key concept in the proof of our main theorem is the following.

Definition 3.3. Let U_1, U_2 be proper open subsets of $(0, 1)$ such that for each $i \in \{1, 2\}$, $\inf(U_i) = 0$, $\sup(U_i) = 1$, and such that the connected components of U_i have pairwise disjoint closures. We say the sets U_1 and U_2 are *staggered* if $U_1 \cup U_2 = (0, 1)$, or equivalently, if $\partial U_1 \cap \partial U_2 = \{0, 1\}$.

The next lemma, which has a straightforward proof, allows us to select connected components from staggered Cantor complements U_1 and U_2 so that the resulting collections have the order type of $\mathbb{Z}$ while still having staggered unions.

Lemma 3.4. *Let U_1 and U_2 be staggered Cantor complements. Then for $i \in \{1, 2\}$ there exists a set of connected components $\mathscr{U}_i$ of U_i such that $\mathscr{U}_i$ has the order type of $\mathbb{Z}$, $\inf(\bigcup \mathscr{U}_i) = 0$, $\sup(\bigcup \mathscr{U}_i) = 1$, and such that $\bigcup \mathscr{U}_1$ and $\bigcup \mathscr{U}_2$ are staggered.*

Proof. Let $\mathscr{U}_1$ be the components of U_1 that are not entirely contained in some component of U_2. Let $\mathscr{U}_2$ be the components of U_2 which are not entirely contained in some element of $\mathscr{U}_1$. Note that $\mathscr{U}_1 \cup \mathscr{U}_2$ still covers $(0, 1)$. Now suppose the set of endpoints of $\mathscr{U}_1$ has an accumulation point, p, that is neither 0 nor 1. Since the components of $\mathscr{U}_1$ are disjoint and open, p is not contained in any element of $\mathscr{U}_1$. Since $\mathscr{U}_1 \cup \mathscr{U}_2$ covers $(0, 1)$, p is contained in some $O \in \mathscr{U}_2$. By construction each element of $\mathscr{U}_2$ can contain at most two endpoints of $\mathscr{U}_1$, which contradicts that p is an accumulation point of the endpoints of $\mathscr{U}_1$.

Since both $\mathscr{U}_1$ and $\mathscr{U}_2$ are nonempty and each component separates $\mathbb{R}$, it must be that the set of components, and thus the set of endpoints, of $\mathscr{U}_1$ is infinite, does not contain 0 or 1, but has both as limit points.

Therefore the set of endpoints of $\mathscr{U}_1$ is discrete and (being a subset of $\mathbb{R}$) linear, and it has limit points at 0 and 1. Hence it must be a bi-infinite sequence and thus have order type $\mathbb{Z}$. A similar argument holds for $\mathscr{U}_2$. □

Lemma 3.5. *Let U be a Cantor complement. For every $\epsilon > 0$, there exists an increasing homeomorphism $f : [0, 1] \to [0, 1]$ such that $\rho(f, \mathrm{id}_{[0,1]}) < \epsilon$ and that U and $f(U)$ are staggered.*

Proof. Let $\epsilon > 0$ and let $\mathscr{U}$ be the set of connected components of U with the induced ordering. Since U is a Cantor complement, $\mathscr{U}$ is densely ordered and we can find a bi-infinite sequence $I_n = (a_n, b_n) \in \mathscr{U}$, $n \in \mathbb{Z}$, such that $\inf(\bigcup_{n\in\mathbb{N}} I_n) = 0$, $\sup\big(\bigcup_{n\in\mathbb{N}} I_n\big) = 1$, and such that for all $n \in \mathbb{Z}$, the segment $[b_n, a_{n+1}]$ has diameter less than $\frac{1}{3}\epsilon$. For each $n \in \mathbb{Z}$, find $J_n = (c_n, d_n) \in \mathscr{U}$ with $J \subseteq (b_n, a_{n+1})$. We will define f by first setting its value on the closures of the intervals in the bi-infinite sequence

$$\cdots < I_{-2} < J_{-2} < I_{-1} < J_{-1} < I_0 < J_0 < I_1 < J_1 < I_2 < J_2 < \cdots .$$

For each $n \in \mathbb{Z}$, choose a subdivision of $[a_n, b_n]$ as follows. Find $a_n < D_{n-1} < A_n < B_n < C_n < b_n$, where the outer four segments $[a_n, D_{n-1}]$, $[D_{n-1}, A_n]$, $[B_n, C_n]$,

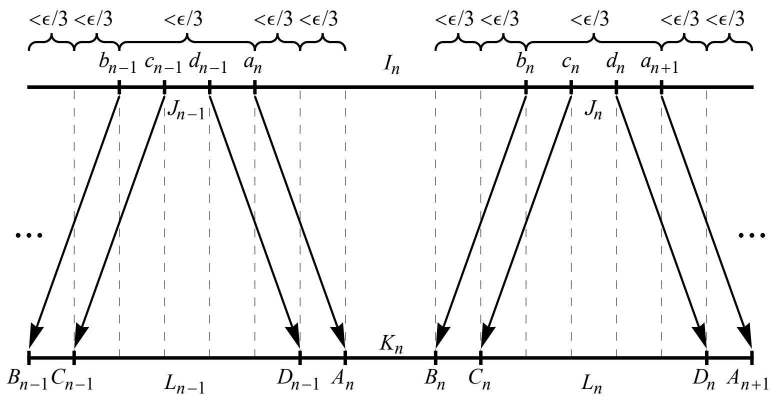


Figure 1. Part of the homeomorphism $f : [0, 1] \to [0, 1]$ that maps I_n to K_n and J_n to L_n in a piecewise linear fashion.

and $[C_n, b_n]$ all have diameter less than $\frac{1}{3}\epsilon$. If $K_n = (A_n, B_n)$ and $L_n = (C_n, D_n)$, then the bi-infinite sequence

$$\cdots < K_{-2} < L_{-2} < K_{-1} < L_{-1} < K_0 < L_0 < K_1 < L_1 < K_2 < L_2 < \cdots$$

limits to 0 on the left and 1 on the right. Moreover, $K_n \subseteq I_n$ and $J_n \subseteq L_n$ for all $n \in \mathbb{Z}$. We define f so that it maps $[a_n, b_n]$ to $[A_n, B_n]$ and $[c_n, d_n]$ to $[C_n, D_n]$ by increasing linear maps. Since the sets I_n, J_n, K_n, L_n limit to 0 as $n \to -\infty$ and to 1 as $n \to \infty$, this definition extends uniquely to an increasing homeomorphism $f : [0, 1] \to [0, 1]$, which is piecewise linear on $[a_{-n}, b_n]$ for all $n \geq 1$ (see Figure 1).

Our choices of the sizes of the intervals $[b_n, a_{n+1}]$ and subdivisions ensure that $\rho(f, \mathrm{id}_{[0,1]}) < \epsilon$. Additionally, if $t \in (0, 1) \setminus U$, then $t \in [b_n, c_n] \cup [d_n, a_{n+1}]$ for some $n \in \mathbb{Z}$. Since $[b_n, c_n] \cup [d_n, a_{n+1}] \subseteq L_n = f(J_n)$ where $J_n \in \mathscr{U}$, we have $t \in f(U)$. We conclude that $U \cup f(U) = (0, 1)$, i.e., U and $f(U)$ are staggered. □

Definition 3.6. We say that two Cantor paths $\alpha, \beta : [0, 1] \to X$ are *staggered* if $\mathcal{O}(\alpha)$ and $\mathcal{O}(\beta)$ are staggered.

Our final lemma of the section allows us to take two nonstaggered Cantor paths and perturb one of them so that the resulting pair is staggered.

Lemma 3.7. *Given any two Cantor paths $\alpha, \beta : [0, 1] \to X$ and $\epsilon, \delta > 0$, there exists an increasing homeomorphism $f : [0, 1] \to [0, 1]$ such that $\rho(f, \mathrm{id}_{[0,1]}) < \delta$, $\rho(\alpha, \alpha \circ f) < \epsilon$, and such that $\alpha \circ f$ and β are staggered.*

Proof. Let $U_1 = \mathcal{O}(\alpha)$ and $U_2 = \mathcal{O}(\beta)$ and note that $V = U_1 \cap U_2$ is also a Cantor complement. Find $0 < \delta' < \delta$ such that $|s - t| < \delta'$ implies $|\alpha(s) - \alpha(t)| < \frac{1}{2}\epsilon$. By Lemma 3.5, we can find an increasing homeomorphism $f : [0, 1] \to [0, 1]$ such that $|f(t) - t| < \delta'$ for all $t \in [0, 1]$ and $V \cup f(V) = (0, 1)$. Hence, $|\alpha(f(t)) - \alpha(t)| < \frac{1}{2}\epsilon$

for all $t \in [0, 1]$, which gives $\rho(f, \mathrm{id}_{[0,1]}) \le \delta' < \delta$ and $\rho(\alpha, \alpha \circ f) < \epsilon$. We also have $V \cup f^{-1}(V) = (0, 1)$ and $\mathcal{O}(\alpha \circ f) = f^{-1}(U_1)$. Thus

$$(0, 1) = V \cup f^{-1}(V) = (U_1 \cap U_2) \cup (f^{-1}(U_1) \cap f^{-1}(U_2)) \subseteq U_2 \cup \mathcal{O}(\alpha \circ f),$$

showing that $\alpha \circ f$ and β are staggered. □

4. Inserting inverse pairs into Cantor paths

If $\mathscr{U}$ is a collection of open intervals in $[0, 1]$ with disjoint closures, then a *$\mathscr{U}$-collapsing map* is a nondecreasing, continuous surjection $\mu : [0, 1] \to [0, 1]$ that is constant on the closure of each $J \in \mathscr{U}$ and bijective on $[0, 1] \setminus \bigcup_{J \in \mathscr{U}} \overline{J}$. Such maps may be constructed canonically using dyadic rational outputs and by enumerating $\mathscr{U}$ by nonincreasing length and the ordering in $[0, 1]$ (lexicographically). If a path $\alpha : [0, 1] \to X$ is not light and $k_\alpha : [0, 1] \to [0, 1]$ is a $\mathrm{lc}(\alpha)$-collapsing map, then there is a unique light path $\alpha^\lambda : [0, 1] \to X$ such that $\alpha^\lambda \circ k_\alpha = \alpha$.

Definition 4.1. We call a loop $\ell : [0, 1] \to X$ an *inverse-pair loop* if there exists a path $\alpha : [0, 1] \to X$ such that $\ell \equiv \alpha\overline{\alpha}$. More specifically:

(1) If α is a Cantor path or has the form $\alpha \equiv \alpha_1\alpha_2$ for Cantor path α_1 and constant path α_2, then, we call ℓ a *Cantor-inverse-pair loop* or *CIP loop* (see Figure 2).

(2) If α is light, we call ℓ a *light-inverse-pair loop* or *LIP loop*.

Remark 4.2. If $\alpha : [0, 1] \to X$ is a CIP loop, then α is a Cantor path. If $\alpha = \alpha^\lambda \circ k_\alpha$, where $k_\alpha : [0, 1] \to [0, 1]$ is an $\mathrm{lc}(\alpha)$-collapsing map $\alpha^\lambda : [0, 1] \to X$ is a light path, then α^λ is a LIP loop.

Definition 4.3. Let $\alpha, \beta : [0, 1] \to X$ be Cantor paths and $\mathscr{U} \subseteq \mathrm{lc}(\alpha)$. We say that β is a *$\mathscr{U}$-extension* of α if

(1) $\beta(t) = \alpha(t)$ for all $t \in [0, 1] \setminus \bigcup \mathscr{U}$, and

(2) for each $J \in \mathscr{U}$, $\beta|_{\overline{J}}$ is a CIP loop.

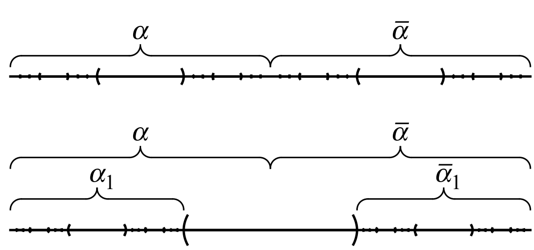


Figure 2. A CIP loop is an inverse pair loop that has either the form of an inverse pair $\alpha\overline{\alpha}$ of Cantor paths (top) or the form $\alpha_1\beta\overline{\alpha}_1$ for a Cantor path α_1 and constant path β (bottom).

If β is a $\mathscr{U}$-extension of α for some subset $\mathscr{U} \subseteq \mathrm{lc}(\alpha)$, we write $\alpha \preceq_* \beta$.

Remark 4.4. If $\alpha \preceq_* \beta$, then $\mathcal{O}(\beta) \subseteq \mathcal{O}(\alpha)$. In particular, if $J \in \mathrm{lc}(\beta)$, then either $J \in \mathrm{lc}(\alpha) \setminus \mathscr{U}$ or $J \subseteq K$ for some $K \in \mathscr{U}$. The loop $\alpha\bar{\beta}$ factors through a dendrite in a highly structured way: namely, there is a dendrite $D(\beta, \alpha)$ constructed by starting with a "base arc" B and attaching an arc to B for each element of $\mathscr{U}$. We have $F \circ (a\bar{b}) = \alpha\bar{\beta}$ where $a : [0, 1] \to D(\beta, \alpha)$ is a monotone map onto B and where $b : [0, 1] \to D(\beta, \alpha)$ maps $[0, 1] \setminus \bigcup \mathscr{U}$ into B and the closure of each element of $\mathscr{U}$ onto the corresponding attached arc in $D(\beta, \alpha)$ by an inverse-pair loop. While $\alpha \simeq_{\mathbb{R}} \beta$ holds, the relation $\alpha \leq_{\mathbb{R}} \beta$ only holds in the trivial case where $\mathscr{U} = \varnothing$. (See the Appendix for details.)

Remark 4.5. The relation $\preceq_*$ on the set of paths in a space X is certainly reflexive and it is straightforward to check that it is antisymmetric. However, $\preceq_*$ is *not* transitive. Rather, it is a very fine relation that generates a partial order relation, which is strictly finer than $\leq_{\mathbb{R}}$.

If we have a sequence $\gamma_1 \preceq_* \gamma_2 \preceq_* \gamma_3 \preceq_* \cdots$ in a space X where, as one proceeds through the sequence, the added out-and-back loops have very quickly shrinking diameters, then $\{\gamma_n\}_{n\in\mathbb{N}}$ should converge uniformly to a path γ. Moreover, $\gamma_1\bar{\gamma}_2$ factors through a dendrite, call it D_1, as described in Remark 4.4. Since γ_{n+1} agrees with γ_n except on portions on which γ_{n+1} is a CIP loop, we may recursively construct a dendrite D_n by attaching arcs to D_{n-1} so that $\gamma_j\bar{\gamma}_{n+1}$ factors through D_n for each $j \in \{1, 2, \ldots, n\}$. In the limit, we find that there is a uniquely determined limit dendrite $D_\infty = \varprojlim_n D_n$ that $\gamma_n\bar{\gamma}$ factors through for all n. Hence, $\gamma \simeq_{\mathbb{R}} \gamma_n$ for all $n \in \mathbb{N}$. The next lemma establishes this conclusion assuming the existence of the limit γ. As one can see from the above proof sketch, this result is fairly intuitive. However, a detailed proof requires careful bookkeeping of parameterizations of paths in inverse limits. We defer the details to Appendix.

Lemma 4.6. *If $\{\gamma_n\}_{n\in\mathbb{N}}$ is a sequence of Cantor paths in a metric space (X, d) such that $\gamma_n \preceq_* \gamma_{n+1}$ for all $n \in \mathbb{N}$ and such that $\{\gamma_n\}_{n\in\mathbb{N}} \to \gamma$ uniformly, then $\gamma_n \simeq_{\mathbb{R}} \gamma$ for all $n \in \mathbb{N}$.*

In the next section, we will form alternating sequences $\alpha_1 \preceq_* \alpha_1' \equiv \alpha_2 \preceq_* \alpha_2' \equiv \cdots$, where the equivalences are given by small perturbations. The next lemma, which is proved using elementary real analysis, allows us to manage all these perturbations simultaneously.

Lemma 4.7. *Suppose $\{f_n\}_{n\in\mathbb{N}}$ is a sequence of increasing homeomorphisms $f_n : [0, 1] \to [0, 1]$ such that $\rho(f_n, \mathrm{id}_{[0,1]}) \leq ar^n$ for some $a > 0$ and $|r| < 1$. If $g_n = f_n^{-1} \circ f_{n-1}^{-1} \circ \cdots \circ f_2^{-1} \circ f_1^{-1}$ for all $n \in \mathbb{N}$, then $\{g_n\}_{n\in\mathbb{N}}$ converges uniformly to a continuous, nondecreasing surjection $g_\infty : [0, 1] \to [0, 1]$.*

Next, we combine the previous two lemmas.

Lemma 4.8. *Suppose* $\{\alpha_n\}_{n\in\mathbb{N}}$ *and* $\{\alpha'_n\}_{n\in\mathbb{N}}$ *are sequences of Cantor paths in a metric space* (X, d), $\gamma : [0, 1] \to X$ *is a path, and* $\{f_n\}_{n\in\mathbb{N}}$ *is a sequence of increasing homeomorphisms* $f_n : [0, 1] \to [0, 1]$ *such that the following hold*:

(1) $\{\alpha_n\}_{n\in\mathbb{N}} \to \gamma$ *uniformly.*

(2) $\alpha_n \preceq_* \alpha'_n$ *for all* $n \in \mathbb{N}$.

(3) $\alpha_{n+1} = \alpha'_n \circ f_n$ *for all* $n \in \mathbb{N}$.

(4) *There exists* $a > 0$ *and* $|r| < 1$ *such that* $\rho(f_n, \mathrm{id}_{[0,1]}) \le ar^n$ *for all* $n \in \mathbb{N}$.

Then $\alpha_n \simeq_{\mathbb{R}} \gamma$ *for all* $n \in \mathbb{N}$.

Proof. Let $g_0 = \mathrm{id}_{[0,1]}$ and $g_n = f_n^{-1} \circ f_{n-1}^{-1} \circ \cdots \circ f_2^{-1} \circ f_1^{-1}$ for all $n \in \mathbb{N}$. Note that $f_n \circ g_n = g_{n-1}$ for all $n \in \mathbb{N}$. By Lemma 4.7, $\{g_n\}_{n\in\mathbb{N}}$ converges uniformly to a continuous, nondecreasing surjection $g_\infty : [0, 1] \to [0, 1]$.

Set $\gamma_1 = \alpha_1$ and $\gamma_n = \alpha_n \circ g_{n-1}$ for $n \ge 2$. For the moment, fix $n \in \mathbb{N}$. Since $\alpha_{n+1} = \alpha'_n \circ f_n$, we have

$$\gamma_{n+1} = \alpha_{n+1} \circ g_n = \alpha'_n \circ f_n \circ g_n = \alpha'_n \circ g_{n-1}.$$

By assumption, $\alpha_n \preceq_* \alpha'_n$. Composing α_n and α'_n with the homeomorphism g_{n-1} gives $\gamma_n \preceq_* \gamma_{n+1}$. Additionally, since $\{\alpha_n\}_{n\in\mathbb{N}} \to \gamma$ and $\{g_{n-1}\}_{n\in\mathbb{N}} \to g_\infty$ uniformly, we have $\{\gamma_n\}_{n\in\mathbb{N}} = \{\alpha_n \circ g_{n-1}\}_{n\in\mathbb{N}} \to \gamma \circ g_\infty$. It now follows from Lemma 4.6 that $\gamma_n \simeq_{\mathbb{R}} \gamma \circ g_\infty$ for all $n \in \mathbb{N}$. Thus $\alpha_n \circ g_{n-1} \simeq_{\mathbb{R}} \gamma \circ g_\infty$ for all $n \in \mathbb{N}$. Since we have Fréchet equivalences (see Definition 2.1) $\alpha_n \approx \alpha_n \circ g_{n-1}$ and $\gamma \approx \gamma \circ g_\infty$, it follows from Lemma 2.4(4) that $\alpha_n \simeq_{\mathbb{R}} \gamma$ for all $n \in \mathbb{N}$. □

5. Proof of the main theorem

To begin, we fix Cantor-path parameterizations of planar line segments. Recall that $\tau : [0, 1] \to [0, 1]$ is the ternary Cantor map.

Definition 5.1. Given points x, y in the closed unit disc $\mathbb{D}^2$, let $L_{x,y} : [0, 1] \to \mathbb{D}^2$ be the path defined by $L_{x,y}(s) = \tau(s)\big(\frac{1}{2}(x + y)\big) + (1 - \tau(s))x$.

Remark 5.2. If $x = y$, then $L_{x,y}$ is constant. If $x \ne y$, then $L_{x,y}$ is a Cantor path that parameterizes the line segment from x to $\frac{1}{2}(x + y)$. Moreover, if $x_1, x_2, y_1, y_2 \in \mathbb{D}^2$ with midpoints $m_i = \frac{1}{2}(x_i + y_i)$, then since the paths L_{x_1,y_1} and $\overline{L}_{y_2,x_2}$ are both parameterized using τ, their sup distance is the maximum distance between the endpoints, that is,

$$\begin{aligned}\rho(L_{x_1,y_1}, \overline{L}_{y_2,x_2}) &\le \max\big\{d(x_1, m_2), d(m_1, y_2)\big\}\\ &\le \max\big\{\tfrac{1}{2}d(x_2, y_2) + d(x_1, x_2), \tfrac{1}{2}d(x_1, y_1) + d(y_1, y_2)\big\}\end{aligned}$$

(see Figure 3). When $x = x_1 = x_2$ and $y = y_1 = y_2$, we have $\rho(L_{x,y}, \overline{L}_{y,x}) = \frac{1}{2}d(x, y)$.

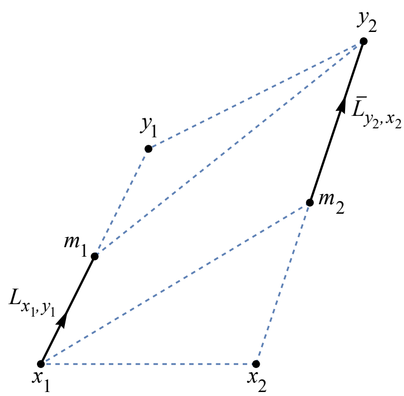


Figure 3. Since L_{x_1,y_1} and $\overline{L}_{y_2,x_2}$ are both parameterized by τ, their metric distance is the maximum length between starting and ending points.

We attribute the next lemma especially to the third author. The main idea is to fix staggered Cantor paths $\alpha, \beta : [0, 1] \to \mathbb{D}^2$ satisfying $\alpha(i) = \beta(i)$, $i \in \{0, 1\}$, and modify both of them by inserting CIP loops on a $\mathbb{Z}$-ordered sequence of elements of $\mathrm{lc}(\alpha)$ and $\mathrm{lc}(\beta)$, respectively, so that $\alpha \preceq_* \alpha'$ and $\beta \preceq_* \beta'$. Specifically, we construct α' from α by inserting a bi-infinite sequence of CIP loops of the form $L_{x,y}\overline{L}_{x,y}$ or $(L_{x,y})c(\overline{L}_{x,y})$ (for a constant path c) on certain elements of $\mathrm{lc}(\alpha)$. We will construct β' from β in an analogous way. However, the two constructions are not symmetric. Rather, they must be done simultaneously in an interlocking fashion. In the end, the resulting paths α' and β' will have $\frac{2}{3}$ the sup distance of the original paths. That one can insert nonconstant portions into both paths and somehow shrink the distance between them is somewhat nonintuitive and is only possible because the paths are *staggered*.

Since the construction of α' and β' will involve an intricate arrangement of overlapping intervals, we employ the following notation: if $I \subseteq [0, 1]$ is an interval, then $\ell(I)$ and $r(I)$ will denote the left and right endpoints of I.

Lemma 5.3. *For staggered Cantor paths $\alpha, \beta : [0, 1] \to \mathbb{D}^2$ such that $\alpha(i) = \beta(i)$, $i \in \{0, 1\}$, there exist Cantor paths $\alpha', \beta' : [0, 1] \to \mathbb{D}^2$ such that*

(1) *$\alpha \preceq_* \alpha'$ and $\rho(\alpha, \alpha') \leq \frac{1}{2}\rho(\alpha, \beta)$,*

(2) *$\beta \preceq_* \beta'$ and $\rho(\beta, \beta') \leq \frac{1}{2}\rho(\alpha, \beta)$, and*

(3) *$\rho(\alpha', \beta') \leq \frac{2}{3}\rho(\alpha, \beta)$.*

Proof. Let $\delta = \rho(\alpha, \beta)$. By assumption, $\mathcal{O}(\alpha)$ and $\mathcal{O}(\beta)$ are staggered Cantor complements. By Lemma 3.4, we may select a set of connected components $\mathscr{U}_1$ of $\mathcal{O}(\alpha)$ and $\mathscr{U}_2$ of $\mathcal{O}(\beta)$ such that for $i \in \{0, 1\}$, $\mathscr{U}_i$ has the order type of $\mathbb{Z}$, $\inf(\bigcup \mathscr{U}_i) = 0$, $\sup(\bigcup \mathscr{U}_i) = 1$, and such that $\bigcup \mathscr{U}_1$ and $\bigcup \mathscr{U}_2$ are staggered. Index the elements of $\mathscr{U}_1$ as $\cdots < A_{-2} < A_{-1} < A_0 < A_1 < A_2 < \cdots$ and the elements of $\mathscr{U}_2$ as $\cdots < C_{-2} < C_{-1} < C_0 < C_1 < C_2 < \cdots$ so that C_n meets A_{n-1} and A_n. Set $w_n = \alpha(\bar{A}_n)$ and $y_n = \beta(\bar{C}_n)$.

For the moment, fix $n \in \mathbb{Z}$. Since $\mathcal{O}(\alpha)$ is a Cantor complement and α is uniformly continuous, we can find a sequence $B_{n,1} < B_{n,2} < \cdots < B_{n,k_n}$ in $\mathcal{O}(\alpha)$ of length $k_n \geq 2$, where each set $B_{n,j}$ is contained in $[r(A_n), \ell(A_{n+1})]$ and such that if I is a connected component of $[r(A_n), \ell(A_{n+1})] \setminus \bigcup_{j=1}^{k_n} B_{n,j}$, then $\operatorname{diam}(\alpha(\bar{I})) < \frac{1}{6}\delta$. Similarly, since $\mathcal{O}(\beta)$ is a Cantor complement, we can find a sequence $D_{n,1} < D_{n,2} < \cdots < D_{n,m_n}$ in $\mathcal{O}(\beta)$ of length $m_n \geq 2$ where each $D_{n,j}$ is contained in $[r(C_{n-1}), \ell(C_n)]$ and such that if I is a connected component of $[r(C_{n-1}), \ell(C_n)] \setminus \bigcup_{j=1}^{m_n} D_{n,j}$, then $\operatorname{diam}(\beta(\bar{I})) < \frac{1}{6}\delta$. Note that $D_{n,j} \subseteq A_n$ and $B_{n,j} \subseteq C_n$ holds whenever these sets are defined (see Figure 4).

Set $x_{n,j} = \alpha(\bar{B}_{n,j})$ and $z_{n,j} = \beta(\bar{D}_{n,j})$ whenever these sets are defined. Let p_n be the midpoint of $[\ell(A_n), r(C_{n-1})]$, q_n be the midpoint of $[\ell(C_n), r(A_n)]$, $\eta_{n,j}$ be the midpoint of $B_{n,j}$, and $\theta_{n,j}$ be the midpoint of $D_{n,j}$.

To begin our definition of α', we set α' to agree with α on $\{0, 1\}$ and on $[r(A_n), \ell(A_{n+1})] \setminus \bigcup_{j=1}^{k_n} B_{n,j}$ for all $n \in \mathbb{Z}$. We complete the definition of α' piecewise by fixing $n \in \mathbb{N}$ and defining α' on $\bar{A}_n$ in three cases and $\bar{B}_{n,j}$ for $1 \leq j \leq k_n$ in a fourth case. It may be helpful to note that

$$\bar{A}_n = [\ell(A_n), p_n] \cup [p_n, r(C_{n-1})] \cup [r(C_{n-1}), \ell(C_n)] \cup [\ell(C_n), q_n] \cup [q_n, r(A_n)]$$

(see Figure 5).

(A1) On $[\ell(A_n), \theta_{n,1}]$, we define α' to be the CIP loop, which is the linear reparameterization of $L_{w_n, y_{n-1}}$ on $[\ell(A_n), p_n]$, the constant path at $\frac{1}{2}(w_n + y_{n-1})$ on $[p_n, \ell(D_{n,1})]$, and the linear reparameterization of $\bar{L}_{w_n, y_{n-1}}$ on $[\ell(D_{n,1}), \theta_{n,1}]$.

(A2) On $[\theta_{n,j}, \theta_{n,j+1}]$ (for each $1 \leq j \leq m_n - 1$), we define α' to be the CIP loop, which is the linear reparameterization of $L_{w_n, z_{n,j}}$ on $[\theta_{n,j}, r(D_{n,j})]$, the constant path at $\frac{1}{2}(w_n + z_{n,j})$ on $[r(D_{n,j}), \ell(D_{n,j+1})]$, and the linear reparameterization of $\bar{L}_{w_n, z_{n,j}}$ on $[\ell(D_{n,j+1}), \theta_{n,j+1}]$.

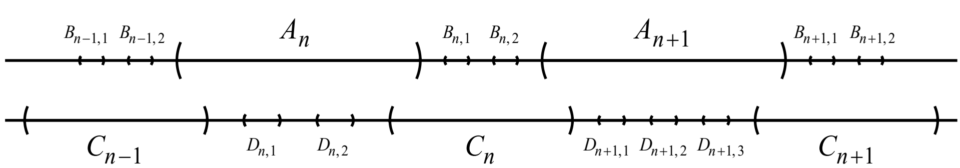


Figure 4. Top: the selected sets A_n and $B_{n,j}$ on which α is constant. Bottom: the selected sets C_n and $D_{n,j}$ on which β is constant.

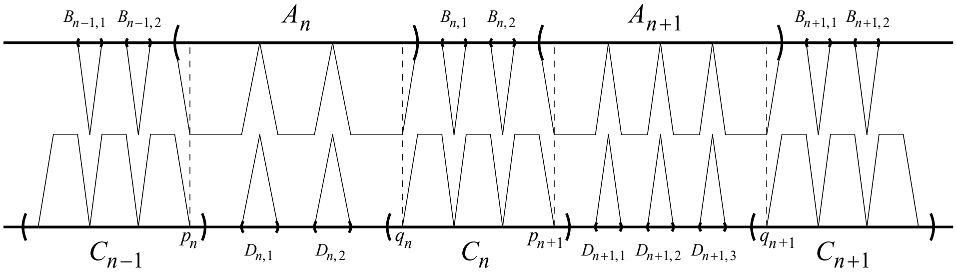


Figure 5. The interlocking pattern that determines the structure of α' and β'. The triangles represent inserted inverse-pair loops and the trapezoids represent an inverse pair loop but with a constant path included in the middle. The three subintervals of $\overline{A}_n$ and the five subintervals of $\overline{C}_n$ partitioned by the trapezoids correspond to the piecewise definitions (A1)-(A3) and (B1)-(B5) respectively.

(A3) On $[\theta_{n,m_n}, r(A_n)]$, we define α' to be the CIP loop, which is the linear reparameterization of $L_{w_n,z_{n,m_n}}$ on $[\theta_{n,m_n}, r(D_{n,m_n})]$, the constant path at $\frac{1}{2}(w_n + z_{n,m_n})$ on $[r(D_{n,m_n}), q_n]$, and the linear reparameterization of $\overline{L}_{w_n,z_{n,m_n}}$ on $[q_n, r(A_n)]$. This completes the definition of α' on $\overline{A}_n$.

(A4) Lastly, on $\overline{B}_{n,j}$, we define α' to be the linear reparameterization of the CIP loop $L_{x_{n,j},y_n}\overline{L}_{x_{n,j},y_n}$.

This completes the definition of α' (compare Figures 5 and 6). Note that α' agrees with α everywhere except on a $\mathbb{Z}$-ordered sequence of elements of $\mathrm{lc}(\alpha)$ on which CIP loops replace constant loops. Hence, it is clear that α' is a well-defined function.

To begin our definition of β', we set β' to agree with β on $\{0, 1\}$ and on $[r(C_{n-1}), \ell(C_n)] \setminus \bigcup_{j=1}^{m_n} D_{n,j}$ for each $n \in \mathbb{Z}$. We complete the definition of β' piecewise by fixing $n \in \mathbb{N}$ and defining β' on $\overline{C}_n$ in five cases and $\overline{D}_{n,j}$ for $1 \le j \le m_n$ in a sixth case.

(B1) On $[\ell(C_n), q_n]$, we define β' to be constant at y_n. (This happens to agree with the value of β.)

(B2) On $[q_n, \eta_{n,1}]$, we define β' to be CIP loops, which is the linear reparameterization of L_{y_n,w_n} on $[q_n, r(A_n)]$, the constant path at $\frac{1}{2}(y_n + w_n)$ on $[r(A_n), \ell(B_{n,1})]$, and $\overline{L}_{y_n,w_n}$ on $[\ell(B_{n,1}), \eta_{n,1}]$.

(B3) On $[\eta_{n,j}, \eta_{n,j+1}]$ (for each $1 \le j \le k_n - 1$), we define β' to be CIP loops, which is the linear reparameterization of $L_{y_n,x_{n,j}}$ on $[\eta_{n,j}, r(B_{n,j})]$, the constant path at $\frac{1}{2}(y_n + x_{n,j})$ on $[r(B_{n,j}), \ell(B_{n,j+1})]$, and the linear reparameterization of $\overline{L}_{y_n,x_{n,j}}$ on $[\ell(B_{n,j+1}), \eta_{n,j+1}]$.

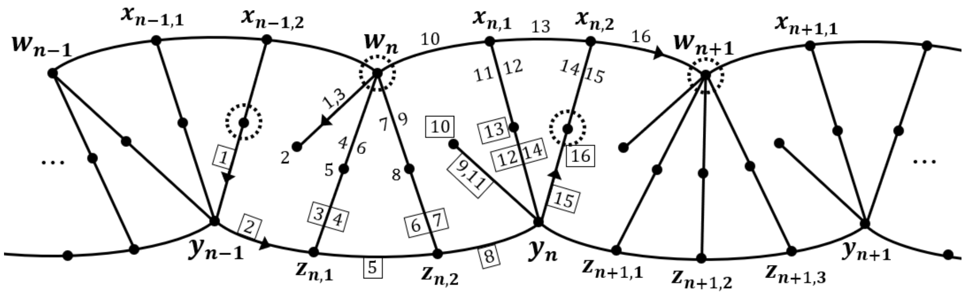


Figure 6. A full period of the construction of α' and β' starting with $\ell(A_n)$ and ending with $\ell(A_{n+1})$. Starting and ending points are circled and the initial and terminal steps are indicated with arrows. The upper and lower curves represent respectively α and β, where each subdivided segment has diameter less than $\frac{1}{6}\delta$. The numbered paths trace out the trajectory of α' and the box-numbered paths trace out the corresponding trajectory of β'. When a number is positioned at a point, the path is constant at that point.

(B4) On $[\eta_{n,k_n}, p_{n+1}]$, we define β' to be the CIP loop given by the linear reparameterization of $L_{y_n,x_{n,k_n}}$ on $[\eta_{n,k_n}, r(B_{n,k_n})]$, the constant path at $\frac{1}{2}(y_n + x_{n,k_n})$ on $[r(B_{n,k_n}), \ell(A_{n+1})]$, and the linear reparameterization of $\overline{L}_{y_n,x_{n,k_n}}$ on $[\ell(A_{n+1}), p_{n+1}]$.

(B5) On $[p_{n+1}, r(C_n)]$, we define β' to be constant at y_n (this happens to agree with the value of β). This completes the definition of β' on $\overline{C}_n$.

(B6) On $\overline{D}_{n,j}$ (for each $1 \le j \le m_n$), we define β' to be the linear reparameterization of the CIP loop $L_{z_{n,j},w_n}\overline{L}_{z_{n,j},w_n}$.

This completes the definition of β', which is a well-defined function (see Figures 5 and 6).

By construction, $\alpha'|_{(0,1)}$ is continuous. If U is a convex neighborhood of $\alpha(0) = \beta(0)$ in $\mathbb{D}^2$, then we can find $N \in \mathbb{Z}$ such that $\alpha(A_n \cup C_n) \cup \beta(A_n \cup C_n) \subseteq U$ for all $n \le N$. Let $t = \sup(A_N)$. Since the CIP loops added to α on $[0, t]$ are contained in line segments with respective endpoints in $\alpha([0, t])$ and $\beta([0, t])$, it follows that $\alpha'([0, t]) \subseteq U$. Thus α' is continuous at 0. A symmetric argument shows that α' is continuous at 1. The construction of α' also ensures that distinct elements of $\mathrm{lc}(\alpha')$ have disjoint closures. Hence, α' is a Cantor path. Since α' is constructed from α only by replacing constant loops with CIP loops on elements of $\mathrm{lc}(\alpha)$ (one on each $B_{n,j}$ and at least three on each A_n), we have $\alpha \preceq_* \alpha'$. Moreover, if $\mu : \overline{I} \to \mathbb{D}^2$ is one of the added CIP loops in the construction of α', then the image of μ is the line segment connecting $\alpha(s)$ and $\frac{1}{2}(\alpha(s) + \beta(s))$ for some $s \in \overline{I}$. Thus $\rho(\alpha, \alpha') \le \frac{1}{2}\rho(\alpha, \beta)$. Since these arguments apply just as well for β', we also conclude that β' is a Cantor path satisfying $\beta \preceq_* \beta'$ and $\rho(\beta, \beta') \le \frac{1}{2}\rho(\alpha, \beta)$.

To complete the proof, we will show that $d(\alpha'(s), \beta'(s)) \le \frac{2}{3}\delta$. We begin by considering the case $s \in \overline{A}_n$.

(1) On $[\ell(A_n), p_n]$, α' parameterizes the line from w_n to $\frac{1}{2}(w_n + y_{n-1})$ (see the first part of (A1)) and β' is a corresponding parameterization of the line from $\frac{1}{2}(x_{n-1,k_{n-1}} + y_{n-1})$ to y_{n-1} (see the last part of (B4)). Thus if $s \in [\ell(A_n), p_n]$, Remark 5.2 gives us the first inequality in the following sequence:

$$\begin{aligned}
&d(\alpha'(s), \beta'(s))\\
&\le \max\left\{\tfrac{1}{2}d(x_{n-1,k_{n-1}}, y_{n-1}) + d(x_{n-1,k_{n-1}}, w_n), \tfrac{1}{2}d(w_n, y_{n-1})\right\}\\
&= \max\left\{\tfrac{1}{2}d\big(\alpha(r(B_{n-1,k_{n-1}})), \beta(r(B_{n-1,k_{n-1}}))\big) + d(x_{n-1,k_{n-1}}, w_n), \tfrac{1}{2}d(\alpha(p_n), \beta(p_n))\right\}\\
&< \max\left\{\tfrac{1}{2}\delta + \tfrac{1}{6}\delta, \tfrac{1}{2}\delta\right\} \le \tfrac{2}{3}\delta.
\end{aligned}$$

(2) On $[p_n, r(C_n)]$, α' is constant at $\frac{1}{2}(w_n + y_{n-1})$ (see middle part of (A1)) and β' is constant at y_{n-1} (see (B5)). Therefore, if $s \in [p_n, r(C_n)]$, we have

$$d(\alpha'(s), \beta'(s)) = \tfrac{1}{2}d(w_n, y_{n-1}) = \tfrac{1}{2}d(\alpha(p_n), \beta(p_n)) \le \tfrac{1}{2}\delta.$$

(3) On $[r(C_n), \ell(D_{n,1})]$, α' is constant at $\frac{1}{2}(w_n + y_{n-1})$ (see middle part of (A1)) and β' agrees with β. Thus if $s \in [r(C_n), \ell(D_{n,1})]$, then

$$\begin{aligned}
d(\alpha'(s), \beta'(s)) &\le d(\alpha'(s), \beta(r(C_{n-1}))) + d(\beta(r(C_{n-1})), \beta'(s))\\
&= d\big(\tfrac{1}{2}w_n + y_{n-1}, y_{n-1}\big) + d(\beta(r(C_{n-1})), \beta(s))\\
&= \tfrac{1}{2}d(\alpha(p_n), \beta(p_n)) + d(\beta(r(C_{n-1})), \beta(s))\\
&< \tfrac{1}{2}\delta + \tfrac{1}{6}\delta \le \tfrac{2}{3}\delta.
\end{aligned}$$

(4) On $[\ell(D_{n,1}), \theta_{n,1}]$, α' parameterizes the line segment from $\frac{1}{2}(w_n + y_{n-1})$ to w_n (see last part of (A1)) and β' parameterizes the line segment from $z_{n,1}$ to $\frac{1}{2}(z_{n,1} + w_n)$. In this case, if $s \in [\ell(D_{n,1}), \theta_{n,1}]$, then

$$\begin{aligned}
d(\alpha'(s), \beta'(s)) &\le \max\left\{d(w_n, \tfrac{1}{2}(w_n + y_{n-1})), d(z_{n,1}, \tfrac{1}{2}(z_{n,1} + w_n))\right\}\\
&= \max\left\{\tfrac{1}{2}d(w_n, y_{n-1}), \tfrac{1}{2}d(z_{n,1}, w_n)\right\}\\
&= \max\left\{\tfrac{1}{2}d(\alpha(p_n), \beta(p_n)), \tfrac{1}{2}d\big(\alpha(\theta_{n,1}), \beta(\theta_{n,1})\big)\right\}\\
&\le \tfrac{1}{2}\delta < \tfrac{2}{3}\delta.
\end{aligned}$$

(5) On $[\theta_{n,1}, r(D_{n,1})]$, α' parameterizes the line segment from w_n to $\frac{1}{2}(z_{n,1} + w_n)$ (see first part of (A2)) and β parameterizes the line segment from $\frac{1}{2}(z_{n,1} + w_n)$ to $z_{n,1}$ (see (B6)). Since these paths move along the same line segment, it follows that if $s \in [\theta_{n,1}, r(D_{n,1})]$, then

$$d(\alpha'(s), \beta'(s)) = \tfrac{1}{2}d(w_n, z_{n,1}) \le \tfrac{1}{2}d(\alpha(\theta_{n,1}), \beta(\theta_{n,1})) \le \tfrac{1}{2}\delta.$$

As we proceed through the remainder of the intervals on which α' and β' are defined piecewise, every remaining case (all of which are illustrated in Figures 5 and 6) may be verified using an argument nearly identical to one of the above five cases. Hence, we omit the remainder of the details. We conclude that $\rho(\alpha', \beta') \le \frac{2}{3}\delta$. □

The construction given in the proof of Lemma 5.3 results in two Cantor paths α' and β', which are not staggered. Hence, to iterate this construction, we must perturb one of these two paths so that they become staggered.

Lemma 5.4. *For given staggered Cantor paths $\alpha, \beta : [0, 1] \to \mathbb{D}^2$ with $\alpha(i) = \beta(i)$ for $i \in \{0, 1\}$ and $\delta = \rho(\alpha, \beta)$, there exists sequences of Cantor paths $\{\alpha_n\}_{n\ge0}$, $\{\alpha'_n\}_{n\ge0}$, $\{\beta_n\}_{n\ge0}$, such that $\alpha_0 = \alpha$, $\beta_0 = \beta$ and that, for all $n \ge 0$,*

(1) $\alpha_n \preceq_* \alpha'_n$ *and* $\beta_n \preceq_* \beta_{n+1}$,

(2) α_n *and* β_n *are staggered*,

(3) $\alpha_{n+1} = \alpha'_n \circ f_n$ *for some increasing homeomorphism* $f_n : [0, 1] \to [0, 1]$ *with* $\rho(f_n, \mathrm{id}_{[0,1]}) < 2^{-n}$, *and*

(4) $\max\left\{\rho(\alpha'_n, \alpha_{n+1}), \rho(\alpha_n, \alpha_{n+1}), \rho(\beta_n, \beta_{n+1}), \rho(\alpha_n, \beta_n)\right\} \le \delta\left(\frac{3}{4}\right)^n$.

Proof. Let $\alpha_0 = \alpha$, $\beta_0 = \beta$, and $\delta = \rho(\alpha_0, \beta_0)$. Suppose α_n and β_n are given staggered Cantor paths that satisfy $\rho(\alpha_n, \beta_n) \le \delta\left(\frac{3}{4}\right)^n$. Applying Lemma 5.3, find Cantor paths $\alpha'_n, \beta'_n : [0, 1] \to \mathbb{D}^2$ such that $\alpha_n \preceq_* \alpha'_n$, $\beta_n \preceq_* \beta'_n$, $\max\left\{\rho(\alpha_n, \alpha'_n), \rho(\beta_n, \beta'_n)\right\} \le \frac{1}{2}\rho(\alpha_n, \beta_n)$, and $\rho(\alpha'_n, \beta'_n) \le \frac{2}{3}\delta$. By Lemma 3.7, there exists an increasing homeomorphism $f_n : [0, 1] \to [0, 1]$ such that $\rho(f_n, \mathrm{id}_{[0,1]}) < 2^{-n}$, $\rho(\alpha'_n \circ f_n, \alpha'_n) < \frac{1}{12}\rho(\alpha_n, \beta_n)$, and that $\alpha'_n \circ f_n$ and β'_n are staggered. Set $\alpha_{n+1} = \alpha'_n \circ f_n$ and $\beta_{n+1} = \beta'_n$. Then α_{n+1} and β_{n+1} are staggered Cantor paths and satisfy the following inequalities:

$$\rho(\alpha'_n, \alpha_{n+1}) < \tfrac{1}{12}\rho(\alpha_n, \beta_n) \le \tfrac{1}{12}\delta\left(\tfrac{3}{4}\right)^n \le \delta\left(\tfrac{3}{4}\right)^{n+1},$$

$$\begin{aligned}\rho(\alpha_n, \alpha_{n+1}) &\le \rho(\alpha_n, \alpha'_n) + \rho(\alpha'_n, \alpha_{n+1})\\ &< \tfrac{1}{2}\rho(\alpha_n, \beta_n) + \tfrac{1}{12}\rho(\alpha_n, \beta_n) < \tfrac{3}{4}\rho(\alpha_n, \beta_n) \le \delta\left(\tfrac{3}{4}\right)^{n+1},\end{aligned}$$

$$\rho(\beta_n, \beta_{n+1}) \le \tfrac{1}{2}\rho(\alpha_n, \beta_n) \le \tfrac{3}{4}\rho(\alpha_n, \beta_n) \le \delta\left(\tfrac{3}{4}\right)^{n+1},$$

$$\begin{aligned}\rho(\alpha_{n+1}, \beta_{n+1}) &\le \rho(\alpha_{n+1}, \alpha'_n) + \rho(\alpha'_n, \beta'_n)\\ &\le \tfrac{1}{12}\rho(\alpha_n, \beta_n) + \tfrac{2}{3}\rho(\alpha_n, \beta_n) = \tfrac{3}{4}\rho(\alpha_n, \beta_n) \le \delta\left(\tfrac{3}{4}\right)^{n+1}.\end{aligned}$$

This completes the inductive construction of the desired sequences. □

In the next two statements, we assume α, β are fixed staggered Cantor paths as given in the hypothesis of Lemma 5.4.

Proposition 5.5. *The sequences $\{\alpha_n\}_{n\geq 0}$ and $\{\beta_n\}_{n\geq 0}$ constructed in the proof of Lemma 5.4 both converge uniformly to a single path $\gamma : [0, 1] \to \mathbb{D}^2$.*

Proof. Recall that $\delta = \rho(\alpha_0, \beta_0)$ is fixed. Since $\rho(\alpha_n, \alpha_{n+1}) \leq \delta\left(\frac{3}{4}\right)^n$ for all $n \geq 0$, $\{\alpha_n\}_{n\geq 0}$ is Cauchy in the sup metric and so converges uniformly to some path $\gamma : [0, 1] \to \mathbb{D}^2$. Since the sequence $\{\beta_n\}_{n\geq 0}$ satisfies the same inequality, $\{\beta_n\}_{n\geq 0}$ also converges uniformly to some path. Additionally, since $\rho(\alpha_n, \beta_n) \leq \delta\left(\frac{3}{4}\right)^n$ for all $n \geq 0$, $\{\alpha_n\}_{n\geq 0}$ and $\{\beta_n\}_{n\geq 0}$ must both converge uniformly to γ. □

Lemma 5.6. *If $\gamma : [0, 1] \to \mathbb{D}^2$ is the uniform limit of the sequences $\{\alpha_n\}_{n\geq 0}$ and $\{\beta_n\}_{n\geq 0}$ as given in the conclusion of Proposition 5.5, then $\alpha \simeq_{\mathbb{R}} \gamma$ and $\beta \simeq_{\mathbb{R}} \gamma$.*

Proof. For given paths α and β, Lemma 5.4 gives sequences $\{\alpha_n\}_{n\geq 0}$, $\{\alpha'_n\}_{n\geq 0}$, $\{f_n\}_{n\in\mathbb{N}}$, and $\{\beta_n\}_{n\geq 0}$ satisfying a variety of relations and inequalities. Proposition 5.5 ensures that $\{\alpha_n\}_{n\geq 0}$ and $\{\beta_n\}_{n\geq 0}$ converge uniformly to a path γ. The sequences $\{\alpha_n\}_{n\geq 0}$, $\{\alpha'_n\}_{n\geq 0}$, $\{f_n\}_{n\in\mathbb{N}}$ and the limit path γ satisfy the hypotheses of Lemma 4.8. It follows that $\alpha_n \simeq_{\mathbb{R}} \gamma$ for all $n \geq 0$. In particular, $\alpha = \alpha_0 \simeq_{\mathbb{R}} \gamma$. Similarly, we may apply Lemma 4.8 to the sequence $\{\beta_n\}_{n\geq 0}$ in the case where $\beta_n = \beta'_n$ and $f_n = \mathrm{id}$ for all $n \geq 0$ (or we could apply Lemma 4.6). Thus, $\beta = \beta_0 \simeq_{\mathbb{R}} \gamma$. □

Proof of Theorem 1.4. First, we prove the special case $X = \mathbb{D}^2$. Respectively, let a and b be the injective paths in $\mathbb{D}^2$ from $(1, 0)$ to $(-1, 0)$ that parameterize the upper and lower semicircles of S^1. Let $\tau : [0, 1] \to [0, 1]$ be the ternary Cantor map and note that $a \circ \tau$ and $b \circ \tau$ are Cantor paths. Set $b_0 = b \circ \tau$. By Lemma 3.7 (taking $\epsilon = 1$), there exists an increasing homeomorphism $f : [0, 1] \to [0, 1]$ such that $a_0 = a \circ \tau \circ f$ and b_0 are staggered Cantor paths. Applying Lemma 5.6 to a_0 and b_0, we obtain the existence of a path $c : [0, 1] \to \mathbb{D}^2$ such that $a_0 \simeq_{\mathbb{R}} c$ and $b_0 \simeq_{\mathbb{R}} c$. Since $a \approxeq a_0$ and $b \approxeq b_0$, Lemma 2.4(4) gives that $a \simeq_{\mathbb{R}} c$ and $b \simeq_{\mathbb{R}} c$. Since a and b are injective paths, it follows that $c \geq_{\mathbb{R}} a$ and $c \geq_{\mathbb{R}} b$.

In the general case, suppose $\alpha, \beta : [0, 1] \to X$ are path homotopic. Find a map $f : \mathbb{D}^2 \to X$ such that $f \circ a = \alpha$ and $f \circ b = \beta$. By Lemma 2.4(1), the path $\gamma = f \circ c$ in X satisfies $\gamma \geq_{\mathbb{R}} \alpha$ and $\gamma \geq_{\mathbb{R}} \beta$. □

Corollary 5.7. *The equivalence relation on the set of paths in a given topological space generated by $\geq_{\mathbb{R}}$ (and $\simeq_{\mathbb{R}}$) coincides with path homotopy.*

If X is one-dimensional, then a loop is null-homotopic if and only if it factors through a loop in a dendrite; see [6, Theorem 3.7] for the nontrivial implication. Thus, for one-dimensional spaces, the relation $\simeq_{\mathbb{R}}$ (see Definition 2.2) is equivalent to the homotopy rel. endpoints relation and hence is transitive. Since $\mathbb{D}^2$ contains simple closed curves (parameterizations of which cannot factor through an $\mathbb{R}$-tree), Theorem 1.4 implies the following.

Corollary 5.8. *If $\mathbb{D}^2$ embeds in X, then the $\mathbb{R}$-tree homotopy relation on the set of paths in X is not transitive.*

6. A Solution to Dydak's Problem

We conclude with a proof of Lemma 1.2 and Theorem 1.1. The following auxiliary result is proved using standard techniques from covering space theory.

Lemma 6.1. *Let $p : (E, e_0) \to (X, x_0)$ be a based map with unique lifting of paths rel. starting point and suppose T is an $\mathbb{R}$-tree. If $f : (T, t_0) \to (X, x_0)$ is a based map, then there exists a unique based map $\tilde{f} : (T, t_0) \to (E, e_0)$ such that $p \circ \tilde{f} = f$.*

Proof of Lemma 1.2. Let $p : E \to X$ be a map where X is first countable, locally path connected and simply connected and such that every path in X has a unique lift in E rel. starting point. Since X is path connected, p is surjective. Suppose that $p(e_1) = p(e_2) = x$ for $e_1, e_2 \in E$. Let $\tilde{\beta} : [0, 1] \to E$ be a path from e_1 to e_2. Then $\beta = p \circ \tilde{\beta}$ is a loop based at x. Let $\alpha : [0, 1] \to X$ be the constant path at x. Since X is simply connected, $\alpha \simeq \beta$ and thus, by Theorem 1.4, there exists a path $\gamma : [0, 1] \to X$ such that $\gamma \geq_{\mathbb{R}} \alpha$ and $\gamma \geq_{\mathbb{R}} \beta$. Let $\tilde{\alpha}, \tilde{\gamma} : ([0, 1], 0) \to (E, e_1)$ be the lifts of α and γ and note that $\tilde{\alpha}$ is constant at e_1. Since $\alpha\bar{\gamma}$ factors through an $\mathbb{R}$-tree, it follows from Lemma 6.1 that $\tilde{\alpha}(1) = \tilde{\gamma}(1)$ in E. Similarly, $\tilde{\beta}(1) = \tilde{\gamma}(1)$. Thus $e_1 = \tilde{\gamma}(1) = e_2$, proving that p is injective. Since X is first countable, it suffices to show $p^{-1} : X \to E$ preserves convergent sequences. If $\{x_n\} \to x$ is a convergent sequence in X, the hypotheses on X allow us to find a path $\alpha : [0, 1] \to X$ with $\alpha(1/n) = x_n$ for all $n \in \mathbb{N}$ and $\alpha(0) = x$. There is one lift $\tilde{\alpha} : [0, 1] \to E$ for which $p \circ \tilde{\alpha} = \alpha$ and it satisfies $\tilde{\alpha}(1/n) = p^{-1}(x_n)$ and $\tilde{\alpha}(0) = p^{-1}(x)$. Since $\tilde{\alpha}$ is continuous, $\{p^{-1}(x_n)\} \to p^{-1}(x)$ in E. Thus p^{-1} is continuous. □

Proof of Theorem 1.1. Suppose that a group G acts on an $\mathbb{R}$-tree T in such a way that the quotient map $p : T \to T/G$ is a UPL map. Suppose that T/G contains a two-dimensional Euclidean disc D. Let E be a path-component of $p^{-1}(D)$. Then $p|_E : E \to D$ is a UPL map over a first countable, locally path-connected, and simply connected space D. Hence $p|_E$ is a homeomorphism, which is impossible since E is a subspace of an $\mathbb{R}$-tree. □

Proof of Corollary 1.3. Lemma 1.2 implies that Dydak's unique lifting problem has a positive answer. It is an exercise to see that the evaluation map from $P(X, x)$ to X is an open surjection, since X is locally path-connected [5, Section 2.1]. Then Corollary 1.3 follows immediately from [5, Corollary 4.10]. □

Appendix: Proof of Lemma 4.6

To prove Lemma 4.6, we must first detail the structure of a single $\mathscr{U}$-extension. For the moment, suppose that $\beta : [0, 1] \to X$ is a given $\mathscr{U}$-extension of $\alpha : [0, 1] \to X$, where both are Cantor paths. Additionally, we fix

(1) an lc(α)-collapsing map k_α and light path α^λ such that $\alpha^\lambda \circ k_\alpha = \alpha$, and

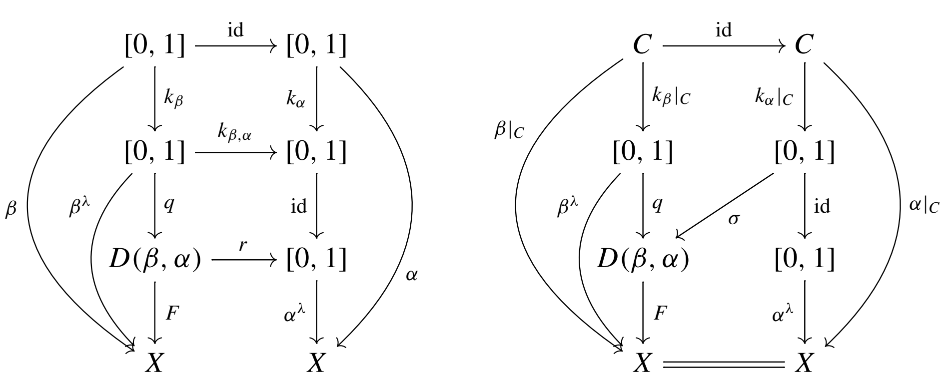
(2) an lc(β)-collapsing map k_β and light path β^λ such that $\beta^\lambda \circ k_\beta = \beta$.

It follows from Remark 4.4 that there is a unique map $k_{\beta,\alpha} : [0, 1] \to [0, 1]$ such that $k_\alpha = k_{\beta,\alpha} \circ k_\beta$. Let $\mathscr{V} = \{k_\beta(J) \mid J \in \mathscr{U}\}$ be the collection of open intervals that are the images of the elements of $\mathscr{U}$. If $J \in \mathscr{U}$ and $(c, d) = k_\beta(J) \in \mathscr{V}$, then $\beta|_{\bar{J}}$ is a CIP loop and $(\beta^\lambda)|_{[c,d]}$ is a LIP loop. Thus we have the equivalence $(\beta^\lambda)|_{[c,d]} \equiv b_J \bar{b}_J$ for a light path b_J. In particular, there exists $m \in (c, d)$ and a "tent map" $\tau_J : [c, d] \to [0, 1]$ which (1) is an increasing homeomorphism $[c, m] \to [0, 1]$ on $[c, m]$, (2) a decreasing homeomorphism $[m, d] \to [0, 1]$ on $[m, d]$, and (3) satisfies the equality $(\beta^\lambda)|_{[c,d]} = b_J \circ \tau_J$.

Let $\sim$ be the smallest equivalence relation on $[0, 1]$ such that $s \sim t$ if there exists $J \in \mathscr{U}$ such that $s, t \in [c, d] = k_\beta(\bar{J})$ and $\tau_J(s) = \tau_J(t)$. Set $D(\beta, \alpha) = [0, 1]/{\sim}$ and let $q : [0, 1] \to D(\beta, \alpha)$ denote the quotient "folding" map.

Note that $D(\beta, \alpha)$ is constructed by folding each interval $[c, d] = k_\beta(\bar{J})$, $J \in \mathscr{U}$ in half according to the tent map τ_J. Hence, $D(\beta, \alpha)$ is a dendrite consisting of the base arc $B = q\big([0, 1]\backslash \bigcup \mathscr{V}\big)$ and possibly infinitely many attached arcs. In particular, the arc $A_J = q([c, d])$ meets B at the point $q(\{c, d\})$ and has free-endpoint $q(m)$ (where m is defined as above).

The definition of $\sim$ ensures that β^λ is constant on the fibers of q and, therefore, there is a unique map $F : D(\beta, \alpha) \to X$ such that $F \circ q = \beta^\lambda$. Recall that each fiber of k_β is contained in (and possibly equal to) a fiber of k_α. If $q(s) = q(t)$ for $s \neq t$, then $k_\beta^{-1}(s)$ and $k_\beta^{-1}(t)$ lie in the same fiber of k_α (the closure of some element of $\mathscr{U}$). Therefore, k_α is constant on the fibers of $q \circ k_\beta$ and there is a unique map $r : D(\beta, \alpha) \to [0, 1]$ such that $k_\alpha = r \circ q \circ k_\beta$. In particular, r maps the base arc B homeomorphically onto $[0, 1]$ and if $(c, d) = k_\beta(J)$ for $J \in \mathscr{U}$, then r maps the arc A_J to the point $k_{\beta,\alpha}([c, d]) = k_\alpha(\bar{J})$ in $[0, 1]$. It follows that r is a monotone map. Finally, since $r \circ q \circ k_\beta = k_\alpha = k_{\beta,\alpha} \circ k_\beta$ where k_β is surjective, we have $r \circ q = k_{\beta,\alpha}$. Overall, the following diagram on the left commutes. We address the diagram on the right in the next proposition.



Proposition A.2. *The map* $r : D(\beta, \alpha) \to [0, 1]$ *is a retraction. The unique section* $\sigma : [0, 1] \to D(\beta, \alpha)$, *which satisfies* $r \circ \sigma = \mathrm{id}$, *parameterizes the base arc* B *and satisfies* $F \circ \sigma = \alpha^\lambda$.

Proof. Let $C = [0, 1] \setminus \bigcup \mathscr{U}$ and note that the restrictions $q \circ k_\beta|_C : C \to B$ and $(k_\alpha)|_C : C \to [0, 1]$ are quotient maps that make the same identifications, namely they collapse intervals $\bar{J}$ for $J \in \mathrm{lc}(\alpha) \setminus \mathscr{U}$ and identify endpoints of intervals $J \in \mathscr{U}$. Thus, there exists a unique homeomorphism $\sigma : [0, 1] \to B$ such that $\sigma \circ (k_\alpha)|_C = q \circ k_\beta|_C$ (see the right diagram on the previous page). Since $r \circ \sigma \circ (k_\alpha)|_C = r \circ q \circ k_\beta|_C = (k_\alpha)|_C$, where $(k_\alpha)|_C$ is surjective, we have $r \circ \sigma = \mathrm{id}_{[0,1]}$. Overall, r is a retraction whose section σ parameterizes B.

One can use the same diagram to confirm that $F \circ \sigma \circ (k_\alpha)|_C = \beta|_C$. By definition of β being a $\mathscr{U}$-extension of α, we have $\alpha|_C = \beta|_C$. Thus $F \circ \sigma \circ (k_\alpha)|_C = \alpha|_C = \alpha^\lambda \circ (k_\alpha)|_C$. Since $(k_\alpha)|_C$ is surjective, we have the desired equality $F \circ \sigma = \alpha^\lambda$. □

Corollary A.3. *If* $\alpha \preceq_* \beta$, *then* $\alpha \simeq_{\mathbb{R}} \beta$ *and* $\alpha^\lambda \leq_{\mathbb{R}} \beta^\lambda$.

Proof. Note that $\sigma \circ k_\alpha$ and $q \circ k_\beta$ are paths in the compact ℝ-tree $D(\beta, \alpha)$ with the same endpoints and satisfying $F \circ (\sigma \circ k_\alpha) = \alpha$ and $F \circ (q \circ k_\beta) = \beta$. Thus $\alpha \simeq_{\mathbb{R}} \beta$. Since $F \circ q = \beta^\lambda$ and $F \circ \sigma = \alpha^\lambda$ where σ is injective, we have $\alpha^\lambda \leq_{\mathbb{R}} \beta^\lambda$. □

Again, we suppose that β is a $\mathscr{U}$-extension of α (where both are Cantor paths) and we reuse the above notation. However, now we suppose also that there exists a dendrite D, a map $H : D \to X$, and a surjective path $Q : [0, 1] \to D$ such that $H \circ Q = \alpha^\lambda$ (surjectivity of Q is not required for the following construction but appears naturally in our recursive application so we assume it). We use this data and the above construction of $D(\beta, \alpha)$ to uniquely determine a factorization of β^λ.

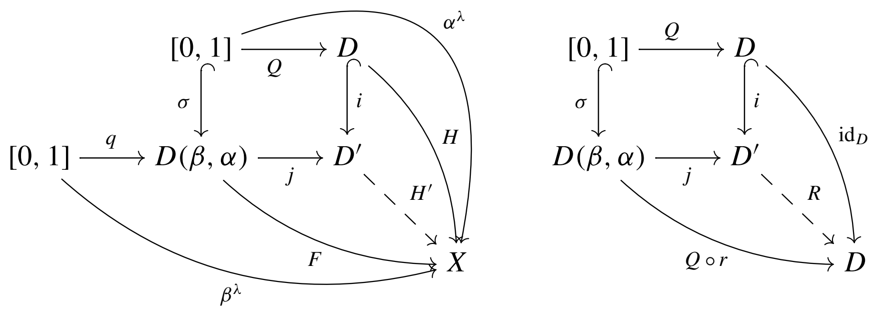
Let D' be the pushout of $Q : [0, 1] \to D$ and the section $\sigma : [0, 1] \to D(\beta, \alpha)$. We have the following pushout square:

$$\begin{array}{ccc} [0,1] & \xrightarrow{\;Q\;} & D \\ {\scriptstyle \sigma}\big\downarrow & & \big\downarrow{\scriptstyle i} \\ D(\beta,\alpha) & \xrightarrow[\;j\;]{} & D' \end{array}$$

Note that i is injective since σ is injective. Also, j is surjective since Q is surjective. All of the domains being compact ensures that i is an embedding and j is a quotient map. Observe that since $i(D)$ is a dendrite and $D' \setminus i(D) \cong D(\beta, \alpha) \setminus \sigma([0, 1])$ is a disjoint union of half-open arcs, D' is a dendrite.

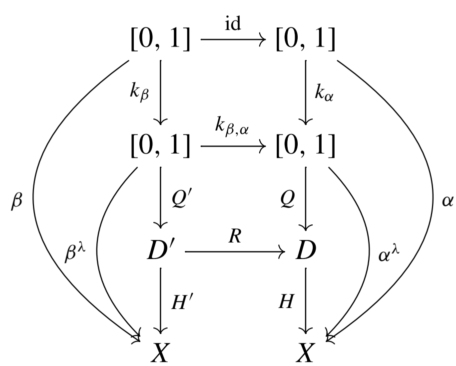
Since $H \circ Q = \alpha^\lambda = F \circ \sigma$, there exists a unique map $H' : D' \to X$ making the left diagram below commute. Set $Q' = j \circ q : [0, 1] \to D'$. Then $H' \circ Q' = H' \circ j \circ q = F \circ q = \beta^\lambda$. Additionally, since $r \circ \sigma = \mathrm{id}_{[0,1]}$ the right diagram below shows that there exists a map $R : D' \to D$ such that $R \circ i = \mathrm{id}_D$ and $R \circ j = Q \circ r$.

Thus R is a retraction with section i. Additionally, note that R is monotone, since r is monotone.



We also have $R \circ Q' = R \circ j \circ q = Q \circ r \circ q = Q \circ k_{\beta,\alpha}$ and $H' \circ i \circ Q = \alpha^{\lambda}$. Finally, note that the paths $i \circ Q \circ k_{\alpha} : [0, 1] \to D'$ and $Q' \circ k_{\beta} : [0, 1] \to D'$ start and end at the same points and that $H' \circ i \circ Q \circ k_{\alpha} = \alpha$ and $H' \circ Q' \circ k_{\beta} = \beta$. This gives a factorization of the loop $\alpha\bar{\beta} : [0, 1] \to X$ through the dendrite D'. Overall, we conclude the following, which employs the notation in the construction of both $D(\beta, \alpha)$ and D'.

Lemma A.4. *Let Cantor path β be a $\mathscr{U}$-extension of another Cantor path α and let k_α and k_β be collapsing maps for these paths respectively. Suppose there exists a dendrite D, a map $H : D \to X$, and a surjective path $Q : [0, 1] \to D$ such that $H \circ Q = \alpha^{\lambda}$. Then there exists a dendrite D' (constructed from the pushout square $i \circ Q = j \circ \sigma$), a map $Q' : [0, 1] \to D'$ defined as $Q = j \circ q$, a monotone retraction $R : D' \to D$ with section i, and a map $H' : D' \to X$ such that the following diagram commutes*:



Moreover, $(i \circ Q \circ k_{\alpha})(\overline{Q' \circ k_{\beta}})$ *is a well-defined loop in D' satisfying*

$$\alpha\bar{\beta} = H' \circ ((i \circ Q \circ k_{\alpha})(\overline{Q' \circ k_{\beta}})).$$

In the proof of Lemma 4.6, we iterate this construction.

Proof of Lemma 4.6. Since $\gamma_n \preceq_* \gamma_{n+1}$, γ_{n+1} is a $\mathscr{U}_n$-extension of γ_n for some

$\mathscr{U}_n \subseteq \mathrm{lc}(\gamma_n)$. Before we inductively apply the construction from Lemma A.4, we fix collapsing functions k_{γ_n} for the paths γ_n. We then have uniquely determined maps $k_{\gamma_{n+1},\gamma_n} : [0, 1] \to [0, 1]$ such that $k_{\gamma_{n+1},\gamma_n} \circ k_{\gamma_{n+1}} = k_{\gamma_n}$ and light paths γ_n^λ. To simplify notation in the inverse systems to come, we write k_n for k_{γ_n} and $K_{n+1,n}$ for $k_{\gamma_{n+1},\gamma_n}$. Recall that since $\gamma_n \preceq_* \gamma_{n+1}$ for all $n \in \mathbb{N}$, we have a dendrite $D(\gamma_{n+1}, \gamma_n)$, which comes equipped with a corresponding folding map $q_{n+1} : [0, 1] \to D(\gamma_{n+1}, \gamma_n)$, retraction $r_{n+1,n} : D(\gamma_{n+1}, \gamma_n) \to [0, 1]$, and embedding $\sigma_{n,n+1} : [0, 1] \to D(\gamma_{n+1}, \gamma_n)$ such that $r_{n+1,n} \circ q_{n+1} = k_{n+1,n}$ and $r_{n+1,n} \circ \sigma_{n,n+1} = \mathrm{id}_{[0,1]}$.

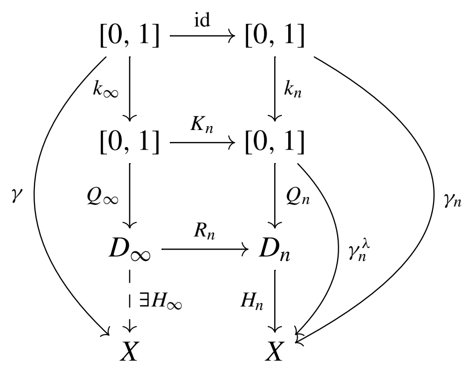
To begin the recursion, set $D_1 = [0, 1]$, $H_1 = \gamma_1^\lambda$, and $Q_1 = \mathrm{id}_{[0,1]}$ so that $H_1 \circ k_1 = \gamma_1$. Suppose that we have given dendrite D_n, map $H_n : D_n \to X$, and quotient map $Q_n : [0, 1] \to D_n$ such that $H_n \circ Q_n = \gamma_n^\lambda$. We apply the construction used in the proof of Lemma A.4 to the case where γ_{n+1} is a $\mathscr{U}_n$-extension of γ_n. We obtain a dendrite D_{n+1} constructed as the pushout of embedding $\sigma_{n,n+1}$ and Q_n. This pushout construction yields a quotient map $j_{n+1} : D(\gamma_{n+1}, \gamma_n) \to D_{n+1}$, and an embedding $i_{n,n+1} : D_n \to D_{n+1}$ such that $i_{n,n+1} \circ Q_n = j_{n+1} \circ \sigma_{n,n+1}$. We also obtain a map $H_{n+1} : D_{n+1} \to X$, a quotient map $Q_{n+1} : [0, 1] \to D_{n+1}$ defined as $Q_{n+1} = j_{n+1} \circ q_{n+1}$, and a retraction $R_{n+1,n} : D_{n+1} \to D_n$. These maps satisfy $H_{n+1} \circ Q_{n+1} = (\gamma_{n+1})^\lambda$, $R_{n+1,n} \circ Q_{n+1} = Q_n \circ k_{n+1,n}$, and $R_{n+1,n} \circ i_{n,n+1} = \mathrm{id}_{D_n}$.

This recursion results in the following infinite diagram where the top three rows form inverse systems. In the n-th column, the vertical composition is γ_n.

$$
\begin{array}{ccccccccccc}
[0,1] & \longrightarrow \cdots & \xrightarrow{\mathrm{id}} & [0,1] & \xrightarrow{\mathrm{id}} & [0,1] & \xrightarrow{\mathrm{id}} & [0,1] & \xrightarrow{\mathrm{id}} & [0,1] & \\
\downarrow{\scriptstyle k_\infty} & & & \downarrow{\scriptstyle k_4} & & \downarrow{\scriptstyle k_3} & & \downarrow{\scriptstyle k_2} & & \downarrow{\scriptstyle k_1} & \\
[0,1] & \longrightarrow \cdots & \xrightarrow{K_{5,4}} & [0,1] & \xrightarrow{K_{4,3}} & [0,1] & \xrightarrow{K_{3,2}} & [0,1] & \xrightarrow{K_{2,1}} & [0,1] & \\
\downarrow{\scriptstyle Q_\infty} & & & \downarrow{\scriptstyle Q_4} & & \downarrow{\scriptstyle Q_3} & & \downarrow{\scriptstyle Q_2} & & \downarrow{\scriptstyle Q_1=\mathrm{id}_{[0,1]}} & \\
D_\infty & \longrightarrow \cdots & \xrightarrow{R_{5,4}} & D_4 & \xrightarrow{R_{4,3}} & D_3 & \xrightarrow{R_{3,2}} & D_2 & \xrightarrow{R_{2,1}} & D_1 & = [0,1] \\
 & & & \downarrow{\scriptstyle H_4} & & \downarrow{\scriptstyle H_3} & & \downarrow{\scriptstyle H_2} & & \downarrow{\scriptstyle H_1=\gamma_1^\lambda} & \\
 & \cdots & & X & & X & & X & & X &
\end{array}
$$

The inverse limit of the top row may be identified with [0, 1] so that the projection maps are also the identity. Since the bonding maps in the second row are nondecreasing continuous surjections, the inverse limit $\varprojlim_n([0, 1], K_{n+1,n})$ may also be identified with [0, 1] and the bonding maps $K_n : [0, 1] \to [0, 1]$ are also nondecreasing continuous surjections, see [7, Theorem 4.8 and Lemma 4.2]. We let $k_\infty = (\varprojlim_n k_n) : [0, 1] \to [0, 1]$ be the inverse limit of the morphisms connecting the first two rows.

In the third row, we have an inverse sequence where the bonding maps $R_{n+1,n}$ are monotone retractions of dendrites. Since any inverse limit of dendrites with monotone bonding maps is a dendrite [17, Theorem 10.36], the inverse limit $D_\infty = \varprojlim_n (D_n, R_{n+1,n})$ is a dendrite and the n-th projection $R_n : D_\infty \to D_n$ is also a retraction. For $m' \geq m$, let $i_{m,m'} : D_m \to D_{m'}$ and $R_{m',m} : D_{m'} \to D_m$ be the respective composition of the sections $i_{n,n+1}$ and retractions $R_{n+1,n}$ (and the identity if $m = m'$). For fixed n, the maps $i_{n,m}$, $m \geq n$ induce a unique map $i_n : D_n \to D_\infty$ such that $R_m \circ i_n = i_{n,m}$. The case $m = n$ shows $R_n \circ i_n = \mathrm{id}_{D_n}$. Finally, let $Q_\infty = \varprojlim_n Q_n$ be the inverse limit of the maps connecting the second and third rows. Then the following diagram commutes for all $n \geq 1$:

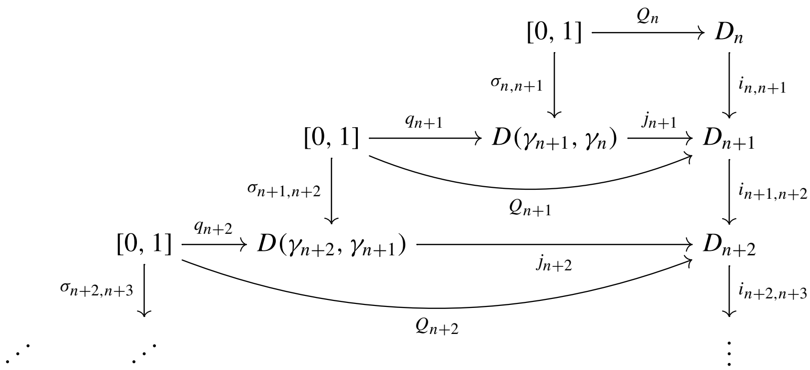


We include γ in the above diagram to indicate that we intend to show that γ is constant on the fibers of $Q_\infty \circ k_\infty$ and therefore induces a unique map $H_\infty : D_\infty \to X$ such that $H_\infty \circ Q_\infty \circ k_\infty = \gamma$. First, we pause to verify that k_∞ and Q_∞ are surjective. In the top two rows, 0 and 1 are identified in the inverse limit with $(0, 0, 0, \dots)$ and $(1, 1, 1, \dots)$ respectively. Since k_∞ is continuous and maps $k_\infty(0) = (k_n(0)) = (0) = 0$ and $k_\infty(1) = (k_n(1)) = (1) = 1$, the connectedness of $[0, 1]$ ensures that k_∞ is surjective.

To check that Q_∞ is surjective, we first show that $i_n(D_n) \subseteq \mathrm{Im}(Q_\infty)$ for all $n \in \mathbb{N}$. If $d_n \in D_n$, we have that $i_n(d_n) = (d_1, d_2, d_3, \dots)$ so that $d_k = i_{n,k}(d_n)$ for $k > n$, and $d_k = R_{n,k}(d_n)$ for $k < n$. Fix $t_n \in [0, 1]$ with $Q_n(t_n) = d_n$. For $k < n$, recursively define $t_{k-1} = K_{k,k-1}(t_k)$. Since $Q_{k-1} \circ K_{k,k-1} = R_{k,k-1} \circ Q_k$, it follows that $Q_k(t_k) = d_k$ for all $1 \leq k < n$. For $k \geq n$, recursively choose points $t_{k+1} \in q_{k+1}^{-1}(\sigma_{k,k+1}(t_k))$ (see the diagram on the next page to trace these choices). From this choice, we have for every $k \geq n$ that

$$Q_{k+1}(t_{k+1}) = j_{k+1} \circ q_{k+1}(t_{k+1}) = j_{k+1}(\sigma_{k,k+1}(t_k)) = i_{k,k+1} \circ Q_k(t_k)$$

and so, by induction, $Q_k(t_k) = d_k$ for all $k \geq n$. Hence $Q_\infty(t_1, t_2, t_3, \dots) = i_n(d_n)$, proving that $i_n(D_n) \subseteq \mathrm{Im}(Q_\infty)$. Now, consider any element $x = (d_1, d_2, d_3, \dots)$



of D_∞. We have

$$x_n = (d_1, \dots, d_{n-1}, d_n, i_{n,n+1}(d_n), i_{n,n+2}(d_n), \dots) \in i_n(D_n)$$

for all $n \in \mathbb{N}$. Since D_∞ is topologized as a subspace of $\prod_{n\in\mathbb{N}} D_n$, we have $\{x_n\}_{n\in\mathbb{N}} \to x$ in D_∞. For each $n \in \mathbb{N}$, find $t_n \in [0, 1]$ such that $Q_\infty(t_n) = x_n$. Since $[0, 1]$ is compact, we can find a subsequence $\{t_{n_m}\}_{m\in\mathbb{N}}$ such that $\{t_{n_m}\}_{m\in\mathbb{N}} \to t$ for some $t \in [0, 1]$. Since $\{Q_\infty(t_{n_m})\}_{m\in\mathbb{N}} = \{x_{n_m}\}_{m\in\mathbb{N}} \to x$ and $\{Q_\infty(t_{n_m})\}_{m\in\mathbb{N}} \to Q_\infty(t)$, it follows that $Q_\infty(t) = x$. Thus Q_∞ is surjective.

