Pacific Journal of Mathematics

Volume 340 No. 2 February 2026

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

EDITORS

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

Matthias Aschenbrenner
Fakultät für Mathematik
Universität Wien
Vienna, Austria
matthias.aschenbrenner@univie.ac.at

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

Atsushi Ichino
Department of Mathematics
Kyoto University
Kyoto 606-8502, Japan
atsushi.ichino@gmail.com

Robert Lipshitz
Department of Mathematics
University of Oregon
Eugene, OR 97403
lipshitz@uoregon.edu

Kefeng Liu
School of Sciences
Chongqing University of Technology
Chongqing 400054, China
liu@math.ucla.edu

Sucharit Sarkar
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
sucharit@math.ucla.edu

Dimitri Shlyakhtenko
Department of Mathematics
University of California
Los Angeles, CA 90095-1555
shlyakht@ipam.ucla.edu

Ruixiang Zhang
Department of Mathematics
University of California
Berkeley, CA 94720-3840
ruixiang@berkeley.edu

PRODUCTION

Silvio Levy, Scientific Editor, production@msp.org

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/

© 2026 Mathematical Sciences Publishers

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.199

FUNCTIONS OF DISSIPATIVE OPERATORS UNDER RELATIVELY BOUNDED AND RELATIVELY TRACE CLASS PERTURBATIONS

A.B. ALEKSANDROV AND V.V. PELLER

We study the behaviour of functions of dissipative operators under relatively bounded and relatively trace class perturbations. We introduce and study the class of analytic relatively operator Lipschitz functions. An essential role is played by double operator integrals with respect to semispectral measures. We also study the class of analytic resolvent Lipschitz functions. Then we obtain a trace formula in the case of relatively trace class perturbations and show that the maximal class of functions for which the trace formula holds in the case of relatively trace class perturbations coincides with the class of analytic relatively operator Lipschitz functions. We also establish the inequality $\int |\xi(t)|(1+|t|)^{-1}\,\mathrm{d}t < \infty$ for the spectral shift function ξ in the case of relatively trace class perturbations.

1. Introduction 200
2. Maximal dissipative operators. Functional calculus 201
3. Trace formulae 204
4. Double operator integrals 206
5. Relatively bounded and relatively trace class perturbations of dissipative operators 207
6. Analytic Schur multipliers 211
7. A representation of operator differences in terms of double operator integrals 212
8. Various characterizations of analytic relatively operator Lipschitz functions 215
9. Differentiaton in the parameter 218
10. The trace formula in the case of relatively trace class perturbations 221
References 226

The research on § 4-6 is supported by Russian Science Foundation [grant number 23-11-00153]. The research on § 7-10 is supported by a grant of the Government of the Russian Federation for the state support of scientific research, carried out under the supervision of leading scientists, agreement 075-15-2025-013.

MSC2020: primary 47A55; secondary 47A20, 47A60, 47B10, 47B15, 47B44.

Keywords: dissipative operator, trace formula, relatively bounded perturbation, relatively trace class perturbation, semispectral measure, double operator integral, relatively operator Lipschitz function.

© 2026 The Authors, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

1. Introduction

In our recent paper [AP7] (see also [AP8]) we studied the behaviour of functions of self-adjoint operators under relatively bounded and relatively compact perturbations. Recall that a linear operator K is called a *relatively bounded perturbation* of an unbounded self-adjoint operator A if the operator $K(A+\mathrm{i}I)^{-1}$ is bounded and it is called a *relatively trace class perturbation* of A if the operator $K(A+\mathrm{i}I)^{-1}$ belongs to the trace class $\boldsymbol{S}_1$. We introduced in [AP7] and [AP8] the class ROL of *relatively operator Lipschitz functions* as the class of functions f on the real line $\mathbb{R}$, for which there is a constant k such that the inequality

$$\|f(B)-f(A)\|\le k\|(B-A)(A+\mathrm{i}I)^{-1}\|$$

holds for arbitrary self-adjoint operators A and B such that $B-A$ is a relatively bounded perturbation of A.

We found in [AP7] and [AP8] various characterizations of the class ROL of relatively operator Lipschitz functions. We also obtained in [AP7] and [AP8] the trace formula

$$\operatorname{trace}\big(f(B)-f(A)\big)=\int_{\mathbb{R}} f'(t)\boldsymbol{\xi}(t)\,\mathrm{d}t \tag{1.1}$$

for arbitrary $f\in\mathrm{ROL}$ and arbitrary pairs of self-adjoint operators A and B such that $B-A$ is a relatively trace class perturbation of A. Here $\boldsymbol{\xi}$ is the *spectral shift function* for the pair (A,B), which satisfies the condition

$$\int\frac{|\boldsymbol{\xi}(t)|}{1+|t|}\,\mathrm{d}t<\infty$$

(the last inequality was also obtained in [CS]). Moreover, it turned out that ROL is the maximal class of functions for which trace formula (1.1) holds for arbitrary pairs (A,B) such that $B-A$ is a relatively trace class perturbation of A; see [AP7] and [AP8].

In this paper we consider similar problems for functions of dissipative operators. In particular, we find several characterizations of the class of analytic relatively operator Lipschitz functions and we obtain a trace formula in the case of relatively trace class perturbations. Note, however, that certain proofs in the dissipative case are considerably harder than in the self-adjoint case.

We express our deep gratitude to the referee for very thorough reading our text and for making many valuable suggestions.

2. Maximal dissipative operators. Functional calculus

In this section we give a brief introduction in dissipative operators and we refer the reader to [AP2] (see also [SNF] for more detail).

Let $\mathscr{H}$ be a Hilbert space. An operator L (not necessarily bounded) with dense domain $\mathrm{Dom}(L)$ in $\mathscr{H}$ is called *dissipative* if

$$\mathrm{Im}(Lu, u) \geq 0, \quad u \in \mathrm{Dom}(L).$$

A dissipative operator is called *maximal dissipative* if it has no proper dissipative extension.

The *Cayley transform* of a dissipative operator L is defined by

$$T \stackrel{\mathrm{def}}{=} (L - \mathrm{i}I)(L + \mathrm{i}I)^{-1}$$

with domain $\mathrm{Dom}(T) = (L+\mathrm{i}I)\mathrm{Dom}(L)$ and range $\mathrm{Range}\, T = (L-\mathrm{i}I)\mathrm{Dom}(L)$ (the operator T is not densely defined in general). T is a contraction, i.e. $\|Tu\| \leq \|u\|$ for $u \in \mathrm{Dom}(T)$; 1 is not an eigenvalue of T; and $\mathrm{Range}(I - T) \stackrel{\mathrm{def}}{=} \{u - Tu : u \in \mathrm{Dom}(T)\}$ is dense.

Conversely, if T is a contraction defined on its domain $\mathrm{Dom}(T)$ such that 1 is not an eigenvalue of T and $\mathrm{Range}(I - T)$ is dense, then it is the Cayley transform of a dissipative operator L and L is the inverse Cayley transform of T:

$$L = \mathrm{i}(I + T)(I - T)^{-1}, \quad \mathrm{Dom}(L) = \mathrm{Range}(I - T).$$

A dissipative operator is maximal if and only if the domain of its Cayley transform is the whole Hilbert space. Clearly, a maximal dissipative operator must be closed.

If L is a dissipative operator which is not maximal, there is a canonical way to extend L to a maximal dissipative operator. Indeed, let T be the Cayley transform of L. Then T extends uniquely to the closure of its domain. If the closure is a proper subspace of our Hilbert space, we can extend T to the whole space by defining it to be the zero operator on the orthogonal complement to its domain. The inverse Cayley transform of this extension is a maximal dissipative operator. It is natural to consider it as the *canonical extension of L to a maximal dissipative operator*.

We proceed now to the construction of functional calculus for dissipative operators. Let L be a maximal dissipative operator and let T be its Cayley transform. Consider its minimal unitary dilation U, i.e. U is a unitary operator defined on a Hilbert space $\mathscr{K}$ that contains $\mathscr{H}$ such that

$$T^n = P_{\mathscr{H}} U^n \big| \mathscr{H}, \quad n \geq 0,$$

and $\mathscr{K} = \mathrm{clos\,span}\{U^n h : h \in \mathscr{H},\ n \in \mathbb{Z}\}$. Since 1 is not an eigenvalue of T, it follows that 1 is not an eigenvalue of U (see [SNF], Ch. II, § 6).

The Sz.-Nagy–Foiaş functional calculus allows us to define a functional calculus for T on the Banach algebra

$$\mathrm{C}_{\mathrm{A},1} \stackrel{\text{def}}{=} \big\{ g \in H^\infty : \ g \quad \text{is continuous on} \quad \mathbb{T} \setminus \{1\} \ \big\}.$$

If $g \in \mathrm{C}_{\mathrm{A},1}$, we put

$$g(T) \stackrel{\text{def}}{=} P_{\mathscr{H}} g(U) \Big| \mathscr{H}.$$

This functional calculus is linear and multiplicative, and

$$\|g(T)\| \le \|g\|_{H^\infty}, \quad g \in \mathrm{C}_{\mathrm{A},1}.$$

We can define now a functional calculus for our dissipative operator on the Banach algebra

$$\mathrm{C}_{\mathrm{A},\infty} \stackrel{\text{def}}{=} \big\{ f \in H^\infty(\mathbb{C}_+) : \ f \quad \text{is continuous on} \quad \mathbb{R} \big\}.$$

Indeed, if $f \in \mathrm{C}_{\mathrm{A},\infty}$, we put

$$f(L) \stackrel{\text{def}}{=} \big(f \circ \omega\big)(T),$$

where ω is the conformal map of $\mathbb{D}$ onto $\mathbb{C}_+$ defined by $\omega(\zeta) \stackrel{\text{def}}{=} \mathrm{i}(1+\zeta)(1-\zeta)^{-1}$, $\zeta \in \mathbb{D}$.

Note that $\|f\|_{H^\infty(\mathbb{C}_+)} \le \|f\|_{L^\infty(\mathbb{R})}$, and so f is uniquely determined by its values on $\mathbb{R}$.

Let us define now a resolvent self-adjoint dilation of a maximal dissipative operator. If L is a maximal dissipative operator in a Hilbert space $\mathscr{H}$, we say that a self-adjoint operator A in a Hilbert space $\mathscr{K}$, $\mathscr{K} \supset \mathscr{H}$, is called a *resolvent self-adjoint dilation* of L if

$$(L - \lambda I)^{-1} = P_{\mathscr{H}} (A - \lambda I)^{-1} \Big| \mathscr{H}, \quad \operatorname{Im} \lambda < 0.$$

The dilation is called *minimal* if

$$\mathscr{K} = \operatorname{clos} \operatorname{span} \big\{ (A - \lambda I)^{-1} v : \ v \in \mathscr{H}, \ \operatorname{Im} \lambda < 0 \big\}.$$

If $f \in \mathrm{C}_{\mathrm{A},\infty}$, then

$$f(L) = P_{\mathscr{H}} f(A) \Big| \mathscr{H}, \quad f \in \mathrm{C}_{\mathrm{A},\infty}.$$

A minimal resolvent self-adjoint dilation of a maximal dissipative operator always exists (and is unique up to a natural isomorphism). Indeed, it suffices to take a minimal unitary dilation of the Cayley transform of this operator and apply the inverse Cayley transform to it.

We can define now the semi-spectral measure $\mathscr{E}_L$ of a maximal dissipative operator L by

$$\mathscr{E}_L(\Delta) = P_{\mathscr{H}} E_A(\Delta)\big|\mathscr{H}, \quad \Delta \quad \text{is a Borel subset of} \quad \mathbb{R},$$

where E_A is the spectral measure of the minimal resolvent self-adjoint dilation of L. Then

$$f(L) = \int_{\mathbb{R}} f(x)\,\mathrm{d}\mathscr{E}_L(x), \quad f \in \mathrm{C}_{\mathrm{A},\infty}. \tag{2.1}$$

Note that the semi-spectral measure of a maximal dissipative operator L can be obtained from the semi-spectral measure of the Cayley transform of L by conformally transferring it from the unit circle to the real line.

We refer the reader to [AP2] for more detail.

We also need functions $f(L)$ of maximal dissipative operators in the case when f is not necessarily a bounded function. Suppose that f is a Lipschitz function in $\operatorname{clos}\mathbb{C}_+$ that is analytic in $\mathbb{C}_+$. We can define the operator $f(L)$ as follows. We have

$$f(\zeta) = \frac{f_{\mathrm{i}}(\zeta)}{(\zeta+\mathrm{i})^{-1}}, \quad \zeta \in \operatorname{clos}\mathbb{C}_+, \qquad \text{where} \quad f_{\mathrm{i}}(\zeta) \stackrel{\text{def}}{=} \frac{f(\zeta)}{\zeta+\mathrm{i}}. \tag{2.2}$$

Clearly, $f_{\mathrm{i}} \in \mathrm{C}_{\mathrm{A},\infty}$ The (possibly unbounded) operator $f(L)$ can be defined by

$$f(L) \stackrel{\text{def}}{=} (L+\mathrm{i}I)\,f_{\mathrm{i}}(L) \tag{2.3}$$

(see [SNF], Ch. IV, § 1). It follows from Th. 1.1 of Ch. IV of [SNF] that

$$f(L) \supset f_{\mathrm{i}}(L)(L+\mathrm{i}I),$$

and so $\operatorname{Dom}(f(L)) \supset \operatorname{Dom}(L)$.

We are ready now to define operator Lipschitz functions analytic in the upper half-plane $\operatorname{clos}\mathbb{C}_+$.

Definition. Suppose that f is a Lipschitz function in the closed upper half-plane $\operatorname{clos}\mathbb{C}_+$ that is analytic in the open half-plane $\mathbb{C}_+$. We say that f is an *operator Lipschitz function* if

$$\|f(L) - f(M)\| \le \operatorname{const}\|L - M\|, \tag{2.4}$$

whenever L and M are maximal dissipative operators with bounded difference.

Note that in the case when $L - M$ is bounded, both operators $f(L)$ and $f(M)$ are defined on the dense subset $\operatorname{Dom}(L) = \operatorname{Dom}(M)$.

We denote by OL_{A} the class of such operator Lipschitz functions. The norm in OL_{A} is, by definition, the smallest possible constant on the right of (2.4).

Note also that a function f analytic in $\mathbb{C}_+$ and continuous in $\operatorname{clos}\mathbb{C}_+$ is operator Lipschitz if and only if

$$\|f(N_1) - f(N_2)\| \le \operatorname{const}\|N_1 - N_2\|,$$

whenever N_1 and N_2 are bounded normal operators with spectra in $\operatorname{clos}\mathbb{C}_+$; see [AP3] and [AP5]. We refer the reader to the survey [AP3] for detailed information on operator Lipschitz functions.

In § 10 of this paper we obtain a trace formula in the case when L and M are maximal dissipative operators such that $M - L$ is a relatively trace class perturbation of L.

3. Trace formulae

Let A be a (not necessarily bounded) self-adjoint operator and $B = A + K$, where K is a trace class self-adjoint operator that is considered as a perturbation of A. Physicist I.M. Lifshits in [L] proposed the following trace formula

$$\operatorname{trace}\big(f(B) - f(A)\big) = \int_{\mathbb{R}} f'(t)\boldsymbol{\xi}(t)\,\mathrm{d}t \tag{3.1}$$

for sufficiently nice functions f on $\mathbb{R}$, where $\boldsymbol{\xi}$ is an integrable function on $\mathbb{R}$. It is called the *spectral shift function associated with* the pair (A, B). Later M.G. Krein found in [Kr] a rigorous mathematical justification of trace formula (3.1). We are going to refer to (3.1) as the *Lifshits–Krein trace formula*. Krein posed in [Kr] the problem to identify the maximal class of functions f on $\mathbb{R}$, for which formula (3.1) holds for all pairs (A, B) of self-adjoint operators with trace class difference. Krein's problem was solved in the paper [Pe5]: the maximal class of such functions f coincides with the class OL of operator Lipschitz functions on $\mathbb{R}$.

If we replace the condition $B - A \in \boldsymbol{S}_1$ with the resolvent condition

$$(B + \mathrm{i}I)^{-1} - (A + \mathrm{i}I)^{-1} \in \boldsymbol{S}_1, \tag{3.2}$$

then trace formula (3.1) holds for the class of *resolvent Lipschitz functions* f, i.e. for functions f satisfying the inequality

$$\|f(B) - f(A)\| \le \operatorname{const}\big\|(B + \mathrm{i}I)^{-1} - (A + \mathrm{i}I)^{-1}\big\|$$

for arbitrary pairs (A, B) of self-adjoint operators. The spectral shift function $\boldsymbol{\xi}$ under condition (3.2) must satisfy the inequality

$$\int_{\mathbb{R}} \frac{|\boldsymbol{\xi}(t)|}{1 + t^2}\mathrm{dt} < \infty. \tag{3.3}$$

This is a well-known fact and can be reduced to the trace formula for unitary operators; see [MNP2] for details. The maximal class of functions f, for which trace

formula (3.1) holds for arbitrary pairs (A, B) of self-adjoint operators satisfying the resolvent condition (3.2) coincides with the class of resolvent Lipschitz functions. The latter follows from the corresponding result for functions of unitary operators [AP4]; see [MNP2] for details. We refer the reader to [AP7] and [AP8] for various characterizations of the class of resolvent Lipschitz functions.

Finally, consider the situation when $K = B - A$ is a relatively trace class perturbation of A. In this case trace formula (3.1) holds for arbitrary relatively operator Lipschitz functions and the spectral shift function $\boldsymbol{\xi}$ in this situation must satisfy the inequality

$$\int_{\mathbb{R}} \frac{|\boldsymbol{\xi}(t)|}{1+|t|}\mathrm{d}t < \infty;$$

see [CS] and [AP8] (see also [AP7]). Moreover, the class ROL of relatively operator Lipschitz functions is the maximal class of functions, for which trace formula (3.1) holds for arbitrary pairs (A, B) of self-adjoint operators such that $B - A$ is a relatively trace class perturbation of A.

We proceed now to trace formulae for functions of dissipative operators. The long-standing problem to generalize the Lifshitz–Krein trace formula to the case of functions of maximal dissipative operators in the case of trace class perturbations and in the case of trace class resolvent perturbations was solved in [MNP1] and [MNP2], see also [MN]. We also refer the reader to [MNP2] for the history of the question.

Suppose that L is a maximal dissipative operator and M is a dissipative operator such that $M - L \in \boldsymbol{S}_1$. Then there exists an integrable function $\boldsymbol{\xi}$ on $\mathbb{R}$ such that for $f \in \mathrm{OL_A}$, the following trace formula holds

$$\operatorname{trace}\big(f(M) - f(L)\big) = \int_{\mathbb{R}} f'(t)\boldsymbol{\xi}(t)\,\mathrm{d}t. \tag{3.4}$$

Such a function $\boldsymbol{\xi}$ (which is never unique) is called a *spectral shift function for the pair* (L, M). As in the case of self-adjoint operators, $\mathrm{OL_A}$ is the maximal class of functions f, for which formula (3.4) holds for all such pairs (L, M).

In the case when (L, M) is a pair of maximal dissipative operators such that M is a resolvent trace class perturbation of L, i.e.

$$(M + \mathrm{i}I)^{-1} - (L + \mathrm{i}I)^{-1} \in \boldsymbol{S}_1, \tag{3.5}$$

trace formula (3.4) holds for a spectral shift function $\boldsymbol{\xi}$ satisfying (3.3) and for functions f in the class $\mathrm{OL_A^{res}}$ of *analytic resolvent Lipschitz functions*, i.e. for functions f in $\mathrm{C_{A,\infty}}$ such that

$$\|f(M) - f(L)\| \le \mathrm{const}\,\|(M + \mathrm{i}I)^{-1} - (L + \mathrm{i}I)^{-1}\|$$

for arbitrary maximal dissipative operators L and M. It was also shown in [MNP2] that $\mathrm{OL}_{\mathrm{A}}^{\mathrm{res}}$ is the maximal class of functions, for which trace formula (3.4) holds for arbitrary pairs (L, M) of maximal dissipative operators satisfying (3.5). We obtain in § 8 of this paper various characterizations of the class $\mathrm{OL}_{\mathrm{A}}^{\mathrm{res}}$. Also, we prove in § 10 a version of the trace formula for analytic relatively operator Lipschitz perturbations.

4. Double operator integrals

Double operator integrals

$$\iint_{\mathscr{X}\times\mathscr{Y}} \Phi(x, y)\, \mathrm{d}E_1(x) Q\, \mathrm{d}E_2(y) \tag{4.1}$$

appeared first in the paper [DK] by Yu.L. Daletskii and S.G. Krein. Later Birman and Solomyak elaborated their beautiful theory of double operator integrals; see [BS1], [BS2] and [BS4] (see also [Pe4], [AP3] and references therein). Here Φ is a bounded measurable function, E_1 and E_2 are spectral measures on Hilbert space defined on σ-algebras of subsets of sets $\mathscr{X}$ and $\mathscr{Y}$ and Q is a bounded linear operator.

The starting point of the Birman–Solomyak approach [BS1] is the case when Q belongs to the Hilbert–Schmidt class $\boldsymbol{S}_2$; they managed to define double operator integrals for arbitrary bounded measurable functions Φ. Moreover, for $Q \in \boldsymbol{S}_2$ and for a bounded measurable function Φ, the operator (4.1) must belong to $\boldsymbol{S}_2$ and the inequality

$$\left\| \iint_{\mathscr{X}\times\mathscr{Y}} \Phi(x, y)\, \mathrm{d}E_1(x) Q\, \mathrm{d}E_2(y) \right\|_{\boldsymbol{S}_2} \le \|\Phi\|_{L^\infty} \|Q\|_{\boldsymbol{S}_2}$$

must hold.

Next, Φ is said to be a *Schur multiplier with respect to* E_1 and E_2 if

$$Q \in \boldsymbol{S}_1 \quad \Longrightarrow \quad \iint_{\mathscr{X}\times\mathscr{Y}} \Phi(x, y)\, \mathrm{d}E_1(x) Q\, \mathrm{d}E_2(y) \in \boldsymbol{S}_1$$

(see [Pe1] and [AP3]) for more detail). In this case one can define by duality the double operator integral (4.1) for an arbitrary bounded linear operator Q and the following inequality must hold

$$\left\| \iint_{\mathscr{X}\times\mathscr{Y}} \Phi(x, y)\, \mathrm{d}E_1(x) Q\, \mathrm{d}E_2(y) \right\| \le \|\Phi\|_{\mathfrak{M}_{E_1,E_2}} \|Q\|,$$

where $\mathfrak{M}_{E_1,E_2}$ stands for the space of Schur multipliers with respect to E_1 and E_2 equipped with the natural norm.

It is well known (see [Pe1], [AP3], [AP6]) that $\Phi \in \mathfrak{M}_{E_1,E_2}$ if and only if Φ belongs to the *Haagerup tensor product* $L^\infty_{E_1} \otimes_{\rm h} L^\infty_{E_2}$ of $L^\infty_{E_1}$ and $L^\infty_{E_2}$ or, in other words, Φ admits a representation

$$\Phi(x, y) = \sum_{n\ge 0} \varphi_n(x)\psi_n(y), \tag{4.2}$$

where the φ_n and ψ_n are measurable functions such that

$$\sum_{n\ge 0} |\varphi_n|^2 \in L^\infty_{E_1} \quad \text{and} \quad \sum_{n\ge 0} |\psi_n|^2 \in L^\infty_{E_2}.$$

In this case

$$\iint \Phi(x, y)\, \mathrm{d}E_1(x)\, Q\, \mathrm{d}E_2(y) = \sum_{n\ge 0} \left(\int \varphi_n(x)\, \mathrm{d}E_1(x) \right) Q \left(\int \psi_n(y)\, \mathrm{d}E_2(y) \right), \tag{4.3}$$

the series on the right converges in the weak operator topology and the right-hand side does not depend on the choice of a representation in (4.2). Note that the sufficiency of the condition $\Phi \in L^\infty_{E_1} \otimes_{\rm h} L^\infty_{E_2}$ for Φ to be a Schur multiplier was established earlier by Birman and Solomyak in [BS1].

We are going to use the notation $\mathfrak{M}(\mathbb{R}^2)$ the class of functions on $\mathbb{R}^2$ that are Schur multipliers with respect to arbitrary Borel spectral measures on $\mathbb{R}$.

We also need double operator integrals with respect to *semi-spectral measures*

$$\iint \Phi(x, y)\, \mathrm{d}\mathscr{E}_1(x)\, Q\, \mathrm{d}\mathscr{E}_2(y). \tag{4.4}$$

Such double operator integrals were introduced in [Pe2] (see also [Pe4]). By analogy with the case of double operator integrals with respect to spectral measures, double operator integrals of the form (4.4) can be defined for arbitrary bounded measurable functions Φ in the case when $Q \in \boldsymbol{S}_2$ and for Schur multipliers Φ with respect to the semi-spectral measures $\mathscr{E}_1$ and $\mathscr{E}_2$ in the case of an arbitrary bounded operator Q. The latter is equivalent to the fact that Φ belongs to the Haagerup tensor product $L^\infty_{\mathscr{E}_1} \otimes_{\rm h} L^\infty_{\mathscr{E}_2}$ in which case the definition of (4.4) is similar to formula (4.3).

5. Relatively bounded and relatively trace class perturbations of dissipative operators

Recall that for a linear operator A in a Hilbert space, its domain is denoted by $\mathrm{Dom}(A)$. Let A and K be densely defined linear operators such that

$$\mathrm{Dom}(A) \subset \mathrm{Dom}(K).$$

It is well known (see [BS5], Theorem 1, Ch. 3, Sec. 4) that if the operator A is closed and the operator K is closable, then K is *dominated by* A, i.e.

$$\|Kv\| \le \text{const}(\|v\| + \|Av\|), \quad v \in \text{Dom}(A). \tag{5.1}$$

Suppose now that L and M are maximal dissipative operators such that $K \stackrel{\text{def}}{=} M - L$ is dominated by L. In this case we say that K is a *relatively bounded perturbation of* L. Note that K is a relatively bounded perturbation of L if and only if the operator

$$C \stackrel{\text{def}}{=} K(L + \mathrm{i}I)^{-1} \tag{5.2}$$

is bounded. Indeed, it is obvious that $K(L + \mathrm{i}I)^{-1}$ is a bounded operator if and only if

$$\|Kv\| \le \text{const}\,\|(L + \mathrm{i}I)v\|, \quad v \in \text{Dom}(L).$$

Clearly,

$$\|(L + \mathrm{i}I)v\| \le \|v\| + \|Lv\|.$$

It remains to observe that for $v \in \text{Dom}(L)$,

$$\begin{aligned}\|(L + \mathrm{i}I)v\|^2 &\ge |((L + \mathrm{i}I)v, (L + \mathrm{i}I)v)| \\ &= \|Lv\|^2 + \|v\|^2 - 2\,\text{Re}(\mathrm{i}(Lv, v)) \ge \|Lv\|^2 + \|v\|^2\end{aligned}$$

because $\text{Im}(Lv, v) \ge 0$.

If the operator C defined by (5.2) is a trace class operator, we say that K is a *relatively trace class perturbation of* L.

It is well known and easy to verify that if K is a relatively trace class perturbation of L, then it is a resolvent trace class perturbation of L, i.e.

$$(M + \mathrm{i}I)^{-1} - (L + \mathrm{i}I)^{-1} \in \boldsymbol{S}_1.$$

Indeed, this follows from the identity

$$(M + \mathrm{i}I)^{-1} - (L + \mathrm{i}I)^{-1} = (M + \mathrm{i}I)^{-1}K(L + \mathrm{i}I)^{-1} \in \boldsymbol{S}_1.$$

The converse is not true which can easily be seen from Theorems 8.1 and 8.3.

We are going to make three elementary (apparently well-known) observations.

Lemma 5.1. *Suppose that*

$$\|Kv\| \le c\|v\| + d\|Lv\|, \quad v \in \text{Dom}(L), \tag{5.3}$$

for some $c > 0$ and $d \in (0, 1)$ (in this case we say that K is strictly dominated by L). Then K is also dominated by $L + tK$ for $t \in (0, 1]$.

Proof. Indeed, let $v \in \mathrm{Dom}(L)$. We have

$$\begin{aligned}\|Kv\| \le c\|v\| + d\|Lv\| &= c\|v\| + d\|(L+tK)v - Kv\| \\ &\le c\|v\| + d\|(L+tK)v\| + dt\|Kv\|.\end{aligned}$$

It follows that

$$\|Kv\| \le (1-dt)^{-1}\big(c\|v\| + d\|(L+tK)v\|\big), \tag{5.4}$$

and so K is dominated by M. ■

We proceed now to the case when K is a *relatively compact perturbation of* L, i.e. the operator $K(A+\mathrm{i}I)^{-1}$ is compact.

Lemma 5.2. *Suppose that K is a relatively compact perturbation of L. Then for any positive number d, there exists a positive number c such that*

$$\|Kv\| \le c\|v\| + d\|Lv\|, \quad v \in \mathrm{Dom}(L),$$

and so, K is strictly dominated by $M = L + K$.

Proof. Let $\varepsilon > 0$. By the assumption, the operator $K(L+\mathrm{i}I)^{-1}$ is compact, and so it can be represented in the form $K(L+\mathrm{i}I)^{-1} = T + R$, where T is a finite rank operator and $\|R\| < \varepsilon$.

Consider the operators K_1 and K_2 defined on $\mathrm{Dom}(A)$ by

$$K_1 = T(L+\mathrm{i}I) \quad \text{and} \quad K_2 = R(L+\mathrm{i}I).$$

Clearly,

$$\|K_2 v\| = \|R(L+\mathrm{i}I)v\| \le \varepsilon\|(L+\mathrm{i}I)v\| \le \varepsilon\|Lv\| + \varepsilon\|v\|, \quad v \in \mathrm{Dom}(L),$$

and so, it suffices to show that for any positive number d, there exists a positive number c such that

$$\|K_1 v\| \le c\|v\| + d\|Lv\|, \quad v \in \mathrm{Dom}(L).$$

Since T is a linear combination of rank one operators, it suffices to consider the case when $\mathrm{rank}\, T = \mathrm{rank}\,\big(K_1(L+\mathrm{i}I)^{-1}\big) = 1$. Suppose that

$$K_1 v = \big((L+\mathrm{i}I)v, \varphi\big)\psi, \quad v \in \mathrm{Dom}(L).$$

since L is densely defined, φ admits a representation $\varphi = \varphi_1 + \varphi_2$, where $\varphi_1 \in \mathrm{Dom}(L)$ and $\|\varphi_2\| < \varepsilon\|\psi\|^{-1}$. We have

$$\begin{aligned}K_1 v &= \big((L+\mathrm{i}I)v, \varphi_1\big)\psi + \big((L+\mathrm{i}I)v, \varphi_2)\psi\big) \\ &= \big(v, (L+\mathrm{i}I)\varphi_1)\psi + \big((L+\mathrm{i}I)v, \varphi_2\big)\psi, \quad v \in \mathrm{Dom}(L),\end{aligned}$$

and so

$$\|Kv\| \le \|(L+\mathrm{i}I)\varphi_1\| \cdot \|\psi\| \cdot \|v\| + \|\varphi_2\| \cdot \|\psi\| \cdot \|v\| + \|\varphi_2\| \cdot \|\psi\| \cdot \|Lv\|$$

$$\le \big(\|(L+\mathrm{i}I)\varphi_1\| \cdot \|\psi\| + \|\varphi_2\| \cdot \|\psi\|\big)\|v\| + \varepsilon\|Lv\|, \quad v \in \mathrm{Dom}(L). \quad \blacksquare$$

Lemma 5.3. *Suppose that K is a relatively trace class perturbation of L. Then K is a relatively trace class perturbation of $M = L + K$.*

Proof. It is easy to see that

$$K(L+K+\mathrm{i}I)^{-1} - K(L+\mathrm{i}I)^{-1} = -K(L+K+\mathrm{i}I)^{-1}K(L+\mathrm{i}I)^{-1} \in \boldsymbol{S}_1.$$

Indeed, by the assumption, $K(L+\mathrm{i}I)^{-1} \in \boldsymbol{S}_1$. On the other hand, by Lemmata 5.1 and 5.2, the operator $K(L+K+\mathrm{i}I)^{-1}$ is bounded. $\blacksquare$

Lemma 5.4. *Suppose that L is a maximal dissipative operator and M is a dissipative operator such that* $\mathrm{Dom}(M) \subset \mathrm{Dom}(L)$ *and $K = M - L$ satisfies* (5.3) *with $d \in (0, \frac{1}{2})$. Then M is a maximal dissipative operator.*

Proof. Let $\varkappa > 0$. It is easy to see that

$$\|(L+\mathrm{i}\varkappa I)^{-1}\| \le \varkappa^{-1}.$$

We have

$$M + \mathrm{i}\varkappa I = K + L + \mathrm{i}\varkappa I = \big(I + K(L+\mathrm{i}\varkappa I)^{-1}\big)(L+\mathrm{i}\varkappa I).$$

Let us show that for some $\varkappa > 0$, $(M+\mathrm{i}\varkappa I)^{-1}$ is a bounded operator defined on the whole Hilbert space. It suffices to show that $\|K(L+\mathrm{i}\varkappa I)^{-1}\| < 1$. Let v be a vector in our Hilbert space. By (5.3), we have

$$\|K(L+\mathrm{i}\varkappa I)^{-1}v\| \le c\|(L+\mathrm{i}\varkappa I)^{-1}v\| + d\|L(L+\mathrm{i}\varkappa I)^{-1}v\|$$

Clearly,

$$L(L+\mathrm{i}\varkappa I)^{-1}v = v - \mathrm{i}\varkappa(L+\mathrm{i}\varkappa I)^{-1}v$$

Thus,

$$\|K(L+\mathrm{i}\varkappa I)^{-1}v\| \le c\varkappa^{-1}\|v\| + d\|v\| + d\|v\| = (2 + c\varkappa^{-1})d\|v\|.$$

It suffices to select $\varkappa > 0$ so that $(2 + c\varkappa^{-1})d < 1$. $\blacksquare$

Remark 1. We conjecture that the conclusion of Lemma 5.4 should hold under the weaker assumption $d < 1$, i.e. under the assumption that K is strictly dominated by L. However, even if it could happen that the dissipative operator $M = L + K$ is not maximal, we can consider instead of M its canonical maximal extension (see § 2). In this paper we consider one-parametric families $L_t = L + tK$, $0 \le t \le 1$. If

L_t is not maximal, we can consider instead its canonical maximal extension. We are going to keep the same notation L_t for the maximal extension.

Corollary 5.5. *Suppose that L, M and K are as in Lemma* 5.4 *and suppose that the operator K is a relatively compact perturbation of L. Then the conclusion of the lemma holds.*

Definition. A function f in $\mathrm{C}_{\mathrm{A},\infty}$ is called an *analytic relatively operator Lipschitz function* if there is a positive number k such that

$$\|f(M)-f(L)\|\le k\|(M-L)(L+\mathrm{i}I)^{-1}\| \tag{5.5}$$

whenever L and M are maximal dissipative operator operators such that $M-L$ is a relatively bounded perturbation of L. *We denote by* $\mathrm{ROL}_{\mathrm{A}}$ *the class of such analytic relatively operator Lipschitz function on* $\mathbb{R}$.

Note that in this definition it is natural to require that f should be bounded. Indeed, if f is a function defined on $\operatorname{clos}\mathbb{C}_+$ and we apply (5.5) to dissipative operators on the one-dimensional space $\mathbb{C}$, we obtain

$$|f(\mu)-f(\lambda)|\le\operatorname{const}|(\mu-\lambda)(\lambda+\mathrm{i})^{-1}|,\quad \lambda,\ \mu\in\operatorname{clos}\mathbb{C}_+,$$

which implies that f should be bounded.

Remark 2. Obviously, if $f\in\mathrm{ROL}_{\mathrm{A}}$, then f is operator Lipschitz, and so, f is differentiable on $\mathbb{R}$; see [AP3], Th. 3.3.3. Indeed, (5.5) implies that

$$\|f(M)-f(L)\|\le k\|(M-L)(L+\mathrm{i}I)^{-1}\|\le k\|M-L\|\cdot\|(L+\mathrm{i}I)^{-1}\|\le k\|M-L\|.$$

6. Analytic Schur multipliers

In this section we quote recent results of [AP9] about analytic Schur multipliers. Recall that $\mathfrak{M}(\mathbb{R}^2)$ is the class of functions on $\mathbb{R}^2$ that are Schur multipliers with respect to arbitrary Borel spectral measures on $\mathbb{R}$.

Definition. Let f be a function that is bounded and analytic in $\mathbb{C}_+\times\mathbb{C}_+$ and suppose that the limit

$$\lim_{r\to0^+}f(x_1+\mathrm{i}r,x_2+\mathrm{i}r)\stackrel{\mathrm{def}}{=}f(x_1,x_2)$$

exists for all $(x_1,x_2)\in\mathbb{R}^2$. We say that f is an *analytic Schur multiplier on* $\mathbb{C}_+\times\mathbb{C}_+$ if the boundary-value function f on $\mathbb{R}^2$ belongs to $\mathfrak{M}(\mathbb{R}^2)$. We denote the class of analytic Schur multipliers by $\mathfrak{M}_{\mathrm{A}}$.

We need the following two results from [AP9] on analytic Schur multipliers (see Th. 2.1 and Th. 2.2):

I. Put

$$\mathrm{C}_{\mathrm{A},\infty}(\mathbb{C}_+^2)=\big\{f\in H^\infty(\mathbb{C}_+^2):\ f(z,\cdot),\quad f(\cdot,z)\in\mathrm{C}_{\mathrm{A},\infty}\quad\text{for all}\quad z\in\operatorname{clos}\mathbb{C}_+\big\}.$$

Suppose that $\Phi\in\mathrm{C}_{\mathrm{A},\infty}(\mathbb{C}_+^2)$ and $\Phi\in\mathfrak{M}_{\mathrm{A}}$. Then Φ belongs to the Haagerup tensor product $\mathrm{C}_{\mathrm{A},\infty}\otimes_{\mathrm{h}}\mathrm{C}_{\mathrm{A},\infty}$. Moreover,

$$\|\Phi\|_{\mathrm{C}_{\mathrm{A},\infty}(\mathbb{C}_+)\otimes_{\mathrm{h}}\mathrm{C}_{\mathrm{A},\infty}(\mathbb{C}_+)}=\big\|\Phi\big|\mathbb{R}^2\big\|_{\mathfrak{M}(\mathbb{R}^2)}.$$

II. Let

$$\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)\stackrel{\text{def}}{=}\big\{f\in\mathrm{C}_{\mathrm{A},\infty}:\ \text{the limit}\ \lim_{|\zeta|\to\infty}f(\zeta)\ \text{exists}\big\}\tag{6.1}$$

and

$$\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+^2)=\big\{f\in H^\infty(\mathbb{C}_+^2):\ f(z,\cdot),\quad f(\cdot,z)\in\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)\quad\text{for all}\quad z\in\operatorname{clos}\mathbb{C}_+\big\}.$$

Suppose that $\Phi\in\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+^2)$ and $\Phi\in\mathfrak{M}_{\mathrm{A}}$. Then Φ belongs to the Haagerup tensor product $\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)\otimes_{\mathrm{h}}\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$. Moreover,

$$\|\Phi\|_{\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)\otimes_{\mathrm{h}}\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)}=\big\|\Phi\big|\mathbb{R}^2\big\|_{\mathfrak{M}(\mathbb{R}^2)}.$$

7. A representation of operator differences in terms of double operator integrals

In this section we obtain a representation of the operator differences $f(M)-f(L)$ in terms of double operator integrals in the case when the function $\mathfrak{D}_{\mathfrak{b}}f$ on $\mathbb{R}^2$ defined by

$$(\mathfrak{D}_{\mathfrak{b}}f)(x,y)\stackrel{\text{def}}{=}\begin{cases}\dfrac{f(x)-f(y)}{x-y}(y+\mathrm{i}), & x\ne y,\\ f'(x)(x+\mathrm{i}), & x=y,\end{cases}\tag{7.1}$$

is a Schur multiplier with respect to arbitrary spectral (or semi-spectral) measures. Recall that the latter means that the function $\mathfrak{D}_{\mathfrak{b}}f$ belongs to the Haagerup tensor product $\mathrm{C}_{\mathrm{A},\infty}\otimes_{\mathrm{h}}\mathrm{C}_{\mathrm{A},\infty}$.

Note that similar results for functions of self-adjoint operators were obtained in [AP7] and [AP8].

Theorem 7.1. *Let L and M be maximal dissipative operators with $M-L$ being a relatively bounded perturbation of L. Suppose that f is a Lipschitz function in* $\operatorname{clos}\mathbb{C}_+$, *analytic in $\mathbb{C}_+$ and such that*

$$\mathfrak{D}_\flat f \in \mathrm{C}_{\mathrm{A},\infty} \otimes_{\mathrm{h}} \mathrm{C}_{\mathrm{A},\infty}.$$

Then the following formula holds

$$f(M)-f(L)=\iint_{\mathbb{R}\times\mathbb{R}} \frac{f(x)-f(y)}{x-y}(y+\mathrm{i})\,\mathrm{d}\mathscr{E}_M(x)(M-L)(L+\mathrm{i}I)^{-1}\,\mathrm{d}\mathscr{E}_L(y), \tag{7.2}$$

and so f must be relatively operator Lipschitz.

Proof. It follows from Lemma 8.3 of [AP8] and from **II** of § 6 that there are sequences $\{\varphi_n\}_{n\ge0}$ and $\{\psi_n\}_{n\ge0}$ of functions in the space $\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$ (see (6.1)) such that the following statements hold:

(a) $\sum\limits_{n\ge0}|\varphi_n|^2 \le$ const everywhere on $\mathbb{C}_+$;

(b) $\sum\limits_{n\ge0}|\psi_n|^2 \le$ const everywhere on $\mathbb{C}_+$;

(c)

$$(w+\mathrm{i})(\mathfrak{D}f)(z,w)=\sum_{n\ge0}\varphi_n(w)\psi_n(z) \quad z,\ w\in\operatorname{clos}\mathbb{C}_+, \tag{7.3}$$

where the divided difference $\mathfrak{D}f$ is defined by

$$(\mathfrak{D}f)(z,w)\stackrel{\text{def}}{=}\begin{cases}\dfrac{f(z)-f(w)}{z-w}, & z\ne w\\[2mm] f'(w), & z=w.\end{cases}$$

We have

$$\iint_{\mathbb{R}\times\mathbb{R}}(\mathfrak{D}_\flat f)(x,y)\,\mathrm{d}E_M(x)(M-L)(L+\mathrm{i}I)^{-1}\,\mathrm{d}E_L(y)$$
$$=\sum_{n\ge0}\varphi_n(M)(M-L)(L+\mathrm{i}I)^{-1}\psi_n(L)\stackrel{\text{def}}{=}Q, \tag{7.4}$$

where Q is a bounded linear operator.

Clearly,

$$\iint\limits_{\mathbb{R}\times\mathbb{R}}\big(f(y)(x+\mathrm{i})^{-1}-f(x)(x+\mathrm{i})^{-1}\big)\,\mathrm{d}\mathscr{E}_M(y)\mathrm{d}\mathscr{E}_L(x)=\big(f(M)-f(L)\big)(L+\mathrm{i}I)^{-1}. \tag{7.5}$$

On the other hand, for $z,\ w \in \operatorname{clos} \mathbb{C}_+$, we have

$$\begin{aligned} f(w)(z+\mathrm{i})^{-1} - f(z)(z+\mathrm{i})^{-1} &= \sum_n \varphi_n(w)(w-z)(w+\mathrm{i})^{-1}(z+\mathrm{i})^{-1}\psi_n(z) \\ &= \sum_n \varphi_n(w)w(w+\mathrm{i})^{-1}(z+\mathrm{i})^{-1}\psi_n(z) \\ &\quad - \sum_n \varphi_n(w)(w+\mathrm{i})^{-1}z(z+\mathrm{i})^{-1}\psi_n(z) \\ &\stackrel{\text{def}}{=} \Omega(z, w). \end{aligned} \tag{7.6}$$

Thus,

$$\begin{aligned} &\iint_{\mathbb{R}\times\mathbb{R}} \sum_n \varphi_n(y)y(y+\mathrm{i})^{-1}(x+\mathrm{i})^{-1}\psi_n(x)\,\mathrm{d}\mathscr{E}_M(y)\,\mathrm{d}\mathscr{E}_L(x) \\ &\qquad = \sum_n \varphi_n(M)\big(M(M+\mathrm{i}I)^{-1}\big)(L+\mathrm{i}I)^{-1}\psi_n(L) \\ &\qquad\qquad = \sum_n \varphi_n(M)\big(I-\mathrm{i}(M+\mathrm{i}I)^{-1}\big)(L+\mathrm{i}I)^{-1}\psi_n(L). \end{aligned}$$

Similarly,

$$\begin{aligned} &\iint_{\mathbb{R}\times\mathbb{R}} \left(\sum_n \varphi_n(y)(y+\mathrm{i})^{-1}x(x+\mathrm{i})^{-1}\psi_n(x)\right) \mathrm{d}\mathscr{E}_M(y)\mathrm{d}\mathscr{E}_L(x) \\ &\qquad\qquad = \sum_n \varphi_n(M)(M+\mathrm{i}I)^{-1}(I-\mathrm{i}(L+\mathrm{i}I)^{-1})\psi_n(L). \end{aligned}$$

It follows that

$$\begin{aligned} \iint_{\mathbb{R}\times\mathbb{R}} \Omega(x, y)\,\mathrm{d}\mathscr{E}_M(y)\mathrm{d}\mathscr{E}_L(x) &= \sum_n \varphi_n(M)(I-(M+\mathrm{i}I)^{-1})(L+\mathrm{i}I)^{-1}\psi_n(L) \\ &\quad - \sum_n \varphi_n(M)(M+\mathrm{i}I)^{-1}(I-(L+\mathrm{i}I)^{-1})\psi_n(L) \\ &= \sum_n \varphi_n(M)\big((L+\mathrm{i}I)^{-1}-(M+\mathrm{i}I)^{-1}\big)\psi_n(L) \\ &= \sum_n \varphi_n(M)(M+\mathrm{i}I)^{-1}(M-L)(L+\mathrm{i}I)^{-1}\psi_n(L), \end{aligned}$$

where the function Ω is defined in (7.6). Combining this equality with (7.6) and (7.5), we find that

$$\big(f(M)-f(L)\big)(L+\mathrm{i}I)^{-1}=\sum_n \varphi_n(M)(M+\mathrm{i}I)^{-1}(M-L)(L+\mathrm{i}I)^{-1}\psi_n(L).$$

Multiplying this identity on the right by $(L+\mathrm{i}I)$, we obtain

$$f(M)-f(L)=\sum_n \varphi_n(M)(M+\mathrm{i}I)^{-1}(M-L)\psi_n(L)=Q$$

on $\operatorname{Range}(L+\mathrm{i}I)=\operatorname{Range}L$ and since the operators on the left and on the right are bounded, it follows that the equality holds on the whole Hilbert space. ■

Corollary 7.2. *Suppose that f satisfies the hypotheses of Theorem* 7.1. *Let L and M be maximal dissipative operators such that $M-L$ is a relatively trace class perturbation of L. Then $f(M)-f(L)\in \boldsymbol{S}_1$.*

The converse of Theorem 7.1 is also true. Indeed, the following result holds.

Theorem 7.3. *Let $f\in \mathrm{ROL_A}$. Then $\mathfrak{D}_\flat f\in \mathrm{C_A}(\mathbb{C}_+)\otimes_{\mathrm{h}}\mathrm{C_A}(\mathbb{C}_+)$, and so* (7.2) *holds whenever L and M are maximal dissipative operators such that $M-L$ is a relatively bounded perturbation of L.*

Proof. By Theorem 6.1 of [AP8], $\mathfrak{D}_\flat f\in\mathfrak{M}(\mathbb{R}^2)$, and so by **II** of § 6 of this paper, the function $\mathfrak{D}_\flat f$ belongs to $\mathrm{C_A}(\mathbb{C}_+)\otimes_{\mathrm{h}}\mathrm{C_A}(\mathbb{C}_+)$ which implies the result. ■

8. Various characterizations of analytic relatively operator Lipschitz functions

Recall that $\mathrm{OL_A}$ denote the space of operator Lipschitz functions on $\operatorname{clos}\mathbb{C}_+$ analytic in $\mathbb{C}_+$. Note that $\mathrm{OL_A}=\mathrm{CL}(\mathbb{C}_+)$, the space of commutator Lipschitz functions on $\mathbb{C}_+$; see [AP3]. Note that if f is a Lipschitz function f on $\operatorname{clos}\mathbb{C}_+$ that is analytic on $\mathbb{C}_+$, then space $f\in\mathrm{OL}_A$ if and only if its restriction $f\big|\mathbb{R}$ is an operator Lipschitz function on $\mathbb{R}$.

We say that a function f on $\operatorname{clos}\mathbb{C}_+$ is a multiplier of the space $\mathrm{OL_A}$ if

$$\varphi\in\mathrm{OL_A}\quad\Longrightarrow\quad\varphi f\in\mathrm{OL_A}. \tag{8.1}$$

We denote by $\mathfrak{M}_{\mathrm{OL}_A}$ the class of all multipliers of $\mathrm{OL_A}$. By the norm $\|\varphi\|_{\mathfrak{M}_{\mathrm{OL}_A}}$ of a function φ in $\mathfrak{M}_{\mathrm{OL}_A}$ we mean the norm of the multiplication operator $\varphi\mapsto\varphi f$ on the space $\mathrm{OL_A}$.

Theorem 8.1. *Let f be a continuous function on* $\operatorname{clos}\mathbb{C}_+$ *that is analytic in* $\mathbb{C}_+$ *and differentiable on* $\mathbb{R}$. *The following are equivalent*:

(a) *f is an analytic relatively operator Lipschitz function*;

(b) $(z+\mathrm{i})(\mathfrak{D}f)(z,w)\in\mathfrak{M}_{\mathrm{A}}$;
(c) $(z+\mathrm{i})f(z)\in\mathrm{OL}_{\mathrm{A}}$;
(d) $f\in\mathfrak{M}_{\mathrm{OL}_A}$.

Proof. The implication (b)⇒(a) follows from Theorem 7.1. The implication (a)⇒(b) follows from Theorem 7.3.

It remains to prove that (b)⇔(c)⇔(d).

Clearly, (b) implies the condition

(b′) $(x+\mathrm{i})(\mathfrak{D}f)(x,y)\in\mathfrak{M}(\mathbb{R}^2)$,

and (*c*) implies the condition

(c′) $(x+\mathrm{i})f(x)\in\mathrm{OL}(\mathbb{R})$.

The implication (b′) ⇒(b) follows from the results of [AP9]. The implication (c′) ⇒(c) is a consequence of Theorem 3.9.5 in [AP3].

The equivalence (b)⇔(c) follows now from the equivalence (b′) ⇔ (c′), which was proved in Theorem 4.1 of [AP8].

The implication (d)⇒(c) is trivial.

It remains to prove that (c)⇒(d). Let $(z+\mathrm{i})f(z)\in\mathrm{OL}_{\mathrm{A}}(\mathbb{C}_+)$. Then $(x+\mathrm{i})f(x)\in\mathrm{OL}(\mathbb{R})$, and $f\in\mathfrak{M}_{\mathrm{OL}}$ by Theorem 4.1 in [AP8]. Thus,

$$\varphi\in\mathrm{OL}(\mathbb{R})\quad\Longrightarrow\quad\varphi f\in\mathrm{OL}(\mathbb{R}).$$

To complete the proof, we observe that f is continuous on $\operatorname{clos}\mathbb{C}_+$ and analytic on $\mathbb{C}_+$. ■

Corollary 8.2. *Let f be a function of class $\mathrm{C}_{\mathrm{A},\infty}$. Suppose that the function*

$$x\mapsto(x+\mathrm{i})f(x),\quad x\in\mathbb{R},$$

belongs to the Besov class $B^1_{\infty,1}(\mathbb{R})$. Then f is an analytic relatively operator Lipschitz function.

Proof. This is an immediate consequence of Theorem 8.1 and the fact that the Besov class $B^1_{\infty,1}(\mathbb{R})$ consists of operator Lipschitz functions; see [Pe1] and [Pe3]. ■

Let us proceed now to various descriptions of the class of analytic resolvent Lipschitz functions.

Theorem 8.3. *Let f be a continuous function on* $\operatorname{clos}\mathbb{C}_+$ *that is analytic in* $\mathbb{C}_+$ *and differentiable on* $\mathbb{R}$. *The following are equivalent*:

(a) $f\in\mathrm{OL}^{\mathrm{res}}_{\mathrm{A}}$, *i.e. f is an analytic resolvent Lipschitz function*;
(b) $(z+\mathrm{i})(w+\mathrm{i})(\mathfrak{D}f)(z,w)\in\mathfrak{M}_{\mathrm{A}}(\mathbb{C}^2_+)$;
(c) $(z+\mathrm{i})^2(f(z)-c)\in\mathrm{OL}_{\mathrm{A}}(\mathbb{C}_+)$ *for some* $c\in\mathbb{C}$;
(d) $(z+\mathrm{i})(f(z)-c)\in\mathfrak{M}_{\mathrm{OL}_A}$ *for some* $c\in\mathbb{C}$.

(e) *the function* φ *on* $\operatorname{clos}\mathbb{D}\setminus\{1\}$ *defined by*

$$\varphi(\zeta)\stackrel{\text{def}}{=}f\left(\mathrm{i}\frac{1+\zeta}{1-\zeta}\right),\quad \zeta\in\operatorname{clos}\mathbb{D}\setminus\{1\},\tag{8.2}$$

extends to a function of class $\mathrm{OL}_{\mathrm{A}}(\mathbb{D})$ *of operator Lipschitz function on* $\operatorname{clos}\mathbb{D}$ *analytic in* $\mathbb{D}$.

If $c\in\mathbb{C}$ *satisfies* (c) *or* (d), *then* c *must be equal to* $\lim_{|z|\to\infty}f(z)$

Proof. First we prove that (a)⇒(e). By (a), we have

$$\|f(M)-f(L)\|\le\operatorname{const}\|(M+\mathrm{i}I)^{-1}-(L+\mathrm{i}I)^{-1}\|\tag{8.3}$$

for all maximal dissipative operators L and M.

Considering dissipative operators on the one-dimensional space, we find that (8.3) implies that

$$|f(w)-f(z)|\le\operatorname{const}|(w+\mathrm{i})^{-1}-(z+\mathrm{i})^{-1}|,\quad z,\ w\in\operatorname{clos}\mathbb{C}_+,$$

and so, the limit $\lim_{|z|\to\infty}f(z)$ exists. Thus, the function φ defined by (8.2) extends to a function in the disc-algebra.

Let U and V be the Cayley transforms of L and M, i.e.

$$U\stackrel{\text{def}}{=}(L-\mathrm{i}I)(L+\mathrm{i}I)^{-1}\quad\text{and}\quad V\stackrel{\text{def}}{=}(M-\mathrm{i}I)(M+\mathrm{i}I)^{-1}.\tag{8.4}$$

Clearly, U and V are unitary operators if L and M are self-adjoint operators. Now applying (8.3) to the bounded self-adjoint operators L and M, we find that

$$\|\varphi(V)-\varphi(U)\|\le\operatorname{const}\|V-U\|\tag{8.5}$$

for all unitary operators U and V with $1\notin\sigma(U)$ and $1\notin\sigma(V)$.

To obtain (8.5) for arbitrary unitary operators U and V, we can construct two sequences of unitary operators $\{U_n\}_{n\ge1}$ and $\{V_n\}_{n\ge1}$ such that $1\notin\sigma(U_n)$ and $1\notin\sigma(V_n)$ for all $n\ge1$, $U_nU=UU_n$ and $V_nV=VV_n$ for all $n\ge1$, $\lim_{n\to\infty}\|U_n-U\|=0$ and $\lim_{n\to\infty}\|V_n-V\|=0$.

Let us prove that (e)⇒(a). By Theorem 3.9.9 of [AP3], inequality (8.5) holds for arbitrary contractions U and V. We want to prove inequality (8.3) for the maximal dissipative operators L and M. It remains to observe that inequality (8.3) is equivalent to inequality (8.5) for the contractions $U=(L-\mathrm{i}I)(L+\mathrm{i}I)^{-1}$ and $V=(M-\mathrm{i}I)(M+\mathrm{i}I)^{-1}$.

The equivalence of (e) and (c) follows from Theorem 3.10.2 in [AP3].

In what follows we need the following identity:

$$\frac{(z+\mathrm{i})^2 f(z) - (w+\mathrm{i})^2 f(w)}{z-w} = (z+\mathrm{i})(w+\mathrm{i})(\mathfrak{D}f)(z,w) + (z+\mathrm{i}) f(z) + (w+\mathrm{i}) f(w). \tag{8.6}$$

Let us show that (b)⇒(c).

To show (c), it suffices by Theorem 7.1 to show that the left-hand side of (8.6) belongs to $\mathfrak{M}_{\mathrm{A}}(\mathbb{C}_+^2)$. Thanks to (8.6) and our assumption that $(z+\mathrm{i})(w+\mathrm{i})(\mathfrak{D}f)(z,w) \in \mathfrak{M}_{\mathrm{A}}(\mathbb{C}_+^2)$, it suffices to show that $(z+\mathrm{i})f(z)$ is bounded. Clearly, $(z+\mathrm{i})(\mathfrak{D}f)(z,w) \in \mathfrak{M}_{\mathrm{A}}(\mathbb{C}_+^2)$. It follows from the previous theorem that $(z+\mathrm{i})f(z) \in \mathrm{OL}_{\mathrm{A}}(\mathbb{C}_+)$. Thus, the limit $c \stackrel{\mathrm{def}}{=} \lim_{|z|\to\infty} f(z)$ exists. Without loss of generality we may assume that $c = 0$. The boundedness of the function $(z+\mathrm{i})f(z)$ follows now from the fact that $\lim_{|z|\to+\infty} f(x) = 0$ and from the fact that the function $(z+\mathrm{i})^2(\mathfrak{D}f)(z,z) = (z+\mathrm{i})^2 f'(z)$ is bounded.

Let us prove that (c)⇒(b). Suppose that $(z+\mathrm{i})^2(f-c) \in \mathrm{OL}_{\mathrm{A}}(\mathbb{R})$ for some $c \in \mathbb{C}$. Let us show that $(z+\mathrm{i})(w+\mathrm{i})(\mathfrak{D}f)(z,w) \in \mathfrak{M}(\mathbb{C}_+^2)$. Again, we may assume that $c = 0$. The result follows from the fact that the function $(z+\mathrm{i})f(z)$ is bounded and from identity (8.6).

The equivalence of (c) and (d) is a consequence of Theorem 8.1.

Finally, it is obvious that (c) can hold only for one value of c. The same is true for (d). ■

9. Differentiaton in the parameter

As before, L is a maximal dissipative operator and M is a dissipative operator such that $K = M - L$ is a relatively bounded perturbation of L. In this section for a function for a relatively operator function f analytic in $\mathbb{C}_+$ we consider the problem of the differentiability of the map $t \mapsto f(L+tK)$. Obviously, for sufficiently small positive t, the operator tK is strictly dominated by L.

Theorem 9.1. *Let L be a maximal dissipative operator and let M be a dissipative operator such that $K = M - L$ is a relatively bounded perturbation of L. Suppose that $f \in \mathrm{ROL}_{\mathrm{A}}$. Then the function $t \mapsto f(L_t) \stackrel{\mathrm{def}}{=} f(L+tK)$ is differentiable at 0 in the strong operator topology and*

$$\frac{\mathrm{d}}{\mathrm{d}t} f(L_t)\Big|_{t=0} = \iint_{\mathbb{R}\times\mathbb{R}} \frac{f(x)-f(y)}{x-y}(y+\mathrm{i})\,\mathrm{d}\mathscr{E}_L(x) C\,\mathrm{d}\mathscr{E}_L(y) \tag{9.1}$$

where the operator C and is defined by (5.2).

Proof. By Theorem 7.3 the function $\mathfrak{D}_\flat f$ belongs to the Haagerup tensor product $\mathrm{C_A}(\mathbb{C}_+) \otimes_\mathrm{h} \mathrm{C_A}(\mathbb{C}_+)$, and so there are sequences φ_n and ψ_n in $\mathrm{C_A}(\mathbb{C}_+)$ such that

$$\sup_{z\in\operatorname{clos}\mathbb{C}_+} \sum_n |\varphi_n(z)|^2 < \infty, \qquad \sup_{z\in\operatorname{clos}\mathbb{C}_+} \sum_n |\psi_n(z)|^2 < \infty$$

and

$$\mathfrak{D}_\flat f(z, w) = \sum_n \varphi_n(z)\psi_(w), \quad z,\ w \in \operatorname{clos}\mathbb{C}_+.$$

Let $K = M - L$. Consider the parametric family $L_t = L + tK$, where t is a sufficiently small positive number. Then L_t is a dissipative operator. Denote by $\mathscr{E}_t$ the semi-spectral measure of L_t. We have

$$\begin{aligned}\frac{1}{t}(f(L_t) - f(L)) &= \iint_{\mathbb{R}\times\mathbb{R}} \frac{f(x) - f(y)}{x - y}(y + \mathrm{i})\,\mathrm{d}\mathscr{E}_t(x)K(L + \mathrm{i}I)^{-1}\,\mathrm{d}\mathscr{E}_0(y)\\ &= \sum_{n\ge 0} \varphi_n(L_t)K(L + \mathrm{i}I)^{-1}\psi_n(L),\end{aligned}$$

We have

$$\lim_{t\to\mathbf{0}} \|\varphi_n(L_t) - \varphi_n(L)\| = 0. \tag{9.2}$$

This follows from the fact that function φ in the class $\mathrm{C_A}(\mathbb{C}_+)$ is uniformly operator continuous, i.e. for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|\varphi_n(L) - \varphi_n(M)\| < \varepsilon \tag{9.3}$$

whenever M is a maximal dissipative operator satisfying $\|L - M\| < \delta$. The latter is a consequence of Theorem 7.2 of [AP2] and an analog of Theorem 8.1 of [AP1] for maximal dissipative operators. Note that the proof of Theorem 8.1 of [AP1] extends very easily to the case of maximal dissipative operators if we use the results of § 7 of [AP2].

We have to prove that

$$\lim_{t\to 0} \sum_{n\ge 0} \varphi_n(L_t)K(L + \mathrm{i}I)^{-1}\psi_n(L) = \sum_{n\ge 0} \varphi_n(L)K(L + \mathrm{i}I)^{-1}\psi_n(L)$$

in the strong operator topology. Thus, we need to prove that for an arbitrary vector u,

$$\lim_{t\to 0} \sum_{n\ge 0} (\varphi_n(L_t) - \varphi_n(L))K(L + \mathrm{i}I)^{-1}\psi_n(L)u = \mathbf{0}$$

where the series is understood in the sense of the weak topology of $\mathscr{H}$ while the limit is taken in the norm of $\mathscr{H}$. Assume that $\|u\| = 1$. Let $u_n \stackrel{\text{def}}{=} K(L + \mathrm{i}I)^{-1}\psi_n(L)u$.

We have

$$\begin{aligned}\sum_{n\ge0}\|u_n\|^2 &\le \|K(L+\mathrm{i}I)^{-1}\|^2\sum_{n\ge0}\|(\psi_n(L))u\|^2\\ &\le \|K(L+\mathrm{i}I)^{-1}\|^2\sum_{n\ge0}\|(\psi_n(A)u\|^2\\ &= \|K(L+\mathrm{i}I)^{-1}\|^2\sum_{n\ge0}(|\psi_n|^2(A)u,u)\\ &\le \sup_{z\in\operatorname{clos}\mathbb{C}_+}\left(\sum_n|\psi_n(z)|^2\right)\|K(L+\mathrm{i}I)^{-1}\|^2\end{aligned}$$

where A is a resolvent self-adjoint dilation of L.

Let $\varepsilon>0$. We can choose a natural number N such that $\sum_{n>N}\|u_n\|^2<\varepsilon^2$. Then it follows from Lemma 3.5.9 in [AP3] that

$$\begin{aligned}\left\|\sum_{n>N}(\varphi_n(L_t)-\varphi_n(L))u_n\right\| &\le \left\|\sum_{n>N}\varphi_n(L_t)u_n\right\|\\ &\quad+\left\|\sum_{n>N}\varphi_n(L)u_n\right\|<2\varepsilon\sup_{z\in\operatorname{clos}\mathbb{C}_+}\left(\sum_n|\varphi_n(z)|^2\right)^{1/2}\end{aligned}$$

for all $t\in[0,1]$. By (9.2),

$$\left\|\sum_{n=0}^{N}(\varphi_n(A+tK)-\varphi_n(A))u_n\right\|<\varepsilon$$

for all t sufficiently close to zero. It follows that

$$\left\|\sum_{n\ge0}(\varphi_n(L_t)-\varphi_n(L))u_n\right\|<\left(2\sup_{z\in\operatorname{clos}\mathbb{C}_+}\left(\sum_n|\varphi_n(z)|^2\right)^{1/2}+1\right)\varepsilon$$

for all t sufficiently close to zero. ■

Theorem 9.2. *Let L be a maximal dissipative operator and let M be a dissipative operator such that $K=M-L$ is strictly dominated by L. Suppose that $f\in\mathrm{ROL}_{\mathrm{A}}$. Then for $t\in(0,1]$, the following identity holds*

$$\frac{\mathrm{d}}{\mathrm{d}s}f(L_s)\Big|_{s=t}=\iint_{\mathbb{R}\times\mathbb{R}}\frac{f(x)-f(y)}{x-y}(y+\mathrm{i})\,\mathrm{d}\mathscr{E}_{L_t}(x)K(L_t+\mathrm{i}I)^{-1}\,\mathrm{d}E_{L_t}(y),$$

in the strong operator topology, where $L_t\stackrel{\mathrm{def}}{=}L+tK$.

Proof. For $t<1$, the formula for the right derivative is an immediate consequence of Theorem 9.1 and Lemma 5.1. On the other hand, for $t\le1$, the formula for the right derivative can be obtained in exactly the same way. ■

10. The trace formula in the case of relatively trace class perturbations

In this section we obtain a trace formula for the trace of $f(M)-f(L)$, where L is a maximal dissipative operator and M is a dissipative operator such that $M-L$ is a relatively trace class perturbation of L. We prove that the pair (L, M) has a spectral shift function ξ such that

$$\int_{\mathbb{R}} \frac{|\xi(t)|}{1+|t|}\,\mathrm{d}t<\infty.$$

We also show in this section that the maximal class of functions f, for which the trace formula holds with arbitrary pairs (L, M) of maximal dissipative operators with $M-L$ being a relatively trace class perturbation of L coincides with $\mathrm{ROL}_{\mathrm{A}}$.

Put $L_t \stackrel{\mathrm{def}}{=} L+t(M-L)$, $0\le t\le 1$. Since a relatively trace class perturbation is relatively compact, it follows from Lemmata 5.1, 5.2, 5.3 and 5.4 that $M-L$ is a relatively trace class perturbation of L_t and L_t is a maximal dissipative operator.

Theorem 10.1. *Let L be a maximal dissipative operator and let M be a dissipative operator such that $M-L$ is a relatively trace class perturbation of L. Then there exists a measurable function ξ such that*

$$\int_{\mathbb{R}} \frac{|\xi(t)|}{1+|t|}\,\mathrm{d}t<\infty \tag{10.1}$$

and the trace formula

$$\operatorname{trace}\big(f(M)-f(L)\big)=\int_{\mathbb{R}} f'(t)\xi(t)\,\mathrm{d}t \tag{10.2}$$

holds for an arbitrary function f in $\mathrm{ROL}_{\mathrm{A}}$. Moreover, $\mathrm{ROL}_{\mathrm{A}}$ is the maximal class of functions f for which trace formula (10.2) *holds for all such pairs (L, M).*

A function ξ on $\mathbb{R}$ is called a *spectral shift function for the pair* (L, M) if it satisfies (10.1) and (10.2).

We need the following lemma, which is similar to Lemma 9.2 in [AP8].

Lemma 10.2. *The function $t\mapsto K(L_t+\mathrm{i}I)^{-1}$ is continuous on* $[0, 1]$ *in the norm of $\boldsymbol{S}_1$, and so $\|K(L_t+\mathrm{i}I)^{-1}\|_{\boldsymbol{S}_1}\le \mathrm{const}$, $0\le t\le 1$.*

Proof. Note first that $(L+\mathrm{i}I)(L_t+\mathrm{i}I)^{-1}$ is a bounded operator. Indeed, since $L=L_t-tK$, we have

$$(L+\mathrm{i}I)(L_t+\mathrm{i}I)^{-1}=I-tK(L_t+\mathrm{i}I)^{-1}$$

and the operator $K(L_t+\mathrm{i}I)^{-1}$ is bounded because by Lemma 5.3, K is a relatively bounded perturbation of L_t.

We have

$$K(L_t+\mathrm{i}I)^{-1}=K(L+\mathrm{i}I)^{-1}(L+\mathrm{i}I)(L_t+\mathrm{i}I)^{-1}.$$

Clearly, it suffices to show that the function

$$t \mapsto \big((L+\mathrm{i}I)(L_t+\mathrm{i}I)^{-1}\big)^{-1} = (L_t+\mathrm{i}I)(L+\mathrm{i}I)^{-1} \tag{10.3}$$

is continuous. This follows immediately from the following obvious equality

$$(L_t+\mathrm{i}I)(L+\mathrm{i}I)^{-1} = I + tK(L+\mathrm{i}I)^{-1}.$$

Note that the operator $(L+\mathrm{i}I)(L_t+\mathrm{i}I)^{-1}$ is invertible and its inverse is the operator on the right of (10.3). ■

Proof of Theorem 10.1. Consider the one-parameter family of operators $L_t \stackrel{\text{def}}{=} L+tK$, $0 \le t \le 1$.

By Theorem 9.1,

$$\frac{\mathrm{d}}{\mathrm{d}s}\big(f(L_s)-f(L_t)\big)\Big|_{s=t} = \iint_{\mathbb{R}\times\mathbb{R}} \frac{f(x)-f(y)}{x-y}(y+\mathrm{i})\,\mathrm{d}\mathscr{E}_{L_t}(x)K(L_t+\mathrm{i}I)^{-1}\,\mathrm{d}\mathscr{E}_{L_t}(y), \tag{10.4}$$

where the derivative on the left is understood in the strong operator topology.

Let Q_t be the operator on the right-hand side of (10.4). Since by Theorem 7.3, the function $\mathfrak{D}_\flat f$ is a Schur multiplier for arbitrary semi-spectral measures, it follows that $Q_t \in \boldsymbol{S}_1$ for every t in $[0, 1]$ and $\sup_t \|Q_t\|_{\boldsymbol{S}_1} < \infty$; see Lemma 10.2.

Since the function $t \mapsto Q_t$ is continuous in the norm of $\boldsymbol{S}_1$ by Lemma 10.2, we obtain

$$f(L+K)-f(L) = \int_0^1 Q_t\,\mathrm{d}t$$

in the sense of integration of continuous functions and

$$\operatorname{trace}\big(f(L+K)-f(L)\big) = \int_0^1 \operatorname{trace} Q_t\,\mathrm{d}t.$$

We need the following lemma.

Lemma 10.3. *Let L be a maximal dissipative operator and let R be a trace class operator. Suppose that $f \in \mathrm{ROL}_{\mathrm{A}}$. Then*

$$\operatorname{trace}\iint_{\mathbb{R}\times\mathbb{R}} (\mathfrak{D}_\flat f)(x,y)\mathrm{d}\mathscr{E}_L(x)R\,\mathrm{d}\mathscr{E}_L(y) = \int_{\mathbb{R}} f'(x)(x+\mathrm{i})\,\mathrm{d}\nu(x),$$

where ν is the complex Borel measure on $\mathbb{R}$ defined by $\nu(\Delta) = \operatorname{trace}(\mathscr{E}_L(\Delta)R)$ for a Borel subset Δ of $\mathbb{R}$.

Proof. Let $\{\varphi_n\}_{n\ge 0}$ and $\{\psi_n\}_{n\ge 0}$ be sequences of functions in $\mathrm{C_A}(\mathbb{C}_+)$ that satisfy conditions a), b) and c) of Theorem 7.1. We have

$$\operatorname{trace}\iint_{\mathbb{R}\times\mathbb{R}}(\mathfrak{D}_\flat f)(x,y)\mathrm{d}\mathscr{E}_L(x)R\,\mathrm{d}\mathscr{E}_L(y)=\sum_n\operatorname{trace}\varphi_n(L)R\varphi_n(L)$$

$$=\operatorname{trace}\left(R\sum_n(\varphi_n\psi_n)(L)\right)=\operatorname{trace}\left(R\int_{\mathbb{R}}\left(\sum_n(\varphi_n\psi_n)\right)(x)\,\mathrm{d}\mathscr{E}_L(x)\right)$$

$$=\operatorname{trace}\left(R\int_{\mathbb{R}}(\mathfrak{D}_\flat f)(x,x)\,\mathrm{d}\mathscr{E}_L(x)\right)=\operatorname{trace}\left(R\int_{\mathbb{R}}f'(x)(x+\mathrm{i})\,\mathrm{d}\mathscr{E}_L(x)\right)$$

$$=\int_{\mathbb{R}}f'(x)(x+\mathrm{i})\,\mathrm{d}\nu(x).\quad\blacksquare$$

Let us return to the proof of Theorem 10.1.

We apply Lemma 10.3 to the dissipative operator L_t and the trace class operator $K(L+\mathrm{i})^{-1}$ and find that

$$\operatorname{trace}Q_t=\int_{\mathbb{R}}f'(x)(x+\mathrm{i})\,\mathrm{d}\nu_t(x),$$

where ν_t is the complex measure defined by $\nu_t(\Delta)=\operatorname{trace}\big(\mathscr{E}_t(\Delta)K(L_t+\mathrm{i}I)^{-1}\big)$ for a Borel subset Δ of $\mathbb{R}$. Clearly $\|\nu_t\|\le\|K(L_t+\mathrm{i}I)^{-1}\|_{\boldsymbol{S}_1}$. Recall that $\mathscr{E}_t=\mathscr{E}_{L_t}$.

We consider the space $\mathcal{M}(\mathbb{R})$ of complex Borel measures on $\mathbb{R}$ as the dual space to the space $\mathrm{C}_0(\mathbb{R})$ of continuous functions on $\mathbb{R}$ vanishing at infinity with respect to the pairing

$$\langle h,\mu\rangle=\int_{\mathbb{R}}h\,\mathrm{d}\mu.\tag{10.5}$$

Consider the subspace $\mathrm{C}_0(\mathbb{R})\cap\mathrm{C_A}(\mathbb{C}_+)$ of $\mathrm{C}_0(\mathbb{R})$ whose dual space can be identified naturally with the quotient space $\mathcal{M}(\mathbb{R})/\big(\mathrm{C}_0(\mathbb{R})\cap\mathrm{C_A}(\mathbb{C}_+)\big)^\perp$ with respect to the pairing (10.5).

Given a measure μ in $\mathcal{M}(\mathbb{R})$, we denote by $\dot\mu$ the coset in $\mathcal{M}(\mathbb{R})/\big(\mathrm{C}_0(\mathbb{R})\cap\mathrm{C_A}(\mathbb{C}_+)\big)^\perp$, which corresponds to μ.

Let us show that the function $t\mapsto\dot\nu_t$ is continuous in the weak-$*$ topology of $\mathcal{M}(\mathbb{R})/\big(\mathrm{C}_0(\mathbb{R})\cap\mathrm{C_A}(\mathbb{C}_+)\big)^\perp$. To this end we establish the following lemma.

Lemma 10.4. *Let L be a maximal dissipative operator and let $R\in\boldsymbol{S}_1$ Consider the complex Borel measure μ on $\mathbb{R}$ defined by*

$$\mu(\Delta)=\operatorname{trace}\big(\mathscr{E}_L(\Delta)R\big),\quad\Delta\subset\mathbb{R}.\tag{10.6}$$

Then

$$\int_{\mathbb{R}} h\,\mathrm{d}\mu = \operatorname{trace}\big(h(L)R\big) \tag{10.7}$$

for an arbitrary function h in $\mathrm{C}_0(\mathbb{R}) \cap \mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$.

Proof. Consider a resolvent self-adjoint dilation A of L and observe that (10.7) is equivalent to the identity

$$\int_{\mathbb{R}} h\,\mathrm{d}\mu = \operatorname{trace}\big(h(A)R\big). \tag{10.8}$$

Here we extend R by the zero operator on the orthogonal complement of the initial Hilbert space. In fact, (10.8) holds for an arbitrary continuous function on $\mathbb{R}$ satisfying $\lim_{|t|\to\infty} f(t) = 0$. Indeed, (10.6) means that (10.8) holds for characteristic functions. It suffices to approximate the function h in (10.8) by linear combinations of characteristic functions. ■

Let us return to the proof of Theorem 10.1. We want prove that the function $t \mapsto \dot{\nu}_t$ is weak-$*$ continuous, i.e. the map

$$t \mapsto \int_{\mathbb{R}} h\,\mathrm{d}\nu_t \tag{10.9}$$

is continuous for an arbitrary function h in $\mathrm{C}_0(\mathbb{R}) \cap \mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$. Suppose that $h \in \mathrm{C}_0(\mathbb{R}) \cap \mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$. Then by Lemma 10.4,

$$\int_{\mathbb{R}} h\,\mathrm{d}\nu_t = \operatorname{trace}\big(h(L_t)K(L_t + \mathrm{i}I)^{-1}\big).$$

The continuity of the map (10.9) follows now from Lemma 10.2 and from inequality (9.3).

Let ν be a measure in $\mathcal{M}(\mathbb{R})$ such that

$$\dot{\nu} = \int_0^1 \nu_t\,\mathrm{d}t.$$

We can also require that

$$\|\nu\|_{\mathcal{M}(\mathbb{R})} \le \max_{0\le t\le 1} \|K(L_t + \mathrm{i}I)^{-1}\|,$$

which is finite by Lemma 10.2.

Next, by Lemma 10.2, the function $t \mapsto Q_t$ is continuous on [0, 1]. On the other hand, the function $x \mapsto f'(x)(x + \mathrm{i})$ is a bounded function of the first Baire class (see the proof of Theorem 9.1 in [AP8]).

Thus, by Lemma 9.3 of [AP8],

$$\operatorname{trace}\big(f(M)-f(L)\big)=\int_0^1 \operatorname{trace} Q_t\,\mathrm{d}t=\int_0^1\left(\int_{\mathbb{R}} f'(x)(x+\mathrm{i})\,\mathrm{d}\nu_t(x)\right)$$

$$=\int_{\mathbb{R}} f'(x)(x+\mathrm{i})\,\mathrm{d}\nu(x).$$

We have mentioned in § 5 that the condition that $M-L$ is a relatively trace class perturbation of L implies that

$$(M+\mathrm{i}I)^{-1}-(L+\mathrm{i}I)^{-1}\in \boldsymbol{S}_1.$$

It was shown in § 7 of [MNP2] that in this case the pair (L, M) has a spectral shift function $\boldsymbol{\eta}$, such that

$$\int_{\mathbb{R}}\frac{|\boldsymbol{\eta}(x)|}{1+x^2}\,\mathrm{d}x<\infty$$

and

$$\operatorname{trace}\big(f(M)-f(L)\big)=\int_{\mathbb{R}} f'(x)\boldsymbol{\eta}(x)\,\mathrm{d}x$$

whenever f is a rational function with poles in $\mathbb{C}_-\stackrel{\text{def}}{=}\{\zeta\in\mathbb{C}:\ \operatorname{Im}\zeta<0\}$. It follows that

$$\int_{\mathbb{R}} f'(x)\boldsymbol{\eta}(x)\,\mathrm{d}x=\int_{\mathbb{R}} f'(x)(x+\mathrm{i})\,\mathrm{d}\nu(x)$$

for an arbitrary rational function f with poles in $\mathbb{C}_-$. Thus,

$$\int_{\mathbb{R}}\frac{1}{(\lambda-x)^2}\boldsymbol{\eta}(x)\,\mathrm{d}x=\int_{\mathbb{R}}\frac{(x+\mathrm{i})}{(\lambda-x)^2}\,\mathrm{d}\nu(x),\quad \operatorname{Im}\lambda<0.$$

Consider the Radon complex measure μ defined by $\mathrm{d}\mu=(x+\mathrm{i})\mathrm{d}\nu-\boldsymbol{\eta}\,\mathrm{d}\boldsymbol{m}$, where $\boldsymbol{m}$ stands for Lebesgue measure on $\mathbb{R}$. Then

$$\int_{\mathbb{R}}\frac{1}{(\lambda-x)^2}\,\mathrm{d}\mu(x)=0,\quad \operatorname{Im}\lambda<0.$$

It can easily be deduced from the brothers Riesz theorem (see e.g., Lemma 3.7 of [MN] for details) that the complex Borel measure ν is absolutely continuous with respect to Lebesgue measure. We can define now the function $\boldsymbol{\xi}$ by

$$\boldsymbol{\xi}(s)\stackrel{\text{def}}{=}(s+\mathrm{i})\frac{\mathrm{d}\nu(x)}{\mathrm{d}x}(s).$$

Clearly, $\boldsymbol{\xi}$ satisfies (10.1) and equality (10.2) holds. ■

Theorem 10.5. *Let f be a function in $\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$ that is differentiable on $\mathbb{R}$ such that trace formula* (10.2) *holds for arbitrary pairs (L, M), where L is a maximal*

dissipative operator and M is a dissipative operator such that $M - L$ is a relatively trace class perturbation of L. Then $f \in \mathrm{ROL}_{\mathrm{A}}$.

Proof. Since each self-adjoint operator is a maximal dissipative operator, the result follows from Theorem 9.4 in [AP8]. Here we use the fact that a function in $\mathrm{C}_{\mathrm{A}}(\mathbb{C}_+)$ is in ROL if and only if it is in $\mathrm{ROL}_{\mathrm{A}}$ which follows from the argument in the proof of Theorem 8.1 of this paper. ■

References

[A] A.B. ALEKSANDROV, *Commutator Lipschitz functions and analytic continuation*, J. Math. Sci. (N. Y.), **215:5** (2016), 543–551.

[AP1] A.B. ALEKSANDROV and V.V. PELLER, *Operator Hölder–Zygmund functions*, Advances in Math. **224** (2010), 910–966.

[AP2] A.B. ALEKSANDROV and V.V. PELLER, *Functions of perturbed dissipative operators*, Algebra i Analiz **23:2** (2011), 9–51.
English Transl.: St.Petersburg Math. J. **23:2** (2011), 209–238.

[AP3] A.B. ALEKSANDROV and V.V. PELLER, *Operator Lipschitz functions*, Uspekhi Mat. Nauk **71:4** (2016), 3–106 (Russian). English transl.: Russian Math. Surveys **71:4**, 605–702.

[AP4] A.B. ALEKSANDROV and V.V. PELLER, *Krein's trace formula for unitary operators and operator Lipschitz functions*, Funkts. Anal. i Ego Pril. **50:3** (2016), 1–11 (Russian). English transl.: Funct. Anal. Appl. **50:4** (2016), 167–175.

[AP5] A.B. ALEKSANDROV and V.V. PELLER, *Dissipative operators and operator Lipschitz functions*, Proc. Amer. Math. Soc. **147** (2019), 2081–2093.

[AP6] A.B. ALEKSANDROV and V.V. PELLER, *Haagerup tensor products and Schur multipliers*, Algebra i Analiz **36:5** (2024), 70–85.

[AP7] A.B. ALEKSANDROV and V.V. PELLER, *Relatively bounded and relatively trace class perturbations*, C.R. Acad. Sci., Paris, 363 (2025), 377–382.

[AP8] A.B. ALEKSANDROV and V.V. PELLER, *Functions of self-adjoint operators under relatively bounded and relatively trace class perturbations*, Math. Nachr., 2025, DOI: 10.1002/mana.70000

[AP9] A.B. ALEKSANDROV and V.V. PELLER, *Analytic Schur multipliers*, arxiv.org/pdf/2506.15173.

[BS1] M.S. BIRMAN and M.Z. SOLOMYAK, *Double Stieltjes operator integrals*, Problems of Math. Phys., Leningrad. Univ. **1** (1966), 33–67 (Russian). English transl.: Topics Math. Physics **1** (1967), 25–54, Consultants Bureau Plenum Publishing Corporation, New York.

[BS2] M.S. BIRMAN and M.Z. SOLOMYAK, *Double Stieltjes operator integrals. II*, Problems of Math. Phys., Leningrad. Univ. **2** (1967), 26–60 (Russian). English transl.: Topics Math. Physics **2** (1968), 19–46, Consultants Bureau Plenum Publishing Corporation, New York.

[BS3] M.S. BIRMAN and M.Z. SOLOMYAK, *Remarks on the spectral shift function*, Zapiski Nauchn. Semin. LOMI **27** (1972), 33–46 (Russian).
English transl.: J. Soviet Math. **3** (1975), 408–419.

[BS4] M.S. BIRMAN and M.Z. SOLOMYAK, *Double Stieltjes operator integrals. III*, Problems of Math. Phys., Leningrad. Univ. **6** (1973), 27–53 (Russian).

[BS5] M.SH. BIRMAN AND M.Z. SOLOMYAK, *Spectral theory of selfadjoint operators in Hilbert space*, Mathematics and its Applications (Soviet Series), D. Reidel Publishing Co., Dordrecht, 1987.

[BS6] M.S. BIRMAN and M.Z. SOLOMYAK, *Tensor product of a finite number of spectral measures is always a spectral measure*, Integral Equations Operator Theory **24** (1996), 179–187.

[CS] A. CHATTOPADHYAY and A. SKRIPKA, *Trace formulas for relative Schatten class perturbations*, J. Funct. Anal. **274** (2018), 3377–3410.

[DK] YU.L. DALETSKII and S.G. KREIN, *Integration and differentiation of functions of Hermitian operators and application to the theory of perturbations* (Russian), Trudy Sem. Functsion. Anal., Voronezh. Gos. Univ. **1** (1956), 81–105.

[F] YU.B. FARFOROVSKAYA, *An example of a Lipschitzian function of selfadjoint operators that yields a nonnuclear increase under a nuclear perturbation*. Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **30** (1972), 146–153 (Russian).

[GK] I.C. Gokhberg and M.G. Krein, Introduction to the theory of linear nonselfadjoint operators in Hilbert space, "Nauka", Moscow, 1965;
English transl., Amer. Math. Soc., Providence, R.I., 1969.

[KSh] E. KISSIN and V. S. SHULMAN, *On fully operator Lipschitz functions*, J. Funct. Anal. **253:2** (2007), 711–728.

[Kr] M.G. KREIN, *On a trace formula in perturbation theory*, Mat. Sbornik **33** (1953), 597–626 (Russian).

[L] I.M. LIFSHITS, *On a problem in perturbation theory connected with quantum statistics*, Uspekhi Mat. Nauk **7** (1952), 171–180 (Russian).

[MN] M. MALAMUD and H. NEIDHARDT, *Trace formulas for additive and non-additive perturbations*, Adv. Math. **274** (2015), 736–832.

[MNP1] M. MALAMUD, H. NEIDHARDT and V.V. PELLER, *Analytic operator Lipschitz functions in the disk and a trace formula for functions of contractions* (Russian), Funkts. Anal. i Pril., **51:3** (2017), 33–55.
English Transl.: Funct. Anal. and its Appl. **51:3** (2017), 185–203.

[MNP2] M.M. MALAMUD, H. NEIDHARDT and V.V. PELLER, *Absolute continuity of spectral shift*, J. Funct. Anal. **276** (2019), 1575–1621.

[Pe1] V.V. PELLER, *Hankel operators in the theory of perturbations of unitary and self-adjoint operators*, Funktsional. Anal. i Prilozhen. **19:2** (1985), 37–51 (Russian). English transl.: Funct. Anal. Appl. **19** (1985) , 111–123.

[Pe2] V.V. PELLER, *For which f does $A - B \in \mathbf{S}_p$ imply that $f(A) - f(B) \in \mathbf{S}_p$?*, Operator Theory, Birkhäuser, **24** (1987), 289–294.

[Pe3] V.V. PELLER, *Hankel operators in the perturbation theory of of unbounded self-adjoint operators*. Analysis and partial differential equations, 529–544, Lecture Notes in Pure and Appl. Math., **122**, Dekker, New York, 1990.

[Pe4] V.V. PELLER, *Multiple operator integrals in perturbation theory*, Bull. Math. Sci. **6** (2016), 15–88.

[Pe5] V.V. PELLER, *The Lifshits–Krein trace formula and operator Lipschitz functions*, Proc. Amer. Math. Soc. **144** (2016), 5207–5215.

[SNF] B. SZ.-NAGY AND C. FOIAŞ, *Harmonic analysis of operators on Hilbert space*, Akadémiai Kiadó, Budapest, 1970.

Received May 6, 2025. Revised September 25, 2025.

A.B. Aleksandrov
St.Petersburg State University
Universitetskaya nab., 7/9
199034 St.Petersburg, Russia

St.Petersburg Department
Steklov Institute of Mathematics
Russian Academy of Sciences
Fontanka 27, 191023 St.Petersburg
Russia
email: alex@pdmi.ras.ru

V.V. Peller
St.Petersburg State University
Universitetskaya nab., 7/9
199034 St.Petersburg, Russia

St.Petersburg Department
Steklov Institute of Mathematics
Russian Academy of Sciences
Fontanka 27, 191023 St.Petersburg
Russia
email: peller@math.msu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.229

ON WATANABE'S THETA GRAPH DIFFEOMORPHISM IN THE 4-SPHERE

DAVID T GAY

Watanabe's theta graph diffeomorphism, constructed using Watanabe's clasper surgery construction which turns trivalent graphs in 4-manifolds into parametrized families of diffeomorphisms of 4-manifolds, is a diffeomorphism of S^4 representing a potentially nontrivial smooth mapping class of S^4. The "(1, 2)-subgroup" of the smooth mapping class group of S^4 is the subgroup represented by diffeomorphisms which are pseudoisotopic to the identity via a Cerf family with only index 1 and 2 critical points. This author and Hartman showed that this subgroup is either trivial or has order 2 and explicitly identified a diffeomorphism that would represent the nontrivial element if this subgroup is nontrivial. Here we show that the theta graph diffeomorphism is isotopic to this one possibly nontrivial element of the (1, 2)-subgroup. To prove this relation we develop a diagrammatic calculus for working in the smooth mapping class group of S^4.

This paper exhibits a relation in the smooth mapping class group of S^4, which we denote $\mathrm{Mod}(S^4) = \pi_0(\mathrm{Diff}^+(S^4))$, where $\mathrm{Diff}^+(S^4)$ is the group of orientation-preserving diffeomorphisms of S^4. In other words, we will exhibit a smooth isotopy from one diffeomorphism of S^4 to another. Part of the point is that the two diffeomorphisms involved are each of independent interest and arise in different contexts and thus exhibiting this relation improves our understanding of both diffeomorphisms. Another point is to illustrate new techniques for understanding relations in the smooth mapping class groups of 4-manifolds which have the potential to be useful in other contexts.

We stress that we do not know whether or not $\mathrm{Mod}(S^4)$ is trivial. Of course, if it is trivial then our relation is trivially true, but it is possible that either $\mathrm{Mod}(S^4)$ is nontrivial or that to prove that $\mathrm{Mod}(S^4)$ is trivial requires tools similar to those developed in this paper.

The relation we prove has the form

$$\mathrm{wat}(\Theta) = \mathcal{H}(\alpha(1)) \in \mathrm{Mod}(S^4),$$

MSC2020: primary 57K40; secondary 57R50, 57R52, 57R65.
Keywords: diffeomorphism, 4-sphere, barbell, pseudoisotopy.

