Pacific Journal of Mathematics

Volume 307 No. 2 August 2020

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2020 is US $520/year for the electronic version, and $705/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 1945-5844 electronic, 0030-8730 printed) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.257

A SPECTRAL APPROACH TO THE LINKING NUMBER IN THE 3-TORUS

ADRIEN BOULANGER

Given a closed Riemannian manifold and a pair of multicurves in it, we give a formula relating the linking number of the latter to the spectral theory of the Laplace operator acting on differential 1-forms. As an application, we compute the linking number of any two multigeodesics of the flat torus of dimension 3, generalising a result of P. Dehornoy.

1. Introduction

Let (M, g) be a closed Riemannian manifold of dimension 3. We call *a curve* an embedding of the oriented circle. A *multicurve* is a finite collection of disjoint curves. We say that a multicurve is *homologically trivial* if its homology class vanishes, as a cycle of M.

Given two homologically trivial multicurves Γ, Υ, one defines their *linking number* by taking any surface S_Γ whose boundary is Γ and algebraically intersecting it with Υ;

$$\operatorname{lk}(\Gamma, \Upsilon) := i(S_\Gamma, \Upsilon).$$

For example, see Figure 1.

It is not immediate that this number is well-defined, because of the choice involved about a surface S_Γ. As a general reference to the notion of linking number, one can recommend [Arnold and Khesin 1998, Section 4 of Chapter III; Bott and Tu 1982, Section 28]. Our main theorem relates the linking number with spectral theory.

Theorem 1.1. *Let (M, g) be a closed Riemannian manifold and Γ, Υ two disjoint homologically trivial multicurves, they link according to the following formula*

$$\operatorname{lk}(\Gamma, \Upsilon) = \lim_{t\to 0} \sum_{k\geq 0} e^{-\lambda_k t} \int_\Gamma \eta_k \int_\Upsilon *\left(\frac{d\eta_k}{\lambda_k}\right), \tag{1.2}$$

MSC2010: 55M99, 58J35, 58J50.
Keywords: linking number, spectral theory, differential geometry.

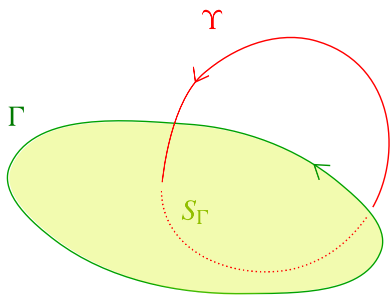


Figure 1. Here, both collections Γ and Υ consist of a single curve. Their linking number is ±1, depending on the global orientation.

where $(\eta_k)_{k\in\mathbb{N}}$ *denotes an eigenvector basis with corresponding eigenvalues* $(\lambda_k)_{k\in\mathbb{N}}$ *of the Laplace operator* Δ *acting on the Hilbert space of square integrable differential* 1*-forms in* $\ker(\Delta)^\perp$.

Note that this theorem relates a topological number with metric quantities. In particular the right member of (1.2) does not depend on the underlying metric g.

Theorem 1.1 can be used if one has enough knowledge of the spectral theory of (M, g), as in the case of the canonical flat torus $\mathbb{T}^3 := \mathbb{R}^3/\mathbb{Z}^3$. We prove a general formula for the linking number of multicurves consisting of geodesics of $\mathbb{T}^3$. However, not to burden this introduction, we postpone the statement to Section 3. Specialising our formula to the case of closed orbits of the geodesic flow on the 2-torus $\mathbb{T}^2$ gives the following corollary.

Corollary 1.3. *Let* $\Gamma = (\gamma^i)_{i\in I}$ *and* $\Upsilon = (\upsilon^j)_{j\in J}$ *be two homologically trivial multicurves in* $\mathbb{T}^3$ *consisting of periodic orbits of the* $\mathbb{T}^2$ *geodesic flow. They link according to the formula*

$$\operatorname{lk}(\Gamma, \Upsilon) = \sum_{i\in I,\, j\in J} \langle \gamma^i, \upsilon^j \rangle \frac{1-\theta_{i,j}/\pi}{2},$$

where $\theta_{i,j}$ *denotes the unique determination in* $[0, 2\pi[$ *of the oriented angle* θ *made at each intersection point* (see Figure 2) *and* $\langle \gamma^i, \upsilon^j \rangle$ *denotes the algebraic intersections between the projections on* $\mathbb{T}^2$ *of the curves* γ^i *and* υ^j.

Another formula was found by P. Dehornoy using different methods in his Ph.D. thesis [2011]. Our formula shows, in a clear way, that the linking number entertains some interactions with the intersection number on the curves projected on the basis.

We now briefly survey old and more recent results about the linking number. The first occurrence of the notion of linking number goes back to Gauss's studies on electromagnetism (see [Ricca and Nipoti 2011]). Gauss noticed that integrating the magnetic field generated by an electric power flowing in a close wire γ — for us a differential form ω_γ — along any closed curve υ gives a number which does

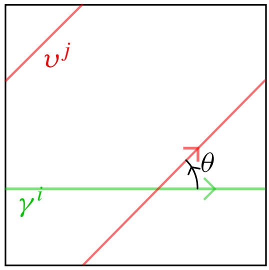


Figure 2. Here, the intersection number is 1. The angle θ defined in Corollary 1.3 is represented in black.

not depend on the homology class of υ in the complement of γ. That is to say, the differential form ω_γ is closed.

In fact, Gauss went further in his study: he gives in $\mathbb{R}^3$ an explicit formula expressing the differential form ω_γ. Let $x \notin \gamma$ and $X(x) \in T_x(\mathbb{R}^3)$, then:

$$(\omega_\gamma)_x(X) = \frac{1}{4\pi} \int_{[0,2\pi]} \det\left(\gamma'(s), X(x), \frac{\gamma(s) - x}{\|\gamma(s) - x\|^3} \right) ds, \tag{1.4}$$

where $\gamma(s)$ denotes any parametrisation compatible with the curve γ orientation, and $\|\cdot\|$ is the Euclidean norm of $\mathbb{R}^3$. Back in these days, there was no topological definition of the linking number so that, following Gauss, one could have defined it setting

$$\begin{aligned} \operatorname{lk}(\gamma, \upsilon) &:= \int_\upsilon \omega_\gamma \\ &= \frac{1}{4\pi} \int_{[0,2\pi]} \int_{[0,2\pi]} \det\left(\gamma'(s), \upsilon'(t), \frac{\gamma(s) - \upsilon(t)}{\|\gamma(s) - \upsilon(t)\|^3} \right) ds\, dt. \end{aligned} \tag{1.5}$$

Gauss's formula was related later with the definition of the linking number introducing this article, see for example [Arnold and Khesin 1998, Section 4 of Chapter 3]. It is still an active research field to try to get Gauss-like formulas for the linking number and its natural generalisations [DeTurck et al. 2008; DeTurck and Gluck 2008].

Formula (1.5) also suggests the existence of a universal object which, integrated over a pair of homologically trivial multicurves, gives back their linking number. A *linking form* Ω is an integrable $(1, 1)$-differential form, satisfying for any two homologically trivial disjoint multicurves Γ and Υ,

$$\operatorname{lk}(\Gamma, \Upsilon) = \int_\Gamma \int_\Upsilon \Omega.$$

One can think of a $(1, 1)$-differential form as a 2-differential form; we will get back on the $(1, 1)$-form precise definition in Section 2B.

The definition of linking form was introduced by Arnold (see [Arnold and Khesin 1998, Section 4 of Chapter III]) to generalise Moffatt's interpretation [1969] of the helicity. Let us recall briefly how the latter is defined.

Let X be a vector field preserving a probability measure μ in the Lebesgue class whose asymptotic cycle vanishes. This assumption implies that the 2-differential form $i_X\mu$ is exact, meaning that there is a differential 1-form α such that $d\alpha = i_X\mu$. One can show that

$$\mathcal{H}(X) := \int_M \alpha \wedge d\alpha$$

does not depend on the choice involving the primitive α. We call this number the helicity of the vector field X. This notion was introduced by [Moreau 1961; Woltjer 1958] to study certain energies associated to vector field solutions of some partial differential equations. Note that the asymptotic cycle assumption is automatically satisfied in some natural situations, for example when the ambient manifold is a homology sphere or if X is the Reeb flow associated to a contact structure.

Arnold interpreted the helicity of a vector field X as some average of the asymptotic linking number of two trajectories of the flow. Given $x, y \in M$, we consider the trajectories starting off x and y of the flow X at times t and s. We close them by gluing their extremities using a small path, we compute the linking number, we divide by the product ts and one would like to let $s, t \to \infty$. To do it, one needs to show this limit to be almost everywhere well-defined; this is one of the reasons why Arnold introduced the notion of linking form. Actually, he showed that the linking form is *integrable*, see Remark 4.14, which allows one to define the previous limit using Birkhoff's ergodic theorem. See [Arnold and Khesin 1998, Section 4 of Chapter III] for more details. This perspective on the helicity was developed in [Vogel 2003; Kotschick and Vogel 2003; DeTurck et al. 2013].

Arnold also noticed that linking forms always exists on compact manifolds. This was made more precise by T. Vogel [2003], relying on G. de Rham's work [1984, Section 28] on Hodge theory. We denote by $g^1(x, y)$ the kernel of the *Green operator*, the inverse operator of the Laplace one. We have

Theorem 1.6 [Vogel 2003, Theorem 3]. *Let (M, g) a compact Riemannian manifold. The $(1, 1)$-differential form*

$$\Omega(x, y) = *_y d_y g^1(x, y) \tag{1.7}$$

is an integrable linking form. We call this linking form the ***de Rham–Vogel linking form****.*

Vogel's proof relies on Arnold's remark that any inverse operator of d gives rise to a linking form (up to some duality). This theorem shows the existence of linking forms on closed manifolds, but does not come with a simple formula like

Gauss's one (1.5). There is, to the author's knowledge, only two others known formulas of this type, found in [DeTurck and Gluck 2013]. The first one holds for the hyperbolic 3-space and the second one for the round 3-sphere. The authors find such a formula in exhibiting a "fundamental solution of Maxwell's equations," meaning in exhibiting the de Rham–Vogel linking form defined above.

Outline of the article. In Section 2 we recall some basics on Laplace operators acting on differential forms and on their inverses, Green operators. We shall also recall that (1, 1)-forms are kernels of operators acting on differential 1-forms as well as how to integrate them along a pair of curves.

With this operator perspective in mind, Theorem 1.6 implies that the operator coming from the de Rham–Vogel linking form (the linking operator) commutes with the Laplace operator. The latter is well known to be diagonalisable on the space of square integrable differential 1-forms. Expanding the linking operator with respect to such a basis of 1-eigenforms will allow us to find a formula relating the spectral theory of differential 1-forms to the linking number.

However, the integration current over a closed curve is not square integrable, which prevents readily obtaining such a spectral-linking formula. To circumvent this difficulty, in order to reach some more regularity, we will use the heat operator to smooth the integration currents. This smoothing is responsible for the limit $t \to 0$ appearing in Theorem 1.1. This is the heart of Section 3 which finishes with the proof of Theorem 1.1.

In Section 4, using Theorem 1.1, we compute the linking numbers of collections of geodesics in the flat 3-torus $\mathbb{T}^3$ for which the spectral theory of differential 1-forms is well known. As an application, we prove Corollary 1.3.

2. Kernels of Green operators and linking forms

This section is devoted to introduce all the objects we will use later on.

2A. ***The Laplace and the Green operators.*** Let (M, g) be a closed manifold of dimension p. We denote by

- μ_g the volume form associated to the metric g;
- $\Omega^*(M) = \bigoplus_{0\leq k\leq p} \Omega^k(M)$ the space of all differential forms, graded with respect to the degree k;
- $*$ the Hodge operator, or Hodge star, which satisfies the following identity

$$**=(-1)^{k(p-k)}; \tag{2.1}$$

Note that we abuse the notation by omitting the degree k of the underlying differential form.

- d the exterior differential operator on $\Omega^*(M)$.

The Hodge star is defined in order to endow the vectorial space $\Omega^k(M)$ with a scalar product

$$\langle \alpha \cdot \beta \rangle = \int_M \alpha \wedge *\beta.$$

With respect to it, the operator d has a unique adjoint operator, denoted by δ, satisfying by definition

$$\langle d\alpha \cdot \beta \rangle = \langle \alpha \cdot \delta\beta \rangle.$$

A straightforward computation involving the Hodge star definition and the Stokes formula gives

$$\delta = (-1)^{p(k+1)+1} * d * . \tag{2.2}$$

We now have all the material required to define the Laplace operator.

Definition 2.3. The *Laplace operator* acting on $\Omega^*(M)$, denoted by Δ, is defined by

$$\Delta := d\delta + \delta d.$$

Note that the Laplace operator stabilises all differential forms spaces of fixed degree, and that it is self-adjoint with respect to the scalar product $\langle \cdot , \cdot \rangle$. We denote by Δ^k its restriction to $\Omega^k(M)$.

A differential form is said to be harmonic if it lies in the kernel of the Laplace operator, denoted by $\ker \Delta$. The space $\mathcal{H}_k$ of all harmonic k-forms being identified, by a famous theorem of Hodge, to the k-th homology group of M (see [Rosenberg 1997, page 46] for example), the Laplace operator Δ^k cannot be invertible in general. However, nothing prevents it from being invertible when restricted to the space orthogonal to its kernel. We denote by $\pi_{\mathcal{H}^k}$ the orthogonal projection on $\mathcal{H}_k = \ker(\Delta^k)$.