Knowing that $Q_\infty \circ k_\infty$ is surjective, we now check that γ is constant on each fiber of $Q_\infty \circ k_\infty$. Suppose $a, b \in [0, 1]$ such that $Q_\infty \circ k_\infty(a) = Q_\infty \circ k_\infty(b)$. Then $Q_n \circ k_n(a) = Q_n \circ k_n(b)$ for all $n \in \mathbb{N}$. Applying H_n gives

$$\gamma_n(a) = H_n \circ Q_n \circ k_n(a) = H_n \circ Q_n \circ k_n(b) = \gamma_n(b)$$

for all $n \in \mathbb{N}$. Since $\{\gamma_n\} \to \gamma$ uniformly, we have $\{\gamma_n(a)\}_{n\in\mathbb{N}} \to \gamma(a)$ and $\{\gamma_n(b)\}_{n\in\mathbb{N}} \to \gamma(b)$, but since these sequences in X are equal, it follows that $\gamma(a) = \gamma(b)$. This completes the check and so we conclude that the desired map H_∞ exists.

With the existence of H_∞ confirmed, we fix $m \in \mathbb{N}$ and check that the equality $H_m = H_\infty \circ i_m$ holds. Since $\{H_n \circ R_n \circ Q_\infty \circ k_\infty\} = \{\gamma_n\} \to \gamma = H_\infty \circ Q_\infty \circ k_\infty$ uniformly and $Q_\infty \circ k_\infty$ is surjective, we have that $\{H_n \circ R_n\} \to H_\infty$ uniformly. Recalling that m is fixed, we have $\{H_n \circ R_n \circ i_m\}_{n>m} \to H_\infty \circ i_m$ uniformly. However, $\{H_n \circ R_n \circ i_m\}_{n>m} = \{H_n \circ i_{m,n}\}_{n>m} = \{H_m\}_{n>m}$ is the constant sequence at H_m. Thus $H_m = H_\infty \circ i_m$ for all $m \in \mathbb{N}$.

We will now consider the endpoints of the paths to complete our proof. For $t \in \{0, 1\}$, set $x_{n,t} = Q_n \circ k_n(t)$. Since $R_{n+1,n}(x_{n+1,t}) = x_{n,t}$, we have $Q_\infty \circ k_\infty(t) = (x_{n,t})_{n\in\mathbb{N}}$. However, recall that our inductive construction ensures that $i_{n,m}(x_{n,t}) = x_{m,t}$ whenever $m > n$. Thus in the limit, we also have $i_m(x_{m,t}) = (x_{n,t})_{n\in\mathbb{N}}$. We conclude that $Q_\infty \circ k_\infty(t) = i_n \circ Q_n \circ k_n(t)$ for all $n \in \mathbb{N}$ and $t \in \{0, 1\}$. Therefore, if we set $g_{n,1} = Q_\infty \circ k_\infty$ and $g_{n,2} = i_n \circ Q_n \circ k_n$, the concatenation $g_{n,1}\bar{g}_{n,2}$ is a well-defined loop in D_∞. Moreover, $H_\infty \circ g_{n,1} = H_\infty \circ Q_\infty \circ k_\infty = \gamma$ and $H_\infty \circ g_{n,2} = H_\infty \circ i_n \circ Q_n \circ k_n = H_n \circ Q_n \circ k_n = \gamma_n$. Thus $H_\infty \circ (g_{n,1}\bar{g}_{n,2}) = \gamma\bar{\gamma}_n$, proving $\gamma \simeq_{\mathbb{R}} \gamma_n$. □

References

[1] V. N. Berestovskii and C. P. Plaut, "Covering ℝ-trees, ℝ-free groups, and dendrites", *Adv. Math.* **224**:5 (2010), 1765–1783. MR

[2] M. Bestvina, "ℝ-trees in topology, geometry, and group theory", pp. 55–91 in *Handbook of geometric topology*, North-Holland, Amsterdam, 2002. MR

[3] M. Bestvina and M. Feighn, "Outer limits", eprint, 1994, https://www.math.utah.edu/~bestvina/eprints/bestvina.feighn..outer_limits.pdf.

[4] M. Bestvina and M. Feighn, "Stable actions of groups on real trees", *Invent. Math.* **121**:2 (1995), 287–321. MR

[5] J. Brazas and A. Mitra, "On maps with continuous path lifting", *Fund. Math.* **261**:3 (2023), 201–234. MR

[6] J. W. Cannon and G. R. Conner, "On the fundamental groups of one-dimensional spaces", *Topology Appl.* **153**:14 (2006), 2648–2672. MR

[7] C. E. Capel, "Inverse limit spaces", *Duke Math. J.* **21** (1954), 233–245. MR

[8] M. L. Curtis and M. K. Fort, Jr., "The fundamental group of one-dimensional spaces", *Proc. Amer. Math. Soc.* **10** (1959), 140–148. MR

[9] J. Dydak, "Coverings and fundamental groups: a new approach", preprint, 2011. arXiv 1108.3253v1

[10] S. Eilenberg, "Sur les transformations continues d'espaces métriques compacts", *Fund. Math.* **22** (1934), 292–296.

[11] R. Engelking, *General topology*, 2nd ed., Sigma Series in Pure Mathematics **6**, Heldermann, Berlin, 1989. MR

[12] D. Gaboriau, G. Levitt, and F. Paulin, "Pseudogroups of isometries of ℝ and Rips' theorem on free actions on ℝ-trees", *Israel J. Math.* **87**:1-3 (1994), 403–428. MR

[13] D. Groves, "Limit groups for relatively hyperbolic groups, I: The basic tools", *Algebr. Geom. Topol.* **9**:3 (2009), 1423–1466. MR

[14] K. A. Hardie, K. H. Kamps, and R. W. Kieboom, "A homotopy 2-groupoid of a Hausdorff space", *Appl. Categ. Structures* **8**:1-2 (2000), 209–234. MR

[15] S.-T. Hu, *Homotopy theory*, Pure and Applied Mathematics **8**, Academic Press, New York, 1959. MR

[16] J. C. Mayer and L. G. Oversteegen, "A topological characterization of ℝ-trees", *Trans. Amer. Math. Soc.* **320**:1 (1990), 395–415. MR

[17] S. B. Nadler, Jr., *Continuum theory*: *an introduction*, Pure and Applied Mathematics **158**, Dekker, New York, 1992. MR

[18] F. Paulin, "Outer automorphisms of hyperbolic groups and small actions on ℝ-trees", pp. 331–343 in *Arboreal group theory* (Berkeley, 1988), Math. Sci. Res. Inst. Publ. **19**, Springer, 1991. MR

[19] Z. Sela, "Diophantine geometry over groups, I: Makanin–Razborov diagrams", *Publ. Math. Inst. Hautes Études Sci.* **93** (2001), 31–105. MR

[20] G. T. Whyburn, “Non-alternating transformations”, *Amer. J. Math.* **56**:1-4 (1934), 294–302. MR

Received March 4, 2025. Revised April 7, 2026.

JEREMY BRAZAS
DEPARTMENT OF MATHEMATICS
WEST CHESTER UNIVERSITY
WEST CHESTER, PA
UNITED STATES
jbrazas@wcupa.edu

GREGORY R. CONNER
DEPARTMENT OF MATHEMATICS
BRIGHAM YOUNG UNIVERSITY
PROVO, UT
UNITED STATES
conner@mathematics.byu.edu

PAUL FABEL
DEPARTMENT OF MATHEMATICS AND STATISTICS
MISSISSIPPI STATE UNIVERSITY
MISSISSIPPI STATE, MS
UNITED STATES
fabel@math.msstate.edu

CURTIS KENT
DEPARTMENT OF MATHEMATICS
BRIGHAM YOUNG UNIVERSITY
PROVO, UT
UNITED STATES
curtkent@mathematics.byu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.343

SEVERI VARIETIES ON RULED SURFACES OVER ELLIPTIC CURVES

XIAOTIAN CHANG, XI CHEN AND ADRIAN ZAHARIUC

We prove that the general members of Severi varieties on an Atiyah ruled surface over a general elliptic curve have nodes and ordinary triple points as singularities.

1. Introduction

Severi varieties of projective surfaces are roughly parameter (moduli) spaces of curves of fixed geometric genus, in a given linear system or homology class on the projective surface. The natural expectation is that a *general* curve in any irreducible component of a Severi variety will have only nodes as singularities. This expectation often turns out to be true under the natural assumption that the surface has general moduli, although there are notable known exceptions, for instance, in the case of abelian surfaces [DS17, Example (4.17)]. This problem has been investigated for many classes of surfaces beyond the classical case of $\mathbb{P}^2$, including K3 surfaces [Che02; Che19], abelian surfaces [KLM19; KL22], Enriques surfaces [CDGK23b], surfaces in $\mathbb{P}^3$ [CC99], and some ruled surfaces [CDGK23a].

In this paper, we consider Severi varieties on a ruled surface over a smooth elliptic curve. We will see that the unexpected phenomenon of general curves with worse than nodal singularities occurs again on this ruled surface.

Let E be a smooth elliptic curve, let $\mathscr{E}$ be a rank 2 vector bundle on E given by a nonzero vector in $\mathrm{Ext}(\mathcal{O}_E, \mathcal{O}_E)$ and let $R = \mathbb{P}\mathscr{E}$. Such surfaces arise naturally when we study the degeneration of abelian and K3 surfaces [Zah22]. We call such a ruled surface the *Atiyah ruled surface* over E.

The main purpose of this note is to prove the following:

Theorem 1.1. *Let E be a smooth elliptic curve, let $\mathscr{E}$ be a rank 2 vector bundle on E given by a nonzero vector in* $\mathrm{Ext}(\mathcal{O}_E, \mathcal{O}_E)$ *and let $R = \mathbb{P}\mathscr{E}$. For a line bundle L on R, let $V_{R,L,g} \subset |L|$ be the locus of integral curves $C \in |L|$ of geometric genus g. Then when E is general, L is ample and $g \geq 1$, for a general member $[C] \in V_{R,L,g}$,*

All authors were partially supported by grants from the Natural Sciences and Engineering Research Council of Canada.
MSC2020: 14C25, 14C30, 14H10, 14H52.
Keywords: Severi variety, deformation theory.

- *if* $L.D \geq 2$, *C is nodal*, *and*
- *if* $L.D = 1$, *C has only nodes and/or ordinary triple points as singularities*,

where D is the unique section of R over E with self-intersection $D^2 = 0$.

Theorem 1.2 below shows that triple points do occur in some cases.

The Picard group $\operatorname{Pic}(R)$ of R is generated by D and $\pi^* \operatorname{Pic}(E)$, where $\pi : R \to E$ is the projection.

For an integral curve $C \subset R$ of geometric genus g with normalization $f : \widehat{C} \to R$, we have

$$\deg(\omega_{\widehat{C}} \otimes f^*\omega_R^\vee) = 2g - 2 - K_R.C = 2(g - 1 + C.D),$$

since $-K_R = 2D$. Based on [DS17, Corollary 2.11] (see also [HM98, Section B, pp. 108–111] and [AC81, Lemma 1.4, p. 345]), we know the following.

- If the degree of $\omega_{\widehat{C}} \otimes f^*\omega_R^\vee$ is at least $2g$, or equivalently,

$$C.D \geq 1, \tag{1-1}$$

  then a general deformation of f is immersive.

- If the degree of $\omega_{\widehat{C}} \otimes f^*\omega_R^\vee$ is at least $2g + 2$, or equivalently,

$$C.D \geq 2, \tag{1-2}$$

  then a general deformation of f has nodal image.

Consequently, our main theorem holds for every $L = mD + \pi^*M$ if $m > 0$ and $\deg M \geq 2$. Therefore, the only remaining case for Theorem 1.1 is $m > 0$ and $\deg M = 1$. Furthermore, we will show that the case $g \geq 2$ can be reduced to $g = 1$ by a degeneration argument. That is, it suffices to prove the theorem for $L = mD + R_p$ and $g = 1$, where $R_p = \pi^* p$ is the fiber of R over a point $p \in E$. Indeed, we have a more precise statement for this case:

Theorem 1.2. *Let E be a smooth elliptic curve*, *let $\mathscr{E}$ be a rank 2 vector bundle on E given by a nonzero vector in* $\operatorname{Ext}(\mathcal{O}_E, \mathcal{O}_E)$ *and let* $R = \mathbb{P}\mathscr{E}$. *When E is general*, *for $L = mD + R_p$ and every* $[C] \in V_{R,L,1}$,

- *if* $4 \nmid m$, *C is nodal*, *and*
- *if* $4 \mid m$, *C has only nodes and/or ordinary triple points as singularities*,

where D is the unique section of R over E with self-intersection $D^2 = 0$ and R_p is the fiber of R over $p \in E$.

In addition, *if* $4 \mid m$, *then there exists at least one irreducible component V of $V_{R,L,1}$ such that the general curve $[C] \in V$ has at least one triple point.*

Such elliptic curves were also studied by E. Sernesi in [Ser23].

Conventions. We work exclusively over $\mathbb{C}$.

2. The Atiyah ruled surface $\mathbb{P}\mathscr{E}$

We start with some basic facts about the Atiyah ruled surface $\mathbb{P}\mathscr{E}$.

Proposition 2.1. *Let E be a smooth elliptic curve, let $\mathscr{E}$ be a rank 2 vector bundle on E given by a nonzero vector in $\mathrm{Ext}(\mathcal{O}_E, \mathcal{O}_E)$, let $R = \mathbb{P}\mathscr{E}$ and let $D \subset R$ be the section of R/E with $D^2 = 0$.*

(1) *For every point $p \in E$, $|D + R_p|$ is a pencil such that every curve $C \neq D \cup R_p \in |D + R_p|$ is a smooth elliptic curve and any pair $C_1 \neq C_2 \in |D + R_p|$ of curves meet only at p (viewed as a point of D) with multiplicity 2, where R_p is the fiber of R over $p \in E$.*

(2) *For every point $p \in E$, $R \setminus (D \cup R_p) \cong (E \setminus \{p\}) \times \mathbb{A}^1$.*

(3) *For every pair of points $p \neq q \in E$, $R \setminus D$ is isomorphic to the gluing of $(E \setminus \{p\}) \times \mathbb{A}^1$ and $(E \setminus \{q\}) \times \mathbb{A}^1$ via an automorphism*

$$(E \setminus \{p, q\}) \times \mathbb{A}^1 \xrightarrow{\ \eta\ } (E \setminus \{p, q\}) \times \mathbb{A}^1$$

given by

$$\eta(z, s) = (z, s + h(z)),$$

where $h(z)$ is a meromorphic function on E with simple poles at p and q.

(4) *There is an exact sequence of group schemes*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathbf{G}_a & \longrightarrow & \mathrm{Aut}(R)_0 & \longrightarrow & \mathrm{Aut}(D)_0 & \longrightarrow & 0 \\ &&&&&& \Vert &&\\ &&&&&& \mathrm{Aut}(E)_0, && \end{array} \tag{2-1}$$

where $\mathbf{G}_a$ is the additive group of $\mathbb{C}$ and $\mathrm{Aut}(R)_0$ and $\mathrm{Aut}(E)_0$ are the connected components of $\mathrm{Aut}(R)$ and $\mathrm{Aut}(E)$, respectively, containing the identity. Every $\phi \in \mathrm{Aut}(R)_0$ is given by

$$\begin{aligned} \phi(z, s) &= (z + \tau, s + b_1(z)) \quad \text{on } (E \setminus \{p, p - \tau\}) \times \mathbb{A}^1, \\ \phi(z, s) &= (z + \tau, s + b_2(z)) \quad \text{on } (E \setminus \{q, q - \tau\}) \times \mathbb{A}^1, \end{aligned} \tag{2-2}$$

where $\tau \in \mathrm{Pic}^0(E) = J(E)$, p and q are two distinct points on E satisfying $p - q \neq \pm\tau$, $b_1(z)$ is a meromorphic function on E with simple poles at p and $p - \tau$, $b_2(z)$ is a meromorphic function on E with simple poles at q and $q - \tau$, and $b_1(z)$ and $b_2(z)$ satisfy

$$b_1(z) + h(z) = b_2(z) + h(z + \tau) \tag{2-3}$$

on $E \setminus \{p, p - \tau, q, q - \tau\}$ with $h(z)$ given in (3).

Proof. By the exact sequence

$$0 \longrightarrow \mathcal{O}_E \longrightarrow \mathscr{E} \longrightarrow \mathcal{O}_E \longrightarrow 0$$

we obtain

$$h^0(\mathscr{E}^\vee \otimes \mathcal{O}_E(p)) = h^0(\mathcal{O}_E(p)) + h^0(\mathcal{O}_E(p)) = 2$$

and hence $|D + R_p|$ is a pencil. Since

$$\mathcal{O}_R(D + R_p)\big|_D = \mathcal{O}_E(p)$$

every $C \in |D + R_p|$ passes through p. If C is reducible, C must contain a section of R/E and hence it must contain D. Consequently, the only reducible member of $|D + R_p|$ is $D \cup R_p$. Every other member of $|D + R_p|$ is a section of R/E. For $C_1 \neq C_2 \in |D + R_p|$, one of C_1 and C_2 must be integral. Let us assume that C_1 is a section of R/E. Then

$$\mathcal{O}_{C_1}(C_2) = \mathcal{O}_{C_1}(D + R_p) = \mathcal{O}_{C_1}(2p).$$

We know that both C_1 and C_2 pass through p and they have intersection number 2. So $C_1.C_2 = p + p'$. Then $p + p' \sim_{\text{rat}} 2p$ on C_1 and hence $p' = p$. That is, C_1 and C_2 meet at p with multiplicity 2 and they do not have any other intersections. This proves (1).

Let $\alpha_p : R \dashrightarrow \mathbb{P}^1$ be the rational map given by the pencil $|D + R_p|$. By (1), the map

$$R \setminus (D \cup R_p) \xrightarrow[\cong]{\pi \times \alpha_p} (E \setminus \{p\}) \times \mathbb{A}^1$$

is an isomorphism, where $\pi : R \to E$ is the projection. This proves (2).

We have

$$R \setminus D = (R \setminus (D \cup R_p)) \cup (R \setminus (D \cup R_q))$$

with $(R \setminus (D \cup R_p))$ and $(R \setminus (D \cup R_q))$ isomorphic to $(E \setminus \{p\}) \times \mathbb{A}^1$ and $(E \setminus \{q\}) \times \mathbb{A}^1$ via $\pi \times \alpha_p$ and $\pi \times \alpha_q$, respectively. So $R \setminus D$ is the gluing of $(E \setminus \{p\}) \times \mathbb{A}^1$ and $(E \setminus \{q\}) \times \mathbb{A}^1$ via an automorphism $\eta \in \mathrm{Aut}(U \times \mathbb{A}^1/U)$

$$U \times \mathbb{A}^1 \xrightarrow{\ \eta\ } U \times \mathbb{A}^1$$

for $U = E \setminus \{p, q\}$. Such an automorphism is given by

$$\eta(z, s) = (z, h(z)s + f(z)),$$

where $h(z)$ and $f(z)$ are meromorphic functions on E such that they are holomorphic on U and $h(z) \neq 0$ on U. So $h(z)$ has zeros and poles only at p and q and $f(z)$ has poles only at p and q.

A member of the pencil $|D + R_p|$ other than $D \cup R_p$ is given by

$$(\pi \times \alpha_p)^{-1}\{s = a\}$$

for $a \in \mathbb{C}$. Similarly, a member of the pencil $|D+R_q|$ other than $D \cup R_q$ is given by

$$(\pi \times \alpha_q)^{-1}\{s=b\}$$

for $b \in \mathbb{C}$. These two curves meet at two points lying in $R \setminus (D \cup R_p \cup R_q)$. Therefore,

$$\{s=a\} \cap \eta^{-1}\{s=b\}$$

has two intersections (counted with multiplicity) in $U \times \mathbb{A}^1$ for all $a, b \in \mathbb{C}$. That is, the function

$$ah(z)+f(z)-b$$

has exactly two zeros over U for all a, b. It follows that $h(z)$ is a nonzero constant and $f(z)$ has simple poles at p and q. We may choose $h(z) \equiv 1$. This proves (3).

Clearly, every automorphism of R preserves the section D. Let $\phi : R \to R$ be an automorphism of R in the kernel of $\mathrm{Aut}(R) \to \mathrm{Aut}(D)$ and let ϕ_1 and ϕ_2 be the restriction of ϕ to $(E \setminus \{p\}) \times \mathbb{A}^1$ and $(E \setminus \{q\}) \times \mathbb{A}^1$, respectively. Suppose that ϕ_1 and ϕ_2 are given by

$$\phi_1(z, s) = (z, a_1(z)s + b_1(z)),$$
$$\phi_2(z, s) = (z, a_2(z)s + b_2(z)),$$

where $a_1(z)$ and $b_1(z)$ are meromorphic functions on E with poles at p, $a_2(z)$ and $b_2(z)$ are meromorphic functions on E with poles at q, $a_1(z) \neq 0$ on $E \setminus \{p\}$ and $a_2(z) \neq 0$ on $E \setminus \{q\}$. Clearly, $a_1(z) \equiv a_1$ and $a_2(z) \equiv a_2$ must be constants. In addition, since $\phi_1 \circ \eta = \eta \circ \phi_2$, we have

$$a_1(s+h(z)) + b_1(z) = a_2 s + b_2(z) + h(z)$$

on $(E \setminus \{p, q\}) \times \mathbb{A}^1$. Obviously, $a_1 = a_2 = a$ and hence

$$b_1(z) - b_2(z) = (1-a)h(z).$$

Since $h(z)$ has simple poles at p and q, $b_1(z)$ has a single pole at p and $b_2(z)$ has a single pole at q, $b_1(z)$ and $b_2(z)$ must have simple poles at p and q, respectively, and hence they must be constant. It follows that $a = 1$ and $b_1(z) \equiv b_2(z) \equiv b$. This proves that

$$\mathbf{G}_a = \ker(\mathrm{Aut}(R) \to \mathrm{Aut}(D)).$$

To complete the proof of (2-1), it remains to prove that the map

$$\mathrm{Aut}(R)_0 \twoheadrightarrow \mathrm{Aut}(D)_0$$

is surjective.

Every automorphism $\lambda \in \mathrm{Aut}(E)_0$ is given by a translation $\lambda(p) = p + \tau$ for some $\tau \in \mathrm{Pic}^0(E) = J(E)$.

For a given $\tau \in J(E)$, if there exists a pair of meromorphic functions $b_1(z)$ and $b_2(z)$ satisfying (2-3), then $\phi \in \mathrm{Aut}(R)_0$ given by (2-2) maps to $\lambda \in \mathrm{Aut}(E)_0$ with $\lambda(p) = p + \tau$. So it suffices to prove the existence of $b_1(z)$ and $b_2(z)$ satisfying (2-3).

If $\tau = 0$, we can simply take $b_1(z) \equiv b_2(z) \equiv b$ to be a constant.

Suppose that $\tau \neq 0$. We lift (2-3) from $E \cong \mathbb{C}/\Lambda$ to $\mathbb{C}$. Then $b_1(z)$, $b_2(z)$ and $h(z)$ are doubly periodic meromorphic functions on $\mathbb{C}$. We choose $b_1(z)$ such that

$$\mathrm{Res}_p\, b_1(z) = -\mathrm{Res}_p\, h(z).$$

Since

$$\mathrm{Res}_p\, b_1(z) + \mathrm{Res}_{p-\tau}\, b_1(z) = 0$$

we have

$$\mathrm{Res}_{p-\tau}\, b_1(z) = \mathrm{Res}_p\, h(z) = \mathrm{Res}_{p-\tau}\, h(z+\tau).$$

So $b_2(z) = b_1(z) + h(z) - h(z+\tau)$ is analytic at p and $p - \tau$. This proves the existence of $b_1(z)$ and $b_2(z)$ satisfying (2-3) and hence (4). □

Let $C \in |mD + R_p|$ be a (possibly singular) elliptic curve on R and let $\nu : \mathscr{C} \to R$ be the normalization of C. We let

$$S = \mathscr{C} \times_E R = \mathbb{P}(\pi \circ \nu)^* \mathscr{E}$$

via the maps $\pi \circ \nu : \mathscr{C} \to E$ and $\pi : R \to E$. Clearly, $(\pi \circ \nu)^* \mathscr{E}$ is a rank 2 vector bundle on $\mathscr{C}$ given by a nonzero vector in $\mathrm{Ext}(\mathcal{O}_{\mathscr{C}}, \mathcal{O}_{\mathscr{C}})$.

The map $g : S \to R$ is induced by $\pi \circ \nu : \mathscr{C} \to E$ and is hence étale. Let us consider the preimage

$$g^{-1}(C) = \mathscr{C} \times_E C$$

of C. It contains the curve $G = \{(s, \nu(s)) : s \in \mathscr{C}\} \cong \mathscr{C}$. It is not hard to see that $G \in |\mathcal{O}_S(\mathscr{D} + S_q)|$, where $\mathscr{D} = g^* D$ is the unique section of $S/\mathscr{C}$ with self-intersection 0, $q \in (\pi \circ \nu)^{-1}(p)$ and S_q is the fiber of $S/\mathscr{C}$ over q.

Since $g : S \to R$ is Galois,

$$g^* C = \sum_{\sigma \in \mathrm{Aut}(S/R)} \sigma(G).$$

The map $g : g^* C \to C$ is étale. So C is nodal if and only if $g^* C$ is, i.e., it has normal crossings.

Since $h = \pi \circ \nu : \mathscr{C} \to E$ is an isogeny, the dual isogeny $h^\vee : E \to \mathscr{C}$ has the property that $h^\vee \circ h : \mathscr{C} \to \mathscr{C}$ is a multiplication map given by $x \to p + n(x - p)$ for some integer n. So the Galois group $\mathrm{Aut}(\mathscr{C}/E)$ is a subgroup of $\mathrm{Aut}(h^\vee \circ h)$. Hence $\mathrm{Aut}(\mathscr{C}/E)$ is given by a finite subgroup of $J(\mathscr{C}) = \mathrm{Pic}^0(\mathscr{C})$. That is, every $\sigma \in \mathrm{Aut}(\mathscr{C}/E)$ is given by a translation $\sigma(x) = x + \tau$ for some torsion element $\tau \in J(\mathscr{C})$.

To prove Theorem 1.1, it suffices to prove the following:

Proposition 2.2. *Let E be a smooth elliptic curve, let $\mathscr{E}$ be a rank 2 vector bundle on E given by a nonzero vector in $\mathrm{Ext}(\mathcal{O}_E, \mathcal{O}_E)$, let $R = \mathbb{P}\mathscr{E}$, let $D \subset R$ be the section of R/E with $D^2 = 0$ and let $A \subset \mathrm{Aut}(R)_0$ be a finite subgroup of $\mathrm{Aut}(R)_0$ acting freely on R. Then when E is general, for every point $p \in E$ and every smooth curve $G \in |D + R_p|$,*

$$\sum_{\sigma \in A} \sigma(G)$$

has normal crossings if A does not contain the subgroup

$$J(E)_2 = \{\tau \in J(E) : 2\tau = 0\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$$

and has only nodes and ordinary triple points as singularities otherwise.

When $C \in |mD + E|$, the Galois group $\mathrm{Aut}(\mathscr{C}/E)$ has order m. If $4 \nmid m$, $\mathrm{Aut}(\mathscr{C}/E)$ does not contain a subgroup of order 4 and hence C is nodal by the above proposition.

Here we let

$$J(E)_n = \{\tau \in J(E) : n\tau = 0\} \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \quad \text{and} \quad J(E)_{\mathrm{tors}} = \bigcup_{n=1}^{\infty} J(E)_n$$

be the torsion subgroups of $J(E)$. For every $\tau \in J(E)_{\mathrm{tors}}$, we define the order $\mathrm{ord}(\tau)$ of τ to be the smallest positive integer n such that $n\tau = 0$ and let $\mathrm{ord}(\tau) = \infty$ if $\tau \notin J(E)_{\mathrm{tors}}$.

Let $\phi \in \mathrm{Aut}(R)_0$ be an automorphism of order n. By (2-2), ϕ is given by a meromorphic function $b_1(z)$ on E with simple poles at p and $p - \tau$ satisfying

$$b_1(z) + b_1(z + \tau) + \cdots + b_1(z + (n-1)\tau) = 0, \tag{2-4}$$

where $\tau \in J(E)_{\mathrm{tors}}$ has order $\mathrm{ord}(\tau) = n$.

To prove that G and $\phi(G)$ intersect transversely, it suffices to prove that $b_1(z)$ does not have a zero of multiplicity 2, i.e.,

$$b_1(p - \eta) \neq 0 \quad \text{for } \tau = 2\eta \tag{2-5}$$

when E is a general elliptic curve.

Let $\phi_1 \neq \phi_2 \in \mathrm{Aut}(R)_0$ be two automorphisms of finite order. Similarly, ϕ_1 and ϕ_2 are given by two meromorphic functions $b_1(z)$ and $b_2(z)$ on E with simple poles at $\{p, p - \tau_1\}$ and $\{p, p - \tau_2\}$, respectively, satisfying

$$b_i(z) + b_i(z + \tau_i) + \cdots + b_i(z + (n_i - 1)\tau_i) = 0 \tag{2-6}$$

for $i = 1, 2$, where $\tau_i \in J(E)_{\mathrm{tors}}$ has order n_i and $\tau_1 \neq \tau_2$. To show that G, $\phi_1(G)$ and $\phi_2(G)$ do not meet at one point, it suffices to show that

$$\{b_1(z) = 0\} \cap \{b_2(z) = 0\} = \varnothing, \tag{2-7}$$

where E is a general elliptic curve. So it remains to prove (2-5) and (2-7).

Let us start with the observation that the meromorphic functions $b_i(z)$ satisfying (2-6) are unique up to a scalar, depending only on p and τ_i.

Proposition 2.3. *Let E be an elliptic curve and let p be a point of E. For every $\tau \in J(E)_{\mathrm{tors}}$ of order n and every meromorphic function $b(z)$ on E with simple poles at p and $p-\tau$ and no other poles,*

$$\sum_{k=0}^{n-1} b(z+k\tau)$$

is constant.

In addition, there is a unique meromorphic function $b(z) = b_{\tau,p}(z)$ on E, up to a scalar, with simple poles at p and $p-\tau$ and no other poles such that

$$\sum_{k=0}^{n-1} b(z+k\tau) = 0. \tag{2-8}$$

Furthermore, for all positive integers m with $n \mid m$ and every meromorphic function $b(z)$ on E with simple poles at p and $p-\tau$ and no other poles,

$$\sum_{\lambda\in J(E)_m} b(z+\lambda) = \frac{m^2}{n}\sum_{k=0}^{n-1} b(z+k\tau). \tag{2-9}$$

Consequently, (2-8) *holds if and only if*

$$\sum_{\lambda\in J(E)_m} b(z+\lambda) = 0 \tag{2-10}$$

for some positive integer m with $n \mid m$.

Proof. Let $\omega \in H^0(\Omega_E)$ be a nonzero holomorphic 1-form on E. Then $b(z)\omega$ is a meromorphic 1-form on E with simple poles at p and $p-\tau$. So

$$\mathrm{Res}_p\, b(z)\omega + \mathrm{Res}_{p-\tau}\, b(z)\omega = 0.$$

It follows that

$$\sum_{k=0}^{n-1} b(z+k\tau)\omega$$

is a holomorphic 1-form on E and hence

$$\sum_{k=0}^{n-1} b(z+k\tau)$$

is constant on E.

Let $V = H^0(\mathcal{O}_E(p_1 + p_2)) \cong \mathbb{C}^2$ be the vector space of meromorphic functions on E with at worst simple poles at $p_1 = p$ and $p_2 = p - \tau$ and let $L : V \to \mathbb{C}$ be the map given by

$$L(b(z)) = \sum_{k=0}^{n-1} b(z + k\tau).$$

Clearly, L is linear. When $b(z) \equiv c$ is constant, $L(b(z)) = nc$ and hence L is surjective. Thus, $\ker(L)$ is a one-dimensional subspace of V. So there exists a unique $b(z) \in V$, up to a scalar, such that

$$\sum_{k=0}^{n-1} b(z + k\tau) = 0.$$

Obviously, $G = \{k\tau : k \in \mathbb{Z}\}$ is a subgroup of $J(E)_m$ for $n \mid m$. So

$$J(E)_m = \bigsqcup_{i=1}^{d} (\lambda_i + G)$$

for some $\lambda_1, \lambda_2, \dots, \lambda_d \in J(E)_m$ and $d = m^2/n$. Then

$$\sum_{\lambda \in J(E)_m} b(z + \lambda) = \sum_{i=1}^{d} \sum_{\lambda \in G} b(z + \lambda_i + \lambda)$$

We have proved that $\sum_{\lambda \in G} b(z + \lambda)$ is constant. Therefore,

$$\sum_{\lambda \in G} b(z + \lambda) \equiv \sum_{\lambda \in G} b(z + \lambda_i + \lambda)$$

for all i and hence

$$\sum_{\lambda \in J(E)_m} b(z + \lambda) = \sum_{i=1}^{d} \sum_{\lambda \in G} b(z + \lambda_i + \lambda) = d \sum_{\lambda \in G} b(z + \lambda).$$

This proves (2-9). □

We formally state the context around (2-5) and (2-7):

Proposition 2.4. *For a general elliptic curve E, every point $p \in E$, every $\tau \in J(E)_{\mathrm{tors}}$ of order $n \geq 2$ and every $\eta \in J(E)_{\mathrm{tors}}$ satisfying $2\eta = \tau$, we have*

$$b_{\tau,p}(p - \eta) \neq 0,$$

where $b_{\tau,p}(z)$ is the meromorphic function on E given in Proposition 2.3.

Proposition 2.5. *Let E be an elliptic curve, let $p \in E$ be a point on E and let $b_{\tau,p}$ be the meromorphic function on E given in Proposition 2.3 for a nonzero torsion point $\tau \in J(E)_{\mathrm{tors}}$.*

For E general and any two torsion points $\tau_1 \neq \tau_2 \in J(E)_{\text{tors}}$ of orders $n_1 \geq 2$ and $n_2 \geq 2$, respectively, one of the following holds:

$$\{b_{\tau_1,p}(z)=0\} \cap \{b_{\tau_2,p}(z)=0\} = \varnothing \tag{2-11}$$

or

$$(n_1, n_2) = (2, 2) \tag{2-12}$$

or

$$(n_1, n_2) = (6, 6),\ \langle \tau_1, \tau_2 \rangle = 3 \text{ in } J(E)_6 \quad \text{and} \quad \operatorname{ord}(\tau_1 - \tau_2) = 6. \tag{2-13}$$

In addition, when $(n_1, n_2) = (2, 2)$,

$$\{b_{\tau_1,p}(z)=0\} \cap \{b_{\tau_2,p}(z)=0\} = \{p - \tau_3\}, \tag{2-14}$$

where $\tau_3 \in J(E)_{\text{tors}}$ is a torsion point of order 2 different from τ_1 and τ_2.

For E general and any three distinct nonzero torsion points $\tau_1, \tau_2, \tau_3 \in J(E)_{\text{tors}}$,

$$\{b_{\tau_1,p}(z)=0\} \cap \{b_{\tau_2,p}(z)=0\} \cap \{b_{\tau_3,p}(z)=0\} = \varnothing. \tag{2-15}$$

The intersection pairing $\langle \cdot, \cdot \rangle$ on $J(E)_n$ will be defined in the next section.

Let us explain how Propositions 2.4 and 2.5 imply Proposition 2.2. Proposition 2.4 implies that any pair of curves among $\{\sigma(G) : \sigma \in A\}$ meet transversely and thus $\sum \sigma(G)$ has only ordinary singularities, i.e., singularities whose local branches are smooth and meet transversely pairwise. Then Proposition 2.4 says that no three curves among $\{\sigma(G) : \sigma \in A\}$ meet at one point with the exceptions (2-12) and (2-13), in which cases no more than three curves among $\{\sigma(G) : \sigma \in A\}$ meet at one point by (2-15). In case (2-12), τ_1 and τ_2 generate $J(E)_2 \subset A$. In case (2-13), τ_1 and τ_2 generate a subgroup of $J(E)_6$ of order 12 contained in A; such a subgroup clearly contains $J(E)_2$.

Finally, let us explicitly illustrate how the considerations above lead to curves with triple points in the case $4 \mid m$. Let us first consider the case $m = 4$. Let $[2] : E \to E$ be the multiplication by 2 map on E relative to the choice of a point on E. It is clear that $[2]^* \mathscr{E} \cong \mathscr{E}$ hence $R \cong E \times_{[2],E,\pi} R$. The group of deck transformations of the projection to the second factor $g : R \to R$ is $\{\mathrm{id}_R, \phi_{\tau_1}, \phi_{\tau_2}, \phi_{\tau_3}\}$, with $\phi_{\tau_i} : R \to R$ lying above $z \mapsto z + \tau_i$, where $J(E)_2 = \{0, \tau_1, \tau_2, \tau_3\}$. If $G \in |D + R_p|$, then (2-14) implies that

$$G \cap \phi_{\tau_1}(G) \cap \phi_{\tau_2}(G) \neq \varnothing$$

by the same reasoning as above. Therefore, the curve $g(G) \in |4D + R_p|$ has a triple point. In general, if $4 \mid m$, consider an isogeny $\mathscr{C} \to E$ of degree $m/4$. As above, let $S = \mathscr{C} \times_E R$, which is the Atiyah surface associated to $\mathscr{C}$, $\mathscr{D} \subset S$ the section of $S/\mathscr{C}$ of self-intersection 0, and $q \in \mathscr{C}$. By the case $m = 4$ discussed above, $|4\mathscr{D} + S_q|$ contains genus 1 curves with triple points. Then, their images in R by $S \to R$

are curves $C \in |mD + R_p|$ with triple points. Furthermore, if $f : \widehat{C} \to R$ is the normalization, then f is immersive and $\deg N_f = 2$, so

$$h^0(\widehat{C}, N_f) = 2 = \dim \operatorname{Aut}(R)_0,$$

which implies that all equigeneric deformations of $C \subset R$ come from automorphisms of R, and thus have triple points as well. (For ease of language, we have included the deformations which change the linear system.) Hence, if $4 \mid m$, there exists at least one irreducible component of the Severi variety $V_{R,mD+R_p,1}$ in which the general curve has triple points, as claimed in Theorem 1.2.

3. Torsion points on generic elliptic curves

We will prove Proposition 2.4 and 2.5 by letting E vary in a complete family of elliptic curves X/B with a unique section P. There are many choices of such X. Let us choose X to be a K3 surface with Picard lattice $\left[\begin{smallmatrix} -2 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. We call such X a *Bryan–Leung K*3 [BL00]. Such X admits an elliptic fibration $\pi : X \to B = \mathbb{P}^1$. For X general, it has 24 nodal fibers over $S \subset B$. The (-2)-curve $P \subset X$ is the only section of π. For each positive integer n, let us consider

$$\Sigma_n = \overline{\{q \in X_b : b \notin S,\ \operatorname{ord}(p-q) = n \text{ for } p = P_b = P \cap X_b\}} \tag{3-1}$$

Clearly, Σ_n is a multisection of X/B of degree

$$n^2 \prod_{\substack{p \text{ prime} \\ p \mid n}} \Big(1 - \frac{1}{p^2}\Big).$$

We claim that Σ_n is irreducible. This is proved by studying the monodromy action of $\pi_1(B \setminus S)$ on Σ_n. Actually, the monodromy action of $\pi_1(B \setminus S)$ on Σ_n is induced by its action on $H^1(X_b, \mathbb{Z})$.

Fix a smooth fiber $E = X_b$ of X over $b \in B^\circ = B \setminus S$ and let us consider the monodromy action of $\pi_1(B^\circ)$ on $J(E)_{\text{tors}}$ and $H^1(E, \mathbb{Z})$. From the exponential sequence, we have the diagram

$$\begin{array}{ccccccccc}
 & & & & & & J(E)_n & & \\
 & & & & & & \big\downarrow & & \\
0 & \longrightarrow & H^1(E,\mathbb{Z}) & \longrightarrow & H^1(\mathcal{O}_E) & \longrightarrow & J(E) & \longrightarrow & 0 \\
 & & \big\downarrow{\scriptstyle \times n} & & \big\|{\scriptstyle \times n} & & \big\downarrow{\scriptstyle \times n} & & \\
0 & \longrightarrow & H^1(E,\mathbb{Z}) & \longrightarrow & H^1(\mathcal{O}_E) & \longrightarrow & J(E) & \longrightarrow & 0 \\
 & & \big\downarrow & & & & & & \\
 & & H^1(E,\mathbb{Z})/nH^1(E,\mathbb{Z}) & & & & & &
\end{array}$$

Thus, we have

$$J(E)_n \cong H^1(E, \mathbb{Z})/nH^1(E, \mathbb{Z})$$

and the action of $\pi_1(B^\circ)$ on $J(E)_{\text{tors}}$ is induced by its action of $H^1(E, \mathbb{Z})$.

The action $\pi_1(B^\circ)$ on $H^1(E, \mathbb{Z})$ preserves the intersection product of $H^1(E, \mathbb{Z})$. Thus, it is given by a group homomorphism

$$\pi_1(B^\circ) \longrightarrow \text{Aut}(H^1(E, \mathbb{Z})) \cong \text{SL}_2(\mathbb{Z}),$$

where $\text{Aut}(H^1(E, \mathbb{Z}))$ is the automorphism group of $H^1(E, \mathbb{Z})$ as a lattice. Thus, the induced action of $\pi_1(B^\circ)$ on Σ_n is given by the group homomorphism

$$\begin{array}{ccc} \pi_1(B^\circ) & \longrightarrow & \text{SL}_2(\mathbb{Z}) \\ & \searrow & \downarrow \\ & & \text{SL}_2(\mathbb{Z}/n\mathbb{Z}) \end{array}$$

Proposition 3.1. *Let $\pi : X \to B = \mathbb{P}^1$ be a Bryan–Leung* K3 *surface with* 24 *nodal fibers. Then the monodromy action $\pi_1(B^\circ) \to \text{SL}_2(\mathbb{Z}/n\mathbb{Z})$ is surjective and Σ_n is irreducible for all $n \in \mathbb{Z}^+$ with $\Sigma_n \subset X$ defined by* (3-1).

The action of $\pi_1(B^\circ)$ on $H^1(E, \mathbb{Z})$ is well understood. At each $b_i \in \{b_1, b_2, \ldots, b_{24}\}$, the loop around b_i acts on $H^1(E, \mathbb{Z})$ by a Lefschetz–Picard transform [Lew99]:

$$T_{\delta_i}(\lambda) = \lambda + \langle \lambda, \delta_i \rangle \delta_i,$$

where $\delta_i \in H^1(E, \mathbb{Z})$ is called the *vanishing cycle* at the node of X_{b_i} for $i = 1, 2, \ldots, 24$ and $\langle \cdot, \cdot \rangle$ is the intersection pairing on $H^1(E, \mathbb{Z})$. The monodromy action of $\pi_1(B^\circ)$ on $H^1(E, \mathbb{Z})$ is the subgroup of $\text{Aut}(H^1(E, \mathbb{Z}))$ generated by $T_{\delta_1}, T_{\delta_2}, \ldots, T_{\delta_{24}}$. Clearly, T_{δ_i} lift to actions on $H^1(E, \mathbb{Z})/nH^1(E, \mathbb{Z})$. We start with a simple observation:

Lemma 3.2. *Let $\delta_1, \delta_2, \ldots, \delta_{24} \in H^1(X_b, \mathbb{Z})$ be the vanishing cycles associated to a Bryan–Leung* K3 *surface $\pi : X \to B = \mathbb{P}^1$ with* 24 *nodal fibers. Then*:

(1) *The δ_i are indivisible, i.e., there do not exist $\eta \in H^1(X_b, \mathbb{Z})$ and an integer $m \geq 2$ such that $\delta_i = m\eta$.*

(2) *For every indivisible $\lambda \in H^1(X_b, \mathbb{Z})$,*

$$\gcd(\langle \lambda, \delta_1 \rangle, \langle \lambda, \delta_2 \rangle, \ldots, \langle \lambda, \delta_{24} \rangle) = 1.$$

Proof. It is well known that the δ_i are indivisible, as a consequence of the smoothness of X. (See [Lew99, Example 6.6, p. 72], for instance.) Here we give another argument based on torsion points.

Suppose that $\delta/m \in H^1(E, \mathbb{Z})$ for some $\delta = \delta_i$ and $m \geq 2$. For simplicity, let us assume that m is prime. Then $H^1(E, \mathbb{Z})/mH^1(E, \mathbb{Z})$ is fixed by T_δ so Σ_m is the union $Q_1 \cup Q_2 \cup \cdots \cup Q_{m^2-1}$ of $m^2 - 1$ local sections over a disk $U \subset B$ around the point $s = b_i \in S$. Since X is smooth, each Q_j meets X_s at a point away from the node x of X_s. Let $f : X \dashrightarrow X$ be the rational map given by $f(q) = p + m(q - p)$ for $q \in X_b$, $b \in B^\circ$ and $p = P \cap X_b$. Then f can be extended to a regular, quasifinite and unramified morphism

$$X \setminus \{x_1, x_2, \ldots, x_{24}\} \xrightarrow{f} X,$$

where $x_1, x_2, \ldots, x_{24}$ are the nodes of the 24 fibers $X_S = \pi^{-1}(S)$. Then

$$X_U \cap f^{-1}(P) = P \cup Q_1 \cup Q_2 \cup \cdots \cup Q_{m^2-1}$$

for $X_U = \pi^{-1}(U)$. Since f is unramified, $P, Q_1, Q_2, \ldots, Q_{m^2-1}$ are disjoint. Therefore, $p = P \cap X_s$ and $q_j = Q_j \cap X_s$ are m^2 distinct points on $X_s \setminus \{x\}$. But there are only m distinct points q on $X_s \setminus \{x\}$ such that $m(q - p) = 0$ in $\mathrm{Pic}^0(X_s) \cong \mathbb{C}^*$, which is a contradiction.

For (2), if

$$\gcd(\langle \lambda, \delta_1 \rangle, \langle \lambda, \delta_2 \rangle, \ldots, \langle \lambda, \delta_{24} \rangle) = m \geq 2,$$

then $\lambda \in H^1(E, \mathbb{Z})/mH^1(E, \mathbb{Z})$ is fixed by T_{δ_i} for all i. Therefore, Σ_m contains a section. But P is the only section of X/B, which is a contradiction. □

Proof of Proposition 3.1. If $n = n_1 n_2$ for two coprime integers n_1 and n_2, then the surjectivity of $\pi_1(B^\circ) \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ follows from those of $\pi_1(B^\circ) \to \mathrm{SL}_2(\mathbb{Z}/n_i\mathbb{Z})$ for $i = 1, 2$ via the group isomorphism

$$\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z}) \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{Z}/n_1\mathbb{Z}) \times \mathrm{SL}_2(\mathbb{Z}/n_2\mathbb{Z})$$

So by induction on the number of prime divisors of n, it suffices to prove the proposition for $n = p^d$ with p prime.

For simplicity, suppose that $\delta_1 = e_1$, where $\{e_1, e_2\}$ is the standard basis of $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. By Lemma 3.2,

$$\gcd(\langle \delta_1, \delta_2 \rangle, \langle \delta_1, \delta_3 \rangle, \ldots, \langle \delta_1, \delta_{24} \rangle) = 1$$

So there exists $2 \leq i \leq 24$ such that $p \nmid \langle \delta_1, \delta_i \rangle$. We may assume that $p \nmid \langle \delta_1, \delta_2 \rangle$. Then $\delta_2 = ae_1 + be_2$ for some $p \nmid b$. Let m be an integer such that $bm \equiv 1 \pmod{n}$. By changing the basis from $\{e_1, e_2\}$ to $\{e_1, ame_1 + e_2\}$, we may assume that $\delta_2 = be_2$.

Clearly,

$$T_{\delta_1}^{-1} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad (T_{\delta_2})^m \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

for $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = xe_1 + ye_2$, and hence $\pi_1(B^\circ) \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ is surjective. □

Let us consider the degeneration of the function $b_{\tau,p}(z)$ when X_t degenerates to X_0 for some $0 \in S$.

Proposition 3.3. *Let $\pi : X \to \Delta$ be a flat projective family of curves over the unit disk Δ such that X is smooth, X_t is a smooth elliptic curve for $t \neq 0$ and X_0 is a rational curve with a node, where X_t is the fiber of X over $t \in \Delta$. Let P and Q be two sections of X/Δ such that $P_t - Q_t$ is a torsion class in $J(X_t)$ of order $n \geq 2$ for $t \neq 0$. Then there exists an integral curve $Z \subset X$ flat of degree 2 over Δ such that Z_0 is supported on the node of X_0 and*

$$\{b_{\tau,p}(z) = 0\} = Z_t \tag{3-2}$$

for $t \neq 0$, where $b_{\tau,p}(z)$ is the meromorphic function on X_t given in Proposition 2.3 with $\tau = P_t - Q_t$ and $p = P_t$.

Proof. Since P and Q are sections of X/Δ and X is smooth, P and Q meet X_0 at smooth points P_0 and Q_0 of X_0. By the argument in the proof of Lemma 3.2, $P_0 - Q_0$ is a torsion class in $\operatorname{Pic}^0(X_0) \cong \mathbb{C}^*$ of order n.

Let us consider $\pi_* \mathcal{O}_X(P + Q)$. This is a rank 2 vector bundle over Δ since $h^0(\mathcal{O}_{X_t}(P + Q)) = 2$ for all t. Therefore,

$$H^0(\pi_* \mathcal{O}_X(P + Q)) = H^0(\mathcal{O}_X(P + Q))$$

is a rank 2 free module over $\mathbb{C}[[t]]$.

Let o be the node of X_0. Then $X_0 \setminus \{o\} \cong \mathbb{C}^*$. We may assume that $P_0 = 1$ and $Q_0 = \eta = \exp(2\pi i/n)$. Then $H^0(\mathcal{O}_{X_0}(P_0 + Q_0))$ is spanned by the constant function 1 and

$$s_0(z) = \frac{z}{(z-1)(z-\eta)}$$

over $\mathbb{C}$. We can choose $s \in H^0(\mathcal{O}_X(P + Q))$ such that s_0 is the restriction of s to X_0, i.e., $s_0(z) = s(0, z)$, where we consider $s = s(t, z)$ as a meromorphic function on X with simple poles along P and Q. Then $H^0(\mathcal{O}_X(P + Q))$ is generated by 1 and s over $\mathbb{C}[[t]]$.

Let $\phi : X \setminus \{o\} \to X \setminus \{o\}$ be the automorphism given by $\phi(z) = z + (p - q)$ for $z \in X_t$, $p = P_t$ and $q = Q_t$. Then

$$\sum_{k=0}^{n-1} s(t, \phi^k(z))$$

is constant for each fixed $t \neq 0$ by Proposition 2.3. For $t = 0$, we have

$$\sum_{k=0}^{n-1} s(0, \phi^k(z)) = \sum_{k=0}^{n-1} \frac{\eta^k z}{(\eta^k z - 1)(\eta^k z - \eta)} = 0.$$

Therefore,

$$f(t) = \sum_{k=0}^{n-1} s(t, \phi^k(z))$$

for some $f(t) \in \mathbb{C}[[t]]$ with $f(0)=0$. Then $ns(t,z)-f(t)$ is a section of $\mathcal{O}_X(P+Q)$ whose restriction to X_t is exactly the function $b_{\tau,p}(z)$.

Let

(3-3) $$Z = \{ns(t,z) - f(t) = 0\}$$

be the vanishing locus of $ns(t,z) - f(t)$. Then (3-2) follows from our choice of $f(t)$. In addition, since $ns(0,z) - f(0) = ns_0(z)$ and s_0 only vanishes at the node o of X_0, we see that Z_0 is supported at o.

We know that Z is a closed subscheme of X of pure dimension one and flat of degree 2 over Δ. So we are in one of the following cases:

- Z is supported on a section of X/Δ with multiplicity 2.
- Z is a union of two distinct sections of X/Δ.
- Z is an irreducible multisection of degree 2 over Δ.

Since Z_0 is supported on the node o of X_0 and X is smooth, Z cannot contain any section of X/Δ. Thus, Z must be an integral curve flat of degree 2 over Δ. □

Proposition 2.4 follows immediately from the above proposition.

Proof of Proposition 2.4. Suppose that $b_{\tau,p}(p-\eta)=0$ on a general elliptic curve E for some torsion class $\tau \in J(E)$ of order $n \geq 2$ and $2\eta = \tau$. Then by Proposition 3.1, this holds for every torsion class τ of order n.

Let $\pi : X \to B = \mathbb{P}^1$ be a Bryan–Leung K3 surface with 24 nodal fibers over $S \subset B$. We choose a point $s \in S$ and let $U \subset B$ be an open disk about s. Then there exists a section Q of $X_U = \pi^{-1}(U)$ over U such that $P_t - Q_t$ is a torsion class of order n for all $t \in U$. It follows from Proposition 3.3 that $b_{\tau,p}(z)$ has two distinct zeros on X_t for $\tau = P_t - Q_t$ and $p = P_t$, which is a contradiction. □

4. Proof of Proposition 2.5

In this section, we will prove Proposition 2.5. Combined with Proposition 2.4, we obtain Proposition 2.2. Then Theorem 1.2 follows.

We will prove the following two statements in sequence:

Proposition 4.1. *For a general elliptic curve E, a point $p \in E$ and a pair $\tau_1 \neq \tau_2 \in J(E)_{\text{tors}}$ of torsion points of orders $n_1 \geq 2$ and $n_2 \geq 2$, respectively, if*

$$\{b_{\tau_1,p}(z)=0\} \cap \{b_{\tau_2,p}(z)=0\} \neq \varnothing$$

then either

(4-1) $$\{p - q : b_{\tau_i,p}(q) = 0\} \subset J(E)_{\text{tors}}$$

for $i = 1, 2$ or

(4-2) $$n_1 = n_2 = 6, \quad \langle \tau_1, \tau_2 \rangle = 3 \text{ in } J(E)_6 \quad \text{and} \quad \operatorname{ord}(\tau_1 - \tau_2) = 6.$$

Proposition 4.2. *For a general elliptic curve E, a point $p \in E$ and a pair $\tau_1 \neq \tau_2 \in J(E)_{\text{tors}}$ of nonzero torsion points, if*

$$\{b_{\tau_1,p}(z)=0\} \cap \{b_{\tau_2,p}(z)=0\} \neq \varnothing \tag{4-3}$$

and

$$\{p-q : b_{\tau_i,p}(q)=0\} \subset J(E)_{\text{tors}} \quad \textit{for } i=1,2,$$

then $\operatorname{ord}(\tau_1)=\operatorname{ord}(\tau_2)=2$.

Our main tool is the monodromy action of $\pi_1(B^\circ)$ on $J(E)_{\text{tors}}$. We fix a Bryan–Leung K3 surface $X \to B = \mathbb{P}^1$ with 24 nodal fibers over $S \subset B$ and a general fiber $E = X_t$ of X/B. We extend the monodromy action on $J(E)_{\text{tors}}$ to the triples (τ, q_1, q_2) with $\tau \in J(E)_{\text{tors}}$ and $\{b_{\tau,p}(z)=0\} = \{q_1, q_2\}$.

Define a curve in $\operatorname{Pic}^0(X/B) \times_B X \times_B X$ by

$$\big\{(\tau, q_1, q_2) : \tau \in J(X_t)_n,\ t \in B \setminus S,\ q_1, q_2 \in X_t, \text{ and } \{b_{\tau,p}(z)=0\} = \{q_1, q_2\} \text{ for } p = P_t\big\}. \tag{4-4}$$

By Proposition 2.4, for each fixed $n \geq 2$, there exists a finite set $S_n \subset B$ such that for every $t \notin S \cup S_n$, $b_{\tau,p}(z)$ has no double zeros on X_t. So the curve defined by (4-4) is unramified over $B \setminus (S \cup S_n)$ and we have a well-defined monodromy action of $\pi_1(B \setminus (S \cup S_n))$ on such triples (τ, q_1, q_2) on a general fiber $E = X_t$. Let us use the notation $\lambda(\tau)$ and $\lambda(\tau, q_1, q_2)$ to denote the action of $\lambda \in \pi_1(B \setminus (S \cup S_n))$ on $\tau \in J(E)_{\text{tors}}$ and (τ, q_1, q_2).

Lemma 4.3. *Let $X \to B = \mathbb{P}^1$ be a Bryan–Leung* K3 *surface with* 24 *nodal fibers and let $E = X_t$ be a general fiber of X/B. Let $\tau \in J(E)_{\text{tors}}$ be a torsion class of order $n \geq 2$ and let $q_1, q_2 \in E$ be two points given by*

$$\{b_{d\tau,p}(z)=0\} = \{q_1, q_2\}$$

for some integer d with $d\tau \neq 0$. If $\lambda \in \pi_1(B \setminus (S \cup S_n))$ acts on $J(E)_n$ by

$$\lambda(\eta) = \eta + \langle \eta, \tau \rangle \tau$$

for all $\eta \in J(E)_n$, then

$$\lambda(d\tau, q_1, q_2) = (d\tau, q_2, q_1).$$

Proof. Fix a point $0 \in S$ and let δ be the vanishing cycle associated to the nodal fiber X_0. If $\tau = \delta$ in $J(E)_n$, then we must have $\lambda = T_\delta$ in $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$, where T_δ is the Picard–Lefschetz transform associated to δ. Since

$$T_\delta(d\tau) = d\tau,$$

there is a local section $Q \subset X_U = X \times_B U$ over a simply connected open neighborhood U of 0 such that $P_t - Q_t = d\tau$. Then the lemma follows from Proposition 3.3.

More generally, by Proposition 3.1, there exists $\alpha \in \pi_1(B \setminus (S \cup S_n))$ such that $\alpha(\delta) = \tau$. Then $T_\delta = \alpha^{-1} \circ \lambda \circ \alpha$ since

$$\begin{aligned}\alpha^{-1} \circ \lambda \circ \alpha(\eta) &= \alpha^{-1}\big(\alpha(\eta) + \langle \alpha(\eta), \alpha(\delta)\rangle \alpha(\delta)\big)\\ &= \alpha^{-1}\big(\alpha(\eta) + \langle \eta, \delta\rangle \alpha(\delta)\big)\\ &= \alpha^{-1} \circ \alpha(\eta + \langle \eta, \delta\rangle \delta) = T_\delta(\eta).\end{aligned}$$

Thus, the lemma follows. □

Lemma 4.4. *Let $X \to B = \mathbb{P}^1$ be a Bryan–Leung* K3 *surface with* 24 *nodal fibers and let $E = X_t$ be a general fiber of X/B. Let τ_1 and $\tau_2 \in J(E)_{tors}$ be two torsion classes of the same order $n \geq 2$ with $m = \langle \tau_1, \tau_2\rangle$ in $J(E)_n$, let n_1, n_2 be two integers such that $n \nmid n_i$ and let*

$$\{b_{n_1\tau_1,p}(z) = 0\} = \{q_1, q_2\}.$$

If $b_{n_2\tau_2,p}(q_1) = 0$, then

$$\begin{aligned} &b_{n_2(\tau_2+km\tau_1),p}(q_1) = 0 \quad \text{if } 2 \mid k,\\ &b_{n_2(\tau_2+km\tau_1),p}(q_2) = 0 \quad \text{if } 2 \nmid k.\end{aligned} \tag{4-5}$$

If, in addition, $(2\gcd(mn_2, n)) \nmid n$, then $n_1\tau_1 = n_2\tau_2$.

Proof. By Proposition 3.1, we can find $\lambda \in \pi_1(B \setminus (S \cup S_n))$ such that

$$\lambda(\alpha) = \alpha + \langle \alpha, \tau_1\rangle \tau_1$$

for all $\alpha \in J(E)_n$. Then $\lambda(\tau_1) = \tau_1$. Hence, by Lemma 4.3, we have

$$\begin{aligned} &\lambda^k(n_1\tau_1, q_1, q_2) = (n_1\tau_1, q_1, q_2) \quad \text{if } 2 \mid k,\\ &\lambda^k(n_1\tau_1, q_1, q_2) = (n_1\tau_1, q_2, q_1) \quad \text{if } 2 \nmid k.\end{aligned} \tag{4-6}$$

Obviously,

$$\lambda^k(\tau_2) = \tau_2 - km\tau_1 \tag{4-7}$$

for all integers k. Combining (4-6) and (4-7), we obtain (4-5).

If $(2\gcd(mn_2, n)) \nmid n$, then $k_0 = n/\gcd(mn_2, n)$ is odd. Setting $k = k_0$ in (4-5), we obtain

$$b_{n_2\tau_2,p}(q_2) = b_{n_2(\tau_2+k_0m\tau_1),p}(q_2) = 0.$$

On the other hand, we assume that $b_{n_2\tau_2,p}(q_1) = 0$. So

$$\{b_{n_i\tau_i,p}(z) = 0\} = \{q_1, q_2\}$$

for $i = 1, 2$. This implies

$$n_1\tau_1 = (p - q_1) + (p - q_2) = n_2\tau_2.$$

□

Lemma 4.5. *Let E be an elliptic curve, let p be a point on E and let $\tau \in J(E)_{\text{tors}}$ be a torsion point of order 2. Then*

$$\{b_{\tau,p}(z) = 0\} = \{q_1, q_2\},$$

where q_1, q_2 are such that τ, $p - q_1$ and $p - q_2$ are the three distinct 2-torsion points.

Proof. Let τ, τ_1 and τ_2 be the three distinct 2-torsion points. Clearly,

$$\tau = \tau_1 + \tau_2.$$

So there exist a rational function $b(z)$ on E with simple poles at p and $p - \tau$ and simple zeros at $p - \tau_1$ and $p - \tau_2$. Note that $b(z + \tau)$ also has simple poles at p and $p - \tau$ and simple zeros at $p - \tau_1$ and $p - \tau_2$. Therefore,

$$b(z + \tau) \equiv cb(z)$$

for a constant c. And since $b(z) + b(z + \tau)$ is a constant by Proposition 2.3, we must have $c = -1$ and

$$b(z) + b(z + \tau) \equiv 0.$$

Therefore, $b_{\tau,p}(z) \equiv \lambda b(z)$ for a constant $\lambda \neq 0$ by the uniqueness of $b_{\tau,p}(z)$ and the lemma follows. □

Lemma 4.6. *Let E be an elliptic curve, let p be a point on E and let $\tau_1 \neq \tau_2 \in J(E)_{\text{tors}}$ be two distinct nonzero torsion classes. If*

$$\{b_{\tau_1,p}(z) = 0\} = \{q_1, q_2\} \quad \textit{and} \quad \{b_{\tau_2,p}(z) = 0\} = \{q_1, q_3\},$$

then

$$b_{\tau_1-\tau_2,p}(q_2) = 0.$$

Proof. From $\tau_1 = (p - q_1) + (p - q_2)$ and $\tau_2 = (p - q_1) + (p - q_3)$ we have $q_2 = q_3 - (\tau_1 - \tau_2)$. Let us consider the meromorphic function $b_{\tau_2,p}(z + (\tau_1 - \tau_2))$. It has simple poles at $p - (\tau_1 - \tau_2)$ and $(p - \tau_2) - (\tau_1 - \tau_2) = p - \tau_1$ and a zero at

$$q_3 - (\tau_1 - \tau_2) = q_2.$$

Therefore,

$$b(z) = b_{\tau_1,p}(z) + cb_{\tau_2,p}(z + (\tau_1 - \tau_2))$$

has simple poles at p and $p - (\tau_1 - \tau_2)$ and a zero at q_2 for the constant c given by

$$c = -\frac{\operatorname{Res}_{p-\tau_1} b_{\tau_1,p}(z)\omega}{\operatorname{Res}_{p-\tau_1} b_{\tau_2,p}(z + (\tau_1 - \tau_2))\omega},$$

where ω is a nonvanishing holomorphic 1-form on E.

Let n be a positive integer such that $\tau_1, \tau_2 \in J(E)_n$. Then

$$\begin{aligned}\sum_{\lambda\in J(E)_n} b(z+\lambda) &= \sum_{\lambda\in J(E)_n} b_{\tau_1,p}(z+\lambda) + c\sum_{\lambda\in J(E)_n} b_{\tau_2,p}(z+(\tau_1-\tau_2)+\lambda)\\ &= \sum_{\lambda\in J(E)_n} b_{\tau_1,p}(z+\lambda) + c\sum_{\lambda\in J(E)_n} b_{\tau_2,p}(z+\lambda) \equiv 0\end{aligned}$$

by Proposition 2.3. Then by the uniqueness of $b_{\tau_1-\tau_2,p}(z)$, we must have

$$b_{\tau_1-\tau_2,p}(z) \equiv ab(z)$$

for some constant $a \neq 0$ and the lemma follows. □

Lemma 4.7. *Let E be an elliptic curve, let n be a positive integer satisfying $4\,|\,n$ and $8\nmid n$ and let $\alpha_1 \neq \alpha_2 \in J(E)_{\mathrm{tors}}$ be two torsion classes of order n. If*

$$\langle\alpha_1,\alpha_2\rangle = n/2 \ \text{ in } J(E)_n \quad \text{and} \quad 4(d_1\alpha_1 - d_2\alpha_2) = 0$$

for some odd integers d_1 and d_2, then

$$\operatorname{ord}(d_1\alpha_1 - d_2\alpha_2) = 2.$$

Proof. Let $m = n/2$. We may assume that $\alpha_1 = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\alpha_2 = \left[\begin{smallmatrix}a\\m\end{smallmatrix}\right]$ in $J(E)_n \cong \mathbb{Z}/n\mathbb{Z}\times\mathbb{Z}/n\mathbb{Z}$, where $\gcd(a,m)=1$ and hence a is odd. Then

$$d_1\alpha_1 - d_2\alpha_2 = \begin{bmatrix} d_1 - ad_2\\ -d_2m\end{bmatrix}$$

and $2m\,|\,4(d_1-ad_2)$. Since $d_1 - ad_2$ is even and $4\nmid m$, we see that $2m\,|\,2(d_1-ad_2)$ and hence $d_1\alpha_1 - d_2\alpha_2$ has order 2. □

Proof of Proposition 4.1. Suppose that E is a general fiber of a Bryan–Leung K3 surface $\pi : X \to B = \mathbb{P}^1$ with 24 nodal fibers. Let

$$n = \operatorname{lcm}(n_1,n_2), \quad d_1 = \frac{n}{n_1} \quad \text{and} \quad d_2 = \frac{n}{n_2}.$$

Suppose that

$$\{b_{\tau_1,p}(z)=0\} = \{q_1,q_2\} \quad \text{and} \quad \{b_{\tau_2,p}(z)=0\} = \{q_1,q_3\}.$$

It suffices to prove that one of $p-q_1$, $p-q_2$ and $p-q_3$ is torsion.

Since $\operatorname{ord}(\tau_i) = n_i$, we have $\tau_i = d_i\alpha_i$ for $i=1,2$ and some $\alpha_i \in J(E)_{\mathrm{tors}}$ of order n. Let $m = \langle\alpha_1,\alpha_2\rangle \in \mathbb{Z}/n\mathbb{Z}$.

By Lemma 4.4,

$$\begin{aligned} b_{\tau_2+kd_2m\alpha_1,p}(q_1) &= 0 \quad \text{if } 2\,|\,k,\\ b_{\tau_2+kd_2m\alpha_1,p}(q_2) &= 0 \quad \text{if } 2\nmid k.\end{aligned}$$

If $k_0 = n/\gcd(d_2m, n)$ is odd, then $\tau_1 = \tau_2$ by Lemma 4.4, which is a contradiction. Therefore, k_0 and n are even. If $k_0 \neq 2$, we have two cases:

Suppose that $4 \mid k_0$. We have

$$b_{\tau_2,p}(q_1) = b_{\tau_2+(k_0/2)d_2m\alpha_1,p}(q_1) = 0.$$

Let

$$\tau_1' = \tau_2 + (k_0/2)d_2m\alpha_1 \quad \text{and} \quad \tau_2' = \tau_2.$$

Suppose that

$$\{b_{\tau_1',p}(z) = 0\} = \{q_1, q_2'\} \quad \text{and} \quad \{b_{\tau_2',p}(z) = 0\} = \{q_1, q_3'\}.$$

By Lemma 4.6,

$$b_{\tau_1'-\tau_2',p}(q_2') = 0.$$

Obviously, $\operatorname{ord}(\tau_1' - \tau_2') = 2$. Therefore, $p - q_2' \in J(E)_{\text{tors}}$ by Lemma 4.5. It follows that $p - q_1 \in J(E)_{\text{tors}}$ and we are done.

Suppose that $4 \nmid k_0$ and $k_0 > 2$. We have

$$b_{\tau_2,p}(q_1) = b_{\tau_2+2d_2m\alpha_1,p}(q_1) = 0.$$

Let

$$\tau_1' = \tau_2 + 2d_2m\alpha_1 \quad \text{and} \quad \tau_2' = \tau_2.$$

We see that $\tau_1' \neq \tau_2'$, $\operatorname{ord}(\tau_1') \mid n_2 = \operatorname{ord}(\tau_2')$ and

$$\langle \tau_1', \tau_2' \rangle = m' = 2(d_2m)^2$$

with $n_2/\gcd(m', n_2)$ odd. Then $\tau_1' = \tau_2'$ by Lemma 4.4, which is a contradiction.

So we have $k_0 = 2$. That is,

$$n = 2\gcd(d_2m, n).$$

Similarly, we have

$$n = 2\gcd(d_1m, n).$$

So we have

$$d_2m \equiv d_1m \equiv \frac{n}{2} \pmod{n}.$$

And since $\gcd(d_1, d_2) = 1$, we conclude that

$$m \equiv \frac{n}{2} \pmod{n}$$

and d_1 and d_2 are both odd. That is, we have reduced the proposition to the case that

$$2 \mid n, \;\; 2 \nmid d_1d_2 \quad \text{and} \quad m = \frac{n}{2}. \tag{4-8}$$

Note that under these assumptions,

$$m\tau_j = d_i m\alpha_j = m\alpha_j$$

for all $i, j = 1, 2$.

If one of the τ_i is 2-torsion, then it follows immediately from Lemma 4.5 that $p - q_1 \in J(E)_{\text{tors}}$ and we are done. So we may assume that $n_i \geq 3$ for $i = 1, 2$.

By Lemma 4.6,

$$b_{\tau_1 - \tau_2, p}(q_2) = 0.$$

If $\tau_1 - \tau_2$ is a 2-torsion class, then $p - q_2 \in J(E)_{\text{tors}}$ by Lemma 4.5. We are again done. So we may assume that none of τ_1, τ_2 and $\tau_1 - \tau_2$ are 2-torsion classes. That is, we may assume that

$$n_1 \geq 3, \quad n_2 \geq 3 \quad \text{and} \quad \operatorname{ord}(\tau_1 - \tau_2) \geq 3, \tag{4-9}$$

in addition to (4-8).

Repeatedly applying Lemma 4.4, we obtain

$$\begin{aligned} \{b_{\tau_1, p}(z) = 0\} &= \{q_1, q_2\}, \\ \{b_{\tau_2, p}(z) = 0\} &= \{q_1, q_3\}, \\ \{b_{\tau_2 + m\alpha_1, p}(z) = 0\} &= \{q_2, q_4\}, \\ \{b_{\tau_1 + m\alpha_2, p}(z) = 0\} &= \{q_3, q_5\}. \end{aligned}$$

Continuing this process, we obtain

$$b_{\tau_1 + m(\alpha_2 + m\alpha_1), p}(q_4) = 0.$$

Suppose that $4 \mid n$, i.e., $2 \mid m$. Then $m(\alpha_2 + m\alpha_1) = m\alpha_2$ and hence

$$b_{\tau_1 + m\alpha_2, p}(q_4) = 0.$$

Since $\{b_{\tau_1 + m\alpha_2, p}(z) = 0\} = \{q_3, q_5\}$, we have either $q_3 = q_4$ or $q_4 = q_5$.

- If $q_3 = q_4$, then

$$\begin{aligned} \{b_{\tau_1, p}(z) = 0\} &= \{q_1, q_2\}, \\ \{b_{\tau_2, p}(z) = 0\} &= \{q_1, q_3\}, \\ \{b_{\tau_2 + m\alpha_1, p}(z) = 0\} &= \{q_2, q_3\}, \end{aligned}$$

and hence

$$\begin{aligned} (p - q_1) + (p - q_2) &= \tau_1 \in J(E)_{\text{tors}}, \\ (p - q_1) + (p - q_3) &= \tau_2 \in J(E)_{\text{tors}}, \\ (p - q_2) + (p - q_3) &= \tau_2 + m\alpha_1 \in J(E)_{\text{tors}}. \end{aligned}$$

It follows that $p - q_1$, $p - q_2$, $p - q_3 \in J(E)_{\text{tors}}$. We are done.

- If $q_4 = q_5$, then

$$\begin{aligned}\{b_{\tau_1,p}(z)=0\} &= \{q_1,q_2\},\\ \{b_{\tau_2,p}(z)=0\} &= \{q_1,q_3\},\\ \{b_{\tau_2+m\alpha_1,p}(z)=0\} &= \{q_2,q_4\},\\ \{b_{\tau_1+m\alpha_2,p}(z)=0\} &= \{q_3,q_4\},\end{aligned}$$

and hence

$$\begin{aligned}&(p-q_1)+(p-q_2)=\tau_1,\\ &(p-q_1)+(p-q_3)=\tau_2,\\ &(p-q_2)+(p-q_4)=\tau_2+m\alpha_1=\tau_2+m\tau_1,\\ &(p-q_3)+(p-q_4)=\tau_1+m\alpha_2=\tau_1+m\tau_2.\end{aligned}$$

It follows that

$$(m-2)(\tau_1-\tau_2)=0 \Rightarrow \gcd(m-2,n)(\tau_1-\tau_2)=0.$$

Since $\gcd(m-2,n)=\gcd(m-2,2m)$ is either 2 or 4, the order of $\tau_1-\tau_2$ is either 2 or 4. By our hypothesis (4-9), $\operatorname{ord}(\tau_1-\tau_2)\neq 2$. So $\operatorname{ord}(\tau_1-\tau_2)=4$. Then $\gcd(m-2,2m)=4$ and $4\nmid m$. This contradicts Lemma 4.7.

So far we have proved the proposition when m is even. Suppose that $2\nmid m$. Then $m(\alpha_2+m\alpha_1)=m(\alpha_1+\alpha_2)$ and hence

$$b_{\tau_1+m(\alpha_1+\alpha_2),p}(q_4)=0.$$

Continuing with the use of Lemma 4.4, we obtain

$$\begin{aligned}\{b_{\tau_1,p}(z)=0\} &= \{q_1,q_2\},\\ \{b_{\tau_2,p}(z)=0\} &= \{q_1,q_3\},\\ \{b_{\tau_2+m\alpha_1,p}(z)=0\} &= \{q_2,q_4\},\\ \{b_{\tau_1+m\alpha_2,p}(z)=0\} &= \{q_3,q_5\},\\ \{b_{\tau_1+m(\alpha_1+\alpha_2),p}(z)=0\} &= \{q_4,q_6\},\\ \{b_{\tau_2+m(\alpha_1+\alpha_2),p}(z)=0\} &= \{q_5,q_7\}.\end{aligned}$$

Applying Lemma 4.4 to $(\tau_1+m(\alpha_1+\alpha_2),\ \tau_2+m\alpha_1)$, we obtain

$$b_{\tau_2+m(\alpha_1+\alpha_2),p}(q_6)=0.$$

Similarly,

$$b_{\tau_1+m(\alpha_1+\alpha_2),p}(q_7)=0.$$

That is, $q_6\in\{q_5,q_7\}$ and $q_7\in\{q_4,q_6\}$. Since $\{q_5,q_7\}\neq\{q_4,q_6\}$, we must have $q_6=q_7$. Then from

$$\begin{aligned}\{b_{\tau_1,p}(z)=0\} &= \{q_1,q_2\},\\ \{b_{\tau_2,p}(z)=0\} &= \{q_1,q_3\},\end{aligned}$$

$$\{b_{\tau_2+m\alpha_1,p}(z)=0\}=\{q_2,q_4\},$$
$$\{b_{\tau_1+m\alpha_2,p}(z)=0\}=\{q_3,q_5\},$$
$$\{b_{\tau_1+m(\alpha_1+\alpha_2),p}(z)=0\}=\{q_4,q_6\},$$
$$\{b_{\tau_2+m(\alpha_1+\alpha_2),p}(z)=0\}=\{q_5,q_6\},$$

we obtain

$$3(\tau_1-\tau_2)=m(\alpha_1-\alpha_2).$$

Hence $\tau_1-\tau_2$ has order 2 or 6.

By our hypothesis (4-9), $\operatorname{ord}(\tau_1-\tau_2)\neq 2$. So $\tau_1-\tau_2$ has order 6. Hence $6\,|\,n$, $3\,|\,m$ and $3\,|\,n_1n_2$.

Since d_1 and d_2 are odd, $n_1=n/d_1$ and $n_2=n/d_2$ are even. So at least one of n_1 and n_2 is divisible by 6. Without loss of generality, let us assume that $6\,|\,n_1$. Then

$$n_1(\tau_1-\tau_2)=0 \implies n_1\tau_2=0 \implies n_2\,|\,n_1 \implies n=n_1.$$

Let

$$\tau_1'=\tau_1 \quad \text{and} \quad \tau_2'=\tau_1-\tau_2.$$

By Lemma 4.6,

$$b_{\tau_1',p}(q_2)=b_{\tau_2',p}(q_2)=0.$$

Applying the whole argument to (τ_1',τ_2'), we again arrive at

$$\operatorname{ord}(\tau_1'-\tau_2')=6.$$

That is, $n_2=\operatorname{ord}(\tau_2)=6$. Then this implies that $\tau_1=\tau_2+(\tau_1-\tau_2)$ also has order 6. So we have (4-2). □

Proof of Proposition 4.2. Let

$$\{b_{\tau_1,p}(z)=0\}=\{q_1,q_2\} \quad \text{and} \quad \{b_{\tau_2,p}(z)=0\}=\{q_1,q_3\},$$

where $\eta_i=p-q_i$ are torsion for $i=1,2,3$.