© 2026 The Author, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

so we now explain what the left and right hand sides of the equation mean. Briefly, for those for whom this is meaningful:

- Θ is the theta graph, i.e., the graph with two vertices and three edges each from one vertex to the other, embedded in S^4,
- $\mathrm{wat}(\,\cdot\,)$ denotes Watanabe's clasper surgery construction [8] for producing families of diffeomorphisms from trivalent graphs,
- $\mathrm{wat}(\Theta)$ is (the smooth isotopy class of) Watanabe's theta graph diffeomorphism [8],
- $\alpha(1)$ is the first in a countable list $\alpha(i)$ of generators of

$$\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3), S^1 \times \{p\})$$

  established by Budney and Gabai in [2], where $\mathrm{Emb}(A, B)$ denotes the space of embeddings of A in B,
- $\mathcal{H}$ is a Cerf-theoretic homomorphism from $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3), S^1 \times \{p\})$ to $\pi_0(\mathrm{Diff}^+(S^4))$ described in [3], and
- $\mathcal{H}(\alpha(1))$ is shown by this author and Hartman in [4] to have order at most 2 and to generate the subgroup $\mathrm{Mod}^{1,2}(S^4)$ of $\mathrm{Mod}(S^4)$ of all diffeomorphisms which are pseudoisotopic to the identity via a pseudoisotopy realizable by a Cerf family involving only critical points of index 1 and 2.

This relation is also proved by Kosanović as Corollary 1.11 in [7]. The current paper was completed after the initial posting of [7] and so is best interpreted as an alternative proof of Kosanović's result, naturally building on many similar techniques, and showcasing a diagrammatic toolbox. In Kosanović's notation, what we call $\mathcal{H}(\alpha(1))$ is denoted $W(1)$.

Given an oriented trivalent graph Γ with $2k$ vertices embedded in a 4-manifold X, with the property that no vertex has all incoming or all outgoing edges, Watanabe [8] describes a "clasper surgery" construction which produces an element $\mathrm{wat}(\Gamma) \in \pi_{k-1}(\mathrm{Diff}^+(X))$. (Watanabe does not use the notation $\mathrm{wat}(\,\cdot\,)$; this is notation we have adopted from [7].) For certain values of k and certain graphs $\Gamma \subset X = B^4$, but *not including* the case $k = 1$, Watanabe is then able to show that $\mathrm{wat}(\Gamma)$ is nontrivial, thus disproving the smooth 4-dimensional Smale conjecture. In the case $k = 1$ the only interesting graph to consider is the theta graph, which we denote Θ, and thus a fundamental question related to Watanabe's work is whether or not $\mathrm{wat}(\Theta) \in \mathrm{Mod}(B^4)$ is trivial. Note that all embeddings of Θ in a simply connected 4-manifold are isotopic so we do not need to specify the embedding. Also note that $\mathrm{Mod}(B^4) = \mathrm{Mod}(S^4)$, so we will work in S^4 rather than B^4.

Of course, if one shows that $\mathrm{wat}(\Theta)$ is nontrivial one will have shown the dramatic result that $\mathrm{Mod}(S^4)$ is nontrivial, but if one shows that $\mathrm{wat}(\Theta)$ is trivial one has only

shown that the clasper surgery construction fails to detect nontriviality of $\mathrm{Mod}(S^4)$. Here we thread the needle between these two possibilities, showing that $\mathrm{wat}(\Theta)$ is equal to another mapping class $\mathcal{H}(\alpha(1))$, about which more is known.

Instead of studying diffeomorphisms up to isotopy one may first study diffeomorphisms up to *pseudoisotopy*: A diffeomorphism $\phi : X \to X$ is pseudoisotopic to the identity if there exists a diffeomorphism $\Phi : [0, 1] \times X \to [0, 1] \times X$ which is the identity on $\{0\} \times X$ and $[0, 1] \times \partial X$ and equals ϕ on $\{1\} \times X$. If Φ is level-preserving then Φ is actually an isotopy. In fact every diffeomorphism of S^4 is pseudoisotopic to the identity; see [3] for a proof that this follows easily from the fact [6] that there are no exotic 5-spheres.

A pseudoisotopy $\Phi : [0, 1] \times X \to [0, 1] \times X$ gives rise to a pair of Morse functions $g_0, g_1 : [0, 1] \times X \to [0, 1]$, where g_0 is simply projection to $[0, 1]$ and $g_1 = g_0 \circ \Phi$. Both g_0 and g_1 are Morse functions without critical points. Cerf's key observation is that if these can be connected by a homotopy g_t such that each g_t does not have critical points then Φ is isotopic to an isotopy, and thus that the problem of turning pseudoisotopies into isotopies can be turned into the problem of simplifying 1-parameter families, called "Cerf families", of Morse functions (allowing births and deaths) between Morse functions without critical points. In our case, when the dimension of X is 4 and X is simply connected, one can always arrange that the critical points that do arise are only of index 2 and 3 [5], and in special cases one can arrange that they are only of index 1 and 2. When the critical points are only of index 1 and 2, the Cerf family is significantly easier to study.

The subgroup of $\mathrm{Mod}(S^4)$ represented by diffeomorphisms which are pseudoisotopic to the identity via a pseudoisotopy with a Cerf family involving only critical points of index 1 and 2 is known as "the (1, 2) subgroup" and denoted $\mathrm{Mod}^{1,2}(S^4)$. In [3] this author exhibited a group homomorphism $\mathcal{H} : \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3), *) \to \mathrm{Mod}(S^4)$ and showed that the image is exactly $\mathrm{Mod}^{1,2}(S^4)$. The basepoint $*$ is the embedding of S^1 as $S^1 \times \{p\}$ for some $p \in S^3$. Budney and Gabai computed $\pi_1(\mathrm{Emb}(S^1, S^1 \times S^3), *)$ and showed that it is free abelian with a countable, explicitly described, list of generators $\alpha(i)$, $i \in \mathbb{N}$. This author and Hartman [4] then showed that in fact $\mathrm{Mod}^{1,2}(S^4)$ is generated just by $\mathcal{H}(\alpha(1))$ and that $\mathcal{H}(\alpha(1))$ is either trivial or of order 2.

From the description of $\mathrm{wat}(\Theta)$ given in the next section, or from a general discussion of pseudosotopies and Watanabe's constructions by Botvinnik and Watanabe [1], it is not hard to see that $\mathrm{wat}(\Theta) \in \mathrm{Mod}^{1,2}(S^4)$. Thus, expressing $\mathrm{wat}(\Theta)$ in terms of $\mathcal{H}(\alpha(1)$ should be a step in the direction of resolving the triviality or nontriviality of $\mathrm{wat}(\Theta)$.

Theorem 1. *We have* $\mathrm{wat}(\Theta) = \mathcal{H}(\alpha(1))$. *In particular, Watanabe's theta graph diffeomorphism generates the* $(1, 2)$*-subgroup* $\mathrm{Mod}^{1,2}(S^4)$ *and has order at most* 2.

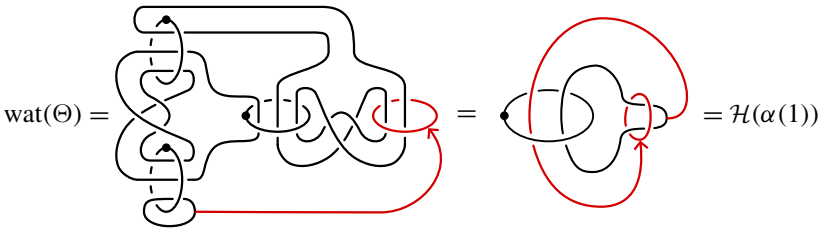


Figure 1. The beginning and the end of our forthcoming sequence of diagrams. Section 1 explains the meaning of the two diagrams in this figure, Section 2 justifies the left-most and right-most equals signs, and Section 3 gives a sequence of intermediate diagrams justifying the middle equals sign.

1. Diagrammatic notation

Central to our argument is something like a Kirby calculus for families of 5-dimensional handlebodies. At the moment this is far from being a complete calculus, but rather an ad hoc collection of tools that works in the context at hand. Consider the two diagrams in Figure 1. The diagram on the left is an illustration of $\mathrm{wat}(\Theta)$ and the diagram on the right is an illustration of $\mathcal{H}(\alpha(1))$; these two statements will be justified in Section 2. The purpose of this paper is to show how to get from the left diagram to the right diagram using various moves. First we explain what both of these diagrams, and the intermediate diagrams in our moves, mean.

1.1. ***Diagrams.*** Each diagram is meant to illustrate a collection of unknotted S^2's in S^4, bounding obvious B^3's, together with a 1-parameter family of embeddings of some number of S^1's in the complement of these S^2's. After explaining how to see this, we will show how such an embedding gives rise to a smooth mapping class of S^4.

Each figure is drawn in $\mathbb{R}^3$ seen as a slice $\{(x_1, x_2, x_3, x_4) \mid x_4 = 0\}$ of $\mathbb{R}^4 = \{(x_1, x_2, x_3, x_4)\}$, which in turn sits in $S^4 = \mathbb{R}^4 \cup \{\infty\}$. The figures have four diagrammatic elements: (1) dotted circles, always unknotted and drawn in black, (2) possibly knotted and linked black circles, without dots, (3) red unknotted circles and (4) red arrows, each pointing from one of the undotted black circles to one of the red circles. Each red circle should have exactly one red arrow pointing to it, but each black undotted circle can have arbitrarily many red arrows attached to it. We explain each of these as follows:

(1) Each black dotted circle bounds an obvious flat disk, and this disk shrinks to a point in $x_4 < 0$ and $x_4 > 0$, thus describing a 3-ball and its boundary S^2. Abusing terminology a bit, we henceforth refer to the black dotted circles themselves as "dotted S^2's".

(2) Each black undotted circle is the starting and ending position of a t-parametrized family, for $t \in [0, 1]$, of embeddings of S^1 into S^4. In other words, for each undotted black circle, we are working towards describing a family of embeddings $\alpha_t : S^1 \hookrightarrow S^4$ such that the drawn black circle is $\alpha_0(S^1) = \alpha_1(S^1)$. For this reason we call these black undotted circles "basepoint circles".

(3) Each red circle represents an embedded S^2 in S^4 given in the usual way: the circle illustrated lies in $\{x_4 = 0\}$ and shrinks to a point for $x_4 > 0$ and $x_4 < 0$. We call each of these red S^2's a "spinning sphere", for reasons which should be clear shortly. This S^2 is actually the trace of a t-parametrized family of arcs, for $t \in [1/4, 3/4]$. To see how to interpret this S^2 as a family of arcs, we first declare that at times $t = 1/4$ and $t = 3/4$ the arc is a small interval neighborhood in the illustrated red circle centered on the point at which the red arrow meets the red circle. Reparametrizing the S^2 so that this short arc is a line of meridian from north to south pole, then as t ranges from $1/4$ to $3/4$ this arc spins around the sphere from west to east. We illustrate this in Figure 2 by redrawing a local picture so that the spinning sphere lies entirely in $\{x_3 = 0\}$ and arcs of basepoint circles that link the spinning sphere appear as points in the interior of the sphere. One should properly adopt an orientation convention, so that an orientation on the red circle unambiguously determines the direction in which the arc spins around the spinning sphere, but in the argument in this paper this will not be important. In other words, our results hold no matter which orientation we choose.

(4) Each arrow represents a blackboard-framed band leading from a basepoint circle to a spinning sphere. This band is the trace of a t-parametrized family of arcs for $t \in [0, 1/4]$ and $t \in [3/4, 1]$. At $t = 0$ and $t = 1$ this arc is a short arc in the basepoint circle centered at the point where the arrow starts, and then as t ranges from 0 to $1/4$ the arc performs a finger move along the band until it meets the spinning sphere; from $3/4$ to 1 the arc simply reverses what it did from 0 to $1/4$. This is illustrated in Figure 3.

The upshot is that a diagram such as one of the ones shown in Figure 1 describes, up to isotopy, an embedding $\beta : \amalg^n B^3 \hookrightarrow S^4$ of some number of 3-balls into S^4, together with a t-parametrized loop of embeddings $\alpha_t : \amalg^m S^1 \hookrightarrow S^4 \setminus \beta(\amalg^n \partial B^3)$ of some number of S^1's into the complement of the boundaries of the 3-balls. The loop α_t is such that each S^1 starts as one of the basepoint circles drawn in the diagram and then sticks out a finger along the various red arrows based at that basepoint circle, and then the tip of each such finger spins once around the spinning sphere at the end of the relevant red arrow, and then the finger retracts back along the red arrow.

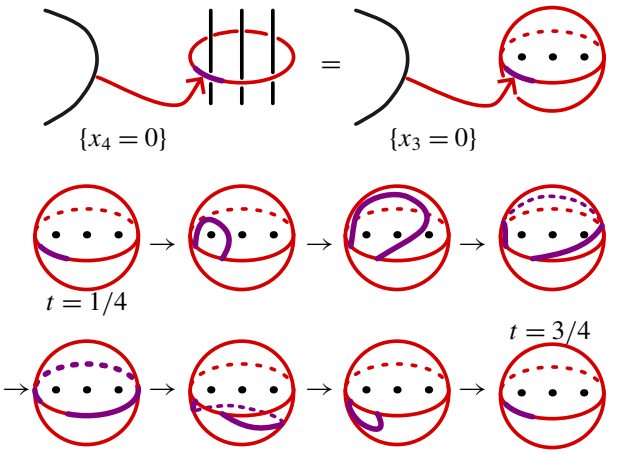


Figure 2. A spinning sphere as a t-parametrized family of arcs, for $t \in [1/4, 3/4]$. In the top line we first draw the figure in the 3-dimensional slice $\{x_4 = 0\}$, so that the sphere appears as a red circle with three linking black arcs, and we then draw the sphere in the slice $\{x_3 = 0\}$ so that we see the entire sphere, but the arcs now appear as three black dots.

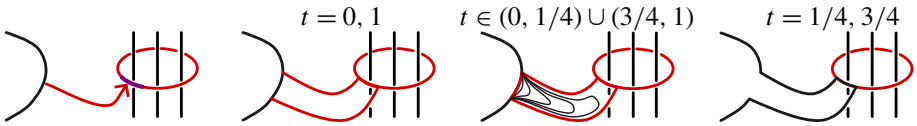


Figure 3. A finger move along an arrow and back.

Note that here we are declaring that when there are multiple spinning spheres, we are still describing a family with one time parameter t. One could adopt the convention that each spinning sphere gives its own time parameter, in which case the parameter space would be $(S^1)^k$, where k is the number of spinning spheres. This is in fact how things work with Watanabe's general construction, as described briefly in the next section. However, in the diagrams in this paper all spinning spheres correspond to the same time parameter. As described above, all spinning spheres spin simultaneously in the time interval $t \in [1/4.3/4]$ but it is clear that these spinnings commute with each other and so in fact we can choose to spin around the spheres one at a time in any order, and this will be important in the main calculation of this paper.

1.2. ***Turning diagrams into bundles and then smooth mapping classes.*** First, given β and α_t, we will build a 6-manifold Z that is a 5-manifold bundle over S^1, with the fiber over $t \in S^1$ being a 5-manifold Z_t. To do this, let W^* be the result of pushing the interior of each component of $\beta(\amalg^n B^3) \subset S^4 \subset B^5$ into the interior of B^5 and carving out a small tubular neighborhood of these B^3's in B^5. Thus

W^* is diffeomorphic to $\natural^n S^1 \times B^4$, the result of attaching n 1-handles to B^5, and $\partial W^* \cong \#^n S^1 \times S^3$. This is the 5-dimensional analogue of dotted circle notation in 4-dimensional Kirby calculus. Now we can view α_t as a family of embeddings of $\amalg^m S^1$ into ∂W^*. Extend this to a family of embeddings $\overline{\alpha_t} : \amalg^m S^1 \times B^3 \hookrightarrow \partial Z^*$ using the 0-framings coming from the original embeddings into S^4; here we need to use the fact that $\pi_2(SO(4))$ is trivial in addition to the usual fact that there are two framings of a circle in a 4-manifold because $\pi_1(SO(4)) = \mathbb{Z}/2\mathbb{Z}$. Let $Z_0 = S^1 \times W^*$. The trace of the family $\overline{\alpha_t}$ can be interpreted as an embedding $\overline{\alpha} : \amalg^m S^1 \times S^1 \times B^3 \hookrightarrow \partial Z_0 = S^1 \times \partial W^*$, with the map on the first S^1 factor being the identity. In other words, $\overline{\alpha}(t, p) = (t, \overline{\alpha_t}(p))$. Use this as the gluing map for an S^1-parametrized family of 5-dimensional 1-handles (equivalently, 6-dimensional round 1-handles) attached to Z_0, and the resulting 6-manifold is Z. The fiber Z_t over $t \in S^1$ is the 5-manifold built by attaching m 5-dimensional 1-handles to W^* with gluing map $\overline{\alpha_t}$. Since these gluing maps vary smoothly with t and since $\overline{\alpha_1} = \overline{\alpha_0}$, this builds Z as a bundle over S^1 with fiber Z_t. Furthermore, ∂Z is a bundle over S^1 with 4-dimensional fiber ∂Z_t. If we let $X = \partial Z_0 = \partial Z_1$, then the monodromy of this bundle is a well defined element of $\mathrm{Mod}(X)$.

Remark. The 5-dimensional 4-manifold bundle ∂Z could be described via parametrized 4-dimensional surgery, rather than via parametrized 5-dimensional handle attachment, and the arguments in this paper could be rewritten in that language. Kosanović's presentation [7] is given in terms of surgery. However, perhaps just due to personal limitations, the author of this paper finds it easier to think in terms of handle attachment than in terms of surgery. Also, the handle description makes the connection to Cerf's approach to pseudoisotopies clearer. Watanabe's description of how to turn a trivalent graph into an element of $\pi_{k-1}(\mathrm{Diff}^+(X))$ is given in terms of surgery, but is easily reinterpreted in terms of handle attachment.

To see that, in our examples, the 4-manifold X sitting over $t = 0$ is diffeomorphic to S^4, note that when $t = 0$ the arrows and the spinning spheres do not come into play. In the two diagrams in Figure 1, ignoring the arrows and the spinning spheres one sees that the 5-dimensional 1- and 2-handles all cancel, so that the 5-manifold Z_0 is diffeomorphic to B^5. We do not need to worry about whether or not the diffeomorphism from $X = \partial Z_0$ to S^4 is canonical because $\mathrm{Mod}(S^4)$ is abelian, and thus the isomorphism from $\mathrm{Mod}(X)$ to $\mathrm{Mod}(S^4)$ given by one diffeomorphism of X to S^4 is equal to that given by any other such diffeomorphism.

Returning to Figure 1, we claim that with the above interpretation of the diagrams, the diagram on the left represents $\mathrm{wat}(\Theta)$ and the diagram on the right represents $\mathcal{H}(\alpha(1))$. In the next section we will explain this, and in the following section we will show how to transform the one diagram into the other using a sequence of moves and intermediate diagrams.

2. The beginning and end diagrams

Watanabe's prescription [8] for turning an oriented trivalent graph Γ in a 4-manifold X into an element of $\pi_{k-1}(\mathrm{Diff}^+(X))$ can be summarized as follows:

- Separate the vertices into those of type I, with 2 incoming edges and 1 outgoing edge, and those of Type II with 1 incoming edge and 2 outgoing edges.
- At each type I vertex place a copy of the 4-dimensional Borromean rings. This is an embedding of $S^2 \amalg S^2 \amalg S^1$ in B^4 characterized by the properties that, ignoring the S^1, the two S^2's bound disjoint B^3's, and that the S^1 reads off the commutator of the meridians of the two S^2's. Arrange for the two incoming edges to land on the two S^2's and for the one outgoing edge to connect to the S^1.
- At each type II vertex also place a copy of the 4-dimensional Borromean rings, but now interpret one of the S^2's as a "spinning sphere" as in Section 1.2, i.e., an S^1-parametrized family of arcs spinning around a north and south pole of the S^2. Arrange for the one incoming edge to connect to the S^2 that does not spin, for one of the outgoing edges to connect to the S^1, and for the other outgoing edge to connect to the spinning sphere — this outgoing edge should really be seen as a band which meets the spinning sphere along the initial arc in the sphere's family of arcs.
- In the middle of each edge place a copy of the 4-dimensional Hopf link. This is an unknotted S^2 and its linking meridian S^1. Arrange that the outgoing half of that edge connects to the S^2 and that the incoming half of that edge connects to the S^1.
- Along the outgoing half of each edge, tube the S^2 from the Hopf link at the middle of the edge to the S^2 from the Borromean rings at the vertex at the end of the edge.
- Along the incoming half of each edge, there is an S^1 coming from the Hopf link at the middle of the edge and either an S^1 or a spinning sphere from a Borromean rings at the incoming end of the edge. In the S^1 case, simply band the two S^1's together. In the spinning sphere case, band the S^1 at the Hopf link to the spinning sphere to create an S^1-parametrized family of S^1's.
- The result is an embedding β of $\amalg^e S^2$ into X, where e is the number of edges, and an $(S^1)^k$-parametrized family $\alpha_{\mathbf{t}}$, with $\mathbf{t} \in (S^1)^k$, of embeddings of $\amalg^e S^1$ into $X \setminus \beta(\amalg^e S^2)$, where k is the number of type II vertices, which is equal to the number of type I vertices. Here each spinning sphere is interpreted as introducing its own independent parameter.

Having done this, Watanabe then wants us to interpret this as an $(S^1)^k$-parametrized family of surgeries along e S^2's and e S^1's, giving a 4-manifold bundle over $(S^1)^k$. With more work, not relevant when $k = 1$, one shows that this bundle is trivial over the $(k-1)$-skeleton of $(S^1)^k$ and thus gives a well defined element of $\pi_{k-1}(\mathrm{Diff}^+(X))$. To turn this 4-dimensional surgery description into a

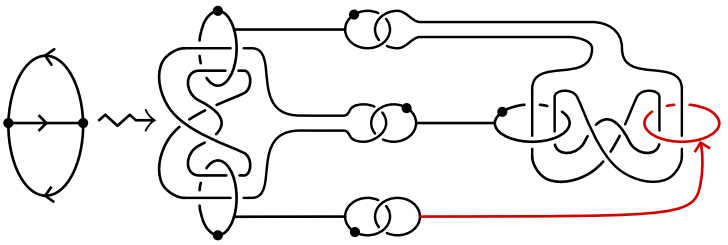


Figure 4. Watanabe's construction applied to the theta graph.

5-dimensional handlebody description, one might think of the S^2's as attaching spheres for 5-dimensional 3-handles, but since the S^2's bound standard B^3's which do not move as the parameters vary (only the S^1's move), these S^2's can just as well be interpreted as dotted S^2's describing 5-dimensional 1-handle attachments, and this is the perspective we take here.

In the case of the theta graph, $k = 1$ so there is only one spinning sphere. Applying the recipe above to the theta graph gives Figure 4, which uses a slight variation of the notation described in Section 1.2, including dotted spheres connected by arcs to indicate spheres that should be tubed together along the arcs. When two dotted spheres are represented by their dotted equatorial circles in our diagrams, and tubed along an arc in the diagram, the resulting sphere is described by the band sum of the two dotted equatorial circles in the diagram. This observation and a simple isotopy turns the diagram on the right of Figure 4 into the diagram on the left of Figure 1. Thus Figure 1 does indeed start with a diagrammatic description of wat(Θ).

To see that the diagram on the right of Figure 1 represents $\mathcal{H}(\alpha(1))$, first note that the diagram does describe some loop $\alpha_t : S^1 \hookrightarrow S^1 \times S^3$ since surgery along the single dotted S^2 turns S^4 into $S^1 \times S^3$. The homomorphism $\mathcal{H} : \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3)) \to \mathrm{Mod}(S^4)$ is nothing more than the same parametrized handle attachment construction described in the preceding section: Think of $S^1 \times S^3$ as the boundary of $S^1 \times B^4$, which is the result of attaching a single 5-dimensional 1-handle to B^5. A loop of S^1's then gives a loop of attaching data with which to attach a 5-dimensional 2-handle, and with this one builds a 5-manifold bundle over S^1 whose boundary is a 4-manifold bundle over S^1 with some monodromy. Because our loop of S^1's starts and ends at $S^1 \times \{p\} \subset S^1 \times S^3$, the fiber of this 4-manifold bundle is diffeomorphic to S^4 and thus the monodromy gives an element of $\mathrm{Mod}(S^4)$. Thus we see that $\mathcal{H}([\alpha])$ is exactly the mapping class described by the figure on the right in Figure 1.

We now explain why $[\alpha_t] = \alpha(1)$, i.e., that the loop α_t of embeddings of $S^1 \times S^3$ is in the homotopy class $\alpha(1)$, or in other words that α_t is homotopic to one of the representatives of $\alpha(1)$ described explicitly in [2], [3] or [4]. Figure 5 illustrates two embeddings of $S^1 \times S^1$ in $S^1 \times S^3$. The left illustration is an adaptation of [4, Figure 6] in the case of $n = 1$ rather than $n = 3$, and the right illustration is an

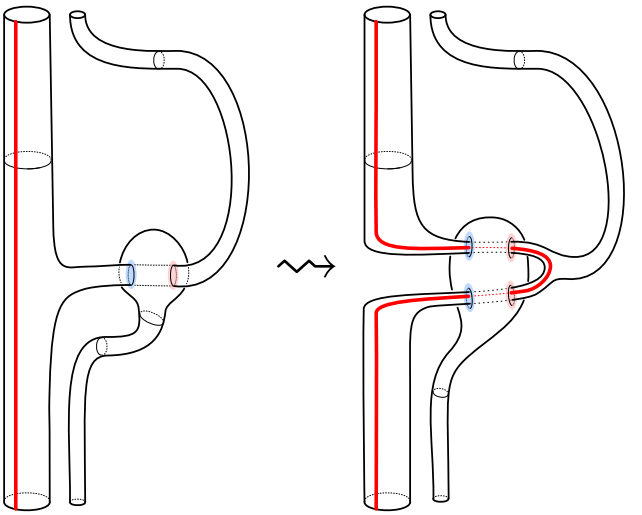


Figure 5. An embedded torus $T(1)$ in $S^1 \times S^3$, which describes a loop of embeddings $\alpha(1) \in \pi_1(\mathrm{Emb}(S^1, S^1 \times S^3))$, and a torus isotopic to $T(1)$. To interpret this figure, the vertical direction is S^1 and the horizontal direction is S^3. The blue and red disks, resp., on the bulbous part of the torus are disks which are pushed into the "past" and "future", resp., so that this torus is in fact embedded and not immersed.

embedding isotopic to the one on the left. These embedded tori are interpreted as loops of embeddings of S^1 with the understanding that the indicated red "longitude" on the torus is the starting embedding of S^1 and that the indicated black "meridians" are the traces of fixed points on this S^1. With this understanding, it is a direct visualization exercise to see that our α_t described by the figure on the right in Figure 1 is homotopic to this loop of embeddings in Figure 5.

3. A sequence of diagrams

We now show how to transform the parametrized 5-dimensional handle diagram on the left in Figure 1 into the diagram on the right in the same figure. We will use parametrized handle cancellation and handleslides as well as homotopies of our families of embeddings of S^1's. By "parametrized handle cancellation" we mean that, if a 2-handle is moving through a 1-parameter family but the attaching circle for the 2-handle goes over a 1-handle geometrically once at all times throughout the entire 1-parameter family, then the 1- and 2-handle can be canceled throughout entire family, i.e., the 1-handle and the 2-handle can be erased from the diagram.

Proof of Theorem 1. Since the diagram on the left in Figure 1 represents $\mathrm{wat}(\Theta)$ and the one on the right represents $\mathcal{H}(\alpha(1))$, by showing how to transform one

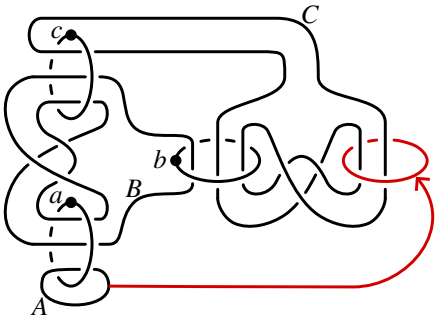


Figure 6. wat(Θ) with labeled handles.

diagram into the other and that the moves we use in this transformation respect the corresponding element of $\mathrm{Mod}(S^4)$, we will complete the proof of Theorem 1.

Figure 6 reproduces our initial diagram for wat(Θ) and introduces labels a, b and c for the three 1-handles and labels A, B and C for the three 2-handles. Recall that this describes a t-parametrized family of 5-dimensional handlebodies, but that the 2-handle A is the only 2-handle that is actually changing with t. Also, note that at any time t, all handles can easily be canceled, but that the cancellation requires handleslides and isotopies to arrange that for each 1-handle there is exactly one 2-handle which goes over it once, i.e., cancels it, and that the other 2-handles do not go over that 1-handle at all. At time $t = 0$, when the attaching circle for 2-handle A is simply the small linking meridian to the dotted 2-sphere for 1-handle a, this is easy to do: Slide 2-handle B over 2-handle A twice to get B off of a, then cancel A and a, then slide 2-handle C over B twice to get C off of b and cancel B and b, and finally cancel C and c. However, once A starts to move, then the slid version of B will move, and so forth, and we will need to be much more careful.

There are three simplifying moves that we can use to keep the handleslide complexity from exploding too quickly. The first is to observe that, when isotoping 1-dimensional elements of our diagrams, we can freely change crossings since this isotopy is happening in $\mathbb{R}^4$ not in $\mathbb{R}^3$. This applies to crossings between "ordinary" S^1's, crossings between arrows, and crossings between "ordinary" S^1's and arrows.

The second move is to observe that a spinning sphere around multiple strands of "ordinary" S^1's can be split into multiple spinning spheres, one around each strand, as illustrated in Figure 7.

The third move is to exploit the fact that when spinning around multiple spheres

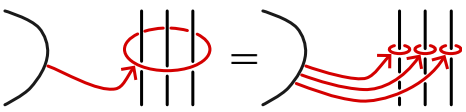


Figure 7. Splitting spinning spheres.

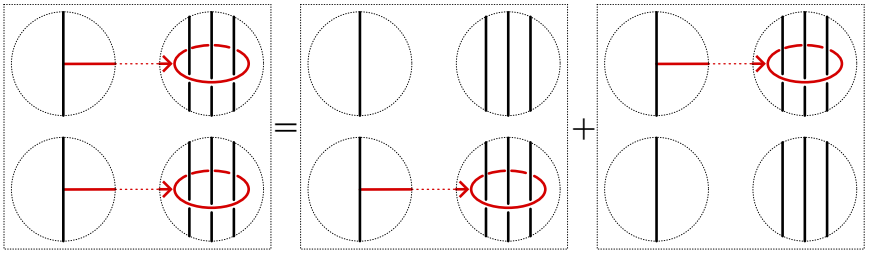


Figure 8. Separating spinning around different spheres in $\mathrm{Mod}(S^4)$.

to describe an S^1-parametrized family, we can spin around them in any order as discussed earlier. Using the fact that $\mathrm{Mod}(S^4)$ is abelian we can use additive notation in $\mathrm{Mod}(S^4)$ and, interpreting each diagram as an element of $\mathrm{Mod}(S^4)$, we derive the diagrammatic move shown in Figure 8. This move is perhaps better described in words: Given a diagram D with two red arrows x and y pointing to red spinning spheres X and Y, respectively, consider the diagram D_1 obtained by erasing x and X and the diagram D_2 obtained by erasing y and Y. Then the mapping class represented by D is the sum of the mapping classes represented by D_1 and D_2.

In Figure 9 we split the spinning sphere for 2-handle A, then simplify with an isotopy, and then decompose $\mathrm{wat}(\Theta)$ as a sum of two mapping classes, one given by the diagram on the left at the bottom and one given by the diagram on the right at the bottom. In the diagram on the left, we can cancel the 1- and 2-handle pairs

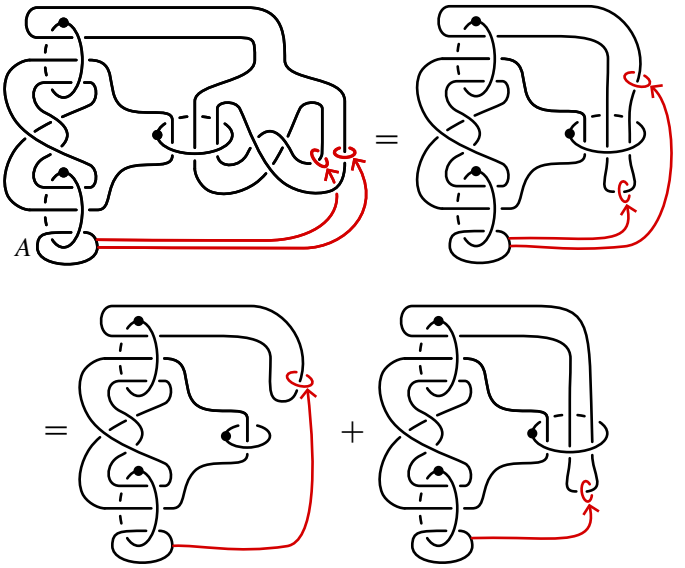


Figure 9. Split the spinning sphere for A, isotope and then decompose additively in $\mathrm{Mod}(S^4)$.

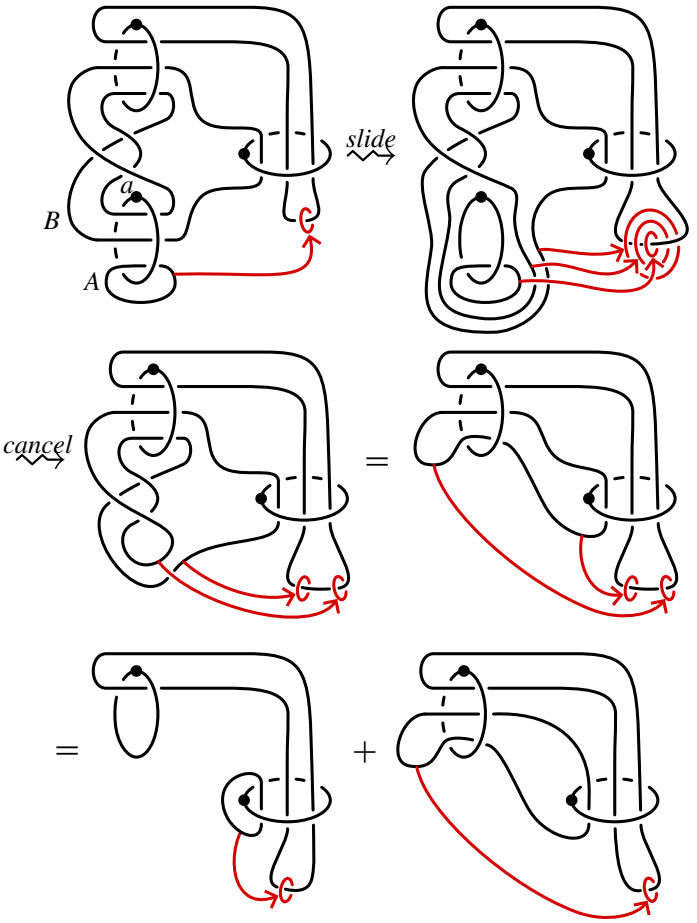


Figure 10. Slide B over A, cancel a and A, isotope, and then decompose.

easily for all t, since there is one dotted circle with only one 2-handle running over it, and we start the cancellation there. Thus in the end $\mathrm{wat}(\Theta)$ is equal to the mapping class given by the diagram on the bottom right.

In Figure 10, we start with the diagram from the bottom right of Figure 9, slide 2-handle B twice over 2-handle A to get B off of a, and then cancel a and A. This is followed by an isotopy and then, again, a decomposition in $\mathrm{Mod}(S^4)$. The diagram on the bottom left easily cancels, so now $\mathrm{wat}(\Theta)$ is described by the diagram on the bottom right.

Next, in Figure 11, we start with the diagram from the bottom right of Figure 10, slide C twice over B to get C off of b, and then cancel B and b. Again we follow this with an isotopy and a decomposition, expressing $\mathrm{wat}(\Theta)$ as a sum of two elements of $\mathrm{Mod}(S^4)$, one of which immediately cancels, leaving $\mathrm{wat}(\Theta)$ equal to the mapping class described by the diagram on the bottom right.

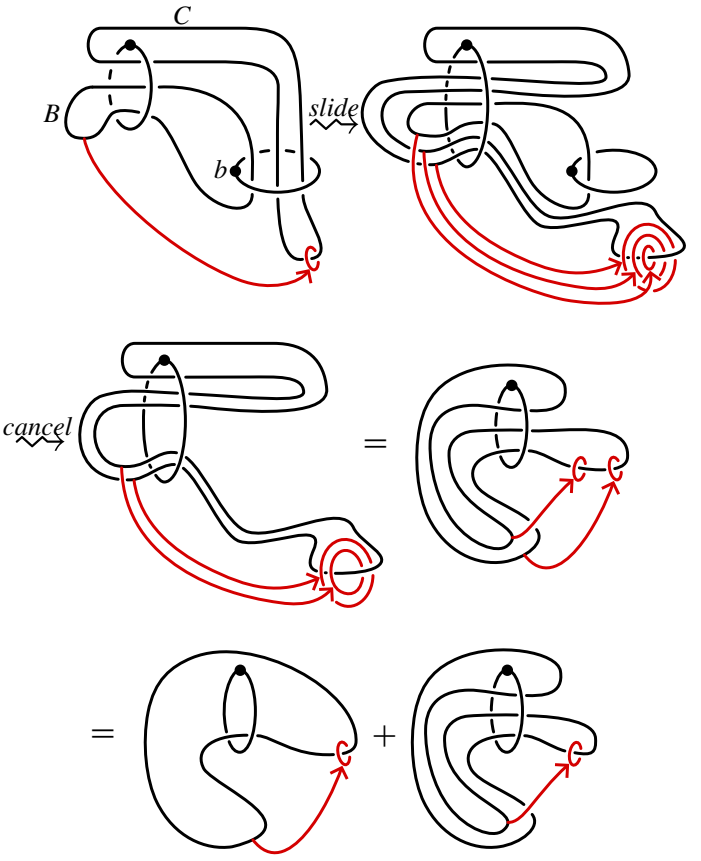


Figure 11. Slide C over B, cancel b and B, isotope, and then decompose.

Figure 12 shows an isotopy followed by the reverse of a cancellation and a decomposition, going from the diagram on the bottom right of Figure 11 to our target diagram on the right of Figure 1, which we have seen is a diagram for $\mathcal{H}(\alpha(1))$. □

References

[1] B. Botvinnik and T. Watanabe, "Families of diffeomorphisms and concordances detected by trivalent graphs", *J. Topol.* **16**:1 (2023), 207–233. MR

[2] R. Budney and D. Gabai, "Knotted 3-balls in S^4", 2019. arXiv 1912.09029

[3] D. T. Gay, "Diffeomorphisms of the 4-sphere, Cerf theory and Montesinos twins", 2021. arXiv 2102.12890

[4] D. T. Gay and D. Hartman, "Relations amongst twists along Montesinos twins in the 4-sphere", *Algebr. Geom. Topol.* **25**:1 (2025), 287–299. MR

[5] A. Hatcher and J. Wagoner, *Pseudo-isotopies of compact manifolds*, vol. No. 6, Astérisque, Société Mathématique de France, Paris, 1973. With English and French prefaces. MR

[6] M. A. Kervaire and J. W. Milnor, "Groups of homotopy spheres. I", *Ann. of Math.* (2) **77** (1963), 504–537. MR

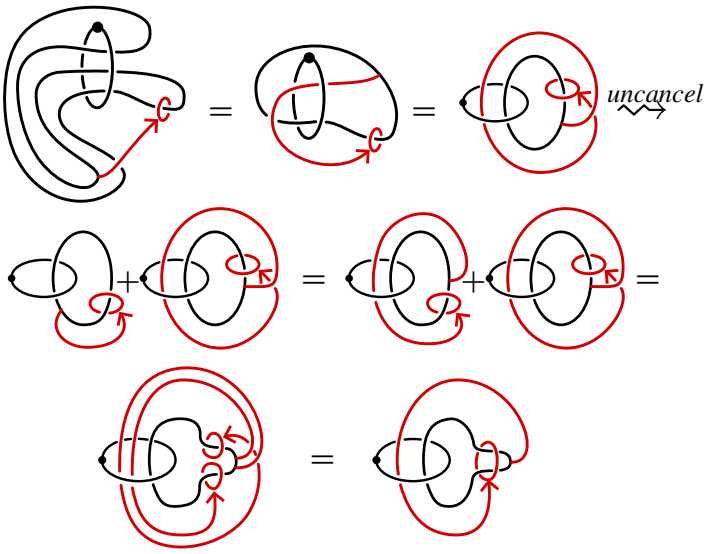


Figure 12. The final steps: An isotopy, an "un-cancellation", another isotopy, an "un-decomposition" and finally an "un-splitting".

[7] D. Kosanović, "Diffeomorphisms of 4-manifolds from graspers", *Proc. Lond. Math. Soc.* (3) **131**:1 (2025), Paper No. e70065, 36. MR

[8] T. Watanabe, "Some exotic nontrivial elements of the rational homotopy groups of Diff(S^4)", 2018. arXiv 1812.02448

Received December 9, 2024. Revised September 19, 2025.

DAVID T GAY
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF GEORGIA
ATHENS, GA 30602
UNITED STATES
dgay@uga.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.245

MUTATION OF SIGNED VALUED QUIVERS AND PRESENTATIONS OF SIMPLE COMPLEX LIE ALGEBRAS

JOSEPH GRANT AND DAVIDE MORIGI

We introduce a signed variant of (valued) quivers and a mutation rule that generalizes the classical Fomin–Zelevinsky mutation of quivers. To any signed valued quiver we associate a matrix that is a signed analogue of the Cartan counterpart appearing in the theory of cluster algebras. From this matrix, we construct a Lie algebra via a "Serre-like" presentation.

In the mutation Dynkin case, we define root systems using the signed Cartan counterpart and show compatibility with mutation of roots as defined by Parsons. Using results from Barot–Rivera and Pérez–Rivera, we show that mutation equivalent signed quivers yield isomorphic Lie algebras, giving presentations of simple complex Lie algebras.

1. Introduction 246
2. Mutation of signed valued quivers 248
2A. Generalized skew-symmetrizable matrices 248
2B. Signed valued quivers 250
2C. Mutation Dynkin type 256
3. Roots 263
3A. The Cartan counterpart 263
3B. Root systems 266
3C. Mutation type A 270
4. Presentations of Lie algebras 273
4A. Presentations 273
4B. Relations 275
4C. Homomorphisms 281
4D. Root space decompositions 287
4E. An example of type A_3 289
Acknowledgements 291
References 291

MSC2020: primary 17B20; secondary 17B22, 13F60.
Keywords: quiver mutation, presentation, Lie algebra, root system.

© 2026 The Authors, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

1. Introduction

The theory of cluster algebras has been one of the mathematical success stories of the 21st century. From early in the theory's development, it was clear that mutation gave new ways to understand finite type phenomena in Lie theory when the Dynkin classification arose in classifying mutation classes of finite type cluster algebras [8]. This led to generalizations of Cartan matrices [1] and simple roots [13], and then to new presentations of finite crystallographic reflection groups [3] and braid groups [9]. As we get new presentations of objects in Lie theory, it is natural to ask whether we get new presentations of the original object of the Dynkin classification: simple complex Lie algebras.

Independently of the theory of cluster algebras, presentations of Lie algebras associated to various unit forms were studied [2] and it was realised that as well as the relations in the classical Serre presentation, one should add further relations according to cycles appearing in a graphical representation of generalized Cartan matrices. This theory was refined, and extended to the setting of quasi-Cartan matrices appearing in the cluster algebras literature [1], in both the simply laced [4] and skew-symmetrizable [14] cases. In these studies, graphs with two kinds of edges play a central role. They have been denoted by solid and broken (dotted) edges, but we will represent them by edges decorated with $-$ and $+$ signs, respectively, to match the signs appearing in the generalized Cartan matrix.

Cluster algebras start with a skew-symmetrizable matrix, and the mutation rule is a case-by-case operation which changes this matrix. In the simpler skew-symmetric case, this can be represented graphically by a quiver where orientation corresponds to sign, and the mutation rule can be described by a three step process. The graphical description of the general skew-symmetrizable matrices has long been known: one should use valued quivers, which have two positive integers attached to each arrow.

In this article we introduce signed variants of quivers and valued quivers, where each arrow has a sign, together with a mutation rule of these objects. As in [7], our theory starts with matrices, but the signs are already used to encode orientation. So, in order to keep track of signs, we introduce a new variable t which squares to 1. We give a mutation rule for these generalized skew-symmetric (gss) matrices and show that everything can be encoded graphically via signed valued quivers in the mutation Dynkin case. We use earlier results on mutation [1], and recover various classical phenomena on specialization to the classical case.

Given a signed valued quiver (Q, v) we define a Cartan counterpart $C(Q, v)$. As with traditional Cartan matrices, this has diagonal entries 2. It is modelled on a definition from cluster algebras [7; 8], where the off-diagonal entries are always nonpositive, but in our definition we get both positive and negative off-diagonal entries. We define a Lie algebra $\mathfrak{g}(Q, v) = \mathfrak{g}(C(Q, v))$ depending on this Cartan

counterpart, and using previous work of Barot–Rivera [4] and Pérez–Rivera [14] we show the following result.

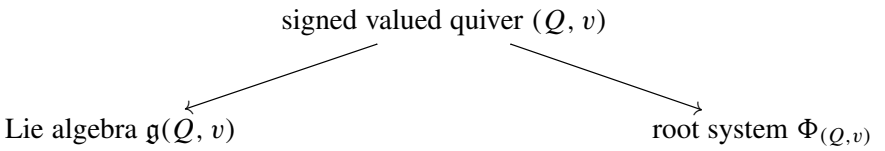
Theorem 4.15. *If (Q, v) is a signed valued quiver of mutation Dynkin type and $(Q', v') = \mu_k(Q, v)$ then we have an isomorphism of Lie algebras*

$$\varphi_k : \mathfrak{g}(Q', v') \xrightarrow{\sim} \mathfrak{g}(Q, v).$$

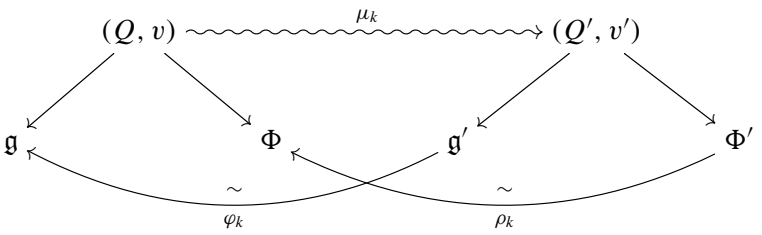
As a corollary (Corollary 4.24) we obtain that each signed valued quiver (Q, v) of mutation Dynkin type Δ gives a presentation of the simple complex Lie algebra $\mathfrak{g}(\Delta)$.

One of the key tools in the study of semisimple Lie algebras is the theory of root systems. These traditionally make use of simple roots, which form a basis of the vector space spanned by all the roots. An alternative basis known as a companion basis, which is adapted to a particular quiver in the mutation class of our Dynkin diagram, was proposed by Parsons in the simply laced case [13], and later extended to the skew-symmetrizable case [3]. We introduce a signed variant of companion bases, and show that our refined variant is preserved by Parsons's mutation rule.

Via iterated mutation of roots, the formal simple basis of our root system is mapped to a signed companion basis in the original Dynkin root system: see Theorem 3.22. Therefore, our theory of signed quiver mutation can be seen as a combinatorial precursor of both of the above theories: presentations of Lie algebras, and companion bases.



If we write $\mathfrak{g} = \mathfrak{g}(Q, v)$, $\mathfrak{g}' = \mathfrak{g}(Q', v')$, $\Phi = \Phi_{(Q,v)}$, and $\Phi' = \Phi_{(Q',v')}$, then we have mutations of all three of the above objects, as illustrated in the following diagram:



This diagram suggests there might be some compatibility between mutation of the Lie algebras and the root systems. We give a compatibility result for root space

decompositions. Let $\mathfrak{h}$, $\mathfrak{h}'$ denote the Cartan subalgebras of $\mathfrak{g}$, $\mathfrak{g}'$, respectively, and let $\mathfrak{g}^\alpha$ denote the root space of $\alpha \in \Phi$, where all $h \in \mathfrak{h}$ act by $\alpha(h)$.

Theorem 4.26. $\varphi_k(\mathfrak{g}'^{\beta'}) = \mathfrak{g}^{\rho_k(\beta')}$.

Notation. Fix $n \geq 1$. All matrices are $n \times n$, and i, j, k are integers between 1 and n. We mainly work in our signed setting, but use a tilde to identify objects coming from the classical setting. So we use $\tilde{B}$ to denote a skew-symmetric (ss) matrix, and B to denote a generalized skew-symmetric (gss) matrix.

2. Mutation of signed valued quivers

2A. ***Generalized skew-symmetrizable matrices.*** An $n \times n$ matrix $\tilde{B} = (\tilde{b}_{ij})$ with integer entries is called skew-symmetrizable if there exist positive integers d_i such that $d_i \tilde{b}_{ij} = -d_j \tilde{b}_{ji}$. The diagonal matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ is called the skew-symmetrizing matrix or skew-symmetrizer, and $D\tilde{B} = -(D\tilde{B})^T$.

Definition 2.1. A *generalized skew-symmetrizable* (*gss*) *matrix* is an $n \times n$ matrix $B = (b_{ij})$ with entries $b_{ij} \in \mathbb{Z}[t]/(t^2 - 1)$ which is skew-symmetrizable, i.e., there exist $d_1, \ldots, d_n \in \mathbb{Z}_{>0}$ such that

$$d_i b_{ij} = -d_j b_{ji}$$

for all $i, j = 1, \ldots, n$. If B is a gss matrix, its *specialization* $B(1)$ is given by evaluating $t = 1$.

Remark 2.2. If B is a gss matrix, its specialization $B(1)$ is also skew-symmetrizable, with the same skew-symmetrizing matrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ as B.

Recall that for a nonzero real number a, its *sign* is $\operatorname{sgn}(a) = a/|a|$, and $\operatorname{sgn}(0) = 0$. We have $\operatorname{sgn}(ab) = \operatorname{sgn}(a)\operatorname{sgn}(b)$, and for a skew-symmetrizable matrix $\tilde{B}$ we have $\operatorname{sgn}(\tilde{b}_{ij}) = -\operatorname{sgn}(\tilde{b}_{ji})$. For every $k \in \{1, \ldots, n\}$, Fomin and Zelevinsky's mutation rule for integer matrices $\tilde{B}' = \mu_k^{\mathrm{FZ}}(\tilde{B}) = (\tilde{b}'_{ij})$ [7, Definition 4.2] can be written as follows [1, (1.1)]:

$$\tilde{b}'_{ij} = \begin{cases} -\tilde{b}_{ij} & \text{if } i = k \text{ or } j = k, \\ \tilde{b}_{ij} + \operatorname{sgn}((\tilde{b}_{ik}))\tilde{b}_{ik}\tilde{b}_{kj} & \text{if } \operatorname{sgn}((\tilde{b}_{ik}\tilde{b}_{kj})) = 1, \\ \tilde{b}_{ij} & \text{otherwise.} \end{cases}$$

We give a mutation rule for generalized matrices.

Definition 2.3. For $a + bt \in \mathbb{Z}[t]/(t^2 - 1)$, define its *sign* as

$$\operatorname{sgn}(a + bt) = \operatorname{sgn}(a + b).$$

The important property of this definition is that if $x = a \in \mathbb{Z}$ then $\operatorname{sgn}(x) = \operatorname{sgn}(a)$, and if $x = tb \in t\mathbb{Z}$ then $\operatorname{sgn}(x) = \operatorname{sgn}(b)$. It is also useful to note that, for $x, y \in \mathbb{Z}[t]/(t^2 - 1)$, we have $\operatorname{sgn}(tx) = \operatorname{sgn}(x)$ and $\operatorname{sgn}(x)\operatorname{sgn}(y) = \operatorname{sgn}(xy)$.

Definition 2.4. Let $B = (b_{ij})$ be a gss matrix and let $k \in \{1, \dots, n\}$. Define the *signed mutation* $B' = \mu_k(B) = (b'_{ij})$ of B as follows:

$$b'_{ij} = \begin{cases} -b_{ij} & \text{if } i = k \text{ and } \operatorname{sgn} b_{ij} = 1, \text{ or } j = k \text{ and } \operatorname{sgn} b_{ij} = -1, \\ -t b_{ij} & \text{if } i = k \text{ and } \operatorname{sgn} b_{ij} = -1, \text{ or } j = k \text{ and } \operatorname{sgn} b_{ij} = 1, \\ t(b_{ij} + \operatorname{sgn}(b_{ik}) b_{ik} b_{kj}) & \text{if } \operatorname{sgn}(b_{ik} b_{kj}) = 1; \\ b_{ij} & \text{otherwise.} \end{cases}$$

Example 2.5. Consider the gss matrix

$$B = \begin{pmatrix} 0 & -t & 0 \\ t & 0 & -2t \\ 0 & t & 0 \end{pmatrix},$$

with skew-symmetrizing matrix $\operatorname{diag}(1, 1, 2)$. Its signed mutation at vertex $k = 2$ is

$$\mu_2(B) = \begin{pmatrix} 0 & t & -2t \\ -t & 0 & 2 \\ t & -1 & 0 \end{pmatrix}.$$

The next result says that our mutation rule specializes to the classical Fomin–Zelevinsky mutation of skew-symmetrizable matrices.

Proposition 2.6. *Let $k \in \{1, \dots, n\}$. Then $\mu_k^{\mathrm{FZ}}(B(1)) = (\mu_k B)(1)$, i.e., the following diagram commutes*:

$$\begin{array}{ccc} B & \xrightarrow{\mu_k} & \mu_k(B) \\ \downarrow & & \downarrow \\ B(1) & \xrightarrow[\mu_k^{\mathrm{FZ}}]{} & (\mu_k B)(1) \end{array}$$

Proof. Note that $\operatorname{sgn}(b_{ij}(1)) = \operatorname{sgn}(b_{ij})$ and compare the definitions of μ_k and μ_k^{FZ}. □

Lemma 2.7. *Let $B = (b_{ij})$ be a gss matrix. Then*:

(i) $\operatorname{sgn}(b_{ij}) = -\operatorname{sgn}(b_{ji})$,

(ii) $\operatorname{sgn}(b'_{ij}) = -\operatorname{sgn}(b_{ij})$ *if* $k = i, j$,

(iii) $\operatorname{sgn}(b_{ij} b_{\ell m}) = \operatorname{sgn}(b_{ij}) \operatorname{sgn}(b_{\ell m})$,

(iv) $\operatorname{sgn}(b_{ik} b_{kj}) = \operatorname{sgn}(b_{jk} b_{ki})$.

Proof. All properties follow from the skew-symmetrizable property and

$$\operatorname{sgn}(x) \operatorname{sgn}(y) = \operatorname{sgn}(xy).$$ □

Fomin–Zelevinsky mutation preserves the property of being skew-symmetrizable; in fact, the skew-symmetrizing matrix remains constant [7, Proposition 4.5]. We have a similar result in our setting.

Lemma 2.8. *If B is a gss matrix with skew-symmetrizer $D = \operatorname{diag}(d_1, \dots, d_n)$ and $B' = \mu_k(B)$, then B' is also a gss matrix with skew-symmetrizer D.*

Proof. We need to check all four cases of Definition 2.4. First, if $k = i$ and $\operatorname{sgn}(b_{ij}) = 1$, then $\operatorname{sgn}(b_{ji}) = -1$, so $d_i b'_{ij} = -d_i b_{ij} = d_j b_{ji} = -d_j b'_{ji}$. The other possibilities in the first two cases are similar. Next suppose $\operatorname{sgn}(b_{ik}b_{kj}) = 1$. We have $d_i b_{ik} b_{kj} = -d_k b_{ki} b_{kj} = d_j b_{jk} b_{ki}$ and $\operatorname{sgn}(b_{ik}) = \operatorname{sgn}(b_{kj}) = -\operatorname{sgn}(b_{jk})$, so $d_i b'_{ij} = t d_i (b_{ij} + \operatorname{sgn}(b_{ik}) b_{ik} b_{kj}) = t d_j (-b_{ji} - \operatorname{sgn}(b_{jk}) b_{jk} b_{ki}) = -d_j b'_{ji}$. Finally, in the fourth case, $\operatorname{sgn}(b_{ik}b_{kj}) \neq 1$, so $\operatorname{sgn}(b_{jk}b_{ki}) \neq 1$, so $d_i b'_{ij} = d_i b_{ij} = -d_j b_{ji} = -d_j b'_{ji}$. □

Famously, Fomin–Zelevinsky mutation is an involution. The same does not hold in our situation, but we do have the following:

Lemma 2.9. *Let B be a gss matrix and let $B'' = \mu_k^2(B)$. Then*

$$b''_{ij} = \begin{cases} t b_{ij} & \text{if } i = k \text{ or } j = k, \\ b_{ij} & \text{otherwise.} \end{cases}$$

Proof. Write $B^{(\ell)} = \mu_k^\ell(B)$. If $i = k$ or $j = k$, Definition 2.4 gives $b^{(2)}_{ij} = t b^{(0)}_{ij}$.

Now let i, j be such that $\operatorname{sgn}(b^{(0)}_{ik} b^{(0)}_{kj}) = 1$. By Lemma 2.7 we have

$$\operatorname{sgn}(b^{(1)}_{ik} b^{(1)}_{kj}) = \operatorname{sgn}(b^{(1)}_{ik}) \operatorname{sgn}(b^{(1)}_{kj}) = (-1)^2 \operatorname{sgn}(b^{(0)}_{ik}) \operatorname{sgn}(b^{(0)}_{kj}) = \operatorname{sgn}(b^{(0)}_{ik} b^{(0)}_{kj}) = 1,$$

so Definition 2.4 and Lemma 2.7 give

$$\begin{aligned} b^{(2)}_{ij} &= t(b^{(1)}_{ij} + \operatorname{sgn}(b^{(1)}_{ik}) b^{(1)}_{ik} b^{(1)}_{kj}) \\ &= t(t(b^{(0)}_{ij} + \operatorname{sgn}(b^{(0)}_{ik}) b^{(0)}_{ik} b^{(0)}_{kj}) + \operatorname{sgn}(b^{(1)}_{ik}) b^{(1)}_{ik} b^{(1)}_{kj}) \\ &= t(t b^{(0)}_{ij} + t \operatorname{sgn}(b^{(0)}_{ik}) b^{(0)}_{ik} b^{(0)}_{kj} + t \operatorname{sgn}(b^{(1)}_{ik}) b^{(0)}_{ik} b^{(0)}_{kj}) \\ &= t(t b^{(0)}_{ij} + t \operatorname{sgn}(b^{(0)}_{ik}) b^{(0)}_{ik} b^{(0)}_{kj} - t \operatorname{sgn}(b^{(0)}_{ik}) b^{(0)}_{ik} b^{(0)}_{kj}) \\ &= t^2 b^{(0)}_{ij} = b^{(0)}_{ij}. \end{aligned}$$

In all other cases we have $b^{(2)}_{ij} = b^{(1)}_{ij} = b^{(0)}_{ij}$, concluding the proof. □

As an immediate consequence, we get:

Proposition 2.10. $\mu_k^4(B) = B$.

2B. ***Signed valued quivers.*** The classical theory of mutation has a graphical interpretation using valued quivers. To develop a corresponding interpretation for mutation of gss matrices, we need the following definition.

Definition 2.11. An $n \times n$ gss matrix B is called *pure* if, for all $i, j \in \{1, \dots, n\}$, either $b_{ij} \in \mathbb{Z}$ or $b_{ij} \in t\mathbb{Z}$.

A quiver is called *simple* if its underlying graph is simple, so it has no loops and has at most one arrow between any two vertices. A quiver is called *cluster* if it satisfies the usual assumptions in cluster theory: it has no loops or 2-cycles. So the Kronecker quiver is cluster, but is not simple.

We recall some standard notions which can be found in, for example, the introduction of [6]. A *valued graph* (Γ, a) is a set Γ of vertices together with nonnegative integers a_{ij} for all $i, j \in \Gamma$ such that $a_{ii} = 0$ and such that there exist positive integers d_i such that $d_i a_{ij} = d_j a_{ji}$. An *orientation* Ω of (Γ, a) is a choice of order $i \to j$ or $j \to i$ for each unordered pair $i, j \in \Gamma$ with $d_{ij} \neq 0$. We can combine these data structures using a *valued quiver*: this is a simple quiver Q together with an ordered pair $\tilde{v}(\alpha) = (a_{ij}, a_{ji})$ of positive integers for each arrow $\alpha : i \to j$, satisfying the $d_i a_{ij} = d_j a_{ji}$ condition.

We will need a signed variant of the previous definition.

Definition 2.12. A *signed valued quiver* is a pair (Q, v) where Q is a simple quiver and

$$v = (v_1, v_2) : Q_1 \to \mathbb{Z} \times \mathbb{Z}$$

is a function on the arrows of Q such that, for all $\alpha \in Q_1$, $\operatorname{sgn}(v_1(\alpha)) = \operatorname{sgn}(v_2(\alpha)) \neq 0$. Moreover, there should exist positive integers d_i, for $i \in Q_0$, such that $d_i v_1(\alpha) = d_j v_2(\alpha)$ for all arrows $\alpha : i \to j$. The ordered pair $v(\alpha)$ is called the *value* of α and the product $v_1(\alpha)v_2(\alpha)$ is called the *weight* of α. We say that an arrow α is *positive* and has *sign* $+1$ if $v_1(\alpha) > 0$, and is *negative* and has *sign* -1 if $v_1(\alpha) < 0$. If (Q, v) is a signed valued quiver, its *specialization* (Q, v^+) is the valued quiver with $v^+(\alpha) = (|v_1(\alpha)|, |v_2(\alpha)|)$ for each $\alpha \in Q_1$.

For an arrow $\alpha : i \to j$ with $v(\alpha) = (a, b)$, we often write $\alpha : i \xrightarrow{(a,b)} j$, or sometimes $\alpha : j \xleftarrow{(a,b)} i$.

It is occasionally convenient to write $\alpha : i \xrightarrow{(0,0)} j$ when there is no arrow between i and j, and $\alpha : i \overset{\pm}{\Rightarrow} j$ when there is an arrow of value $(\pm 1, \pm 1)$.

If all the arrows in (Q, v) have value $(\pm 1, \pm 1)$, so a sign on the arrows is enough to determine the values, we will simply talk about a *signed quiver*.

We now explain how to associate a signed valued quiver to a pure gss matrix. The rough idea is that $+, -$ signs of matrix entries correspond to orientation of arrows, as in the classical setting, and $1, t$ coefficients correspond to positive/negative sign of arrows.

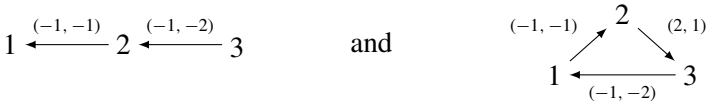
Definition 2.13. Let $B = (b_{ij})$ be a pure gss matrix. Define its *associated signed valued quiver* to be the pair $(Q, v) = (Q_B, v_B)$ where

- Q has vertex set $\{1, \ldots, n\}$;
- there is an arrow $i \to j$ if and only if $\operatorname{sgn}(b_{ij}) = 1$;
- $v : Q_1 \to \mathbb{Z} \times \mathbb{Z}$ is given by $v(\alpha) = (b_{ij}(-1), -b_{ji}(-1))$.

Given a signed valued quiver (Q, v), we define its *associated gss matrix* to be the pure gss matrix $B = B(Q, v) = (b_{ij})$, where

$$b_{ij} = \begin{cases} v_1(\alpha) & \text{if } \exists \text{ positive } \alpha : i \to j, \\ -v_2(\alpha) & \text{if } \exists \text{ positive } \alpha : j \to i, \\ -t v_1(\alpha) & \text{if } \exists \text{ negative } \alpha : i \to j, \\ t v_2(\alpha) & \text{if } \exists \text{ negative } \alpha : j \to i, \\ 0 & \text{if there is no arrow between } i \text{ and } j. \end{cases}$$

Example 2.14. The signed valued quivers associated to the matrices B and $\mu_2(B)$ in Example 2.5 are



Proposition 2.15. *The two constructions in Definition 2.13 are mutually inverse.*

Proof. Start with a pure gss matrix $B = (b_{ij})$. Let (Q, v) be its associated signed valued quiver, and let $A = (a_{ij})$ be the associated gss matrix of (Q, v). If $b_{ij} = 0$ then also $b_{ji} = 0$, so there is no arrow between i and j, so $a_{ij} = 0$. If $b_{ij} \neq 0$ and $\operatorname{sgn}(b_{ij}) = 1$, we get an arrow $\alpha : i \to j$ in Q with $v(\alpha) = (b_{ij}(-1), -b_{ji}(-1))$. If $b_{ij} \in \mathbb{Z}$ then $v_1(\alpha) = b_{ij}$. As $\operatorname{sgn}(b_{ij}) = 1$ we have $b_{ij} > 0$, so α is positive, so $a_{ij} = b_{ij}$. If $b_{ij} = bt \in t\mathbb{Z}$ then $v_1(\alpha) = -b < 0$ because $\operatorname{sgn}(b_{ij}) = 1$. So α is negative, so $a_{ij} = (-t)(-b) = bt = b_{ij}$. If $b_{ij} \neq 0$ and $\operatorname{sgn}(b_{ij}) = -1$, just swap i and j.

Now start with a signed valued quiver (Q, v). Construct the associated gss matrix $B = (b_{ij})$ and its associated signed valued quiver (R, w). If Q has no arrow between i and j then $b_{ij} = 0$, so neither does R. If $\alpha : i \to j$ in Q is positive then $b_{ij} = v_1(\alpha)$ is a positive integer, and $b_{ji} = -v_2(\alpha) \in \mathbb{Z}$. So we get an arrow $\alpha : i \to j$ in R and $w(\alpha) = (v_1(\alpha), v_2(\alpha))$. If instead $\alpha : i \to j$ in Q is negative then $b_{ij} = -t v_1(\alpha)$ is a positive multiple of t, and $b_{ji} = t v_2(\alpha) \in \mathbb{Z}$. So we get an arrow $\alpha : i \to j$ in R and again $w(\alpha) = (v_1(\alpha), v_2(\alpha))$. □

We would like to define mutation of signed valued quivers by first constructing the associated (pure) gss matrix, then applying Definition 2.4, and then taking the associated signed valued quiver. However, as the following example shows, the mutation of a pure gss matrix may not be pure.

Example 2.16. Let Q be an oriented 3-cycle and give all arrows value $(-1, -1)$. Then the associated gss matrix B and the signed mutation $B' = \mu_2(B)$ are

$$B = \begin{pmatrix} 0 & t & -t \\ -t & 0 & t \\ t & -t & 0 \end{pmatrix} \quad \text{and} \quad B' = \begin{pmatrix} 0 & -1 & -1+t \\ 1 & 0 & -t \\ 1-t & t & 0 \end{pmatrix}.$$

Definition 2.17. We say that a pure gss matrix B satisfies the *positive* 3*-cycle condition at* k if, for all $i, j \neq k$ with $\operatorname{sgn}(b_{ik}b_{kj}) = 1$, we have

$$b_{ij}b_{ik}b_{kj} \in \mathbb{Z}.$$

Remark 2.18. In terms of the associated signed valued quiver of B, the positive 3-cycle condition at k says that, whenever we have arrows $i \to k \to j$ and an arrow between i and j in either direction, the product of the signs of the three arrows is positive. We will apply the condition to signed valued quivers below.

Lemma 2.19. *Let* $B = (b_{ij})$ *be a pure gss matrix. Then* $B' = \mu_k(B) = (b'_{ij})$ *is pure if and only if* B *satisfies the positive* 3*-cycle condition at* k.

Proof. Since B is pure, we have $b'_{ij} \in \mathbb{Z} \cup t\mathbb{Z}$ if $k = i, j$ by definition of μ_k. This is also true if $\operatorname{sgn}(b_{ik}b_{kj}) \neq 1$, and if $\operatorname{sgn}(b_{ik}b_{kj}) = 1$ and $b_{ij} = 0$. So assume $\operatorname{sgn}(b_{ik}b_{kj}) = 1$ and $b_{ij} \neq 0$. Then B' is pure if and only if

$$b'_{ij} = t(b_{ij} + \operatorname{sgn}(b_{ik})b_{ik}b_{kj}) \in \mathbb{Z} \cup t\mathbb{Z}$$

for all i, j such that $\operatorname{sgn}(b_{ik}b_{kj}) = 1$. Consider the $\mathbb{Z}$ and $t\mathbb{Z}$ situations separately:

- $b'_{ij} \in \mathbb{Z}$ if and only if ($b_{ij}, b_{ik} \in t\mathbb{Z}$ and $b_{kj} \in \mathbb{Z}$) or ($b_{ij}, b_{kj} \in t\mathbb{Z}$ and $b_{ik} \in \mathbb{Z}$);
- $b'_{ij} \in t\mathbb{Z}$ if and only if ($b_{ij}, b_{ik}, b_{kj} \in \mathbb{Z}$) or ($b_{ij} \in \mathbb{Z}$ and $b_{ik}, b_{kj} \in t\mathbb{Z}$).

The cases described above reduce to $b_{ij}b_{ik}b_{kj} \in \mathbb{Z}$. □

Remark 2.20. Under the assumptions of Lemma 2.19, the pure matrix $B' = \mu_k(B)$ also satisfies the positive 3-cycle condition at k. This is a straightforward consequence of the proof above. For example, if $b'_{ij} \in \mathbb{Z}$ and $b_{ij}, b_{ik} \in t\mathbb{Z}$ and $b_{kj} \in \mathbb{Z}$, Definition 2.4 yields $b'_{ik}b'_{kj} \in \mathbb{Z}$, and so $b'_{ij}b'_{ik}b'_{kj} \in \mathbb{Z}$.

Before describing mutation of signed valued quivers, it helps to introduce some terminology.

Definition 2.21. For arrows $\alpha = i \xrightarrow{(a,b)} j$ and $\beta = j \xrightarrow{(c,d)} h$, we define the following operations:

- *reverse* α to get $\alpha^* : j \xrightarrow{(b,a)} i$;
- *negate* α to get $-\alpha : i \xrightarrow{(-a,-b)} j$;
- *compose* α and β to get $[\alpha\beta] : i \xrightarrow{(ac,\,bd)} h$.

Definition 2.22. For arrows $\alpha : i \xrightarrow{(a,b)} j$ and $\beta : j \xrightarrow{(c,d)} i$, we say that α is:

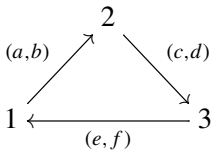
- *smaller* than β if $|a| < |d|$ and $|b| < |c|$;
- *bigger* than β if $|a| > |d|$ and $|b| > |c|$;
- *the same size* as β if $a = d$ and $b = c$.

We say that α is *comparable* to β if one of the three conditions above hold.

For the rest of this section, let (Q, v) denote a signed valued quiver.

Lemma 2.23. *For an oriented* 3*-cycle* $1 \xrightarrow{\alpha} 2 \xrightarrow{\beta} 3 \xrightarrow{\gamma} 1$ *in* (Q, v), *where* $[\alpha\beta\gamma]$ *is positive*, *the composed arrow* $[\alpha\beta]$ *is comparable to* γ.

Proof. Suppose our arrows have the following values:



Let $D = \operatorname{diag}(d_1, d_2, d_3)$ be the skew-symmetrizing matrix of Q, so that

$$d_1 a = b d_2, \quad d_2 c = d d_3, \quad d_3 e = f d_1,$$

with $d_1, d_2, d_3 > 0$. We have

$$|ac| - |f| = |bd| d_3/d_1 - |e| d_3/d_1 = \frac{d_3}{d_1}(|bd| - |e|),$$

so $|ac| > |f| \iff |bd| > |e|$ and $|ac| < |f| \iff |bd| < |e|$. Moreover, if $|ac| = |f|$ and $|bd| = |e|$ then, because $[\alpha\beta\gamma]$ is positive, $[\alpha\beta]$ and γ must have the same sign, so $ac = f$ and $bd = e$. Therefore we are in one of the cases of Definition 2.22. □

We now give the mutation procedure for signed valued quivers. Lemma 2.23 ensures that the second part of Step 3 always makes sense.

Definition 2.24. Let (Q, v) be a signed valued quiver which satisfies the positive 3-cycle condition at $k \in Q_0$. Define the *signed quiver mutation* $\mu_k(Q, v)$ as follows:

Step 1: If there are arrows $i \xrightarrow{\alpha} k \xrightarrow{\beta} j$, add a new arrow $-[\alpha\beta] : i \to j$ and replace any arrow between i and j with its negation.

Step 2: Reverse any arrow going out of k, and reverse and negate any arrow going into k.

Step 3: If we have two arrows γ, δ between i and j as a result of Step 1, then:

- if $\gamma, \delta : i \to j$, replace them with a single arrow having value $v(\gamma) + v(\delta)$;
- if $\gamma : j \to i$ and $\delta : i \to j$, reverse and negate the smaller (or same size) arrow, and then combine them as above.

The diagram in Figure 1 exhibits the mutation explicitly; we assume $a, b, c, d \neq 0$.

Recall that $B(Q, v)$ denotes the associated gss matrix of a signed valued quiver (Q, v).

Proposition 2.25. *If* (Q, v) *satisfies the positive* 3*-cycle condition at* k, *then*

$$\mu_k B(Q, v) = B \mu_k(Q, v).$$

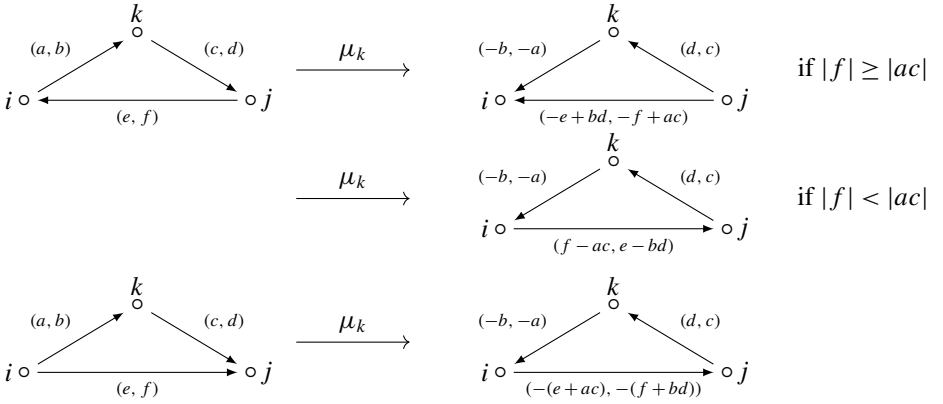


Figure 1. The signed quiver mutation $\mu_k(Q, v)$.

Proof. This is a case-by-case check using Definitions 2.4, 2.24, and 2.13. We do not write out all the details; instead we write out some checks in full and then list what other checks need to be done.

If Q has a positive arrow $\alpha : i \to k$ with $v(\alpha) = (a, b)$, then $b_{ik} = v_1(\alpha) = a > 0$ and $b_{ki} = -v_2(\alpha) = -b < 0$. So $b'_{ik} = -tb_{ik} = -ta$ and $b'_{ki} = -tb_{ki} = tb$. As $\operatorname{sgn}(b'_{ki}) = +1$ we get an arrow $k \to i$ with value $(b'_{ki}(-1),\ -b'_{ik}(-1)) = (-b, -a)$. So α has been negated and reversed, as instructed by Step 2. One checks that the same thing happens if α is negative. Then one can check that arrows $\beta : k \to j$, of positive or negative sign, are reversed. Next one can check that if Q has an arrow $\gamma : i \to j$ and there do not exist paths $i \to k \to j$ or $j \to k \to i$ in Q, then γ is unchanged.

Now suppose that Q has arrows $\alpha : i \to k$, $\beta : k \to j$, and $\gamma : i \to j$, all positive, with $v(\alpha) = (a, b)$, $v(\beta) = (c, d)$, and $v(\gamma) = (e, f)$. Then

$$b'_{ij} = t(b_{ij} + b_{ik}b_{kj}) = t(e + ac);$$
$$b'_{ji} = t(b_{ji} - b_{jk}b_{ki}) = -t(f + bd).$$

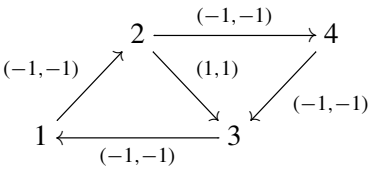
As $\operatorname{sgn}(b'_{ij}) = +1$ we get an arrow $i \to j$ with value $(-e - ac,\ -f - bd)$. This matches the first part of Step 3: we have replaced the arrows $-\gamma$ and $-[\alpha\beta]$ from Step 1 with a single arrow, summing their values. One can do similar checks where α, β, γ have positive or negative sign, subject to the positive 3-cycle condition, and get the same result in all cases. This argument also works if there are no arrows between i and j.

Finally suppose that Q has arrows $\alpha : i \to k$, $\beta : k \to j$, and $\gamma : j \to i$. If γ is positive then, in the mutation of the associated gss matrix, $b'_{ij} = t(ac - f)$ and $b'_{ji} = t(e - bd)$. If instead γ is negative then $b'_{ij} = f - ac$ and $b'_{ji} = bd - e$. The associated signed valued quiver depends on whether $|ac| > |f|$ or $|ac| \leq |f|$: in

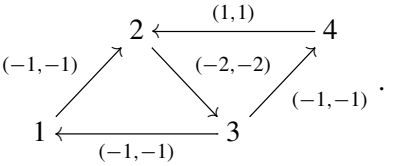
the former case we get an arrow $i \to j$ with value $(f - ac, e - bd)$, irrespective of the sign of γ, and in the latter case we get $j \to i$ with value $(bd - e, ac - f)$. This matches the second part of Step 3. □

Example 2.26. Mutating the first signed valued quiver in Example 2.14 at vertex 2 gives the second.

Example 2.27. Consider the signed valued quiver



The positive 3-cycle condition is satisfied at all vertices. We mutate at vertex 4 to get the following signed valued quiver:



.

Now the 3-cycle $1 \to 2 \to 3 \to 1$ has negative product of signs, so the above signed valued quiver does not satisfy the positive 3-cycle condition at vertices 1, 2, or 3.

2C. *Mutation Dynkin type.*

Definition 2.28. We say that two gss matrices B and B' are *mutation equivalent* if there exists a sequence of indices $i_1, \ldots, i_r$ such that $B' = \mu_{i_r} \ldots \mu_{i_1} B$. The *mutation class* of B is the set of all gss matrices that are mutation equivalent to B.

If (Q, v) is a signed valued quiver, we can apply the previous definition by taking the associated gss matrix $B = B(Q, v)$. But, in general, the mutation class of a signed valued quiver may include gss matrices which are not pure, and so have no associated signed valued quiver. We will soon see that this problem does not occur for the appropriate notion of Dynkin signed valued quivers.

Proposition 2.29. *Let (Q, v) be a signed valued quiver whose underlying graph is a tree. Let (Q', v') be obtained by reversing and/or negating some arrows of (Q, v). Then (Q, v) and (Q', v') are mutation equivalent.*

Proof. It is enough to show we can reverse or negate a single arrow. We start with negation. As the underlying graph of Q is a tree, there is a unique (undirected) path between any two vertices. Given an arrow $\alpha : i \to j$ in Q, let

$$P_-(\alpha) = \{k \in Q_0 \mid \text{the unique path from } k \text{ to } i \text{ does not pass through } j\}$$

and write this as the disjoint union

$$P_-(\alpha) = P_-^0(\alpha) \cup P_-^1(\alpha),$$

where $P_-^0(\alpha)$ and $P_-^1(\alpha)$ denote the subsets where the path has even and odd length, respectively. By Lemma 2.9, μ_k^2 just changes the sign of all arrows incident to k. Note that $\mu_h^2\mu_k^2 = \mu_k^2\mu_h^2$ for any $h, k \in Q_0$. So

$$\Big(\prod_{k\in P_-(\alpha)} \mu_k^2\Big)(Q, v) = \Big(\prod_{k\in P_-^1(\alpha)} \mu_k^2\Big)\Big(\prod_{k\in P_-^0(\alpha)} \mu_k^2\Big)(Q, v)$$

is (Q, v) with α negated, because $\prod_{k\in P_-^0(\alpha)} \mu_k^2$ changes the sign of all arrows incident with $P_-(\alpha)$, and $\prod_{k\in P_-^1(\alpha)} \mu_k^2$ changes the sign of all arrows incident with $P_-(\alpha) - \{i\}$.

Now we consider reversing arrows. If Q' is another orientation of Q, we know from classical results that the (unsigned) quivers Q and Q' are mutation equivalent via mutations only at sources (see, e.g., [11, Proposition 2.2.8] for a proof). From Definition 2.24, source mutations just reverse arrows, so this sequence of signed mutations gives the result. □

We want a data structure that includes the classical Dynkin diagrams. We will use graphs with a choice of ordering $i < j$, $i = j$, or $i > j$ whenever i and j have at least 2 edges between them.

Definition 2.30. The *unsigned diagram* $\tilde{\Gamma} = \tilde{\Gamma}(Q, v)$ of a signed valued quiver (Q, v) has the same vertex set as Q. If Q has an arrow $\alpha : i \to j$ then $\tilde{\Gamma}$ has $v_1(\alpha)v_2(\alpha)$ edges between i and j, with ordering $i > j \iff |v_1(\alpha)| > |v_2(\alpha)|$ and $i < j \iff |v_1(\alpha)| < |v_2(\alpha)|$.

Recall the Dynkin diagrams of types A_n, B_n, C_n, D_n, E_n, F_4, and G_2: see, e.g., [15, Section V.15].

Definition 2.31. A gss matrix is called *Dynkin* if it is pure and the unsigned diagram of its associated signed valued quiver is a Dynkin diagram. A gss matrix is called *mutation Dynkin* if it is mutation equivalent to a Dynkin gss matrix.

Proposition 2.29 tells us that any two gss matrices which are Dynkin and of the same type are mutation equivalent. The mutation classes of different Dynkin types are distinct: this follows from the classical result (see [8, Theorem 1.7]) together with Proposition 2.6.

Example 2.32. The signed valued quiver

$$1 \xrightarrow{(1,1)} 2 \xleftarrow{(-1,-1)} 3 \xleftarrow{(1,2)} 4$$

has unsigned diagram

$$1 \text{ —— } 2 \text{ —— } 3 \Longrightarrow 4$$

and so is Dynkin of type B_4. The signed valued quiver

$$1 \xrightarrow{(2,2)} 2$$

is not mutation Dynkin.

Definition 2.33. Let B be a gss matrix. A *p-cycle* in B is a p-tuple $(i_1, \dots, i_p)$ with $i_j \in \{1, \dots, n\}$ distinct such that $b_{i_j, i_{j+1}} \neq 0$ for all $j \in \{1, \dots, p\}$ (we interpret all subscripts modulo p).

The cycle is called:

- *oriented* if $\operatorname{sgn}(b_{i_j, i_{j+1}}) = 1$ for all $j \in \{1, \dots, p\}$;
- *chordless* if for all $j \in \{1, \dots, p\}$ and all $k \in \{1, \dots, p\} \setminus \{j+1, j-1\}$ we have $b_{i_j i_k} = 0$.

Note that if B is pure then a chordless cycle in B corresponds to a chordless cycle in Q_B, as defined in the introduction of [1].

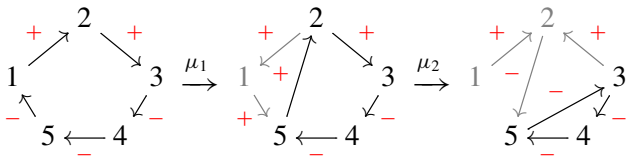
The following definition is reminiscent of the condition in [1, Proposition 1.4].

Definition 2.34. A p-cycle $(i_1, \dots, i_p)$ in a pure gss matrix B is called *dangerous* if it is chordless and

$$\prod_{j=1}^{p} b_{i_j, i_{j+1}} \in t^p \mathbb{Z}.$$

Remark 2.35. A dangerous 3-cycle must contain a vertex which does not satisfy the positive 3-cycle condition, and therefore has a mutation which is not pure. We will prove that a similar phenomenon holds for all dangerous p-cycles, but first we illustrate this graphically.

Example 2.36. The first quiver below has a dangerous oriented 5-cycle $(1, 2, 3, 4, 5)$. We construct a dangerous oriented 3-cycle $(3, 4, 5)$ by mutating twice:



Lemma 2.37. *Let B be a pure gss matrix with a dangerous oriented p-cycle $(i_1, \dots, i_p)$. Then*

$$\mu_{i_{p-2}} \mu_{i_{p-3}} \dots \mu_{i_1}(B)$$

is not pure.

Proof. Let $B^{(h)} = (b_{ij}^{(h)}) = \mu_{i_h} \dots \mu_{i_1}(B)$ for $h = 0, \dots, p-3$. We want to show $B^{(p-2)}$ is not pure.

We claim that $B^{(h)}$ has a dangerous oriented $(p-h)$-cycle $(i_{h+1}, \dots, i_{p-1}, i_p)$ for all $h = 0, \dots, p-3$. The base case for $B^{(0)}$ is true by assumption, so assume it holds for $B^{(h)}$, and we need to show it holds for $B^{(h+1)} = \mu_{h+1} B^{(h)}$.

As the cycle $(i_{h+1}, \dots, i_{p-1}, i_p)$ in $B^{(h)}$ is chordless, we have $b^{(h+1)}_{i_j i_{j+1}} = b^{(h)}_{i_j i_{j+1}}$ for $j = h+1, \dots, p-1$. as we are always in the fourth case of Definition 2.4. Then using the third case we get

$$b^{(h+1)}_{i_p i_{h+2}} = t(0 + 1(b^{(h)}_{i_p i_{h+1}} b^{(h)}_{i_{h+1} i_{h+2}})) = t b^{(h)}_{i_p i_{h+1}} b^{(h)}_{i_{h+1} i_{h+2}}$$

so we have shown that $(i_{h+2}, \dots, i_{p-1}, i_p)$ is a chordless oriented $(p-h-1)$-cycle, and it follows directly that

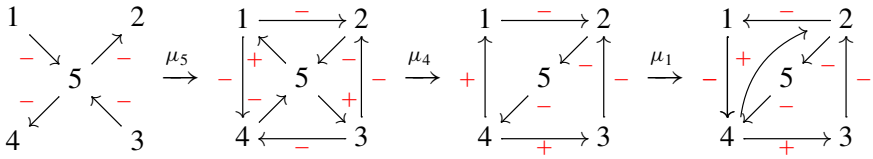
$$\left(\prod_{j=h+2}^{p-1} b^{(h+1)}_{i_j, i_{j+1}} \right) b^{(h+1)}_{i_p i_{h+2}} = t \left(\prod_{j=h+1}^{p-1} b^{(h)}_{i_j, i_{j+1}} \right) b^{(h)}_{i_p i_{h+1}} \in t(t^{p-h}\mathbb{Z}) = t^{p-h-1}\mathbb{Z},$$

showing that our cycle is dangerous.

Thus $B^{(p-3)}$ has a dangerous 3-cycle so, by Lemma 2.19, $B^{(p-2)}$ is not pure. □

Remark 2.38. One can show that any pure gss matrix with a dangerous p-cycle is mutation equivalent to a gss matrix which is not pure, whether the cycle is oriented or not. First, by mutating if necessary, we get some $1 \le j \le p$ with $\operatorname{sgn}(b_{i_{j-1} i_j}) = \operatorname{sgn}(b_{i_j i_{j+1}}) = 1$. Then, as in the proof of Lemma 2.37, we mutate at j to create a dangerous $(p-1)$-cycle, and continue this process.

Example 2.39. It is not necessary to have a dangerous cycle in order to be mutation equivalent to a gss matrix which is not pure. The following valued quiver is acyclic, but mutating at the vertex 5 creates a dangerous 4-cycle (1, 2, 3, 4). Then, following the procedure described in Remark 2.38, we mutate at 4 to give us composable arrows passing through 1, then mutate at 1 to get our dangerous (unoriented) 3-cycle (2, 3, 4). Then the vertex 3 does not satisfy the positive 3-cycle condition, so mutating at 3 would give a gss matrix which is not pure.



We now record some useful results about cycles in mutation Dynkin valued quivers.

Lemma 2.40. *Let (Q, v) be a mutation Dynkin (unsigned) valued quiver. Then*:

(1) *Every chordless cycle in Q is oriented.*

(2) *For any 3-cycle $C = (i, j, k)$ in Q, the values of the arrows in C are either* $\{(1, 1), (1, 1), (1, 1)\}$ *or* $\{(1, 1), (1, 2), (2, 1)\}$, *ordered clockwise or anticlockwise.*

Proof. For (1), by [1, Theorem 1.2] we know that every chordless cycle in a skew-symmetrizable matrix of mutation Dynkin type is oriented. As for (2), by [8, Lemma 7.6] we know that the edge weights for a 3-cycle are $\{1, 1, 1\}$ or $\{1, 2, 2\}$. But the values $\{(1, 1), (1, 2), (1, 2)\}$ are not skew-symmetrizable. □

Proposition 2.41. *If the specialization $B(1)$ of B is mutation Dynkin, and B contains no dangerous cycles, then every gss matrix mutation equivalent to B is pure.*

Proof. Suppose, for contradiction, that there is a mutation sequence $(k_1, \ldots, k_\ell)$ such that

$$B^{(\ell)} := \mu_{k_\ell} \ldots \mu_{k_1} B = (b_{ij}^{(\ell)})$$

is not pure. Let ℓ be minimal, so any mutation sequence of length smaller than ℓ yields a pure gss matrix. We will prove by downward induction on $m = \ell - 1, \ldots, 0$ that $B^{(m)}$ has a dangerous oriented $(\ell - m + 2)$-cycle. The case $m = 0$ will give our contradiction.

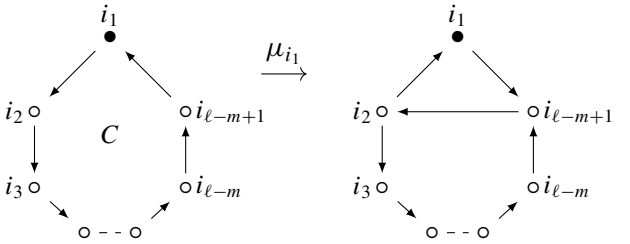
For the base case, take $m = \ell - 1$. As $B^{(\ell)} = \mu_{k_\ell} B^{(\ell-1)}$ is not pure, then $B^{(\ell-1)}$ does not satisfy the positive 3-cycle condition at k_ℓ, and so it contains a dangerous 3-cycle by Remark 2.35. Such a 3-cycle must be oriented by Lemma 2.40.

Now suppose that $m < \ell - 1$ and assume that

$$B^{(m+1)} \text{ contains a dangerous oriented } (\ell - m + 1)\text{-cycle } C = (i_1, \ldots, i_{\ell-m+1}).$$

Our proof has two steps: first we will show that the next vertex we mutate at does not belong to C, i.e., that $k_{m+1} \notin \{i_1, \ldots, i_{\ell-m+1}\}$. Then we show that $B^{(m)}$ has a dangerous oriented $(\ell - m + 2)$-cycle on the vertices of C together with k_{m+1}.

Step 1: Suppose for contradiction that $k_{m+1} \in \{i_1, \ldots, i_{\ell-m+1}\}$ and, without loss of generality, $k_{m+1} = i_1$. By Lemma 2.40(i), C is oriented, and we can represent the specializations $C(1)$ and $\big(\mu_{i_1}(C)\big)(1)$ as in the following diagram:



By Lemma 2.37, we have that $\mu_{i_{\ell-m-1}} \ldots \mu_{i_1}(B^{(m+1)})$ is not pure. Since $B^{(m+1)} = \mu_{k_{m+1}} B^{(m)}$ and $k_{m+1} = i_1$, we have

$$\mu_{i_{\ell-m-1}} \ldots \mu_{i_1}(B^{(m+1)}) = \mu_{i_{\ell-m-1}} \ldots \mu_{i_2}(\mu_{i_1})^2(B^{(m)}).$$

But since $\mu_{i_1}^2$ just changes the sign of arrows incident to i_1 by Lemma 2.9, we also

have that $\mu_{i_{\ell-m-1}} \dots \mu_{i_2}(B^{(m)})$ is not pure. But now

$$\mu_{i_{\ell-m-1}} \dots \mu_{i_2}(B^{(m)}) = \mu_{i_{\ell-m-1}} \dots \mu_{i_2} \mu_{k_m} \dots \mu_{k_1}(B)$$

exhibits a mutation sequence of length $\ell - 2$ from B to a non-pure matrix, contradicting the minimality of ℓ.