Definition 2.4. A *Green operator*, denoted by G^k, is any operator satisfying the following equation on the space of smooth differential forms of degree k:

$$G^k \circ \Delta^k = \Delta^k \circ G^k = \mathrm{Id} - \pi_{\mathcal{H}^k} \tag{2.5}$$

Such an operator always exists, provided that M is closed. There is a slight ambiguity about G^k, which is fully determined up to its restriction on the space $\ker(\Delta^k)$. From now on, we will suppose that $G(\ker(\Delta^k)) = \{0\}$, allowing one to speak of *the* Green operator. One can recommend [de Rham 1984, Chapter 3] for a general introduction to Green operators (and their kernels).

Green operators are *kernel operators*, meaning that there is a smooth family of endomorphisms — what we call a (1, 1)-form — $g^k(x, y) : \Lambda^k(T_xM) \to \Lambda^k(T_yM)$, indexed by $M \times M \setminus \mathrm{Diag}$ such that for all smooth differential forms α of degree k

$$G^k(\alpha)_y = \int_{x \in M} g^k(x, y)(\alpha_x)\, d\mu_g(x).$$

2B. ***Differential* (1, 1)*-forms.*** We give in this subsection the precise definition of a (1, 1)-form. We also explain how to integrate them over a pair of multicurves. Given an Euclidean space E, we denote by $\sharp$ the musical endomorphism which maps some vector $X \in E$ on its dual linear form, so in E^*, according to the Euclidean structure on E.

Definition 2.6. Let M be a manifold, We call $(1, 1)$-*form* a family of morphisms $T_x^*(M) \to T_y^*(M)$ indexed by $M \times M$.

Let γ and υ two curves parametrised by s and t. We define the integral over the pair of curves (γ, ν) of a (1, 1)-form Ω as

$$\int_\gamma \int_\upsilon \Omega := \int_\gamma \left(\int_0^1 \Omega(\upsilon(s), y)(\upsilon'(s)^\sharp)\, ds \right).$$

Moreover, the following integral — an element of $T_y^*(M)$ —

$$\int_0^1 \Omega(\upsilon(s), y)(\upsilon'(s)^\sharp)\, ds$$

does not depend on a choice of parametrisation, since $\Omega(x, y)$ is linear. So that we will prefer to denote it for short as

$$\int_\upsilon \Omega((\cdot)^\sharp, y),$$

omitting the underlying parametrisation.

This formula clearly shows that the linking form enjoys some bilinearity. In fact, if we denote by $\Upsilon = \bigcup_{i \in I} \upsilon^i$ and $\Gamma = \bigcup_{j \in J} \gamma^j$ we have

$$\mathrm{lk}(\Gamma, \Upsilon) = \int_{\Gamma \times \Upsilon} \Omega = \int_{\bigcup \gamma_i \times \bigcup \gamma_j} \Omega = \sum_{i \in I,\, j \in J} \int_{\gamma_i \times \gamma_j} \Omega.$$

Note that we did not require that either of the curves γ_i or υ_j be homologically trivial.

2C. ***The de Rham–Vogel linking form.*** Recall that the de Rham–Vogel linking form is defined as

$$*_y d_y g^1(x, y),$$

which may be slightly confusing at first. What does it mean to consider the image by $*_y d_y$ of a family of morphism from $T_x^* M$ to $T_y^* M$?

Given $\alpha \in T_x^*(M)$, the Green kernel defines a differential form by

$$\alpha_y := y \mapsto g(x, y)(\alpha(x)) \in T_y^*(M).$$

This differential form is smooth on $M \setminus \{x\}$, which allows one to take its image by the operator $*d$ wherever it makes sense. This gives rise to another linear morphism

$$T_x^*(M) \to T_y^*(M), \quad \alpha \mapsto \alpha_y,$$

which turns out to correspond to the kernel of the operator $\alpha \mapsto (*dG^1)(\alpha)$. Then, in the end, de Rham's notation $*_y d_y g^1(x, y)$ is to be understood as the kernel of the operator $\alpha \mapsto (*dG^1)(\alpha)$ that we call the *linking operator*.

Remark 2.7. (1) As pointed out by Arnold, any kernel associated to the inverse operator of the exterior differential d is a linking form. The operator $\alpha \mapsto (*dG)(\alpha)$ is actually one of them, up to Hodge duality. See [Vogel 2003, Lemma 2].

(2) The singularity of the $(1, 1)$-form $g^1(x, y)$ along the diagonal is roughly equivalent to r^{-1}. Thus, after one differentiation, this singularity turns to be in r^{-2}, which is still integrable in dimension three, see [de Rham 1984, Theorem 23 page 134]. So that what we meant by integrable is that for every x the function

$$y \mapsto \| *_y d_y g(x, y)\|$$

is integrable on M with respect to μ_g. The notation $\| \cdot \|$ stands for the linear morphism norm induced by the metric g.

2D. ***Behaviour of the linking form under isometries.*** The de Rham–Vogel linking form being constructed from a metric, it is natural to wonder how it behaves under an isometry Φ. The isometry Φ commutes with the Hodge star as well as with the exterior differential d. In particular, it commutes with every operators made out of this two ones, as the Laplace operator and its inverse, the Green operator. Looking at the kernel of the latter, this commutation relation can be read as

$$(\Phi_1)_* g^1(x, y) = (\Phi_2)^* g^1(x, y),$$

where Φ_1 and Φ_2 denote the Φ-action on the first and second factor, respectively, of the product $M \times M$. In particular, the diagonal action of Φ on the product $M \times M \setminus \mathrm{Diag}$ preserves the Green kernel, and thus the de Rham–Vogel linking form. Since we will use this remark to simplify a bit the computations performed in Section 4, we present it as:

Proposition 2.8. *Let γ and υ two curves* (not necessary homologically trivial) *and Φ an isometry of* (M, g), *then*

$$\int_\gamma \int_\upsilon *_y d_y g^1 = \int_{\Phi^{-1}(\gamma)} \int_{\Phi^{-1}(\upsilon)} *_y d_y g^1.$$

3. The spectral-linking formula

Let us recall that Δ is self-adjoint with respect to $\langle \cdot, \cdot \rangle$. It is well known that self-adjoint operators are diagonalisable in finite dimension; it is actually still the case for the Laplace operator, provided that the underlying manifold is closed.

Theorem 3.1 [Rosenberg 1997, Theorem 1.30]. *Let (M, g) be a closed Riemannian manifold. There is a orthonormal basis $(\eta_n)_{n\in\mathbb{N}}$ of differential 1-forms, meaning that $\langle \eta_i, \eta_j \rangle = \delta_i(j)$, and a sequence of nonnegative numbers $(\lambda_n)_{n\in\mathbb{N}}$ such that*

$$\Delta \eta_n = \lambda_n \eta_n.$$

In particular, if $\alpha \in \ker(\Delta)^{\perp}$ we have

$$\alpha = \sum_{n\in\mathbb{N}} \langle \eta_n \cdot \alpha \rangle \eta_n.$$

Formally, one would like to write the Green operator as

$$g^1(x, y) := \sum_{n\in\mathbb{N}} \frac{1}{\lambda_n} \eta_n(x) \otimes \eta_n(y),$$

which gives the following expression for the de Rham–Vogel linking form:

$$*_y d_y g^1(x, y) := \sum_{n\in\mathbb{N}} \eta_n(x) \otimes *d\left(\frac{\eta_n(y)}{\lambda_n}\right).$$

Keeping it formal, one would like then to integrate each factor along γ and υ to get

$$\mathrm{lk}(\gamma, \upsilon) = \sum_{n\in\mathbb{N}} \int_\gamma \eta_n \int_\upsilon *\left(\frac{d\eta_n}{\lambda_n}\right).$$

However, the previous series does not converge a priori. In fact, an integration current over a curve is not square integrable and therefore cannot be decomposed with respect to the orthonormal basis (η_n). To circumvent this difficulty, we will regularise them thanks to the use of the heat kernel, from which the term $e^{-\lambda_n t}$ of formula (1.2) comes from. As a corollary of this approach, we are able to prove the following stronger version of Theorem 1.1.

Theorem 3.2. *Let (M, g) be a closed Riemannian manifold and Ω the de Rham–Vogel linking form, then for all pairs of curves γ and υ* (not necessary homologically trivial) *we have*

$$\int_\gamma \int_\upsilon \Omega = \lim_{t\to 0} \sum_{k>0} e^{-\lambda_k t} \int_\gamma \eta_k \int_\upsilon *\left(\frac{d\eta_k}{\lambda_k}\right), \tag{3.3}$$

where $(\eta_k)_{k\in\mathbb{N}}$ *denotes an eigenvector basis of the Laplace operator* Δ *acting on the Hilbert space* $\ker(\Delta)^\perp$ *— viewed as a subspace of square integrable differential forms — and* (λ_k) *the associated eigenvalues.*

All the rest of this section is dedicated to the proof of the above theorem.

3A. ***The heat operator on 1-differential forms.*** The following definition is the key to regularise the integration currents. More details about generalised heat kernels can be found in [Berline et al. 1992, Section 2.3].

Definition 3.4. Let (M, g) be a closed Riemannian manifold and η be a continuous, bounded differential 1-form. The following Cauchy problem of unknown $(\eta_t)_{t\in\mathbb{R}_+}$:

$$\begin{cases} \Delta\eta_t + \partial_t\eta_t = 0, \\ \eta_0 = \eta, \end{cases}$$

has a unique solution. We denote by $e^{-t\Delta^1}$ the *heat operator* which maps η to the time t solution of the above Cauchy problem.We denote by p_t^1 the *heat kernel*, which satisfies by definition

$$(\eta_t)_y = \int_M p_t^1(x, y)(\eta_x)\, d\mu_g(x).$$

Moreover, one has

$$e^{-t\Delta^1}(\eta) \xrightarrow{t\to 0} \eta$$

for the uniform convergence topology.

In particular, if U and V are two closed disjoin subsets of M, one has

$$p_t^1(x, y) \xrightarrow{t\to 0} 0$$

uniformly on $U \times V$.

The heat kernel has the interesting property of being smooth for all $t > 0$, as opposed to the Green operator. In particular, it can be decomposed according to an orthonormal basis of eigenforms.

3B. ***The diffused curves.*** Let γ be a curve of M. We denote by $L^1(\Omega^1(M))$ the space of integrable differential 1-form, meaning forms whose punctual norm is integrable over M with respect to the Riemannian measure.

Definition 3.5. We call the γ*-diffused curve*, denoted by $e^{-t\Delta^1}(\gamma)$, the following family of linear forms indexed by $t > 0$:

$$e^{-t\Delta^1}(\gamma) : L^1(\Omega^1(M)) \to \mathbb{R}, \quad \beta \mapsto \int_\gamma e^{-t\Delta^1}(\beta).$$

This diffusing process associates to each $t > 0$ a differential form approximating the integration current over the curve γ: the smaller t, the better the approximation.

Lemma 3.6. *For all $\beta \in L^1(\Omega^1(M))$ continuous on a neighbourhood of U of the curve γ, we have*

$$e^{-t\Delta^1}(\gamma)(\beta) \xrightarrow{t\to 0} \int_\gamma \beta.$$

Proof. We have been careful to consider a differential form β integrable. So, since the heat kernel converges uniformly to 0 away from the diagonal, we have

$$\left| e^{-t\Delta^1}(\gamma)(\beta) - \int_\gamma \int_U p_t^1(x,y)(\beta_x)\, d\mu_g(x) \right| \xrightarrow{t\to 0} 0.$$

The differential form β being continuous on U, from the very definition on the heat kernel we have

$$\int_U p_t^1(x,y)(\beta_x)\, d\mu_g(x) \xrightarrow{t\to 0} \beta_y$$

uniformly. Therefore, one is allowed to permute limit and integral to get

$$\int_\gamma \int_U p_t^1(x,y)(\beta_x)\, d\mu_g(x) \xrightarrow{t\to 0} \int_\gamma \beta,$$

which is the expected result. □

If now υ is a curve disjoint to γ, recall that the differential 1-form

$$(\omega_\upsilon)_y := \int_\upsilon \Omega((\cdot)^\sharp, y)$$

is integrable, where $\Omega = *_1 d_1 g_1$ is the de Rham–Vogel linking form. Applying Lemma 3.6 readily gives:

Corollary 3.7. *For any two curves γ and υ we have*

$$e^{-t\Delta^1}(\gamma)(\omega_\upsilon) \xrightarrow{t\to 0} \int_\gamma \int_\upsilon \Omega.$$

The goal is now to identify, $t > 0$ being fixed, the left member of the above equation to the series appearing in (3.3). We will conclude by using the above corollary to recover Theorem 3.2 by letting $t \to 0$.

3C. ***The approximating series.*** The benefits of having diffused the integration current is to allow one to write the left member of (3.3) as a scalar product of two smooth differential 1-forms. We will conclude by using Plancherel's formula, allowing one to write down this scalar product with respect to an orthonormal basis.

Lemma 3.8. *For all differential forms $\beta \in L^1(\Omega^1(M))$ and all $t > 0$ we have*

$$e^{-t\Delta^1}(\gamma)(\beta) = \left\langle \beta \cdot \int_\gamma p_t^1((\cdot)^\sharp, y) \right\rangle.$$

Note the scalar product is well-defined, since the differential form $\int_\gamma p_t^1((\,\cdot\,)^\sharp, y)$ is smooth.

Proof. The operator $e^{-t\Delta}$ being self-adjoint and since $i_X(\alpha)(x) = g_x(X^\sharp \cdot \alpha)$, we have the following identity for any differential 1-form β and any vector field X:

$$i_X(y)(p_t(x, y)\beta_x) = g\big(\beta_y \cdot (p_t(x, y)(X_x^\sharp))\big).$$

Therefore, setting $X_x = \gamma'(s)$ and integrating along γ, one gets

$$\int_\gamma p_t^1(x, \,\cdot\,)(\beta_x) = g_y\bigg(\beta_y, \int_\gamma p_t^1((\,\cdot\,)^\sharp, y)\bigg),$$

which gives, after integration over M with respect to μ_g,

$$\int_M \int_\gamma p_t^1(x, y)(\beta_x)\, d\mu_g(y) = \bigg\langle \beta \cdot \int_\gamma p_t^1((\,\cdot\,)^\sharp, y)\bigg\rangle.$$

We conclude recalling that the form β is integrable, which allows one to switch both integrals of the above equation left member, recovering our definition of a diffused curve. □

We conclude the proof of Theorem 3.2 as announced by identifying the right member of (3.3) with some series.