Suppose that $n=\operatorname{lcm}(\operatorname{ord}(\tau_1),\operatorname{ord}(\eta_1),\operatorname{ord}(\eta_2))$ and that $\tau_1=d\alpha_1$ for some $\alpha_1\in J(E)_{\text{tors}}$ of order n. Let E be a general fiber of a Bryan–Leung K3 surface $X\to B=\mathbb{P}^1$ with 24 nodal fibers. Clearly, each $\lambda\in\pi_1(B\setminus(S\cup S_n))$ acts on (τ_1,q_1,q_2) by

$$\lambda(\tau_1,q_1,q_2)=(\lambda(\tau_1),\,p-\lambda(\eta_1),\,p-\lambda(\eta_2)).$$

On the other hand, for $\lambda(\eta)=\eta+\langle\eta,\alpha_1\rangle\alpha_1$,

$$\lambda(\tau_1,q_1,q_2)=(\lambda(\tau_1),q_2,q_1)$$

by Lemma 4.3. Therefore,

$$\eta_2=\lambda(\eta_1)=\eta_1+m\alpha_1$$

for $m = \langle \eta_1, \alpha_1 \rangle$. And since $\tau_1 = \eta_1 + \eta_2$, we have

$$\tau_1 = 2\eta_1 + m\alpha_1 \implies \langle (d-m)\alpha_1, \alpha_1 \rangle = \langle 2\eta_1, \alpha_1 \rangle \implies 2m = 0$$

in $\mathbb{Z}/n\mathbb{Z}$. If $m = 0$, then $\eta_1 = \eta_2$, which contradicts Proposition 2.4. So n is even and $m = n/2$. Therefore, we have

$$\operatorname{ord}(\eta_1 - \eta_2) = \operatorname{ord}(\tau_1 - 2\eta_1) = 2. \tag{4-10}$$

Similarly,

$$\operatorname{ord}(\eta_1 - \eta_3) = \operatorname{ord}(\tau_2 - 2\eta_1) = 2. \tag{4-11}$$

It follows that $\tau_1 - \tau_2$ is a 2-torsion class as well. By Lemma 4.6,

$$b_{\tau_1 - \tau_2, p}(q_2) = 0.$$

Hence η_2 is a 2-torsion class by Lemma 4.7. Together with (4-10) and (4-11), we see that all of $\tau_1, \tau_2, \eta_1, \eta_2, \eta_3$ are 2-torsion classes. □

To finish the proof of Proposition 2.5, it remains to justify (2-15).

Proof of Proposition 2.5. We have proved (2-11) with two exceptions outlined in the proposition.

If (2-15) fails, we must have one of the following:

A. τ_1, τ_2, τ_3 are three distinct 2-torsion points.

B. τ_1, τ_2, τ_3 are three distinct 6-torsion points satisfying that $\langle \tau_i, \tau_j \rangle = 3$ and $\operatorname{ord}(\tau_i - \tau_j) = 6$ for all $1 \le i < j \le 3$.

In case A, by (2-14),

$$\begin{aligned} \{b_{\tau_1,p}(z) = 0\} \cap \{b_{\tau_2,p}(z) = 0\} &= \{p - \tau_3\}, \\ \{b_{\tau_2,p}(z) = 0\} \cap \{b_{\tau_3,p}(z) = 0\} &= \{p - \tau_1\}, \end{aligned}$$

and (2-15) follows.

In case B, we may assume that

$$\tau_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in J(E)_6 \cong \mathbb{Z}/6\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$$

Since $\langle \tau_i, \tau_j \rangle = 3$ for $i \neq j$, we must have $\tau_2 = \begin{bmatrix} a \\ 3 \end{bmatrix}$ and $\tau_3 = \begin{bmatrix} b \\ 3 \end{bmatrix}$ for some $a, b \in \mathbb{Z}$ satisfying $3 \nmid ab$ and $2 \nmid (a - b)$.

Since $\operatorname{ord}(\tau_1 - \tau_2) = \operatorname{ord}(\tau_1 - \tau_3) = 6$, $3 \nmid (a-1)(b-1)$. Together with $3 \nmid ab$, we must have

$$a \equiv b \equiv 2 \pmod 3.$$

Then $\operatorname{ord}(\tau_2 - \tau_3) = 2$. Therefore, there are no such triples (τ_1, τ_2, τ_3). □

5. Proof of Theorem 1.1 for $g \geq 2$

It remains to prove Theorem 1.1 for $g \geq 2$. As mentioned in Section 1, we will reduce it to the case $g = 1$ by a degeneration argument.

Let E be a smooth elliptic curve. We first construct a smooth projective family X of surfaces over $\Delta = \mathbb{A}^1$ such that $X_0 \cong E \times \mathbb{P}^1$ and $X_t \cong \mathbb{P}\mathscr{E}$ for $t \neq 0$, where $\mathscr{E}$ is the rank 2 vector bundle on E given by a nonzero vector in $\operatorname{Ext}(\mathcal{O}_E, \mathcal{O}_E)$.

Let $\mathcal{V}$ be a rank 2 vector bundle over $E \times \Delta$ given by

$$t \in \operatorname{Ext}(\mathcal{O}_{E\times\Delta}, \mathcal{O}_{E\times\Delta}) = H^1(\mathcal{O}_{E\times\Delta}) = \mathbb{C}[t]$$

and let $X = \mathbb{P}\mathcal{V}$. Clearly, X is such a family.

There is an effective divisor $D \subset X$, flat over Δ, such that D_t is the section of X_t/E with $D_t^2 = 0$. Fix a point $p \in E$ and let $L = mD + \pi^* p$, where π is the projection $X \to E$.

For $t \neq 0$, the Severi variety $V_{X_t,L,g}$ has expected dimension g. If we fix g general points on X_t, there exist finitely many $[C] \in V_{X_t,L,g}$ such that C passes through these points. Let us fix g general sections $P_1, P_2, \ldots, P_g \subset X$ of X/Δ. Then after a base change, there exists a flat projective family $C \subset X$ of curves over Δ such that C_t is an integral curve in $|L|$ on X_t passing through $P_i \cap X_t$ for $i = 1, 2, \ldots, g$ and $t \neq 0$. Here we replace Δ by an analytic disk or a smooth affine curve finite over $\mathbb{A}^1$.

We may choose the base change in such a way that there exists a family of stable maps $\varphi : \mathscr{C} \to X$ over Δ such that φ maps $\mathscr{C}$ birationally onto C.

On X_0, the linear system $|L|$ is completely reducible in the sense that

$$H^0(\mathcal{O}_{X_0}(L)) = \operatorname{Sym}^m H^0(\mathcal{O}_{X_0}(D)) \otimes H^0(\mathcal{O}_{X_0}(\pi^* p)).$$

Therefore,

$$C_0 = m_1 D_1 + m_2 D_2 + \cdots + m_g D_g + F,$$

where D_i are the sections of X_0/E passing through $P_i \cap X_0$ for $i = 1, 2, \ldots, g$, F is the fiber of $\pi : X_0 \to E$ over p and the m_i are positive integers such that $\sum m_i = m$.

Clearly, C_t only has singularities in open neighborhoods of D_i. So it suffices to show that C_t has only nodes and ordinary triple points as singularities in an analytic neighborhood of each D_i for $i = 1, 2, \ldots, g$, if E is general.

Since $\mathscr{C}_t$ is a smooth projective curve of genus g for $t \neq 0$, there are exactly g irreducible components $\Gamma_1, \Gamma_2, \ldots, \Gamma_g$ of $\mathscr{C}_0$ such that each Γ_i is a smooth elliptic curve dominating D_i for $i = 1, 2, \ldots, g$.

Let us fix i. If $m_i = 1$, there is nothing to do. Otherwise, suppose that $m_i \geq 2$. Let $\psi : \widehat{X} \to X$ be the blowup of X along D_i. Then the central fiber $\widehat{X}_0 = S \cup R$ is a union of two smooth projective surfaces S and R, where S is the proper transform of X_0, R is the exceptional divisor of ψ and S and R meet transversely along a curve over D_i, which we still denote by D_i. Let $\widehat{C}$ be the proper transform of C under ψ.

The rational map $\psi^{-1} \circ \varphi : \mathscr{C} \dashrightarrow \widehat{X}$ is regular at a general point of Γ_i. We claim that

$$\psi^{-1} \circ \varphi(\Gamma_i) \not\subset D_i = S \cap R.$$

Otherwise, we choose a local section Q of $\mathscr{C}/\Delta$ passing through a general point of Γ_i. Then $\varphi(Q)$ is a local section of $\widehat{X}/\Delta$ meeting $D_i = S \cap R$, which is impossible since $\widehat{X}$ is smooth. So $\psi^{-1} \circ \varphi$ maps Γ_i to an irreducible curve on R other than D_i. That is, $\widehat{C}_0$ does not contain D_i.

We have either $R \cong \mathbb{P}\mathscr{E}$ or $R \cong E \times \mathbb{P}^1$.

A. If $R \cong \mathbb{P}\mathscr{E}$, then $\widehat{C} \cap R$ must be an integral curve in $|m_i \widehat{D} + \hat{\pi}^* p|$ of geometric genus 1, where $\widehat{D}$ is the proper transform of D and $\hat{\pi} = \pi \circ \psi$ is the projection $\widehat{X} \to E$. Then by Theorem 1.2, $\widehat{C} \cap R$ has only nodes and ordinary triple points as singularities and the same holds for C_t in an open neighborhood of D_i.

B. If $R \cong E \times \mathbb{P}^1$, then $\widehat{C} \cap R = m_i \widehat{D}_i + \widehat{F}$, where $\widehat{D}_i$ is the section R/E passing through the point $\widehat{P}_i \cap R$ with $\widehat{P}_i$ being the proper transform of P_i under ψ and $\widehat{F}$ is the fiber of R over $p \in E$. So we continue to blow up $\widehat{X}$ along $\widehat{D}_i$. By embedded resolution of singularities, there exists a sequence of blowups over D_i, say $f : X' \to X$, such that the proper transform C' of C is smooth over a general point of D_i. Then by Zariski's main theorem, the map $f^{-1} \circ \varphi : \mathscr{C} \dashrightarrow X'$ has connected fiber over $f^{-1}(D_i)$. This means that C'_0 is smooth over a general point of D_i. So we will eventually end up in case A after a sequence of blowups over D_i.

References

[AC81] E. Arbarello and M. Cornalba, "Footnotes to a paper of Beniamino Segre: 'On the modules of polygonal curves and on a complement to the Riemann existence theorem'", *Math. Ann.* **256**:3 (1981), 341–362. MR

[BL00] J. Bryan and N. C. Leung, "The enumerative geometry of $K3$ surfaces and modular forms", *J. Amer. Math. Soc.* **13**:2 (2000), 371–410. MR

[CC99] L. Chiantini and C. Ciliberto, "On the Severi varieties of surfaces in $\mathbf{P}^3$", *J. Algebraic Geom.* **8**:1 (1999), 67–83. MR

[CDGK23a] C. Ciliberto, T. Dedieu, C. Galati, and A. L. Knutsen, "Severi varieties on blow-ups of the symmetric square of an elliptic curve", *Math. Nachr.* **296**:2 (2023), 574–587. MR

[CDGK23b] C. Ciliberto, T. Dedieu, C. Galati, and A. L. Knutsen, "Nonemptiness of Severi varieties on Enriques surfaces", *Forum Math. Sigma* **11** (2023), art. id. e52, 32 pp. MR

[Che02] X. Chen, "A simple proof that rational curves on $K3$ are nodal", *Math. Ann.* **324**:1 (2002), 71–104. MR

[Che19] X. Chen, "Nodal curves on K3 surfaces", *New York J. Math.* **25** (2019), 168–173. MR

[DS17] T. Dedieu and E. Sernesi, "Equigeneric and equisingular families of curves on surfaces", *Publ. Mat.* **61**:1 (2017), 175–212. MR

[HM98] J. Harris and I. Morrison, *Moduli of curves*, Graduate Texts in Mathematics **187**, Springer, 1998. MR

[KL22] A. L. Knutsen and M. Lelli-Chiesa, "Genus two curves on abelian surfaces", *Ann. Sci. Éc. Norm. Supér.* (4) **55**:4 (2022), 905–918. MR

[KLM19] A. L. Knutsen, M. Lelli-Chiesa, and G. Mongardi, "Severi varieties and Brill–Noether theory of curves on abelian surfaces", *J. Reine Angew. Math.* **749** (2019), 161–200. MR

[Lew99] J. D. Lewis, *A survey of the Hodge conjecture*, 2nd ed., CRM Monograph Series **10**, Amer. Math. Soc., 1999. MR

[Ser23] E. Sernesi, "On Treibich–Verdier curves", *Pure Appl. Math. Q.* **20**:5 (2024), 2453–2468. MR

[Zah22] A. Zahariuc, "The Severi problem for abelian surfaces in the primitive case", *J. Math. Pures Appl.* (9) **158** (2022), 320–349. MR

Received December 31, 2024. Revised February 16, 2026.

XIAOTIAN CHANG
DIVISION OF MATHEMATICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE
xiaotian.chang@ntu.edu.sg

XI CHEN
DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES
UNIVERSITY OF ALBERTA
EDMONTON, AB
CANADA
xichen@math.ualberta.ca

ADRIAN ZAHARIUC
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WINDSOR
WINDSOR, ON
CANADA
adrian.zahariuc@uwindsor.ca

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.371

SPUN NORMAL SURFACES IN 3-MANIFOLDS II: THE TOROIDAL CASE

ENSIL KANG AND J. HYAM RUBINSTEIN

Spun normal surfaces are a useful way of representing proper essential surfaces using ideal triangulations for 3-manifolds with tori boundaries. Here we consider spinning surfaces in the case of a 3-manifold with a nontrivial JSJ decomposition, where each of the JSJ components is hyperbolic. We prove that a proper essential surface Σ can be spun, so long as none of the JSJ components are bundles with fiber a subsurface of Σ and the ideal triangulation satisfies similar properties to a taut structure.

1. Introduction

Definition 1. A *crushing* of a space X is an epimorphism $f : X \to Y$ that is cell-like, i.e., $f^{-1}(y)$ is a cell for each $y \in Y$.

It is well-known that such a map is homotopic to a homeomorphism; see [Sie]. We will be interested in crushings associated to triangulations and normal surfaces. In particular, if S is a closed 2-sided normal surface in a triangulation $\mathcal{T}$ we will be concerned with the process of cutting M open along S and then crushing each component of S to a point. If we delete the image points of S after crushing, we obtain a new triangulation which is ideal. We want to crush the cell structure of the cut open triangulation into a new triangulation. This crushing process is detailed in [JR1] in the case of 2-spheres and also will be described in Section 2 for our case of tori.

In [KR2], it is shown that a properly embedded nonfibered incompressible and ∂-incompressible 2-sided surface in M can be spin normalized along boundary tori of M, if M is a suitable 3-manifold with 1-efficient ideal triangulation.

Definition 2. An ideal triangulation $\mathfrak{I}$ of a 3-manifold M is 1-*efficient* if it satisfies the following conditions:

- There are no embedded normal 2-spheres, projective planes, or Klein bottles.
- Every embedded normal torus is peripheral.

The authors would like to thank Seokbeom Yoon for helpful conversations and Juhan Shin for help with diagrams. Kang was supported by research fund from Chosun University (2024). Rubinstein was partially supported by ARC grant DP19012363.
MSC2020: primary 57K35; secondary 57K32.
Keywords: 3-manifold, spun normal surface, toroidal case, ideal triangulation.

Equivalently, there are no normal surfaces with nonnegative Euler characteristic, except for peripheral tori and Klein bottles.

[KR2, **Theorem 8**]. *Let M be an anannular, atoroidal, irreducible and P^2-irreducible 3-manifold with tori boundary components and $\Im$ be a 1-efficient ideal triangulation of M. If F is a properly embedded, incompressible and ∂-incompressible 2-sided surface in M which is not a fiber, then F can be spin normalized in $(M, \Im)$ with 2^r choices of spinning direction, where r is the number of boundary components of M containing a curve of ∂F.*

We generalize this result by allowing JSJ-normal tori in M. To deal with such tori, we need to restrict how they can be realized as normal surfaces.

Definition 3. An ideal triangulation $\Im$ of an irreducible and P^2-irreducible 3-manifold M is *T-efficient* if it satisfies the following conditions:

- There are no embedded normal 2-spheres or projective planes.
- Every embedded normal torus or Klein bottle is essential, i.e., π_1-injective.
- For each isotopy class of essential tori or Klein bottles in M, there is exactly one embedded normal torus or Klein bottle representing that class.

Our main theorem is the following.

Theorem 6. *Suppose that M is compact, irreducible, and P^2-irreducible and has essential tori boundary components. Assume that each of the JSJ components M_i of M is hyperbolic for $1 \leq i \leq k$. Let Σ be any properly embedded 2-sided essential surface in M with the property that no subsurface of Σ is a fiber of a bundle structure for one of the JSJ components.*

Assume that M has an ideal triangulation $\Im$ which is T-efficient. Then Σ can be spin normalized with boundary curves spinning in all possible combinations of directions.

To prove Theorem 6, we carry out the crushing process along JSJ-normal tori T_i, cutting M open along those T_i and then crushing and deleting the boundaries to have an ideal triangulation for the interior of each JSJ component of M. This new ideal triangulation of each of the JSJ components is 1-efficient, so that we can apply the results of [KR2]. Also, in Section 5, we will show that there always exists an ideal triangulation that satisfies all the conditions described in Theorem 6, so it is reasonable to impose the above restrictions.

Another generalization would be to replace “1-efficiency” with “0-efficiency” for the ideal triangulation of the JSJ components we are considering. However, in this case, it is difficult to avoid inessential normal tori in spinning. Whether a properly embedded surface can spin along an inessential normal torus is not clear to us.

Spun normal surfaces and generalized spun normal surfaces play a central role in the 3d index of Dimofte, Gaiotto, and Gukov [DGG]. The relationships between spun and generalized spun normal surfaces and the 3d index are discussed in [GHRH]. To define the 3d index, a 1-efficient ideal triangulation is required. In [GHRS] it is shown that the 3d index is independent of the choice of triangulation, so long as a suitable class of triangulations are used, which come from the hyperbolic structure on the 3-manifold. It is conjectured that the 3d index counts surfaces in 3-manifolds without reference to a triangulation. Moreover, an important problem is to understand if the 3d index can be defined for toroidal manifolds. Exploring the behaviour of spun normal surfaces in toroidal manifolds may give some insight into this issue.

2. Crushing JSJ tori in an ideal triangulation

When a triangulation is cut open along a normal surface, we get a cell decomposition. The cells are either truncated tetrahedra, truncated (triangular) prisms or normal prisms (quadrilateral normal prisms and triangular normal prisms) with base a quadrilateral or triangular disk included in the normal surface. As the normal surface consists of quadrilateral and triangular disks, when these are crushed, we need to extend the crushing to these cells to produce a new triangulation. In this section, we will discuss such crushing. A detailed background and discussion is given in [JR1].

Most of the arguments in this section will be developed under the conditions given in the assumptions of the main theorem (Theorem 6) in Section 1. Although there are some arguments that hold true in more general situations, this article will focus on normal spinning a properly embedded incompressible and ∂-incompressible surface, rather than on crushing a more general triangulation.

Let M be a compact, irreducible, and P^2-irreducible 3-manifold with essential tori boundary components, and $\Im$ be an ideal T-efficient triangulation of $\mathring{M}$. Assume that M is toroidal and each of the JSJ components M_i of M is hyperbolic, for $1 \leq i \leq k$. We ultimately want to find a 1-efficient ideal triangulation for each JSJ component M_i. To make further arguments easier, we will truncate $(\mathring{M}, \Im)$ by removing an open regular neighborhood of ideal vertices whose boundaries are peripheral normal tori, and denote the resulting manifold by $\hat{M}$.

Let S be the union of all the JSJ tori that are in (unique) normal form for $(\mathring{M}, \Im)$ and all the boundary components of $\hat{M}$. We assume that each JSJ component $\hat{M}_i$ is obtained by cutting $\hat{M}$ open along the JSJ-normal tori in S. Then all the boundary components of $\hat{M}_i$ are in normal form. Since we truncated $(\mathring{M}, \Im)$, $\hat{M}_i$ has no ideal vertices and has the induced cell decomposition C_i from the ideal triangulation $\Im$, which consists of four types of cells: truncated tetrahedra, truncated

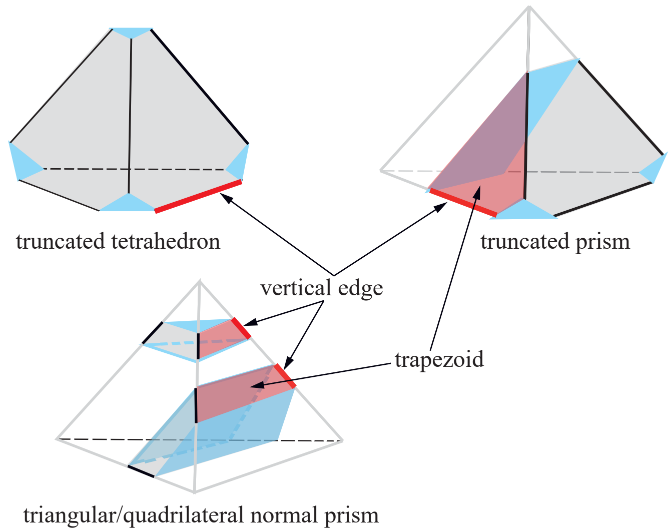


Figure 1. Four types of cells in C_i.

prisms, triangular normal prisms, and quadrilateral normal prisms (see Figure 1).

We will define a nice crushing $f_i : \hat{M}_i \to \overline{M}_i$, in Theorem 1, which satisfies the following conditions; f_i maps each component of $\partial\hat{M}_i$ to a point, $\overline{M}_i^*$ obtained by deleting vertices from $\overline{M}_i$ is homeomorphic to $\mathring{M}_i$, and the induced cell decomposition of $\overline{M}_i$ is an ideal triangulation of $\mathring{M}_i$ after deleting the vertices. We will show that the new induced ideal triangulation of $\overline{M}_i^*$, or equivalently $\mathring{M}_i$, is 1-efficient (Theorem 5).

Let us define a crushing $f_i : \hat{M}_i \to \overline{M}_i$ by collapsing each cell of C_i as shown in Figure 2, and call it a *canonical crushing* on $\hat{M}_i$.

We will investigate if f_i is cell-like (see Figures 2 and 3). We denote the induced cell decomposition of $\overline{M}_i$ by $\Im_i$. Note that each boundary component of $\hat{M}_i$ crushes to a vertex of $\Im_i$. But there is no guarantee that the new manifold $\overline{M}_i^*$ given by deleting vertices from $\overline{M}_i$ is homeomorphic to $\mathring{M}_i$, or that the induced cell structure of $\overline{M}_i^*$ is a well-defined ideal triangulation. If f_i is cell-like, the first assertion follows from [Sie]. We will find that this is the case under our assumptions.

To investigate this, we need to look at the cycles of truncated prisms or trapezoids. Here a *trapezoid* is a rectangular face of C_i bounded by two parallel (i.e., normal isotopic) normal arcs and two vertical edges, where a *vertical edge* is an edge of C_i whose interior is inside $\hat{M}_i$ and both of whose boundary points are in $\partial\hat{M}_i$, i.e., in S (see Figure 1). A *trapezoidal cycle* is an annulus or a Möbius band properly embedded in M, formed by gluing trapezoids together along their vertical edges.

Note that a truncated prism crushes to a triangle and a trapezoid to an edge by the canonical crushing f_i. To achieve that $\overline{M}_i^*$ is homeomorphic to $\mathring{M}_i$, we will verify

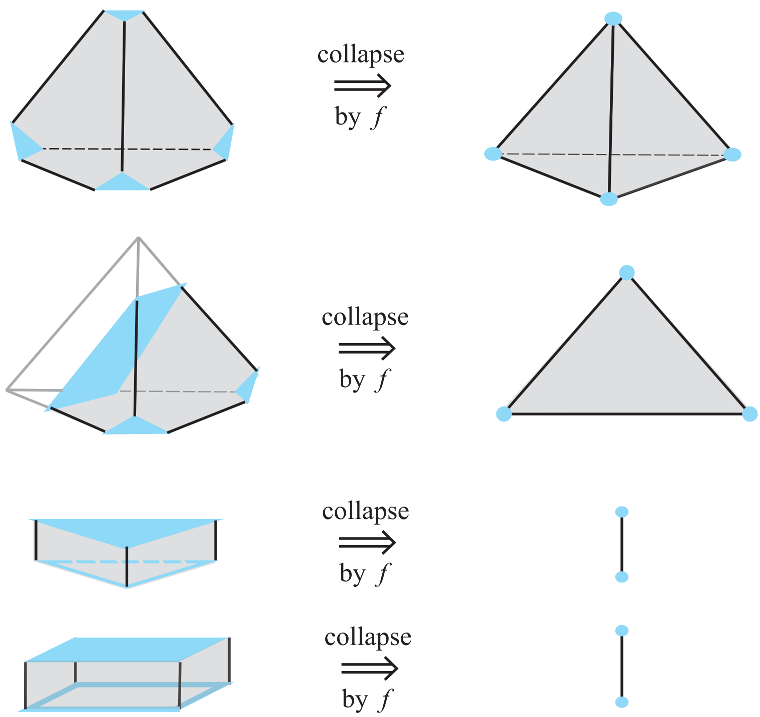


Figure 2. Canonical crushing $f = f_i$.

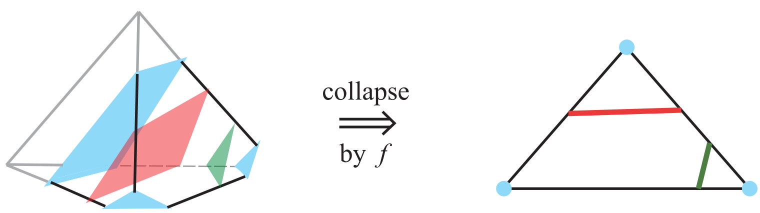


Figure 3. Crushings of quadrilateral and triangular disks by $f = f_i$.

that the regions surrounded by the cycles of trapezoids are simply connected, so that such regions can be crushed into the edges. By modifying f_i in this way, we are able to achieve that the modified f_i is cell-like and so the topology of $\hat{M}_i$ is that of M_i (Lemma 2). We also need to verify that there are no cycles of truncated prisms (Lemma 3) which collapse to a triangle by f_i, to finish the verification that the modified f_i is cell-like and that the induced cell structure of $\overline{M}_i^*$ is a well-defined ideal triangulation. This is the main topic of discussion in this section.

Let $P(C_i)$ be the collection of all cells which crush to edges of $\mathfrak{T}_i$ by the canonical crushing f_i. Then $P(C_i) = \{$all normal prisms of $C_i\} \cup \{$all trapezoids of $C_i\} \cup \{$all vertical edges of truncated tetrahedra and truncated prisms of $C_i\}$. We call $P(C_i)$

the *combinatorial product region* of C_i. Each component of $P(C_i)$ crushes to an edge of $\Im_i$ by the canonical crushing $f_i : \hat{M}_i \to \overline{M}_i$, and for $\overline{M}_i^*$ to be topologically identical to $\mathring{M}_i$, the inverse image of an edge must be simply connected. If each component of $P(C_i)$ is simply connected, the canonical crushing f_i is what we need for spinning. If not, we will modify f_i by further crushing as follows: each region enclosed by the cycles of trapezoids (together with normal disks in $\partial \hat{M}_i$) crushes into an edge (Figure 4, right, shows an example of such a region — in this case, a truncated prism that is crushed to an edge under modified crushing), so that the enlarged crushing region is simply connected. We denote the union of the new enlarged regions by $P(\hat{M}_i)$ and call it the *induced product region*, and the modified crushing is called a *combinatorial crushing*. Under suitable conditions, the desired $P(\hat{M}_i)$ can be constructed with properties required to achieve spinning. In our case (assuming the hypotheses of the main theorem (Theorem 6), $P(C_i)$ has simply connected components so that $P(C_i) = P(\hat{M}_i)$ (Lemma 4) and furthermore the canonical crushing f_i itself is a combinatorial crushing and does not change the topology of the original manifold. Under what general circumstances $P(C_i)$ is simply connected, and under what conditions $P(\hat{M}_i)$ exists, is a subject for further study.

The following theorem states that the canonical crushing in our case induces a suitable cell decomposition of the resulting manifold.

Theorem 1. *Let M be compact, irreducible, and P^2-irreducible with essential tori boundary components. Assume that each of the JSJ components M_i of M is hyperbolic for $1 \le i \le k$ and the ideal triangulation $\Im$ of $\mathring{M}$ is T-efficient. Then for each JSJ component M_i, $1 \le i \le k$, the canonical crushing $f_i : \hat{M}_i \to \overline{M}_i$ gives that $\mathring{M}_i$ is homeomorphic to $\overline{M}_i^*$, with vertices deleted from $\overline{M}_i$. Moreover the induced cell decomposition $\Im_i^*$ of $\overline{M}_i^*$ is an ideal triangulation of $\mathring{M}_i$.*

We will prove this by using Lemmas 2, 3 and 4 below and [JR1, Theorem 4.1].

Lemma 2. *With the same hypotheses as Theorem 1, any trapezoidal cycle in $\hat{M}_i$ is an annulus and together with two disks in S bounds a 3-cell, where $\hat{M}_i$ is a truncated JSJ component obtained by cutting $\hat{M}$ open along the normal representatives of the JSJ tori.*

Note. The hypothesis of “unique normal representative of each JSJ torus” in Theorem 1 could be replaced by “least-weight normal representative” in this lemma.

Proof. Let T be a trapezoidal cycle in $\hat{M}_i$. Then ∂T is on S, which is the union of all normal JSJ tori of M and the normal boundary components of $\hat{M}$ and T is a properly embedded annulus or Möbius band in $\hat{M}_i$. By the assumptions of Theorem 1, $\hat{M}_i$ is hyperbolic and so there are no properly embedded essential surfaces with nonnegative Euler characteristic. If T is an inessential Möbius band, then T is either ∂-parallel or has a boundary compression to S. But since $\hat{M}_i$ is hyperbolic and has tori boundary, neither can occur. Therefore T is an inessential annulus with both boundary curves on S. Here since S is incompressible, either both boundary curves of T are essential in S or both are not.

If both curves of ∂T are essential, then T must be boundary parallel, since $\hat{M}_i$ is hyperbolic. Let S_1 be the component of S containing ∂T and A the annulus in S_1 which is parallel to T. Note that T is 0-weight (meaning that its interior does not intersect any edges of the triangulation). The new torus $T \cup (S_1 \setminus A)$ is a barrier (see [JR1]) so normalizing produces a topologically boundary parallel normal torus with less weight than S_1. This contradicts the fact that there is a unique normal representation of each JSJ torus in S. Therefore the only possibility is that the curves of ∂T are inessential and each bounds a disk in S_1, say D_1 and D_2 respectively. If say $D_1 \subset D_2$ then T must be ∂-parallel (to the annulus $D_2 \setminus \mathring{D}_1$). But then we get a contradiction to the unique normal representation of JSJ tori again by a similar argument to the previous paragraph. Hence $D_1 \cap D_2 = \varnothing$ and D_1 and D_2 together with T form a 2-sphere. Therefore T together with two disks in S bounds a 3-cell in $\hat{M}_i$ since M is irreducible, and the proof is complete. □

The following remark comes from the proofs of Lemmas 3 and 2.

Remark 1. (1) No 0-weight annulus properly embedded in $\hat{M}_i$ can be ∂-parallel to an annulus with positive weight in S.

(2) Any non-∂-parallel 0-weight annulus properly embedded in $\hat{M}_i$ has inessential ∂-curves in S and together with two disks in S bounds a 3-cell in $\hat{M}_i$.

Since any trapezoidal cycle together with two disks in S bounds a 3-cell by Lemma 2, we can define $P(\hat{M}_i)$ by adding to $P(C_i)$ these 3-cells enclosed by trapezoidal cycles of $P(C_i)$. Note that $P(\hat{M}_i)$ is obtained by gluing a collection of project regions of the form $D \times [0, 1]$, where $D \times 0$ and $D \times 1$ are disks in S, to $P(C_i)$, and so the components of $P(\hat{M}_i)$ are simply connected.

The following lemma says that the components of $P(C_i)$ itself are simply connected with our hypotheses and so $P(C_i) = P(\hat{M}_i)$ and the canonical crushing f_i is a combinatorial crushing. This plays an important role in proving Theorem 1, and also in Theorem 5, which asserts the 1-efficiency of the induced ideal triangulation $\mathfrak{I}_i^*$ of $\mathring{M}_i$.

Lemma 3. *With the same hypotheses as Theorem 1, there is no cycle of truncated prisms in C_i, the induced cell decomposition of $\hat{M}_i$.*

Proof. We mostly follow the ideas of Theorem 5.5 in [JR1]. But in the details, the argument is simpler because of our strong hypotheses.

Let X be a cycle of truncated prisms of C_i. We can exclude the case that the cycle X is formed along a single edge, i.e., the hexagonal faces of all truncated prisms of X are glued along a single edge. For in that case, there is a properly embedded

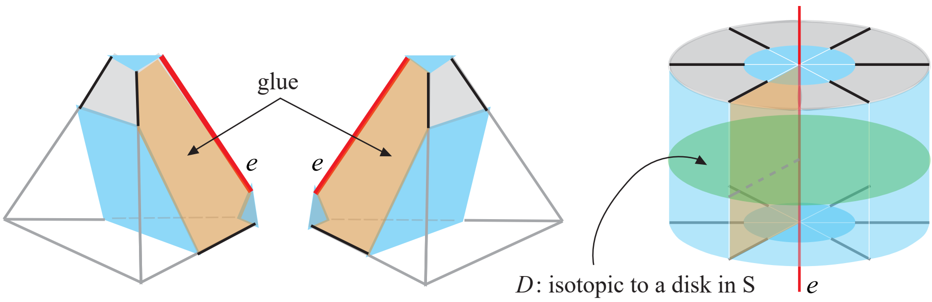


Figure 4. A cycle of truncated prisms glued along a single edge.

disk D in M_i which is isotopic to a disk in S (due to the incompressibility of S) and meeting edges of C_i in precisely one point (see Figure 4). By replacing the disk in S with D and normalizing, we obtain a new normal representative of a JSJ torus, contradicting the uniqueness assumption.

Hence we may assume that X is a cycle about more than one edge and so is a solid torus with boundary identifications along some trapezoids (if any). Figure 5 describes X without any identifications along trapezoids.

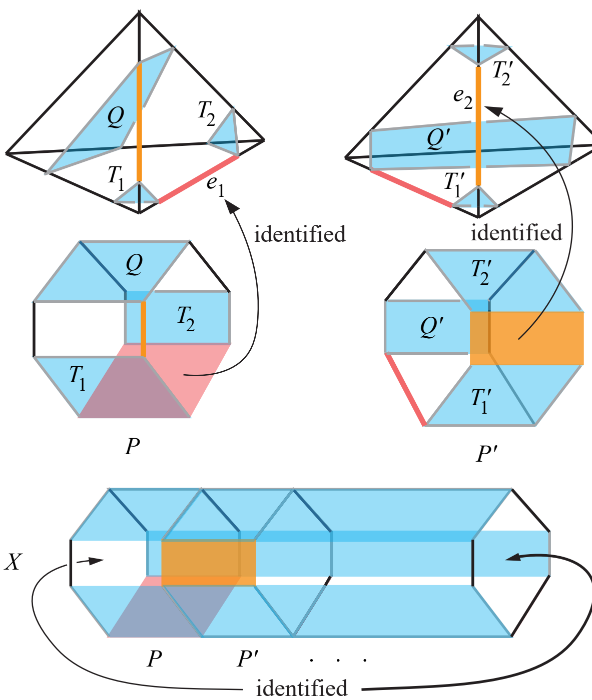


Figure 5. A cycle of truncated prisms without any trapezoidal identification.

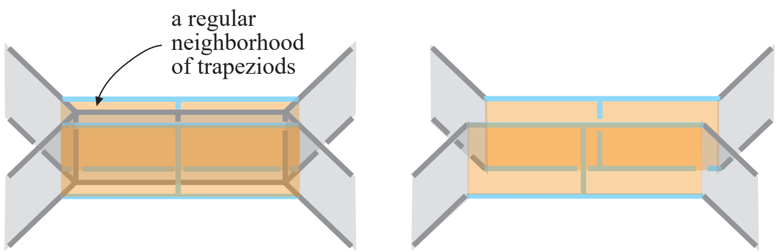


Figure 6. Truncating prisms along a trapezoidal identification.

The first case is of six strips in ∂X, three from S and three from trapezoidal cycles, and the second case is of two strips in ∂X, one from S and the other from trapezoidal cycles. If there are identifications along some trapezoids, we will truncate an open regular neighborhood of the trapezoidal cycles at X (see Figure 6) so that the new cycle of truncated prisms, again denoted by X, is a solid torus without any boundary identifications. In this case, the boundary of the solid torus is covered by either six strips (three from S and three 0-weight annuli parallel to the original trapezoidal cycles in ∂X) or two strips (one from S and one 0-weight annulus parallel to the original trapezoidal cycles in ∂X).

Case 1. Assume that the boundary of X has three strips, say A_i $(i = 1, 2, 3)$, from S and three 0-weight strips, say T_i $(i = 1, 2, 3)$, from trapezoidal cycles. Figure 7 describes this case.

By Lemma 2 and Remark 1, T_i with two disks D_i and D_i' in S bounds a 3-cell Δ_i, and each boundary curve of T_i (and so of A_i) bounds a disk in S. Fix an annulus A_i, say A_1, which has boundary curves, say C_1 and C_2, and let D_1 and D_2 be disks in S bounded by C_1 and C_2 respectively. Since D_1 and D_2 are glued along ∂A_1,

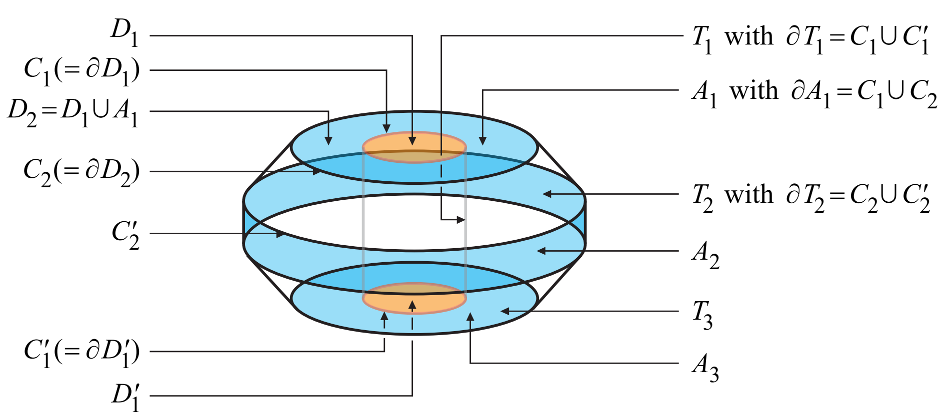


Figure 7. A cycle X of truncated prisms with six strips in its boundary.

these are all contained in the same component of S, and so either $D_2 = D_1 \cup A_1$ or $D_1 = D_2 \cup A_1$, say $D_2 = D_1 \cup A_1$ which means that $A_1 \subset D_2$. (Otherwise, $D_1 \cup D_2 \cup A_1$ is a 2-sphere.)

On the other hand, two 0-weight annuli, say T_1 and T_2, from trapezoidal cycles are adjacent to A_1 along the boundary curves C_1 and C_2. Let C_1' and C_2' be the remaining boundary curves of T_1 and T_2. By Remark 1, C_1' and C_2' also bound disks in S, say D_1' and D_2', and $T_1 \cup D_1 \cup D_1'$ and $T_2 \cup D_2 \cup D_2'$ bound 3-cells Δ_1 and Δ_2. Then the entirety of the cycle X is contained in Δ_2 and so in $P(\hat{M}_i)$ since $A_1 \subset D_2 \subset \Delta_2$ and so truncated prisms along A_1 are contained in Δ_2. Furthermore $\Delta_1 \subset \Delta_2$ (because $\partial \Delta_1 \subset \Delta_2$) and $A_2, A_3 \subset D_2'$ (because $\partial \Delta_2 = D_2 \cup T_2 \cup D_2'$ and $A_2, A_3 \subset \partial \Delta_2$).

Therefore, the remaining trapezoidal cycle T_3 adjacent to A_2 and A_3 must be parallel to an annulus in D_2'. This contradicts Remark 1(1).

Case 2. Assume that the boundary of the cycle X has two strips: one from S, say A, and the other from a trapezoidal cycle, say T. Let C_1 and C_2 be the curves in ∂X bounding A and T. Since T is a 0-weight annulus properly embedded in $\hat{M}_i$, by Remark 1(2), T with two disks D_1 and D_2 in S bounds a 3-cell Δ, where $\partial D_1 = C_1$ and $\partial D_2 = C_2$. Then $A \cup D_1 \cup D_2$ is a 2-sphere in S, which gives a contradiction. □

Lemma 4. *With the same hypotheses as Theorem 1, there are no truncated prisms or truncated tetrahedra in $P(\hat{M}_i)$, i.e., $P(\hat{M}_i) = P(C_i)$ and so the components of $P(C_i)$ are simply connected.*

Proof. Suppose that there is a truncated prism or truncated tetrahedron Δ in $P(\hat{M}_i)$. The closure of each component of $P(\hat{M}_i) \setminus P(C_i)$ is a product $D \times [0, 1]$, where $D \times 0$ and $D \times 1$ are disks in S. Since Δ is contained in such a closure of a component of $P(\hat{M}_i) \setminus P(C_i)$, at least two normal disks in $\partial \Delta$ belong to the same $D \times \epsilon$ ($\epsilon =$ either 0 or 1) and so the vertical arc α connecting these two normal disks is isotopic to an arc β of S along a disk in $D \times [0, 1] \subset P(\hat{M}_i)$ (see Figure 8).

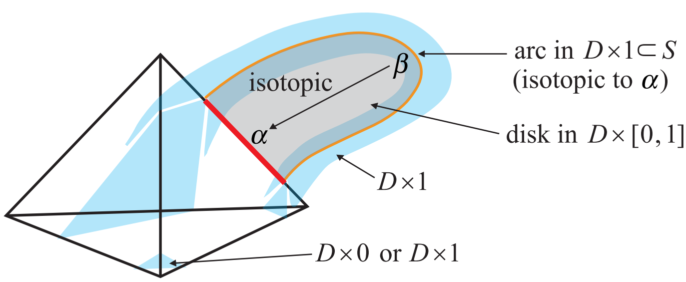


Figure 8. A vertical edge isotopic to an arc in S.

Now we can isotope S so that it doesn't meet the vertical edge and obtain a new normal representative of S which contradicts the uniqueness of normal representatives of each JSJ torus. (The new normal representation here may not be least-weight.)

This completes the claim that $P(\hat{M}_i)$ has no truncated prisms or tetrahedra and so $P(\hat{M}_i) = P(C_i)$. Thus the components of $P(C_i)$ are simply connected. □

Remark 2. From the arguments of the above lemmas, we can say that Lemma 2 (as already observed) and Lemma 3 hold even if the "unique normal representation of each JSJ torus" hypothesis is replaced by "unique least-weight normal representation". On the other hand, the proof of Lemma 4 required the assumption that there is a unique normal representation of each JSJ torus.

Now we are ready to prove that $\overline{M}_i^*$ is homeomorphic to $\mathring{M}_i$ and that $\Im_i^*$ is an ideal triangulation of $\mathring{M}_i$. In the general case of $P(\hat{M}_i) \neq P(C_i)$, an additional crushing is required and the 3 cells enclosed by the trapezoidal cycles of $P(\hat{M}_i)$ are further crushed to the edges. This modified crushing $\bar{f}_i : \hat{M}_i \to \overline{M}_i$ will be a combinatorial crushing. But in our situation described in Theorem 1, since the components of $P(C_i)$ are simply connected, no further crushing to edges is required and the canonical crushing will be a combinatorial crushing.

The following theorem from [JR1] provides sufficient conditions for $P(\hat{M}_i)$ to ensure that the manifold $\overline{M}_i^*$ obtained by a combinatorial crushing map is homeomorphic to $\mathring{M}_i$ and that the induced cell-structure $\Im_i^*$ is an ideal triangulation of $\mathring{M}_i$. In our case of Theorem 1, it is enough to show that $P(\hat{M}_i)$ satisfies the three conditions given by the theorem.

[JR1, **Theorem 4.1**]. *Suppose $\Im$ is a triangulation of a closed, orientable 3-manifold or an ideal triangulation of the interior of a compact, orientable 3-manifold M. Suppose S is a normal surface embedded in M, X is the closure of a component of the complement of S and X does not contain any vertices of $\Im$. Let $P(X)$ be the induced product region for X. Suppose the following conditions are met*:

- *$X \neq P(X)$.*
- *$P(X)$ is a trivial product region for X, i.e., it has simply connected components.*
- *There are no cycles of truncated prisms in X that are not in $P(X)$.*

Then the triangulation $\Im$ can be crushed along S into an ideal triangulation $\Im^*$ of $\mathring{X}$.

Proof of Theorem 1. We will show that the three properties described in Theorem 4.1 of [JR1] hold for $X = \hat{M}_i$.

Since $P(\hat{M}_i) = P(C_i)$ and $P(C_i)$ itself has simply connected components, we can work with the canonical crushing map $f_i : \hat{M}_i \to \overline{M}_i$, which crushes each cell of C_i, as shown in Figure 2. Here C_i is the cut-open cell decomposition of $\hat{M}_i$ induced from $\Im$.

- If $\hat{M}_i = P(\hat{M}_i)$, by Lemma 4, $\hat{M}_i$ is simply connected. But this contradicts that M_i is a JSJ component.
- Since $P(C_i)$ has simply connected components, $P(\hat{M}_i)$ $(= P(C_i))$ is a trivial product region.
- There are no cycles of truncated prisms outside $P(\hat{M}_i)$ by Lemma 3. □

Remark 3. (1) In [JR1], the induced product region $P(M_i)$ is defined in a different way; each component $D \times [0, 1]$ of $P(M_i)$ does not have to be simply connected, but the inclusion homomorphism $\pi_1(D \times \epsilon)$ into $\pi_1(S)$ must be injective for $\epsilon = 0, 1$. But in our case, we are able to satisfy the [JR1] requirements above and deduce Theorem 1. Therefore, Theorem 4.1 of [JR1] is also valid in our case.

(2) In [JR2] and [BJR], any crushing satisfying the three conditions above is called a "*combinatorial crushing*". In our case, the canonical crushing f_i satisfies those conditions, so it is also a combinatorial crushing in the sense of [JR2] and [BJR].

Now we are ready to prove that the constructed ideal triangulation $\Im_i^*$ of $\overline{M}_i^*$ is 1-efficient.

Theorem 5. *Let M be compact, irreducible, and P^2-irreducible with essential tori boundary components. Assume that the JSJ components M_i of M are hyperbolic for $1 \leq i \leq k$. If the ideal triangulation $\Im$ of $\mathring{M}$ is T-efficient, then for each JSJ component M_i, $1 \leq i \leq k$, the ideal triangulation $\Im_i^*$ of $\mathring{M}_i$ induced by the canonical crushing $f_i : \hat{M}_i \to \overline{M}_i$ is 1-efficient.*

Proof. By Theorem 1, the cell decomposition $\Im_i^*$ induced by the canonical crushing $f_i : \hat{M}_i \to \overline{M}_i$ is an ideal triangulation of $\mathring{M}_i$. To prove the 1-efficiency of $(\mathring{M}_i, \Im_i^*)$ (which is hyperbolic), we need to show that there is no normal 2-sphere or non-vertex-linking normal torus in $(\overline{M}_i^*, \Im_i^*)$. Suppose that there is one, denoted by T^*. It is enough to show that the inverse image of T^* by f_i, denoted by T, is a normal surface homeomorphic to T^* in (M_i, C_i) and so in $(\mathring{M}, \Im)$.

(i) If T is a 2-sphere then 0-efficiency is contradicted.

(ii) If T is a torus then, by uniqueness of normal tori, T must be vertex-linking in $\Im$ or boundary-linking normally parallel to a boundary component of M_i that crushes onto a vertex-linking normal torus in $(\overline{M}_i^*, \Im_i^*)$. This contradicts that T^* is not vertex-linking.

We will follow the proof ideas of Lemma 3.4 and Theorem 3.5 in [BJR]. (In [BJR], the cell decomposition of $\hat{M}_i$ must be an inflation of $\Im_i^*$ which is a minimal-vertex triangulation of $\hat{M}_i$ (in the sense of [BJR]) crushing to $\Im_i^*$. So direct application of the proof technique is problematic. However, the proof ideas are valid in our case using the cell decomposition C_i of $\hat{M}_i$.)

Note that the canonical crushing $f_i : \hat{M}_i \to \overline{M}_i$ is a combinatorial crushing and T crushes onto T^* by f_i. Since f_i is a combinatorial crushing and $P(M_i) = P(C_i)$, there is a 1-1 correspondence between the tetrahedra of $\Im_i^*$ and the truncated tetrahedra of C_i, and a 1-1 correspondence between the normal disks in a tetrahedron of $\Im_i^*$ and the normal disks in the corresponding truncated tetrahedron of C_i. Therefore, when comparing T and T^*, we need only consider the inverse image of T^* intersecting with 2-simplices or 1-simplices of $\Im_i^*$.

We first show that T and T^* are homeomorphic. In the proof of [BJR, Lemma 3.4], which uses the approximation theorem in [Sie], it is only necessary to prove that the canonical crushing f_i gives a proper cell-like map from T to T^*, i.e., for each point y of T^* the inverse image $f_i^{-1}(y)$ is compact and contractible. Let y be a point of T^* contained in a tetrahedron Δ of $\Im_i^*$. If y is placed in the interior of Δ, the inverse image $f_i^{-1}(y)$ is a point in a truncated tetrahedron of C_i. Assume that y is in the interior of a face of Δ. The inverse image of a face of a tetrahedron via f_i is either a face between two truncated tetrahedra or a chain of truncated prisms that is not a cycle. So the inverse image $f_i^{-1}(y)$ in this case is either a point or a long arc inside the chain of truncated prisms. Finally, assume that y is in an edge of Δ. The inverse image of an edge of $\Im_i^*$ by f_i is a product component $D \times [0, 1]$ of $P(C_i)$, so the inverse image $f_i^{-1}(y)$ is a horizontal cross section $D \times t$ which is simply connected. Therefore, $f_i^{-1}(y)$ is contractible for all cases and by the approximation theorem in [Sie], T and T^* are homeomorphic. (For the crushing f_i, see Figures 2 and 3.)

Next, we will prove that T, the inverse image of T^* by f_i, is a normal surface. We will prove that the inverse image of each normal disk, which is a quadrilateral or a triangle, must be in normal form, i.e., a union of normal disks. Let A^* be a normal disk of T^*. Since there is a 1-1 correspondence between the tetrahedra of $\Im_i^*$ and the truncated tetrahedra of C_i, there must be a unique normal disk A, which crushes onto A^*, in a truncated tetrahedron of C_i.

We need now consider inverse images of edges and vertices of A^*. Let α be an edge of A^*; then α is an arc inside a face of a tetrahedron in $\Im_i^*$. As we described above, the inverse image of a face of a tetrahedron via f_i is either a face between two truncated tetrahedra or a chain of truncated prisms that is not a cycle. Hence the inverse image of α must be an arc in a face of truncated tetrahedra (so it is an edge of A in this case) or a strip inside the chain and further normally parallel to a horizontal boundary strip of the chain (so in this case it is a union of normal disks parallel to normal disks in the boundary of truncated prisms). Now we will look at the inverse image of a vertex v of A^*. Since v is a point in an edge of $\Im_i^*$, the inverse image of v is a horizontal cross section of a product component $D \times [0, 1]$ of $P(C_i)$, so it is either a vertex of A or a disk that is a union of normal disks parallel to those in D. This completes the proof. □

Remark 4. From the reasoning in this proof, we can derive the following results.

(1) The inverse image under the combinatorial crushing f_i of a normal surface T in $(\overline{M}_i^*, \Im_i^*)$, or equivalently $(\mathring{M}_i, \Im_i^*)$, is a normal surface T^* in $(\hat{M}_i, C_i)$, and further T and T^* are homeomorphic.

(2) The inverse image under the combinatorial crushing f_i of a spun normal surface in $(\overline{M}_i^*, \Im_i^*)$ is a surface that spins in normal form along some JSJ tori and some boundary tori of M. We call such a surface a *pseudo-spun normal surface*. Such a surface can spin along nonperipheral essential normal tori embedded in the interior of M, and thus may contain infinitely many quadrilateral normal disks. However there are only finitely many quadrilateral normal disks which are not contained in the infinite annuli spinning around nonperipheral essential normal tori.

(3) The combinatorial crushing $f_i : \hat{M}_i \to \overline{M}_i$ is a proper cell-like map, i.e., for each point y of $\overline{M}_i$, the inverse image $f_i^{-1}(y)$ is compact and contractible.

3. Review of the construction of a spun normal surface in the case of 1-efficient ideal triangulations

The way that normal surfaces with nonnegative Euler characteristic can be an obstruction in normal spinning of a properly embedded incompressible and ∂-incompressible surfaces has been described in [KR2]. We will briefly review the construction of a spun normal surface in the absence of normal surfaces with nonnegative Euler characteristic.

Definition 4. Let M be a 3-manifold with an ideal triangulation $\Im$ and let $\hat{M}$ be the compact 3-manifold obtained by deleting regular neighborhoods of ideal vertices of $\Im$ from M with the induced truncated triangulation $\hat{\Im}$.

(1) A normal surface embedded in M is a closed surface that intersects each tetrahedron in finitely many elementary disks.

(2) A spun normal surface S in M is an embedded surface formed from elementary disks in the tetrahedra satisfying the following conditions:

- S consists of finitely many quadrilaterals and infinitely many triangular disks.
- S has a finite collection of disjoint infinite cylinders that spiral around ideal vertices.
- Each such cylinder consists of an infinite subset of the triangular disks in S.
- The remainder of S outside these cylinders is compact and consists of finitely many quadrilateral and triangular disks.

(3) Suppose that F is a properly embedded incompressible and ∂-incompressible surface in $\hat{M}$. We say that F spin normalizes (or normally spins) if there is a spun normal surface S in M such that $S \cap \hat{M}$ is isotopic to F.

[KR2, **Theorem 8**]. *Let M be an anannular, atoroidal, irreducible and P^2-irreducible 3-manifold with tori boundary components and $\Im$ be a 1-efficient ideal triangulation of M. If F is a properly embedded, incompressible and ∂-incompressible 2-sided surface in M which is not a fiber, then F can be spin normalized in $(M, \Im)$ with 2^r choices of spinning direction, where r is the number of boundary components of M containing a curve of ∂F.*

According to this theorem, if the manifold M we are considering is equipped with a 1-efficient ideal triangulation then we can always normally spin a properly embedded 2-sided essential surface F if F is not a fiber. To prove this, we truncated M at an open regular neighborhood of the ideal vertices, to work on a compact manifold with a truncated triangulation, and applied normal surface theory adapted to the truncated triangulation.

Here is a rough sketch of the proof of [KR2, Theorem 8]. Each step will be discussed in some detail later.

Step 1. Find a sequence $< F_k >$ of topological spinnings of F, which contains an infinite number of isotopy classes with a fixed boundary.

Step 2. Normalize each F_k to obtain a sequence $\langle \hat{F}_k \rangle$ of normal surfaces, which also has an infinite number of normal isotopy classes.

Step 3. Find a fixed common core surface S of an infinite number of normal surfaces $\hat{F}_k$, which means that the $\hat{F}_k$ differ only by annuli (with only triangular normal disks) parallel to boundary tori or Klein bottles of M.

Remark 5. (1) A sequence of "topological spinnings" in Step 1 requires that there are infinitely many different isotopy classes in the sequence. These are compact properly embedded surfaces with the same boundary curves and isotopies must keep the boundaries fixed.

(2) The statement of Step 3 means that F is spin normalized. The reason is that if a surface $\hat{F}_k$ spins once along a boundary torus or Klein bottle then we can attach an infinite normal annulus along the surface without introducing any new folds, and obtain a spun normal surface where the core surface S is isotopic to F.

(3) Step 3 is the only place that requires 1-efficiency in the proof of the theorem.

(4) All steps are performed on the truncated manifold and triangulation $(\hat{M}, \hat{\Im})$.

To apply this process to the proof of our main theorem in the next section, we need additional information on how to construct the surfaces mentioned in each step. Without loss of generality, we can assume that F is properly embedded in the truncated manifold $(\hat{M}, \hat{\Im})$ and its boundary curves are normal curves on the boundary of $\hat{M}$ and have the least intersection with the 1-simplices of $\partial \hat{M}$ in their isotopy classes.

Step 1: *Constructing F_k, the k-th spinning of F.* The *k-th spinning* F_k of F is a surface obtained by spinning F k times for each boundary curve, along the boundary tori or Klein bottles of M. That is, we attach a boundary parallel annulus of length k along each boundary curve of F so that the boundary slope of F_k is exactly the same as that of F.

Step 2: *Normalizing F_k with the boundary fixed.* Here, we perform the normalizing process on the truncated triangulation of M. Since F, and so F_k, are essential, we can always normalize F_k keeping the boundary fixed, so that the boundary curve of the resulting normal surface $\hat{F}_k$ is identical to the boundary curve of F and $\hat{F}_k$ is in normal form.

Step 3: *Finding a core surface S.* There is a finite number of fundamental surfaces from which any closed normal surface can be represented by geometric sums. Since the original ideal triangulation is 1-efficient, there are no closed normal surfaces with nonnegative Euler characteristic, so all fundamental surfaces except the peripheral tori or Klein bottles have negative Euler characteristic. Therefore, for a fixed Euler characteristic $\chi(F) = \chi(\hat{F}_k)$, only a finite number of possibilities for A and C arise when we write $\hat{F}_k = A + B + C$, where A is a proper normal surface with the same boundary as $\hat{F}_k$, B is a multiple of a peripheral normal surface and C is a closed non-peripheral normal surface. (Here we assume that A cannot be written as a sum $A = A' + B$.) This allows us to choose an infinite number of $\hat{F}_k$'s with a fixed core, say S:

$$\hat{F}_k = S + \sum_{j=1}^{r} l_{k_i,j} T_j$$

where the $l_{k,j}$ are nonnegative integers, and the T_j are boundary components of $\hat{M}$ containing some curves of ∂F. Furthermore, the boundary slope of S is the same as the boundary slope of F.

4. Spinning for suitable ideal triangulations in the toroidal case

Now we prove the main theorem, which guarantees the existence of a spun normal surface for the case with only hyperbolic JSJ components. As mentioned in the introduction, we first crush the JSJ tori to points by an appropriate cell-like crushing map, yielding a 1-efficient ideal triangulation of each JSJ component, and then apply the result of the 1-efficient case in [KR2].

Theorem 6. *Suppose that M is compact, irreducible and P^2-irreducible and has essential tori boundary components. Assume that each of the JSJ components M_i of M is hyperbolic, for $1 \le i \le k$. Let Σ be any properly embedded 2-sided essential surface in M with the property that no subsurface of Σ is a fiber of a bundle structure for one of the JSJ components.*

Assume that M has a T-efficient ideal triangulation $\mathcal{T}$. Then Σ can be spin normalized with boundary curves spinning in all possible combinations of directions.

Proof. Let $\mathcal{C}$ be the collection of all JSJ tori which are the only normal surfaces that have nonnegative Euler characteristic, except for boundary parallel normal tori, and let S be the union of tori in the collection $\mathcal{C}$. Since S is a collection of closed 2-sided essential normal surfaces, crushing each of these tori to a new ideal vertex produces a 1-efficient ideal triangulation for each JSJ piece M_i of M, for $1 \leq i \leq k$, by Theorem 5 in Section 2.

Now, given an essential proper surface Σ embedded in M, we can isotope Σ until it meets each of the tori of $\mathcal{C}$ in a minimal set of essential loops. (We may assume that Σ meets all JSJ tori in $\mathcal{C}$. The reason is that if there are JSJ tori and JSJ components that Σ does not meet, we can cut and remove those JSJ components along the JSJ tori and then work on the remaining M.) In particular, it then follows that $\Sigma \cap M_i = \Sigma_i$ is a properly embedded essential surface in M_i. We can further isotope Σ to have minimal intersection with the edges of tori in $\mathcal{C}$, which is required in the spin normalizing process of each piece Σ_i. This immediately implies that Σ intersects tori of $\mathcal{C}$ in normal curves. So when we crush all the tori of $\mathcal{C}$ to ideal vertices, Σ_i becomes an essential surface Σ_i^* with the same boundary slope as Σ_i at each vertex-linking torus of $\Im_i^*$ and properly embedded in $(\overline{M}_i^*, \Im_i^*)$, where $\overline{M}_i^*$ is obtained by deleting vertices from the result of the canonical crushing $f_i : \hat{M}_i \to \overline{M}_i$ and $\Im_i^*$ is the induced ideal triangulation.

Since, by Theorem 5, $\Im_i^*$ is 1-efficient, Σ_i^* can be spin normalized by [KR2] with any combination of directions of spinning for the boundary curves. We will use these spun normal surfaces, each representing Σ_i^* in $\overline{M}_i^*$ $(1 \leq i \leq k)$, as barriers when normalizing a spinning of Σ, and do this in a covering space of M. Here is some notation for the construction.

- Σ^s (resp. Σ_i^s): the spinning of Σ (resp. Σ_i) obtained by attaching infinite annuli parallel to the boundary tori of M (resp. M_i) along $\partial\Sigma$ (resp. $\partial\Sigma_i$).
- Σ^{ns} (resp. Σ_i^{ns}): the normal spinning of Σ (resp. Σ_i) obtained by normalizing Σ^s (resp. Σ_i^s) in $(\mathring{M}, \Im^*)$ (resp. $(\overline{M}_i^*, \Im_i^*)$). (Here, Σ_i^{ns} is a pseudo-spun normal surface in M.)
- $\tilde{M}$: the covering space of M corresponding to $\pi_1(\Sigma)$.
- $\tilde{M}_i$: the lift of M_i to $\tilde{M}$, which is the covering space of M_i corresponding to $\pi_1(\Sigma_i)$.
- $\tilde{\Sigma}$: the lift of Σ to $\tilde{M}$.
- $\tilde{\Sigma}'$ and $\tilde{\Sigma}''$: parallel copies of $\tilde{\Sigma}$ placed on opposite sides of $\tilde{\Sigma}$.

- $\tilde{\Sigma}_i = \tilde{\Sigma} \cap \tilde{M}_i$: the lift of Σ_i to $\tilde{M}_i$.
- $\tilde{\Sigma}^s$ (resp. $\tilde{\Sigma}_i^s$): the spinning of $\tilde{\Sigma}$ (resp. $\tilde{\Sigma}_i$) obtained by attaching infinite annuli parallel to the boundary infinite annuli of $\tilde{M}$ (resp. $\tilde{M}_i$) along $\partial\tilde{\Sigma}$ (resp. $\partial\tilde{\Sigma}_i$).

Let $\tilde{M}$ be the covering space of M corresponding to $\pi_1(\Sigma)$ and $\tilde{\Sigma}$ the lift of Σ to $\tilde{M}$. It is obvious that the covering space $\tilde{M}_i$ of M_i corresponding to $\pi_1(\Sigma_i)$ is embedded in $\tilde{M}$, and each piece $\tilde{\Sigma}_i = \tilde{\Sigma} \cap \tilde{M}_i$ is an essential surface properly embedded in $\tilde{M}_i$. Since Σ is 2-sided, we can take a parallel copy of $\tilde{\Sigma}$ on each side and denote the copies by $\tilde{\Sigma}'$ and $\tilde{\Sigma}''$. All three copies are parallel, i.e., normally isotopic in $\tilde{M}$ with the induced ideal triangulation. (Here, the ideal triangulation of $\tilde{M}$ contains an infinite number of ideal tetrahedra.) Let $\tilde{\Sigma}_i' = \tilde{\Sigma}' \cap \tilde{M}_i$ and $\tilde{\Sigma}_i'' = \tilde{\Sigma}'' \cap \tilde{M}_i$. Now we will spin all these $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ surfaces by attaching infinite annuli parallel to $\partial\tilde{M}_i$ along the boundary curves of $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ (see Figure 9).

The spinning direction for each boundary curve will be chosen so that the spinning does not meet $\tilde{\Sigma}$. Note that this choice of spinning direction causes $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ to spin in opposite directions in $\tilde{M}_i$ $(1 \le i \le k)$. Now using the methods of [KR2], we claim that the spinnings of $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ can be pseudo-spin normalized as follows: Let Σ_i' and Σ_i'' be the image of $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ into M, and $(\Sigma_i')^*$ and $(\Sigma_i'')^*$ be the corresponding essential surfaces properly embedded in $\overline{M}_i^*$. As mentioned above, the induced cell-structure $\Im_i^*$ of $\overline{M}_i^*$ is a 1-efficient ideal triangulation by Theorem 5, so any spinnings of $(\Sigma_i')^*$ and $(\Sigma_i'')^*$ can be spin normalized by [KR2] and the inverse images of the resulting pseudo-spun normal surfaces are also pseudo-spun normal surfaces (by Remark 4) in M_i. Finally the latter surfaces can be lifted to pseudo-spun normal surfaces, say $(\tilde{\Sigma}_i')^{ns}$ and $(\tilde{\Sigma}_i'')^{ns}$, representing $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ in $\tilde{M}_i$. This establishes the claim that the spinnings of $\tilde{\Sigma}_i'$ and $\tilde{\Sigma}_i''$ can be pseudo-spin normalized.

Now we will use all the surfaces $(\tilde{\Sigma}_i')^{ns}$ and $(\tilde{\Sigma}_i'')^{ns}$, which are lifts of spun normal surfaces representing Σ_i' and Σ_i'', as a barrier when we normalize a spinning of $\tilde{\Sigma}$ in $\tilde{M}$ (see Figure 10).

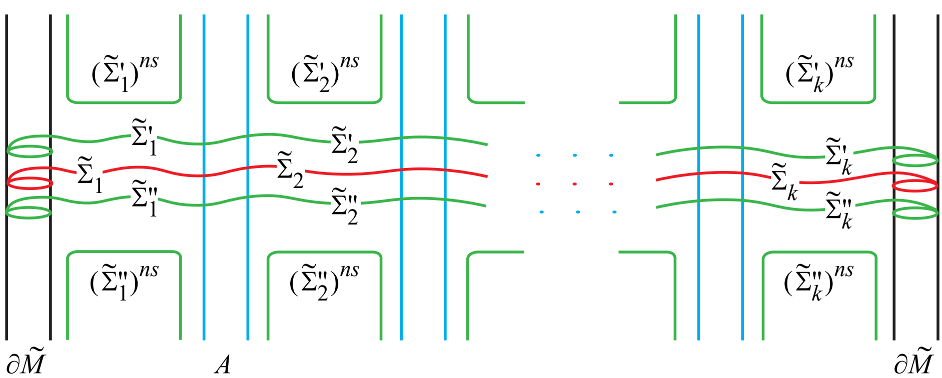


Figure 9. Both way spinnings of $\tilde{\Sigma}_i$ in $\tilde{M}$.

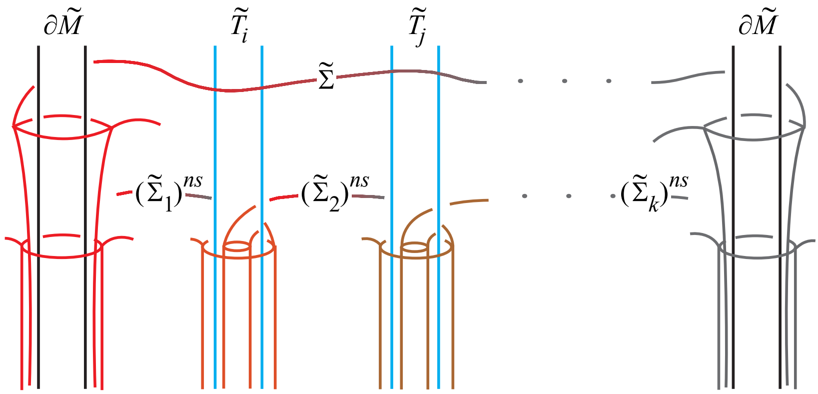


Figure 10. A barrier for spun normalizing $\tilde{\Sigma}$.

Note that the pseudo-spun normal surfaces $(\tilde{\Sigma}'_i)^{ns}$ and $(\tilde{\Sigma}''_i)^{ns}$ are inverse images of spun normal surfaces obtained by normalizing the corresponding spinnings of $\tilde{\Sigma}'_i$ and $\tilde{\Sigma}''_i$, and the direction of spinnings of $\tilde{\Sigma}'_i$ and $\tilde{\Sigma}''_i$ were chosen to not meet $\tilde{\Sigma}$. This implies that $\tilde{\Sigma}$ does not meet any of $(\tilde{\Sigma}'_i)^{ns}$ and $(\tilde{\Sigma}''_i)^{ns}$ for $1 \leq i \leq k$, and any combination of spinning directions for the boundary curves of $\tilde{\Sigma}$ can be chosen so that the spinnings with arbitrarily long annuli attached are all disjoint from the barriers $(\tilde{\Sigma}'_i)^{ns}$ and $(\tilde{\Sigma}''_i)^{ns}$ for $1 \leq i \leq k$.

The normalizing process does not introduce any new intersections with the 1-skeleton of the triangulation; it just removes them by isotopies. With this observation, the union of pseudo-spun normal surfaces $(\tilde{\Sigma}'_i)^{ns}$ and $(\tilde{\Sigma}''_i)^{ns}$ acts as a *barrier*, so that the normalization of a spinning of $\tilde{\Sigma}$ in the complement of this union does not touch the barrier (see Section 3 of [JR1] for details on barriers) and $\tilde{\Sigma}$ must normally spin in $\tilde{M}$ exactly as in the argument in [KR2]. What we need to additionally check here, unlike [KR2], is that the normal spinning of $\tilde{\Sigma}$ does not spin along the infinite annuli lifted from JSJ tori of M. Suppose that $\tilde{\Sigma}$ normally spins along the infinite annulus A lifted from a JSJ torus T bounding two JSJ components M_i and M_j. Here, there are two possible ways of spinning $\tilde{\Sigma}$ along A; spinning between $(\tilde{\Sigma}'_i)^{ns}$ and $(\tilde{\Sigma}'_j)^{ns}$, or spinning between $(\tilde{\Sigma}''_i)^{ns}$ and $(\tilde{\Sigma}''_j)^{ns}$ (see Figure 11).

However, we obtain the normal spinning $(\tilde{\Sigma})^{ns}$ of $\tilde{\Sigma}$ by normalizing a spinning $(\tilde{\Sigma})^{s}$, which is obtained by attaching infinitely long annuli along $\partial\tilde{\Sigma}$ lying on the annuli of $\partial\tilde{M}$ (note that none of these were lifted from JSJ tori), without touching the barriers. Therefore, neither $(\tilde{\Sigma})^{s}$ nor $(\tilde{\Sigma})^{ns}$ can meet the 1-skeleton located between the long annuli of $(\tilde{\Sigma}'_i)^{ns}$ (resp. $(\tilde{\Sigma}''_i)^{ns}$) and $(\tilde{\Sigma}'_j)^{ns}$ (resp. $(\tilde{\Sigma}''_j)^{ns}$). This establishes the claim that the spun normal surface $(\tilde{\Sigma})^{ns}$ does not spin along any lifts of JSJ tori of M.

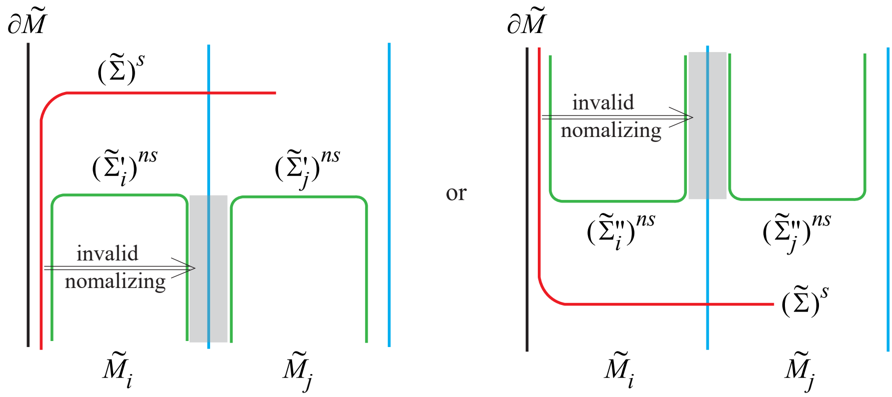


Figure 11. Spinnings of $\tilde{\Sigma}$ along a lift of a JSJ torus in $\tilde{M}$.

Now, $(\tilde{\Sigma})^{ns}$, which normally spins along the boundary annuli of $\tilde{M}$, can be projected to M to obtain a normal spinning of Σ in M. This completes the proof. □

5. Constructing suitable triangulations

Among the assumptions of Theorem 6, the condition of T-efficiency of the triangulation, that the boundary of the JSJ components are the only normal tori, seems rather strong. However, in this section we show that such an ideal triangulation can always be constructed if the manifold under consideration admits a taut ideal triangulation. Moreover, Lackenby showed in [Lac] that taut ideal triangulations are very common; it is sufficient to assume the manifold is irreducible, P^2-irreducible and anannular to achieve a taut triangulation. The assumption that all JSJ components are hyperbolic implies these conditions and hence the existence of a T-efficient taut triangulation.

Lemma 7. *Suppose that M is compact, irreducible and P^2-irreducible and has essential tori and Klein bottle boundary components. Further suppose that M is toroidal and admits a taut ideal triangulation $\mathcal{T}$. Then every essential torus or Klein bottle admits a unique normal representative in its isotopy class.*

Proof. The argument follows by the same method as in [DGR]. We first consider the case of 2-sided surfaces. Namely, if there were two normal essential 2-sided tori or 2-sided Klein bottles isotopic to each other [DGR, Theorem 5.5], then we can find two disjoint such normal surfaces bounding a product region [Wal, Lemma 5.3]. So the surfaces are topologically but not normally parallel. However, by standard sweepout theory from [Rub] and [Sto], there must be an almost normal torus or Klein bottle in this product region, and this contradicts the taut structure of $\mathcal{T}$. According to the Euler characteristic calculation using the angles induced by the taut structure (see Remark 6), almost normal surfaces with a nonnegative Euler characteristic are not allowed in such a structure.

To deal with the case of 1-sided surfaces, assume that is a normal essential 1-sided torus or Klein bottle. The boundary of a small regular neighborhood is then a 2-sided normal torus or Klein bottle. By the 2-sided case, this surface is unique in its isotopy class. but then it follows immediately that so is also the 1-sided surface. □

Remark 6. Let Σ be a normal or almost normal surface in M with an ideal triangulation $\mathcal{T}$.

(1) By Gauss–Bonnet,

$$\begin{aligned} 2\pi\chi(\Sigma) &= \sum_T (\text{vertex angle sum of a triangle } T - \pi) \\ &\quad + \sum_Q (\text{vertex angle sum of a quadrilateral } Q - 2\pi) \\ &\quad + \sum_O (\text{vertex angle sum of an octagon } O - 6\pi). \end{aligned}$$

(2) In a taut structure, the vertex angle sum is π for a triangle, 0 or 2π for a quadrilateral, and 2π or 4π for an octagon. For details, see [KR1].

Theorem 8. *Suppose that M is compact, irreducible and P^2-irreducible and has essential tori boundary components. Further suppose that M has a JSJ decomposition with only hyperbolic pieces. Then M has a taut ideal triangulation $\mathcal{T}$ which is T-efficient. Consequently, any surface Σ in M satisfying the same conditions as Theorem 6 can be spin normalized with boundary curves spinning in all possible combinations of directions.*

Proof. By Theorem 2.6 of [KR1], any taut ideal triangulation is 0-efficient; that is, there are no embedded normal 2-spheres, projective planes, or Klein bottles, and moreover any normal torus or Klein bottle is essential. By Lemma 7, every essential torus or Klein bottle admits a unique normal representative in its isotopy class. Therefore, it suffices to establish the existence of a taut ideal triangulation.

According to Theorem 1 of [Lac], a compact, irreducible, P^2-irreducible, and anannular 3-manifold with incompressible torus boundary admits a taut ideal triangulation. The assumption that the 3-manifold has a JSJ decomposition with all components hyperbolic implies that the manifold is anannular. Indeed, suppose there exists an essential annulus or Möbius band B. Then B can be isotoped to intersect the JSJ tori in parallel essential curves. An innermost annulus between two such curves would lie in a single JSJ component and must be inessential, since all JSJ components are hyperbolic. Consequently, B can be isotoped to remove all intersections with the JSJ tori. This contradicts the assumption that B is essential, as it would then be properly embedded in a hyperbolic component. Hence, all such 3-manifolds admit taut ideal triangulations. This completes the proof. □

Example 1. Suppose that M is a punctured surface bundle over a circle. Assume that the fiber is a union of subsurfaces glued along their boundary curves. Finally suppose that the monodromy is a collection of pseudo-Anosov maps on these subsurfaces and the boundary curves of the subsurfaces are fixed under the monodromy. If we take a layered ideal triangulation of M, this clearly has a taut structure satisfying the conditions of Theorem 6 and the JSJ decomposition is given by the surface bundles over a circle with fibers given by the subsurface system. In general such bundles have many essential surfaces which do not have subsurfaces in the fiber bundle structure. These surfaces can all be spin normalized by Theorem 8.

Remark 7. In the proof of Theorem 8, it is sufficient to assume that all JSJ components of M adjacent to the boundary components of M are hyperbolic in order to conclude that M is anannular and hence admits a taut triangulation. Therefore, it may be possible to extend the main results of this paper to this more general setting. However, for such 3-manifolds, in addition to the tori defining the JSJ decomposition, further essential tori are allowed inside JSJ components that are not adjacent to the boundary of M. This must be taken into account in the construction, and as a consequence, applying the methods of [KR2] becomes more difficult.

References

[BJR] B. Bryant, W. Jaco, and J. H. Rubinstein, "Efficient triangulations and boundary slopes", *Topology Appl.* **297** (2021), art. id. 107689, 18 pp. MR

[DGG] T. Dimofte, D. Gaiotto, and S. Gukov, "3-manifolds and 3d indices", *Adv. Theor. Math. Phys.* **17**:5 (2013), 975–1076. MR

[DGR] N. M. Dunfield, S. Garoufalidis, and J. H. Rubinstein, "Counting essential surfaces in 3-manifolds", *Invent. Math.* **228**:2 (2022), 717–775. MR

[GHRH] S. Garoufalidis, C. D. Hodgson, N. R. Hoffman, and J. H. Rubinstein, "The 3D-index and normal surfaces", *Illinois J. Math.* **60**:1 (2016), 289–352. MR

[GHRS] S. Garoufalidis, C. D. Hodgson, J. H. Rubinstein, and H. Segerman, "1-efficient triangulations and the index of a cusped hyperbolic 3-manifold", *Geom. Topol.* **19**:5 (2015), 2619–2689. MR

[JR1] W. Jaco and J. H. Rubinstein, "0-efficient triangulations of 3-manifolds", *J. Differential Geom.* **65**:1 (2003), 61–168. MR

[JR2] W. Jaco and J. H. Rubinstein, "Inflations of ideal triangulations", *Adv. Math.* **267** (2014), 176–224. MR

[KR1] E. Kang and J. H. Rubinstein, "Ideal triangulations of 3-manifolds, II: Taut and angle structures", *Algebr. Geom. Topol.* **5** (2005), 1505–1533. MR

[KR2] E. Kang and J. H. Rubinstein, "Spun normal surfaces in 3-manifolds, I: 1-efficient triangulations", *Algebr. Geom. Topol.* **22**:8 (2022), 3533–3576. MR

[Lac] M. Lackenby, "Taut ideal triangulations of 3-manifolds", *Geom. Topol.* **4** (2000), 369–395. MR

[Rub] J. H. Rubinstein, "Polyhedral minimal surfaces, Heegaard splittings and decision problems for 3-dimensional manifolds", pp. 1–20 in *Geometric topology* (Athens, GA, 1993), vol. 1, AMS/IP Stud. Adv. Math. **2.1**, Amer. Math. Soc., 1997. MR

[Sie] L. C. Siebenmann, "Approximating cellular maps by homeomorphisms", *Topology* **11** (1972), 271–294. MR

[Sto] M. Stocking, "Almost normal surfaces in 3-manifolds", *Trans. Amer. Math. Soc.* **352**:1 (2000), 171–207. MR

[Wal] F. Waldhausen, "On irreducible 3-manifolds which are sufficiently large", *Ann. of Math.* (2) **87** (1968), 56–88. MR

Received July 16, 2025. Revised January 29, 2026.