Step 2: Next we want to show that $B^{(m)}$ has a dangerous oriented $(\ell - m + 2)$-cycle. This is equivalent to such a cycle existing for $\mu_{k_{m+1}}(B^{(m+1)}) = \mu^2_{k_{m+1}}(B^{(m)})$, because applying $\mu_{k_{m+1}}$ twice has the only effect of changing the sign of the arrows incident with k_{m+1} by Lemma 2.9.

To simplify notation, let

$$k = k_{m+1}, \quad B^{(m+1)} = (b_{ij}), \quad \mu_k(B^{(m+1)}) = (b'_{ij}), \quad \text{and} \quad a_h := b_{i_h,k}$$

for $1 \le h \le \ell - m + 1$. We will show that exactly two a_h are nonzero, and they are consecutive, of the form a_j and a_{j+1}. We do this by ruling out all other possibilities. Then we will show that $(i_1, \dots, i_j, k_{m+1}, i_{j+1}, \dots i_{\ell-m+1})$ is a dangerous oriented $(\ell - m + 2)$-cycle in $B^{(m)}$.

• First suppose for contradiction that at most one a_h is nonzero. Then, for $h \neq j$, we have $b_{i_h,i_k} b_{i_k,i_j} = -a_h a_j = 0$. Consider the values of $b'_{i_h,i_{h+1}}$. We showed in Step 1 that $k \neq i_h$ for all h, so we are never in the first two cases of Definition 2.4, and as $a_h a_j = 0$ we are never in the third case. Thus

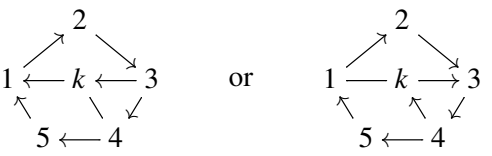
$$b'_{i_h,i_{h+1}} = b_{i_h,i_{h+1}} \quad \text{for all } 1 \le h \le \ell - m + 1. \tag{1}$$

By assumption, $B^{(m+1)}$ contains a dangerous oriented $(\ell - m + 1)$-cycle $C = (i_1, \dots, i_{\ell-m+1})$, and therefore $\mu_k(B^{(m+1)}) = \mu^2_k(B^{(m)})$ contains the same dangerous cycle by (1). But then, by Lemma 2.9, $B^{(m)}$ contains a dangerous oriented $(\ell - m + 1)$-cycle, so by Lemma 2.37 the matrix

$$\mu_{i_{\ell-m-1}} \dots \mu_{i_1}(B^{(m)}) = \mu_{i_{\ell-m-1}} \dots \mu_{i_1} \mu_{k_m} \dots \mu_{k_1}(B)$$

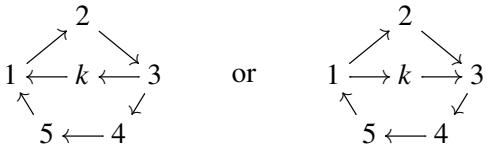
is not pure. But this mutation sequence has length $\ell - 1$, contradicting the minimality of ℓ. Thus we can assume at least two of the a_h's are nonzero.

• Now suppose for contradiction that there exist $h_1 < h_2 < h_3$ in C such that $a_{h_1}, a_{h_2}, a_{h_3} \neq 0$ and $a_j = 0$ for all j such that $h_1 < j < h_2$ and $h_2 < j < h_3$. Then at least one of the chordless cycles $(i_{h_1}, \dots, i_{h_2}, k)$ or $(i_{h_2}, \dots, i_{h_3}, k)$ is not oriented, as illustrated in the following picture:



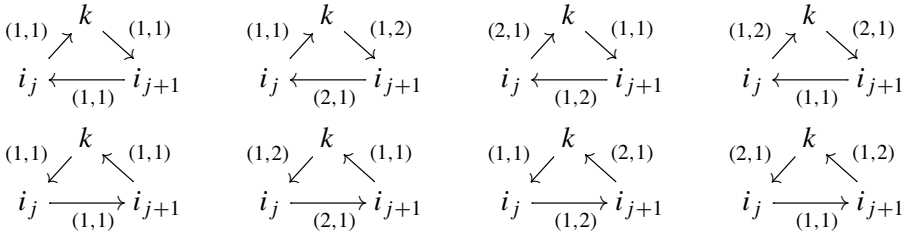
This contradicts Lemma 2.40, and therefore cannot happen.

• The only case left to exclude is when there is some $h_1 < h_2$ with $h_2 - h_1 > 1$ such that $a_{h_1}, a_{h_2} \neq 0$ and $a_j = 0$ for all $j \neq h_1, h_2$. This case is similar to the one examined above: as shown in the figure below, at least one of the cycles $(i_{h_1}, i_{h_1+1}, \ldots, i_{h_2}, k)$ or $(i_{h_2}, i_{h_2+1}, \ldots, i_{h_3}, k)$ is nonoriented in $B^{(m+1)}$, and therefore this case can never happen.



So we have shown that $a_j, a_{j+1} \neq 0$, and $a_h = 0$ for $h \neq j, j+1$.

Hence, by Lemma 2.40(ii), the valued quiver associated to the cycle (i_j, i_{j+1}, k) in $B^{(m+1)}(1)$ is one of the following:



We can assume that $b_{i_j,i_{j+1}} b_{i_j,k} b_{k,i_{j+1}} \in \mathbb{Z}$, since otherwise we could use Lemma 2.37 to contradict the minimality of ℓ. Furthermore, the two rightmost cycles (i_j, i_{j+1}, k) in the picture cannot appear in $B^{(m+1)}$, since otherwise $b'_{i_j,i_{j+1}} = t b_{i_j,i_{j+1}}$, and the chordless cycle $(i_1, \ldots, i_{\ell-m+1})$ in $\mu_k(B^{(m+1)})$ would not be oriented, contradicting Lemma 2.40(i). The other valued quivers yield $b'_{i_j,i_{j+1}} = 0$, and therefore $(i_1, \ldots, i_j, k, i_{j+1}, \ldots, i_{\ell-m+1})$ is a chordless $(\ell - m + 2)$-cycle in $\mu_k(B^{(m+1)})$, and thus in $B^{(m)}$.

Now we have found our cycle in $B^{(m)}$. It just remains to show that the product of its signs is $(-1)^{\ell-m}$.

By the usual minimality argument, the 3-cycle (i_j, k, i_{j+1}) in $B^{(m+1)}$ must be oriented. So we have $b_{i_j,k} b_{k,i_{j+1}} b_{i_{j+1},i_j} \in \mathbb{Z}$. By assumption, we have

$$\prod_{j=1}^{\ell-m+1} b_{i_j,i_{j+1}} \in t^{\ell-m+1}\mathbb{Z};$$

so together we get

$$\left(\prod_{h=1}^{j-1} b_{i_h,i_{h+1}}\right) b_{i_j,k} b_{k,i_{j+1}} \left(\prod_{h=j+1}^{\ell-m-1} b_{i_h,i_{h+1}}\right) \in t^{\ell-m+1}\mathbb{Z}.$$

Next, since $a_h = 0$ for $h \neq j, j+1$ we have $b'_{i_h,i_{h+1}} = b_{i_h,i_{h+1}}$ when $h \neq j$, and Definition 2.4 gives $b'_{i_j,k} = -tb_{i_j,k}$ and $b'_{k,i_{j+1}} = -b_{k,i_{j+1}}$.

Therefore

$$\left(\prod_{h=1}^{j-1} b'_{i_h,i_{h+1}}\right) b'_{i_j,k} b'_{k,i_{j+1}} \left(\prod_{h=j+1}^{\ell-m-1} b'_{i_h,i_{h+1}}\right) = t\left(\prod_{h=1}^{j-1} b_{i_h,i_{h+1}}\right) b_{i_j,k} b_{k,i_{j+1}} \left(\prod_{h=j+1}^{\ell-m-1} b_{i_h,i_{h+1}}\right) \in t^{\ell-m+1}\mathbb{Z};$$

so our chordless $(\ell-m+2)$-cycle in $B^{(m)}$ is dangerous and our downwards induction is complete. □

Theorem 2.42. *Every mutation Dynkin gss matrix is pure.*

Proof. By definition, a mutation Dynkin gss matrix is mutation equivalent to a pure gss matrix B such that $B(1)$ is Dynkin. As every Dynkin diagram is a tree, B has no dangerous cycles, so the result follows by Proposition 2.41. □

Theorem 2.43. *The mutation class of Dynkin type Δ consists of all gss matrices B such that $B(1)$ is of (unsigned) mutation type Δ and B contains no dangerous cycles.*

Proof. If B is mutation Dynkin then $B(1)$ must be of (unsigned) mutation type Δ by Proposition 2.6. Then B cannot contain a dangerous cycle, as the cycle would be oriented by Lemma 2.40, and Lemma 2.37 would say that B is mutation equivalent to a non-pure matrix, contradicting Theorem 2.42.

Now suppose that B is a gss matrix with no dangerous cycles and that $\tilde{B} = B(1)$ is mutation Dynkin. Then there exists a sequence $i_1, \ldots, i_n$ such that $\tilde{B}' = \mu^{\mathrm{FZ}}_{i_k} \ldots \mu^{\mathrm{FZ}}_{i_1} \tilde{B}$ is Dynkin. Define the gss matrix $B' = \mu_{i_k} \ldots \mu_{i_1} B$, so $B'(1) = \tilde{B}'$. Then B' is pure by Proposition 2.41, so it is mutation Dynkin, and therefore so is $\mu^3_{i_1} \ldots \mu^3_{i_k} B'$. But, by Proposition 2.10, this is B. □

3. Roots

3A. ***The Cartan counterpart.*** Given a skew-symmetrizable matrix $\tilde{B}$, its *Cartan counterpart* is the $n \times n$ matrix $\tilde{C}(\tilde{B}) = (\tilde{c}_{ij})$ with $\tilde{c}_{ii} = 2$ and $\tilde{c}_{ij} = -|\tilde{b}_{ij}|$: see [8, Section 1.3], following [7, Remark 4.6].

We revise some definitions from the introduction of [1]. Let $\tilde{C}$ denote a symmetrizable matrix (so $D\tilde{C} = (D\tilde{C})^T$ for D diagonal) and $\tilde{B}$ a skew-symmetrizable matrix. We say $\tilde{C}$ is

- *quasi-Cartan* if its diagonal entries $\tilde{c}_{ii}$ are all equal to 2;
- *positive* if the symmetrized matrix $D\tilde{C}$ is positive definite;
- a *quasi-Cartan companion* of $\tilde{B}$ if it is quasi-Cartan and $|\tilde{c}_{ij}| = |\tilde{b}_{ij}|$ for all $i \neq j$.

The Cartan counterpart of a skew-symmetrizable matrix $\tilde{B}$ is an example of a quasi-Cartan companion of $\tilde{B}$, but it will not be positive in general, even if $\tilde{B}$ is mutation Dynkin: for example, if we start with a linearly oriented A_3 quiver and mutate to get a 3-cycle, the Cartan counterpart has determinant 0 and so cannot be positive.

In this section we give a signed variant of the Cartan counterpart.

Definition 3.1. Given a gss matrix $B = (b_{ij})$, its *Cartan counterpart* is the $n \times n$ integer matrix $C(B) = (c_{ij})$, where

$$c_{ij} = \begin{cases} 2 & \text{if } i = j, \\ |a| - |b| & \text{if } i \neq j \text{ and } b_{ij} = a + bt. \end{cases}$$

For $C = C(B)$, note that $DB = -(BD)^T$ implies $DC = (CD)^T$.

The following two results are immediate from the definitions.

Lemma 3.2. *If B is a pure gss matrix, $C(B)$ is a quasi-Cartan companion of $B(1)$.*

Proposition 3.3. *If $\tilde{B}$ is a skew-symmetrizable matrix then $\tilde{C}(\tilde{B}) = C(t\tilde{B})$.*

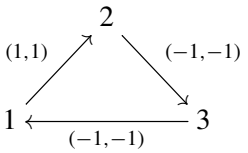
Remark 3.4. We interpret Proposition 3.3 as saying that the canonical way to embed valued quivers into signed valued quivers is to give all arrows negative sign.

If (Q, v) is a signed valued quiver, write $C(Q, v) = C(B(Q, v))$. The following result is immediate from Definitions 2.13 and 3.1.

Lemma 3.5. *If (Q, v) is a signed valued quiver then $C(Q, v) = (c_{ij})$, where*

$$c_{ij} = \begin{cases} 2 & \text{if } i = j, \\ v_1(\alpha) & \text{if } \exists \alpha : i \to j, \\ v_2(\alpha) & \text{if } \exists \alpha : j \to i, \\ 0 & \text{if there is no arrow between } i \text{ and } j. \end{cases}$$

Example 3.6. Consider the following signed valued quiver of mutation type A_3:



Its associated gss matrix and Cartan counterpart are as follows.

$$B = \begin{pmatrix} 0 & 1 & -t \\ -1 & 0 & t \\ t & -t & 0 \end{pmatrix}, \qquad C(B) = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

We give a signed variant of Definition 2.30.

Definition 3.7. The *signed diagram* $\Gamma = \Gamma(Q, v)$ of a signed valued quiver (Q, v) is the unsigned diagram $\tilde{\Gamma}(Q, v)$ together with a sign $\varepsilon_{ij} = \varepsilon_{ji}$ attached to each unordered pair $\{i, j\}$ of vertices which are joined by at least one edge. If Q has an arrow $\alpha : i \to j$ then the sign of $\{i, j\}$ is $\varepsilon_{ij} = \operatorname{sgn}(v_1(\alpha))$.

For a signed valued quiver of mutation type Dynkin, the next result shows that its Cartan counterpart and the associated signed diagram capture the same information. Note that a similar definition of the diagram of a quasi-Cartan matrix appears in [1, Definition 2.2], without recording the ordering on multiple edges.

Lemma 3.8. *Suppose B is a mutation Dynkin gss matrix with associated signed valued quiver (Q, v). Let $\Gamma = \Gamma(Q, v)$ and $C = C(Q, v)$. Then, for $i \neq j$, the number of edges between i and j in Γ is $c_{ij}c_{ji}$, with $i > j$ if and only if $|c_{ij}| > |c_{ji}|$. Moreover, if ε_{ij} denotes the sign of an edge between i and j in Γ, then*

$$c_{ij} = \begin{cases} 0 & \text{if there are no edges between } i \text{ and } j, \\ \varepsilon_{ij} & \text{if there is one edge between } i \text{ and } j, \\ \varepsilon_{ij} & \text{if there are } d > 1 \text{ edges between } i \text{ and } j, \text{ and } i < j \text{ in } \Gamma, \\ \varepsilon_{ij} d & \text{if there are } d > 1 \text{ edges between } i \text{ and } j, \text{ and } i > j \text{ in } \Gamma. \end{cases}$$

Proof. This is a case-by-case check according to whether we have arrows $\alpha : i \to j$, $\alpha : j \to i$, or no such arrows.

The idea is to use Lemma 3.5 together with the fact that the (unsigned) valued quiver associated to $B(1)$ is "2-finite" (see the comment after Theorem 8.6 in [8]): by Proposition 2.6 any arrow in (Q, v) has weight ≤ 3. This property implies that, for i, j vertices of $\Gamma(Q, v)$, the data given in Definition 3.7 uniquely determines the (i, j)-th and (j, i)-th entries of $C(Q, v)$, and vice versa. □

We now describe how Cartan counterparts change under mutation.

Lemma 3.9. *Let (Q, v) be mutation Dynkin and let $(Q', v') = \mu_k(Q, v)$. Let $C = (c_{ij})$ and $C' = (c'_{ij})$ be their respective Cartan counterparts. Then*

$$c'_{ij} = \begin{cases} c_{ij} & \text{if neither or both of } i, j \to k \text{ in } Q, \\ c_{ij} - c_{ik}c_{kj} & \text{if exactly one of } i, j \to k \text{ in } Q. \end{cases}$$

Proof. Write $B = B(Q, v)$ and $B' = B(Q', v')$. By Definition 2.4 of signed mutation, if neither or both $i, j \to k$, then $b'_{ij} = b_{ij}$ if $i, j \neq k$ or $b'_{ij} = -b_{ij}$ if $i \leftarrow k = j$ or $k = i \to j$, and therefore in this case the statement follows trivially.

If $i \to k = j$, then $b'_{ik} = -tb_{ik}$, so $c'_{ij} = -c_{ij} = c_{ik} - c_{ik}c_{kj}$ as $c_{kj} = c_{kk} = 2$. A similar argument works for $j \to k = i$.

The only case left to examine is if there is exactly one arrow $i, j \to k$ and $k \neq i, j$. Without loss of generality, suppose $i \to k$.

- If $b_{kj} = 0$, then we have $b'_{ij} = b_{ij}$. Therefore:

$$c'_{ij} = c_{ij} - c_{ik}c_{kj},$$

since $c_{kj} = 0$.

- If $b_{kj} \neq 0$, then we have $i \to k \to j$. The signed mutation rule yields $b'_{ij} = t(b_{ij} + b_{ik}b_{kj})$.
  - If $b_{ij} = 0$ then $c_{ij} = 0$ and $b'_{ij} = tb_{ik}b_{kj}$. So, as B is pure, one checks that in each case we have $c'_{ij} = -c_{ik}c_{kj} = c_{ij} - c_{ik}c_{kj}$.
  - If $b_{ij} \neq 0$ then, by Lemma 2.40, $\operatorname{sgn}(b_{ij}) = -1$ since all 3-cycles in B are oriented. Now we use that B is pure and satisfies the positive 3-cycle condition at k. If $b_{ij}, b_{ik}b_{kj} \in \mathbb{Z}$ then $c_{ij} = -b_{ij}$ and $c_{ik}c_{kj} = b_{ik}b_{kj}$. If (i, j, k) is a 3-cycle in B', then it is oriented by Lemma 2.40. Therefore $b'_{ij} = t(b_{ij} + b_{ik}b_{kj})$ implies $b_{ij} + b_{ik}b_{kj} \geq 0$. So

$$c'_{ij} = -|b_{ij} + b_{ik}b_{kj}| = -(b_{ij} + b_{ik}b_{kj}) = c_{ij} - c_{ik}c_{kj}.$$

The case $b_{ij}, b_{ik}b_{kj} \in t\mathbb{Z}$ can be treated similarly.

Therefore we get the statement. □

3B. ***Root systems.*** Given an $n \times n$ gss matrix B, let $S_B = \{\alpha_1^B, \ldots, \alpha_n^B\}$, where α_i^B are formal simple roots and let V_B be the real vector space with basis S_B. We sometimes drop the super/subscript B when it is clear from context.

Definition 3.10. Let $C = C(B) = (c_{ij})$ be the Cartan counterpart of B. Define linear functions $s_i = s_{\alpha_i^B} : V_B \to V_B$ called *simple reflections* by letting

$$s_i(\alpha_j) = \alpha_j - c_{ij}\alpha_i$$

and extending linearly to V_B. Let W_B denote the subgroup of $\mathrm{GL}(V_B)$ generated by s_i, $1 \leq i \leq n$.

Definition 3.11. Suppose B is mutation Dynkin. The *root system* of B is the orbit $\Phi_B = W_B \cdot S_B$ of S_B under W_B.

If B is Dynkin then this is exactly the classical root system, by Theorem 2 of [15, Section V.10]. In particular, if B is Dynkin then W_B is finite. We will show in Theorem 3.21 that the name "root system" is justified in the mutation Dynkin case.

Recall that we have a skew-symmetrizing matrix D such that DB is skew-symmetric. If $C = C(B)$ then DC is symmetric, and so $DC = (DC)^T = C^T D$.

Definition 3.12. Let $M = DC$ and define an inner product (i.e., a symmetric, positive-definite bilinear form) on V_B by $(u, v) = u^T M v$.

For $\beta \in \Phi_B$, define $\beta^\vee = 2\beta/(\beta, \beta)$. Then

$$(\alpha_i, \alpha_j^\vee) = \frac{2(\alpha_i, \alpha_j)}{(\alpha_j, \alpha_j)} = \frac{2m_{ij}}{m_{jj}} = \frac{2c_{ji}d_j}{c_{jj}d_j} = c_{ji}$$

and the length of α_i is $|\alpha_i| = (\alpha_i, \alpha_i) = m_{ii} = 2d_i$.

Proposition 3.13. *The inner product $(-,-)$ is W_B-invariant.*

Proof. Since $M = C^T D = DC$, then $m_{ij} = c_{ji}d_j = d_i c_{ij}$ for $i, j = 1, \ldots, n$. Using this and $c_{ii} = 2$, we get

$$\begin{aligned}
(s_i(\alpha_j),\ s_i(\alpha_k)) &= (\alpha_j - c_{ij}\alpha_i,\ \alpha_k - c_{ik}\alpha_i) \\
&= m_{jk} - c_{ik}m_{ji} - c_{ij}m_{ik} + c_{ij}c_{ik}m_{ii} \\
&= m_{jk} - c_{ik}c_{ij}d_i - c_{ij}c_{ki}d_k + 2c_{ij}c_{ik}d_i \\
&= m_{jk} + c_{ij}c_{ik}d_i - c_{ij}c_{ki}d_k \\
&= m_{jk} + c_{ij}(c_{ik}d_i - c_{ki}d_k) \\
&= m_{jk} = (\alpha_j,\ \alpha_k).
\end{aligned}$$

□

Remark 3.14. Our convention in Definition 3.10 differs from some other texts, such as [10], where $s_i(\alpha_j)$ is defined using c_{ji} and not c_{ij}. Later, in Definition 4.1, we will define a Lie algebra relation (R4) which specialises to $(\operatorname{ad} e_i)^{1-c_{ij}}(e_j) = 0$ for a classical Cartan matrix: we use this convention in order to match certain other results in the literature [2; 4; 14]. Then, if we want the root space decomposition of our Lie algebra to be compatible with Definition 3.11, this forces us to use the conventions in Definition 3.10.

For convenience, the following table collects in one place the construction of classical Lie-theoretic objects from the Cartan matrix C when using our conventions.

$C \leftrightarrow \mathfrak{g}$	$[h_i, e_j] = c_{ij}e_j$ and $(\operatorname{ad} e_i)^{1-c_{ij}}(e_j) = 0$
$C \leftrightarrow \Gamma$	$\lvert c_{ij}\rvert > \lvert c_{ji}\rvert \Longleftrightarrow i > j$ in Γ
$C \leftrightarrow (-,-)$	$c_{ij} = (\alpha_j,\ \alpha_i^\vee)$
$C \leftrightarrow W$	$s_i(\alpha_j) = \alpha_j - c_{ij}\alpha_i$

The following definition was given in [13, Definition 4.1] for the simply laced case, and extended to the skew-symmetrizable case in [3, Section 6]. Let Φ_Δ be a root system of Dynkin type Δ with rank n and let $\tilde{B}$ be a skew-symmetrizable matrix of mutation class Δ. A subset $\{\gamma_i\}_{i=1}^n$ of Φ_Δ is called a *companion basis for* $\tilde{B}$ if

(i) $\{\gamma_i\}_{i=1}^n$ is a $\mathbb{Z}$-basis of $\mathbb{Z}\Phi_\Delta$, and

(ii) the matrix A with $a_{ij} = (\gamma_j, \gamma_i^\vee)$ is a positive quasi-Cartan companion of $\tilde{B}$.

Note that [3] asks for $a_{ij} = (\gamma_i, \gamma_j^\vee)$ in the above definition, but we have adapted the condition to match our convention as outlined in Remark 3.14.

Definition 3.15. Let B be a mutation Dynkin gss matrix of Dynkin type Δ. We say that $\{\gamma_i\}_{i=1}^n \subset \Phi_\Delta$ is a *signed companion basis for* B if:

(i) $\{\gamma_i\}_{i=1}^n$ is a $\mathbb{Z}$-basis of $\mathbb{Z}\Phi_\Delta$, and

(ii) the matrix C with $c_{ij} = (\gamma_j, \gamma_i^\vee)$ is equal to $C(B)$.

Lemma 3.16. *If $\{\gamma_i\}_{i=1}^n \subset \Phi_\Delta$ is a signed companion basis for B, then it is a companion basis for $B(1)$.*

Proof. By Theorem 2.42 and Lemma 3.2, C is a quasi-Cartan companion of $B(1)$. As $\{\gamma_i\}_{i=1}^n \subset \Phi_\Delta$ is a $\mathbb{Z}$-basis for $\mathbb{Z}\Phi$, it is related to the Cartan matrix of Δ by a change of basis, so positivity of the Cartan matrix implies positivity of C. □

The following lemma is useful. For a proof, see for example [10, III.9.2].

Lemma 3.17. *If $f : V_B \to V_B$ is an isomorphism such that $f(\Phi) = \Phi$ then $s_{f(\delta)} f = f s_\delta$ for all $\delta \in \Phi$.*

Now let $B = B(Q, v)$, $B' = \mu_k B$, $\Phi = \Phi_B = \{\alpha_i\}$, and $\Phi' = \Phi_{B'} = \{\alpha_i'\}$. The following formulas send our base roots to the inward mutation of roots as defined by Parsons [13, Theorem 6.1]. Note that $\rho_k(\Phi') \in \Phi$ follows by Lemma 3.17 and Definition 3.11.

Definition 3.18. The *mutation of roots* at k is the function $\rho_k : \Phi' \to \Phi$ defined by extending the following rule linearly:

$$\rho_k : \ \alpha_i' \mapsto \begin{cases} s_k(\alpha_i) & \text{if } i \to k \text{ in } Q, \\ \alpha_i & \text{otherwise.} \end{cases}$$

Notice that the direction of the morphism $\Phi' \to \Phi$ is opposite to what one might at first expect.

The formula in the next result comes from the outward mutation of roots [13, Theorem 6.1].

Lemma 3.19. *Mutation of roots is a bijection with inverse $\rho_k^{-1} : \Phi \to \Phi'$ given by*

$$\rho_k^{-1} : \ \alpha_j \mapsto \begin{cases} s_k'(\alpha_j') & \text{if } k \to j \text{ in } Q', \\ \alpha_j' & \text{otherwise.} \end{cases}$$

Proof. We check that the maps from Definitions 3.18 and 3.19 are inverse. Note that we have an arrow $i \to k$ in Q if and only if we have an arrow $k \to i$ in Q'. We just check the composition $\Phi' \to \Phi \to \Phi'$; the other composition is similar. If we do not have an arrow $i \to k$ in Q, then $\alpha_i' \mapsto \alpha_i \mapsto \alpha_i'$. If we do have $i \to k$ in Q, then $\alpha_i' \mapsto s_k(\alpha_i) = \alpha_i - c_{ki}\alpha_k \mapsto (\alpha_i' - c_{ki}'\alpha_k') - c_{ki}\alpha_k'$. But $c_{ki}' = -c_{ki}$ by Lemma 3.9, so this is α_i'. □

Lemma 3.20. *Write $\gamma_i = \alpha_i'$. Then*

$$(\rho_k(\gamma_i),\ \rho_k(\gamma_j)^\vee) = \begin{cases} (\alpha_i,\ \alpha_j^\vee) & \text{if neither or both of } i,\ j \to k \text{ in } Q_B, \\ (\alpha_i,\ \alpha_j^\vee) - (\alpha_i,\ \alpha_k^\vee)(\alpha_k,\ \alpha_j^\vee) & \text{if exactly one of } i,\ j \to k \text{ in } Q_B. \end{cases}$$

Proof. The check for the case where neither of i, j has an arrow to k is trivial. For the remaining checks, we first note that the $W_{B'}$-invariance of $(-,-)$ implies $W_{B'}$-invariance of $(-,(-)^\vee)$.

If both i, j have arrows to k then $(\rho_k(\gamma_i),\ \rho_k(\gamma_j)^\vee) = (s_{\alpha_k}\alpha_i,\ (s_{\alpha_k}\alpha_j)^\vee) = (\alpha_i,\ \alpha_j^\vee)$ as required.

The check for the case where only $i \to k$ is an easy computation, using that $(-,(-)^\vee)$ is linear in the first component:

$$(\rho_k(\gamma_i),\ \rho_k(\gamma_j)^\vee) = (s_k\alpha_i,\ \alpha_j^\vee) = (\alpha_i - c_{ki}\alpha_k,\ \alpha_j^\vee) = (\alpha_i,\ \alpha_j^\vee) - (\alpha_i,\ \alpha_k^\vee)(\alpha_k,\ \alpha_j^\vee).$$

The check for $j \to k$ follows from the previous check and reflection invariance: $(\rho_k(\gamma_i),\ \rho_k(\gamma_j)^\vee) = (\alpha_i,\ (s_k\alpha_j)^\vee) = (s_k\alpha_i,\ \alpha_j^\vee) = (\alpha_i,\ \alpha_j^\vee) - (\alpha_i,\ \alpha_k^\vee)(\alpha_k,\ \alpha_j^\vee)$. □

Recall the definition of a (reduced) root system R in a real vector space V with inner product (see [15, V.2] or [10, III.9.2]):

(1) R is finite, spans V, and $0 \notin R$.

(2) For all $\alpha \in R$ we have $s_\alpha(R) = R$.

(3) For all $\alpha, \beta \in R$ we have $(\alpha, \beta^\vee) \in \mathbb{Z}$.

(4) For all $\alpha \in R$ we have $R \cap \mathbb{R}\alpha = \{\pm\alpha\}$.

A vector space isomorphism $f : V \to V'$ is an isomorphism $R \xrightarrow{\sim} R'$ of root systems if $f(R) = R'$ and for all $\alpha, \beta \in \mathbb{R}$ we have $(f(\alpha),\ f(\beta)^\vee)_{V'} = (\alpha, \beta^\vee)_V$.

Theorem 3.21. *If B is mutation Dynkin then Φ_B is a root system and $\rho_k : \Phi'_{B'} \to \Phi_B$ extends to an isomorphism of root systems.*

Proof. By induction, we just need to check that Φ_B is a root system when $B' = \mu_k B$ is Dynkin. Property 1 holds by Lemma 3.19, property 2 holds by Lemma 3.17, property 3 holds by Lemma 3.20, and property 4 holds by linearity of ρ_k. Then Lemmas 3.9 and 3.20 show that ρ_k preserves the inner products, and so is an isomorphism of root systems. □

Recall that $S_B = \{\alpha_1, \ldots, \alpha_n\}$ is a basis for V_B.

Theorem 3.22. *Suppose B is mutation Dynkin, so $B = \mu_{k_\ell} \ldots \mu_{k_1} B_\Delta$ where $\tilde{B}_\Delta$ is a skew-symmetric matrix of Dynkin type and $B_\Delta = t\tilde{B}_\Delta$. Write $\Phi = \Phi_B$ and $\Phi_\Delta = \Phi_{B_\Delta}$. Let $\rho = \rho_{k_1} \ldots \rho_{k_\ell} : \Phi \to \Phi_\Delta$. Then $\rho(S_B)$ is a signed companion basis for B.*

Proof. We use induction on ℓ. The base case $\ell = 0$ is clear. So suppose the statement is true for mutation sequences of length ℓ and let $B' = \mu_{k_{\ell+1}} B$. The mutation in Definition 3.18 clearly preserves $\mathbb{Z}$-bases of $\mathbb{Z}\Phi_\Delta$. By induction we know that $C(B)_{ij} = (\rho(\alpha_i), \rho(\alpha_j)^\vee)$, then comparing Lemmas 3.9 and 3.20 shows that $C(B')_{ij} = (\rho_{k_{\ell+1}}\rho(\alpha_i), \rho_{k_{\ell+1}}\rho(\alpha_j)^\vee)$, so we are done. □

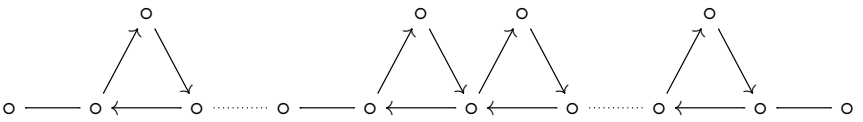
Corollary 3.23. *If (Q, v) is mutation Dynkin then $C(Q, v)$ is a positive quasi-Cartan companion of $B(1)$.*

Proof. Combine Theorem 3.22 and Lemma 3.16. □

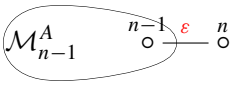
Given Corollary 3.23, we could apply results of [4; 14] to get presentations of simple Lie algebras from mutation Dynkin signed valued quivers. We want to understand homomorphisms between these presentations, so we conduct a more in-depth analysis in Section 4.

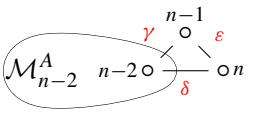
3C. ***Mutation type A.*** In this subsection we give a recursive description of the root system associated to a gss matrix of mutation Dynkin type A.

Recall from [5, Section 2] the classification of (unsigned) quivers of mutation type A. Roughly, all such quivers look like the following diagram:



Let $\mathcal{M}_n^A$ denote the set of signed valued quivers of mutation type A_n and let $(Q, v) \in \mathcal{M}_n^A$. By Theorem 2.43, we know that (Q, v) is one of the following shapes, with signs of certain arrows labelled in red by $\gamma, \delta, \varepsilon \in \{\pm 1\}$.

(a) $Q =$ 

(b) $Q =$ 

By Theorem 2.43, the product of the signs of the arrows in the 3-cycle in (b) is $\gamma\delta\varepsilon = 1$.

Let $\Phi = \Phi(Q, v)$ with simple roots $\{\alpha_1, \ldots, \alpha_n\}$. If $\beta = \sum_{i=1}^n \lambda_i\alpha_i \in \Phi$, let $m_i : \Phi \to \mathbb{Z}$ pick out the coefficient of α_i, so $m_i(\beta) = \lambda_i$.

For a full subquiver Q' of Q on vertices $\{i_1, \ldots, i_k\}$, let $\Phi_{\{i_1,\ldots,i_k\}}$ denote the root system associated to $(Q', v|_{Q'})$.

Proposition 3.24. *Let $(Q, v) \in \mathcal{M}_n^A$, and ε be the sign of the arrow between $n-1$ and n in Q. The root system $\Phi_{(Q,v)}$ is as follows*:

(a) *If* (Q, v) *is of type* (a), *then*

$$\Phi_{(Q,v)} = \Phi_{\{1,\dots,n-1\}} \cup \{\beta - \varepsilon m_{n-1}(\beta)\alpha_n \mid \beta \in \Phi_{\{1,\dots,n-1\}}\} \cup \{\pm\alpha_n\}. \tag{2}$$

(b) *If* (Q, v) *is of type* (b), *then*

$$\Phi_{(Q,v)} = \Phi_{\{1,\dots,n-1\}} \cup \Phi_{\{1,\dots,n-2,n\}} \cup \{\pm(\alpha_{n-1} - \varepsilon\alpha_n)\}. \tag{3}$$

Proof. Let $\Theta_{(Q,v)}$ denote the right hand side of (2) if Q is of type (a) or the right hand side of (3) if Q is of type (b).

We proceed by induction on n. If $n = 1$ then (Q, v) is of type (a) and $\Phi_{(Q,v)} = \{\pm\alpha_1\}$, so the claim is true.

For $n > 1$, we start by showing that $\Theta_{(Q,v)} \subseteq \Phi_{(Q,v)}$.

(a) Suppose Q is of type (a). Then $\Phi_{\{1,\dots,n-1\}} \subseteq \Phi_{(Q,v)}$ by definition of root system. Furthermore, for $\beta \in \Phi_{\{1,\dots,n-1\}}$ we have

$$s_n(\beta) = \beta - \sum_{i=1}^{n} m_i(\beta)c_{ni}\alpha_n = \beta - \varepsilon m_{n-1}(\beta)\alpha_n.$$

Hence, since $\pm\alpha_n \in \Phi_{(Q,v)}$ by definition, we get $\Theta_{(Q,v)} \subseteq \Phi_{(Q,v)}$.

(b) Suppose Q is of type (b). Once again, by definition of root system, we have $\Phi_{\{1,\dots,n-1\}}, \Phi_{\{1,\dots,n-2,n\}} \subseteq \Phi_{(Q,v)}$. Furthermore

$$s_n(\pm\alpha_{n-1}) = \pm(\alpha_{n-1} - \varepsilon\alpha_n),$$

and therefore $\Theta_{(Q,v)} \subseteq \Phi_{(Q,v)}$.

To prove the inverse inclusion, it is enough to show that $\Theta_{(Q,v)}$ is closed under reflections.

(a) Let $\beta \in \Phi_{\{1,\dots,n-1\}}$.

First, we need to show that $s_i(\beta) \in \Theta_{(Q,v)}$ for $1 \le i \le n$. But $s_i(\beta) \in \Phi_{\{1,\dots,n-1\}} \subseteq \Theta_{(Q,v)}$ for $i = 1, \dots, n-1$ by definition, and $s_n(\beta) = \beta - \varepsilon m_{n-1}(\beta)\alpha_n \in \Theta_{(Q,v)}$.

Second, consider $\beta - \varepsilon m_{n-1}(\beta)\alpha_n$. If $i \neq n-1$ then $m_{n-1}(s_i(\beta)) = m_{n-1}(\beta)$, so

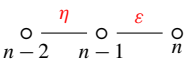
$$s_i(\beta - \varepsilon m_{n-1}(\beta)\alpha_n) = s_i(\beta) - \varepsilon m_{n-1}(s_i(\beta))\alpha_n \in \Theta_{(Q,v)}.$$

For $i = n-1$ we split into two cases.

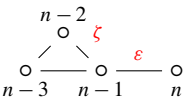
Case 1: $\beta = \pm\alpha_{n-1}$. If $\beta = \alpha_{n-1}$, then $s_{n-1}(\beta - \varepsilon m_{n-1}(\beta)\alpha_n) = s_{n-1}(\alpha_{n-1} - \varepsilon\alpha_n) = -\alpha_{n-1} - \varepsilon(\alpha_n - \varepsilon\alpha_{n-1}) = -\varepsilon\alpha_n \in \Theta_{(Q,v)}$. The check for $\beta = -\alpha_{n-1}$ is analogous.

Case 2: $\beta \in \Phi_{\{1,\dots,n-1\}}$ with $\beta \neq \pm\alpha_{n-1}$. If $m_{n-1}(\beta) = 0$ then we have already checked that $W \cdot (\beta - \varepsilon m_{n-1}(\beta)\alpha_n) = W \cdot \beta \subseteq \Theta_{(Q,v)}$, so we can assume that $m_{n-1}(\beta) \neq 0$.

We claim that there is exactly one neighbour $j \neq n$ of $n-1$ such that $m_j(\beta) \neq 0$. To see this we use the inductive hypothesis. If $Q_{\{1,\dots,n-1\}}$ is of type (a) then the only neighbour $j \neq n$ of $n-1$ is $j = n-2$, and as $\beta \neq \pm\alpha_{n-1}$ the only possibility given the inductive hypothesis is that $m_{n-2}(\beta) \neq 0$. We write η for the sign of the arrow between $n-2$ and $n-1$.



If $Q_{\{1,\dots,n-1\}}$ is of type (b) then we have two neighbours $j \neq n$ of $n-1$: these are $j = n-2$ and $j = n-3$.

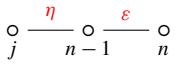


By the inductive hypothesis, β belongs to the union

$$\Phi_{\{1,\dots,n-2\}} \cup \Phi_{\{1,\dots,n-3,n-1\}} \cup \{\pm(\alpha_{n-2} - \zeta\alpha_{n-1})\}.$$

We have $m_{n-1}(\beta) \neq 0$, so β cannot belong to the first of these sets. If it belongs to the second, we have $m_{n-2}(\beta) = 0$ and we set $j = n-3$ and let η denote the sign of the arrow between $n-3$ and $n-1$. If it belongs to the third, we have $m_{n-3}(\beta) = 0$; we set $j = n-2$ and $\eta = \zeta$.

So we have the diagram



with $m_j(\beta) \neq 0$. By checking each of the above cases, we know that $m_{n-1}(\beta) = -\eta m_j(\beta)$. Therefore

$$\begin{aligned} s_{n-1}(\beta) &= \sum_{i \neq j, n-1} m_i(\beta)\alpha_i + m_j(\beta)s_{n-1}(\alpha_j) + m_{n-1}(\beta)s_{n-1}(\alpha_{n-1}) \\ &= \sum_{i \neq j, n-1} m_i(\beta)\alpha_i + m_j(\beta)\left(\alpha_j - \eta\alpha_{n-1}\right) - m_{n-1}(\beta)\alpha_{n-1} \\ &= \beta - m_{n-1}(\beta)\alpha_{n-1}, \end{aligned}$$

so we have $s_{n-1}(\beta - \varepsilon m_{n-1}(\beta)\alpha_n) = \beta - m_{n-1}(\beta)\alpha_{n-1} - \varepsilon m_{n-1}(\beta)\,(\alpha_n - \varepsilon\alpha_{n-1}) = \beta - \varepsilon m_{n-1}(\beta)\alpha_n \in \Theta_{(Q,v)}$.

Third, consider $\pm\alpha_n$. As $\Theta_{(Q,v)} = -\Theta_{(Q,v)}$, it is enough to consider α_n. If $i \neq n-1$ then $s_i(\alpha_n) \in \{\pm\alpha_n\}$. For $i = n-1$ we have $s_{n-1}(\alpha_n) = \alpha_n - \varepsilon\alpha_{n-1}$. Let $\beta = -\varepsilon\alpha_{n-1} \in \Phi_{\{1,\dots,n-1\}}$. Then

$$\beta - \varepsilon m_{n-1}(\beta)\alpha_n = -\varepsilon\alpha_{n-1} + \alpha_n = s_{n-1}(\alpha_n)$$

so $s_i(\alpha_n) \in \Theta_{(Q,v)}$.

(b) Let $\beta \in \Phi_{\{1,\dots,n-1\}}$. For $i = 1, \dots, n-1$ we have $s_i(\beta) \in \Phi_{\{1,\dots,n-1\}} \subseteq \Theta_{(Q,v)}$ by definition.

Further, if $m_{n-1}(\beta) = 0$, then $\beta \in \Phi_{\{1,\dots,n-2\}}$ and so $s_n(\beta) \in \Phi_{\{1,\dots,n-2,n\}} \subseteq \Theta_{(Q,v)}$.

Therefore, we can assume $m_{n-1}(\beta) \neq 0$.

If $\beta = \pm\alpha_{n-1}$, then $s_n(\beta) = \pm(\alpha_{n-1} - \varepsilon\alpha_n) \in \Theta_{(Q,v)}$. Otherwise, using the same argument as for case (a), we have $m_{n-1}(\beta) = -\gamma m_{n-2}(\beta)$. Thus

$$s_n(\beta) = \beta - \delta m_{n-2}(\beta)\alpha_n - \varepsilon m_{n-1}(\beta)\alpha_n = \beta - (\delta - \varepsilon\gamma)m_{n-2}(\beta)\alpha_n = \beta \in \Theta_{(Q,v)},$$

where the last equality follows since $\gamma\delta\varepsilon = 1$.

One proves similarly that $W \cdot (\Phi_{\{1,\dots,n-2,n\}}) \subseteq \Theta_{(Q,v)}$.

Finally, one checks that s_i fixes $\pm(\alpha_{n-1} - \varepsilon\alpha_n)$ for $i = 1, \dots, n-2$, and that $s_{n-1}(\alpha_{n-1} - \varepsilon\alpha_n) = -\varepsilon\alpha_n$ and $s_n(\alpha_{n-1} - \varepsilon\alpha_n) = \alpha_{n-1}$. So $s_i(\pm(\alpha_{n-1} - \varepsilon\alpha_n)) \in \Theta_{(Q,v)}$ for all $i = 1, \dots, n$.

Hence $\Phi_{(Q,v)} \subseteq \Theta_{(Q,v)}$, and we are done. □

4. Presentations of Lie algebras

4A. ***Presentations.*** In this section we work with Lie algebras over the complex numbers. We use iterated bracket notation:

$$[x_1, x_2, \dots, x_n] := [x_1, [x_2, [\dots, x_n]]].$$

We first consider four types of relations:

Definition 4.1. Let $C = (c_{ij})$ be a quasi-Cartan matrix. Define the Lie algebra $\mathfrak{g}_4(C)$ with generators $\{h_i, e_{\varepsilon i} \mid \varepsilon \in \{\pm 1\}, 1 \leq i \leq n\}$ subject to the following relations, for $\varepsilon, \delta \in \{\pm 1\}$:

(R1) $[h_i, h_j] = 0$;

(R2) $[e_i, e_{-i}] = h_i$;

(R3) $[h_i, e_{\varepsilon j}] = \varepsilon c_{ij} e_{\varepsilon j}$;

(R4) $(\operatorname{ad} e_{\varepsilon i})^{-M+1}(e_{\delta j}) = 0$, where $M = \min\{0, \varepsilon\delta c_{ij}\}$.

Note that if C is a Cartan matrix of Dynkin type Δ, then Definition 4.1 reduces to the classical Serre presentation of the Lie algebra $\mathfrak{g}(\Delta)$ [15].

We record two useful calculational results which always hold in $\mathfrak{g}_4(C)$. We will use both often.

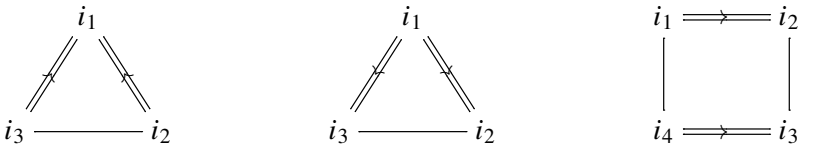
Lemma 4.2. $[h_i, e_{\varepsilon_1 j_1}, \dots e_{\varepsilon_\ell j_\ell}] = \sum_{k=1}^{\ell} \varepsilon_k c_{ik} [e_{\varepsilon_1 j_1}, \dots e_{\varepsilon_\ell j_\ell}]$.

Proof. Use (R3), induction, and the Jacobi identity. □

Lemma 4.3. *If* $\delta = \operatorname{sgn}(c_{ij})$ *then* $[e_i, e_{\delta j}] = [e_{-i}, e_{-\delta j}] = 0$.

Proof. This follows from (R4). To show that the first term is zero, set $\varepsilon = 1$ and $\delta = \operatorname{sgn}(c_{ij})$, so $\delta c_{ij} \geq 0$ and thus $M = 0$. For the second, set $\varepsilon = -1$ and $\delta = -\operatorname{sgn}(c_{ij})$. □

The next set of relations depends on chordless cycles of C. If C is mutation Dynkin then, by [1, Proposition 2.6], there are only four possibilities for the unsigned diagram of the cycle. Either every edge has weight 1, or we are in one of the following three cases:



These cycles occur in mutation types B_n, C_n, and F_4, respectively. Note that the B_3-cycle and the C_3-cycle also appear in mutation type F_4.

In cycles with double edges, the vertices correspond to roots which are either long or short: $i \Longrightarrow j$ tells us that i is long and j is short. So in the B_3-cycle displayed above, the vertex i_1 is short and the vertices i_2 and i_3 are long.

Definition 4.4. Define the Lie algebra $\mathfrak{g}_5(C)$ as $\mathfrak{g}_4(C)$ subject to the following extra relations. To each chordless cycle $i_1, i_2, \ldots, i_t$ where every edge has weight 1 we impose the relations:

$$\text{(R5)}\ \ [e_{\varepsilon_1 i_1},\ e_{\varepsilon_2 i_2},\ \ldots,\ e_{\varepsilon_t i_t}] = 0 \quad \text{where} \quad \varepsilon_{q+1} = -\operatorname{sgn}(c_{i_q, i_{q+1}})\varepsilon_q$$

where $\varepsilon_1 \in \{\pm 1\}$.

Notice that for each chordless cycle as above we impose $4t$ relations: $2t$ going clockwise (two for every starting point i_j) and $2t$ going anticlockwise.

To each chordless cycle with an edge of higher weight, we only impose the subset of these relations where i_1 and i_t are not both long.

We unpack Definition 4.4 for the three cycles with double edges, using the three diagrams above.

For the B_3-cycle we impose only 3-cycle relations which do not go "long, short, long". Explicitly, these are

$$[e_{\varepsilon_1 i_{\sigma(1)}},\ e_{\varepsilon_2 i_{\sigma(2)}},\ e_{\varepsilon_3 i_{\sigma(3)}}], \quad \text{where } \varepsilon_{q+1} = -\operatorname{sgn}(c_{i_{\sigma(q)}, i_{\sigma(q+1)}})\varepsilon_q$$

for $\varepsilon_1 \in \{\pm 1\}$ and $\sigma \in \operatorname{Dih}(3)$ such that $\{\sigma(1), \sigma(3)\} \neq \{2, 3\}$, where $\operatorname{Dih}(p)$ denotes the dihedral group of order $2p$.

For the C_3-cycle there is no path of length 3 which starts and ends at a long root, so we impose all twelve 3-cycle relations. Explicitly, these are

$$[e_{\varepsilon_1 i_{\sigma(1)}},\ e_{\varepsilon_2 i_{\sigma(2)}},\ e_{\varepsilon_3 i_{\sigma(3)}}], \quad \text{where } \varepsilon_{q+1} = -\operatorname{sgn}(c_{i_{\sigma(q)}, i_{\sigma(q+1)}})\varepsilon_q$$

for $\sigma \in \mathrm{Dih}(3)$.

For the F_4-cycle we impose only 4-cycle relations which do not go "long, short, short, long". Explicitly, these are

$$[e_{\varepsilon_1 i_{\sigma(1)}},\ e_{\varepsilon_2 i_{\sigma(2)}},\ e_{\varepsilon_3 i_{\sigma(3)}},\ e_{\varepsilon_4 i_{\sigma(4)}}], \quad \text{where } \varepsilon_{q+1} = -\mathrm{sgn}(c_{i_{\sigma(q)}, i_{\sigma(q+1)}})\varepsilon_q$$

for $\sigma \in \mathrm{Dih}(4)$ such that $\{\sigma(1), \sigma(4)\} \neq \{1, 4\}$.

In the introduction of [14], following [4], a Lie algebra which we denote $\hat{\mathfrak{g}}_4(C)$ is defined. We write its generators as $\hat{h}_i$ and $\hat{e}_{\varepsilon i}$. The relations are similar to our relations in Definition 4.1, except that the second and third relations are swapped and their relation $[\hat{h}_i, \hat{e}_{\varepsilon j}] = -\varepsilon c_{ij} \hat{e}_{\varepsilon j}$ has a minus sign.

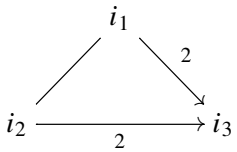
Lemma 4.5. *The map*

$$\hat{e}_{\varepsilon i} \mapsto \varepsilon e_{\varepsilon i}, \quad \hat{h}_i \mapsto -h_i,$$

extends to an isomorphism $\hat{\mathfrak{g}}_4(C) \xrightarrow{\sim} \mathfrak{g}_4(C)$.

Proof. Since the map is clearly bijective, the proof consists of straightforward checks that the relations (R1) to (R4) are preserved by the map defined above. □

Remark 4.6. To state their fifth relation in non-simply laced types and define their Lie algebra $\hat{\mathfrak{g}}_5(C)$, Pérez and Rivera use a 3-cycle denoted $\mathcal{C}_{\mathbb{C}}(i_1, i_2, i_3)$, but the diagram which defines it seems to contain a typo as only directed edges should be weighted. If we replace their diagram by



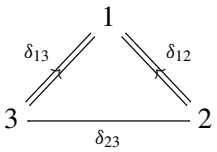
and interpret their statements correctly, our relations for the B_3- and C_3-cycles are consistent with theirs, and the isomorphism of Lemma 4.5 extends to an isomorphism $\hat{\mathfrak{g}}_5(C) \xrightarrow{\sim} \mathfrak{g}_5(C)$ for all positive quasi-Cartan matrices C which are not of mutation type F_4.

However, Pérez and Rivera include all potential relations in F_4-cycles, while we only include 12 of the 16 possible relations. We will show in Lemma 4.8 below that imposing all 16 relations would give the zero Lie algebra. We believe there is an error in [14, Lemma 5.3]: the final paragraph of their proof states that $[e_{-\varepsilon i}, e_{-\varepsilon s}, \tilde{e}_{\varepsilon r}, e_{\varepsilon j}] = 0$ but we think this should be nonzero.

4B. ***Relations.*** In this section we first explain why we excluded certain relations in Definition 4.4, and then we show that we could have imposed fewer relations in Definition 4.4 to get the same Lie algebra.

Here we explain why we do not include relations starting and ending at a long root. To simplify notation, we replace the vertices i_j by j in our diagrams.

We start by considering the B_3-cycle and its Cartan counterpart,



$$C = \begin{pmatrix} 2 & \delta_{12} & \delta_{13} \\ 2\delta_{12} & 2 & \delta_{23} \\ 2\delta_{13} & \delta_{23} & 2 \end{pmatrix},$$

where the signs $\delta_{12}, \delta_{23}, \delta_{13} \in \{\pm 1\}$ are such that $\delta_{12}\delta_{23}\delta_{13} = 1$. The relations we did not include in Definition 4.4 are

$$[e_{-\varepsilon\delta_{12}\delta_{13}2}, e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] = 0, \tag{4}$$

$$[e_{-\varepsilon\delta_{13}\delta_{12}3}, e_{\varepsilon\delta_{12}1}, e_{-\varepsilon 2}] = 0. \tag{5}$$

Lemma 4.7. *If C is a quasi-Cartan matrix coming from a B_3-cycle, then the quotient of $\mathfrak{g}_5(C)$ by the ideal generated by any of the relation* (4) *or* (5) *is the zero Lie algebra.*

Proof. We show that if we impose a relation from (4) to $\mathfrak{g}_5(C)$ we get the zero Lie algebra, as the other check is similar.

Since $\operatorname{ad}(e_{\varepsilon\delta_{12}\delta_{13}2})$ is a derivation, we have

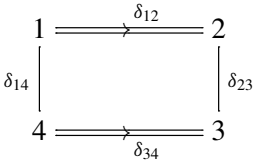
$$\begin{aligned} 0 &= [e_{\varepsilon\delta_{12}\delta_{13}2}, e_{-\varepsilon\delta_{12}\delta_{13}2}, e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] \\ &= \varepsilon\delta_{12}\delta_{13}[h_2, e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] + [e_{-\varepsilon\delta_{12}\delta_{13}2}, e_{\varepsilon\delta_{12}\delta_{13}2}, e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] \\ &= \varepsilon\delta_{12}\delta_{13}(2\varepsilon\delta_{12}\delta_{13} - \varepsilon\delta_{23})[e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] - [e_{-\varepsilon\delta_{12}\delta_{13}2}, e_{\varepsilon\delta_{12}\delta_{13}2}, e_{-\varepsilon 3}, e_{\varepsilon\delta_{13}1}]. \end{aligned}$$

(R5) yields $[e_{\varepsilon\delta_{12}\delta_{13}2}, e_{-\varepsilon 3}, e_{\varepsilon\delta_{13}1}] = 0$, since $-\operatorname{sgn}(c_{23})\varepsilon\delta_{12}\delta_{13} = -\varepsilon\delta_{12}\delta_{23}\delta_{13} = -\varepsilon$ and $-\operatorname{sgn}(c_{31})(-\varepsilon) = \varepsilon\delta_{13}$. Therefore, as $\varepsilon\delta_{12}\delta_{13}(2\varepsilon\delta_{12}\delta_{13} - \varepsilon\delta_{23}) \neq 0$, we get $[e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] = 0$. As a consequence, Lemma 4.2 yields

$$0 = [e_{-\varepsilon\delta_{13}1}, e_{\varepsilon\delta_{13}1}, e_{-\varepsilon 3}] = -\varepsilon\delta_{13}[h_1, e_{-\varepsilon 3}] = e_{-\varepsilon 3}.$$

Hence we get $e_3 = e_{-3} = 0$. Similar computations yield $e_1 = e_{-1} = e_2 = e_{-2} = 0$, and thus $h_i = 0$ by (R2). Hence $\mathfrak{g}_5(C)$ with an additional relation (4) is the zero Lie algebra. □

Now consider the F_4-cycle and associated Cartan counterpart,



$$C = \begin{pmatrix} 2 & 2\delta_{12} & 0 & \delta_{14} \\ \delta_{12} & 2 & \delta_{23} & 0 \\ 0 & \delta_{23} & 2 & \delta_{34} \\ \delta_{14} & 0 & 2\delta_{34} & 2 \end{pmatrix},$$

where the signs $\delta_{12}, \delta_{14}, \delta_{23}, \delta_{34} \in \{\pm 1\}$ are such that $\delta_{12}\delta_{23}\delta_{34}\delta_{14} = -1$. The relations we did not include in Definition 4.4 are

$$[e_{-\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] = 0, \tag{6}$$

$$[e_{-\varepsilon\delta_{12}\delta_{23}\delta_{34}1}, e_{\varepsilon\delta_{23}\delta_{34}2}, e_{-\varepsilon\delta_{34}3}, e_{\varepsilon 4}] = 0. \tag{7}$$

Lemma 4.8. *If C is a quasi-Cartan matrix coming from a F_4-cycle, then the quotient of $\mathfrak{g}_5(C)$ by the ideal generated by any of the relation* (6) *or* (7) *is the zero Lie algebra.*

Proof. Once again, we show that if we impose one of the relations from (6) to $\mathfrak{g}_5(C)$ we get the zero Lie algebra, as the other check is similar.

We have

$$\begin{aligned}0 &= [e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{-\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] \\ &= \varepsilon\delta_{12}\delta_{23}\delta_{34}[h_4, e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] + [e_{-\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}]\end{aligned}$$

We will show the second summand is zero. By (R4) we have $[e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon\delta_{12}\delta_{23}3}] = 0$ and $[e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{-\varepsilon\delta_{12}2}] = 0$. Therefore

$$\begin{aligned}[e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] &= [e_{\varepsilon\delta_{12}\delta_{23}3}, e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] \\ &= [e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}, e_{\varepsilon 1}] \\ &= -[e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}, e_{\varepsilon\delta_{12}\delta_{23}\delta_{34}4}] \\ &= 0,\end{aligned}$$

where the final equality is by (R5). Therefore

$$\begin{aligned}0 &= \varepsilon\delta_{12}\delta_{23}\delta_{34}[h_4, e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] \\ &= \varepsilon\delta_{12}\delta_{23}\delta_{34}(2\varepsilon\delta_{12}\delta_{23}\delta_{34} + \varepsilon\delta_{14})[e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}]\end{aligned}$$

so since $\varepsilon\delta_{12}\delta_{23}\delta_{34}(2\varepsilon\delta_{12}\delta_{23}\delta_{34} + \varepsilon\delta_{14}) \neq 0$ we get $[e_{\varepsilon\delta_{12}\delta_{23}3}, e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] = 0$. By applying $[e_{-\varepsilon\delta_{12}\delta_{23}3}, -]$ and using a similar strategy to above we get $[e_{-\varepsilon\delta_{12}2}, e_{\varepsilon 1}] = 0$. Finally, applying $[e_{\varepsilon\delta_{12}2}, -]$ yields $e_{\varepsilon 1} = 0$. Similar computations yield $e_{\varepsilon i} = 0$ for all i, and therefore also $h_i = 0$ by (R2). Hence $\mathfrak{g}_5(C)$ with an additional relation (6) is the zero Lie algebra. □

Now we want to show that, for any chordless cycle in a signed valued quiver of mutation Dynkin type, most of the (R5) relations are redundant. If our cycle is of type B_3 or F_4 we need to take at most two of the (R5) relations; in other cases just one (R5) relation is enough.

First we present a lemma that holds for arbitrary Lie algebras, and will be useful in the following (see [4, Lemma 4.2]).

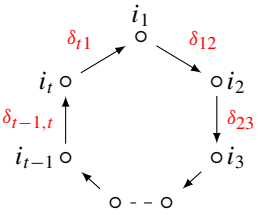
Lemma 4.9. *Let $\mathfrak{g}$ be a Lie algebra, and $x_1, \ldots, x_t \in \mathfrak{g}$ such that $[x_i, x_j] = 0$ for all $|i - j| \neq 1$. Then $[x_t, \ldots, x_1] = (-1)^{t+1}[x_1, \ldots, x_t]$.*

We start by analysing the simply laced case.

Lemma 4.10. *Let (Q, v) be a signed quiver of mutation Dynkin type with a chordless cycle $i_1 \to i_2 \to \dots i_t \to i_1$ where every edge has weight 1 and $i_q \to i_{q+1}$ has sign $\delta_{q,q+1}$. Let $\varepsilon_1 \in \{\pm 1\}$ and quotient $\mathfrak{g}_4(C)$ by the relation $[e_{\varepsilon_1 i_1}, e_{\varepsilon_2 i_2}, \dots, e_{\varepsilon_t i_t}] = 0$, where $\varepsilon_{q+1} = -\delta_{q,q+1}\varepsilon_q$. Then*

$$[e_{\varepsilon_2 i_2}, \dots, e_{\varepsilon_t i_t}, e_{\varepsilon_{t+1} i_1}] = 0,$$

where $\varepsilon_{t+1} = -\delta_{t1}\varepsilon_t$.



Proof. To simplify the notation we let $i_q = q$. Let $C = C(Q, v)$ be the Cartan counterpart of (Q, v) and note that $\delta_{q,q+1} = c_{q,q+1}$, as every edge has weight 1.

Suppose $[e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] = 0$ and apply $[e_{-\varepsilon_1 1}, -]$. Using the Jacobi identity, (R2), and Lemma 4.2, we get

$$\begin{aligned} 0 &= [e_{-\varepsilon_1 1}, e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] \\ &= -\varepsilon_1 [h_1, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] + [e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] \\ &= -\varepsilon_1 \lambda [e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] + [e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] \end{aligned}$$

where $\lambda = \sum_{j=2}^{t} \varepsilon_j c_{1j}$. As our cycle is chordless, $c_{1j} = 0$ for $j = 3, \dots, t-1$, and the weight 1 assumption implies $c_{12} = \delta_{12}$ and $c_{1t} = \delta_{t1}$, so

$$\lambda = \varepsilon_2 c_{12} + \varepsilon_t c_{1t} = \varepsilon_1(-\delta_{12}^2 + (-1)^{t-1}\delta_{12}\delta_{23}\cdots\delta_{t-1,t}\delta_{t1})$$

Now, since Q is mutation Dynkin, the cycle C is not dangerous, and therefore $\delta_{12}\dots\delta_{t-1,t}\delta_{1t} = (-1)^{t+1}$, which implies $\lambda = 0$. So $[e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_t t}] = 0$.

We have $[e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}] = [e_{-\varepsilon_1 1}, e_{-\varepsilon_1 \delta_{12} 2}]$, so by Lemma 4.3 this is zero. Also, as our cycle is chordless, we have $[e_{-\varepsilon_1 1}, e_{\varepsilon_j j}] = 0$ for $j = 3, \dots, t-1$. Therefore, repeatedly using the Jacobi identity gives

$$[e_{\varepsilon_1 1}, \dots, e_{\varepsilon_{t-1} t-1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] = 0.$$

Applying $[e_{-\varepsilon_1 1}, -]$ once again, we get

$$\begin{aligned} 0 &= [e_{-\varepsilon_1 1}, e_{\varepsilon_1 1}, \dots, e_{\varepsilon_{t-1} t-1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] \\ &= -\varepsilon_1 [h_1, e_{\varepsilon_2 2}, \dots, e_{\varepsilon_{t-1} t-1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] + [e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, \dots, e_{\varepsilon_{t-1} t-1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] \\ &= -\varepsilon_1 \mu [e_{\varepsilon_2 2}, \dots, e_{\varepsilon_{t-1} t-1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] + [e_{\varepsilon_1 1}, \dots, e_{\varepsilon_{t-1} t-1}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] \end{aligned}$$

where $\mu = \varepsilon_2 c_{12} - \varepsilon_1 c_{11} + \varepsilon_t c_{1t} = \lambda - 2\varepsilon_1 = -2\varepsilon_1 \neq 0$. Also, the second summand of the above equation is zero, since $[e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_t t}] = 0$ by (R4). As a consequence, by skew-symmetry we have

$$[e_{\varepsilon_2 2}, e_{\varepsilon_3 3}, \ldots, e_{\varepsilon_{t-1} t-1}, e_{\varepsilon_t t}, e_{-\varepsilon_1 1}] = 0$$

which proves the claim since $\varepsilon_{t+1} = (-1)^t \delta_{12}\delta_{23}\cdots\delta_{t-1,t}\delta_{t1}\varepsilon_1 = -\varepsilon_1$. □

Corollary 4.11. *Under the assumptions of Lemma 4.10, all the relations of type* (R5) *associated to the cycle* $i_1 \to i_2 \to \ldots i_t \to i_1$ *are satisfied.*

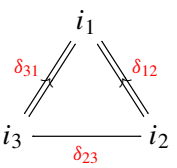
Proof. Notice that $\varepsilon_{t+1} = -\delta_{t1}\varepsilon_t = -(-1)^{t+1}\delta_{12}\ldots\delta_{t-1,t}\delta_{t1}\varepsilon_1 = -\varepsilon_1$, and therefore applying Lemma 4.10 $2t$ times we get all the possible clockwise assignments of signs that satisfy the conditions of (R5).

The above computation also shows that $\varepsilon_t = \delta_{t1}\varepsilon_1$. Therefore, using (R4) and Lemma 4.3 one sees that $e_{\varepsilon_1 i_1}$ commutes with $e_{\varepsilon_j i_j}$ for all $|i - j| \neq 1$. Thus Lemma 4.9 implies that

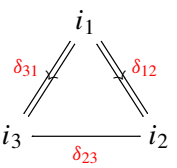
$$[e_{\varepsilon_t i_t}, \ldots, e_{\varepsilon_1 i_1}] = (-1)^{t+1}[e_{\varepsilon_1 i_1}, \ldots, e_{\varepsilon_t i_t}] = 0.$$

Hence, using the same argument as above we also get all possible anticlockwise assignments of signs that satisfy the conditions of (R5). □

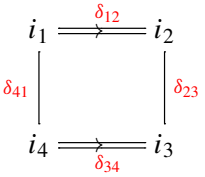
We now give the statement for B_3, C_3 and F_4-cycles. We consider signs on such cycles as follows.



B_3-cycle



C_3-cycle



F_4-cycle

Lemma 4.12. *Let* (Q, v) *be a signed valued quiver of mutation Dynkin type.*

(1) *Suppose* Q *contains a* B_3*-cycle. Then the following two relations imply all the other* (R5) *relations*

$$[e_{\varepsilon i_1}, e_{-\delta_{12}\varepsilon i_2}, e_{\delta_{12}\delta_{23}\varepsilon i_3}] = 0$$

for all $\varepsilon \in \{\pm 1\}$.

(2) *Suppose* (Q, v) *contains a* C_3*-cycle. Then any one* (R5) *relation for this cycle implies all other* (R5) *relations for this cycle.*

(3) *Suppose* Q *contains an* F_4*-cycle. Then the following two relations imply all the other* (R5) *relations*

$$[e_{\varepsilon i_2}, e_{-\delta_{23}\varepsilon i_3}, e_{\delta_{23}\delta_{34}\varepsilon i_4}, e_{-\delta_{23}\delta_{34}\delta_{41}\varepsilon i_1}] = 0$$

for all $\varepsilon \in \{\pm 1\}$.

Proof. Again, let $i_j = j$. For the C_3-cycle, the same argument as in Lemma 4.10 goes through in all cases except one: that $[e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] = 0$ implies $[e_{\varepsilon_2 2}, e_{\varepsilon_3 3}, e_{\varepsilon_4 1}] = 0$.

Therefore, suppose $[e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] = 0$. As in the proof of Lemma 4.10, we get that $[e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] = 0$. Our Cartan matrix is

$$C = \begin{pmatrix} 2 & 2\delta_{12} & 2\delta_{31} \\ \delta_{12} & 2 & \delta_{23} \\ \delta_{31} & \delta_{23} & 2 \end{pmatrix}$$

so, applying $[e_{-\varepsilon_1 1}, -]$ once more, we get

$$\begin{aligned} 0 &= [e_{-\varepsilon_1 1}, e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}], \\ 0 &= -\varepsilon_1 [h_1, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] + [e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}], \\ 0 &= -\varepsilon_1 \mu [e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] + [e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}], \end{aligned}$$

where $\mu = 2\delta_{12}(-\delta_{12}\varepsilon) - 2\varepsilon_1 + 2\delta_{13}\delta_{23}\delta_{12}\varepsilon_1 = -2\varepsilon_1$. Notice that, in contrast to the simply laced case, relation (R4) does not imply that the second summand is 0. In any case, we have

$$2[e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] + [e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] = 0.$$

Applying $[e_{-\varepsilon_1 1}, -]$ again, we get

$$\begin{aligned} 0 &= 2[e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] + [e_{-\varepsilon_1 1}, e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] \\ &= 2[e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] - \varepsilon_1 [h_1, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] \\ &\qquad + [e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] \\ &= (2 - \varepsilon_1 \nu)[e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] + [e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] \\ &= (2 - \varepsilon_1 \nu)[e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}], \end{aligned}$$

where the last identity follows from (R4), since $[e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] = 0$. Here

$$\begin{aligned} \nu &= \varepsilon_2 c_{12} - 2\varepsilon_1 c_{11} + \varepsilon_3 c_{13} = 2\varepsilon_2 \delta_{12} - 4\varepsilon_1 + 2\varepsilon_3 \delta_{31} \\ &= -2\varepsilon_1 - 4\varepsilon_1 + 2\varepsilon_1 \delta_{12}\delta_{23}\delta_{31} = -4\varepsilon_1, \end{aligned}$$

so $2 - \varepsilon_1 \nu = 6 \neq 0$, and therefore we get

$$[e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] = 0.$$

Finally, if we apply $[e_{\varepsilon_1 1}, -]$ we get

$$\begin{aligned} 0 &= [e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] = [e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] \\ &= \varepsilon_1 [h_1, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] + [e_{-\varepsilon_1 1}, e_{\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] \\ &= \varepsilon_1 \lambda_1 [e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] + \varepsilon_1 [e_{-\varepsilon_1 1}, h_1, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] + [e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] \\ &= \varepsilon_1 \lambda_1 [e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] + \varepsilon_1 \lambda_2 [e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] + [e_{-\varepsilon_1 1}, e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_1 1}, e_{\varepsilon_3 3}]. \end{aligned}$$

Now, $e_{\varepsilon_1 1}$ and $e_{\varepsilon_3 3}$ commute by (R4), and therefore the last term is zero. Direct computation shows that $\lambda_1 = -\varepsilon_1 c_{11} + \varepsilon_2 c_{12} + \varepsilon_3 c_{13} = -2\varepsilon_1(\delta_{12}\delta_{23} - \delta_{12} - 1) \neq 0$ and $\lambda_2 = \varepsilon_2 c_{12} + \varepsilon_3 c_{13} = 2\varepsilon_2(\delta_{12} - \delta_{31}\delta_{23}) = 0$. Therefore

$$0 = [e_{-\varepsilon_1 1}, e_{\varepsilon_2 2}, e_{\varepsilon_3 3}] = [e_{\varepsilon_2 2}, e_{-\varepsilon_1 1}, e_{\varepsilon_3 3}] = -[e_{\varepsilon_2 2}, e_{\varepsilon_3 3}, e_{-\varepsilon_1 1}]$$

and, as $\varepsilon_4 = -\delta_{31}\varepsilon_3 = -\varepsilon_1$, we have the claim.

In particular, for type C_3-cycles, a result that is analogous to Corollary 4.11 follows, and therefore we just need one relation from (R5) to get all the others.

For B_3-cycles, Definition 4.4 specifies that we should introduce 8 relations. However, the argument of Lemma 4.10 works in all cases but for

$$[e_{\varepsilon i_2}, e_{-\delta_{23}\varepsilon i_3}, e_{\delta_{23}\delta_{31}\varepsilon i_1}] \Longrightarrow [e_{-\delta_{23}\varepsilon i_3}, e_{\delta_{23}\delta_{31}\varepsilon i_1}, e_{-\varepsilon i_2}].$$

Thus we can reduce (R4) to 4 relations. But Lemma 4.9 allows us to get the anticlockwise relations in term of the clockwise ones, and therefore we just need the two relations from the statement.

For F_4-cycles, Definition 4.4 specifies that we should introduce 12 relations. However, the argument of Lemma 4.10 works in all cases but for

$$[e_{\varepsilon i_4}, e_{-\delta_{41}\varepsilon i_1}, e_{\delta_{41}\delta_{12}\varepsilon i_2}, e_{\delta_{41}\delta_{12}\delta_{13}\varepsilon i_3}] \Longrightarrow [e_{-\delta_{41}\varepsilon i_1}, e_{\delta_{41}\delta_{12}\varepsilon i_2}, e_{\delta_{41}\delta_{12}\delta_{13}\varepsilon i_3}, e_{-\varepsilon i_4}].$$

Thus we can reduce (R4) to 4 relations. But Lemma 4.9 allows us to get the anticlockwise relations in term of the clockwise ones, and therefore we just need the two relations from the statement. □

Remark 4.13. Computer calculations with GAP (http://gap-system.org) suggest that even in types B and F, in the Lie algebra $\mathfrak{g}_4(C)$ any one of our relations from (R5) implies all the others.

4C. *Homomorphisms.*

Definition 4.14. Let (Q, v) be a signed valued quiver of mutation Dynkin type. We define $\mathfrak{g}(Q, v) = \mathfrak{g}_5(C)$, where $C = C(Q, v)$ is the Cartan counterpart of (Q, v).

Fix $k \in Q_0$, and let $C' = C(\mu_k(Q, v))$ and $\mathfrak{g}(\mu_k(Q, v)) = \mathfrak{g}_5(C')$, with generators of $\mathfrak{g}_5(C')$ written as h'_i and $e'_{\varepsilon i}$.

In the following formulas we write $i \xrightarrow{\delta} k$, where $\delta \in \{\pm 1\}$, to denote the sign of an arrow $i \to k$. Write $C = (c_{ij})$.

Theorem 4.15. *There is an isomorphism of Lie algebras*

$$\varphi_k : \mathfrak{g}(\mu_k(Q, v)) \xrightarrow{\sim} \mathfrak{g}(Q, v)$$

given by the formulas

$$e'_{\varepsilon i} \mapsto \begin{cases} \dfrac{(-\varepsilon)^{c_{ki}}}{|c_{ki}|!} \operatorname{ad}^{|c_{ki}|}(e_{-\delta\varepsilon k})(e_{\varepsilon i}) & \text{if } i \xrightarrow{\delta} k \text{ in } Q, \\ e_{\varepsilon i} & \text{otherwise;} \end{cases}$$

$$h'_i \mapsto \begin{cases} h_i - c_{ik} h_k & \text{if } i \to k \text{ in } Q, \\ h_i & \text{otherwise.} \end{cases}$$

Its inverse $\psi_k : \mathfrak{g}(Q, v) \xrightarrow{\sim} \mathfrak{g}(\mu_k(Q, v))$ *is given by*

$$e_{\varepsilon i} \mapsto \begin{cases} \dfrac{(-\varepsilon)^{c_{ki}}}{|c_{ki}|!} \operatorname{ad}^{|c_{ki}|}(e'_{\delta\varepsilon k})(e'_{\varepsilon i}) & \text{if } i \xrightarrow{\delta} k \text{ in } Q, \\ e_{\varepsilon i} & \text{otherwise;} \end{cases}$$

$$h_i \mapsto \begin{cases} h'_i + c_{ik} h'_k & \text{if } i \to k \text{ in } Q, \\ h'_i & \text{otherwise.} \end{cases}$$

The rest of this section is devoted to proving Theorem 4.15. Our proof will rely on the isomorphisms of Lie algebras given in [4] (for simply laced types) and [14] (for nonsymmetric Cartan matrices).

We recall some notation from [14]. We have elementary matrices

$$E^{\sigma}_{sr} = I + \sigma e_s e_r^T$$

where I is the $n \times n$ identity matrix, $\sigma \in \mathbb{Z}$, e_r is the r-th basis vector of the space of column vectors $\mathbb{R}^n$, and e_r^T is its transpose. Write $I_r = E^{-2}_{rr}$, so that conjugating by I_r negates the rth row and column. If C is a quasi-Cartan matrix with symmetrizer D and $s \neq r$, let

$$T^{\sigma}_{sr}(C) = D^{-1} E^{\sigma}_{rs} D C E^{\sigma}_{sr} \quad \text{and} \quad J_r(C) = D^{-1} I_r D C I_r.$$

Note that $T^0_{sr}(C) = C$ and J_r is self-inverse.

We now state the fundamental results of Barot–Rivera and Pérez–Rivera in terms of our notation and conventions.

Proposition 4.16 (Barot–Rivera, Pérez–Rivera). *For any positive quasi-Cartan matrix* C *and* $1 \leq r \leq n$, *there is an isomorphism of Lie algebras* $\mathfrak{g}_5(J_r(C)) \to \mathfrak{g}_5(C)$ *sending*

$$e'_{\varepsilon i} \mapsto \begin{cases} e_{\varepsilon i} & \text{if } i \neq r, \\ -e_{-\varepsilon r} & \text{if } i = r; \end{cases} \qquad h'_i \mapsto \begin{cases} h_i & \text{if } i \neq r, \\ -h_r & \text{if } i = r. \end{cases}$$

Proof. By [14, Proposition 2.12], following [4, Proposition 2.6] in simply laced types, there is an isomorphism $\hat{\mathfrak{g}}_5(J_r(C)) \to \hat{\mathfrak{g}}_5(C)$ sending $\hat{e}'_{\varepsilon r}$ to $\hat{e}_{-\varepsilon r}$, $\hat{h}'_r$ to $-\hat{h}_r$, and for $i \neq r$, $\hat{e}'_{\varepsilon i} \mapsto \hat{e}_{\varepsilon i}$ and $\hat{h}'_i \mapsto \hat{h}_i$. We compose this with isomorphisms from Lemma 4.5 to get the isomorphisms $\mathfrak{g}_5(J_r(C)) \to \hat{\mathfrak{g}}_5(J_r(C)) \to \hat{\mathfrak{g}}_5(C) \to \mathfrak{g}_5(C)$ in the given formula. □

Proposition 4.17 (Barot–Rivera, Pérez–Rivera). *For any positive quasi-Cartan matrix C and any $1 \le r, s \le n$ with $r \ne s$, if $c_{sr} = -1$ or -2 then there is an isomorphism of Lie algebras $\mathfrak{g}_5(T_{sr}^{-c_{sr}}(C)) \to \mathfrak{g}_5(C)$ sending*

$$e'_{\varepsilon i} \mapsto \begin{cases} e_{\varepsilon i} & \text{if } i \ne r, \\ \dfrac{(-\varepsilon)^{|c_{sr}|}}{|c_{sr}|} \operatorname{ad}^{|c_{sr}|}(e_{\varepsilon s})(e_{\varepsilon r}) & \text{if } i = r; \end{cases} \qquad h'_i \mapsto \begin{cases} h_i & \text{if } i \ne r, \\ h_r - c_{rs} h_s & \text{if } i = r. \end{cases}$$

Proof. By [14, Proposition 2.14], following [4, Proposition 2.8] in simply laced types, there is an isomorphism $\widehat{\mathfrak{g}}_5(J_r(C)) \to \hat{\mathfrak{g}}_5(C)$. Note the typo in the statement of [14, Proposition 2.14]: the superscript on T should have a minus sign. According to equations (4.2) and (4.3) of [14], this isomorphism sends $\hat{e}'_{\varepsilon i} \mapsto \hat{e}_{\varepsilon i}$ and $\hat{h}'_i \mapsto \hat{h}_i$ for $i \ne r$, and for $i = r$ it sends:

$$\hat{e}'_{\varepsilon i} \mapsto \begin{cases} [\hat{e}_{\varepsilon r}, \hat{e}_{\varepsilon s}] & \text{if } c_{sr} = -1, \\ \frac{1}{2}[\hat{e}_{\varepsilon s}, \hat{e}_{\varepsilon s}, \hat{e}_{\varepsilon r}] & \text{if } c_{sr} = -2; \end{cases} \qquad \hat{h}'_i \mapsto \begin{cases} \hat{h}_r - c_{rs}\hat{h}_s & \text{if } c_{sr} = -1, \\ \hat{h}_r + \hat{h}_s & \text{if } c_{sr} = -2. \end{cases}$$

By the usual Dynkin classification, together with Lemma 2.40, we have that $c_{sr} = -2$ implies $c_{rs} = -1$. So we can unify the formulas for $i = r$ as follows:

$$\hat{e}'_{\varepsilon r} \mapsto \frac{(-1)^{c_{sr}}}{|c_{sr}|} \operatorname{ad}^{|c_{sr}|}(\hat{e}_{\varepsilon s})(\hat{e}_{\varepsilon r}), \quad \hat{h}'_r \mapsto \hat{h}_r - c_{rs}\hat{h}_s.$$

Now conjugate by the isomorphisms from Lemma 4.5. For $e'_{\varepsilon i}$ we get $e'_{\varepsilon i} \mapsto \varepsilon \hat{e}'_{\varepsilon i}$, giving one ε, then from $\operatorname{ad}^{|c_{sr}|}(\hat{e}_{\varepsilon s})(\hat{e}_{\varepsilon i})$ we pick up another $\varepsilon^{1+c_{sr}}$. This gives the formula in the statement. □

Lemma 4.18. *For any positive quasi-Cartan matrix C and any $1 \le r, s \le n$ with $r \ne s$, if $c_{sr} = -3$ then there is an isomorphism of Lie algebras $\mathfrak{g}_5(T_{sr}^3(C)) \to \mathfrak{g}_5(C)$ sending*

$$e'_{\varepsilon i} \mapsto \begin{cases} e_{\varepsilon i} & \text{if } i \ne r, \\ \dfrac{-\varepsilon}{6} \operatorname{ad}^3(e_{\varepsilon s})(e_{\varepsilon r}) & \text{if } i = r; \end{cases} \qquad h'_i \mapsto \begin{cases} h_i & \text{if } i \ne r, \\ h_r + h_s & \text{if } i = r. \end{cases}$$

Proof. By Lemma 2.40, the weight 3 cannot occur in a chordless cycle, so the only possibility is classical type G_2. Up to relabelling, this means we must have

$$C = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix} \quad \text{and} \quad T_{12}^3 C = \begin{pmatrix} 2 & 3 \\ C & 2 \end{pmatrix}.$$

Then our candidate isomorphism acts as follows:

$$h'_1 \mapsto h_1, \qquad h'_2 \mapsto h_1 + h_2, \quad e'_1 \mapsto e_1,$$
$$e'_2 \mapsto -\tfrac{1}{6}[e_1, e_1, e_1, e_2], \quad e'_{-1} \mapsto e_{-1}, \qquad e'_{-2} \mapsto -\tfrac{1}{6}[e_{-1}, e_{-1}, e_{-1}, e_{-2}].$$

The checks that this preserves relations (R1)-(R4) are straightforward, except for the relation $[e'_1, e'_{-1}] = h'_1$, but a careful calculation shows that

$$\begin{aligned}\big[[e_1, e_1, e_1, e_2], [e_{-1}, e_{-1}, e_{-1}, e_{-2}]\big] &= 3\,(c_{11}+c_{12})\,(c_{11}+2c_{12})\,(3c_{21}h_1 + c_{12}h_2)\\ &= -36(h_1+h_2).\end{aligned}$$

Then the inverse is easy to write down directly. □

We will consider cases where c_{sr} is positive, as well as negative, so we define:

$$U_{sr}(C) = \begin{cases} T_{sr}^{-c_{sr}}(C) & \text{if } c_{sr} \le 0,\\ J_r T_{sr}^{c_{sr}} J_r(C) & \text{if } c_{sr} > 0.\end{cases}$$

In Lemma 4.20 we will see that we can interpret U_{sr} in a more straightforward way just in terms of the transformation $T_{sr}^{-c_{sr}}$, regardless of whether c_{sr} is positive or negative. For the moment we stick to the definition above because the setting of the proof of the next result requires $c_{sr} < 0$.

Proposition 4.19. *For any positive quasi-Cartan matrix C and any $1 \le r, s \le n$ with $r \ne s$, there is an isomorphism of Lie algebras $\mathfrak{g}_5(U_{sr}(C)) \to \mathfrak{g}_5(C)$ sending*

$$e'_{\varepsilon i} \mapsto \begin{cases} e_{\varepsilon i} & \text{if } i \ne r,\\ \dfrac{(-\varepsilon)^{|c_{sr}|}}{|c_{sr}|!}\,\mathrm{ad}^{|c_{sr}|}(e_{-\delta\varepsilon s})(e_{\varepsilon r}) & \text{if } i = r;\end{cases} \qquad h'_i \mapsto \begin{cases} h_{\varepsilon i} & \text{if } i \ne r,\\ h_r - c_{rs}h_s & \text{if } i = r,\end{cases}$$

where $\delta = \mathrm{sgn}(c_{sr})$.

Proof. When $c_{sr} = 0$ we have $U_{sr}(C) = C$ and we interpret our homomorphism as the identity map. When $c_{sr} < 0$ we have $\delta = -1$ and the formulas reduce to those in Proposition 4.17 (or Lemma 4.18); this also works for $c_{sr} = 0$. When $c_{sr} > 0$ we compose the isomorphisms from Propositions 4.16, 4.17 and Lemma 4.18:

$$\mathfrak{g}(U_{sr}(C)) = \mathfrak{g}(J_r T_{sr}^{c_{sr}} J_r(C)) \xleftarrow{4.16} \mathfrak{g}(T_{sr}^{c_{sr}} J_r(C)) \xleftarrow{4.17} \mathfrak{g}(J_r(C)) \xleftarrow{4.16} \mathfrak{g}(C).$$

Write $C^{(3)} = J_r T_{sr}^{c_{sr}} J_r(C)$, $C^{(2)} = T_{sr}^{c_{sr}} J_r(C)$, etc., using the same superscripts for the generators of the Lie algebras. Then the isomorphism sends $h'_i = h_i^{(3)} \mapsto h_i$ for $i \ne r$, and

$$h'_r = h_r^{(3)} \mapsto -h_r^{(2)} \mapsto -h_r^{(1)} + c_{rs}^{(1)} h_s^{(1)} \mapsto h_r + c_{rs}^{(1)} h_s = h_r - c_{rs}h_s$$

where $C^{(1)} = J_r(C)$ gives $c_{rs}^{(1)} = -c_{rs}$. So the result for h_i does not depend on δ.

The isomorphism sends $e_{\varepsilon i}^{(3)} \mapsto e_{\varepsilon i}$ for $i \ne r$, and

$$e'_{\varepsilon r} \mapsto -e_{-\varepsilon r}^{(2)} \mapsto \frac{(-\varepsilon)^{|c_{sr}^{(1)}+1|}}{|c_{sr}^{(1)}|!}\,\mathrm{ad}^{|c_{sr}^{(1)}|}(e_{-\varepsilon s}^{(1)})(e_{-\varepsilon r}^{(1)}) \mapsto \frac{(-\varepsilon)^{|c_{sr}|}}{|c_{sr}|!}\,\mathrm{ad}^{|c_{sr}|}(e_{-\varepsilon s})(e_{\varepsilon r})$$

as $|c^{(1)}_{sr}| = |c_{sr}|$. So when $\delta = 1$ we have an $\mathrm{ad}^{|c_{sr}|}(e_{-\varepsilon s})$; when $\delta = -1$ we have an $\mathrm{ad}^{|c_{sr}|}(e_{\varepsilon s})$. The $-\delta$ in our formula covers both cases. □

Lemma 4.20. $U_{sr}(C) = T_{sr}^{-c_{sr}}(C)$.

Proof. As noted in the proof of [4, Lemma 2.4], it is easily verified that $E^{\sigma}_{rs} I_i = I_i E^{-\sigma}_{rs}$ if $i \in \{r, s\}$, and $E^{\sigma}_{rs} I_i = I_i E^{\sigma}_{rs}$ otherwise. From this, it follows from the definitions that

$$T^{\sigma}_{rs} J_i = \begin{cases} J_i T^{-\sigma}_{rs} & \text{if } i \in \{r, s\}, \\ J_i T^{\sigma}_{rs} & \text{if } i \notin \{r, s\} \end{cases}$$

and so we get $J_r T^{-\sigma}_{sr} J_r = T^{\sigma}_{sr}$, and the statement follows. □

We get the next result by using the definition of U_{sr} in terms of E^{σ}_{sr}.

Lemma 4.21. $U_{sr_1} U_{sr_2} = U_{sr_2} U_{sr_1}$.

Now we use our signed valued quiver (Q, v). Given the previous lemma, we can unambiguously write the product in the following lemma.

Lemma 4.22. $C(\mu_k(Q, v)) = \left(\prod_{i \to k} U_{ki}\right) C(Q, v)$.

Proof. We understand $C(\mu_k(Q, v))$ by Lemma 3.9, so we just need to understand $\left(\prod_{i \to k} U_{kk}\right) C(Q, v)$. If $c_{sr} < 0$ then a useful formula for $C' = U_{sr}(C) = T^{-c_{sr}}_{sr}(C)$, when $c_{sr} < 0$, is given in [14, equation (4.1)]. First note that their three-case equation can be expressed as the following two-case equation:

$$c'_{xy} = \begin{cases} c_{xy} & \text{if } (x = r \text{ and } y = r) \text{ or } (x \neq r \text{ and } y \neq r), \\ c_{xy} - c_{xs} c_{sy} & \text{if } (x = r \text{ and } y \neq r) \text{ or } (x \neq r \text{ and } y = r). \end{cases}$$

A similar computation shows that this formula also holds when $c_{sr} \geq 0$.

Suppose that the vertices i with arrows $i \to k$ are precisely $i = r_1, \ldots, r_\ell$. Let $C^{(0)} = C$ and $C^{(q+1)} = U_{kr_q} C^{(q)}$, and write $C' = C^{(\ell)}$. We use the above formula to calculate c'_{xy} in four cases, depending on whether or not $x, y \in \{r_1, \ldots, r_\ell\}$.

- If $x, y \notin \{r_1, \ldots, r_\ell\}$ then $c^{(q+1)}_{xy} = c^{(q)}_{xy}$ at each stage, so $c'_{xy} = c_{xy}$.
- If $x = r_u$ and $y \notin \{r_1, \ldots, r_\ell\}$ then $c'_{xy} = c^{(u+1)}_{xy} = c^{(u)}_{xy} - c^{(u)}_{xs} c^{(u)}_{sy} = c_{xy} - c_{xs} c_{sy}$.
- The case $x \notin \{r_1, \ldots, r_\ell\}$ and $y = r_u$ is exactly the same.
- Suppose $x = r_u$ and $y = r_v$ with $u < v$; the case $v < u$ is similar. Then

$$c^{(u+1)}_{xy} = c^{(u)}_{xy} - c^{(u)}_{xs} c^{(u)}_{sy} \quad \text{and} \quad c^{(v+1)}_{xy} = c^{(v)}_{xy} - c^{(v)}_{xs} c^{(v)}_{sy}$$

  with $c^{(p+1)}_{xy} = c^{(p)}_{xy}$ for $p \neq u, v$. So, using $c_{ss} = 2$, we have

$$\begin{aligned} c'_{xy} &= c^{(v)}_{xy} - c^{(v)}_{xs} c^{(v)}_{sy} = c^{(u+1)}_{xy} - c^{(u+1)}_{xs} c^{(u+1)}_{sy} \\ &= \left(c^{(u)}_{xy} - c^{(u)}_{xs} c^{(u)}_{sy}\right) - \left(c^{(u)}_{xs} - c^{(u)}_{xs} c^{(u)}_{ss}\right) c^{(u)}_{sy} \\ &= c^{(u)}_{xy} - c^{(u)}_{xs} c^{(u)}_{sy} + c^{(u)}_{xs} c^{(u)}_{sy} = c_{xy} \end{aligned}$$

and our answer agrees with Lemma 3.9 in all cases. □

Lemma 4.23. *Let $k, i \in Q_0$, and $c_{ki} = \delta\lambda$, where $\delta \in \{\pm 1\}$. Then*

$$[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] = 2\lambda(\lambda - 1)e_{\varepsilon i}.$$

Proof. Using the Jacobi identity, (R2), (R3) and (R4) we have

$$\begin{aligned}
&[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] \\
&\qquad = \delta\varepsilon[e_{\delta\varepsilon k}, h_k, e_{-\delta\varepsilon k}, e_{\varepsilon i}] + [e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] \\
&\qquad = \delta\varepsilon(-2\delta\varepsilon + \lambda\delta\varepsilon)[e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] + \delta\varepsilon[e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, h_k, e_{\varepsilon i}] \\
&\qquad = (-2 + \lambda)[e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] + \lambda[e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] \\
&\qquad = (-2 + \lambda + \lambda)\delta\varepsilon[h_k, e_{\varepsilon i}] \\
&\qquad = 2\lambda(\lambda - 1)e_{\varepsilon i}.
\end{aligned}$$

□

Proof of Theorem 4.15. If (Q, v) has arrows $i \to k$ precisely for vertices $i = r_1, \ldots, r_\ell$, we get the Lie algebra homomorphism

$$\mathfrak{g}(\mu_k(Q, v)) \to \mathfrak{g}(Q, v)$$

by composing the morphisms from Proposition 4.19, using Lemma 4.22, as follows:

$$\begin{aligned}
\varphi_k : \mathfrak{g}(\mu_k(Q, v)) = \mathfrak{g}(U_{sr_\ell} U_{sr_{\ell-1}} \cdots U_{sr_1} C) \to \\
\mathfrak{g}(U_{sr_{\ell-1}} \cdots U_{sr_1} C) \to \cdots \to \mathfrak{g}(U_{sr_1} C) \to \mathfrak{g}(C) = \mathfrak{g}(Q, v).
\end{aligned}$$

This gives the formula for φ_k in the statement of the theorem.