Lemma 3.9. *For all $t > 0$ we have*

$$e^{-t\Delta^1}(\gamma)(\omega_\upsilon) = \sum_{k>0} e^{-\lambda_k t} \int_\gamma \eta_k \int_\upsilon *\bigg(\frac{d\eta_k}{\lambda_k}\bigg).$$

Proof. We start by using the semigroup property of the heat operator $e^{-t\Delta}$,

$$e^{-t\Delta^1}(\gamma)(\omega_\upsilon) = e^{-t\Delta^1/2}(\gamma)(e^{-t\Delta^1/2}(\omega_\upsilon)),$$

for which we apply Lemma 3.8 to get

$$e^{-t\Delta^1/2}(\gamma)(e^{-t\Delta^1/2}(\omega_\upsilon)) = \bigg\langle \int_\gamma p_{t/2}^1((\,\cdot\,)^\sharp, y) \cdot (e^{-t\Delta^1/2}(\omega_\upsilon))\bigg\rangle.$$

Both differential 1-forms appearing in the above equation being smooth, one is able to write down this scalar product with respect to an orthonormal basis consisting of the Laplace operator eigenforms:

$$e^{-t\Delta^1}(\gamma)(\omega_\upsilon) = \sum_{k\in\mathbb{N}} \bigg\langle \bigg[\int_\gamma p_{t/2}^1((\,\cdot\,)^\sharp, y)\bigg] \cdot \eta_k \bigg\rangle \big\langle e^{-t\Delta^1/2}(\omega_\upsilon) \cdot \eta_k \big\rangle.$$

It remains then to prove both the two following identities:

$$e^{-\lambda_k t/2}\int_\gamma \eta_k = \left\langle \left[\int_\gamma p^1_{t/2}((\cdot)^\sharp, y)\right]\cdot \eta_k \right\rangle, \tag{3.10}$$

$$\frac{e^{-\lambda_k t/2}}{\lambda_k}\int_\upsilon *d\eta_k = \left\langle e^{-t\Delta^1/2}(\omega_\upsilon)\cdot \eta_k\right\rangle \tag{3.11}$$

We start off with the right member of (3.10). Recalling Lemma 3.8 the other way around, one gets

$$\left\langle \left[\int_\gamma p^1_{t/2}((\cdot)^\sharp, y)\right]\cdot \eta_k \right\rangle = \int_\gamma e^{-t\Delta^1/2}(\eta_k).$$

Then, η_k being an eigenform of eigenvalue λ_k, we have

$$e^{-t\Delta^1/2}(\eta_k) = e^{-t\lambda_k/2}\int_\gamma \eta_k,$$

which proves that (3.10) holds.

Let us show in the same way that (3.11) occurs as well. We start again from the right member:

$$\left\langle e^{-t\Delta^1/2}(\omega_\upsilon)\cdot \eta_k\right\rangle.$$

The differential 1-form ω_υ being integrable and the operator $e^{-t\Delta/2}$ being self-adjoint we have:

$$\left\langle e^{-t\Delta^1/2}(\omega_\upsilon)\cdot \eta_k\right\rangle = \left\langle \omega_\upsilon \cdot e^{-t\Delta^1/2}(\eta_k)\right\rangle.$$

Therefore, η_k being an eigenform of eigenvalue λ_k, we have

$$e^{-t\Delta^1/2}(\eta_k) = e^{-t\lambda_k/2}\eta_k,$$

and thus

$$\left\langle e^{-t\Delta^1/2}(\omega_\upsilon)\cdot \eta_k\right\rangle = e^{-\lambda_k t/2}\langle \omega_\upsilon \cdot \eta_k\rangle.$$

The linking form Ω being integrable, one can use Fubini's theorem again to get

$$\langle \omega_\upsilon \cdot \eta_k\rangle = \int_\upsilon \left[\int_M \Omega((\cdot)^\sharp, y)(\eta_k)_x \, d\mu_g(x)\right].$$

But, by construction of the linking form as the operator $*dG$ kernel, we have

$$\int_M \Omega(x, y)(\eta_k)_x \, d\mu_g(y) = ((*dG)(\eta_k))_y.$$

The operator G commutes with d, in particular all terms of this series corresponding to a closed differential form vanish. The remaining terms are given by

$$*dG(\eta_k) = \frac{*d\eta_k}{\lambda_k}.$$

Finally we have

$$\langle \omega_\upsilon \cdot \eta_k \rangle = \int_\upsilon * \left(\frac{d\eta_k}{\lambda_k} \right),$$

which gives

$$\left\langle e^{-t\Delta^1/2}(\omega_\upsilon) \cdot \eta_k \right\rangle = e^{-t\lambda_k/2} \int_\upsilon * \left(\frac{d\eta_k}{\lambda_k} \right), \tag{3.12}$$

the expected result. □

4. Application to torus geodesics linking

This section is devoted to the use of Theorem 3.2 to compute the linking number of homologically trivial multigeodesics of the canonical 3-tore torus $\mathbb{T}^3 := \mathbb{R}^3/\mathbb{Z}^3$. The spectral theory of $\mathbb{T}^3$ is fully understood: we will describe it in Section 4B. One would like to use it to give a more or less explicit expression to the following series, $t > 0$ being fixed:

$$\sum_{k>0} e^{-\lambda_k t} \int_\gamma \eta_k \int_\upsilon * \left(\frac{d\eta_k}{\lambda_k} \right).$$

Right after, we will identify the limit of this series when $t \to 0$ to the Fourier development of some function. In the meantime, we will recall Theorem 3.2, which guarantees that this sequence of series actually converges to the linking number.

4A. ***Statement of the generalised torus linking theorem.*** We call a *multigeodesic* a multicurve consisting of geodesics. This subsection goal is to state a formula giving the linking number of any two collections of multigeodesics of $\mathbb{T}^3$.

Let us fix some notation. Given a closed geodesic γ of $\mathbb{T}^3$, we parametrise it as

$$\gamma : \mathbb{R}/\mathbb{Z} \to \mathbb{T}^3, \quad t \mapsto \begin{pmatrix} \gamma_1 t + \nu_1 \\ \gamma_2 t + \nu_2 \\ \gamma_3 t + \nu_3 \end{pmatrix} \quad \text{mod } \mathbb{Z}^3,$$

where $\gamma' = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{Z}^3$ is the *parametrised slope* of γ. With the above notation $\gamma(0) = \gamma(1)$. Note that these curves are automatically oriented by the parametrisation. Note also that we did not require that the geodesic γ is primitive.

We call the point $\nu = \gamma(0) = (\nu_1, \nu_2, \nu_3) \in \mathbb{T}^3$ the *origin* of γ. Such a choice is not canonical since any point $\nu \in \mathrm{Im}(\gamma)$ can also define the origin. See Figure 3.

The parametrisation proposed above is not the arc-length one, but has the benefit to be very closely related to the homology class that γ defines. The 3-torus fundamental group being Abelian, one can check that the vector $\gamma' \in \mathbb{Z}^3$ is canonically identified to the homology class of the closed curve γ in $\mathbb{Z}^3$. We denote by $[\gamma]$ the vector γ' to emphasis its topological flavour. From that remark comes the

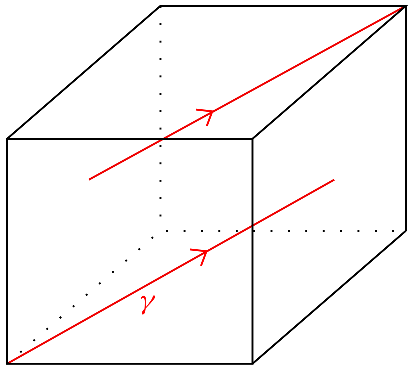


Figure 3. The cube is identified to $\mathbb{T}^3$ by gluing opposed faces with translations. The red curve γ has parametrised slope $\gamma' = (2, 1, 1)$. The origin of γ can be taken anywhere on the image of γ. For example, one can choose $\nu = 0_{\mathbb{T}^3}$ here.

following necessary and sufficient condition for a multigeodesics $\Gamma = (\gamma^i)_{i\in I}$ to be homologically trivial:

$$\sum_{i\in I} [\gamma^i] = 0_{\mathbb{R}^3}.$$

The following construction is needed to state our theorem.

Given two vectors $u, v \in \mathbb{Z}^3$, we define the vector $\beta^{u,v} \in \mathbb{Z}^3$ as the unique one verifying the following conditions:

- $\beta^{u,v} \in \text{vect}(u, v)^{\perp}$;
- $\det(u, v, \beta^{u,v}) > 0$;
- its Euclidean norm $\|\beta^{u,v}\|$ is minimal for the two first properties.

Given two geodesics γ and ν, we still simply denote by $\beta^{\gamma,\nu}$ the vector $\beta^{[\gamma],[\nu]}$. Our torus linking theorem can then be stated as follows.

Theorem 4.1. *Let $\Gamma = (\gamma^i)_{i\in I}$ and $\Upsilon = (\upsilon^j)_{j\in J}$ two homologically trivial multigeodesics of $\mathbb{T}^3$. They link according to the following formula*:

$$\text{lk}(\Gamma, \Upsilon) = \sum_{i\in I, j\in J} \det\left([\gamma^i], [\upsilon^j], \frac{\beta^{i,j}}{\|\beta^{i,j}\|}\right) \frac{1 - 2\lfloor(\nu^{i,j}\cdot\beta^{i,j})\rfloor}{2\|\beta^{i,j}\|}, \tag{4.2}$$

where $\nu^{i,j} = \gamma^i(0) - \upsilon^j(0)$ is the difference between the two origins and $\lfloor\alpha\rfloor$ denotes the unique representative in $[0, 1[$ *of the class* $(\alpha \mod \mathbb{Z})$.

Remark 4.3. (1) One can define the linking number in every dimension n, provided that we consider two homologically disjoint submanifolds of dimension p and q

satisfying $p+q=n-1$. Our method is likely to be generalised for a flat torus in any dimension.

(2) A priori, (4.2) depends on a choice of parametrisation. We will clarify this point later on with Remark 4.14.

4B. ***Spectral theory of differential 1-forms of*** $\mathbb{T}^3$**.** We start by introducing some notation.

- We denote by a lower index i the i-th coordinate of a vector and by an upper index its belonging to a family of vectors. For example, γ_i^j denotes the i-th coordinate of the j-th vector of a family indexed by $j \in J$.
- Given a vector
$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \in \mathbb{R}^3,$$
we denote by v^* the differential form $v_1\, dx_1 + v_2\, dx_2 + v_3\, dx_3$. This one being invariant by translations, it defines a harmonic differential form on $\mathbb{T}^3$. *We continue to denote by* v^* *the induced-on-*$\mathbb{T}^3$ *differential form.*
- The scalar product of two vectors a and b in $\mathbb{R}^3$ is denoted by $(a \cdot b)$ and the associated Euclidean norm by $\|\cdot\|$.
- The $\mathbb{R}^3$ vectorial product is denoted by $\wedge$.

Let us describe the differential 1-forms spectral theory of $\mathbb{T}^3$ thanks to the following set of datum:

- a vector
$$\boldsymbol{k} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} \in \mathbb{Z}^3,$$
- an orthonormal basis (v^1, v^2, v^3) of $\mathbb{R}^3$.
- a function $f \in \{\cos, \sin\}$

Note that here we have the choice of an orthonormal basis of $\mathbb{R}^3$.

Lemma 4.4. *The differential* 1*-form of* $\mathbb{T}^3$

$$\eta(x) = \sqrt{2} f(2\pi(\boldsymbol{k} \cdot x))(v^i)^* \tag{4.5}$$

is an eigenform of Δ^1 *with associated eigenvalue* $\lambda = (2\pi\|\boldsymbol{k}\|)^2$.

Proof. We start by showing that these forms are of unit norm:

$$\begin{aligned}\|\eta\|_{L^2} &:= \int_{\mathbb{T}^3} \eta \wedge *\eta \\ &= \int_{\mathbb{T}^3} 2f^2(2\pi(\boldsymbol{k}\cdot x))(v^i)^* \wedge *(v^i)^* \\ &= \int_{\mathbb{T}^3} 2f^2(2\pi(\boldsymbol{k}\cdot x))\, d\mathrm{vol} = 1,\end{aligned}$$

since $f^2 = (1 \pm f(2\cdot))/2$.