ENSIL KANG
DEPARTMENT OF MATHEMATICS
COLLEGE OF NATURAL SCIENCES
CHOSUN UNIVERSITY
GWANGJU
SOUTH KOREA
ekang@chosun.ac.kr

J. HYAM RUBINSTEIN
SCHOOL OF MATHEMATICS AND STATISTICS
THE UNIVERSITY OF MELBOURNE
MELBOURNE VIC
AUSTRALIA
joachim@unimelb.edu.au

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.395

RESOLVENT BOUNDS FOR REPULSIVE POTENTIALS

ANDRÉS LARRAÍN-HUBACH, YULONG LI,
JACOB SHAPIRO AND JOSEPH TILLER

We prove limiting absorption resolvent bounds for the semiclassical Schrödinger operator with a repulsive potential in dimensions $n \geq 3$, which may have a singularity at the origin. As an application, we obtain time decay for the weighted energy of the solution to the associated wave equation with a short range repulsive potential and compactly supported initial data.

1. Introduction and statement of results

We establish limiting absorption resolvent bounds for the semiclassical Schrödinger operator with a repulsive potential in dimensions three and higher. The one-dimensional case was studied in [ChDa21, Section 2]. As an application, we obtain time decay of a weighted energy for the solution to the associated wave equation with a short range repulsive potential and compactly supported initial data.

Let $\Delta := \sum_{j=1}^{n} \partial_j^2 \leq 0$ be the Laplacian on $\mathbb{R}^n$. We use $(r, \theta) = (|x|, x/|x|) \in (0, \infty) \times \mathbb{S}^{n-1}$ for polar coordinates on $\mathbb{R}^n \setminus \{0\}$. Throughout, we equip $\mathbb{S}^{n-1}$ with the standard Borel measure $d\theta$ such that the product measure $r^{n-1} dr \times d\theta$ gives Lebesgue measure on $(0, \infty) \times \mathbb{S}^{n-1}$. Put $B(0, r_0) := \{x \in \mathbb{R}^n : |x| < r_0\}$. For a function u defined on a subset of $\mathbb{R}^n$, we write $u(r, \theta) := u(r\theta)$, and denote the radial derivative by $u' := \partial_r u$. If $E \subseteq \mathbb{R}^n$ is a Borel set, $\mathbf{1}_E$ stands for its indicator function.

Our Schrödinger operator takes the form

$$P(h) := -h^2 \Delta + V(x) : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n), \quad x \in \mathbb{R}^n, \tag{1}$$

where $h > 0$ is the semiclassical parameter. The conditions we place on the Borel measurable potential $V : \mathbb{R}^n \to \mathbb{R}$ are as follows. We suppose $V = V^+ - V^-$ with $V^+ = \max(V, 0)$ and $V^- = -\min(V, 0)$; moreover

$$V^- \in L^\infty(\mathbb{R}^n), \tag{2}$$

MSC2020: 35P25.
Keywords: resolvent estimate, Schrödinger operator, repulsive potential, local energy decay.

$$r^{\rho(n)}V(x) \text{ is bounded for } r \le 1, \tag{3}$$

$$V(x) \text{ is bounded for } r \ge 1, \tag{4}$$

$$\text{for each } \theta \in \mathbb{S}^{n-1},\ (0,\infty) \ni r \mapsto V(r,\theta) := V(r\theta) \text{ has bounded variation.} \tag{5}$$

Here, $\rho(n) > 0$ depends on the dimension n:

$$\rho(n) < \begin{cases} \frac{3}{2} & \text{if } n = 3, \\ 2 & \text{if } n \ge 4. \end{cases} \tag{6}$$

Recall that a function f of locally bounded variation on an interval $I \subseteq \mathbb{R}$ has distributional derivative equal to a locally finite signed Borel measure, which we denote by df. In particular

$$\int \varphi\, df = -\int f\varphi' dx, \quad \varphi \in C_0^\infty(I), \tag{7}$$

where the dx on the right side denotes Lebesgue measure on I; df further satisfies

$$\int_{(a,b]} df = f^R(b) - f^R(a), \tag{8}$$

for any interval $(a,b] \subseteq \mathring{I}$, where $f^R(x) := \lim_{\delta \to 0^+} f(x+\delta)$.

The last condition we impose on V is that there exists $C_V > 0$ such that for all $\theta \in \mathbb{S}^{n-1}$ and every bounded Borel set $E \subseteq (0,\infty)$,

$$\int_E dV(r,\theta) \le -C_V \int_E (r+1)^{-1} V^+(r,\theta) dr. \tag{9}$$

We emphasize that since the inequality (9) is one sided, each measure $dV(\,\cdot\,,\theta)$ is allowed to have negative point masses. Moreover, because only the positive part V^+ appears on the right side, the potential is allowed to approach a negative constant as $r \to \infty$. A prototype potential satisfying the above conditions is

$$V(r,\theta) = g(\theta)\big(\mathbf{1}_{B(0,1)} r^{-\rho(n)} - 2^{-1}\mathbf{1}_{\mathbb{R}^n \setminus B(0,1)}(r^{-\delta} - 2^{-1})\big)$$

for some $\delta > 0$ and $g \ge 0$ a bounded measurable function on $\mathbb{S}^{n-1}$. Note that for $V \in C^1(\mathbb{R}^n; [0,\infty))$, (9) implies each $V(\,\cdot\,,\theta)$ is repulsive in sense of classical mechanics, i.e., that $V(r,\theta) > 0$ implies $V'(r,\theta) < 0$. The local bound (3) allows for mild singularities at the origin, most notably the repulsive Coulomb behavior.

For a real-valued potential $V \in L^p(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)$ with $p \ge 2$, $p > n/2$, $P(h)$ is self-adjoint if one takes the Sobolev space $H^2(\mathbb{R}^n)$ as the domain [Ne64, Theorem 8]. The conditions (2), (3), and (4) imply that $V \in L^p(\mathbb{R}^n) + L^\infty(\mathbb{R}^n)$ for some such p.

Our main results are the following weighted limiting absorption resolvent bounds for $P(h)$.

Theorem 1.1. *Suppose $n \ge 3$ and V satisfies* (2)–(5) *and* (9). *Define $P(h)$ by* (1) *and equip it with the domain $H^2(\mathbb{R}^n)$. For all $s > \frac{1}{2}$ and $z = E \pm i\varepsilon$ with $E > 0$*

fixed, there is a constant $\mathfrak{C}(E, s, \varepsilon, \|V^-\|_{L^\infty}) > 0$ *defined in* (40) *such that*

$$\|(r+1)^{-s}(P(h)-z)^{-1}(r+1)^{-s}\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)} \le \frac{\mathfrak{C}(E, s, \varepsilon, \|V^-\|_{L^\infty})}{h}. \tag{10}$$

When $V^- = 0$, we prove stronger estimates.

Theorem 1.2. *Suppose* $n \ge 3$ *and* V *satisfies* (2)–(5) *and* (9) *with* $V^- = 0$. *Define* $P(h)$ *by* (1) *and equip it with the domain* $H^2(\mathbb{R}^n)$. *For all* s, s_1, $s_2 > \frac{1}{2}$, *with* $s_1 + s_2 > 2$, *there is* $C > 0$ *such that for all* $z \in \mathbb{C} \setminus [0, \infty)$ *and* $h > 0$,

$$\left\|(r+1)^{-s}(P(h)-z)^{-1}(r+1)^{-s}\right\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)} \le \frac{C}{h|z|^{1/2}}, \tag{11}$$

$$\left\|(r+1)^{-s_1}(P(h)-z)^{-1}(r+1)^{-s_2}\right\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)} \le \frac{C}{h^2}. \tag{12}$$

Remark 1.3. From (42), we see how the constant C in (11) depends on s and C_V. The dependence of C in (12) on s_1, s_2 and C_V can be deduced from (49).

Section 2 is devoted to the proof of Theorem 1.2, which builds on [ChDa21, Theorem 1.2], where the authors obtained (11) and (12) for bounded, repulsive potentials, nonnegative potentials on the half-line. The novelty of our paper is that it extends these bounds to dimensions $n \ge 3$ for repulsive potentials that may be singular as $r \to \infty$. The corresponding problem in dimension two appears to be delicate, and to our knowledge remains open; we comment below on a possible approach.

Remark 1.4. In Appendix D, we recall how for the case $V = 0$ and $n = 3$, the conditions on s, s_1 and s_2 in Theorem 1.2, as well as the h- and z-dependencies of the right sides of (11) and (12), are nearly optimal in a suitable sense.

Remark 1.5. The weighted estimates underlying (10), (11), and (12) hold under a condition weaker than (3), namely that $|r^{n-1}V(r,\theta)| \to 0$ as $r \to 0$; see (32). (The factor r^{n-1} reflects the volume element of Lebesgue measure in polar coordinates.) These estimates are derived for test functions in $C_0^\infty(\mathbb{R}^n)$ and are transferred to resolvent bounds via the density argument in Appendix C. This step uses that the operator domain is $H^2(\mathbb{R}^n)$ and that $C_0^\infty(\mathbb{R}^n)$ is dense in $H^2(\mathbb{R}^n)$, which is why we impose the stronger hypothesis (3). We note, however, that self-adjoint realizations of $-\Delta + V$ exist under weaker local assumptions on V. For example, if $V \in L^2_{\text{loc}}(\mathbb{R}^n)$ and $V \ge 0$, then $-\Delta + V$ is essentially self-adjoint on $C_0^\infty(\mathbb{R}^n)$ [ReSi75, Theorem X.28]; essential self-adjointness on $C_0^\infty(\mathbb{R}^n \setminus \{0\})$ holds under inverse square lower bounds [ReSi75, Theorem X.30]. It would be interesting to formulate the weighted estimates in such frameworks, which would allow more singular potentials. However, for the wave decay results discussed below, a restriction comes from the need to control rV near the origin. We thus adopt the more streamlined approach here.

We prove Theorems 1.1 and 1.2 using the *spherical energy method*, a widely used technique for deriving weighted estimates for Schrödinger operators. The approach is based on separation of variables and the classical identity

$$r^{\frac{n-1}{2}}(-\Delta)r^{-\frac{n-1}{2}} = -\partial_r^2 + r^{-2}\Lambda, \tag{13}$$

where

$$\Lambda := -\Delta_{\mathbb{S}^{n-1}} + \tfrac{1}{4}(n-1)(n-3), \tag{14}$$

and $\Delta_{\mathbb{S}^{n-1}}$ denotes the negative Laplace–Beltrami operator on $\mathbb{S}^{n-1}$. The repulsivity condition is sufficiently advantageous to allow the use of a relatively simple weight — specifically, the same weight employed in [ChDa21] — to obtain (35) and (46). For more general potentials, it is usually necessary to instead conjugate the Laplacian by $e^{\varphi/h}r^{(n-1)/2}$ (see, e.g., [CaVo02; Da14; GaSh22]) for a suitable phase φ. This results in a Carleman estimate with exponential losses as $h \to 0^+$.

We use in a crucial way that $\Lambda \geq 0$ on $L^2(\mathbb{S}^{n-1})$; see (30). This is why our approach does not cover the case $n = 2$ where the effective potential $-1/(4r^2)$ has a strong negative singularity as $r \to 0$. We expect that repulsive potentials in dimension two can be treated by adapting the Mellin transform methods used in [DGS23; Ob24].

As an application of (12), we prove weighted energy decay for the solution to the wave equation

$$\begin{cases} (\partial_t^2 - \Delta + V(x))u(t,x) = 0 & \text{for } (t,x) \in \mathbb{R}\times\mathbb{R}^n,\ n \geq 3, \\ u(0,x) = u_0(x),\ \partial_t u(0,x) = u_1(x) & \text{for } x \in \mathbb{R}^n, \end{cases} \tag{15}$$

where $u_0 \in H^1(\mathbb{R}^n)$ and $u_1 \in L^2(\mathbb{R}^n)$ have compact support. The potential obeys $V \geq 0$, (3), (5) with $\rho = \rho(n) = 1$, (9), and the extra short range condition

$$\mathbf{1}_{\mathbb{R}^n\setminus B(0,1)} V \leq C(r+1)^{-\delta(n)}, \tag{16}$$

for some $C > 0$ and $\delta(n) > 0$ such that

$$\delta(n) > \begin{cases} \frac{1}{2} + \frac{n+3}{4} & \text{if } n \neq 8, \\ \frac{1}{2} + 3 & \text{if } n = 8. \end{cases} \tag{17}$$

Since $P := P(1) = -\Delta + V$ is self-adjoint (and nonnegative) under such conditions, we may use the spectral theorem for self-adjoint operators to represent the solution to (15) by

$$u(t,\cdot) = \cos(t\sqrt{P})u_0 + \frac{\sin(t\sqrt{P})}{\sqrt{P}}u_1. \tag{18}$$

For $s > 0$ fixed, define the weighted energy of the solution u to (15) to be

$$E_s[u](t) = E_s(t) := \int_{\mathbb{R}^n} \langle x\rangle^{-2s}(|\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2 + |u(t,x)|^2)\,dx.$$

Set also

$$E(0) := \|\nabla u_0\|_{L^2}^2 + \|u_1\|_{L^2}^2.$$

Theorem 1.6. *Suppose* $V \geq 0$ *satisfies* (3), (5) *with* $\rho = 1$, (9), *as well as* (16). *Let* $u_0 \in H^1(\mathbb{R}^n)$ *and* $u_1 \in L^2(\mathbb{R}^n)$ *have compact support. For each s such that*

$$s > \begin{cases} \frac{n+3}{4} & \text{if } n \neq 8, \\ 3 & \text{if } n = 8, \end{cases} \tag{19}$$

there exists $C_s > 0$ *depending on s but independent of* t, u_0, *and* u_1 *so that*

$$E_s(t) \leq C_s \langle t \rangle^{-2} E(0), \tag{20}$$

where $\langle t \rangle := (1 + |t|^2)^{1/2}$.

Remark 1.7. Since $u(-t, \cdot) = \cos(t\sqrt{P})u_0 + (\sin(t\sqrt{P})/\sqrt{P})(-u_1)$, to prove (20) it suffices to suppose $t \geq 0$.

Remark 1.8. We expect that, by adapting the arguments of [Vo04b, Section 3], the assumption of compact support on the initial data can be relaxed. Specifically, the decay should continue to hold for initial conditions lying in an appropriate weighted Sobolev spaces. However, to keep the presentation technically streamlined, we suppose u_0 and u_1 have compact support.

Remark 1.9. The condition (17) has a quirk in dimension eight compared to other dimensions. This appears to be an artifact of our approach, which analyzes the low-frequency behavior of the resolvent kernel in terms of Hilbert–Schmidt norms (Appendix E).

For smooth, nonnegative potentials of compact support, the local energy

$$E_{r_0}(t) := \int_{B(0,r_0)} |\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2 + |u(t,x)|^2 dx, \quad r_0 > 0,$$

obeys

$$E_{r_0}(t) = \begin{cases} O(e^{-ct}) \text{ for some } c > 0 & \text{if } n \geq 3 \text{ is odd,} \\ O(t^{-2n}) & \text{if } n \geq 4 \text{ is even.} \end{cases} \tag{21}$$

Indeed, Vainberg [Va75] showed (21) for compactly supported perturbations of the Laplacian satisfying the generalized Huygens principle (as defined in [Vo04c]). A result of Melrose and Sjöstrand on the propagation of singularities [MeSj78; MeSj82] implies that this principle holds for a broad class of nontrapping perturbations of the Laplacian, which includes smooth, nonnegative, compactly supported potentials. The study of energy decay for nontrapping perturbations has a long history, going back to the works of Lax, Morawetz, and Phillips [LMP63; Mo66; Mo75].

Bounds similar to (20) were obtained in previous works for various classes of short range potentials. In [Za04], Zappacosta considered potentials $V \in C^1(\mathbb{R}^3; (0, \infty))$

with $\partial_x^\alpha V = O(\langle x\rangle^{-\delta-|\alpha|})$ for all $0 \le |\alpha| \le 1$ and some $\delta > 2$. For each $\chi \in C_0^\infty(\mathbb{R}^3)$, the bound

$$\|\chi\sqrt{V}(\sin(t\sqrt{P})/\sqrt{P})\sqrt{V}\chi\|^2_{L^2\to L^2} = O(t^{-2})$$

was proved. In [Vo04a], Vodev showed $E_{r_0}(t) = O(t^{-2})$ in dimension $n \ge 3$, where $V \in C^1(\mathbb{R}^n; [0,\infty))$ obeys

$$V = O(\langle x\rangle^{-\delta_0}) \qquad \text{for some } \delta_0 > 2, \text{ and} \tag{22}$$

$$2V + r\partial_r V \le C\langle x\rangle^{-\delta} \quad \text{for some } C > 0 \text{ and some } \delta > 1. \tag{23}$$

Additionally, it is assumed that V has no resonance at zero energy, a condition closely related to the validity of a bound such as (12) for $|z| \ll 1$. Vodev also obtained weighted energy decay for a class of long-range, nontrapping perturbations of the Laplacian that includes perturbations by a nonnegative long-range potential, provided the initial conditions are spectrally localized away from $[0,a]$ for $a > 0$ sufficiently large [Vo04b]. (In our approach, the positivity of the potential appears crucial for obtaining (12). Indeed, the constant $\mathfrak{C}(E, s, \varepsilon, \|V^-\|_{L^\infty})$ in (40) blows up as $E \to 0^+$ unless $\|V^-\|_{L^\infty} = 0$.)

We note the connection between (23) and our repulsiveness condition. If $V \in C^1(\mathbb{R}^n; (0,\infty))$ satisfies (9), then $\partial_r(\log V(\,\cdot\,,\theta)) \le -C_V(r+1)^{-1}$ uniformly in θ. Integration in r yields $0 \le V(r,\theta) \le C(r+1)^{-C_V}$, so $2V + r\partial_r V \le 2V \le C(r+1)^{-C_V}$, which is (23) provided $C_V > 1$.

In the absence of a bound like (12), or a condition excluding a resonance or eigenvalue at zero, decay of wave equation solutions generally cannot be expected. Thus, in the present work the positivity of the potential serves as a sufficient condition to rule out threshold obstructions. Other sufficient conditions have been obtained in previous works; for example, smallness of the potential in the Rollnik and global Kato norms [RoSc04]. For related discussions and further references, see [JeNe01; ChDa25; CDY25].

The proof of Theorem 1.6 follows the strategy of [Vo04a, Section 3], with modifications to account for the possible singularity of V at the origin. The key step is to establish

$$t^2 E_s(t) \le Ct^2 \int_t^\infty E_s(\tau)\,d\tau \le CE(0), \quad t \ge 1. \tag{24}$$

Using Duhamel's formula and the Fourier transform (t dual to λ), the Fourier transform of u can be expressed in terms of $(P-\lambda^2)^{-1}$; see (64). Because the initial data are compactly supported, finite speed of propagation allows insertion of a cutoff function η, and Plancherel's theorem then reduces control of $E_s(t)$ to bounds on $\langle x\rangle^{-s}(P-\lambda^2)^{-1}\eta$, which comes from (12); see also Lemma 3.1.

However, the factor t^2 in (24) corresponds to differentiation with respect to λ, so it is also necessary to control

$$\langle x\rangle^{-s}\frac{d}{d\lambda}(P-\lambda^2)^{-1}\eta.$$

This is achieved under the stronger assumptions (3) with $\rho = 1$ (compare with (6)) and (16). In particular, the restriction $\rho = 1$ arises from the need for a uniform bound on $\lambda V(-\Delta-\lambda^2)^{-1}\langle x\rangle^{-s}$ for $\operatorname{Im}\lambda > 0$, see (58). It would be interesting to determine whether more singular potentials could be accommodated by controlling the λ-derivative in a less perturbative manner.

We expect that for certain potentials with mild spatial decay depending on the dimension, t^{-2} is the optimal decay for the local energy. A sharper description should depend on a low-frequency expansion of the resolvent around $\lambda = 0$ (see, e.g., [JeNe01]), rather than a bound alone.

Another energy studied is the quantity

$$E_K^{(1)}[u](t) := \int_K |\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2 + V(x)|\,u(t,x)|^2\,dx,$$

where $K \subseteq \mathbb{R}^n$ is a region of interest. In [Vo04a, Theorem 1.1], Vodev studied the case $K = B(0,\gamma_0 t) \subseteq \mathbb{R}^n$ for $n \geq 3$ and suitable $0 < \gamma_0 < 1$. Under the assumption of no resonance at zero, along with (22) for some $\delta_0 > 1$ and (23) for constants $C > 0$ and $\delta > 1$, he proved that $E_K^{(1)}[u](t) = O(t^{-1})$. In [Ik23, Theorems 1.1 and 1.2], Ikehata considered exterior subdomains Ω of $\mathbb{R}^n$, $n \geq 2$, excluding the origin. For compact subsets $K \subseteq \Omega$, he established the same decay rate assuming V is nonnegative, C^1, and obeys $x\cdot\nabla V + 2V \leq 0$. The $O(t^{-1})$-decay was first showed by Morawetz [Mo61] for $V = 0$ in the exterior of a three-dimensional star-shaped obstacle (later improved to exponential decay in [LMP63]).

There is an extensive body of literature on wave decay for higher-order perturbations. For general second-order perturbations that may exhibit trapping, logarithmic decay — rather than polynomial decay — is more typical. For historical background and related developments, see [Bu98; Vo99; Bu02; Bo11; CaVo04; Sh18; ChIk20].

Outline. In Section 2, we prove Theorem 1.2. In Section 3, we apply Theorem 1.2 to prove norm bounds for $\langle x\rangle^{-s}(P-\lambda^2)^{-1}\langle x\rangle^{-s}$ and its λ-derivative. In Section 4 we prove Theorem 1.6. Finally, we include several appendices of technical results that assist with the proofs of earlier sections.

2. Proof of Theorem 1.2

Throughout this section, we take $P(h)$ as in (1) and assume the potential V satisfies (2)–(5) and(9).

By (13),

$$P^{\pm}(h) := r^{\frac{n-1}{2}}\,(P(h) - E \pm i\varepsilon)\,r^{-\frac{n-1}{2}} = -h^2\partial_r^2 + h^2 r^{-2}\Lambda + V - E \pm i\varepsilon, \tag{25}$$

where we let E and ε vary in $[0,\infty)$. Let $u \in r^{(n-1)/2}C_0^\infty(\mathbb{R}^n)$. Define a spherical energy functional $F[u](r)$,

$$F(r) = F[u](r) := \|hu'(r,\cdot)\|^2 - \langle (h^2 r^{-2}\Lambda + V^R(r,\cdot) - E)u(r,\cdot),\, u(r,\cdot)\rangle, \tag{26}$$

where $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ denote the norm and inner product on $L^2(\mathbb{S}_\theta^{n-1})$. We take complex conjugation to occur in the first argument of $\langle\cdot,\cdot\rangle$. Here, $V^R(r,\theta)$ is the measurable function defined by $V^R(r,\theta) := \lim_{k\to\infty} V((r+(1/k))\theta)$ for $k \in \mathbb{N}$. The limit exists for each r and θ since each $V(\cdot,\theta)$ has bounded variation. Each $V^R(\cdot,\theta)$ is decreasing thanks to (2), (8) and (9).

For a weight $w(r)$ which is absolutely continuous, nonnegative, and increasing, we compute the derivative of wF in the sense of distributions on $(0,\infty)$. For this we need the following technical lemma, whose proof we give in Appendix A.

Lemma 2.1. *The function* $r \mapsto \int_{\mathbb{S}^{n-1}} V^R(r,\theta)|u(r,\theta)|^2\,d\theta$ *has locally bounded variation and its derivative in the sense of distributions on* $(0,\infty)$ *is the element of* $C_0^\infty(0,\infty)$ *given by*

$$\varphi \mapsto \int_0^\infty \varphi(r)\int_{\mathbb{S}^{n-1}} V(r,\theta)\,2\,\mathrm{Re}(\bar{u}u')\,d\theta\,dr + \int_{\mathbb{S}^{n-1}}\int_0^\infty \varphi(r)|u(r,\theta)|^2\,dV(r,\theta)\,d\theta. \tag{27}$$

Note that, aside from the term in $F(r)$ involving $V^R(r,\theta)$, the remaining terms are functions of r with locally bounded variation on $(0,\infty)$. Thus by Lemma 2.1, $F(r)$ itself is of locally bounded variation on $(0,\infty)$.

Using (27) we have, in the sense of distributions on $(0,\infty)$:

$$\begin{aligned}(wF)' &= wF' + w'F \\ &= w(-2\,\mathrm{Re}\langle(-h^2\partial_r^2 + h^2r^{-2}\Lambda + V - E)u, u'\rangle \\ &\qquad + 2h^2r^{-3}\langle\Lambda u, u\rangle - \textstyle\int_{\mathbb{S}^{n-1}} |u(r,\theta)|^2\,dV(r,\theta)\,d\theta) \\ &\qquad + w'(\|hu'\|^2 - \langle h^2r^{-2}\Lambda u, u\rangle + \langle(E-V)u, u\rangle \\ &= -2w\,\mathrm{Re}\langle P^{\pm}(h)u, u'\rangle \pm 2\varepsilon w\,\mathrm{Im}\langle u, u'\rangle + w'\|hu'\|^2 \\ &\qquad + (2wr^{-1} - w')\langle h^2r^{-2}\Lambda u, u\rangle + Ew'\|u\|^2 \\ &\qquad - \textstyle\int_{\mathbb{S}^{n-1}} |u(r,\theta)|^2(w(r)\,dV(r,\theta) + w'(r)V(r,\theta))\,d\theta.\end{aligned} \tag{28}$$

First we show (11). Since increasing s decreases the left side of (11), without loss of generality we may take $0 < \delta := 2s - 1 < 1$. We will show that the last line of (28) can be made to have a suitable lower bound, using

$$w(r) := 1 - \frac{C_V}{C_V+\delta}(1+r)^{-\delta}. \tag{29}$$

For such w, we clearly have

$$w'(r) = \frac{\delta C_V}{C_V+\delta}(r+1)^{-1-\delta}.$$

Therefore, on the one hand,

$$2wr^{-1} - w' = 2r^{-1}(r+1)^{-\delta}\left((r+1)^{\delta} - \frac{C_V}{C_V+\delta}\left(1+\frac{\delta r}{2(r+1)}\right)\right) \geq 0,$$

where we used

$$(r+1)^{\delta} - \frac{C_V}{C_V+\delta} \geq \frac{\delta C_V}{C_V+\delta}\int_1^{r+1} s^{\delta-1}\,ds \geq \frac{\delta C_V r}{(C_V+\delta)(r+1)},$$

since $\delta < 1$. On the other hand, using (9), we have, in the sense of measures on bounded Borel subsets of $(0,\infty)$,

$$w\,dV + w'V = \frac{\delta C_V V}{(C_V+\delta)(r+1)^{1+\delta}} + w\,dV \leq \frac{C_V V^+}{1+r}((r+1)^{-\delta}-1) \leq 0.$$

Thus, the last two estimates and (28) imply, for any $\varphi \in C_0^\infty((0,\infty);[0,\infty))$,

$$\begin{aligned}\int \varphi\, d(wF) &= -\int (wF)\varphi' dr \\ &\geq \int \left(-2w\operatorname{Re}\langle P^{\pm}(h)u, u'\rangle \pm 2\varepsilon w \operatorname{Im}\langle u, u'\rangle + w'\|hu'\|^2 + Ew'\|u\|^2\right)\varphi\, dr,\end{aligned} \tag{30}$$

where in the first line of (30) we applied (7). Now, take a sequence of $\varphi_k \in C_0^\infty((0,\infty);[0,1])$ that converges pointwise to the indicator function $\mathbf{1}_{(r_0,r_1]}$ with $0 < r_0 \ll 1$ and r_1 large enough so that $u(r,\theta) = 0$ for near $[r_1,\infty)\times\mathbb{S}^{n-1}$. Substituting $\varphi = \varphi_k$ in (30), sending $k \to \infty$, and applying the dominated convergence theorem and (8) gives

$$\begin{aligned}\int_{r_0}^{\infty} Ew'\|u\|^2 + w'\|hu'\|^2\,dr + w(r_0)F^R(r_0) \\ \leq \int_{r_0}^{\infty} 2w\operatorname{Re}\langle P^{\pm}(h)u, u'\rangle \mp 2\varepsilon w\operatorname{Im}\langle u, u'\rangle\, dr.\end{aligned} \tag{31}$$

Since $u = r^{(n-1)/2}v$ for some $v \in C_0^\infty(\mathbb{R}^n)$, we recognize that

$$\begin{aligned}F^R(r_0) = \|hu'(r_0,\cdot)\|^2 &+ r_0^{n-3}\langle h^2\Delta_{\mathbb{S}^{n-1}}v(r_0,\cdot), v(r_0,\cdot)\rangle \\ &+ \left(Er_0^{n-1} - h^2 4^{-1}(n-1)(n-3)r_0^{n-3}\right)\|v(r_0,\cdot)\|^2 \\ &+ r_0^{n-1}\langle V^R(r_0,\cdot)v(r_0,\cdot), v(r_0,\cdot)\rangle.\end{aligned} \tag{32}$$

We rewrite the term in (32) involving $\Delta_{\mathbb{S}^{n-1}}$ using the formula for the Laplacian in spherical coordinates:

$$r^{-2}\Delta_{\mathbb{S}^{n-1}} = \Delta - \partial_r^2 - (n-1)r^{-1}\partial_r.$$

This leads to

$$\begin{aligned}(33)\quad & r_0^{n-3}\langle h^2\Delta_{\mathbb{S}^{n-1}} v(r_0,\cdot), v(r_0,\cdot)\rangle \\ &= h^2 r_0^{n-1}\langle(\Delta v)(r_0,\cdot), v(r_0,\cdot)\rangle - h^2 r_0^{n-1}\langle(\partial_r^2 v)(r_0,\cdot), v(r_0,\cdot)\rangle \\ &\qquad - h^2(n-1)r_0^{n-2}\langle(\partial_r v)(r_0,\cdot), v(r_0,\cdot)\rangle.\end{aligned}$$

We can express the differential operators ∂_r and ∂_r^2 with respect to the Euclidean coordinate system as

$$(34)\qquad \partial_r = r^{-1}\sum_{j=1}^n x_j\partial_{x_j}, \qquad \partial_r^2 = r^{-2}\sum_{k=1}^n x_k \sum_{j=1}^n x_j \partial_{x_k}\partial_{x_j}.$$

By (3), (33) and (34), all terms in (32) tend to zero as $r_0 \to 0$, except possibly for $\|hu'(r,\cdot)\|^2$ in dimension three, which in that case tends to $|v(0)|^2 \int_{\mathbb{S}^{n-1}} d\theta$. We conclude that

$$\lim_{r_0\to 0} w(r_0)F(r_0) = w(0)F(0) = \begin{cases}\omega_{n-1}w(0)|v(0)|^2 & \text{if } n=3,\\ 0 & \text{if } n\ge 4,\end{cases}$$

where ω_{n-1} is the $(n-1)$-dimensional volume of $\mathbb{S}^{n-1}$.

Thus, in view of (31) and $0 < w \le 1$,

$$\begin{aligned}(35)\quad & \int_0^\infty Ew'\|u\|^2 + w'\|hu'\|^2\,dr \\ &\le 2\Big(\int_0^\infty \frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}\Big(\int_0^\infty w'\|hu'\|^2\,dr\Big)^{1/2} \\ &\qquad + \frac{2\varepsilon}{h}\Big(\int_0^\infty \|u\|^2\,dr\Big)^{1/2}\Big(\int_0^\infty \|hu'\|^2\,dr\Big)^{1/2}.\end{aligned}$$

We now estimate

$$\begin{aligned}&\int_0^\infty \|hu'\|^2\,dr \\ &= \operatorname{Re}\int_0^\infty \langle u, -h^2u''\rangle\,dr \\ &= \operatorname{Re}\Big(\int_0^\infty \langle u, P^\pm(h)u\rangle\,dr + \int_0^\infty \langle u, (E-V-h^2r^{-2}\Lambda)u\rangle\,dr \mp i\varepsilon\int_0^\infty \|u\|^2\,dr\Big) \\ &= \operatorname{Re}\int_0^\infty \langle u, P^\pm(h)u\rangle\,dr + \int_0^\infty \langle u, (E-V-h^2r^{-2}\Lambda)u\rangle\,dr \\ &\le \Big(\int_0^\infty \frac{1}{w'}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}\Big(\int_0^\infty w'\|u\|^2\,dr\Big)^{1/2} + (E+\|V^-\|_{L^\infty})\int_0^\infty \|u\|^2\,dr\end{aligned}$$

and

$$\begin{aligned}\varepsilon\int_0^\infty \|u\|^2\,dr &= \varepsilon\|v\|_{L^2}^2\\ &= \big|\mathrm{Im}\langle(P(h)-E\pm i\varepsilon)v,v\rangle_{L^2}\big| = \Big|\mathrm{Im}\int_0^\infty\langle P^\pm(h)u,u\rangle\,dr\Big|\\ &\le \Big(\int_0^\infty \frac{1}{w'}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}\Big(\int_0^\infty w'\|u\|^2\,dr\Big)^{1/2}.\end{aligned}$$

Combining these estimates gives

$$\begin{aligned}&\frac{\varepsilon^2}{h^2}\int_0^\infty\|u\|^2\,dr\int_0^\infty\|hu'\|^2\,dr\\ &\qquad\le (E+\varepsilon+\|V^-\|_{L^\infty})\int_0^\infty\frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr\cdot\int_0^\infty w'\|u\|^2\,dr.\end{aligned}$$

Plugging this into (35) yields

$$\begin{aligned}&\int_0^\infty Ew'\|u\|^2+w'\|hu'\|^2\,dr \qquad (36)\\ &\le 2\Big(\int_0^\infty\frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}\\ &\cdot\Big(\Big(\int_0^\infty w'\|hu'\|^2\,dr\Big)^{1/2}+\big(E+\varepsilon+\|V^-\|_{L^\infty}\big)^{1/2}\Big(\int_0^\infty w'\|u\|^2\,dr\Big)^{1/2}\Big).\end{aligned}$$

Now restrict to $E>0$ and complete the square in (36) to find

$$\begin{aligned}&\Big(\sqrt{E}\Big(\int_0^\infty w'\|u\|^2\,dr\Big)^{\frac12}-\frac{\sqrt{E+\varepsilon+\|V^-\|_{L^\infty}}}{\sqrt{E}}\Big(\int_0^\infty\frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr\Big)^{\frac12}\Big)^2 \qquad (37)\\ &+\Big(\Big(\int_0^\infty w'\|hu'\|^2\,dr\Big)^{\frac12}-\Big(\int_0^\infty\frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr\Big)^{\frac12}\Big)^2\\ &\qquad\le\frac{2E+\varepsilon+\|V^-\|_{L^\infty}}{E}\int_0^\infty\frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr.\end{aligned}$$

Dropping the second term on the left side of (37) implies, for all $E>0$ and $\varepsilon\ge 0$,

$$\begin{aligned}&\sqrt{E}\Big(\int_0^\infty w'\|u\|^2\,dr\Big)^{\frac12} \qquad (38)\\ &\le\Big(\frac{\sqrt{E+\varepsilon+\|V^-\|_{L^\infty}}}{\sqrt{E}}+\frac{\sqrt{2E+\varepsilon+\|V^-\|_{L^\infty}}}{\sqrt{E}}\Big)\Big(\int_0^\infty\frac{1}{h^2w'}\|P^\pm(h)u\|^2\,dr\Big)^{\frac12}.\end{aligned}$$

Recall that $w'=C_V\delta(C_V+\delta)^{-1}(r+1)^{-1-\delta}$ and $\delta=2s-1$. From (38) and the density argument in Appendix C we get, rewriting $z=E\pm i\varepsilon$, that

$$\|(1+r)^{-2s}(P-z)^{-1}(1+r)^{-2s}\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)}\le\frac{\mathfrak{C}(E,s,\varepsilon,\|V^-\|_{L^\infty})}{h}, \qquad (39)$$

where

$$\mathfrak{C}(E, s, \varepsilon, \|V^-\|_{L^\infty}) = \frac{C_V+\delta}{E\delta C_V}\big((E+\varepsilon+\|V^-\|_{L^\infty})^{1/2}+(2E+\varepsilon+\|V^-\|_{L^\infty})^{1/2}\big). \tag{40}$$

From this point on, we assume $V^- = 0$. Fix $0<\alpha<1$. From (38), we get for all $h>0$, all $E\pm i\varepsilon$ in the sector $\{z\in\mathbb{C} : |\operatorname{Im} z| < \alpha \operatorname{Re} z\}$, and $u \in r^{n-1/2}C_0^\infty(\mathbb{R}^n)$,

$$\begin{aligned}\big|(E\pm i\varepsilon)^{1/2}\big|\Big(\int_0^\infty (r+1)^{-2s}\|u\|^2\,dr\Big)^{1/2} &\le h^{-1}(1+\alpha^2)^{1/4}\Big(\frac{1}{\delta}+\frac{1}{C_V}\Big)\big((1+\alpha)^{1/2}+(2+\alpha)^{1/2}\big)\\ &\quad\cdot\Big(\int_0^\infty (r+1)^{2s}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}.\end{aligned} \tag{41}$$

Here, our branch of the complex square root is chosen so that $\operatorname{Im}(E\pm i\varepsilon)^{1/2}>0$, and we used that $|(E\pm i\varepsilon)|^{1/2} = (E^2+\varepsilon^2)^{1/4} \le E^{1/2}(1+\alpha^2)^{1/4}$ because $E\pm i\varepsilon \in \{z\in\mathbb{C} : |\operatorname{Im} z| < \alpha\operatorname{Re} z\}$. Since $u\in r^{(n-1)/2}C_0^\infty(\mathbb{R}^n)$, a standard density argument, which we review in Appendix C, shows that (41) implies

$$\begin{aligned}\big\|z^{1/2}(r+1)^{-s}(P(h)-z)^{-1}(r+1)^{-s}\big\|_{L^2\to L^2} \le h^{-1}(1+\alpha^2)^{1/4}\Big(\frac{1}{\delta}+\frac{1}{C_V}\Big)\big((1+\alpha)^{1/2}+(2+\alpha)^{1/2}\big),\end{aligned} \tag{42}$$

on $\{z\in\mathbb{C} : |\operatorname{Im} z| < \alpha\operatorname{Re} z\}$ and for any $0<\alpha<1$. To extend this bound to all $z\in\mathbb{C}\setminus[0,\infty)$, we will use the Phragmén–Lindelöf method [EM] in the following way. For $u, v\in L^2(\mathbb{R}^n)$, put

$$U(z) := z^{1/2}\big\langle (r+1)^{-s}(P(h)-z)^{-1}(r+1)^{-s}u, v\big\rangle_{L^2}.$$

Then $U(z)$ in analytic in $\Omega_\alpha := \{z\in\mathbb{C} : \alpha\operatorname{Re} z < |\operatorname{Im} z|\}$. By (42), on $\partial\Omega_\alpha\setminus\{0\}$ we have

$$|U(z)| \le h^{-1}(1+\alpha^2)^{1/4}\Big(\frac{1}{\delta}+\frac{1}{C_V}\Big)((1+\alpha)^{1/2}+(2+\alpha)^{1/2})\|u\|_{L^2}\|v\|_{L^2}. \tag{43}$$

On the other hand, in Ω_α, we have the standard bound

$$|U(z)| \le \frac{|z|^{1/2}\|u\|_{L^2}\|v\|_{L^2}}{\operatorname{dist}(z,[0,\infty))} = \begin{cases} \dfrac{\|u\|_{L^2}\|v\|_{L^2}}{|z|^{1/2}} & \text{if } \operatorname{Re} z<0,\\[2ex] \dfrac{|z|^{1/2}\|u\|_{L^2}\|v\|_{L^2}}{|\operatorname{Im} z|} & \text{if } \operatorname{Re} z\ge 0,\ z\in\Omega_\alpha,\end{cases} \tag{44}$$

where we used

$$\frac{1}{\operatorname{dist}(z,[0,\infty))} = \frac{1}{\inf_{r\ge 0}((\operatorname{Re} z-r)^2+(\operatorname{Im} z)^2)^{1/2}} = \begin{cases}|z|^{-1} & \text{if } \operatorname{Re} z<0,\\ |\operatorname{Im} z|^{-1} & \text{if } \operatorname{Re} z\ge 0,\ z\in\Omega_\alpha.\end{cases}$$

Finally, define, $g(z) = e^{i(z^{-1})^{1/2}}$, where our branch of the square root is as above. In Ω_α, $|g(z)| \le e^{-c_\alpha |z|^{-1/2}}$ for some $0 < c_\alpha < 1$ depending on α; this is because, from the definition of Ω_α, there exists $\theta_\alpha \in (0, \pi/4)$ such that any $z \in \Omega_\alpha$ takes the form $|z|e^{i\theta}$ with $\theta_\alpha < \theta < 2\pi - \theta_\alpha$. Hence $\operatorname{Re}(i(z^{-1})^{1/2}) = -|z|^{-1/2}\sin(\theta/2) \le -|z|^{-1/2}\sin(\theta_\alpha/2)$. Combining with (44) gives

$$\limsup_{z\to 0,\, z\in\Omega_\alpha} |g(z)|^\sigma |U(z)| = 0, \quad \sigma > 0. \tag{45}$$

Therefore, from (43) and (45), the Phragmén–Lindelöf theorem (Theorem B.1 in Appendix B) implies that (42) holds for all $z \in \Omega_\alpha$ too. Sending $\alpha \to 0^+$ completes the proof of (11).

To prove (12), start again at (37) and drop the first term on the left-hand side. Still working on $\{z \in \mathbb{C} : |\operatorname{Im} z| < \alpha \operatorname{Re} z\}$, some manipulations give

$$\Big(\int_0^\infty w' \|hu'\|^2\,dr\Big)^{1/2} \le (1+\sqrt{2+\alpha})\Big(\int_0^\infty \frac{1}{h^2 w'}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}. \tag{46}$$

By integration by parts, we have

$$\begin{aligned}\int_0^\infty &(r+1)^{-3-\delta}\|u\|^2\,dr \\ &= \frac{2}{2+\delta}\int_0^\infty (r+1)^{-2-\delta}\operatorname{Re}\langle u, u'\rangle\,dr \\ &\le h^{-1}\Big(\int_0^\infty (r+1)^{-1-\delta}\|hu'\|^2\,dr\Big)^{1/2}\Big(\int_0^\infty (r+1)^{-3-\delta}\|u\|^2\,dr\Big)^{1/2},\end{aligned}$$

which implies

$$\Big(\int_0^\infty (r+1)^{-3-\delta}\|u\|^2\,dr\Big)^{1/2} \le h^{-1}\Big(\int_0^\infty (r+1)^{-1-\delta}\|hu'\|^2\,dr\Big)^{1/2}. \tag{47}$$

From (46), (47) and $w' = C_V\delta(C_V+\delta)^{-1}(r+1)^{-1-\delta}$, we obtain

$$\begin{aligned}\Big(\int_0^\infty &(r+1)^{-3-\delta}\|u\|^2\,dr\Big)^{1/2} \\ &\le h^{-2}\Big(\frac{1}{\delta}+\frac{1}{C_V}\Big)(1+\sqrt{2+\alpha})\Big(\int_0^\infty (r+1)^{1+\delta}\|P^\pm(h)u\|^2\,dr\Big)^{1/2}.\end{aligned}$$

Using again the density argument in Appendix C, for $0 < \delta < 1$, we get

$$\big\|(1+r)^{-\frac{3+\delta}{2}}(P(h)-z)^{-1}(1+r)^{-\frac{1+\delta}{2}}\big\|_{L^2\to L^2} \le h^{-2}\big(\delta^{-1}+C_V^{-1}\big)(1+\sqrt{2+\alpha}) \tag{48}$$

on $\{z \in \mathbb{C} : |\operatorname{Im} z| < \alpha \operatorname{Re} z\}$. Then, as above, we can use (44) and the Phragmén–Lindelöf theorem, and take the limit $\alpha \to 0^+$, to obtain

$$\big\|(1+r)^{-\frac{3+\delta}{2}}(P(h)-z)^{-1}(1+r)^{-\frac{1+\delta}{2}}\big\|_{L^2\to L^2} \le h^{-2}\big(\delta^{-1}+C_V^{-1}\big)(1+\sqrt{2}),$$

for $z \in \mathbb{C} \setminus [0, \infty)$. Since the norm of an operator coincides with that of its adjoint, we reach

$$\left\| (1+r)^{-\frac{1+\delta}{2}} (P(h)-z)^{-1} (1+r)^{-\frac{3+\delta}{2}} \right\|_{L^2 \to L^2} \le h^{-2} \big(\delta^{-1} + C_V^{-1} \big) (1+\sqrt{2}),$$

for $z \in \mathbb{C} \setminus [0, \infty)$. The three lines lemma then says that for fixed $z \in \mathbb{C} \setminus [0, \infty)$, the analytic mapping

$$\lambda \mapsto (1+r)^{-\frac{3+\delta}{2}+\lambda} (P(h)-z)^{-1} (1+r)^{-\frac{1+\delta}{2}-\lambda}, \quad 0 < \operatorname{Re} \lambda < 1,$$

with values in the space of bounded operators $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$, obeys

$$\left\| (1+r)^{-\frac{3+\delta}{2}+\theta} (P(h)-z)^{-1} (1+r)^{-\frac{1+\delta}{2}-\theta} \right\|_{L^2 \to L^2} \le h^{-2} \big(\delta^{-1} + C_V^{-1} \big) (1+\sqrt{2}), \quad \theta \in [0, 1]. \tag{49}$$

Having established (49), to finish, we need to see that we can choose δ and θ appropriately to arrive at (12). That is, we need to attain the more general weights characterized by $s_1, s_2 > \frac{1}{2}$ and $s_1 + s_2 > 2$. However, because we have the restrictions $\delta \in (0, 1)$ and $\theta \in [0, 1]$, we first need to make reductions as follows. Since decreasing s_1 or s_2 in (12) increases the left side, it suffices to suppose $s_1, s_2 > \frac{1}{2}$ and $2 < s_1 + s_2 < 3$. Furthermore, by taking the adjoint, it is no restriction to have $s_1 \le s_2$. If we write $s_1 = (1 + 2\delta_2)/2$ for some $\delta_2 > 0$, then we may replace s_2 by $\min(s_2, (3 + \delta_2)/2)$. Having made these reductions, (12) follows from (49) by setting $\delta = s_1 + s_2 - 2 < 1$ and

$$\theta = \tfrac{1}{2}(s_2 - s_1 + 1) \le \tfrac{1}{4}(4 - \delta_2) < 1.$$

Remark 2.2. The bound (38) with $\varepsilon = 0$ rules out $P(h)$ having an eigenvalue $E > 0$. When $V^- = 0$, a zero eigenvalue is ruled out by combining (36) (with $E = \varepsilon = 0$) with (47).

3. Resolvent bounds for wave decay

In this section, we consider the operator $P := P(1) = -\Delta + V$, with $P(h)$ as in (1) and V obeying (2)–(5) and (9). As a consequence of Theorem 1.2, we prove several more resolvent bounds for P, which enable us in Section 4 to establish weighted energy decay for the solution to the wave equation (15). Throughout this section, C denotes a positive constant whose precise value may change, but is always independent of λ, which plays the role of our spectral parameter.

Lemma 3.1. *Fix $s_1, s_2 > \frac{1}{2}$ with $s_1 + s_2 > 2$. There exist $C > 0$ such that for all $\lambda \in \mathbb{C}$ with $0 < |\operatorname{Im} \lambda| \le 1$ and for all multiindices α_1, α_2 with $|\alpha_1| + |\alpha_2| \le 2$,*

$$\left\| \langle x \rangle^{-s_1} \partial_x^{\alpha_1} (P - \lambda^2)^{-1} \partial_x^{\alpha_2} \langle x \rangle^{-s_2} \right\|_{L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)} \le C (1 + |\operatorname{Re} \lambda|)^{|\alpha_1| + |\alpha_2| - 1}. \tag{50}$$

Proof. Since $((P-\lambda^2)^{-1})^* = (P-\bar{\lambda}^2)^{-1}$, to prove (50) is suffices to assume $\operatorname{Im}\lambda > 0$.

First, we treat the case $|\alpha_2| = 0$. Using (11) if $|\operatorname{Re}\lambda| > 1$ or (12) if $|\operatorname{Re}\lambda| \le 1$, we get

$$\big\| \langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2}\big\|_{L^2\to L^2} \le C(1+|\operatorname{Re}\lambda|)^{-1}, \quad 0<\operatorname{Im}\lambda\le 1, \tag{51}$$

Recall from standard elliptic theory that for all $f \in H^2(\mathbb{R}^n)$ and all $\gamma > 0$,

$$\begin{aligned} &\|f\|_{H^2} \le C(\|f\|_{L^2} + \|\Delta f\|_{L^2}),\\ &\|f\|^2_{H^1} \le C\|f\|_{L^2}\|f\|_{H^2} \le C(\gamma^{-1}\|f\|^2_{L^2} + \gamma\|\Delta f\|^2_{L^2}). \end{aligned} \tag{52}$$

Therefore, for any $f \in L^2(\mathbb{R}^n)$,

$$\begin{aligned} &\|\langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{H^2(\mathbb{R}^n)}\\ &\le C(\|\langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{L^2(\mathbb{R}^n)} + \|(-\Delta)\langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{L^2(\mathbb{R}^n)})\\ &\le C(\|\langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{H^1(\mathbb{R}^n)} + \|\langle x\rangle^{-s_1}(-\Delta)(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{L^2(\mathbb{R}^n)})\\ &\le C(\gamma^{-1} + |\operatorname{Re}\lambda|^2)\|\langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{L^2(\mathbb{R}^n)}\\ &\qquad + C\gamma\|\Delta\langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\|_{L^2(\mathbb{R}^n)} + C\|f\|_{L^2(\mathbb{R}^n)}. \end{aligned}$$

Selecting γ sufficiently small depending on C and applying (51) yields

$$\big\| \langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2} f\big\|_{H^2(\mathbb{R}^n)} \le C(1+|\operatorname{Re}\lambda|)\|f\|_{L^2(\mathbb{R}^n)}, \tag{53}$$

as desired. This confirms (50) for $|\alpha_1| = 2$. For $|\alpha_1| = 1$ (still with $|\alpha_2| = 0$), combine (51) and (53) via the second line of (52).

If $|\alpha_2| > 0$, let $f \in C_0^\infty(\mathbb{R}^n)$, and put

$$u = \langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2}\partial_x^{\alpha_2} f.$$

We need to show

$$\|u\|_{H^{|\alpha_1|}} \le C(1+|\operatorname{Re}\lambda|)^{|\alpha_1|+|\alpha_2|-1}\|f\|_{L^2}, \quad H^0 := L^2(\mathbb{R}^n).$$

If $|\alpha_1| = 0$, we use self-adjointness and the already proved estimate

$$\|\langle x\rangle^{-s_2}(P-\lambda^2)^{-1}\langle x\rangle^{-s_1} f\|_{H^j} \le C(1+|\operatorname{Re}\lambda|)^{j-1}\|f\|_{L^2}, \quad j\in\{0,1,2\},$$

to get

$$\begin{aligned} \|u\|^2_{L^2} &= \langle u, \langle x\rangle^{-s_1}(P-\lambda^2)^{-1}\langle x\rangle^{-s_2}\partial_x^{\alpha_2} f\rangle_{L^2}\\ &\le \|\partial_x^{\alpha_2}\langle x\rangle^{-s_2}(P-\bar{\lambda}^2)^{-1}\langle x\rangle^{-s_1} u\|_{L^2}\|f\|_{L^2}\\ &\le C(1+|\operatorname{Re}\lambda|)^{|\alpha_2|-1}\|u\|_{L^2}\|f\|_{L^2}. \end{aligned}$$

If $|\alpha_1| = 1$, we recognize that

$$(P-\lambda^2)u = \langle x\rangle^{-s_1-s_2}\partial_x^{\alpha_2} f + [-\Delta, \langle x\rangle^{-s_1}]\langle x\rangle^{s_1} u.$$

Then multiply by $\bar{u}$, integrate over $\mathbb{R}^n$, and integrate by parts as appropriate:

$$\begin{aligned}\|\nabla u\|_{L^2}^2 &= \int (\lambda^2 - V)|u|^2 - \int \partial_x^{\alpha_2}(\langle x\rangle^{-s_1-s_2}\bar{u}) f + \int \bar{u}[-\Delta, \langle x\rangle^{-s_1}]\langle x\rangle^{s_1} u \\ &\le \lambda^2 \int |u|^2 - \int \partial_x^{\alpha_2}(\langle x\rangle^{-s_1-s_2}\bar{u}) f + \int \bar{u}[-\Delta, \langle x\rangle^{-s_1}]\langle x\rangle^{s_1} u.\end{aligned}$$

Because both $\partial_x^{\alpha_2}\langle x\rangle^{-s_1-s_2}$ and

$$[-\Delta, \langle x\rangle^{-s_1}]\langle x\rangle^{s_1} = (-\Delta\langle x\rangle^{-s_1})\langle x\rangle^{s_1} - 2(\nabla\langle x\rangle^{-s_1})\cdot\nabla\langle x\rangle^{s_1}$$

are first-order differential operators with bounded coefficients, we conclude, for all $\gamma > 0$,

$$\begin{aligned}\|\nabla u\|_{L^2}^2 &\le C_\gamma((1 + |\operatorname{Re}\lambda|)^2\|u\|_{L^2}^2 + \|f\|_{L^2}^2) + \gamma\|\nabla u\|_{L^2}^2 \\ &\le C_\gamma(1 + |\operatorname{Re}\lambda|)^2\|f\|_{L^2}^2 + \gamma\|\nabla u\|_{L^2}^2,\end{aligned}$$

for some $C_\gamma > 0$ depending on γ.

Fixing γ small enough, we absorb the second term on the right side into the left side, confirming (50) when $|\alpha_1| = |\alpha_2| = 1$. □

Next, we prove an estimate for the λ-derivative of the weighted resolvent, which requires the extra short range conditions (16) and (17) on the potential. As input we need the following bound for the weighted square of the free resolvent, which we prove in Appendix E.

Lemma 3.2. *Let $n \ge 3$. Suppose s satisfies* (19). *There exists $C > 0$ such that for all $\lambda \in \mathbb{C}$ with $0 < |\operatorname{Im}\lambda| \le 1$, and any multiindex α such that $|\alpha| \le 1$,*

$$\left\|\lambda\langle x\rangle^{-s}\partial_x^\alpha(-\Delta - \lambda^2)^{-2}\langle x\rangle^{-s}\right\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)} \le C(1 + |\lambda|)^{|\alpha|-1}. \tag{54}$$

Remark 3.3. In [Vo04a], the estimate (54) is stated to hold in any dimension $n \ge 3$ provided $s > \frac{3}{2}$. However, our proof of Lemma 3.2 in dimension $n \ge 4$ needs s larger if (54) is to hold uniformly as $|\lambda| \to 0$. In our approach, we use the integral kernel of $\lambda\langle x\rangle^{-s}(-\Delta - \lambda^2)^{-2}\langle x\rangle^{-s}$ to assess L^2-boundedness as $|\lambda| \to 0$. The kernel is given in terms of the Macdonald function [DLMF, 10.27.4, 10.27.5] of order $(n/2) - 2$, along with other factors. We are able to conclude boundedness on $L^2(\mathbb{R}^n)$ for s as in (19).

Lemma 3.4. *Let $n \ge 3$ and suppose s is as in* (19). *Assume V obeys* (2)–(5) *and* (9), *as well as* (16) *and* (17). *There exists $C > 0$ such that for all $\lambda \in \mathbb{C}$ with $0 < |\operatorname{Im}\lambda| \le 1$, and any $j \in \{0, 1\}$ and multiindex α such that $j + |\alpha| \le 1$,*

$$\left\|\frac{d}{d\lambda}\langle x\rangle^{-s}\lambda^j\partial_x^\alpha(P - \lambda^2)^{-1}\langle x\rangle^{-s}\right\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)} \le C. \tag{55}$$

Proof. Without loss of generality, we take s sufficiently close to (but larger than) $\frac{1}{4}(n+3)$ when $n \ne 8$, or sufficiently close to (but larger than) 3 when $n = 8$, so that by (17) we may fix $s' > \frac{1}{2}$ such that $s + s' < \delta$. With these choices we have

$s' + s > 2$, so we are permitted to apply (50) as the need arises.

We begin from the resolvent identity

$$(P - \lambda^2)^{-1} \langle x \rangle^{-s} (I + K(\lambda)) = R_0(\lambda) \langle x \rangle^{-s}, \tag{56}$$

where

$$K(\lambda) := V(x) \langle x \rangle^{s+s'} \langle x \rangle^{-s'} R_0(\lambda) \langle x \rangle^{-s} \quad \text{and} \quad R_0(\lambda) := (-\Delta - \lambda^2)^{-1}.$$

We know that $\langle x \rangle^{-s'} R_0(\lambda) \langle x \rangle^{-s} : L^2(\mathbb{R}^n) \to H^2(\mathbb{R}^n)$ has a continuous extension from either half-plane ($\pm \operatorname{Im} \lambda > 0$) to $\mathbb{R}$ [GiMo74, Proposition 2.4]. Let us denote this extension by $R^{\pm}_{0,s',s}(\lambda)$ and put $K^{\pm}(\lambda) = V(x) \langle x \rangle^{s+s'} R^{\pm}_{0,s',s}(\lambda)$.

We now show that $K^{\pm}(\lambda)$ is a compact operator $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. To see this, write $K^{\pm}(\lambda)$ as the sum

$$K^{\pm}(\lambda) = (\chi \langle x \rangle^{s+s'} V) R^{\pm}_{0,s',s}(\lambda) \; + \; ((1 - \chi) V \langle x \rangle^{\delta}) \langle x \rangle^{s+s'-\delta} R^{\pm}_{0,s',s}(\lambda).$$

where $\chi \in C_0^\infty(\mathbb{R}^n; [0,1])$ is supported in $B(0,1)$ and $\chi \equiv 1$ near the origin in $\mathbb{R}^n$. The second operator on the right side of the display is compact by [DyZw19, Theorem B.4]). The first operator on the right is compact, as follows: It is the composition of bounded $R^{\pm}_{0,s',s}(\lambda) : L^2(\mathbb{R}^n) \to H^2(\mathbb{R}^n)$ with multiplication by $\chi \langle x \rangle^{s+s'} V$. Due to (3) and Lemma F.1, we have $\|\chi \langle x \rangle^{s+s'} V u\|_{L^2(\mathbb{R}^n)} \leq C \|u\|_{H^1(B(0,1))}$ for some $C > 0$ and all $u \in H^2(\mathbb{R}^n)$. By the Kondrachov embedding theorem the inclusion $H^2(B(0, 1)) \to H^1(B(0, 1))$ is compact. So compactness of $(\chi \langle x \rangle^{s+s'} V) R^{\pm}_{0,s',s}(\lambda)$ holds as desired.

We claim further that $I + K^{\pm}(\lambda) : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is invertible for all λ with $\pm \operatorname{Im} \lambda \geq 0$. By compactness of $K^{\pm}(\lambda)$ and the Fredholm alternative [ReSi80, Theorem VI.14], $I + K^{\pm}(\lambda)$ will be invertible if we can show that for $g \in L^2(\mathbb{R}^n)$, $(I + K^{\pm}(\lambda))g = 0$ implies $g = 0$. To this end, put $u := \langle x \rangle^{s'} R^{\pm}_{0,s',s}(\lambda) g$, which belongs to $\langle x \rangle^{s'} H^2(\mathbb{R}^n)$. If we can show $u = 0$, then in fact $g = 0$. This is because $(-\Delta - \lambda^2) u = \langle x \rangle^{-s} g$ in the distributional sense.

Now let us show $u = 0$. If $\lambda^2 \in \mathbb{C} \setminus [0, \infty)$ (so that $K^{\pm}(\lambda) = K(\lambda)$), this follows immediately from $(P - \lambda^2) u = \langle x \rangle^{-s} g + V R_0(\lambda) \langle x \rangle^{-s} g = \langle x \rangle^{-s} (I + K(\lambda)) g = 0$. If $\lambda^2 \in [0, \infty)$, the idea is the same, but we incorporate a limiting step. Set $u_{\pm,\varepsilon} = (-\Delta - \lambda^2 \pm i\varepsilon)^{-1} \langle x \rangle^{-s} g$. Proposition 2.4 of [GiMo74] implies that $\langle x \rangle^{-s'} u_{\pm,\varepsilon}$ converges to $\langle x \rangle^{-s'} u$ in $H^2(\mathbb{R}^n)$ as $\varepsilon \to 0^+$. We also have

$$\begin{aligned} u_{\pm,\varepsilon} &= (-\Delta - \lambda^2 \pm i\varepsilon)^{-1} \langle x \rangle^{-s} g \\ &= (P - \lambda^2 \pm i\varepsilon)^{-1} (P - \lambda^2 \pm i\varepsilon)(-\Delta - \lambda^2 \pm i\varepsilon)^{-1} \langle x \rangle^{-s} g \\ &= (P - \lambda^2 \pm i\varepsilon)^{-1} \langle x \rangle^{-s} (I + V \langle x \rangle^{s} (-\Delta - \lambda^2 \pm i\varepsilon)^{-1} \langle x \rangle^{-s}) g. \end{aligned}$$

Therefore, by (12),

$$\begin{aligned}\|\langle x\rangle^{-s'}u\|_{L^2} &= \lim_{\varepsilon\to 0^+}\|\langle x\rangle^{-s'}u_{\pm,\varepsilon}\|_{L^2}\\ &\le C\lim_{\varepsilon\to 0^+}\|(I+V\langle x\rangle^{s}(-\Delta-\lambda^2\pm i\varepsilon)^{-1}\langle x\rangle^{-s})g\|_{L^2}\\ &= \|(I+K^{\pm}(\lambda))g\|_{L^2}=0.\end{aligned}$$

Thus we have demonstrated that $I+K^{\pm}(\lambda)$ is invertible for $\pm\operatorname{Im}\lambda\ge 0$. As $\lambda\to\infty$, $\|K(\lambda)\|_{L^2\to L^2}\to 0$ thanks to (50); hence we can compute $(I+K^{\pm}(\lambda))^{-1}$ by a Neumann series, thanks to (50). Therefore

$$\|(I+K^{\pm}(\lambda))^{-1}\|_{L^2\to L^2}\le C. \tag{57}$$

Now for $0<|\operatorname{Im}\lambda|\le 1$ the λ-derivative of (56) is

$$\begin{aligned}&\Big(\frac{d}{d\lambda}\langle x\rangle^{-s}\lambda^{j}\partial_x^{\alpha}(P-\lambda^2)^{-1}\langle x\rangle^{-s}\Big)(I+K(\lambda))\\ &\quad=\frac{d}{d\lambda}\langle x\rangle^{-s}\lambda^{j}\partial_x^{\alpha}R_0(\lambda)\langle x\rangle^{-s}\\ &\qquad-2\langle x\rangle^{-s}\lambda^{j}\partial_x^{\alpha}(P-\lambda^2)^{-1}\langle x\rangle^{-s'}V\langle x\rangle^{s+s'}\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s},\end{aligned} \tag{58}$$

where we used

$$\frac{d}{d\lambda}\langle x\rangle^{-s'}\partial_x^{\alpha}R_0(\lambda)\langle x\rangle^{-s}=2\lambda\langle x\rangle^{-s'}\partial_x^{\alpha}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s},\quad j\in\{0,1\}. \tag{59}$$

The operator norm $L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)$ of the term in the second line of (58) is bounded above by a constant, due to (54) and (59). As for the third line, $\big\|\langle x\rangle^{-s}\lambda^{j}\partial_x^{\alpha}(P-\lambda^2)^{-1}\langle x\rangle^{-s'}\big\|_{L^2\to L^2}\le C$ by (50). Moreover,

$$\big\|V\langle x\rangle^{s+s'}\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}\big\|_{L^2\to L^2}\le C\,\big\|\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}\big\|_{H^1\to L^2},$$

since multiplication by $V\langle x\rangle^{s+s'}$ is a bounded operator $H^1(\mathbb{R}^n)\to L^2(\mathbb{R}^n)$ (see (3) and Lemma F.1). Finally, because $\big\|\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}\big\|_{H^1\to L^2}\le C$ by (54), the proof of (55) is complete. □

4. Proof of Theorem 1.6

In this section we prove Theorem 1.6 by combining the resolvent bounds of the previous section with an argument from [Vo04a, Section 3]. As before we set $P=-\Delta+V:L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)$, $n\ge 3$, where V obeys (2)–(9), (16) and (17).

In several steps below, we use that for all $0\le\alpha\le 1$ there exists $C>0$ such that, for any $f\in H^1(\mathbb{R}^n)$,

$$\|V^{\alpha}f\|_{L^2}^2\le C(\|\nabla f\|_{L^2}^2+\|f\|_{L^2}^2)\le C\|\nabla f\|_{L^2}^2. \tag{60}$$

The first inequality follows from (3), (16) and Lemma F.1, while the second follows from the Poincaré inequality (as we work in dimension $n\ge 3$).

Given $s>0$ and u as in (18) solving the wave equation (15), with compactly supported initial conditions $u(0,x)=u_0(x)\in H^1(\mathbb{R}^n)$, $\partial_t u(0,x)=u_1(x)\in L^2(\mathbb{R}^n)$, define

$$E_s(t) := \int_{\mathbb{R}^n} \langle x\rangle^{-2s}(|\partial_t u(t,x)|^2 + |\nabla u(t,x)|^2 + |u(t,x)|^2)\,dx,$$
$$E(0) := \|\nabla u_0\|_{L^2}^2 + \|u\|_{L^2}^2.$$

Lemma 4.1. *If $s > \frac{1}{2}$ and V satisfies* (2)–(9), *there exists $C > 0$ such that*

$$\int_0^\infty E_s(\tau)\,d\tau \le C E(0). \tag{61}$$

If in addition s satisfies (19) *and V* (16) *and* (17), *there exists $C > 0$ such that*, *for $t \ge 1$*,

$$\int_t^\infty E_s(\tau)\,d\tau \le C t^{-2} E(0). \tag{62}$$

Proof. Choose $\phi \in C^\infty(\mathbb{R})$ with $\phi \ge 0$, $\phi(t) = 0$ near $\left(-\infty, \frac{1}{2}\right]$, and $\phi(t) = 1$ near $[1,\infty)$. Since $(\partial_t^2 + P)u = 0$, where $P = -\Delta + V$, we have

$$(\partial_t^2 + P)\phi u = (\phi'' + 2\phi'\partial_t)u := v(t). \tag{63}$$

By Duhamel's formula for the solution to an inhomogeneous wave equation with zero initial conditions,

$$\phi u(t) = \int_0^t \frac{\sin(t-\tau)\sqrt{P}}{\sqrt{P}} v(\tau)\,d\tau.$$

On the other hand,

$$(P - (\lambda - i\varepsilon)^2)^{-1} = \int_0^\infty e^{-it(\lambda - i\varepsilon)} \frac{\sin(t\sqrt{P})}{\sqrt{P}} dt, \quad \varepsilon > 0.$$

It follows from the last two identities that the Fourier transform $\widehat{\phi u}$ of ϕu satisfies

$$\widehat{\phi u}(\lambda - i\varepsilon) := \int_{-\infty}^\infty e^{-it(\lambda - i\varepsilon)}\phi(t) u(\,\cdot\,, t)\,dt = (P - (\lambda - i\varepsilon)^2)^{-1}\hat{v}(\lambda - i\varepsilon). \tag{64}$$

By finite propagation speed for the wave equation, and because $v(t)$ is compactly supported in t, $\operatorname{supp}_x v(t)$, and thus also $\operatorname{supp}_x \hat{v}(\lambda)$, is contained in some compact subset of $\mathbb{R}^n$ independent of t. Choose $\eta \in C_0^\infty(\mathbb{R}^n)$ such that $\eta = 1$ near $\operatorname{supp}_x v(t)$ for all $t \in \mathbb{R}$. By (64),

$$\langle x\rangle^{-s}\widehat{\phi u}(\lambda - i\varepsilon) = \langle x\rangle^{-s}(P - (\lambda - i\varepsilon)^2)^{-1}\eta\hat{v}(\lambda - i\varepsilon),$$
$$\langle x\rangle^{-s}\widehat{\partial_t(\phi u)}(\lambda - i\varepsilon) = \langle x\rangle^{-s}(\lambda - i\varepsilon)(P - (\lambda - i\varepsilon)^2)^{-1}\eta\hat{v}(\lambda - i\varepsilon),$$
$$\langle x\rangle^{-s}\nabla\widehat{\phi u}(\lambda - i\varepsilon) = \langle x\rangle^{-s}\nabla(P - (\lambda - i\varepsilon)^2)^{-1}\eta\hat{v}(\lambda - i\varepsilon).$$

Therefore, by (50), for $s > \frac{1}{2}$ and V obeying (2)–(9), there is $C > 0$ independent of λ and ε and such that for all $\lambda \in \mathbb{R}$, $0 < \varepsilon \le 1$, we have

$$\left\| \frac{d^k}{d\lambda^k} \langle x\rangle^{-s} \widehat{\partial_t(\phi u)}(\lambda - i\varepsilon)\right\|_{L^2} + \left\| \frac{d^k}{d\lambda^k} \langle x\rangle^{-s} \nabla\widehat{\phi u}(\lambda - i\varepsilon)\right\|_{L^2} + \left\| \frac{d^k}{d\lambda^k} \langle x\rangle^{-s} \widehat{\phi u}(\lambda - i\varepsilon)\right\|_{L^2} \le C\|\hat{v}(\lambda - i\varepsilon)\|_{L^2} + Ck\|\widehat{tv}(\lambda - i\varepsilon)\|_{L^2} \tag{65}$$

for $k = 0$. If in addition we suppose s satisfies (19) and V satisfies (16) and (17), then by (55), (65) holds for $k \in \{0, 1\}$. Here when $k = 1$ we used the product rule and the identity $d\hat{v}(\lambda - i\varepsilon)/d\lambda = -i\widehat{tv}(\lambda - i\varepsilon)$.

Next, by (65) and Plancherel's theorem, there exist $C_1, C_2, C_3, C > 0$ independent of ε and such that

$$\begin{aligned} &\int_{-\infty}^{\infty} (\|\langle x\rangle^{-s}\partial_t(\phi u)\|^2_{L^2} + \|\langle x\rangle^{-s}\nabla(\phi u)\|^2_{L^2} + \|\langle x\rangle^{-s}\phi u\|^2_{L^2})e^{-2\varepsilon t}\,dt \\ &= C_1 \int_{-\infty}^{\infty} (\|\langle x\rangle^{-s}\widehat{\partial_t(\phi u)}(\lambda - i\varepsilon)\|^2_{L^2} + \|\langle x\rangle^{-s}\nabla\widehat{\phi u}(\lambda - i\varepsilon)\|^2_{L^2} \\ &\qquad + \|\langle x\rangle^{-s}\widehat{\phi u}(\lambda - i\varepsilon)\|^2_{L^2})\,d\lambda \\ &\le C_2 \int_{-\infty}^{\infty} \|\hat{v}(\lambda - i\varepsilon)\|^2_{L^2} d\lambda = C_3 \int_{-\infty}^{\infty} \|v(t)\|^2_{L^2} e^{-2\varepsilon t}\,dt \le C \sup_{t\in\mathbb{R}} \|v(t)\|^2. \end{aligned} \tag{66}$$

The last constant C is independent of ε because $v(t)$ has compact support in t; see (63). The proof of (61) is completed by sending $\varepsilon \to 0$ in (66) and observing that

$$\begin{aligned} \|v(t)\|_{L^2} &\le C(\|u_0\|_{L^2} + \|\sqrt{P}u_0\|_{L^2} + \|u_1\|_{L^2}) \\ &\le C(\|\nabla u_0\|_{L^2} + \|u_1\|_{L^2}) = C\sqrt{E(0)}. \end{aligned} \tag{67}$$

In the last inequality we used that for any $f \in H^2(\mathbb{R}^n)$ (and thus any $f \in H^1(\mathbb{R}^n)$, since $H^2(\mathbb{R}^n)$ is dense in $H^1(\mathbb{R}^n)$),

$$\|\sqrt{P}f\|^2_{L^2} = \langle f, Pf\rangle_{L^2} = \|\nabla f\|^2_{L^2} + \|\sqrt{V}f\|^2_{L^2} \le C\|\nabla f\|^2_{L^2},$$

with the inequality being due to (60).

To prove (62), we again use Plancherel's theorem with (65), so that for all $0 < \varepsilon \le 1$ and $T \ge 1$,

$$\begin{aligned} &T^2 \int_T^{\infty} \big(\|\langle x\rangle^{-s}\partial_t(\phi u)\|^2_{L^2} + \|\langle x\rangle^{-s}\nabla(\phi u)\|^2_{L^2} + \|\langle x\rangle^{-s}\phi u\|^2_{L^2}\big)e^{-2\varepsilon t}\,dt \\ &\le \int_{-\infty}^{\infty} \big(\|\langle x\rangle^{-s}t\partial_t(\phi u)\|^2_{L^2} + \|\langle x\rangle^{-s}t\nabla(\phi u)\|^2_{L^2} + \|\langle x\rangle^{-s}t\phi u\|^2_{L^2}\big)e^{-2\varepsilon t}\,dt \\ &= C_1 \int_{-\infty}^{\infty} \Big(\Big\| \frac{d}{d\lambda}\langle x\rangle^{-s}\widehat{\partial_t(\phi u)}(\lambda - i\varepsilon)\Big\|^2_{L^2} + \Big\| \frac{d}{d\lambda}\langle x\rangle^{-s}\nabla\widehat{\phi u}(\lambda - i\varepsilon)\Big\|^2_{L^2} \\ &\qquad + \Big\| \frac{d}{d\lambda}\langle x\rangle^{-s}\widehat{\phi u}(\lambda - i\varepsilon)\Big\|^2_{L^2}\Big)\,d\lambda \\ &\le C_2 \int_{-\infty}^{\infty} \big(\|\hat{v}(\lambda - i\varepsilon)\|^2_{L^2} + \|\widehat{tv}(\lambda - i\varepsilon)\|^2_{L^2}\big)\,d\lambda \\ &= C_3 \int_{-\infty}^{\infty} \big(\|v(t)\|^2_{L^2} + \|tv(t)\|^2_{L^2}\big)\,e^{-2\varepsilon t}\,dt \le C \sup_{t\in\mathbb{R}} \|v(t)\|^2 \le CE(0). \end{aligned} \tag{68}$$

Once again sending $\varepsilon \to 0^+$ concludes the proof of (62). □

The local energy decay (20) follows from (62) and the following:

Lemma 4.2. *If* $s > 0$ *and* $V \geq 0$ *satisfies* (3) *and* (4), *there exists* $C > 0$ *such that, for all* $t \geq 1$,

$$E_s(t) \leq C \int_t^\infty E_s(\tau)\, d\tau. \tag{69}$$

Proof. The strategy is the same as that of [Vo04a, Lemma 3.2]. Computing $\frac{d}{dt} E_s(t)$, one finds

$$\begin{aligned}\frac{d}{dt} E_s(t) = -2\,\mathrm{Re} \int_{\mathbb{R}^n} \partial_r u(t,x) \overline{\partial_t u(t,x)} \partial_r \langle x\rangle^{-2s}\, dx \\ + 2\,\mathrm{Re} \int_{\mathbb{R}^n} (-V u(t,x) \overline{\partial_t u(t,x)} + u(t,x) \overline{\partial_t u(t,x)}) \langle x\rangle^{-2s}\, dx.\end{aligned} \tag{70}$$

By (60),

$$\begin{aligned}\|V\langle x\rangle^{-s} u(t,x)\|_{L^2} &\leq C\|\nabla \langle x\rangle^{-s} u(t,x)\|_{L^2} \\ &\leq C\,\|\langle x\rangle^{-s} \nabla u(t,x)\|_{L^2} + C\,\|\langle x\rangle^{-s} u(t,x)\|_{L^2}.\end{aligned}$$

for $C > 0$ independent of t, and whose precise value may change between lines. Thus we can bound the right side of (70) from above by Cauchy–Schwarz:

$$\begin{aligned}&\frac{d}{dt} E_s(t) \\ &\leq C \left\| \langle x\rangle^{-s} \partial_r u(t,x) \right\|_{L^2} \left\| \langle x\rangle^{-s} \partial_t u(t,x) \right\|_{L^2} + C \left\| V\langle x\rangle^{-s} u(t,x) \right\|_{L^2} \left\| \langle x\rangle^{-s} \partial_t u(t,x) \right\|_{L^2} \\ &\qquad + C \left\| \langle x\rangle^{-s} u(t,x) \right\|_{L^2} \left\| \langle x\rangle^{-s} \partial_t u(t,x) \right\|_{L^2} \\ &\leq C E_s(t).\end{aligned}$$

We then have, for all $T > t \geq 1$,

$$E_s(t) \leq E_s(T) + C_s \int_t^T E_s(\tau)\, d\tau. \tag{71}$$

From (61), we also have a sequence $T_j \to \infty$ such that $\lim_{T_j \to \infty} E_s(T_j) = 0$. So setting $T = T_j$ in (71) and sending $T_j \to \infty$ completes the proof of (69). □

Appendix A. Proof of Lemma 2.1

First we check that

$$C_0^\infty(0,\infty) \ni \varphi \mapsto \int_{\mathbb{S}^{n-1}} \int_0^\infty \varphi(r) |u(r,\theta)|^2\, dV(r,\theta)\, d\theta \tag{72}$$

(part of the expression in (27)) is well defined as a distribution on $(0,\infty)$. Indeed, by (7), for each $\theta \in \mathbb{S}^{n-1}$,

$$\int_0^\infty \varphi(r)|u(r,\theta)|^2\, dV(r,\theta) = -\int_0^\infty V(r,\theta)(|u(r,\theta)|^2 \varphi(r))' dr.$$

By Fubini's theorem the expression on the right side belongs to $L^1(\mathbb{S}^{n-1})$. Hence the quantity in (72) is well defined as a distribution on $(0, \infty)$.