Now we prove that φ_k and ψ_k are inverse. For simplicity we omit the index k. We will prove that $\varphi \circ \psi = \mathrm{id}$ on the generators. For generators at vertices i, where i does not have an arrow going into k, the identity follows by definition of ψ and φ. Therefore, we can assume there is an arrow $i \to k$ in Q.

We have

$$\varphi \circ \psi(h_i) = \varphi(h'_i + c_{ik}h'_k) = h_i - c_{ik}h_k + c_{ik}h_k = h_i.$$

To show that $\varphi \circ \psi(e_{\varepsilon i}) = e_{\varepsilon i}$ we divide into cases according to the value of c_{ki}.

- If $c_{ki} = \delta \in \{\pm 1\}$ then $\varphi \circ \psi(e_{\varepsilon i}) = -\varepsilon\varphi([e'_{\delta\varepsilon k}, e'_{\varepsilon i}]) = -\varepsilon \cdot -\varepsilon[e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] = \delta\varepsilon[h_k, e_{\varepsilon i}] = \delta\varepsilon\delta\varepsilon e_{\varepsilon i} = e_{\varepsilon i}$, where we used $[e_{-\delta\varepsilon k}, e_{\varepsilon i}] = 0$ from Lemma 4.3.
- If $c_{ki} = 2\delta \in \{\pm 2\}$ then

$$\varphi \circ \psi(e_{\varepsilon i}) = \tfrac{1}{2}\varphi([e'_{\delta\varepsilon k}, e'_{\delta\varepsilon k}, e'_{\varepsilon i}]) = \tfrac{1}{4}[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] = e_{\varepsilon i},$$

where the last equality follows from Lemma 4.23.

- If $c_{ki} = 3\delta \in \{\pm 3\}$ then

$$
\begin{aligned}
&\varphi \circ \psi(e_{\varepsilon i}) \\
&= -\tfrac{1}{6}\varepsilon\varphi([e'_{\delta\varepsilon k}, e'_{\delta\varepsilon k}, e'_{\delta\varepsilon k}, e'_{\varepsilon i}]) = \tfrac{1}{36}[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] \\
&= \tfrac{1}{36}(\delta\varepsilon[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, h_k, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] + [e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}]) \\
&= \tfrac{1}{36}(\delta\varepsilon(-2\delta\varepsilon - 2\delta\varepsilon + 3\delta\varepsilon)[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] \\
&\quad + \delta\varepsilon[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, h_k, e_{-\delta\varepsilon k}, e_{\varepsilon i}] + [e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}]) \\
&= \tfrac{1}{36}(-[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] + \delta\varepsilon(-2\delta\varepsilon + 3\delta\varepsilon)[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}] \\
&\quad + \delta\varepsilon[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, h_k, e_{\varepsilon i}]) \\
&= \tfrac{1}{36}(\delta\varepsilon(3\delta\varepsilon)[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}]) \\
&= \tfrac{1}{36}(3[e_{\delta\varepsilon k}, e_{\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{-\delta\varepsilon k}, e_{\varepsilon i}]) = e_{\varepsilon i},
\end{aligned}
$$

where the last equality follows from Lemma 4.23.

One checks $\psi \circ \varphi = \mathrm{id}$ similarly. □

Corollary 4.24. *If the signed valued quiver (Q, v) is mutation Dynkin of type Δ then $\mathfrak{g}(Q, v) \cong \mathfrak{g}(\Delta)$.*

Remark 4.25. By considering mutation theories from higher cluster categories, one expects to find other presentations of semisimple Lie algebras, in analogy with [12].

4D. ***Root space decompositions.*** Let (Q, v) be a signed valued quiver of mutation Dynkin type, and (Q', v') its mutation at k. In this section we give a compatibility result between root space decompositions of the Lie algebras $\mathfrak{g} = \mathfrak{g}(Q, v)$ and $\mathfrak{g}' = \mathfrak{g}'(Q', v')$.

For basics about root space decomposition of simple Lie algebras we refer to [10, Section II.8].

Let $\mathfrak{h}, \mathfrak{h}'$ denote the Cartan subalgebras of $\mathfrak{g}, \mathfrak{g}'$ generated by h_i and h'_i, respectively. Denote by Φ (resp. Φ') the root system associated to (Q, v) (resp. (Q', v')) according to Definition 3.11. We define a map $\Phi \times \mathfrak{h} \to \mathbb{C}$ by $\alpha_i(h_j) = c_{ji}$ [15, Section VI.4].

By Corollary 4.24 we know that $\mathfrak{g}$ and $\mathfrak{g}'$ are simple complex Lie algebras. By Theorem 3.21 their root systems are Φ and Φ'. So we have root space decompositions

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\beta \in \Phi} \mathfrak{g}^{\beta}, \quad \mathfrak{g}' = \mathfrak{h}' \oplus \bigoplus_{\beta' \in \Phi'} \mathfrak{g}'^{\beta'},$$

where

$$g^{\beta} = \{x \in \mathfrak{g} \mid [h, x] = \beta(h)x \text{ for all } h \in \mathfrak{h}\}.$$

By (R4) we have $[h_i, e_j] = c_{ij}e_j = \alpha_j(h_i)e_j$, so $e_j \in \mathfrak{g}^{\alpha_j}$.

Recall the isomorphism of root systems $\rho_k : \Phi' \to \Phi$ from Definition 3.18.

Theorem 4.26. $\varphi_k(\mathfrak{g}'^{\beta'}) = \mathfrak{g}^{\rho_k(\beta')}$.

Proof. Suppose $x' \in \mathfrak{g}'^{\beta'}$, so that $[h', x'] = \beta'(h')x'$ for all $h' \in \mathfrak{h}'$. Therefore $\varphi_k(x') \in \mathfrak{g}^{\rho_k(\beta')}$ if and only if

$$[h, \varphi_k(x')] = \rho_k(\beta')(h)\varphi_k(x')$$

for all $h \in \mathfrak{h}$.

Now, fix $h \in \mathfrak{h}$. Notice that φ_k induces an isomorphism $\varphi_{k|\mathfrak{h}'} : \mathfrak{h}' \xrightarrow{\sim} \mathfrak{h}$ by definition, and therefore we can write $h = \varphi_k(h')$ for some $h' \in \mathfrak{h}'$. In particular, since φ_k is a Lie algebra homomorphism, then we can rewrite the left hand side of the above equation as

$$[h, \varphi_k(x')] = \varphi_k([h', x']) = \varphi_k(\beta'(h')x') = \beta'(h')\varphi_k(x').$$

Hence the result follows if we prove that

$$\beta'(h') = \rho_k(\beta')(h), \tag{8}$$

where $h = \varphi_k(h')$.

It is enough to prove (8) for $h = h'_i$ and $\beta' = \alpha'_j$, where h'_i is a generator of $\mathfrak{h}'$ and α'_j is a simple root; the result will then follow by linearity.

Let $C = C(Q, v) = (c_{ij})$ (resp. $C' = C(Q', v') = c'_{ij}$) denote the Cartan counterpart of (Q, v) (resp. (Q', v')). Then Definition 3.18 and Theorem 4.15 imply

$$\varphi_k(h'_i) = \begin{cases} h_i - c_{ik}h_k & \text{if } i \to k, \\ h_i & \text{otherwise;} \end{cases} \qquad \rho_k(\alpha'_j) = \begin{cases} \alpha_j - c_{kj}\alpha_k & \text{if } j \to k, \\ \alpha_j & \text{otherwise.} \end{cases}$$

Therefore, using $\alpha_\ell(h_m) = c_{m\ell}$ and $c_{kk} = 2$ we get

$$\rho_k(\alpha'_j)\varphi_k(h'_i) = \begin{cases} c_{ij} - c_{kj}c_{ik} - c_{ik}c_{kj} + c_{ik}c_{kj}c_{kk} = c_{ij} & \text{if } i, j \to k, \\ c_{ij} - c_{ik}c_{kj} & \text{if } i \to k \text{ and } j \not\to k, \\ c_{ij} - c_{kj}c_{ik} & \text{if } i \not\to k \text{ and } j \to k, \\ c_{ij} & \text{if } i, j \not\to k. \end{cases}$$

Therefore, by Lemma 3.9 we have $\rho_k(\alpha'_j)\varphi_k(h'_i) = c'_{ij} = \alpha'_j(h'_i)$.

Thus we have the inclusion $\varphi_k(\mathfrak{g}'^{\beta'}) \subseteq \mathfrak{g}^{\rho_k(\beta')}$. The other inclusion follows from the classical result that root spaces are 1-dimensional (see, e.g., Theorem 2(b), Chapter VI in [15]), together with the fact that the map φ_k is an isomorphism by Theorem 4.15. □

4E. ***An example of type A_3.*** We conclude with an example involving the signed quiver

$$Q = 1 \leftarrow 2 \leftarrow 3,$$

which has associated gss matrix and Cartan counterpart

$$B = \begin{pmatrix} 0 & -t & 0 \\ t & 0 & -t \\ 0 & t & 0 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Since B is skew-symmetric, D is the identity matrix. Its simple roots are $S = \{\alpha_1, \alpha_2, \alpha_3\}$ and the group W is generated by $\{s_1, s_2, s_3\}$. Its root system is

$$\Phi = W \cdot S = \pm\{\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_1+\alpha_2,\ \alpha_2+\alpha_3,\ \alpha_1+\alpha_2+\alpha_3\}.$$

The Lie algebra $\mathfrak{g}$ is generated by $\{e_i, f_i, h_i\}_{i=1,2,3}$ where we write $f_i = e_{-i}$. We have the usual (R1), (R2), and (R3) relations, and our (R4) relations are

$$\begin{aligned} &\varepsilon = 1,\ \delta = 1: \quad [e_1, e_1, e_2] = [e_1, e_3] = [e_2, e_2, e_1] = [e_2, e_2, e_3] = [e_3, e_3, e_2] = 0, \\ &\varepsilon = 1,\ \delta = -1: \ [e_1, f_2] = [e_1, f_3] = [e_2, f_3] = 0, \end{aligned}$$

together with the relations obtained by interchanging e's and f's and some redundant relations. There are no (R5) relations. So $\mathfrak{g}$ has dimension 15 with basis

$$\begin{aligned} &h_1,\ h_2,\ h_3, \\ &e_1,\ e_2,\ e_3,\ [e_1, e_2],\ [e_2, e_3],\ [e_1, e_2, e_3], \\ &f_1,\ f_2,\ f_3,\ [f_1, f_2],\ [f_2, f_3],\ [f_1, f_2, f_3]. \end{aligned}$$

Classically, it is isomorphic to the Lie algebra $\mathfrak{sl}_4(\mathbb{C})$ of traceless 4×4 matrices by sending the above ordered basis to

$$\begin{aligned} &x_{11} - x_{22},\ x_{22} - x_{33},\ x_{33} - x_{44}, \\ &x_{12},\ x_{23},\ x_{34},\ x_{13},\ x_{24},\ x_{14}, \\ &x_{21},\ x_{32},\ x_{43},\ -x_{31},\ -x_{42},\ x_{43}, \end{aligned}$$

where x_{ij} denotes a matrix with 1 in the (i, j) entry and 0 everywhere else. We define an action of α_j on h_i by $\alpha_j(h_i) = c_{ij}$. Consider weight spaces $\mathfrak{g}^\beta$ for $\beta \in \Phi$. By construction we have $[h_i, e_j] = c_{ij} e_j$, so $e_{\varepsilon j} \in \mathfrak{g}^{\varepsilon\alpha_j}$, and Lemma 4.2 gives the other weight spaces, such as $[e_1, e_2] \in \mathfrak{g}^{\alpha_1+\alpha_2}$.

Now we perform signed mutation of Q at the vertex 2 to get

$$Q' = \begin{array}{ccc} & 2 & \\ {\scriptstyle -}\nearrow & & \searrow{\scriptstyle +} \\ 1 & \xleftarrow[-]{} & 3 \end{array}$$

which has associated gss matrix and Cartan counterpart

$$B' = \begin{pmatrix} 0 & t & -t \\ -t & 0 & 1 \\ t & -1 & 0 \end{pmatrix} \quad \text{and} \quad C' = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & 1 \\ -1 & 1 & 2 \end{pmatrix}.$$

Its simple roots are $S' = \{\alpha'_1, \alpha'_2, \alpha'_3\}$ and the group W' is generated by $\{s'_1, s'_2, s'_3\}$. We get roots

$$s'_1(\alpha'_2) = \alpha'_1 + \alpha'_2, \;\; s'_1(\alpha'_3) = \alpha'_1 + \alpha'_3, \;\; s'_2(\alpha'_3) = -\alpha'_2 + \alpha'_3,$$

so we get

$$\Phi' = W' \cdot S' = \pm\{\alpha'_1,\ \alpha'_2,\ \alpha'_3,\ \alpha'_1 + \alpha'_2,\ \alpha'_1 + \alpha'_3,\ -\alpha'_2 + \alpha'_3\}.$$

The mutation of roots function acts as follows:

$$\begin{aligned} \rho_2 : \Phi' &\to \Phi, \\ \alpha'_1 &\mapsto \alpha_1, \\ \alpha'_2 &\mapsto \alpha_2, \\ \alpha'_3 &\mapsto \alpha_2 + \alpha_3, \\ \alpha'_1 + \alpha'_2 &\mapsto \alpha_1 + \alpha_2, \\ \alpha'_1 + \alpha'_3 &\mapsto \alpha_1 + \alpha_2 + \alpha_3, \\ -\alpha'_2 + \alpha'_3 &\mapsto \alpha_3. \end{aligned}$$

Our Lie algebra $\mathfrak{g}$ is generated by $\{e'_i, f'_i, h'_i\}_{i=1,2,3}$ where we write $f'_i = e'_{-i}$. Our (R4) relations are

$$\begin{aligned} &\varepsilon = 1, \delta = 1: \quad [e'_1, e'_1, e'_2] = [e'_1, e'_1, e'_3] = [e'_2, e'_2, e'_1] = [e'_2, e'_3] = [e'_3, e'_3, e'_2] = 0, \\ &\varepsilon = 1, \delta = -1: \quad [e'_1, f'_2] = [e'_1, f'_3] = [e'_2, e'_2, f'_3] = [e'_3, e'_3, f'_2] = 0, \end{aligned}$$

together with the relations obtained by interchanging e's and f's and some redundant relations. Our (R5) relations are

$$[e'_1, e'_2, f'_3] = [e'_2, f'_3, f'_1] = [e'_3, e'_1, e'_2] = 0$$

together with the relations obtained by interchanging e's and f's and the (redundant) relations going the other way round the cycle. So $\mathfrak{g}$ has dimension 15 with basis

$$h'_1, h'_2, h'_3, e'_1, e'_2, e'_3, [e'_1, e'_2], [e'_1, e'_3], [e'_2, f'_3], f_1, f_2, f_3, [f'_1, f'_2], [f'_1, f'_3], [f'_2, e'_3].$$

Lemma 4.2 gives the weight spaces: for example, $[e'_2, f'_3] \in \mathfrak{g}^{\alpha'_2 - \alpha'_3}$.

The isomorphism $\varphi_2 : \mathfrak{g}' \to \mathfrak{g}$ of Lie algebras acts as follows:

$$\begin{array}{lll} h'_1 \mapsto h_1, & e'_1 \mapsto e_1, & f_1 \mapsto f_1, \\ h'_2 \mapsto h_2, & e'_2 \mapsto e_2, & f_2 \mapsto f_2, \\ h'_3 \mapsto h_2 + h_3, & e'_3 \mapsto -[e_2, e_3], & f_3 \mapsto [f_2, f_3]. \end{array}$$

Note that it preserves weight spaces, as predicted by Theorem 4.26. For example,

$$\begin{array}{ccc} \Phi & \xleftarrow{\rho_2} & \Phi' \\ \downarrow & & \downarrow \\ \{\mathfrak{g}^{\beta} \mid \beta \in \Phi\} & \xleftarrow{\phi_2} & \{\mathfrak{g}'^{\beta'} \mid \beta' \in \Phi'\} \end{array} \qquad \begin{array}{ccc} -\alpha_3 & \longmapsto\!\shortmid & \alpha'_2 - \alpha'_3 \\ \downarrow & & \downarrow \\ -f_3 \in \mathfrak{g}^{-\alpha_3} & \longmapsto\!\shortmid & \mathfrak{g}'^{\alpha'_2 - \alpha'_3} \ni [e'_2, f'_3] \end{array}$$

Composing this isomorphism with the classical realization of $\mathfrak{g}$ as $\mathfrak{sl}_4(\mathbb{C})$, we get

$$e'_1 \mapsto x_{12}, \quad e'_2 \mapsto x_{23}, \quad e'_3 \mapsto -x_{24}.$$

Acknowledgements

Thanks to Bethany Marsh for pointers to the literature and to the referee for a careful reading of the paper.

Thanks to the authors of the computer algebra system GAP (Groups Algorithms Programming), which was very useful in the exploratory and checking stages of this research.

This article was completed while Grant visited Aarhus University. He thanks The Danish National Research Foundation (grant no. DNRF156) and The Independent Research Fund Denmark (grant no. 1026-00050B) for their support, and thanks Peter Jørgensen and the Aarhus Homological Algebra Group for their hospitality.

Morigi was supported by a research grant (VIL42076) from VILLUM FONDEN.

References

[1] M. Barot, C. Geiss, and A. Zelevinsky, "Cluster algebras of finite type and positive symmetrizable matrices", *J. London Math. Soc.* (2) **73**:3 (2006), 545–564. MR

[2] M. Barot, D. Kussin, and H. Lenzing, "The Lie algebra associated to a unit form", *J. Algebra* **296**:1 (2006), 1–17. MR

[3] M. Barot and B. R. Marsh, "Reflection group presentations arising from cluster algebras", *Trans. Amer. Math. Soc.* **367**:3 (2015), 1945–1967. MR

[4] M. Barot and D. Rivera, "Generalized Serre relations for Lie algebras associated with positive unit forms", *J. Pure Appl. Algebra* **211**:2 (2007), 360–373. MR

[5] A. B. Buan and D. F. Vatne, "Derived equivalence classification for cluster-tilted algebras of type A_n", *J. Algebra* **319**:7 (2008), 2723–2738. MR

[6] V. Dlab and C. M. Ringel, *Indecomposable representations of graphs and algebras*, Mem. Amer. Math. Soc. **173**, Amer. Math. Soc., 1976. MR

[7] S. Fomin and A. Zelevinsky, "Cluster algebras, I: Foundations", *J. Amer. Math. Soc.* **15**:2 (2002), 497–529. MR

[8] S. Fomin and A. Zelevinsky, "Cluster algebras, II: Finite type classification", *Invent. Math.* **154**:1 (2003), 63–121. MR

[9] J. Grant and B. R. Marsh, "Braid groups and quiver mutation", *Pacific J. Math.* **290**:1 (2017), 77–116. MR

[10] J. E. Humphreys, *Introduction to Lie algebras and representation theory*, Graduate Texts in Mathematics **9**, Springer, 1978. MR

[11] P. Lampe, "Cluster algebras", lecture notes, 2013, available at https://www.math.uni-bielefeld.de/~lampe/teaching/cluster/cluster.pdf.

[12] D. Morigi, "Presentations of braid groups of type A arising from $(m+2)$-angulations of regular polygons", *Algebr. Represent. Theory* **27**:2 (2024), 1237–1265. MR

[13] M. J. Parsons, "Companion bases for cluster-tilted algebras", *Algebr. Represent. Theory* **17**:3 (2014), 775–808. MR

[14] C. Pérez and D. Rivera, "Serre type relations for complex semisimple Lie algebras associated to positive definite quasi-Cartan matrices", *Linear Algebra Appl.* **567** (2019), 14–44. MR

[15] J.-P. Serre, *Complex semisimple Lie algebras*, Springer, 1987. MR

Received July 18, 2024. Revised April 29, 2025.

JOSEPH GRANT
joseph.grant@gmail.com

DAVIDE MORIGI
davide.morigi14@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.293

AN AXIOMATIC APPROACH TO SYMMETRIC EXTENSIONS

PETER HOLY

We provide a collection of natural axioms centered around the symmetric forcing theorem, which yield the concept of symmetric extensions, avoiding the technicalities involved in standard presentations.

1. Introduction

Symmetric extensions are an important concept in set theory, originating from Paul Cohen's famous proof of the independence of the axiom of choice from ZF, and they have since proven to be the key tool to obtain independence and consistency results over ZF, in the absence of the axiom of choice. In their usual presentation, they are based on technicalities like the concepts of genericity, forcing names and their evaluations, and on the recursively defined forcing predicates, the definition of which is particularly intricate for the basic case of atomic first order formulas.

In [1], Rodrigo Freire provided an axiomatic framework for set forcing over models of ZFC that is a collection of guiding principles for extensions over which one still has *control* from the ground model. Freire showed that his axioms necessarily lead to the usual concepts of genericity and of forcing extensions, and that one can infer from them the usual recursive definition of the forcing predicates. In [2], this was extended to class forcing by Freire and the author. Building on some of the basic ideas of Freire, we introduce here an axiomatic framework for symmetric extensions over models of ZF, that also avoids the technicalities connected with any usual standard setup for symmetric extensions. In particular, the concepts of genericity and, perhaps most importantly, the recursively defined forcing predicates are avoided. Instead, we provide a natural collection of axioms centered around the symmetric forcing theorem. That is, the conjunction of the definability of the symmetric forcing relations and the truth lemma, stating that anything that holds true in a symmetric extension is forced by a condition in the relevant (symmetrically) generic filter. We show that this collection of axioms essentially induces the common standard concepts: namely, we derive the relevant concept of genericity, the usual

The author thanks the referees for their many very useful comments, suggestions and corrections, which significantly helped to improve this paper.
MSC2020: 03E40, 03E30, 03E25, 03A05.
Keywords: forcing, symmetric extension.

© 2026 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

recursive definitions of forcing predicates, an analogue of the structure of names for elements of symmetric extensions and their evaluations, and the preservation of the axioms of ZF to symmetric extensions.

The aim of this paper is twofold. First, it provides a new viewpoint on a central technical tool in modern set theory: within a suitable basic setup, requiring the symmetric forcing theorem is sufficient to yield exactly the concept of symmetric extensions. Second, it provides a self-contained way of introducing symmetric extensions axiomatically. The only point in the paper where it is strictly necessary to refer to some sort of standard setup is when we briefly argue that our axioms are actually consistent, in Section 5.

In this introductory section, we provide a rough description of our axiomatic framework, which will be followed with formal definitions in Sections 2 and 3.

In the standard setup for symmetric extensions, they are based on so-called *symmetric systems*, that is, triples $\mathbb{S} = (\mathbb{P}, \mathcal{G}, \mathcal{F})$ where $\mathbb{P}$ is a forcing notion (i.e., a preorder, meaning a reflexive and transitive binary relation), $\mathcal{G}$ is a group of automorphisms of $\mathbb{P}$, and $\mathcal{F}$ is a normal filter on the set of subgroups of $\mathcal{G}$. Models of set theory have a large variety of symmetric systems, and these symmetric systems usually give rise to a vast number of different symmetric extensions. Symmetric systems themselves may already seem like a fairly technical notion, but in order to capture the magnitude of complex possibilities offered by the technique of symmetric extensions, it seems necessary for our basic setup to contain these notions. Thus, just like usual (class) forcing notions were the basis of the axiomatic description of (class) forcing in [1] (and [2]), we will make the usual notion of symmetric system the basis of our symmetric extensions.

Let us fix a transitive ground model $\mathcal{M} \in V$ for this discussion, and a symmetric system $\mathbb{S} = (\mathbb{P}, \mathcal{G}, \mathcal{F}) \in M$.[1] For the sake of simplicity, we require that $\mathcal{M} \models \mathrm{ZF}$.[2] We think of conditions (that is, elements of the domain P) of $\mathbb{P}$ as having partial information on properties of our extensions. If $q \leq p$ in $\mathbb{P}$, we say that q is stronger than p, and we think of stronger conditions as having more information. The automorphisms $\pi \in \mathcal{G}$ will naturally extend to maps $\Omega(\pi)$ on M, and we consider $x \in M$ to be *symmetric* in case it is mapped to itself by a large number of maps $\Omega(\pi)$, namely whenever π comes from a certain set in $\mathcal{F}$. Any particular *symmetric extension* is built out of such symmetric elements,[3] serving as *names* for the elements of the symmetric extension, together with a choice of filter on $\mathbb{P}$. We think of such a filter as a selection of conditions which have *correct* information about our symmetric extension, and we will refer to such conditions as being *correct*.

[1] As is common practice, we will use M for the domain of $\mathcal{M}$, P for the domain of $\mathbb{P}$, etc.

[2] The usual ways of avoiding this extra consistency assumption apply; see for example [5].

[3] In fact, we only use *hereditarily symmetric* elements later on.

The motivation for using a *filter* of conditions can be explained exactly as in [2]:

- If we consider the information that a condition q has to be correct, then any weaker condition p has less information than q, and this information should therefore also be correct. This corresponds to the upwards closure property of filters.
- If p and q are correct conditions, we consider the information that is jointly collected by p and q to be correct. We require that there is a condition that collects this joint information and that we consider to be correct. This corresponds to the property of a filter that any two of its elements are compatible, as witnessed by yet another element of the filter.

We require that for any condition $p \in P$, there exists a filter G of correct conditions of which p is an element, so that no condition is a priori incorrect. A number of natural axioms will make sure that we have *ground model control* over our symmetric extension, which we denote as $\mathcal{M}_{\mathbb{S}}[G]$, in a sufficiently simple way. We require the existence of a definable relation on our ground model which, following [1], we call the $\mathbb{P}$-*membership relation*. It is supposed to relate to partial knowledge about the membership relation in symmetric extensions. If $a, b \in M$ and $p \in P$, we say that a is an element of b according to p, and write $a \in_p b$ in case the triple (p, a, b) stands in this relation.[4] If $X \subseteq P$, and $a, b \in M$, we write $a \in_X b$ to abbreviate the statement $\exists p \in X\ a \in_p b$. We want to use $\in_G$ as the membership relation of $\mathcal{M}_{\mathbb{S}}[G]$. In order to be able to obtain a transitive model $\mathcal{M}_{\mathbb{S}}[G]$, we require the relation $\in_P$ to be well-founded. The relation $\in_G$ will usually not be extensional, but we nevertheless obtain a transitive $\in$-structure $\mathcal{M}_{\mathbb{S}}[G]$ as the image of the homomorphism that is our *evaluation map* F_G, recursively defined by setting $F_G(b) = \{F_G(a) \mid a \in_G b\}$ for every symmetric b in M. In order to be able to show that $\mathcal{M}_{\mathbb{S}}[G]$ is well-defined and satisfies the axioms of ZF, we will need to require the following:

- Set-likeness of the $\mathbb{P}$-membership relation: for any symmetric $b \in M$, $\{a \mid a \in_P b\}$ is a set in M.
- *High degrees of freedom* for the $\mathbb{P}$-membership relation: for any symmetric relation S on $M \times P$ in M, we find $b \in M$ for which $\{(a, p) \mid a \in_p b\} = S$.

Furthermore, we also require the existence of forcing predicates definably over $\mathcal{M}$, individually for each first order formula. We do not require any particular defining instances for these predicates, we only require them to be connected to truth in

[4]This relation corresponds to the relation that $(a, p) \in b$ in the standard setup, given that a, b are usual (hereditarily symmetric) forcing names. In this sense, the symmetric ground model objects that we will make use of as names can immediately be seen to be very similar to the usual (hereditarily) symmetric names for elements of symmetric extensions.

generic extensions by the following two axioms (these requirements correspond to what is usually known as the *forcing theorem* in any standard setup):

- Whatever holds in $\mathcal{M}_{\mathbb{S}}[G]$ is forced by some condition in G.
- Whatever is forced by some condition in G holds true in $\mathcal{M}_{\mathbb{S}}[G]$.

Let us mention two additional observations that this paper will help us make: First, our framework will help us to establish what the right notion of genericity with respect to symmetric systems is, namely that of symmetric rather than full genericity, a notion that was only recently introduced in [4]. Second, we will make the easy observation that one of the axioms in [1] and [2] was in fact unnecessary, as it easily follows from the remaining axioms, namely the requirement that any condition in P forces at least as much as any weaker condition in P does.

2. The basic setup

Let $\mathcal{L}(\in)$ denote the collection of first order formulas in the language with the binary elementhood relation $\in$. We consider equality between sets to abbreviate the statement that they have the same elements. We start by providing the definition of a *symmetric framework*, which will be the basic formal concept in our approach.

Definition 2.1. A *symmetric framework* is a tuple of the form

$$\left(\mathcal{M}, \mathbb{S}, R, \Omega, (\Vdash_\varphi)_{\varphi\in\mathcal{L}(\in)}, \mathfrak{G}\right)$$

with the following properties.

- $\mathcal{M}$ is a transitive set-size model of ZF.
- $\mathbb{S} = (\mathbb{P}, \mathcal{G}, \mathcal{F}) \in M$.[5]
  - $\mathbb{P} = (P, \leq)$ is a preorder with weakest element 1.[6]
  - $\mathcal{G}$ is a group of automorphism of $\mathbb{P}$.
  - $\mathcal{F}$ is a normal filter on the set of subgroups of $\mathcal{G}$.
- The $\mathbb{P}$-*membership relation* R is a definable (over $\mathcal{M}$) relation on $P \times M \times M$. We denote the property $R(p, a, b)$ as $a \in_p b$.[7] We also write

$$b =_R \{(a, p) \mid a \in_p b\}.$$[8]

[5] We call a triple $\mathbb{S}$ with the next three properties a *symmetric system*.

[6] We use preorders rather than (the perhaps more common restriction to) partial orders, dropping the requirement of antisymmetry (this more general context naturally appears for example in the case of (symmetric) iterations of forcing notions).

[7] In the standard forcing setup, this corresponds to the property that $(a, p) \in b$.

[8] We emphasize here that $=_R$ is not a symmetric relation in general, contrary to what one might expect from the use of the $=$ symbol. It is so, however, in the standard forcing setup, where $=_R$ is just $=$, at least when we restrict our attention to subsets of $M \times P$.

- Ω is a map with domain $\mathcal{G}$, and for $\pi \in \mathcal{G}$, $\Omega(\pi)\colon M \to M$ is such that $\{(\pi, x) \mid x \in \Omega(\pi)\}$ is definable (over $\mathcal{M}$).
- $\mathfrak{G}$ is a second order unary predicate on P, i.e., a unary predicate on $\mathcal{P}(P)$, and we require that $\mathfrak{G}(G)$ implies that $G \subseteq P$ is a filter. If $\mathfrak{G}(G)$ holds, we say that G is a *symmetrically generic filter*. Whenever we quantify over G in the following, we tacitly assume that we quantify over G's such that $\mathfrak{G}(G)$ holds.
- For every $\varphi \in \mathcal{L}(\in)$, $\Vdash_\varphi$ is a definable (over $\mathcal{M}$) predicate (which we also call a *forcing relation for* φ) on $P \times M^m$, where m denotes the number of free first order variables of φ.

  If $(q, a_0, \ldots, a_{m-1}) \in \Vdash_\varphi$, we also write $q \Vdash \varphi(a_0, \ldots, a_{m-1})$.

3. The basic axioms

In this section, we present our basic axioms for symmetric frameworks.

(1) ***Existence of generic filters:*** $\forall p \in P\ \exists G\ p \in G$.[9]

(2) ***Well-foundedness:*** The relation $\in_P$ on M is well-founded.

Using axiom (2), we can define a notion of *name rank*, letting, for $a \in M$,

$$\operatorname{rank} a = \sup\{\operatorname{rank}(b) + 1 \mid b \in_P a\}.$$

Our next axiom describes how the maps $\Omega(\pi)$ work.

(3) ***Extension:*** For all $b \in M$, and $\pi \in \mathcal{G}$,

$$\Omega(\pi)(b) =_R \{(\Omega(\pi)(a), \pi(p)) \mid a \in M,\ p \in P,\ a \in_p b\}.$$

Note that the above induces an action of $\mathcal{G}$ on M, that is, if $\pi_0, \pi_1 \in \mathcal{G}$ and $a \in M$, then, as can easily be verified by induction,

$$\Omega(\pi_0)(\Omega(\pi_1)(a)) = \Omega(\pi_0\pi_1)(a).$$

Let the *symmetry group* of $a \in M$ be

$$\operatorname{sym}(a) = \{\pi \in \mathcal{G} \mid \Omega(\pi)(a) = a\}.$$

We say that $a \in M$ is *symmetric* if $\operatorname{sym}(a) \in \mathcal{F}$, and we let N denote the collection of all symmetric objects (from M). We say (inductively, using the axiom of well-foundedness) that $b \in N$ is *hereditarily symmetric* if a is hereditarily symmetric whenever $a \in_P b$. We let HS denote the collection of all hereditarily symmetric objects (from M). Elements of HS will serve as *names* for the elements of our symmetric extensions defined below. Note that since $\mathcal{F}$ is normal, if a is symmetric,

[9] Remember from the penultimate item in Definition 2.1 that we require that $\mathfrak{G}(G)$ holds.

then $\Omega(\pi)(a)$ is symmetric with $\mathrm{sym}(\Omega(\pi)(a)) = \pi\,\mathrm{sym}(a)\,\pi^{-1} \in \mathcal{F}$. This easy fact will often be used without mention.

Assume that G is such that $\mathfrak{G}(G)$ holds. Using axiom (2), the relation $\in_G$ on HS is well-founded, and since $\mathrm{HS} \in V$, we may thus recursively define our *evaluation function* F_G along the relation $\in_G$, letting $F_G(b) = \{F_G(a) \mid a \in_G b\}$ for each $b \in \mathrm{HS}$.[10] Let $\mathcal{M}_{\mathbb{S}}[G]$ denote the $\in$-structure on the transitive set $F_G[\mathrm{HS}]$:[11] That is, let $\mathcal{M}_{\mathbb{S}}[G] = (M_{\mathbb{S}}[G], \in)$, where $M_{\mathbb{S}}[G] = F_G[\mathrm{HS}] = \{F_G(a) \mid a \in \mathrm{HS}\}$. We refer to $\mathcal{M}_{\mathbb{S}}[G]$ as a *symmetric extension* of M.

The next two axioms should be seen as the most crucial ones, and they state that a natural form of the forcing theorem holds, that is based on our forcing relations. Given a finite tuple $\vec{a} = \langle a_i \mid i < n\rangle$ of elements of HS (we simply say $\vec{a} \in \mathrm{HS}$ in this case), let $F_G(\vec{a}) = \langle F_G(a_i) \mid i < n\rangle$. We use analogous notation also for functions other than F_G.

(4) ***Truth lemma:*** For all $\varphi \in \mathcal{L}(\in)$, all $\vec{a} \in \mathrm{HS}$ and all G,

$$\mathcal{M}_{\mathbb{S}}[G] \models \varphi(F_G(\vec{a})) \text{ if and only if } \exists p \in G\ p \Vdash \varphi(\vec{a}).$$

(5) ***Definability lemma:*** For all $\varphi \in \mathcal{L}(\in)$, all $\vec{a} \in \mathrm{HS}$ and $p \in P$,

$$p \Vdash \varphi(\vec{a}) \text{ if and only if } \forall G \ni p\ \ \mathcal{M}_{\mathbb{S}}[G] \models \varphi(F_G(\vec{a})).$$[12]

Our final two axioms make sure that our setup is reasonable: the former assumes that names have set-like properties with respect to the $\mathbb{P}$-membership relation, and the next makes sure that we have a sufficient amount of names available.[13]

(6) ***Set-likeness:*** If $b \in \mathrm{HS}$, then $\{a \mid a \in_P b\} \in M$.

If $S \in M$ is a subset of $\mathrm{HS} \times P$, we say that S is a *symmetric* subset of $\mathrm{HS} \times P$ if

$$\exists F \in \mathcal{F}\, \forall \pi \in F\, \forall (a, p) \in S\ (\Omega(\pi)(a), \pi(p)) \in S.$$

(7) ***Universality:*** There is a map $\Gamma\colon M \to \mathrm{HS}$ that is definable over $\mathcal{M}$, such that if $S \in M$ is a *symmetric* subset of $\mathrm{HS} \times P$, then $\Gamma(S)$ is the unique $T \in \mathrm{HS}$ for which $T =_R S$.

[10] It may seem like we are taking some sort of transitive collapse of the structure $(\mathcal{M}[G], \in_G)$, however note that there is no reason to assume that $\in_G$ is extensional, or that $\in_G$ can be factorized in order to obtain an extensional relation.

[11] For the moment, this notation is ambiguous, for $\mathcal{M}_{\mathbb{S}}[G]$ may not only depend on $\mathcal{M}$, $\mathbb{S}$ and G, but also on the $\mathbb{P}$-membership relation. We will however show at the end of this section that under additional assumptions, $\mathcal{M}_{\mathbb{S}}[G]$ is uniquely determined.

[12] Note that we already required the forcing relations to be predicates of our model in our basic setup, however this axiom connects them with their intended meaning, and it thus seems justified to consider it to be our version of the *definability lemma*.

[13] The uniqueness requirement in axiom (7) below could be avoided, however it is very natural and easily available in any sort of setup for symmetric extensions.

The statement of the following lemma was taken to be an axiom in [1] and [2]. However, it is easily provable (and would also have been easily provable in [1] or [2]) from axiom (5), which has been overlooked in earlier work on the subject.

Lemma 3.1. *For all $\varphi \in \mathcal{L}(\in)$, for all $\vec{a} \in \mathrm{HS}$, and $p, q \in P$, if $p \Vdash \varphi(\vec{a})$ and $q \leq p$, then $q \Vdash \varphi(\vec{a})$.*

Proof. Assume $p \Vdash \varphi(\vec{a})$ and $q \leq p$. Then, axiom (5) implies that

$$\forall G \ni p \;\; \mathcal{M}_{\mathbb{S}}[G] \models \varphi(F_G(\vec{a})).$$

But since any G above is a filter, it contains p whenever it contains q, hence it clearly follows that $\forall G \ni q \;\; \mathcal{M}_{\mathbb{S}}[G] \models \varphi(F_G(\vec{a}))$, which again by axiom (5) is equivalent to $q \Vdash \varphi(\vec{a})$, as desired. □

We close this section with a lemma which in particular shows that $\mathcal{M}_{\mathbb{S}}[G]$ does not depend on the choice of the $\mathbb{P}$-membership relation.

Lemma 3.2. *Assume that we have two symmetric frameworks*

$$\big(\mathcal{M}, \mathbb{S}, R, \Omega, (\Vdash_\varphi)_{\varphi\in\mathcal{L}(\in)}, \mathfrak{G}\big) \quad \textit{and} \quad \big(\mathcal{M}, \mathbb{S}, R', \Omega', (\Vdash'_\varphi)_{\varphi\in\mathcal{L}(\in)}, \mathfrak{G}'\big),$$

which are based on the same model $\mathcal{M}$ and symmetric system $\mathbb{S} = (\mathbb{P}, \mathcal{G}, \mathcal{F})$, and that G is such that both $\mathfrak{G}(G)$ and $\mathfrak{G}'(G)$ hold. We will write $a \in_p b$ and $a \in'_p b$ in case $R(p, a, b)$ or $R'(p, a, b)$ hold. We will use HS′ *to denote the version of* HS, *we use F'_G to denote the version of F_G, and we use Γ' to denote the version of Γ provided by the latter symmetric framework.*

If $a \in \mathrm{HS}$, then there is $b \in \mathrm{HS}'$ such that $F_G(a) = F'_G(b)$.

Proof. Making use of the map Γ', we define a *translation function* $h\colon \mathrm{HS} \to \mathrm{HS}'$ by induction on name rank, and simultaneously show that for any $c \in \mathrm{HS}$ and $\pi \in \mathcal{G}$, $\Omega'(\pi)(h(c)) = h(\Omega(\pi)(c))$. For $c \in \mathrm{HS}$, consider the set $C = \{(d,\, p) \mid d \in_p c\}$, and let $F = \mathrm{sym}(c) \in \mathcal{F}$. Let $C' = \{(h(d),\, p) \mid (d,\, p) \in C\} \subseteq \mathrm{HS}' \times P$. Let $\pi \in F$ and pick $(h(d),\, p) \in C'$. Then, $(\Omega'(\pi)(h(d)), \pi(p)) = (h(\Omega(\pi)(d)), \pi(p)) \in C'$, and thus we may invoke axiom (7), letting

$$h(c) = \Gamma'(C') \in \mathrm{HS}',$$

since $\mathrm{sym}(h(c)) \geq F$ by the above. Note that clearly, $d \in_p c$ if and only if $h(d) \in'_p h(c)$. Now, if $\pi \in \mathcal{G}$, then

$$\begin{aligned}\Omega'(\pi)(h(c)) &=_{R'} \{(\Omega'(\pi)(h(d)), \pi(p)) \mid d \in_p c\}\\ &= \{(h(\Omega(\pi)(d)), \pi(p)) \mid d \in_p c\}.\end{aligned}$$

On the other hand, $h(\Omega(\pi)(c)) =_{R'} \{(h(\Omega(\pi)(d)), \pi(p)) \mid d \in_p c\}$ as well, thus $\Omega'(\pi)(h(c)) = h(\Omega(\pi)(c))$ by the uniqueness requirement in axiom (7).

Concluding the proof of the lemma, we show by induction on name rank that for any $c \in \mathrm{HS}$, $F'_G(h(c)) = F_G(c)$. Let $c \in \mathrm{HS}$. Inductively,

$$F'_G(h(c)) = \{F'_G(h(d)) \mid h(d) \in_G h(c)\} = \{F_G(d) \mid d \in_G c\} = F_G(c),$$

as desired. □

4. Forcing predicates, density, and symmetry

We will use our axioms to verify some of the basic properties of forcing, including that the forcing predicates satisfy their usual defining clauses, by arguments that are partially similar to those in [1, Section 4] or [2, Section 4]. However, we are making strong use of the universality axiom, the analogue of which was only introduced much later in both [1] and [2], already in the proof of Lemma 4.2. We will also show how automorphisms are to be applied to forcing statements. We start by observing that we obtain the usual defining clause for the forcing relation for negated formulae.

Lemma 4.1. *For all $\varphi \in \mathcal{L}(\in)$, $p \in P$ and $\vec{a} \in \mathrm{HS}$, we have*

$$p \Vdash \neg\varphi(\vec{a}) \text{ if and only if } \forall q \leq p \; q \not\Vdash \varphi(\vec{a}).$$

Proof. Let us assume that

(i) $p \Vdash \neg\varphi(\vec{a})$.

By axiom (5), equivalently

(ii) $\forall G \ni p \;\; \mathcal{M}_{\mathbb{S}}[G] \models \neg\varphi(F_G(\vec{a}))$.

By axiom (4), this is equivalent to

(iii) $\forall G \ni p \; \forall q \in G \; q \not\Vdash \varphi(\vec{a})$.

We want to argue that this in turn is equivalent to our desired statement that

(iv) $\forall q \leq p \; q \not\Vdash \varphi(\vec{a})$.

Thus, assume first that (iii) holds, and let $q \leq p$. By axiom (1), we may pick a generic filter $G \ni q$, which will thus also contain p as an element. By (iii), we thus have that $q \not\Vdash \varphi(\vec{a})$, as desired.

Conversely, assume that (iv) holds. Let G be a generic filter that contains p as an element, and assume for a contradiction that there is $r \in G$ such that $r \Vdash \varphi(\vec{a})$. Since G is a filter, we may pick q below both p and r. By Lemma 3.1, it follows that $q \Vdash \varphi(\vec{a})$, contradicting (iv). □

The next lemma provides us with objects that represent ground model elements in our symmetric extensions.

Lemma 4.2 (ground model elements). *There is a definable map* $\check{}: M \to \mathrm{HS}$, $a \mapsto \check{a}$, *such that* $\forall a \in M\ \forall G$

$$F_G(\check{a}) = a \ \wedge\ \mathrm{sym}(\check{a}) = \mathcal{G}.$$

Proof. Using axiom (7), by recursion on von Neumann rank in M, for $b \in M$, let $\check{b} = \Gamma(\{(\check{a}, 1) \mid a \in b\})$. Now, for any $b \in M$, $F_G(\check{b}) = b$ and $\mathrm{sym}(\check{b}) = \mathcal{G}$ is easily shown by induction on the rank of $\check{b}$, using that $\pi(1) = 1$ for the latter. □

If $D \subseteq P$, we let $\mathrm{sym}^*(D) = \{\pi \in \mathcal{G} \mid \pi[D] = D\}$. A subset D of P is *symmetrically dense* if it is dense (i.e., $\forall p \in P\ \exists q \leq p\ q \in D$) and $\mathrm{sym}^*(D) \in \mathcal{F}$. We will show that our axioms imply generic filters to intersect all symmetrically dense subsets of $\mathbb{P}$ in M.[14]

Lemma 4.3. *Let $D \in M$ be such that $D \subseteq P$ is symmetrically dense, and let $G \in \mathfrak{G}$. Then $G \cap D \neq \varnothing$.*

Proof. Let $\dot{D} = \Gamma(\{(\check{\varnothing}, d) \mid d \in D\})$. Clearly, $\dot{D}$ is symmetric with $\mathrm{sym}(\dot{D}) = \mathrm{sym}^*(D)$. Let $p \in \mathbb{P}$ and assume $p \Vdash \dot{D} = \varnothing$. Then $\exists q \leq p\ q \in D$, hence $\check{\varnothing} \in_q \dot{D}$, so $F_G(\dot{D}) \neq \varnothing$ whenever $q \in G$, that is, by axiom (5), $q \Vdash \dot{D} \neq \varnothing$, contradicting Lemma 4.1. Thus, again by Lemma 4.1, $1 \Vdash \dot{D} \neq \varnothing$. It follows that for all G, $F_G(\dot{D}) \neq \varnothing$, and hence $D \cap G \neq \varnothing$. □

We next need another auxiliary result on symmetric open dense sets (which could easily be extended to arbitrary dense sets, but the current version is sufficient for our purposes). We say that a subset A of a preorder P is *open* if it is downward closed, that is if $p \in A$ and $q \leq p$, then also $q \in A$.

Lemma 4.4. *If $D \subseteq P$ is open and symmetric, with $D \in M$, then D is dense below p if and only if*

$$(\dagger)\ \forall G \ni p\ D \cap G \neq \varnothing.$$

Proof. Assume first that (†) holds. Let $r \leq p$, and using axiom (1), let G be a generic filter with $r \in G$. It follows that also $p \in G$, and thus using (†), we obtain $s \in D \cap G$. Since D is open and G is a filter, we obtain q below both r and s that is an element of $D \cap G$, showing that D is dense below p.

On the other hand, assume that D is dense below p, and let G be a generic filter containing p as an element. Let $E = D \cup \{q \in P \mid \forall r \leq q\ r \notin D\}$. Then E is clearly dense, as for any $q \in P$, either some $r \leq q$ is in D or, if not, then $q \in E$. But E is also symmetric. Let π be such that $\pi[D] = D$. If $q \in D$, then $\pi(q) \in D \subseteq E$. If $q \in E \setminus D$, this is because no $r \leq q$ is in D. But then no $r \leq \pi(q)$ is in $\pi[D] = D$;

[14] It may be surprising that we do not obtain that our generic filters intersect *all* dense subsets of $\mathbb{P}$. However, it was already noted in [4] that intersecting all symmetrically dense sets seems to be the right notion of genericity in the context of symmetric extensions. This could be seen to be further supported by Lemma 4.3, the proof of which does not extend to dense subsets of P.

that is, $\pi(q) \in E$. By Lemma 4.3, it follows that $G \cap E \neq \varnothing$. Since $p \in G$ and G is a filter, it thus follows that $G \cap D \neq \varnothing$, as desired. □

It is now possible to show that the usual defining clauses for the forcing relation can be recovered from our basic axioms, as well as how automorphisms from $\mathcal{G}$ can be applied to forcing statements. This is shown by a simultaneous induction on formula complexity, and for atomic formulas, on name rank (or, more precisely, by induction on pairs of name ranks, with the componentwise ordering[15]). For $a, b \in \mathrm{HS}$ and $p \in P$, let $a \mathrel{\overline{\in}}_p b$ abbreviate the statement that $\exists q \geq p\ a \in_q b$.

Theorem 4.5. *For any $p \in P$, $\varphi, \psi \in \mathcal{L}(\in)$, and $a, b, \vec{a} \in \mathrm{HS}$, we have*:

(i) $p \Vdash a \in b$ *if and only if*

$$\forall r \leq p\, \exists s \leq r\, \exists x \in \mathrm{HS}\ [x \mathrel{\overline{\in}}_s b \wedge s \Vdash a = x].$$

(ii) $p \Vdash a \subseteq b$ *if and only if*

$$\forall x \in \mathrm{HS}\ \forall r \in P\ [x \mathrel{\overline{\in}}_r a \rightarrow \forall q \leq p, r\, \exists s \leq q\ s \Vdash x \in b].$$

(iii) $p \Vdash a = b$ *if and only if* $[p \Vdash a \subseteq b \wedge p \Vdash b \subseteq a]$.

(iv) $p \Vdash [\varphi \wedge \psi](\vec{a})$ *if and only if* $p \Vdash \varphi(\vec{a}) \wedge p \Vdash \psi(\vec{a})$.

(v) $p \Vdash [\varphi \vee \psi](\vec{a})$ *if and only if* $\forall r \leq p\, \exists q \leq r\ [q \Vdash \varphi(\vec{a}) \vee q \Vdash \psi(\vec{a})]$.

(vi) $p \Vdash \exists x \varphi(x, \vec{a})$ *if and only if* $\forall r \leq p\, \exists q \leq r\, \exists x \in \mathrm{HS}\ q \Vdash \varphi(x, \vec{a})$.

(vii) $p \Vdash \forall x \varphi(x, \vec{a})$ *if and only if* $\forall x \in \mathrm{HS}\ p \Vdash \varphi(x, \vec{a})$.

(viii) *For any $\pi \in \mathcal{G}$, $p \Vdash \varphi(\vec{a})$ if and only if $\pi(p) \Vdash \varphi(\Omega(\pi)(\vec{a}))$.*

Proof. (i) In order to start the induction, if $a = b = \check{\varnothing}$, note that by axiom (5), no condition forces $a \in b$, and also the statement on the right hand side of our claimed equivalence is always wrong. Let us assume that $p \Vdash a \in b$. By axiom (5), this is equivalent to

$$\forall G \ni p\ F_G(a) \in F_G(b).$$

By the definition of F_G, this in turn is equivalent to

$$\forall G \ni p\, \exists x \in \mathrm{HS}\ [F_G(a) = F_G(x) \wedge x \in_G b].$$

Using axiom (4), we obtain the following equivalent form.

$$\forall G \ni p\, \exists x \in \mathrm{HS}\ [\exists r \in G\ r \Vdash a = x \wedge x \in_G b].$$

Now we make use of Lemma 3.1, equivalently obtaining that

$$\forall G \ni p\, \exists s \in G\, \exists x \in \mathrm{HS}\ [s \Vdash a = x \wedge x \mathrel{\overline{\in}}_s b].$$

[15]That is, $(r_0, s_0) \leq (r_1, s_1)$ if and only if $r_0 \leq r_1$ and $s_0 \leq s_1$, and for the strict ordering, we require that one of these inequalities is strict.

Now note that $D := \{s \in P \mid \exists x \in \mathrm{HS}\ [s \Vdash a = x \wedge x \mathrel{\overline{\in}}_s b]\}$ is open with symmetry group $\mathrm{sym}(a) \cap \mathrm{sym}(b) \in \mathcal{F}$: Inductively, noting that $\mathrm{rank}(x) < \mathrm{rank}(b)$, by (viii), for any $\pi \in \mathrm{sym}(a)$, $s \Vdash a = x$ if and only if $\pi(s) \Vdash \Omega(\pi)(a) = a = \Omega(\pi)(x)$, and for any $\pi \in \mathrm{sym}(b)$, $x \mathrel{\overline{\in}}_s b$ if and only if $\Omega(\pi)(x) \mathrel{\overline{\in}}_{\pi(s)} \Omega(\pi)(b) = b$. Therefore, if $s \in D$ and $\pi \in \mathrm{sym}(a) \cap \mathrm{sym}(b)$, also $\pi(s) \in D$, as witnessed by $\Omega(\pi)(x) \in \mathrm{HS}$. Thus, as desired, Lemma 4.4 equivalently yields:

$$\forall r \leq p\ \exists s \leq r\ \exists x \in \mathrm{HS}\ [x \mathrel{\overline{\in}}_s b \wedge s \Vdash a = x].$$

(ii) In order to start the induction, if $a = b = \check{\varnothing}$, note that by axiom (5), any condition forces $a \subseteq b$, and also the statement on the right hand side of our claimed equivalence is always true. Let us assume that $p \Vdash a \subseteq b$. By axiom (5), this is equivalent to

$$\forall G \ni p\ F_G(a) \subseteq F_G(b).$$

By the definition of F_G and axiom (4), this in turn is equivalent to

$$\forall G \ni p\ \forall x \in \mathrm{HS}\ \forall r \in P\ [(x \mathrel{\overline{\in}}_r a \wedge r \in G) \to \exists s \in G\ s \Vdash x \in b].$$

Since all relevant r will be compatible with p, we may equivalently assume that $r \leq p$, and thus obtain the following equivalent form.

$$\forall x \in \mathrm{HS}\ \forall r \leq p\ \forall G \ni r\ [x \mathrel{\overline{\in}}_r a \to \exists s \in G\ s \Vdash x \in b].$$

Now note that $\{s \in P \mid s \Vdash x \in b\}$ is open with symmetry group $\mathrm{sym}(x) \cap \mathrm{sym}(b) \in \mathcal{F}$: Inductively, noting that $\mathrm{rank}(x) < \mathrm{rank}(a)$, by (viii), we argue analogous to the proof of (i). Thus, Lemma 4.4 equivalently yields:

$$\forall x \in \mathrm{HS}\ \forall r \leq p\ [x \mathrel{\overline{\in}}_r a \to \forall q \leq r\ \exists s \leq q\ s \Vdash x \in b].$$

Finally, it is easy to check that we equivalently obtain our desired statement below.

$$\forall x \in \mathrm{HS}\ \forall r \in P\ [x \mathrel{\overline{\in}}_r a \to \forall q \leq r,\ p\ \exists s \leq q\ s \Vdash x \in b].$$

(iii) and (iv) are very easy, using axioms (5) and (4), and (viii) inductively in the case of (iv), as in the verification of the next item.

(v) Assume that

$$p \Vdash (\varphi \vee \psi)(\vec{a}).$$

By axiom (5), this is equivalent to

$$\forall G \ni p\ \ \mathcal{M}_{\mathbb{S}}[G] \models (\varphi \vee \psi)(F_G(\vec{a})).$$

This in turn is equivalent to

$$\forall G \ni p\ \ \big[\mathcal{M}_{\mathbb{S}}[G] \models \varphi(\vec{a}) \vee \mathcal{M}_{\mathbb{S}}[G] \models \psi(\vec{a})\big].$$

By axiom (4), we obtain the following equivalent form.

$$\forall G \ni p\ [\exists q \in G\ q \Vdash \varphi(\vec{a}) \vee \exists q \in G\ q \Vdash \psi(\vec{a})].$$

Now note that $\{q \in P \mid q \Vdash \varphi(\vec{a}) \vee q \Vdash \psi(\vec{a})\}$ is open with symmetry group $\mathrm{sym}(a) \in \mathcal{F}$, using (viii) inductively for φ and ψ. Thus, Lemma 4.4 equivalently yields our desired equivalent form:

$$\forall q \leq p\, \exists r \leq q\ r \Vdash \varphi(\vec{a}) \vee r \Vdash \psi(\vec{a}).$$

(vi) Assume that $p \Vdash \exists x \varphi(x, \vec{a})$. By axiom (5), this is equivalent to

$$\forall G \ni p\ \mathcal{M}_{\mathbb{S}}[G] \models \exists x\, \varphi(x, F_G(\vec{a})).$$

This in turn is equivalent to

$$\forall G \ni p\, \exists x \in \mathrm{HS}\ \mathcal{M}_{\mathbb{S}}[G] \models \varphi(F_G(x), F_G(\vec{a})).$$

By axiom (4), we equivalently obtain that

$$\forall G \ni p\, \exists x \in \mathrm{HS}\, \exists q \in G\ q \Vdash \varphi(x, \vec{a}).$$

Now note that $\{q \in P \mid \exists x \in \mathrm{HS}\ q \Vdash \varphi(x, \vec{a})\}$ is open and symmetric, using (viii) inductively for φ. Then, as desired, Lemma 4.4 equivalently yields

$$\forall r \leq p\, \exists q \leq r\, \exists x \in \mathrm{HS}\ q \Vdash \varphi(x, \vec{a}).$$

(vii) is again an easy consequence of axioms (5) and (4), inductively using (viii) for φ.

(viii) By (i), $p \Vdash a \in b$ if and only if

$$\forall r \leq p\, \exists s \leq r\, \exists x \in M\ [x \mathrel{\overline{\in}_s} b \wedge s \Vdash a = x].$$

Note that $\mathrm{rank}(x) < \mathrm{rank}(b)$ for any possible witness x in the above. Using axiom (3), and by (viii) inductively, we obtain that

$$\Omega(\pi)(x) \mathrel{\overline{\in}_{\pi(s)}} \Omega(\pi)(b) \wedge \pi(s) \Vdash \Omega(\pi)(a) = \Omega(\pi)(x).$$

Using (i) again, it follows that $\pi(p) \Vdash \Omega(\pi)(a) \in \Omega(\pi)(b)$, as witnessed by $\Omega(\pi)(x)$. The reverse direction follows making use of $\pi^{-1} \in \mathcal{G}$. The argument for $\subseteq$ is analogous, using (ii).

For the case of negations, assume that $p \Vdash \neg\varphi(\vec{a})$. By Lemma 4.1, equivalently, $\forall q \leq p\ q \nVdash \varphi(\vec{a})$. Applying π inductively, and using Lemma 4.1 again, we obtain that

$$\pi(p) \Vdash \neg\varphi(\Omega(\pi)(\vec{a})).$$

The reverse direction is again obtained by simply using π^{-1} in the same way (note that $\Omega(\pi^{-1})\Omega(\pi) = \mathrm{id}_M$). The remaining cases are essentially analogous to the case of negations, using (iv)–(vii) instead of Lemma 4.1, and using (viii) inductively in each case. □

5. ZF in symmetric extensions

In this section, we consider the following statement.

(∗) ***Preservation of axioms:*** $\forall G \ \mathcal{M}_{\mathbb{S}}[G] \models \mathrm{ZF}$.[16]

Let us start with the important remark that our axioms (1)–(7), as well as (∗), hold in the standard setup for symmetric extensions (and are thus, in particular, consistent relative to ZF), as described, for example, in [4]:[17] Given a countable transitive model $\mathcal{M}$ of ZF and a symmetric system $\mathbb{S} \in M$, interpreting $a \in_p b$ as $(a, p) \in b$, letting $\mathfrak{G}(G)$ hold if and only if G is a filter on $\mathbb{P}$ that intersects every symmetrically dense subset of P, and using the standard inductive definitions for the forcing predicates (which are exactly the ones we derived in Section 4), we arrive at a symmetric framework. The easy standard result known as the Rasiowa-Sikorski lemma implies that (1) for every $p \in P$, there is a (fully) $\mathbb{P}$-generic filter over $\mathcal{M}$ that contains p as an element. Axioms (2) and (3) are immediate from our above definitions. Verifying axioms (4) and (5) amounts to the proof of the forcing theorem in the standard setup for symmetric extensions (see for example [4]).[18]Axiom (6) is immediate from our definitions, and axiom (7) follows taking, in the notation of that axiom, $T = S$, by a straightforward calculation using axiom (3). It is well-known [4] how to verify (∗) with respect to $\mathcal{M}$ and $\mathbb{S}$ in this context.

We can however also derive axiom (∗) from axioms (1)–(7). There are two different ways to do so. The first possibility is to make use of Lemma 3.2, showing that $\mathcal{M}_{\mathbb{S}}[G]$ is just the standard symmetric extension of $\mathcal{M}$ by the $\mathbb{S}$-generic filter G, and thus, again by the same standard arguments [4], $\mathcal{M}_{\mathbb{S}}[G] \models \mathrm{ZF}$. The second possibility is to actually verify the axioms of ZF in $\mathcal{M}_{\mathbb{S}}[G]$ using axioms (1)–(7). The advantage of this second option, which we choose in the below, is that the argument is self-contained.

Theorem 5.1. *Axioms* (1)–(7) *imply axiom* (∗).

Proof. Since $\mathcal{M}_{\mathbb{S}}[G]$ is a transitive $\in$-structure, it clearly satisfies Regularity and Extensionality. Using axiom (7), $\mathcal{M}_{\mathbb{S}}[G]$ satisfies Pairing: If $a, b \in \mathrm{HS}$, let $c =_R \{(a, 1), (b, 1)\}$. Then $\mathrm{sym}(c) \geq \mathrm{sym}(a) \cap \mathrm{sym}(b) \in \mathcal{F}$, and $F_G(c) = \{F_G(a), F_G(b)\}$. By Lemma 4.2, $\mathcal{M}_{\mathbb{S}}[G]$ satisfies Infinity.

[16]Due to axiom (5), this statement could equivalently be replaced by a scheme of axioms, consisting of statements of the form $1 \Vdash \varphi$ for every $\varphi \in \mathrm{ZF}$.

[17]Earlier references tend to make use of fully generic rather than just symmetrically generic filters, leading to a somewhat more restricted setting which however simplifies the verification of (∗), as it is possible to make use of the fact that symmetric extensions are submodels of fully generic extensions in this setting (for example in [3]).

[18]Since [4] is not yet published, let us remark here that the arguments needed here are very similar to the arguments for the standard forcing theorem for forcing.

Let us treat the union axiom: Let $a \in \mathrm{HS}$. We need to show that for some $b \in \mathrm{HS}$, $\bigcup F_G(a) \subseteq F_G(b)$. Let $X = \{c \mid c \in_P a\} \in M$ by axiom (6). Let $Y = \{d \mid \exists c \in X\, d \in_P c\} \in M$ by axiom (6). Using axiom (7), let $b =_R \{(d, 1) \mid d \in Y\}$. It is straightforward to check that $F_G(b) \supseteq \bigcup F_G(a)$. It remains to show that $b \in \mathrm{HS}$. Let $\pi \in \mathrm{sym}(a)$. It follows that $\Omega(\pi)[X] = X$, which in turn implies that $\Omega(\pi)[Y] = Y$. It clearly follows that $\Omega(\pi)(b) = b$, and hence that $\mathrm{sym}(b) \supseteq \mathrm{sym}(a) \in \mathcal{F}$.

Next, $\mathcal{M}_{\mathbb{S}}[G]$ satisfies collection: Let $a, t \in \mathrm{HS}$, let φ be a first order formula, and assume $p \Vdash \forall x \in a\, \exists y\, \varphi(x, y, t)$. We will find $b \in \mathrm{HS}$ such that $p \Vdash \forall x \in a\, \exists y \in b\, \varphi(x, y, t)$. Let $X = \{c \mid c \in_P a\} \in M$. Using the axiom of collection in $\mathcal{M}$, let $Y \subseteq \mathrm{HS}$, $Y \in M$, be such that whenever $c \in X$ and $s \in P$ are such that $s \leq p$ and $s \Vdash c \in a$, if there is $y \in \mathrm{HS}$ such that $s \Vdash \varphi(c, y, t)$, then there is $y \in Y$ such that $s \Vdash \varphi(c, y, t)$. Let $Y^* = \{\Omega(\pi)(y) \mid y \in Y \wedge \pi \in \mathcal{G}\} \in M$. Let $b =_R \{(y, 1) \mid y \in Y^*\} \in \mathrm{HS}$. Now if $c \in X$ and $q \leq p$ forces that $c \in a$, by Theorem 4.5, there is $s \leq q$ and $y \in \mathrm{HS}$ such that $s \Vdash \varphi(c, y, t)$. This shows that $p \Vdash \forall x \in a\, \exists y \in b\, \varphi(x, y, t)$, as desired.

Let us show that $\mathcal{M}_{\mathbb{S}}[G]$ satisfies separation: Let $a, t \in \mathrm{HS}$ and let φ be a first order formula. Let $X = \{c \mid c \in_P a\} \in M$. Let

$$b =_R \{(c, p) \mid c \in X \wedge p \Vdash [c \in a \wedge \varphi(c, t)]\}.$$

Clearly, $b \in \mathrm{HS}$ for $\mathrm{sym}(b) \supseteq \mathrm{sym}(a) \cap \mathrm{sym}(t)$, since $\pi[X] = X$ for $\pi \in \mathrm{sym}(a)$, and by Theorem 4.5 (viii). From the definition of b, it easily follows that $F_G(b) = \{x \mid x \in F_G(a) \wedge \mathcal{M}_{\mathbb{S}}[G] \models \varphi(x, F_G(t))\}$.

Finally, we argue that the power set axiom is preserved to $\mathcal{M}_{\mathbb{S}}[G]$: Let $a \in \mathrm{HS}$, and let $X = \{c \mid c \in_P a\} \in M$. For symmetric $d \subseteq \mathrm{HS} \times P$ in M, let $x_d =_R d$. Let

$$b =_R \{(x_d, 1) \mid d \subseteq X \times P \text{ is a symmetric subset of } \mathrm{HS} \times P \text{ in } M\}.$$

If $\pi \in \mathcal{G}$ and $d \subseteq X \times P$ in M, let $\pi^*(d) = \{(\Omega(\pi)(c), \pi(p)) \mid (c, p) \in d\}$. If $\pi \in \mathrm{sym}(a)$ and d is a symmetric subset of $\mathrm{HS} \times P$, then $\Omega(\pi)[X] = X$ and $\Omega(\pi)(x_d) = x_{\pi^*(d)}$, with $\mathrm{sym}^*(\pi^*(d)) \geq \pi\, \mathrm{sym}^*(d)\, \pi^{-1} \in \mathcal{F}$ by the normality of $\mathcal{F}$, and thus it follows that $\Omega(\pi)(b) = b$. Now assume that $e \in \mathrm{HS}$ is such that $F_G(e) \subseteq F_G(a)$ in $\mathcal{M}_{\mathbb{S}}[G]$. Using axiom (4), let $p \in G$ force that $e \subseteq a$. Let

$$d = \{(c, r) \in X \times P \mid \exists q, f\ f \in_q e \wedge r \leq q \wedge r \Vdash f = c\}$$

and let $x = x_d =_R d$. Note that $\mathrm{sym}(x) = \mathrm{sym}(e)$ by Theorem 4.5(viii). Since $F_G(x) \in F_G(b)$ by the definition of b, it suffices to check that $F_G(e) = F_G(x)$. Assume first that $y \in F_G(e)$. Then there is $q \in G$ and $f \in \mathrm{HS}$ such that $f \in_q e$ and $y = F_G(f)$. Hence there is $r \leq p, q$ in G and $c \in X$ such that $r \Vdash f = c$, i.e., $(c, r) \in d$. This however implies that $y = F_G(f) \in F_G(x)$. On the other hand, if $y \in F_G(x)$, then there is $(c, r) \in d$ with $r \in G$ and $y = F_G(c)$. But then there is

$q \in G$ and $f \in_q e$ such that $r \Vdash f = c$. This implies that $y = F_G(c) \in F_G(e)$, as desired. □

References

[1] R. A. Freire, "An axiomatic approach to forcing and generic extensions", *C. R. Math. Acad. Sci. Paris* **358**:6 (2020), 757–775. MR

[2] R. A. Freire and P. Holy, "An axiomatic approach to forcing in a general setting", *Bull. Symb. Log.* **28**:3 (2022), 427–450. MR

[3] T. J. Jech, *The axiom of choice*, Studies in Logic and the Foundations of Mathematics **75**, North-Holland and Elsevier, 1973. Reprinted Dover, 2008. MR

[4] A. Karagila and J. Schilhan, "Towards a theory of symmetric extensions", draft notes, 2024.

[5] K. Kunen, *Set theory*: *an introduction to independence proofs*, Studies in Logic and the Foundations of Mathematics **102**, North-Holland, 1980. MR

Received November 8, 2024. Revised September 23, 2025.

PETER HOLY
INSTITUT FÜR DISKRETE MATHEMATIK UND GEOMETRIE
TU WIEN
1040 WIEN
AUSTRIA
peter.holy@tuwien.ac.at

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.309

ON BRAVERMAN–KAZHDAN–NGÔ PAIRS

DIHUA JIANG, ZHAOLIN LI AND GUODONG XI

The Braverman–Kazhdan program, later refined by Ngô, aims to understand Langlands L-functions attached to a reductive group G and a representation $\rho : {}^L G \to \mathrm{GL}_{V_\rho}$ of its L-group, plus certain additional desiderata. Such pairs (G, ρ) are called Braverman–Kazhdan–Ngô (BKN) pairs. We explain in this paper how it is enough to consider BKN pairs (G, ρ), in order to understand general Langlands L-functions. A key tool in the approach of Braverman and Kazhdan is a certain reductive monoid attached to ρ. There are two methods of constructing such a reductive monoid in the literature. We prove that the two methods yield the same monoid when (G, ρ) is a BKN pair.

1. Introduction 309
2. Braverman–Kazhdan–Ngô pairs 312
3. Borel's conjecture 319
4. L-monoids and BKN-pairs 321
5. Examples 329
Acknowledgement 335
References 335

1. Introduction

Let G be a split reductive algebraic group defined over a number field k, and $G^\vee$ be the dual group over $\mathbb{C}$ with the root datum dual to G. For an irreducible finite-dimensional algebraic representation (ρ, V_ρ) of $G^\vee$ and an irreducible cuspidal automorphic representation π of $G(\mathbb{A})$, where $\mathbb{A}$ is the ring of adeles of k, R. Langlands introduced in [9] the automorphic L-function $L(s, \pi, \rho)$ associated with (π, ρ) and proved that $L(s, \pi, \rho)$ can be defined as an Euler product of local L-factors

$$L(s, \pi, \rho) = \prod_{v \in |k|} L_v(s, \pi_v, \rho)$$

The research in this paper is supported in part by the NSF Grant DMS-2200890 and the Simons Fellow: SFI-MPS-SFM-00005659.

MSC2020: primary 11F66, 22E50; secondary 11F70.

Keywords: reductive monoid, automorphic L-function, Braverman–Kazhdan proposal, Borel conjecture.

© 2026 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

when $s \in \mathbb{C}$ with $\mathrm{Re}(s)$ sufficiently positive, where $|k|$ is the set of all local places of k, $\pi = \otimes_v \pi_v$ (the restricted tensor product decomposition), and the local L-factors $L_v(s, \pi_v, \rho)$ can be defined through the local Langlands conjecture. It is well known that the local L-factors are well-defined when v is archimedean or any finite local place at which π_v is unramified. The Langlands conjecture predicts that $L(s, \pi, \rho)$ admits a meromorphic continuation to $s \in \mathbb{C}$ and satisfies the standard global functional equation. One of the central problems in the theory of automorphic forms and in the Langlands program is to prove the Langlands conjecture for general automorphic L-functions $L(s, \pi, \rho)$.

The Langlands–Shahidi method and the Rankin–Selberg method established the Langlands conjecture for long lists of cases. However, the general situation of the Langlands conjecture is still widely open. Around 2000, A. Braverman and D. Kazhdan [4] proposed a framework to establish the Langlands conjecture for general L-functions $L(s, \pi, \rho)$, when (G, ρ) is a Braverman–Kazhdan–Ngô pair as defined below, via local and global harmonic analysis on G. It generalizes the method of J. Tate's thesis [28] and the work of R. Godement and H. Jacquet [6].

Definition 1.1 (Braverman–Kazhdan–Ngô pair). Let G be a split reductive group over a field of characteristic zero and let $\rho : G^\vee \to \mathrm{GL}_{V_\rho}$ be a representation of its dual group. The pair (G, ρ) is a BKN pair if it satisfies the following conditions.

(1) The kernel of $\rho : G^\vee \to \mathrm{GL}_{d_\rho}$ is trivial, where d_ρ is the dimension of ρ.

(2) There exists a character $\mathfrak{d} : G \to \mathbb{G}_m$, such that the exact sequence

$$1 \to G_0 \to G \xrightarrow{\mathfrak{d}} \mathbb{G}_m \to 1$$

holds with $G_0 = [G, G]$ the derived group of G.

(3) The derived group $G_0 = [G, G]$ is simply connected.

(4) The dual morphism $\mathfrak{d}^\vee : \mathbb{G}_m \to G^\vee$ composed with the given representation ρ is a scalar multiplication of $\mathbb{G}_m$ on the vector space V_ρ, i.e., the identity $\rho(\mathfrak{d}^\vee(a)) = a \cdot \mathrm{I}_{V_\rho}$ holds for any $a \in \mathbb{G}_m(\mathbb{C}) = \mathbb{C}^\times$.

Note that in Definition 1.1, we take any field of characteristic zero as the ground field, which puts the notion in algebraic terms, and keep in mind that the notion of the BKN pairs may have a connection to the relative Langland duality of D. Ben-Zvi, Y. Sakellaridis and A. Venkatesh [5].

The local conjecture in the Braverman–Kazhdan proposal for any local field F of characteristic zero can be found in [13, Section 4] and [11, Section 2.1]. The assumptions for a BKN pair are natural but restrictive. For the Langlands conjecture on automorphic L-function, we may consider the following pairs (G, ρ), without loss of generality.

Definition 1.2 (L-pair). Let G be a reductive group that splits over a number field k and let $\rho : G^\vee \to \mathrm{GL}_{V_\rho}$ be a representation of its dual group. We say (G, ρ) is an *L-pair* if the restriction of ρ to any essentially simple component of the derived group of $G^\vee$ is nontrivial.

Theorem 1.3. *Let F be a local field of characteristic zero, let G be a split reductive F-group, and let (G, ρ) be an L-pair. Then there is a canonical BKN pair $(H_\rho, \rho_{\mathrm{fp}})$ attached to (G, ρ) with the following property: Assume that the local Langlands correspondence is known for $G(F)$ and $H_\rho(F)$. If π is an irreducible admissible representation of $G(F)$, then there is an irreducible admissible representation σ of $H_\rho(F)$ such that the local L-factors and local ϵ-factors*

$$L(s, \pi, \rho) = L(s, \sigma, \rho_{\mathrm{fp}}) \quad \textit{and} \quad \epsilon(s, \pi, \rho, \psi) = \epsilon(s, \sigma, \rho_{\mathrm{fp}}, \psi)$$

hold as meromorphic functions of s for all nontrivial characters $\psi : F \to \mathbb{C}^\times$.

Theorem 1.3 is one of the main results of this paper, which will be proved in Section 2. The BKN-pair assertion is proved in Section 2A (Theorem 2.4) and the assertion on local L-factors is proved in Section 2B. With the help of the Borel conjecture [1; 2; 26], one can make the representation σ more explicitly in terms of the given representation π in Theorem 1.3, which will be discussed in Section 3. Theorem 1.3 suggests that in order to study the Langlands conjecture on general automorphic L-functions via global zeta integral methods, such as the Rankin–Selberg method, the Langlands–Shahidi method, the Miller–Schmid method and the Bravermen–Kazhdan–Ngô method, the harmonic analysis can be taken over the BKN-pairs that have the extra nice structures as in Definition 1.1. This will be further explained by the examples in Section 5, and in our forthcoming work.

As explained in [3; 4; 13], the Braverman–Kazhdan–Ngô approach to the Langlands conjecture for automorphic L-functions and other related geometric problems requires a good understanding of the L-monoids associated with the given group G as its unit group. For a given BKN-pair (G, ρ), one has two different constructions of L-monoids: one is from the Vinberg method [12; 14; 29], which is denoted by $\mathcal{M}_\rho$, and the other is from the general theory of M. Putcha and L. Renner [21], which is denoted by $\mathcal{M}^\rho$. Note that the two methods construct the monoids with G as the unit groups under the different assumptions on the structure of the algebraic groups G. When G is from a BKN-pair, the two methods work and hence we have both $\mathcal{M}_\rho$ and $\mathcal{M}^\rho$. It is important to know whether $\mathcal{M}_\rho \cong \mathcal{M}^\rho$ since both have different geometric features.

Theorem 1.4. *For any given* BKN*-pair (G, ρ), let $\mathcal{M}_\rho$ be the L-monoid as constructed from the Vinberg method and let $\mathcal{M}^\rho$ be the L-monoid as constructed from the Putcha–Renner method. Then the two reductive monoids are isomorphic, i.e., $\mathcal{M}_\rho \cong \mathcal{M}^\rho$.*

The proof of Theorem 1.4 will be given in Section 4C.

Outline. Section 2 is devoted to the construction of the BKN-pair associated with any given L-pair and proves Theorem 1.3. In Section 3, we discuss the Borel conjecture on the local L-packets, which are important for us to understand the known construction of global zeta integrals for families of L-functions. Section 4 is to discuss the relation between the monoid constructed via the Putcha–Renner theory and the monoid constructed through the Vinberg universal monoid for any given BKN-pair. Finally, in Section 5, we present a series of examples of the BKN-pairs.

2. Braverman–Kazhdan–Ngô pairs

Let k be a field of characteristic zero, which can be a number field or a local field of characteristic zero — the latter being the only ones considered in this paper. Let G be a reductive algebraic group that is k-split. Take $G^\vee$ to be the k-split reductive algebraic group with its root datum dual to that of G Let ρ be a finite-dimensional k-rational representation of $G^\vee$. For a given L-pair (G, ρ), in Section 2A we show that a particular pair constructed using a fiber product is a BKN-pair. For convenience, we write $(G, \rho)_{\mathrm{BKN}}$ for a BKN-pair and $(G, \rho)_L$ for an L-pair.

For the given L-pair $(G, \rho)_L$ over k, let T be a maximal k-split torus of G and $T^\vee$ be its dual, which is a maximal k-split torus of $G^\vee$. Let $X^*(T)$ be the k-rational characters of T and $X_*(T)$ be the k-rational cocharacters of T. Then we have that $X^*(T) = X_*(T^\vee)$ and $X^*(T^\vee) = X_*(T)$. Let $\Phi = \Phi(G, T)$ be the set of roots of G with respect to T and $\Phi^\vee = \Phi(G^\vee, T^\vee)$ be the set of roots of $G^\vee$ with respect to $T^\vee$, which is dual to Φ. Let $(X, \Phi; X^\vee, \Phi^\vee)$ be the root data as in [27]. Let (B, T) be a fixed Borel pair of G that determines the subset Φ^+ of positive roots in Φ.

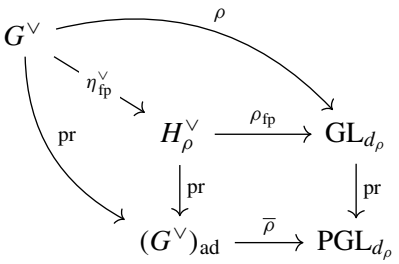
2A. ***BKN-pair via fiber product.*** For a given L-pair $(G, \rho)_L$ as in Definition 1.2, in order to prove Theorem 1.3, we are going to construct a BKN-pair $(G_{\mathfrak{d}}, \rho^{\mathfrak{d}})_{\mathrm{BKN}}$ via a fiber product associated to $(G, \rho)_L$. The construction is purely algebraic and we take any field k of characteristic zero as the ground field.

We first construct the dual group $(G_{\mathfrak{d}})^\vee$ by considering the fiber product diagram

$$\begin{array}{ccc} G^\vee & \xrightarrow{\ \rho\ } & \mathrm{GL}_{d_\rho} \\ \big\downarrow{\scriptstyle \mathrm{pr}} & & \big\downarrow{\scriptstyle \mathrm{pr}} \\ (G^\vee)_{\mathrm{ad}} & \xrightarrow{\ \overline{\rho}\ } & \mathrm{PGL}_{d_\rho} \end{array} \tag{2-1}$$

where d_ρ is the dimension of the representation ρ and $\overline{\rho}$ is the canonical morphism induced from ρ via the canonical projection. Let $H_\rho^\vee$ be the fiber product of GL_{d_ρ} and $(G^\vee)_{\mathrm{ad}}$ over PGL_{d_ρ} defined by the diagram (2-1), which is the universal object

fits into the commutative diagram

(2-2)


The morphisms

$$\eta_{\mathrm{fp}}^\vee : G^\vee \to H_\rho^\vee, \quad \rho_{\mathrm{fp}} : H_\rho^\vee \to \mathrm{GL}_{d_\rho}, \quad \text{and} \quad \mathrm{pr} : H_\rho^\vee \to (G^\vee)_{\mathrm{ad}}$$

are given by the universality of the fiber product, so that diagram (2-2) commutes. The fiber product fits into the following commutative diagram with exact rows:

(2-3)
$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathbb{G}_m & \xrightarrow{\mathfrak{d}_{\mathrm{fp}}^\vee} & H_\rho^\vee & \xrightarrow{\mathrm{pr}} & (G^\vee)_{\mathrm{ad}} & \longrightarrow & 1 \\ & & \downarrow{=} & & \downarrow{\rho_{\mathrm{fp}}} & & \downarrow{\overline{\rho}} & & \\ 1 & \longrightarrow & \mathbb{G}_m & \longrightarrow & \mathrm{GL}_{d_\rho} & \xrightarrow{\mathrm{pr}} & \mathrm{PGL}_{d_\rho} & \longrightarrow & 1 \end{array}$$

which implies that the composition $\rho_{\mathrm{fp}} \circ \mathfrak{d}_{\mathrm{fp}}^\vee(z) = z\mathrm{I}_{d_\rho}$ for any $z \in \mathbb{G}_m$, which is Condition (4) of Definition 1.1. It is easy to see that $H_\rho^\vee$ is a k-split reductive algebraic group over k. We are going to prove that $H_\rho^\vee$ is a candidate for $(G_{\mathfrak{d}})^\vee$ and its dual $(H_\rho^\vee)^\vee$ is the candidate for the group $G_{\mathfrak{d}}$ for the BKN pair $(G_{\mathfrak{d}}, \rho^{\mathfrak{d}})$.

To make the notion clearer, we take H_ρ to be the k-split reductive algebraic group whose root datum is dual to that of $H_\rho^\vee$. In this way, we have that $H_\rho^\vee = (H_\rho)^\vee$. It is clear that H_ρ is equipped with a morphism

$$\mathfrak{d}_{\mathrm{fp}} : H_\rho \to \mathbb{G}_m, \tag{2-4}$$

which is dual to $\mathfrak{d}_{\mathrm{fp}}^\vee : \mathbb{G}_m \to H_\rho^\vee$ as in (2-3). Thus, we construct a pair which is denoted by

$$(H_\rho, \rho_{\mathrm{fp}}). \tag{2-5}$$

In order to show that the constructed pair $(H_\rho, \rho_{\mathrm{fp}})$ is a BKN-pair, it remains to show that the exact sequence

$$1 \to H_\rho^\circ \to H_\rho \xrightarrow{\mathfrak{d}_{\mathrm{fp}}} \mathbb{G}_m \to 1 \tag{2-6}$$

holds, with the kernel H_ρ° of the morphism $\mathfrak{d}_{\mathrm{fp}}$ equal to the derived group of H_ρ, i.e., $H_\rho^\circ = [H_\rho, H_\rho]$, the kernel $\ker(\rho_{\mathrm{fp}})$ being trivial, and H_ρ° simply connected according to Definition 1.1.

Proposition 2.1. *With notation as given by the fiber product diagram in* (2-2)*, the kernel of the morphism* $\rho_{\mathrm{fp}} : H_\rho^\vee \to \mathrm{GL}_{d_\rho}$ *is trivial.*

Proof. Let $\lambda \in X^*(T^\vee)$ be the highest weight of ρ. Let $Z^\vee$ be the center of $G^\vee$. Then $\lambda : Z^\vee \to \mathbb{G}_m$ is a character of $Z^\vee$. We claim that

$$H_\rho^\vee \cong (\mathbb{G}_m \times G^\vee)/\{(\lambda(z)^{-1}, z) \mid z \in Z^\vee\} =: \mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee. \tag{2-7}$$

In order to prove the isomorphism in (2-7), it suffices to show that the group on the right-hand side $\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$ is also a universal object for the fiber product diagram in (2-1). First of all, we extend the representation ρ of $G^\vee$ to that of the group $\mathbb{G}_m \times G^\vee$ by defining the scalar multiplication of $\mathbb{G}_m$:

$$\rho^{\mathfrak{e}}(a, g) = a \cdot \rho(g) \quad \text{for } a \in \mathbb{G}_m \text{ and } g \in G^\vee. \tag{2-8}$$

When $(a, g) = (\lambda^{-1}(z), z) \in \mathbb{G}_m \times G^\vee$ for $z \in Z^\vee$, we have

$$\rho^{\mathfrak{e}}((\lambda^{-1}(z), z)) = \lambda^{-1}(z) \cdot \rho(z) = 1.$$

The extended representation $\rho^{\mathfrak{e}}$ factors through the quotient group $\mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee$:

$$\mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee \xrightarrow{\rho^{\mathfrak{e}}} \mathrm{GL}_{d_\rho} : [(a, g)] \mapsto a \cdot \rho(g). \tag{2-9}$$

On the other hand, we have the natural projection

$$\mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee \xrightarrow{\mathrm{pr}} (G^\vee)_{\mathrm{ad}} : [(a, g)] \mapsto [g]. \tag{2-10}$$

It is clear that the above construction fits into the fiber product diagram as in (2-1):

$$\begin{array}{ccc} \mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee & \xrightarrow{\rho^{\mathfrak{e}}} & \mathrm{GL}_{d_\rho} \\ \downarrow{\scriptstyle \mathrm{pr}} & & \downarrow{\scriptstyle \mathrm{pr}} \\ (G^\vee)_{\mathrm{ad}} & \xrightarrow{\bar{\rho}} & \mathrm{PGL}_{d_\rho} \end{array} \tag{2-11}$$

It remains to show the universality of the group $\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$ with respect to the fiber product diagram in (2-11).

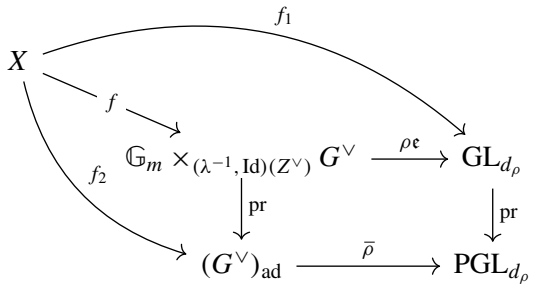
Let X be any variety fitting into the commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{f_1} & \mathrm{GL}_{d_\rho} \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle \mathrm{pr}} \\ (G^\vee)_{\mathrm{ad}} & \xrightarrow{\bar{\rho}} & \mathrm{PGL}_{d_\rho} \end{array} \tag{2-12}$$

We have to show that there exists a unique morphism $f : X \to \mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$

such that the following diagram commutes:

(2-13)



We first construct the map f and then prove that it is an algebraic morphism.

We first consider the morphism $f_2 : X \to (G^\vee)_{\mathrm{ad}}$ and the short exact sequence

$$1 \to Z^\vee \to G^\vee \xrightarrow{\mathrm{pr}} (G^\vee)_{\mathrm{ad}} \to 1.$$

For any $x \in X$, we take a local section $\mathfrak{s}_x$ of the projection pr at $f_2(x)$, namely, take an open affine neighborhood U_x of $f_2(x)$ together with a morphism $\mathfrak{s}_x : U_x \to V_x \subset G^\vee$, where V_x is some open affine subvariety of $G^\vee$, such that $(\mathrm{pr} \circ \mathfrak{s}_x)(y) = y$ for any $y \in U_x$. Let W_x be an open affine neighborhood of x in X such that $f_2(W_x) \subset U_x$ since f_2 is a morphism.

Write $g_x = \mathfrak{s}_x(f_2(x)) \in G^\vee$. According to the commutative diagram in (2-13), there is some $a_x \in \mathbb{G}_m$ such that

$$f_1(x) = a_x \rho(g_x) = \rho^{\mathfrak{e}}(a_x, g_x), \quad \text{with } (a_x, g_x) \in \mathbb{G}_m \times G^\vee. \tag{2-14}$$

Since the local section is not unique, if we take another local section $\mathfrak{s}'_x$ at $f_2(x)$ with similar open affine neighborhoods, then we have that $g'_x = \mathfrak{s}'_x(x)$ and

$$f_1(x) = a'_x \rho(g'_x) = \rho^{\mathfrak{e}}(a'_x, g'_x), \quad \text{with } (a'_x, g'_x) \in \mathbb{G}_m \times G^\vee.$$

Hence

$$f_1(x) = a_x \rho(g_x) = a'_x \rho(g'_x),$$

which can be rewritten as

$$\rho(g_x (g'_x)^{-1}) = a_x^{-1} a'_x.$$

Since (G, ρ) is an L-pair as in Definition 1.2, and $\mathrm{pr} \circ \rho = \bar{\rho} \circ \mathrm{pr}$ by (2-1), there is some $z \in Z^\vee$ such that $g_x (g'_x)^{-1} = z$, which implies that

$$(a_x, g_x) = (a'_x, g'_x)(\lambda(z)^{-1}, z).$$

Hence the image of $(a_x, g_x) \in \mathbb{G}_m \times G^\vee$ in the quotient group $\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$, which is denoted by $[a_x, g_x]$, does not depend on the choice of the local sections

and is uniquely determined by the given $x \in X$. Therefore, we construct a map

$$x \mapsto f(x) = [a_x, g_x] \tag{2-15}$$

from X to the quotient group $\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$. It is clear from the construction that the map f fits into the commutative diagram (2-13). From the construction of the map f, we have

$$g_x = \mathfrak{s}_x(f_2(x)) \quad \text{and} \quad a_x = f_1(x) \cdot \rho(g_x)^{-1} = f_1(x) \cdot \rho(\mathfrak{s}_x(x))^{-1}$$

according to (2-14). Hence the map f is locally explicitly given by

$$x \mapsto f(x) = [a_x, g_x] = [f_1(x) \cdot \rho(\mathfrak{s}_x(x))^{-1}, \mathfrak{s}_x(f_2(x))]$$

which is clearly algebraic and thus, a morphism.