Recall the Laplace operator definition

$$\Delta\eta = (d\delta + \delta d)\eta.$$

Because $\delta = -*d*$ in dimension 3, one gets

$$d\delta\eta = d(-*d*)\eta = -\sqrt{2}(d*d)\big(f(2\pi(\boldsymbol{k}\cdot x)) * (v^i)^*\big). \tag{4.6}$$

By the Hodge star definition we have

$$*(v^i)^* = (v^j)^* \wedge (v^t)^*,$$

where (i, j, t) is a circular permutation of $(1, 2, 3)$, so that

$$d\delta\eta = -\sqrt{2} d*d\big(f(2\pi(\boldsymbol{k}\cdot x)) \wedge (v^j)^* \wedge (v^t)^*\big). \tag{4.7}$$

And then,

$$\begin{aligned}d\delta\eta &= -\sqrt{2}(d*)\big(2\pi k_i f'(2\pi(\boldsymbol{k}\cdot x))\big((v^i)^* \wedge (v^j)^* \wedge (v^t)^*\big)\big) \\ &= -2\sqrt{2}\pi k_i\, d\big(f'(2\pi(\boldsymbol{k}\cdot x))\big) \\ &= -2\sqrt{2}\pi k_i\, df'(2\pi(\boldsymbol{k}\cdot x)) \\ &= -4\sqrt{2}\pi^2\Big(k_i^2 f''(2\pi(\boldsymbol{k}\cdot x))\, dx_i + k_i k_j f''(2\pi(\boldsymbol{k}\cdot x))\, dx_j \\ &\qquad + k_i k_t f''(2\pi(\boldsymbol{k}\cdot x))\, dx_t\Big).\end{aligned}$$

We compute $\delta d\eta$ in a similar way to get

$$\begin{aligned}\delta d\eta = &-4\sqrt{2}\pi^2\big(k_j^2 f''(2\pi(\boldsymbol{k}\cdot x))\, dx_i + k_t^2 f''(2\pi(\boldsymbol{k}\cdot x))\big) \\ &+ 4\sqrt{2}\pi^2\big(k_i k_j f''(2\pi(\boldsymbol{k}\cdot x))\, dx_j - k_i k_t f''(2\pi(\boldsymbol{k}\cdot x))\, dx_t\big),\end{aligned}$$

Summing both terms gives

$$\Delta\eta = -4\sqrt{2}\pi^2\big(k_i^2 f''(2\pi(\boldsymbol{k}\cdot x))\, dx_i + k_j^2 f''(2\pi(\boldsymbol{k}\cdot x))\, dx_i + k_t^2 f''(2\pi(\boldsymbol{k}\cdot x))\big)\, dx_i.$$

Since $f'' = -f$ one has

$$\Delta\eta = 4\pi^2(k_1^2 + k_2^2 + k_3^2)\eta,$$

the expected outcome. □

To use Theorem 3.2 we need a basis of eigenforms. Fixing an orthonormal basis of $\mathbb{R}^3$, the family issued from all $\boldsymbol{k} \in \mathbb{Z}^3$ and both the function cos and sin forms a generating family. To see it, one can decomposes in Fourier series the coefficients of a differential 1-form ω written as

$$\omega(x, y, z) = f_1(x, y, z) \cdot (v^1)^* + f_2(x, y, z) \cdot (v^2)^* + f_3(x, y, z) \cdot (v^3)^*.$$

Moreover, this family is free up to the trivial relations $\cos(-\boldsymbol{k} \cdot x) = \cos(\boldsymbol{k} \cdot x)$ and $\sin(-\boldsymbol{k} \sin x) = -\sin(\boldsymbol{k} \cdot x)$.

4C. ***Computation of the approximating series.*** Recall that we parametrised both geodesics γ and υ as

$$\gamma : \mathbb{R}/\mathbb{Z} \to \mathbb{T}^3, \quad t \mapsto \begin{pmatrix} \gamma_1 t + \nu_1 \\ \gamma_2 t + \nu_2 \\ \gamma_3 t + \nu_3 \end{pmatrix}, \qquad \upsilon : \mathbb{R}/\mathbb{Z} \to \mathbb{T}^3, \quad t \mapsto \begin{pmatrix} \upsilon_1 t + \mu_1 \\ \upsilon_2 t + \mu_2 \\ \upsilon_3 t + \mu_3 \end{pmatrix},$$

where $\gamma_i, \upsilon_j \in \mathbb{Z}$ and $\mu_j, \nu_j \in [0, 1]$.

First, note that we can assume $\nu = 0$. In fact, since the translation of $\mathbb{R}^3$

$$\tau_\nu := x \to x + \nu,$$

descends to an isometry of $\mathbb{T}^3$, using Proposition 2.8 one has

$$\int_\gamma \int_\upsilon \Omega = \int_{\tau^{-1}(\gamma)} \int_{\tau^{-1}(\upsilon)} \Omega,$$

where now $(0, 0, 0)^\top$ belongs to $\tau^{-1}(\gamma)$. *In order not to burden the notation we will still denote by μ the new origin* (keeping in mind that it actually corresponds to $\mu(\gamma, \upsilon) = \mu - \nu$) of the translated curve υ.

We saw that, given an orthonormal basis of $\mathbb{R}^3$, one can build an orthonormal eigenforms basis of the Laplace operator. To simplify the computation we will perform in (3.3) we make a choice of this orthonormal basis adapted to the curve γ: the first vector is chosen to be $v^1 = [\gamma]/\|[\gamma]\|$, and we arbitrarily complete it to get an orthonormal basis:

$$\left(\left(v^1 = \frac{[\gamma]}{\|[\gamma]\|} \right)^*, v^2, v^3 \right).$$

Recall that we want to compute the following series;

$$\sum_{k>0} e^{-\lambda_k s} \int_\gamma \eta_k \int_\upsilon * \left(\frac{d\eta_k}{\lambda_k} \right). \tag{4.8}$$

We will compute all terms involved in this series separately and we will sum them in the next subsection. These terms are the product of two integrals that we compute independently.

We start with the integral involving the operator $*d$. Let η be an eigenform. One has

$$\int_\gamma \eta = \int_{[0,1]} \sqrt{2} f(2\pi \boldsymbol{k} \cdot \gamma(t))(v^i)^*([\gamma])\, dt, \tag{4.9}$$

where $f \in \{\cos, \sin\}$ and $\boldsymbol{k}$ is a vector of $\mathbb{Z}^3$. The above integral vanishes whenever f is a sinus since the curve γ passes by 0;

$$\begin{aligned}\int_\gamma \eta &= \int_{[0,1]} \sqrt{2} \sin(2\pi \boldsymbol{k} \cdot \gamma(t))(v^i)^*([\gamma])\, dt \\ &= C_1 \int_{[0,1]} \sin(2\pi C_2 t)\, dt = 0\end{aligned}$$

because $C_2 \in \mathbb{Z}$. *We can assume then that f is a cosine.* We keep computing in considering the eigenforms

$$\eta_{\boldsymbol{k},i} = \sqrt{2} \cos(2\pi(\boldsymbol{k} \cdot x))(v^i)^*$$

only, where $\boldsymbol{k} \in \mathbb{Z}^3$ and $i \in \{1, 2, 3\}$. Which, looking backward to (4.9), gives

$$\int_\gamma \eta_{\boldsymbol{k},i} = \int_{t=0}^{1} \cos(2\pi t(\boldsymbol{k} \cdot [\gamma]))(v^i)^*([\gamma])\, dt,$$

where $(v^i)^*([\gamma]) = ([\gamma] \cdot v^i) = \|[\gamma]\| \delta_{i,1}$.

The above integral therefore vanishes whenever

- $(\boldsymbol{k} \cdot [\gamma]) \neq 0$;
- $i \neq 1$.

Moreover, in the case where it does not, the function $t \mapsto \cos(2\pi(\boldsymbol{k} \cdot [\gamma])t)$ is constant, so that

$$\int_\gamma \eta_{\boldsymbol{k},i} = \sqrt{2} \|[\gamma]\|. \tag{4.10}$$

Differential forms giving a nonvanishing term of the series (4.8) are therefore

$$\eta_{\boldsymbol{k},1} = \sqrt{2} \cos(2\pi \boldsymbol{k} \cdot x) \left(\frac{[\gamma]}{\|[\gamma]\|} \right)^*,$$

with $\boldsymbol{k} \in \mathbb{Z}^3$ and $\boldsymbol{k} \cdot [\gamma] = 0$.

From now on, we will denote $\eta_{\boldsymbol{k},1}$ by $\eta_{\boldsymbol{k}}$. We now compute the second term of the series:

$$\int_\upsilon *d\eta_{\boldsymbol{k}},$$

starting off computing

$$\begin{aligned}*d\eta_{\boldsymbol{k}} &= *d\big(\sqrt{2}\cos(2\pi(x\cdot\boldsymbol{k}))(v^1)^*\big)\\ &= -2\sqrt{2}\pi\,\sin(2\pi(x\cdot\boldsymbol{k}))*\big(k_1\,dx_1\wedge(v^1)^*+k_2\,dx_2\wedge(v^1)^*+k_3\,dx_3\wedge(v^1)^*\big)\\ &= -2\sqrt{2}\pi\,\sin(2\pi(x\cdot\boldsymbol{k}))(\boldsymbol{k}\wedge v^1)^*.\end{aligned}$$

We then get

$$\begin{aligned}\int_\upsilon *d\eta_{\boldsymbol{k}} &= \int_{t=0}^1 -2\sqrt{2}\pi\,\sin\big(2\pi t([\upsilon]\cdot\boldsymbol{k})+2\pi(\mu\cdot\boldsymbol{k})\big)(\boldsymbol{k}\wedge v^1)^*([\upsilon])\,dt\\ &= -2\sqrt{2}\pi\,\det\left(\frac{[\gamma]}{\|[\gamma]\|},[\upsilon],\boldsymbol{k}\right)\int_0^1 \sin\big(2\pi t([\upsilon]\cdot\boldsymbol{k})+2\pi(\mu\cdot\boldsymbol{k})\big)\,dt.\end{aligned}$$

As before, this integral vanishes if one of these conditions holds:

- the vectors $[\gamma]$ and $[\upsilon]$ are collinear;
- $(\boldsymbol{k}\cdot[\upsilon])\neq 0$.

Moreover if $\int_\nu *d\eta_{\boldsymbol{k}}\neq 0$, we have

$$\int_\upsilon *d\eta_{\boldsymbol{k}} = -2\sqrt{2}\pi\,\det\left(\frac{[\gamma]}{\|[\gamma]\|},[\upsilon],\boldsymbol{k}\right)\sin(2\pi(\mu\cdot\boldsymbol{k})). \tag{4.11}$$

Multiplying (4.10) and (4.11) one has:

$$\int_\gamma \eta_{\boldsymbol{k}}\int_\upsilon *d\eta_{\boldsymbol{k}} = \begin{cases}4\pi\,\det([\gamma],[\upsilon],\boldsymbol{k})\sin(2\pi(\mu\cdot\boldsymbol{k})) & \text{if } \boldsymbol{k}\in\operatorname{Span}([\gamma],[\upsilon])^\perp,\\ 0 & \text{otherwise.}\end{cases}$$

This leads us to characterise elements of $\operatorname{Span}([\gamma],[\upsilon])^\perp$ with integer coefficients.

Lemma 4.12. *Let b_1 and b_2 two nonzero vectors of $\mathbb{Z}^3$. Then the group*

$$\operatorname{Span}(b_1,b_2)^\perp\cap\mathbb{Z}^3$$

is cyclic. We note by $\pm\beta$ one of these two possible generators.

Proof. As a set it is nonempty; the vector $b_1\wedge b_2$ belongs to $\mathbb{Z}^3$ and is orthogonal to both b_1 and b_2. As the intersection of two subgroups, $\mathbb{Z}^3$ and $\mathbb{R}\cdot b_1\wedge b_2$, it is a subgroup of $\mathbb{R}$. The neutral element of $\operatorname{Span}(b_1,b_2)^\perp\cap\mathbb{Z}^3$ must be isolated since $\mathbb{Z}^3$ is discrete. By characterisation of $\mathbb{R}$ subgroups, this group is cyclic. □

We apply the previous lemma to the pair $([\gamma],[\upsilon])$ to get the following description of elements $\boldsymbol{k}\in\mathbb{Z}^3$ giving a nonvanishing term in the series of (3.3):

$$\operatorname{Span}([\gamma],[\upsilon])^\perp\cap\mathbb{Z}^3 = \{k\beta,\ k\in\mathbb{Z}\}.$$

Among both possible generators, we choose β such that the family $([\gamma],[\upsilon],\beta)$ is positively oriented.

The only nonvanishing terms of the series appearing in (4.8) correspond to the differential forms

$$\eta_{(k\beta)} = \sqrt{2}\cos((k\beta)\cdot x)\left(\frac{[\gamma]}{\|\gamma\|}\right)^{*},$$

and in this case we have

$$\int_{\gamma} \eta_k \int_{\upsilon} *d\eta_k = -4\pi k \det([\gamma], [\upsilon], \beta) \sin(2\pi k(\mu\cdot\beta)).$$

From now on we denote by η_k *the differential form* $\eta_{(k\beta)}$. Recall that the differential forms η_k and η_{-k} are collinear. To get a free family of eigenforms one needs to choose the sign of the integers k: we take them nonnegative. The series of Equation (3.3) then becomes

$$-\sum_{k>0} \frac{e^{-(2\pi\|\beta\|)^2 n^2 s}}{\pi k\|\beta\|} \det\left([\gamma], [\upsilon], \frac{\beta}{\|\beta\|}\right) \sin(2\pi k(\mu\cdot\beta)). \tag{4.13}$$

Remark 4.14. As noticed in Remark 4.3, (4.13) is not independent of the parametrisations involved a priori. In fact, the point $\mu \in \mathbb{T}^3$ appearing in $\sin(2\pi k(\mu\cdot\beta))$ depends of an origin choice for υ. Let us thus check that $k(\mu\cdot\beta)$ actually doesn't, modulo $\mathbb{Z}$. Let $\mu_2 \in \upsilon$ be another origin of υ, by definition there is $t \in \mathbb{R}$ and $\alpha \in \mathbb{Z}^3$ such that

$$\mu_2 - \mu = t[\upsilon] + \alpha,$$

thus

$$(\mu_2\cdot\beta) = (\mu_2 - \mu + \mu\cdot\beta) = (\mu\cdot\beta) + (\alpha\cdot\beta),$$

since $\beta \in [\upsilon]^{\perp}$. We conclude reducing the above formula modulo $\mathbb{Z}$ to get

$$(\mu_2\cdot\beta) = (\mu\cdot\beta),$$

since $(\alpha\cdot\beta) \in \mathbb{Z}$.