Next we demonstrate that the function $r \mapsto \int_{\mathbb{S}^{n-1}} V^R(r,\theta)\,|u(r,\theta)|^2\,d\theta$ has locally bounded variation. Suppose $[a,b] \subseteq (0,\infty)$ and $a = r_0 < r_1 < \cdots < r_N = b$. We have

$$\begin{aligned}\sum_{k=1}^{N} \Big| \int_{\mathbb{S}^{n-1}} V^R(r_k,\theta)\,|u(r_k,\theta)|^2\,d\theta - \int_{\mathbb{S}^{n-1}} V^R(r_{k-1},\theta)\,|u(r_{k-1},\theta)|^2\,d\theta \Big| \\ \le \int_{\mathbb{S}^{n-1}} \sum_{k=1}^{N} |u(r_k,\theta)|^2\,|V^R(r_k,\theta) - V^R(r_{k-1},\theta)|\,d\theta \\ + \int_{\mathbb{S}^{n-1}} \sum_{k=1}^{N} |V^R(r_{k-1},\theta)|\,\big||u(r_k,\theta)|^2 - |u(r_{k-1},\theta)|^2\big|\,d\theta.\end{aligned}$$

For each θ, the function $r \mapsto V^R(r,\theta)$ is decreasing, so

$$\begin{aligned}\int_{\mathbb{S}^{n-1}} \sum_{k=1}^{N} |u(r_k,\theta)|^2 |V^R(r_k,\theta) - V^R(r_{k-1},\theta)|\,d\theta \\ \le \|u\|^2_{L^\infty([a,b]\times\mathbb{S}^{n-1})} \int_{\mathbb{S}^{n-1}} \sum_{k=1}^{N} |V^R(r_k,\theta) - V^R(r_{k-1},\theta)|\,d\theta \\ \le \|u\|^2_{L^\infty([a,b]\times\mathbb{S}^{n-1})} \int_{\mathbb{S}^{n-1}} (V^R(a,\theta) - V^R(b,\theta))\,d\theta \\ \le \omega_{n-1} \|u\|^2_{L^\infty([a,b]\times\mathbb{S}^{n-1})} \|V\|_{L^\infty([a,b]\times\mathbb{S}^{n-1})},\end{aligned}$$

where ω_n is the $(n-1)$-dimensional volume of $\mathbb{S}^{n-1}$. On the other hand,

$$\begin{aligned}\sum_{k=1}^{N} \int_{\mathbb{S}^{n-1}} |V^R(r_{k-1},\theta)|\,\big||u(r_k,\theta)|^2 - |u(r_{k-1},\theta)|^2\big|\,d\theta \\ \le \|V\|_{L^\infty([a,b]\times\mathbb{S}^{n-1})} \sum_{k=1}^{N} \int_{\mathbb{S}^{n-1}} \int_{r_{k-1}}^{r_k} |\partial_r |u(r,\theta)|^2|\,dr\,d\theta \\ \le \omega_{n-1}(b-a) \|V\|_{L^\infty([a,b]\times\mathbb{S}^{n-1})}\, \|\partial_r |u|^2\|_{L^\infty([a,b]\times\mathbb{S}^{n-1})}.\end{aligned}$$

Taken together, the previous two estimates show that $r \mapsto \int_{\mathbb{S}^{n-1}} V^R(r,\theta)|u(r,\theta)|^2\,d\theta$ has locally bounded variation.

We finish by confirming (27). Let $\varphi \in C_0^\infty(0,\infty)$. In the following chain of equalities, the first uses Fubini's theorem, the second arises because for each $\theta \in \mathbb{S}^{n-1}$, $V^R(\,\cdot\,,\theta) = V(\,\cdot\,,\theta)$ almost everywhere with respect to the measure dr, and the last uses (7):

$$\begin{aligned}&-\int_0^\infty \varphi'(r) \int_{\mathbb{S}^{n-1}} V^R(r,\theta)|u(r,\theta)|^2\,d\theta\,dr \\ &\quad= -\int_{\mathbb{S}^{n-1}} \int_0^\infty \varphi'(r) V^R(r,\theta)|u(r,\theta)|^2\,dr\,d\theta \\ &\quad= -\int_{\mathbb{S}^{n-1}} \int_0^\infty \varphi'(r) V(r,\theta)|u(r,\theta)|^2\,dr\,d\theta \\ &\quad= -\int_{\mathbb{S}^{n-1}} \int_0^\infty \big((\varphi(r)\,|u(r,\theta)|^2)' - \varphi(r)\,2\,\mathrm{Re}(\bar{u}(r,\theta)u'(r,\theta))\big)\,V(r,\theta)\,dr\,d\theta\end{aligned}$$

$$= \int_{\mathbb{S}^{n-1}} \int_0^\infty \varphi'(r)|u(r,\theta)|^2 \, dV(r,\theta)\, d\theta + 2\int_{\mathbb{S}^{n-1}} \int_0^\infty \varphi(r)\,\mathrm{Re}(\bar{u}(r,\theta)u'(r,\theta)))V(r,\theta)\, dr\, d\theta.$$

Appendix B. The Phragmén–Lindelöf theorem

Let $f(z)$ be a holomorphic function in a domain D of the complex plane with boundary Γ. We say that $f(z)$ does not exceed a number $M \geq 0$ in modulus at a boundary point $\zeta \in \Gamma$ if $\limsup_{z\to\zeta,\, z\in D} |f(z)| \leq M$.

Theorem B.1 (Phragmén–Lindelöf [EM]). *Suppose that $E \subseteq \Gamma$ and that $f : D \to \mathbb{C}$ is analytic and does not exceed M in modulus at any point of $\Gamma \setminus E$. Suppose also there is a function $g(z)$ with the following properties*:

(1) *$g(z)$ is analytic in D.*

(2) *$|g(z)| < 1$ in D.*

(3) *$g(z) \neq 0$ in D.*

(4) *For every $\sigma > 0$, the function $|g(z)|^\sigma\, |f(z)|$ does not exceed M in modulus at any $\zeta \in E$.*

Then $|f(z)| \leq M$ everywhere in D.

Appendix C. Density argument: proof of (42) and (48)

Estimates (42) and (48) are consequences of the following fact:

Lemma C.1. *Fix $h,\ s_1 > 0$, $0 < s_2 < 1$, and $z \in \mathbb{C} \setminus [0,\infty)$. Let $P(h)$ be as in* (1) *with $V : \mathbb{R}^n \to \mathbb{R}$ obeying* (3) *and* (4) (*so that $P(h)$ is self-adjoint with respect to the domain $H^2(\mathbb{R}^n)$*). *Suppose there exists $C > 0$ so that for all $v \in C_0^\infty(\mathbb{R}^n)$,*

$$\|\langle x\rangle^{-s_1} v\|_{L^2}^2 \leq C \|\langle x\rangle^{s_2}(P(h)-z)v\|_{L^2}^2. \tag{73}$$

Then

$$\|\langle x\rangle^{-s_1}(P(h)-z)^{-1}\langle x\rangle^{-s_2}\|_{L^2\to L^2} \leq C. \tag{74}$$

Proof. The operator

$$[P(h), \langle x\rangle^{s_2}]\langle x\rangle^{-s_2} = \left(-h^2(\Delta\langle x\rangle^{s_2}) - 2h^2(\nabla\langle x\rangle^{s_2})\cdot\nabla\right)\langle x\rangle^{-s_2}$$

is bounded $H^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. So, for $v \in H^2(\mathbb{R}^n)$ such that $\langle x\rangle^{s_2} v \in H^2(\mathbb{R}^n)$,

$$\begin{aligned} \|\langle x\rangle^{s_2}(P(h)-z)v\|_{L^2} &\leq \|(P(h)-z)\langle x\rangle^{s_2} v\|_{L^2} + \|[P(h), \langle x\rangle^{s_2}]\langle x\rangle^{-s_2}\langle x\rangle^{s_2} v\|_{L^2} \\ &\leq C_{z,h}\|\langle x\rangle^{s_2} v\|_{H^2}, \end{aligned} \tag{75}$$

for some constant $C_{z,h} > 0$ depending on z and h.

Given $f \in L^2(\mathbb{R}^n)$, the function $u = \langle x\rangle^{s_2}(P(h)-z)^{-1}\langle x\rangle^{-s_2} f \in H^2(\mathbb{R}^n)$ because

$$u = (P(h)-z)^{-1}(f+w), \quad w = [P(h), \langle x\rangle^{s_2}]u,$$

with $[P(h), \langle x\rangle^{s_2}]$ being bounded $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ since $s_2 < 1$.

Now, choose a sequence $v_k \in C_0^\infty$ such that $v_k \to \langle x\rangle^{s_2}(P(h)-z)^{-1}\langle x\rangle^{-s_2} f$ in $H^2(\mathbb{R}^n)$. Define $\tilde{v}_k := \langle x\rangle^{-s_2} v_k$. Then, as $k \to \infty$,

$$\big\| \langle x\rangle^{-s_1}\tilde{v}_k - \langle x\rangle^{-s_1}(P(h)-z)^{-1}\langle x\rangle^{-s_2} f \big\|_{L^2} \le \big\| v_k - \langle x\rangle^{s_2}(P(h)-z)^{-1}\langle x\rangle^{-s_2} f \big\|_{H^2},$$

which tends to 0. Also, applying (75),

$$\|\langle x\rangle^{s_2}(P(h)-z)\tilde{v}_k - f\|_{L^2} \le C_{z,h}\|v_k - \langle x\rangle^{s_2}(P(h)-E\pm i\varepsilon)^{-1}\langle x\rangle^{-s_2} f\|_{H^2} \to 0.$$

Thus (74) follows by replacing v by $\tilde{v}_k$ in (73) and sending $k \to \infty$. □

Appendix D. Justification of Remark 1.4

In the setting of Theorem 1.2, consider the case of $V = 0$ and $n = 3$. In that scenario the integral kernel of $(P(h)-z)^{-1} = (-h^2\Delta - z)^{-1}$ with $z \in \mathbb{C}\setminus[0,\infty)$ is given by

$$R_0(x,y,z) := h^{-2}\frac{e^{i\frac{\sqrt{z}}{h}|x-y|}}{4\pi|x-y|}, \quad \operatorname{Im}\sqrt{z} > 0.$$

Looking at the operator

$$\langle\cdot\rangle^{-s_1}(-h^2\Delta - z)^{-1}\langle\cdot\rangle^{-s_2} : L^2(\mathbb{R}^3) \to L^2(\mathbb{R}^3) \tag{76}$$

we recall why having a bound on it like (11) requires $s_1,\ s_2 > \frac{1}{2}$.

Since the norm of an operator and its adjoint coincide, it suffices to show that $s_1 > \frac{1}{2}$ is necessary. Use $\sqrt{z}$ of the form $\sqrt{z} = E + i\varepsilon$ for $E > 0$ fixed and $\varepsilon > 0$ tending to zero. As calculated in the proof of [DyZw19, Theorem 3.5], for $f \in C_0^\infty(\mathbb{R}^3)$,

$$\begin{aligned}\langle x\rangle^{-s_1}\int_{\mathbb{R}^3} R_0(x,y,z) f(y)\,dy \\ = h^{-2}\frac{\langle x\rangle^{-s_1}}{4\pi|x|} e^{\frac{i}{h}(E+i\varepsilon)|x|}\Big(\hat{f}\Big(\frac{E}{h}\frac{x}{|x|}\Big) + o(1)\Big) + O(|x|^{-2})\end{aligned} \tag{77}$$

as $\varepsilon \to 0^+$ and $|x| \to \infty$. If f is chosen so that $|\hat{f}| > c$ for some $c > 0$ on $\{|x| = E/h\}$, then (77) and $s_1 \le \frac{1}{2}$ imply $\|\langle x\rangle^{-s_1}\int_{\mathbb{R}^3} R_0(x,y,z)f(y)\,dy\|_{L^2} \to \infty$ as $\varepsilon \to 0^+$.

Next, supposing $s_1,\ s_2 > \frac{1}{2}$, we show why a bound like (12) on (76) requires additionally that $s_1 + s_2 \ge 2$, which is nearly the condition we impose for (12). Using $\sqrt{z} = i\varepsilon$ for $\varepsilon > 0$ tending to zero, and $f_\eta(y) = \langle y\rangle^{-\eta-\frac{3}{2}}$, $\eta > 0$, we see as in [BoHa10, proof of Remark 2] that

$$\begin{aligned}\langle x\rangle^{-s_1}\int_{\mathbb{R}^3} &R_0(x,y,z)\langle y\rangle^{-s_2} f(y)\\ &= h^{-2}\langle x\rangle^{-s_1}\int_{\mathbb{R}^3}\frac{e^{-\frac{\varepsilon}{h}|x-y|}}{4\pi|x-y|}\langle y\rangle^{-s_2} f(y)\,dy\\ &\gtrsim h^{-2}e^{-\frac{3\varepsilon}{2h}|x|}\langle x\rangle^{-s_1-1}\int_{|y|\le\frac{|x|}{2}}\langle y\rangle^{-s_2-\eta-\frac{3}{2}}\,dy \gtrsim h^{-2}e^{-\frac{3\varepsilon}{2h}|x|}\langle x\rangle^{-s_1-s_2-\eta+\frac{1}{2}},\end{aligned}$$

where the implicit constants indicated by $\gtrsim$ are independent of ε and η. First sending $\varepsilon\to 0^+$ gives $s_1+s_2\ge 2-\eta$, but since $\eta>0$ is arbitrary, we in turn get $s_1+s_2\ge 2$.

To see that the $O(|z|^{-\frac{1}{2}}h^{-1})$-dependence of the right side of (11) is optimal, consider the function $u=e^{i\frac{\sqrt{z}}{h}x_1}\chi$ for nontrivial $\chi\in C_0^\infty(\mathbb{R}^3;[0,1])$. We have

$$\langle x\rangle^s(-h^2\Delta-z)u=-i\sqrt{z}h\langle x\rangle^s\partial_{x_1}\chi-h^2e^{i\frac{\sqrt{z}}{h}x_1}\langle x\rangle^s\Delta\chi=:f,$$

whence $\langle\cdot\rangle^{-s}(-h^2\Delta-z)^{-1}\langle\cdot\rangle^{-s}f=\langle\cdot\rangle^{-s}u$ and thus, as $h\to 0$,

$$\frac{\|\langle\cdot\rangle^{-s}(-h^2\Delta-z)^{-1}\langle\cdot\rangle^{-s}f\|_{L^2}}{\|f\|_{L^2}}=\frac{\|\langle\cdot\rangle^{-s}u\|_{L^2}}{\|f\|_{L^2}}\gtrsim|z|^{-\frac{1}{2}}h^{-1}.$$

Finally, we argue why the $O(h^{-2})$-dependence of the right-hand side of (12) is sharp. As noted before, Proposition 2.4 or [GiMo74] gives that $R_{0,s_1,s_2}(\lambda)=\langle\cdot\rangle^{-s_1}(-h^2\Delta-\lambda^2)^{-1}\langle\cdot\rangle^{-s_2}$, with $s_1,\ s_2>\frac{1}{2}$ and $s_1+s_2>2$, extends continuously from $\operatorname{Im}\lambda>0$ to $\mathbb{R}$ in the space of bounded operators $\mathbb{R}^3\to\mathbb{R}^3$. In this case,

$$h^{-2}\Big\|\langle x\rangle^{-s_1}\int_{\mathbb{R}^3}\frac{1}{4\pi|x-y|}\langle y\rangle^{-s_2}dy\Big\|_{L^2\to L^2}=\lim_{\varepsilon\to 0}\|R_{0,s_1,s_2}(i\varepsilon)\|_{L^2\to L^2}.$$

Appendix E. Proof of Lemma 3.2

Lemma 3.2. *Let* $n\ge 3$. *Suppose* s *satisfies* (19). *There exists* $C>0$ *such that for all* $\lambda\in\mathbb{C}$ *with* $0<|\operatorname{Im}\lambda|\le 1$, *and any multiindex* α *such that* $|\alpha|\le 1$,

$$\big\|\lambda\langle x\rangle^{-s}\partial_x^\alpha(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}\big\|_{L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)}\le C(1+|\lambda|)^{|\alpha|-1}. \tag{78}$$

Proof of Lemma 3.2. Initially we take $|\alpha|=0$ in (54), so may assume without loss of generality that $\operatorname{Im}\lambda>0$. We treat the $|\alpha|=1$ case at the end of the proof. Observe that

$$\frac{d}{d\lambda}\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-1}\langle x\rangle^{-s}=2\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s},$$

so we can bound the $L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)$ norm of either quantity.

We now consider two cases for $|\lambda|$.

Case $|\lambda| \geq 1$. If we assume $|\lambda| \geq 1$, then

$$(79)\quad -\lambda^2\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s} \\ = -\tfrac{1}{2}\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-1}(-2\Delta)(-\Delta-\lambda^2)^{-1}\langle x\rangle^{-s} + \langle x\rangle^{-s}(-\Delta-\lambda^2)^{-1}\langle x\rangle^{-s}.$$

By (50), the $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ norm of the second line of (79) is bounded by $C(1+|\operatorname{Re}\lambda|)^{-1}$. So it suffices to investigate the first summand on the right side of (79). For notational brevity, put $R_0(\lambda) := (-\Delta-\lambda^2)^{-1}$. We show that, for all $f \in C_0^\infty(\mathbb{R}^n)$,

$$(80)\quad \langle x\rangle^{-s}R_0(\lambda)(-2\Delta)R_0(\lambda)\langle x\rangle^{-s}f \\ = -\langle x\rangle^{-s}R_0(\lambda)\partial_r(r\langle x\rangle^{-s}f) + \langle x\rangle^{-s}R_0(\lambda)\langle x\rangle^{-s}f + \langle x\rangle^{-s}r\partial_r R_0(\lambda)\langle x\rangle^{-s}f.$$

Since $s > \frac{3}{2}$ by (19), the $L^2(\mathbb{R}^n)$-norm of the right side of (80) is bounded by $C\|f\|_{L^2}$, thanks to (50). So it remains to show (80).

Recall the formula

$$\Delta = \partial_r^2 + (n-1)r^{-1}\partial_r + r^{-2}\Delta_{\mathbb{S}^{n-1}}$$

for the Laplacian in polar coordinates, which implies the commutator identity

$$(81)\qquad [r\partial_r, \Delta] := r\partial_r(\Delta) - \Delta(r\partial_r) = -2\Delta.$$

Fix $f \in C_0^\infty(\mathbb{R}^n)$, $g \in L^2(\mathbb{R}^n)$, and put

$$u := R_0(\lambda)\langle x\rangle^{-s}f \in H^2(\mathbb{R}^n).$$

Let $\{u_k\}_{k=1}^\infty \subseteq C_0^\infty(\mathbb{R}^n)$ be a sequence converging to u in $H^2(\mathbb{R}^n)$. Starting from the left side of (80) and applying (81),

$$(82)\quad \langle g, \langle x\rangle^{-s}R_0(\lambda)(-2\Delta)R_0(\lambda)\langle x\rangle^{-s}f\rangle_{L^2} \\ = \lim_{k\to\infty}\langle g, \langle x\rangle^{-s}R_0(\lambda)[r\partial_r, \Delta]u_k\rangle_{L^2} \\ = \langle g, \langle x\rangle^{-s}r\partial_r u\rangle_{L^2} - \lim_{k\to\infty}\big\langle g, \langle x\rangle^{-s}R_0(\lambda)r\partial_r(-\Delta-\lambda^2)u_k\big\rangle_{L^2}.$$

The purpose of the following calculations is to show that the limit on the last line of (82) equals $\langle g, \langle x\rangle^{-s}R_0(\lambda)r\partial_r\langle x\rangle^{-s}f\rangle_{L^2}$. First, for any $v \in L^2(\mathbb{R}^n)$, we have $rR_0(\lambda)\langle x\rangle^{-1}v \in H^1(\mathbb{R}^n)$, for the following reason. Setting $w := \langle x\rangle R_0(\lambda)\langle x\rangle^{-1}v$, then $rR_0(\lambda)\langle x\rangle^{-1}v = r\langle x\rangle^{-1}w$ and

$$(-\Delta-\lambda^2)w = [-\Delta, \langle x\rangle]R_0(\lambda)\langle x\rangle^{-1}v + v \implies \\ w = R_0(\lambda)\big([-\Delta, \langle x\rangle]R_0(\lambda)\langle x\rangle^{-1}v + v\big) \in H^2(\mathbb{R}).$$

Furthermore, for any $w, v \in C_0^\infty(\mathbb{R}^n)$,

$$\langle w, \partial_r v\rangle_{L^2} = \langle \partial_r^* w, v\rangle_{L^2} := (1-n)\langle r^{-1}w, v\rangle_{L^2} - \langle \partial_r w, v\rangle_{L^2}.$$

Hence, by the density of $C_0^\infty(\mathbb{R}^n)$ in $H^1(\mathbb{R}^n)$, and setting $\tilde{u}_k = (-\Delta - \lambda)u_k$, we get

$$\begin{aligned}(83)\qquad \lim_{k\to\infty}\big\langle g, \langle x\rangle^{-s} R_0(\lambda)(r\partial_r)\tilde{u}_k\big\rangle_{L^2} &= \lim_{k\to\infty}\langle(\partial_r)^* r R_0(\bar\lambda)\langle x\rangle^{-s} g, \tilde{u}_k\rangle_{L^2}\\ &= \langle(\partial_r)^* r R_0(\bar\lambda)\langle x\rangle^{-s} g, \langle x\rangle^{-s} f\rangle_{L^2}\\ &= \langle g, \langle x\rangle^{-s} R_0(\lambda) r\partial_r \langle x\rangle^{-s} f\rangle_{L^2}.\end{aligned}$$

as desired. Taken together, (82) and (83) confirm (80).

Case $|\lambda| \le 1$. Now we turn to the case $|\lambda| \le 1$, and utilize the integral kernel of the free resolvent, which is given by [JeNe01, Section 3],

$$(84)\quad (-\Delta-\lambda^2)^{-1}(|x-y|) = \frac{1}{2\pi}\left(\frac{-i\lambda}{2\pi|x-y|}\right)^{\frac{n}{2}-1} K_{\frac{n}{2}-1}(-i\lambda|x-y|), \qquad \operatorname{Im}\lambda > 0,$$

where $K_\nu(z)$ is the Macdonald function of order ν [DLMF, 10.27.4, 10.27.5]. Now, if $n=3$, the integral kernel of

$$\langle x\rangle^{-s}\frac{d}{d\lambda}(-\Delta-\lambda^2)^{-1}\langle x\rangle^{-s}$$

is given by $i(4\pi)^{-1}\langle x\rangle^{-s} e^{i\lambda|x-y|}\langle y\rangle^{-s}$, which has Hilbert–Schmidt norm bounded uniformly in $|\lambda| \le 1$ provided $s > \frac{3}{2}$. Moving on to $n \ge 4$, by [DLMF, 10.29.2],

$$\frac{d}{d\lambda}\left(\frac{-i\lambda}{2\pi|x-y|}\right)^{\frac{n}{2}-1} K_{\frac{n}{2}-1}(-i\lambda|x-y|) = \frac{-(-i)^{\frac{n}{2}}\lambda^{\frac{n}{2}-1}}{(2\pi)^{\frac{n}{2}-1}|x-y|^{\frac{n}{2}-2}} K_{\frac{n}{2}-2}(-i\lambda|x-y|).$$

The Macdonald function satisfies [DLMF, 10.25.3, 10.30.2, 10.30.3]

$$(85)\qquad |K_\nu(z)| \le \begin{cases} C|z|^{-\nu} & \text{if } 0<|z|\le 1,\ \nu>0,\\ C\big|\ln|z|\big| & \text{if } 0<|z|\le 1,\ \nu=0,\\ C|z|^{-1/2} & \text{if } |z|\ge 1,\ \operatorname{Re} z\ge 0,\ \nu\ge 0,\end{cases}$$

for $C > 0$ a constant independent of z. Therefore, there is a constant $C > 0$ independent of λ such that, if $n=4$,

$$\begin{aligned}(86)\quad &\left|\frac{\lambda^{\frac{n}{2}-1}\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{\frac{n}{2}-2}} K_{\frac{n}{2}-2}(-i\lambda|x-y|)\right|\\ &\le C|\lambda|\langle x\rangle^{-s}\langle y\rangle^{-s}\big|\ln(|\lambda|\,|x-y|)\big|\,\mathbf{1}_{\{|\lambda||x-y|\le 1\}} + C\frac{|\lambda|^{\frac{n}{2}-\frac{3}{2}}\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{\frac{n}{2}-\frac{3}{2}}}\mathbf{1}_{\{|\lambda||x-y|>1\}}\end{aligned}$$

and, if $n > 4$,

$$\begin{aligned}(87)\quad &\left|\frac{\lambda^{\frac{n}{2}-1}\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{\frac{n}{2}-2}} K_{\frac{n}{2}-2}(-i\lambda|x-y|)\right|\\ &\le C\frac{|\lambda|\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{n-4}}\mathbf{1}_{\{|\lambda||x-y|\le 1\}} + C\frac{|\lambda|^{\frac{n}{2}-\frac{3}{2}}\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{\frac{n}{2}-\frac{3}{2}}}\mathbf{1}_{\{|\lambda||x-y|>1\}}.\end{aligned}$$

As preparation for the conclusions we draw in the next paragraph, we observe that the first summand on the right-hand side of (87) has the bound, for $|\lambda| \le 1$ and $0 < \alpha \le 1$,

$$\begin{aligned}\frac{|\lambda|\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{n-4}}\mathbf{1}_{\{|\lambda||x-y|\le 1\}} &= \frac{|\lambda||x-y|^{\alpha}\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{n-4+\alpha}}\mathbf{1}_{\{|\lambda||x-y|\le 1\}} \\ &\le \frac{\langle x\rangle^{-s}\langle y\rangle^{-s}}{|x-y|^{n-4+\alpha}}\mathbf{1}_{\{|\lambda||x-y|\le 1\}}.\end{aligned} \tag{88}$$

In what follows we make repeated use of Lemma F.3 to verify that a given kernel is Hilbert–Schmidt. In both (86) and (87), the last summand on the right side is uniformly bounded in Hilbert–Schmidt norm for $|\lambda| \le 1$, provided $s > \frac{1}{4}(n+3)$. This also holds for the first summand on the right side of (86) if $s > \frac{3}{2}$. In addition, to address the first summand on the right side of (87), we utilize (88) in combination with Lemma F.3. Taken together this means that $\langle x\rangle^{-s}\langle y\rangle^{-s}|x-y|^{-n+4-\alpha}$ is Hilbert–Schmidt

when $n = 5$, if $\alpha = 1$ and $s > \frac{3}{2}$;
when $n = 6$, if $0 < \alpha < 1$ and $s > 2 - \frac{1}{2}\alpha$; and
when $n = 7$, if $0 < \alpha < \frac{1}{2}$ and $s > 2 - \frac{1}{2}\alpha$.

Thus, when $n = 6$ it is enough to take $s > \frac{3}{2}$, while when $n = 7$, $s > \frac{7}{4}$ suffices.

Finally, if $n \ge 8$, the first summand on the right in (87) is uniformly bounded as an operator $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ for $|\lambda| \le 1$ provided $s > 3$. This is due to (88) with $\alpha = 1$ and the Schur test; see Lemma F.2.

We finish by resolving the $|\alpha| = 1$ case for (54). By (54) in the $|\alpha| = 0$ case, and by (52), we need to show that

$$\left\|\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f\right\|_{H^2} \le O(1+|\lambda|)\|f\|_{L^2}.$$

According to (89) below,

$$\begin{aligned}&\left\|\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f\right\|_{H^2} \\ &\quad\le C\left\|\lambda\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f\right\|_{L^2} + C\left\|\lambda\langle x\rangle^{-s}(-\Delta)(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f\right\|_{L^2} \\ &\quad= C\|f\|_{L^2} + C\left\|\lambda\langle x\rangle^{-s}(-\Delta)(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f\right\|_{L^2}.\end{aligned}$$

Then use

$$\begin{aligned}&\langle x\rangle^{-s}(-\Delta)(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f \\ &\qquad\qquad= \langle x\rangle^{-s}(-\Delta-\lambda^2)^{-1}\langle x\rangle^{-s}f + \lambda^2\langle x\rangle^{-s}(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f,\end{aligned}$$

which in combination with (50), as well as (54) in the $|\alpha| = 0$ case, yields

$$\left\|\lambda\langle x\rangle^{-s}(-\Delta)(-\Delta-\lambda^2)^{-2}\langle x\rangle^{-s}f\right\|_{L^2} \le C(1+|\lambda|)\|f_{L^2},$$

completing the proof. □

Appendix F. Useful lemmas

Lemma F.1 [Fa67, Proposition 6]. *Let $n \geq 3$. Then*

$$\|r^{-1}u\|_{L^2}^2 \leq \left(\frac{2}{n-2}\right)^2 \|\nabla u\|_{L^2}^2, \quad u \in H^1(\mathbb{R}^n).$$

Lemma F.2 (Schur's test [DyZw19, Section A.5]). *Suppose $K(x, y)$ is measurable on $\mathbb{R}^n \times \mathbb{R}^n$ and satisfies*

$$\sup_x \int |K(x, y)|\, dy \leq C \quad \text{and} \quad \sup_y \int |K(x, y)|\, dx \leq C.$$

Then the linear operator

$$TF(x) = \int K(x, y) f(y)\, dy,$$

obeys the estimate

$$\|Tf\|_{L^2} \leq C\|f\|_{L^2}.$$

Lemma F.3 [Pe24]. *The necessary and sufficient conditions for*

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \langle x\rangle^{-s}\langle y\rangle^{-t}|x-y|^{-p}\, dx\, dy < \infty,$$

are

$$s+p > n, \quad t+p > n, \quad s+p+t > 2n, \quad p < n.$$

Lemma F.4. *Suppose $T : L^2(\mathbb{R}^n) \to H^2(\mathbb{R}^n)$ is a bounded operator. For any $s > 0$, there exists $C > 0$ such that*

$$\|\langle x\rangle^{-s}T\|_{L^2\to H^2} \leq C(\|\langle x\rangle^{-s}T\|_{L^2\to L^2} + \|\langle x\rangle^{-s}\Delta T\|_{L^2\to L^2}). \tag{89}$$

Proof. Let $f \in L^2(\mathbb{R}^n)$ and put $u = TF$. By the first line of (52), there exists $C > 0$, whose precise value may change from line to line, such that

$$\|\langle x\rangle^{-s}u\|_{H^2} \leq C\|\langle x\rangle^{-s}u\|_{L^2} + C\|\Delta\langle x\rangle^{-s}u\|_{L^2}, \quad \tilde{u} \in H^2(\mathbb{R}^n). \tag{90}$$

Then use the second line of (52):

$$\begin{aligned}\|\Delta\langle x\rangle^{-s}u\|_{L^2} &\leq \|[\Delta, \langle x\rangle^{-s}]u\|_{L^2} + \|\langle x\rangle^{-s}\Delta u\|_{L^2}\\ &\leq C\|\langle x\rangle^{-s}u\|_{H^1} + \|\langle x\rangle^{-s}\Delta u\|_{L^2}\\ &\leq C\gamma^{-1}\|\langle x\rangle^{-s}u\|_{L^2} + C\gamma\|\Delta\langle x\rangle^{-s}u\|_{L^2}) + \|\langle x\rangle^{-s}\Delta u\|_{L^2}, \quad \gamma > 0.\end{aligned}$$

Fixing γ small enough yields

$$\|\Delta\langle x\rangle^{-s}u\|_{L^2} \leq C(\|\langle x\rangle^{-s}u\|_{L^2} + \|\langle x\rangle^{-s}\Delta u\|_{L^2}),$$

which in combination with (90) implies (89). □

Acknowledgements

We thank Kiril Datchev and Georgi Vodev for helpful discussions. We also thank the referee for careful reading and helpful comments that improved the paper. Shapiro and Larraín-Hubach gratefully acknowledge support from NSF DMS-2204322. Shapiro was also supported by a University of Dayton Research Council Seed Grant.

References

[Bo11] J.-M. Bouclet, "Low frequency estimates and local energy decay for asymptotically Euclidean Laplacians", *Comm. Partial Differential Equations* **36**:7 (2011), 1239–1286. MR

[BoHa10] J.-F. Bony and D. Häfner, "Low frequency resolvent estimates for long range perturbations of the Euclidean Laplacian", *Math. Res. Lett.* **17**:2 (2010), 303–308. MR

[Bu02] N. Burq, "Lower bounds for shape resonances widths of long range Schrödinger operators", *Amer. J. Math.* **124**:4 (2002), 677–735. MR

[Bu98] N. Burq, "Décroissance de l'énergie locale de l'équation des ondes pour le problème extérieur et absence de résonance au voisinage du réel", *Acta Math.* **180**:1 (1998), 1–29. MR

[CaVo02] F. Cardoso and G. Vodev, "Uniform estimates of the resolvent of the Laplace–Beltrami operator on infinite volume Riemannian manifolds, II", *Ann. Henri Poincaré* **3**:4 (2002), 673–691. MR

[CaVo04] F. Cardoso and G. Vodev, "High frequency resolvent estimates and energy decay of solutions to the wave equation", *Canad. Math. Bull.* **47**:4 (2004), 504–514. MR

[CDY25] T. J. Christiansen, K. Datchev, and M. Yang, "From resolvent expansions at zero to long time wave expansions", *Comm. Partial Differential Equations* **50**:4 (2025), 477–492. MR

[ChDa21] T. J. Christiansen and K. Datchev, "Resolvent estimates on asymptotically cylindrical manifolds and on the half line", *Ann. Sci. Éc. Norm. Supér.* (4) **54**:4 (2021), 1051–1088. MR

[ChDa25] T. J. Christiansen and K. Datchev, "Low energy resolvent expansions in dimension two", *Commun. Am. Math. Soc.* **5** (2025), 48–80. MR

[ChIk20] R. C. Charão and R. Ikehata, "A note on decay rates of the local energy for wave equations with Lipschitz wavespeeds", *J. Math. Anal. Appl.* **483**:2 (2020), art. id. 123636, 14 pp. MR

[Da14] K. Datchev, "Quantitative limiting absorption principle in the semiclassical limit", *Geom. Funct. Anal.* **24**:3 (2014), 740–747. MR

[DGS23] K. Datchev, J. Galkowski, and J. Shapiro, "Semiclassical resolvent bounds for compactly supported radial potentials", *J. Funct. Anal.* **284**:7 (2023), art. id. 109835, 28 pp. MR

[DLMF] F. W. J. Olver, A. B. O. Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, , and M. A. McClain (editors), "NIST digital library of mathematical functions", http://dlmf.nist.gov. Release 1.2.4 of 2025-03-15.

[DyZw19] S. Dyatlov and M. Zworski, *Mathematical theory of scattering resonances*, Graduate Studies in Mathematics **200**, Amer. Math. Soc., 2019. MR

[EM] "Phragmén–Lindelöf theorem", Encyclopedia of Mathematics, http://encyclopediaofmath.org/index.php?title=Phragmén-Lindelöf_theorem.

[Fa67] W. G. Faris, "The product formula for semigroups defined by Friedrichs extensions", *Pacific J. Math.* **22** (1967), 47–70. MR

[GaSh22] J. Galkowski and J. Shapiro, "Semiclassical resolvent bounds for long-range Lipschitz potentials", *Int. Math. Res. Not.* **2022**:18 (2022), 14134–14150. MR

[GiMo74] J. Ginibre and M. Moulin, "Hilbert space approach to the quantum mechanical three-body problem", *Ann. Inst. H. Poincaré Sect. A* (*N.S.*) **21** (1974), 97–145. MR

[Ik23] R. Ikehata, "A note on local energy decay results for wave equations with a potential", *Asymptot. Anal.* **134**:1-2 (2023), 281–295. MR

[JeNe01] A. Jensen and G. Nenciu, "A unified approach to resolvent expansions at thresholds", *Rev. Math. Phys.* **13**:6 (2001), 717–754. MR

[LMP63] P. D. Lax, C. S. Morawetz, and R. S. Phillips, "Exponential decay of solutions of the wave equation in the exterior of a star-shaped obstacle", *Comm. Pure Appl. Math.* **16** (1963), 477–486. MR

[MeSj78] R. B. Melrose and J. Sjöstrand, "Singularities of boundary value problems, I", *Comm. Pure Appl. Math.* **31**:5 (1978), 593–617. MR

[MeSj82] R. B. Melrose and J. Sjöstrand, "Singularities of boundary value problems, II", *Comm. Pure Appl. Math.* **35**:2 (1982), 129–168. MR

[Mo61] C. S. Morawetz, "The decay of solutions of the exterior initial-boundary value problem for the wave equation", *Comm. Pure Appl. Math.* **14** (1961), 561–568. MR

[Mo66] C. S. Morawetz, "Exponential decay of solutions of the wave equation", *Comm. Pure Appl. Math.* **19** (1966), 439–444. MR

[Mo75] C. S. Morawetz, "Decay for solutions of the exterior problem for the wave equation", *Comm. Pure Appl. Math.* **28** (1975), 229–264. MR

[Ne64] E. Nelson, "Feynman integrals and the Schrödinger equation", *J. Mathematical Phys.* **5** (1964), 332–343. MR

[Ob24] D. Obovu, "Resolvent bounds for Lipschitz potentials in dimension two and higher with singularities at the origin", *J. Spectr. Theory* **14**:1 (2024), 163–183. MR

[Pe24] F. Petrov, "Show that the kernel $|x-y|^{-1}$ on $\mathbb{R}^3 \times \mathbb{R}^3$ is Hilbert–Schmidt with respect to a weighted L^2 space", Math Overflow posting, https://mathoverflow.net/q/478823.

[ReSi75] M. Reed and B. Simon, *Methods of modern mathematical physics*, *II*: *Fourier analysis*, *self-adjointness*, Academic Press, New York, 1975. MR

[ReSi80] M. Reed and B. Simon, *Methods of modern mathematical physics*, *I*: *Functional analysis*, revised ed., Academic Press, San Diego, 1980.

[RoSc04] I. Rodnianski and W. Schlag, "Time decay for solutions of Schrödinger equations with rough and time-dependent potentials", *Invent. Math.* **155**:3 (2004), 451–513. MR

[Sh18] J. Shapiro, "Local energy decay for Lipschitz wavespeeds", *Comm. Partial Differential Equations* **43**:5 (2018), 839–858. MR

[Va75] B. R. Vainberg, "On the short wave asymptotic behaviour of solutions of stationary problems and the asymptotic behaviour as $t \to \infty$ of solutions of non-stationary problems", *Uspehi Mat. Nauk* **30**:2(182) (1975), 3–55. In Russian; translated in *Russ. Math. Surv.* **30**:2 (1975), 1–58. MR

[Vo04a] G. Vodev, "Local energy decay of solutions to the wave equation for short-range potentials", *Asymptot. Anal.* **37**:2 (2004), 175–187. MR

[Vo04b] G. Vodev, "Local energy decay of solutions to the wave equation for nontrapping metrics", *Ark. Mat.* **42**:2 (2004), 379–397. MR

[Vo04c] G. Vodev, "Local energy decay of solutions to the wave equation for nontrapping metrics", *Mat. Contemp.* **26** (2004), 129–133. MR

[Vo99] G. Vodev, "On the uniform decay of the local energy", *Serdica Math. J.* **25**:3 (1999), 191–206. MR

[Za04] S. Zappacosta, "Resolvent estimates and applications to wave equations with potential", *Asymptot. Anal.* **38**:3-4 (2004), 221–239. MR

Received May 9 49, 2025. Revised March 5, 2026.

ANDRÉS LARRAÍN-HUBACH
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF DAYTON
DAYTON, OH
UNITED STATES
alarrainhubach1@udayton.edu

YULONG LI
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF DAYTON
DAYTON, OH
UNITED STATES
yli004@udayton.edu

JACOB SHAPIRO
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF DAYTON
DAYTON, OH9-2316
UNITED STATES
jshapiro1@udayton.edu

JOSEPH TILLER
PEERLESS TECHNOLOGIES
BEAVERCREEK, OH
UNITED STATES
joseph.tiller@epeerless.us

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.427

DIFFERENTIABLE SPHERE THEOREMS FOR COMPACT SUBMANIFOLDS

JUAN LI, HONGWEI XU AND ENTAO ZHAO

We investigate the differentiable structure on compact simply connected submanifolds in Riemannian manifolds under curvature pinching conditions. We prove a sharp differentiable sphere theorem that an n-dimensional compact simply connected submanifold M^n ($n \geq 5, n \neq 7, 8$) in the sphere $\mathbb{S}^N(1/\sqrt{c})$ ($c > 0$) with the second fundamental form A and the mean curvature vector H satisfying $|A|^2 \leq 4c + \frac{|H|^2}{n-2}$ is diffeomorphic to the standard sphere. The similar differentiable sphere theorem also holds for compact simply connected submanifolds in the space form $\mathbb{F}^N(c)$ with $c \leq 0$.

1. Introduction

The sphere theorems characterize the geometric and topological properties of compact Riemannian manifolds through curvature pinching conditions, representing a forefront topic in global Riemannian geometry. Rauch [30] first introduced the concept of curvature pinching for Riemannian manifolds. A Riemannian manifold M^n is δ-pinched (globally) for $\delta > 0$ if the sectional curvature K_M of M satisfies $\delta < K_M \leq 1$. Rauch [30] proved a topological sphere theorem for compact simply connected δ-pinched Riemannian manifolds with $\delta \approx 3/4$. Berger [3] and Klingenberg [17] provided a topological sphere theorem under the $1/4$-curvature pinching condition. Subsequently, Brendle and Schoen [6] proved the following differentiable sphere theorem by using Ricci flow techniques.

Theorem A [6]. *Let M be an n-dimensional ($n \geq 4$) complete and simply connected Riemannian manifold. If $1/4 < K_M \leq 1$, then M is diffeomorphic to the standard sphere $\mathbb{S}^n$.*

In fact, Brendle and Schoen proved the differentiable sphere theorem for pointwise $1/4$-pinched Riemannian manifolds. They proved in [5] that a compact simply connected weakly pointwise $1/4$-pinched Riemannian manifold is diffeomorphic to the standard sphere $\mathbb{S}^n$ or isometric to a compact rank one symmetric space (CROSS).

MSC2020: 53C40.

Keywords: differentiable sphere theorem, compact submanifold, curvature pinching, mean curvature flow, homology vanishing theorem.

Since any Riemannian manifold can be regarded as a zero-codimensional submanifold of itself, it is an interesting problem whether it is possible to extend the Brendle–Schoen differentiable sphere theorem to the case of submanifolds of arbitrary codimension in a general Riemannian manifold.

Let M^n be an n-dimensional submanifold in an N-dimensional Riemannian manifold $\overline{M}^N$. Denote by A and H the second fundamental form and the mean curvature vector of M, respectively. Lawson and Simons [19] proved a topological sphere theorem for compact submanifolds in the unit sphere by combining the homology vanishing theorem for compact submanifolds with Smale's [33] proof of the generalized Poincaré conjecture in dimensions $n \geq 5$.

Theorem B [19]. *Let M^n be an n-dimensional oriented compact submanifold in the unit sphere $\mathbb{S}^N$.*

(i) *If $n \neq 3, 4$ and $|A|^2 < 2\sqrt{n-1}$, then M is homeomorphic to a sphere.*

(ii) *If $n = 3, 4$ and $|A|^2 < n-1$, then M is a homotopy sphere.*

Inspired by the rigidity theorems for submanifolds with parallel mean curvature (see [32; 36; 37; 38]), Shiohama and Xu [31] improved and extended the theorem of Lawson and Simons [19]. They proved the following optimal topological sphere theorem for complete submanifolds in space forms.

Theorem C [31]. *Let M^n be an n-dimensional oriented complete submanifold in a simply connected space form $\mathbb{F}^N(c)$ with nonnegative constant curvature c. Assume*

$$\sup_M \big(|A|^2 - \alpha(n, |H|, c)\big) < 0,$$

where

$$\alpha(n, |H|, c) = nc + \frac{n}{2(n-1)}|H|^2 - \frac{n-2}{2(n-1)}\sqrt{|H|^4 + 4(n-1)c|H|^2}.$$

(i) *If $n \neq 3$, then M is homeomorphic to an n-dimensional sphere.*

(ii) *If $n = 3$, then M is diffeomorphic to a 3-dimensional spherical space form.*

For compact submanifolds in the hyperbolic space, a similar topological sphere theorem was proved by Fu and Xu [11].

Xu and Zhao [40] were the first to apply the Ricci flow to prove differentiable sphere theorems for compact submanifolds in general Riemannian manifolds. In particular, they proved the following theorem.

Theorem D [40]. *Let M^n be an n-dimensional $(n \geq 4)$ oriented complete submanifold in the unit sphere $\mathbb{S}^N$.*

(i) *If $n = 4, 5, 6$ and $\sup_M |A|^2 < 2\sqrt{n-1}$, then M is diffeomorphic to $\mathbb{S}^n$.*

(ii) *If $n \geq 7$ and $|A|^2 < 2\sqrt{2}$, then M is diffeomorphic to $\mathbb{S}^n$.*

Xu and Zhao [40] also obtained a topological sphere theorem for compact simply connected submanifolds in a Riemannian manifold under the assumption that

$$|A|^2 < \tfrac{16}{3}\left(\overline{K}_{\min} - \tfrac{1}{4}\overline{K}_{\max}\right) + \frac{|H|^2}{n-2}.$$

Here $\overline{K}_{\min}$ and $\overline{K}_{\max}$ are the minimum and the maximum of the sectional curvatures of $\overline{M}$ at a point.

Xu and Gu [39] further advanced the field by proving an optimal differential sphere theorem for complete submanifolds in space forms.

Theorem E [39]. *Let M^n be an n-dimensional $(n \geq 2)$ oriented complete submanifold in a simply connected space form $\mathbb{F}^N(c)$ with nonnegative constant curvature c. If*

$$\sup_M \left(|A|^2 - \frac{|H|^2}{n-1} - 2c\right) < 0,$$

then M is diffeomorphic to $\mathbb{S}^n$.

Many other differentiable sphere theorems for compact submanifolds were obtained by using Ricci flow and mean curvature flow techniques [1; 2; 12; 13; 22; 21; 23; 24; 26]. For instance, Lei and Xu [22; 21; 23] proved several optimal or sharp smooth convergence theorems for the mean curvature flow of submanifolds in space forms, which imply optimal or sharp differentiable sphere theorems for submanifolds in space forms.

Inspired by these developments, we prove the following differentiable sphere theorem for compact submanifolds in Riemannian manifolds.

Theorem 1.1. *Let M^n $(n \geq 4, n \neq 7, 8)$ be an n-dimensional compact simply connected submanifold in an N-dimensional Riemannian manifold $\overline{M}^N$. If*

$$|A|^2 < \frac{16}{3}\left(\overline{K}_{\min} - \frac{1}{4}\overline{K}_{\max}\right) + \frac{|H|^2}{n-2},$$

then M is diffeomorphic to $\mathbb{S}^n$.

Theorem 1.1 refines Xu and Zhao's topological sphere theorem in [40] to a differentiable sphere theorem. More differentiable sphere theorems for compact submanifolds in Riemannian manifolds will be proved in Section 6.

If the ambient space is a space form of constant sectional curvature c, then the condition in Theorem 1.1 simplifies to $|A|^2 < 4c + \frac{|H|^2}{n-2}$. With the aid of the homology vanishing theorem for compact submanifolds and the mean curvature flow of arbitrary codimension, we prove the following differentiable sphere theorem for compact submanifolds in space forms under a weak pinching condition.

Theorem 1.2. *Let M^n be an n-dimensional compact simply connected submanifold in a simply connected space form $\mathbb{F}^N(c)$ of constant sectional curvature c.*

(i) *If* $c \geq 0$, $n \geq 5$, $n \neq 7, 8$, $|H| > 0$ *for* $c = 0$, *and*

$$|A|^2 \leq 4c + \frac{|H|^2}{n-2},$$

then M *is diffeomorphic to* $\mathbb{S}^n$.

(ii) *If* $c < 0$, $n \geq 9$, *and*

$$|A|^2 \leq 4c + \frac{|H|^2}{n-2},$$

then M *is diffeomorphic to* $\mathbb{S}^n$.

For $c > 0$ and $n \geq 5$, consider

$$M^n(\varepsilon) := \mathbb{S}^2(\varepsilon) \times \mathbb{S}^{n-2}\big(\sqrt{1/c - \varepsilon^2}\big) \subset \mathbb{S}^{n+1}\big(1/\sqrt{c}\big),$$

with $0 < \varepsilon < 1/\sqrt{c}$. Intuitively, $M^n(\varepsilon)$ is simply connected and not homeomorphic to the sphere. By a direct computation, we have

$$|A|^2 - \frac{|H|^2}{n-2} = \frac{2(n-4)}{n-2}\frac{1}{\varepsilon^2} + \frac{2n}{n-2}c.$$

For $n \geq 5$, one has

$$\frac{2(n-4)}{n-2}\frac{1}{\varepsilon^2} + \frac{2n}{n-2}c \to +\infty \quad \text{as } \varepsilon \to 0,$$

$$\frac{2(n-4)}{n-2}\frac{1}{\varepsilon^2} + \frac{2n}{n-2}c \to 4c \quad \text{as } \varepsilon \to \frac{1}{\sqrt{c}}.$$

Therefore, for any constant $C > 4c$, there is $0 < \varepsilon < 1/\sqrt{c}$ such that $M^n(\varepsilon)$ satisfies $|A|^2 \leq C + \frac{|H|^2}{n-2}$. So for $c > 0$, Theorem 1.2 is sharp in the sense that the constant $4c$ in the pinching condition is the largest constant such that the differentiable sphere theorem holds. For $c > 0$ and $n = 7, 8$, M is homeomorphic to the sphere under the same pinching condition (see Theorem 4.1 in Section 4).

The paper is organized as follows. In Section 2, we introduce the relevant concepts and some curvature inequalities for Riemannian manifolds. In Section 3, we prove Theorem 1.1 with the aid of the classification theorem for Riemannian manifolds proved by using Ricci flow techniques. In Section 4, we prove a topological sphere theorem for compact submanifolds in space forms $\mathbb{F}^N(c)$ with $c \geq 0$, utilizing the homology vanishing theorem for compact submanifolds. In Section 5, we prove Theorem 1.2 by applying the mean curvature flow techniques. In Section 6, more differentiable sphere theorems for compact submanifolds under different curvature pinching conditions are proved.

2. Notation and formulas

Let (M^n, g) be an n-dimensional Riemannian submanifold in a Riemannian manifold $\overline{M}^N$ of dimension N with metric $\overline{g}$. We shall make use of the following convention on the range of indices:

$$1 \le i, j, k, \ldots \le n, \quad 1 \le A, B, C, \ldots \le N, \quad \text{and} \quad n+1 \le \alpha, \beta, \gamma, \ldots \le N.$$

We choose a local orthonormal frame $\{e_i\}$ for the tangent bundle and a local orthonormal frame $\{e_\alpha\}$ for the normal bundle. Let $\{\omega_A\}$ be the dual frame field corresponding to $\{e_A\}$. Denote by Rm and $\overline{Rm}$ the Riemannian curvature tensors of M and $\overline{M}$. Then we have

$$Rm = \sum_{i,j,k,l} R_{ijkl}\omega_i \otimes \omega_j \otimes \omega_k \otimes \omega_l,$$
$$\overline{Rm} = \sum_{A,B,C,D} \bar{R}_{ABCD}\omega_A \otimes \omega_B \otimes \omega_C \otimes \omega_D.$$

Let A and H be the second fundamental form and the mean curvature vector of M, given by

$$A = \sum_{i,j,\alpha} A^\alpha_{ij}\omega_i \otimes \omega_j \otimes e_\alpha, \quad H = \sum_\alpha H^\alpha e_\alpha, \quad H^\alpha = \sum_i A^\alpha_{ii}.$$

We have the Gauss equation

$$R_{ijkl} = \bar{R}_{ijkl} + \sum_\alpha \big(A^\alpha_{ik}A^\alpha_{jl} - A^\alpha_{il}A^\alpha_{jk}\big).$$

The trace-free second fundamental form $\mathring{A}$ is defined by $\mathring{A} = A - \frac{1}{n} g \otimes H$. Then

$$\mathring{A} = \sum_{i,j,\alpha} \mathring{A}^\alpha_{ij}\omega_i \otimes \omega_j \otimes e_\alpha, \quad \mathring{A}^\alpha_{ij} = A^\alpha_{ij} - \frac{1}{n}H^\alpha \delta_{ij}.$$

The sectional curvature, Ricci curvature, and scalar curvature on M and $\overline{M}$ are defined as K, $\bar{K}$, Ric, $\overline{\mathrm{Ric}}$, R, and $\bar{R}$, respectively. Thus, we have

$$\mathrm{Ric}(e_i) = \sum_j R_{ijij}, \quad \overline{\mathrm{Ric}}(e_A) = \sum_B \bar{R}_{ABAB}, \quad R = \sum_{i,j} R_{ijij}, \quad \bar{R} = \sum_{A,B} \bar{R}_{ABAB}.$$

The normalized scalar curvature $\bar{R}_0$ on $\overline{M}$ is defined as

$$\bar{R}_0 = \frac{\bar{R}}{N(N-1)}.$$

Set

$$\bar{K}_{\min}(x) = \min_{\pi \subset T_x\overline{M}} \bar{K}(\pi), \quad \overline{\mathrm{Ric}}_{\min}(x) = \min_{u \in U_x\overline{M}} \overline{\mathrm{Ric}}(u),$$
$$\bar{K}_{\max}(x) = \max_{\pi \subset T_x\overline{M}} \bar{K}(\pi), \quad \overline{\mathrm{Ric}}_{\max}(x) = \max_{u \in U_x\overline{M}} \overline{\mathrm{Ric}}(u).$$

We have Berger's inequalities:

$$|\bar{R}_{ACBC}| \le \tfrac{1}{2}(\bar{K}_{\max} - \bar{K}_{\min}) \quad \text{for all distinct indices } A, B, C,$$
$$|\bar{R}_{ABCD}| \le \tfrac{2}{3}(\bar{K}_{\max} - \bar{K}_{\min}) \quad \text{for all distinct indices } A, B, C, D.$$

For any unit tangent vector $u \in U_x\overline{M}$ at point $x \in \overline{M}$, let V_x^k be a k-dimensional subspace of $T_x\overline{M}$ satisfying $u \perp V_x^k$. Choose an orthonormal basis $\{e_A\}$ in $T_x\overline{M}$ such that for distinct indices $1 \le A_0, A_1, \dots, A_k \le N$, we have

$$e_{A_0} = u, \quad \operatorname{span}\{e_{A_1}, \dots, e_{A_k}\} = V_x^k.$$

The k-th Ricci curvature on $\overline{M}$ is defined as

$$\overline{\mathrm{Ric}}^{(k)}\left(u; V_x^k\right) = \overline{\mathrm{Ric}}^{(k)}\left([e_{A_0}, e_{A_1}, \dots, e_{A_k}]\right) = \sum_{p=1}^{k} \bar{R}_{A_0A_pA_0A_p}.$$

Set

$$\overline{\mathrm{Ric}}^{(k)}_{\min}(x) = \min_{u \in U_x\overline{M}} \min_{u \perp V_x^k \subset T_x\overline{M}} \overline{\mathrm{Ric}}^{(k)}(u; V_x^k),$$
$$\overline{\mathrm{Ric}}^{(k)}_{\max}(x) = \max_{u \in U_x\overline{M}} \max_{u \perp V_x^k \subset T_x\overline{M}} \overline{\mathrm{Ric}}^{(k)}(u; V_x^k).$$

Extend the orthonormal s-frame $\{e_{A_0}, \dots, e_{A_{s-1}}\}$ on $T_x\overline{M}$ to an orthonormal $(k+1)$-frame $\{e_{A_0}, \dots, e_{A_k}\}$ for $1 \le s \le k+1 \le N$. The (k, s)-curvature on $\overline{M}$ is defined as

$$\bar{R}^{(k,s)}([e_{A_0}, e_{A_1}, \dots, e_{A_k}]) = \sum_{p=0}^{s-1} \sum_{q=0}^{k} \bar{R}_{A_pA_qA_pA_q}.$$

Set

$$\bar{R}^{(k,s)}_{\min}(x) = \min_{\{e_{A_0}, e_{A_1}, \dots, e_{A_k}\} \subset T_x\overline{M}} \bar{R}^{(k,s)}([e_{A_0}, e_{A_1}, \dots, e_{A_k}]),$$
$$\bar{R}^{(k,s)}_{\max}(x) = \max_{\{e_{A_0}, e_{A_1}, \dots, e_{A_k}\} \subset T_x\overline{M}} \bar{R}^{(k,s)}([e_{A_0}, e_{A_1}, \dots, e_{A_k}]).$$

The s-th weak Ricci curvature on $\overline{M}$ is defined as

$$\begin{aligned}\overline{\mathrm{Ric}}^{[s]}([e_{A_0}, e_{A_1}, \dots, e_{A_{N-1}}]) &= \bar{R}^{(N-1,s)}([e_{A_0}, e_{A_1}, \dots, e_{A_{N-1}}]) \\ &= \sum_{p=0}^{s-1} \sum_{q=0}^{N-1} \bar{R}_{A_pA_qA_pA_q}.\end{aligned}$$

Set

$$\overline{\mathrm{Ric}}^{[s]}_{\min}(x) = \min_{\{e_{A_0}, e_{A_1}, \dots, e_{A_{N-1}}\} \subset T_x\overline{M}} \overline{\mathrm{Ric}}^{[s]}([e_{A_0}, e_{A_1}, \dots, e_{A_{N-1}}]),$$
$$\overline{\mathrm{Ric}}^{[s]}_{\max}(x) = \max_{\{e_{A_0}, e_{A_1}, \dots, e_{A_{N-1}}\} \subset T_x\overline{M}} \overline{\mathrm{Ric}}^{[s]}([e_{A_0}, e_{A_1}, \dots, e_{A_{N-1}}]).$$

The k-th scalar curvature on $\overline{M}$ is defined as

$$\overline{R}^{(k)}([e_{A_0}, e_{A_1}, \ldots, e_{A_k}]) = \overline{R}^{(k,k+1)}([e_{A_0}, e_{A_1}, \ldots, e_{A_k}]) = \sum_{p=0}^{k}\sum_{q=0}^{k} \overline{R}_{A_pA_qA_pA_q}.$$

Set

$$\overline{R}^{(k)}_{\min} = \min_{\{e_{A_0}, e_{A_1}, \ldots, e_{A_k}\}\subset T_x\overline{M}} \overline{R}^{(k)}([e_{A_0}, e_{A_1}, \ldots, e_{A_k}]),$$

$$\overline{R}^{(k)}_{\max} = \max_{\{e_{A_0}, e_{A_1}, \ldots, e_{A_k}\}\subset T_x\overline{M}} \overline{R}^{(k)}([e_{A_0}, e_{A_1}, \ldots, e_{A_k}]).$$

Remark 2.1. Based on the definitions, the Ricci curvature of $\overline{M}$ is equivalent to its $(N-1, 1)$-curvature, its $(N-1)$-th Ricci curvature, and its 1-st weak Ricci curvature. The scalar curvature of $\overline{M}$ is equivalent to its $(N-1, N)$-curvature, its N-th weak Ricci curvature, and its $(N-1)$-th scalar curvature.

Without loss of generality, all manifolds and submanifolds in this paper are assumed to be connected and without boundary.

3. Differentiable sphere theorem for compact submanifolds in Riemannian manifolds

To prove the sphere theorem, we need the classification theorem for compact Riemannian manifolds with positive isotropic curvature. For an n-dimensional $(n \geq 4)$ Riemannian manifold M^n, if the Riemannian curvature tensor R of M^n satisfies $R_{1313} + R_{1414} + R_{2323} + R_{2424} - 2R_{1234} > 0$ for every orthonormal four-frame $\{e_1, e_2, e_3, e_4\}$, then we say that M has positive isotropic curvature. Recently, Brendle [4] utilized curvature pinching estimates under Ricci flow to demonstrate that compact Riemannian manifolds with positive isotropic curvature in dimensions $n \geq 12$ maintain the pinching condition under Ricci flow. This led to the development of a classification theorem for compact Riemannian manifolds with positive isotropic curvature, thereby strengthening the intrinsic connection between curvature conditions and differential structure, and extending the applicability of classical differentiable sphere theorems. Chen [7] further extended the theorem to dimensions $n \geq 9$.

Theorem F [4; 7]. *Let M^n be an n-dimensional $(n \geq 9)$ compact manifold with positive isotropic curvature. If M does not contain any essential incompressible $(n-1)$-dimensional space forms, then M is diffeomorphic to a connected sum of finitely many spaces, each of which is a quotient of $\mathbb{S}^n$ or $\mathbb{S}^{n-1} \times \mathbb{R}$ by standard isometries.*

In the above theorem, an incompressible space form N^{n-1} in M^n is an $(n-1)$-dimensional submanifold diffeomorphic to $\mathbb{S}^{n-1}/\,\Gamma$ such that the fundamental group

$\pi_1(N)$ injects into $\pi_1(M)$. The space form is said to be essential unless Γ is trivial, or $\Gamma = \mathbb{Z}_2$ and the normal bundle is nonorientable.

Proposition 3.1. *Let M be an n-dimensional $(n \geq 4)$ compact simply connected Riemannian manifold. Then there is no essential $(n-1)$-dimensional incompressible space form in M.*

Proof. Let N be an $(n-1)$-dimensional submanifold that is diffeomorphic to $\mathbb{S}^{n-1}/\Gamma$ in the Riemannian manifold M. By applying the Killing–Hopf theorem [15; 16], we deduce that $\pi_1(N) = \pi_1(\mathbb{S}^{n-1}/\Gamma) \approx \Gamma$ as $\mathbb{S}^{n-1}$ is simply connected. Clearly, if Γ is nontrivial, then there is no injective homomorphism from $\pi_1(N)$ to $\pi_1(M)$, since Γ contains more than one element, while $\pi_1(M)$ is trivial. Therefore, there is no essential $(n-1)$-dimensional incompressible space form in M. □

Using the above proposition, we present the proof of Theorem 1.1.

Proof of Theorem 1.1. From the definition of the second fundamental form, we have

$$|A|^2 = \sum_{i,j=1}^{n} (A_{ij}^{\alpha})^2 = \sum_{i=1}^{n} (A_{ii}^{\alpha})^2 + \sum_{i \neq j} (A_{ij}^{\alpha})^2.$$

For all distinct indices p, q, k, l, applying the Cauchy inequality yields

$$\begin{aligned}\Big(\sum_{i=1}^{n} A_{ii}^{\alpha}\Big)^2 &\leq (n-2)\Big[(A_{pp}^{\alpha} + A_{qq}^{\alpha})^2 + (A_{kk}^{\alpha} + A_{ll}^{\alpha})^2 + \sum_{i \neq p,q,k,l} (A_{ii}^{\alpha})^2\Big] \\ &= (n-2)\Big(\sum_{i=1}^{n} (A_{ii}^{\alpha})^2 + 2A_{pp}^{\alpha} A_{qq}^{\alpha} + 2A_{kk}^{\alpha} A_{ll}^{\alpha}\Big).\end{aligned}$$

This inequality implies

$$\begin{aligned}2A_{pp}^{\alpha} A_{qq}^{\alpha} + 2A_{kk}^{\alpha} A_{ll}^{\alpha} &\geq \frac{\big(\sum_{i=1}^{n} A_{ii}^{\alpha}\big)^2}{n-2} - \sum_{i=1}^{n} (A_{ii}^{\alpha})^2 \\ &= \frac{\big(\sum_{i=1}^{n} A_{ii}^{\alpha}\big)^2}{n-2} + \sum_{i \neq j} (A_{ij}^{\alpha})^2 - \sum_{i,j=1}^{n} (A_{ij}^{\alpha})^2.\end{aligned} \tag{3-1}$$

Suppose $\{e_1, e_2, e_3, e_4\}$ is an orthonormal four-frame. By the Gauss equation, we have

$$\begin{aligned} &R_{1313} + R_{1414} + R_{2323} + R_{2424} - 2R_{1234} \\ &\quad = \bar{R}_{1313} + \bar{R}_{1414} + \bar{R}_{2323} + \bar{R}_{2424} - 2\bar{R}_{1234} \\ &\qquad + \sum_{\alpha} \big[A_{11}^{\alpha} A_{33}^{\alpha} + A_{11}^{\alpha} A_{44}^{\alpha} + A_{22}^{\alpha} A_{33}^{\alpha} + A_{22}^{\alpha} A_{44}^{\alpha} - (A_{13}^{\alpha})^2 - (A_{14}^{\alpha})^2 \\ &\qquad\qquad - (A_{23}^{\alpha})^2 - (A_{24}^{\alpha})^2 - 2(A_{13}^{\alpha} A_{24}^{\alpha} - A_{14}^{\alpha} A_{23}^{\alpha})\big]. \end{aligned} \tag{3-2}$$

Applying Berger's inequality yields

$$\bar{R}_{1234} \le \tfrac{2}{3}(\bar{K}_{\max} - \bar{K}_{\min}).$$

Hence,

$$\bar{R}_{1313} + \bar{R}_{1414} + \bar{R}_{2323} + \bar{R}_{2424} - 2\bar{R}_{1234} \ge 4\bar{K}_{\min} - \tfrac{4}{3}(\bar{K}_{\max} - \bar{K}_{\min}). \tag{3-3}$$

By (3-1), we obtain the following estimate for the first four terms in brackets on the right-hand side of (3-2):

$$\begin{aligned} A^{\alpha}_{11}A^{\alpha}_{33} + A^{\alpha}_{11}A^{\alpha}_{44} + A^{\alpha}_{22}A^{\alpha}_{33} + A^{\alpha}_{22}A^{\alpha}_{44} \\ \ge \frac{\left(\sum_{i=1}^{n} A^{\alpha}_{ii}\right)^2}{n-2} + \sum_{i\ne j}(A^{\alpha}_{ij})^2 - \sum_{i,j=1}^{n}(A^{\alpha}_{ij})^2. \end{aligned} \tag{3-4}$$

Combining (3-2)–(3-4), we obtain

$$\begin{aligned} &R_{1313} + R_{1414} + R_{2323} + R_{2424} - 2R_{1234} \\ &\quad \ge \tfrac{16}{3}(\bar{K}_{\min} - \tfrac{1}{4}\bar{K}_{\max}) \\ &\qquad + \sum_{\alpha}\Bigg[\frac{\left(\sum_{i=1}^{n} A^{\alpha}_{ii}\right)^2}{n-2} + \sum_{i\ne j}(A^{\alpha}_{ij})^2 - \sum_{i,j=1}^{n}(A^{\alpha}_{ij})^2 \\ &\qquad\qquad - 2(A^{\alpha}_{13})^2 - 2(A^{\alpha}_{14})^2 - 2(A^{\alpha}_{23})^2 - 2(A^{\alpha}_{24})^2\Bigg] \\ &\quad \ge \tfrac{16}{3}\left(\bar{K}_{\min} - \tfrac{1}{4}\bar{K}_{\max}\right) + \frac{|H|^2}{n-2} - |A|^2. \end{aligned} \tag{3-5}$$

Therefore, under the assumption of Theorem 1.1, M has positive isotropic curvature.

When $n = 4$, by [8; 9; 14], M is diffeomorphic to the standard sphere.

When $n = 5, 6$, it is known from [27] that M is homeomorphic to a sphere. As the differentiable structure on the topological sphere of dimension 5 or 6 is unique, M is diffeomorphic to the standard sphere.

When $n \ge 9$, Proposition 3.1 implies that M contains no essential $(n-1)$-dimensional incompressible spatial form. By Theorem F, M is diffeomorphic to the standard sphere. □

4. Topological sphere theorem for compact submanifolds in space forms

In this section, we prove a topological sphere theorem for compact submanifolds in space forms of nonnegative sectional curvature.

Theorem 4.1. *Let M^n be an n-dimensional ($n \geq 5$) compact simply connected submanifold in a simply connected space $\mathbb{F}^N(c)$ with $c \geq 0$. Assume that $|H| > 0$ for $c = 0$. If*

$$|A|^2 \leq 4c + \frac{|H|^2}{n-2},$$

then M is homeomorphic to $\mathbb{S}^n$.

Theorem 4.1 provides a proof of the case $n = 5, 6$ of (i) in Theorem 1.2.

We need the following homology vanishing theorem for compact submanifolds in space forms to prove Theorem 4.1. Lawson and Simons [19] initially established the homology vanishing theorem for compact submanifolds in spheres under a strict pointwise pinching condition. Subsequently, Xin [35] generalized this theorem to compact submanifolds in the Euclidean space. It was observed by Elworthy and Rosenberg [10] that the Lawson–Simons theorem still holds under a weak pinching condition that is strict at some point. This generalization is also true for compact submanifolds in the Euclidean space.

Theorem 4.2 [10; 19; 35]. *Let M^n be an n-dimensional ($n \geq 5$) compact submanifold in a simply connected space form $\mathbb{F}^N(c)$ with nonnegative constant curvature c. Assume that, for an integer $0 < q < n$,*

$$\sum_{k=q+1}^{n} \sum_{i=1}^{q} \big(2|A(e_i, e_k)|^2 - \langle A(e_i, e_i), A(e_k, e_k)\rangle\big) \leq q(n-q)c \tag{4-1}$$

holds for any orthonormal basis $\{e_i\}$ of T_xM at any point $x \in M$. If there is a point such that (4-1) *is strict for any orthonormal basis $\{e_i\}$ at that point, then there do not exist any stable q-currents, and*

$$H_q(M; \mathbb{Z}) = H_{n-q}(M; \mathbb{Z}) = 0,$$

where $H_i(M; \mathbb{Z})$ is the i-th homology group of M with integer coefficients.

Proof of Theorem 4.1. We follow the computation in [31]. For any orthonormal basis $\{e_\alpha\}$ of the normal space at a point,

$$\begin{aligned} &\sum_{k=q+1}^{n} \sum_{i=1}^{q} \big(2|A(e_i, e_k)|^2 - \langle A(e_i, e_i), A(e_k, e_k)\rangle\big) \\ &\qquad = 2\sum_{\alpha} \sum_{k=q+1}^{n} \sum_{i=1}^{q} (A_{ik}^{\alpha})^2 - \sum_{\alpha} \sum_{k=q+1}^{n} \sum_{i=1}^{q} A_{ii}^{\alpha} A_{kk}^{\alpha} \\ &\qquad = \sum_{\alpha} \bigg[2 \sum_{k=q+1}^{n} \sum_{i=1}^{q} (A_{ik}^{\alpha})^2 - \Big(\sum_{i=1}^{q} A_{ii}^{\alpha} \Big) \Big(H^{\alpha} - \sum_{i=1}^{q} A_{ii}^{\alpha} \Big) \bigg]. \end{aligned} \tag{4-2}$$

We take $r = n - q$ and set

$$Z_\alpha := -\Big(\sum_{i=1}^{q} A_{ii}^\alpha\Big)\Big(H^\alpha - \sum_{i=1}^{q} A_{ii}^\alpha\Big), \quad S_\alpha := \sum_{i,j=1}^{n} (A_{ij}^\alpha)^2, \quad \tilde{S}_\alpha := \sum_{i=1}^{n} (A_{ii}^\alpha)^2.$$

Applying the Cauchy inequality, we have

$$\begin{aligned} qr\tilde{S}_\alpha &= qr\sum_{i=1}^{q}(A_{ii}^\alpha)^2 + qr\sum_{k=q+1}^{n}(A_{kk}^\alpha)^2 \\ &\geq r\Big(\sum_{i=1}^{q} A_{ii}^\alpha\Big)^2 + q\Big(\sum_{k=q+1}^{n} A_{kk}^\alpha\Big)^2 \\ &= (r+q)\Big(\sum_{i=1}^{q} A_{ii}^\alpha\Big)^2 - 2qH^\alpha\Big(\sum_{i=1}^{q} A_{ii}^\alpha\Big) + q(H^\alpha)^2. \end{aligned} \tag{4-3}$$

This inequality implies

$$nZ_\alpha + (r-q)H^\alpha\Big(\sum_{i=1}^{q} A_{ii}^\alpha\Big) + q(H^\alpha)^2 - qr\tilde{S}_\alpha \leq 0. \tag{4-4}$$

By the definition and applying the Cauchy inequality, we have

$$\begin{aligned} \tilde{S}_\alpha - \frac{1}{n}(H^\alpha)^2 &= \sum_{i=1}^{q}(\mathring{A}_{ii}^\alpha)^2 + \sum_{k=q+1}^{n}(\mathring{A}_{kk}^\alpha)^2 \geq \frac{1}{q}\Big(\sum_{i=1}^{q}\mathring{A}_{ii}^\alpha\Big)^2 + \frac{1}{r}\Big(\sum_{k=q+1}^{n}\mathring{A}_{kk}^\alpha\Big)^2 \\ &= \Big(\frac{1}{q}+\frac{1}{r}\Big)\Big(\sum_{i=1}^{q}\mathring{A}_{ii}^\alpha\Big)^2. \end{aligned}$$

Hence,

$$\Big|\sum_{i=1}^{q}\mathring{A}_{ii}^\alpha\Big| \leq \sqrt{\frac{qr}{n}\big(\tilde{S}_\alpha - \tfrac{1}{n}(H^\alpha)^2\big)}.$$

Combining this inequality with (4-4), we get

$$Z_\alpha \leq \frac{qr}{n}\tilde{S}_\alpha - \Big(\frac{q(r-q)}{n^2} + \frac{q}{n}\Big)(H^\alpha)^2 + \frac{|r-q|}{n}|H^\alpha|\sqrt{\frac{qr}{n}\big(\tilde{S}_\alpha - \tfrac{1}{n}(H^\alpha)^2\big)}.$$

Substituting this inequality into (4-2) yields

$$\begin{aligned} &\sum_{k=q+1}^{n}\sum_{i=1}^{q}(2|A(e_i,e_k)|^2 - \langle A(e_i,e_i), A(e_k,e_k)\rangle) \\ &\quad\leq \sum_\alpha\bigg\{2\sum_{k=q+1}^{n}\sum_{i=1}^{q}(A_{ik}^\alpha)^2 + \frac{qr}{n}\tilde{S}_\alpha - \Big(\frac{q(r-q)}{n^2}+\frac{q}{n}\Big)(H^\alpha)^2 \\ &\qquad + \frac{|r-q|}{n}|H^\alpha|\sqrt{\frac{qr}{n}\big(\tilde{S}_\alpha - \tfrac{1}{n}(H^\alpha)^2\big)}\bigg\}. \end{aligned}$$

For $n \geq 5$ and $2 \leq q \leq n-2$, one has $qr > n$. So for any $n-1 \leq \alpha \leq N$,

$$
\begin{aligned}
(4\text{-}5)\quad & 2\sum_{k=q+1}^{n}\sum_{i=1}^{q}(A_{ik}^{\alpha})^2 + \frac{qr}{n}\sum_{i=1}^{n}(A_{ii}^{\alpha})^2 \\
&= \frac{qr}{n}\sum_{i,j=1}^{n}(A_{ij}^{\alpha})^2 - \Big(\frac{qr}{n}-1\Big)\sum_{i\neq j}(A_{ij}^{\alpha})^2 - \sum_{\substack{i,j=1\\ i\neq j}}^{q}(A_{ij}^{\alpha})^2 - \sum_{\substack{i,j=q+1\\ i\neq j}}^{n}(A_{ij}^{\alpha})^2 \\
&\leq \frac{qr}{n}\sum_{i,j=1}^{n}(A_{ij}^{\alpha})^2.
\end{aligned}
$$

Applying the Cauchy inequality, we have

$$
(4\text{-}6)\qquad \sum_{\alpha}|H^{\alpha}|\sqrt{\tilde{S}_{\alpha}-\frac{1}{n}(H^{\alpha})^2} \leq \sqrt{\sum_{\alpha}(H^{\alpha})^2\cdot\sum_{\alpha}\Big(\tilde{S}_{\alpha}-\frac{1}{n}(H^{\alpha})^2\Big)}.
$$

Hence,

$$
\begin{aligned}
(4\text{-}7)\ & \sum_{k=q+1}^{n}\sum_{i=1}^{q}(2|A(e_i,e_k)|^2 - \langle A(e_i,e_i), A(e_k,e_k)\rangle) \\
&\leq \sum_{\alpha}\Big(\frac{qr}{n}S_{\alpha} - \frac{2qr}{n^2}(H^{\alpha})^2\Big) + \frac{|r-q|}{n}\sqrt{\frac{qr}{n}\sum_{\alpha}(H^{\alpha})^2\cdot\sum_{\alpha}\Big(\tilde{S}_{\alpha}-\frac{1}{n}(H^{\alpha})^2\Big)} \\
&\leq \frac{qr}{n}\bigg(|A|^2 - \frac{2}{n}|H|^2 + \frac{|r-q|}{\sqrt{nqr}}\,|H|\sqrt{|A|^2-\tfrac{1}{n}|H|^2}\bigg) \\
&\leq \frac{qr}{n}\bigg(|A|^2 - \frac{2}{n}|H|^2 - nc + \frac{n-4}{\sqrt{2n(n-2)}}\,|H|\sqrt{|A|^2-\tfrac{1}{n}|H|^2}\bigg) + qrc.
\end{aligned}
$$

Now we take the case $c > 0$. By a direct computation, we have, for $2 \leq q \leq n-2$ and $n \geq 5$,

$$
|A|^2 - \frac{2}{n}|H|^2 - nc + \frac{n-4}{\sqrt{2n(n-2)}}\,|H|\sqrt{|A|^2-\tfrac{1}{n}|H|^2} < 0,
$$

provided $|A|^2 \leq 4c + \frac{|H|^2}{n-2}$. Therefore,

$$
\sum_{k=q+1}^{n}\sum_{i=1}^{q}\big(2|A(e_i,e_k)|^2 - \langle A(e_i,e_i), A(e_k,e_k)\rangle\big) < qrc
$$

for $2 \leq q \leq n-2$. By Theorem 4.2, $H_q(M;\mathbb{Z}) = H_{n-q}(M;\mathbb{Z}) = 0$ for $2 \leq q \leq n-2$.

Next, we consider the case $c = 0$. As $n \geq 5$ and $2 \leq q \leq n-2$, by a direct computation we have

$$\sum_{k=q+1}^{n}\sum_{i=1}^{q}\big(2|A(e_i,e_k)|^2 - \langle A(e_i,e_i), A(e_k,e_k)\rangle\big) \leq \frac{qr}{n}\Big(|A|^2 - \frac{2}{n}|H|^2 + \frac{n-4}{\sqrt{2n(n-2)}}|H|\sqrt{|A|^2 - \tfrac{1}{n}|H|^2}\Big) \leq 0, \tag{4-8}$$

provided $|A|^2 \leq |H|^2/(n-2)$. Moreover, the second inequality becomes equality at a point if and only if $|A|^2 = |H|^2/(n-2)$ at this point.

If the second inequality in (4-8) is strict at a point for any orthonormal basis $\{e_i\}$ at that point, then by Theorem 4.2, $H_q(M;\mathbb{Z}) = H_{n-q}(M;\mathbb{Z}) = 0$ for $2 \leq q \leq n-2$.

Now we consider the case that the equality holds for the second inequality in (4-8) at any point for a certain orthonormal basis $\{e_i\}$ at that point. Then at any point, all inequalities become equalities. In particular, from (4-8), there holds $|A|^2 - |H|^2/(n-2) \equiv 0$.

At any point, given a corresponding $\{e_i\}$ and for any $\{e_\alpha\}$, the equality case of (4-3) implies that for any $n+1 \leq \alpha \leq N$,

$$A^\alpha_{11} = \cdots = A^\alpha_{qq}, \quad A^\alpha_{q+1q+1} = \cdots = A^\alpha_{nn}.$$

Moreover, the equality case of (4-5) gives that for any $n+1 \leq \alpha \leq N$,

$$A^\alpha_{ij} = 0, \quad \text{for any } i \neq j.$$

This implies $S_\alpha = \tilde{S}_\alpha$ for any $n+1 \leq \alpha \leq N$.

From the equality case of (4-6) we conclude that there exists a constant ρ for this point such that

$$\tilde{S}_\alpha = \rho\cdot(H^\alpha)^2$$

for any $n+1 \leq \alpha \leq N$. Since $S_\alpha = \tilde{S}_\alpha$ and $|A|^2 - \dfrac{|H|^2}{n-2} \equiv 0$, one has $\rho = \dfrac{1}{n-2}$.

As $|H| > 0$ and $|A|^2 - |H|^2/n > 0$, the equality case of (4-7) implies

$$\frac{|r-q|}{\sqrt{nqr}} = \frac{n-4}{\sqrt{2n(n-2)}}.$$

Hence, $q = 2$ or $q = n-2$. Without loss of generality, we choose $q = 2$.