It remains to prove the uniqueness of the morphism $f : X \to \mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$ that fits into the commutative diagram in (2-13). Suppose that

$$X \xrightarrow{f'} \mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee : x \mapsto [b_x, h_x]$$

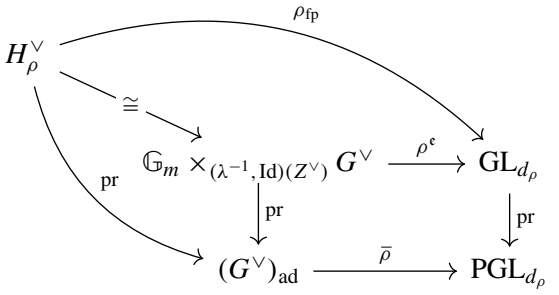
is another morphism fitting into the commutative diagram in (2-13), where $b_x \in \mathbb{G}_m$ and $h_x \in G^\vee$. For any $x \in X$, since $\mathrm{pr}(h_x) = [h_x] = f_2(x) \in (G^\vee)_{\mathrm{ad}}$, $h_x \in G^\vee$ is a pre-image of $f_2(x)$ (with a choice of a section). With h_x chosen, b_x must satisfy the equation

$$b_x \rho(h_x) = \rho^{\mathfrak{e}}(b_x, h_x) = f_1(x)$$

according to the commutative diagram in (2-13). From the discussion above, the morphism f is independent of the choice of the local sections. Hence we must have that $f = f'$, which proves the uniqueness of the morphism f.

Finally, it is easy to verify that the kernel of the morphism $\rho_{\mathrm{fp}} : H_\rho^\vee \to \mathrm{GL}_{d_\rho}$ is trivial from the commutative diagram

(2-16)



In fact, if $h \in H_\rho^\vee$ belongs to the kernel $\ker(\rho_{\mathrm{fp}})$, then $f(h) = [a, g]$ belongs to the $\ker(\rho^{\mathfrak{e}})$. This means that $1 = \rho^{\mathfrak{e}}(a, g) = a\rho(g)$ and $\rho(g) = a^{-1} \in \mathbb{G}_m$. Since (G, ρ) is an L-pair, we must obtain that $g = z \in Z^\vee$, which implies that $a = \rho(z)^{-1} = \lambda(z)^{-1}$ and $[a, g] = [\lambda(z)^{-1}, z]$ must be the identity element of

$\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$. Therefore, the given element h must be the identity element in $H_\rho^\vee$. We are done with $(H_\rho, \rho_{\mathrm{fp}}) = (G_{\mathfrak{d}}, \rho^{\mathfrak{d}})$ being the candidate for the canonical BKN pair associated with the given L-pair $(G, \rho)_L$. □

The proof of Proposition 2.1 yields the following structure of the group $H_\rho^\vee$.

Corollary 2.2. *The isomorphism $H_\rho^\vee \overset{\varsigma}{\cong} \mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} G^\vee$ as in the commutative diagram in* (2-16) *is explicitly given by*

$$h \in H_\rho^\vee \mapsto \varsigma(h) = [\rho_{\mathrm{fp}}(h) \cdot \rho(\mathfrak{s}_h(h))^{-1}, \mathfrak{s}_h(\mathrm{pr}(h))] \in \mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee,$$

which is independent of the choice of the local section $\mathfrak{s}_h$ near $\mathrm{pr}(h)$ *from* $(G^\vee)_{\mathrm{ad}}$ *to* $G^\vee$.

Proposition 2.3. (1) *The group H_ρ° defined via* (2-6) *is the derived group of H_ρ, i.e., $H_\rho^\circ = [H_\rho, H_\rho]$.*

(2) *The group H_ρ, which is dual to $H_\rho^\vee$ as in* (2-2)*, has the following properties*:

(i) *$H_\rho \cong \mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}}$, whose multiplication is given as in* (2-17).

(ii) *Its derived group $H_\rho^\circ = [H_\rho, H_\rho] \cong G_0^{\mathrm{sc}}$ is simply connected.*

Proof. Under the isomorphism of Corollary 2.2, we may identify $\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z^\vee)} T^\vee$ with a maximal torus of $H_\rho^\vee$. One has an isomorphism

$$\mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} T^\vee \cong \mathbb{G}_m \times (T^\vee)_{\mathrm{ad}} : [a, t] \mapsto (\lambda(t)a, [t]).$$

Let $(X, \Phi, X^\vee, \Phi^\vee)$ be the root datum of G and denote the root lattice, the coroot lattice, and the weight lattice by Q, $Q^\vee$ and P respectively. We may denote by $\alpha \mapsto \bar{\alpha}$ the map $X \to P$, which is dual to $Q^\vee \hookrightarrow X^\vee$, and identify a maximal split torus of $H_\rho^\vee$ with $\mathbb{G}_m \times (T^\vee)_{\mathrm{ad}}$. Then the root datum of $H_\rho^\vee$ is given by

$$(\mathbb{Z} \oplus Q^\vee, \{0\} \oplus \Phi^\vee, \mathbb{Z} \oplus P, (\{\langle \alpha, \lambda \rangle, \alpha) \mid \alpha \in \Phi\}).$$

Associated with the natural morphism $G^\vee \to H_\rho^\vee$, one has the map between the root data given by

$$\mathbb{Z} \oplus Q^\vee \to X^\vee : (n, \alpha^\vee) \mapsto n\lambda + \alpha^\vee,$$

with its dual

$$X \to \mathbb{Z} \oplus P : \alpha \mapsto (\langle \alpha, \lambda \rangle, \bar{\alpha}).$$

On the other hand, let G_0^{sc} be the simply connected cover of G_0. We consider the semi-direct product $\mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}}$ with the multiplication given by

(2-17)
$$(a_1, g_1)(a_2, g_2) = (a_1 a_2, \lambda(a_2)^{-1} g_1 \lambda(a_2) g_2) \quad \text{for } a_1, a_2 \in \mathbb{G}_m, g_1, g_2 \in G_0^{\mathrm{sc}}.$$

Here the highest weight $\lambda \in X^*(T^\vee)$ is used as a cocharacter $\lambda \in X_*(T)$ of T. We may use the same λ to denote its image in $X_*(T_{\mathrm{ad}})$ via the projection from T to T_{ad}.

Since G and G_0^{sc} share the same adjoint group, we have $T_0^{\mathrm{sc}}/Z_0^{\mathrm{sc}} = T_{\mathrm{ad}}$, where T_0^{sc} is the maximal torus of G_0^{sc} and Z_0^{sc} is the center of G_0^{sc}. Hence the formula for the multiplication in (2-17) is well-defined. Take the maximal split torus $\mathbb{G}_m \ltimes_\lambda T_0^{\mathrm{sc}}$ of $\mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}}$. Since $\lambda(\mathbb{G}_m)$ commutes with T_0^{sc}, we have $\mathbb{G}_m \ltimes_\lambda T_0^{\mathrm{sc}} = \mathbb{G}_m \times T_0^{\mathrm{sc}}$. Thus, the root datum of $\mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}}$ with respect to $\mathbb{G}_m \times T_0^{\mathrm{sc}}$ is

$$\big(\mathbb{Z} \oplus P, \{(\langle \alpha, \lambda\rangle, \alpha) \mid \alpha \in \Phi\}, \mathbb{Z} \oplus Q^\vee, \{0\} \oplus \Phi^\vee\big).$$

Denote the covering $G_0^{\mathrm{sc}} \to G_0$ by $g \mapsto \bar{g}$, and consider the homomorphism

$$\mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}} \to G : (a, g) \mapsto \lambda(a)\bar{g}. \tag{2-18}$$

Then the associated map on the root data is given by

$$X \to \mathbb{Z} \oplus P : \alpha \mapsto (\langle \alpha, \lambda\rangle, \bar{\alpha}),$$

with its dual

$$\mathbb{Z} \oplus Q^\vee \to X^\vee : (n, \alpha^\vee) \mapsto n\lambda + \alpha^\vee.$$

This proves that $H_\rho \cong \mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}}$. Hence the derived subgroup $[H_\rho, H_\rho]$ is isomorphic to the derived subgroup of $\mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}}$, which is clearly $\{1\} \times G_0^{\mathrm{sc}}$.

Finally, from the above discussion, the projection

$$\mathbb{G}_m \underset{(\lambda^{-1}, \mathrm{Id})(Z^\vee)}{\times} G^\vee = \mathbb{G}_m \times G^\vee / \{(\lambda(z)^{-1}, z) \mid z \in Z^\vee\} \to (G^\vee)_{\mathrm{ad}} : [a, g] \mapsto [g]$$

is the dual of the embedding

$$G_0^{\mathrm{sc}} \to \mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}} : g \mapsto (1, g),$$

which implies the kernel H_ρ° is also the subgroup $\{1\} \times G_0^{\mathrm{sc}}$. Therefore we prove that $H_\rho^\circ \cong [H_\rho, H_\rho]$ is simply connected. □

By the above discussion with Propositions 2.1 and 2.3, we obtain the following result over any field k of characteristic zero, which implies the first part of Theorem 1.3.

Theorem 2.4. *Let k be a field of characteristic zero. For any given L-pair $(G, \rho)_L$, the pair $(H_\rho, \rho_{\mathrm{fp}})$ of* (2-5) *is the canonical BKN-pair associated with $(G, \rho)_L$.*

2B. ***Local factors.*** Toward the proof of Theorem 1.3,we consider local L-factors and local ϵ-factors in the sense of Langlands over a local field F of characteristic zero.

We are going to prove a relation of the local L-factors and of the local ϵ-factors, whose definitions are given under the assumption that the local Langlands conjecture holds for the relevant group $G(F)$. More precisely, they are defined by

$$L(s, \pi, \rho) := L(s, \rho \circ \phi_\pi) \quad \text{and} \quad \epsilon(s, \pi, \rho, \psi) := \epsilon(s, \rho \circ \phi_\pi, \psi) \tag{2-19}$$

for any nontrivial character $\psi : F \to \mathbb{C}^\times$, if ϕ_π is the local L-parameter of π. Note that this definition of local L-factors is assumption-free when either π is unramified

or F is archimedean. When F is non-archimedean and π is ramified, the local Langlands conjecture for (G, F) remains to be proved in general, although many cases have been proved.

To complete the proof of Theorem 1.3, we recall from the fiber product diagram in (2-2) that there is a morphism

$$\eta_{\mathrm{fp}}^{\vee} : G^{\vee} \to H_{\rho}^{\vee}. \tag{2-20}$$

Let $\mathcal{L}_F$ be the local Langlands group of F. Then the local L-parameters of $G(F)$ are the admissible homomorphisms

$$\phi : \mathcal{L}_F \to G^{\vee}(\mathbb{C})$$

up to conjugation by $G^{\vee}(\mathbb{C})$. Let $\Pi_F(G)$ be the set of equivalence classes of irreducible admissible representations of $G(F)$, which are assumed to be of the Casselman-Wallach type if F is archimedean.

By the assumption in Theorem 1.3, the local Langlands conjecture holds for the L-pair $(G, \rho)_L$ and the BKN-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ over F. For any $\pi \in \Pi_F(G)$, there is a local Langlands parameter ϕ_π associated with π. It is clear that the composition $\eta_{\mathrm{fp}}^{\vee} \circ \phi_\pi$ is a local Langlands parameter for the group $H_\rho(F)$. Hence by the local Langlands conjecture for $H_\rho(F)$, we obtain that $\eta_{\mathrm{fp}}^{\vee} \circ \phi_\pi$ is the local Langlands parameter for some $\sigma \in \Pi_F(H_\rho)$. In other words, we have the diagram

$$\begin{array}{ccccc} \mathcal{L}_F & \xrightarrow{\phi_\pi} & G^{\vee}(\mathbb{C}) & \xrightarrow{\quad\rho\quad} & \mathrm{GL}_{d_\rho}(\mathbb{C}) \\ & \searrow^{\phi_\sigma} & \downarrow{\scriptstyle \eta_{\mathrm{fp}}^{\vee}} & \nearrow_{\rho_{\mathrm{fp}}} & \\ & & H_{\rho}^{\vee}(\mathbb{C}) & & \end{array} \tag{2-21}$$

with $\phi_\sigma = \eta_{\mathrm{fp}}^{\vee} \circ \phi_\pi$. From the definition of local L-factors in (2-19), we obtain

$$L(s, \pi, \rho) = L(s, \rho \circ \phi_\pi) = L(s, \rho_{\mathrm{fp}} \circ \eta_{\mathrm{fp}}^{\vee} \circ \phi_\pi) = L(s, \rho_{\mathrm{fp}} \circ \phi_\sigma) = L(s, \sigma, \rho_{\mathrm{fp}}).$$

Similarly, we obtain $\epsilon(s, \pi, \rho, \psi) = \epsilon(s, \sigma, \rho_{\mathrm{fp}}, \psi)$ for any nontrivial character $\psi : F \to \mathbb{C}^{\times}$. Combining with Theorem 2.4, this completes the proof of Theorem 1.3.

3. Borel's conjecture

To provide more explicit information about the representation σ for the given π in Theorem 1.3, we recall the following conjecture by A. Borel in [2, 10.3.5].

Conjecture 3.1 (Borel). *Let H be a connected reductive group over a local field F and $\eta : H \to G$ a homomorphism with commutative kernel and cokernel. Denote the induced map on the L-parameters by $\Phi(\eta)$. Then for any $\pi \in \Pi_F(G)$ with an L-parameter ϕ, the pullback $\eta^*\pi$ of π is a finite direct sum of irreducible admissible representations $H(F)$ with L-parameter $\Phi(\eta)(\phi)$.*

This conjecture is refined and proved for several cases in [26]. The refined conjecture requires the kernel of $\mathrm{d}\eta : \mathrm{Lie}(H) \to \mathrm{Lie}(G)$, the corresponding Lie algebra of H and G, respectively, is central and the cokernel of η is commutative. We intend to show that the morphism

$$\eta_{\mathrm{fp}} : H_\rho \to G \tag{3-1}$$

which is dual to the morphism $\eta_{\mathrm{fp}}^\vee : G^\vee \to H_\rho^\vee$ as defined in (2-2), satisfies this condition.

Lemma 3.2. *The morphism $\eta_{\mathrm{fp}} : H_\rho \to G$ in* (3-1) *enjoys the following properties.*

(1) *The kernel of $\mathrm{d}\eta_{\mathrm{fp}}$ is central.*

(2) *The cokernel of η_{fp} is a commutative group defined over F.*

Proof. According to the proof of Proposition 2.3, the morphism in (2-18) and Proposition 2.3, the morphism η_{fp} is explicitly given by

$$H_\rho \cong \mathbb{G}_m \ltimes_\lambda G_0^{\mathrm{sc}} \to G : (a, g) \mapsto \lambda(a)\bar{g}.$$

Let $\mathfrak{g}$ be the Lie algebra of G and $\mathfrak{g}_0 := [\mathfrak{g}, \mathfrak{g}]$, which is the Lie algebra of G_0^{sc}. Let $\mathfrak{g}_a$ be the Lie algebra of $\mathbb{G}_m$, then the morphism $\mathrm{d}\eta$ is given by

$$\mathrm{d}\eta_{\mathrm{fp}} : \mathfrak{g}_a \oplus \mathfrak{g}_0 \to \mathfrak{g} : (A, X) \mapsto \mathrm{d}\lambda(A) + X,$$

where $\mathrm{d}\lambda : \mathfrak{g}_a \to \mathfrak{g}$ is the differential of $\lambda : \mathbb{G}_m \to G$. Hence we have the kernel of $\mathrm{d}\eta$ given by

$$\{(A, -\mathrm{d}\lambda(A)) \mid A \in \mathfrak{g}_a\}.$$

According to the multiplication of H_ρ given in Proposition 2.3, the Lie bracket in $\mathfrak{g}_a \oplus \mathfrak{g}_0$ is given as

$$[(A_1, X_1), (A_2, X_2)]_{H_\rho} = (0, -[[\mathrm{d}\lambda(A_2), X_1], X_2]),$$

where the brackets on the right-hand side of the equality is the Lie bracket in $\mathfrak{g}$. Then, for any $A \in \mathfrak{g}_a$ and $(A', X') \in \mathfrak{g}_a \oplus \mathfrak{g}_0$, we have

$$[(A, -\mathrm{d}\lambda(A)), (A', X')]_{H_\rho} = (0, [[\mathrm{d}\lambda(A'), \mathrm{d}\lambda(A)], X']) = 0,$$

since $[\mathrm{d}\lambda(A'), \mathrm{d}\lambda(A)] = \mathrm{d}\lambda([A, A']) = 0$, where $[A, A']$ denotes the trivial Lie bracket in $\mathfrak{g}_a$. This proves part (1).

As for part (2), note that $G/\eta_{\mathrm{fp}}(H_\rho) = G/\lambda(\mathbb{G}_m)G_0$, and since G/G_0 is already abelian, we obtain $G/\eta_{\mathrm{fp}}(H_\rho)$ is also abelian and defined over F since every group is defined over F. We are done. □

Assume that the general Borel conjecture holds for the morphism $\eta_{\mathrm{fp}} : H_\rho \to G$, which under the above Condition of the Lemma 3.2 is proved for many cases in [26, Theorem 3]. We deduce that the σ appearing in Theorem 1.3 occurs as a

direct summand in the direct sum decomposition of the pull-back of π under the morphism η_{fp}, with expected multiplicity [26, Conjecture 2]. Hence Theorem 1.3 can be improved:

Theorem 3.3. *Assume that the local Langlands conjecture over F holds for a given L-pair $(G, \rho)_L$ and its associated* BKN*-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$. Assume that the Borel conjecture holds for the morphism $\eta_{\mathrm{fp}} : H_\rho \to G$ as in* (3-1)*. Then for any $\pi \in \Pi_F(G)$, there exists a $\sigma \in \Pi_F(H_\rho)$, which occurs in the direct decomposition of the pull-back of π via the morphism η_{fp}, such that*

$$L(s, \pi, \rho) = L(s, \sigma, \rho_{\mathrm{fp}}) \quad \textit{and} \quad \epsilon(s, \pi, \rho, \psi) = \epsilon(s, \sigma, \rho_{\mathrm{fp}}, \psi)$$

for any nontrivial character $\psi : F \to \mathbb{C}^\times$.

4. *L*-monoids and BKN-pairs

By Theorem 1.3, for any given L-pair $(G, \rho)_L$, in order to understand the Langlands L-function associated with $(G, \rho)_L$, locally and globally, it is enough to work with the associated BKN-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$, which is the pair given by Definition 1.1. In this section, we may take any BKN-pair $(G, \rho)_{\mathrm{BKN}}$ from Definition 1.1.

As explained in [13], the Schwartz space in the Braverman–Kazhdan proposal may depend on the geometry of the L-monoid $\mathcal{M}$ that contains G as the unit group, which is the open subset of all invertible elements. From [13, Proposition 5.1], one can deduce the existence of such L-monoid $\mathcal{M}^\rho$, following the general theory of M. Putcha and L. Renner [21]. For the BKN-pair $(G, \rho)_{\mathrm{BKN}}$, since the derived group $G_0 = [G, G]$ is simply connected, one may also construct a monoid $\mathcal{M}_\rho$ from the construction of the Vinberg universal monoid [12; 14; 29] that also contains G as the unit group. The goal of this section is to prove Theorem 1.4, which states that these two constructions yield the same monoid.

4A. ***Construction via the Vinberg method.*** We now explain how to construct the Vinberg L-monoid $\mathcal{M}_\rho$ of a BKN pair $(G, \rho)_{\mathrm{BKN}}$.

Take k to be a field of characteristic zero and fix an k-Borel pair (B, T) of G. By Definition 1.1, the derived group $G_0 = [G, G]$ is simply connected. Let $(B_0, T_0) = (B \cap G_0, T \cap G_0)$ be the corresponding k-Borel pair of G_0. Let Z be the center of G and Z_0 the center of G_0. Then we have the adjoint group $G_{\mathrm{ad}} = G/Z = G_0/Z_0$, and in particular, $T_{\mathrm{ad}} = T_0/Z_0 = T/Z$.

Associated with (G_0, B_0, T_0), we take Δ to be the set of simple roots and $\hat{\Delta}$ to be the set of the associated fundamental weights. For each fundamental weight $\omega \in \hat{\Delta}$, denote by ρ_ω the associated irreducible rational representation of G_0 with highest weight ω. Denote by V_ω the underlying space of ρ_ω. Define $G_0^+ := (T_0 \times G_0)/Z_0^\Delta$, where Z_0^Δ denotes the diagonal embedding of the center Z_0 of G_0 into $T_0 \times G_0$.

We may extend the representation ρ_ω of G_0 to G_0^+ by

$$\rho_\omega^+(t, g) = \omega(w_0(t^{-1}))\rho_\omega(g)' for (t, g) \in T_0 \times G_0, \tag{4-1}$$

where w_0 is the longest Weyl element in the Weyl group $W = W(G_0, T_0)$. Each root $\alpha \in \Delta$ can also be extended to G_0^+ via $\alpha^+(t, g) := \alpha(t)$.

Following [29], we construct the Vinberg universal monoid. Define a map ι by composing the representation

$$\Big(\prod_{\alpha\in\Delta} \alpha^+\Big)\times\Big(\prod_{\omega\in\hat{\Delta}} \rho_\omega^+\Big)\colon G_0^+ \to \Big(\prod_{\alpha\in\Delta} \mathbb{G}_m\Big)\times\Big(\prod_{\omega\in\hat{\Delta}} \mathrm{GL}(V_\omega)\Big) \tag{4-2}$$

with the embedding

$$\Big(\prod_{\alpha\in\Delta} \mathbb{G}_m\Big)\times\Big(\prod_{\omega\in\hat{\Delta}} \mathrm{GL}(V_\omega)\Big)\to \Big(\prod_{\alpha\in\Delta} \mathbb{A}^1\Big)\times\Big(\prod_{\omega\in\hat{\Delta}} \mathrm{End}(V_\omega)\Big).$$

Take $\mathcal{M}^+$ to be the normalization of the closure of the image $\iota(G_0^+)$ of G_0^+, which is the *Vinberg universal monoid* associated with the given reductive group G from $(G, \rho)_{\mathrm{BKN}}$. Denote by G^+ the unit group of the monoid $\mathcal{M}^+$. It is clear that $G^+ \times G^+$ acts on $\mathcal{M}^+$ by

$$m(g_1, g_2) = g_1^{-1} m g_2 \quad \text{for } m \in \mathcal{M}^+, (g_1, g_2) \in G^+ \times G^+.$$

Let $G^+_{\mathrm{der}} = [G^+, G^+]$ be the derived group of the unit group G^+. The abelianization morphism from $\mathcal{M}^+$ to the GIT quotient $\mathcal{M}^+/\!\!/(G^+_{\mathrm{der}} \times G^+_{\mathrm{der}})$ by $G^+_{\mathrm{der}} \times G^+_{\mathrm{der}}$ is denoted by

$$\pi^+ : \mathcal{M}^+ \to \mathcal{M}^+/\!\!/(G^+_{\mathrm{der}} \times G^+_{\mathrm{der}}) = \prod_{\alpha\in\Delta} \mathbb{A}^1. \tag{4-3}$$

It is a flat morphism as algebraic varieties with equidimensional reduced fibers [12; 14; 29].

Since the derived group G_0 of G is split and simply connected, its dual group $G_0^\vee$ is of adjoint type. Let $G_0^{\vee,\mathrm{sc}}$ be the simply connected cover of $G_0^\vee$. Then $G_0^{\vee,\mathrm{sc}}$ is the dual group of the adjoint group G_{ad} of G. Let $T_0^\vee$ be the maximal torus of $G_0^\vee$ that is dual to T_0 and $T_0^{\vee,\mathrm{sc}}$ be the maximal torus of $G_0^{\vee,\mathrm{sc}}$ that is dual to the torus T_{ad}. It is clear that $T_{\mathrm{ad}} = T_0/Z_0 = T/Z$ and we have the following exact sequences

$$1 \to Z_0 \to T_0 \to T_0/Z_0 \to 1 \quad \text{and} \quad 1 \to Z_0^{\vee,\mathrm{sc}} \to T_0^{\vee,\mathrm{sc}} \to T_0^\vee \to 1,$$

where $Z_0^{\vee,\mathrm{sc}}$ is the center of $G_0^{\vee,\mathrm{sc}}$ and is the Cartier dual of Z_0. Let $T^\vee$ be the maximal torus of $G^\vee$ that is dual to T. Then $X^*(T) = X_*(T^\vee)$ and $X^*(T^\vee) = X_*(T)$.

Let λ be the highest weight of ρ. In this case, we may write $\rho = \rho_\lambda$. This dominant weight λ can be viewed as a dominant cocharacter in $X_*(T)$ of T. By composing with the canonical map $T \to T/Z = T_0/Z_0 = T_{0,\mathrm{ad}}$, we regard such

$\lambda \in X_*(T)$ as an element in $X_*(T_{0,\mathrm{ad}})$, which is still denoted by λ if it does not cause any confusion. Then we obtain its composition

$$\mathbb{G}_m \to T_0/Z_0 \to \prod_{\alpha\in\Delta} \mathbb{G}_m, \tag{4-4}$$

which extends to $\mathbb{A}^1 \to \prod_{\alpha\in\Delta} \mathbb{A}^1$ and is still denoted by λ. The monoid $\mathcal{M}_\lambda$ associated with λ or the representation ρ is defined to be the fiber product

$$\begin{array}{ccc} \mathcal{M}_\lambda & \longrightarrow & \mathcal{M}^+ \\ \downarrow & & \downarrow{\scriptstyle \pi^+} \\ \mathbb{A}^1 & \xrightarrow{\lambda} & \prod_{\alpha\in\Delta} \mathbb{A}^1 \end{array} \tag{4-5}$$

where the abelianization morphism π^+ is as given in (4-3). The normality of $\mathcal{M}_\lambda$ follows from [22, Proposition 2, Theorem 9].

Let G_λ be the unit group of $\mathcal{M}_\lambda$, which fits into an exact sequence

$$1 \to G_0 \to G_\lambda \xrightarrow{m_\lambda} \mathbb{G}_m \to 1 \tag{4-6}$$

where m_λ is the abelianization morphism of G_λ. More precisely, by the definition of G_0^+ and by (4-2), (4-4) and (4-5), the unit group G_λ fits into the fiber product

$$\begin{array}{ccc} G_\lambda & \longrightarrow & G_0^+ = (T_0 \times G_0)/Z_0^\Delta \\ \downarrow & & \downarrow{\scriptstyle \mathrm{pr}_1} \\ \mathbb{G}_m & \xrightarrow{\lambda} & T_{\mathrm{ad}} = T_0/Z_0 \end{array} \tag{4-7}$$

and the multiplication of the group G_λ is induced from these of $\mathbb{G}_m$ and G_0^+.

Lemma 4.1. *We may write $G_\lambda = \mathbb{G}_m \ltimes_\lambda G_0$, with the multiplication given by*

$$(a_1, g_1)(a_2, g_2) = (a_1a_2, \lambda(a_2)^{-1} g_1 \lambda(a_2) g_2) \quad \text{for } a_1, a_2 \in \mathbb{G}_m,\ g_1, g_2 \in G_0. \tag{4-8}$$

Since $\lambda(a_1) \in T_0/Z_0$, the product $\lambda(a_2)g_1\lambda(a_2)^{-1}$ is well-defined.

Proof. We have an isomorphism of varieties

$$(T_0 \times G_0)/Z_0^\Delta \cong (T_0/Z_0) \times G_0 : [(t, g)] \mapsto ([t], t^{-1}g)$$

with inverse given by $(T_0/Z_0) \times G_0 \ni ([t], g) \mapsto [(t, tg)] \in (T_0 \times G_0)/Z_0^\Delta$. Then the projection $\mathrm{pr}_1 : (T_0 \times G_0)/Z_0^\Delta$ transfers to the canonical projection

$$(T_0/Z_0) \times G_0 \to T_0/Z_0.$$

Then $\mathbb{G}_m \times G_0$ is the fiber product of $(T_0/Z_0) \times G_0$ with $\mathbb{G}_m$ over T_0/Z_0 according to the transitivity of the product (see the second to last paragraph in [7, p. 89], for

instance):

$$\begin{array}{ccc} \mathbb{G}_m \times G_0 & \xrightarrow{\lambda\times \mathrm{Id}} & (T_0/Z_0)\times G_0 \\ \downarrow & & \downarrow{\scriptstyle \mathrm{pr}_1} \\ \mathbb{G}_m & \xrightarrow{\lambda} & T_{\mathrm{ad}} = T_0/Z_0 \end{array} \tag{4-9}$$

with morphisms

$$q_1' = \lambda \times \mathrm{Id} : \mathbb{G}_m \times G_0 \to (T_0/Z_0)\times G_0 : (a, g) \mapsto (\lambda(a), g)$$

and

$$q_2 = \mathrm{pr}_1 : \mathbb{G}_m \times G_0 \to \mathbb{G}_m : (a, g) \mapsto a.$$

Moreover, for any variety X with $f_1 = (f_1^{(1)}, f_1^{(2)}) : X \to (T_0/Z_0)\times G_0$ and $f_2 : X \to \mathbb{G}_m$ such that $\mathrm{pr}_1 \circ f_1 = \lambda \circ f_2$, there is a unique morphism $f : X \to \mathbb{G}_m \times G_0$ such that $(\lambda \times \mathrm{Id}) \circ f = f_1$ and $q_2 \circ f = f_2$, which is given by $x \mapsto (f_2(x), f_1^{(2)}(x))$.

Under the isomorphism $(T_0/Z_0)\times G_0 \cong (T_0 \times G_0)/Z_0^{\Delta}$, the morphism q_1' induces the corresponding morphism $q_1 : \mathbb{G}_m \times G_0 \to (T_0 \times G_0)/Z_0^{\Delta}$ is given by $(a, g) \mapsto [(\lambda(a), \lambda(a)g)]$. Then we obtain the fiber product diagram

$$\begin{array}{ccc} \mathbb{G}_m \times G_0 & \xrightarrow{q_1} & (T_0 \times G_0)/Z_0^{\Delta} \\ \downarrow{\scriptstyle q_2} & & \downarrow{\scriptstyle \mathrm{pr}_1} \\ \mathbb{G}_m & \xrightarrow{\lambda} & T_{\mathrm{ad}} = T_0/Z_0 \end{array} \tag{4-10}$$

with $\mathbb{G}_m \times G_0$ as the fiber product. Moreover, using the above identification, for any variety X with $f_1 = (f_1^{(1)}, f_1^{(2)}) : X \to (T_0 \times G_0)/Z_0^{\Delta}$ and $f_2 : X \to \mathbb{G}_m$ such that $\mathrm{pr}_1 \circ f_1 = \lambda \circ f_2$, there is a unique morphism $f : X \to \mathbb{G}_m \times G_0$ such that $q_1 \circ f = f_1$ and $q_2 \circ f = f_2$ and is given by $x \mapsto (f_2(x), f_1^{(1)}(x)^{-1} \cdot f_1^{(2)}(x))$. It remains to determine the multiplication structure explicitly.

For given $(a_1, g_1), (a_2, g_2) \in \mathbb{G}_m \times G_0$, we have

$$q_2((a_1, g_1)) \cdot q_2((a_2, g_2)) = a_1 a_2$$

and

$$\begin{aligned} q_1((a_1, g_1)) \cdot q_1((a_2, g_2)) &= [(\lambda(a_1), \lambda(a_1)g_1)] \cdot [(\lambda(a_2), \lambda(a_2)g_2)] \\ &= [(\lambda(a_1)\lambda(a_2), \lambda(a_1)g_1\lambda(a_2)g_2)], \end{aligned}$$

where the first multiplication is in $\mathbb{G}_m$ and the second multiplication is in $(T_0 \times G_0)/Z_0^{\Delta}$. In this way, we get the morphisms

$$(\mathbb{G}_m \times G_0) \times (\mathbb{G}_m \times G_0) \to \mathbb{G}_m : ((a_1, g_1), (a_2, g_2)) \mapsto a_1 a_2$$

and

$$\begin{aligned} &(\mathbb{G}_m \times G_0) \times (\mathbb{G}_m \times G_0) \to (T_0 \times G_0)/Z_0^{\Delta} \\ &\qquad ((a_1, g_1), (a_2, g_2)) \mapsto [(\lambda(a_1)\lambda(a_2), \lambda(a_1)g_1\lambda(a_2)g_2)], \end{aligned}$$

whose corresponding morphism $(\mathbb{G}_m \times G_0) \times (\mathbb{G}_m \times G_0) \to (\mathbb{G}_m \times G_0)$ is given by

(4-11)
$$(a_1, g_1) \cdot (a_2, g_2) = (a_1 a_2, \lambda(a_1 a_2)^{-1} \lambda(a_1) g_1 \lambda(a_2) g_2) = (a_1 a_2, \lambda(a_2)^{-1} g_1 \lambda(a_2) g_2)$$

according to the last paragraph.

As the fiber product of the fiber product diagram in (4-10), the group $\mathbb{G}_m \times G_0$ with the multiplication given in (4-11) is isomorphic to the group G_λ because of the universality. □

From the proof of Lemma 4.1, it is easy to obtain that when the cocharacter $\lambda : \mathbb{G}_m \to T_0/Z_0$ factors through $\mathbb{G}_m \to T_0 \to T_0/Z_0$, or equivalently, the representation ρ_λ can factor through the adjoint group $G^\vee_{\mathrm{ad}}$, the unit group G_λ is isomorphic to the direct product $\mathbb{G}_m \times G_0$, which is explicitly given by

$$G_\lambda \cong \mathbb{G}_m \times G_0, \quad \text{with } (a, g) \mapsto (a, \lambda(a) g). \tag{4-12}$$

For convenience, we state it as a corollary.

Corollary 4.2. *If the representation ρ_λ can factor through the adjoint group $(G^\vee)_{\mathrm{ad}}$, then the unit group G_λ is isomorphic to the direct product $\mathbb{G}_m \times G_0$, which is explicitly given as in* (4-12).

For any BKN-pair $(G, \rho)_{\mathrm{BKN}}$ with $\rho = \rho_\lambda$, one checks easily that the natural homomorphism

$$G^\vee \to H^\vee_\rho \cong (\mathbb{G}_m \times G^\vee)/\{(\lambda(z)^{-1}, z) \mid z \in Z^\vee\} : g \mapsto [(1, g)]$$

is an isomorphism. By Proposition 2.3 and Lemma 4.1, we obtain the following

Corollary 4.3. *For a given* BKN*-pair $(G, \rho)_{\mathrm{BKN}}$ with $\rho = \rho_\lambda$, the group G is isomorphic to the unit group $G_\lambda = G_\rho = G(\mathcal{M}_\rho)$, where $\mathcal{M}_\rho = \mathcal{M}_\lambda$ is the L-monoid as constructed via the Vinberg method in* (4-5).

4B. ***Construction via the Putcha–Renner method.*** Given a BKN-pair $(G, \rho)_{\mathrm{BKN}}$ with $\rho = \rho_\lambda$ as before, we are going to recall the construction of the monoid $\mathcal{M}^\rho$ via the Putcha–Renner method and show that $G = G(\mathcal{M}^\rho)$, the unit group of the monoid $\mathcal{M}^\rho$.

A series of papers by Putcha [15; 16] and Renner [17; 18; 19; 20] classified the reductive monoid characterized by the toric embedding of maximal tori. The survey [21] is also a good reference. An explicit construction over not necessarily algebraically closed fields is given in [14].

Following the notation in [14], we denote by $\Omega(\rho)$ the set of weights of ρ and by $\xi(\rho)$ the convex cone generated by $\Omega(\rho)$. According to [13, Proposition 5.1], since the central character of ρ is the scalar multiplication, it follows that $\xi(\rho)$ is strictly

convex. Let $\xi(\rho)^\vee$ be the sub-monoid of $X^*(T)$ consisting of $\alpha \in X^*(T)$ such that the restriction of α to $\Omega(\rho)$ takes non-negative values. Let $\alpha_1, \ldots, \alpha_k \in X^*(T)$ such that their orbits under the Weyl group W generate $\xi(\rho)^\vee$. By replacing α_i by a W-conjugate if necessary, we may assume that each α_i lies in the positive Weyl chamber. Let $\omega_{\alpha_i} : G \to \mathrm{GL}(V_{d_i})$ be the representation of G with α_i being the highest weight, where d_i is the dimension of the representation. Then the monoid $\mathcal{M}^\rho$ is defined as the normalization of the closure of the image G under

$$\prod_{i=1}^{k} \omega_{\alpha_i} : G \to \prod_{i=1}^{k} \mathrm{End}(V_{d_i}).$$

According to [14, Proposition 5.1], the monoid $\mathcal{M}^\rho$ is independent of the choice of the generators $\alpha_1, \ldots, \alpha_k$ and is characterized by the property that for every maximal torus T of G, the normalization of its closure in $\mathcal{M}^\rho$ corresponds to the strictly convex cone generated by the set of weights $\Omega(\rho)$ in the sense that the ring of regular functions on this toric embedding is $k[\xi(\rho)^\vee]$.

4C. ***Proof of Theorem 1.4.*** For a given BKN-pair $(G, \rho)_{\mathrm{BKN}}$ with $\rho = \rho_\lambda$ as before, we constructed in Section 4A a monoid $\mathcal{M}_\rho$ with $G = G(\mathcal{M}_\rho)$, the unit group of $\mathcal{M}_\rho$, and in Section 4B a monoid $\mathcal{M}^\rho$ with $G = G(\mathcal{M}^\rho)$, the unit group of $\mathcal{M}^\rho$. To prove Theorem 1.4, it is enough to prove that $\mathcal{M}_\rho \cong \mathcal{M}^\rho$.

We regard λ as a dominant element in $X_*(T_{0,\mathrm{ad}}) = X^*(T_0^{\vee,\mathrm{sc}})$ as in Section 4A. Recall from Corollary 4.3 that $G \cong G_\lambda = \mathbb{G}_m \ltimes_\lambda G_0$. When $\lambda = 0$, it is clear that ρ does not satisfy (4) in Definition 1.1. Hence we are able to the condition that $\lambda \neq 0$.

Take the maximal split torus $T_\lambda = \mathbb{G}_m \times T_0$ and recall the commutative diagram

$$\begin{array}{ccccccc}
T_\lambda = \mathbb{G}_m \times T_0 & \longrightarrow & G_\lambda = \mathbb{G}_m \ltimes_\lambda G_0 & \longrightarrow & \mathcal{M}_\rho & \longrightarrow & \mathbb{A}^1 \\
\downarrow & & \downarrow & & \downarrow & & \downarrow{\scriptstyle \lambda} \\
T_0^+ = (T_0 \times T_0)/Z_0^\Delta & \longrightarrow & G_0^+ & \longrightarrow & \mathcal{M}^+ & \xrightarrow[\pi^+]{} & \prod_{\alpha\in\Delta} \mathbb{A}^1
\end{array}$$

where every square is Cartesian and the left vertical arrow is given by

$$\mathbb{G}_m \times T_0 \to (T_0 \times G_0)/Z_0^\Delta : (a, t) \mapsto [\lambda(a), \lambda(a)t].$$

Recall $\mathcal{M}^+$ is the normalization of the image of the following map

$$\begin{aligned}
\iota : G_0^+ = (T_0 \times G_0)/Z_0^\Delta \to \Big(\prod_{\alpha\in\Delta} \mathbb{A}^1\Big) \times \Big(\prod_{\omega\in\hat{\Delta}} \mathrm{End}(V_\omega)\Big), \\
[t, g] \longmapsto \Big(\prod_{\alpha\in\Delta} \alpha(t)\Big) \times \Big(\prod_{\omega\in\hat{\Delta}} \omega(w_0(t^{-1}))\rho_\omega(g)\Big).
\end{aligned}$$

Regard ι as a representation of G_0^+ and denote the convex cone generated by all of the weights in ι by $\xi^+ \subset X^*(T_0^+) \otimes \mathbb{R}$. Let $\overline{T}_0^+$ denote the Zariski closure of T_0^+ in $\mathcal{M}^+$. According to [21, Theorem 5.4], $\overline{T}_0^+$ is a normal affine toric variety. It follows from [21, Section 3.3] that $\overline{T}_0^+ = \operatorname{Spec} k[X(\overline{T}_0^+)]$, where the character lattice $X(\overline{T}_0^+)$ of $\overline{T}_0^+$ is given by

$$X(\overline{T}_0^+) = \xi^+ \cap X^*(T_0^+).$$

Let $\overline{T}_\lambda$ denote the Zariski closure of T_λ in $\mathcal{M}_\rho$. Note $\lambda \neq 0$. Then the commutative diagram

$$\begin{array}{ccc} \overline{T}_\lambda & \longrightarrow & \mathbb{A}^1 \\ \downarrow & & \downarrow{\scriptstyle \lambda} \\ \overline{T}_0^+ & \longrightarrow & \prod_{\alpha\in\Delta} \mathbb{A}^1 \end{array}$$

is Cartesian.

Identify the character lattice $X^*(T_\lambda)$ of $T_\lambda = \mathbb{G}_m \times T_0$ with $\mathbb{Z} \oplus X^*(T_0)$. Denote by $\xi_\lambda \subset X^*(T_\lambda) \otimes_{\mathbb{Z}} \mathbb{R}$ the convex cone generated by the characters

$$(1, 0) \quad \text{and} \quad (\langle \omega' - w_0\omega, \lambda\rangle, \omega'), \ \ \omega \in \hat{\Delta}.$$

Here ω' denotes any weight in ρ_ω and $\langle\ ,\ \rangle$ denotes the pairing between $X^*(T_{0,\mathrm{ad}})$ and $X_*(T_{0,\mathrm{ad}})$. Note $\omega' - w_0\omega \in X^*(T_{0,\mathrm{ad}})$, so the pairing is well-defined.

By the Cartesian diagram above, one calculates directly that $\overline{T}_\lambda = \operatorname{Spec} k[X(\overline{T}_\lambda)]$, where the character lattice of $\overline{T}_\lambda$ is given by

$$X(\overline{T}_\lambda) = \xi_\lambda \cap X^*(T_\lambda).$$

On the other hand, by the same arguments as in Proposition 2.3, the Langlands dual group of $G \cong G_\lambda$ is

$$G_\lambda^\vee = \mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z_0^{\vee,\mathrm{sc}})} G_0^{\vee,\mathrm{sc}} \cong G^\vee.$$

The isomorphism is induced from $\mathfrak{d}^\vee : \mathbb{G}_m \to G^\vee$ and the homomorphism $G_0^{\vee,\mathrm{sc}} \to G^\vee$, which is dual to the quotient $G \to G_{\mathrm{ad}} = G_{0,\mathrm{ad}}$. Under this isomorphism, we may write the representation ρ of $G^\vee$ as the representation of $G_\lambda^\vee$ given by

$$G_\lambda^\vee = \mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z_0^{\vee,\mathrm{sc}})} G_0^{\vee,\mathrm{sc}} \to \mathrm{GL}(V_\rho) : (a, g) \mapsto a\mathrm{I}_{V_\rho} \cdot \rho'_\lambda(g),$$

where ρ'_λ denotes the highest weight representation of $G_0^{\vee,\mathrm{sc}}$ associated with $\lambda \in X^*(T_0^{\vee,\mathrm{sc}})$.

Take the maximal torus $\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z_0^{\vee,\mathrm{sc}})} T_0^{\vee,\mathrm{sc}}$, which has the isomorphism

$$\mathbb{G}_m \times_{(\lambda^{-1}, \mathrm{Id})(Z_0^{\vee,\mathrm{sc}})} T_0^{\vee,\mathrm{sc}} \cong \mathbb{G}_m \times (T_0^{\vee,\mathrm{sc}})_{\mathrm{ad}} = \mathbb{G}_m \times T_0^\vee = T_\lambda^\vee :$$
$$[a, t] \mapsto (\lambda(t)a, [t]).$$

Identify the character lattice $X^*(\mathbb{G}_m \times_{(\lambda^{-1},\mathrm{Id})(Z_0^{\vee,\mathrm{sc}})} T_0^{\vee,\mathrm{sc}})$ with $\mathbb{Z} \oplus X^*(T_0^\vee)$. Then the weights of ρ_λ are given by

$$(1, \lambda' - \lambda)$$

where $\lambda' \in X^*(T_0^{\vee,\mathrm{sc}})$ denotes any weight ρ'_λ. Note $\lambda' - \lambda \in X^*(T_0^\vee)$. So $\xi(\rho) \subset X_*(T_\lambda) \otimes_{\mathbb{Z}} \mathbb{R}$ is equal to the convex cone generated by the weights above. Let $\xi(\lambda)^\vee$ be its dual cone.

Let $\overline{T}^\lambda$ denote the Zariski closure of T_λ in $\mathcal{M}^\rho$. Similarly $\overline{T}^\lambda$ is the normal affine toric variety Spec $k[X(\overline{T}^\lambda)]$, where

$$X(\overline{T}^\lambda) = \xi(\lambda)^\vee \cap X^*(T_\lambda) = \xi(\rho)^\vee.$$

We next prove that $\xi_\lambda = \xi(\lambda)^\vee$, which implies that $\overline{T}_\lambda \cong \overline{T}^\lambda$. Then by [21, Theorem 5.4] this isomorphism extends to an isomorphism $\mathcal{M}_\rho \cong \mathcal{M}^\rho$.

Proposition 4.4. *With notation as above, the cones coincide*: $\xi_\lambda = \xi(\lambda)^\vee$.

For the proof, we notice that the isogeny $T_0 \to T_{0,\mathrm{ad}}$ induces isomorphisms

$$X^*(T_{0,\mathrm{ad}}) \otimes_{\mathbb{Z}} \mathbb{R} \cong X^*(T_0) \otimes_{\mathbb{Z}} \mathbb{R} \quad \text{and} \quad X_*(T_0) \otimes_{\mathbb{Z}} \mathbb{R} \cong X_*(T_{0,\mathrm{ad}}) \otimes_{\mathbb{Z}} \mathbb{R}.$$

So we can use the same notation $\langle\ ,\ \rangle$ to denote the pairing between $X^*(T_{0,\mathrm{ad}}) \otimes_{\mathbb{Z}} \mathbb{R} \cong X^*(T_0) \otimes_{\mathbb{Z}} \mathbb{R}$ and $X_*(T_0) \otimes_{\mathbb{Z}} \mathbb{R} \cong X_*(T_{0,\mathrm{ad}}) \otimes_{\mathbb{Z}} \mathbb{R}$. If the context is clear, we may also use $\langle\ ,\ \rangle$ to denote the pairing between

$$X^*(T_\lambda) \otimes_{\mathbb{Z}} \mathbb{R} = \mathbb{R} \oplus X^*(T_0) \otimes_{\mathbb{Z}} \mathbb{R} \quad \text{and} \quad X_*(T_\lambda) \otimes_{\mathbb{Z}} \mathbb{R} = \mathbb{R} \oplus X_*(T_0) \otimes_{\mathbb{Z}} \mathbb{R}.$$

Proof. We first prove that $\xi_\lambda \subset \xi(\lambda)^\vee$. By definition,

$$\langle (1, 0), (1, \lambda' - \lambda) \rangle = 1 > 0,$$

and

$$\begin{aligned} \big\langle (\langle \omega' - w_0\omega, \lambda \rangle, \omega'), (1, \lambda' - \lambda) \big\rangle &= \langle \omega' - w_0\omega, \lambda \rangle + \langle \omega', \lambda' - \lambda \rangle \\ &= \langle \omega', \lambda' \rangle - \langle w_0\omega, \lambda \rangle. \end{aligned}$$

Any weight in the highest weight representation associated with λ lies in the convex hull of $\{W\lambda\}$. It follows that for a fixed ω', one has

$$\langle \omega', \lambda' \rangle \geq \langle \omega', w\lambda \rangle = \langle w^{-1}\omega', \lambda \rangle$$

for certain $w \in W$, which implies that

$$\big\langle (\langle \omega' - w_0\omega, \lambda \rangle, \omega'), (1, \lambda' - \lambda) \big\rangle \geq \langle w_0(w_0 w^{-1}\omega' - \omega), \lambda \rangle.$$

Since ω is the highest weight in ρ_ω, we have

$$\omega - w_0 w^{-1}\omega' = \sum_{\alpha \in \Delta} n_\alpha \alpha, \quad n_\alpha \geq 0,$$

and

$$w_0(w_0w^{-1}\omega' - \omega) = \sum_{\alpha\in\Delta} n_\alpha(-w_0\alpha).$$

Since $-w_0\Delta = \Delta$ and λ is dominant we obtain $\langle w_0(w_0w^{-1}\omega' - \omega), \lambda\rangle \geq 0$, which implies that $\xi_\lambda \subset \xi(\lambda)^\vee$.

It remains to prove that $\xi(\lambda)^\vee \subset \xi_\lambda$. Suppose that $(a, \mu') \in \xi(\lambda)^\vee$. Then $\mu' = w\mu$ for certain $w \in W$ and dominant $\mu \in X_*(T_0)\otimes_{\mathbb{Z}} \mathbb{R}$. It follows that

$$\langle (a, \mu'), (1, \lambda' - \lambda)\rangle = a + \langle w\mu, \lambda' - \lambda\rangle = a + \langle \mu, w^{-1}\lambda'\rangle - \langle w\mu, \lambda\rangle \geq 0.$$

By taking $\lambda' = ww_0\lambda$, we obtain

$$a \geq \langle w\mu, \lambda\rangle - \langle \mu, w_0\lambda\rangle = \langle w\mu - w_0\mu, \lambda\rangle.$$

Since μ is dominant, we write $\mu = \sum_{\omega\in\hat{\Delta}} m_\omega\omega$ for $m_\omega \geq 0$. This implies

$$(a, \mu') = (a - \langle w\mu - w_0\mu, \lambda\rangle)(1, 0) + \sum_{\omega\in\hat{\Delta}} m_\omega(\langle w\omega - w_0\omega, \lambda\rangle, w\omega) \in \xi_\lambda. \quad \square$$

5. Examples

We provide the explicit BKN-pairs $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ for some examples of L-pairs $(G, \rho)_L$ in order to illustrate the discussions above.

5A. ***Symmetric powers of*** $\mathbf{GL_2}$**.** We take the L-pair $(G, \rho)_L$ to be $(\mathrm{GL}_2, \mathrm{sym}^n)_L$. Since $\rho = \mathrm{sym}^n$, the symmetric n-th power of GL_2, we must have that the highest weight $\lambda_n = n\omega_1$, where ω_1 is the first fundamental weight of $\mathrm{GL}_2(\mathbb{C})$.

In the case $G = \mathrm{GL}_2$, we have $G_0 = \mathrm{SL}_2$. We take T to be the maximal torus of GL_2 consisting of all diagonal matrices, then $T_0 = T \cap \mathrm{SL}_2$ consisting of all diagonal matrices in SL_2, and then $Z_0 = \{\pm \mathrm{I}_2\} \subset T_0$. Since the cocharacter is given by

$$\lambda = \lambda_n : \mathbb{G}_m \to T_0/Z_0 \cong \mathbb{G}_m, \quad x \mapsto x^n.$$

We have $H_\rho = \mathbb{G}_m \times_\lambda \mathrm{SL}_2$. We conclude that $H_\rho \cong \mathbb{G}_m \times \mathrm{SL}_2$ when n is even and $H_\rho \cong \mathrm{GL}_2$ when n is odd, which can be verified as follows.

When n is even, since λ could lift to a cocharacter

$$\lambda : \mathbb{G}_m \to T_0 : x \mapsto \begin{pmatrix} x^{n/2} & 0 \\ 0 & x^{-n/2} \end{pmatrix},$$

by Corollary 4.2, which is a special case of Lemma 4.1, we have $G_\lambda \cong \mathbb{G}_m \times \mathrm{SL}_2$.

When n is odd, there is an algebraic group isomorphism

$$\mathbb{G}_m \ltimes_\lambda \mathrm{SL}_2 \to \mathrm{GL}_2 : (a, g) \mapsto \begin{pmatrix} a^{(n+1)/2} & 0 \\ 0 & a^{(1-n)/2} \end{pmatrix} g$$

with inverse given by

$$g \mapsto \left(\det g, \begin{pmatrix} \det^{(-n-1)/2} g & 0 \\ 0 & \det^{(n-1)/2} g \end{pmatrix} g\right).$$

On the dual side, by Corollary 2.2, we have

$$G_\lambda^\vee = \mathbb{G}_m \times \mathrm{SL}_2/\{(\pm 1)^n, \pm \mathrm{I}_2\}.$$

When n is even, we have $(\pm 1)^n = 1$ and hence $G_\lambda^\vee$ is isomorphic to $\mathbb{G}_m \times \mathrm{PGL}_2$. When n is odd, we have the following isomorphism of algebraic groups:

$$G_\lambda^\vee \cong \mathrm{GL}_2 : (a, g) \mapsto a\mathrm{I}_2 \cdot g. \tag{5-1}$$

Finally, we consider the representation ρ_{fp}. When n is even, the representation is given by

$$\rho_{\mathrm{fp}} = \mathrm{st}_{\mathbb{G}_m} \otimes \mathrm{sym}_n^*, \tag{5-2}$$

where $\mathrm{st}_{\mathbb{G}_m}$ is the standard representation of $\mathbb{G}_m$ acting by scalar, and sym_n^* is the induced symmetric n-th power on PGL_2. When n is odd, according to the isomorphism (5-1), the representation ρ_{fp} is given as follows. For any $g \in \mathrm{GL}_2$, choose $a \in \mathbb{G}_m$ such that $a^2 = \det g$, which implies that $a^{-1}g \in \mathrm{SL}_2$. Hence

$$\rho_{\mathrm{fp}}(g) = a \cdot \mathrm{sym}^n(a^{-1}g). \tag{5-3}$$

Note that this is not the symmetric n-th power of GL_2. We summarize the discussion as a proposition.

Proposition 5.1. *For any integer $n \geq 1$, let $\lambda = n\omega_1$ be the dominant weight for the symmetric n-th power representation ρ_λ of GL_2. Then the* BKN*-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ associated to the given L-pair $(\mathrm{GL}_2, \rho_\lambda)_L$ is explicitly given as follows.*

(1) *The group H_ρ is given by*

$$H_\rho = \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_2 \cong \begin{cases} \mathbb{G}_m \times \mathrm{SL}_2 & \text{if } n \text{ is even,} \\ \mathrm{GL}_2 & \text{if } n \text{ is odd.} \end{cases}$$

(2) *The dual group $H_\rho^\vee$ is given by*

$$H_\rho^\vee \cong \begin{cases} \mathbb{G}_m \times \mathrm{PGL}_2 & \text{if } n \text{ is even,} \\ \mathrm{GL}_2 & \text{if } n \text{ is odd.} \end{cases}$$

(3) *The representation ρ_{fp} is given by*

$$\rho_{\mathrm{fp}} = \begin{cases} \mathrm{st}_{\mathbb{G}_m} \otimes \mathrm{sym}_n^* \text{ as in (5-2)} & \text{if } n \text{ is even,} \\ \rho_{\mathrm{fp}}(g) = a \cdot \mathrm{sym}^n(a^{-1}g) \text{ as in (5-3)} & \text{if } n \text{ is odd.} \end{cases}$$

(4) *The generalized determinant* $\mathfrak{d}_{\mathrm{fp}} : H_\rho \to \mathbb{G}_m$ *is given by*

$$H_\rho \cong \begin{cases} \mathbb{G}_m \times \mathrm{SL}_2 : (a, g) \mapsto a & \text{if } n \text{ is even,} \\ \mathrm{GL}_2 : g \mapsto \det g & \text{if } n \text{ is odd.} \end{cases}$$

We also refer to [10; 24; 25] for discussions of this case.

5B. ***The standard representation of*** $\mathbf{GL}_n$**.** In this case, the L-pair is $(\mathrm{GL}_n, \rho_\lambda)_L$, with ρ_λ being the standard representation of GL_n associated to the first fundamental weight ω_1. To explicate the associated BKN-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$, we take T to be the maximal torus of GL_n consisting of all diagonal matrices, and Z be the center of GL_n consisting of all scalar matrices. Then $T_0 = T \cap \mathrm{SL}_n$ and $Z_0 = Z \cap \mathrm{SL}_n = \mu_n$, where μ_n is the group of the n-th roots of unity. Consider the isomorphism

$$T_0/Z_0 \cong T/Z \cong \mathbb{G}_m^{n-1} : \mathrm{diag}(t_1, \ldots, t_n) \mapsto \left(\frac{t_1}{t_n}, \ldots, \frac{t_{n-1}}{t_n}\right).$$

The highest weight $\lambda = \omega_1$ of the standard representation defines a cocharacter

$$\lambda : \mathbb{G}_m \to T/Z \cong \mathbb{G}_m^{n-1}, \quad t \mapsto \mathrm{diag}(t, 1, \ldots, 1) \mapsto (t, 1, \ldots, 1).$$

Then $H_\rho = \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n$, with group law given by

$$(\mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n) \times (\mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n) \to \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n :$$
$$((a, g), (a', g')) \mapsto (aa', \mathrm{diag}((a')^{-1}, 1, \ldots, 1) g \cdot \mathrm{diag}(a', 1, \ldots, 1) \cdot g').$$

As algebraic groups, we have the isomorphism

$$\mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n \cong \mathrm{GL}_n : (a, g) \mapsto \mathrm{diag}(a, 1, \ldots, 1) \cdot g,$$

with the inverse given by

$$\mathrm{GL}_n \ni g \mapsto (\det g, \mathrm{diag}(\det{}^{-1} g, 1, \ldots, 1) \cdot g) \in \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n.$$

On the dual side, we have

$$\begin{aligned} G_\lambda^\vee &= \mathbb{G}_m \times \mathrm{SL}_n / \{(\omega_\lambda(z)^{-1}, z) \mid z \in \mu_n \subset \mathrm{SL}_n\} \\ &= \mathbb{G}_m \times \mathrm{SL}_n / \{(z^{-1}, z) \mid z \in \mu_n \subset \mathrm{SL}_n\}, \end{aligned}$$

which is isomorphic to GL_n via the map

$$G_\lambda^\vee \to \mathrm{GL}_n : [(a, g)] \mapsto ag.$$

Finally, according to the construction of ρ_{fp} in this case, it is still the standard representation of GL_n.

To summarize the discussion:

Proposition 5.2. *Let λ be the highest weight for the standard representation ρ_λ of GL_n. Then the* BKN*-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ associated to the L-pair $(\mathrm{GL}_n, \rho_\lambda)_L$ is explicitly given as follows.*

(1) *The group H_ρ is given by*

$$H_\rho = \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n \cong \mathrm{GL}_n.$$

(2) *The dual group $H_\rho^\vee$ is given by*

$$H_\rho^\vee \cong \mathrm{GL}_n.$$

(3) *The representation ρ_{fp} is given by*

$$\rho_{\mathrm{fp}} = \rho_\lambda = \mathrm{st}_{\mathrm{GL}_n},$$

the standard representation of GL_n.

(4) *The generalized determinant $\mathfrak{d}_{\mathrm{fp}} : H_\rho \to \mathbb{G}_m$ is given by*

$$H_\rho \cong \mathrm{GL}_n \to \mathbb{G}_m : g \mapsto \det(g).$$

5C. *The adjoint representation of* $\mathbf{GL}_n$. Let $G = \mathrm{GL}_n$ and ρ_λ be the adjoint representation of $\mathrm{GL}_n(\mathbb{C})$ and λ be its highest weight. In this case, we have $G_0 = \mathrm{SL}_n$. Take T and T_0 as in Section 5B.

Since the character λ can be lifted to a cocharacter

$$\mathbb{G}_m \to T_0 : a \mapsto \mathrm{diag}(a, 1, \ldots, 1, a^{-1}),$$

by Corollary 4.2, which is a special case of Lemma 4.1, we obtain $H_\rho = \mathbb{G}_m \times \mathrm{SL}_n$.

On the dual side, we have

$$H_\rho^\vee = \mathbb{G}_m \times \mathrm{SL}_n / \{(\omega_\lambda(z))^{-1}, z) \mid z \in \mu_n \subset \mathrm{SL}_n\} \cong \mathbb{G}_m \times (\mathrm{SL}_n/\mu_n),$$

where μ_n is the finite group scheme of n-th roots of unity, and the last equality is because $\omega_\lambda(z) = 1$ for all $z \in \mu_n$. As algebraic groups, we obtain $G_\lambda^\vee = \mathbb{G}_m \times \mathrm{PGL}_n$.

Finally the corresponding representation is $\rho_{\mathrm{fp}} = \mathrm{st}_{\mathbb{G}_m} \otimes \mathrm{Ad}$, where Ad is the adjoint representation of PGL_n induced from that of SL_n. To summarize:

Proposition 5.3. *Let λ be the highest weight for the adjoint representation ρ_λ of GL_n. Then the* BKN*-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ associated to the L-pair $(\mathrm{GL}_n, \rho_\lambda)_L$ is explicitly given as follows.*

(1) *The group H_ρ is given by*

$$H_\rho = \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n \cong \mathbb{G}_m \times \mathrm{SL}_n.$$

(2) *The dual group* $H_\rho^\vee$ *is given by*

$$H_\rho^\vee \cong \mathbb{G}_m \times \mathrm{PGL}_n.$$

(3) *The representation* ρ_{fp} *is given by*

$$\rho_{\mathrm{fp}} = \mathrm{st}_{\mathbb{G}_m} \otimes \mathrm{Ad},$$

where Ad *is the adjoint representation of* PGL_n *induced from that of* SL_n.

(4) *The generalized determinant* $\mathfrak{d}_{\mathrm{fp}} : H_\rho \to \mathbb{G}_m$ *is given by*

$$\mathbb{G}_m \times \mathrm{SL}_n \to \mathbb{G}_m : (a, g) \mapsto a.$$

5D. ***The symmetric square representation of*** $\mathbf{GL}_n$**.** In this case, we consider the L-pair $(\mathrm{GL}_n, \rho_\lambda)_L$ with $G = \mathrm{GL}_n$ and ρ_λ being the symmetric square representation $\mathrm{GL}_n \to \mathrm{GL}_{n(n+1)/2}$, associated to the highest weight $2\omega_1$. Take T and T_0 as in Section 5B.

The highest weight $\lambda = 2\omega_1 \in X^*(T^\vee) = X_*(T)$ of the symmetric square representation of $G^\vee$ gives a cocharacter

$$\lambda : \mathbb{G}_m \to T, \quad a \mapsto \mathrm{diag}(a^2, 1, \ldots, 1).$$

Similar to the previous case, we can write H_ρ as the image of the following homomorphism

$$\mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n \to \mathbb{G}_m \times \mathrm{GL}_n : (a, g) \mapsto (a, \mathrm{diag}(a^2, 1, \ldots, 1) \cdot g)$$

which is clearly equal to

$$\{(a, g) \in \mathbb{G}_m \times \mathrm{GL}_n \mid \det(g) = a^2\}.$$

More specifically, when $n = 2l$, we have the following isomorphism

$$\mathbb{G}_m \times \mathrm{SL}_n/\{(z^{-1}, z\mathrm{I}_n \mid z \in \mu_l\} \cong H_\rho, \quad [(a, g)] \mapsto (a^l, g \cdot a\mathrm{I}_n),$$

which is the same as in [23].

On the dual side, we have

$$H_\rho^\vee = \mathbb{G}_m \times \mathrm{SL}_n/\{(z^{-2}, z\mathrm{I}_n) \mid z \in \mu_n\},$$

with the isomorphisms

$$\mathbb{G}_m \times \mathrm{SL}_n/\{(z^{-2}, z\mathrm{I}_n) \mid z \in \mu_n\} \cong \mathbb{G}_m \times \mathrm{GL}_n/\{(z^{-2}, z\mathrm{I}_n) \mid z \in \mathbb{G}_m\} :$$
$$[(a, g)] \mapsto [(a, g)]$$

and

$$\mathrm{GL}_n/\{\pm\mathrm{I}_n\} \cong \mathbb{G}_m \times \mathrm{GL}_n/\{(z^{-2}, z\mathrm{I}_n) \mid z \in \mathbb{G}_m\} : [g] \mapsto [(1, g)].$$

Then the unique morphism $\mathrm{GL}_n \to H_\rho^\vee$ is given by the central isogeny

$$\mathrm{GL}_n \to \mathrm{GL}_n/\{\pm \mathrm{I}_n\} : g \mapsto [g].$$

Since the symmetric square representation has its kernel equal to $\{\pm \mathrm{I}_n\}$, the given representation ρ_λ descents to a representation $\rho^{\mathfrak{n}}$ of $\mathrm{GL}_n/\{\pm \mathrm{I}_n\}$.

Dual to the morphism $G^\vee \to G_\lambda^\vee$, we have the morphism $G_\lambda \to G$, which is explicitly given by

$$\{(a, g) \in \mathbb{G}_m \times \mathrm{GL}_n \mid \det(g) = a^2\} \to \mathrm{GL}_n : (a, g) \mapsto g.$$

To summarize:

Proposition 5.4. *Let $\lambda = 2\omega_1$ be the highest weight for the symmetric square representation ρ_λ of GL_n. Then the* BKN*-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ associated to the given L-pair $(\mathrm{GL}_n, \rho_\lambda)_L$ is explicitly given as follows.*

(1) *The group H_ρ is given by*

$$H_\rho = \mathbb{G}_m \ltimes_\lambda \mathrm{SL}_n \cong \{(a, g) \in \mathbb{G}_m \times \mathrm{GL}_n \mid \det(g) = a^2\}$$

(2) *The dual group $H_\rho^\vee$ is given by*

$$H_\rho^\vee \cong \mathrm{GL}_n/\{\pm \mathrm{I}_n\}$$

(3) *The representation ρ_{fp} is given by the descending of ρ_λ from GL_n to $H_\rho^\vee$.*

(4) *The generalized determinant $\mathfrak{d}_{\mathrm{fp}} : H_\rho \to \mathbb{G}_m$ is given by*

$$H_\rho \cong \{(a, g) \in \mathbb{G}_m \times \mathrm{GL}_n \mid \det(g) = a^2\} \to \mathbb{G}_m : (a, g) \mapsto a.$$

5E. ***The doubling case of $\mathbf{Sp_{2n}}$.*** In this case, the given L-pair is $(\mathrm{Sp}_{2n}, \rho_\lambda)_L$ with $G = \mathrm{Sp}_{2n} = G_0$. Let $T = T_0$ be a maximal torus of G. Let ρ_λ be the standard representation of SO_{2n+1} with λ its highest weight. Since $T = T_0$ and λ is already a cocharacter of T_0, by Corollary 4.2, which is a special case of Lemma 4.1, we have in this case $H_\rho = \mathbb{G}_m \times \mathrm{Sp}_{2n}$. On the dual side,

$$H_\rho^\vee = \mathbb{G}_m \times \mathrm{Spin}_{2n+1}/\{(1, \pm \mathrm{I}_{2n+1})\} \cong \mathbb{G}_m \times \mathrm{SO}_{2n+1}.$$

Finally due to the construction of ρ_{fp}, we see in this case that $\rho_{\mathrm{fp}} = \mathrm{st}_{\mathbb{G}_m} \otimes \mathrm{st}_{\mathrm{SO}_{2n+1}}$, where $\mathrm{st}_{\mathbb{G}_m}$ is the standard representation of $\mathbb{G}_m$ and $\mathrm{st}_{\mathrm{SO}_{2n+1}}$ is the standard presentation of SO_{2n+1}. In summary:

Proposition 5.5. *Let λ be the highest weight for the standard representation ρ_λ of SO_{2n+1}. Then the* BKN*-pair $(H_\rho, \rho_{\mathrm{fp}})_{\mathrm{BKN}}$ associated to the L-pair $(\mathrm{Sp}_{2n}, \rho_\lambda)_L$ is explicitly given as follows.*

(1) *The group H_ρ is given by*

$$H_\rho = \mathbb{G}_m \ltimes_\lambda \mathrm{Sp}_{2n} \cong \mathbb{G}_m \times \mathrm{Sp}_{2n}.$$

(2) *The dual group $H_\rho^\vee$ is given by*

$$H_\rho^\vee \cong \mathbb{G}_m \times \mathrm{SO}_{2n+1}.$$

(3) *The representation ρ_{fp} is given by*

$$\rho_{\mathrm{fp}} = \mathrm{st}_{\mathbb{G}_m} \otimes \mathrm{st}_{\mathrm{SO}_{2n+1}},$$

where $\mathrm{st}_{\mathbb{G}_m}$ is the standard representation of $\mathbb{G}_m$ and $\mathrm{st}_{\mathrm{SO}_{2n+1}}$ is the standard presentation of SO_{2n+1}.

(4) *The generalized determinant $\mathfrak{d}_{\mathrm{fp}} : H_\rho \to \mathbb{G}_m$ is given by*

$$H_\rho \cong \mathbb{G}_m \times \mathrm{Sp}_{2n} \to \mathbb{G}_m : (a, g) \mapsto a.$$

This example explains that in [8], the authors must take the group $\mathbb{G}_m \times \mathrm{Sp}_{2n}$ for the local theory of the Braverman–Kazhdan–Ngô proposal in the case of the standard L-functions of Sp_{2n}.

Acknowledgement

We would like to thank Shengkai Mao for numerous helpful discussions. We are also very grateful to X. Griffin Wang for providing a helpful reference that improves the proof of Theorem 1.4. Finally, we express our sincere appreciation to the referee for valuable comments and suggestions, which have greatly improved the clarity and quality of this paper.

References

[1] J. D. Adler and D. Prasad, "Multiplicity upon restriction to the derived subgroup", *Pacific J. Math.* **301**:1 (2019), 1–14. MR

[2] A. Borel, "Automorphic L-functions", pp. 27–61 in *Automorphic forms, representations and L-functions* (Corvallis, OR, 1977), Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, RI, 1979. MR

[3] A. Bouthier, B. C. Ngô, and Y. Sakellaridis, "On the formal arc space of a reductive monoid", *Amer. J. Math.* **138**:1 (2016), 81–108. MR

[4] A. Braverman and D. Kazhdan, "γ-functions of representations and lifting", *Geom. Funct. Anal.* (2000), 237–278. MR

[5] Y. S. D. Ben-Zvi and A. Venkatesh, "Relative Langlands duality", preprint, 2024. arXiv 2409.04677

[6] R. Godement and H. Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Mathematics **260**, Springer, 1972. MR

[7] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, 1977. MR

[8] D. Jiang, Z. Luo, and L. Zhang, *Harmonic analysis and gamma functions on symplectic groups*, Mem. Amer. Math. Soc. **1473**, American Mathematical Society, 2024. MR

[9] R. P. Langlands, "Problems in the theory of automorphic forms", pp. 18–61 in *Lectures in modern analysis and applications*, Lecture Notes in Math. **170**, Springer, 1970. MR

[10] Z. Luo, "An introduction to the proposal of Braverman and Kazhdan", pp. 379–436 in *On the Langlands program*: *endoscopy and beyond*, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. **43**, World Scientific, Hackensack, NJ, 2024. MR

[11] Z. Luo and N. B. Chau, "Nonabelian Fourier Kernels on SL_2 and GL_2", 2024. arXiv 2409.14696

[12] B. C. Ngô, "On a certain sum of automorphic L-functions", pp. 337–343 in *Automorphic forms and related geometry*: *assessing the legacy of I. I. Piatetski-Shapiro*, Contemp. Math. **614**, Amer. Math. Soc., Providence, RI, 2014. MR

[13] B. C. Ngô, "Hankel transform, Langlands functoriality and functional equation of automorphic L-functions", *Jpn. J. Math.* **15**:1 (2020), 121–167. MR

[14] B. C. Ngô, Letter to Y. Sakellaridis on semigroups and basic functions, available at https://personal.math.ubc.ca/~cass/beijing/pdf/ngo-semigroup-test.pdf.

[15] M. S. Putcha, "Reductive groups and regular semigroups", *Semigroup Forum* **30**:3 (1984), 253–261. MR

[16] M. S. Putcha, *Linear algebraic monoids*, London Mathematical Society Lecture Note Series **133**, Cambridge University Press, 1988. MR

[17] L. E. Renner, "Reductive monoids are von Neumann regular", *J. Algebra* **93**:2 (1985), 237–245. MR

[18] L. E. Renner, "Classification of semisimple algebraic monoids", *Trans. Amer. Math. Soc.* **292**:1 (1985), 193–223. MR

[19] L. E. Renner, "Classification of semisimple varieties", *J. Algebra* **122**:2 (1989), 275–287. MR

[20] L. E. Renner, "Algebraic varieties and semigroups", pp. 81–101 in *The analytical and topological theory of semigroups*, Expositions in Mathematics **1**, de Gruyter, Berlin, 1990. MR

[21] L. E. Renner, *Linear algebraic monoids*, Encyclopaedia of Mathematical Sciences **134**, Springer, Berlin, 2005. MR

[22] A. Rittatore, "Very flat reductive monoids", *Publ. Mat. Urug.* **9** (2001), 93–121. MR

[23] F. Shahidi, "On non-vanishing of twisted symmetric and exterior square L-functions for $\mathrm{GL}(n)$", *Pacific J. Math.* (1997), 311–322. MR

[24] F. Shahidi, "Local factors, reciprocity and Vinberg monoids", preprint, 2017. arXiv 1710.04285

[25] F. Shahidi and W. Sokurski, "On the resolution of reductive monoids and multiplicativity of γ-factors", *J. Number Theory* **240** (2022), 404–438. MR

[26] M. Solleveld, "Langlands parameters, functoriality and Hecke algebras", *Pacific J. Math.* **304**:1 (2020), 209–302. MR

[27] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progress in Mathematics **9**, Birkhäuser, 1998. MR

[28] J. T. Tate, "Fourier analysis in number fields, and Hecke's zeta-functions", pp. 305–347 in *Algebraic number theory* (Brighton, 1965), Academic Press, London, 1967. MR

[29] E. B. Vinberg, "On reductive algebraic semigroups", pp. 145–182 in *Lie groups and Lie algebras*: *E. B. Dynkin's Seminar*, Amer. Math. Soc. Transl. Ser. 2 **169**, Amer. Math. Soc., Providence, RI, 1995. MR

Received December 12, 2024. Revised August 4, 2025.

DIHUA JIANG
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455
UNITED STATES
jiang034@umn.edu

ZHAOLIN LI
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455
UNITED STATES
li001870@umn.edu

GUODONG XI
SCHOOL OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455
UNITED STATES
xi000023@umn.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.339

ARITHMETIC OF CUTS IN ORDERED ABELIAN GROUPS AND OF IDEALS OVER VALUATION RINGS

FRANZ-VIKTOR KUHLMANN AND KATARZYNA KUHLMANN

We investigate the existence, uniqueness and maximality of solutions T for equations $S_1 + T = S_2$ and inequalities $S_1 + T \subseteq S_2$, where S_1 and S_2 are final segments of ordered abelian groups. Since cuts are determined by their upper cut sets, which are final segments, this gives information about the corresponding equalities and inequalities for cuts. We apply our results to investigate the existence, uniqueness and maximality of solutions J for equations $I_1 J = I_2$ and inequalities $I_1 J \subseteq I_2$, where I_1 and I_2 are ideals of valuation rings. This enables us to compute the annihilators of quotients of the form I_1/I_2.

1. Introduction

For the notions and notation we use in this paper, see Sections 2.1, 2.2, 3.1 and 3.2.

One main purpose of this article is to address a problem that comes up in the papers [3; 2]. There we work with valued field extensions $(L|K, v)$ of prime degree with unique extension of the valuation v from K to L. We show that the module of relative differentials (Kähler differentials) for the extension $\mathcal{O}_L|\mathcal{O}_K$ of the respective valuation rings is isomorphic, as $\mathcal{O}_L$-module, to a quotient U/UV of ideals U and V of $\mathcal{O}_L$. We wish to compute the annihilator of such a quotient U/UV. It will contain V, but there are cases where it is bigger.

In the present article, we will consider the more general case of quotients I_1/I_2 of $\mathcal{O}_v$-ideals I_1, I_2 (see Section 3.5). To this end, we consider equations $I_1 J = I_2$ and inequalities $I_1 J \subseteq I_2$, as we will describe in detail below.

Our approach is to translate questions about ideals to corresponding questions about final segments in the value group vK of (K, v). The latter are connected with the investigation of cuts in ordered abelian groups which has appeared at several points in the literature. For a survey and a list of references, see [8]. The present paper uses, and improves on, parts of the unpublished manuscript [9]. Let us describe the background on cuts first.

We thank the referee for many very helpful corrections and comments.

MSC2020: primary 06F20, 13F30; secondary 12J25, 13A15.

Keywords: ordered abelian group, cut, final segment, valuation ring, ideal, quotients of ideals over valuation rings, annihilator.

© 2026 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Take an ordered abelian group $(G, \leq)$. By a *cut* in $(G, \leq)$ we mean a pair

$$\Lambda = (\Lambda^L, \Lambda^R),$$

where Λ^L is an initial segment of G and Λ^R is a final segment of G such that $\Lambda^L \cup \Lambda^R = G$ and $\Lambda^L \cap \Lambda^R = \varnothing$. We call Λ^L the *lower cut set* of Λ, and Λ^R the *upper cut set* of Λ. For basic information on cuts and final segments, see Sections 2.1 and 2.2. *When we speak of final segments we will always assume that they are proper nonempty subsets of G.*

In the present paper we are particularly interested in the *addition of cuts*. The two obvious ways to define the sum $\Lambda_1 + \Lambda_2$ of two cuts Λ_1 and Λ_2 are following:

(1) the lower cut set of $\Lambda_1 + \Lambda_2$ is $\Lambda_1^L + \Lambda_2^L$, and

(2) the upper cut set of $\Lambda_1 + \Lambda_2$ is $\Lambda_1^R + \Lambda_2^R$,

where for any $S_1, S_2 \subseteq G$, $S_1 + S_2 := \{\alpha + \beta \mid \alpha \in S_1, \beta \in S_2\}$. The two additions usually do not yield the same results, but their properties are very similar. Under both additions, the set of all cuts in G is a commutative monoid. However, in general it is not a group, since for fixed Λ the function

$$\Lambda' \mapsto \Lambda + \Lambda'$$

may not be injective. There may even be idempotents; see Section 2.2.

In the present paper, we will exclusively work with the second definition, which we call the *upper cut set addition*. Every cut is uniquely determined by its upper cut set, so instead of cuts we will just work with final segments (for a reason that will be explained later). The set of final segments of G is linearly ordered by inclusion, and unions, intersections and sums of any collections of final segments are again final segments, as long as they are proper nonempty subsets of G.

From our observations about cut addition, the following questions arise. Take final segments S_1 and S_2 of G.

(QFS1) *Is there a final segment T of G such that*

$$S_1 + T = S_2 \tag{1}$$

holds?

(QFS2) *If yes, is T uniquely determined*? *If so, determine it.*

(QFS3) *If T exists but is not uniquely determined, determine the largest final segment $T_{\max}$ such that $S_1 + T_{\max} = S_2$.*

(QFS4) *If T does not exist, compute the largest final segment $T_{\max}$ such that $S_1 + T_{\max} \subset S_2$.*

One aim of this paper is to answer these questions, which will be done in Section 2.7.

Obviously,

$$S_2 -_{\mathrm{fs}} S_1 := \{\alpha \in G \mid \alpha + S_1 \subseteq S_2\} \tag{2}$$

is the largest solution $T_{\max}$ for the inequality $S_1 + T \subseteq S_2$; hence if a solution of (1) exists, then $S_2 -_{\mathrm{fs}} S_1$ is the largest. However, in the literature there does not seem to exist any recipe for computing $S_2 -_{\mathrm{fs}} S_1$ in terms of (elementwise) sums and differences of final segments derived from S_1 and S_2 that would help solve the problems that came up in the papers [3; 2] and which we will explain below.

A key to the answers to the above questions is the notion of *invariance group* of a final segment, which always is a convex subgroup of G (see Section 2.4). This sheds light on the important role that convex subgroups play in nonarchimedean ordered groups for the answers to the above questions. The case of archimedean ordered groups, which have no proper nontrivial convex subgroups, is significantly easier. Our answers to the above questions are given in Theorems 2.33 and 2.34.

Our interest in these questions arose from our work with cuts and its applications to ordered abelian groups and ordered fields; see the survey article [8], which lists many references to background literature. Recently, it gained even more importance for us because of applications to the theory of ideals over valuation rings.

Take a valued field (K, v) with value group vK and valuation ring $\mathcal{O}_v$. For $a \in K$ we write va in place of $v(a)$. When we talk of $\mathcal{O}_v$-*ideals* we will always assume them to be nonzero, and we also include fractional ideals; in other words, *we talk about nonzero* $\mathcal{O}_v$-*modules* $I \subsetneq K$. Just like the set of final segments of vK, the set of $\mathcal{O}_v$-ideals is linearly ordered by inclusion, and unions, intersections and products of any collections of $\mathcal{O}_v$-ideals are again $\mathcal{O}_v$-ideals, as long as they are nonzero $\mathcal{O}_v$-modules properly contained in K. The function

$$v : I \mapsto vI := \{vb \mid 0 \neq b \in I\} \tag{3}$$

is an order-preserving bijection from the set of all $\mathcal{O}_v$-ideals onto the set of all final segments of vK: $J \subseteq I$ holds if and only if $vJ \subseteq vI$ holds. The inverse of this function is the order-preserving function

$$S \mapsto I_S := (a \in K \mid va \in S) = \{a \in K \mid va \in S\} \cup \{0\}. \tag{4}$$

Further, the function (3) is a homomorphism from the multiplicative monoid of $\mathcal{O}_v$-ideals onto the additive monoid of final segments:

$$vIJ = vI + vJ \quad \text{and} \quad I_{S+S'} = I_S I_{S'}. \tag{5}$$

Via the function (3), the following questions translate to questions (QFS1)–(QFS4). Take nonzero $\mathcal{O}_v$-ideals I_1 and I_2.

(QID1) *Is there an* $\mathcal{O}_v$-*ideal* J *such that*

$$I_1 J = I_2 \tag{6}$$

holds?

(QID2) *If yes, is J uniquely determined*? *If so, compute it.*

(QID3) *If J exists but is not uniquely determined, compute the largest $\mathcal{O}_v$-ideal $J_{\max}$ such that $I_1 J_{\max} = I_2$.*

(QID4) *If J does not exist, compute the largest $\mathcal{O}_v$-ideal $J_{\max}$ such that $I_1 J_{\max} \subset I_2$.*

We will derive answers to these questions in Section 3.4, using the results of Section 2.7. Here, a key role is played by the notion of the *invariance valuation ring* of an ideal I, which corresponds to the invariance group of the final segment vI.

Obviously,

$$I_2 : I_1 := \{a \in K \mid aI_1 \subseteq I_2\} \tag{7}$$

is the largest solution $J_{\max}$ for the inequality $I_1 J \subseteq I_2$. Again, in the literature there does not seem to exist any recipe for computing $I_1 : I_2$ from I_1 and I_2 that would help solve the problems addressed in Section 3.5. This section is devoted to the computation of the annihilators $I_2 : I_1$ of quotients I_1/I_2 of $\mathcal{O}_v$-ideals I_1, I_2 with $I_2 \subseteq I_1$. The results are then applied in Section 3.6 to compute the annihilators of some special quotients that appear in [3] and [2] and to answer the question in what cases they are equal to the maximal ideal $\mathcal{M}_v$ of $\mathcal{O}_v$. Our answers to the above questions are given in Theorems 3.13 and 3.14.

Ideals over valuation domains have been studied extensively; see for instance [7, Chapter II, §4]. This article lists a number of properties of such ideals; in Section 3.3 we present the corresponding properties of final segments of the value group and use them to give alternate proofs for the properties of ideals I by deriving them from the properties of the final segments vI.

Equations of the form $I_1 J = I_2$ are implicitly related to groups in the semigroups formed by the isomorphy classes of the ideals, which for instance are studied in [1]. Further, the connection between ideals of valuation rings and the final segments of value groups is described and used in [1], and it is pointed out that something similar was already done by Paulo Ribenboim in [11]. In fact, Ribenboim arrives at the same criteria that answer questions (QFS1) and (QID1) as we give in Theorems 2.33 and 3.13. Also invariance groups appear in [11], in a somewhat disguised form. However, the machinery used by Ribenboim and cited in [1] is heavy. We show in this paper that embedding the value group in a Hahn product and even working with the natural valuation of an ordered abelian group is not necessary. The fact that final segments in ordered abelian groups of higher rank may not have infima in suitably defined extensions of the group may be scary, but it really is no obstacle.

2. Final segments

2.1. ***Preliminaries on cuts and final segments.*** Take $(\Gamma, \leq)$ to be the underlying ordered set of any nonzero ordered abelian group (*by "ordered", we will always mean "totally ordered"*). Note that $(\Gamma, \leq)$ does not have a smallest or largest element, and the definitions in this section also work without the presence of the group structure. If M_1, M_2 are nonempty subsets of Γ and $\alpha \in \Gamma$, we will write $\alpha < M_2$ if $\alpha < \beta$ for all $\beta \in M_2$, and we will write $M_1 < M_2$ if $\alpha < M_2$ for all $\alpha \in M_1$. Similarly, we use the relations $>$, $\leq$ and $\geq$ in place of $<$.

An ordered set $(\Gamma, \leq)$ is *discretely ordered* if every element has an immediate successor, and it is *densely ordered* if for every $\alpha, \beta \in \Gamma$ with $\alpha < \beta$ there is $\gamma \in \Gamma$ with $\alpha < \gamma < \beta$.

A subset M of Γ is called *convex in* $(\Gamma, \leq)$ if for all $\alpha, \beta \in M$ and all $\gamma \in \Gamma$ such that $\alpha \leq \gamma \leq \beta$, it follows that $\gamma \in M$. A subset S of Γ is an *initial segment of* Γ if for every $\alpha \in S$ and every $\gamma \in \Gamma$ with $\gamma \leq \alpha$, it follows that $\gamma \in S$. Symmetrically, S is a *final segment of* Γ if for every $\alpha \in S$ and every $\gamma \in \Gamma$ with $\gamma \geq \alpha$, it follows that $\gamma \in S$. *Throughout, we will assume initial and final segments to be nonempty and not equal to* Γ. Then S is a final segment of Γ if and only if $\Gamma \setminus S$ is an initial segment of Γ. We will denote the set of all final segments of $(\Gamma, \leq)$ by

$$\Gamma^{\uparrow}.$$

If $\varnothing \neq M_1, M_2 \subsetneq \Gamma$ are such that $M_1 \leq M_2$ and $\Gamma = M_1 \cup M_2$, then we call (M_1, M_2) a *quasicut* in Γ, and we write $\Lambda^L = M_1$, $\Lambda^R = M_2$, and $\Lambda = (\Lambda^L, \Lambda^R)$. It follows that M_1 is an initial segment of Γ, M_2 is a final segment of Γ, and the intersection of M_1 and M_2 consists of at most one element. If this intersection is empty, then (M_1, M_2) is called a (*Dedekind*) *cut* in Γ.

For any nonempty subset $M \subseteq \Gamma$, we set

$$M^+ := \{\gamma \in \Gamma \mid \gamma > M\}.$$

If this is nonempty, it is the largest final segment having empty intersection with M. Similarly, we set

$$M^- := \{\gamma \in \Gamma \mid \exists m \in M : \gamma \geq m\}.$$

That is, if $M^- \neq \Gamma$, then it is the smallest final segment of Γ which contains M.

We will write γ^+ instead of $\{\gamma\}^+$ and γ^- instead of $\{\gamma\}^-$. A final segment of the form γ^- will be called *principal*; it has smallest element γ. The principal final segments are exactly the final segments that contain smallest elements. Analogously, we define an initial segment to be *principal* if it contains a largest element.

For every subset M of Γ, we will denote by M^c its complement in Γ, i.e.,

$$M^c = \Gamma \setminus M.$$

We note that if $S \subsetneq \Gamma$ is a final segment of Γ, then $S^c = \{\gamma \in \Gamma \mid \gamma < S\}$ is an initial segment and (S^c, S) is a cut.

The proof of the following lemma is straightforward.

Lemma 2.1. *Take a final segment S of the ordered set Γ.*

(1) *If S is nonprincipal but has infimum γ in Γ, then S^c is principal with largest element γ.*

(2) *We have that S^c has a supremum if and only if S has an infimum in Γ.*

(3) *If Γ is densely ordered, then an element $\gamma \in \Gamma$ is an infimum of S if and only if it is a supremum of S^c.* □

Every nonempty ordered set $(\Gamma, \leq)$ without smallest or largest element carries a topology with the open intervals (α, β), $\alpha, \beta \in \Gamma$ with $\alpha < \beta$, as its basic open sets. We leave the proof of the following lemma to the reader.

Lemma 2.2. *Take a final segment S in Γ.*

(1) *If S is principal, then it is closed, and it is clopen if and only if Γ is discretely ordered.*

(2) *If S does not have an infimum in Γ, then it is clopen.*

(3) *If S is nonprincipal but has an infimum in Γ, then S is open, but not closed.*

(4) *The final segment S is closed if and only if S is principal or does not have an infimum in Γ.* □

Given a final segment S in Γ, we set

$$\widehat{S} := \begin{cases} S \cup \{\gamma\} & \text{if } S \text{ has infimum } \gamma \text{ in } \Gamma, \\ S & \text{otherwise.} \end{cases}$$

Then $\widehat{S}$ is the closure of S in Γ. We will also write $S\widehat{\ }$ for $\widehat{S}$ in case the expression S is too long to accommodate a hat.

The function

$$\Gamma \ni \gamma \mapsto \gamma^- \in \Gamma^\uparrow \tag{8}$$

is an embedding of $(\Gamma, \leq)$ in $(\Gamma^\uparrow, \subseteq)$. Also sending γ to γ^+ produces such an embedding, but throughout this paper we will work with the embedding (8).

2.2. ***Final segments in ordered abelian groups.*** From now on, we will work with ordered abelian groups $(G, \leq)$ and their final segments. The definitions and results of the previous section remain valid for $\Gamma = G$. We note that $(G, \leq)$ is discretely ordered if and only if it contains a smallest positive element, and it is densely ordered otherwise.

If M, M_1 and M_2 are nonempty subsets of G, then we set $M_1+M_2=\{\alpha+\beta \mid \alpha \in M_1, \beta \in M_2\}$ and $M_1-M_2=\{\alpha-\beta \mid \alpha \in M_1, \beta \in M_2\}$. From this, the sets $\alpha+M$, $\alpha-M$ and $M-\alpha$ are obtained via replacing α by $\{\alpha\}$. We set $-M:=0-M$.

If S_1 and S_2 are final segments, then so is S_1+S_2. With this addition, $G\uparrow$ is a commutative monoid with neutral element 0^-. It is in general not a group and can even contain idempotents other than 0^-; see Lemma 2.3(1) and Lemma 2.6(3). For example, the lexicographically ordered group $G=\mathbb{Q}\times\mathbb{Z}$ has the proper nontrivial subgroup $H=\{0\}\times\mathbb{Z}$, which gives rise to two idempotent final segments. Indeed, since $H+H=H$, both

$$\{\alpha \in G \mid \forall \beta \in H : \alpha > \beta\} \quad \text{and} \quad \{\alpha \in G \mid \exists \beta \in H : \alpha \geq \beta\}$$

are idempotent, and it can be shown that they are the only idempotent final segments in G.

For $\Gamma = G$ the embedding (8) is a homomorphism, that is,

$$(\gamma_1+\gamma_2)^- = \gamma_1^- + \gamma_2^-. \tag{9}$$

We note that

$$\gamma^- + 0^+ = \gamma^+. \tag{10}$$

Further, the reader may prove:

Lemma 2.3. (1) *G is densely ordered if and only if $0^+ + 0^+ = 0^+$.*

(2) *Assume that G is discretely ordered. If the final segment S of G has an infimum, then this is the smallest element of S and S is principal.* □

Proposition 2.4. *Take final segments S_1 and S_2 of G.*

(1) *The sum of two final segments is again a final segment. The sum of two principal final segments is again a principal final segment. The sum of two final segments of which at least one is nonprincipal is a nonprincipal final segment.*

(2) *If S_2 is principal but S_1 is not, then* (1) *has no solution.*

(3) *If S_1 has infimum α_1 and S_2 has infimum α_2, then S_1+S_2 has infimum $\alpha_1+\alpha_2$.*

(4) *If S is a final segment of G and $\alpha \in G$, then*

$$\alpha + S = \alpha^- + S. \tag{11}$$

Proof. The proof of parts (1) and (4) are straightforward. Part (2) follows immediately from part (1).