4D. ***A uniformly converging family of functions.*** Let us now look into the series (4.13) more in detail. If one is able to let $t \to 0$ within all terms of this series one would get

$$-C \sum_{k>0} \frac{1}{k} \sin(2\pi kx),$$

with $C = (1/(\pi\|\beta\|)) \det([\gamma], [\upsilon], \beta/\|\beta\|)$ and $x = (\mu\cdot\beta)$.

One can recognise here the Fourier series development of the defined-on-the-circle-$\mathbb{R}/\mathbb{Z}$ function

$$x \mapsto \begin{cases} 0 & \text{if } x = 0, \\ \frac{\pi}{2}(1-2x) & \text{on } (0, 1). \end{cases} \tag{4.15}$$

So that we would have

$$\int_\gamma \int_\upsilon \Omega = \frac{1}{2\|\beta\|} \det\left([\gamma], [\upsilon], \frac{\beta}{\|\beta\|}\right)(1 - 2(\mu \cdot \beta)), \tag{4.16}$$

which is precisely what is expected. To justify the term-by-term convergence of (4.13), we use the following lemma.

Lemma 4.17. *Let $a_k(t)$ and $b_k(t)$ two sequences of functions defined on an interval I containing 0 such that*

1. $\left(\sum_{k \le n} a_k(t)\right)_{n \in \mathbb{N}}$ *is uniformly bounded with respect to t;*
2. *the sequence of functions $b_k(t)$ is nonincreasing with respect to t and converges uniformly, with respect to k, to 0.*

Then the series of functions $\sum_{k \in \mathbb{N}} a_k(t) b_k(t)$ converges uniformly on I.

We omit the proof, which consists of a discrete integration by parts of the series.

We set $a_k(t) = \sin(2\pi kx)$, $b_k(t) = e^{-atk^2}/k$ and $I = [0, +\infty]$. One can then check that for all $a \in \mathbb{R}^+$ and $x > 0$, all assumptions of Lemma 4.17 hold. We deduce that the series of functions

$$\sum_{k>0} \frac{e^{-atk^2}}{k} \sin(2\pi kx)$$

converges uniformly on $]0, +\infty]$. One is therefore allowed to switch limits and sum in (4.13) to get

$$\begin{aligned} \lim_{t \to 0} \sum_{k>0} \frac{e^{-atk^2}}{k} \sin(2\pi kx) &= \sum_{k>0} \lim_{t \to 0} \frac{e^{-atk^2}}{k} \sin(2\pi kx) \\ &= \sum_{k>0} \frac{\sin(2\pi kx)}{k}, \end{aligned}$$

which concludes the proof.

4E. ***The $\mathbb{T}^2$-geodesic flow special case.*** Particularly interesting collections of multi-geodesics of $\mathbb{T}^3$ arise as periodic orbits of the $\mathbb{T}^2$-geodesic flow. More generally, linking number of collections of periodic orbits have been studied by E. Ghys [2007] and Dehornoy [2017; 2011] for dynamical purposes. The latter showed that, for a large class of examples given by geodesic flows on surfaces, these collections all link positively, up to a choice of global orientation. This implies the existence of Birkhoff sections and, as a corollary, that periodic orbits of this flows display fibred knots. In the setting of $\mathbb{T}^2$, we will see that Theorem 4.1 specifies easily giving a new linking number formula.

We start by noticing that $\mathbb{T}^3$ is identified to the unitary tangent bundle $U\mathbb{T}^2$ of the 2-torus $\mathbb{T}^2 := \mathbb{R}^2/\mathbb{Z}^2$. In fact, the unitary tangent bundle of $\mathbb{T}^2$ is trivial, $\mathbb{T}^2$

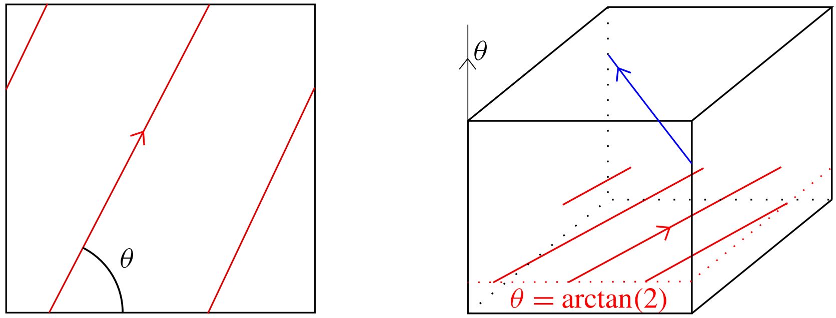


Figure 4. The red curves on the left represents a closed geodesic of $\mathbb{T}^2$. This curve lifts canonically to the red right one on the unitary tangent bundle. This lifted curve remains in the leaf $\theta = \arctan(2)$. The blue curve represents another lifted geodesic.

being a Lie group. One trivialisation consists to choose a direction of $\mathbb{R}^2$, which induces one on $\mathbb{T}^2$, from which one is able to assign an angle to any vector of $U\mathbb{T}^2$. That is to say the map

$$U\mathbb{T}^2 \to \mathbb{T}^3, \quad u \mapsto ((x, y), \theta)$$

is an actual trivialisation. With the unitary tangent bundle of a Riemannian manifold comes always a flow: the geodesic flow. In the case of $\mathbb{T}^2$, one can fully write down the flow in the trivialisation given above

$$\Phi_t : \mathbb{T}^3 \to \mathbb{T}^3, \quad (x, y, \theta) \mapsto (x + t\cos\theta, y + t\sin\theta, \theta).$$

Note that periodic orbits of a flow are naturally parametrised and oriented by the flow itself:

$$\gamma : \mathbb{S}^1 \to \mathbb{T}^3, \quad t \mapsto (x + t\cos\theta, y + t\sin\theta, \theta).$$

See Figure 4.

Remark 4.18. The fact that orbits of the geodesic flow are still geodesics on the unitary tangent bundle is more general, providing that one endows the latter with the right metric; the so called Sasaki metric. In our case, it turns out that the Sasaki metric coincides with the $\mathbb{T}^3$ flat one.

In this setting, one can readily specifies Theorem 3.2 to get:

Corollary 4.19 [Dehornoy 2011, page 11]. *Let $\Gamma = (\gamma^i)_{i\in I}$ and $\Upsilon = (\upsilon^j)_{j\in J}$ be two homologically trivial multigeodesics of $\mathbb{T}^2$. In the $\mathbb{T}^2$-unitary tangent bundle* *they link according to the following formula*

$$\mathrm{lk}(\Gamma, \Upsilon) = \sum_{i\in I,\, j\in J} \langle \gamma^i, \upsilon^j \rangle \frac{1-\theta_{i,j}/\pi}{2},$$

where $\theta_{i,j}$ *denotes the unique determination in* $[0, 2\pi[$ *of the oriented angle* θ *made at any intersections points* (see Figure 2), *and* $\langle \gamma^i, \upsilon^j \rangle$ *denotes the algebraic intersections between* γ^i *and* υ^j *on* $\mathbb{T}^2$.

Proof. As previously noticed, the orbits of this flow remain in the leaves $\theta = \mathrm{cst}$, so that the vectors $[\gamma^i]$ and $[\upsilon^j]$ belong $\mathbb{R}^2 \subset \mathbb{R}^3$. Our vector $\beta^{i,j}$ defined in Theorem 4.1 becomes

$$\beta^{i,j} = \begin{pmatrix} 0 \\ 0 \\ \pm 1 \end{pmatrix}$$

for all pairs (i, j), the sign depending whether or not the angle between the curves γ_i and υ_j is greater than π. In particular we have $\|\beta^{i,j}\| = 1$. Moreover, the determinant $\det([\gamma^i], [\upsilon^j], \beta^{i,j})$ becomes $\det_{\mathbb{R}^2}([\gamma_i], [\upsilon_j])$, which corresponds to the algebraic intersection number between γ^i and υ^j seen as curve of $\mathbb{T}^2$. To conclude, the quantity $(\beta^{i,j} \cdot \mu^{i,j})$ turns out to be interpreted as the difference between the angle made by the curve, i.e,

$$(\pi - (\mu^{i,j} \cdot \beta^{i,j})) = (\pi - (\theta_{i,j})). \qquad \square$$

Acknowledgements

I would like to thank P. Dehornoy for the many interesting conversations we had around this article and more. I also want to thank my Ph.D. advisors, G. Courtois and F. Dal'Bo, for pointing out the question as well as for their comments about the redaction of this article.

References

[Arnold and Khesin 1998] V. I. Arnold and B. A. Khesin, *Topological methods in hydrodynamics*, Appl. Math. Sci. **125**, Springer, 1998. MR Zbl

[Berline et al. 1992] N. Berline, E. Getzler, and M. Vergne, *Heat kernels and Dirac operators*, Grundlehren der Math. Wissenschaften **298**, Springer, 1992. MR Zbl

[Bott and Tu 1982] R. Bott and L. W. Tu, *Differential forms in algebraic topology*, Grad. Texts in Math. **82**, Springer, 1982. MR Zbl

[Dehornoy 2011] P. Dehornoy, *Invariants topologiques des orbites périodiques d'un champ de vecteurs*, Ph.D. thesis, Université de Lyon, 2011, Available at https://tinyurl.com/dehorphd.

[Dehornoy 2017] P. Dehornoy, "Which geodesic flows are left-handed?", *Groups Geom. Dyn.* **11**:4 (2017), 1347–1376. MR Zbl

[DeTurck and Gluck 2008] D. DeTurck and H. Gluck, "Linking integrals in the n-sphere", *Mat. Contemp.* **34** (2008), 239–249. MR Zbl

[DeTurck and Gluck 2013] D. DeTurck and H. Gluck, "Linking, twisting, writhing, and helicity on the 3-sphere and in hyperbolic 3-space", *J. Differential Geom.* **94**:1 (2013), 87–128. MR Zbl

[DeTurck et al. 2008] D. DeTurck, H. Gluck, R. Komendarczyk, P. Melvin, C. Shonkwiler, and D. S. Vela-Vick, "Triple linking numbers, ambiguous Hopf invariants and integral formulas for three-component links", *Mat. Contemp.* **34** (2008), 251–283. MR Zbl

[DeTurck et al. 2013] D. DeTurck, H. Gluck, R. Komendarczyk, P. Melvin, H. Nuchi, C. Shonkwiler, and D. S. Vela-Vick, "Generalized Gauss maps and integrals for three-component links: toward higher helicities for magnetic fields and fluid flows, II", *Algebr. Geom. Topol.* **13**:5 (2013), 2897–2923. MR Zbl

[Ghys 2007] É. Ghys, "Knots and dynamics", pp. 247–277 in *International Congress of Mathematicians, I*, edited by M. Sanz-Solé et al., Eur. Math. Soc., Zürich, 2007. MR Zbl

[Kotschick and Vogel 2003] D. Kotschick and T. Vogel, "Linking numbers of measured foliations", *Ergodic Theory Dynam. Systems* **23**:2 (2003), 541–558. MR Zbl

[Moffatt 1969] H. K. Moffatt, "The degree of knottedness of tangled vortex lines", *J. Fluid Mech.* **35**:1 (1969), 117–129. Correction in **830** (2017), 821–822. Zbl

[Moreau 1961] J.-J. Moreau, "Constantes d'un îlot tourbillonnaire en fluide parfait barotrope", *C. R. Acad. Sci. Paris* **252** (1961), 2810–2812. MR Zbl

[de Rham 1984] G. de Rham, *Differentiable manifolds: forms, currents, harmonic forms*, Grundlehren der Math. Wissenschaften **266**, Springer, 1984. MR Zbl

[Ricca and Nipoti 2011] R. L. Ricca and B. Nipoti, "Gauss' linking number revisited", *J. Knot Theory Ramifications* **20**:10 (2011), 1325–1343. MR Zbl

[Rosenberg 1997] S. Rosenberg, *The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds*, Lond. Math. Soc. Student Texts **31**, Cambridge Univ. Press, 1997. MR Zbl

[Vogel 2003] T. Vogel, "On the asymptotic linking number", *Proc. Amer. Math. Soc.* **131**:7 (2003), 2289–2297. MR Zbl

[Woltjer 1958] L. Woltjer, "A theorem on force-free magnetic fields", *Proc. Nat. Acad. Sci. U.S.A.* **44** (1958), 489–491. MR Zbl

Received February 4, 2019. Revised February 5, 2020.

ADRIEN BOULANGER
INSTITUT MATHÉMATIQUE DE MARSEILLE
FRANCE
adrien.boulanger@univ-amu.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.283

CLUSTER AUTOMORPHISM GROUPS AND AUTOMORPHISM GROUPS OF EXCHANGE GRAPHS

WEN CHANG AND BIN ZHU

For a coefficient-free cluster algebra $\mathcal{A}$, we study the cluster automorphism group $\mathrm{Aut}(\mathcal{A})$ and the automorphism group $\mathrm{Aut}(E_{\mathcal{A}})$ of its exchange graph $E_{\mathcal{A}}$. We show that these two groups are isomorphic with each other, if $\mathcal{A}$ is of finite type excepting types of rank 2 and type F_4, or if $\mathcal{A}$ is of skew-symmetric finite mutation type.

1. Introduction

Cluster algebras were introduced by Sergey Fomin and Andrei Zelevinsky [2002]. In this paper we consider cluster algebras with trivial coefficients, which can be defined through a skew-symmetrizable square matrix. Such a cluster algebra is a $\mathbb{Z}$-subalgebra of a rational function field with n indeterminates. More precisely, a *seed* is a pair consisting of a set (*cluster*) of n indeterminates (*cluster variables*) in the field and a skew-symmetrizable square matrix (*exchange matrix*) of size n. Starting from an initial seed, we get a new seed by an operation called mutation. Then the cluster algebra is algebraic-generated by all the cluster variables obtained by iterated mutations. The cluster algebra has nice combinatorial structures which are (in some sense) given by mutations, and these structures are captured by its exchange graph, which is a graph with seeds as vertices and with mutations as edges.