As $|H| > 0$ on M, at any point we can choose $\{e_\alpha\}$ such that $e_{n+1} = H/|H|$. Then $H^{n+1} = |H|$ and $H^\alpha = 0$ for any $\alpha = n+2, \dots, N$. Since $A^\alpha_{ij} = 0$ for any α and $i \neq j$, and $\tilde{S}_\alpha = \frac{1}{n-2}(H^\alpha)^2$ for any α, we conclude that $A^\alpha = 0$ for $\alpha = n+2, \dots, N$. Thus,

$$S_{n+1} = \sum_{i,j}(A^{n+1}_{ij})^2 = |A|^2 = \frac{|H|^2}{n-2}.$$

This implies $A^{n+1}_{ij} \leq |A| = |H|/\sqrt{n-2} < |H|$. Therefore, by Proposition 2.5 in [28], M lies in an $(n+1)$-dimensional affine subspace of $\mathbb{R}^N$. Since M is a hypersurface in $\mathbb{R}^{n+1}$, its second fundamental form can be considered as a symmetric 2-tensor $A = \{A_{ij}\}$ and the mean curvature H of M is essentially a scalar function. We have $|A|^2 - H^2/(n-2) \equiv 0$, and at any point there is an orthonormal basis $\{e_i\}$ such that

$$A_{11} = A_{22}, \qquad A_{33} = \cdots = A_{nn}, \qquad A_{ij} = 0 \quad \text{for all } i \neq j.$$

Therefore, M is a compact hypersurface in $\mathbb{R}^{n+1}$ with two distinct principal curvatures of multiplicities 2 and $n-2$. We denote the two principal curvatures by λ and μ, respectively. Then λ and μ are smooth functions on M. We may assume $H > 0$. Since

$$2\lambda + (n-2)\mu = H \quad \text{and} \quad 2\lambda^2 + (n-2)\mu^2 = \frac{H^2}{n-2},$$

we have, by a direct computation,

$$\lambda = 0 \quad \text{and} \quad \mu = \frac{H}{n-2},$$

or

$$\lambda = \frac{2}{n}H \quad \text{and} \quad \mu = \frac{n-4}{n(n-2)}H.$$

In the first subcase, one has $\lambda\mu = 0$. By the Gauss equation, at any point of M there exists a tangent plane with zero sectional curvature. However, since M is a compact hypersurface in $\mathbb{R}^{n+1}$, there exists a point of M at which all the sectional curvatures are positive. Therefore, this subcase is impossible.

For the second subcase, we will prove that M is isoparametric. By Theorem 2 in [29] and its corollary, the distribution of the space of the principal vectors corresponding to λ (resp., μ) is completely integrable, and λ (resp., μ) is constant on each integral submanifold of the corresponding distribution of the space of the principal vectors. At any point of M, there exists a corresponding $\{e_i\}$ such that e_1, e_2 are principal vectors of λ and $e_3, \dots, e_n$ are principal vectors of μ. Putting

$$d\lambda = \sum_{k=1}^{n} \lambda_{,k}\omega_k, \qquad d\mu = \sum_{k=1}^{n} \mu_{,k}\omega_k,$$

one has at this point that

$$\lambda_{,1} = \lambda_{,2} = 0, \qquad \mu_{,3} = \cdots = \mu_{,n} = 0.$$

Putting

$$dH = \sum_{k=1}^{n} H_k\omega_k,$$

we have

$$H_i = \frac{n}{2}\lambda_{,i} = 0, \qquad \text{for } \ i = 1, 2,$$

and

$$H_j = \frac{n(n-2)}{n-4}\mu_{,j} = 0, \qquad \text{for } \ j = 3, \dots, n.$$

Consequently, $dH = 0$ at this point. Since the point is arbitrary, one gets $dH = 0$ on M. This implies that the mean curvature is constant on M and the principal curvatures λ and μ are both constant. Therefore, M is an isoparametric hypersurface in $\mathbb{R}^{n+1}$ with two distinct principal curvatures λ and μ. By Cartan's formula, $\lambda\mu = 0$. This leads to a contradiction.

In summary, we conclude that there always exists a point of M such that the second inequality in (4-8) is strict at this point for any orthonormal basis $\{e_i\}$. Therefore, $H_q(M;\mathbb{Z}) = H_{n-q}(M;\mathbb{Z}) = 0$ for $2 \le q \le n-2$.

Finally, we give the proof of Theorem 4.1. Since M is simply connected, it is oriented (see Theorem 15.43 in [20] for a proof). By the Hurewicz theorem (see Theorem 5 in Chapter 7 of [34]), the homotopy groups of M satisfy $\pi_i(M) = 0$ for $i = 1, \dots, n-1$. So M is a homotopy sphere. By the proof of the generalized Poincaré conjecture [33], M is a topological sphere. □

5. Differentiable sphere theorem for compact submanifolds in space forms

We will use the mean curvature flow as a tool to prove Theorem 1.2.

Let M be an n-dimensional compact submanifold in a simply connected space form $\mathbb{F}^N(c)$ of constant sectional curvature c. Denote by F the immersion. We deform M by the mean curvature flow $F : M^n \times [0, T) \to \mathbb{F}^N(c)$ that satisfies

$$\frac{\partial}{\partial t} F(x,t) = H(x,t),$$

where $H(x,t)$ is the mean curvature vector of the submanifold $M_t = F(M,t)$.

Lemma 5.1. *Let $F : M^n \times [0, T) \to \mathbb{F}^N(c)$ be a mean curvature flow of compact submanifolds of dimension $n \ge 8$ in a simply connected space form $\mathbb{F}^N(c)$. Assume M_0 satisfies $|A|^2 \le 4c + \frac{|H|^2}{n-2}$, and $|H| > 0$ for $c = 0$. Then M_t satisfies $|A|^2 < 4c + \frac{|H|^2}{n-2}$ for $t > 0$.*

Proof. As in [1; 2; 26], we set $Q = |A|^2 - a|H|^2 - bc$. Then

$$\text{(5-1)} \quad \Big(\frac{\partial}{\partial t} - \Delta\Big) Q = -2(|\nabla A|^2 - a|\nabla H|^2) + 2R_1 - 2aR_2 - 2nc|\mathring{A}|^2 - 2n\Big(a - \frac{1}{n}\Big)c|H|^2,$$

where

$$R_1 = \sum_{\alpha,\beta}\Big(\sum_{i,j} A^\alpha_{ij} A^\beta_{ij}\Big)^2 + |R^\perp|^2 \quad \text{and} \quad R_2 = \sum_{i,j}\Big(\sum_\alpha H^\alpha A^\alpha_{ij}\Big)^2,$$

with

$$|R^\perp|^2 = \sum_{i,j,\alpha,\beta}\Big(\sum_k (A^\alpha_{ik} A^\beta_{jk} - A^\beta_{ik} A^\alpha_{jk})\Big)^2 = \sum_{\alpha,\beta} N(A^\alpha A^\beta - A^\beta A^\alpha).$$

Here $A^\alpha = (A^\alpha_{ij})$ and $N(\cdot)$ denotes the squared Frobenius norm of a matrix.

At the point where $|H| \neq 0$, we choose $\{e_\alpha\}$ such that $e_{n+1} = H/|H|$ at this point. Let $A_H = \sum_{i,j} A_{ij}^{n+1} \omega^i \otimes \omega^j$, $A_I = A - A_H$, $\mathring{A}_H = A_H - \frac{1}{n} H \otimes g$, $\mathring{A}_I = \mathring{A} - \mathring{A}_H$. From the definitions, it follows that

$$|A_H|^2 = |A^{n+1}|^2, \quad |A_I|^2 = \sum_{\alpha > n+1} |A^\alpha|^2 = |A|^2 - |A_H|^2,$$
$$|\mathring{A}_H|^2 = |\mathring{A}^{n+1}|^2, \quad |\mathring{A}_I|^2 = \sum_{\alpha > n+1} |A^\alpha|^2 = |\mathring{A}|^2 - |\mathring{A}_H|^2.$$

Notice that $|A_H|^2 = |\mathring{A}_H|^2 + \frac{1}{n}|H|^2$ and $|\mathring{A}_I|^2 = |A_I|^2$. By the calculations in [1], one has

$$\begin{aligned}(5\text{-}2)\quad 2R_1 - 2aR_2 = {} & 2|\mathring{A}_H|^4 - 2\Big(a - \frac{2}{n}\Big)|\mathring{A}_H|^2|H|^2 - \frac{2}{n}\Big(a - \frac{1}{n}\Big)|H|^4 \\ & + 4\sum_{\alpha > n+1}\Big(\sum_{i,j} \mathring{A}_{ij}^{n+1}\mathring{A}_{ij}^{\alpha}\Big)^2 + 4\sum_{\alpha > n+1} N(A^{n+1}\mathring{A}^\alpha - \mathring{A}^\alpha A^{n+1}) \\ & + 2\sum_{\alpha,\beta > n+1}\Big(\sum_{i,j} \mathring{A}_{ij}^{\alpha}\mathring{A}_{ij}^{\beta}\Big)^2 + 2\sum_{\alpha,\beta > n+1} N(\mathring{A}^\alpha \mathring{A}^\beta - \mathring{A}^\beta \mathring{A}^\alpha).\end{aligned}$$

For the first line of the right-hand side of (5-2), we replace $|H|^2$ with

$$\frac{|\mathring{A}_H|^2 + |\mathring{A}_I|^2 - bc - Q}{a - \frac{1}{n}}$$

and get

$$\begin{aligned}(5\text{-}3)\quad & 2|\mathring{A}_H|^4 - 2\Big(a - \frac{2}{n}\Big)|\mathring{A}_H|^2|H|^2 - \frac{2}{n}\Big(a - \frac{1}{n}\Big)|H|^4 \\ & = -\bigg(2 + \frac{2}{n\big(a - \frac{1}{n}\big)}\bigg)|\mathring{A}_H|^2|\mathring{A}_I|^2 + \bigg(2 + \frac{2}{n\big(a - \frac{1}{n}\big)}\bigg)|\mathring{A}_H|^2 Q \\ & \quad + \bigg(2b + \frac{2b}{n\big(a - \frac{1}{n}\big)}\bigg)c|\mathring{A}_H|^2 - \frac{2}{n\big(a - \frac{1}{n}\big)}|\mathring{A}_I|^4 + \frac{4}{n\big(a - \frac{1}{n}\big)}|\mathring{A}_I|^2 Q \\ & \quad + \frac{4bc}{n\big(a - \frac{1}{n}\big)}|\mathring{A}_I|^2 - \frac{4bc}{n\big(a - \frac{1}{n}\big)} Q - \frac{2b^2c^2}{n\big(a - \frac{1}{n}\big)} - \frac{2}{n\big(a - \frac{1}{n}\big)} Q^2.\end{aligned}$$

For the last two lines of the right side of (5-2), the computations in [1] give

$$(5\text{-}4)\quad \sum_{\alpha > n+1}\Big(\sum_{i,j} \mathring{A}_{ij}^{n+1}\mathring{A}_{ij}^{\alpha}\Big)^2 + \sum_{\alpha > n+1} N(A^{n+1}\mathring{A}^\alpha - \mathring{A}^\alpha A^{n+1}) \le 2|\mathring{A}_H|^2|\mathring{A}_I|^2$$

and

$$2\sum_{\alpha,\beta>n+1}\Big(\sum_{i,j}\mathring{A}^{\alpha}_{ij}\mathring{A}^{\beta}_{ij}\Big)^2+2\sum_{\alpha,\beta>n+1}N(\mathring{A}^{\alpha}\mathring{A}^{\beta}-\mathring{A}^{\beta}\mathring{A}^{\alpha})\le 3|\mathring{A}_I|^4. \tag{5-5}$$

Combining (5-2)–(5-5), we have

$$\begin{aligned}2R_1-2aR_2\le{}&\bigg(6-\frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)|\mathring{A}_H|^2|\mathring{A}_I|^2+\bigg(3-\frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)|\mathring{A}_I|^4\\&+\bigg(2+\frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)|\mathring{A}_H|^2Q+\frac{4}{n\big(a-\frac{1}{n}\big)}|\mathring{A}_I|^2Q\\&+\bigg(2b+\frac{2b}{n\big(a-\frac{1}{n}\big)}\bigg)c|\mathring{A}_H|^2+\frac{4bc}{n\big(a-\frac{1}{n}\big)}|\mathring{A}_I|^2\\&-\frac{4bc}{n\big(a-\frac{1}{n}\big)}Q-\frac{2b^2c^2}{n\big(a-\frac{1}{n}\big)}-\frac{2}{n\big(a-\frac{1}{n}\big)}Q^2.\end{aligned}\tag{5-6}$$

We also calculate that

$$\begin{aligned}-2nc|\mathring{A}|^2-2n\Big(a-\frac{1}{n}\Big)c|H|^2&=-2nc|\mathring{A}|^2-2nc(|\mathring{A}|^2-bc-Q)\\&=-4nc|\mathring{A}_H|^2-4nc|\mathring{A}_I|^2+2nc^2b+2ncQ.\end{aligned}\tag{5-7}$$

Substituting (5-6) and (5-7) into (5-1) yields

$$\begin{aligned}\Big(\frac{\partial}{\partial t}-\Delta\Big)Q\\ \le -2\big(|\nabla A|^2-a|\nabla H|^2\big)&\\+\bigg(6-\frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)&|\mathring{A}_H|^2|\mathring{A}_I|^2+\bigg(3-\frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)|\mathring{A}_I|^4\\+\bigg(2b+\frac{2b}{n\big(a-\frac{1}{n}\big)}-4n\bigg)&c|\mathring{A}_H|^2+\bigg(\frac{4b}{n\big(a-\frac{1}{n}\big)}-4n\bigg)c|\mathring{A}_I|^2\\+\bigg(\frac{2\big(a-\frac{2}{n}\big)}{a-\frac{1}{n}}&|\mathring{A}_H|^2+\frac{4}{n\big(a-\frac{1}{n}\big)}(|\mathring{A}|^2-bc)+2nc\bigg)Q\\+\bigg(2nb-\frac{2b^2}{n\big(a-\frac{1}{n}\big)}\bigg)&c^2-\frac{2}{n\big(a-\frac{1}{n}\big)}Q^2.\end{aligned}\tag{5-8}$$

For $n\ge 8$, we choose $a=\frac{1}{n-2}$ and $b=4$, satisfying the condition $a<\frac{3}{n+2}$. We also have the following gradient inequality for submanifolds in $\mathbb{F}^N(c)$:

$$|\nabla A|^2\ge\frac{3}{n+2}|\nabla H|^2. \tag{5-9}$$

From this inequality, we derive $-2\big(|\nabla A|^2 - \frac{1}{n-2}|\nabla H|^2\big) \le 0$. Consequently, we discard this term in (5-8).

We will analyze the following three cases separately: $c = 0$, $c < 0$ and $c > 0$ for the constant curvature c of the space form $\mathbb{F}^N(c)$.

Case $c = 0$: Since $|H| > 0$ at $t = 0$, there is a $t_0 > 0$ such that $|H| > 0$ for $t \in [0, t_0]$. At any point in M_t for $t \in [0, t_0]$, (5-8) implies

$$\Big(\frac{\partial}{\partial t} - \Delta\Big)Q \le \bigg(6 - \frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)|\mathring{A}_H|^2|\mathring{A}_I|^2 + \bigg(3 - \frac{2}{n\big(a-\frac{1}{n}\big)}\bigg)|\mathring{A}_I|^4 - \frac{2}{n\big(a-\frac{1}{n}\big)}Q^2 + \bigg(\frac{2\big(a-\frac{2}{n}\big)}{a-\frac{1}{n}}|\mathring{A}_H|^2 + \frac{4}{n\big(a-\frac{1}{n}\big)}|\mathring{A}|^2\bigg)Q.$$

Since $a = \frac{1}{n-2}$, one has

$$6 - \frac{2}{n\big(a-\frac{1}{n}\big)} = -(n-8) \le 0, \qquad -\frac{2}{n\big(a-\frac{1}{n}\big)} = -(n-2) < 0.$$

Therefore,

$$\Big(\frac{\partial}{\partial t} - \Delta\Big)Q \le -(n-5)|\mathring{A}_I|^4 + \big(-(n-4)|\mathring{A}_H|^2 + 2(n-2)|\mathring{A}|^2\big)Q. \tag{5-10}$$

By the maximum principle, (5-10) implies that $Q \le 0$ is preserved on $[0, t_0]$.

By (5-10) we have, on $[0, t_0]$,

$$\Big(\frac{\partial}{\partial t} - \Delta\Big)Q \le -(n-5)|\mathring{A}_I|^4 - (n-4)|\mathring{A}_H|^2 Q. \tag{5-11}$$

Since $-(n-5)|\mathring{A}_I|^4 \le 0$, by the strong maximum principle, either $Q < 0$ for all $t \in (0, t_0]$, or $Q \equiv 0$ for all $t \in [0, t_0]$.

Recall that we discarded the nonpositive term $-2\big(|\nabla A|^2 - \frac{1}{n-2}|\nabla H|^2\big)$ in the above computations. If we retain this item, then (5-11) becomes

$$\Big(\frac{\partial}{\partial t} - \Delta\Big)Q \le -2\Big(|\nabla A|^2 - \frac{1}{n-2}|\nabla H|^2\Big) - (n-5)|\mathring{A}_I|^4 - (n-4)|\mathring{A}_H|^2 Q.$$

If $Q \equiv 0$ for all $t \in [0, t_0]$, this inequality implies $|\nabla A|^2 - \frac{1}{n-2}|\nabla H|^2 \equiv 0$ and $\mathring{A}_I \equiv 0$. As $\frac{1}{n-2} < \frac{3}{n+2}$ for $n \ge 8$, it follows from (5-9) that $\nabla A \equiv 0$, i.e., the second fundamental form is parallel. Therefore, M lies in an $(n+1)$-dimensional affine subspace of $\mathbb{R}^N$, and the second fundamental form of the hypersurface M in $\mathbb{R}^{n+1}$ is parallel. Then by [18], $M = \mathbb{S}^k(r) \times \mathbb{R}^{n-k}$ for $k = 0, 1, 2, 3, \dots, n$. Since M is compact, we conclude that $k = n$ and $M = \mathbb{S}^n(r)$. This implies $|A|^2 \equiv \frac{1}{n}|H|^2$, which is a contradiction to that $Q \equiv 0$. Therefore, $Q < 0$ for all $t \in (0, t_0]$. By the maximum principle, $Q < 0$ for all $t > 0$.

Case $c < 0$: At $t = 0$, the pinching condition implies that $|H| > 0$. Then there is a $t_0 > 0$ such that $|H| > 0$ for $t \in [0, t_0]$.

As $a = \frac{1}{n-2}$ and $b = 4$, (5-8) implies that, at any point in M_t for $t \in [0, t_0]$,

$$\left(\frac{\partial}{\partial t} - \Delta\right) Q \le \left[-(n-4)|\mathring{A}_H|^2 + 2(n-2)(|\mathring{A}|^2 - bc) + 2nc\right] Q - 8(n-4)c^2. \tag{5-12}$$

As $-8(n-4)c^2 < 0$, by the maximum principle, $Q \le 0$ is preserved on $[0, t_0]$. By the strong maximum principle, either $Q < 0$ for all $t \in (0, t_0]$, or $Q \equiv 0$ for all $t \in [0, t_0]$. For the second case, (5-12) implies

$$0 \le -8(n-4)c^2 < 0,$$

which is impossible. Therefore, $Q < 0$ for all $t \in (0, t_0]$. By the maximum principle, $Q < 0$ for all $t > 0$.

<u>Case $c > 0$</u>: We first consider the point $|H| \ne 0$. For $x, y \ge 0$, define the function

$$\begin{aligned} F(x, y) = {} & \left(6 - \frac{2}{n\left(a - \frac{1}{n}\right)}\right) xy + \left(3 - \frac{2}{n\left(a - \frac{1}{n}\right)}\right) y^2 \\ & + \left(2b + \frac{2b}{n\left(a - \frac{1}{n}\right)} - 4n\right) cx + \left(\frac{4b}{n\left(a - \frac{1}{n}\right)} - 4n\right) cy + \left(2nb - \frac{2b^2}{n\left(a - \frac{1}{n}\right)}\right) c^2. \end{aligned}$$

As $a = \frac{1}{n-2}$ and $b = 4$, $F(x, y)$ can be expressed as

$$F(x, y) = -(n-8)xy - (n-5)y^2 + 4(n-4)cy - [8(n-4) - \epsilon]c^2 - \epsilon c^2.$$

As $n \ge 8$, we have $-(n-8) \le 0$. Hence, we only need to consider the function

$$f(y) = -(n-5)y^2 + 4(n-4)cy - [8(n-4) - \epsilon]c^2.$$

The discriminant of this quadratic satisfies

$$\triangle = 16(n-4)^2 - 4(n-5)[8(n-4) - \epsilon] < 0$$

for a sufficiently small $\epsilon > 0$. So $f(y) < 0$ for this ϵ. Therefore,

$$F\left(|\mathring{A}_H|^2, |\mathring{A}_I|^2\right) < -\epsilon c^2.$$

Thus, (5-8) implies

$$\left(\frac{\partial}{\partial t} - \Delta\right) Q \le \left[-(n-4)|\mathring{A}_H|^2 + 2(n-2)(|\mathring{A}|^2 - bc) + 2nc\right] Q - \epsilon c^2. \tag{5-13}$$

If $|H| = 0$ at a point, then $Q = |A|^2 - bc$, which implies $|\mathring{A}|^2 = |A|^2 = Q + bc$. Therefore, by the Li–Li inequality [25],

$$\left(\frac{\partial}{\partial t} - \Delta\right) Q \le 3|\mathring{A}|^4 - 2nc|\mathring{A}|^2 = 3|\mathring{A}|^2(Q + bc) - 2nc|\mathring{A}|^2 3|\mathring{A}|^2 Q + (3b - 2n)c|\mathring{A}|^2.$$

As $b = 4$, $3b - 2n = -2(n-6) < 0$. Hence,

$$\left(\frac{\partial}{\partial t} - \Delta\right) Q \le \left(3|\mathring{A}|^2 - 2(n-6)c\right) Q - 8(n-6)c^2. \tag{5-14}$$

Therefore, by the maximum principle, we see from (5-13) and (5-14) that $Q \le 0$ is preserved.

Moreover, by (5-13) and (5-14), one always has

$$\Big(\frac{\partial}{\partial t} - \Delta\Big) Q \le -\gamma_1 (|\mathring{A}|^2 + c) Q - \gamma_2$$

for positive constants γ_1 and γ_2. Therefore, by the strong maximum principle, either $Q < 0$ for all $t > 0$, or $Q \equiv 0$ for all $t \ge 0$. However, for the second case, the above inequality implies

$$0 \le -\gamma_1 < 0,$$

which is impossible. Therefore, $Q < 0$ for all $t > 0$. □

Remark 5.2. For $4 \le n \le 7$, by a similar computation, the pinching condition $|A|^2 \le a|H|^2 + bc$ for certain $\frac{1}{n-1} < a < \frac{1}{n-2}$ and $2 < b < 4$ is also preserved.

Proof of Theorem 1.2.. When $n = 5, 6$, if $c > 0$, or $c = 0$ and $|H| > 0$, by Theorem 4.1, M is a topological sphere. Since the differentiable structure on a topological sphere of dimension 5 or 6, M is diffeomorphic to the standard sphere.

When $n \ge 9$, combining Theorem 1.1 and Lemma 5.1, we conclude that M is diffeomorphic to the standard sphere. □

6. More differentiable sphere theorems for compact submanifolds

We present two more differentiable sphere theorems for compact submanifolds in Riemannian manifolds.

Theorem 6.1. *Let M^n $(n \ge 4, n \ne 7, 8)$ be an n-dimensional compact simply connected submanifold in an N-dimensional Riemannian manifold $\overline{M}^N$. Suppose that one of the following conditions holds*:

(i) $|A|^2 < \frac{10}{3}\big(\overline{\mathrm{Ric}}_{\min} - \frac{5N-11}{5}\overline{K}_{\max}\big) + \frac{|H|^2}{n-2}$.

(ii) $|A|^2 < \frac{7N}{6s}\big(\overline{\mathrm{Ric}}^{[s]}_{\min} - \frac{s(7N^2-7N-24)}{7N}\overline{K}_{\max}\big) + \frac{|H|^2}{n-2}$ *for some* $1 \le s \le N$.

(iii) $|A|^2 < \frac{7N(N-1)}{6}\big(\overline{R}_0 - \frac{7N^2-7N-24}{7N(N-1)}\overline{K}_{\max}\big) + \frac{|H|^2}{n-2}$.

Then M is diffeomorphic to $\mathbb{S}^n$.

Proof. As in Theorem 1.1, we need only show that M has positive isotropic curvature.

(i) For any $x \in \overline{M}$, suppose $u, v \in U_x\overline{M}$ be two orthonormal vectors such that $\overline{K}(\pi) = \overline{K}_{\min}(x)$, where $\pi = \mathrm{span}\{u, v\}$. Let $V_x^k \subset T_x\overline{M}$ be a k-dimensional subspace such that $v \in V_x^k$ and $u \perp V_x^k$. Define $e_{A_0} = u,\ e_{A_1} = v$, and let $\{e_{A_1}, \dots, e_{A_k}\}$

be an orthonormal basis of V_x^k. We have

$$\overline{\mathrm{Ric}}^{(k)}\big([e_{A_0}, e_{A_1}, \dots, e_{A_k}]\big) = \sum_{p=1}^{k} \bar{R}_{A_0A_pA_0A_p} = \bar{R}_{A_0A_1A_0A_1} + \sum_{p=2}^{k} \bar{R}_{A_0A_pA_0A_p} \le \bar{K}_{\min} + (k-1)\bar{K}_{\max}.$$

By the definition of $\overline{\mathrm{Ric}}^{(k)}_{\min}$, one has

$$\bar{K}_{\min} \ge \overline{\mathrm{Ric}}^{(k)}_{\min} - (k-1)\bar{K}_{\max}. \tag{6-1}$$

Suppose $\{e_1, e_2, e_3, e_4\}$ is an orthonormal four-frame. Extend $\{e_1, e_2, e_3, e_4\}$ to an orthonormal $(k+1)$-frame $\{e_1, \dots, e_{k+1}\}$. Then

$$\overline{\mathrm{Ric}}^{(k)}([e_1, e_2, \dots, e_{k+1}]) = \sum_{p=2}^{k+1} \bar{R}_{1p1p} = \bar{R}_{1313} + \bar{R}_{1414} + \Big(\bar{R}_{1212} + \sum_{p=5}^{k+1} \bar{R}_{1p1p}\Big) \le \bar{R}_{1313} + \bar{R}_{1414} + (k-2)\bar{K}_{\max}.$$

This implies

$$\bar{R}_{1313} + \bar{R}_{1414} \ge \overline{\mathrm{Ric}}^{(k)}_{\min} - (k-2)\bar{K}_{\max}. \tag{6-2}$$

Combining (6-1) and (6-2) with Berger's inequality, we obtain

$$\begin{aligned} \bar{R}_{1313} + \bar{R}_{1414} - \bar{R}_{1234} &\ge \overline{\mathrm{Ric}}^{(k)}_{\min} - (k-2)\bar{K}_{\max} - \tfrac{2}{3}\big(k\bar{K}_{\max} - \overline{\mathrm{Ric}}^{(k)}_{\min}\big) \\ &= \tfrac{5}{3}\big[\overline{\mathrm{Ric}}^{(k)}_{\min} - \big(k - \tfrac{6}{5}\big)\,\bar{K}_{\max}\big]. \end{aligned} \tag{6-3}$$

Similarly,

$$\bar{R}_{2323} + \bar{R}_{2424} - \bar{R}_{1234} \ge \tfrac{5}{3}\big[\overline{\mathrm{Ric}}^{(k)}_{\min} - \big(k - \tfrac{6}{5}\big)\,\bar{K}_{\max}\big]. \tag{6-4}$$

Combining (3-2), (6-3) and (6-4), we obtain

$$\begin{aligned} &R_{1313} + R_{1414} + R_{2323} + R_{2424} - 2R_{1234} \\ &\quad\ge \tfrac{10}{3}\big[\overline{\mathrm{Ric}}^{(k)}_{\min} - \big(k - \tfrac{6}{5}\big)\,\bar{K}_{\max}\big] + \sum_{\alpha}\bigg[\frac{\big(\sum_{i=1}^{n} A^{\alpha}_{ii}\big)^2}{n-2} + \sum_{i\ne j}(A^{\alpha}_{ij})^2 - \sum_{i,j=1}^{n}(A^{\alpha}_{ij})^2 \\ &\qquad\qquad - 2(A^{\alpha}_{13})^2 - 2(A^{\alpha}_{14})^2 - 2(A^{\alpha}_{23})^2 - 2(A^{\alpha}_{24})^2\bigg] \\ &\quad\ge \tfrac{10}{3}\Big[\overline{\mathrm{Ric}}^{(k)}_{\min} - \big(k - \tfrac{6}{5}\big)\,\bar{K}_{\max}\Big] + \frac{|H|^2}{n-2} - |A|^2. \end{aligned}$$

Choosing $k = N-1$ and combining the assumption, we know that M has positive isotropic curvature.

(ii) For any $x \in \overline{M}$, let $u, v \in U_x\overline{M}$ be two orthonormal vectors such that $\overline{K}(\pi) = \overline{K}_{\min}(x)$ with $\pi = \operatorname{span}\{u, v\}$. Let $\{e_{A_0}, e_{A_1}, \ldots, e_{A_k}\}$ be an orthonormal $(k+1)$-frame such that $e_{A_0} = u$, $e_{A_1} = v$. Then

$$\begin{aligned}\overline{R}^{(k)}\big([e_{A_0}, e_{A_1}, \ldots, e_{A_k}]\big) &= \sum_{p=0}^{k}\sum_{q=0}^{k} \overline{R}_{A_pA_qA_pA_q}\\ &= \overline{R}_{A_0A_1A_0A_1} + \overline{R}_{A_1A_0A_1A_0}\\ &\qquad + \sum_{p=0}^{k}\sum_{q=0}^{k} \overline{R}_{A_pA_qA_pA_q} - (\overline{R}_{A_0A_1A_0A_1} + \overline{R}_{A_1A_0A_1A_0})\\ &\le 2\overline{K}_{\min} + [k(k+1)-2]\,\overline{K}_{\max}.\end{aligned}$$

This implies

$$\overline{K}_{\min} \ge \tfrac{1}{2}\big(\overline{R}^{(k)}_{\min} - [k(k+1)-2]\,\overline{K}_{\max}\big). \tag{6-5}$$

By the definition of $\overline{R}^{(k,s)}$, one has

$$\frac{\overline{R}^{(k)}_{\min}}{k(k+1)} \ge \frac{\overline{R}^{(k,s)}_{\min}}{ks}. \tag{6-6}$$

Combining (6-5) and (6-6), we have

$$\overline{K}_{\min} \ge \tfrac{1}{2}\big(\tfrac{k+1}{s}\overline{R}^{(k,s)}_{\min} - [k(k+1)-2]\,\overline{K}_{\max}\big). \tag{6-7}$$

Suppose $\{e_1, e_2, e_3, e_4\}$ is an orthonormal four-frame. Extend $\{e_1, e_2, e_3, e_4\}$ to a $(k+1)$-frame $\{e_1, \ldots, e_{k+1}\}$. As $\overline{R}^{(k)} = \overline{R}^{(k,k+1)}$, we have

$$\begin{aligned}\overline{R}^{(k)}([e_1, \ldots, e_{k+1}]) &= \sum_{i=1}^{k+1}\sum_{j=1}^{k+1} \overline{R}_{ijij}\\ &= 2\big(\overline{R}_{1313} + \overline{R}_{1414} + \overline{R}_{2323} + \overline{R}_{2424}\big)\\ &\qquad + \sum_{i=1}^{k+1}\sum_{j=1}^{k+1} \overline{R}_{ijij} - 2(\overline{R}_{1313} + \overline{R}_{1414} + \overline{R}_{2323} + \overline{R}_{2424})\\ &\le 2\big(\overline{R}_{1313} + \overline{R}_{1414} + \overline{R}_{2323} + \overline{R}_{2424}\big) + [k(k+1)-8]\,\overline{K}_{\max}.\end{aligned}$$

By (6-6), this implies that for $1 \le s \le N$, there holds

$$\overline{R}_{1313} + \overline{R}_{1414} + \overline{R}_{2323} + \overline{R}_{2424} \ge \tfrac{1}{2}\big(\tfrac{k+1}{s}\overline{R}^{(k,s)}_{\min} - [k(k+1)-8]\overline{K}_{\max}\big). \tag{6-8}$$

Combining (3-5), (6-7) and (6-8) with Berger's inequality, we obtain

$$\begin{aligned} R_{1313}+R_{1414}+R_{2323}+R_{2424}-2R_{1234} \\ \geq \tfrac{1}{2}\big\{\tfrac{k+1}{s}\overline{R}^{(k,s)}_{\min}-[k(k+1)-8]\,\overline{K}_{\max}\big\} \\ -\tfrac{2}{3}\big[k(k+1)\overline{K}_{\max}-\tfrac{(k+1)}{s}\overline{R}^{(k,s)}_{\min}\big]+\tfrac{|H|^2}{n-2}-|A|^2 \\ \geq \tfrac{7}{6}\big[\tfrac{k+1}{s}\overline{R}^{(k,s)}_{\min}-\big(k^2+k-\tfrac{24}{7}\big)\,\overline{K}_{\max}\big]+\tfrac{|H|^2}{n-2}-|A|^2. \end{aligned}$$

Choosing $k=N-1$ and combining the assumption, we know that M has positive isotropic curvature.

(iii) Substituting $s=N$ into (ii), we derive that M has positive isotropic curvature. □

Theorem 6.2. *Let M^n $(n\geq 4, n\neq 7,8)$ be an n-dimensional compact simply connected submanifold in an N-dimensional Riemannian manifold $\overline{M}^N$. Suppose that one of the following conditions holds*:

(i) $|A|^2<\frac{4(N+2)}{3}\big(\overline{K}_{\min}-\frac{1}{N+2}\overline{\mathrm{Ric}}_{\max}\big)+\frac{|H|^2}{n-2}$.

(ii) $|A|^2<\frac{2(sN-s+6)}{3}\big(\overline{K}_{\min}-\frac{1}{sN-s+6}\overline{\mathrm{Ric}}^{[s]}_{\max}\big)+\frac{|H|^2}{n-2}$ *for some* $2\leq s\leq N$.

(iii) $|A|^2<\frac{2(N^2-N+6)}{3}\big(\overline{K}_{\min}-\frac{N^2-N}{N^2-N+6}\overline{R}_0\big)+\frac{|H|^2}{n-2}$.

Then M is diffeomorphic to $\mathbb{S}^n$.

Proof. As in Theorem 1.1, we need only show that M has positive isotropic curvature.

(i) As in the proof Theorem 6.1(i), one has

$$\overline{K}_{\max}\leq\overline{\mathrm{Ric}}^{(k)}_{\max}-(k-1)\overline{K}_{\min}.$$

Suppose $\{e_1,e_2,e_3,e_4\}$ is an orthonormal four-frame. Combining (3-5) with the above inequality, we obtain

$$\begin{aligned} R_{1313}+R_{1414}+R_{2323}+R_{2424}-2R_{1234} &\geq \tfrac{16}{3}\big(\overline{K}_{\min}-\tfrac{1}{4}\overline{K}_{\max}\big)+\tfrac{|H|^2}{n-2}-|A|^2 \\ &\geq \tfrac{4}{3}\big((k+3)\overline{K}_{\min}-\overline{\mathrm{Ric}}^{(k)}_{\max}\big)+\tfrac{|H|^2}{n-2}-|A|^2. \end{aligned}$$

Choosing $k=N-1$ and combining with the assumption, we see that M has positive isotropic curvature.

(ii) As in the proof of Theorem 6.1(ii), one has that for $2\leq s\leq k+1$,

$$\overline{K}_{\max}\leq\tfrac{1}{2}\big(\overline{R}^{(k,s)}_{\max}-(ks-2)\overline{K}_{\min}\big).$$

Combining this with (3-5), we obtain

$$\begin{aligned} R_{1313}+R_{1414}+R_{2323}+R_{2424}-2R_{1234} &\geq \tfrac{16}{3}\big(\overline{K}_{\min}-\tfrac{1}{4}\overline{K}_{\max}\big)+\tfrac{|H|^2}{n-2}-|A|^2 \\ &\geq \tfrac{2}{3}\big((ks+6)\overline{K}_{\min}-\overline{R}^{(k,s)}_{\max}\big)+\tfrac{|H|^2}{n-2}-|A|^2. \end{aligned}$$

Choosing $k = N - 1$ and combining with the assumption, we know that M has positive isotropic curvature.

(iii) Substituting $s = N$ into (ii), we derive that M has positive isotropic curvature. □

Remark 6.3. From the proofs of Theorems 6.1 and 6.2, similar differentiable sphere theorems can be established for compact submanifolds under pinching conditions involving the k-th Ricci curvature or the (k, s)-curvature.

Acknowledgements

The authors thank the referees for their valuable comments and suggestions. This research was supported by the National Natural Science Foundation of China, Grant Nos. 12471051, 12171423, 12071424.

References

[1] B. Andrews and C. Baker, "Mean curvature flow of pinched submanifolds to spheres", *J. Differential Geom.* **85**:3 (2010), 357–395. MR

[2] C. Baker, "The mean curvature flow of submanifolds of high codimension", preprint, 2011. arXiv 1104.4409

[3] M. Berger, "Les variétés Riemanniennes (1/4)-pincées", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (3) **14** (1960), 161–170. MR

[4] S. Brendle, "Ricci flow with surgery on manifolds with positive isotropic curvature", *Ann. of Math.* (2) **190**:2 (2019), 465–559. MR

[5] S. Brendle and R. M. Schoen, "Classification of manifolds with weakly 1/4-pinched curvatures", *Acta Math.* **200**:1 (2008), 1–13. MR

[6] S. Brendle and R. Schoen, "Manifolds with 1/4-pinched curvature are space forms", *J. Amer. Math. Soc.* **22**:1 (2009), 287–307. MR

[7] Z. Chen, "Manifolds with positive isotropic curvature of dimension at least nine", preprint, 2024. arXiv 2410.21078

[8] B. L. Chen and X. P. Zhu, "Ricci flow with surgery on four-manifolds with positive isotropic curvature", *J. Differential Geom.* **74**:2 (2006), 177–264. MR

[9] B. L. Chen, S. H. Tang, and X. P. Zhu, "Complete classification of compact four-manifolds with positive isotropic curvature", *J. Differential Geom.* **91**:1 (2012), 41–80. MR

[10] K. D. Elworthy and S. Rosenberg, "Homotopy and homology vanishing theorems and the stability of stochastic flows", *Geom. Funct. Anal.* **6**:1 (1996), 51–78. MR

[11] H. P. Fu and H. W. Xu, "Vanishing and topological sphere theorems for submanifolds in a hyperbolic space", *Internat. J. Math.* **19**:7 (2008), 811–822. MR

[12] J. R. Gu, *Some results on geometry and topology of manifolds*, Ph.D. dissertation, Zhejiang University, 2010.

[13] J. R. Gu and H. W. Xu, "The sphere theorems for manifolds with positive scalar curvature", *J. Differential Geom.* **92**:3 (2012), 507–545. MR

[14] R. S. Hamilton, "Four-manifolds with positive isotropic curvature", *Comm. Anal. Geom.* **5**:1 (1997), 1–92. MR

[15] H. Hopf, “Zum Clifford–Kleinschen Raumproblem”, *Math. Ann.* **95**:1 (1926), 313–339. MR

[16] W. Killing, “Ueber die Clifford–Klein’schen Raumformen”, *Math. Ann.* **39**:2 (1891), 257–278. MR

[17] W. Klingenberg, “Über Riemannsche Mannigfaltigkeiten mit positiver Krümmung”, *Comment. Math. Helv.* **35** (1961), 47–54. MR

[18] H. B. Lawson, Jr., “Local rigidity theorems for minimal hypersurfaces”, *Ann. of Math.* (2) **89** (1969), 187–197. MR

[19] H. B. Lawson, Jr. and J. Simons, “On stable currents and their application to global problems in real and complex geometry”, *Ann. of Math.* (2) **98** (1973), 427–450. MR

[20] J. M. Lee, *Introduction to smooth manifolds*, 2nd ed., Graduate Texts in Mathematics **218**, Springer, 2013. MR

[21] L. Lei and H. W. Xu, “Mean curvature flow of arbitrary codimension in spheres and sharp differentiable sphere theorem”, preprint, 2015. arXiv 1506.06371

[22] L. Lei and H. W. Xu, “An optimal convergence theorem for mean curvature flow of arbitrary codimension in hyperbolic spaces”, preprint, 2015. arXiv 1503.06747

[23] L. Lei and H. W. Xu, “New developments in mean curvature flow of arbitrary codimension inspired by Yau rigidity theory”, pp. 327–348 in *Proceedings of the Seventh International Congress of Chinese Mathematicians*, Adv. Lect. Math. **43**, International Press, Somerville, MA, 2019. MR

[24] S. Y. Li, *Studies on curvature flows in geometry and differentiable sphere theorems*, Ph.D. dissertation, Zhejiang University, 2024.

[25] A. M. Li and J. M. Li, “An intrinsic rigidity theorem for minimal submanifolds in a sphere”, *Arch. Math.* (*Basel*) **58**:6 (1992), 582–594. MR

[26] K. F. Liu, H. W. Xu, F. Ye, and E. T. Zhao, “Mean curvature flow of higher codimension in hyperbolic spaces”, *Comm. Anal. Geom.* **21**:3 (2013), 651–669. MR

[27] M. J. Micallef and J. D. Moore, “Minimal two-spheres and the topology of manifolds with positive curvature on totally isotropic two-planes”, *Ann. of Math.* (2) **127**:1 (1988), 199–227. MR

[28] K. Naff, “A planarity estimate for pinched solutions of mean curvature flow”, *Duke Math. J.* **171**:2 (2022), 443–482. MR

[29] T. Ôtsuki, “Minimal hypersurfaces in a Riemannian manifold of constant curvature”, *Amer. J. Math.* **92** (1970), 145–173. MR

[30] H. E. Rauch, “A contribution to differential geometry in the large”, *Ann. of Math.* (2) **54** (1951), 38–55. MR

[31] K. Shiohama and H. W. Xu, “The topological sphere theorem for complete submanifolds”, *Compositio Math.* **107**:2 (1997), 221–232. MR

[32] K. Shiohama and H. W. Xu, “A general rigidity theorem for complete submanifolds”, *Nagoya Math. J.* **150** (1998), 105–134. MR

[33] S. Smale, “Generalized Poincaré’s conjecture in dimensions greater than four”, *Ann. of Math.* (2) **74** (1961), 391–406. MR

[34] E. H. Spanier, *Algebraic topology*, McGraw-Hill, New York, 1966. MR

[35] Y. L. Xin, “An application of integral currents to the vanishing theorems”, *Sci. Sinica Ser. A* **27**:3 (1984), 233–241. MR

[36] H. W. Xu, *Pinching theorems*, *global pinching theorems*, *and eigenvalues for Riemannian submanifolds*, Ph.D. dissertation, Fudan University, 1990.

[37] H. W. Xu, "A pinching constant of Simons' type and the problem of isometric immersion", *Chinese Ann. Math. Ser. A* **12**:3 (1991), 261–269. MR

[38] H. W. Xu, "A rigidity theorem for submanifolds with parallel mean curvature in a sphere", *Arch. Math.* (*Basel*) **61**:5 (1993), 489–496. MR

[39] H. W. Xu and J. R. Gu, "An optimal differentiable sphere theorem for complete manifolds", *Math. Res. Lett.* **17**:6 (2010), 1111–1124. MR

[40] H. W. Xu and E. T. Zhao, "Topological and differentiable sphere theorems for complete submanifolds", *Comm. Anal. Geom.* **17**:3 (2009), 565–585. MR

Received September 5, 2025. Revised January 31, 2026.

JUAN LI
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
juanli@zju.edu.cn

HONGWEI XU
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
xuhw@zju.edu.cn

ENTAO ZHAO
CENTER OF MATHEMATICAL SCIENCES
ZHEJIANG UNIVERSITY
HANGZHOU
CHINA
zhaoet@zju.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.453

QUASITRIANGULAR AND FACTORIZABLE POISSON BIALGEBRAS

YUANCHANG LIN AND DILEI LU

We introduce the notions of quasitriangular and factorizable Poisson bialgebras. A factorizable Poisson bialgebra induces a factorization of the underlying Poisson algebra. We prove that the Drinfeld classical double of a Poisson bialgebra naturally admits a factorizable Poisson bialgebra structure. Furthermore, we introduce the notion of quadratic Rota–Baxter Poisson algebras and show that a quadratic Rota–Baxter Poisson algebra of zero weight induces a triangular Poisson bialgebra. Moreover, we establish a one-to-one correspondence between factorizable Poisson bialgebras and quadratic Rota–Baxter Poisson algebras of nonzero weights. Finally, we establish the quasitriangular and factorizable theories for differential antisymmetric infinitesimal (ASI) bialgebras, and construct quasitriangular and factorizable Poisson bialgebras from quasitriangular and factorizable (commutative and cocommutative) differential ASI bialgebras respectively.

1. Introduction

Poisson algebras serve as fundamental structures in various areas of mathematics and mathematical physics, including Poisson geometry [28; 29], classical and quantum mechanics [3; 9; 22], algebraic geometry [13; 23], quantization theory [14; 17] and quantum groups [8; 12]. A Poisson algebra is both a Lie algebra and a commutative associative algebra which are compatible in a certain sense.

MSC2020: primary 16S32, 16T10, 17A30, 17B62, 17B63; secondary 17B38, 81R60.
Keywords: Poisson algebra, Poisson Yang–Baxter equation, quasitriangular Poisson bialgebra, factorizable Poisson bialgebra, quadratic Rota–Baxter Poisson algebra, differential antisymmetric infinitesimal bialgebra.

Definition 1.1 [19; 29]. A *Poisson algebra* is a triple $(A, [\ ,\ ], \cdot)$, where $(A, [\ ,\ ])$ is a Lie algebra and $(A, \cdot)$ is a commutative associative algebra satisfying

$$[a, b \cdot c] = [a, b] \cdot c + b \cdot [a, c], \quad \forall a, b, c \in A.$$

Both Lie algebras and associative algebras possess well-developed theories of bialgebras, which have found extensive applications in various mathematical fields. In the Lie algebra setting, Lie bialgebras, introduced by Drinfeld [11], play a fundamental role as the infinitesimalization of Poisson Lie groups [8; 16]. On the side of associative algebras, the notion of infinitesimal bialgebras was introduced by Joni and Rota to provide an algebraic framework for the calculus of divided differences [15]. Variants of this notion, such as balanced infinitesimal bialgebras (also termed antisymmetric infinitesimal bialgebras or associative D-bialgebras in different contexts [2; 4; 30]), have been systematically studied by Aguiar [1; 2], who established their close analogy with Lie bialgebras. These structures have been found significant applications in combinatorial mathematics.

Building on these ideas, a unified bialgebra theory for Poisson algebras, called Poisson bialgebras, was later developed in [21], combining aspects of both Lie and infinitesimal bialgebras.

Meanwhile, In the realm of Lie bialgebras, quasitriangular Lie bialgebra structures have been pivotal objects in mathematical physics [12; 26]. Among these quasitriangular Lie bialgebra structures, factorizable Lie bialgebras constitute a particularly important subclass, linking classical r-matrices to factorization problems and playing a key role in integrable systems [5; 24; 25]. Recently, quasitriangular and factorizable theories has been extended to antisymmetric infinitesimal bialgebras with [27] introducing quasitriangular and factorizable antisymmetric infinitesimal bialgebras.

Naturally, we try to establish the quasitriangular and factorizable theories in the context of Poisson bialgebras, synthesizing concepts from both quasitriangular Lie bialgebras and quasitriangular antisymmetric infinitesimal bialgebras. More precisely, we introduce the notion of quasitriangular Poisson bialgebras based on the (ad, L)-invariant condition. In particular, if the symmetric part of the solution of the Poisson Yang–Baxter equation in a quasitriangular Poisson bialgebra is nondegenerate, then a factorizable Poisson bialgebra is obtained. We prove that every factorizable Poisson bialgebra induces a factorization of its underlying Poisson algebra. Furthermore, we establish that the Drinfeld classical double of a Poisson bialgebra is automatically endowed with a canonical factorizable Poisson bialgebra structure.

Recent results have shown that factorizable Lie bialgebras and factorizable antisymmetric infinitesimal bialgebras could be characterized by quadratic Rota–Baxter Lie algebras of nonzero weights and symmetric Rota–Baxter Frobenius algebras of nonzero weights respectively [18; 27]. This motivates our investigation of analogous Rota–Baxter characterizations of factorizable Poisson bialgebras. For this purpose, we introduce the notion of a quadratic Rota–Baxter Poisson algebra by equipping a quadratic Poisson algebra with a Rota–Baxter operator satisfying a compatibility condition. We show that a quadratic Rota–Baxter Poisson algebra of zero weight can give rise to a triangular Poisson bialgebra. Moreover, we establish a one-to-one correspondence between factorizable Poisson bialgebras and quadratic Rota–Baxter Poisson algebras of nonzero weights.

Building on the fact that a Poisson algebra can be obtained from a commutative differential algebra with two commuting derivations [7], Lin, Liu and Bai [20] extended such a connection to the context of bialgebras, utilizing the theory of differential antisymmetric infinitesimal (ASI) bialgebras to construct Poisson bialgebras from commutative and cocommutative differential ASI bialgebras. In this paper, we further investigate this relationship in greater depth. Specifically, we develop the theories of quasitriangular and factorizable differential ASI bialgebras, and apply them to the study of their Poisson bialgebra counterparts. We establish the constructions of quasitriangular and factorizable Poisson bialgebras from quasitriangular and factorizable (commutative and cocommutative) differential ASI bialgebras respectively.

Outline. Section 2 introduces the notion of quasitriangular (triangular) Poisson bialgebras as a special class of coboundary Poisson bialgebras. Section 3 presents the concept of factorizable Poisson bialgebras, which is a distinguished subclass of quasitriangular Poisson bialgebras. We demonstrate that a factorizable Poisson bialgebra induces a factorization of the underlying Poisson algebra. Furthermore, we prove that the Drinfeld classical double of a Poisson bialgebra is naturally endowed with a factorizable Poisson bialgebra structure. Section 4 establishes the Rota–Baxter characterization of factorizable Poisson bialgebras. We introduce the notion of quadratic Rota–Baxter Poisson algebras and establish a one-to-one correspondence between quadratic Rota–Baxter Poisson algebras of nonzero weights and factorizable Poisson bialgebras. Moreover, we show that a quadratic Rota–Baxter Poisson algebra of zero weight can give rise to a triangular Poisson bialgebra. In Section 5, we introduce the notions of quasitriangular and factorizable differential ASI bialgebras, and give Rota–Baxter characterization of factorizable differential ASI bialgebras. The constructions of quasitriangular and factorizable Poisson bialgebras from quasitriangular and factorizable (commutative and cocommutative) differential ASI bialgebras are illustrated respectively.

Throughout this paper, we work over a base field $\mathbb{K}$ of characteristic 0, and all vector spaces and algebras are assumed to be finite-dimensional. We adopt the following conventions and notations.

(1) Let $(A, \diamond)$ be a vector space equipped with a bilinear operation $\diamond : A \otimes A \to A$. Let $L_\diamond(a)$ and $R_\diamond(a)$ denote the left and right multiplication operators, that is,

$$L_\diamond(a)b = R_\diamond(b)a = a \diamond b, \quad \forall a, b \in A.$$

We also simply denote them by $L(a)$ and $R(a)$ without confusion. If $(A, [\ ,\ ])$ is a Lie algebra, we let $\mathrm{ad}_{[,]}(a) = \mathrm{ad}(a)$ denote the adjoint operator, that is,

$$\mathrm{ad}_{[,]}(a)b = \mathrm{ad}(a)b = [a, b], \quad \forall a, b \in A.$$

(2) Let V be a vector space. Denote the flip operator by $\tau : V \otimes V \to V \otimes V$:

$$\tau(u \otimes v) = v \otimes u, \quad \forall u, v \in V.$$

(3) Let $(A, \diamond)$ be a vector space equipped with a bilinear operation $\diamond : A \otimes A \to A$. Let $r = \sum_i a_i \otimes b_i \in A \otimes A$. Set

$$r_{12} = \sum_i a_i \otimes b_i \otimes 1, \quad r_{13} = \sum_i a_i \otimes 1 \otimes b_i, \quad r_{23} = \sum_i 1 \otimes a_i \otimes b_i,$$

where 1 is the unit if $(A, \diamond)$ is unital or a symbol playing a similar role as the unit for the nonunital cases. Further define compound symbols such as $r_{12} \diamond r_{13}$ by

$$r_{12} \diamond r_{13} = \sum_{i,j} a_i \diamond a_j \otimes b_i \otimes b_j, \quad r_{13} \diamond r_{23} = \sum_{i,j} a_i \otimes a_j \otimes b_i \diamond b_j,$$
$$r_{23} \diamond r_{12} = \sum_{i,j} a_i \otimes a_j \diamond b_i \otimes b_j, \quad r_{12} \diamond r_{23} = \sum_{i,j} a_i \otimes b_i \diamond a_j \otimes b_j.$$

(4) Denote the standard pairing between the dual space V^* and V by $\langle\ ,\ \rangle$, so that

$$\langle f, v\rangle := f(v) =: \langle v, f\rangle, \quad \forall f \in V^*,\ v \in V.$$

(5) Let V, W be two vector spaces and $T : V \to W$ be a linear map. Denote the dual map by $T^* : W^* \to V^*$:

$$\langle v, T^*(w^*)\rangle = \langle T(v), w^*\rangle, \quad \forall v \in V, w^* \in W^*.$$

(6) Let A, V be vector spaces. For a linear map $\mu : A \to \mathrm{End}(V)$, define a linear map $\mu^* : A \to \mathrm{End}(V^*)$ by $\mu^*(a) = (\mu(a))^*$, or, more explicitly,

$$\langle \mu^*(a)v^*, u\rangle = \langle v^*, \mu(a)u\rangle, \quad \forall a \in A, u \in V, v^* \in V^*,$$

(7) Let $\Pi_1 = \{\alpha_k : V_1 \to V_1\}_{k=1}^m$ and $\Pi_2 = \{\beta_k : V_2 \to V_2\}_{k=1}^m$ be two m-tuples of commuting linear maps. Then obviously $\{\alpha_k + \beta_k\}_{k=1}^m$ is still an m-tuple of commuting linear maps, which is denoted by $\Pi_1 + \Pi_2$.

2. Quasitriangular Poisson bialgebras

In this section, we recall the notion of coboundary Poisson bialgebras and introduce the notion of quasitriangular Poisson bialgebras as a special case of the coboundary Poisson bialgebras, based on the notion of (ad, L)-invariance. The notion of a quadratic Poisson algebra is also introduced, which gives rise to an (ad, L)-invariant 2-tensor and serves as the foundation for the subsequent notion of a quadratic Rota–Baxter Poisson algebra.

Definition 2.1 [21]. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra, V a vector space and $\rho, \mu : A \to \mathrm{End}(V)$ two linear maps. The triple (V, ρ, μ) is called a *representation* of the Poisson algebra $(A, [\ ,\ ]_A, \cdot_A)$ if the following conditions hold:

(1) (V, ρ) is a representation of the Lie algebra $(A, [\ ,\ ]_A)$, that is, $\rho([a, b]_A) = \rho(a)\rho(b) - \rho(b)\rho(a)$ for all $a, b \in A$.

(2) (V, μ) is a representation of $(A, \cdot_A)$, that is, $\mu(a \cdot_A b) = \mu(a)\mu(b)$ for all $a, b \in A$.

(3) The following equations hold:

$$\rho(a \cdot_A b) = \mu(b)\rho(a) + \mu(a)\rho(b),$$
$$\mu([a, b]_A) = \rho(a)\mu(b) - \mu(b)\rho(a), \quad \forall a, b \in A.$$

Proposition 2.2 [21]. *Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be a Poisson algebra*, V *a vector space and* $\rho, \mu : A \to \mathrm{End}(V)$ *two linear maps. Then* (V, ρ, μ) *is a representation of* $(A, [\ ,\ ]_A, \cdot_A)$ *if and only if there is a Poisson algebra structure on* $A \oplus V$ *with the bilinear operations* $[\ ,\ ]$ *and* $\cdot$ *defined as follows, for all* $a_1, a_2 \in A$ *and* $v_1, v_2 \in V$:

$$[a_1 + v_1, a_2 + v_2] = [a_1, a_2]_A + \rho(a_1)v_2 - \rho(a_2)v_1,$$
$$(a_1 + v_1) \cdot (a_2 + v_2) = a_1 \cdot_A a_2 + \mu(a_1)v_2 + \mu(a_2)v_1,$$

The resulting Poisson algebra structure on $A \oplus V$ *is denoted by* $(A \ltimes_{\rho,\mu} V, [\ ,\ ], \cdot)$ *and called the* ***semidirect product Poisson algebra*** *by* $(A, [\ ,\ ]_A, \cdot_A)$ *and* (V, ρ, μ).

Example 2.3 [21]. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra. Then (A, ad, L) is a representation of $(A, [\ ,\ ]_A, \cdot_A)$, called the *adjoint representation*, and $(A^*, -\mathrm{ad}^*, L^*)$ is also a representation of $(A, [\ ,\ ]_A, \cdot_A)$, called the *coadjoint representation*.

A *Lie bialgebra* is a pair of Lie algebras $(A, [\ ,\ ]_A)$ and $(A^*, [\ ,\ ]_{A^*})$ such that

$$\delta([a, b]_A) = (\mathrm{ad}(a) \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{ad}(a))\delta(b) - (\mathrm{ad}(b) \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{ad}(b))\delta(a), \quad \forall a, b \in A,$$

where $\delta : A \to \mathrm{Alt}^2(A)$ is defined by $\langle \delta(a), x^* \otimes y^* \rangle = \langle a, [x^*, y^*]_{A^*} \rangle$. A Lie bialgebra is denoted by $\big((A, [\ ,\ ]_A), (A^*, [\ ,\ ]_{A^*})\big)$ or $(A, [\ ,\ ]_A, \delta)$.

An *infinitesimal bialgebra* is a pair of associative algebras $(A, \cdot_A)$ and $(A^*, \cdot_{A^*})$ such that

$$\Delta(a \cdot_A b) = (L(a) \otimes \mathrm{id})\Delta(b) + (\mathrm{id} \otimes R(b))\Delta(a), \quad \forall a, b \in A,$$

where $\Delta : A \to A \otimes A$ is defined by $\langle \Delta(a), x^* \otimes y^* \rangle = \langle a, x^* \cdot_{A^*} y^* \rangle$. An infinitesimal bialgebra is denoted by $\big((A, \cdot_A), (A^*, \cdot_{A^*})\big)$ or $(A, \cdot_A, \Delta)$.

Definition 2.4 [21]. A *Poisson bialgebra* is a pair of Poisson algebras $(A, [\ ,\ ]_A, \cdot_A)$ and $(A^*, [\ ,\ ]_{A^*}, \cdot_{A^*})$ such that

(1) $\big((A, [\ ,\ ]_A), (A^*, [\ ,\ ]_{A^*})\big)$ is a Lie bialgebra,

(2) $\big((A, \cdot_A), (A^*, \cdot_{A^*})\big)$ is an infinitesimal bialgebra, and

(3) δ and Δ are compatible in the sense that, for all $a, b \in A$,

$$\delta(a \cdot_A b) = (L(a) \otimes \mathrm{id})\delta(b) + (L(b) \otimes \mathrm{id})\delta(a) + (\mathrm{id} \otimes \mathrm{ad}(a))\Delta(b) + (\mathrm{id} \otimes \mathrm{ad}(b))\Delta(a),$$

and

$$\Delta([a, b]_A) = (\mathrm{ad}(a) \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{ad}(a))\Delta(b) + (L(b) \otimes \mathrm{id} - \mathrm{id} \otimes L(b))\delta(a),$$

where δ, Δ are the linear duals of $[\ ,\ ]_{A^*}$ and $\cdot_{A^*}$ respectively.

A Poisson bialgebra is denoted by $\big((A, [\ ,\ ]_A, \cdot_A),\ (A^*, [\ ,\ ]_{A^*}, \cdot_{A^*})\big)$ or, in full, $(A, [\ ,\ ]_A, \cdot_A, \delta, \Delta)$.

Poisson bialgebras can be equivalently characterized by Manin triples of Poisson algebras [21]. Notably, for a Poisson bialgebra $\big((A, [\ ,\ ]_A, \cdot_A), (A^*, [\ ,\ ]_{A^*}, \cdot_{A^*})\big)$, the pair $\big((A^*, [\ ,\ ]_{A^*}, \cdot_{A^*}), (A, [\ ,\ ]_A, \cdot_A)\big)$ forms a Poisson bialgebra as well.

Definition 2.5. Let $(A, [\ ,\ ]_A, \cdot_A, \delta_A, \Delta_A)$ and $(B, [\ ,\ ]_B, \cdot_B, \delta_B, \Delta_B)$ be Poisson bialgebras. A *homomorphism* of Poisson bialgebras is a homomorphism of Poisson algebras $\varphi : A \to B$ such that

$$(\varphi \otimes \varphi)\delta_A = \delta_B \circ \varphi, \quad (\varphi \otimes \varphi)\Delta_A = \Delta_B \circ \varphi.$$

If $\varphi : A \to B$ is a linear isomorphism of vector spaces, then $\varphi : A \to B$ is called an *isomorphism* of Poisson bialgebras.

Proposition 2.6. *Let* $((A, [\ ,\ ]_A, \cdot_A), (A^*, [\ ,\ ]_{A^*}, \cdot_{A^*}))$ *be a Poisson bialgebra and* B *be a vector space. Suppose that* $\varphi : A \to B$ *is a linear isomorphism of vector spaces. Define brackets* $[\ ,\ ]_B : B \otimes B \to B$ *and* $[\ ,\ ]_{B^*} : B^* \otimes B^* \to B^*$ *by*

$$[a, b]_B = \varphi([\varphi^{-1}(a), \varphi^{-1}(b)]_A), \quad [x^*, y^*]_{B^*} = (\varphi^*)^{-1}([\varphi^*(x^*), \varphi^*(y^*)]_{A^*}),$$

and multiplications $\cdot_B : B \otimes B \to B$ *and* $\cdot_{B^*} : B^* \otimes B^* \to B^*$ *by*

$$a \cdot_B b = \varphi(\varphi^{-1}(a) \cdot_A \varphi^{-1}(b)), \quad x^* \cdot_{B^*} y^* = (\varphi^*)^{-1}(\varphi^*(x^*) \cdot_{A^*} \varphi^*(y^*)),$$

the equalities holding, as the case may be, for all $a, b \in B$ *and all* $x^*, y^* \in B^*$. *Then* $\big((B, [\ ,\ ]_B, \cdot_B), (B^*, [\ ,\ ]_{B^*}, \cdot_{B^*})\big)$ *is a Poisson bialgebra, and* φ *is an isomorphism of Poisson bialgebras.*

Proof. This follows from a straightforward verification. □

Definition 2.7 [21]. A Poisson bialgebra $(A, [\ ,\ ]_A, \cdot_A, \delta, \Delta)$ is called *coboundary* if there exists $r \in A \otimes A$ such that, for all $a \in A$,

$$\delta(a) = (\mathrm{id} \otimes \mathrm{ad}(a) + \mathrm{ad}(a) \otimes \mathrm{id})(r), \tag{2-1}$$

$$\Delta(a) = (\mathrm{id} \otimes L(a) - L(a) \otimes \mathrm{id})(r). \tag{2-2}$$

A coboundary Poisson bialgebra is denoted by $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$.

Theorem 2.8 [21, Theorem 2]. *Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be a Poisson algebra and* $r \in A \otimes A$. *Write* r *as* $r = S + \Lambda$ *with* $S \in \mathrm{Sym}^2(A)$ *and* $\Lambda \in \mathrm{Alt}^2(A)$. *Let* $\delta : A \to A \otimes A$ *and* $\Delta : A \to A \otimes A$ *be the linear maps defined by* (2-1) *and* (2-2)*.Then* $(A^*, \delta^*, \Delta^*)$ *is a Poisson algebra such that* $(A, [\ ,\ ]_A, \cdot_A, \delta, \Delta)$ *is a Poisson bialgebra if and only if the following conditions are satisfied, for all* $a \in A$:

(1) $(\mathrm{ad}(a) \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{ad}(a))S = 0$.

(2) $(L(a) \otimes \mathrm{id} - \mathrm{id} \otimes L(a))S = 0$.

(3) $(\mathrm{ad}(a) \otimes \mathrm{id} \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{ad}(a) \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{id} \otimes \mathrm{ad}(a))\boldsymbol{C}(r) = 0$.

(4) $(L(a) \otimes \mathrm{id} \otimes \mathrm{id} - \mathrm{id} \otimes \mathrm{id} \otimes L(a))\boldsymbol{A}(r) = 0$.

(5) $(\mathrm{ad}(a) \otimes \mathrm{id} \otimes \mathrm{id})\boldsymbol{A}(r) - (\mathrm{id} \otimes L(a) \otimes \mathrm{id} - \mathrm{id} \otimes \mathrm{id} \otimes L(a))\boldsymbol{C}(r) = 0$.

Here

$$\boldsymbol{C}(r) := [r_{12}, r_{13}]_A + [r_{13}, r_{23}]_A + [r_{12}, r_{23}]_A,$$
$$\boldsymbol{A}(r) := r_{12} \cdot_A r_{13} + r_{13} \cdot_A r_{23} - r_{23} \cdot_A r_{12}.$$

Definition 2.9 [21]. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra and $r \in A \otimes A$. Then r is called a solution of the *Poisson Yang–Baxter equation* in $(A, [\ ,\ ]_A, \cdot_A)$ if $\boldsymbol{C}(r) = \boldsymbol{A}(r) = 0$.

More precisely, $\boldsymbol{C}(r) = 0$ is called the *classical Yang–Baxter equation* in the Lie algebra $(A, [\ ,\ ]_A)$ and $\boldsymbol{A}(r) = 0$ is called the *associative Yang–Baxter equation* in the associative algebra $(A, \cdot_A)$.

Let A be a vector space. Any $r \in A \otimes A$ can be identified with the pair of maps $r_+, r_- : A^* \to A$ defined by

$$\langle r_+(x^*), y^*\rangle = -\langle x^*, r_-(y^*)\rangle = \langle r, x^* \otimes y^*\rangle, \quad \forall x^*, y^* \in A^*.$$

Note that $(\tau(r))_+ = -r_-$ and $(\tau(r))_- = -r_+$. The bracket and multiplication on A^*

defined by (2-1)–(2-2) (as the duals) are given by, for all $x^*, y^* \in A^*$,

$$[x^*, y^*]_r = -\mathrm{ad}^*(r_+(x^*))y^* + \mathrm{ad}^*(r_-(y^*))x^*, \tag{2-3}$$

$$x^* \cdot_r y^* = L^*(r_+(x^*))y^* + L^*(r_-(y^*))x^*. \tag{2-4}$$

Lemma 2.10. *Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra and $r \in A \otimes A$. Then the following statements are equivalent*:

(1) *r is a solution of the Poisson Yang–Baxter equation in $(A, [\ ,\ ]_A, \cdot_A)$.*

(2) *$\tau(r)$ is a solution of the Poisson Yang–Baxter equation in $(A, [\ ,\ ]_A, \cdot_A)$.*

(3) *The following equations hold*, *for all $x^*, y^* \in A^*$*:

$$[r_+(x^*), r_+(y^*)]_A = r_+\left(-\mathrm{ad}^*(r_+(x^*))y^* + \mathrm{ad}^*(r_-(y^*))x^*\right), \tag{2-5}$$

$$r_+(x^*) \cdot_A r_+(y^*) = r_+\left(L^*(r_+(x^*))y^* + L^*(r_-(y^*))x^*\right). \tag{2-6}$$

(4) *The following equations hold*, *for all $x^*, y^* \in A^*$*:

$$[r_-(x^*), r_-(y^*)]_A = r_-\left(-\mathrm{ad}^*(r_-(x^*))y^* + \mathrm{ad}^*(r_+(y^*))x^*\right), \tag{2-7}$$

$$r_-(x^*) \cdot_A r_-(y^*) = r_-\left(L^*(r_-(x^*))y^* + L^*(r_+(y^*))x^*\right). \tag{2-8}$$

Proof. (1) $\Longleftrightarrow$ (2): Let $r = \sum_i a_i \otimes b_i \in A \otimes A$. Then we have

$$\begin{aligned}
\boldsymbol{C}(\tau(r)) &= \sum_{i,j} [b_i, b_j]_A \otimes a_i \otimes a_j + b_i \otimes [a_i, b_j]_A \otimes a_j + b_i \otimes b_j \otimes [a_i, a_j]_A \\
&= \sigma_{13}\big(\sum_{i,j} a_j \otimes a_i \otimes [b_i, b_j]_A + a_j \otimes [a_i, b_j]_A \otimes b_i + [a_i, a_j]_A \otimes b_j \otimes b_i\big) \\
&= -\sigma_{13}(\boldsymbol{C}(r)),
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{A}(\tau(r)) &= \sum_{i,j} b_i \cdot_A b_j \otimes a_i \otimes a_j + b_i \otimes b_j \otimes a_i \cdot_A a_j - b_i \otimes b_j \cdot_A a_i \otimes a_j \\
&= \sigma_{13}\big(\sum_{i,j} a_j \otimes a_i \otimes b_i \cdot_A b_j + a_i \cdot_A a_j \otimes b_j \otimes b_i - a_j \otimes b_j \cdot_A a_i \otimes b_i\big) \\
&= \sigma_{13}\big(\sum_{i,j} a_j \otimes a_i \otimes b_j \cdot_A b_i + a_j \cdot_A a_i \otimes b_j \otimes b_i - a_j \otimes a_i \cdot_A b_j \otimes b_i\big) \\
&= \sigma_{13}(\boldsymbol{A}(r)),
\end{aligned}$$

where $\sigma_{13} \in \mathrm{End}(A \otimes A \otimes A)$ is defined by $\sigma_{13}(a \otimes b \otimes c) = c \otimes b \otimes a$ for all $a, b, c \in A$. Thus, r is a solution of the Poisson Yang–Baxter equation if and only if $\tau(r)$ is a solution of the Poisson Yang–Baxter equation.

(1) $\Longleftrightarrow$ (3): This follows from [5, Proposition 3.8] and [6, Theorem 3.5].

(2) $\Longleftrightarrow$ (4): This is similar to the proof of (1) $\Longleftrightarrow$ (3). □

We now turn to the definition of quasitriangular Poisson bialgebras as a special case of the coboundary Poisson bialgebras. The notion of (ad, L)-invariance of a 2-tensor in $A\otimes A$ is the main ingredient employed, and is motivated by Theorem 2.8.

Definition 2.11. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra. An element $r \in A\otimes A$ is called (ad, L)-*invariant* if, for all $a \in A$, we have

$$(\mathrm{ad}(a)\otimes \mathrm{id} + \mathrm{id}\otimes \mathrm{ad}(a))(r) = 0, \tag{2-9}$$

$$(L(a)\otimes \mathrm{id} - \mathrm{id}\otimes L(a))(r) = 0. \tag{2-10}$$

Obviously, if r is (ad, L)-invariant, then so is $\tau(r)$. Denote by I_r the operator

$$I_r = r_+ - r_- : A^* \to A. \tag{2-11}$$

Note that $r_- = -r_+^*$ and hence $I_r^* = I_r$.

Letting S be the symmetric part of r, we have $S_+ = \frac{1}{2}I_r = \frac{1}{2}I_{\tau(r)}$. In particular, if r is antisymmetric, then $I_r = 0$.

Definition 2.12. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra. If r is a solution of the Poisson Yang–Baxter equation in $(A, [\ ,\ ]_A, \cdot_A)$ and the symmetric part of $r \in A\otimes A$ is (ad, L)-invariant, then the coboundary Poisson bialgebra $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ induced by r is called a *quasitriangular Poisson bialgebra*. If r is also antisymmetric, then $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ is called a *triangular Poisson bialgebra*.

Proposition 2.13. *Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra and $r \in A\otimes A$. Then $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ is a quasitriangular Poisson bialgebra if and only if*

$$(A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)})$$

is a quasitriangular Poisson bialgebra.

Proof. This follows from Lemma 2.10. □

Lemma 2.14. *Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra and $r \in A\otimes A$. Let S be the symmetric part of r. Then the following conditions are equivalent*:

(1) *S is* (ad, L)*-invariant.*

(2) *The following equations hold*, *for all $a \in A$ and $x^* \in A^*$*:

$$S_+(\mathrm{ad}^*(a)x^*) + [a, S_+(x^*)]_A = 0, \tag{2-12}$$

$$S_+(L^*(a)x^*) - a\cdot_A S_+(x^*) = 0. \tag{2-13}$$

(3) *The following equations hold*, *for all $x^*, y^* \in A^*$*:

$$\mathrm{ad}^*(S_+(x^*))y^* + \mathrm{ad}^*(S_+(y^*))x^* = 0, \tag{2-14}$$

$$L^*(S_+(x^*))y^* - L^*(S_+(y^*))x^* = 0. \tag{2-15}$$

Proof. (1) $\Longleftrightarrow$ (2): For all $a \in A$ and $x^*, y^* \in A^*$, we have

$$\begin{aligned}\langle S_+(\mathrm{ad}^*(a)x^*) + [a, S_+(x^*)], y^*\rangle &= \langle S, \mathrm{ad}^*(a)x^* \otimes y^* + x^* \otimes \mathrm{ad}^*(a)y^*\rangle \\ &= \langle (\mathrm{ad}(a) \otimes \mathrm{id} + \mathrm{id} \otimes \mathrm{ad}(a))S, x^* \otimes y^*\rangle\,, \\ \langle S_+(L^*(a)x^*) - a \cdot S_+(x^*), y^*\rangle &= \langle S, L^*(a)x^* \otimes y^* - x^* \otimes L^*(a)y^*\rangle \\ &= \langle (L(a) \otimes \mathrm{id} - \mathrm{id} \otimes L(a))S, x^* \otimes y^*\rangle\,,\end{aligned}$$

which shows that S is (ad, L)-invariant if and only if equations (2-12)–(2-13) hold.

(2) $\Longleftrightarrow$ (3): For all $a \in A$ and $x^*, y^* \in A^*$, we have

$$\begin{aligned}\langle \mathrm{ad}^*(S_+(x^*))y^* + \mathrm{ad}^*(S_+(y^*))x^*, a\rangle &= \langle y^*, [S_+(x^*), a]_A\rangle + \langle x^*, [S_+(y^*), a]_A\rangle \\ &= -\langle y^*, [a, S_+(x^*)]_A + S_+(\mathrm{ad}^*(a)(x^*))\rangle\,, \\ \langle L^*(S_+(x^*))y^* - L^*(S_+(y^*))x^*, a\rangle &= \langle y^*, S_+(x^*) \cdot_A a\rangle - \langle x^*, S_+(y^*) \cdot_A a\rangle \\ &= -\langle y^*, -a \cdot_A S_+(x^*) + S_+(L^*(a)x^*)\rangle\,,\end{aligned}$$

which shows that equations (2-12)–(2-13) hold if and only if (2-14)–(2-15) do. □

Theorem 2.15. *Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be a Poisson algebra and* $r = S + \Lambda \in A \otimes A$, *with* $S \in \mathrm{Sym}^2(A)$ *and* $\Lambda \in \mathrm{Alt}^2(A)$. *Suppose that* S *is* (ad, L)*-invariant. Then* r *is a solution of the Poisson Yang–Baxter equation in* $(A, [\ ,\ ]_A, \cdot_A)$ *if and only if* $(A^*, [\ ,\ ]_r, \cdot_r)$ *is a Poisson algebra and the linear maps* $r_+, r_- : (A^*, [\ ,\ ]_r, \cdot_r) \to (A, [\ ,\ ]_A, \cdot_A)$ *are both Poisson algebra homomorphisms, where* $[\ ,\ ]_r : A^* \otimes A^* \to A^*$ *and* $\cdot_r : A^* \otimes A^* \to A^*$ *are defined by equations* (2-3) *and* (2-4).

Proof. ($\Longrightarrow$) By Definition 2.12, $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ is a quasitriangular Poisson bialgebra where δ_r, Δ_r are defined by (2-1) and (2-2). Thus, $(A^*, [\ ,\ ]_r, \cdot_r)$ is a Poisson algebra where $[\ ,\ ]_r$ and $\cdot_r$ are given by (2-3) and (2-4). By Lemmas 2.10 and 2.14, for all $x^*, y^* \in A^*$, we have

$$\begin{aligned}r_+([x^*, y^*]_r) &= r_+(-\mathrm{ad}^*(r_+(x^*))y^* + \mathrm{ad}^*(r_-(y^*))x^*) = [r_+(x^*), r_+(y^*)]_A, \\ r_+(x^* \cdot_r y^*) &= r_+(L^*(r_+(x^*))y^* + L^*(r_-(y^*))x^*) = r_+(x^*) \cdot_A r_+(y^*), \\ r_-([x^*, y^*]_r) &= r_-(-\mathrm{ad}^*(r_+(x^*))y^* + \mathrm{ad}^*(r_-(y^*))x^*) \\ &= r_-\big(-2\mathrm{ad}^*(S_+(x^*))y^* - \mathrm{ad}^*(r_-(x^*))y^* + \mathrm{ad}^*(r_+(y^*))x^* \\ &\qquad - 2\mathrm{ad}^*(S_+(y^*))x^*\big) \\ &= r_-\big(-\mathrm{ad}^*(r_-(x^*))y^* + \mathrm{ad}^*(r_+(y^*))x^*\big) \\ &= [r_-(x^*), r_-(y^*)]_A, \\ r_-(x^* \cdot_r y^*) &= r_-\big(L^*(r_+(x^*))y^* + L^*(r_-(y^*))x^*\big) \\ &= r_-\big(2L^*(S_+(x^*))y^* + L^*(r_-(x^*))y^* + L^*(r_+(y^*))x^* \\ &\qquad - 2L^*(S_+(y^*))x^*\big) \\ &= r_-\big(L^*(r_-(x^*))y^* + L^*(r_+(y^*))x^*\big) \\ &= r_-(x^*) \cdot_A r_-(y^*),\end{aligned}$$

which shows that r_+, r_- are both Poisson algebra homomorphisms.

($\Longleftarrow$) This follows from Lemma 2.10. □

Recall that a bilinear form $\mathfrak{B} \in \otimes^2 A^*$ on a Poisson algebra $(A, [\ ,\ ]_A, \cdot_A)$ is called *invariant* if

$$\mathfrak{B}([a,b]_A, c) = \mathfrak{B}(a, [b,c]_A), \quad \mathfrak{B}(a \cdot_A b, c) = \mathfrak{B}(a, b \cdot_A c), \quad \forall a,b,c \in A.$$

A *quadratic Poisson algebra* is a quadruple

$$(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B}),$$

where $(A, [\ ,\ ]_A, \cdot_A)$ is a Poisson algebra and $\mathfrak{B} \in \otimes^2 A^*$ is a nondegenerate symmetric invariant bilinear form on $(A, [\ ,\ ]_A, \cdot_A)$.