(3) If both S_1 and S_2 are principal, then the assertion follows from (9). If at least one of them is nonprincipal, then by Lemma 2.3(2), G must be densely ordered. Now we have to consider the following two cases (up to symmetry). If $S_1=\alpha_1^-$

and $S_2 = \alpha_2^+$, then by (10) and (9),

$$S_1 + S_2 = \alpha_1^- + \alpha_2^+ = \alpha_1^- + \alpha_2^- + 0^+ = (\alpha_1 + \alpha_2)^- + 0^+ = (\alpha_1 + \alpha_2)^+. \quad (12)$$

If $S_1 = \alpha_1^+$ and $S_2 = \alpha_2^+$, then by (10) and (9) together with Lemma 2.3(1),

$$\begin{aligned} S_1 + S_2 &= \alpha_1^+ + \alpha_2^+ = \alpha_1^- + 0^+ + \alpha_2^- + 0^+ \\ &= (\alpha_1 + \alpha_2)^- + 0^+ = (\alpha_1 + \alpha_2)^+. \end{aligned}$$ □

Via the embedding (8) we can view $(G, \leq)$ as an ordered subgroup of the ordered monoid $(G^\uparrow, \subseteq)$, consisting of all principal final segments in $G^\uparrow$. From this, we obtain the following positive answers to questions (QFS1) and (QFS2) in the case of principal final segments. The proofs are straightforward.

Proposition 2.5. *Take elements α, α_1^- and α_2^- in G.*

(1) *The function*

$$G^\uparrow \ni T \mapsto \alpha^- + T \in G^\uparrow \quad (13)$$

is a bijection with inverse $G^\uparrow \ni T \mapsto (-\alpha)^- + T \in G^\uparrow$.

(2) *If S is any final segment of G, then the unique final segment T of G such that $\alpha^- + T = S$ is $T = S + (-\alpha)^- = S - \alpha$. Consequently, if S_1 and S_2 are final segments of G and S_1 is principal, say $S_1 = \alpha^-$, then* (1) *has the unique solution*

$$T = S_2 -_{\mathrm{fs}} S_1 = S_2 - \alpha. \qquad \square$$

Part (1) of this proposition shows that the function

$$G^\uparrow \ni T \mapsto S + T \in G^\uparrow \quad (14)$$

is injective if $S \in G^\uparrow$ is principal. Part (3) of the following Lemma shows that if S is nonprincipal, then the function (14) is never injective, provided that G is nontrivial.

Lemma 2.6. *Take final segments S and T of G.*

(1) *We have that $0^+ + T = \{\beta \in T \mid \alpha < \beta \text{ for some } \alpha \in T\}$.*

(2) *If there is $\alpha \in G$ such that $T = \alpha^-$, then $0^+ + T = \alpha^+$.*

(3) *If T is nonprincipal, then $0^+ + T = T$.*

(4) *If S is nonprincipal, then $S + T = S + \widehat{T}$.*

Proof. The proof of statements (1) and (2) is straightforward. If T is nonprincipal, then for every $\beta \in T$ there is some $\alpha \in T$ such that $\alpha < \beta$. Therefore, part (3) follows from part (1).

(4) Since S is nonprincipal, part (3) yields $S = S + 0^+$. If T is closed in G, then $\widehat{T} = T$ and there is nothing to show. If T is not closed in G, then by part (4) of Lemma 2.2, T has an infimum $\alpha \in G \setminus T$. Thus, $\widehat{T} = \alpha^-$ and $0^+ + \widehat{T} = \alpha^+ = T$, so that $S + \widehat{T} = S + 0^+ + \widehat{T} = S + T$. □

Corollary 2.7. *For every $\alpha \in G$ and nonprincipal final segment T of G,*

$$\alpha + T = \alpha^- + T = \alpha^+ + T. \tag{15}$$

Hence the function

$$T \mapsto \alpha^+ + T \tag{16}$$

is a bijection on the set of all nonprincipal final segments of G, and there it is equal to the function (13).

Proof. The first equality of (15) follows from (11). The second equality holds since $\alpha^- + T = \alpha^- + (0^+ + T) = (\alpha^- + 0^+ + T) = \alpha^+ + T$ by Lemma 2.6(3) and equation (10). The final statement follows from the second equality of (15) together with Proposition 2.5(1) and Proposition 2.4(1). □

For the conclusion of this section, we note the following fact; its proof is straightforward. If $M \subseteq G$ and $\alpha \in G$, then

$$\alpha^- + M = (\alpha + M)^-. \tag{17}$$

2.3. ***Subtraction of final segments.*** The notion of subtraction of Dedekind cuts has been defined in articles in the literature, such as [5] and [6]. When upper cut set addition is used, then for two Dedekind cuts $\Lambda_1 = (\Lambda_1^L, \Lambda_1^R)$ and $\Lambda_2 = (\Lambda_2^L, \Lambda_2^R)$, their difference $\Lambda_2 - \Lambda_1$ has been defined to be the cut with upper cut set $\Lambda_2^R - \Lambda_1^L$, which is the sum of the final segments Λ_2^R and $-\Lambda_1^L$ and thus again a final segment.

Analogously, we can define a difference of two final segments S_2 and S_1 of G as

$$S_2 - S_1^c = S_2 + (-S_1^c) = S_2 + \{\gamma \in G \mid \gamma > -S_1\}. \tag{18}$$

Note that $(-S)^c = -(S^c)$ since $\gamma \in (-S)^c \Leftrightarrow \gamma > -S \Leftrightarrow -\gamma < S \Leftrightarrow -\gamma \in S^c \Leftrightarrow \gamma \in -(S^c)$. Therefore we will omit the brackets and just write "$-S^c$".

Definition (18) has for instance been used in [5; 6]. However, it does not extend the difference operation of the embedded copy of G. Indeed, for every $\alpha \in G$ we have

$$\alpha^- - (\alpha^-)^c = \alpha^- + \{\gamma \in G \mid \gamma > -\alpha^-\} = \alpha^- + \{\gamma \in G \mid \gamma > -\alpha\} = 0^+.$$

In order to obtain a difference operation that extends the one of G, replacing "$>$" by "$\geq$" in (18) we define the (*modified*) *final segment subtraction* of two final segments S_2 and S_1 as

$$S_2 -_{\mathrm{mc}} S_1 := S_2 + \{\gamma \in G \mid \gamma \geq -S_1\}.$$

In other words, we are replacing the complement S_1^c in (18) by the "modified complement" $\{\gamma' \in G \mid \gamma' \leq S_1\}$. This subtraction extends the subtraction of G: for arbitrary $\alpha, \beta \in G$,

$$\alpha^- -_{\mathrm{mc}} \beta^- = \alpha^- + \{\gamma \mid \gamma \geq -\beta^-\} = \alpha^- + \{\gamma \mid \gamma \geq -\beta\} = (\alpha - \beta)^- . \quad (19)$$

To simplify notation, for a final segment S of G we set

$$\Delta S := \{\gamma \in G \mid \gamma \geq -S\} = \widehat{-S^c}.$$

Lemma 2.8. (1) *For each $\alpha \in G$,*

$$\Delta\alpha^- = (-\alpha)^- . \quad (20)$$

(2) *If S is a nonprincipal final segment of G, then $\Delta S = -S^c$.*

(3) *For every final segment S of G, we have $\Delta(\Delta S) = \widehat{S}$.*

Proof. (1) We have $\Delta\alpha^- = \{\gamma \in G \mid \gamma \geq -\alpha^-\} = \{\gamma \in G \mid \gamma \geq -\alpha\} = (-\alpha)^-$.

(2) If S is a nonprincipal final segment of G, then $\gamma \geq -S \Leftrightarrow -\gamma \leq S \Leftrightarrow -\gamma < S \Leftrightarrow -\gamma \in S^c \Leftrightarrow \gamma \in -S^c$, so $\Delta S = -S^c$.

(3) If $S = \alpha^-$ for some $\alpha \in G$, then $\Delta(\Delta S) = \Delta((-\alpha)^-) = \alpha^- = S = \widehat{S}$, where we have used part (1) twice. If S is a nonprincipal final segment of G, then $\Delta(\Delta S) = \Delta(-S^c) = (-(-S^c)^c)\widehat{\ } = (-(-S))\widehat{\ } = \widehat{S}$, where we have used part (2) and the definition of Δ. □

Remark 2.9. The final segment α^+ is nonprincipal if and only if G is densely ordered; if G has smallest positive element γ, then $\alpha^+ = (\alpha + \gamma)^-$. Therefore, we have

$$\Delta\alpha^+ = \begin{cases} (-\alpha)^- & \text{if } G \text{ is densely ordered,} \\ (-\alpha - \gamma)^- & \text{if } G \text{ is discretely ordered.} \end{cases}$$

Lemma 2.10. *Take final segments S_1 and S_2 of G.*

(1) *We have that*

$$S_2 -_{\mathrm{mc}} S_1 = S_2 + \Delta S_1 .$$

In particular, $0^- -_{\mathrm{mc}} S_1 = \Delta S_1$.

(2) *The sets ΔS_1, $S_2 -_{\mathrm{mc}} S_1$, $-S_1^c$ and $S_2 - S_1^c$ are again final segments of G.*

(3) *If S_1 is nonprincipal, then*

$$S_2 -_{\mathrm{mc}} S_1 = S_2 - S_1^c . \quad (21)$$

(4) *If S_1 is principal, say $S_1 = \alpha^-$ for some $\alpha \in G$, then*

$$S_2 -_{\mathrm{mc}} S_1 = S_2 + (-\alpha)^- = S_2 - \alpha .$$

(5) *The following holds*:

$$S_1 + (S_2 -_{\mathrm{mc}} S_1) = (S_1 -_{\mathrm{mc}} S_1) + S_2\,. \tag{22}$$

(6) *The following holds*:

$$S_2 -_{\mathrm{mc}} (0^- -_{\mathrm{mc}} S_1) = S_2 + \widehat{S_1}\,.$$

In particular,

$$0^- -_{\mathrm{mc}} (0^- -_{\mathrm{mc}} S_1) = \widehat{S_1}\,. \tag{23}$$

Proof. (1) This follows from the definitions of $S_2 -_{\mathrm{mc}} S_1$ and ΔS_1.

(2) For ΔS_1 this holds by definition, so for $S_2 -_{\mathrm{mc}} S_1$ it follows from part (1). For $-S_1^c$ we observe that S_1^c is an initial segment, hence $-S_1^c$ is a final segment and the same holds for $S_2 - S_1^c = S_2 + (-S_1^c)$.

(3) This follows from part (1) together with Lemma 2.8(2).

(4) This follows from part (1) together with (20) and Proposition 2.4(4).

(5) The proof is straightforward, using part (1).

(6) Using parts (1) and part (3) of Lemma 2.8, we compute

$$S_2 -_{\mathrm{mc}} (0^- -_{\mathrm{mc}} S_1) = S_2 -_{\mathrm{mc}} \Delta S_1 = S_2 + \Delta(\Delta S_1) = S_2 + \widehat{S_1}\,. \qquad \square$$

Remark 2.11. The final segment $S_2 -_{\mathrm{mc}} S_1$ is closed, except in the following cases. If $S_1 = \alpha^-$ and $S_2 = \beta^+$ for some $\alpha, \beta \in G$ and β has no immediate successor in G, then $\beta - \alpha$ has no immediate successor in G and therefore, $S_2 -_{\mathrm{mc}} S_1 = (\beta - \alpha)^+$ is not closed. If S_2 and S_1 have no infimum in G, then $S_2 -_{\mathrm{mc}} S_1$ may or may not have an infimum in G. For instance, if $G = \mathbb{Q}$ and $S_1 = S_2 = \{\alpha \in \mathbb{Q} \mid \alpha > \pi\}$, then $\Delta S_1 = \{\alpha \in \mathbb{Q} \mid \alpha > -\pi\}$ and $S_2 -_{\mathrm{mc}} S_1 = 0^+$ which is not closed although S_1 and S_2 are.

The question arises whether the difference operation $-_{\mathrm{mc}}$ provides some positive answer to our questions (QFS1) and (QFS2). As Proposition 2.5(2) in conjunction with Lemma 2.10(4) shows, it indeed gives a complete answer to these questions in the case where S_1 is principal. It now remains to deal with the case where S_1 is nonprincipal. This case is much more complex and we need some preparation.

2.4. ***Invariance groups.*** Throughout, we let G be an arbitrary ordered abelian group. For a subset M of G, we define its *invariance group* to be

$$\mathcal{G}(M) := \{\gamma \in G \mid M + \gamma = M\}.$$

Lemma 2.12. *Take any nonempty subset $M \subseteq G$.*

(1) *$\mathcal{G}(M)$ is a subgroup of G.*

(2) *We have*

$$\mathcal{G}(M) = \{\gamma \in G \mid M + \gamma \subseteq M \text{ and } M - \gamma \subseteq M\}. \tag{24}$$

(3) *The following holds*:

$$\mathcal{G}(M) = \mathcal{G}(-M) = \mathcal{G}(M^c).$$

(4) *We have* $0 \in M$ *if and only if* $\mathcal{G}(M) \subseteq M$.

(5) *If* M *is convex, then* $\mathcal{G}(M)$ *is a convex subgroup of* G.

(6) *If* M *has an infimum in* G, *then* $\mathcal{G}(M) = \{0\}$. *In particular, for every* $\alpha \in G$,

$$\mathcal{G}(\alpha^-) = \mathcal{G}(\alpha^+) = \mathcal{G}(\{\alpha\}) = \{0\}. \tag{25}$$

Proof. (1) Take $\alpha, \beta \in \mathcal{G}(M)$. Then $M+\alpha+\beta = M+\beta = M$, whence $\alpha+\beta \in \mathcal{G}(M)$. Further, $M - \alpha = M + \alpha - \alpha = M$, whence $-\alpha \in \mathcal{G}(M)$. This proves that $\mathcal{G}(M)$ is a group.

(2) The inclusion "$\subseteq$" in (24) follows from the definition of the invariance group and part (1) of our lemma. If $M + \gamma \subseteq M$ and $M - \gamma \subseteq M$, then $M + \gamma \subseteq M = M - \gamma + \gamma \subseteq M + \gamma$, whence $M + \gamma = M$. This proves the inclusion "$\supseteq$" in (24).

(3) The first equality holds since $\gamma \in \mathcal{G}(M) \Leftrightarrow -\gamma \in \mathcal{G}(M) \Leftrightarrow M - \gamma = M \Leftrightarrow -M + \gamma = -M \Leftrightarrow \gamma \in \mathcal{G}(-M)$, where we used that $\mathcal{G}(M)$ is a group.

To prove the second equality, we observe that if $\gamma \notin \mathcal{G}(M)$, then there exists some $\alpha \in M$ such that $\alpha + \gamma \notin M$, whence $\alpha + \gamma \in M^c$. Then $\alpha + \gamma - \gamma = \alpha \notin M^c$, so $\gamma \notin \mathcal{G}(M^c)$ by (24). This shows that $\mathcal{G}(M^c) \subseteq \mathcal{G}(M)$. The reverse inclusion follows by substituting M^c for M in our argument.

(4) If $0 \in M$, then $\mathcal{G}(M) = 0 + \mathcal{G}(M) \subseteq M + \mathcal{G}(M) = M$. The converse is trivial.

(5) By part (1), it suffices to show that $\mathcal{G}(M)$ is convex. Take $\alpha, \beta \in \mathcal{G}(M)$ and $\gamma \in G$ with $\alpha \leq \gamma \leq \beta$. Then for all $\delta \in M$, $\delta+\alpha \leq \delta+\gamma \leq \delta+\beta$ and $\delta+\alpha, \delta+\beta \in M$. Hence $\delta + \gamma \in M$ by convexity of M. This proves that $M + \gamma \subseteq M$. In the same way, we also obtain that $M - \gamma \subseteq M$ because $-\alpha, -\beta \in \mathcal{G}(M)$ (since $\mathcal{G}(M)$ is a group) and $-\beta \leq -\gamma \leq -\alpha$. By part (2) it follows that $\gamma \in \mathcal{G}(M)$.

(6) Assume that $\alpha \in G$ is the infimum of M in G. Then for every positive $\gamma \in G$ there is $\beta \in M$ such that $0 \leq \beta - \alpha < \gamma$. Hence $\beta - \gamma < \alpha$ so that $\beta - \gamma \notin M$ and thus $\gamma \notin \mathcal{G}(M)$ by (24). This proves that $\mathcal{G}(M) = \{0\}$. □

Observe that in the linear ordering given by inclusion on the set of all convex subgroups,

$$H_1 \cup H_2 = H_1 + H_2 = \max\{H_1, H_2\}$$

for every two convex subgroups H_1 and H_2 of G.

Lemma 2.13. (1) *If M_1, M_2 are arbitrary subsets of G, then*

$$\mathcal{G}(M_1) \cup \mathcal{G}(M_2) \subseteq \mathcal{G}(M_1 + M_2).$$

(2) *If both S_1, S_2 are initial segments or final segments of G, then*

$$\mathcal{G}(S_1 + S_2) = \mathcal{G}(S_1) + \mathcal{G}(S_2) = \mathcal{G}(S_1) \cup \mathcal{G}(S_2).$$

(3) *If $\alpha \in G$ and S is a final segment of G, then*

$$\mathcal{G}(\alpha + S) = \mathcal{G}(S).$$

(4) *If S is a final segment and H a convex subgroup of G, then $\mathcal{G}(H) = H$ and*

$$\mathcal{G}(S + H) = \mathcal{G}(S) \cup H.$$

(5) *For each $\alpha \in G$ and every convex subgroup H of G,*

$$\mathcal{G}(\alpha + H) = \mathcal{G}((\alpha + H)^-) = \mathcal{G}((\alpha + H)^+) = H.$$

(6) *For each $\alpha \in G$ and every final segment S of G, we have*

$$\alpha + S \subseteq S \Leftrightarrow \alpha \in \mathcal{G}(S)^- \text{ and } \alpha + S \subsetneq S \Leftrightarrow \alpha \in \mathcal{G}(S)^+.$$

Proof. (1) Take $\gamma_1 \in \mathcal{G}(M_1)$ and $\gamma_2 \in \mathcal{G}(M_2)$. Then for all $\alpha_1 \in M_1$ and $\alpha_2 \in M_2$, $(\alpha_1 + \alpha_2) + \gamma_1 = (\alpha_1 + \gamma_1) + \alpha_2 \in M_1 + M_2$ and $(\alpha_1 + \alpha_2) + \gamma_2 = \alpha_1 + (\alpha_2 + \gamma_2) \in M_1 + M_2$. Hence, $\gamma_1, \gamma_2 \in \mathcal{G}(M_1 + M_2)$.

(2) In view of part (1), it suffices to show that $\mathcal{G}(S_1 + S_2) \subseteq \mathcal{G}(S_1) \cup \mathcal{G}(S_2)$. Let S_1 and S_2 be final segments of G; the proof for initial segments is similar. Take a positive $\gamma \in G$ such that $\gamma \notin \mathcal{G}(S_1) \cup \mathcal{G}(S_2)$. Then there are $\alpha_1 \in S_1$ and $\alpha_2 \in S_2$ such that $\alpha_1 - \gamma \notin S_1$ and $\alpha_2 - \gamma \notin S_2$. Since S_1 and S_2 are final segments of G, this means that $\alpha_1 - \gamma < S_1$ and $\alpha_2 - \gamma < S_2$. Thus $\alpha_1 + \alpha_2 - 2\gamma < S_1 + S_2$, which means that $2\gamma \notin \mathcal{G}(S_1 + S_2)$ and hence $\gamma \notin \mathcal{G}(S_1 + S_2)$.

(3) We compute $\mathcal{G}(\alpha + S) = \mathcal{G}(\alpha^- + S) = \mathcal{G}(\alpha^-) \cup \mathcal{G}(S) = \{0\} \cup \mathcal{G}(S) = \mathcal{G}(S)$, where we have used part (2) of our lemma together with (11) and (25)

(4) If H' is any convex subgroup of G, then it is comparable to H by inclusion. Hence $H + H' = H$ holds if and only if $H' \subseteq H$. This shows that $\mathcal{G}(H) = H$. Thus we obtain from part (1) that $\mathcal{G}(S) \cup H = \mathcal{G}(S) \cup \mathcal{G}(H) \subseteq \mathcal{G}(S + H)$. Take a positive $\gamma \in G$ such that $\gamma \notin \mathcal{G}(S) \cup H$. Then there are $\alpha_1 \in S$ and $\alpha_2 \in H$ such that $\alpha_1 - \gamma < S$ and $\alpha_2 - \gamma < H$. Thus $\alpha_1 + \alpha_2 - 2\gamma < S + H$, which means that $2\gamma \notin \mathcal{G}(S + H)$ and hence $\gamma \notin \mathcal{G}(S + H)$.

(5) We have $\mathcal{G}(\alpha + H) = \{\gamma \in G \mid \gamma + \alpha + H = \alpha + H\} = \{\gamma \in G \mid \gamma + H = H\} = H$. By (17), $(\alpha + H)^- = \alpha^- + H$. Hence, by part (4) and (25) of Lemma 2.12, $\mathcal{G}((\alpha + H)^-) = \mathcal{G}(\alpha^- + H) = \mathcal{G}(\alpha^-) + H = H$. We observe that the complement

of $-(\alpha+H)^+$ in G is $(-\alpha+H)^-$. Applying Lemma 2.12(3) twice, we obtain $\mathcal{G}((\alpha+H)^+) = \mathcal{G}(-(\alpha+H)^+) = \mathcal{G}((-\alpha+H)^-) = H$.

(6) Since $\mathcal{G}(S)$ is a convex subgroup of G, $\mathcal{G}(S)^-$ is the disjoint union of $\mathcal{G}(S)$ and $\mathcal{G}(S)^+$. If $\alpha \in \mathcal{G}(S)$, then $\alpha + S = S$. If $\alpha \in \mathcal{G}(S)^+$, then $\alpha > 0$ and therefore, $\alpha + S \subseteq S$; but as $\alpha \notin \mathcal{G}(S)$, $\alpha + S \neq S$. This proves the implications "$\Leftarrow$" in both statements.

If $\alpha \notin \mathcal{G}(S)^-$, then $\alpha + S \neq S$ since $\alpha \notin \mathcal{G}(S)$, hence $S \subsetneq \alpha + S$ since $\alpha < 0$. If $\alpha \notin \mathcal{G}(S)^+$, then $\alpha \in \mathcal{G}(S)$ and $\alpha + S = S$, or $\alpha < 0$ and $S \subseteq \alpha + S$. This proves the implications "$\Rightarrow$" in both statements. □

For a final segment S, we define

$$nS := \{\alpha_1 + \ldots + \alpha_n \mid \alpha_i \in S\}.$$

Note that the set $\{n\alpha \mid \alpha \in S\}$ is a final segment if and only if G is n-divisible; in general, nS is the smallest final segment that contains $\{n\alpha \mid \alpha \in S\}$. By a repeated application of Lemma 2.13(2), we obtain:

Corollary 2.14. *Take a final segment S of G. Then*

$$\mathcal{G}(nS) = \mathcal{G}(S).$$

From Lemma 2.13(2) and the fact that the convex subgroups of G are linearly ordered by inclusion, we obtain:

Proposition 2.15. *Take final segments S_1, S_2 and T such that $S_1 + T = S_2$. Then $\mathcal{G}(S_1) \subseteq \mathcal{G}(S_1) \cup \mathcal{G}(T) = \mathcal{G}(S_2)$, and if $\mathcal{G}(S_1) \subsetneq \mathcal{G}(S_2)$, then $\mathcal{G}(T) = \mathcal{G}(S_2)$.*

Lemma 2.16. *Take a final segment S of G.*

(1) *The following holds*:

$$\mathcal{G}(-S) = \mathcal{G}(S) = \mathcal{G}(S^c) = \mathcal{G}(\Delta S).$$

(2) *If $\mathcal{G}(S) \neq \{0\}$ then S is nonprincipal.*

Proof. (1) The first and second equality are special cases of Lemma 2.12(3). The equality $\mathcal{G}(S) = \mathcal{G}(\Delta S)$ is seen as follows. If S is nonprincipal, then by Lemma 2.8(2), $\Delta S = -S^c$ and the equality follows from what we have already proved. If S is principal, say $S = \alpha^-$ for some $\alpha \in G$, then $\Delta S = (-\alpha)^-$ by (20), whence $\mathcal{G}(\Delta S) = \{0\} = \mathcal{G}(S)$ by (25).

(2) If $\mathcal{G}(S) \neq \{0\}$, then there is a negative element $\alpha \in \mathcal{G}(S)$ so that for every $\beta \in S$, $\beta > \alpha + \beta \in S$; hence S has no smallest element. □

This lemma enables us to also define the invariance groups of cuts and quasicuts.

Proposition 2.17. *If* (Λ^L, Λ^R) *is a quasicut in* G, *then*

$$\mathcal{G}(\Lambda^L) = \mathcal{G}(\Lambda^R).$$

If (Λ^L, Λ^R) *is not a cut, then these groups are trivial.*

Proof. The first statement follows from the fact that for every $\gamma \in G$,

$$\Lambda^L + \gamma = \Lambda^L \Leftrightarrow \Lambda^R + \gamma = \Lambda^R.$$

If (Λ^L, Λ^R) is not a cut, then $\Lambda^L \cap \Lambda^R = \{\gamma\}$ for some $\gamma \in G$. It follows that $\Lambda^R = \gamma^-$, hence by Lemma 2.16(2), $\mathcal{G}(\Lambda^L) = \mathcal{G}(\Lambda^R) = \{0\}$. □

For every quasicut $\Lambda = (\Lambda^L, \Lambda^R)$ we can now define

$$\mathcal{G}(\Lambda) := \mathcal{G}(\Lambda^L) = \mathcal{G}(\Lambda^R).$$

In order to give answers to questions (QFS1), (QFS2) and (QFS3) in the case where $\mathcal{G}(S_1) = \mathcal{G}(S_2) = \{0\}$ in Section 2.5 below, we need some more preparation. We set $G^{>0} := \{\gamma \in G \mid \gamma > 0\} = 0^+$.

Lemma 2.18. *Take a final segment* S *of* G. *Then* $\mathcal{G}(S)$ *is the largest of all subgroups* H *of* G *satisfying* $H \cap (S - S^c) = \varnothing$. *Further*,

$$S - S^c = G^{>0} \setminus \mathcal{G}(S) = \mathcal{G}(S)^+. \tag{26}$$

Proof. If $\gamma = \alpha - \beta$ with $\alpha \in S$ and $\beta \in S^c$, then $\alpha - \gamma = \beta \notin S$, so $-\gamma \notin \mathcal{G}(S)$. Since $\mathcal{G}(S)$ is a subgroup of G, this implies that $\gamma \notin \mathcal{G}(S)$. Conversely, if H is a subgroup of G properly containing $\mathcal{G}(S)$, then there is $\gamma \in H^{>0} \setminus \mathcal{G}(S)$. Then $-\gamma \notin \mathcal{G}(S)$ and there is $\alpha \in S$ such that $\alpha - \gamma \notin S$. This implies $\beta = \alpha - \gamma \in S^c$, whence $\gamma = \alpha - \beta \in S - S^c$.

Now we prove (26). Since $\alpha > \beta$ whenever $\alpha \in S$ and $\beta \in S^c$, it follows that $S - S^c \subseteq G^{>0} \setminus \mathcal{G}(S)$. For the converse, assume that $\gamma \in G^{>0} \setminus \mathcal{G}(S)$. Then there is some $\beta \in S^c$ such that $\beta + \gamma \in S$ and therefore, $\gamma \in S - S^c$. □

Corollary 2.19. *Let* S_1 *and* S_2 *be final segments of* G, *with* S_1 *nonprincipal.*

(1) *We have*

$$S_1 \ -_{\mathrm{mc}} \ S_1 = S_1 - S_1^c = \mathcal{G}(S_1)^+,$$

so that

$$S_1 + (S_2 - S_1^c) = S_1 + (S_2 \ -_{\mathrm{mc}} \ S_1) = \mathcal{G}(S_1)^+ + S_2 \subseteq S_2. \tag{27}$$

(2) *If* $\mathcal{G}(S_1) = \{0\}$ *and* S_2 *is nonprincipal, then* $S_1 + (S_2 - S_1^c) = S_2$.

(3) *If* $\mathcal{G}(S_1) \subsetneq \mathcal{G}(S_2)$, *then* $S_1 + (S_2 - S_1^c) = S_2$.

Proof. (1) This follows from (21) in Lemma 2.10(3) together with Lemma 2.18. The equalities in (27) follow from part (1) together with parts (3) and (5) of Lemma 2.10. The inclusion holds since $\mathcal{G}(S_1)^+$ only contains positive elements.

(2) This follows from part (1) together with Lemma 2.6(3).

(3) Since $\mathcal{G}(S_1) \subsetneq \mathcal{G}(S_2)$, there is some $\alpha \in \mathcal{G}(S_1)^+ \cap \mathcal{G}(S_2)$. We obtain that $S_2 = \alpha + S_2 \subseteq \mathcal{G}(S_1)^+ + S_2$. In view of (27), this proves statement (3). □

Finally, for application in Section 3.3 we prove the following facts:

Proposition 2.20. *Take final segments S, S_1 and S_2 of G.*

(1) *We have that $\mathcal{G}(S)^+$ does not contain S if and only if $0 \in S$.*

(2) *The set $\{\alpha \in G \mid \alpha + \mathcal{G}(S)^+ \subseteq S\}$ is equal to S if there is no $\gamma \in G$ such that $\mathcal{G}(S)^+ = \gamma + S$; otherwise, it properly contains S.*

(3) *If $\mathcal{G}(S_2) \subsetneq \mathcal{G}(S_1)$, then there is $\alpha \in G$ such that $S_1 + S_2 = S_1 + \alpha$.*

(4) *The set of all $\alpha \in G$ such that $\alpha + S \subseteq G^{>0}$ is equal to $-S^c$.*

(5) *The final segment $\mathcal{G}(S)^+$ is the union of all $\alpha + S \subseteq G^{>0}$, $\alpha \in S$.*

Proof. (1) If $\mathcal{G}(S)^+$ does not contain S, then $S \setminus \mathcal{G}(S)^+ \neq \varnothing$, so there is some $\alpha \in S \cap \mathcal{G}(S)$. Since $\mathcal{G}(S)$ is a group, it follows that $-\alpha \in \mathcal{G}(S)$. Therefore $S = -\alpha + S$, whence $0 = \alpha - \alpha \in S$. The converse holds since $0 \in \mathcal{G}(S)^+$ is impossible.

(2) Since $\mathcal{G}(S)^+$ only contains positive elements, S is always contained in the set $\{\alpha \in G \mid \alpha + \mathcal{G}(S)^+ \subseteq S\}$. Assume there is no $\gamma \in G$ such that $\mathcal{G}(S)^+ = \gamma + S$. Take any $\alpha \in G$ such that $\alpha + \mathcal{G}(S)^+ \subseteq S$, i.e., $\mathcal{G}(S)^+ \subseteq -\alpha + S$. Then by assumption, $\mathcal{G}(S)^+ \subsetneq -\alpha + S$. Since $\mathcal{G}(-\alpha + S) = \mathcal{G}(S)$ by Lemma 2.13(3), we have $\mathcal{G}(-\alpha + S)^+ = \mathcal{G}(S)^+$. Then $0 \in -\alpha + S$ by part (1) of our Proposition, so $\alpha \in S$.

Now assume that $\mathcal{G}(S)^+ = \gamma + S$ for some $\gamma \in G$. Then $-\gamma + \mathcal{G}(S)^+ \subseteq S$, but $-\gamma \notin S$ since otherwise $0 \in \mathcal{G}(S)^+$, which is impossible.

(3) Take $\beta \in \mathcal{G}(S_1) \setminus \mathcal{G}(S_2)$ with $\beta > 0$. Then $\beta + S_2 \subsetneq S_2$ and there is $\alpha \in S_2 \setminus \beta + S_2$, so that $\beta + S_2 \subseteq \alpha^- \subseteq S_2$. Now $S_1 + S_2 = (S_1 + \beta) + S_2 = S_1 + (\beta + S_2) \subseteq S_1 + \alpha^- \subseteq S_1 + S_2$, whence $S_1 + S_2 = S_1 + \alpha^- = S_1 + \alpha$.

(4) This holds since $\alpha + S \subseteq G^{>0} \Leftrightarrow \alpha + S > 0 \Leftrightarrow S > -\alpha \Leftrightarrow -\alpha \in S^c \Leftrightarrow \alpha \in -S^c$.

(5) By part (4), $\bigcup\{\alpha + S \mid \alpha \in G \text{ with } \alpha + S \subseteq G^{>0}\} = \bigcup\{\alpha + S \mid \alpha \in -S^c\} = -S^c + S$. By (26) in Lemma 2.18, this is equal to $\mathcal{G}(S)^+$. □

2.5. *Partial answers to questions (*QFS1*), (*QFS2*) and (*QFS3*).* We are now in the position to give an answer to questions (QFS1), (QFS2) and (QFS3) in the important special case where $\mathcal{G}(S_1) = \mathcal{G}(S_2) = \{0\}$. We will use this answer later together with the techniques which we will develop in Section 2.6 to settle the general case where $\mathcal{G}(S_1)$ can be any proper convex subgroup of G.

Lemma 2.21. *Take final segments S_1 and S_2 of G and assume that S_1 is nonprincipal. If $\mathcal{G}(S_2) = \{0\}$ and $T = T_0$ is a solution of* (1)*, then $T = \widehat{T_0}$ is the largest solution of* (1)*.*

Proof. Since $S_1 + T_0 = S_2$, by Lemma 2.6(4) also $S_1 + \widehat{T_0} = S_2$ holds. Suppose that T' is a larger solution, and pick some $\alpha' \in T' \setminus \widehat{T_0}$. If $\widehat{T_0}$ has a minimum, then denote it by α. If $\widehat{T_0}$ has no minimum, then $\widehat{T_0} = T_0$ has no infimum and $G \setminus T_0$ has no supremum, so there is $\alpha \in G$ such that $\alpha' < \alpha < T_0$. In both cases, we have that $\alpha \leq \widehat{T_0}$. Consequently,

$$S_2 = S_1 + \widehat{T_0} \subseteq S_1 + \alpha \subseteq S_1 + \alpha' \subseteq S_1 + T' = S_2,$$

so equality holds everywhere. It follows that $S_2 + \alpha' - \alpha = S_1 + \alpha + \alpha' - \alpha = S_1 + \alpha' = S_2$, whence $\alpha' - \alpha \in \mathcal{G}(S_2) = \{0\}$. Thus $\alpha' = \alpha$, a contradiction. □

Remark 2.22. If S_1 is principal, then $\widehat{T_0}$ will not always be a solution, even if $\mathcal{G}(S_2)$ is trivial. As a simple counterexample, assume that G is densely ordered and consider the equation $0^- + T = 0^+$. Then $T = 0^+$ is a solution, but $T = 0^-$ is not.

Proposition 2.23. *Take final segments S_1 and S_2 of G. Assume that $\mathcal{G}(S_1) = \mathcal{G}(S_2) = \{0\}$ and that S_2 is nonprincipal if S_1 is nonprincipal. Then $T = S_2 -_{\mathrm{fs}} S_1$ is the largest solution of* (1)*, and we have*

$$S_2 -_{\mathrm{fs}} S_1 = \begin{cases} S_2 - \alpha & \text{if } S_1 = \alpha^- \text{ for some } \alpha \in G, \\ (S_2 - S_1^c)^\frown & \text{if } S_1 \text{ is nonprincipal.} \end{cases} \tag{28}$$

Proof. As stated in the introduction, $T = S_2 -_{\mathrm{fs}} S_1$ is the largest solution of $S_1 + T \subseteq S_2$; hence if a solution of equation (1) exists, then $T = S_2 -_{\mathrm{fs}} S_1$ is the largest solution of it. For $S_1 = \alpha^-$, our assertion follows from Proposition 2.5. If S_1 is nonprincipal, then by assumption, S_2 is nonprincipal, and it follows from Corollary 2.19(2) that $S_2 - S_1^c$ is a solution of (1). Hence our statement for this case follows from Lemma 2.21. □

2.6. ***Ordered abelian groups modulo convex subgroups.*** *In what follows, G will be an ordered abelian group with arbitrary convex subgroup H.* The quotient G/H carries a canonical ordering such that for all $\alpha, \beta \in G$,

$$\alpha \leq \beta \Rightarrow \alpha + H \leq \beta + H. \tag{29}$$

Note that under this induced ordering on G/H, we have $\alpha + H < \beta + H$ if and only if it holds for $\alpha + H$ and $\beta + H$ as subsets of G as defined at the beginning of Section 2.1.

For the image of an element $\gamma \in G$ under the canonical epimorphism

$$\varphi_H : G \to G/H,$$

i.e., the coset $\gamma + H$, we prefer to write γ/H since we also have to deal with subsets of the form $\gamma + H$ in G. With this notation, $M/H = \{\gamma/H \mid \gamma \in M\}$ for every subset $M \subseteq G$. Since the epimorphism preserves $\leq$, we have that M/H is convex if M is convex, and S/H is a final segment (or initial segment) of G/H if S is a

final segment (or initial segment, respectively) of G. In what follows, we will freely make use of the following fact, whose proof is straightforward:

Lemma 2.24. *For each subset M of G the preimage $\varphi_H^{-1}(M/H)$ of M/H under φ_H is $M+H$. If M is a final segment of G, then so is $M+H$.*

Proposition 2.25. *Take a subset $M \subseteq G$, a final segment S of G, and a convex subgroup H of G.*

(1) *If $H \subseteq \mathcal{G}(M)$, then*
$$\mathcal{G}(M/H) = \mathcal{G}(M)/H\,.$$

(2) *If $\mathcal{G}(S) \subseteq H$, then $\mathcal{G}(S/H)$ is trivial.*

(3) *If $\mathcal{G}(S)$ is trivial, then also $\mathcal{G}(S/H)$ is trivial.*

(4) *If $\mathcal{G}(S) \subsetneq H$, then S/H is principal.*

(5) *If $G/\mathcal{G}(S)$ is discretely ordered, then $S/\mathcal{G}(S)$ is principal.*

Proof. (1) Assume that $H \subseteq \mathcal{G}(M)$. Then $M+H=M$. It follows that $M+H+\gamma+H=M+\gamma$ for every $\gamma \in G$, so we have $M/H+\gamma/H=M/H$ if and only if $M+\gamma=M$. Hence, $\mathcal{G}(M/H)=\mathcal{G}(M)/H$.

(2) Assume that $\mathcal{G}(S) \subseteq H$. Take a convex subgroup H' of G containing H such that $\mathcal{G}(S/H)=H'/H$. That is, $S/H+H'/H=S/H$. It follows that $S+H+H'+H=S+H$, showing that H' is a subgroup of $\mathcal{G}(S+H)$. By Lemma 2.13(4), $\mathcal{G}(S+H)=\mathcal{G}(S)\cup H=H$. We conclude that $H'=H$ and thus, $\mathcal{G}(S/H)=H'/H$ is trivial.

(3) This follows from part (2) since the trivial subgroup is contained in H.

(4) If $\mathcal{G}(S) \subsetneq H$, then we have $S+H \neq S$, so there is some $\alpha \in S+H$ such that $\alpha < S$ (since S is a final segment). As $\alpha \in S+H$, it follows that $\alpha/H \in S/H$. We wish to show that α/H is the smallest element of S/H. Take any $\beta \in G$ such that $\beta/H < \alpha/H$. This means that $\beta+H < \alpha < S$, so $S \cap (\beta+H)=\varnothing$ and therefore, $\beta/H \notin S/H$. This proves our claim.

(5) Take $\gamma \in G$ such that $\gamma/\mathcal{G}(S)$ is the smallest positive element of $G/\mathcal{G}(S)$ and therefore, $(\gamma/\mathcal{G}(S))\mathbb{Z}$ is a convex subgroup of $G/\mathcal{G}(S)$ and we have $(\gamma/\mathcal{G}(S))^- = (0/\mathcal{G}(S))^+$. Suppose that $S/\mathcal{G}(S)$ is nonprincipal. Then $(\gamma/\mathcal{G}(S))^- + S/\mathcal{G}(S) = (0/\mathcal{G}(S))^+ + S/\mathcal{G}(S) = (0/\mathcal{G}(S))^- + S/\mathcal{G}(S) = S/\mathcal{G}(S)$ by (15), which implies that $(\gamma/\mathcal{G}(S))\mathbb{Z} \subseteq \mathcal{G}(S/\mathcal{G}(S))$, contradicting part (2) of our proposition. □

Lemma 2.26. *Take a final segment S of G. Then $S+\mathcal{G}(S)^+ \neq S$ if and only if $S/\mathcal{G}(S)$ is principal.*

Proof. We have that $\mathcal{G}(S)^+/\mathcal{G}(S) = (0/\mathcal{G}(S))^+$. Hence it follows from (10) together with Lemma 2.6(3) that $S/\mathcal{G}(S)+\mathcal{G}(S)^+/\mathcal{G}(S) \neq S/\mathcal{G}(S)$ if and only if $S/\mathcal{G}(S)$ is principal. If inequality holds, then we also must have $S+\mathcal{G}(S)^+ \neq S$. Now

assume that equality holds. Then it must also hold for the respective preimages of $S/\mathcal{G}(S)$ and $\mathcal{G}(S)^+/\mathcal{G}(S)$ under $\varphi_{\mathcal{G}(S)}$. By Lemma 2.24, $\varphi^{-1}_{\mathcal{G}(S)}(S/\mathcal{G}(S)) = S + \mathcal{G}(S) = S$. By Lemma 2.13(5) with $\alpha = 0$, $\mathcal{G}(\mathcal{G}(S)^+) = \mathcal{G}(S)$, which shows that $\varphi^{-1}_{\mathcal{G}(S)}(\mathcal{G}(S)^+/\mathcal{G}(S)) = \mathcal{G}(S)^+ + \mathcal{G}(S) = \mathcal{G}(S)^+$. Therefore, the equality $S/\mathcal{G}(S) + \mathcal{G}(S)^+/\mathcal{G}(S) = S/\mathcal{G}(S)$ implies the equality $S + \mathcal{G}(S)^+ = S$. □

We note that if (Λ^L, Λ^R) is a cut in G, then $\Lambda^L/H \leq \Lambda^R/H$, hence

$$\Lambda/H := (\Lambda^L/H, \Lambda^R/H) \tag{30}$$

is a quasicut in G/H.

Proposition 2.27. *The induced quasicut* (30) *in G/H is a cut if and only if $H \subseteq \mathcal{G}(\Lambda)$. The invariance group of the cut $\Lambda/\mathcal{G}(\Lambda)$ in $G/\mathcal{G}(\Lambda)$ is trivial.*

Proof. The quasicut Λ/H is a cut in G/H if and only if $\Lambda^L/H \cap \Lambda^R/H = \varnothing$. This holds if and only if $\Lambda^R + H \subseteq \Lambda^R$. By the definition of $\mathcal{G}(\Lambda)$, this in turn holds if and only if $H \subseteq \mathcal{G}(\Lambda)$.

The last statement follows from the definition of $\mathcal{G}(\Lambda)$ and Proposition 2.25(2). □

2.7. *The equation $S_1 + T = S_2$ in ordered abelian groups of arbitrary rank.* In this section we will first assume that S_1, S_2 are final segments of the ordered abelian group G and that the equation (1) has a solution $T = T_0$. We determine whether it is unique, and if not, compute the largest solution. We will use the following crucial fact, whose proof is straightforward:

Lemma 2.28. *Take final segments S_1, S_2, T_0 and a proper convex subgroup H of G. If $\overline{T} = \overline{T}_0$ is the largest solution of $S_1/H + \overline{T} = S_2/H$, then its preimage $T = \varphi_H^{-1}(\overline{T}_0)$ is the largest solution of $S_1 + T = S_2$. In particular, if $\overline{T} = T_0/H$ is the largest solution of $S_1/H + \overline{T} = S_2/H$, then its preimage $T = T_0 + H$ under φ_H is the largest solution of $S_1 + T = S_2$.*

Take a final segment T of G. We define

$$T^\diamond$$

to be the preimage of $\widehat{T/\mathcal{G}(T)}$ under $\varphi_{\mathcal{G}(T)}$. Note that $T^\diamond$ is a final segment of G. Observe that $T^\diamond = T$ holds if and only if $T/\mathcal{G}(T)$ is closed in $G/\mathcal{G}(T)$. If $\mathcal{G}(T)$ is trivial, then $T^\diamond = \widehat{T}$. It follows from Lemma 2.12(6) that $T^\diamond$ is closed also if $\mathcal{G}(T)$ is not trivial. We view $T^\diamond$ as the "deep closure" of T.

Lemma 2.29. *Take a final segment T of G.*

(1) *The preimage of $-(T/\mathcal{G}(T))^c$ under $\varphi_{\mathcal{G}(T)}$ is $-T^c$.*

(2) *For each $\alpha \in G$, the preimage of $(\alpha/\mathcal{G}(T))^-$ under $\varphi_{\mathcal{G}(T)}$ is $\alpha^- + \mathcal{G}(T)$.*

(3) *If there is $\alpha \in G$ such that $\alpha/\mathcal{G}(T)$ is the infimum of $T/\mathcal{G}(T)$ in $G/\mathcal{G}(T)$, then*

$$T^\diamond = \alpha^- + \mathcal{G}(T);$$

otherwise, $T^\diamond = T$.

Proof. (1) Lemma 2.16(1) gives $\mathcal{G}(-T^c) = \mathcal{G}(T^c) = \mathcal{G}(T)$; hence the preimage of $-T^c/\mathcal{G}(T)$ under $\varphi_{\mathcal{G}(T)}$ is $-T^c+\mathcal{G}(T) = -T^c$. It remains to show that $-T^c/\mathcal{G}(T) = -(T/\mathcal{G}(T))^c$. Since $\mathcal{G}(T^c) = \mathcal{G}(T)$, the sets $T/\mathcal{G}(T)$ and $T^c/\mathcal{G}(T)$ are disjoint. On the other hand, their union is $G/\mathcal{G}(T)$, so $T^c/\mathcal{G}(T) = (T/\mathcal{G}(T))^c$, showing that $-T^c/\mathcal{G}(T) = -(T/\mathcal{G}(T))^c$ holds.

(2) This follows from Lemma 2.24 since $\alpha^-/\mathcal{G}(T) = (\alpha/\mathcal{G}(T))^-$.

(3) If there is $\alpha \in G$ such that $\alpha/\mathcal{G}(T)$ is the infimum of $T/\mathcal{G}(T)$ in $G/\mathcal{G}(T)$, then $\widehat{T/\mathcal{G}(T)} = (\alpha/\mathcal{G}(T))^-$, whose preimage under $\varphi_{\mathcal{G}(T)}$ is $\alpha^- + \mathcal{G}(T)$. Otherwise, $\widehat{T/\mathcal{G}(T)} = T/\mathcal{G}(T)$, whose preimage is $T + \mathcal{G}(T) = T$. □

Proposition 2.30. *Take final segments S_1 and S_2 of G and assume that $T = T_0$ is a solution of* (1)*. Set $T_1 := T_0 + \mathcal{G}(S_2)$. Then*:

(1) *If $S_1/\mathcal{G}(S_2)$ is principal, then $T = T_1$ is the largest solution of* (1)*.*

(2) *If $S_1/\mathcal{G}(S_2)$ is nonprincipal, then $T = T_1^\diamond$ is the largest solution of* (1)*.*

Proof. Since $S_1 + T_1 = S_1 + T_0 + \mathcal{G}(S_2) = S_2 + \mathcal{G}(S_2) = S_2$, we have $S_1/\mathcal{G}(S_2) + T_1/\mathcal{G}(S_2) = S_2/\mathcal{G}(S_2)$.

(1) If $S_1/\mathcal{G}(S_2)$ is principal, then by Proposition 2.5(2), $\overline{T} = T_1/\mathcal{G}(S_2)$ is the unique solution of $S_1/\mathcal{G}(S_2) + \overline{T} = S_2/\mathcal{G}(S_2)$. Hence by Lemma 2.28, $T = T_1 + \mathcal{G}(S_2) = T_1$ is the largest solution of (1).

(2) We observe that $\mathcal{G}(S_2/\mathcal{G}(S_2))$ is trivial by Proposition 2.25(2). By assumption, $S_1/\mathcal{G}(S_2)$ is nonprincipal, so we can apply Lemma 2.21 to conclude that $\overline{T} = \widehat{T_1/\mathcal{G}(S_2)}$ is the largest solution of $S_1/\mathcal{G}(S_2) + \overline{T} = S_2/\mathcal{G}(S_2)$. Therefore, from its definition and Lemma 2.28 we obtain that $T = T_1^\diamond$ is the largest solution of (1). □

Remark 2.31. Remark 2.22 shows that $T = T_1^\diamond$ is not always a solution of (1).

For an arbitrary final segment S of G, the following application of the foregoing proposition gives an interesting characterization of $S^\diamond$:

Corollary 2.32. *Take a final segment S of G. Then $T = S^\diamond$ is the largest solution of the equation*

$$S + T = S + S.$$

Proof. We apply Proposition 2.30 with $T_0 = S$. Since $\mathcal{G}(S + S) = \mathcal{G}(S)$ by Corollary 2.14 and $S + \mathcal{G}(S) = S$, we obtain $T_1 = S$. If $S/\mathcal{G}(S)$ is nonprincipal, then our claim follows from Proposition 2.30(2). Now assume that $S/\mathcal{G}(S)$ is principal. Then by Proposition 2.30(1), S is the largest solution, and we need to show that

$S = S^{\diamond}$ in this case. As $S/\mathcal{G}(S)$ is principal, it is closed in $G/\mathcal{G}(S)$ and we indeed have $S^{\diamond} = S$ by our observation before Lemma 2.29. □

Now we answer questions (QFS1), (QFS2) and (QFS3) in full generality.

Theorem 2.33. *Take final segments S_1 and S_2 of G.*

(1) *Assume that*

$$\mathcal{G}(S_1) \subseteq \mathcal{G}(S_2), \text{ and } S_2/\mathcal{G}(S_2) \text{ is nonprincipal if } S_1/\mathcal{G}(S_2) \text{ is nonprincipal.} \quad (*)$$

Then $T = S_2 -_{\mathrm{fs}} S_1$ is the largest solution of (1); *moreover we have $T = S_2 - \alpha$ if $S_1/\mathcal{G}(S_2) = (\alpha/\mathcal{G}(S_2))^{-}$, and $T = (S_2 - S_1^c)^{\diamond}$ if $S_1/\mathcal{G}(S_2)$ is nonprincipal.*

(2) *Assume instead that condition* (∗) *does not hold*; *equivalently*, *that either*

(i) $\mathcal{G}(S_2) \subsetneq \mathcal{G}(S_1)$ *or*

(ii) $\mathcal{G}(S_1) = \mathcal{G}(S_2)$ *and $S_1/\mathcal{G}(S_2)$ is nonprincipal*, *but $S_2/\mathcal{G}(S_2)$ is principal.*

Then equation (1) *has no solution.*

Proof. To simplify notation, we set $\mathcal{G}_2 := \mathcal{G}(S_2)$.

(1) Assume (∗) holds. From Proposition 2.25(2) we know that $\mathcal{G}(S_1/\mathcal{G}_2) = \mathcal{G}(S_2/\mathcal{G}_2) = \{0\}$. Hence by Proposition 2.23, the largest solution of $S_1/\mathcal{G}_2 + \overline{T} = S_2/\mathcal{G}_2$ is $\overline{T} = S_2/\mathcal{G}_2 - \alpha/\mathcal{G}_2 = (S_2 - \alpha)/\mathcal{G}_2$ if $S_1/\mathcal{G}_2 = (\alpha/\mathcal{G}_2)^{-}$, and $\overline{T} = (S_2/\mathcal{G}_2 - (S_1/\mathcal{G}_2)^c)\widehat{\ }$ if $S_1/\mathcal{G}_2$ is nonprincipal.

By Lemma 2.28 the preimage T of $\overline{T}$ under $\varphi_{\mathcal{G}_2}$ is the largest solution of (1). As stated in the introduction, $S_2 -_{\mathrm{fs}} S_1$ is the largest solution of $S_1 + T \subseteq S_2$, so it follows that $T = S_2 -_{\mathrm{fs}} S_1$.

If $\overline{T} = (S_2 - \alpha)/\mathcal{G}_2$, then $T = (S_2 - \alpha) + \mathcal{G}_2 = S_2 - \alpha$ by Lemma 2.24 and Lemma 2.13(3). Assume $S_1/\mathcal{G}_2$ is nonprincipal and $\overline{T} = (S_2/\mathcal{G}_2 - (S_1/\mathcal{G}_2)^c)\widehat{\ }$. By Proposition 2.25(4), the former implies that $\mathcal{G}_2 \subseteq \mathcal{G}(S_1)$, whence $\mathcal{G}(S_1) = \mathcal{G}_2$. Therefore, $\varphi_{\mathcal{G}_2}^{-1}(-(S_1/\mathcal{G}_2)^c) = -S_1^c$ by Lemma 2.29(1). By Lemma 2.24, $\varphi_{\mathcal{G}_2}^{-1}(S_2/\mathcal{G}_2) = S_2 + \mathcal{G}_2 = S_2$, hence $\varphi_{\mathcal{G}_2}^{-1}(S_2/\mathcal{G}_2 - (S_1/\mathcal{G}_2)^c) = S_2 - S_1^c$. Thus the preimage of $\overline{T}$ is $T = (S_2 - S_1^c)^{\diamond}$.

(2) Now assume that condition (∗) does not hold. Then one of the cases (i) or (ii) must hold; for if $\mathcal{G}(S_1) \subsetneq \mathcal{G}_2$, then by Proposition 2.25(4), $S_1/\mathcal{G}_2$ is principal and condition (∗) holds.

In both cases (i) and (ii), S_1 is nonprincipal. In case (i), Proposition 2.15 shows that (1) has no solution. Assume that case (ii) holds and that the final segment T is a solution of (1). Then $S_1/\mathcal{G}_2 + T/\mathcal{G}_2 = S_2/\mathcal{G}_2$; but this contradicts Proposition 2.4(2) applied to the group $G/\mathcal{G}_2$. □

The following theorem addresses question (QFS4) in the case where no solution to (1) exists:

Theorem 2.34. *Take final segments S_1 and S_2 of G.*

(1) *Assume that $\mathcal{G}(S_2) \subsetneq \mathcal{G}(S_1)$. Then there exists an element $\alpha \in S_2$ such that $S_2' := (\alpha + \mathcal{G}(S_1))^+ \subsetneq S_2$ is the largest final segment contained in S_2 having invariance group $\mathcal{G}(S_1)$.*

(2) *Assume that $\mathcal{G}(S_1) = \mathcal{G}(S_2)$ and $S_1/\mathcal{G}(S_1)$ is nonprincipal, but $S_2/\mathcal{G}(S_1)$ is principal, say $S_2/\mathcal{G}(S_1) = (\alpha/\mathcal{G}(S_1))^-$ for some $\alpha \in S_2$. Then $S_2' := (\alpha+\mathcal{G}(S_1))^+ \subsetneq S_2$ is the largest final segment contained in S_2 such that $S_2'/\mathcal{G}(S_1)$ is nonprincipal.*

In both cases, condition $(*)$ of Theorem 2.33 is satisfied with S_2' in place of S_2 and $T = S_2' -_{\mathrm{fs}} S_1$ is the largest solution of $S_1 + T = S_2'$. Further, $S_2' -_{\mathrm{fs}} S_1 = S_2 -_{\mathrm{fs}} S_1$.

Proof. To simplify notation, we set $\mathcal{G}_1 := \mathcal{G}(S_1)$.

(1) By the definition of $\mathcal{G}(S_2)$, there is some $\alpha \in S_2$ such that $\alpha+\mathcal{G}_1$ is not contained in S_2. We set $S_2' = (\alpha + \mathcal{G}_1)^+$. By definition, S_2' is a final segment, and since $\alpha \in S_2 \setminus S_2'$, we have $S_2' \subsetneq S_2$. From Lemma 2.13(5) we know that $\mathcal{G}(S_2') = \mathcal{G}_1$.

Suppose that there is a final segment S of S_2 properly containing S_2' and such that $\mathcal{G}_1 \subseteq \mathcal{G}(S)$. Then there would exist an element $\beta \in (\alpha + \mathcal{G}_1) \cap S_2$ such that $\beta + \mathcal{G}_1 \subseteq \beta + \mathcal{G}(S) \subseteq S \subseteq S_2$. However, it would follow that $\beta - \alpha \in \mathcal{G}_1$ and $\alpha+\mathcal{G}_1 = \beta+\mathcal{G}_1 \subseteq S_2$, contradiction. This shows that S_2' is the largest final segment contained in S_2 having an invariance group containing $\mathcal{G}_1$.

Now if $S_1 + T \subseteq S_2$, then $\mathcal{G}_1 \subseteq \mathcal{G}(S_1 + T)$ by Lemma 2.13(2); hence, by what we have just shown, $S_1 + T \subseteq S_2'$.

Assume that $S_2'/\mathcal{G}_1 = (\alpha/\mathcal{G}_1)^+$ is principal; it then follows that $G/\mathcal{G}_1$ is discretely ordered. Hence by Proposition 2.25(5), also $S_1/\mathcal{G}_1$ is principal. This shows that condition $(*)$ of Theorem 2.33 holds with S_2 replaced by S_2'. Now it follows from Theorem 2.33 that $T = S_2' -_{\mathrm{fs}} S_1$ is the largest solution of $S_1 + T = S_2'$. Since $S_2' \subset S_2$, we have $S_2' -_{\mathrm{fs}} S_1 \subseteq S_2 -_{\mathrm{fs}} S_1$. On the other hand, $\mathcal{G}_1 \subseteq \mathcal{G}(S_1 + (S_2 -_{\mathrm{fs}} S_1))$ which implies that $S_1 + (S_2 -_{\mathrm{fs}} S_1) \subseteq S_2'$ and therefore $S_2 -_{\mathrm{fs}} S_1 \subseteq S_2' -_{\mathrm{fs}} S_1$.

(2) As in the proof of part (1) it is shown that $S_2' \subsetneq S_2$ and that $\mathcal{G}(S_2') = \mathcal{G}_1$. Also as in the proof of part (1) it is shown that if $S_2'/\mathcal{G}_1 = (\alpha/\mathcal{G}_1)^+$ is principal, then so is $S_1/\mathcal{G}_1$. But this contradicts our assumption, so $S_2'/\mathcal{G}_1$ is nonprincipal.

Any final segment S that properly contains $S_2' = (\alpha + \mathcal{G}_1)^+$ will contain an element $\beta \in \alpha + \mathcal{G}_1$. This implies that $S/\mathcal{G}_1$ contains $\alpha/\mathcal{G}_1$ and thus contains $(\alpha/\mathcal{G}(S_1))^- = S_2/\mathcal{G}_1$. Hence if S is contained in S_2, $S/\mathcal{G}_1$ is contained in and therefore equal to $S_2/\mathcal{G}_1$, which is principal. This proves that S_2' is the largest final segment contained in S_2 such that $S_2'/\mathcal{G}_1$ is nonprincipal.

Again, condition $(*)$ of Theorem 2.33 holds with S_2 replaced by S_2' so that $T = S_2' -_{\mathrm{fs}} S_1$ is the largest solution of $S_1 + T = S_2'$ and as before, $S_2' -_{\mathrm{fs}} S_1 \subseteq S_2 -_{\mathrm{fs}} S_1$. To show the reverse inclusion, suppose that $S_2' \subsetneq S_1 + (S_2 -_{\mathrm{fs}} S_1)$. Then $(S_1 + (S_2 -_{\mathrm{fs}} S_1))/\mathcal{G}_1$ contains $\alpha/\mathcal{G}_1$ and thus $S_1 + (S_2 -_{\mathrm{fs}} S_1)$ must be equal to S_2,

which is impossible by Theorem 2.33. Hence again, $S_1 + (S_2 -_{\mathrm{fs}} S_1) \subseteq S_2'$ and therefore $S_2 -_{\mathrm{fs}} S_1 \subseteq S_2' -_{\mathrm{fs}} S_1$. □

Remark 2.35. As mentioned in the introduction, a survey and a list of references on the arithmetic of cuts is given in [8]. However, there we missed (at least) one important reference, namely to the work of Schikhof; see e.g. [10]. The solution of equations of the form $S_1 + T = S_2$ is implicitly connected with the investigation of the set $S_1 + G^{\uparrow}$, which is done in [10]. There too, invariance groups, appearing under the name of *stabilizer*, play a crucial role, as well as the final segments H^- and H^+. At this point, let us mention that in our terminology indicated in [8], the cuts determined by H^- and H^+ are called *group cuts*, and the *shifted* cuts determined by $\gamma + H^-$ and $\gamma + H^+$, with $\gamma \in G$, are called *ball cuts*.

3. $\mathcal{O}$-ideals

For general background from valuation theory, we recommend [4; 12]. Throughout, we consider a valued field (K, v) with valuation ring $\mathcal{O}_v$ and denote its maximal ideal by $\mathcal{M}_v$. Its value group will be denoted by vK, and its residue field $\mathcal{O}_v/\mathcal{M}_v$ by Kv. For $a \in K$ we denote its value under v by va.

Recall that by an $\mathcal{O}_v$-ideal I we mean a nonzero $\mathcal{O}_v$-module $I \subsetneq K$. We consider (possibly fractional) $\mathcal{O}_v$-ideals I, I_1, I_2, J, etc. We are interested in the inequality

$$I_1 J \subseteq I_2 \tag{31}$$

and the corresponding equality (6). We note that $I_2 : I_1$ as defined in (7) is an $\mathcal{O}_v$-ideal, which may be fractional, and recall that $J = I_2 : I_1$ is the largest $\mathcal{O}_v$-ideal that satisfies (31). Further, we recall:

Proposition 3.1. *The map*

$$v : I \mapsto vI = \{vb \mid 0 \neq b \in I\}$$

establishes an order-preserving bijection between the $\mathcal{O}_v$-ideals and the final segments of vK; its inverse map is

$$S \mapsto I_S = (a \in K \mid va \in S).$$

The proofs of the following facts are straightforward.

Lemma 3.2. (1) *For $\mathcal{O}_v$-ideals I_1, I_2 and J, the equation* (6) *holds if and only if*

$$vI_1 + vJ = vI_2. \tag{32}$$

(2) *An $\mathcal{O}_v$-ideal I is principal if and only if the final segment vI of vK is principal.*

We will also need:

Lemma 3.3. *Take $\mathcal{O}_v$-ideals I_1 and I_2.*

(1) *We have*

$$v(I_2 : I_1) = \{\alpha \in vK \mid \alpha + vI_1 \subseteq vI_2\} = vI_2 \; -_{\rm fs} \; vI_1 . \tag{33}$$

In particular,

$$v(\mathcal{O}_v : I_1) = \{\alpha \in vK \mid \alpha + vI_1 \geq 0\} = \Delta vI_1 . \tag{34}$$

and

$$vI_2(\mathcal{O}_v : I_1) = vI_2 + \Delta vI_1 = vI_2 \; -_{\rm mc} \; vI_1 . \tag{35}$$

(2) *We have*

$$vI_2(\mathcal{O}_v : I_1) = \begin{cases} vI_2 - va & \text{if } I_1 = a\mathcal{O}_v , \\ vI_2 - (vI_1)^c & \text{if } I_1 \text{ is nonprincipal.} \end{cases} \tag{36}$$

Proof. (1) Equation (33) follows via Proposition 3.1 from the definitions of $I_2 : I_1$ and $vI_2 \; -_{\rm fs} \; vI_1$. Equation (34) follows from (33) since $\alpha + vI_1 \geq 0 \Leftrightarrow \alpha \geq -vI_1$. Equation (35) follows from (34) and Lemma 2.10(1).

(2) If $I_1 = a\mathcal{O}_v$, then $vI_1 = (va)^-$ and $\Delta vI_1 = (-va)^-$ by Lemma 2.8, hence by (35) and (15), $vI_2(\mathcal{O}_v : I_1) = vI_2 + \Delta vI_1 = vI_2 - va$. If I_1 is nonprincipal, then so is vI_1 and again by Lemma 2.8, $\Delta vI_1 = -(vI_1)^c$, whence $vI_2(\mathcal{O}_v : I_1) = vI_2 - (vI_1)^c$ by (35). □

3.1. ***Overrings of $\mathcal{O}_v$ and their ideals.*** Every overring of $\mathcal{O}_v$ is again a valuation ring. Every such overring $\mathcal{O}$ is obtained from $\mathcal{O}_v$ by localization with respect to a prime ideal of $\mathcal{O}_v$; this prime ideal then becomes the maximal ideal $\mathcal{M}$ of $\mathcal{O}$. The set $\mathcal{R}$ of all valuation rings which contain $\mathcal{O}_v$ is linearly ordered by inclusion. Also the set of maximal ideals of these valuation rings, i.e., the set of all prime ideals of $\mathcal{O}_v$, is ordered by inclusion; if $\mathcal{O}_1, \mathcal{O}_2 \in \mathcal{R}$ and $\mathcal{M}_1, \mathcal{M}_2$ are their respective maximal ideals, then $\mathcal{O}_1 \subseteq \mathcal{O}_2 \Leftrightarrow \mathcal{M}_2 \subseteq \mathcal{M}_1$. Note that for $\mathcal{O} \in \mathcal{R}$, an $\mathcal{O}_v$-ideal I is an $\mathcal{O}$-ideal if and only if $I\mathcal{O} = I$, and since all $\mathcal{O} \in \mathcal{R}$ are rings with 1, this holds if and only if $I\mathcal{O} \subseteq I$.

For every $\mathcal{O} \in \mathcal{R}$, we set

$$H(\mathcal{O}) := v\mathcal{O} \cap -v\mathcal{O} = v\mathcal{O}^\times .$$

This is a convex subgroup of the value group vK of (K, v). The valuation w associated with $\mathcal{O}$ is (up to equivalence) given by

$$wa = va/H(\mathcal{O}) \tag{37}$$

for every $a \in K$, the value group of w is canonically isomorphic to $vK/H(\mathcal{O})$, and the value group of the valuation induced by v on the residue field Kw is canonically isomorphic to $H(\mathcal{O})$ (cf. [12]).

Conversely, for every convex subgroup H of vK,

$$\mathcal{O}(H) := \{b \in K \mid \exists \alpha \in H : \alpha \leq vb\} \in \mathcal{R}. \tag{38}$$

We have

$$v\mathcal{O}(H) = H^-. \tag{39}$$

Note that

$$v\mathcal{O}(H)^\times = H, \tag{40}$$

or in other words, $\mathcal{O}(H)^\times = v^{-1}(H)$.

We recall the following facts from general valuation theory (cf. [12]):

Proposition 3.4. *The map $H \mapsto \mathcal{O}(H)$ is an order-preserving bijection from the set of all convex subgroups of vK onto the set $\mathcal{R}$ of all valuation rings which contain $\mathcal{O}_v$. Its inverse is the order preserving map $\mathcal{R} \ni \mathcal{O} \mapsto H(\mathcal{O})$. Thus,*

$$\mathcal{O}(H(\mathcal{O})) = \mathcal{O} \quad \textit{and} \quad H(\mathcal{O}(H)) = H. \tag{41}$$

We denote the maximal ideal of the valuation ring $\mathcal{O}(H)$ by $\mathcal{M}(H)$. Then

$$\mathcal{M}(H) = \{b \in K \mid vb > H\} \quad \text{and} \quad v\mathcal{M}(H) = H^+. \tag{42}$$

We will read off information about an $\mathcal{O}_v$-ideal I from the invariance group of the associated final segment vI. We start with the case of $I = \mathcal{O} \in \mathcal{R}$. Note that for every $\mathcal{O} \in \mathcal{R}$ with maximal ideal $\mathcal{M}$, both $\mathcal{O}$ and $\mathcal{M}$ are $\mathcal{O}_v$-ideals.

Lemma 3.5. *For every $\mathcal{O} \in \mathcal{R}$ with maximal ideal $\mathcal{M}$,*

$$H(\mathcal{O}) = \mathcal{G}(v\mathcal{O}) = \mathcal{G}(v\mathcal{M}) \quad \textit{and} \quad \mathcal{O} = \mathcal{O}(\mathcal{G}(v\mathcal{O})) = \mathcal{O}(\mathcal{G}(v\mathcal{M})). \tag{43}$$

Further, $v\mathcal{O}$ is the disjoint union of $H(\mathcal{O}) = v\mathcal{O}^\times$ and $H(\mathcal{O})^+ = v\mathcal{M}$, with $v\mathcal{O}^\times < v\mathcal{M}$.

Proof. Denote the valuation associated with $\mathcal{O}$ by w. For every $a \in \mathcal{O}^\times$ and every $b \in \mathcal{M}$, we have $wa = 0 < wb$. By the contrapositive of implication (29), this implies that $H(\mathcal{O}) = v\mathcal{O}^\times < v\mathcal{M}$, which shows that $v\mathcal{O}$ is the disjoint union of $v\mathcal{O}^\times = H(\mathcal{O})$ and $v\mathcal{M}$. Thus, $v\mathcal{O} = H(\mathcal{O})^-$ and $v\mathcal{M} = H(\mathcal{O})^+$. Now (43) follows from Proposition 3.4 and Lemma 2.13(5), where $H = H(\mathcal{O})$ and $\alpha = 0$. □

3.2. *The invariance valuation ring of an $\mathcal{O}_v$-ideal.* Take any $\mathcal{O}_v$-ideal I. We set

$$\mathcal{O}(I) := \{b \in K \mid bI \subseteq I\} \quad \text{and} \quad \mathcal{M}(I) = \{b \in K \mid bI \subsetneq I\}. \tag{44}$$

For example, for every $\mathcal{O} \in \mathcal{R}$ with maximal ideal $\mathcal{M}$,

$$\mathcal{O}(\mathcal{O}) = \mathcal{O}(\mathcal{M}) = \mathcal{O} \quad \text{and} \quad \mathcal{M}(\mathcal{O}) = \mathcal{M}(\mathcal{M}) = \mathcal{M}. \tag{45}$$

Further, we define

$$H(I) := H(\mathcal{O}(I)) = v\mathcal{O}(I) \cap -v\mathcal{O}(I) = v\mathcal{O}(I)^\times.$$

Theorem 3.6. (1) *For every $\mathcal{O}_v$-ideal I, $\mathcal{O}(I)$ is a valuation ring of K containing $\mathcal{O}_v$, with maximal ideal $\mathcal{M}(I)$, which is a prime $\mathcal{O}_v$-ideal. It is the largest of all valuation rings $\mathcal{O}$ of K for which I is an $\mathcal{O}$-ideal.*

(2) *For every $\mathcal{O}_v$-ideal I, $\mathcal{O}(I)^\times = \{b \in K \mid bI = I\}$ and $H(I)$ is a convex subgroup of vK.*

(3) *For every $\mathcal{O}_v$-ideal I,*

$$\mathcal{O}(I) = \mathcal{O}(\mathcal{G}(vI)) \text{ and } \mathcal{M}(I) = \mathcal{M}(\mathcal{G}(vI)). \tag{46}$$

as well as

$$H(I) = H(\mathcal{O}(I)) = \mathcal{G}(vI). \tag{47}$$

(4) *For every $\mathcal{O}_v$-ideal I,*

$$\mathcal{M}(I) = \{b \in K \mid vb > \mathcal{G}(vI)\} \quad \text{and} \quad v\mathcal{M}(I) = \mathcal{G}(vI)^+. \tag{48}$$

Proof. (1) It is straightforward to prove that $\mathcal{O}(I)$ is a ring. As I is an $\mathcal{O}_v$-ideal, we have $\mathcal{O}_v I = I$. Hence, $\mathcal{O}_v \subseteq \mathcal{O}(I)$. By general valuation theory, it follows that $\mathcal{O}(I)$ is a valuation ring.

We have $\{b \in K \mid bI = I\} \subseteq \mathcal{O}(I)^\times$ since $bI = I \Leftrightarrow I = b^{-1}I$ for $b \neq 0$. The converse inclusion holds since if $b, b^{-1} \in \mathcal{O}(I)$, then $I = bb^{-1}I \subseteq bI \subseteq I$ and thus, $bI = I$. We have shown that $\mathcal{O}(I)^\times = \{b \in K \mid bI = I\}$. In every valuation ring, the unique maximal ideal consists precisely of all nonunits. Hence the maximal ideal of $\mathcal{O}(I)$ is $\mathcal{O}(I) \setminus \{b \in K \mid bI = I\} = \mathcal{M}(I)$.

Finally, it follows directly from the definition that if $\mathcal{O}$ is a valuation ring of K such that I is an $\mathcal{O}$-ideal, then $\mathcal{O} \subseteq \mathcal{O}(I)$. On the other hand, it also follows from the definition that I is an $\mathcal{O}(I)$-ideal; therefore, $\mathcal{O}(I)$ is the largest of all valuation rings of K for which I is an $\mathcal{O}$-ideal.

(2) The first assertion has been shown in the proof of part (1). The second assertion follows from the definition of $H(I)$ and the fact that $\mathcal{O}(I)$ is a valuation ring.

(3) By Lemma 2.13(6), we compute $\mathcal{O}(I)=\{b\in K\,|\,bI\subseteq I\}=\{b\in K\,|\,vb+vI\subseteq vI\}=\{b\in K\,|\,vb\in\mathcal{G}(vI)^-\}=\mathcal{O}(\mathcal{G}(vI))$; and similarly, we have $\mathcal{M}(I)=\{b\in K\,|\,bI\subsetneq I\}=\{b\in K\,|\,vb+vI\subsetneq vI\}=\{b\in K\,|\,vb\in\mathcal{G}(vI)^+\}=\mathcal{M}(\mathcal{G}(vI))$. From what we have shown, $H(I)=v\mathcal{O}(I)^\times=v\{b\in K\,|\,bI=I\}=\mathcal{G}(vI)^-\setminus\mathcal{G}(vI)^+=\mathcal{G}(vI)$.

(4) Since $\mathcal{M}(I)=\mathcal{M}(\mathcal{G}(vI))$, this follows from part (3) and (42). □

Lemma 3.7. (1) *Take $\mathcal{O}_v$-ideals I and J. Then*

$$\mathcal{O}(IJ) = \mathcal{O}(I) \cup \mathcal{O}(J) = \max\{\mathcal{O}(I), \mathcal{O}(J)\}.$$

(2) *For every $\mathcal{O}_v$-ideal I and $n \in \mathbb{N}$, $\mathcal{O}(I^n) = \mathcal{O}(I)$.*

(3) *For every $\mathcal{O}_v$-ideal I and $0 \neq a \in K$, $\mathcal{O}(aI) = \mathcal{O}(I)$.*

Proof. (1) We have $\mathcal{G}(vIJ) = \mathcal{G}(vI + vJ) = \mathcal{G}(vI) \cup \mathcal{G}(vJ)$ by Lemma 2.13(2). Our statement follows from this by means of (46) and Proposition 3.4.

(2) This follows from part (2) by induction on n.

(3) We have $\mathcal{G}(vaI) = \mathcal{G}(va + vI) = \mathcal{G}(vI)$ by Lemma 2.13(3). Our statement follows from this by means of (46). □

Lemma 3.8. *Take $\mathcal{O} \in \mathcal{R}$ and an $\mathcal{O}_v$-ideal I.*

(1) *$I\mathcal{O}$ is an $\mathcal{O}$-ideal with*

$$vI\mathcal{O} = vI + H(\mathcal{O}). \tag{49}$$

Consequently, I is an $\mathcal{O}$-ideal if and only if $H(\mathcal{O}) \subseteq \mathcal{G}(vI)$.

(2) *If $0 \neq a \in K$, then*

$$va\mathcal{O}(I) = va^- + \mathcal{G}(vI).$$

(3) *Denote by w the valuation associated with $\mathcal{O}$. Then the preimage of $wI = vI/H(\mathcal{O})$ in vK is equal to $vI\mathcal{O}$.*

Proof. (1) We compute

$$vI\mathcal{O} = vI + v\mathcal{O} = vI + (H(\mathcal{O}) \cup H(\mathcal{O})^+) = (vI + H(\mathcal{O})) \cup (vI + H(\mathcal{O})^+),$$

where we have used Lemma 3.5. Since $H(\mathcal{O})^+$ consists only of positive elements, we have $vI + H(\mathcal{O})^+ \subseteq vI$. Since $0 \in H(\mathcal{O})$, we have $vI \subseteq vI + H(\mathcal{O})$. Thus, $(vI + H(\mathcal{O})) \cup (vI + H(\mathcal{O})^+) = vI + H(\mathcal{O})$.

Further, I is an $\mathcal{O}$-ideal if and only if $I\mathcal{O} = I$, or equivalently, $vI + H(\mathcal{O}) = vI$. This in turn holds if and only if $H(\mathcal{O}) \subseteq \mathcal{G}(vI)$.

(2) By (46) and (39), $v\mathcal{O}(I) = v\mathcal{O}(\mathcal{G}(vI)) = \mathcal{G}(vI)^-$. It follows that $va\mathcal{O}(I) = va + v\mathcal{O}(I) = va + 0^- + \mathcal{G}(vI) = va^- + \mathcal{G}(vI)$, where we have used (17) twice.

(3) By (37) and Lemma 2.24, the preimage of wI under $\varphi_{H(\mathcal{O})}$ is $vI + H(\mathcal{O})$. By part (1) of our lemma, this is equal to $vI\mathcal{O}$. □

3.3. *Properties of $\mathcal{M}(I)$.* For $\mathcal{O}_v$-ideals I, the ideals $\mathcal{M}(I)$ appear in [7, Chapter II, Section 4] under the notation $I^\sharp$. There, properties of $\mathcal{M}(I) = I^\sharp$ are listed. In what follows, we will present several of them and show how they follow from the work we have done so far, in particular from our results on final segments in ordered abelian groups in Section 2.

Proposition 3.9. *Take $\mathcal{O}_v$-ideals I and J. Then*:

(1) *$\mathcal{M}(I) = \mathcal{M}_v$ if I is a principal ideal.*

(2) *For $a \in K$, $\mathcal{M}(aI) = \mathcal{M}(I)$.*

(3) *$\mathcal{O}(I) \subseteq I$ if and only if $\mathcal{M}(I)$ does not contain I.*

(4) $\mathcal{M}(I) = I$ *whenever* I *is a prime* $\mathcal{O}_v$*-ideal.*

(5) *If* I *and* J *are prime* $\mathcal{O}_v$*-ideals and* $aI = J$ *for some* $a \in K$*, then* $I = J$*.*

(6) $\mathcal{M}(\mathcal{O}_v : I) = \mathcal{M}(I)$*.*

(7) $I : \mathcal{M}(I) = I$ *if there is no* $a \in K$ *such that* $\mathcal{M}(I) = aI$*; otherwise,* $I : \mathcal{M}(I)$ *properly contains* I*.*

(8) $I\mathcal{M}(I) \subsetneq I$ *if and only if* I *is a principal* $\mathcal{O}(I)$*-ideal.*

(9) *If* $\mathcal{M}(I) \subsetneq \mathcal{M}(J)$*, then* $IJ = aI$ *for some* $a \in K$*.*

(10) *If* $\mathcal{M}(I) = \mathcal{M}(J)$ *and* $J\mathcal{M}(J) \subsetneq \mathcal{M}(J)$*, then* $IJ = aI$ *for some* $a \in K$*.*

(11) $\mathcal{M}(I)$ *is the union of all proper integral ideals* aI*,* $a \in K$*.*

(12) $I(\mathcal{O}_v : I) = \mathcal{M}(I)$ *if* I *is a nonprincipal* $\mathcal{O}_v$*-ideal, and* $I(\mathcal{O}_v : I) = \mathcal{O}_v$ *otherwise.*

(13) $\mathcal{M}(IJ) = \mathcal{M}(I) \cap \mathcal{M}(J) = \min\{\mathcal{M}(I), \mathcal{M}(J)\}$.

(14) *For each* $n \in \mathbb{N}$*,* $\mathcal{M}(I^n) = \mathcal{M}(I)$*.*

Proof. Throughout, we will use (48) and Proposition 3.1 freely.

(1) If $I = a\mathcal{O}_v$ with $a \in K$, then $vI = (va)^-$ and $\mathcal{G}(vI) = \{0\}$ by (25). Hence $\mathcal{M}(I) = \{b \in K \mid vb > 0\} = \mathcal{M}_v$.

(2) Since $v(aI) = va + vI$, we have $\mathcal{G}(v(aI)) = \mathcal{G}(vI)$ by Lemma 2.13(3), whence $\mathcal{G}(v(aI))^+ = \mathcal{G}(vI)^+$ and $\mathcal{M}(aI) = \mathcal{M}(I)$.

(3) By Theorem 3.6, I is an $\mathcal{O}(I)$-ideal and $\mathcal{M}(I)$ is the maximal ideal of $\mathcal{O}(I)$. Hence $\mathcal{O}(I) \subseteq I$ if and only if $\mathcal{O}(I)^\times \cap I$ is nonempty, and this holds if and only if I is not contained in $\mathcal{O}(I) \setminus \mathcal{O}(I)^\times = \mathcal{M}(I)$.

(4) If I is a prime $\mathcal{O}_v$-ideal, then it is the maximal ideal $\mathcal{M}$ of a valuation ring $\mathcal{O} \in \mathcal{R}$ and $\mathcal{M}(I) = \mathcal{M}(\mathcal{M}) = \mathcal{M} = I$ by (45).

(5) By parts (2) and (4), the assumptions imply $I = \mathcal{M}(I) = \mathcal{M}(aI) = \mathcal{M}(J) = J$.

(6) From (34) we know that $v(\mathcal{O}_v : I) = \Delta vI$. By Lemma 2.16(1), $\mathcal{G}(\Delta vI) = \mathcal{G}(vI)$. Using (46) twice, we obtain $\mathcal{M}(\mathcal{O}_v : I) = \mathcal{M}(\mathcal{G}(v(\mathcal{O}_v : I))) = \mathcal{M}(\mathcal{G}(\Delta vI)) = \mathcal{M}(\mathcal{G}(vI)) = \mathcal{M}(I)$.

(7) Since $v(I : \mathcal{M}(I)) = \{va \mid va + v\mathcal{M}(I) \subseteq vI\}$ by (33) of Lemma 3.3 and $v\mathcal{M}(I) = \mathcal{G}(vI)^+$, the statement follows from Proposition 2.20(2).

(8) $I\mathcal{M}(I) \subsetneq I$ if and only if $vI + \mathcal{G}(vI)^+ \neq vI$. By Lemma 2.26, this holds if and only if $vI/\mathcal{G}(vI)$ has a smallest element. Denote the valuation associated with $\mathcal{O}(I)$ by w. Since $\mathcal{G}(vI) = H(\mathcal{O}(I))$ by (47), this in turn holds if and only if $wI = vI/H(\mathcal{O}(I)) = vI/\mathcal{G}(vI)$ has a smallest element, which means that I is a principal $\mathcal{O}(I)$-ideal.

(9) The assumption $\mathcal{M}(I) \subsetneq \mathcal{M}(J)$ is equivalent to $\mathcal{O}(J) \subsetneq \mathcal{O}(I)$. By (46) and Proposition 3.4, this is equivalent to $\mathcal{G}(vJ) \subsetneq \mathcal{G}(vI)$. By Proposition 2.20(3), there is $\alpha \in vK$ such that $vI + vJ = vI + \alpha$. Choosing $a \in K$ with $va = \alpha$, we obtain that $IJ = aI$.

(10) If $\mathcal{M}(I) = \mathcal{M}(J)$, then $\mathcal{O}(I) = \mathcal{O}(J)$. Hence, by part (8), $J\mathcal{M}(J) \subsetneq \mathcal{M}(J)$ implies that J is a principal $\mathcal{O}(J)$-ideal and hence a principal $\mathcal{O}(I)$-ideal, say $J = a\mathcal{O}(I)$. Consequently, $IJ = Ia\mathcal{O}(I) = aI\mathcal{O}(I) = aI$.

(11) Since $v\mathcal{M}(I) = \mathcal{G}(vI)^+$, this follows from Proposition 2.20(5).

(12) By (35), $vI(\mathcal{O}_v : I) = vI -_{\mathrm{mc}} vI$. If vI is nonprincipal, then Corollary 2.19(1) gives $vI -_{\mathrm{mc}} vI = \mathcal{G}(vI)^+$, so $I(\mathcal{O}_v : I) = \mathcal{M}(I)$. If vI is principal, then from (19) with $\alpha = \beta$ we obtain $vI -_{\mathrm{mc}} vI = 0^-$, so $I(\mathcal{O}_v : I) = \mathcal{O}_v$.

(13) and (14) follow from parts (1) and (2) of Lemma 3.7. □

3.4. ***The equation $I_1 J = I_2$.*** We will use Propositions 3.4 and 3.1 to deduce the results in this section from our results in Section 2. Our aim is to use these results to compute $I_2 : I_1$ and to determine under which conditions the equation $I_1(I_2 : I_1) = I_2$ holds.

For every $\mathcal{O} \in \mathcal{R}$ with associated valuation w and every $\mathcal{O}$-ideal J, we define

$$\hat{J}_{\mathcal{O}} := \begin{cases} a\mathcal{O} & \text{if there exists } a \in K \text{ such that } wa \text{ is the infimum of } wJ, \text{ and} \\ J & \text{otherwise.} \end{cases}$$

This is well defined because if $a, b \in K$ with $wa = wb$, then $a\mathcal{O} = b\mathcal{O}$. We will call $\hat{J}_{\mathcal{O}}$ the *closure of J as an $\mathcal{O}$-ideal*. If wJ is closed in wK, then $\hat{J}_{\mathcal{O}} = J$.

Proposition 3.10. *For every $\mathcal{O}$-ideal J,*

$$\mathcal{O} : (\mathcal{O} : J) = \hat{J}_{\mathcal{O}}.$$

Proof. Define w as above. Applying equality (34) (see Lemma 3.3) twice, we obtain $w(\mathcal{O} : (\mathcal{O} : J)) = \Delta w(\mathcal{O} : J) = \Delta(\Delta wJ)$. By Lemma 2.8(3), this is equal to the closure of wJ in wK. Our assertion now follows by Proposition 3.1. □

For every $\mathcal{O}_v$-ideal J, we define

$$J^{\diamond} := \hat{J}_{\mathcal{O}(J)}.$$

We have the following characterizations of $J^{\diamond}$:

Proposition 3.11. *Take any $\mathcal{O}_v$-ideal J. Then*:

(1) $J^{\diamond} = I_{(vJ)^{\diamond}}$.

(2) $J^2 : J = J^{\diamond}$.

Hence, if $vJ/\mathcal{G}(vJ)$ is closed in $vK/\mathcal{G}(vJ)$, then $(J^2 : J) = J$.

Proof. (1) By Theorem 3.6, J is an $\mathcal{O}(J)$-ideal. Denote by w the valuation associated with $\mathcal{O}(J)$, so that $wa = va/H(\mathcal{O}(J))$ for every $a \in K$. By (47), $H(\mathcal{O}(J)) = \mathcal{G}(vJ)$. By definition, $(vJ)^\diamond$ is the preimage of $\widehat{vJ/\mathcal{G}(vJ)} = \widehat{vJ/H(\mathcal{O}(J))} = \widehat{wJ}$ under $\varphi_{H(\mathcal{O}(J))}$. If wJ is closed in wK, then $\widehat{wJ} = wJ = vJ/\mathcal{G}(vJ)$ whose preimage is $vJ + \mathcal{G}(vJ) = vJ$ and therefore, $I_{(vJ)^\diamond} = I_{vJ} = J = J^\diamond$. Now assume that wJ is not closed in wK and take $a \in K$ such that wa is the infimum of wJ in wK. Then $\widehat{wJ} = (wa)^-$ whose preimage is $va^- + \mathcal{G}(vJ)$. By Lemma 3.8(2), this is equal to $va\mathcal{O}(I)$. It follows that $I_{(vJ)^\diamond} = I_{va\,O(I)} = a\,O(I) = J^\diamond$. This proves our assertion.

(2) This follows from part (1) and Corollary 2.32. □

We now consider solutions J of (6).

Proposition 3.12. *Take $\mathcal{O}_v$-ideals I_1 and I_2 and assume that $J = J_0$ is a solution of* (6). *Set $J_1 := J_0\mathcal{O}(I_2)$. Then the following statements hold.*

(1) *If $I_1\mathcal{O}(I_2)$ is a principal $\mathcal{O}(I_2)$-ideal, then $I_2 : I_1 = J_1$.*

(2) *If $I_1\mathcal{O}(I_2)$ is a nonprincipal $\mathcal{O}(I_2)$-ideal, then $I_2 : I_1 = J_1^\diamond$.*

Proof. Denote by w the valuation associated with $\mathcal{O}(I_2)$. We have $H(\mathcal{O}(I_2)) = \mathcal{G}(vI_2)$ by (47). From Lemma 3.5 it follows that $H(\mathcal{O}(I_2)) = v\mathcal{O}(I_2)^\times$ and therefore, $v\mathcal{O}(I_2)/H(\mathcal{O}(I_2)) = (0/H(\mathcal{O}(I_2))^-$. Hence,

$$\begin{aligned} wI_1\mathcal{O}(I_2) &= v(I_1\mathcal{O}(I_2))/H(\mathcal{O}(I_2)) \\ &= vI_1/H(\mathcal{O}(I_2)) + v\mathcal{O}(I_2)/H(\mathcal{O}(I_2)) \\ &= vI_1/H(\mathcal{O}(I_2)) + (0/H(\mathcal{O}(I_2))^- \\ &= vI_1/H(\mathcal{O}(I_2)) = vI_1/\mathcal{G}(vI_2)\,. \end{aligned}$$

By Lemma 3.2(2), $I_1\mathcal{O}(I_2)$ is a principal $\mathcal{O}(I_2)$-ideal if and only if $wI_1\mathcal{O}(I_2)$ is principal. Further, using (49), we find that $vJ_0\mathcal{O}(I_2) = vJ_0 + H(\mathcal{O}(I_2)) = vJ_0 + \mathcal{G}(vI_2)$. Now statements 1) and 2) follow from Proposition 3.11(1) together with Proposition 2.30 where $S_1 = vI_1$, $S_2 = vI_2$, and $T_0 = vJ_0$. □

From Theorem 2.33 we deduce:

Theorem 3.13. *Take $\mathcal{O}_v$-ideals I_1 and I_2.*

(1) *Assume that*

$$\begin{aligned} &\mathcal{O}(I_1) \subseteq \mathcal{O}(I_2), \text{ and } I_2 \text{ is a nonprincipal} \\ &\mathcal{O}(I_2)\text{-ideal if } I_1 \text{ is a nonprincipal } \mathcal{O}(I_2)\text{-ideal.} \end{aligned} \tag{$*$}$$

Then $J = I_2 : I_1$ is the largest solution of (6); *moreover $T = a^{-1}I_2$ if $I_1\mathcal{O}(I_2) = a\mathcal{O}(I_2)$, and $T = (I_2(\mathcal{O}_v : I_1))^\diamond$ if $I_1\mathcal{O}(I_2)$ is a nonprincipal $\mathcal{O}(I_2)$-ideal.*

(2) *Assume instead that condition* ($*$) *does not hold*; *equivalently*, *that either*

(i) $\mathcal{O}(I_2) \subsetneq \mathcal{O}(I_1)$ *or*

(ii) $\mathcal{O}(I_1) = \mathcal{O}(I_2)$ *and* I_1 *is a nonprincipal* $\mathcal{O}(I_1)$*-ideal, but* I_2 *is a principal* $\mathcal{O}(I_1)$*-ideal.*

Then (6) *has no solution.*

Proof. Since $\mathcal{O}(I) = \mathcal{O}(\mathcal{G}(vI))$ for every $\mathcal{O}_v$-ideal I by (46), we know that $\mathcal{O}(I_1) \subseteq \mathcal{O}(I_2)$ if and only if $\mathcal{G}(vI_1) \subseteq \mathcal{G}(vI_2)$. Part (2) of Lemma 3.2, with the valuation w associated with $\mathcal{O}(I_2)$ in place of v, shows for $i = 1, 2$ that I_i is a principal $\mathcal{O}(I_2)$-ideal if and only if $wI_i = vI_i/\mathcal{G}(I_2)$ is principal. Further, if $vI_1/\mathcal{G}(I_2)$ is nonprincipal, then also I_1 is nonprincipal and therefore, $vI_2(\mathcal{O}_v : I_1) = vI_2 - (vI_1)^c$ by (36).

By Lemma 3.2(1), $I_1 J = I_2$ holds if and only if $vI_1 + vJ = vI_2$. Now all of our assertions follow from Theorem 2.33, where $S_i = vI_i$ for $i = 1, 2$. □

The following theorem addresses question (QID4) in the case where no solution to (6) exists.