We focus in this paper on two special types of cluster algebras: the *finite type* and the *finite mutation type*. Cluster algebras of finite type are those algebras with a finite number of clusters. They are classified in [Fomin and Zelevinsky 2003a], which corresponds to the Killing–Cartan classification of complex semisimple Lie algebras, or, equivalently, corresponds to the classification of root systems in Euclidean space. If there are finitely many matrix classes in the seeds of a cluster algebra, then we say it is of finite mutation type, where two matrices are in the same class if one of them can be obtained from the other by simultaneous relabeling of the rows and columns. The cluster algebras of finite mutation type with skew-symmetric

Wen Chang was supported by Shaanxi Province, Shaanxi Normal University and the NSF of China (Grant 11601295) and Bin Zhu was supported by the NSF of China (Grant 11671221).
MSC2010: primary 13F60; secondary 05E40.
Keywords: cluster algebra, cluster automorphism group, exchange graph.

exchange matrices are classified in [Felikson et al. 2012b]; a large class of them arises from marked Riemann surfaces (possibly with boundary) [Fomin et al. 2008], and there are 11 exceptional ones. The classification of skew-symmetrizable cluster algebras of finite mutation type is given in [Felikson et al. 2012a] via operations called unfoldings upon the skew-symmetric cluster algebras of finite mutation type.

We consider the relations in this paper between two groups associated to the cluster algebras. One is the cluster automorphism group consisting of *cluster automorphisms*, which are permutations of the clusters that commute with mutations. This group is introduced in [Assem et al. 2012] for a coefficient-free cluster algebra, and in [Chang and Zhu 2016b] for a cluster algebra with coefficients, it reveals the combinatorial and algebraic symmetries of the cluster algebra. Another is the automorphism group of the exchange graph, which consists of graph automorphism of the exchange graph. This group describes the symmetries of the exchange graph; in other words, it describes combinatorial symmetries of the cluster algebra. The problem that considers the relations between these two groups is stated in [Saleh 2014].

The exchange graph is a fairly coarse invariant of a cluster algebra, e.g., all infinite type cluster algebras of rank 2 have the same exchange graph. This article suggests that, nonetheless, the exchange graph is already rich enough to capture most of the symmetries of the cluster algebra.

For a coefficient-free cluster algebra $\mathcal{A}$ with exchange graph $E_{\mathcal{A}}$, we write the cluster automorphism group of $\mathcal{A}$ and the automorphism group of $E_{\mathcal{A}}$ as $\mathrm{Aut}(\mathcal{A})$ and $\mathrm{Aut}(E_{\mathcal{A}})$, respectively. In general, $\mathrm{Aut}(\mathcal{A})$ is a subgroup of $\mathrm{Aut}(E_{\mathcal{A}})$, and may be a proper subgroup; see Examples 3.3 and 3.5. The main result of this paper is that these two groups are isomorphic with each other if $\mathcal{A}$ is of finite type, excepting types of rank two and type F_4 (Theorem 3.16), or $\mathcal{A}$ is of skew-symmetric finite mutation type (Theorem 3.18). Therefore in some degree, for these cluster algebras, the algebraic symmetries are also captured by the exchange graphs. In particular, we compute the automorphism group of the exchange graph of a finite type cluster algebra in Table 1; see Remark 3.17.

To prove these results, we describe $E_{\mathcal{A}}$ more precisely. In Section 3A, we define layers of geodesic loops of $E_{\mathcal{A}}$ by using the distance of a vertex to a fixed vertex on $E_{\mathcal{A}}$. An easy observation is that an isomorphism of exchange graphs should maintain the combinatorial numbers of the layers of geodesic loops based on the corresponding vertices; see Remark 3.2(4). By this observation, we directly show in Examples 3.6, 3.7, 3.10 and 3.11 that for a cluster algebra of type A_3, B_3, C_3, $\tilde{A}_2$ or T_3 (the cluster algebra from a once-punctured torus), we have $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$. For the general cases we reduce them to above five cases (Theorems 3.16 and 3.18).

The paper is organized as follows: we recall preliminaries on cluster algebras, cluster algebras of finite mutation type and cluster automorphisms in Section 2, then we prove the main theorems in Section 3.

Dynkin type	automorphism group $\mathrm{Aut}(E_{\mathcal{A}})$
$A_n (n \geqslant 2)$	$\mathbb{D}_{n+3}$
B_2	$\mathbb{D}_6$
$B_n (n \geqslant 3)$	$\mathbb{D}n + 1$
C_2	$\mathbb{D}_6$
$C_n (n \geqslant 3)$	$\mathbb{D}_{n+1}$
D_4	$\mathbb{D}_4 \times S_3$
$D_n (n \geqslant 5)$	$\mathbb{Z}_2$
E_6	$\mathbb{D}_{14}$
E_7	$\mathbb{D}_{10}$
E_8	$\mathbb{D}_{16}$
F_4	$\mathbb{D}_7 \rtimes \mathbb{Z}_2$
G_2	$\mathbb{D}_8$

Table 1. Automorphism groups of exchange graphs of cluster algebras of finite type.

2. Preliminaries

2A. *Cluster algebras.*

Definition 2.1. [Fomin and Zelevinsky 2002] (labeled seeds). A *labeled seed* is a pair $\Sigma = (\boldsymbol{x}, B)$, where

- $\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}$ is an ordered set of n indeterminates;
- $B = (b_{x_j x_i})_{n\times n} \in M_{n\times n}(\mathbb{Z})$ is a skew-symmetrizable matrix labeled by $\boldsymbol{x} \times \boldsymbol{x}$; that is, there exists a diagonal matrix D with positive integer entries such that DB is skew-symmetric.

The set $\boldsymbol{x}$ is called the *cluster* with elements the *cluster variables*, and B is called the *exchange matrix*. An element $b_{x_j x_i}$ in B is also written as b_{ji} for brevity. We assume throughout the paper that B is indecomposable; that is, for any $1 \leqslant i, j \leqslant n$, there is a sequence $i_0 = i, i_1, \dots, i_m, i_{m+1} = j$, such that $b_{i_k, i_{k+1}} \neq 0$ for any $0 \leqslant k \leqslant m$. We also assume that $n > 1$ for convenience. One may produce a new labeled seed by a mutation at direction k for any cluster variable x_k.

Definition 2.2. [Fomin and Zelevinsky 2002] (seed mutations). The labeled seed $\mu_k(\Sigma) = (\mu_k(\boldsymbol{x}), \mu_k(B))$ obtained by the *mutation* of Σ in the direction k is given by:

- $\mu_k(\boldsymbol{x}) = (\boldsymbol{x} \setminus \{x_k\}) \sqcup \{\mu_{x_k,\boldsymbol{x}}(x_k)\}$ where

$$x_k \mu_{x_k,\boldsymbol{x}}(x_k) = \prod_{\substack{1\leqslant j\leqslant n;\\ b_{jk}>0}} {x_j}^{b_{jk}} + \prod_{\substack{1\leqslant j\leqslant n;\\ b_{jk}<0}} {x_j}^{-b_{jk}}.$$

- $\mu_k(B) = (b'_{ji})_{n\times n} \in M_{n\times n}(\mathbb{Z})$ is given by

$$b'_{ji} = \begin{cases} -b_{ji} & \text{if } i = k \text{ or } j = k; \\ b_{ji} + \frac{1}{2}(|b_{ji}|b_{ik} + b_{ji}|b_{ik}|) & \text{otherwise.} \end{cases}$$

It is easy to check that a mutation is an involution; that is, $\mu_k\mu_k(\Sigma) = \Sigma$.

Definition 2.3. [Fomin and Zelevinsky 2007] (n-cluster patterns). An n-regular tree $\mathbb{T}_n$ is a diagram, whose edges are labeled by $1, 2, \dots, n$, such that the n edges emanating from each vertex receive different labels. A *n-cluster pattern* is an assignment of a labeled seed $\Sigma_t = (\boldsymbol{x}_t, B_t)$ to every vertex $t \in \mathbb{T}_n$, so that the labeled seeds assigned to the endpoints of any edge labeled by k are obtained from each other by the seed mutation in direction k. The elements of Σ_t are written as follows:

$$\boldsymbol{x}_t = (x_{1;t}, \dots, x_{n;t}), \quad B_t = (b^t_{ij}). \tag{1}$$

Note that $\mathbb{T}_n$ is in fact determined by any fixed labeled seed on it. Now we are ready to define cluster algebras.

Definition 2.4. [Fomin and Zelevinsky 2007] (cluster algebras). Given a seed Σ and a cluster pattern $\mathbb{T}_n$ associated to it, we denote

$$\mathscr{X} = \bigcup_{t\in\mathbb{T}_n} \boldsymbol{x}_t = \{x_{i,t} : t \in \mathbb{T}_n,\ 1 \le i \le n\}, \tag{2}$$

the union of clusters of all the seeds in the pattern. We call the elements $x_{i,t} \in \mathscr{X}$ the *cluster variables*. The *cluster algebra* $\mathcal{A}$ associated with Σ is the $\mathbb{Z}$-subalgebra of the rational function field $\mathcal{F} = \mathbb{Q}(x_1, x_2, \dots, x_n)$, generated by all cluster variables, $\mathcal{A} = \mathbb{Z}[\mathscr{X}]$.

To a skew-symmetrizable matrix $B = (b_{ji})_{n\times n}$, one can associate a *valued quiver* (quiver for brevity) $Q = (Q_0, Q_1, \upsilon)$ as follows: $Q_0 = \{1, 2, \dots, n\}$ is a set of vertices. For any two vertices j and i, if $b_{ji} > 0$, then there is an arrow α from j to i to which we assign a pair of values $(\upsilon_1(\alpha), \upsilon_2(\alpha)) = (b_{ji}, -b_{ij})$. These arrows form the set Q_1. Since B is an indecomposable skew-symmetrizable matrix, the defined valued quiver Q is connected and there are no loops nor 2-cycles in Q. Then we can define a mutation of the valued quiver by the mutation of the matrix; we refer to [Fomin and Zelevinsky 2002; Keller 2012] for details. We say two quivers Q and Q' are *mutation equivalent* if the corresponding matrices are mutation equivalent; that is, one of them can be obtained from the other one by a finite sequence of mutations. We also write $(\boldsymbol{x}, Q)$ for the labeled seed $(\boldsymbol{x}, B)$, and write $\mathcal{A}_Q$ for the cluster algebra defined by Σ. The quiver and the defined cluster algebra are called skew-symmetric if the corresponding matrix is skew-symmetric.

If the cluster algebra is of finite type [Fomin and Zelevinsky 2003a] or of skew-symmetric type, then the cluster determines the quiver [Gekhtman et al. 2008], and we denote the quiver of a cluster $\boldsymbol{x}$ by $Q(\boldsymbol{x})$.

Example 2.5. Let B be the following skew-symmetrizable matrix with skew-symmetrizer $D = \mathrm{diag}\{2, 2, 1, 1\}$:

$$B = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & -2 & 0 \end{pmatrix}.$$

The quiver corresponding to B is Q, where we always delete the trivial pairs of values $(1, 1)$, and replace a arrow assigning pair (m, m) by m arrows:

$$Q : 1 \longrightarrow 2 \xleftarrow{(2,1)} 3 \rightrightarrows 4.$$

Definition 2.6. [Fomin and Zelevinsky 2007] (seeds). Given two labeled seeds $\Sigma = (\boldsymbol{x}, B)$ and $\Sigma' = (\boldsymbol{x}', B')$, we say that they define the same *seed* if Σ' is obtained from Σ by simultaneous relabeling of the sets $\boldsymbol{x}$ and the corresponding relabeling of the rows and columns of B.

We denote by $[\Sigma]$ the seed represented by a labeled seed Σ. The cluster $\boldsymbol{x}$ of a seed $[\Sigma]$ is an unordered n-element set. For any $x \in \boldsymbol{x}$, there is a well-defined mutation $\mu_x([\Sigma]) = [\mu_k(\Sigma)]$ of $[\Sigma]$ at direction x, where $x = x_k$. For two same rank skew-symmetrizable matrices B and B', we say $B \cong B'$ if B' is obtained from B by simultaneous relabeling of the rows and columns of B. Then the exchange matrices in any two labeled seeds representing a same seed are isomorphic. The isomorphism of two exchange matrices induces an isomorphism of corresponding quivers. For convenience, in the rest of the paper, we also denote by Σ the seed $[\Sigma]$ represented by Σ.

Definition 2.7. [Fomin and Zelevinsky 2007] (exchange graphs). The *exchange graph* of a cluster algebra is the n-regular graph whose vertices are the seeds of the cluster algebra and whose edges connect the seeds related by a single mutation. We denote by $E_{\mathcal{A}}$ the exchange graph of a cluster algebra $\mathcal{A}$.

Clearly, the exchange graph of a cluster algebra is a quotient graph of the n-regular tree; its vertices are equivalent classes of labeled seeds. The exchange graph need not be a finite graph; if it is finite, then we say the corresponding cluster algebra (and its cluster pattern) are of *finite type*.

Definition 2.8. [Fomin and Zelevinsky 2003a, page 70] (cluster complexes). A cluster complex Δ of $\mathcal{A}$ is a simplicial complex on the ground set $\mathscr{X}$ with the clusters as the maximal simplices.