Proposition 2.16. *Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be a Poisson algebra. Then*

$$\big(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot, \mathfrak{B}_d\big)$$

is a quadratic Poisson algebra, where $\mathfrak{B}_d$ *is defined by*

$$\mathfrak{B}_d\left(a + x^*, b + y^*\right) = \langle a, y^* \rangle + \langle b, x^* \rangle, \quad \forall a, b \in A,\ x^*, y^* \in A^*. \tag{2-16}$$

Proof. This follows from a straightforward computation. □

Let A be a vector space and $\mathfrak{B}$ be a nondegenerate bilinear form. Denote by $I_{\mathfrak{B}} : A^* \to A$ the induced linear isomorphism defined by

$$\langle I_{\mathfrak{B}}^{-1}(a), b \rangle := \mathfrak{B}(a,b), \quad \forall a, b \in A. \tag{2-17}$$

Denote by $r_{\mathfrak{B}} \in A \otimes A$ the 2-tensor form of $I_{\mathfrak{B}}$, that is,

$$\langle r_{\mathfrak{B}}, x^* \otimes y^* \rangle := \langle I_{\mathfrak{B}}(x^*), y^* \rangle, \quad \forall x^*, y^* \in A^*. \tag{2-18}$$

Proposition 2.17. *Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be a Poisson algebra and* $\mathfrak{B}$ *a nondegenerate bilinear form. Let* $I_{\mathfrak{B}} : A^* \to A$ *be the linear isomorphism induced by* $\mathfrak{B}$ *and* $r_{\mathfrak{B}} \in A \otimes A$ *be the 2-tensor form of* $I_{\mathfrak{B}}$ *given by* (2-18). *Then* $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ *is a quadratic Poisson algebra if and only if* $r_{\mathfrak{B}} \in A \otimes A$ *is symmetric and* (ad, L)*-invariant.*

Proof. Suppose that $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ is a quadratic Poisson algebra. For all $a, b, c \in A$, there exist $x^*, y^*, z^* \in A^*$ such that $I_{\mathfrak{B}}(x^*) = a$, $I_{\mathfrak{B}}(y^*) = b$ and $I_{\mathfrak{B}}(z^*) = c$. Then we have

$$\begin{aligned}\langle r_{\mathfrak{B}} - \tau(r_{\mathfrak{B}}), x^* \otimes y^* \rangle &= \langle I_{\mathfrak{B}}(x^*), I_{\mathfrak{B}}^{-1}(b) \rangle - \langle I_{\mathfrak{B}}(y^*), I_{\mathfrak{B}}^{-1}(a) \rangle \\ &= \mathfrak{B}(b,a) - \mathfrak{B}(a,b) = 0.\end{aligned}$$

Thus, $r_{\mathfrak{B}}$ is symmetric. Moreover, we have

$$\begin{aligned}
&\langle (\mathrm{ad}(c)\otimes \mathrm{id}+\mathrm{id}\otimes \mathrm{ad}(c))(r_{\mathfrak{B}}), x^*\otimes y^*\rangle \\
&\qquad = \langle r_{\mathfrak{B}}, \mathrm{ad}^*(c)(x^*)\otimes y^*\rangle + \langle r_{\mathfrak{B}}, x^*\otimes \mathrm{ad}^*(c)(y^*)\rangle \\
&\qquad = \langle I_{\mathfrak{B}}(y^*), \mathrm{ad}^*(c)(x^*)\rangle + \langle I_{\mathfrak{B}}(x^*), \mathrm{ad}^*(c)(y^*)\rangle \\
&\qquad = \mathfrak{B}([c,b]_A, a)+\mathfrak{B}([c,a]_A, b)=0,\\
&\langle (L(c)\otimes \mathrm{id}-\mathrm{id}\otimes L(c))(r_{\mathfrak{B}}), x^*\otimes y^*\rangle \\
&\qquad = \langle r_{\mathfrak{B}}, L^*(c)(x^*)\otimes y^*\rangle - \langle r_{\mathfrak{B}}, x^*\otimes L^*(c)(y^*)\rangle \\
&\qquad = \langle I_{\mathfrak{B}}(y^*), L^*(c)(x^*)\rangle - \langle I_{\mathfrak{B}}(x^*), L^*(c)(y^*)\rangle \\
&\qquad = \mathfrak{B}(c\cdot_A b, a)-\mathfrak{B}(c\cdot_A a, b)=0,
\end{aligned}$$

which shows that $r_{\mathfrak{B}}\in A\otimes A$ is (ad, L)-invariant. The converse statement can be proved by reversing the argument. □

The next theorem demonstrates that every quadratic Poisson algebra naturally induces an isomorphism between the adjoint and coadjoint representations of the corresponding Poisson algebra.

Definition 2.18. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra, and (V_1, ρ_1, μ_1) and (V_2, ρ_2, μ_2) be two representations of $(A, [\ ,\ ]_A, \cdot_A)$. A *homomorphism of representations* from (V_1, ρ_1, μ_1) to (V_2, ρ_2, μ_2) is a linear map $\varphi : V_1 \to V_2$ such that

$$\varphi\circ\rho_1(a)=\rho_2(a)\circ\varphi,\quad \varphi\circ\mu_1(a)=\mu_2(a)\circ\varphi,\quad \forall a\in A.$$

When $\varphi : V_1 \to V_2$ is a vector space isomorphism satisfying this same condition, φ is called an *isomorphism of representations* from (V_1, ρ_1, μ_1) to (V_2, ρ_2, μ_2).

Theorem 2.19. *Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra. If there is a bilinear form $\mathfrak{B}$ such that $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ is a quadratic Poisson algebra, then the linear map $I_{\mathfrak{B}}^{-1} : A \to A^*$ defined by $\langle I_{\mathfrak{B}}^{-1}(a), b\rangle = \mathfrak{B}(a, b)$ is an isomorphism from the adjoint representation (A, ad, L) to the coadjoint representation $(A^*, -\mathrm{ad}^*, L^*)$.*

Conversely, if $I^{-1} : A \to A^*$ is an isomorphism from the adjoint representation (A, ad, L) to the coadjoint representation $(A^*, -\mathrm{ad}^*, L^*)$, then the bilinear form $\mathfrak{B}$ defined by $\mathfrak{B}(a, b) = \langle I^{-1}(a), b\rangle$ is nondegenerate invariant on $(A, [\ ,\ ]_A, \cdot_A)$.

Proof. Suppose that $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ is a quadratic Poisson algebra. For all $a, b, c \in A$, we have

$$\langle I_{\mathfrak{B}}^{-1}\mathrm{ad}(a)b, c\rangle = -\mathfrak{B}([b,a]_A, c) = -\mathfrak{B}(b, [a,c]_A) = \langle -\mathrm{ad}^*(a)I_{\mathfrak{B}}^{-1}(b), c\rangle,$$

$$\langle I_{\mathfrak{B}}^{-1}L(a)b, c\rangle = \mathfrak{B}(a\cdot_A b, c) = \mathfrak{B}(b\cdot_A a, c) = \mathfrak{B}(b, a\cdot_A c) = \langle L^*(a)I_{\mathfrak{B}}^{-1}(b), c\rangle,$$

that is,

$$I_{\mathfrak{B}}^{-1}\mathrm{ad}(a) = -\mathrm{ad}^*(a)I_{\mathfrak{B}}^{-1},\quad I_{\mathfrak{B}}^{-1}L(a) = L^*(a)I_{\mathfrak{B}}^{-1}.$$

Hence, $I_{\mathfrak{B}}^{-1} : A \to A^*$ is an isomorphism from the adjoint representation (A, ad, L)

to the coadjoint representation $(A^*, -\mathrm{ad}^*, L^*)$. The converse can be proved by a similar argument. □

We conclude this section with a proposition that will be valuable in Section 4, in obtaining Rota–Baxter characterizations of factorizable Poisson bialgebras.

Proposition 2.20. *Let $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ be a quadratic Poisson algebra and $I_{\mathfrak{B}} : A^* \to A$ be the induced linear isomorphism by $\mathfrak{B}$. Suppose that $r \in A \otimes A$. Then r is a solution of the Poisson Yang–Baxter equation in $(A, [\ ,\ ]_A, \cdot_A)$ if and only if the linear map $P := r_+ \circ I_{\mathfrak{B}}^{-1} : A \to A$ satisfies the following equations, for all $a, b \in A$:*

$$[P(a), P(b)]_A = P\big([P(a), b]_A + [a, P(b)]_A - [a, (I_r \circ I_{\mathfrak{B}}^{-1})(b)]_A\big), \tag{2-19}$$

$$P(a) \cdot_A P(b) = P\big(P(a) \cdot_A b + a \cdot_A P(b) - a \cdot_A (I_r \circ I_{\mathfrak{B}}^{-1})(b)\big), \tag{2-20}$$

Proof. By Theorem 2.19, we have

$$I_{\mathfrak{B}}^{-1}\mathrm{ad}(a) = -\mathrm{ad}^*(a) I_{\mathfrak{B}}^{-1}, \quad I_{\mathfrak{B}}^{-1} L(a) = L^*(a) I_{\mathfrak{B}}^{-1}, \quad \forall a \in A.$$

For all $a, b \in A$, there exist $x^*, y^* \in A^*$ such that $I_{\mathfrak{B}}(x^*) = a$ and $I_{\mathfrak{B}}(y^*) = b$. Then we have

$$\begin{aligned}
[P(a), P(b)]_A &= [(r_+ \circ I_{\mathfrak{B}}^{-1})(a), (r_+ \circ I_{\mathfrak{B}}^{-1})(b)]_A = [r_+(x^*), r_+(y^*)]_A,\\
P([P(a), b]_A) &= (r_+ \circ I_{\mathfrak{B}}^{-1})([(r_+ \circ I_{\mathfrak{B}}^{-1})(a), b]_A)\\
&= (r_+ \circ I_{\mathfrak{B}}^{-1})(\mathrm{ad}(r_+(x^*)) I_{\mathfrak{B}}(y^*)) = -r_+(\mathrm{ad}^*(r_+(x^*)) y^*),\\
P([a, P(b)]_A) &= (r_+ \circ I_{\mathfrak{B}}^{-1})([a, (r_+ \circ I_{\mathfrak{B}}^{-1})(b)]_A)\\
&= -(r_+ \circ I_{\mathfrak{B}}^{-1})(\mathrm{ad}(r_+(y^*)) I_{\mathfrak{B}}(x^*)) = r_+(\mathrm{ad}^*(r_+(y^*)) x^*),\\
-P([a, (I_r \circ I_{\mathfrak{B}}^{-1})(b)]_A) &= -(r_+ \circ I_{\mathfrak{B}}^{-1})([a, (I_r \circ I_{\mathfrak{B}}^{-1})(b)]_A)\\
&= (r_+ \circ I_{\mathfrak{B}}^{-1})(\mathrm{ad}(I_r(y^*)) I_{\mathfrak{B}}(x^*)) = -r_+(\mathrm{ad}^*(I_r(y^*)) x^*),\\
P(a) \cdot_A P(b) &= (r_+ \circ I_{\mathfrak{B}}^{-1})(a) \cdot_A (r_+ \circ I_{\mathfrak{B}}^{-1})(b) = r_+(x^*) \cdot_A r_+(y^*),\\
P(P(a) \cdot_A b) &= (r_+ \circ I_{\mathfrak{B}}^{-1})((r_+ \circ I_{\mathfrak{B}}^{-1})(a) \cdot_A b)\\
&= (r_+ \circ I_{\mathfrak{B}}^{-1})(L(r_+(x^*)) I_{\mathfrak{B}}(y^*)) = r_+(L^*(r_+(x^*)) y^*),\\
P(a \cdot_A P(b)) &= (r_+ \circ I_{\mathfrak{B}}^{-1})(a \cdot_A (r_+ \circ I_{\mathfrak{B}}^{-1})(b))\\
&= (r_+ \circ I_{\mathfrak{B}}^{-1})(L(r_+(y^*)) I_{\mathfrak{B}}(x^*)) = r_+(L^*(r_+(y^*)) x^*),\\
-P(a \cdot_A (I_r \circ I_{\mathfrak{B}}^{-1})(b)) &= -(r_+ \circ I_{\mathfrak{B}}^{-1})(a \cdot_A (I_r \circ I_{\mathfrak{B}}^{-1})(b))\\
&= -(r_+ \circ I_{\mathfrak{B}}^{-1})(L(I_r(y^*)) I_{\mathfrak{B}}(x^*)) = -r_+(L^*(I_r(y^*)) x^*).
\end{aligned}$$

By (2-11) and Lemma 2.10, the conclusion follows. □

3. Factorizable Poisson bialgebras

A factorizable Poisson bialgebra is a special quasitriangular Poisson bialgebra such that the map $I_r : A^* \to A$ is a linear isomorphism of vector spaces. We will show that the Drinfeld classical double of a Poisson bialgebra is naturally a factorizable Poisson bialgebra.

Definition 3.1. A quasitriangular Poisson bialgebra $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ is called *factorizable* if the symmetric part S of r is nondegenerate, which means that the linear map $I_r : A^* \to A$ defined by (2-11) is a linear isomorphism of vector spaces.

Proposition 3.2. *Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be a Poisson algebra and* $r \in A \otimes A$. *Then* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ *is a factorizable Poisson bialgebra if and only if the same is true of* $(A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)})$.

Proof. This follows from Proposition 2.13. □

Consider the map

$$A^* \xrightarrow{r_+ \oplus r_-} A \oplus A \xrightarrow{(a,b)\mapsto a-b} A.$$

The next result justifies the term “factorizable Poisson bialgebra”.

Proposition 3.3. *Let* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ *be a factorizable Poisson bialgebra. Then* $\mathrm{Im}(r_+ \oplus r_-)$ *is a Poisson subalgebra of the direct sum Poisson algebra* $A \oplus A$, *which is isomorphic to the Poisson algebra* $(A^*, [\ ,\ ]_r, \cdot_r)$, *where* $[\ ,\ ]_r$ *and* $\cdot_r$ *are defined by* (2-3) *and* (2-4). *Any* $a \in A$ *has a unique decomposition* $a = a_+ - a_-$ *with* $(a_+, a_-) \in \mathrm{Im}(r_+ \oplus r_-)$.

Proof. By Theorem 2.15, both r_+ and r_- are Poisson algebra homomorphisms. Therefore, $\mathrm{Im}(r_+ \oplus r_-)$ is a Poisson subalgebra of the Poisson algebra $A \oplus A$. Since $I_r = r_+ - r_-$ is a linear isomorphism of vector spaces, it follows that the Poisson algebra $\mathrm{Im}(r_+ \oplus r_-)$ is isomorphic to the Poisson algebra $(A^*, [\ ,\ ]_r, \cdot_r)$. Moreover, we have

$$r_+ I_r^{-1}(a) - r_- I_r^{-1}(a) = (r_+ - r_-) I_r^{-1}(a) = a,$$

which shows that $a = a_+ - a_-$ with $a_+ = r_+ I_r^{-1}(a)$ and $a_- = r_- I_r^{-1}(a)$. Uniqueness also follows from the fact that $I_r : A^* \to A$ is a linear isomorphism of vector spaces. □

Let $\big((A, [\ ,\ ]_A, \cdot_A), (A^*, [\ ,\ ]_{A^*}, \cdot_{A^*})\big)$ be an arbitrary Poisson bialgebra. We endow $\mathfrak{A} = A \oplus A^*$ with a bracket

$$[(a, x^*), (b, y^*)]_{\mathfrak{A}}$$
$$= \big([a, b]_A - \mathrm{ad}^*_{[,]_{A^*}}(x^*)b + \mathrm{ad}^*_{[,]_{A^*}}(y^*)a, [x^*, y^*]_{A^*} - \mathrm{ad}^*_{[,]_A}(a)y^* + \mathrm{ad}^*_{[,]_A}(b)x^*\big)$$

and a multiplication

$$(a, x^*) \cdot_{\mathfrak{A}} (b, y^*) \\ = \big(a \cdot_A b + L^*_{\cdot_{A^*}}(x^*)b + L^*_{\cdot_{A^*}}(y^*)a,\ x^* \cdot_{A^*} y^* + L^*_{\cdot_A}(a)y^* + L^*_{\cdot_A}(b)x^*\big),$$

where $a, b \in A$, $x^*, y^* \in A^*$. Then $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}})$ is a Poisson algebra, called the *Drinfeld classical double* of the Poisson bialgebra.

Let $\{e_1, e_2, \dots, e_n\}$ be a basis of A and $\{e_1^*, e_2^*, \dots, e_n^*\}$ be the dual basis of A^*. Then

$$r = \sum_i e_i \otimes e_i^* \in A \otimes A^* \subset \mathfrak{A} \otimes \mathfrak{A}$$

induces a coboundary Poisson bialgebra structure $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)$ (see [21, Theorem 3] for details).

Theorem 3.4. *The Poisson bialgebra* $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)$ *is a factorizable quasi-triangular Poisson bialgebra.*

Proof. First, we need to verify that the symmetric part $S = \frac{1}{2}(e_i \otimes e_i^* + e_i^* \otimes e_i)$ of r is $(\mathrm{ad}_{[,]_{\mathfrak{A}}}, L_{\cdot_{\mathfrak{A}}})$-invariant. For all $(x^*, a) \in \mathfrak{A}^*$, we have $S_+(x^*, a) = \frac{1}{2}(a, x^*) \in \mathfrak{A}$. A straightforward computation gives

$$[(a, x^*), S_+(y^*, b)]_{\mathfrak{A}} \\ = \tfrac{1}{2}\big([a, b]_A - \mathrm{ad}^*_{[,]_{A^*}}(x^*)b + \mathrm{ad}^*_{[,]_{A^*}}(y^*)a,\ [x^*, y^*]_{A^*} - \mathrm{ad}^*_{[,]_A}(a)y^* + \mathrm{ad}^*_{[,]_A}(b)x^*\big),$$

$$\mathrm{ad}^*_{[,]_{\mathfrak{A}}}(a, x^*)(y^*, b) \\ = -\big([x^*, y^*]_{A^*} - \mathrm{ad}^*_{[,]_A}(a)y^* + \mathrm{ad}^*_{[,]_A}(b)x^*,\ [a, b]_A - \mathrm{ad}^*_{[,]_{A^*}}(x^*)b + \mathrm{ad}^*_{[,]_{A^*}}(y^*)a\big),$$

$$(a, x^*) \cdot_{\mathfrak{A}} S_+(y^*, b) \\ = \tfrac{1}{2}\big(a \cdot_A b + L^*_{\cdot_{A^*}}(x^*)b + L^*_{\cdot_{A^*}}(y^*)a,\ x^* \cdot_{A^*} y^* + L^*_{\cdot_A}(a)y^* + L^*_{\cdot_A}(b)x^*\big),$$

$$L^*_{\cdot_{\mathfrak{A}}}(a, x^*)(y^*, b) \\ = \big(x^* \cdot_{A^*} y^* + L^*_{\cdot_A}(a)y^* + L^*_{\cdot_A}(b)x^*,\ a \cdot_A b + L^*_{\cdot_{A^*}}(x^*)b + L^*_{\cdot_{A^*}}(y^*)a\big).$$

Thus, we have

$$S_+(\mathrm{ad}^*_{[,]_{\mathfrak{A}}}(a, x^*)(y^*, b)) + [(a, x^*), S_+(y^*, b)]_{\mathfrak{A}} = 0,$$
$$S_+(L^*_{\cdot_{\mathfrak{A}}}(a, x^*)(y^*, b)) - (a, x^*) \cdot_{\mathfrak{A}} S_+(y^*, b) = 0.$$

Therefore, by Lemma 2.14, the symmetric part S of r is $(\mathrm{ad}_{[,]_{\mathfrak{A}}}, L_{\cdot_{\mathfrak{A}}})$-invariant. On the other hand, by [21, Theorem 3], r is a solution of the Poisson Yang–Baxter equation in $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}})$. Hence, the Poisson bialgebra $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)$ is a quasitriangular Poisson bialgebra.

That $I_r : \mathfrak{A}^* \to \mathfrak{A}$ is a linear isomorphism of vector spaces follows from the equality $I_r(x^*, a) = 2S_+(x^*, a) = (a, x^*)$. Therefore, the Poisson bialgebra $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)$ is factorizable. □

4. Quadratic Rota–Baxter Poisson algebras

A linear map $P : A \to A$ is called a *Rota–Baxter operator of weight* λ on a Poisson algebra $(A, [\ ,\ ]_A, \cdot_A)$ if, for all $a, b \in A$, we have

$$[P(a), P(b)]_A = P([P(a), b]_A + [a, P(b)]_A + \lambda[a, b]_A),$$
$$P(a) \cdot_A P(b) = P(P(a) \cdot_A b + a \cdot_A P(b) + \lambda a \cdot_A b).$$

A *Rota–Baxter Poisson algebra* $(A, [\ ,\ ]_A, \cdot_A, P)$ *of weight* λ is a Poisson algebra $(A, [\ ,\ ]_A, \cdot_A)$ equipped with a Rota–Baxter operator P of weight λ.

Example 4.1. Let $(A, [\ ,\ ]_A, \cdot_A)$ be a three-dimensional Poisson algebra with a basis $\{e_1, e_2, e_3\}$ whose nonzero products are given by

$$[e_1, e_2]_A = e_3, \quad e_1 \cdot_A e_2 = e_3. \tag{4-1}$$

Let $P : A \to A$ be the linear map given by

$$P(e_1) = e_1, \quad P(e_2) = 2e_2, \quad P(e_3) = \tfrac{1}{2}e_3.$$

Then $(A, [\ ,\ ]_A, \cdot_A, P)$ is a Rota–Baxter Poisson algebra of weight 1.

Lemma 4.2. *Let* $(A, [\ ,\ ]_A, \cdot_A, P)$ *be a Rota–Baxter Poisson algebra of weight* λ. *Define a bracket* $[\ ,\ ]_P : A \otimes A \to A$ *by*

$$[a, b]_P = [P(a), b]_A + [a, P(b)]_A + \lambda[a, b]_A, \tag{4-2}$$

and a product $\cdot_P : A \otimes A \to A$ *by*

$$a \cdot_P b = P(a) \cdot_A b + a \cdot_A P(b) + \lambda a \cdot_A b. \tag{4-3}$$

Then $(A, [\ ,\ ]_P, \cdot_P)$ *is a Poisson algebra, called the* ***descendent Poisson algebra*** *of* $(A, [\ ,\ ]_A, \cdot_A, P)$, *and* P *is a Poisson algebra homomorphism from* $(A, [\ ,\ ]_P, \cdot_P)$ *to* $(A, [\ ,\ ]_A, \cdot_A)$.

Proof. The proof is straightforward. □

Equipping quadratic Poisson algebras with Rota–Baxter operators satisfying compatibility conditions, we now introduce the notion of quadratic Rota–Baxter Poisson algebras. We then show that a quadratic Rota–Baxter Poisson algebra of zero weight induces a triangular Poisson bialgebra, and that there is a one-to-one correspondence between factorizable Poisson bialgebras and quadratic Rota–Baxter Poisson algebras of nonzero weight.

Definition 4.3. The triple $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ is called a *quadratic Rota–Baxter Poisson algebra of weight* λ if $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ is a quadratic Poisson algebra and $(A, [\ ,\ ]_A, \cdot_A, P)$ is a Rota–Baxter Poisson algebra of weight λ satisfying the compatibility condition

$$\mathfrak{B}(a, P(b)) + \mathfrak{B}(P(a), b) + \lambda\mathfrak{B}(a, b) = 0, \quad \forall a, b \in A. \tag{4-4}$$

Proposition 4.4. *Let* $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ *be a quadratic Poisson algebra and let* $P : A \to A$ *be a linear map. Then* $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ *is a quadratic Rota–Baxter Poisson algebra of weight* λ *if and only if* $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, \tilde{P})$ *is a quadratic Rota–Baxter Poisson algebra of weight* λ, *where* $\tilde{P} := -\lambda\,\mathrm{id} - P$.

Proof. It is easy to show that P is a Rota–Baxter operator of weight λ if and only if $\tilde{P}$ is a Rota–Baxter operator of weight λ. Moreover, we have, for all $a, b \in A$,

$$\mathfrak{B}(\tilde{P}(a), b) + \mathfrak{B}(a, \tilde{P}(b)) + \lambda\mathfrak{B}(a, b) = -\mathfrak{B}(P(a), b) - \mathfrak{B}(a, P(b)) - \lambda\mathfrak{B}(a, b).$$

Thus, P satisfies (4-4) if and only if $\tilde{P}$ satisfies (4-4). □

Proposition 4.5. *Let* $(A, [\ ,\ ]_A, \cdot_A, P)$ *be a Rota–Baxter Poisson algebra of weight* λ. *Then* $\big((A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot), \mathfrak{B}_d, P + \tilde{P}^*\big)$ *is a quadratic Rota–Baxter Poisson algebra of weight* λ, *where* $\mathfrak{B}_d$ *is defined by* (2-16) *and* $\tilde{P} := -\lambda\,\mathrm{id} - P$.

Proof. The proof is straightforward. □

Example 4.6. Building on Example 4.1, let $(A, [\ ,\ ]_A, \cdot_A, P)$ be a Rota–Baxter Poisson algebra of weight 1. Let $\{e_1^*, e_2^*, e_3^*\}$ be the dual basis of $\{e_1, e_2, e_3\}$. Then $(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot)$ is a Poisson algebra whose nonzero products are given by

$$\begin{aligned} &[e_1, e_2] = e_3, \quad [e_3^*, e_1] = e_2^*, \quad [e_3^*, e_2] = -e_1^*, \\ &e_1 \cdot e_2 = e_3, \quad\ \ e_3^* \cdot e_1 = e_2^*, \quad\ \ e_3^* \cdot e_2 = e_1^*. \end{aligned} \tag{4-5}$$

By Proposition 4.5, we obtain a quadratic Rota–Baxter Poisson algebra

$$\big((A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot), \mathfrak{B}_d, P + \tilde{P}^*\big)$$

of weight 1, where $\mathfrak{B}_d$ is defined by

$$\mathfrak{B}_d(e_i, e_j) = \mathfrak{B}_d(e_i^*, e_j^*) = 0, \quad \mathfrak{B}_d(e_i^*, e_j) = \mathfrak{B}_d(e_i, e_j^*) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

for $i, j = 1, 2, 3$, and $\tilde{P}^*$ is given by

$$\tilde{P}^*(e_1^*) = -2e_1^*, \quad \tilde{P}^*(e_2^*) = -3e_2^*, \quad \tilde{P}^*(e_3^*) = -\tfrac{3}{2}e_3.$$

Lemma 4.7. *Let* A *be a vector space and* $\mathfrak{B}$ *be a nondegenerate symmetric bilinear form. Let* $I_{\mathfrak{B}} : A^* \to A$ *be the induced linear isomorphism by* $\mathfrak{B}$ *and* $r_{\mathfrak{B}} \in A \otimes A$ *be the* 2*-tensor form of* $I_{\mathfrak{B}}$ *given by* (2-18). *Suppose that* $r \in A \otimes A$. *Then* $P_r := r_+ \circ I_{\mathfrak{B}}^{-1}$ *satisfies* (4-4) *if and only if* $r + \tau(r) = -\lambda r_{\mathfrak{B}}$ *where* $\lambda \in \mathbb{K}$.

Proof. For all $a, b \in A$, there exist x^*, $y^* \in A^*$ such that $I_{\mathfrak{B}}(x^*) = a$, $I_{\mathfrak{B}}(y^*) = b$. Then

$$\mathfrak{B}\,(P_r(a), b) = \mathfrak{B}\,(b, P_r(a)) = \langle I_{\mathfrak{B}}^{-1}(b), (r_+ \circ I_{\mathfrak{B}}^{-1})(a)\rangle = \langle r, x^* \otimes y^*\rangle\,,$$
$$\mathfrak{B}(a, P_r(b)) = \langle I_{\mathfrak{B}}^{-1}(a), (r_+ \circ I_{\mathfrak{B}}^{-1})(b)\rangle = \langle r, y^* \otimes x^*\rangle = \langle \tau(r), x^* \otimes y^*\rangle\,,$$
$$\lambda\mathfrak{B}(a, b) = \lambda\mathfrak{B}(b, a) = \lambda\langle I_{\mathfrak{B}}^{-1}(b), (I_{\mathfrak{B}} \circ I_{\mathfrak{B}}^{-1})(a)\rangle = \lambda\langle r_{\mathfrak{B}}, x^* \otimes y^*\rangle\,.$$

Hence, $r + \tau(r) = -\lambda r_{\mathfrak{B}}$ if and only if P_r satisfies (4-4). □

As a direct consequence, a quadratic Rota–Baxter Poisson algebra of zero weight gives rise to a triangular Poisson bialgebra in the following sense.

Proposition 4.8. *Let* $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ *be a quadratic Rota–Baxter Poisson algebra of weight* 0 *and* $I_{\mathfrak{B}} : A^* \to A$ *be the induced linear isomorphism by* $\mathfrak{B}$. *Then* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ *is a triangular Poisson bialgebra where* $r \in A \otimes A$ *is the* 2*-tensor form of* $P \circ I_{\mathfrak{B}}$ *given by*

$$\langle r, x^* \otimes y^*\rangle := \langle (P \circ I_{\mathfrak{B}})(x^*), y^*\rangle\,, \quad \forall x^*, y^* \in A^*. \tag{4-6}$$

Proof. This follows from Proposition 2.20 and Lemma 4.7 with $r + \tau(r) = 0$. □

The following theorem shows that a factorizable Poisson bialgebra naturally gives rise to a quadratic Rota–Baxter Poisson algebra of nonzero weight.

Theorem 4.9. *Let* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ *be a factorizable Poisson bialgebra with* $I_r = r_+ - r_-$. *Then* $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ *is a quadratic Rota–Baxter Poisson algebra of weight* λ, *where* $P : A \to A$ *is given by*

$$P = -\lambda r_+ \circ I_r^{-1}, \quad \lambda \neq 0,$$

and the bilinear form $\mathfrak{B} \in \otimes^2 A^*$ *is defined by*

$$\mathfrak{B}(a, b) = -\lambda\langle I_r^{-1}(a), b\rangle\,, \quad \forall a, b \in A.$$

Proof. Clearly, $\mathfrak{B}$ is a nondegenerate symmetric bilinear form and $I_{\mathfrak{B}}^{-1} = -\lambda I_r^{-1}$. Immediately, we have $I_r = -\lambda I_{\mathfrak{B}}$ and thus $r + \tau(r) = -\lambda r_{\mathfrak{B}}$. Noting that the symmetric part of r is (ad, L)-invariant, we have $r_{\mathfrak{B}}$ is (ad, L)-invariant and thus by Proposition 2.17, $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B})$ is a quadratic Poisson algebra. Moreover, it is clear that P is a Rota–Baxter operator of weight λ on the Poisson algebra $(A, [\ ,\ ]_A, \cdot_A)$ by Proposition 2.20 and P satisfies (4-4) by Lemma 4.7. Thus, $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ is a quadratic Rota–Baxter Poisson algebra of weight λ. □

Corollary 4.10. *Let* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ *be a factorizable Poisson bialgebra with* $I_r = r_+ - r_-$, *and* $P = -\lambda r_+ \circ I_r^{-1}$ *be the induced Rota–Baxter operator of weight* $\lambda \neq 0$. *Let* $(A, [\ ,\ ]_P, \cdot_P)$ *be the descendent Poisson algebra of the Rota–Baxter algebra* $(A, [\ ,\ ]_A, \cdot_A, P)$. *Then* $((A, [\ ,\ ]_P, \cdot_P), (A^*, [\ ,\ ]_{I_r}, \cdot_{I_r}))$ *is a Poisson* *bialgebra, where, for all* $x^*, y^* \in A^*$,

$$[x^*, y^*]_{I_r} := -\lambda I_r^{-1}\big([\lambda^{-1} I_r(x^*), \lambda^{-1} I_r(y^*)]_A\big),$$
$$x^* \cdot_{I_r} y^* := -\lambda I_r^{-1}\big((\lambda^{-1} I_r(x^*)) \cdot_A (\lambda^{-1} I_r(y^*))\big).$$

Moreover, $-\lambda^{-1} I_r : A^* \to A$ *is a Poisson bialgebra isomorphism*

$$\big((A^*, [\ ,\ ]_r, \cdot_r), (A, [\ ,\ ]_A, \cdot_A)\big) \to \big((A, [\ ,\ ]_P, \cdot_P), (A^*, [\ ,\ ]_{I_r}, \cdot_{I_r})\big),$$

where $[\ ,\ ]_r,\ \cdot_r : A^* \otimes A^* \to A^*$ *are defined in* (2-3) *and* (2-4).

Proof. For all $x^*, y^* \in A^*$, setting $a = I_r(x^*)$ and $b = I_r(y^*)$, we have

$$\begin{aligned}
&-\lambda^{-1} I_r([x^*, y^*]_r)\\
&\quad\overset{(2\text{-}3)}{=} -\lambda^{-1} I_r\big(-\mathrm{ad}^*(r_+(x^*))y^* + \mathrm{ad}^*(r_-(y^*))x^*\big)\\
&\quad\overset{(2\text{-}12)}{=} -\lambda^{-1}\big([r_+(x^*), I_r(y^*)]_A - [r_-(y^*), I_r(x^*)]_A\big)\\
&\quad= \lambda^{-2}\big([P I_r(x^*), I_r(y^*)]_A + [I_r(x^*), P I_r(y^*)]_A + \lambda[I_r(x^*), I_r(y^*)]_A\big)\\
&\quad\overset{(4\text{-}2)}{=} [-\lambda^{-1} I_r(x^*), -\lambda^{-1} I_r(y^*)]_P.
\end{aligned}$$

Similarly, we have

$$-\lambda^{-1} I_r(x^* \cdot_r y^*) = (-\lambda^{-1} I_r(x^*)) \cdot_P (-\lambda^{-1} I_r(y^*)).$$

Hence, $-\lambda^{-1} I_r : (A^*, [\ ,\ ]_r, \cdot_r) \to (A, [\ ,\ ]_P, \cdot_P)$ is a Poisson algebra homomorphism.

Noting that $\lambda^{-1} I_r^* = \lambda^{-1} I_r$, we have

$$-\lambda^{-1} I_r^*([x^*, y^*]_{I_r}) = \big[-\lambda^{-1} I_r(x^*), -\lambda^{-1} I_r(y^*)\big]_A = \big[-\lambda^{-1} I_r^*(x^*), -\lambda^{-1} I_r^*(y^*)\big]_A,$$
$$-\lambda^{-1} I_r^*(x^* \cdot_{I_r} y^*) = (-\lambda^{-1} I_r(x^*)) \cdot_A (-\lambda^{-1} I_r(y^*)) = (-\lambda^{-1} I_r^*(x^*)) \cdot_A (-\lambda^{-1} I_r^*(y^*)),$$

which means that $-\lambda^{-1} I_r^* : (A^*, [\ ,\ ]_{I_r}, \cdot_{I_r}) \to (A, [\ ,\ ]_A, \cdot_A)$ is also a Poisson algebra homomorphism. Since $\big((A^*, [\ ,\ ]_r, \cdot_r), (A, [\ ,\ ]_A, \cdot_A)\big)$ is a Poisson bialgebra, so is $((A, [\ ,\ ]_P, \cdot_P), (A^*, [\ ,\ ]_{I_r}, \cdot_{I_r}))$, by Proposition 2.6. Clearly, $-\lambda^{-1} I_r$ is an isomorphism of Poisson bialgebras. □

As a counterpart to Theorem 4.9, the following theorem shows that a quadratic Rota–Baxter Poisson algebra of nonzero weight induces a factorizable Poisson bialgebra, thereby refining the one-to-one correspondence between factorizable Poisson bialgebras and quadratic Rota–Baxter Poisson algebras of nonzero weight.

Theorem 4.11. *Let* $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ *be a quadratic Rota–Baxter Poisson algebra of weight* $\lambda \neq 0$, *and* $I_{\mathfrak{B}} : A^* \to A$ *be the induced linear isomorphism by* $\mathfrak{B}$. *Let* $r \in A \otimes A$ *be the* 2*-tensor form of* $P \circ I_{\mathfrak{B}}$ *given by* (4-6). *Then* r *is a solution of the Poisson Yang–Baxter equation in* $(A, [\ ,\ ]_A, \cdot_A)$ *and gives rise to a factorizable Poisson bialgebra* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$.

Proof. Let $r_{\mathfrak{B}} \in A \otimes A$ be the 2-tensor form of $I_{\mathfrak{B}}$ in (2-18). By Proposition 2.17 and Lemma 4.7, $r_{\mathfrak{B}}$ is (ad, L)-invariant and $r + \tau(r) = -\lambda r_{\mathfrak{B}}$, which shows that the symmetric part of r is also (ad, L)-invariant and $I_r = -\lambda I_{\mathfrak{B}}$ is a linear isomorphism. By Proposition 2.20, r satisfies the Poisson Yang–Baxter equation, since P is a Rota–Baxter operator of weight λ on $(A, [\ ,\ ]_A, \cdot_A)$. Thus, $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ is a factorizable Poisson bialgebra. □

Corollary 4.12. *Let $(A, [\ ,\ ]_A, \cdot_A)$ be a Poisson algebra, $\{e_1, e_2, \dots, e_n\}$ be a basis of A and $\{e_1^*, e_2^*, \dots, e_n^*\}$ be the dual basis of A^*. Then $r = \sum_i e_i^* \otimes e_i$ is a solution of the Poisson Yang–Baxter equation in $(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot)$ and gives rise to a factorizable Poisson bialgebra $(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot, \delta_r, \Delta_r)$.*

Proof. Clearly, $\mathrm{id}: A \to A$ is a Rota–Baxter operator of weight -1 on $(A, [\ ,\ ]_A, \cdot_A)$. By Proposition 4.5, $\big((A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot), \mathfrak{B}_d, \mathrm{id} + 0^*\big)$ is then a quadratic Rota–Baxter Poisson algebra of weight -1, where $\mathfrak{B}_d$ is defined by (2-16). Now the conclusion follows from Theorem 4.11. □

Example 4.13. Let $(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot)$ be the Poisson algebra in Example 4.6. By Corollary 4.12, $r = \sum_{i=1}^3 e_i^* \otimes e_i$ is a solution of the Poisson Yang–Baxter equation in $(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot)$ and then $(A \ltimes_{-\mathrm{ad}^*, L^*} A^*, [\ ,\ ], \cdot, \delta_r, \Delta_r)$ is a factorizable Poisson bialgebra where δ_r and Δ_r are explicitly defined by

$$\begin{aligned} &\delta_r(e_1) = \delta_r(e_2) = \delta_r(e_3) = 0, \qquad \delta_r(e_1^*) = \delta_r(e_2^*) = 0, \\ &\delta_r(e_3^*) = e_1^* \otimes e_2^* - e_2^* \otimes e_1^*, \end{aligned} \tag{4-7}$$

$$\begin{aligned} &\Delta_r(e_1) = \Delta_r(e_2) = \Delta_r(e_3) = 0, \quad \Delta_r(e_1^*) = \Delta_r(e_2^*) = 0, \\ &\Delta_r(e_3^*) = e_1^* \otimes e_2^* + e_2^* \otimes e_1^*. \end{aligned} \tag{4-8}$$

Proposition 4.14. *Let $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ be a factorizable Poisson bialgebra, which corresponds to a quadratic Rota–Baxter Poisson algebra $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ of weight $\lambda \neq 0$ via Theorems 4.9 and 4.11. Then the factorizable Poisson bialgebra $(A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)})$ corresponds to the quadratic Rota–Baxter Poisson algebra $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, \tilde{P})$ of weight $\lambda \neq 0$ where $\tilde{P} := -\lambda \mathrm{id} - P$. In conclusion, we have the following commutative diagram:*

$$\begin{array}{ccc} (A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r) & \xleftrightarrow{\text{Proposition 3.2}} & (A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)}) \\ \text{Theorem 4.11} \uparrow \downarrow \text{Theorem 4.9} & & \text{Theorem 4.11} \uparrow \downarrow \text{Theorem 4.9} \\ ((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P) & \xleftrightarrow{\text{Proposition 4.4}} & ((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, \tilde{P}) \end{array}$$

Proof. By Theorem 4.9, the factorizable Poisson bialgebra $(A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)})$ induces a quadratic Rota–Baxter Poisson algebra $(A, [\ ,\ ]_A, \cdot_A, \mathfrak{B}', P')$ of weight

$\lambda \neq 0$, where

$$\mathfrak{B}'(a, b) = -\lambda \langle I_{\tau(r)}^{-1}(a), b\rangle = -\lambda \langle I_r^{-1}(a), b\rangle = \mathfrak{B}(a, b), \quad \forall a, b \in A,$$

and

$$P' = -\lambda(\tau(r))_+ \circ I_{\tau(r)}^{-1} = \lambda r_- \circ I_r^{-1} = \lambda r_+ \circ I_r^{-1} - \lambda I_r \circ I_r^{-1} = -P - \lambda \mathrm{id}_A = \tilde{P}.$$

Thus, $(A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)})$ induces a quadratic Rota–Baxter Poisson algebra $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, \tilde{P})$ of weight λ by Theorem 4.9. By a similar argument, we show the converse: the quadratic Rota–Baxter Poisson algebra $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, \tilde{P})$ of weight λ induces the factorizable Poisson bialgebra $(A, [\ ,\ ]_A, \cdot_A, \delta_{\tau(r)}, \Delta_{\tau(r)})$ via Theorem 4.11. □

5. Quasitriangular Poisson bialgebras via quasitriangular differential ASI bialgebras

In this section, we generalize the construction of Poisson algebras from commutative algebras with a pair of commuting derivations to the context of quasitriangular bialgebras. We establish the quasitriangular and factorizable theories for differential ASI bialgebras, and then construct quasitriangular and factorizable Poisson bialgebras from quasitriangular and factorizable (commutative and cocommutative) differential ASI bialgebras respectively.

Recall that an *associative coalgebra* (A, Δ) is a vector space A with a linear map $\Delta : A \to A \otimes A$ satisfying the coassociative law

$$(\Delta \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \Delta)\Delta.$$

An associative coalgebra (A, Δ) is called *cocommutative* if $\Delta = \tau\Delta$.

Definition 5.1 [4]. An *antisymmetric infinitesimal bialgebra* or simply an *ASI bialgebra* is a triple $(A, \cdot_A, \Delta)$ consisting of a vector space A and linear maps $\cdot_A : A \otimes A \to A$ and $\Delta : A \to A \otimes A$ such that

(1) $(A, \cdot_A)$ is an associative algebra,

(2) (A, Δ) is an associative coalgebra, and

(3) the following equations hold for all $a, b \in A$:

$$\Delta(a \cdot_A b) = (R(b) \otimes \mathrm{id})\Delta(a) + (\mathrm{id} \otimes L(a))\Delta(b),$$
$$(L(a) \otimes \mathrm{id} - \mathrm{id} \otimes R(a))\Delta(b) = \tau\left((\mathrm{id} \otimes R(b) - L(b) \otimes \mathrm{id})\Delta(a)\right).$$

Definition 5.2. Let $(A, \cdot_A)$ be an associative algebra. A linear map $\partial : A \to A$ is called a *derivation* if the Leibniz rule is satisfied:

$$\partial(a \cdot_A b) = \partial(a) \cdot_A b + a \cdot_A \partial(b), \quad \forall a, b \in A.$$

A *differential algebra* is a triple $(A, \cdot_A, \Phi)$, where $(A, \cdot_A)$ is an associative algebra and $\Phi = \{\partial_i : A \to A\}_{i=1}^m$ is a tuple of commuting derivations. A differential algebra $(A, \cdot_A, \Phi)$ is called *commutative* if $(A, \cdot_A)$ is commutative.

Definition 5.3 [10]. Let (A, Δ) be an associative coalgebra. A linear map $\eth : A \to A$ is called a *coderivation* on (A, Δ) if

$$\Delta\eth = (\eth \otimes \mathrm{id} + \mathrm{id} \otimes \eth)\Delta. \tag{5-1}$$

A *differential coalgebra* is a triple (A, Δ, Ψ), consisting of an associative coalgebra (A, Δ) and a tuple of commuting coderivations $\Psi = \{\eth_k : A \to A\}_{k=1}^m$. A differential coalgebra (A, Δ, Ψ) is called *cocommutative* if (A, Δ) is cocommutative.

Definition 5.4 [20]. A *differential antisymmetric infinitesimal bialgebra* or simply a *differential ASI bialgebra* is a quintuple $(A, \cdot_A, \Delta, \Phi, \Psi)$ with these properties:

(1) $(A, \cdot_A, \Delta)$ is an ASI bialgebra.

(2) $(A, \cdot_A, \Phi = \{\partial_k\}_{k=1}^m)$ is a differential algebra.

(3) $(A, \Delta, \Psi = \{\eth_k\}_{k=1}^m)$ is a differential coalgebra.

(4) $(A, \cdot_A, \Phi)$ is Ψ*-admissible*, that is, for all $a, b \in A$ and $k = 1, \dots, m$ we have

$$\eth_k(a) \cdot_A b = a \cdot_A \partial_k(b) + \eth_k(a \cdot_A b), \tag{5-2}$$

$$a \cdot_A \eth_k(b) = \partial_k(a) \cdot_A b + \eth_k(a \cdot_A b). \tag{5-3}$$

(5) (A, Δ^*, Ψ^*) is Φ^**-admissible*, that is, for all $k = 1, \dots, m$ we have

$$(\partial_k \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \eth_k)\Delta + \Delta\partial_k, \tag{5-4}$$

$$(\mathrm{id} \otimes \partial_k)\Delta = (\eth_k \otimes \mathrm{id})\Delta + \Delta\partial_k. \tag{5-5}$$

A differential ASI bialgebra $(A, \cdot_A, \Delta, \Phi, \Psi)$ is called *commutative and cocommutative* if $(A, \cdot_A, \Phi)$ is commutative and (A, Δ, Ψ) is cocommutative.

Definition 5.5. Let $(A, \cdot_A, \Phi = \{\partial_k\}_{k=1}^m)$ be a differential algebra. Suppose that $r \in A \otimes A$ and that $\Psi = \{\eth_k : A \to A\}_{k=1}^m$ is a set of commuting linear maps. The Ψ*-admissible associative Yang–Baxter equation* or Ψ*-admissible AYBE* in $(A, \cdot_A, \Phi)$ is the set of equations

$$\boldsymbol{A}(r) = 0,$$

$$(\partial_k \otimes \mathrm{id} - \mathrm{id} \otimes \eth_k)(r) = 0, \quad \forall k = 1, \dots, m, \tag{5-6}$$

$$(\eth_k \otimes \mathrm{id} - \mathrm{id} \otimes \partial_k)(r) = 0, \quad \forall k = 1, \dots, m. \tag{5-7}$$

Lemma 5.6. *Let $(A, \cdot_A, \Phi = \{\partial_k\}_{k=1}^m)$ be a differential algebra. Suppose that $r \in A \otimes A$ and $\Psi = \{\eth_k : A \to A\}_{k=1}^m$ is a set of commuting linear maps. Then equations* (5-6)–(5-7) *hold if and only if*

$$\partial_i r_+ = r_+\eth_i^*, \quad \partial_i r_- = r_-\eth_i^*, \quad \forall i = 1, 2, \dots, m.$$

Proof. For all $x^*, y^* \in A^*$, we have

$$\begin{aligned}
\langle \partial_i r_+(x^*), y^*\rangle &= \langle r, x^* \otimes \partial_i^*(y^*)\rangle = \langle (\mathrm{id} \otimes \partial_i)(r), x^* \otimes y^*\rangle\,,\\
\langle r_+(\partial_i^*(x^*)), y^*\rangle &= \langle r, \partial_i^*(x^*) \otimes y^*\rangle = \langle (\partial_i \otimes \mathrm{id})(r), x^* \otimes y^*\rangle\,,\\
\langle \partial_i r_-(x^*), y^*\rangle &= -\langle r, \partial_i^*(y^*) \otimes x^*\rangle = -\langle (\partial_i \otimes \mathrm{id})(r), y^* \otimes x^*\rangle\,,\\
\langle r_-(\partial_i^*(x^*)), y^*\rangle &= -\langle r, y^* \otimes \partial_i^*(x^*)\rangle = -\langle (\mathrm{id} \otimes \partial_i)(r), y^* \otimes x^*\rangle\,,
\end{aligned}$$

which completes the proof. □

Lemma 5.7. *Let* $(A, \cdot_A, \Phi)$ *be a* Ψ*-admissible differential algebra and* $r \in A \otimes A$. *If* $r \in A \otimes A$ *is a solution of the* Ψ*-admissible AYBE in* $(A, \cdot_A, \Phi)$ *and the symmetric part of* r *satisfies* (2-10), *then* $(A, \cdot_A, \Delta, \Phi, \Psi)$ *is a differential ASI bialgebra*, *where* $\Delta : A \to A \otimes A$ *is defined by*

$$\Delta(a) = (\mathrm{id} \otimes L(a) - R(a) \otimes \mathrm{id})(r), \quad \forall a \in A. \tag{5-8}$$

Proof. This follows from [20, Corollary 4.4]. □

Definition 5.8. Let $(A, \cdot_A, \Phi)$ be a differential algebra and $r \in A \otimes A$. Then r is called *L-invariant* if the following equation holds:

$$(L(a) \otimes \mathrm{id} - \mathrm{id} \otimes R(a))(r) = 0, \quad \forall a \in A. \tag{5-9}$$

Recall that a differential ASI bialgebra $(A, \cdot_A, \Delta, \Phi, \Psi)$ is called *coboundary* if Δ is defined by (5-8) for some $r \in A \otimes A$. A coboundary differential ASI bialgebra is denoted by $(A, \cdot_A, \Delta_r, \Phi, \Psi)$.

Definition 5.9. Let $(A, \cdot_A, \Phi)$ be a Ψ-admissible differential algebra. If r is a solution of the Ψ-admissible AYBE in $(A, \cdot_A, \Phi)$ and the symmetric part of $r \in A \otimes A$ is L-invariant, then the coboundary differential ASI bialgebra $(A, \cdot_A, \Delta_r, \Phi, \Psi)$ induced by r is called a *quasitriangular differential ASI bialgebra*. In particular, if r is antisymmetric, then $(A, \cdot_A, \Delta_r, \Phi, \Psi)$ is called a *triangular differential ASI bialgebra*. A quasitriangular differential ASI bialgebra $(A, \cdot_A, \Delta_r, \Phi, \Psi)$ is called *factorizable* if the symmetric part of r is nondegenerate.

Example 5.10. By Example 4.6, there is a six-dimensional commutative algebra $(A \oplus A^*, \cdot)$ on $A \oplus A^*$ with the nonzero products of $\cdot$ given by (4-5). Let $\partial_1, \partial_2 : A \oplus A^* \to A \oplus A^*$ be linear maps given by

$$\partial_1(e_1) = e_1, \quad \partial_1(e_2) = 0, \quad \partial_1(e_3) = e_3, \quad \partial_1(e_1^*) = 0, \quad \partial_1(e_2^*) = e_2^*, \quad \partial_1(e_3^*) = 0,$$
$$\partial_2(e_1) = 0, \quad \partial_2(e_2) = e_2, \quad \partial_2(e_3) = e_3, \quad \partial_2(e_1^*) = e_1^*, \quad \partial_2(e_2^*) = 0, \quad \partial_2(e_3^*) = 0.$$

Let $\Phi = \{\partial_1, \partial_2\}$ and $\Psi = \{-\partial_1, -\partial_2\}$. Then $(A \oplus A^*, \cdot, \Phi)$ is a Ψ-admissible differential algebra. Moreover, $r = \sum_{i=1}^{3} e_i^* \otimes e_i$ is a solution of the Ψ-admissible AYBE in $(A \oplus A^*, \cdot)$ and $(A \oplus A^*, \cdot, \Delta_r, \Phi, \Psi)$ is a commutative and cocommutative factorizable differential ASI bialgebra with Δ_r given by (4-8).

The next proposition justifies the term "factorizable differential ASI bialgebra."

Definition 5.11. Let $(A, \cdot_A, \Phi_A = \{\partial_{A,i}\}_{i=1}^m)$ and $(B, \cdot_B, \Phi_B = \{\partial_{B,i}\}_{i=1}^m)$ be two differential algebras. A linear map $\varphi : (A, \cdot_A, \Phi_A) \to (B, \cdot_B, \Phi_B)$ is called a *homomorphism* of differential algebras if $\varphi : (A, \cdot_A, \Phi_A) \to (B, \cdot_B, \Phi_B)$ is a homomorphism of associative algebras satisfying

$$\varphi \circ \partial_{A,i} = \partial_{B,i} \circ \varphi, \quad \forall i = 1, 2, \dots, m.$$

If in addition φ is a linear isomorphism, φ is called an *isomorphism* of differential algebras.

Proposition 5.12. *Let $(A, \cdot_A, \Delta_r, \Phi, \Psi)$ be a factorizable differential ASI bialgebra. Then $\mathrm{Im}(r_+ \oplus r_-)$ is a differential subalgebra of the differential algebra $A \oplus A$, which is isomorphic to the differential algebra $(A^*, \cdot_r, \Psi^*)$, where $\cdot_r : A^* \otimes A^* \to A^*$ is defined by*

$$x^* \cdot_r y^* = R^*(r_+(x^*))y^* + L^*(r_-(y^*))x^*, \quad \forall x^*, y^* \in A^*. \tag{5-10}$$

Every $a \in A$ has a unique decomposition $a = a_+ - a_-$ with $(a_+, a_-) \in \mathrm{Im}(r_+ \oplus r_-)$.

Proof. By [27, Proposition 3.2], both $r_+, r_- : (A^*, \cdot_r) \to (A, \cdot_A)$ are associative algebra homomorphisms, $\mathrm{Im}(r_+ \oplus r_-)$ is an associative subalgebra of the direct sum associative algebra $A \oplus A$, which is isomorphic to the associative algebra $(A^*, \cdot_r)$, and every $a \in A$ has a unique decomposition $a = a_+ - a_-$ with $(a_+, a_-) \in \mathrm{Im}(r_+ \oplus r_-)$. By Lemma 5.6, we know that both $r_+, r_- : (A^*, \cdot_r, \Psi^*) \to (A, \cdot_A, \Phi)$ are differential algebra homomorphisms. Therefore, $\mathrm{Im}(r_+ \oplus r_-)$ is a differential subalgebra of the direct sum differential algebra $A \oplus A$, which is isomorphic to the differential algebra $(A^*, \cdot_r, \Psi^*)$. □

Let $(A, \cdot_A, \Delta, \Phi, \Psi)$ be an arbitrary differential ASI bialgebra and $(\mathfrak{A}, \cdot_{\mathfrak{A}})$ be the associative algebra structure on $A \oplus A^*$ obtained from the matched pair of associative algebras $(A, A^*, R^*_{\cdot_A}, L^*_{\cdot_A}, R^*_{\cdot_{A^*}}, L^*_{\cdot_{A^*}})$ [20]. Let $\{e_1, e_2, \dots, e_n\}$ be a basis of A and $\{e_1^*, e_2^*, \dots, e_n^*\}$ be the dual basis of A^*. Define

$$r = \sum_i e_i \otimes e_i^* \in A \otimes A^* \subset \mathfrak{A} \otimes \mathfrak{A}$$

and

$$\Delta_r(u) = (\mathrm{id} \otimes L_{\cdot_{\mathfrak{A}}}(u) - R_{\cdot_{\mathfrak{A}}}(u) \otimes \mathrm{id})(r), \quad \forall u \in \mathfrak{A}.$$

Then $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Delta_r, \Phi + \Psi^*, \Psi + \Phi^*)$ is a coboundary differential ASI bialgebra (see [20, Theorem 4.5] for details).

Theorem 5.13. *The differential ASI bialgebra $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Delta_r, \Phi + \Psi^*, \Psi + \Phi^*)$ is a factorizable quasitriangular differential ASI bialgebra.*

Proof. The proof of Theorem 4.5 in [20] implies that r is a solution of the $(\Psi+\Phi^*)$-admissible AYBE in $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Phi+\Psi^*)$. By suitably modifying the same proof, one can show that the symmetric part of r is L-invariant. By the proof of Theorem 3.4, I_r is an isomorphism of vector spaces. Hence, the differential ASI bialgebra $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Delta_r, \Phi+\Psi^*, \Psi+\Phi^*)$ is factorizable. □

Definition 5.14. A bilinear form $\mathfrak{B}(\ ,\ )$ on an associative algebra $(A, \cdot_A)$ is called *invariant* if

$$\mathfrak{B}(a\cdot_A b, c)=\mathfrak{B}(a, b\cdot_A c),\quad \forall a,b,c\in A.$$

A *Frobenius algebra* $(A, \cdot_A, \mathfrak{B})$ is an associative algebra $(A, \cdot_A)$ with a nondegenerate invariant bilinear form $\mathfrak{B}(\ ,\ )$. A Frobenius algebra $(A, \cdot_A, \mathfrak{B})$ is called *symmetric* if $\mathfrak{B}(\ ,\ )$ is symmetric.

Definition 5.15. A *differential Frobenius algebra* is a quadruple $(A, \cdot_A, \Phi, \mathfrak{B})$, where $(A, \cdot_A, \Phi=\{\partial_k\}_{k=1}^m)$ is a differential algebra and $(A, \cdot_A, \mathfrak{B})$ is a Frobenius algebra. It is called *symmetric* if $\mathfrak{B}$ is symmetric. For all $k=1,\dots,m$, let $\hat{\partial}_k$ be the adjoint linear operator of ∂_k under the nondegenerate bilinear form $\mathfrak{B}$:

$$\mathfrak{B}(\partial_k(a), b)=\mathfrak{B}(a, \hat{\partial}_k(b)),\quad \forall a,b\in A.$$

We call $\hat{\Phi}:=\{\hat{\partial}_k\}_{k=1}^m$ the *adjoint of* $\Phi=\{\partial_k\}_{k=1}^m$ *with respect to* $\mathfrak{B}$.

Note that $\hat{\Phi}$ is admissible to $(A, \cdot_A, \Phi)$, by [20, Proposition 3.3].

Definition 5.16. The triple $((A, \cdot_A, \Phi=\{\partial_k\}_{k=1}^m), \mathfrak{B}, P)$ is called a *symmetric Rota–Baxter differential Frobenius algebra of weight* λ if $(A, \cdot_A, \Phi, \mathfrak{B})$ is a symmetric differential Frobenius algebra and $(A, \cdot_A, P)$ is a Rota–Baxter algebra of weight λ satisfying the compatibility condition given by (4-4) such that

$$\partial_i P=P\partial_i,\quad \forall i=1,2,\dots,m.$$

Lemma 5.17. *Let* $\big((A, \cdot_A, \Phi=\{\partial_k\}_{k=1}^m), \mathfrak{B}, P\big)$ *be a symmetric Rota–Baxter differential Frobenius algebra of weight* λ, *and* $I_{\mathfrak{B}}: A^*\to A$ *be the induced linear isomorphism by* $\mathfrak{B}$. *Let* $r\in A\otimes A$ *be the* 2*-tensor form of* $P\circ I_{\mathfrak{B}}$ *given by* (4-6) *and* $\hat{\Phi}=\{\hat{\partial}_k\}_{k=1}^m$ *be the adjoint of* Φ *with respect to* $\mathfrak{B}$. *Then*

$$\partial_i I_{\mathfrak{B}}=I_{\mathfrak{B}}\hat{\partial}_i^*,\quad \partial_i r_+=r_+\hat{\partial}_i^*,\quad \forall i=1,2,\dots,m.$$

Proof. For all $a\in A$ and $x^*\in A^*$, we have

$$\mathfrak{B}(\partial_i I_{\mathfrak{B}}(x^*), a)=\mathfrak{B}(I_{\mathfrak{B}}(x^*), \hat{\partial}_i(a))=\langle x^*, \hat{\partial}_i(a)\rangle=\langle \hat{\partial}_i^*(x^*), a\rangle=\mathfrak{B}(I_{\mathfrak{B}}\hat{\partial}_i^*(x^*), a),$$

that is, $\partial_i I_{\mathfrak{B}}=I_{\mathfrak{B}}\hat{\partial}_i^*$. Therefore,

$$\partial_i r_+=\partial_i P I_{\mathfrak{B}}=P\partial_i I_{\mathfrak{B}}=P I_{\mathfrak{B}}\hat{\partial}_i^*=r_+\hat{\partial}_i^*,$$

completing the proof. □

Proposition 5.18. *Let $((A, \cdot_A, \Phi), \mathfrak{B}, P)$ be a symmetric Rota–Baxter differential Frobenius algebra of weight 0 and $I_{\mathfrak{B}} : A^* \to A$ be the induced linear isomorphism by $\mathfrak{B}$. Then $(A, \cdot_A, \Delta_r, \Phi, \hat{\Phi})$ is a triangular differential ASI bialgebra where $r \in A \otimes A$ is the 2-tensor form of $P \circ I_{\mathfrak{B}}$ given by* (4-6) *and $\hat{\Phi}$ is the adjoint of Φ with respect to $\mathfrak{B}$.*

Proof. It follows from Lemma 4.7 that $r + \tau(r) = 0$. Similarly to Proposition 2.20, we show that $\boldsymbol{A}(r) = 0$. On the other hand, setting $\Phi = \{\partial_k\}_{k=1}^m$, by Lemma 5.17, we have $\partial_i r_+ = r_+ \hat{\partial}_i^*$. Therefore, by Lemma 5.6, we show that r is a solution of the $\hat{\Phi}$-admissible AYBE in $(A, \cdot_A, \Phi)$. Noting that $\hat{\Phi}$ is admissible to $(A, \cdot_A, \Phi)$, we complete the proof. □

Theorem 5.19. *Let $(A, \cdot_A, \Delta_r, \Phi = \{\partial_k\}_{k=1}^m, \Psi = \{\eth_k\}_{k=1}^m)$ be a factorizable differential ASI bialgebra with $I_r = r_+ - r_-$. Then $((A, \cdot_A, \Phi), \mathfrak{B}, P)$ is a symmetric Rota–Baxter differential Frobenius algebra of weight λ such that the adjoint of Φ with respect to $\mathfrak{B}$ is Ψ, where*

$$P = -\lambda r_+ \circ I_r^{-1}, \quad \lambda \neq 0$$

and the bilinear form $\mathfrak{B} \in \otimes^2 A^*$ is defined by

$$\mathfrak{B}(a, b) = -\lambda \langle I_r^{-1}(a), b \rangle, \quad \forall a, b \in A.$$

Proof. It follows from [27, Corollary 4.8] that $(A, \cdot_A, \mathfrak{B})$ is a symmetric Frobenius algebra and $(A, \cdot_A, P)$ is a Rota–Baxter algebra of weight λ such that (4-4) holds. For all $a, b \in A$, we have

$$\mathfrak{B}(\partial_i(a), b) = -\lambda \langle I_r^{-1}(\partial_i(a)), b \rangle = -\lambda \langle I_r^{-1} \partial_i I_r I_r^{-1}(a), b \rangle,$$
$$\mathfrak{B}(a, \eth_i(b)) = -\lambda \langle I_r^{-1}(a), \eth_i(b) \rangle = -\lambda \langle I_r^{-1} I_r \eth_i^* I_r^{-1}(a), b \rangle,$$

and by Lemma 5.6, we have

$$\partial_i I_r = \partial_i(r_+ - r_-) = (r_+ - r_-)\eth_i^* = I_r \eth_i^*.$$

Therefore, $\mathfrak{B}(\partial_i(a), b) = \mathfrak{B}(a, \eth_i(b))$. Furthermore, we have

$$\partial_i P = -\lambda \partial_i r_+ I_r^{-1} = -\lambda r_+ \eth_i^* I_r^{-1} = -\lambda r_+ I_r^{-1} \partial_i = P \partial_i.$$

The proof is complete. □

Theorem 5.20. *Let $((A, \cdot_A, \Phi = \{\partial_k\}_{k=1}^m), \mathfrak{B}, P)$ be a symmetric Rota–Baxter differential Frobenius algebra of weight $\lambda \neq 0$, and $I_{\mathfrak{B}} : A^* \to A$ be the induced linear isomorphism by $\mathfrak{B}$. Let $r \in A \otimes A$ be the 2-tensor form of $P \circ I_{\mathfrak{B}}$ given by* (4-6) *and $\hat{\Phi}$ be the adjoint of Φ with respect to $\mathfrak{B}$. Then r is a solution of the $\hat{\Phi}$-admissible AYBE in $(A, \cdot_A, \Phi)$ and gives rise to a factorizable differential ASI bialgebra $(A, \cdot_A, \Delta_r, \Phi, \hat{\Phi})$.*

Proof. By Lemma 4.7, $r+\tau(r)$ is equal to $-\lambda r_{\mathfrak{B}}$ and $I_r=-\lambda I_{\mathfrak{B}}$. Then by Lemma 5.17, we have

$$\partial_i r_+ = r_+\hat{\partial}_i^* \quad \text{and} \quad \partial_i r_- = \partial_i(r_+ - I_r) = r_+\hat{\partial}_i^* - I_r\hat{\partial}_i^* = r_-\hat{\partial}_i^*.$$

Similarly to Theorem 4.11, we can show that $\boldsymbol{A}(r)=0$ and the symmetric part of r is L-invariant. Hence, by Lemma 5.6, r is a solution of the $\hat{\Phi}$-admissible AYBE in $(A,\cdot_A,\Phi)$, which completes the proof. □

We now turn to the constructions of quasitriangular and factorizable Poisson bialgebras from quasitriangular and factorizable (commutative and cocommutative) differential ASI bialgebras.

Proposition 5.21 [7]. *Let $(A,\cdot_A,\Phi=\{\partial_1,\partial_2\})$ be a commutative differential algebra. Then $(A,[\ ,\ ]_A,\cdot_A)$ is a Poisson algebra, called the **induced Poisson algebra** of $(A,\cdot_A,\Phi)$, where $[\ ,\ ]: A\otimes A\to A$ is defined by*

$$[a,b]_A := \partial_1(a)\cdot_A\partial_2(b) - \partial_2(a)\cdot_A\partial_1(b),\quad \forall a,b\in A. \tag{5-11}$$

Lemma 5.22 [20]. *Let $(A,\cdot_A,\Phi=\{\partial_1,\partial_2\})$ be a $\Psi=\{\vartheta_1,\vartheta_2\}$-admissible commutative differential algebra and $(A,[\ ,\ ]_A,\cdot_A)$ be the induced Poisson algebra of $(A,\cdot_A,\Phi)$. Suppose that*

$$\vartheta_2(\partial_1(a))\cdot b = \vartheta_1(\partial_2(a))\cdot b,\quad \forall a,b\in A. \tag{5-12}$$

Then every solution of the Ψ-admissible AYBE in $(A,\cdot_A,\Phi)$ is a solution of the Poisson Yang–Baxter equation in the Poisson algebra $(A,[\ ,\ ]_A,\cdot_A)$.

Theorem 5.23. *If $(A,\cdot_A,\Delta_r,\Phi,\Psi)$ is a quasitriangular (resp. triangular, factorizable) commutative and cocommutative differential ASI bialgebra, and equation* (5-12) *holds, then $(A,[\ ,\ ]_A,\cdot_A,\delta_r,\Delta_r)$ is a quasitriangular (resp. triangular, factorizable) Poisson bialgebra through r, where $(A,[\ ,\ ]_A,\cdot_A)$ is the induced Poisson algebra of $(A,\cdot_A,\Phi)$ and $\delta_r: A\to A\otimes A$ is defined by*

$$\delta_r = (\vartheta_1\otimes\vartheta_2 - \vartheta_2\otimes\vartheta_1)\Delta_r. \tag{5-13}$$

Proof. For all $a\in A$, we have

$$\begin{aligned}&(\mathrm{ad}(a)\otimes\mathrm{id}+\mathrm{id}\otimes\mathrm{ad}(a))(r+\tau(r))\\ &= \big((L(\partial_1(a))\partial_2 - L(\partial_2(a))\partial_1)\otimes\mathrm{id} + \mathrm{id}\otimes(L(\partial_1(a))\partial_2 - L(\partial_2(a))\partial_1)\big)(r+\tau(r)).\end{aligned}$$

By (5-2), this equals

$$\begin{aligned}&\big(L(\partial_1(a))\otimes\vartheta_2 - L(\partial_2(a))\otimes\vartheta_1 - \mathrm{id}\otimes\vartheta_2 L(\partial_1(a)) + \mathrm{id}\otimes\vartheta_1 L(\partial_2(a))\big)(r+\tau(r))\\ &\qquad + \big(\mathrm{id}\otimes L(\vartheta_2\partial_1(a)) - \mathrm{id}\otimes L(\vartheta_1\partial_2(a))\big)(r+\tau(r)).\end{aligned}$$

Using (5-12) this becomes

$$\big((\mathrm{id}\otimes\vartheta_2)(L(\partial_1(a))\otimes\mathrm{id}-\mathrm{id}\otimes L(\partial_1(a))) \\ +(\mathrm{id}\otimes\vartheta_1)(L(\partial_2(a))\otimes\mathrm{id}-\mathrm{id}\otimes L(\partial_2(a)))\big)(r+\tau(r))=0.$$

Hence, the symmetric part of r is (ad, L)-invariant. By Lemma 5.22, r is a solution of the Poisson Yang–Baxter equation in $(A,[\ ,\ ]_A,\cdot_A)$. Furthermore,

$$\begin{aligned}\delta_r(a) &= (\vartheta_1\otimes\vartheta_2-\vartheta_2\otimes\vartheta_1)\Delta_r(a)=(\vartheta_1\otimes\vartheta_2-\vartheta_2\otimes\vartheta_1)(\mathrm{id}\otimes L(a)-L(a)\otimes\mathrm{id})(r)\\ &\overset{(5\text{-}6)}{=} (\mathrm{id}\otimes\vartheta_2L(a)\partial_1-\mathrm{id}\otimes\vartheta_1L(a)\partial_2)(r)-(\vartheta_1L(a)\partial_2\otimes\mathrm{id}-\vartheta_2L(a)\partial_1\otimes\mathrm{id})(r)\\ &\overset{(5\text{-}2)}{=} (\mathrm{ad}(a)\otimes\mathrm{id}+\mathrm{id}\otimes\mathrm{ad}(a))(r).\end{aligned}$$

It is now obvious that the theorem holds. □

Example 5.24. Continuing with Example 5.10, there is a commutative and cocommutative factorizable differential ASI bialgebra $(A\oplus A^*,\cdot,\Delta_r,\Phi,\Psi)$. Note that (5-12) holds automatically. Thus by Theorem 5.23, it induces a factorizable Poisson bialgebra $(A\oplus A^*,[\ ,\ ],\cdot,\delta_r,\Delta_r)$ where the nonzero product of $[\ ,\ ]$ is defined by $[e_1,e_2]=e_3$ and δ_r is explicitly given by (4-8).

Let $\big(A,\cdot_A,\Delta_r,\Phi=\{\partial_1,\partial_2\},\Psi=\{\vartheta_1,\vartheta_2\}\big)$ be a commutative and cocommutative factorizable differential ASI bialgebra. Using Proposition 5.12, let

$$\big(\mathrm{Im}(r_+\oplus r_-),\cdot,\Phi+\Phi=\{\partial_1+\partial_1,\partial_2+\partial_2\}\big)$$

be the differential subalgebra of the differential algebra $A\oplus A$, which is isomorphic to the differential algebra $(A^*,\cdot_r,\Psi^*)$, where $\cdot_r: A^*\otimes A^*\to A^*$ is defined by (2-4). On the other hand, suppose that (5-12) holds. Let $(A,[\ ,\ ]_A,\cdot_A,\delta_r,\Delta_r)$ be the induced factorizable Poisson bialgebra in Theorem 5.23. Then by Proposition 3.3, let $(\mathrm{Im}(r_+\oplus r_-),[\ ,\ ],\cdot)$ be the Poisson subalgebra of the direct sum Poisson algebra $A\oplus A$, which is isomorphic to the Poisson algebra $(A^*,[\ ,\ ]_r,\cdot_r)$, where $[\ ,\ ]_r,\cdot_r: A^*\otimes A^*\to A^*$ are respectively defined by (2-3) and (2-4).

Corollary 5.25. *With the same conditions as above, the induced Poisson algebras of* $(\mathrm{Im}(r_+\oplus r_-),\cdot,\Phi+\Phi)$ *and* $(A^*,\cdot_r,\Psi^*)$ *are exactly* $(\mathrm{Im}(r_+\oplus r_-),[\ ,\ ],\cdot)$ *and* $(A^*,[\ ,\ ]_r,\cdot_r)$. *Thus we have the commutative diagram*

$$\begin{array}{ccccc}(A,\cdot_A,\Delta_r,\Phi,\Psi) & \xrightarrow{\text{Proposition 5.12}} & (\mathrm{Im}(r_+\oplus r_-),\cdot,\Phi+\Phi) & \xrightarrow{\simeq} & (A^*,\cdot_r,\Psi^*)\\ \text{Theorem 5.23}\big\downarrow & & \text{Proposition 5.21}\big\downarrow & & \text{Proposition 5.21}\big\downarrow\\ (A,[\ ,\ ]_A,\cdot_A,\delta_r,\Delta_r) & \xrightarrow{\text{Proposition 3.3}} & (\mathrm{Im}(r_+\oplus r_-),[\ ,\ ],\cdot) & \xrightarrow{\simeq} & (A^*,[\ ,\ ]_r,\cdot_r)\end{array}$$

Proof. By Theorem 5.23, the induced Poisson algebra of $(A^*,\cdot_r,\Psi^*)$ is exactly $(A^*,[\ ,\ ]_r,\cdot_r)$.Noting that $(A,[\ ,\ ]_A,\cdot_A)$ is the induced Poisson algebra of $(A,\cdot_A,\Phi)$, it is straightforward that the induced Poisson algebra of $(\mathrm{Im}(r_+\oplus r_-),\cdot,\Phi+\Phi)$ is exactly $(\mathrm{Im}(r_+\oplus r_-),[\ ,\ ],\cdot)$. The proof is complete. □

Let $(A, \cdot_A, \Delta, \Phi = \{\partial_1, \partial_2\}, \Psi = \{\eth_1, \eth_2\})$ be a commutative and cocommutative differential ASI bialgebra. Let $(A, [\ ,\ ]_A, \cdot_A)$ be the induced Poisson algebra of $(A, \cdot_A, \Phi)$. Suppose that (5-12) and the following equation hold:

$$(\eth_2\partial_1 \otimes \mathrm{id})\Delta = (\eth_1\partial_2 \otimes \mathrm{id})\Delta. \tag{5-14}$$

Then $(A, [\ ,\ ]_A, \cdot_A, \delta, \Delta)$ is a Poisson bialgebra where $\delta : A \to A \otimes A$ is defined by $\delta = (\eth_1 \otimes \eth_2 - \eth_2 \otimes \eth_1)\Delta$ [20, Theorem 5.24]. Let $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Delta_r, \Phi + \Psi^*, \Psi + \Phi^*)$ be the factorizable differential ASI bialgebra obtained in Theorem 5.13 arising from $(A, \cdot_A, \Delta, \Phi, \Psi)$ and let $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)$ be the factorizable Poisson bialgebra of Theorem 3.4 arising from $(A, [\ ,\ ]_A, \cdot_A, \delta, \Delta)$. Since the Poisson algebra structure $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}})$ is the induced Poisson algebra of $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Phi + \Psi^*)$ [20, Remark 5.25] and equation (5-12) also holds for the $(\Psi + \Phi^*)$-admissible commutative differential algebra $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Phi + \Psi^*)$, we have the following conclusion.