Theorem 3.14. *Take* $\mathcal{O}$*-ideals* I_1 *and* I_2*.*

(a) *Assume that* $\mathcal{O}(I_2) \subsetneq \mathcal{O}(I_1)$*. Then there exists* $a \in I_2$ *such that* $I_2' := a\mathcal{M}(I_1) \subsetneq I_2$ *is the largest* $\mathcal{O}(I_1)$*-ideal contained in* I_2*.*

(b) *Assume that* $\mathcal{O}(I_1) = \mathcal{O}(I_2)$ *and* I_1 *is a nonprincipal* $\mathcal{O}(I_1)$*-ideal, but* I_2 *is a principal* $\mathcal{O}(I_1)$*-ideal, say* $I_2 = a\mathcal{O}(I_1)$ *for some* $a \in K$*. In this case,* $I_2' := a\mathcal{M}(I_1) \subsetneq I_2$ *is the largest nonprincipal* $\mathcal{O}(I_1)$*-ideal contained in* I_2*.*

In both cases, $J = I_2' : I_1$ *is the largest solution of* $I_1 J = I_2'$*, and* $I_2' : I_1 = I_2 : I_1$*.*

Proof. This is a consequence of Theorems 2.34 and 3.13. We just need to observe that $(va + \mathcal{G}(vI_1))^+ = va + \mathcal{G}(vI_1)^+$, whence $I_{(va+\mathcal{G}(vI_1))^+} = aI_{\mathcal{G}(vI_1)^+} = a\mathcal{M}(I_1)$, where the last equality holds since $\mathcal{G}(vI_1)^+ = v\mathcal{M}(I_1)$ by (48). □

3.5. ***Annihilators of quotients*** $\boldsymbol{I_1/I_2}$**.** Take $\mathcal{O}_v$-ideals I_1 and I_2. The *annihilator* of the quotient I_1/I_2 with $I_2 \subsetneq I_1$ is the $\mathcal{O}_v$-ideal $I_2 : I_1$. In general, the equation $I_1 J = I_2$ is not solvable and $J = I_2 : I_1$ will not be a solution, but it will be the largest $\mathcal{O}_v$-ideal such that $I_1 J \subseteq I_2$. However, in the situation appearing in the papers [3; 2], we have $I_1 = U$ and $I_2 = UV$ for $\mathcal{O}_v$-ideals U, V. Then $J = V$ is a solution of $I_1 J = I_2$ and our task is just to compute the largest solution $I_2 : I_1$, which we accomplish in the next theorem.

Theorem 3.15. *Take* $\mathcal{O}_v$*-ideals* I *and* J*. Then*

$$IJ : I = \begin{cases} J\mathcal{O}(IJ) & \text{if } I\mathcal{O}(IJ) \text{ is a principal } \mathcal{O}(IJ)\text{-ideal,} \\ (J\mathcal{O}(IJ))^{\diamond} & \text{if } I\mathcal{O}(IJ) \text{ is a nonprincipal } \mathcal{O}(IJ)\text{-ideal.} \end{cases}$$

In the special case where $\mathcal{O}(I) = \mathcal{O}(J)$, we have $\mathcal{O}(IJ) = \mathcal{O}(I)$, $I\mathcal{O}(IJ) = I$ and $J\mathcal{O}(IJ) = J$, so that $IJ : I = J$ if I is a principal $\mathcal{O}(I)$-ideal, and $IJ : I = J^\diamond$ otherwise.

Proof. Our assertions follow from Proposition 3.12 with $I_1 = I$, $I_2 = IJ$ and $J_0 = J$. If $\mathcal{O}(I) = \mathcal{O}(J)$, then $\mathcal{O}(IJ) = \mathcal{O}(I) \cup \mathcal{O}(J) = \mathcal{O}(I) = \mathcal{O}(J)$ by Lemma 3.7(1), which yields our last assertion. □

Lemma 3.16. *Take $\mathcal{O}_v$-ideals I_1 and I_2 with $I_2 \subsetneq I_1$. If the annihilator of I_1/I_2 is $\mathcal{M}_v$, then I_1 is a principal $\mathcal{O}_v$-ideal.*

Proof. If I_1 is nonprincipal, then vI_1 is nonprincipal and by Lemma 2.6(3) we obtain $vI_1 = vI_1 + 0^+ = vI_1 + v\mathcal{M}_v$, whence $I_1 = I_1\mathcal{M}_v$. As $\mathcal{M}_v$ annihilates I_1/I_2, it follows that $I_1 = I_1\mathcal{M}_v \subseteq I_2$, which contradicts our assumption. □

Proposition 3.17. (1) *Take $\mathcal{O}_v$-ideals I_1 and I_2 with $I_2 \subsetneq I_1$. Then the annihilator of I_1/I_2 is equal to $\mathcal{M}_v$ if and only if I_1 is principal and $I_2 = I_1\mathcal{M}_v$.*

(2) *Take $\mathcal{O}_v$-ideals I and J with $J \subsetneq \mathcal{O}_v$. Then the annihilator of I/IJ is equal to $\mathcal{M}_v$ if and only if I is principal and $J = \mathcal{M}_v$.*

Proof. (1) Assume that I_1 is principal and $I_2 = I_1\mathcal{M}_v$. Then $\mathcal{M}_v$ annihilates I_1/I_2. On the other hand, $I_2 = I_1\mathcal{M}_v \subsetneq I_1$, so I_1/I_2 is not zero and $\mathcal{O}_v$ does not annihilate it. Thus, $\mathcal{M}_v$ is the annihilator of I_1/I_2.

In order to prove the converse, assume that $\mathcal{M}_v$ is the annihilator of I_1/I_2. From Lemma 3.16 we infer that I_1 must be principal, say $I_1 = (a)$ for some $a \in K$. As $\mathcal{M}_v$ annihilates I_1/I_2, we have $a\mathcal{M}_v = I_1\mathcal{M}_v \subseteq I_2$. We cannot have $a\mathcal{M}_v \subsetneq I_2$ since then $I_1 = (a) \subseteq I_2$, which contradicts our assumption. Thus we have $I_2 = a\mathcal{M}_v = I_1\mathcal{M}_v$.

(2) By part (1), applied to $I_1 = I$ and $I_2 = IJ$, the annihilator of I/IJ is equal to $\mathcal{M}_v$ if and only if I is principal and $IJ = I\mathcal{M}_v$. As I is principal, the latter is equivalent to $J = \mathcal{M}_v$. □

Finally, we treat the general case.

Theorem 3.18. *Take $\mathcal{O}_v$-ideals I_1, I_2 with $I_2 \subsetneq I_1$.*

(1) *Assume that* (6) *has a solution. Then the annihilator of I_1/I_2 is $a^{-1}I_2$ if $I_1\mathcal{O}(I_2) = a\mathcal{O}(I_2)$, and is $I_2(\mathcal{O}_v : I_1)^\diamond$ if $I_1\mathcal{O}(I_2)$ is a nonprincipal $\mathcal{O}(I_2)$-ideal.*

(2) *Assume that* (6) *does not have a solution, and let I_2' be as defined in Theorem 3.14. Then the annihilator of I_1/I_2 is $a^{-1}I_2'$ if $I_1\mathcal{O}(I_2) = a\mathcal{O}(I_2)$, and is $I_2'(\mathcal{O}_v : I_1)^\diamond$ if $I_1\mathcal{O}(I_2)$ is a nonprincipal $\mathcal{O}(I_2)$-ideal.*

Proof. (1) This follows from Theorem 3.13.

(2) This follows from Theorems 3.14 and 3.13. □

3.6. ***Application to the computation of annihilators in special cases.*** For an $\mathcal{O}_v$-module M we denote by $\operatorname{ann} M$ the annihilator of M.

Proposition 3.19. *Take an $\mathcal{O}_v$-ideal $I \subsetneq \mathcal{O}_v$ and $n \in \mathbb{N}$, and set $J = bI^{n-1}$, where $b \in \mathcal{O}_v$. If $n = 1$, assume that $b \in \mathcal{M}_v$. Then $I/IJ = I/bI^n$ is not zero, and the following statements hold.*

(1) *If there is $a \in K$ such that $vJ/\mathcal{G}(vJ)$ has infimum $va/\mathcal{G}(vJ)$ in $vK/\mathcal{G}(vJ)$ but does not contain this infimum, then*

$$\operatorname{ann} I/IJ = J^{\diamond} = a\mathcal{O}(J), \tag{50}$$

which properly contains J. In all other cases, $\operatorname{ann} I/IJ = J$.

In all cases, $\mathcal{M}(J) \operatorname{ann} I/IJ \subseteq J$.

(2) *We have that* $\operatorname{ann} I/IJ = \mathcal{M}_v$ *if and only if $\mathcal{M}_v$ is principal and either $I = J = \mathcal{M}_v$ (which holds if and only if $n = 2$ and $b \in \mathcal{O}_v^{\times}$), or $I = \mathcal{O}_v$ and $J = \mathcal{M}_v = (b)$ (which holds if and only if $n = 1$ and $b \notin \mathcal{O}_v^{\times}$).*

Proof. Our assumption that $b \in \mathcal{M}_v$ in the case of $n = 1$ and that $b \in \mathcal{O}_v$ and $I \subsetneq \mathcal{O}_v$ in the case of $n > 1$ implies that $J \subsetneq \mathcal{O}_v$. This yields that I/IJ is not zero.

Since $vJ = vb + (n-1)vI$, Lemma 2.13(3) together with Corollary 2.14 shows that $\mathcal{G}(vI) = \mathcal{G}(vJ)$, which implies that $\mathcal{O}(I) = \mathcal{O}(J)$.

(1) From Theorem 3.15 we infer that $\operatorname{ann} I/IJ = J$ if I is a principal $\mathcal{O}(J)$-ideal, and $\operatorname{ann} I/IJ = J^{\diamond}$ if I is a nonprincipal $\mathcal{O}(J)$-ideal. Assume first that there is $a \in K$ such that $vJ/\mathcal{G}(vJ)$ has infimum $va/\mathcal{G}(vJ)$ in $vK/\mathcal{G}(vJ)$ but does not contain this infimum. Then $vJ/\mathcal{G}(vJ)$ is nonprincipal, and the same must hold for $vI/\mathcal{G}(vJ)$, so that I is a nonprincipal $\mathcal{O}(J)$-ideal. It follows that $\operatorname{ann} I/IJ = J^{\diamond} = a\mathcal{O}(J)$. Now assume the remaining case, that is, $vJ/\mathcal{G}(vJ)$ is closed in $G/\mathcal{G}(vJ)$ and therefore, $J^{\diamond} = J$. In this case we obtain that $\operatorname{ann} I/IJ = J$.

Finally, we prove the last assertion. Since $\mathcal{M}(J) \subseteq \mathcal{M}_v$, it is trivial when $\operatorname{ann} I/IJ = J$. In the remaining case, $\operatorname{ann} I/IJ = a\mathcal{O}(J)$, whence $\mathcal{M}(J) \operatorname{ann} I/IJ = a\mathcal{M}(J)$. Since $va/\mathcal{G}(vJ) = wa$ is the infimum of $vJ/\mathcal{G}(vJ) = wJ$ in $vK/\mathcal{G}(vJ) = wK$ and $wc > 0$ for every $c \in \mathcal{M}(J)$ and therefore $wac > wa$, it follows that $wa\mathcal{M}(J) \subseteq wJ$. Hence $w\mathcal{M}(J) \operatorname{ann} I/IJ \subseteq wJ$, so $\mathcal{M}(J) \operatorname{ann} I/IJ \subseteq J$.

(2) By Proposition 3.17(2), the annihilator of I/IJ is equal to $\mathcal{M}_v$ if and only if I is principal and $J = \mathcal{M}_v$. Assume first that I is principal and $J = \mathcal{M}_v$. Since $J = bI^{n-1}$, it follows that J and $\mathcal{M}_v$ are principal. Since I is principal, vI is principal and $\mathcal{G}(vJ) = \mathcal{G}(vI) = \{0\}$, so $vJ/\mathcal{G}(vI) = vJ$ is principal and from part (1) we obtain that $\mathcal{M}_v = \operatorname{ann} I/IJ = J = bI^{n-1} \subseteq I$. Thus either $I = \mathcal{M}_v = J$ or $I = \mathcal{O}_v$ and $J = \mathcal{M}_v = (b)$. Since $\mathcal{M}_v$ is principal, the former holds if and only if $n = 2$ and $b \in \mathcal{O}_v^{\times}$, and the latter holds if and only if $n = 1$ and $b \notin \mathcal{O}_v^{\times}$.

Now assume that $\mathcal{M}_v$ is principal and either $I = \mathcal{M}_v = J$ or $I = \mathcal{O}_v$ and $J = \mathcal{M}_v = (b)$. If the former holds and $\mathcal{M}_v = (c)$ for some $c \in \mathcal{M}_v$, then $I/IJ = (c)/(c^2)$ and $\operatorname{ann} I/IJ = (c) = \mathcal{M}_v$. If the latter holds, then $I/IJ = \mathcal{O}/(b)$ and $\operatorname{ann} I/IJ = (b) = \mathcal{M}_v$. □

The following facts are used in [2].

Lemma 3.20. *Take $n \geq 2$, any valuation ring $\mathcal{O}$ with maximal ideal $\mathcal{M}$, and $a \in \mathcal{O}$. Then*:

(1) *$a\mathcal{M} = \mathcal{M}$ if and only if $a \notin \mathcal{M}$.*

(2) *$\mathcal{M}^n = \mathcal{M}$ if and only if $\mathcal{M}$ is a nonprincipal $\mathcal{O}$-ideal.*

(3) *$(a\mathcal{M})^n = a\mathcal{M}$ if and only if $a \notin \mathcal{M}$ and $\mathcal{M}$ is a nonprincipal $\mathcal{O}$-ideal.*

Proof. (1) We have $a \notin a\mathcal{M}$, hence if $a \in \mathcal{M}$, then $a\mathcal{M} \neq \mathcal{M}$. If $a \notin \mathcal{M}$, then a is a unit in $\mathcal{O}$, so $a\mathcal{M} = \mathcal{M}$.

(2) The value group of w is not discrete, and hence dense, if and only if $\mathcal{M}$ is a nonprincipal $\mathcal{O}$-ideal. If it is discrete and γ is its smallest positive element, then $\mathcal{M} = \{b \in K \mid wb \geq \gamma\}$ and $\mathcal{M}^n = \{c \in K \mid wc \geq n\gamma\} \subsetneq \mathcal{M}$ since $n\gamma > \gamma$. If it is dense, then for every $b \in \mathcal{M}$ there is $c \in K$ such that $0 < nwc < wb$, whence $b \in \mathcal{M}^n$; therefore, $\mathcal{M}^n \subseteq \mathcal{M} \subseteq \mathcal{M}^n$ and consequently, $\mathcal{M}^n = \mathcal{M}$.

(3) If $a \notin \mathcal{M}$ and $\mathcal{M}$ is a nonprincipal $\mathcal{O}$-ideal, then by parts (1) and (2), $(a\mathcal{M})^n = \mathcal{M}^n = \mathcal{M} = a\mathcal{M}$. If $a \in \mathcal{M}$, then $wa > 0$, whence $a\mathcal{M} = \{c \in K \mid wc > wa\}$ and $(a\mathcal{M})^n \subseteq a^n\mathcal{M} = \{c \in K \mid wc > nwa\} \subsetneq a\mathcal{M}$ since $nwa > wa$. If $\mathcal{M}$ is a principal $\mathcal{O}$-ideal, say $\mathcal{M} = b\mathcal{O}$ with $b \in \mathcal{M}$, then $a\mathcal{M} = ab\mathcal{O} = \{c \in K \mid wc \geq wab\}$ and $(a\mathcal{M})^n = (ab\mathcal{O})^n = \{c \in K \mid wc \geq nwab\} \subsetneq a\mathcal{M}$ since $nwab > wab$ and $ab \in a\mathcal{M}$. □

Proposition 3.21. *Take $n \geq 2$, $a \in \mathcal{O}_v$ and $\mathcal{O} \in \mathcal{R}$ with maximal ideal $\mathcal{M}$. Assume that $(a\mathcal{M})^n \neq a\mathcal{M}$. Then the following statements hold.*

(1) $\operatorname{ann} a\mathcal{M}/(a\mathcal{M})^n = \begin{cases} (a\mathcal{M})^{n-1} & \text{if } \mathcal{M} \text{ is a principal } \mathcal{O}\text{-ideal,} \\ (a\mathcal{O})^{n-1} = a^{n-1}\mathcal{O} & \text{if } \mathcal{M} \text{ is a nonprincipal } \mathcal{O}\text{-ideal.} \end{cases}$

(2) *The annihilator is equal to $\mathcal{M}_v$ if and only if $n = 2$, $a \notin \mathcal{M}_v$ and $\mathcal{M} = \mathcal{M}_v$ is a principal $\mathcal{O}_v$-ideal.*

Proof. (1) We apply Theorem 3.15 with $I = a\mathcal{M}$ and $J = (a\mathcal{M})^{n-1}$. Then $\mathcal{O}(IJ) = \mathcal{O}((a\mathcal{M})^n) = \mathcal{O}(a\mathcal{M}) = \mathcal{O}(\mathcal{M}) = \mathcal{O}$ by parts (2) and (3) of Lemma 3.7 and (45). Since I and J are already $\mathcal{O}$-ideals, we obtain that $I\mathcal{O}(IJ) = I$ and $J\mathcal{O}(IJ) = J$. We observe that $I = a\mathcal{M}$ is a principal $\mathcal{O}$-ideal if and only if the same is true for $\mathcal{M}$.

It remains to show that $J\mathcal{O}(IJ)^\diamond = J^\diamond$ is equal to $(a\mathcal{O})^{n-1}$ if $\mathcal{M}$ is a nonprincipal $\mathcal{O}$-ideal. Denote the valuation associated with $\mathcal{O}$ by w. In the present case, $J =$

$(a\mathcal{M})^{n-1} = a^{n-1}\mathcal{M}$, and 0 is the infimum of $w\mathcal{M}$, not contained in $w\mathcal{M}$. Therefore, $wa^{n-1} \notin wa^{n-1} + w\mathcal{M} = wa^{n-1}\mathcal{M}$ is the infimum of $wa^{n-1}\mathcal{M}$. Hence by definition, $J^{\diamond} = \{c \mid wc \geq wa^{n-1}\} = a^{n-1}\mathcal{O} = (a\mathcal{O})^{n-1}$. This completes the proof of part (1).

(2) Assume first that $\operatorname{ann} a\mathcal{M}/(a\mathcal{M})^n = \mathcal{M}_v$. Then $a\mathcal{M}/(a\mathcal{M})^n$ is nontrivial, whence $a\mathcal{M} \neq (a\mathcal{M})^n$, which by Lemma 3.20(3) means that $a \in \mathcal{M}$ or $\mathcal{M}$ is a principal $\mathcal{O}$-ideal. If the latter holds, then by part (1), $a\mathcal{M} : (a\mathcal{M})^n = (a\mathcal{M})^{n-1}$. Since $a\mathcal{M} \subseteq \mathcal{M}_v$, this is equal to $\mathcal{M}_v$ if and only if $n-1 = 1$, $\mathcal{M} = \mathcal{M}_v$, and $va = 0$, where we have used parts (1) and (3) of Lemma 3.20. On the other hand, if the latter does not hold, then $a \in \mathcal{M}$ and by part (1), $\operatorname{ann} a\mathcal{M}/(a\mathcal{M})^n = (a\mathcal{O})^{n-1} \subseteq a\mathcal{O}$. The latter is contained in $\mathcal{M}$ but is not equal to $\mathcal{M}$ since $\mathcal{M}$ is a nonprincipal $\mathcal{O}$-ideal. Thus $\operatorname{ann} a\mathcal{M}/(a\mathcal{M})^n = a\mathcal{M} : (a\mathcal{M})^n \subsetneq \mathcal{M} \subseteq \mathcal{M}_v$ and consequently, $\operatorname{ann} a\mathcal{M}/(a\mathcal{M})^n \neq \mathcal{M}_v$.

For the converse, assume that $n = 2$, $a \notin \mathcal{M}_v$ and $\mathcal{M} = \mathcal{M}_v$ is a principal $\mathcal{O}_v$-ideal. Then part (1) together with Lemma 3.20(1) yield $\operatorname{ann} a\mathcal{M}/(a\mathcal{M})^n = \operatorname{ann} a\mathcal{M}_v/(a\mathcal{M}_v)^n = (a\mathcal{M}_v)^{n-1} = a\mathcal{M}_v = \mathcal{M}_v$. □

References

[1] S. Bazzoni and L. Salce, "Groups in the class semigroups of valuation domains", *Israel J. Math.* **95** (1996), 135–155. MR

[2] S. D. Cutkosky and F.-V. Kuhlmann, "Kähler differentials of extensions of valuation rings and deeply ramified fields", 2023. arXiv 2306.04967

[3] S. D. Cutkosky, F.-V. Kuhlmann, and A. Rzepka, "On the computation of Kähler differentials and characterizations of Galois extensions with independent defect", *Math. Nachr.* **298**:5 (2025), 1549–1577. MR

[4] A. J. Engler and A. Prestel, *Valued fields*, Springer, 2005. MR

[5] A. Fornasiero and M. Mamino, "Arithmetic of Dedekind cuts of ordered abelian groups", *Ann. Pure Appl. Logic* **156**:2-3 (2008), 210–244. MR

[6] A. Fornasiero, F.-V. Kuhlmann, and S. Kuhlmann, "Towers of complements to valuation rings and truncation closed embeddings of valued fields", *J. Algebra* **323**:3 (2010), 574–600. MR

[7] L. Fuchs and L. Salce, *Modules over non-Noetherian domains*, Mathematical Surveys and Monographs **84**, Amer. Math. Soc., Providence, RI, 2001. MR

[8] F.-V. Kuhlmann, "Selected methods for the classification of cuts, and their applications", pp. 85–106 in *Algebra, logic and number theory*, Banach Center Publ. **121**, Polish Acad. Sci., Warsaw, 2020. MR

[9] F.-V. Kuhlmann, "Invariance group and invariance valuation ring of a cut", preprint, available at https://www.fvkuhlmann.de/CUTS.pdf.

[10] E. Olivos and W. H. Schikhof, "Extending the multiplication of a totally ordered group to its completion", pp. 231–242 in *Advances in non-Archimedean analysis*, Contemp. Math. **551**, Amer. Math. Soc., Providence, RI, 2011. MR

[11] P. Ribenboim, "Sur les groupes totalement ordonnés et l'arithmétique des anneaux de valuation", *Summa Brasil. Math.* **4** (1958), 1–64. MR

[12] O. Zariski and P. Samuel, *Commutative algebra*, vol. 2, Van Nostrand, Princeton, 1960. MR

Received April 9, 2025. Revised October 6, 2025.

FRANZ-VIKTOR KUHLMANN
INSTITUTE OF MATHEMATICS
UNIVERSITY OF SZCZECIN
70-451 SZCZECIN
POLAND
fvk@usz.edu.pl

KATARZYNA KUHLMANN
PATER-HERTLE-WEG 5A
83727 SCHLIERSEE
GERMANY
katka314@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.375

THE RICCI–DETURCK FLOW FROM AN INITIAL METRIC WITH A MORREY-TYPE INTEGRABILITY CONDITION

MAN-CHUN LEE AND STEPHEN SHANG YI LIU

We study the short-time existence theory of Ricci–DeTurck flow starting from rough metrics that satisfy a Morrey-type integrability condition. Using the rough existence theory, we show the preservation and improvement of distributional scalar curvature lower bounds provided the singular set for such metrics is not too large. As an application, we use Ricci flow smoothing to study the removable singularity related to scalar curvature. Our result supplements those of Jiang, Sheng and Zhang.

1. Introduction

Let M^n be a Riemannian manifold. The Ricci flow is a family of metrics $\hat{g}(t)$ evolving in the direction of their Ricci tensors,

$$\frac{\partial}{\partial t}\hat{g}(t) = -2\,\mathrm{Ric}(\hat{g}(t)).$$

Introduced by Hamilton in [7], the Ricci flow has seen successful applications to a number of problems in geometry, most famously in Perelman's resolution of Thurston's geometrization conjecture. It is a weakly parabolic system, and in [5], DeTurck showed that it is diffeomorphic to a strictly parabolic system, the Ricci–DeTurck h-flow. Precisely, let $\mathrm{Sym}_2(T^*M)$ denote the space of symmetric 2-tensors on M and let $X : \mathrm{Sym}_2(T^*M) \to TM$ be given by

$$X_h(g)^k := g^{ij}(\tilde{\Gamma}^k_{ij} - \Gamma^k_{ij}),$$

where $\tilde{\Gamma}$ and Γ are the Christoffel symbols of h and g respectively. Then the Ricci–DeTurck h-flow (here using the terminology as in [18] to emphasize the dependence on the background metric h) is given by the equation

$$\frac{\partial}{\partial t}g(t) = -2\,\mathrm{Ric}(g(t)) - \mathcal{L}_{X_h(g(t))}g(t) \tag{1.0.1}$$

MSC2020: 53E20.
Keywords: Ricci flow, Ricci–DeTurck flow, scalar curvature rigidity, singular metrics.

© 2026 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Solutions to the Ricci–DeTurck h-flow are related to a Ricci flow via pullback by diffeomorphisms, that is, if $g(t)$ solves (1.0.1) and $\chi_t : M \to M$ is a family of diffeomorphisms satisfying

$$\begin{cases} X_{h(t)}(g(t))f = \dfrac{\partial}{\partial t}(f \circ \chi_t) \text{ for all } f \in C^\infty(M), \\ \chi_0 = \mathrm{id}, \end{cases} \tag{1.0.2}$$

then $\chi_t^* g(t)$ is a Ricci flow with initial data $g(0)$.

In local coordinates, (1.0.1) is given by

$$\begin{aligned} \frac{\partial}{\partial t} g_{ij} = {} & g^{pq}\tilde{\nabla}_p \tilde{\nabla}_q g_{ij} - g^{k\ell} g_{ip} h^{pq} \tilde{R}_{jkq\ell} - g^{k\ell} g_{jp} h^{pq} \tilde{R}_{ikq\ell} \\ & + \tfrac{1}{2} g^{k\ell} g^{pq} \big(\tilde{\nabla}_i g_{pk} \tilde{\nabla}_j g_{q\ell} + 2\tilde{\nabla}_k g_{jp} \tilde{\nabla}_q g_{i\ell} - 2\tilde{\nabla}_k g_{jp} \tilde{\nabla}_\ell g_{iq} \\ & - 2\tilde{\nabla}_j g_{pk} \tilde{\nabla}_\ell g_{iq} - 2\tilde{\nabla}_i g_{pk} \tilde{\nabla}_\ell g_{qj} \big), \end{aligned} \tag{1.0.3}$$

where $\tilde{\nabla}$ and $\tilde{R}$ denote respectively the Levi-Civita connection and Riemann curvature tensor of h and we are suppressing the dependence on t for $g(t)$ in the notation above.

Since the flow is a parabolic system, one usually expects that it will improve the regularity of initial data. In view of this, there has been a wide variety of literature extending the existence theory of Ricci Flow and Ricci–DeTurck h-flow to settings with less regularity. The foundational result in this area is by Shi in [16], who showed the short-time existence of a solution to the Ricci–DeTurck flow from a metric that is complete with bounded curvature. A metric g is called L^∞ or bi-Lipschitz if $\Lambda^{-1}h \leq g \leq \Lambda h$ almost everywhere on M with $\Lambda > 1$ and h is some fixed smooth background metric. Building on the result of Shi, Simon in [18] established the following regularization result: if $(1+\delta)^{-1}h \leq g_0 \leq (1+\delta)h$ for some sufficiently small dimensional constant δ, then there exists a short-time solution to the Ricci–DeTurck h-flow on $M \times (0, T]$ which is smooth for $t > 0$ with convergence back to g_0 in C^0_{loc} if g_0 was additionally in C^0_{loc}. In Euclidean space, Koch and Lamm in [10] used heat kernel estimates to show the existence and uniqueness of a global and analytic solution to the Ricci–DeTurck h-flow on $\mathbb{R}^n$ from L^∞ initial metrics that are close in L^∞ to the Euclidean metric. Building on these results, Burkhardt–Guim in [3] further studied Ricci flow from C^0 initial metrics on closed manifolds, in particular relying on heat kernel estimates to establish an iteration scheme to construct a short-time solution and applied this Ricci flow existence theory to study scalar curvature lower bounds and a question of Gromov. In the direction of more general L^∞ initial data, Lamm and Simon in [11] showed the short-time existence for the Ricci–DeTurck h-flow on complete four-manifolds for rough initial metrics that are in $L^\infty \cap W^{2,2}$. They showed that solutions are uniformly smooth for positive time and converge back to the initial data in $W^{2,2}_{\mathrm{loc}}$

sense as $t \to 0$. In [4], Chu and the first named author developed this further and considered the case where $g_0 \in L^\infty \cap W^{1,n}$, showing the existence of the flow starting from metrics which are L^∞ and satisfy a small local gradient concentration in a $W^{1,n}$ sense, and study a number of applications.

Motivated by the work [4], and the heat kernel based approach of Burkhardt–Guim [3], in this work we consider the existence theory for metrics that are in L^∞ and satisfy a L^p ($p \geq 1$) Morrey-type integrability condition (condition (iii) of Theorem 1.1 below) and study a removable singularity theorem related to scalar curvature. We will see that the Morrey-type condition allows for the use of heat kernel estimates to establish a priori estimates that are important for the existence theory. Our main theorem is the following existence theorem for the smooth h-flow with quantitative estimates.

Theorem 1.1. *Suppose (M^n, h) is a complete Riemannian manifold satisfying* (2.0.1). *Suppose g_0 is a $C^0_{\mathrm{loc}} \cap W^{1,2}_{\mathrm{loc}}$ Riemannian metric on M and $\Sigma \subseteq M$ is a compact set so that the following holds*:

(i) *g_0 is smooth on $M \setminus \Sigma$.*

(ii) *g_0 is globally uniformly bi-Lipschitz on M; that is, there exists $\Lambda_0 > 1$ such that*

$$\Lambda_0^{-1} h \leq g_0 \leq \Lambda_0 h \quad \text{on } M.$$

(iii) *There exist $L_0, \delta, r_0 > 0$ and $p \geq 1$ such that for every $x_0 \in M$ and $0 < r < r_0$*

$$\fint_{B(x_0,r)} |\tilde{\nabla} g_0|^p \, d\mu_h \leq L_0 r^{-p+\delta}.$$

Then there are $T, L > 0$ depending only on $n, \Lambda_0, L_0, p, \delta, r_0, h$ and a smooth solution $g(t)$ to the Ricci–DeTurck h-flow on $M \times (0, T]$ such that:

(a) *for any $k \in \mathbb{N}$, there is $B_k(n, k, r_0, \Lambda_0, L_0, p, \delta, h) > 0$ such that for all $t \in (0, T]$, $x_0 \in M$,*

$$\sup_M |\tilde{\nabla}^k g(t)| \leq \frac{B_k}{t^{\frac{1}{2}(k-\delta/p)}};$$

(b) *$g(t) \to g_0$ in $C^0_{\mathrm{loc}}(M)$ as $t \to 0$;*

(c) *$g(t) \to g_0$ in $C^\infty_{\mathrm{loc}}(M \setminus \Sigma)$ as $t \to 0$;*

Moreover, the solution is unique within the class of solutions satisfying (a).

The result is new even if $\Sigma = \varnothing$. Our work is motivated by the removable singularity in view of scalar curvature. In [13], Lee and LeFloch introduced a notion of distributional scalar curvature for metrics with lower regularity. In [8], the authors showed that for metrics $g \in C^0 \cap W^{1,p}_{\mathrm{loc}}(M)$ where M is complete asymptotically flat with $n \leq p \leq \infty$ satisfying $R_g \geq 0$ outside some compact singular set Σ also has nonnegative ADM mass (with rigidity) provided the singular

set is not too large in terms of Hausdorff measure. In the compact case, the authors in [9] also show that distributional scalar curvature lower bounds can be preserved and improved to classical scalar curvature lower bounds along Ricci flow for metrics in $W^{1,p}(M)$ for $n < p \leq \infty$. These results are related to positive mass theorems for lower regularity metrics and a conjecture of Schoen regarding the removable singularity of L^∞ metrics on $\mathbb{T}^n$ with codimension three singularity and $R_g \geq 0$ on the regular part. In this work, we are primarily interested in the following question: if we strengthen the regularity from C^0 slightly to the Morrey-type condition on the gradient, then under what conditions on the singular set could we expect to have scalar curvature rigidity. In Theorem 4.6 and Corollary 4.7, we use the rough existence theory established in Theorem 1.1 to obtain similar results for metrics which satisfy the Morrey-type condition, as long as the singular set is small in terms of a notion of codimension considered in [14].

We now outline the rest of this work. In Section 2, we establish the a priori estimates that will be used later. In Section 3, we prove Theorem 1.1. Finally, in Section 4, we study the application outlined above.

2. A priori estimates

In this section, we establish quantitative estimates that will be important in the proof of the main theorem. In general, we consider rough metrics g_0 inside a smooth background manifold (M, h) with $\sup_M |\mathrm{Rm}(h)| < +\infty$. Thanks to Shi's smoothing result [16], in our content we will as well assume

$$\sup_M |\tilde{\nabla}^i \mathrm{Rm}(h)| := k_i < +\infty \ \ \forall i \in \mathbb{N}, \tag{2.0.1}$$

where $\tilde{\nabla}$ denotes the connection of h. In the following and the rest of this work, we denote by $k_i := \sup_M |\tilde{\nabla}^i \mathrm{Rm}(h)|$ and we will specify using parentheses the quantities that certain numbered constants depend upon, e.g., $C_j = C_j(n, k_1, \ldots, k_i) > 0$ denotes a positive constant C_j that depends on the dimension and quantities k_1, up to k_i. These constants may be changing line by line throughout and we will only re-introduce the parentheses when the dependency changes. All distances, norms and connection are measured with respect to the metric h, unless noted otherwise. We also use $a \wedge b$ to denote $\min\{a, b\}$.

2.1. ***Parabolic boot-strapping.*** Following [16] and [18], the following lemma says that given a first order interior estimate, we may parabolically bootstrap to obtain higher order interior estimates.

Lemma 2.2. *Suppose $g(t)$ is a smooth solution to the Ricci–DeTurck h-flow on $M \times [0, T]$ for some smooth background metric h satisfying* (2.0.1) *so that*

(i) $\Lambda^{-1} h \leq g(t) \leq \Lambda h$ *on* $M \times [0, T]$ *for some* $\Lambda > 1$;

(ii) *there exists* $x_0 \in M$, $B_1, \gamma > 0$ *such that*

$$|\tilde{\nabla} g(x,t)| \le \frac{B_1}{t^{\frac{1}{2}-\frac{\delta}{2p}}}$$

for all $x \in B\left(x_0, 1+\frac{\gamma}{2}\right), t \in (0, T]$.

Then for any $m \in \mathbb{N}$, *there are* $B_m(n, \gamma, m, \Lambda, k_0, k_1, \ldots, k_m) > 0$ *such that*

$$\sup_{B\left(x_0, 1+\gamma/2^{m+1}\right)} |\tilde{\nabla}^m g(x,t)| \le \frac{B_m}{t^{\frac{1}{2}(m-\delta/p)}} \tag{2.2.1}$$

for all $t \in (0, T \wedge 1]$.

Sketch of Proof. The proof of the above Lemma is a standard induction argument, using Bernstein and Shi's trick and the nonscaling invariant order on t given in assumption (ii) above instead of the $t^{-1/2}$ found in Lemmas 4.1 and 4.2 of [18]. For $m \in \mathbb{N}$, we obtain (2.2.1) by considering the evolution equation under the operator $\partial_t - g^{ij}\tilde{\nabla}_i\tilde{\nabla}_j$ of

$$t^{2m-1-2\delta/p}|\tilde{\nabla}^m g|^2\left(|\tilde{\nabla}^{m-1} g|^2 + Lt^{-m+1+\delta/p}\right),$$

together with standard cut-off argument. □

2.3. ***Gradient estimate.*** We now show that the interior gradient estimate assumed in Lemma 2.2 can be obtained for smooth solutions of the Ricci–DeTurck h-flow under the L^∞ and Morrey-type integrability assumption. Once the C^1 estimate has been established, Lemma 2.2 gives us all higher-order estimates.

Proposition 2.4. *Suppose* $g(t)$ *is a smooth solution to the Ricci–DeTurck* h*-flow on* $M \times [0, T]$ *for some smooth initial background metric* h *satisfying* (2.0.1) *so that*

(i) $\Lambda^{-1}h \le g(t) \le \Lambda h$ *on* $M \times [0, T]$ *for some* $\Lambda > 1$;

(ii) *there exists* $p \ge 1$, $L_0, r_0, \delta > 0$, *such that for all* $x_0 \in M$ *and* $0 < r < r_0$,

$$⨍_{B(x_0,r)} |\tilde{\nabla} g(x,0)|^p \, d\mu_h(x) \le L_0 r^{-p+\delta};$$

(iii) $\sup_{M\times[0,T]} |\tilde{\nabla} g(x,t)| < +\infty$.

Then there are $B_1(n, \Lambda, L_0, k_0, k_1, k_2)$ *and* $T_1(n, \Lambda, L_0, p, \delta, k_0, k_1, k_2) > 0$ *such that*

$$\sup_M |\tilde{\nabla} g(x,t)| \le \frac{B_1}{t^{\frac{1}{2}-\frac{\delta}{2p}}} \tag{2.4.1}$$

for all $x \in M, t \in (0, T_1 \wedge T]$.

Proof. For the sake of convenience, we will suppress the dependence on t and denote $g(t)$ by g unless explicitly stated otherwise. We might assume $T \leq 1$, $p > \delta$ and $r_0 \leq 1$. We will use C to denote any constants depending only on $n, k_0, k_1, k_2, \Lambda, L_0$ which might change line by line.

Let $B_1 \geq 1$ be a constant to be chosen. We let T_1 to be the maximal time on which (2.4.1) holds. We have $T_1 > 0$ thanks to (iii). Moreover, if $T_1 < T$, then there is a $x_0 \in M$ such that

$$\begin{cases} \sup_M |\tilde{\nabla} g(x,t)| \leq \dfrac{B_1}{t^{\frac{1}{2}-\frac{\delta}{2p}}} & \text{on } (0, T_1), \\ |\tilde{\nabla} g(x_0, T_1)| \geq \dfrac{B_1}{2T_1^{\frac{1}{2}-\frac{\delta}{2p}}}. \end{cases} \tag{2.4.2}$$

It suffices to estimate T_1 from below. We assume $T_1 \leq T$; otherwise the result holds trivially.

Our first step is to compute the evolution equation of $|\tilde{\nabla} g|^2$ under the operator $\Box_g := \partial/\partial t - \Delta^g - \nabla_X$ where $\Delta^g = g^{ij}\nabla_i \nabla_j$ consists in taking derivatives with respect to the metric g instead of h. By (1.0.1) and Young's inequality,

$$\frac{\partial}{\partial t}|\tilde{\nabla} g|^2 \leq g^{ij}\tilde{\nabla}_i \tilde{\nabla}_j |\tilde{\nabla} g|^2 - C^{-1}|\tilde{\nabla}^2 g|^2 + C|\tilde{\nabla} g|^4 + C|\tilde{\nabla} g|. \tag{2.4.3}$$

Since $|\tilde{\nabla} g|$ may not be smooth, define the smooth function Q by

$$Q := \left(|\tilde{\nabla} g|^2 + 1\right)^{\frac{1}{2}},$$

which satisfies

$$\Box_g Q \leq C Q^3 + C. \tag{2.4.4}$$

Let $f(t) = \exp\left(-A(t^{\delta/p} + t)\right)$ for a constant $A > 0$ to be determined. For all $t < T_1$, (2.4.4) and (2.4.2) imply

$$\Box_g \left(f(t)Q - Ct\right) = f(t)\left(-\frac{A\delta}{p} t^{\frac{\delta}{p}-1} - A\right) Q + f(t)\Box_g Q - C \leq 0, \tag{2.4.5}$$

by choosing A sufficiently large in the second inequality above and using the fact that $f(t) \leq 1$ for all $t \geq 0$ for the last inequality. Hence, by the maximum principle we obtain

$$\tfrac{1}{4} f(T_1) B_1 T_1^{-\frac{1}{2}+\frac{\delta}{2p}} - C T_1 \leq \int_M K(x_0, T_1; y, 0) f(0) Q(y, 0)\, d\mu_h(y), \tag{2.4.6}$$

where K is the heat kernel for the heat operator $\Box_g$.

We now wish to obtain an estimate for the heat kernel K. We do this by modifying an estimate established by Bamler, Cabezas-Rivas and Wilking in [2]. Let

$\tilde{G}(x, t; y, s)$ denote the heat kernel associated to the backwards heat equation coupled with a Ricci flow $h(t)$. That is,

$$\left(\partial_s + \Delta_{y,s}^{h_t}\right)\tilde{G}(x, t; \cdot, \cdot) = 0 \quad \text{and} \quad \lim_{s \to t^-} \tilde{G}(x, t; \cdot, s) = \delta_x(\cdot).$$

Then for all $(y, s) \in M \times [0, T]$, $\tilde{G}(\cdot, \cdot; y, s)$ is the heat kernel associated to the conjugate equation

$$\left(\partial_t - \Delta_{x,t}^{h_t} - R_{h_t}\right)\tilde{G}(\cdot, \cdot; y, s) = 0 \quad \text{and} \quad \lim_{t \to s^+} \tilde{G}(\cdot, t; y, s) = \delta_y(x).$$

Proposition 3.1 of [2] gives the following estimate: for any $n, A > 0$, there is a $C(n, A) < \infty$ such that the following holds: Let $(M, h(t))$, $t \in [0, T]$ be a complete Ricci flow satisfying

$$|\operatorname{Rm}(x, t)| \leq At^{-1} \quad \text{and} \quad \operatorname{Vol}_{h_t}\left(B_{h_t}(x, \sqrt{t})\right) \geq A^{-1}t^{n/2} \tag{2.4.7}$$

for all $(x, t) \in M \times (0, T]$. Then

$$\tilde{G}(x, t; y, s) \leq \frac{C}{t^{n/2}} \exp\left(-\frac{d_s^2(x, y)}{Ct}\right) \tag{2.4.8}$$

for any $0 \leq 2s \leq t$.

Let $G(x, t; y, s)$ denote the heat kernel associated with the heat equation coupled with the Ricci Flow $h(t)$

$$\left(\partial_t - \Delta_{x,t}^{h_t}\right)G(\cdot, \cdot; y, s) = 0 \quad \text{and} \quad \lim_{t \to s^+} G(\cdot, t; y, s) = \delta_y(x).$$

Choosing the Ricci flow $h(t)$ above to be the one related to $g(t)$ via pullback by diffeomorphisms generated by X, that is, $h(t) = \chi_t^* g(t)$, then the heat kernel K that we are interested is related to the heat kernel G also by pulling back the diffeomorphisms generated by X. We first show that under our assumptions, (2.4.8) can be used to obtain an estimate for $G(x, t; y, 0)$. By our assumptions and Lemma 2.2, we have the bound

$$\sup_M |R_{h_t}| = \sup_M |R_{g_t}| \leq B_2 t^{-1+\alpha}$$

for some constant $B_2(n, \Lambda, k_0, k_1, k_2, B_1) < \infty$ and some small $\alpha = \alpha(p, \delta) > 0$. Then

$$\left(\partial_t - \Delta_{x,t}^{h_t}\right)\tilde{G} = R_{h_t}\tilde{G} \geq -B_2 t^{\alpha-1}\tilde{G},$$

which gives

$$\begin{aligned}\left(\partial_t - \Delta_{x,t}^{h_t}\right)(\exp(\alpha^{-1}B_2t^{\alpha})\tilde{G}) &\geq -B_2t^{\alpha-1}\exp(\alpha^{-1}B_2t^{\alpha})\tilde{G} + B_2t^{\alpha-1}\exp(\alpha^{-1}B_2t^{\alpha})\tilde{G} \\ &\geq 0.\end{aligned}$$

Since $G(x,t;y,s)=\tilde{G}(x,t;y,s)=\delta_y(x)$ at time $t=s>0$, the maximum principle gives

$$G(x,t;y,s)\le \exp(\alpha^{-1}B_2t^{\alpha})\tilde{G}(x,t;y,s)\le 2\tilde{G}(x,t;y,s)$$

if we shrink T_1 depending on B_2,δ,p. Note that our assumptions also ensure that the Ricci flow $h(t)=\chi_t^*g(t)$ satisfies the assumptions (2.4.7) above. Finally, writing $K(x,t;y,s)=G(\chi_t^{-1}(x),t;\chi_s^{-1}(y),s)$ and after computations similar to that of Lemma 2.9 of [3], we obtain the following estimate for K: there is a $C=C(n,\Lambda)>0$ such that

$$K(x,t;y,s)\le \frac{C}{t^{n/2}}\exp\Big(-\frac{d_h^2(x,y)}{Ct}\Big) \tag{2.4.9}$$

for any $t\in(0,T]$.

By (2.4.9), the co-area formula and Stokes' theorem, we have on the right-hand side of (2.4.6)

$$\begin{aligned}
&\int_M K(x_0,T_1;y,0)f(0)Q(y,0)\,d\mu_h(y) \\
&\quad\le \int_M \frac{C_0}{T_1^{\frac{n}{2}}}\exp\left(-\frac{d_{h_0}^2(x_0,y)}{C_0T_1}\right)\big(|\tilde{\nabla}g_0|^2+1\big)^{\frac12}\,d\mu_h(y) \\
&\quad= \int_0^\infty \frac{C_0}{T_1^{\frac{n}{2}}}\exp\left(-\frac{r^2}{C_0T_1}\right)\Big(\int_{\partial B(x_0,r)}\big(|\tilde{\nabla}g_0|^2+1\big)^{\frac12}dS(y)\Big)\,dr \\
&\quad= \int_0^\infty \frac{C_0}{T_1^{\frac{n}{2}+1}}\exp\left(-\frac{r^2}{C_0T_1}\right)\frac{2r}{C_0}\Big(\int_{B(x_0,r)}\big(|\tilde{\nabla}g_0|^2+1\big)^{\frac12}\,d\mu_h(y)\Big)\,dr,
\end{aligned} \tag{2.4.10}$$

where dS is the surface area measure induced by $d\mu_h$ and C_0 above is the constant obtained from (2.4.9). For $p\ge1$, assumption (ii) and Hölder's inequality gives

$$\int_{B(x_0,r)}\big(|\tilde{\nabla}g_0|^2+1\big)^{1/2}\,d\mu_h(y)\le CL_0r^{n-1+\delta/p}$$

for all $0<r<r_0$. If $r>r_0$, we estimate it using trivial bound from (2.4.2) and volume comparison:

$$\int_{B(x_0,r)}\big(|\tilde{\nabla}g_0|^2+1\big)^{1/2}\,d\mu_h(y)\le B_1t^{\frac{\delta}{2p}-\frac12}\cdot\exp(Cr).$$

Substituting this into (2.4.10) we obtain

$$\int_0^\infty \frac{C_0}{T_1^{\frac{n}{2}+1}}\exp\left(-\frac{r^2}{C_0T_1}\right)\frac{2r}{C_0}\Big(\int_{B(x_0,r)}\big(|\tilde{\nabla}g_0|^2+1\big)^{\frac12}\,d\mu_h(y)\Big)\,dr$$

$$\leq \int_0^{r_0} \frac{C}{T_1^{\frac{n}{2}+1}} \exp\left(-\frac{r^2}{C_0 T_1}\right) r^{n+\frac{\delta}{p}} dr + \int_{r_0}^{\infty} \frac{2B_1}{T_1^{\frac{n}{2}+\frac{3}{2}-\frac{\delta}{p}}} \exp\left(Cr - \frac{r^2}{C_0 T_1}\right) dr$$

$$\leq C T_1^{-\frac{1}{2}+\frac{\delta}{2p}} \int_0^{\infty} \exp\left(-\frac{s^2}{C_0}\right) s^{n+\frac{\delta}{p}} ds + \int_{\frac{1}{\sqrt{T_1}}}^{\infty} \frac{C B_1}{T_1^{1+\frac{n}{2}}} \exp\left(-\frac{r^2}{2C_0}\right) dr$$

$$\leq C T_1^{-\frac{1}{2}+\frac{\delta}{2p}} + \frac{C B_1}{T_1^{1+\frac{n}{2}}} \exp\left(-\frac{1}{2C_0 T_1}\right).$$

So we have

$$\frac{1}{4} e^{-AT_1^{\frac{\delta}{p}} - AT_1} B_1 T_1^{-\frac{1}{2}+\frac{\delta}{2p}} \leq C_1 T_1^{-\frac{1}{2}+\frac{\delta}{2p}} + \frac{C B_1}{T_1^{1+\frac{n}{2}}} \exp\left(-\frac{1}{2C_0 T_1}\right) + C T_1. \tag{2.4.11}$$

for some $C_1(n, \Lambda, k_0, k_1, k_2, L_0) > 0$. By choosing $B_1 = 8C_1$, we see that T_1 is bounded from below by a constant depending only on δ, p, k_0, k_1, k_2, n, Λ, L_0. □

3. Short-time existence

In this section, we will prove short-time existence on metrics which are possibly singular and satisfy the Morrey-type integrability condition. We first consider the case when $\Sigma = \varnothing$, that is, when the initial data is smooth and satisfies the uniform Morrey-type integrability condition. We note that since M is possibly noncompact, the uniform short-time existence is not covered by the work of Simon [18].

Proof of Existence in Theorem 1.1 when $\Sigma = \varnothing$. Without loss of generality, we assume M to be noncompact. By [19] and $|\operatorname{Rm}(h)| \leq k_0$, there is $\rho \in C^{\infty}_{\mathrm{loc}}(M)$ such that $|\tilde{\nabla}\rho|^2 + |\tilde{\nabla}^2\rho| \leq 1$ and

$$C(n, k_0)^{-1}(d_h(\cdot, p) + 1) \leq \rho(\cdot) \leq C(n, k_0)(d_h(\cdot, p) + 1)$$

for some $C(n, k_0) > 0$. Let ϕ be a smooth function on $[0, +\infty)$ such that $\phi \equiv 1$ on $[0, 1]$, $\phi \equiv 0$ on $[2, +\infty)$ and $0 \leq -\phi' \leq 10$. Fix $R_i \to +\infty$ and define a smooth metric

$$g_{i,0} = \phi(\rho/R_i) g_0 + (1 - \phi(\rho/R_i)) h$$

on M which coincide with h at spatial infinity of M and coincide with g_0 on compact subset of M as $i \to +\infty$. Then we might apply Theorem A.1 of [11] (which is a modification of Shi's classical existence theory in [16]), to find short-time solution to the Ricci–DeTurck h-flow $g_i(t)$ on $M \times [0, S_i]$ for some maximal existence time $S_i > 0$ such that $g_i(0) = g_{i,0}$ and $\sup_M |\tilde{\nabla}^m g_i(\cdot, t)| < +\infty$ for all $t \in [0, S_i]$ and all $m \in \mathbb{N}$.

Since

$$|\tilde{\nabla} g_{i,0}| \leq \frac{C(n,k_0)}{R_i}|g_0 - h| + \phi \cdot |\tilde{\nabla} g_0|, \tag{3.0.1}$$

the metric $g_{i,0}$ satisfies the assumption of Proposition 2.4 uniformly for all $i \to +\infty$. Hence, Proposition 2.4 and Lemma 2.2 apply to show that there is a $T(n, \Lambda_0, L_0, p, \delta, h) > 0$ such that $T \leq S_i$ with the solution $g_i(t)$ obtained above also satisfying (a) for all $t \in (0, T]$ uniformly for all $i \to +\infty$. We therefore obtain a smooth Ricci–DeTurck h-flow $g(t)$ on $M \times (0, T]$ by taking subsequence $i \to +\infty$ using the Arzelà–Ascoli theorem. Since $|\tilde{\nabla} g_i|$ has a uniform integrable bound in t, we see that $g(t)$ exists as a C^0_{loc} metric on $M \times [0, T]$. Indeed, by shrinking T we have

$$(2\Lambda)^{-1} h \leq g(t) \leq 2\Lambda h \tag{3.0.2}$$

on $M \times [0, T]$ by integrating $\partial_t g$ in time. □

We now want to consider the case when Σ is a nonempty compact subset of M. We start with constructing a C^∞ approximation of g_0, see also [12; 17; 14; 6] for the other similar approximation schemes.

Lemma 3.1. *Suppose g_0 is a C^0_{loc} metric on M and is smooth away from a compact set Σ. Moreover, suppose that*

(i) *there is a $\Lambda_0 > 1$ such that on M,*

$$\Lambda_0^{-1} h \leq g_0 \leq \Lambda_0 h;$$

(ii) *there exist $L_0, \delta, r_0 > 0$ and $p \geq 1$ such that for every $x_0 \in M$, $0 < r < r_0$,*

$$⨍_{B(x_0,r)} |\tilde{\nabla} g_0|^p \, d\mu_h \leq L_0 r^{-p+\delta}.$$

Then there is a sequence of smooth metrics $g_{i,0}$ on M satisfying the following conditions:

(a) *For i sufficiently large,*

$$(2\Lambda_0)^{-1} h \leq g_{i,0} \leq 2\Lambda_0 h \quad \text{on } M.$$

(b) *There exist $L_1, r_1 > 0$ such that for every $x_0 \in M$ and $0 < r < r_1$,*

$$⨍_{B(x_0,r)} |\tilde{\nabla} g_{i,0}|^p \, d\mu_h \leq L_1 r^{-p+\delta}.$$

(c) *$g_{i,0} \to g_0$ in C^0_{loc} on M as $i \to \infty$.*

(d) *$g_{i,0} \to g_0$ in C^∞_{loc} on $M \setminus \Sigma$ as $i \to \infty$.*

(e) *$g_{i,0} \to g_0$ in $W^{1,p}_{\text{loc}}(M)$ as $i \to +\infty$.*

Proof. Without loss of generality, we will assume M to be noncompact. By the embedding theorem of Morrey (see for example "Morrey's Lemma" [1, (1.3)]), assumption (ii) implies that g_0 is locally Hölder continuous with exponent $\delta/p < 1$ on M. Since Σ is compact, there is a point $p \in M$ and $R > 0$ such that $\Sigma \Subset B(p, R)$.

Note that $K = \overline{B(p, R)}$ is compact. We now cover K by finitely many open coordinate charts $\{U_k\}_{k=1}^N$. Also let $U_0 = M \setminus K$. Let φ_k be a partition of unity subordinate to $U_0 \cup \bigcup_{k=1}^N U_k$. Then we can decompose the metric $g_0 = \sum_{k=0}^N g_0^k$ on each chart by $g_0^k = \varphi_k g_0$. We might assume each U_k is diffeomorphic to unit ball $B_{\mathbb{R}^n}(1)$ in $\mathbb{R}^n$. For $k = 1, \ldots, N$, let η be the standard mollifier with compact support inside $B_{\mathbb{R}^n}(1)$ and $\int_{U_k} \eta(y)\, dy = 1$. For $k \geq 1$, we define

$$g_{i,0}^k(x) := \int_{U_k} g_0^k(x - i^{-1}y)\eta(y)\, dy.$$

for $x \in U_k$. Since φ_k is compactly supported in U_k for $k \geq 1$, we see that $g_{i,0}^k(x) = 0$ for $x \to \partial U_k$ and sufficiently large i. Hence, $g_{i,0}^k$ extends trivially on M. Thus we define $g_{i,0} := \sum_{k=1}^N g_{i,0}^k + g_0^0$ to be a metric on M. Near the infinity of M, $g_{i,0} = g_0^0 = g_0$, while on compact set, $g_{i,0}$ is a mollification of g_0. It suffices to check that the above properties are satisfied on compact sets. For (a) and (b) above, it suffices to show that they are satisfied on compact set. Clearly by the standard mollification, we have (e) and (d) since $g_0 \in W^{1,p}_{\mathrm{loc}}(M) \cap C^\infty_{\mathrm{loc}}(M \setminus \Sigma)$ so that $g_{i,0}^k \to g_0^k$ as $i \to +\infty$ in $W^{1,p}_{\mathrm{loc}}(M) \cap C^\infty_{\mathrm{loc}}(M \setminus \Sigma)$ for each k.

To see (c), If $x \notin \cup_{k=1}^N U_k$, then the result holds directly. If $x \in K$, then $x \in U_k$ for some $k = 1, \ldots, N$. Fix any $\varepsilon > 0$. By the Hölder continuity of g_0, for i sufficiently large, $|g_0^k(x - i^{-1}y) - g_0^k(x)| < C|i^{-1}y|^\alpha < \varepsilon$ for any $x, y \in U_k$. Here C depends also on the choice of partition of unity. Hence, since $\int_{U_k} \eta(y)\, dy = 1$,

$$\begin{aligned}
|g_{i,0}^k(x) - g_0^k(x)| &= \left| \int_{U_k} g_0^k(x - i^{-1}y)\eta(y)\, dy - g_0^k(x) \right| \\
&\leq \int_{U_k} |g_0^k(x - i^{-1}y) - g_0^k(x)|\eta(y)\, dy \\
&\leq \int_{U_k} |i^{-1}y|^\alpha \eta(y)\, dy \\
&\leq \int_{U_k} \varepsilon\eta(y)\, dy = \varepsilon.
\end{aligned}$$

From this, (a) also follows for i sufficiently large.

It remains to prove (b). It suffices to show (b) for each $g_{i,0}^k$. We choose $r_1 < r_0$ so that for each $x_0 \in \mathrm{supp}(\varphi_k)$, $B(x_0, r_1) \Subset U_k$. For fixed y with $|y| \leq 1$, note that $z := x - i^{-1}y$ is a translation with determinant of Jacobian uniformly bounded from above and below. We use the fact that mollification and differentiation commute,

i.e., $\partial g_i(x) = (\eta_i * \partial g)(x)$, then by Minkowski's integral inequality,

$$\begin{aligned}\Big(\int_{B(x_0,r)} |\partial g_{i,0}^k(x)|^p d\mu_h(x)\Big)^{\frac{1}{p}} &= \Big(\int_{B(x_0,r)} \Big|\int_{U_k} |\partial g_0^k(x - i^{-1}y)\eta(y)\, dy\Big|^p d\mu_h(x)\Big)^{\frac{1}{p}} \\ &\le \int_{U_k} \Big(\int_{B(x_0,r)} |\partial g_0^k(x - i^{-1}y)\eta(y)|^p dx\Big)^{\frac{1}{p}} dy \\ &\le \int_{U_k} \eta(y)\, dy \Big(\int_{B(x_0,r)} |\partial g_0^k(x - i^{-1}y)|^p\, d\mu_h(x)\Big)^{\frac{1}{p}} \\ &\le \Big(C \int_{B(x_0+i^{-1}y,r)} |\partial g_0^k(z)|^p dz\Big)^{\frac{1}{p}} \\ &\le (C L_0 r^{n-p+\delta})^{\frac{1}{p}}.\end{aligned}$$

Here C depends on the choice of a partition of unity. So (b) is satisfied. □

Now we are ready to prove Theorem 1.1 under the presence of singular sets Σ.

Proof of Theorem 1.1 when $\Sigma \neq \varnothing$. Let $g_{i,0}$ be the smooth approximation of g_0 obtained from Lemma 3.1. By properties (a), (b) in Lemma 3.1 above, we know that for i sufficiently large, that

$$(2\Lambda_0)^{-1} h \le g_{i,0} \le 2\Lambda_0 h, \tag{3.1.1}$$

and for all $x_0 \in M$ and $0 < r < r_1$,

$$⨍_{B(x_0,r)} |\tilde{\nabla} g_{i,0}|^p\, d\mu_h \le L r^{-p+\delta} \tag{3.1.2}$$

for some $L > 0$.

Since each $g_{i,0}$ is smooth and satisfies (3.1.1) and (3.1.2) above for i large enough, by fullrefon, we obtain constants $T(n, h, \Lambda_0, p, \delta, h, L, r_1) > 0$ and a sequence of $B_k(n, h, \Lambda_0, p, \delta, h, L, r_1) > 0$ and solutions $g_i(t)$ to the Ricci–DeTurck h-flow on $M \times [0, T]$ such that $g_i(0) = g_{i,0}$ satisfying

$$\sup_M |\tilde{\nabla}^k g_i(t)| \le \frac{B_k}{t^{\frac{1}{2}(m-\frac{\delta}{p})}}$$

on $M \times (0, T]$.

So by the Arzelà–Ascoli theorem and taking a subsequence, we obtain smooth $g(t) = \lim_{i\to\infty} g_i(t)$ on $M \times (0, S]$ with $g(t)$ satisfying (a) and (b). Since $g_{i,0} \to g_0$ in $C^\infty_{\mathrm{loc}}(M \setminus \Sigma)$ as $i \to +\infty$, $g(t) \in C^\infty_{\mathrm{loc}}((M \setminus \Sigma) \times [0, T])$ using [4, Proposition 2.2].

It remains to show the uniqueness within the class of solution satisfying the property (a). If M is compact, it follows from [3] since $g(t)$ attains g_0 in C^0_{loc} sense. It remains to consider the case when M is noncompact. It follows almost identically to the proof of [4, Theorem 3.1] with modification. We give a sketch here. Suppose

$g(t)$ and $\hat{g}(t)$ are two solution on $M \times (0, T]$ satisfying derivatives estimates in (a). Clearly, both $g(t)$ and $\hat{g}(t)$ converge to g_0 in C^0_{loc} as $t \to 0$. Instead of using the concentration of gradient, we use the fact that $|\tilde{\nabla} g| + |\tilde{\nabla} \hat{g}| \leq C t^{-\frac{1}{2}+\frac{\delta}{2p}}$ to show that the tensor $\alpha = g - \hat{g}$ satisfies

$$\frac{\partial}{\partial t} \int_M \eta^4 |\alpha|^2 \, d\mu_h \leq C \int_M \eta^2 |\alpha|^2 \, d\mu_h + C t^{-\frac{1}{2}+\frac{\delta}{2p}} \int_M \eta^4 |\alpha|^2 \, d\mu_h \tag{3.1.3}$$

which is analogous to [4, (3.20)]. Here η is a smooth cut-off function on M with compact support on $B(x, 2)$. Since $t^{-\frac{1}{2}+\frac{\delta}{2p}}$ is integrable in t near $t = 0$, the argument can be carried over. □

4. Applications: preserving the distributional scalar curvature lower bound

In this section, we use the existence theory established above to study the preserving of weak scalar curvature lower bound along the Ricci flow smoothing. In [13], Lee–LeFloch introduced a notion of distributional scalar curvature that is defined for metrics with low regularity. In particular, for any $g \in L^\infty_{\text{loc}} \cap W^{1,2}_{\text{loc}}$ with locally bounded inverse $g^{-1} \in L^\infty_{\text{loc}}$, they define the scalar curvature distribution R_g by

$$\langle R_g, u \rangle := \int_M \left(-V \cdot \tilde{\nabla} \left(u \frac{d\mu_g}{d\mu_h} \right) + F u \frac{d\mu_g}{d\mu_h} \right) d\mu_h \tag{4.0.1}$$

for every compactly supported smooth test function $u : M \to \mathbb{R}$ and where

$$\begin{aligned}
\Psi^k_{ij} &:= \frac{1}{2} g^{k\ell} \left(\tilde{\nabla}_i g_{j\ell} + \tilde{\nabla}_j g_{i\ell} - \tilde{\nabla}_\ell g_{ij} \right) \\
V^k &:= g^{ij} \Psi^k_{ij} - g^{ik} \Psi^j_{ji} \\
F &:= \operatorname{tr}_g \widetilde{Ric} - \tilde{\nabla}_k g^{ij} \Psi^k_{ij} + \tilde{\nabla}_k g^{ik} \Psi^j_{ji} + g^{ij} \left(\Psi^k_{k\ell} \Psi^\ell_{ij} - \Psi^k_{j\ell} \Psi^\ell_{ik} \right)
\end{aligned}$$

and $d\mu_g / d\mu_h \in L^\infty_{\text{loc}} \cap W^{1,2}_{\text{loc}}$ is the density of $d\mu_g$ with respect to $d\mu_h$. Let a be a constant. We say $R_g \geq a$ in the distributional sense when $\langle R_g - a, u \rangle \geq 0$ for every nonnegative smooth test function u. Clearly, when g is C^2, the distributional scalar curvature R_g coincides with the classical scalar curvature.

Our main motivation is to study metrics on compact manifolds which are smooth outside some singular sets. In other words, we want to consider the question of whether a scalar curvature lower bound can be extended to a distributional scalar curvature lower bound across a set where the metric is singular. In [8], the authors consider this for when $g \in C^0 \cap W^{1,p}_{\text{loc}}(M)$ for $n \leq p \leq \infty$ and the singular set is small in the sense of Hausdorff dimension, see [8, Lemma 2.7]. In this work, we use a notion of codimension for compact subsets $\Sigma \subset M$ based on the volume growth of tubular neighborhoods of Σ.

Definition 4.1. For a complete smooth Riemannian manifold M^n with smooth background metric h, a compact set Σ of M is said to have codimension at least $d_0 > 0$ if there exist $b > 0$ and $C > 0$ such that for all $0 < \varepsilon \leq b$

$$\mathrm{Vol}_h\,(\Sigma(\varepsilon)) = \mathrm{Vol}_h\,(\{x \in M : d_h(x, \Sigma) < \varepsilon\}) \leq C\varepsilon^{d_0}.$$

We show that when g satisfies the L^∞ and the Morrey-type condition, and satisfies a scalar curvature lower bound outside of a compact set Σ, then the corresponding distributional scalar curvature lower bound holds when Σ is not too large in the sense of Definition 4.1.

Proposition 4.2. *Let M^n be an n-dimensional manifold with smooth background metric h satisfying* (2.0.1)*. Suppose the metric g satisfies the following conditions*:

(i) *There is a $\Lambda > 0$ such that $\Lambda^{-1}h \leq g \leq \Lambda h$ on M.*

(ii) *For $p \geq 2$, $\delta > 0$, $r_0 > 0$ there is a L such that for all $x_0 \in M$ and $0 < r < r_0$,*

$$⨍_{B(x_0,r)} |\tilde{\nabla} g|^p \, d\mu_h \leq L r^{-p+\delta}. \tag{4.2.1}$$

(iii) *There is a compact $\Sigma \subset M$ with codimension d at least $2 - \frac{\delta'}{p}$ for some $0 < \delta' < \delta$ such that g is smooth on $M \setminus \Sigma$ and $R_g \geq a$ holds in the classical sense on $M \setminus \Sigma$.*

Then for all smooth compactly supported nonnegative test functions u, the distributional scalar curvature satisfies $\langle R_g - a, u\rangle \geq 0$.

Proof. Fix a smooth compactly supported nonnegative function u. We will use C to denote any constants depending only on n, Λ, p, δ, r_0, $L > 0$ and u which might be adjusted from line to line. As in [8], for any $\varepsilon > 0$, we let η_ε be a smooth nonnegative function so that $\eta_\varepsilon \equiv 1$ on $\Sigma(\varepsilon)$ and $\eta_\varepsilon \equiv 0$ on $M \setminus \Sigma$ with $|\tilde{\nabla}\eta_\varepsilon| \leq C\varepsilon^{-1}$. Then

$$\langle R_g - a, u\rangle = \langle R_g - a, \eta_\varepsilon u\rangle + \langle R_g - a, (1 - \eta_\varepsilon)u\rangle.$$

Since the support of $(1 - \eta_\varepsilon)u$ is outside Σ, we have

$$\langle R_g - a, (1 - \eta_\varepsilon)u\rangle = \int_{M\setminus\Sigma(\varepsilon)} (R_g - a)(1 - \eta_\varepsilon)u \, d\mu_g \geq 0,$$

because g is smooth and satisfies $R_g \geq a$ in the classical sense outside of $\Sigma(\varepsilon)$. So it suffices to show

$$\lim_{\varepsilon\to 0} \left|\langle R_g - a, \eta_\varepsilon u\rangle\right| = 0.$$

Then, again as in Lemma 2.7 of [8], we have by the definition of V, F and the Hölder inequality that

(4.2.2)

$$\begin{aligned}\left|\langle R_g - a, \eta_\varepsilon u\rangle\right| &\le \int_M |V| \cdot \left|\tilde{\nabla}\left(\eta_\varepsilon u \frac{d\mu_g}{d\mu_h}\right)\right| d\mu_h + \int_M |F-a| \cdot \eta_\varepsilon u \frac{d\mu_g}{d\mu_h}\, d\mu_h \\ &\le I + II + III + IV,\end{aligned}$$

where

$$\begin{aligned} I &:= C\Big(\int_{\Sigma(\varepsilon)} |\tilde{\nabla} g|^p d\mu_h\Big)^{\frac{1}{p}} \operatorname{Vol}_h(\Sigma(\varepsilon))^{1-\frac{1}{p}}, \\ II &:= C\Big(\int_{\Sigma(\varepsilon)} |\tilde{\nabla} g|^p d\mu_h\Big)^{\frac{1}{p}} \Big(\int_{\Sigma(\varepsilon)} |\tilde{\nabla}\eta_\varepsilon|^{\frac{p}{p-1}} d\mu_h\Big)^{\frac{p-1}{p}}, \\ III &:= C\Big(\int_{\Sigma(\varepsilon)} |\tilde{\nabla} g|^p d\mu_h\Big)^{\frac{2}{p}} \operatorname{Vol}_h(\Sigma(\varepsilon))^{1-\frac{2}{p}}, \\ IV &:= C \operatorname{Vol}_h(\Sigma(\varepsilon)), \end{aligned}$$

where the constants depend on n, Λ and the C^1 norm of u.

Since Σ has codimension at least d, by definition there is $C > 0$ such that

$$\operatorname{Vol}_h(\Sigma(\varepsilon)) \le C\varepsilon^d$$

for all ε sufficiently small. Since Σ is compact and M carries the structure of a metric space with distance function d_h, Σ is totally bounded, that is, for every fixed radius, it can be covered by a finite number of balls of that radius measured with respect to d_h. In particular, it can be covered by a finite number of balls of radius $\varepsilon/2$. Let N be the minimal such number of balls, $B(x_k, \frac{\varepsilon}{2})$, $k = 1, \ldots, N$. We first claim that

$$\Sigma \subset \bigcup_{k=1}^{N} B\left(x_k, \frac{\varepsilon}{2}\right) \subset \Sigma(\varepsilon).$$

The first inclusion is obvious. Now suppose $x \in B(x_k, \frac{\varepsilon}{2})$ for some k. By minimality of N, $B(x_k, \frac{\varepsilon}{2}) \cap \Sigma \neq \varnothing$ and so $d_h(\Sigma, x_k) < \frac{\varepsilon}{2}$. Then by the triangle inequality we have

$$d_h(\Sigma, x) \le d_h(\Sigma, x_k) + d_h(x_k, x) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

So the second inequality holds. Hence, we have

$$\sum_{k=1}^{N} \operatorname{Vol}_h\left(B\left(x_k, \frac{\varepsilon}{2}\right)\right) \le \operatorname{Vol}_h(\Sigma(\varepsilon)) \le C\varepsilon^d.$$

By taking ε small enough, we can also assume there is a constant $D = D(n, h)$

such that

$$D^{-1}\varepsilon^n \leq \mathrm{Vol}_h(B(\varepsilon)) \leq D\varepsilon^n.$$

So we have

$$N \leq C\varepsilon^{d-n}. \tag{4.2.3}$$

Similarly by the triangle inequality, we have that

$$\Sigma(\varepsilon) \subset \bigcup_{k=1}^{N} B\left(x_k, \frac{3\varepsilon}{2}\right)$$

since if $x \in \Sigma(\varepsilon)$, there is $y \in \Sigma$ such that $d_h(y, x) < \varepsilon$, and by $\Sigma \subset \bigcup_{k=1}^{N} B\left(x_k, \frac{\varepsilon}{2}\right)$, there is an x_k such that $d_h(x_k, y) < \frac{\varepsilon}{2}$, so $d_h(x, x_k) \leq d_h(x, y) + d_h(y, x_k) < \frac{3\varepsilon}{2}$.

Clearly, we have from definition that

$$IV \leq C\varepsilon^d.$$

With (4.2.3) and the Morrey assumption (ii), we can estimate the terms on the right-hand side of (4.2.2) above. For I, we have

$$\begin{aligned} I &\leq C\left(\sum_{k=1}^{N} \int_{B(x_k, \frac{3\varepsilon}{2})} |\tilde{\nabla} g|^p d\mu_h\right)^{\frac{1}{p}} \mathrm{Vol}_h(\Sigma(\varepsilon))^{1-\frac{1}{p}} \\ &\leq C\left(NL\varepsilon^{n-p+\delta}\right)^{\frac{1}{p}} \mathrm{Vol}_h(\Sigma(\varepsilon))^{1-\frac{1}{p}} \\ &\leq C(\varepsilon^{d-n+n-p+\delta})^{\frac{1}{p}} \varepsilon^{d\left(1-\frac{1}{p}\right)} \\ &= C\varepsilon^{d-1+\frac{\delta}{p}}. \end{aligned}$$

Arguing in the same way, we have $III \leq C\varepsilon^{d-2+\frac{2\delta}{p}}$. Finally, using $|\tilde{\nabla}\eta| \leq C\varepsilon^{-1}$, we get

$$II \leq C\varepsilon^{d-2+\frac{\delta}{p}} \tag{4.2.4}$$

so that $II \to 0$ as $\varepsilon \to 0$ since $d \geq 2 - \frac{\delta'}{p}$ for some $\delta' < \delta$. Substituting these back into (4.2.2) and by the value of d, we get that the right-hand side of (4.2.2) vanishes as $\varepsilon \to 0^+$, and so we are done. □

4.3. ***Preservation of the distributional lower bound.*** We now consider the preservation of distributional scalar curvature lower bounds along the Ricci flow. In [9], the authors showed that scalar curvature lower bounds in the distributional sense as above are preserved along the Ricci flow for initial metrics $g \in W^{1,p}(M^n)$ for $3 \leq n < p \leq \infty$. In this section, we improve the result by relaxing the regularity to Morrey-type condition with $2 \leq p \leq n$. We follow the approach of [9] with modifications.

Lemma 4.4. *Let (M^n, h) be a closed manifold and g_0 is a C^0 metric on M satisfying. Let g_0 and $g(t)$ be as in Theorem 1.1. Then there is a constant $C = C(n, M, h, \Lambda, L_0, p, \delta, r_0) < +\infty$ such that we have*

$$\int_0^T \int_M |\tilde{\nabla}^2 g(t)|^2 d\mu_h dt \leq C.$$

Proof. Recall that from (1.0.1), standard computations and the Cauchy–Schwarz inequality yields

$$\frac{\partial}{\partial t}|\tilde{\nabla} g|^2 \leq g^{ij}\tilde{\nabla}_i\tilde{\nabla}_j|\tilde{\nabla} g|^2 - C_1|\tilde{\nabla}^2 g|^2 + C_2|\tilde{\nabla} g|^4 + C_3$$

for constants C_1, C_2, C_3 that are only depending on n, h where we have gradient estimate:

$$C_2|\tilde{\nabla} g|^4 \leq C_2 B_1 |\tilde{\nabla} g|^2 t^{-1+\frac{\delta}{p}},$$

so defining $f(x,t) = |\tilde{\nabla} g|^2 \exp(-At^{\delta/p})$ for some constant $A > 0$ to be determined, we have

$$\frac{\partial}{\partial t} f(t) \leq \exp(-At^{\delta/p}) g^{ij}\tilde{\nabla}_i\tilde{\nabla}_j|\tilde{\nabla} g|^2 - C_1|\tilde{\nabla}^2 g|^2 + C_2 f(t) t^{-1+\frac{\delta}{p}} - Af(t)t^{-1+\frac{\delta}{p}} + C_3.$$

Here we used the fact that $\exp(-At^{\delta/p})$, for $t \in [0, T]$, is uniformly bounded above. Integrating by parts and Young's inequality yields

$$\frac{\partial}{\partial t}\int_M f d\mu_h \leq \int_M -C_1|\tilde{\nabla}^2 g|^2 d\mu_h + \int_M (C_2 + C_4 - A)\, f(t) t^{-1+\frac{\delta}{p}} d\mu_h + C_3.$$

Now we choose A large enough so that it dominates $C_2 + C_4$ and we obtain

$$\frac{\partial}{\partial t}\int_M f d\mu_h + \int_M C_1|\tilde{\nabla}^2 g|^2 d\mu_h \leq C_3.$$

Since M is compact, by covering argument we see that

$$\int_M |\tilde{\nabla} g_0|^2 d\mu_h \leq C(n, M, h, r_0, L_0, p, \delta). \tag{4.4.1}$$

With this, we integrate in time from 0 to T to obtain the result. □

Now let M be a closed manifold and let $g(t)$ be the Ricci–DeTurck h-flow on M. Consider the heat kernel:

$$(\partial_t - \Delta_x - \nabla_X) K(\cdot, \cdot; y, s) = 0, \quad \lim_{t \to s^+} K(\cdot, t; y, s) = \delta_y(\cdot)$$

and moreover satisfies for the conjugate heat operator:

$$\left(-\partial_s - \Delta_y - \nabla_X + R\right) K(x, t; \cdot, \cdot) = 0, \quad \lim_{s \to t^-} K(x, t; \cdot, s) = \delta_x(\cdot),$$

where the Laplacian operators are taken with respect to $g(t)$. Let $\tilde{u}$ be an arbitrary nonnegative C^∞ function on M. We consider the conjugate heat equation with $\tilde{u}$ as final data, that is,

$$\begin{cases} \dfrac{\partial}{\partial t} u = -\Delta u - \nabla_X u + Ru & \text{on } M \times [0, T], \\ u|_{t=T} = \tilde{u}, \end{cases} \tag{4.4.2}$$

where R is the scalar curvature with respect to $g(t)$. By the maximum principle, the solution is nonnegative and given by convolving with the fundamental solution:

$$u(x, t) = \int_M K(y, T; x, t)\tilde{u}(y)\, d\mu_{g(T)}(y).$$

Similar to Proposition 4.1 of [9], we have the estimates for u in the following lemma.

Lemma 4.5. *Let (M, h) be a closed Riemannian manifold and g_0 is a $C^0_{\text{loc}} \cap W^{1,p}_{\text{loc}}$ metric on M satisfying assumptions in Theorem 1.1 with $p \geq 2$. Suppose $g(t)$ is the unique solution obtained from Theorem 1.1 and u is a solution obtained as above. Then there exists $C_0 > 0$ depending only on $n, M, h, \Lambda, L_0, p, \delta, \tilde{u}$ such that*

(a) $u(\cdot, t) \leq C_0$, *for all* $t \in [0, T]$.

(b) $\int_M |\nabla_{g(t)} u(\cdot, t)|^2 d\mu_{g(t)} \leq C_0$, *for all* $t \in [0, T]$.

(c) *For any* $a \in \mathbb{R}$, $\int_M (R_{g(t)} - a)u\, d\mu_{g(t)}$ *is monotonically nondecreasing with respect to* t.

Proof. The proof is identical to that of [9, Proposition 4.1] using Lemma 4.4 and the integrable interior estimates from Theorem 1.1. □

With these two lemmata, we have the following main application for the compact case.

Theorem 4.6. *Let (M^n, h) be a closed manifold and g_0 be a C^0 metric satisfying*

(i) *there is a $\Lambda > 0$ such that $\Lambda^{-1} h \leq g_0 \leq \Lambda h$ on M;*

(ii) *there exist $p \geq 2$, $\delta > 0$, r_0, $L > 0$ such that for all $x_0 \in M$ and $0 < r < r_0$,*

$$⨍_{B(x_0, r)} |\tilde{\nabla} g_0|^p\, d\mu_h \leq L r^{-p+\delta}; \tag{4.6.1}$$

(iii) *$R_{g_0} \geq a$ in the distributional sense on M for some $a \in \mathbb{R}$.*

Let $g(t)$, $t \in (0, T]$ be the Ricci–DeTurck h-flow from g_0 obtained in Theorem 1.1 with $2 \leq p \leq n$. Then for any $t \in (0, T]$, $R_{g(t)} \geq a$ in the classical sense on M.

Proof. Fix any $t > 0$. It suffices to show

$$\langle R_{g(t)} - a, \tilde{u} \rangle \geq 0$$

for any arbitrary nonnegative function $\tilde{u} \in C^\infty(M)$, where $\langle R, u\rangle$ is defined as in (4.0.1). We follow the strategy as in Theorem 5.1 of [9], but modify it for our setting.

We pay extra attention to the construction of $g(t)$. Let $g_{i,0}$ be as in Lemma 3.1 and $g_i(t)$ be solutions to the Ricci–DeTurck h-flow starting from $g_{i,0}$ as in Theorem 1.1 above so that $g(t)$ is the smooth limit of $g_i(t)$ as $i \to +\infty$ on $M \times (0, T]$.

Let u_i be the solution to (4.4.2) (with respect to $g_i(t)$) with $u_{i,t} = \tilde{u}$ and $u_{i,0}(x) = u_i(x, 0)$. Note that u_i satisfies the conclusion in Lemma 4.5 above uniformly as $i \to +\infty$. In particular, we have

$$\int_M (R_{g_i(t)} - a)\tilde{u}\, d\mu_{g_i(t)} \geq \int_M (R_{g_{i,0}} - a)u_{i,0}\, d\mu_{g_{i,0}}. \tag{4.6.2}$$

Since by hypothesis we have that $\langle R_{g_0} - a, u_{i,0}\rangle \geq 0$, what remains is to examine the difference

$$\left|\langle R_{g_{i,0}} - a, u_{i,0}\rangle - \langle R_{g_0} - a, u_{i,0}\rangle\right|.$$

By the triangle inequality,

$$\begin{aligned}\left|\langle R_{g_{i,0}} - a, u_{i,0}\rangle - \langle R_{g_0} - a, u_{i,0}\rangle\right| &\\ \leq \left|\langle R_{g_{i,0}}, u_{i,0}\rangle - \langle R_{g_0}, u_{i,0}\rangle\right| + |a| &\left|\int_M u_{i,0}\, d\mu_{g_{i,0}} - \int_M u_{i,0}\, d\mu_{g_0}\right|\end{aligned}$$

Since $g_{i,0} \to g_0$ globally in C^0, we observe that

$$\lim_{i\to\infty} \left\| \frac{d\mu_{g_{i,0}}}{d\mu_{g_0}} - 1 \right\|_{C^0(M)} = 0 \quad \text{and} \quad \lim_{i\to\infty} \left\| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right\|_{C^0(M)} = 0.$$

To handle the second term above, by (a) of Lemma 4.5, we have

$$\begin{aligned}|a| \left|\int_M u_0\, d\mu_{g_{i,0}} - \int_M u_0\, d\mu_{g_0}\right| &\leq |a| \left|\int_M u_0 \left(\frac{d\mu_{g_{i,0}}}{d\mu_{g_0}} - 1\right) d\mu_{g_0}\right| \\ &\leq C \left\| \frac{d\mu_{g_{i,0}}}{d\mu_{g_0}} - 1 \right\|_{C^0(M)}\end{aligned}$$

for some positive constant C depending on $a, n, k_1, k_2, \Lambda, p, \delta, \tilde{u}$.

For the first term, we expand out the definition of distributional scalar curvature. Let V, F be the vector and scalar fields associated with g_0, and let V_i, F_i be the vector and scalar fields associated with $g_{i,0}$ that appear in the definition for distributional scalar curvature (4.0.1). Then by triangle inequality, we have

$$\left|\langle R_{g_{i,0}}, u_{i,0}\rangle - \langle R_{g_0}, u_{i,0}\rangle\right| \leq I + II + III + IV,$$

where

$$I := \int_M |V - V_i| \left| \tilde{\nabla} \left(u_{i,0} \frac{d\mu_{g_{i,0}}}{d\mu_h} \right) \right| d\mu_h,$$

$$II := \int_M |V| \left| \tilde{\nabla} \left(u_{i,0} \frac{d\mu_{g_0}}{d\mu_h} - u_{i,0} \frac{d\mu_{g_{i,0}}}{d\mu_h} \right) \right| d\mu_h,$$

$$III := \int_M |F_i u_{i,0} - F u_{i,0}| \left| \frac{d\mu_{g_{i,0}}}{d\mu_h} \right| d\mu_h,$$

$$IV := \int_M |F u_{i,0}| \left| \frac{d\mu_{g_{i,0}}}{d\mu_h} - \frac{d\mu_{g_0}}{d\mu_h} \right| d\mu_h,$$

and we estimate each term separately. In the following, the constants C_m will come from (a) and (b) of Lemma 4.5 and the Morrey condition on g_0. In other words, C_m will at most depend on $n, k_1, k_2, \Lambda, p, \delta, \tilde{u}, L, h$. The constants may be changing line by line.

Before handling each term, observe that since M is compact, we have $g_{i,0} \to g_0$ in $W^{1,p}(M)$ by (e) of Lemma 3.1 above.

By Hölder's inequality,

$$\begin{aligned}
I &= \int_M |V - V_i| |\tilde{\nabla} u_{i,0}| \left| \frac{d\mu_{g_{i,0}}}{d\mu_h} \right| d\mu_h + \int_M |V - V_i| |u_{i,0}| \left| \tilde{\nabla} \frac{d\mu_{g_{i,0}}}{d\mu_h} \right| d\mu_h \\
&\leq C_1 \int_M |\tilde{\nabla} g_0 - \tilde{\nabla} g_{i,0}| |\tilde{\nabla} u_{i,0}| \, d\mu_h + C_2 \int_M |\tilde{\nabla} g_0 - \tilde{\nabla} g_{i,0}| |\tilde{\nabla} g_{i,0}| \, d\mu_h \\
&\leq C_1 \left(\int_M |\tilde{\nabla} g_{i,0} - \tilde{\nabla} g_0|^p d\mu_h \right)^{\frac{1}{p}} \left(\int_M |\tilde{\nabla} u_{i,0}|^2 d\mu_h \right)^{\frac{1}{2}} \mathrm{Vol}_h(M)^{\frac{p-2}{2p}} \\
&\quad + C_2 \left(\int_M |\tilde{\nabla} g_{i,0} - \tilde{\nabla} g_0|^p d\mu_h \right)^{\frac{1}{p}} \left(\int_M |\tilde{\nabla} g_{i,0}|^p d\mu_h \right)^{\frac{1}{p}} \mathrm{Vol}_h(M)^{\frac{p-2}{p}} \\
&\leq C_1 \| g_{i,0} - g_0 \|_{W^{1,p}(M)} + C_2 \| g_{i,0} - g_0 \|_{W^{1,p}(M)},
\end{aligned}$$

where for the last inequality we are using the fact that $g_{i,0}$ satisfies the Morrey-type condition and a covering argument from the fact that M is compact. Next we have

$$\begin{aligned}
II &= \int_M |V| |\tilde{\nabla} u_{i,0}| \left| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right| d\mu_h + \int_M |V| |u_{i,0}| \left| \tilde{\nabla} \left(\frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right) \right| d\mu_h \\
&\leq C_3 \int_M |\tilde{\nabla} g_0| |\tilde{\nabla} u_{i,0}| \left| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right| d\mu_h + C_4 \int_M |\tilde{\nabla} g_0| |\tilde{\nabla} g_0 - \tilde{\nabla} g_{i,0}| \, d\mu_h \\
&\leq C_3 \left\| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right\|_{C^0(M)} \left(\int_M |\tilde{\nabla} g_0|^p d\mu_h \right)^{\frac{1}{p}} \left(\int_M |\tilde{\nabla} u_{i,0}|^2 d\mu_h \right)^{\frac{1}{2}} \mathrm{Vol}_h(M)^{\frac{p-2}{p}} \\
&\quad + C_4 \left(\int_M |\tilde{\nabla} g_0 - \tilde{\nabla} g_{i,0}|^p d\mu_h \right)^{\frac{1}{p}} \left(\int_M |\tilde{\nabla} g_0|^p d\mu_h \right)^{\frac{1}{p}} \mathrm{Vol}_h(M)^{\frac{p-2}{p}} \\
&\leq C_3 \left\| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right\|_{C^0(M)} + C_4 \| g_{i,0} - g_0 \|_{W^{1,p}(M)}.
\end{aligned}$$

For the last two terms, we have

$$\begin{aligned} III &= \int_M |F_i u_{i,0} - F u_{i,0}| \left| \frac{d\mu_{g_{i,0}}}{d\mu_h} \right| d\mu_h \le C_5 \int_M |F_i - F| d\mu_h \\ &\le C_5 \Big(\int_M |F_i - F|^{\frac{p}{2}} \, d\mu_h \Big)^{\frac{2}{p}} \operatorname{Vol}_h(M)^{\frac{p-2}{p}} \le C_5 \| g_{i,0} - g_0 \|^2_{W^{1,p}(M)}, \\ IV &= \int_M |F u_{i,0}| \left| \frac{d\mu_{g_{i,0}}}{d\mu_h} - \frac{d\mu_{g_0}}{d\mu_h} \right| d\mu_h \\ &\le C_6 \left\| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right\|_{C^0(M)} \Big(\int_M |\tilde{\nabla} g_0|^p d\mu_h \Big)^{\frac{1}{p}} \operatorname{Vol}_h(M)^{\frac{p-2}{p}} \\ &\le C_6 \left\| \frac{d\mu_{g_0}}{d\mu_h} - \frac{d\mu_{g_{i,0}}}{d\mu_h} \right\|_{C^0(M)}. \end{aligned}$$

Combining all of the above, we get

$$\big| \langle R_{g_{i,0}} - a, u_{i,0} \rangle - \langle R_{g_0} - a, u_{i,0} \rangle \big| \le C b_i$$

for some constant $C = C(n, k_1, k_2, \Lambda, p, \delta, \tilde{u}, a, L, \operatorname{Vol}_h(M))$ and b_i is a positive function of i which satisfies $\lim_{i\to\infty} b_i = 0$. Combining this with (4.6.2), we obtain

$$\int_M (R_{g_i(t)} - a) \tilde{u} \, d\mu_{g_i(t)} \ge \int_M (R_{g_{i,0}} - a) u_{i,0} \, d\mu_{g_{i,0}} \ge -C b_i.$$

Finally, observing that $g_i(t)$ converges smoothly to $g(t)$ as $i \to \infty$ for each $t > 0$, we let $i \to \infty$ in the inequality above to obtain the desired result. □

Together with Proposition 4.2, we have a removable singularity result as an application.

Corollary 4.7. *Suppose (M^n, h) is a closed manifold and g_0 is a $C^0 \cap W^{1,p}$ metric on M satisfying the assumption in Proposition 4.2 with $a = 0$. If the Yamabe invariant of M is nonpositive or equivalently M does not admit metric with positive scalar curvature, then g_0, then g_0 is Ricci-flat outside Σ.*

Proof. By Proposition 4.2 and Theorem 4.6, there exists a Ricci–DeTurck h-flow on M starting from g_0 in sense of $C^0(M) \cap C^\infty_{\text{loc}}(M \setminus \Sigma)$ so that $R_{g(t)} \ge 0$. It follows from the strong maximum principle that $\operatorname{Ric}_{g(t)} \equiv 0$ for all $t > 0$. The result follows from smooth convergence away from Σ by letting $t \to 0$. □

Remark 4.8. In case the metric g is Lipschitz across the singularity, we are free to choose $\delta = p$. Consider the case where the singularity is in the form of a hypersurface, it was well-known that without additional assumption on mean curvature, the scalar curvature rigidity is in general false while Miao [15] and Shi and Tam [17] showed that the mean curvature inequality is indeed necessary and sufficient. With this in mind, we see that $\delta' < \delta$ is necessary.

Acknowledgement

Man-Chun Lee was partially supported by Hong Kong RGC grant (Early Career Scheme) of Hong Kong, nos. 24304222 and 14300623, and an NSFC grant, no. 12222122.

References

[1] D. Adams, *Morrey spaces*, Birkhäuser, 2015.

[2] R. H. Bamler, E. Cabezas-Rivas, and B. Wilking, "The Ricci flow under almost non-negative curvature conditions", *Invent. Math.* **217**:1 (07 2019), 95–126.

[3] P. Burkhardt-Guim, "Pointwise lower scalar curvature bounds for C^0 metrics via regularizing Ricci flow", *Geom. Funct. Anal.* **29**:6 (2019), 1703–1772.