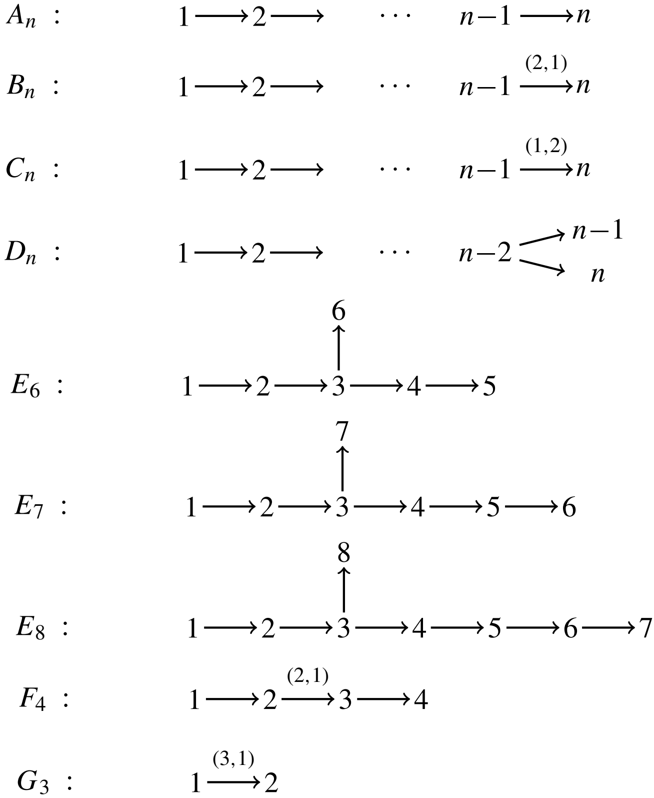
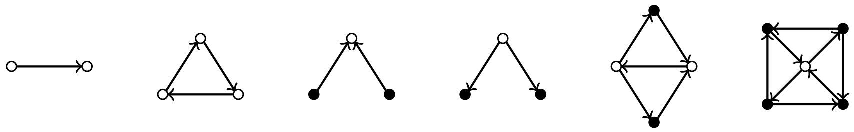


Figure 1. Quivers of finite type.

Then Δ is an n-dimensional complex. In particular, if $\mathcal{A}$ is of finite type or skew-symmetric, then the vertices of $E_{\mathcal{A}}$ are clusters, so the dual graph of Δ is $E_{\mathcal{A}}$.

2B. ***Finite types and finite mutation types.*** By the classification of cluster algebras of finite type [Fomin and Zelevinsky 2003a], a cluster algebra is of finite type if and only if there is a seed whose quiver is one of the quivers depicted in Figure 1. Note that the underlying graphs of quivers in Figure 1 are trees, thus any two quivers with the same underlying graph are mutation-equivalent.

Definition 2.9. [Fomin et al. 2008; Felikson et al. 2012b] A *block* is a quiver isomorphic to one of the quivers with black or white colored vertices shown in Figure 2. Vertices marked in white are called *outlets*. A connected quiver Q is called *block-decomposable* (*decomposable* for brevity) if it can be obtained from a collection of blocks by identifying outlets of different blocks along some partial matching (matching of outlets of the same block is not allowed), where two arrows with the same endpoints and opposite directions cancel out. If Q is not block-decomposable then we call Q *nondecomposable*.



I II IIIa IIIb IV V

Figure 2. Blocks. Outlets are colored white, dead ends are black.

It is proved in [Fomin et al. 2008, Theorem 13.3] that a quiver is decomposable if and only if it is a quiver of a triangulation of an oriented marked Riemann surface, and thus a quiver mutation equivalent to a decomposable quiver is also decomposable. Note that all arrow multiplicities of a decomposable quiver are 1 or 2. Therefore decomposable quivers are mutation finite. It is clear that a quiver of rank 2, that is, a quiver with two vertices, is mutation finite. Besides these two kinds of quivers, there are exactly 11 exceptional skew-symmetric quivers of finite mutation type; see Theorem 6.1 in [Felikson et al. 2012b]. We list the exceptional quivers in Figure 3.

2C. ***Automorphism groups.*** In this section, we recall the cluster automorphism group [Assem et al. 2012] of a cluster algebra, and the automorphism group of the corresponding exchange graph [Chang and Zhu 2016b].

Definition 2.10. [Assem et al. 2012] *(*cluster automorphisms). For a cluster algebra $\mathcal{A}$ and a $\mathbb{Z}$-algebra automorphism $f : \mathcal{A} \to \mathcal{A}$, we call f a *cluster automorphism* if there exists a labeled seed $(\boldsymbol{x}, B)$ of $\mathcal{A}$ such that the following conditions are satisfied:

(1) $f(\boldsymbol{x})$ is a cluster.

(2) f is compatible with mutations; that is, for every $x \in \boldsymbol{x}$ and $y \in \boldsymbol{x}$, we have

$$f(\mu_{x,\boldsymbol{x}}(y)) = \mu_{f(x),f(\boldsymbol{x})}(f(y)).$$

Then a cluster automorphism maps a labeled seed $\Sigma = (\boldsymbol{x}, B)$ to a labeled seed $\Sigma' = (\boldsymbol{x}', B')$. Under our assumption that B is indecomposable, we have the following:

Lemma 2.11 [Assem et al. 2012]. *A $\mathbb{Z}$-algebra automorphism $f : \mathcal{A} \to \mathcal{A}$ is a cluster automorphism if and only if there exists a labeled seed $\Sigma = (\boldsymbol{x}, B)$ of $\mathcal{A}$, such that $f(\boldsymbol{x})$ is the cluster in a labeled seed $\Sigma' = (\boldsymbol{x}', B')$ of $\mathcal{A}$ with $B' = B$ or $B' = -B$.*

We call the cluster automorphism such that $B = B'$ (resp. $B = -B'$) a *direct cluster automorphism* (resp. an *inverse cluster automorphism*). Clearly, all the cluster automorphisms of a cluster algebra $\mathcal{A}$ form a group with homomorphism

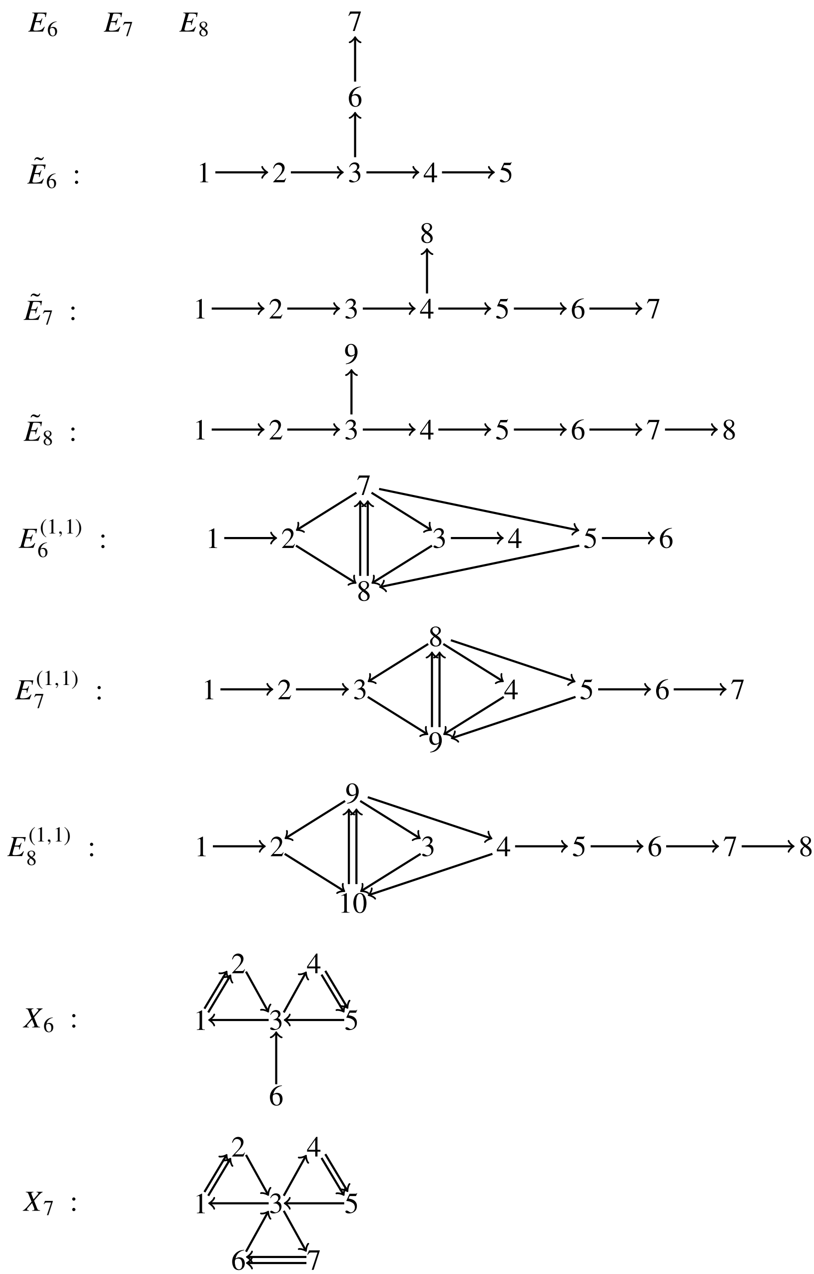


Figure 3. Representatives of nondecomposable quivers of finite mutation type.

composition as multiplication. We call this group the *cluster automorphism group* of $\mathcal{A}$, and denote it by $\mathrm{Aut}(\mathcal{A})$. We call the group $\mathrm{Aut}^+(\mathcal{A})$ consisting of the direct cluster automorphisms of $\mathcal{A}$ the *direct cluster automorphism group* of $\mathcal{A}$, which is a subgroup of Aut($\mathcal{A}$) with index at most two; see [Assem et al. 2012].

Definition 2.12. [Saleh 2014; Chang and Zhu 2016b] (automorphism of exchange graphs). An automorphism of the exchange graph $E_{\mathcal{A}}$ of a cluster algebra $\mathcal{A}$ is an automorphism of $E_{\mathcal{A}}$ as a graph, that is, a permutation σ of the vertex set, such that the pair of vertices (u, v) forms an edge if and only if the pair $(\sigma(u), \sigma(v))$ also forms an edge.

Clearly, the natural composition of two automorphisms of $E_{\mathcal{A}}$ is again an automorphism. We define an *automorphism group* $\mathrm{Aut}(E_{\mathcal{A}})$ of $E_{\mathcal{A}}$ as a group consisting of automorphisms of $E_{\mathcal{A}}$. It is clear that a cluster automorphism induces a unique automorphism of the exchange graph. Thus Aut($\mathcal{A}$) is a subgroup of $\mathrm{Aut}(E_{\mathcal{A}})$; see [Chang and Zhu 2016b]. By the definition, an automorphism σ of an exchange graph maps clusters to clusters, and induces an automorphism of its dual graph, the cluster complex Δ; we denote this automorphism by σ_Δ. Then σ_Δ is a permutation of cluster variables in $\mathscr{X}$, which maps a maximal simplex to a maximal simplex, but the map may not be compatible with the algebra relations among cluster variables in $\mathcal{A}$, thus it is not necessarily a cluster automorphism. In fact, Aut($\mathcal{A}$) may be a proper subgroup of $\mathrm{Aut}(E_{\mathcal{A}})$; see Examples 3.3 and 3.5. The following lemma can be viewed as a description of Aut($\mathcal{A}$) as a subgroup of $\mathrm{Aut}(E_{\mathcal{A}})$, as those exchange graph automorphisms which happen to preserve B-matrices (perhaps up to global reversal of sign) up to simultaneously relabeling of the rows and columns. In this point of view, the main thrust of this paper is to show that, typically for the cluster algebras we consider, any graph automorphism has the property of preserving B-matrices.

Lemma 2.13. *Let $\Phi : E_{\mathcal{A}} \to E_{\mathcal{A}}$ be an automorphism which maps a seed $\Sigma = (\boldsymbol{x}, B)$ to a seed $\Sigma' = (\boldsymbol{x}', B')$. If $B \cong B'$ or $B \cong -B'$ under the correspondence $\boldsymbol{x} \to \boldsymbol{x}'$, then the map $\boldsymbol{x} \to \boldsymbol{x}'$ induces a cluster automorphism Ψ of $\mathcal{A}$ and the induced automorphism $\Psi_E : E_{\mathcal{A}} \to E_{\mathcal{A}}$ coincides with Φ.*

Proof. Since $B \cong B'$ or $B \cong -B'$, the map $\boldsymbol{x} \to \boldsymbol{x}'$ induces a cluster automorphism Ψ of $\mathcal{A}$ by Lemma 2.11. Notice that $\Phi(\boldsymbol{x}) = \Psi(\boldsymbol{x})$; then by inductions on the mutations, we have $\Phi = \Psi$ on each cluster of $E_{\mathcal{A}}$, so $\Phi = \Psi_E$ as automorphisms of the exchange graph $E_{\mathcal{A}}$. □

3. Automorphism groups of exchange graphs

In this section we consider relations between the groups Aut($\mathcal{A}$) and $\mathrm{Aut}(E_{\mathcal{A}})$ for a cluster algebra $\mathcal{A}$ of finite type or of skew-symmetric finite mutation type. For this, we need to describe $E_{\mathcal{A}}$ more precisely. In the following we will recall the basic structures of $E_{\mathcal{A}}$ from [Fomin and Zelevinsky 2002; 2003a], and then introduce layers of geodesic loops on $E_{\mathcal{A}}$.