Corollary 5.26. *With the same conditions as above*, $(\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)$ *is the induced factorizable Poisson bialgebra of the factorizable differential ASI bialgebra* $(\mathfrak{A}, \cdot_{\mathfrak{A}}, \Delta_r, \Phi + \Psi^*, \Psi + \Phi^*)$. *Thus, we have the following commutative diagram*:

$$\begin{array}{ccc}
(A, \cdot_A, \Delta, \Phi, \Psi) & \xrightarrow{\text{Theorem 5.13}} & (\mathfrak{A}, \cdot_{\mathfrak{A}}, \Delta_r, \Phi + \Psi^*, \Psi + \Phi^*) \\
\downarrow & & \downarrow{\scriptstyle\text{Theorem 5.23}} \\
(A, [\ ,\ ]_A, \cdot_A, \delta, \Delta) & \xrightarrow{\text{Theorem 3.4}} & (\mathfrak{A}, [\ ,\ ]_{\mathfrak{A}}, \cdot_{\mathfrak{A}}, \delta_r, \Delta_r)
\end{array}$$

Proposition 5.27. *Let* $(A, \cdot_A, \Delta_r, \Phi, \Psi)$ *be a commutative and cocommutative factorizable differential ASI bialgebra, which corresponds to a symmetric Rota–Baxter differential Frobenius algebra* $((A, \cdot_A, \Phi), \mathfrak{B}, P)$ *of weight* $\lambda \neq 0$. *Suppose that* (5-12) *holds. Let* $(A, [\ ,\ ]_A, \cdot_A)$ *be the induced Poisson algebra of* $(A, \cdot_A, \Phi)$. *Then* $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ *is a quadratic Rota–Baxter Poisson algebra of weight* λ, *to which the induced factorizable Poisson bialgebra* $(A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r)$ *in Theorem 5.23 corresponds. Thus, we have the following commutative diagram*:

$$\begin{array}{ccc}
(A, \cdot_A, \Delta_r, \Phi, \Psi) & \xrightarrow{\text{Theorem 5.23}} & (A, [\ ,\ ]_A, \cdot_A, \delta_r, \Delta_r) \\
{\scriptstyle\text{Theorem 5.20}}\uparrow\ \downarrow{\scriptstyle\text{Theorem 5.19}} & & {\scriptstyle\text{Theorem 4.11}}\uparrow\ \downarrow{\scriptstyle\text{Theorem 4.9}} \\
((A, \cdot_A, \Phi), \mathfrak{B}, P) & \longrightarrow & ((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)
\end{array}$$

Proof. For all $a, b, c \in A$, we have

$$\begin{aligned}
&\mathfrak{B}([a, b]_A, c) \\
&\quad = \mathfrak{B}(\partial_1(a) \cdot_A \partial_2(b) - \partial_2(a) \cdot_A \partial_1(b), c) = \mathfrak{B}(a, \eth_1(\partial_2(b) \cdot_A c) - \eth_2(\partial_1(b) \cdot_A c)) \\
&\quad = \mathfrak{B}(a, \eth_1(\partial_2(b)) \cdot_A c - \partial_2(b) \cdot_A \partial_1(c) - \eth_2(\partial_1(b)) \cdot_A c + \partial_1(b) \cdot_A \partial_2(c)) \\
&\quad = \mathfrak{B}(a, [b, c]_A)
\end{aligned}$$

and

$$
\begin{aligned}
[P(a), P(b)]_A &= \partial_1(P(a)) \cdot_A \partial_2(P(b)) - \partial_2(P(a)) \cdot_A \partial_1(P(b)) \\
&= P(\partial_1(a)) \cdot_A P(\partial_2(b)) - P(\partial_2(a)) \cdot_A P(\partial_1(b)) \\
&= P\big(P(\partial_1(a)) \cdot_A \partial_2(b) + \partial_1(a) \cdot_A P(\partial_2(b)) + \lambda\partial_1(a) \cdot_A \partial_2(b)\big) \\
&\quad - P\big(P(\partial_2(a)) \cdot_A \partial_1(b) + \partial_2(a) \cdot_A P(\partial_1(b)) + \lambda\partial_2(a) \cdot_A \partial_1(b)\big) \\
&= P\big(\partial_1(P(a)) \cdot_A \partial_2(b) + \partial_1(a) \cdot_A \partial_2(P(b)) + \lambda\partial_1(a) \cdot_A \partial_2(b)\big) \\
&\quad - P\big(\partial_2(P(a)) \cdot_A \partial_1(b) + \partial_2(a) \cdot_A \partial_1(P(b)) + \lambda\partial_2(a) \cdot_A \partial_1(b)\big) \\
&= P\big([P(a), b]_A + [a, P(b)]_A + \lambda[a, b]_A\big).
\end{aligned}
$$

Therefore, $((A, [\ ,\ ]_A, \cdot_A), \mathfrak{B}, P)$ is a quadratic Rota–Baxter Poisson algebra of weight λ. Thus the proposition is true. □

References

[1] M. Aguiar, "Infinitesimal Hopf algebras", pp. 1–29 in *New trends in Hopf algebra theory* (La Falda, 1999), Contemp. Math. **267**, Amer. Math. Soc., 2000. MR

[2] M. Aguiar, "On the associative analog of Lie bialgebras", *J. Algebra* **244**:2 (2001), 492–532. MR

[3] V. I. Arnold, *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics **60**, Springer, 1978. MR

[4] C. Bai, "Double constructions of Frobenius algebras, Connes cocycles and their duality", *J. Noncommut. Geom.* **4**:4 (2010), 475–530. MR

[5] C. Bai, L. Guo, and X. Ni, "Nonabelian generalized Lax pairs, the classical Yang–Baxter equation and PostLie algebras", *Comm. Math. Phys.* **297**:2 (2010), 553–596. MR

[6] C. Bai, L. Guo, and X. Ni, "$\mathcal{O}$-operators on associative algebras and associative Yang–Baxter equations", *Pacific J. Math.* **256**:2 (2012), 257–289. MR

[7] K. H. Bhaskara and K. Viswanath, *Poisson algebras and Poisson manifolds*, Pitman Research Notes in Mathematics **174**, Longman, Harlow, UK, 1988. MR

[8] V. Chari and A. Pressley, *A guide to quantum groups*, Cambridge University Press, 1994. MR

[9] P. A. M. Dirac, *Lectures on quantum mechanics*, Belfer Graduate School of Science Monographs Series **2**, Academic Press, New York, 1967. MR

[10] Y. Doi, "Homological coalgebra", *J. Math. Soc. Japan* **33**:1 (1981), 31–50. MR

[11] V. G. Drinfeld, "Hamiltonian structures on Lie groups, Lie bialgebras and the geometric meaning of classical Yang–Baxter equations", *Dokl. Akad. Nauk SSSR* **268**:2 (1983), 285–287. In Russian; translated in *Soviet Math. Dokl.* **27** (1983), 68–71. MR

[12] V. G. Drinfeld, "Quantum groups", pp. 798–820 in *Proceedings of the International Congress of Mathematicians* (Berkeley, 1986), vol. 1, Amer. Math. Soc., 1987. MR

[13] V. Ginzburg and D. Kaledin, "Poisson deformations of symplectic quotient singularities", *Adv. Math.* **186**:1 (2004), 1–57. MR

[14] J. Huebschmann, "Poisson cohomology and quantization", *J. Reine Angew. Math.* **408** (1990), 57–113. MR

[15] S. A. Joni and G.-C. Rota, "Coalgebras and bialgebras in combinatorics", *Stud. Appl. Math.* **61**:2 (1979), 93–139. MR

[16] C. Kassel, *Quantum groups*, Graduate Texts in Mathematics **155**, Springer, 1995. MR

[17] M. Kontsevich, "Deformation quantization of Poisson manifolds", *Lett. Math. Phys.* **66**:3 (2003), 157–216. MR

[18] H. Lang and Y. Sheng, "Factorizable Lie bialgebras, quadratic Rota–Baxter Lie algebras and Rota–Baxter Lie bialgebras", *Comm. Math. Phys.* **397**:2 (2023), 763–791. MR

[19] A. Lichnerowicz, "Les variétés de Poisson et leurs algèbres de Lie associées", *J. Differential Geometry* **12**:2 (1977), 253–300. MR

[20] Y. Lin, X. Liu, and C. Bai, "Differential antisymmetric infinitesimal bialgebras, coherent derivations and Poisson bialgebras", *SIGMA Symmetry Integrability Geom. Methods Appl.* **19** (2023), art. id. 018, 47 pp. MR

[21] X. Ni and C. Bai, "Poisson bialgebras", *J. Math. Phys.* **54**:2 (2013), art. id. 023515, 14 pp. MR

[22] A. Odzijewicz, "Hamiltonian and quantum mechanics", pp. 385–472 in *Lectures on Poisson geometry*, Geom. Topol. Monogr. **17**, Geom. Topol. Publ., Coventry, UK, 2011. MR

[23] A. Polishchuk, "Algebraic geometry of Poisson brackets", *J. Math. Sci.* (*New York*) **84**:5 (1997), 1413–1444. Algebraic geometry, 7. MR

[24] N. Y. Reshetikhin and M. A. Semenov-Tian-Shansky, "Quantum R-matrices and factorization problems", *J. Geom. Phys.* **5**:4 (1988), 533–550. MR

[25] M. A. Semenov-Tian-Shansky, "Integrable systems and factorization problems", pp. 155–218 in *Factorization and integrable systems* (Faro, Portugal, 2000), Oper. Theory Adv. Appl. **141**, Birkhäuser, Basel, 2003. MR

[26] M. A. Semenov-Tyan-Shansky, "What is a classical r-matrix?", *Funktsional. Anal. i Prilozhen.* **17**:4 (1983), 17–33. In Russian; translated in *Funct. Anal. Appl.* **17** (1983), 259–272. MR

[27] Y. Sheng and Y. Wang, "Quasi-triangular and factorizable antisymmetric infinitesimal bialgebras", *J. Algebra* **628** (2023), 415–433. MR

[28] I. Vaisman, *Lectures on the geometry of Poisson manifolds*, Progress in Mathematics **118**, Birkhäuser, Basel, 1994. MR

[29] A. Weinstein, *Lectures on symplectic manifolds*, CBMS Regional Conference Series in Mathematics **29**, Amer. Math. Soc., 197k. MR

[30] V. N. Zhelyabin, "Jordan bialgebras and their connection with Lie bialgebras", *Algebra i Logika* **36**:1 (1997), 3–25, 117. In Russian; translated in *Algebra Logic* **36**:1 (1997), 1–15. MR

Received October 7, 2025. Revised February 20, 2026.

YUANCHANG LIN
SCHOOL OF MATHEMATICS
NORTH UNIVERSITY OF CHINA
TAIYUAN
CHINA
linyuanchang@mail.nankai.edu.cn

DILEI LU
COLLEGE OF APPLIED SCIENCE
BEIJING INFORMATION SCIENCE AND TECHNOLOGY UNIVERSITY
BEIJING, 100192
CHINA
ludyray@bistu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 343, No. 2, 2026
https://doi.org/10.2140/pjm.2026.343.485

SYMPLECTIC SEMI-CHARACTERISTICS

HAO ZHUANG

We define and study the symplectic semi-characteristic of a closed $4n$-dimensional symplectic manifold, based on the even-degree part of the primitive cohomology. Using a vector field with nondegenerate zero points, we prove a counting formula for the symplectic semi-characteristic. As corollaries of the counting formula, we obtain a vanishing property and the fact that the definition of the symplectic semi-characteristic is independent of the choice of symplectic forms.

1. Introduction

In three consecutive papers [18; 19; 20], Tsai, Tseng, and Yau introduced and studied the p-filtered cohomology groups $F^pH^k(M,\omega)$ $(0 \leq k \leq 1+2p+\dim M)$ of a symplectic manifold (M,ω). The $p=0$ case in [18, (3.14), (3.22)] and [19, (1.5), (1.6)] is called the *primitive cohomology* of (M,ω), and the $p \geq 1$ case in [20, (1.2), Theorem 3.1] is generalized from it. Different from the classical de Rham cohomology, this p-filtered cohomology includes information about the symplectic form ω. Thus, an important application of it is in distinguishing among different symplectic structures; examples can be found in [19, Section 4] (computing the primitive cohomology groups) and [20, Section 6] (computing the product structures). Tanaka and Tseng [16, Theorem 1.1] proved that the mapping cone complex determined by the map $\wedge\omega^{p+1}$ between de Rham complexes computes the p-filtered cohomology.

MSC2020: primary 58J20; secondary 53D05.
Keywords: mod 2 index, primitive cohomology, symplectic semi-characteristic.

In this paper, we focus on the primitive cohomology. Our work starts from an interesting fact: When the symplectic manifold (M, ω) is closed, the Euler characteristic

$$\dim F^0 H^{\mathrm{even}}(M, \omega) - \dim F^0 H^{\mathrm{odd}}(M, \omega) \tag{1-1}$$

of the primitive cohomology is equal to zero. This was originally proved in [18, Proposition 3.26] and [19, Proposition 3.7] by means of duality between cohomology groups. Recently, using Tanaka and Tseng's mapping cone complex, Clausen, Tang, and Tseng proved the symplectic Morse inequality [7, Theorem 1.4], also showing that the value of (1-1) is zero. Intuitively speaking, we may say that the even-degree part of the primitive cohomology contains the same amount of information as the odd-degree part.

We then ask whether $F^0 H^{\mathrm{even}}(M, \omega)$ is an obstruction to some geometric object. Here, to some extent, we are motivated by the Kervaire semi-characteristic in a similar scenario. Recall that for any odd-dimensional closed oriented manifold N, the Euler characteristic of the de Rham cohomology of N is zero. Let b_k be the dimension of the k-th de Rham cohomology group of N. The $\mathbb{Z}_2$-valued Kervaire semi-characteristic of N [11, Introduction; 12, Section 1; 1, Section 4] is defined to be

$$\sum_{k \text{ even}} b_k \mod 2. \tag{1-2}$$

When $\dim N = 4n+1$, the $\mathbb{Z}_2$-valued Kervaire semi-characteristic satisfies Atiyah's vanishing theorem [1, Theorem 4.1] and Zhang's counting formula [23, Theorem 1.3]. Both the vanishing theorem and the counting formula involve two vector fields, showing that the $\mathbb{Z}_2$-valued Kervaire semi-characteristic is an obstruction to a certain pair of vector fields. Now, back to our symplectic situation, we state our main question:

Question 1.1. For a closed symplectic manifold (M, ω), what geometric object(s) on M does the even-degree part of the primitive cohomology of (M, ω) obstruct? Is it a pair of vector fields or something else?

Here we answer this question for $4n$-dimensional closed symplectic manifolds.

Assumption 1.2. Throughout this paper, when not stated otherwise, (M, ω) is a $4n$-dimensional closed symplectic manifold M equipped with a symplectic form ω.

We recall Tanaka and Tseng's mapping cone complex $(C^*(M, \omega), \partial_C)$, which computes the primitive cohomology of (M, ω). Let $\Omega^k(M)$ be the space of smooth k-forms on M. The space of k-cochains [16, Section 3.1; 7, Definition 1.1] is

$$C^k(M, \omega) := \Omega^k(M) \oplus \Omega^{k-1}(M) \quad (k = 0, 1, \ldots, 4n+1).$$

Let d be de Rham exterior differentiation. The boundary map is

$$\partial_C : C^k(M,\omega) \to C^{k+1}(M,\omega) \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \mapsto \begin{bmatrix} d & \omega \\ 0 & -d \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} d\alpha + \omega \wedge \beta \\ -d\beta \end{bmatrix}.$$

Here, we write the pair $(\alpha, \beta) \in \Omega^k(M) \oplus \Omega^{k-1}(M)$ as a column for the convenience of using matrices and operators later.

The k-th cohomology group of $(C^*(M,\omega), \partial_C)$ is exactly $F^0H^k(M,\omega)$.

Definition 1.3. Let b_k^ω be the dimension of $F^0H^k(M,\omega)$. The *symplectic semi-characteristic* of (M,ω) is the $\mathbb{Z}_2$-valued number

$$\kappa(M,\omega) := \sum_{k \text{ even}} b_k^\omega \bmod 2. \tag{1-3}$$

To state our main result, we review the definition of nondegenerate vector fields. Let V be a smooth vector field on M. Following [5, Section 1.6], at each zero point p of V, we define a homomorphism

$$\Phi_p : T_pM \to T_pM, \quad v \mapsto [V, \tilde{v}](p),$$

where $\tilde{v}$ is a vector field extending the tangent vector v and $[\,\cdot\,,\cdot\,]$ is the Lie bracket between vector fields. This Φ_p is independent of the extension $\tilde{v}$ since $V(p) = 0$.

Definition 1.4. A smooth vector field V on M is called *nondegenerate* if either V vanishes nowhere or Φ_p is invertible for each zero point p of V.

Such a nondegenerate vector field always exists because by [13, Theorem 6.6], there is always a Morse function on M. Now, our main result is as follows:

Theorem 1.5 (compare [23, Theorem 1.3]). *Let V be a smooth nondegenerate vector field on M. The counting formula for the symplectic semi-characteristic under Assumption 1.2 is*

$$\kappa(M,\omega) = \textit{the number of zero points of } V \bmod 2. \tag{1-4}$$

Remark 1.6. By the Poincaré–Hopf index formula [24, Theorem 4.5], formula (1-4) also equals (mod 2) the Euler characteristic of the de Rham cohomology of M.

Remark 1.7. A special situation is when the de Rham cohomology class of ω is integral. Then, by [16, Theorem 7.1], Theorem 1.5 computes the classical Kervaire semi-characteristic (1-2) of the circle bundle over M induced by the line bundle associated with ω.

The main idea of the proof is that we find a skew-adjoint operator, as in Zhang's construction [23, (1.1)]. Then, we show that $\kappa(M,\omega)$ is equal to the Atiyah–Singer mod 2 index Definition 2.5 of this operator. Next, as in [23, (2.1)], we apply a Witten deformation [22, Section 2] and Bismut and Lebeau's asymptotic analysis

[6, Chapters VIII–X; 24, Chapters 4-7] to the operator, compute its mod 2 index, and then obtain Theorem 1.5.

A corollary of Theorem 1.5 is an Atiyah-type vanishing property:

Corollary 1.8 (compare [1, Theorem 4.1]). *The semi-characteristic $\kappa(M, \omega)$ vanishes when there is a nonvanishing smooth vector field on M.*

Another way to prove Corollary 1.8 without using Theorem 1.5 is described in Remark 4.2.

The converse of Corollary 1.8 is not true; see Example 5.2. This contrasts with the Euler characteristic of the de Rham cohomology of M.

Finally, although we have used the symplectic form ω to define $\kappa(M, \omega)$, a nondegenerate vector field always exists and is independent of ω. Thus:

Corollary 1.9. *The definition of $\kappa(M, \omega)$ is independent of the chosen symplectic form.*

Remark 1.10. The definition of the symplectic semi-characteristic can be assigned to any closed symplectic manifold without assuming that $\dim M = 4n$. However, by Example 5.4, the counting formula does not work when the dimension is $4n+2$. Thus, the $(4n+2)$-dimensional part of Question 1.1 is still open.

Outline. In Section 2, we review Clifford actions and find a skew-adjoint operator so that $\kappa(M, \omega)$ equals the Atiyah–Singer mod 2 index of this operator. In Section 3, we carry out necessary analytic details about this operator. In Section 4, we prove Theorem 1.5 based on those analytic details. In Section 5, we give some examples, and Section 6 describes further developments.

2. Clifford actions and operators

In this section, we clarify technical details about the Clifford actions of tangent vectors and introduce the skew-adjoint operator that we will work with.

After choosing an almost complex structure J on M, we let g be the Riemannian metric on M:

$$g(\cdot\,,\cdot) = \omega(\cdot\,, J\cdot).$$

We equip M with the orientation $\omega \wedge \cdots \wedge \omega$ and let $\star$ be the Hodge star operator. Letting $\mathrm{dvol} = \star 1$ be the volume form of M, we define the L^2-norm (and inner product) by

$$\|\alpha\| = \left(\int_M g(\alpha, \alpha)\,\mathrm{dvol}\right)^{1/2} \tag{2-1}$$

on $\Omega^k(M)$. We require that $\Omega^k(M) \perp \Omega^\ell(M)$ when $k \neq \ell$. For pairs of forms on

$C^k(M,\omega) = \Omega^k(M) \oplus \Omega^{k-1}(M)$, following [7, (2.2)], we define

$$g\left(\begin{bmatrix}\alpha\\ \beta\end{bmatrix}, \begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}\right) = g(\alpha,\alpha') + g(\beta,\beta').$$

As in (2-1), we have the L^2-norm (and inner product)

$$\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| = \left(\int_M g(\alpha,\alpha)\,\mathrm{dvol} + \int_M g(\beta,\beta)\,\mathrm{dvol}\right)^{1/2} \tag{2-2}$$

on $C^k(M,\omega)$. We require $C^k(M,\omega) \perp C^\ell(M,\omega)$ when $k \neq \ell$.

Let d^* be the formal adjoint of d with respect to the inner product induced by (2-1), and

$$\omega^*\lrcorner : \Omega^k(M) \to \Omega^{k-2}(M)$$

be the adjoint of

$$\omega\wedge : \Omega^k(M) \to \Omega^{k+2}(M), \quad \alpha \mapsto \omega \wedge \alpha,$$

with respect to the same inner product. For convenience, we will omit the "$\lrcorner$" after ω^* and the "$\wedge$" after ω when there is no ambiguity. Recall the mapping cone complex

$$\partial_C : \Omega^k(M) \oplus \Omega^{k-1}(M) \to \Omega^{k+1}(M) \oplus \Omega^k(M)(\alpha,\beta) \mapsto \begin{bmatrix} d & \omega \\ 0 & -d\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}. \tag{2-3}$$

The formal adjoint of ∂_C is

$$\partial_C^* = \begin{bmatrix} d^* & 0 \\ \omega^* & -d^*\end{bmatrix}$$

with respect to the inner product induced by (2-2).

Proposition 2.1. *The kernel of the Dirac type operator $\partial_C + \partial_C^*$ and that of the Laplacian $(\partial_C + \partial_C^*)^2$ are isomorphic to the primitive cohomology of (M,ω). In particular,*

$$\ker\left((\partial_C + \partial_C^*)^2 : C^k(M,\omega) \to C^k(M,\omega)\right)$$

is isomorphic to the k-th primitive cohomology group.

Proof. We can check that (2-3) defines an elliptic complex [14, Definition 10.4.28]. By the properties of such a complex [14, Theorem 10.4.30], the complex defined by (2-3) satisfies the Hodge decomposition theorem. Therefore, the kernel of $(\partial_C + \partial_C^*)^2|_{C^k(M,\omega)}$ is isomorphic to the k-th primitive cohomology group. □

For any (globally or locally defined) vector field Y on M, we have two Clifford actions:

$$\hat{c}(Y) = Y^* \wedge + Y\lrcorner \quad \text{and} \quad c(Y) = Y^* \wedge - Y\lrcorner.$$

Given any oriented local orthonormal frame $e_1, \dots, e_{4n}$ of TM, the Clifford action of the volume form dvol is expressed as

$$\hat{c}(\mathrm{dvol}) = \hat{c}(e_1) \cdots \hat{c}(e_{4n}).$$

This is independent of the choice of oriented local orthonormal frames. Following [1, Section 3], we lay out some interactions between the Hodge star, Clifford actions, and differential forms. Recall that the dimension of M is $4n$.

Lemma 2.2. *For all $\alpha \in \Omega^k(M)$, we have $\hat{c}(\mathrm{dvol})\alpha = (-1)^{k(k+1)/2} \star \alpha$.*

Proof. Let $e_1, \dots, e_{4n}$ be an oriented local orthonormal frame. Suppose $\alpha = e_{i_1}^* \wedge \cdots \wedge e_{i_k}^*$ and choose indices $j_1, \dots, j_{4n-k}$ so that

$$e_{i_1}^* \wedge \cdots \wedge e_{i_k}^* \wedge e_{j_1}^* \wedge \cdots \wedge e_{j_{4n-k}}^* = e_1^* \wedge \cdots \wedge e_{4n}^*.$$

Then we have

$$\begin{aligned}
\hat{c}(\mathrm{dvol})\alpha &= \hat{c}(e_{i_1})\hat{c}(e_{i_2}) \cdots \hat{c}(e_{i_k})\hat{c}(e_{j_1}) \cdots \hat{c}(e_{j_{4n-k}})\alpha \\
&= (-1)^{k(4n-k)} \hat{c}(e_{j_1}) \cdots \hat{c}(e_{j_{4n-1}})\hat{c}(e_{i_1})\hat{c}(e_{i_2}) \cdots \hat{c}(e_{i_k})\alpha \\
&= (-1)^{k(4n-k)+\frac{1}{2}(0+k-1)k} \star (e_{i_1}^* \wedge \cdots \wedge e_{i_k}^*) \\
&= (-1)^{k(k+1)/2} \star \alpha.
\end{aligned}$$

The case of general α is straightforward. □

Lemma 2.3. *For all $\alpha \in \Omega^k(M)$, we have $\hat{c}(\mathrm{dvol})(\omega^* \lrcorner \alpha) = -\omega \wedge \hat{c}(\mathrm{dvol})\alpha$.*

Proof. Using the equality $\omega^* \lrcorner \alpha = (-1)^k \star \omega \star \alpha$ from [7, Section 2.1], we find

$$\begin{aligned}
\hat{c}(\mathrm{dvol})(\omega^* \lrcorner \alpha) &= \hat{c}(\mathrm{dvol})((-1)^k \star \omega \star \alpha) \\
&= (-1)^{\frac{1}{2}(k-2+1)(k-2)} \star ((-1)^k \star \omega \star \alpha) \quad \text{(by Lemma 2.2)} \\
&= (-1)^{\frac{1}{2}(k-2+1)(k-2)} \star \big((-1)^k \star \big(\omega \wedge (-1)^{\frac{1}{2}k(k+1)} \hat{c}(\mathrm{dvol})\alpha\big)\big) \\
&= (-1)^{k^2+1} \star\star \big(\omega \wedge (\hat{c}(\mathrm{dvol})\alpha)\big).
\end{aligned}$$

When k is odd, $\omega \wedge \hat{c}(\mathrm{dvol})\alpha$ is an odd-degree form, making $\star\star = -1$ and then $(-1)^{k^2+1}\star\star = -1$. Similarly, when k is even, we have $\star\star = 1$ and then $(-1)^{k^2+1}\star\star = -1$. Thus, we obtain $\hat{c}(\mathrm{dvol})(\omega^* \lrcorner \alpha) = -\omega \wedge (\hat{c}(\mathrm{dvol})\alpha)$. □

We set $\Omega^{\mathrm{even}}(M) := \bigoplus_{k=0}^{2n} \Omega^{2k}(M)$ and $\Omega^{\mathrm{odd}}(M) := \bigoplus_{k=0}^{2n} \Omega^{2k-1}(M)$. We also set

$$\underline{\Omega}(M) := \Omega^{\mathrm{even}}(M) \oplus \Omega^{\mathrm{odd}}(M) \quad \text{and} \quad \overline{\Omega}(M) := \Omega^{\mathrm{odd}}(M) \oplus \Omega^{\mathrm{even}}(M). \tag{2-4}$$

Using Lemmas 2.2 and 2.3, we obtain a skew-adjoint operator as follows.

Proposition 2.4. *Abbreviate $\hat{c}(\mathrm{dvol})$ as $\hat{c}_v$. The operator*

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{c}_v & \\ & \hat{c}_v \end{bmatrix} \begin{bmatrix} d+d^* & \omega \\ \omega^* & -d-d^* \end{bmatrix} \tag{2-5}$$

on $\underline{\Omega}(M)$ is skew-adjoint.

Proof. Using Lemma 2.2 and [21, Definition 6.1(2)], for all $\alpha \in \Omega^k(M)$, we have

$$\begin{aligned} \hat{c}_v(d+d^*)\alpha &= (-1)^{\frac{1}{2}(k+1)(k+2)} \star d\alpha + (-1)^{\frac{1}{2}(k-1)k} \star (-1) \star d \star \alpha \\ &= (-1)^{\frac{1}{2}(k+1)(k+2)} \star d\alpha + (-1)^{\frac{1}{2}(k-1)k+1} (-1)^{(4n-k+1)(4n-4n+k-1)} d \star \alpha \\ &= (-1)^{\frac{1}{2}(k+1)(k+2)} \star d\alpha + (-1)^{\frac{1}{2}k(k+1)} d \star \alpha. \end{aligned}$$

Similarly,

$$\begin{aligned} (d+d^*)\hat{c}_v\alpha &= d(-1)^{\frac{1}{2}k(k+1)} \star \alpha + d^*(-1)^{\frac{1}{2}k(k+1)} \star \alpha \\ &= (-1)^{\frac{1}{2}k(k+1)} d \star \alpha + (-1) \star d \star (-1)^{\frac{1}{2}k(k+1)} \star \alpha \\ &= (-1)^{\frac{1}{2}k(k+1)} d \star \alpha + (-1)^{\frac{1}{2}(k+1)(k+2)} \star d\alpha. \end{aligned}$$

Thus, we have

$$(d+d^*)\hat{c}_v = \hat{c}_v(d+d^*).$$

Now, since $\hat{c}_v^* = \hat{c}(e_{4n}) \cdots \hat{c}(e_1) = \hat{c}(e_1) \cdots \hat{c}(e_{4n}) = \hat{c}_v$, we find

$$\begin{aligned} \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{c}_v & \\ & \hat{c}_v \end{bmatrix} \begin{bmatrix} d+d^* & \omega \\ \omega^* & -d-d^* \end{bmatrix} \right)^* &= \begin{bmatrix} \hat{c}_v\omega^* & -\hat{c}_v(d+d^*) \\ \hat{c}_v(d+d^*) & \hat{c}_v\omega \end{bmatrix}^* \\ &= \begin{bmatrix} \omega\hat{c}_v & (d+d^*)\hat{c}_v \\ -(d+d^*)\hat{c}_v & \omega^*\hat{c}_v \end{bmatrix} \\ &= -\begin{bmatrix} \hat{c}_v\omega^* & -\hat{c}_v(d+d^*) \\ \hat{c}_v(d+d^*) & \hat{c}_v\omega \end{bmatrix} \\ &= -\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{c}_v & \\ & \hat{c}_v \end{bmatrix} \begin{bmatrix} d+d^* & \omega \\ \omega^* & -d-d^* \end{bmatrix}, \end{aligned}$$

where the second-to-last equality is justified by Lemma 2.3. Thus, the operator (2-5) is skew-adjoint. □

We recall the definition of the Atiyah–Singer mod 2 index. In [2, Theorem A], this invariant was defined for real Fredholm skew-adjoint operators. However, by functional calculus [10, Definition 1.13], one derives a version for real elliptic skew-adjoint operators:

Definition 2.5 [24, (7.5)]. Given a real elliptic skew-adjoint operator D, its Atiyah–Singer mod 2 index is the $\mathbb{Z}_2$-valued number

$$\mathrm{ind}_2 D := \dim\ker D \mod 2.$$

According to the definition of $\kappa(M, \omega)$ and the identification between kernels and cohomology groups, we have:

Corollary 2.6. *Keep the notation $\hat{c}_v := \hat{c}(\mathrm{dvol})$. The Atiyah–Singer mod 2 index of*

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \hat{c}_v & \\ & \hat{c}_v \end{bmatrix} \begin{bmatrix} d+d^* & \omega \\ \omega^* & -d-d^* \end{bmatrix} \tag{2-6}$$

on $\underline{\Omega}(M)$ is equal to $\kappa(M, \omega)$.

Proof. This is verified using Definition 1.3 and Propositions 2.1 and 2.4. With $[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}]$, the operator (2-6) preserves the parity of the grading, mapping $\underline{\Omega}(M)$ into $\underline{\Omega}(M)$ and $\overline{\Omega}(M)$ into $\overline{\Omega}(M)$. When we restrict (2-6) to $\underline{\Omega}(M)$, its kernel counts half of the b_k^{ω}. □

Similar to the Fredholm index [8, Theorem 3.11], and as stated in [2, Proposition 5.1] and [3, Section 2], the mod 2 index of a real skew-adjoint elliptic operator on a compact manifold is homotopy invariant:

Proposition 2.7. *Given a real skew-adjoint elliptic operator D on M, the index $\mathrm{ind}_2 D$ is invariant under a continuous deformation of D.*

Remark 2.8. The invariance of ind_2 in [2, Proposition 5.1] is for bounded real skew-adjoint Fredholm operators. It is extended to the elliptic operators on M as follows [1, Section 4]: for D as in Proposition 2.7, the operator $1 + (-D^2)$ is self-adjoint and positive. We have the compact operator

$$(1 + (-D^2))^{-1/2}$$

defined by functional calculus [10, Definition 1.13]. Then,

$$D \circ (1 + (-D^2))^{-1/2}$$

is a bounded real skew-adjoint Fredholm operator.

The next proposition gives us the skew-adjoint operator similar to [23, (1.1)].

Proposition 2.9. *The Atiyah–Singer mod 2 index of the skew-adjoint operator*

$$\begin{bmatrix} \frac{1}{2}(\omega^* - \omega) & -d-d^* \\ d+d^* & \frac{1}{2}(\omega - \omega^*) \end{bmatrix} \tag{2-7}$$

on $\underline{\Omega}(M)$ is equal to $\kappa(M, \omega)$.

Proof. We keep the notation $\hat{c}_v := \hat{c}(\mathrm{dvol})$. By Lemma 2.3, the operator

$$\frac{1}{2} \begin{bmatrix} \hat{c}_v(\omega^* + \omega) & \\ & \hat{c}_v(\omega + \omega^*) \end{bmatrix}$$

is skew-adjoint on $\underline{\Omega}(M)$. Then, by Corollary 2.6, we find

$$\begin{aligned}\kappa(M,\omega) &= \mathrm{ind}_2\left(\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}\hat{c}_v&\\&\hat{c}_v\end{bmatrix}\begin{bmatrix}d+d^*&\omega\\\omega^*&-d-d^*\end{bmatrix}\right)\\ &= \mathrm{ind}_2\begin{bmatrix}\hat{c}_v\omega^*&-\hat{c}_v(d+d^*)\\\hat{c}_v(d+d^*)&\hat{c}_v\omega\end{bmatrix}\\ &= \mathrm{ind}_2\left(\begin{bmatrix}\hat{c}_v\frac{1}{2}(\omega^*-\omega)&-\hat{c}_v(d+d^*)\\\hat{c}_v(d+d^*)&\hat{c}_v\frac{1}{2}(\omega-\omega^*)\end{bmatrix}+\frac{1}{2}\begin{bmatrix}\hat{c}_v(\omega^*+\omega)&\\&\hat{c}_v(\omega+\omega^*)\end{bmatrix}\right)\\ &= \mathrm{ind}_2\begin{bmatrix}\hat{c}_v\frac{1}{2}(\omega^*-\omega)&-\hat{c}_v(d+d^*)\\\hat{c}_v(d+d^*)&\hat{c}_v\frac{1}{2}(\omega-\omega^*)\end{bmatrix} = \mathrm{ind}_2\begin{bmatrix}\frac{1}{2}(\omega^*-\omega)&-d-d^*\\d+d^*&\frac{1}{2}(\omega-\omega^*)\end{bmatrix}.\end{aligned}$$

The second-to-last equality is justified by Proposition 2.7. □

3. Symplectic Witten deformation

In this section, we study the symplectic Witten deformation of the skew-adjoint operator (2-7) on $C^{\text{even}}(M,\omega) := \underline{\Omega}(M)$.

We let V be a nondegenerate smooth vector field on M. This means around each zero point p of V, given any small local chart with coordinates $x_1, y_1, \dots, x_{2n}, y_{2n}$ satisfying $x_1(p) = \cdots = y_{2n}(p) = 0$, there is an $\mathbb{R}^{4n}$-valued smooth function B on the chart with order

$$O(x_1^2+y_1^2+\cdots+x_{2n}^2+y_{2n}^2)$$

and a matrix $A \in \mathrm{GL}_{4n}(\mathbb{R})$ such that

$$V(x_1,y_1,\dots,x_{2n},y_{2n}) = \left[\partial_{x_1},\partial_{y_1},\dots,\partial_{x_{2n}},\partial_{y_{2n}}\right]\left(A\begin{bmatrix}x_1\\y_1\\\vdots\\x_{2n}\\y_{2n}\end{bmatrix}+B\right). \tag{3-1}$$

Here, ∂_{x_i} and ∂_{y_i} are the local coordinate vector fields. For convenience, we let

$$\boldsymbol{x} = \begin{bmatrix}x_1\\y_1\\\vdots\\x_{2n}\\y_{2n}\end{bmatrix}, \quad \boldsymbol{x}^{\mathsf{T}} = [x_1,y_1,\dots,x_{2n},y_{2n}],$$

$$|\boldsymbol{x}| = \sqrt{x_1^2+y_1^2+\cdots+x_{2n}^2+y_{2n}^2}.$$

Alternatively, if we write $\boldsymbol{x} = [\varphi_1,\dots,\varphi_{4n}]^{\mathsf{T}}$, we have $|\boldsymbol{x}| = \sqrt{\varphi_1^2+\cdots+\varphi_{4n}^2}$.

Lemma 3.1. *There is a smooth vector field X on M such that the zero set of X is the same as the zero set of V, and*

$$X = \left[\partial_{x_1},\partial_{y_1},\dots,\partial_{x_{2n}},\partial_{y_{2n}}\right]A\boldsymbol{x}$$

near each zero point p.

Proof. We will use a cutoff function to modify the vector field V near its zeros so that it agrees with a standard local model. We find a constant $C > 0$ such that

$$|B| \leq C\,|\boldsymbol{x}|^2 \tag{3-2}$$

on the local chart centered at p. Viewing A as an operator on the linear space $\mathbb{R}^{4n}$, we let $\|A\|$ be its operator norm. We choose a bump function σ such that

$$\operatorname{supp}(\sigma) \subseteq \left\{(x_1, \dots, y_{2n}) : (x_1^2 + \cdots + y_{2n}^2)^{1/2} < \|A^{-1}\|^{-1} \cdot C^{-1}\right\} \tag{3-3}$$

and $\sigma = 1$ near p. Now, we show that

$$\begin{aligned} X &= \sigma V + (1-\sigma)\left[\partial_{x_1}, \partial_{y_1}, \dots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{x} \\ &= \left[\partial_{x_1}, \partial_{y_1}, \dots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] (A\boldsymbol{x} + (1-\sigma)B) \end{aligned}$$

is the vector field we need. Indeed, by (3-2) and (3-3),

$$|\boldsymbol{x} + (1-\sigma)A^{-1}B| \geq |\boldsymbol{x}| - \|A^{-1}\| \cdot C \cdot |\boldsymbol{x}|^2 \geq 0.$$

The last "$\geq$" becomes "$=$" if and only if $\boldsymbol{x}$ is zero. Thus,

$$A\boldsymbol{x} + (1-\sigma)B = A \cdot (\boldsymbol{x} + (1-\sigma)A^{-1}B) = 0$$

if and only if we are at a zero point p of V. Therefore, the zero set of X coincides that of V. □

Inspired by [22, Section 2; 6, Chapters VIII–X; 24, Section 7.3; 23, Section 2], for a parameter $T > 0$, we use the vector field X to set up the Witten deformation

$$\mathbb{D}_T := \begin{bmatrix} \frac{1}{2}(\omega^* - \omega) & -d - d^* - T\hat{c}(X) \\ d + d^* + T\hat{c}(X) & \frac{1}{2}(\omega - \omega^*) \end{bmatrix} \tag{3-4}$$

of the operator (2-7) on $\underline{\Omega}(M)$. Let $\varepsilon > 0$ be a sufficiently small number. Around each zero point p of X, we choose a chart

$$U = \left\{(x_1, \dots, y_{2n}) : x_1^2 + y_1^2 + \cdots + x_{2n}^2 + y_{2n}^2 < (4\varepsilon)^2\right\} \tag{3-5}$$

centered at p and satisfying

(1) $\omega|_U = dx_1 \wedge dy_1 + \cdots + dx_{2n} \wedge dy_{2n}$,

(2) $g(\cdot\,,\cdot)|_U = dx_1^2 + dy_1^2 + \cdots + dx_{2n}^2 + dy_{2n}^2$, where g is the metric, and

(3) $X|_U = \left[\partial_{x_1}, \partial_{y_1}, \dots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{x}$.

We can obtain (1)–(3) as follows: Using [15, Theorem 8.1], we first choose a Darboux chart U centered at the zero point p with coordinates $x_1, y_1, \dots, x_{2n}, y_{2n}$ such that

$$\omega|_U = dx_1 \wedge dy_1 + \cdots + dx_{2n} \wedge dy_{2n}.$$

We then construct a metric g' on M such that

$$g'(\,\cdot\,,\cdot\,)|_U = dx_1^2 + dy_1^2 + \cdots + dx_{2n}^2 + dy_{2n}^2.$$

Next, following the proof of [15, Proposition 12.3], we use the polar decomposition together with g' to construct the almost complex structure J. Then, we let $g(\,\cdot\,,\cdot\,) = \omega(\,\cdot\,, J\cdot\,)$. The two metrics $g(\,\cdot\,,\cdot\,)$ and $g'(\,\cdot\,,\cdot\,)$ are different, but checking the polar decomposition, we have

$$g(\,\cdot\,,\cdot\,)|_U = g'(\,\cdot\,,\cdot\,)|_U = dx_1^2 + dy_1^2 + \cdots + dx_{2n}^2 + dy_{2n}^2.$$

Finally, the vector field X is guaranteed by Lemma 3.1.

Let

$$\boldsymbol{e}_i = \big[\underbrace{0,\ldots,0}_{2i-2}, 1, 0, 0, \ldots, 0\big]^{\mathsf{T}} \quad\text{and}\quad \boldsymbol{f}_i = \big[\underbrace{0,\ldots,0,0}_{2i-1}, 1, 0, \ldots, 0\big]^{\mathsf{T}}$$

Inside U, we find that

$$\begin{aligned}(d+d^*+T\hat{c}(X))^2 = -\sum_{i=1}^{2n} \partial_{x_i}^2 - \sum_{i=1}^{2n} \partial_{y_i}^2 + T\sum_{i=1}^{2n} c(\partial_{x_i})\hat{c}\left(\left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{e}_i\right)\\ + T\sum_{i=1}^{2n} c(\partial_{y_i})\hat{c}\left(\left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{f}_i\right) + T^2\boldsymbol{x}^{\mathsf{T}}A^*A\boldsymbol{x}.\end{aligned}$$

Now, on $\mathbb{R}^{4n}$ with coordinates denoted by $x_1, y_1, \ldots, x_{2n}, y_{2n}$, we let

$$X_0 = \left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{x}.$$

Meanwhile, using the standard Euclidean metric

$$g_0 := dx_1^2 + dy_1^2 + \cdots + dx_{2n}^2 + dy_{2n}^2$$

on $\mathbb{R}^{4n}$, we have the L^2-norm (and inner product)

$$\|\alpha\| = \left(\int_{\mathbb{R}^{4n}} g_0(\alpha,\alpha)\, dx_1\wedge dy_1\wedge\cdots\wedge dx_{2n}\wedge dy_{2n}\right)^{1/2} \tag{3-6}$$

on the space $\Omega^k(\mathbb{R}^{4n})$ of smooth k-forms on $\mathbb{R}^{4n}$. For the standard symplectic form

$$\omega_0 = dx_1\wedge dy_1 + \cdots + dx_{2n}\wedge dy_{2n}$$

on $\mathbb{R}^{4n}$, we let

$$\omega_0^*\lrcorner = \partial_{y_1}\lrcorner\,\partial_{x_1}\lrcorner + \cdots + \partial_{y_{2n}}\lrcorner\,\partial_{x_{2n}}\lrcorner$$

be the adjoint of $\omega_0\wedge$.

Let L be the operator with the expression

$$-\sum_{i=1}^{2n} \partial_{x_i}^2 - \sum_{i=1}^{2n} \partial_{y_i}^2 + T \sum_{i=1}^{2n} c(\partial_{x_i})\hat{c}\left(\left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{e}_i\right) + T \sum_{i=1}^{2n} c(\partial_{y_i})\hat{c}\left(\left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{f}_i\right) + T^2 \boldsymbol{x}^{\mathsf{T}} A^* A\boldsymbol{x}$$

but defined on the space $\bigoplus_{k=0}^{4n} \Omega^k(\mathbb{R}^{4n})$ of smooth forms on $\mathbb{R}^{4n}$. As in [24, (4.23)], we let

$$L' = -\sum_{i=1}^{2n} \partial_{x_i}^2 - \sum_{i=1}^{2n} \partial_{y_i}^2 - T \cdot \mathrm{trace}(\sqrt{A^*A}\,) + T^2 \boldsymbol{x}^{\mathsf{T}} A^* A\boldsymbol{x}$$

and

$$L'' = \mathrm{trace}(\sqrt{A^*A}\,) + \sum_{i=1}^{2n} c(\partial_{x_i})\hat{c}\left(\left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{e}_i\right) + \sum_{i=1}^{2n} c(\partial_{y_i})\hat{c}\left(\left[\partial_{x_1}, \partial_{y_1}, \ldots, \partial_{x_{2n}}, \partial_{y_{2n}}\right] A\boldsymbol{f}_i\right).$$

Then, $L = L' + T \cdot L''$. Indeed, L' is the (rescaled) harmonic oscillator [17, Chapter 8, Section 6] on the space of square-integrable functions on $\mathbb{R}^{4n}$, and L'' is a nonnegative operator on the (real) vector space spanned by

$$\{dx_{i_1} \wedge \cdots \wedge dx_{i_r} \wedge dy_{j_1} \wedge \cdots \wedge dy_{j_s} : 0 \le i_1 < \cdots < i_r \le 2n,\ 0 \le j_1 < \cdots < j_s \le 2n,\ 0 \le r, s \le 2n\}. \tag{3-7}$$

The following result was proved using the properties of harmonic oscillators [loc. cit.] and equations (4.23)–(4.25) and Lemma 4.8 of [24].

Proposition 3.2 [24, Proposition 4.9]. *For any $T > 0$, the kernel of L is one-dimensional and generated by*

$$\rho = \exp\left(-\tfrac{1}{2} T \boldsymbol{x}^{\mathsf{T}} \sqrt{A^*A}\boldsymbol{x}\right) \cdot \delta, \tag{3-8}$$

where δ is a certain linear combination (with real coefficients independent of T) of elements of (3-7). *The grading of δ is even if* $\det A > 0$ *and odd if* $\det A < 0$. *Each nonzero eigenvalue of L has the expression $\alpha \cdot T$, where α is a positive constant independent of T.*

Noticing that $\omega_0^* \lrcorner - \omega_0 \wedge$ is skew-symmetric, we have:

Proposition 3.3. *There exists a unique smooth form η on $\mathbb{R}^{4n}$ such that*

$$\tfrac{1}{2}(\omega_0^* - \omega_0)\rho = \left(d + d^* + T\hat{c}(X_0)\right)\eta$$

and $\eta \perp \rho$. Here, $d + d^*$ is defined on $\bigoplus_{k=0}^{4n} \Omega^k(\mathbb{R}^{4n})$according to the L^2-norm of forms.

Proof. Recall the definition (3-8) of ρ. We notice that

$$g_0\left((\omega_0^* - \omega_0)\rho, \rho\right) = 0.$$

Therefore, $\frac{1}{2}(\omega_0^* - \omega_0)\rho$ is orthogonal to the kernel of L on $\bigoplus_{k=0}^{4n} \Omega^k(\mathbb{R}^{4n})$. Since $d + d^* + T\hat{c}(X_0)$ preserves the eigenspaces of L, we find

$$\eta = L^{-1} \circ \left(d + d^* + T\hat{c}(X_0)\right)\left(\tfrac{1}{2}(\omega_0^* - \omega_0)\rho\right). \tag{3-9}$$

Here, L^{-1} is the inverse of L restricted to the orthogonal complement of $\ker(L)$. The kernel of L is given by Proposition 3.2. See [6, (10.17)] for details about the inverse map L^{-1}. □

The next proposition will be used in the estimates of the spectrum of $-\mathbb{D}_T^2$.

Proposition 3.4. *There is a constant $C_1 \geq 0$ independent of T such that*

$$\|\eta\| = C_1 T^{-1/2} \cdot \|\rho\|. \tag{3-10}$$

Here, *the L^2-norm is that on the space of forms on $\mathbb{R}^{4n}$.*

Proof. If $\eta = 0$, we choose $C_1 = 0$. If $\eta \neq 0$, by Proposition 3.3, $\frac{1}{2}(\omega_0^* - \omega_0)\rho$ is nonzero, and we look at (3-9). We write $\frac{1}{2}(\omega_0^* - \omega_0)\rho$ as a finite sum of eigenvectors of L:

$$\tfrac{1}{2}(\omega_0^* - \omega_0)\rho = \sum_i K_i \cdot \exp\left(-\tfrac{1}{2}T\boldsymbol{x}^{\mathsf{T}}\sqrt{A^*A}\boldsymbol{x}\right)\cdot \delta_i,$$

where each K_i is a constant independent of T and each δ_i is an eigenvector of L'' in the span of (3-7), associated with an eigenvalue $\lambda_i > 0$. These δ_i and λ_i satisfy

$$g_0(\delta_i, \delta_j) = 0 \quad \text{and} \quad \lambda_i \neq \lambda_j \quad \text{when } i \neq j.$$

Then, we apply $L^{-1} \circ (d + d^* + T\hat{c}(X_0))$ to $\frac{1}{2}(\omega_0^* - \omega_0)\rho$. Since $d + d^* + T\hat{c}(X_0)$ preserves the eigenspaces of L, we obtain

$$\begin{aligned} &L^{-1} \circ (d + d^* + T\hat{c}(X_0))\left(\tfrac{1}{2}(\omega_0^* - \omega_0)\rho\right) \\ &\qquad = \sum_i \frac{1}{\lambda_i T} \cdot (d + d^* + T\hat{c}(X_0))\left(K_i \cdot \exp\left(-\tfrac{1}{2}T\boldsymbol{x}^{\mathsf{T}}\sqrt{A^*A}\boldsymbol{x}\right)\cdot \delta_i\right). \end{aligned}$$

One step further, considering the effect of $d + d^* + T\hat{c}(X_0)$, we find

$$\begin{aligned} \eta &= L^{-1} \circ (d + d^* + T\hat{c}(X_0))\left(\tfrac{1}{2}(\omega_0^* - \omega_0)\rho\right) \\ &= \sum_i \frac{1}{\lambda_i T} \cdot T \cdot \exp\left(-\tfrac{1}{2}T\boldsymbol{x}^{\mathsf{T}}\sqrt{A^*A}\boldsymbol{x}\right)\cdot\left(\sum_{j=1}^{2n} K_{ij}\cdot x_j \cdot \delta_{ij} + \sum_{j=1}^{2n} \tilde{K}_{ij}\cdot y_j \cdot \tilde{\delta}_{ij}\right) \\ &= \sum_i \frac{1}{\lambda_i} \cdot \exp\left(-\tfrac{1}{2}T\boldsymbol{x}^{\mathsf{T}}\sqrt{A^*A}\boldsymbol{x}\right)\cdot\left(\sum_{j=1}^{2n} K_{ij}\cdot x_j \cdot \delta_{ij} + \sum_{j=1}^{2n} \tilde{K}_{ij}\cdot y_j \cdot \tilde{\delta}_{ij}\right), \end{aligned}$$

where the K_{ij} and $\tilde{K}_{ij}$ are constants independent of T and the δ_{ij} and $\tilde{\delta}_{ij}$ are certain linear combinations (with real coefficients independent of T) of elements of (3-7). Thus, (3-10) is essentially the relation between

$$\left(\int_{\mathbb{R}^{4n}} x_i^2 \exp\left(-T|\boldsymbol{x}|^2\right) dx_1\, dy_1 \cdots dx_{2n}\, dy_{2n}\right)^{1/2} = \frac{\pi^n}{T^n} \cdot \frac{1}{\sqrt{2T}}$$

and

$$\left(\int_{\mathbb{R}^{4n}} \exp\left(-T|\boldsymbol{x}|^2\right) dx_1\, dy_1 \cdots dx_{2n}\, dy_{2n}\right)^{1/2} = \frac{\pi^n}{T^n}.$$

Their ratio gives us the factor $T^{-1/2}$. □

Remark 3.5. In the standard Witten deformation, the form ρ functions as a model for eigenforms associated with small eigenvalues of the deformed Laplacian. In this paper, the pair (ρ, η) plays a similar role in the mapping cone Witten deformation.

Now, using (3-6), we define the L^2-norm (and inner product)

$$\left\| \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right\| = \left(\|\alpha\|^2 + \|\beta\|^2\right)^{1/2} \tag{3-11}$$

on $\underline{\Omega}(\mathbb{R}^{4n})$. (See (2-4) and recall the matrix A in (3-1) associated with the zero point p.) When $\det A > 0$, we study the orthogonal complement of $\operatorname{span}_{\mathbb{R}}\left(\begin{bmatrix} \rho \\ \eta \end{bmatrix}\right)$ in $\underline{\Omega}(\mathbb{R}^{4n})$ under the inner product induced by (3-11). Let $\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \in \underline{\Omega}(\mathbb{R}^{4n})$ be an L^2-element orthogonal to $\begin{bmatrix} \rho \\ \eta \end{bmatrix}$. We write

$$\alpha = r\rho + \alpha' \quad \text{and} \quad \beta = s\eta + \beta',$$

with $\alpha' \perp \rho$ and $\beta' \perp \eta$. Then, we have

$$r\|\rho\|^2 + s\|\eta\|^2 = 0. \tag{3-12}$$

Let $\|\cdot\|_1$ be the first Sobolev norm (see [14, Definition 10.2.7]) induced by (3-6). If $\|\alpha\|_1 < \infty$ and $\|\beta\|_1 < \infty$, we find that $\|\alpha'\|_1 < \infty$, $\|\beta'\|_1 < \infty$, and then

$$\begin{aligned} &\left\| \begin{bmatrix} \frac{1}{2}(\omega_0^* - \omega_0) & -d - d^* - T\hat{c}(X_0) \\ d + d^* + T\hat{c}(X_0) & \frac{1}{2}(\omega_0 - \omega_0^*) \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right\| \quad (3\text{-}13) \\ &\geq \left\| \begin{bmatrix} & -d - d^* - T\hat{c}(X_0) \\ d + d^* + T\hat{c}(X_0) & \end{bmatrix} \begin{bmatrix} r\rho + \alpha' \\ s\eta + \beta' \end{bmatrix} \right\| \\ &\qquad - \left\| \frac{1}{2} \begin{bmatrix} \omega_0^* - \omega_0 & \\ & \omega_0 - \omega_0^* \end{bmatrix} \begin{bmatrix} r\rho + \alpha' \\ s\eta + \beta' \end{bmatrix} \right\| \\ &\geq \left\| \begin{bmatrix} (-d - d^* - T\hat{c}(X_0))(s\eta + \beta') \\ (d + d^* + T\hat{c}(X_0))(r\rho + \alpha') \end{bmatrix} \right\| - C_2 \left\| \begin{bmatrix} r\rho + \alpha' \\ s\eta + \beta' \end{bmatrix} \right\| \\ &= \left(\|(d + d^* + T\hat{c}(X_0))(s\eta + \beta')\|^2 + \|(d + d^* + T\hat{c}(X_0))\alpha'\|^2\right)^{1/2} - C_2 \left\| \begin{bmatrix} r\rho + \alpha' \\ s\eta + \beta' \end{bmatrix} \right\|, \end{aligned}$$

since $(d+d^*+T\hat{c}(X_0))\rho=0$. Using $\|\alpha'\|_1<\infty$, $\|\beta'\|_1<\infty$, and Proposition 3.2, we see that the right-hand side of (3-13) is

$$\begin{aligned}&\geq C_3\sqrt{T}\|s\eta+\beta'\|+C_3\sqrt{T}\|\alpha'\|-C_2\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|\\ &=C_3\sqrt{T}\sqrt{\|s\eta\|^2+\|\beta'\|^2}+C_3\sqrt{T}\|\alpha'\|-C_2\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|\\ &\geq \mathfrak{X},\end{aligned}$$

where

$$\mathfrak{X}:=C_4\sqrt{T}\|s\eta\|+C_4\sqrt{T}\|\beta'\|+C_3\sqrt{T}\|\alpha'\|-C_2\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|. \tag{3-14}$$

We estimate $\mathfrak{X}$ in two complementary cases. If $\eta=0$, then by (3-12), we find $r=0$ and then

$$\mathfrak{X}=C_4\sqrt{T}\|\beta'\|+C_3\sqrt{T}\|\alpha'\|-C_2\left\|\begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}\right\|\geq C_5\sqrt{T}\left\|\begin{bmatrix}\alpha'\\ \beta'\end{bmatrix}\right\|=C_5\sqrt{T}\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|.$$

If $\eta\neq 0$, then

$$\begin{aligned}\mathfrak{X}&=\frac{1}{2}C_4\sqrt{T}\|s\eta\|+\frac{1}{2}C_4\sqrt{T}\frac{|r|\cdot\|\rho\|^2}{\|\eta\|^2}\|\eta\|+C_4\sqrt{T}\|\beta'\|+C_3\sqrt{T}\|\alpha'\|-C_2\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|\\ &=\frac{1}{2}C_4\sqrt{T}\|s\eta\|+\frac{1}{2}C_4\sqrt{T}\|r\rho\|C_1^{-1}\sqrt{T}+C_4\sqrt{T}\|\beta'\|+C_3\Big[\sqrt{T}\|\alpha'\|-C_2\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|\\ &\geq C_5\sqrt{T}\left\|\begin{bmatrix}r\rho+\alpha'\\ s\eta+\beta'\end{bmatrix}\right\|.\end{aligned}$$

This takes care of the case $\det A>0$. When $\det A<0$, we replace $\left[\begin{smallmatrix}\rho\\ \eta\end{smallmatrix}\right]$ by $\left[\begin{smallmatrix}\eta\\ \rho\end{smallmatrix}\right]$ and repeat the argument. Then, we summarize:

Proposition 3.6. *There exists a constant* $C_5>0$ *such that when* $\det A>0$ (*resp. when* $\det A<0$), *for all sufficiently large* T, *we have*

$$\left\|\begin{bmatrix}\frac{1}{2}(\omega_0^*-\omega_0) & -d-d^*-T\hat{c}(X_0)\\ d+d^*+T\hat{c}(X_0) & \frac{1}{2}(\omega_0-\omega_0^*)\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\|\geq C_5\sqrt{T}\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\|$$

whenever $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]\in\underline{\Omega}(\mathbb{R}^{4n})$ *is orthogonal to* $\left[\begin{smallmatrix}\rho\\ \eta\end{smallmatrix}\right]$ $\big($*resp.* $\left[\begin{smallmatrix}\eta\\ \rho\end{smallmatrix}\right]\big)$ *and satisfies* $\|\alpha\|_1<\infty$ *and* $\|\beta\|_1<\infty$.

Following [6] and [24], based on Propositions 3.2–3.6, we apply the asymptotic analysis to carry out the estimates about $\mathbb{D}_T$. Recall from (3-5) the chart U around each zero point p of X. For each zero point p, we pick a bump function $\gamma: M\to\mathbb{R}$ such that

$$\operatorname{supp}(\gamma)\subseteq U(2\varepsilon):=\big\{(x_1,\ldots,y_{2n}):x_1^2+\cdots+y_{2n}^2<(2\varepsilon)^2\big\},$$

and $\gamma = 1$ on

$$U(\varepsilon) := \left\{(x_1, \dots, y_{2n}) : x_1^2 + \cdots + y_{2n}^2 < \varepsilon^2\right\}.$$

For each zero point p, we let

$$\rho_p = \gamma \cdot \exp\left(-\tfrac{1}{2} T \boldsymbol{x}^\top \sqrt{A^* A} \boldsymbol{x}\right) \cdot \delta$$

and $\eta_p = \gamma \cdot \eta$. As in [6, Definition 9.4] and [24, (4.36)], we let

$$\begin{aligned} E_{T,0} &:= \operatorname{span}_{\mathbb{R}}\left\{\left[\begin{smallmatrix}\rho_p \\ \eta_p\end{smallmatrix}\right] : p \text{ is a zero point of } X\right\}, \\ E_{T,1} &:= \operatorname{span}_{\mathbb{R}}\left\{\left[\begin{smallmatrix}\eta_p \\ \rho_p\end{smallmatrix}\right] : p \text{ is a zero point of } X\right\}, \\ E_T &:= E_{T,0} \oplus E_{T,1}. \end{aligned}$$

Let $E_T^\perp$ be the orthogonal complement of E_T in $\underline{\Omega}(M)$ and p_T (resp. $p_T^\perp$) be the orthogonal projection from $\underline{\Omega}(M)$ to E_T (resp. $E_T^\perp$).

Recall the operator

$$\mathbb{D}_T := \begin{bmatrix} \frac{1}{2}(\omega^* - \omega) & -d - d^* - T\hat{c}(X) \\ d + d^* + T\hat{c}(X) & \frac{1}{2}(\omega - \omega^*) \end{bmatrix}$$

on $\underline{\Omega}(M)$. There is a constant $C_6 > 0$ such that

$$\begin{aligned} \left\| \mathbb{D}_T \begin{bmatrix} \rho_p \\ \eta_p \end{bmatrix} \right\| &= \left\| \begin{bmatrix} \frac{1}{2}(\omega^* - \omega) & -d - d^* - T\hat{c}(X) \\ d + d^* + T\hat{c}(X) & \frac{1}{2}(\omega - \omega^*) \end{bmatrix} \begin{bmatrix} \gamma\rho \\ \gamma\eta \end{bmatrix} \right\| \\ &= \left\| \begin{bmatrix} -c(d\gamma)\eta \\ c(d\gamma)\rho + \frac{1}{2}(\omega - \omega^*)\gamma\eta \end{bmatrix} \right\| \\ &\le \left\| \begin{bmatrix} c(d\gamma)\eta \\ c(d\gamma)\rho \end{bmatrix} \right\| + \left\| \begin{bmatrix} 0 \\ \frac{1}{2}(\omega - \omega^*)\gamma\eta \end{bmatrix} \right\| \le C_6 \left\| \begin{bmatrix} \rho_p \\ \eta_p \end{bmatrix} \right\| \end{aligned}$$

when T is sufficiently large. Summarizing this estimate, we get:

Proposition 3.7. *There is a constant $C_6 > 0$ such that, when T is sufficiently large,*

$$\left\| \mathbb{D}_T \left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \right\| \le C_6 \left\| \left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \right\|$$

for all $\left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \in E_T$.

Now, if $\left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \in E_T^\perp$, we have the following estimate similar to those in Theorem 9.11 of [6] and Proposition 4.12 of [24]:

Proposition 3.8. *There exists a constant $C_7 > 0$ such that when T is sufficiently large,*

$$\left\| \mathbb{D}_T \left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \right\| \ge C_7 \sqrt{T} \left\| \left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \right\|$$

for all $\left[\begin{smallmatrix}\alpha \\ \beta\end{smallmatrix}\right] \in E_T^\perp$.

Proof. We perform three steps:

Step 1: If $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ is supported outside all the $U(2\varepsilon)$'s, the minimum of $g(X, X)$ is greater than 0. Then, as in [24, Proposition 4.7], we find

$$\begin{aligned}\left\|\mathbb{D}_T\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| &\geq \left\|\begin{bmatrix} & -d-d^*-T\hat{c}(X)\\ d+d^*+T\hat{c}(X) & \end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| - \left\|\begin{bmatrix}\frac{1}{2}(\omega^*-\omega) & \\ & \frac{1}{2}(\omega-\omega^*)\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| \\ &\geq C_8 T\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| - C_9\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\|.\end{aligned}$$

Step 2: If $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ is supported inside the chart U centered at some zero point p, we view $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]$ as an element in $\underline{\Omega}(\mathbb{R}^{4n})$. Let p'_T be the orthogonal projection from $\underline{\Omega}(\mathbb{R}^{4n})$ to the one-dimensional space generated by $\left[\begin{smallmatrix}\rho\\ \eta\end{smallmatrix}\right]$. Letting $\langle\cdot,\cdot\rangle$ denote the inner product induced by (3-11), we have

$$\begin{aligned}p'_T\begin{bmatrix}\alpha\\ \beta\end{bmatrix} &= \frac{1}{\sqrt{\|\rho\|^2+\|\eta\|^2}}\left\langle\begin{bmatrix}\rho\\ \eta\end{bmatrix},\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\rangle\cdot\left(\frac{1}{\sqrt{\|\rho\|^2+\|\eta\|^2}}\begin{bmatrix}\rho\\ \eta\end{bmatrix}\right) \\ &= \frac{1}{\|\rho\|^2+\|\eta\|^2}\left\langle\begin{bmatrix}\rho\\ \eta\end{bmatrix},\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\rangle\cdot\begin{bmatrix}\rho\\ \eta\end{bmatrix} \\ &= \frac{1}{\|\rho\|^2+\|\eta\|^2}\int_M(1-\gamma)\cdot g\left(\begin{bmatrix}\rho\\ \eta\end{bmatrix},\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right)\mathrm{dvol}\cdot\begin{bmatrix}\rho\\ \eta\end{bmatrix},\end{aligned}$$

since $\left\langle\left[\begin{smallmatrix}\gamma\rho\\ \gamma\eta\end{smallmatrix}\right],\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]\right\rangle = 0$. Then, we find, by Cauchy–Schwarz and comparing $\|\rho\|$ with $\exp(-C_{10}\varepsilon^2 T)$,

$$\begin{aligned}\left\|p'_T\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| &= \frac{1}{\sqrt{\|\rho\|^2+\|\eta\|^2}}\left|\int_M(1-\gamma)\cdot g\left(\begin{bmatrix}\rho\\ \eta\end{bmatrix},\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right)\mathrm{dvol}\right| \\ &\leq \frac{1}{\sqrt{\|\rho\|^2+\|\eta\|^2}}\cdot\exp(-C_{10}\varepsilon^2 T)\int_{|x|\leq 4\varepsilon} g\left(\begin{bmatrix}\alpha\\ \beta\end{bmatrix},\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right)^{1/2}\mathrm{dvol} \\ &\leq \frac{\sqrt{T}}{\sqrt{T+C_1^2}}\cdot\|\rho\|^{-1}\cdot\exp(-C_{10}\varepsilon^2 T)\cdot\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| \\ &\leq \exp(-C_{11}T)\cdot\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\|.\end{aligned}$$

By Proposition 3.6, we find

$$\left\|\mathbb{D}_T\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| \geq C_5\sqrt{T}\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}-p'_T\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\| \geq C_5\sqrt{T}\cdot(1-\exp(-C_{11}T))\cdot\left\|\begin{bmatrix}\alpha\\ \beta\end{bmatrix}\right\|.$$

Step 3: For a general $\left[\begin{smallmatrix}\alpha\\ \beta\end{smallmatrix}\right]\in E_T^{\perp}$ supported on M, we combine what we have shown in Steps 1 and 2, following the standard procedure used in Step 3 of the proof of [24, Proposition 4.12]. □

Noticing that $\mathbb{D}_T$ is skew-adjoint, we have:

Proposition 3.9. *The operator* $-\mathbb{D}_T^2$ *is self-adjoint and nonnegative. When* T *is sufficiently large, the eigenvalues of* $-\mathbb{D}_T^2$ *lie in the union* $[0, C_6^2] \cup [C_7^2 T, +\infty)$.

Proof. This is a combination of Propositions 3.7 and 3.8, following the same pattern as in the proof of [25, Lemma 5.3]. Since there is no essential spectrum here, we only need a simplified procedure as in the proof of [26, Proposition 6.18]. □

4. The counting formula

We now prove the counting formula (1-4) stated in Theorem 1.5. Let $\tilde{E}_T$ be the sum of eigenspaces of $-\mathbb{D}_T^2$ on $\underline{\Omega}(M)$ associated with eigenvalues in $[0, C_6^2]$. Then,

$$\begin{aligned}\kappa(M, \omega) &= \operatorname{ind}_2\big(\mathbb{D}_T \text{ on } \underline{\Omega}(M)\big)\\ &= \dim\ker\big(-\mathbb{D}_T^2 \text{ on } \underline{\Omega}(M)\big) \bmod 2\\ &= \dim\ker\big(\mathbb{D}_T : \tilde{E}_T \to \tilde{E}_T\big) \bmod 2,\end{aligned}$$

since each eigenspace of $-\mathbb{D}_T^2$ is invariant under $\mathbb{D}_T$. By [9, Section 8.16], every $r \times r$ skew-symmetric matrix has Atiyah–Singer mod 2 index equal to the parity of r. Thus,

$$\kappa(M, \omega) = \dim \tilde{E}_T \bmod 2.$$

Now, to prove Theorem 1.5, we only need to show that $\dim E_T = \dim \tilde{E}_T$.

Proposition 4.1. *We have* $\dim E_T = \dim \tilde{E}_T$ *when* T *is sufficiently large.*

Proof. Recall the L^2-norm (2-2) on $\underline{\Omega}(M)$. We let

$$\tilde{P}_T : \underline{\Omega}(M) \to \tilde{E}_T$$

be the orthogonal projection to $\tilde{E}_T$. Then, for any $h \in E_T$, we obtain

$$\begin{aligned}\|h - \tilde{P}_T h\| &\le \frac{1}{C_7\sqrt{T}} \|\mathbb{D}_T(h - \tilde{P}_T h)\| && \text{(by Proposition 3.9)}\\ &\le \frac{1}{C_7\sqrt{T}} \big(\|\mathbb{D}_T h\| + \|\mathbb{D}_T \tilde{P}_T h\|\big)\\ &\le \frac{1}{C_7\sqrt{T}} \cdot C_6 \cdot (\|h\| + \|h\|) && \text{(by Proposition 3.9).}\end{aligned}$$

Thus, when T is large, $\tilde{P}_T$ maps E_T injectively into $\tilde{E}_T$, meaning that $\dim \tilde{E}_T \ge \dim E_T$.

We prove the opposite inequality by contradiction. As in [24, (5.32)], suppose that $\dim \tilde{E}_T > \dim E_T$. We pick some $\varphi \in \tilde{E}_T$ such that φ is orthogonal to the space $\tilde{P}_T E_T$. Let $\langle\cdot,\cdot\rangle$ denote the inner product induced by (2-2). For any zero point p of X and the associated $\left[\begin{smallmatrix}\rho_p\\ \eta_p\end{smallmatrix}\right]$ (or $\left[\begin{smallmatrix}\eta_p\\ \rho_p\end{smallmatrix}\right]$, depending on the sign of $\det A$), we have

$$\left\langle \varphi, \begin{bmatrix}\rho_p\\ \eta_p\end{bmatrix}\right\rangle = \left\langle \varphi, \begin{bmatrix}\rho_p\\ \eta_p\end{bmatrix}\right\rangle - \left\langle \varphi, \tilde{P}_T \begin{bmatrix}\rho_p\\ \eta_p\end{bmatrix}\right\rangle = \left\langle \varphi, \begin{bmatrix}\rho_p\\ \eta_p\end{bmatrix}\right\rangle - \left\langle \varphi, \begin{bmatrix}\rho_p\\ \eta_p\end{bmatrix}\right\rangle = 0,$$

the middle equality being a consequence of $\varphi \in \tilde{E}_T$. Thus, $\varphi \in E_T^{\perp}$. Using Proposition 3.8, we get

$$\|\mathbb{D}_T\varphi\| \geq C_7\sqrt{T}\,\|\varphi\|,$$

contradicting to the fact that $\varphi \in \tilde{E}_T$ (this space, we recall, is the sum of the eigenspaces of $-\mathbb{D}_T^2$ associated with eigenvalues in $[0, C_6^2]$.) Therefore, $\tilde{E}_T$ is isomorphic to E_T when T is sufficiently large. □

Recall that X is an adjusted version of V. By Proposition 4.1, we finally have

$$\begin{aligned}\kappa(M,\omega) &= \dim \tilde{E}_T \mod 2\\ &= \dim E_T \mod 2\\ &= \text{the number of zero points of the adjusted vector field } X \mod 2\\ &= \text{the number of zero points of the original vector field } V \mod 2,\end{aligned}$$

and this completes the proof of Theorem 1.5.

Remark 4.2. We get Corollary 1.8 from Theorem 1.5. An alternative to using Theorem 1.5 involves applying Atiyah's perturbation technique from [1, Section 4] to prove Corollary 1.8 directly. Let V be a vector field with $g(V,V)=1$ on M. We perturb the operator

$$D = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\hat{c}(\mathrm{dvol}) & \\ & \hat{c}(\mathrm{dvol})\end{bmatrix}\begin{bmatrix}d+d^* & \omega\\ \omega^* & -d-d^*\end{bmatrix}$$

on $\underline{\Omega}(M)$ into the operator

$$D' = D + \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\hat{c}(V) & \\ & -\hat{c}(V)\end{bmatrix} D \begin{bmatrix}\hat{c}(V) & \\ & -\hat{c}(V)\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}.$$

Once we verify that $\mathrm{ind}_2 D = \mathrm{ind}_2 D'$ and that $\ker D'$ admits a complex structure given by

$$\left(\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\hat{c}(V) & \\ & -\hat{c}(V)\end{bmatrix}\right)^2 = \begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix},$$

we conclude that $\dim\ker D'$ is even, and therefore $\kappa(M,\omega)=0$.

5. Examples

We illustrate with some examples, most of which have already been studied in other papers; we just adapt them to the computation of symplectic semi-characteristics.

Example 5.1. We equip $M = \mathbb{C}\mathrm{P}^2$ with the Fubini–Study form [15, Homework 12]. According to [7, Example 4.2],

$$b_0^{\omega} = 1, \quad b_2^{\omega} = 0, \quad b_4^{\omega} = 0,$$

meaning that $\kappa(M,\omega)=1$. These b_i^ω are computed using a Morse function with three critical points, together with the associated cone Morse cochain complex [7, Definition 1.2]. By the counting formula (1-4), we can also use the three critical points of this perfect Morse function to find $\kappa(M,\omega)=1$.

Example 5.2. Let $M=\mathbb{S}^2\times\mathbb{S}^2$ equipped with the standard symplectic structure. Recall that we have a height function h (see [4, Example 3.4]) on $\mathbb{S}^2$ with two critical points. Then,

$$f:\mathbb{S}^2\times\mathbb{S}^2\to\mathbb{R},\quad (p,q)\mapsto h(p)+h(q),$$

is a Morse function on $\mathbb{S}^2$ with four critical points. Thus, in this case, $\kappa(M,\omega)=0$.

As we know, the Euler characteristic of the de Rham cohomology of $\mathbb{S}^2\times\mathbb{S}^2$ is 4, meaning that $\mathbb{S}^2\times\mathbb{S}^2$ does not admit a nonvanishing vector field. However, as we see, its symplectic semi-characteristic is 0. Thus, in terms of judging the existence of nonvanishing vector fields, the symplectic semi-characteristic (1-3) of the primitive cohomology is a weak substitute for the Euler characteristic of the de Rham cohomology.

Example 5.3. As in [18, Section 3.4] and [16, (5.3)], we let $\sim$ be the identification

$$(x_1,x_2,x_3,x_4)\sim(x_1+a,x_2+b,x_3+c,x_4+d-bx_3)\quad(\text{when } a,b,c,d\in\mathbb{Z})$$

on $\mathbb{R}^4$. Then, the Kodaira–Thurston fourfold is equal to $\mathbb{R}^4/{\sim}$. Let $M=\mathbb{R}^4/{\sim}$ equipped with the symplectic form ω given in [18, (3.26)] and [16, (5.4)]. We know that $\kappa(M,\omega)=0$ from the tables of primitive cohomology groups in [18, Section 3.4] and [16, Section 5.4]. Since $\mathbb{R}^4/{\sim}$ has a globally defined tangent vector field ∂_{x_1}, we also obtain $\kappa(M,\omega)=0$ according to Corollary 1.8.

Example 5.4. We briefly mention the $(4n+2)$-dimensional case. Let $M=\mathbb{T}^2$ be equipped with the standard symplectic form. Since $\mathbb{T}^2$ is Kähler, we use the formula [7, (4.4)] to find $b_0^\omega=1$, $b_2^\omega=2$, and so $\kappa(M,\omega)=1$.

However, we know there is a height function [13, Part I, Section 1] with 4 nondegenerate critical points on $\mathbb{T}^2$. This means our Theorem 1.5 does not apply to the $(4n+2)$-dimensional case.

6. Discussion and perspectives

We now mention some possible extensions of this project. The deformation replaces $d+d^*$ by $d+d^*+T\hat{c}(V)$ and does not perturb ω. This preserves the symplectic information and relates $\kappa(M,\omega)$ to the primitive forms given by the Lefschetz decomposition. However, as $\kappa(M,\omega)$ is unchanged when replacing ω by another symplectic form, it is also natural to consider replacing ω by $\omega\wedge\cdots\wedge\omega$ or by other forms.

The form $\omega\wedge\cdots\wedge\omega$ gives the semi-characteristic of the 1-filtered cohomology [16; 20]. In a follow-up work [27], we use $\omega\wedge\omega$ on a $(4n+2)$-dimensional closed symplectic manifold. Then, the associated semi-characteristic vanishes, which is exactly the parity of the de Rham Euler characteristic. This provides more evidence that κ relies more on M than on the form.

Also, from an index-theoretic perspective and without involving the primitive cohomology, we may perturb ω into $s\omega$ and let s change from 0 to 1. The study could thus be done for any closed orientable manifold equipped with a closed homogeneous form. For this case, we will give the formula and the analysis in a future joint work with S. Xu, with assumptions on both $\dim M$ and the degree of the form. A similar perturbation using s^{-1} instead of s was carried out in a recent work [28] on the mapping cone Morse theory for any closed oriented manifold equipped with a closed homogeneous smooth form. This s^{-1} preserves the cohomology.

To conclude, we briefly discuss the K-theoretic background of this study. By [3, Theorem 2.3], the mod 2 index is equivalent to the map

$$KO^{-1}(TM) \to KO^{-1}(\text{point}) \cong \mathbb{Z}_2. \tag{6-1}$$

When $\dim M = 4n$, we have a skew-adjoint elliptic operator (3-4) whose skew-symbol class is in $KO^{-1}(TM)$ and then mapped to $\kappa(M,\omega)$. In the $(4n+2)$-dimensional case, if we continue the current pattern of construction, we cannot obtain a skew-adjoint elliptic operator on $\underline{\Omega}(M)$ whose skew-symbol class is mapped to $\kappa(M,\omega)$. Thus, in the $4n+2$ case, it could be worthwhile to study the geometric meaning of the image of the skew-symbol class of (2-7) under (6-1) [23, Theorem 3.2], and whether (6-1) can be generalized to give the $(4n+2)$-dimensional case a KO-valued index result of the semi-characteristic.

Acknowledgments

I thank my PhD supervisor, Prof. Xiang Tang, for introducing me to the analytic method in symplectic Morse theory and offering practical suggestions to simplify several steps. I thank Prof. Li-Sheng Tseng for mentioning the relations between mapping cone complexes and odd sphere bundles, and Dr. David Clausen for his talk on the symplectic Morse inequality. I thank Prof. Yi Lin for pointing out the influence of the Kähler condition, and Prof. Christopher Seaton for asking about the relations between the classical Euler characteristic and the symplectic semi-characteristic. I thank the referees for the helpful comments and suggestions on mathematics and writing. Finally, I thank Washington University in St. Louis for supporting my PhD studies during the completion of this project.

References

[1] M. F. Atiyah, *Vector fields on manifolds*, Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen **200**, Westdeutscher Verlag, Cologne, 1970. MR

[2] M. F. Atiyah and I. M. Singer, "Index theory for skew-adjoint Fredholm operators", *Inst. Hautes Études Sci. Publ. Math.* **37** (1969), 5–26. MR

[3] M. F. Atiyah and I. M. Singer, "The index of elliptic operators, V", *Ann. of Math.* (2) **93** (1971), 139–149. MR

[4] A. Banyaga and D. Hurtubise, *Lectures on Morse homology*, Texts in the Mathematical Sciences **29**, Kluwer, Dordrecht, 2004. Reprinted Springer, 2013. MR

[5] N. Berline, E. Getzler, and M. Vergne, *Heat kernels and Dirac operators*, Grundl. Math. Wissen. **298**, Springer, 1992. MR

[6] J.-M. Bismut and G. Lebeau, "Complex immersions and Quillen metrics", *Inst. Hautes Études Sci. Publ. Math.* **74** (1991), 1–298. MR

[7] D. Clausen, X. Tang, and L.-S. Tseng, "Symplectic Morse theory and Witten deformation", *Math. Z.* **312**:3 (2026), art. id. 94, 16 pp. MR

[8] J. B. Conway, *A course in functional analysis*, Graduate Texts in Mathematics **96**, Springer, 1985. MR

[9] W. H. Greub, *Linear algebra*, 3rd ed., Grundl. Math. Wissen. **97**, Springer, 1967. MR

[10] N. Higson and J. Roe, *Analytic K-homology*, Oxford University Press, Oxford, 2000. MR

[11] M. Kervaire, "Courbure intégrale généralisée et homotopie", *Math. Ann.* **131** (1956), 219–252. MR

[12] G. Lusztig, J. Milnor, and F. P. Peterson, "Semi-characteristics and cobordism", *Topology* **8** (1969), 357–359. MR

[13] J. Milnor, *Morse theory*, Annals of Mathematics Studies **51**, Princeton University Press, 1963. MR

[14] L. I. Nicolaescu, *Lectures on the geometry of manifolds*, 3rd ed., World Scientific, Hackensack, NJ, 2021. MR

[15] A. Cannas da Silva, *Lectures on symplectic geometry*, Lecture Notes in Mathematics **1764**, Springer, 2001 (reprinted with corrections 2008). MR

[16] H. L. Tanaka and L.-S. Tseng, "Odd sphere bundles, symplectic manifolds, and their intersection theory", *Camb. J. Math.* **6**:3 (2018), 213–266. MR

[17] M. E. Taylor, *Partial differential equations II: Qualitative studies of linear equations*, 3rd ed., Applied Mathematical Sciences **116**, Springer, 2023. MR

[18] L.-S. Tseng and S.-T. Yau, "Cohomology and Hodge theory on symplectic manifolds, I", *J. Differential Geom.* **91**:3 (2012), 383–416. MR

[19] L.-S. Tseng and S.-T. Yau, "Cohomology and Hodge theory on symplectic manifolds, II", *J. Differential Geom.* **91**:3 (2012), 417–443. MR

[20] C.-J. Tsai, L.-S. Tseng, and S.-T. Yau, "Cohomology and Hodge theory on symplectic manifolds, III", *J. Differential Geom.* **103**:1 (2016), 83–143. MR

[21] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*, Scott, Foresman, London, 1971. Reprinted Springer, 2013. MR

[22] E. Witten, "Supersymmetry and Morse theory", *J. Differential Geometry* **17**:4 (1982), 661–692. MR

[23] W. Zhang, "A counting formula for the Kervaire semi-characteristic", *Topology* **39**:3 (2000), 643–655. MR

[24] W. Zhang, *Lectures on Chern–Weil theory and Witten deformations*, Nankai Tracts in Mathematics **4**, World Scientific, River Edge, NJ, 2001. MR

[25] W. Zhang and H. Feng, 流形上的几何与分析 [*Geometry and analysis on manifolds*], Higher Education Press, Beijing, 2022.

[26] H. Zhuang, "Analytic and topological realizations of the invariant Thom–Smale complex", preprint, 2025. arXiv 2311.10417v4

[27] H. Zhuang, "A vanishing property about the 1-filtered cohomology groups of $(4n+2)$-dimensional closed symplectic manifolds", preprint, 2025. arXiv 2510.10630

[28] H. Zhuang, "Instanton construction of the mapping cone Thom–Smale complex", preprint, 2026. arXiv 2603.08404

Received October 30, 2025. Revised March 10, 2026.

HAO ZHUANG
BEIJING INTERNATIONAL CENTER FOR MATHEMATICAL RESEARCH
PEKING UNIVERSITY
BEIJING
CHINA

hzhuang@pku.edu.cn

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to journals+pjm@msp.org 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.

At submission time a PDF file is required, and it must be derived from a LaTeX source file in which the command `\documentclass[12pt]{amsart}` is active. No changes to horizontal or vertical margins, or line spacing, should be made. Please include in the preamble the command `\usepackage[pagewise]{lineno}\linenumbers` — this will spare referees tedious effort in line counting.

Carefully preserve all related files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to journals+pjm@msp.org.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text (“the curve looks like this:”). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as “Place Figure 1 here”. The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal’s preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 343 No. 2 August 2026