[4] J. Chu and M.-C. Lee, "Ricci–DeTurck flow from rough metrics and applications to scalar curvature problems", *J. Funct. Anal.* **289**:2 (2025), art. id. 110916, 41 pp.

[5] D. M. DeTurck, "Deforming metrics in the direction of their Ricci tensors", *J. Differ. Geom.* **18**:1 (1983-01), 157–162.

[6] J. D. E. Grant and N. Tassotti, "A positive mass theorem for low-regularity Riemannian metrics", preprint, 2014. arXiv 1408.6425

[7] R. S. Hamilton, "Three-manifolds with positive Ricci curvature", *J. Differ. Geom.* **17**:2 (1982-01), 255–306.

[8] W. Jiang, W. Sheng, and H. Zhang, "Removable singularity of positive mass theorem with continuous metrics", *Math. Z.* **302**:2 (2022), 839–874.

[9] W. Jiang, W. Sheng, and H. Zhang, "Weak scalar curvature lower bounds along Ricci flow", *Sci. China Math.* **66** (2023), 1141–1160.

[10] H. Koch and T. Lamm, "Geometric flows with rough initial data", *Asian J. Math.* **16**:2 (2012-06), 209–235.

[11] T. Lamm and M. Simon, "Ricci flow of $W^{2,2}$-metrics in four dimensions", *Comment. Math. Helv.* **98** (09 2023), 261–364.

[12] D. A. Lee, "A positive mass theorem for Lipschitz metrics with small singular sets", *Proc. Amer. Math. Soc.* **141**:11 (2013), 3997–4004.

[13] D. A. Lee and P. G. LeFloch, "The Positive Mass Theorem for Manifolds with Distributional Curvature", *Commun. Math. Phys.* **339**:1 (2015), 99–120.

[14] M.-C. Lee and L.-F. Tam, "Continuous metrics and a conjecture of Schoen", *Trans. Amer. Math. Soc.* **378**:3 (2025), 1531–1550.

[15] P. Miao, "Positive mass theorem on manifolds admitting corners along a hypersurface", *Adv. Theor. Math. Phys.* **6**:6 (2003), art. id. 4.

[16] W.-X. Shi, "Deforming the metric on complete Riemannian manifolds", *J. Differ. Geom.* **30**:1 (1989-01), 223–301.

[17] Y. Shi and L.-F. Tam, "Scalar curvature and singular metrics", *Pac. J. Math.* **293**:2 (2016), 427–470.

[18] M. Simon, "Deformation of C^0 Riemannian metrics in the direction of their Ricci curvature", *Comm. Anal. Geom.* **10**:5 (2002), 1033–1074.

[19] L.-F. Tam, "Exhaustion functions on complete manifolds", pp. 211–215 in *Recent advances in geometric analysis*, Advanced Lectures in Mathematics **11**, International Press, Somerville, MA, 2010. MR

Received July 13, 2024. Revised September 19, 2025.

MAN-CHUN LEE
DEPARTMENT OF MATHEMATICS
CHINESE UNIVERSITY OF HONG KONG
HONG KONG
mclee@math.cuhk.edu.hk

STEPHEN SHANG YI LIU
DEPARTMENT OF MATHEMATICS
HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY
HONG KONG
masyliu@ust.hk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.399

THE GEOMETRIC CAUCHY PROBLEM FOR CONSTANT-RANK SUBMANIFOLDS

MATTEO RAFFAELLI

Given a smooth s-dimensional submanifold S of $\mathbb{R}^{m+c}$ and a smooth distribution $\mathcal{D} \supset TS$ of rank m along S, we study the following geometric Cauchy problem: to find an m-dimensional rank-s submanifold M of $\mathbb{R}^{m+c}$ (that is, an m-submanifold with constant index of relative nullity $m-s$) such that $M \supset S$ and $TM|_S = \mathcal{D}$. In particular, under some reasonable assumption and using a constructive approach, we show that a solution exists and is unique in a neighborhood of S.

1. Introduction and main result

Given a manifold $Q^{q\geq 3}$ and some class $\mathscr{A}$ of $(m<q)$-dimensional submanifolds, the *geometric Cauchy problem for the class* $\mathscr{A}$ is to find all members of $\mathscr{A}$ passing from a given $(s<m)$-dimensional submanifold S of Q with prescribed tangent bundle along S. This problem, which represents a far-reaching generalization of the classical Björling problem for minimal surfaces in $\mathbb{R}^3$ [18, Section 22.6], has recently been examined, always with $s=1$, for several combinations of Q and $\mathscr{A}$; see, e.g., [24; 5; 3; 16; 9; 23; 7; 2; 1].

In this paper, we are interested in the geometric Cauchy problem for *constant-rank submanifolds*, that is, smooth embedded submanifolds of $\mathbb{R}^{m+c}$ whose Gauss map has constant rank, or, equivalently, whose second fundamental form has constant index of relative nullity [8]. These constitute one of the simplest and most natural classes of submanifolds and, for this reason, also one of the most studied; see, e.g., [26; 25; 14; 17; 15; 29; 12; 10] and [13, Chapter 7].

Problem 1.1 (geometric Cauchy problem for constant-rank submanifolds). Let S be a connected s-dimensional smooth embedded submanifold of $\mathbb{R}^{m+c}$, and let $\mathcal{D} \supset TS$ be a smooth distribution of rank m along S, that is, a smooth rank-m subbundle of the ambient tangent bundle over S. Find all rank-s submanifolds of $\mathbb{R}^{m+c}$ containing S and whose tangent bundle along S is precisely $\mathcal{D}$.

MSC2020: primary 53A07; secondary 53B25, 58A30.

Keywords: constant nullity, distributions along submanifolds, index of relative nullity, ruled submanifold, vector cross product.

© 2026 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Some special cases of Problem 1.1 have already been considered: the case $m = 2$ is classical [6, pp. 198–200]; the case of rank-one hypersurfaces ($s = c = 1$) was first addressed by Markina and the author in [22]; and the rank-one case in arbitrary codimension was solved by the author in [27]. The interest in Problem 1.1 is motivated by the search for new approaches, alternative to the Gauss parametrization [11; 28], to describe constant-rank submanifolds. The purpose of this paper is to present a solution of Problem 1.1 without any assumption on s, m, or c.

To state our main result, let $\mathcal{D}^\perp$ be the distribution of rank c whose fiber at $p \in S$ is the orthogonal complement of $\mathcal{D}_p$ in $\mathbb{R}^{m+c}$, let $\pi^\top$ denote orthogonal projection onto $\mathcal{D}$, and let ϕ_p be the map

$$\begin{aligned} T_pS \times \mathcal{D}_p^\perp &\to \mathcal{D}_p \\ (v, n) &\mapsto \pi^\top \nabla_v N, \end{aligned}$$

where ∇ is the Euclidean connection, and where N is any extension of n to a local section of $\mathcal{D}^\perp$. It is easy to see that ϕ_p is well defined.

Theorem 1.2. *Suppose that there exists a section N^* of $\mathcal{D}^\perp$ such that the shape operator $A^* := A_{N^*}$ of S in direction N^* is nonsingular, i.e.,*

$$\operatorname{rank} A^*|_p = s \quad \forall p \in S.$$

The geometric Cauchy problem for rank-s submanifolds of $\mathbb{R}^{m+c}$ has a solution if and only if $\phi_p(T_pS, N_p^*) = \operatorname{Im}\phi_p$ for all $p \in S$. Moreover:

(i) *The solution is locally unique, i.e., any two solutions coincide on an open set containing S.*

(ii) *In this open set, the unique solution is given by*

$$\left\{p + x \mid p \in S,\, x \in \mathcal{D}_p \cap (\operatorname{Im}\phi_p)^\perp\right\},$$

where $^\perp$ denotes orthogonal complement in $\mathbb{R}^{m+c}$.

(iii) *Given a local parametrization $\xi : A \to \mathbb{R}^{m+c}$ of S, the solution can be locally parametrized as follows. Let $(E_1, \dots, E_m)$ be a smooth orthonormal frame for $\mathcal{D}|_{\xi(A)}$ whose first s elements span TS. For any $j = 1, \dots, m-s$, let G_i be the group of permutations of the set $\{1, \dots, \widehat{i}, \dots, s, s+j\}$ obtained by omitting the element i, and, denoting by ϕ_i^k the coordinate function $\langle \phi(E_i, N^*), E_k\rangle = \langle \nabla_{E_i} N^*, E_k\rangle$, let*

$$X_j = \sum_{i \in \{1,\dots,s,s+j\}} (-1)^h \sum_{\lambda \in G_i} \operatorname{sgn}(\lambda) \phi_1^{\lambda(1)} \cdots \phi_s^{\lambda(s+j)} E_i, \tag{1}$$

where $h = i$ for $i = 1, \dots, s$ and $h = s+1$ otherwise. Then

$$\sigma(a, b^1, \dots, b^{m-s}) = \xi(a) + b^1 X_1(a) + \cdots + b^{m-s} X_{m-s}(a)$$

is the desired parametrization.

Remark 1.3. $A^* = \pi_S \circ \phi(\cdot, N^*)$, where π_S denotes orthogonal projection onto TS.

Remark 1.4. The section N^* does not need to be global. The theorem remains valid if one merely assumes the existence of a local section $N^*(p)$ in a neighborhood U_p of each point $p \in S$, along with the compatibility condition

$$N^*(p) = \pm N^*(q) \quad \forall p, q \text{ such that } U_p \cap U_q \neq \varnothing.$$

Remark 1.5. If the distribution $\mathcal{D}$ is obtained by restricting to S the tangent bundle of an arbitrary submanifold $M \supset S$, then the map $\phi_p(\cdot, n)$ coincides with the restriction to T_pS of the shape operator of M with respect to n. In this case the theorem provides conditions for the existence of a constant-rank, first-order approximation of M along S; cf. [27, section 6].

Remark 1.6. When $s = 1$, formula (1) reduces to $X_j = \phi^1 E_{1+j} - \phi^{1+j} E_1$, and we thus retrieve the third item in [27, Theorem 1.3]. The formula is new starting from $s = 2$, when it reduces to $X_j = -(\phi_1^2 \phi_2^{2+j} - \phi_1^{2+j} \phi_2^2) E_1 + (\phi_1^1 \phi_2^{2+j} - \phi_1^{2+j} \phi_2^1) E_2 - (\phi_1^1 \phi_2^2 - \phi_1^2 \phi_2^1) E_{j+2}$.

Clearly, when $c = 1$, the condition on ϕ_p is automatically satisfied, and Problem 1.1 becomes generically well posed.

Corollary 1.7. *Suppose that $c = 1$. If for each $p \in S$ the shape operator of S in direction $n \in \mathcal{D}_p^{\perp}$ is nonsingular, then the geometric Cauchy problem for rank-s hypersurfaces of $\mathbb{R}^{m+1}$ has a solution.*

The proof of Theorem 1.2 will be given in Section 3. It relies on the well-known fact that any m-dimensional rank-s submanifold of $\mathbb{R}^{m+c}$ admits a foliation by totally geodesic leaves of dimension $m - s$, along which the tangent space is constant; the converse holds provided that the foliation is unique. In particular, as shown in the next section, the tangent space is constant along the rulings precisely when the ambient covariant derivative of any tangent vector field along S has vanishing normal component (Lemma 2.5). This extends a classical result about developable surfaces in $\mathbb{R}^3$ [20, Theorem 54.1]; see also [30].

2. Constant-rank submanifolds

The purpose of this section is to generalize [27, Lemma 3.7] to $(m-s)$-*ruled* submanifolds.

Definition 2.1. Let M be an m-dimensional embedded submanifold of $\mathbb{R}^{m+c}$. We say that M is $(m-s)$-*ruled* if it admits a foliation by $(m-s)$-dimensional totally geodesic submanifolds of $\mathbb{R}^{m+c}$, called *rulings*. In particular, we say that M is *properly* $(m-s)$-*ruled* when the foliation is unique.

Given a s-dimensional embedded submanifold S of $\mathbb{R}^{m+c}$, the following lemma is the key to extending S to an $(m-s)$-ruled submanifold. Recall that two embedded submanifolds S, S' of a smooth manifold are said to *intersect transversely* if for each $p \in S \cap S'$, the tangent spaces T_pS and T_pS' together span the ambient tangent space.

Lemma 2.2. *Let $\mathcal{F}$ be a distribution of rank $m-s$ along S such that*

$$\dim(T_pS \oplus \mathcal{F}_p) = m \quad \forall p \in S.$$

Then the subset

$$\{p + x \mid p \in S,\, x \in \mathcal{F}_p\}$$

defines an $(m-s)$-ruled submanifold $M_{\mathcal{F}}$ in a neighborhood of S. Conversely, any $(m-s)$-ruled submanifold containing S and whose rulings intersect S transversely admits a description of this form.

Proof. We first prove a local version of the lemma. Let $\xi : A \to \mathbb{R}^{m+c}$ be a local parametrization of S, let $(X_1, \dots, X_{m-s})$ be a local frame for $\mathcal{F}$ over A, and, for $B \subset \mathbb{R}^{m-s}$ open, define $\sigma : A \times B \to \mathbb{R}^{m+c}$ by

$$\sigma(a, b^1, \dots, b^{m-s}) = \xi(a) + b^1 X_1(a) + \cdots + b^{m-s} X_{m-s}(a).$$

Since the partial derivatives of σ are linearly independent at $(a, 0)$, they remain linearly independent in some neighborhood U_a of $(a, 0)$ in $A \times B$. Moreover, after shrinking U_a and B if necessary, we may assume that $\sigma(U_a)$ is embedded and there exists a neighborhood V_a of $\xi(A)$ in $\mathbb{R}^{m+c}$ such that

$$\sigma(U_a) = V_a \cap \sigma(A \times B);$$

to see this, note that otherwise, by making B arbitrarily close to 0, the assumption that S is embedded would be violated. Let $M := \bigcup_{a \in A} U_a$, and let $q \in M$. Since M is an immersed submanifold, the rank theorem [21, Theorem 4.12] implies the existence of a slice chart for M in $\mathbb{R}^{m+c}$ around q. Hence M satisfies the local slice criterion for embedded submanifolds [21, Theorem 5.8], which is the desired local conclusion.

To prove the full statement, let $\{\xi_i : A_i \to \mathbb{R}^{m+c}\}_{i=1}^r$ be a set of local parametrizations of S covering S, and let $\{\sigma_i : A_i \times B_i \to \mathbb{R}\}_{i=1}^r$ be the corresponding maps σ, as defined earlier. Then we know that for each i there exists a neighborhood U_i of $(A_i, 0)$ in $A_i \times B_i$ such that $\sigma_i(U_i)$ is embedded. Shrinking $U_1, \dots, U_r$ along the

second factor if necessary, we may assume that there are neighborhoods $V_1, \dots, V_r$ of $\xi_1(A_1), \dots, \xi_r(A_r)$ in $\mathbb{R}^{m+c}$ such that

$$\sigma_i(U_i) = V_i \cap \Big(\bigcup_{i=1}^{r} \sigma_i(U_i)\Big) \quad \forall i = 1, \dots, r.$$

As before, this implies the existence of a slice chart for $M_i = \sigma_i(U_i)$ in $\mathbb{R}^{m+c}$ around every $q \in M_i$. Hence $M := \bigcup_i M_i$ is the desired $(m-s)$-ruled submanifold. □

Next we define constant-rank submanifolds and give two alternative characterizations of them.

Definition 2.3. Let M be an m-dimensional embedded submanifold of $\mathbb{R}^{m+c}$. The *index of relative nullity* of M at a point $p \in M$ is the dimension of the kernel Δ of the second fundamental form α at p:

$$\Delta = \{x \in T_pM \mid \alpha(x, \cdot) = 0\}.$$

We say that M is a *constant-rank* submanifold if the index of relative nullity is constant. In particular, we say that M is *rank-s* when the index is equal to $m - s$.

Theorem 2.4 [19, Lemma 3.1]. *The following statements are equivalent*:

(1) *M is rank-s.*

(2) *The Gauss map $M \to Gr(m, m+c)$ has rank s everywhere*; *here $Gr(m, m+c)$ denotes the Grassmannian of m-dimensional subspaces of $\mathbb{R}^{m+c}$.*

(3) *M is properly $(m-s)$-ruled with constant tangent space along the $(m-s)$-rulings.*

Thus, rank-s submanifolds containing S are locally in one-to-one correspondence with distributions $\mathcal{F}$ of rank $m - s$ along S making $M_{\mathcal{F}}$ properly ruled and satisfy

$$T_pM_{\mathcal{F}} = T_{p+x}M_{\mathcal{F}} \quad \forall x \in \mathcal{F}_p \text{ such that } p + x \in M_{\mathcal{F}}.$$

The next lemma expresses the last condition in terms of the ambient covariant derivative along S and represents the sought generalization of [27, Lemma 3.7].

Lemma 2.5. *The tangent space of $M_{\mathcal{F}}$ is constant along $x \in \mathcal{F}_p$ if and only if*

$$\nabla_v X \in T_pM_{\mathcal{F}} \quad \forall v \in T_pS,$$

where X is any extension of x to a local section of $\mathcal{F}$.

Proof. Let $\xi : A \to \mathbb{R}^{m+c}$ be a local parametrization of S around p, and let $(X_j)_{j=1}^{m-s}$ be a local frame for $\mathcal{F}$ over A. Then the map $\sigma : A \times \mathbb{R}^{m-s} \to \mathbb{R}^{m+n}$ defined by

$$\sigma(a, b^1, \dots, b^{m-s}) = \xi(a) + b^1 X_1(a) + \cdots + b^{m-s} X_{m-s}(a)$$

is a local parametrization of $M_{\mathcal{F}}$ in a neighborhood of p. Since the tangent space $T_{(a,0,\dots,0,b^j,0,\dots,0)}\sigma$ of σ at $\sigma(a,0,\dots,0,b^j,0,\dots,0)$ is spanned by

$$\partial_1\xi(a)+b^j\partial_1 X_j(a),\ \dots,\ \partial_s\xi(a)+b^j\partial_s X_j(a),\ X_1(a),\ \dots,\ X_{m-s}(a),$$

it follows that

$$T_{(a,0,\dots,0,b^j,0,\dots,0)}\sigma = T_{(a,0)}\sigma = \operatorname{span}\big(\partial_1\xi(a),\dots,\partial_s\xi(a),X_1(a),\dots,X_{m-s}(a)\big)$$

if and only if $\partial_i X_j(a)\in T_{(a,0)}\sigma$ for every $i=1,\dots,s$; by linearity, this is equivalent to all directional derivatives of X_j being in $T_{(a,0)}\sigma$ at a. Now, to complete the proof, it is enough to note that any of the elements of the frame $(X_j)_{j=1}^{m-s}$ may be chosen to coincide with x at p. □

Remark 2.6. It is easy to see, using orthogonality of $T_pM_{\mathcal{F}}$ and $(T_pM_{\mathcal{F}})^{\perp}$, that the condition $\nabla_v X\in T_pM_{\mathcal{F}}$ in Lemma 2.5 only depends on X through $X_p=x\in T_pM_{\mathcal{F}}$. Hence the lemma continues to hold if "local section of $\mathcal{F}$" is replaced by "local section of $TM_{\mathcal{F}}$".

Remark 2.7. By the properties of the covariant derivative, Lemma 2.5 implies that the tangent space of $M_{\mathcal{F}}$ is constant along every $x\in\mathcal{F}$ if and only if

$$\nabla_{v_i}X_j\in T_pM_{\mathcal{F}}\quad\forall i,j,$$

where $(v_1,\dots,v_s)$ is a basis of T_pS and $(X_1,\dots,X_{m-s})$ a local frame for $\mathcal{F}$ around p.

3. Proof of the main result

In this section we prove Theorem 1.2. The proof of the first two items is based on the following lemma.

Lemma 3.1. *Suppose that there exists a section N^* of $\mathcal{D}^{\perp}$ such that the shape operator A^* of S in direction N^* is nonsingular at p. Then*

$$T_pS\cap\phi_p(T_pS,N_p^*)^{\perp}=\{0\}.$$

Proof. Let π_S denote orthogonal projection onto TS, so that

$$A_N=\pi_S\circ\phi(\cdot,N).$$

Suppose, towards a contradiction, that there exists a *nonzero* vector $v\in T_pS$ such that

$$\langle v,\phi(w,N^*)\rangle=0\quad\forall w\in T_pS.$$

Since $\phi(\cdot,N)$ is self-adjoint with respect to the metric, this implies

$$A^*(v)=0,$$

contradicting the assumption that A^* has full rank at p. □

Proof of Theorem 1.2(i)–(ii). According to Lemmas 2.2 and 2.5, we need to find an $(m-s)$-dimensional subspace $\mathcal{F}_p$ of $\mathcal{D}_p$ satisfying $T_pS \cap \mathcal{F}_p = \{0\}$ and

$$\nabla_v x \in \mathcal{D}_p \quad \forall (v, x) \in T_pS \times \mathcal{F}_p,$$

where x is being extended arbitrarily to a local section X of $\mathcal{D}$. Note that the latter condition is equivalent to

$$\langle \nabla_v x, n \rangle = 0 \quad \forall (v, x, n) \in T_pS \times \mathcal{F}_p \times \mathcal{D}_p^{\perp}; \tag{2}$$

using orthogonality of x and n, we can rewrite (2) as

$$\langle x, \phi_p(v, n) \rangle = 0 \quad \forall (v, x, n) \in T_pS \times \mathcal{F}_p \times \mathcal{D}_p^{\perp}$$

or, more succinctly, as

$$\mathcal{F}_p \subset (\operatorname{Im} \phi_p)^{\perp}.$$

Suppose that there exists a section N^* of $\mathcal{D}^{\perp}$ such that the shape operator A^* of S in direction N^* is nonsingular at p, and set

$$\mathcal{F}_p^* := \mathcal{D}_p \cap \phi_p(T_pS, N_p^*)^{\perp}.$$

Then, since $A_{N^*} = \pi_S \circ \phi(\,\cdot\,, N^*)$, we deduce that $\mathcal{F}_p^*$ is a subspace of dimension $m - s$; besides, it has trivial intersection with T_pS by Lemma 3.1. Observing that

$$(\operatorname{Im} \phi_p)^{\perp} \subset \phi_p(T_pS, N_p^*)^{\perp},$$

we conclude that $\mathcal{F}^*$ can only define a solution of our problem when $\operatorname{Im} \phi_p = \phi_p(T_pS, N_p^*)$.

Next we check that $\mathcal{F}^*$ does indeed define a solution. To this end, let $M^* := M_{\mathcal{F}^*}$ be an embedded submanifold containing S, whose existence is guaranteed by Lemma 2.2. First note that, by construction, the shape operator A^* of S is nothing but the restriction to TS of a shape operator of M^*. Hence, as the tangent space is constant along $\mathcal{F}^*$, the index of relative nullity of M^* equals $m - s$ along S. By the same reason, extending N^* via parallel translation in the ambient space to a unit normal vector field along M^*, we have that

$$A^*_{p+x}(v) = A^*_p(v) \quad \forall v \in \mathcal{D}_p \equiv T_{p+x}M^*,$$

which implies that the index of relative nullity of M^* is constant, as claimed. □

It remains to prove Theorem 1.2(iii). To set the stage, we first recall the notion of ($(m-1)$-fold) vector cross product.

Let V be an m-dimensional oriented inner product space, and let $(e_1, \dots, e_m)$ be an orthonormal basis of V. A *vector cross product on* V is a multilinear map $\cdot \times \cdots \times \cdot : V^{m-1} \to V$ such that

$$\begin{aligned}
&\langle e_1 \times \cdots \times e_{m-1}, e_1 \rangle = \cdots = \langle e_1 \times \cdots \times e_{m-1}, e_{m-1} \rangle = 0,\\
&\langle e_1 \times \cdots \times e_{m-1}, e_1 \times \cdots \times e_{m-1} \rangle = \det(\langle e_i, e_j \rangle)_{i,j=1}^{m-1},\\
&(e_1, \dots, e_{m-1}, e_1 \times \cdots \times e_{m-1}) \text{ is positively oriented.}
\end{aligned}$$

It is well known that there is only one vector cross product on V. It is given by

$$v_1 \times \cdots \times v_{m-1} = \star(v_1 \wedge \cdots \wedge v_{m-1}),$$

where $\star$ denotes the Hodge star operator and $\wedge$ the wedge product on V; see [4]. More generally, if $v_1, \dots, v_{m-1}$ are continuous sections of a smooth oriented vector bundle over a smooth manifold, then both $\star$ and $\wedge$ are globally defined, and the previous equation defines their cross product.

Our proof of Theorem 1.2(iii) is based on the following lemma.

Lemma 3.2. *Let W be an s-dimensional subspace of V, and let ρ be a linear map $W \to V$; without loss of generality, we may assume that the first s elements of $(e_1, \dots, e_m)$ span W and the first r elements span* $\operatorname{coim}\rho = (\ker\rho)^\perp$. *Suppose that the orthogonal complement $\varphi(W)^\perp$ of $\varphi(W)$ in V has trivial intersection with* $\operatorname{coim}\rho$. *Then*

$$x_j = \varphi(e_1) \times \cdots \times \varphi(e_r) \times e_{r+1} \times \cdots \times \widehat{e_{r+j}} \times \cdots \times e_m, \quad j = 1, \dots, m-r$$

defines an orthogonal basis of $\varphi(W)^\perp$; here the hat indicates that e_{r+j} is omitted.

Proof. By construction, the vector x_j belongs to the intersection between $\varphi(W)^\perp$ and the subspace spanned by $e_1, \dots, e_r, e_{r+j}$. Since $\varphi(W)^\perp$ has trivial intersection with $\operatorname{span}(e_1, \dots, e_r)$, it follows that $x_1, \dots, x_{m-r}$ all lie in orthogonal subspaces, showing that they indeed span $\varphi(W)^\perp$. □

Proof of Theorem 1.2(iii). Let $(E_1, \dots, E_m)$ be a smooth orthonormal frame for $\mathcal{D}|_{\xi(A)}$ whose first s elements span TS. Equipping $\mathcal{D}|_{\xi(A)}$ with the orientation naturally induced by that frame, the vector cross product becomes well defined on $\mathcal{D}|_{\xi(A)}$. Then, for each $j = 1, \dots, m-s$, let

$$X_j = \phi(E_1, N^*) \times \cdots \times \phi(E_s, N^*) \times E_{s+1} \times \cdots \times \widehat{E_{s+j}} \times \cdots \times E_m, \tag{3}$$

where the hat indicates that E_{s+j} is omitted. Since $\operatorname{coim}\phi = TS$ by assumption, we deduce from Lemmas 3.1 and 3.2 that $(X_1, \dots, X_{m-s})$ is an orthogonal frame for $\phi(TS, N^*)^\perp$.

It remains to compute (3) in coordinates. Let $\phi_i^k = \langle \phi(E_i, N^*), E_k \rangle$. Then

$$\phi(E_i, N^*) = \phi_i^1 E_1 + \cdots + \phi_i^m E_m.$$

Substitution into (3) gives

$$X_j = \big(\phi_1^1 E_1 + \cdots + \phi_1^s E_s + \phi_1^{s+j} E_{s+j}\big) \times \cdots \times (\phi_s^1 E_1 + \cdots + \phi_s^s E_s + \phi_s^{s+j} E_{s+j}) \times E_{s+1} \times \cdots \times \widehat{E_{s+j}} \times \cdots \times E_m,$$

which implies (1). □

References

[1] J. A. Aledo, J. A. Gálvez, and P. Mira, "A D'Alembert formula for flat surfaces in the 3-sphere", *J. Geom. Anal.* **19**:2 (2009), 211–232. MR

[2] D. Brander and M. Svensson, "The geometric Cauchy problem for surfaces with Lorentzian harmonic Gauss maps", *J. Differential Geom.* **93**:1 (2013), 37–66. MR

[3] D. Brander and P. Wang, "On the Björling problem for Willmore surfaces", *J. Differential Geom.* **108**:3 (2018), 411–457. MR

[4] R. B. Brown and A. Gray, "Vector cross products", *Comment. Math. Helv.* **42** (1967), 222–236. MR

[5] A. Bueno, "The Björling problem for prescribed mean curvature surfaces in $\mathbb{R}^3$", *Ann. Global Anal. Geom.* **56**:1 (2019), 87–96. MR

[6] M. do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall, Englewood Cliffs, NJ, 1976. Reprinted 2016, Dover, Mineola, NY. MR

[7] R. M. B. Chaves, M. P. Dussan, and M. Magid, "Björling problem for timelike surfaces in the Lorentz–Minkowski space", *J. Math. Anal. Appl.* **377**:2 (2011), 481–494. MR

[8] S.-s. Chern and N. H. Kuiper, "Some theorems on the isometric imbedding of compact Riemann manifolds in euclidean space", *Ann. of Math.* (2) **56** (1952), 422–430. MR

[9] A. A. Cintra, F. Mercuri, and I. I. Onnis, "The Björling problem for minimal surfaces in a Lorentzian three-dimensional Lie group", *Ann. Mat. Pura Appl.* (4) **195**:1 (2016), 95–110. MR

[10] M. Dajczer and L. A. Florit, "Genuine rigidity of Euclidean submanifolds in codimension two", *Geom. Dedicata* **106** (2004), 195–210. MR

[11] M. Dajczer and D. Gromoll, "Gauss parametrizations and rigidity aspects of submanifolds", *J. Differential Geom.* **22**:1 (1985), 1–12. MR

[12] M. Dajczer and P. Morais, "Isometric rigidity in codimension 2", *Michigan Math. J.* **58**:3 (2009), 759–770. MR

[13] M. Dajczer and R. Tojeiro, *Submanifold theory: beyond an introduction*, Springer, 2019. MR

[14] M. Dajczer and T. Vlachos, "The infinitesimally bendable Euclidean hypersurfaces", *Ann. Mat. Pura Appl.* (4) **196**:6 (2017), 1961–1979. MR

[15] M. Dajczer, L. Florit, and R. Tojeiro, "Euclidean hypersurfaces with genuine deformations in codimension two", *Manuscripta Math.* **140**:3-4 (2013), 621–643. MR

[16] M. P. Dussan, A. P. Franco Filho, and M. Magid, "The Björling problem for timelike minimal surfaces in $\mathbb{R}^4_1$", *Ann. Mat. Pura Appl.* (4) **196**:4 (2017), 1231–1249. MR

[17] L. A. Florit and G. M. de Freitas, "Classification of codimension two deformations of rank two Riemannian manifolds", *Comm. Anal. Geom.* **25**:4 (2017), 751–797. MR

[18] A. Gray, E. Abbena, and S. Salamon, *Modern differential geometry of curves and surfaces with Mathematica*, 3rd ed., CRC, Boca Raton, FL, 2006. MR

[19] P. Hartman, "On isometric immersions in Euclidean space of manifolds with non-negative sectional curvatures", *Trans. Amer. Math. Soc.* **115** (1965), 94–109. MR

[20] E. Kreyszig, *Introduction to differential geometry and Riemannian geometry*, Mathematical Expositions **16**, University of Toronto Press, 1968. MR

[21] J. M. Lee, *Introduction to smooth manifolds*, 2nd ed., Graduate Texts in Mathematics **218**, Springer, 2013. MR

[22] I. Markina and M. Raffaelli, "Flat approximations of hypersurfaces along curves", *Manuscripta Math.* **160**:3-4 (2019), 315–325. MR

[23] A. Martínez and F. Milán, "The geometric Cauchy problem for the hyperbolic Hessian one equation", *Nonlinear Anal.* **125** (2015), 323–333. MR

[24] F. Milán, "The indefinite affine Cauchy problem", *Nonlinear Anal.* **196** (2020), art. id. 111819, 11 pp. MR

[25] C.-R. Onti, "On complete conformally flat submanifolds with nullity in Euclidean space", *Results Math.* **75**:3 (2020), art. id. 106, 7 pp. MR

[26] M. Raffaelli, "Classification of rank-one submanifolds", *Results Math.* **78**:6 (2023), art. id. 208, 13 pp. MR

[27] M. Raffaelli, "The geometric Cauchy problem for rank-one submanifolds", *Comm. Anal. Geom.* **32**:1 (2024), 323–342. MR

[28] U. Sbrana, "Sulla deformazione infinitesima delle ipersuperficie", *Ann. Mat. Pura Appl.* **15** (1908), 329–348.

[29] F. Vittone, "On the nullity distribution of a submanifold of a space form", *Math. Z.* **272**:1-2 (2012), 1–16. MR

[30] K. Yano, "On the torse-forming directions in Riemannian spaces", *Proc. Imp. Acad. Tokyo* **20** (1944), 340–345. MR

Received September 9, 2024. Revised September 9, 2025.

MATTEO RAFFAELLI

raffaelli@math.gatech.edu
SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GA 30332
UNITED STATES

PACIFIC JOURNAL OF MATHEMATICS
Vol. 340, No. 2, 2026
https://doi.org/10.2140/pjm.2026.340.409

ON THE PRIMITIVE IDEAL SPACE OF RADIAL REPRESENTATIONS OF FREE GROUPS

SHIGERU YAMAGAMI

Dedicated to the memory of Shukoh Ishikawa

Radial representations of finitely generated free groups are studied. The associated C*-algebra is located between the reduced and full group C*-algebras and its primitive ideal space is described concretely as a topological space.

Introduction

Spherical functions on finitely generated free groups have been investigated from various points of view. Among them, fundamental is the framework of Poisson boundaries, which allows us to take analogies with the case of semisimple Lie groups. In fact, studies of spherical functions go more or less around it and, as their basic properties, irreducibility and inequivalence of the associated spherical representations are established under this background.

Other than spherical representations, a series of nonirreducible representations are associated with radial functions. Within that class interesting are positive definite functions of Haagerup, which include the standard trace as a limit case. For nonirreducible representations associated with positive definite radial functions, a generalization of Plancherel formula is described in [22] via spectral decomposition of radial functions based on the radial algebra which is commutative. Notice here that, different from the semisimple case, noncommutative free groups are not type I and the uniqueness of decomposition into irreducible representations breaks down.

In other words, to obtain a Plancherel formula, we need to specify a maximal commutative subalgebra first and the above-mentioned Plancherel formula is based on the algebra of radial functions. As an instance, the standard trace belongs to the radial class and its Plancherel measure turns out to be the so-called Kesten measure, which is supported by the regular spectrum and equivalent to the Lebesgue measure.

In this paper, we look into the primitive ideal space of radial representations and show that it is given by primitive ideals of spherical representations. Furthermore,

MSC2020: 22D25, 46L45.
Keywords: primitive ideal, free group, radial function.

© 2026 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

its hull-kernel topology is described explicitly in terms of the spectral parameter which distinguishes equivalence classes of spherical representations.

During the process of description, we also see that the kernel of the regular representation is represented by a CCR algebra under the universal radial representation and a complete description of the primitive ideal space of the radial representation is obtained in terms of primitive ideals of this CCR algebra.

The author is grateful to Tomohiro Hayashi for many discussions on the subject.

1. Background review

We shall work with the countable discrete group G generated freely by a finite set $\{s_1, \ldots, s_l\}$ and representations of G by bounded linear operators on a Hilbert space $\mathcal{H}$. By freeness, such a representation is specified by assigning a finite family $(S_i)_{1\le i\le l}$ of bounded invertible operators on $\mathcal{H}$ so that a unitary representation corresponds to a family of unitary operators.

Let $\mathbb{C}G = \sum_{g\in G} \mathbb{C}g$ be the algebraic group algebra of G, which is naturally identified with the convolution algebra of functions on G having finite supports. The completion $\ell^1(G)$ of $\mathbb{C}G$ with respect to the ℓ^1-norm consists of summable sequences labeled by elements in G so that the convolution product makes $\ell^1(G)$ a unital Banach algebra.

The *-operation $g \mapsto g^* = g^{-1}$ on $\mathbb{C}G$ is obviously extended to $\ell^1(G)$ isometrically. Thus unitary representations of G are in one-to-one correspondence with *-representations of $\mathbb{C}G$ or $\ell^1(G)$ on Hilbert spaces. The associated universal C*-algebra is called the group C*-algebra of G and denoted by $C^*(G)$. The *parity automorphism* ϖ of $C^*(G)$ is defined by $\varpi(g) = (-1)^{|g|}g$, where $|g|$ denotes the length of $g \in G$ with respect to the generator $\{s_1, \ldots, s_l\}$.

Among *-representations, the regular representation on $\ell^2(G)$ plays a central role in what follows: the left regular representation is denoted by λ. The image $\lambda(C^*(G))$, also denoted by $C_r^*(G)$, is the reduced group C*-algebra and known to be simple ([13]). Recall that the regular representation is related with the standard trace τ on $C^*(G)$ specified by $\tau(g) = \delta_{g,e}$ for $g \in \mathbb{C}G \subset C^*(G)$: Let $\tau^{1/2}$ be the GNS vector of τ. Then, for $f = \sum_{g\in G} f(g)g \in \mathbb{C}G$, $f\tau^{1/2} = \tau^{1/2}f$ is identified with the function $f \in \ell^2(G)$ so that $\lambda(f)$ is realized by left multiplication of f on the GNS representation space of τ.

The parity automorphism ϖ is implemented on $C_r^*(G)$ by the parity operator P on $\ell^2(G)$ defined by $P(g\tau^{1/2}) = \varpi(g)\tau^{1/2}$.

We shall now review (more or less well-known) relevant facts on spherical representations of free groups together with some comments.

Spherical functions on free groups are introduced and studied as counterparts of the semisimple case (see [3; 6; 9; 16] for example).

A function of $g \in G$ is said to be *radial* if it depends only through the length $|g|$. Radial functions supported by finite sets constitute a commutative *-subalgebra $\mathcal{A}$ of $\mathbb{C}G$. Let $G_n = \{g \in G;\ |g| = n\}$ with its indicator function denoted by 1_{G_n}. Note that the number of elements in G_n is $|G_n| = 2l(2l-1)^{n-1}$ $(n \geq 1)$. As a linear base of $\mathcal{A}$, elementary radial functions are introduced by

$$h_n = \frac{1}{|G_n|} 1_{G_n} \quad (n = 0, 1, 2, \dots),$$

which are hermitian $h_n^* = h_n$ with h_0 the unit element of $\mathbb{C}G$ and fulfill the recurrence relation

$$h_1 h_n = r h_{n-1} + (1-r) h_{n+1} \quad (n \geq 1)$$

with $r = 1/2l$ satisfying $0 < r \leq 1/2$.

If we introduce a polynomial sequence $p_n(t)$ of indeterminate t by $p_0(t) = 1$, $p_1(t) = t$ and $t p_n(t) = r p_{n-1}(t) + (1-r) p_{n+1}(t)$ $(n \geq 1)$, then $h_n = p_n(h_1)$ and

$$\sum_{n=0}^{\infty} p_n(t) z^n = \frac{1 - r - rtz}{1 - r - tz + rz^2}.$$

Consequently the correspondence $p_n(t) \mapsto h_n$ gives rise to an algebra homomorphism of $\mathbb{C}[t]$ onto $\mathcal{A}$, which is in fact an isomorphism because $\mathcal{A}$ is infinite-dimensional.

The spectrum of $\lambda(h_1)$ is exactly

$$\sigma_r = [-2\sqrt{r(1-r)}, 2\sqrt{r(1-r)}]$$

(called the *regular spectrum*) without eigenvalues inside ([8; 4]).

Let $E : \mathbb{C}G \to \mathcal{A}$ be the averaging map defined linearly by $E(g) = h_{|g|}$ $(g \in G \subset \mathbb{C}G)$, which satisfies algebraic properties of conditional expectation: $E(a) = a$, $E(f^*) = E(f)^*$ and $E(af) = aE(f)$ for $a \in \mathcal{A}$ and $f \in \mathbb{C}G$. A radial function ϕ is then in one-to-one correspondence with a linear functional φ of $\mathcal{A}$ by the relation $\phi(g) = \varphi \circ E(g)$ $(g \in G)$. The averaging map E is also characterized by the equality $\tau(f\phi) = \tau(E(f)\phi)$ for $f \in \mathbb{C}G$ and a radial function ϕ.

Given a complex number $c \in \mathbb{C}$, let δ_c be the multiplicative linear functional of $\mathcal{A}$ evaluated at c, i.e., $\delta_c(h_n) = p_n(c)$ $(n \geq 0)$. When $\varphi = \delta_c$, the associated radial function ε_c is called a *spherical function* (this being justified by a characterization in [6, Lemma 1.5]) with c referred to as the *spectral parameter*.

A multiplicative functional δ_c of spectral parameter $c \in \mathbb{C}$ is ℓ^1-bounded if and only if $c = a + ib$ $(a, b \in \mathbb{R})$ is in the elliptic disk

$$a^2 + \frac{b^2}{(1-2r)^2} \leq 1$$

(when $r = 1/2$, this shrinks to the interval $[-1, 1] \subset \mathbb{R}$) and δ_c is C*-bounded if and only if $-1 \leq c \leq 1$ ([3; 14]). Notice that C*-bounded multiplicative functionals are automatically *-preserving.

With this spectral property of c in hand, we see that a spherical function ε_c of spectral parameter c is positive definite if and only if $-1 \leq c \leq 1$.

Let A be the closure of $\mathcal{A}$ in the full group C*-algebra $C^*(G)$. Then, by Haagerup (see [20; 22]), E is extended to a conditional expectation (also denoted by E) of $C^*(G)$ onto $A \subset C^*(G)$.

Since E preserves positivity, any positive functional φ of A induces a positive functional $\varphi \circ E$ of $C^*(G)$. If we take an evaluation functional δ_t on A at $t \in [-1, 1]$ as a φ, it gives rise to a spherical state on $C^*(G)$, which is also denoted by ε_t, i.e., $\varepsilon_t = \delta_t \circ E$ and the spectrum of A is identified with $[-1, 1]$, i.e., $A \cong C([-1, 1])$ ($C(K)$ standing for the continuous function algebra of a compact set K).

Remark that the spherical state ε_t is the obvious extension of a positive definite spherical function of spectral parameter $-1 \leq t \leq 1$ with $\epsilon_{\pm 1}(g) = (\pm 1)^{|g|}$ multiplicative on G, i.e., the trivial/parity character (one-dimensional representation) of G.

The GNS-representation of ε_t is called a *spherical representation*.

Theorem 1.1 [6; 10; 20]. *Spherical representations are irreducible and mutually disjoint for different* $t \in [-1, 1]$.

Spherical representations of spectral parameter in σ_r appear as irreducible components of the regular representation in a form of Plancherel formula and are referred to as being *principal*, whereas ones parametrized by $[-1, 1] \setminus \sigma_r$ are said to be *complementary*.

As will be reviewed below, these series of representations (except for the residual values $t = \pm 1$) are realized as an analytic family of representations on a single Hilbert space.

We now introduce another analytic family (λ_z) of representations on $\ell^2(G)$ due to Pytlik and Szwarc. (The original notation is changed to λ_z in view of the fact that this is a deformation of λ.) Here are basic properties: (λ_z) is a family of bounded representations of G on $\ell^2(G)$ parametrized by a complex number z satisfying $z^2 \notin (1, \infty)$.

Theorem 1.2 [16]. (i) *For each* $g \in G$, $\lambda_z(g)$ *is a finite-rank perturbation of* $\lambda(g)$: *there is a finite dimensional subspace* L_g *of* $\ell^2(G)$ *satisfying* $\lambda_z(g) = \lambda(g)$ *on* $L_g^{\perp}$ *for each* z.

(ii) *For any* $g \in G$, $\lambda_z(g)$ *is continuous in* z *and holomorphic on*

$$\left\{z \in \mathbb{C} \setminus ((-\infty, -1] \sqcup [1, \infty))\right\}.$$

(iii) *The equality* $\lambda_z(g)^* = \lambda_{\bar{z}}(g^{-1})$ *holds for any* $g \in G$ *and any* z. *Consequently* λ_z *is unitary if* z *is real, i.e., if* $z \in [-1, 1]$.

(iv) *For* $z^2 \notin [1, \infty)$, $\tau^{1/2}$ *is a cyclic vector of* λ_z *and satisfies*

$$(\tau^{1/2} | \lambda_z(g) \tau^{1/2}) = z^{|g|} \quad (g \in G).$$

In particular, the function $z^{|g|}$ *is positive definite for* $-1 \le z \le 1$ (*see* [7]) *and* $\lambda_0 = \lambda$.

(v) *Limits* $\lambda_{\pm 1} = \lim_{z \to \pm 1} \lambda_z$ *are unitary representations of* G *on* $\ell^2(G)$ *satisfying*

$$\lambda_{\pm 1}(a) g \tau^{1/2} = \begin{cases} \pm \tau^{1/2} & (g = e), \\ \mp a \tau^{1/2} & (g = a^{-1}), \\ ag\tau^{1/2} & \text{otherwise}, \end{cases}$$

for $a \in G_1$ *and* $g \in G$.

Corollary 1.3. *According to an orthogonal decomposition* $\ell^2(G) = \mathbb{C}\tau^{1/2} \oplus \ell^2(G(1)) \oplus \cdots \oplus \ell^2(G(l))$, *where* $G(i)$ *consists of words whose right ends are in* $\{s_i, s_i^{-1}\}$, $\lambda_{\pm 1}$ *is unitarily equivalent to a direct sum of* λ *by multiplicity* l *and the parity/trivial character of* G.

Proof. Let $G^* = G \setminus \{e\}$. Since $\lambda_{-1}(g)(G^* \tau^{1/2}) = G^* \tau^{1/2}$ for $g \in G$ by (v), $\lambda_{-1}(g)$ induces a free action of G on G^* with $G^* = G(1) \sqcup \cdots \sqcup G(l)$ the decomposition of G^* into orbits. Thus λ_{-1} on $\ell^2(G^*)$ is decomposed into the direct sum of subrepresentations on $\ell^2(G(i))$ $(1 \le i \le l)$ so that the restriction of λ_{-1} on $\ell^2(G(i))$ is unitarily equivalent to the regular representation of G.

Likewise λ_1 and the parity operator P leave $\ell^2(G(i))$ invariant so that $P\lambda_1(x)P = \lambda_{-1}(\varpi(x))$ $(x \in C^*(G))$. From the decomposition of λ_{-1}, this implies that λ_1 on $\ell^2(G(i))$ is unitarily equivalent to $\lambda \circ \varpi = P\lambda P \cong \lambda$ as well. □

Remark 1. The above corollary is taken from [16] 2.4 Remark (3), where it is pointed out that λ_{-1} on $\ell^2(G^*)$ is considered by J. Cuntz to illustrate K-amenability of free groups.

The following supplements to [16] are extracted from [19; 20] with the case of critical values $\upsilon = \pm\sqrt{r/(1-r)}$ added in [22, Corollary 4.4].

Theorem 1.4. *Let* λ_υ *be a unitary representation, i.e.,* $\upsilon \in [-1, 1]$.

(i) *If* $|\upsilon| \le \sqrt{r/(1-r)}$, λ_υ *is unitarily equivalent to* λ.

(ii) *If* $\sqrt{r/(1-r)} < |\upsilon| < 1$, λ_υ *is unitarily equivalent to a direct sum of* λ *and the spherical representation of spectral parameter*

$$c_r(\upsilon) = \frac{r}{\upsilon} + (1-r)\upsilon$$

with the spectrum of $\lambda_\upsilon(h_1)$ *equal to* $\sigma_r \sqcup \{c_r(\upsilon)\}$.

Given a positive functional φ of A, the GNS representation space

$$\mathcal{R}_\varphi \equiv \overline{C^*(G)(\varphi \circ E)^{1/2}}$$

($(\varphi \circ E)^{1/2}$ being the GNS-cyclic vector) turns out to be a $C^*(G)$-A-bimodule: The left action of $C^*(G)$ is just the (left) GNS-representation based on $\varphi \circ E$ and the right action of A is given by $(x(\varphi \circ E)^{1/2})a = (xa)(\varphi \circ E)^{1/2}$ ($x \in C^*(G)$, $a \in A$).

If ψ is another positive functional of A, the space $\mathrm{Hom}(\mathcal{R}_\varphi, \mathcal{R}_\psi)$ of intertwiners is naturally isomorphic to $\mathrm{Hom}(L^2(\varphi), L^2(\psi))$ ([22]). Here $L^2(\varphi)$ denotes the GNS representation space of φ and $\mathrm{Hom}(L^2(\varphi), L^2(\psi))$ denotes the space of intertwiners between A-modules $L^2(\varphi)$ and $L^2(\psi)$.

In view of $L^2(\varphi) = L^2(A)[\varphi]$ with $L^2(A)$ the standard space of the second dual W*-algebra A^{**} and $[\varphi]$ the support projection of φ in A^{**}, we have

$$\mathrm{Hom}(L^2(\varphi), L^2(\psi)) \cong A^{**}[\varphi][\psi].$$

Thus $\mathrm{Hom}(L^2(\varphi), L^2(\psi))$ is isomorphic to the L^∞-space on $[-1, 1]$ with respect to the common measure class of φ and ψ in the Lebesgue decomposition.

When $\varphi = \delta_t$ with $\varepsilon_t = \varphi \circ E$ a spherical state on $C^*(G)$, $\mathcal{R}_\varphi$ is simply denoted by $\mathcal{R}_t$. As observed in [22], $(\mathcal{R}_t)_{-1 \le t \le 1}$ is a continuous family of $C^*(G)$-A-bimodules and therefore it provides a Borel field structure so that the following holds.

Theorem 1.5 (Plancherel Formula). *Under the identification of φ with the associated Radon measure $\varphi(dt)$ on the spectrum $[-1, 1]$ of A, we have a natural isometric isomorphism between $\mathcal{R}_\varphi$ and the direct integral $\oint_{[-1,1]} \mathcal{R}_t \sqrt{\varphi(dt)}$ specified by the correspondence*

$$\mathcal{R}_\varphi \ni (\varphi \circ E)^{1/2} \mapsto \oint_{[-1,1]} \varepsilon_t^{1/2} \sqrt{\varphi(dt)} \in \oint_{[-1,1]} \mathcal{R}_t \sqrt{\varphi(dt)}.$$

For the positive definite function $\phi_\upsilon = \varphi_\upsilon \circ E$ ($-1 \le \upsilon \le 1$) of Haagerup, where $\varphi_\upsilon(h_n) = \upsilon^n$ ($n \ge 0$), the accompanied Radon measure $\varphi_\upsilon(dt)$ takes the following form ([22]):

(i) If $|\upsilon| \le \sqrt{r/(1-r)}$, $\varphi_\upsilon(dt)$ is supported by the interval σ_r and of the form

$$\varphi_\upsilon(dt) = \frac{\upsilon^{-1} - \upsilon}{2\pi} \frac{\sqrt{4r(1-r) - t^2}}{(1-t^2)(c_r(\upsilon) - t)}\, dt.$$

(ii) If $\sqrt{r/(1-r)} < |\upsilon| \le 1$, adding to the continuous measure in (i), there appears an atomic measure of the form

$$\frac{1 - c_r(\upsilon^2)}{1 - c_r(\upsilon)^2} \delta(t - c_r(\upsilon)).$$

In the special case $\upsilon = 0$ for which $\phi_0 = \tau$, $\varphi_0(dt)$ is reduced to the Kesten measure

$$\varphi_0(dt) = \lim_{\upsilon \to 0} \varphi_\upsilon(dt) = \frac{1}{2\pi r} \frac{\sqrt{4r(1-r) - t^2}}{1 - t^2}\, dt$$

on $\sigma_r = [-2\sqrt{r(1-r)}, 2\sqrt{r(1-r)}]$ as stated in [3] (see also [18; 14]).

Let π be the GNS-representation of the spherical state ε_s for a spectral parameter $s = 2\sqrt{r(1-r)}$ and $\mathcal{H} = \overline{C^*(G)\varsigma}$ be the representation space of π with the GNS-vector denoted by $\varsigma = \varepsilon_s^{1/2}$. Notice that the spectral parameter $s = 2\sqrt{r(1-r)}$ is critical in the sense that it is located at the boundary of principal and complementary series. It is also critical from the viewpoint of C*-completion relative to $\ell^p(G)$ (see [2; 11]).

As a deformation of π, an analytic family (π_c) of bounded representations on $\mathcal{H}$ is constructed in [15] for $c \in \mathbb{C}$ satisfying $c^2 \notin (1, \infty)$ in such a way that $\pi_s = \pi$ and the following properties hold.

(i) For each $g \in G$, there is a finite dimensional subspace $\mathcal{H}_g$ of $\mathcal{H}$ so that $\pi_c(g) = \pi(g)$ on $\mathcal{H}_g^\perp$ for any c, i.e., $\pi_c(g) - \pi(g)$ is a finite-rank operator.

(ii) For any $g \in G$, $\pi_c(g)$ is continuous in c and holomorphic on

$$\big\{c \in \mathbb{C} \setminus ((-\infty, -1] \sqcup [1, \infty))\big\}.$$

(iii) We have $\pi_c(g)^* = \pi_{\bar{c}}(g^{-1})$ for $g \in G$ and any parameter c. Consequently π_c is unitary if $c \in [-1, 1]$.

(iv) If $c \neq \pm 1$, ς is a cyclic vector of π_c and satisfies $(\varsigma | \pi_c(g)\varsigma) = \varepsilon_c(g) = \delta_c(h_{|g|})$ $(g \in G)$.

(v) For $c = \pm 1$, we have $\pi_{\pm 1}(g)\varsigma = \varepsilon_{\pm 1}(g)\varsigma$ $(g \in G)$ but ς is not cyclic.

From property (iv), we see $(\varsigma | (\pi_c(h_1) - c)^2 \varsigma) = 0$, i.e., $\pi_c(h_1)\varsigma = c\varsigma$ for $c \in (-1, 1)$, and then, by analytic continuation on c, $\pi_c(h_1)\varsigma = c\varsigma$ for any c. Thus ς is an eigenvector of $\pi_c(h_1)$ of eigenvalue c.

More is known on the spectrum of $\pi_c(h_1)$:

Theorem 1.6 [19, Theorem 5; 15, Lemma 3]. *Any spectral parameter $c \in \mathbb{C}$ in the closed elliptic disk of ℓ^1-boundedness is an eigenvalue of $\pi_c(h_1)$ with ς its eigenvector so that $\pi_c(h_1)\varsigma^\perp \subset \varsigma^\perp$ and the spectrum of $\pi_c(h_1)$ as a bounded operator on the reducing subspace $\varsigma^\perp \subset \mathcal{H}$ is contained in the regular spectrum $\sigma_r = [-2\sqrt{r(1-r)}, 2\sqrt{r(1-r)}]$.*

Remark 2. A common-space-realization Π_z of the complementary component of λ_z in [19] turns out to be extended to the region $z^2 \notin [1, \infty)$ by an analytic continuation, which seems to be globally similar to (π_z) in [15].

2. The primitive ideal space

We begin with a universal construction of radial bimodules. For two positive functionals φ and ψ of A satisfying $\varphi \leq \psi$, the spherical decomposition in the Plancherel formula enables us to define an isometric embedding by

$$\oint_{[-1,1]} x\sqrt{\varphi(dt)} \mapsto \oint_{[-1,1]} x\sqrt{\frac{\varphi(dt)}{\psi(dt)}}\sqrt{\psi(dt)}$$

($\varphi(dt)/\psi(dt)$ being the Radon–Nikodym derivative), which in turn is converted to an isometric embedding of $\mathcal{R}_\varphi$ into $\mathcal{R}_\psi$ thanks to the identity $(\varphi \circ E)^{1/2} = \sqrt{d\varphi/d\psi}(\psi \circ E)^{1/2}$ in the standard representation space $L^2(C^*(G))$ of the second dual $C^*(G)^{**}$ (see [21]).

By the way of this construction, embeddings $\mathcal{R}_\varphi \to \mathcal{R}_\psi$ give an inductive system of $C^*(G)$-A-bimodules and we obtain the universal $C^*(G)$-A-bimodule $\mathcal{R}$ as an inductive limit. The image of the left action of $C^*(G)$ on $\mathcal{R}$ is then called the radial C*-algebra of G and denoted by $C^*_{\mathrm{rad}}(G)$, which is minimally universal with respect to radial representations in the sense that it includes all $\mathcal{R}_\varphi$ and satisfies the property stated in Proposition 2.1 below. Our main concern is in describing the primitive ideal space of $C^*_{\mathrm{rad}}(G)$.

Note that $C^*_{\mathrm{rad}}(G)$ itself is realized without minimality by a direct sum representation on $\bigoplus_\varphi \mathcal{R}_\varphi$. Remark also that we have a series of natural *-homomorphisms

$$\mathbb{C}G \to C^*(G) \to C^*_{\mathrm{rad}}(G) \to C^*_r(G)$$

with the total composition $\mathbb{C}G \to C^*_r(G)$ being injective because so is the regular representation of G. As a result, the obvious *-homomorphism $\mathbb{C}G \to C^*_{\mathrm{rad}}(G)$ is injective as well.

In terms of this universal radial bimodule, irreducibility of spherical representations is now rephrased in the following manner:

Proposition 2.1. *The von Neumann algebra* $\mathrm{End}({}_{C^*(G)}\mathcal{R}_A)$ *of self-intertwiners is isomorphic to* A^{**} *through the right multiplication of* A^{**}, *i.e.*, $\mathrm{End}({}_{C^*(G)}\mathcal{R}_A)$ *is generated by the right action of* A *and isomorphic to the second dual W*-algebra* A^{**} *associated with* A.

Here ${}_{C^*(G)}\mathcal{R}_A$ *indicates that* $\mathcal{R}$ *is a* $C^*(G)$-A-*bimodule and a bounded linear operator* T *on* $\mathcal{R}$ *is called a self-intertwiner if* $T(x\xi a) = xT(\xi)a$ *for* $x \in C^*(G)$, $a \in A$ *and* $\xi \in \mathcal{R}$.

Proof. By [22], the space $\mathrm{End}(\mathcal{R}_\varphi)$ of self-intertwiners is generated by the right action of A and $\mathcal{R}$ contains $L^2(A)$ as an A-A-subbimodule, whence the problem is reduced to showing that $\mathrm{End}(\mathcal{R})$ is generated by the right action of A as a von Neumann algebra.

To see this, observe first that the projection e_φ to the subspace $\mathcal{R}_\varphi$ belongs to $\mathrm{End}(_{C^*(G)}\mathcal{R}_A)$ and satisfies $\lim_{\varphi\uparrow\infty} e_\varphi = 1_{\mathcal{R}}$ in the σ-strong operator topology. Let $T \in \mathrm{End}(_{C^*(G)}\mathcal{R}_A)$. If its reduction $e_\varphi T e_\varphi \in \mathrm{End}(_{C^*(G)}\mathcal{H}_A)$ commutes with $e_\varphi \mathrm{End}(\mathcal{R}_A)e_\varphi$ for any $\varphi \in A_+^*$, $a' \in \mathrm{End}(\mathcal{R}_A)$ and $\xi, \eta \in \mathcal{R}_\psi$ with $\psi \in A_+^*$ satisfy

$$(\xi|Te_\varphi a'\eta) = (\xi|e_\varphi Te_\varphi a'e_\varphi\eta) = (\xi|e_\varphi a'e_\varphi Te_\varphi\eta) = (\xi|a'e_\varphi T\eta) \quad (\varphi \geq \psi)$$

and then we have $(\xi|Ta'\eta) = (\xi|a'T\eta)$ by taking limit $\varphi \uparrow \infty$. Since the inductive (algebraic) limit of $C^*(G)(\psi\circ E)^{1/2}$ for $\psi \uparrow \infty$ is dense in $\mathcal{R}$, this implies $Ta' = a'T$, i.e., $T \in \mathrm{End}(\mathcal{R}_A)'$. As a final step, apply the double commutant theorem of von Neumann. □

Corollary 2.2. *For a positive functional φ of A, $\mathrm{End}(\mathcal{R}_\varphi) \cong A^{**}[\varphi]$, where $[\varphi]$ denotes the support projection of φ in A^{**}.*

We use the notation $\mathcal{C}(\mathcal{H})$ to stand for the compact operator algebra on a Hilbert space $\mathcal{H}$.

Proposition 2.3. *For a continuous function $f \in C([-1, 1])$ vanishing on $\sigma_r \subset [-1, 1]$, we have $\pi_t(f(h_1)) = f(t)|\varsigma)(\varsigma|$ $(-1 \leq t \leq 1)$ and the representation π_t for $t \in (-1, 1) \setminus \sigma_r$ is well-behaved in the sense that $\mathcal{C}(\mathcal{H}) \subset \pi_t(C^*(G))$.*

Proof. By the spectral property in Theorem 1.6, $\pi_t(f(h_1)) = f(t)|\varsigma)(\varsigma| = 0$ if $t \in \sigma_r$.

For $t \in [-1, 1] \setminus \sigma_r$, the spectrum of $\pi_t(f(h_1))$ is supported by the point t with $|\varsigma)$ a unique eigenvector and hence $\pi_t(f(h_1)) = f(t)|\varsigma)(\varsigma|$ by functional calculus.

Thus $\pi_t(C^*(G))$ contains a rank-one projection $|\varsigma)(\varsigma|$ and the irreducibility of π_t for $-1 < t < 1$ implies $\mathcal{C}(\mathcal{H}) \subset \pi_t(C^*(G))$. □

Remark 3. Properties of complementary series representations concerning compact operator algebras are also pointed out in [20].

Proposition 2.4. *Primitive ideals of $C^*(G)$ associated to pure states ε_t are different for $t \in [-1, 1] \setminus \sigma_r$ and coincide with the kernel of the regular representation λ for $t \in \sigma_r$.*

Proof. Clearly characters $\varepsilon_{\pm 1}$ give rise to ideals different from those for ε_t $(-1 < t < 1)$ and we focus on irreducible representations π_t $(-1 < t < 1)$.

If two primitive ideals coincide for spectral parameters in $(-1, 1)$, the associated C*-algebras $\pi_t(C^*(G))$ for various $t \in (-1, 1)$ are canonically isomorphic and give rise to the same spectrum of $\pi_t(h_1)$. In the case when one of t is outside of σ_r, this necessitates the coincidence of t by Theorem 1.6.

Let $t \in \sigma_r$. If $x \in C^*(G)$ is in the kernel of λ, then the Plancherel formula ensures $\pi_{t'}(x^*x) = 0$ for almost any $t' \in \sigma_r$ with respect to the Kesten measure (being

equivalent to the Lebesgue measure on σ_r) and we can find a sequence $t_n \in \sigma_r$ converging to t so that $\pi_{t_n}(x^*x) = 0$. Since $\varepsilon_{t'}$ is weak*-continuous in t', we have

$$\varepsilon_t(y^*x^*xy) = \lim_{n\to\infty} \varepsilon_{t_n}(y^*x^*xy) = 0$$

for any $y \in C^*(G)$. Thus, by cyclicity of ς, we have $\pi_t(x^*x) = 0$, i.e., $\ker\lambda \subset \ker\pi_t$ with $\lambda(\ker\pi_t)$ a closed ideal of a simple C*-algebra $C_r^*(G)$, which proves that $\ker\pi_t = \ker\lambda$. □

Corollary 2.5. (i) *We have* $\pi_t(C^*(G)) \cap \mathcal{C}(\mathcal{H}) = \{0\}$ *for* $t \in \sigma_r$.

(ii) *For* $f \in C([-1, 1])$, *the condition* $f(h_1) \in \ker\pi$ *is equivalent to* $f|_{\sigma_r} = 0$.

Proof. (i) If $\pi_t(C^*(G)) \cap \mathcal{C}(\mathcal{H}) \neq 0$, $\mathcal{C}(\mathcal{H}) \subset \pi_t(C^*(G))$ by irreducibility of π_t. In view of $\ker\pi_{t'} = \ker\pi_t$ for $t' \in \sigma_r$, this implies $\pi_{t'} \cong \pi_t$ ([5] Corollary 4.1.10), which contradicts with the disjointness of π_t and $\pi_{t'}$ for $t \neq t'$ in σ_r.

(ii) From the kernel coincidence, the spectrum of $\pi(h_1)$ is

$$\sigma_{C^*(G)/\ker\pi}(h_1 + \ker\pi) = \sigma_{C^*(G)/\ker\lambda}(h_1 + \ker\lambda) = \sigma(\lambda(h_1)) = \sigma_r,$$

whence $\pi(f(h_1)) = f(\pi(h_1)) = 0$ if and only if $f|_{\sigma_r} = 0$. □

Proposition 2.6. *Let* $-1 \le t \le 1$. *In the orthogonal decomposition* $\pi_t(h_1) = t|\varsigma)(\varsigma| \oplus \pi_t(h_1)|_{\varsigma^\perp}$, *the spectrum of* $\pi_t(h_1)|_{\varsigma^\perp}$ *is* σ_r *and therefore the spectrum of* $\pi_t(h_1)$ *is* $\{t\} \cup \sigma_r$.

Proof. Let σ (σ') be the (essential) spectrum of $\pi_t(h_1)$ on $\mathcal{H}$, i.e., σ' consists of accumulation points of σ or eigenvalues of infinite multiplicity of $\pi_t(h_1)$. Since $\pi_t(h_1)$ for various t coincide up to finite-rank operators, σ' does not depend on $t \in [-1, 1]$ (Weyl's stability theorem, see [17] XIII.4 for example).

By Proposition 2.4, $\ker\pi_t = \ker\lambda$ for $t \in \sigma_r$ and the spectrum of $\pi_t(h_1)$ in $\pi_t(C^*(G))$ is equal to that of $\lambda(h_1)$ in $C_r^*(G)$. Consequently, if $t \in \sigma_r$, $\sigma = \sigma_r$ and hence $\sigma' = \sigma_r$.

Now $\sigma(\pi_t(h_1)|_{\varsigma^\perp}) \subset \sigma_r$ (Theorem 1.6) is combined with $\sigma_r = \sigma' \subset \sigma(\pi_t(h_1)|_{\varsigma^\perp})$ to conclude that $\sigma(\pi_t(h_1)|_{\varsigma^\perp}) = \sigma_r$. □

Proposition 2.7. *The continuous family* π_t ($-1 \le t \le 1$) *of representations gives rise to the same *-homomorphism* $C^*(G) \to \mathcal{B}(\mathcal{H})/\mathcal{C}(\mathcal{H})$ *by taking the quotient to the Calkin algebra* $\mathcal{B}(\mathcal{H})/\mathcal{C}(\mathcal{H})$.

In particular, we have $\pi_t(\ker\pi) \subset \mathcal{C}(\mathcal{H})$ *for* $-1 \le t \le 1$. *Recall that* $\pi = \pi_s$ *with* $s = 2\sqrt{r(1-r)}$ *and* $\ker\pi_t = \ker\pi$ *for* $t \in \sigma_r$.

Proof. Since $\pi_t(g) - \pi(g)$ is a finite-rank operator for each $g \in G$, the same holds for $\pi_t(x) - \pi(x)$ if $x \in \mathbb{C}G$. Since $x \in C^*(G)$ is norm-approximated by a sequence x_n in $\mathbb{C}G$ and $\pi_t : C^*(G) \to \mathcal{B}(\mathcal{H})$ is contractive in norm, $\lim_{n\to\infty} \|\pi_t(x_n) - \pi_t(x)\| = 0$ shows that $\pi_t(x) - \pi(x)$ is norm-approximated by a sequence $(\pi_t(x_n) - \pi(x_n))_{n\ge1}$ of finite-rank operators. Thus $\pi_t(x) - \pi(x)$ is a compact operator on $\mathcal{H}$. □

Corollary 2.8. *For $-1 \le t \le 1$, we have $\ker\pi = \pi_t^{-1}(\mathcal{C}(\mathcal{H})) \supset \ker\pi_t$ and hence $\pi_t(\ker\pi) = \pi_t(C^*(G)) \cap \mathcal{C}(\mathcal{H})$.*

Proof. Since $\pi(C^*(G)) \cap \mathcal{C}(\mathcal{H}) = 0$ (Corollary 2.5) and

$$\pi_t(C^*(G)) + \mathcal{C}(\mathcal{H}) = \pi(C^*(G)) + \mathcal{C}(\mathcal{H}),$$

we see

$$\begin{aligned} C^*(G)/\pi_t^{-1}(\mathcal{C}(\mathcal{H})) &\cong \pi_t(C^*(G))/\big(\pi_t(C^*(G)) \cap \mathcal{C}(\mathcal{H})\big) \\ &\cong \big(\pi_t(C^*(G)) + \mathcal{C}(\mathcal{H})\big)/\mathcal{C}(\mathcal{H}) \\ &\cong \pi(C^*(G)) \cong C^*(G)/\ker\pi, \end{aligned}$$

which means that $\ker\pi = \pi_t^{-1}(\mathcal{C}(\mathcal{H})) \supset \ker\pi_t$. □

By the continuity of $\pi_t(g)$ in $t \in [-1, 1]$ for each $g \in G$, a norm-continuous family $(\pi_t(x))_{-1\le t\le 1}$ of operators in $\mathcal{B}(\mathcal{H})$ is associated to any $x \in C^*(G)$ as a uniform limit of continuous functions and the correspondence $x \mapsto (\pi_t(x))$ defines a *-homomorphism Π of $C^*(G)$ into $C([-1, 1]) \otimes \mathcal{B}(\mathcal{H})$ so that

$$\ker\Pi = \bigcap_{-1\le t\le 1} \ker\pi_t = \bigcap_{t\in(-1,1)\setminus\sigma_r} \ker\pi_t$$

and $\Pi(\ker\pi) \subset C([-1, 1]) \otimes \mathcal{C}(\mathcal{H})$ by Corollary 2.8.

In view of $\ker\pi_t = \ker\pi$ $(t \in \sigma_r)$, $\Pi(\ker\pi)$ is in fact included in the ideal $C_{\sigma_r}([-1, 1]) \otimes \mathcal{C}(\mathcal{H})$ of $C([-1, 1]) \otimes \mathcal{C}(\mathcal{H})$, where

$$C_{\sigma_r}([-1, 1]) = \{f \in C([-1, 1]);\ f|_{\sigma_r} = 0\} \cong C_0([-1, 1] \setminus \sigma_r)$$

is the ideal of $C([-1, 1])$ vanishing on $\sigma_r \subset [-1, 1]$. Recall also that π_t is irreducible for $-1 < t < 1$, whereas $\pi_{\pm1}$ is decomposed as $\varepsilon_{\pm1} \oplus \pi'_\pm$ according to the orthogonal decomposition $\mathcal{H} = \mathbb{C}\varsigma \oplus \varsigma^\perp$.

Choose $f \in C_{\sigma_r}([-1, 1])$ so that $f(\pm1) = 1$. By Theorem 1.6 or Proposition 2.6, $f(h_1) \in A$ satisfies $\pi_{\pm1}(f(h_1)) = |\varsigma)(\varsigma|$ and then

$$\pi_{\pm1}\big(zf(h_1) + x(1 - f(h_1))\big) = z|\varsigma)(\varsigma| + \pi_{\pm1}(x)(1 - |\varsigma)(\varsigma|) = z|\varsigma)(\varsigma| + \pi'_\pm(x)$$

$(z \in \mathbb{C}, x \in C^*(G))$ reveals that $\pi_{\pm1}(C^*(G)) = \mathbb{C}|\varsigma)(\varsigma| \oplus \pi'_\pm(C^*(G))$.

Consequently, Corollary 2.8 is used again to see

$$\begin{aligned} \ker\pi &= \pi_{\pm1}^{-1}(\mathcal{C}(\mathcal{H})) = \pi_{\pm1}^{-1}\big(\mathbb{C}|\varsigma)(\varsigma| \oplus \pi'_\pm(C^*(G)) \cap \mathcal{C}(\varsigma^\perp)\big) \\ &= (\pi'_\pm)^{-1}(\mathcal{C}(\varsigma^\perp)) \supset \ker\pi'_\pm \end{aligned}$$

and hence $\ker\pi/\ker\pi'_\pm \cong \pi'_\pm(\ker\pi) = \pi'_\pm(C^*(G)) \cap \mathcal{C}(\varsigma^\perp)$.

Define the residual representation of $C^*(G)$ on $\varsigma^\perp \oplus \varsigma^\perp$ to be a direct sum $\pi' = \pi'_+ \oplus \pi'_-$, which satisfies $\ker\pi' = \ker\pi'_+ \cap \ker\pi'_- \subset \ker\pi$, $\pi'(\ker\pi) \subset$

$\pi'(C^*(G)) \cap \big(\mathcal{C}(\varsigma^\perp) \oplus \mathcal{C}(\varsigma^\perp)\big)$ and $\Pi(\ker \pi') \subset C$ with

$$C = \{f \in C_{\sigma_r}([-1,1]) \otimes \mathcal{C}(\mathcal{H});\ f(\pm 1) \in \mathbb{C}|\varsigma)(\varsigma|\}$$

a C*-subalgebra of $C_{\sigma_r}([-1,1]) \otimes \mathcal{C}(\mathcal{H})$.

Proposition 2.9. *We have the equality* $\Pi(\ker \pi') = C$ *and the associated isomorphism* $C \cong \ker \pi' / \ker \Pi$ *induces a bijection between* $\mathrm{Prim}(C)$ *and*

$$\mathrm{Prim}(\ker \pi') \setminus \mathrm{Prim}(\ker \Pi) \subset \mathrm{Prim}(C^*(G))$$

so that $\mathrm{Prim}(C)$ *is the set of primitive ideals of spherical representations in the complementary series, i.e.,*

$$\{\ker \pi_t;\ t \in (-1,1) \setminus \sigma_r\} \sqcup \{\ker \varepsilon_{\pm 1}\}.$$

Proof. We first describe pure states of C. Let $C_0 = \{f \in C;\ f(\pm 1) = 0\}$. Since a closed ideal of C_0 is isomorphic to $C_0((-1,1) \setminus \sigma_r) \otimes \mathcal{C}(\mathcal{H})$ and satisfies $C/C_0 \cong \mathbb{C} \oplus \mathbb{C}$, irreducible representations of C are given by evaluation at $t \in [-1,1] \setminus \sigma_r$ and hence pure states of C are evaluations followed by applying pure states of fibers ($\mathcal{C}(\mathcal{H})$ or $\mathbb{C}|\varsigma)(\varsigma|$). In particular, C is a CCR (completely continuous representation) algebra.

The equality then follows from the Stone–Weierstrass theorem on CCR C*-algebras due to Kaplansky (see [1] for a survey) if these pure states are separated by elements in $\Pi(\ker \pi')$.

To see this, we recall Proposition 2.3 that $\pi_t(f(h_1)) = f(t)|\varsigma)(\varsigma|$ holds for $f \in C_{\sigma_r}([-1,1])$ and $t \in [-1,1]$, whence $f(h_1) \in \ker \pi'$.

Let φ and ψ be pure states of C given by $\varphi(x) = (\xi|x(s)\xi)$ and $\psi(x) = (\eta|x(t)\eta)$, where $s, t \in [-1,1] \setminus \sigma_r$ and ξ, η are unit vectors in $\mathcal{H}$.

If $s \neq t$ with $t \in (-1,1)$, π_t is irreducible and we can find $y \in C^*(G)$ such that $\pi_t(y) = |\eta)(\varsigma|$ by Kadison's transitivity[1]. Then $yf(h_1)y^* \in \ker \pi'$ is represented by

$$\pi_s(yf(h_1)y^*) = f(s)|\pi_s(y)\varsigma)(\pi_s(y)\varsigma|, \quad \pi_t(yf(h_1)y^*) = f(t)|\eta)(\eta|.$$

Thus, for a choice $f \in C_{\sigma_r}([-1,1])$ satisfying $f(s) = 0$ and $f(t) = 1$, φ and ψ are separated by $\Pi(yf(h_1)y^*)$:

$$\varphi(\Pi(yf(h_1)y^*)) = (\xi|\pi_s(yf(h_1)y^*)\xi) = f(s)(\xi|\pi_s(y)\varsigma)(\pi_s(y)\varsigma|\xi) = 0,$$
$$\psi(\Pi(yf(h_1)y^*)) = (\eta|\pi_t(yf(h_1)y^*)\eta) = f(t)(\eta|\pi_t(y)\varsigma)(\pi_t(y)\varsigma|\eta) = 1.$$

When both $s \neq t$ come from $\{\pm 1\}$, we may assume that $s = -1$ and $t = 1$, with $\varphi(x) = (\varsigma|x(-1)\varsigma)$ and $\psi(x) = (\varsigma|x(1)\varsigma)$ for $x = (x(t))_{-1 \le t \le 1} \in C$. Then

[1]We can dispense with Kadison's transitivity if η is approximated by $\pi_t(y)\varsigma$.

$\varphi(\Pi(f(h_1))) = f(-1)$ and $\psi(\Pi(f(h_1))) = f(1)$ and we see that φ and ψ are separated by $f(h_1) \in \ker \pi'$ if $f \in C_{\sigma_r}([-1, 1])$ satisfies $f(-1) \neq f(1)$.

Finally consider the case $s = t$. Since states are unique for $t = \pm 1$, we assume $-1 < t < 1$ with $\varphi(x) = (\xi | x(t)\xi)$ and $\psi(x) = (\eta | x(t)\eta)$. Then the condition $\varphi \neq \psi$ is equivalent to $|(\xi|\eta)| < 1$ and $yf(h_1)y^* \in \ker \pi'$ described above is evaluated by

$$\varphi(\Pi(yf(h_1)y^*)) = f(t)(\xi|\eta)(\eta|\xi), \quad \psi(\Pi(yf(h_1)y^*)) = f(t).$$

Thus φ and ψ are separated by $yf(h_1)y^* \in \ker \pi'$ if $f(t) \neq 0$. □

Lemma 2.10. *We have* $\ker \pi = \ker \pi'$.

Proof. To see this, we use the continuous deformation $(\lambda_\upsilon)_{-1 \le \upsilon \le 1}$ of the regular representation $\lambda = \lambda_0$ in [16]. Recall that these are unitary representations of G with the following properties (Theorem 1.4):

(i) λ_υ is unitarily equivalent to λ if $\upsilon \in \sigma_r$, to $\pi_\upsilon \oplus \lambda$ if $\upsilon \in (-1, 1) \setminus \sigma_r$ and to $\epsilon_{\pm 1} \oplus \overbrace{\lambda \oplus \cdots \oplus \lambda}^{l \text{ times}}$ if $\upsilon = \pm 1$.

(ii) $\ker \lambda_\upsilon = \ker \pi_{c_r(\upsilon)} \cap \ker \lambda = \ker \pi_{c_r(\upsilon)} \cap \ker \pi = \ker \pi_{c_r(\upsilon)}$ if $|\upsilon| > \sqrt{r/(1-r)}$, $\ker \lambda_\upsilon = \ker \lambda$ if $|\upsilon| \le \sqrt{r/(1-r)}$, and finally $\ker \lambda_{\pm 1} = \ker \epsilon_{\pm 1} \cap \ker \lambda = \ker \epsilon_{\pm 1} \cap \ker \pi$. Notice that $c_r(\upsilon) \notin \sigma_r$ for $\upsilon \neq \pm\sqrt{r/(1-r)}$.

($\pi'_\pm$ would be unitarily equivalent to $\lambda \otimes 1_{\mathbb{C}^{l-1}}$ but its validity is irrelevant here.)

As observed for (π_t) before, the family $(\lambda_\upsilon)_{-1 \le \upsilon \le 1}$ of *-representations of $C^*(G)$ satisfies $\lambda_\upsilon(x) - \lambda(x) \in \mathcal{C}(\ell^2(G))$ for $x \in C^*(G)$ and gives rise to a *-homomorphism $\Lambda : C^*(G) \to C([-1, 1]) \otimes \mathcal{B}(\ell^2(G))$ in such a way that $\Lambda(\ker \lambda) = \Lambda(\ker \pi) \subset C([-1, 1]) \otimes \mathcal{C}(\ell^2(G))$. In view of norm-continuity of $\lambda_\upsilon(x)$ in $\upsilon \in [-1, 1]$ and Theorem 1.4,

$$\ker \Lambda = \bigcap_{\sqrt{r/(1-r)} < |\upsilon| < 1} \ker \lambda_\upsilon = \bigcap_{4r(1-r) < t^2 < 1} \ker \pi_t = \ker \Pi$$

and $\Lambda(\ker \lambda) = \Lambda(\ker \pi) \subset C([-1, 1]) \otimes \mathcal{C}(\ell^2(G))$. Note here that

$$\{c_r(\upsilon); \sqrt{r/(1-r)} < |\upsilon| < 1\} = \{t \in \mathbb{R}; 4r(1-r) < t^2 < 1\}.$$

Thus pure states of $\ker \pi / \ker \Pi = \ker \lambda / \ker \Lambda \cong \Lambda(\ker \lambda)$ are given by pure states of $\lambda_\upsilon(\ker \lambda)$ after the evaluation at υ satisfying $|\upsilon| > \sqrt{r/(1-r)}$. Here recall that

$$\lambda_\upsilon(\ker \lambda) \cong \begin{cases} 0 & (|\upsilon| \le \sqrt{r/(1-r)}), \\ \pi_{c_r(\upsilon)}(\ker \pi) & (\sqrt{r/(1-r)} < |\upsilon| < 1), \\ \mathbb{C} & (\upsilon = \pm 1). \end{cases}$$

Hence the set of associated primitive ideals of $\ker \lambda / \ker \Lambda = \ker \pi / \ker \Pi$ is

$$\{\ker \pi_t; t \in (-1, 1) \setminus \sigma_r\} \cup \{\ker \epsilon_{\pm 1}\}.$$

At this point, there might be overlapping in the union but the comparison of this with Proposition 2.9 enables us to conclude that these are in fact distinct and $\ker\pi/\ker\pi' = 0$. □

Remark 4. We have an inclusion $\overline{C^*(G)f(h_1)C^*(G)} \subset \ker\pi$ for $f \in C_{\sigma_r}([-1,1])$ and the equality $\ker\pi' = \ker\pi$ is rephrased by

$$\ker\pi = \overline{\bigcup_{f\uparrow[-1,1]\setminus\sigma_r} C^*(G)f(h_1)C^*(G)}.$$

Here $f \uparrow [-1,1]\setminus\sigma_r$ means an inductive limit on $f \in C_{\sigma_r}([-1,1])$ satisfying $0 \le f \le 1$.

We summarize our considerations so far:

Theorem 2.11. *Primitive ideals of the radial C*-algebra $C^*_{\mathrm{rad}}(G)$ are exactly kernels of spherical representations of G:*

$$\mathrm{Prim}(C^*_{\mathrm{rad}}(G)) = \{\ker\pi\} \sqcup \{\ker\pi_t;\, t \in (-1,1)\setminus\sigma_r\} \sqcup \{\ker\varepsilon_{\pm1}\}$$

with the primitive ideal space of $\ker\lambda/\ker\Lambda = \ker\pi/\ker\Pi$ *identified with*

$$\{\ker\pi_t;\, t \in (-1,1)\setminus\sigma_r\} \sqcup \{\ker\varepsilon_{\pm1}\}.$$

We now look into the topology of the primitive ideal space $\Delta \equiv \mathrm{Prim}(C^*_{\mathrm{rad}}(G))$, which is a closed subset of $\mathrm{Prim}(C^*(G))$. Since primitive ideals in Δ are of the form $[t] = \ker\pi_t$ $(-1 < t < 1)$ or $[\pm1] = \ker\varepsilon_{\pm1}$, Δ is identified with a quotient of $[-1,1]$ in such a way that σ_r is shrunken to one point in Δ.

We first check the continuity of $[t] \in \Delta$ in $t \in [-1,1]$: If a sequence (t_n) in $[-1,1]$ converges to $t \in [-1,1]$, then $\bigcap_{n\ge1}\ker\pi_{t_n} \subset \ker\pi_t$ because $x \in \bigcap_{n\ge1}\ker\pi_{t_n}$ and $y \in C^*(G)$ satisfy

$$\|\pi_t(xy)\varsigma\| = \|xy\epsilon_t^{1/2}\| = \lim_{n\to\infty}\|xy\epsilon_{t_n}^{1/2}\| = \lim_{n\to\infty}\|\pi_{t_n}(x)\pi_{t_n}(y)\varsigma\| = 0.$$

Theorem 2.12. *For a nonempty subset T of $(-1,1)\setminus\sigma_r$, we have*

$$\overline{\{[t];\, t \in T\}} = \{[t];\, t \in \overline{T}\} \cup \{\ker\pi\}$$

in $\Delta = \mathrm{Prim}(C^*_{\mathrm{rad}}(G))$. Here the left hand side is the closure in Δ and $\overline{T}$ denotes the closure of T in $[-1,1]$.

Since $\ker\varepsilon_{\pm1}$ *and* $\ker\pi$ *are maximal ideals, they are closed as one-point sets. Thus the closure operation in Δ is completely described by this.*

Proof. Let us begin with showing that Δ contains relatively open intervals in $[-1,1]\setminus\sigma_r$ as open subsets of Δ: For a continuous function $f \in C([-1,1])$,

Proposition 2.6 gives

$$\|\pi_t(f(h_1))\| = \begin{cases} \max\{|f(t')|;\, t' \in \{t\} \cup \sigma_r\} & (-1 < t < 1), \\ |\varepsilon_t(f(h_1))| = |f(\pm 1)| & (t = \pm 1). \end{cases}$$

Since relatively open subsets of $[-1, 1] \setminus \sigma_r$ are then realized in the form

$$\{t \in [-1, 1] \setminus \sigma_r;\, \|\pi_t(f(h_1))\| > 0\}$$

with $f = 0$ on σ_r (relatively open subsets being disjoint unions of countably many relatively open intervals), they are open in Δ as well because $\Delta \ni [t] \mapsto \|\pi_t(f(h_1)\|$ is a lower semicontinuous function ([5] Proposition 3.3.2).

Thus a relatively open subset $[-1, 1] \setminus (\sigma_r \cup \overline{T})$ of $[-1, 1] \setminus \sigma_r$ is open in Δ and hence its complement $[\overline{T}] \cup \{\ker \pi\}$ in Δ is closed. Consequently

$$[\overline{T}] \subset \overline{[T]} \subset [\overline{T}] \cup \{\ker \pi\}.$$

Here the first inclusion is due to the continuity of $[t]$ in $t \in [-1, 1]$. Since T is nonempty, its closure $\overline{[T]}$ in Δ contains $\ker \pi$ (Corollary 2.8) and the assertion is proved. □

Corollary 2.13. *Open sets of* Δ *are exactly of the following form*:

(i) $[U]$ *with* $U \subset [-1, 1] \setminus \sigma_r$ *an open subset in the relative topology of* $[-1, 1]$.

(ii) $\Delta \setminus F$ *with* F *a subset* (*including the empty set*) *of* $\{[1], [-1]\}$.

Corollary 2.14. *The radial C*-algebra* $C^*_{\mathrm{rad}}(G)$ *of* G *is isomorphic to* $C^*(G)/\ker \Pi$ *and the center of* $C^*_{\mathrm{rad}}(G)$ *is trivial.*

Proof. Let $f : \Delta \to \mathbb{C}$ be a continuous function. Since $\ker \pi \in \overline{\{\ker \pi_t\}}$ $(-1 < t < 1)$, $f([t]) = f(\ker \pi)$ for $-1 < t < 1$ and then for $-1 \le t \le 1$ by continuity of $[t]$ in t. Thus f is constant on Δ and the assertion follows from the Dauns–Hofmann theorem ([12], §4.4). □

References

[1] C. A. Akemann and J. Anderson, "The Stone–Weierstrass problem for C^*-algebras", pp. 15–32 in *Invariant subspaces and other topics* (Timişoara and Herculane, Romania, 1981), Oper. Theory Adv. Appl. **6**, Birkhäuser, Basel, 1982. MR

[2] N. P. Brown and E. P. Guentner, "New C*-completions of discrete groups and related spaces", *Bull. Lond. Math. Soc.* **45**:6 (2013), 1181–1193. MR

[3] P. Cartier, "Harmonic analysis on trees", pp. 419–424 in *Harmonic analysis on homogeneous spaces* (Williamstown, MA, 1972), Proc. Sympos. Pure Math. **26**, Amer. Math. Soc., Providence, RI, 1973. MR

[4] J. M. Cohen, "Operator norms on free groups", *Boll. Un. Mat. Ital. B* (6) **1**:3 (1982), 1055–1065. MR

[5] J. Dixmier, *C^*-algebras*, North-Holland Mathematical Library **15**, North-Holland, Amsterdam, 1977. MR

[6] A. Figà-Talamanca and M. A. Picardello, *Harmonic analysis on free groups*, Lecture Notes in Pure and Appl. Math. **87**, Dekker, New York, 1983. MR

[7] U. Haagerup, "An example of a nonnuclear C^*-algebra, which has the metric approximation property", *Invent. Math.* **50**:3 (1979), 279–293. MR

[8] H. Kesten, "Symmetric random walks on groups", *Trans. Amer. Math. Soc.* **92** (1959), 336–354. MR

[9] A. M. Mantero and A. Zappa, "The Poisson transform and representations of a free group", *J. Funct. Anal.* **51**:3 (1983), 372–399. MR

[10] A. M. Mantero and A. Zappa, "Irreducibility of the analytic continuation of the principal series of a free group", *J. Austral. Math. Soc. Ser. A* **43**:2 (1987), 199–210. MR

[11] R. Okayasu, "Free group C^*-algebras associated with ℓ_p", *Internat. J. Math.* **25**:7 (2014), art. id. 1450065, 12 pp. MR

[12] G. K. Pedersen, *C^*-algebras and their automorphism groups*, London Mathematical Society Monographs **14**, Academic Press, London, 1979. MR

[13] R. T. Powers, "Simplicity of the C^*-algebra associated with the free group on two generators", *Duke Math. J.* **42** (1975), 151–156. MR

[14] T. Pytlik, "Radial functions on free groups and a decomposition of the regular representation into irreducible components", *J. Reine Angew. Math.* **326** (1981), 124–135. MR

[15] T. Pytlik, "Spherical functions and uniformly bounded representations of free groups", *Studia Math.* **100**:3 (1991), 237–250. MR

[16] T. Pytlik and R. Szwarc, "An analytic family of uniformly bounded representations of free groups", *Acta Math.* **157**:3-4 (1986), 287–309. MR

[17] M. Reed and B. Simon, *Analysis of operators*, Academic Press, 1978.

[18] S. Sawyer, "Isotropic random walks in a tree", *Z. Wahrsch. Verw. Gebiete* **42**:4 (1978), 279–292. MR

[19] R. Szwarc, "An analytic series of irreducible representations of the free group", *Ann. Inst. Fourier* (*Grenoble*) **38**:1 (1988), 87–110. MR

[20] R. Szwarc, "Banach algebras associated with spherical representations of the free group", *Pacific J. Math.* **143**:1 (1990), 201–207. MR

[21] S. Yamagami, "Algebraic aspects in modular theory", *Publ. Res. Inst. Math. Sci.* **28**:6 (1992), 1075–1106. MR

[22] S. Yamagami, "On a moment problem of analytic functionals of polynomial hypergroups", preprint, 2019. To appear in J. Math. Soc. Japan. arXiv 1912.03895

Received April 11, 2025. Revised September 27, 2025.

SHIGERU YAMAGAMI
GRADUATE SCHOOL OF MATHEMATICS
NAGOYA UNIVERSITY
NAGOYA 464-8602
JAPAN

yamagami@math.nagoya-u.ac.jp

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 340 No. 2 February 2026

Functions of dissipative operators under relatively bounded and relatively trace class perturbations 199
A.B. ALEKSANDROV and V.V. PELLER

On Watanabe's theta graph diffeomorphism in the 4-sphere 229
DAVID T GAY

Mutation of signed valued quivers and presentations of simple complex Lie algebras 245
JOSEPH GRANT and DAVIDE MORIGI

An axiomatic approach to symmetric extensions 293
PETER HOLY

On Braverman–Kazhdan–Ngô pairs 309
DIHUA JIANG, ZHAOLIN LI and GUODONG XI

Arithmetic of cuts in ordered abelian groups and of ideals over valuation rings 339
FRANZ-VIKTOR KUHLMANN and KATARZYNA KUHLMANN

The Ricci–DeTurck flow from an initial metric with a Morrey-type integrability condition 375
MAN-CHUN LEE and STEPHEN SHANG YI LIU

The geometric Cauchy problem for constant-rank submanifolds 399
MATTEO RAFFAELLI

On the primitive ideal space of radial representations of free groups 409
SHIGERU YAMAGAMI