3A. ***Layers of geodesic loops.*** Let $\Sigma = (\boldsymbol{x}, B)$ be a labeled seed on the cluster pattern of $\mathcal{A}$. Let $\boldsymbol{x}'$ be a proper subset of $\boldsymbol{x}$, then $\boldsymbol{x}'$ is a nonmaximal simplex in the cluster complex Δ. We denote by $\Delta_{\boldsymbol{x}'}$ the *link* of $\boldsymbol{x} \setminus \boldsymbol{x}'$, which is the simplicial complex on the ground set

$$\mathscr{X}_{\boldsymbol{x}'} = \{\alpha \in \mathscr{X} - (\boldsymbol{x} \setminus \boldsymbol{x}') : (\boldsymbol{x} \setminus \boldsymbol{x}') \cup \{\alpha\} \in \Delta\},$$

such that $\boldsymbol{x}''$ is a simplex in $\Delta_{\boldsymbol{x}'}$ if and only if $\boldsymbol{x} \setminus \boldsymbol{x}' \cup \boldsymbol{x}''$ is a simplex in Δ. Let $\Gamma_{\boldsymbol{x}'}$ be the dual graph of $\Delta_{\boldsymbol{x}'}$. We view $\Gamma_{\boldsymbol{x}'}$ as a subgraph of $E_{\mathcal{A}}$ whose vertices are the maximal simplices in Δ that contain $\boldsymbol{x} \setminus \boldsymbol{x}'$. In fact, as we explain now, $\Gamma_{\boldsymbol{x}'}$ is the exchange graph of a cluster algebra $\mathcal{A}_f$ defined by a frozen seed

$$\Sigma_f = (\boldsymbol{x}', \boldsymbol{x} \setminus \boldsymbol{x}', B_f),$$

which is the freezing of Σ at $\boldsymbol{x} \setminus \boldsymbol{x}'$ (see [Chang and Zhu 2016c, Definition 2.25]), where B_f is obtained from B by deleting the columns labeled by variables in $\boldsymbol{x} \setminus \boldsymbol{x}'$. Then elements in $\boldsymbol{x} \setminus \boldsymbol{x}'$ are coefficients of $\mathcal{A}_f$ (we refer to [Fomin and Zelevinsky 2002; 2007] for a cluster algebra with coefficients). Let $\mathcal{A}'$ be a cluster algebra defined by a seed $\Sigma' = (\boldsymbol{x}', B')$, where B' is obtained from B by deleting rows and columns labeled by variables in $\boldsymbol{x} \setminus \boldsymbol{x}'$. In our setting, that is, where cluster algebras are of finite type or of skew-symmetric finite type, the exchange graph of a cluster algebra (with coefficients) only depends on the principal part of the exchange matrix (see [Fomin and Zelevinsky 2003a; Cerulli Irelli et al. 2013]) which is the submatrix labeled by $\boldsymbol{x} \setminus \boldsymbol{x}' \times \boldsymbol{x} \setminus \boldsymbol{x}'$; thus the graph $\Gamma_{\boldsymbol{x}'}$ coincides with the exchange graph $E_{\mathcal{A}'}$.

For a 2-dimensional subcomplex $\boldsymbol{x}'$ of Δ, we call the dual graph $\Gamma_{\boldsymbol{x}'}$ a *geodesic loop* of $E_{\mathcal{A}}$. We mention that the definition of geodesic loop is slightly different from the definition used in [Fomin and Zelevinsky 2003a], where a line is not a geodesic loop. If $\mathcal{A}$ is of finite type, then $E_{\mathcal{A}}$ is a finite graph, and $\Gamma_{\boldsymbol{x}'}$ is a polygon. Notice that in the seed $\Sigma' = (\boldsymbol{x}', B')$ constructed above, B' is of Dynkin type, that is, one of types A_2, B_2, C_2 or G_2. Therefore $\Gamma_{\boldsymbol{x}'}$ is a $(h+2)$-polygon, where h is the Coxeter number of the corresponding Dynkin type; see [Fomin and Zelevinsky 2003a]. If $\mathcal{A}$ is of finite mutation type, then $\Gamma_{\boldsymbol{x}'}$ may be a line. We fix a basepoint $\Sigma = (\boldsymbol{x}, B)$ and introduce the following concept.

Definition 3.1. (1) Letting Σ' be a point of $E_{\mathcal{A}}$, the *distance* $\ell(\Sigma, \Sigma')$ between Σ and Σ' is the minimal length of paths between Σ and Σ'.

(2) Letting L be a geodesic loop of $E_{\mathcal{A}}$, the *distance* $\ell_{\Sigma}(L)$ between Σ and L is the minimal length $\min\{\ell(\Sigma, \Sigma'), \Sigma' \in L\}$.

(3) Letting $m \in \mathbb{Z}_{\geqslant 0}$ be a nonnegative integer, denote by ℓ_{Σ}^{m} the set of geodesic loops whose distance to Σ is m. We call it the *m-layer* of geodesic loops of $E_{\mathcal{A}}$ based on Σ.

(4) For any $m \in \mathbb{Z}_{\geqslant 0}$, denote by $N(\ell_{\Sigma}^{m})$ the set of amounts of edges belonging to geodesic loops in the *m-layer* ℓ_{Σ}^{m}.

Remark 3.2. The following observations are directly derived from the definitions:

(1) The elements in ℓ^0_Σ are those geodesic loops $\Gamma_{\boldsymbol{x}'}$ for the 2-dimensional subcomplex $\boldsymbol{x}'$ of Δ, where $\boldsymbol{x}'$ is a subset of the cluster $\boldsymbol{x}$ in Σ.

(2) For $m_1 \neq m_2$, $\ell^{m_1}_\Sigma \cap \ell^{m_2}_\Sigma = \varnothing$.

(3) The disjoint union $\bigsqcup_{m\geqslant 0} \ell^m_\Sigma$ is the set of all the geodesic loops of $E_{\mathcal{A}}$.

(4) If $\sigma : E_{\mathcal{A}} \to E_{\mathcal{A}'}$ is an isomorphism of graphs such that the image of Σ is Σ', then for every $m \in \mathbb{Z}_{\geqslant 0}$, $N(\ell^m_\Sigma) = N(\ell^m_{\Sigma'})$ as sets.

3B. ***Cases of rank 2 and rank 3.*** In this subsection, we consider the relations between $\mathrm{Aut}(\mathcal{A})$ and $\mathrm{Aut}(E_{\mathcal{A}})$ for a cluster algebra $\mathcal{A}$ of rank 2 or rank 3.

Example 3.3. For a finite type cluster algebra $\mathcal{A}$ of rank 2, that is, one of types A_2, B_2, C_2 or G_2, its exchange graph $E_{\mathcal{A}}$ is a $(h+2)$-polygon, thus $\mathrm{Aut}(E_{\mathcal{A}})$ is isomorphic to the dihedral group $\mathbb{D}_{h+2}$, where h is the Coxeter number. If $\mathcal{A}$ is of type A_2, then $\mathrm{Aut}(\mathcal{A}) \cong \mathbb{D}_5$ [Assem et al. 2012], thus $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$. If $\mathcal{A}$ is of type B_2, C_2 or G_2, [Chang and Zhu 2016a, Theorem 3.5] shows that $\mathrm{Aut}(\mathcal{A}) \cong \mathbb{D}_{(h+2)/2}$, thus $\mathrm{Aut}(\mathcal{A}) \subsetneq \mathrm{Aut}(E_{\mathcal{A}})$.

Example 3.4. For an infinite type skew-symmetric cluster algebra $\mathcal{A}$ of rank 2, its exchange graph $E_{\mathcal{A}}$ is a line, thus $\mathrm{Aut}(E_{\mathcal{A}}) = \langle s\rangle \times \langle r\rangle \cong \mathbb{Z} \rtimes \mathbb{Z}_2 = \mathbb{D}_\infty$, where s is a left shift of $E_{\mathcal{A}}$ which maps a cluster to the left adjacent cluster and r is a reflection with respect to a fixed cluster. Then s corresponds to a direct cluster automorphism of $\mathcal{A}$ and r corresponds to an inverse cluster automorphism of $\mathcal{A}$; thus by Lemma 2.13, $\mathrm{Aut}(E_{\mathcal{A}}) \subseteq \mathrm{Aut}(\mathcal{A})$. Therefore $\mathrm{Aut}(E_{\mathcal{A}}) \cong \mathrm{Aut}(\mathcal{A}) \cong \mathbb{D}_\infty$.

Example 3.5. For an infinite type non-skew-symmetric cluster algebra $\mathcal{A}$ of rank 2, its exchange graph $E_{\mathcal{A}}$ is also a line, thus as shown in Example 3.4, $\mathrm{Aut}(E_{\mathcal{A}}) = \langle s\rangle \times \langle r\rangle \cong \mathbb{Z} \rtimes \mathbb{Z}_2$, where s corresponds to a direct cluster automorphism of $\mathcal{A}$, while r dose not correspond to any cluster automorphism of $\mathcal{A}$, since there is no nontrivial symmetry of the quiver in any seed of $\mathcal{A}$. Thus $\mathrm{Aut}(\mathcal{A}) \cong \mathbb{Z} \subsetneq \mathrm{Aut}(E_{\mathcal{A}})$.

Example 3.6. We consider the cluster algebra $\mathcal{A}$ of type A_3 with an initial labeled seed $\Sigma_0 = (\{x_1, x_2, x_3\}, Q)$, where Q is $1 \to 2 \leftarrow 3$. Its exchange graph $E_{\mathcal{A}}$ is depicted in Figure 4. Note that there are three quadrilaterals and six pentagons in $E_{\mathcal{A}}$. Then as shown in [Chang and Zhu 2016a, Example 3], $\mathrm{Aut}(\mathcal{A}) = \langle f_-, f_+\rangle \cong \mathbb{D}_6$, where f_- is defined by

$$f_- : \begin{cases} x_1 \mapsto x_1, \\ x_2 \mapsto \mu_2(x_2), \\ x_3 \mapsto x_3. \end{cases} \tag{3}$$

It maps Σ_0 to Σ_1, and induces a reflection with respect to the horizontal central

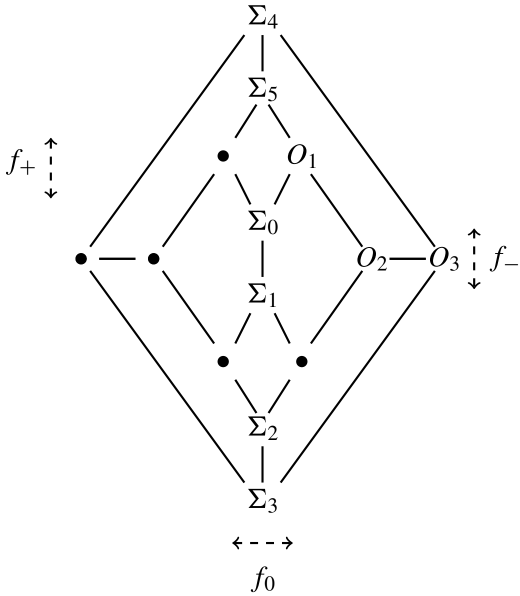


Figure 4. The exchange graph of a cluster algebra of type A_3.

axis of $E_{\mathcal{A}}$. The cluster automorphism f_+ is defined by

$$f_+ : \begin{cases} x_1 \mapsto \mu_1(x_1), \\ x_2 \mapsto x_2, \\ x_3 \mapsto \mu_3(x_3). \end{cases} \tag{4}$$

It gives a reflection on $E_{\mathcal{A}}$, which maps Σ_0 to Σ_5. In fact, as shown in [Chang and Zhu 2016a], a direct cluster automorphism of $\mathcal{A}$ is of the form $(f_+f_-)^m$, $0 \le m \le 5$, which induces a rotation of seeds in $\{\Sigma_0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4, \Sigma_5\}$, thus $\mathrm{Aut}^+(\mathcal{A})$ can be viewed as the symmetry group of the *bipartite belt* consisting of seeds in $\{\Sigma_0, \Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4, \Sigma_5\}$, where the quivers in these seeds are the bipartite quivers isomorphic to Q.

We will prove that any automorphism of $E_{\mathcal{A}}$ is induced from an element in $\mathrm{Aut}(\mathcal{A})$, and thus $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$. For this purpose, we show the following claims:

(1) There exists no automorphism of $E_{\mathcal{A}}$ which maps Σ_0 to a vertex except for Σ_i, $0 \le i \le 5$.

(2) If an automorphism of $E_{\mathcal{A}}$ maps Σ_0 to Σ_i, $0 \le i \le 5$, then it is induced from a cluster automorphism of $\mathcal{A}$.

Let σ be an automorphism of $E_{\mathcal{A}}$. Due to symmetries of $E_{\mathcal{A}}$, we only show that $\sigma(\Sigma) \ne O_i$, $i = 1, 2, 3$. By a direct computation, the sets of numbers for the layers of geodesic loops based on these vertices are

$$N(\ell_{\Sigma}^{0})=\{4,5,5\},\quad N(\ell_{\Sigma}^{1})=\{4,5,5\},\quad N(\ell_{\Sigma}^{2})=\{5,5\},\quad N(\ell_{\Sigma}^{3})=\{4\};$$
$$N(\ell_{O_1}^{0})=\{4,5,5\},\quad N(\ell_{O_1}^{1})=\{5,5,5\},\quad N(\ell_{O_1}^{2})=\{4,4\},\quad N(\ell_{O_1}^{3})=\{5\};$$
$$N(\ell_{O_2}^{0})=\{5,5,5\},\quad N(\ell_{O_1}^{1})=\{4,4,4\},\quad N(\ell_{O_1}^{2})=\{5,5,5\};$$
$$N(\ell_{O_3}^{0})=\{4,5,5\},\quad N(\ell_{O_3}^{1})=\{5,5,5\},\quad N(\ell_{O_3}^{2})=\{4,4\},\quad N(\ell_{O_3}^{3})=\{5\}.$$

Then by Remark 3.2(4), $\sigma(\Sigma_0)\neq O_i,\ i=1,2,3$. So the first claim is affirmed.

Now we consider the second claim. Still due to the symmetries of the graph, we may assume that $\sigma(\Sigma_0)=\Sigma_0$. Since σ is a graph automorphism, it can be seen that there are two possibilities for σ; one is the identity, the other is the reflection f_0 with respect to the vertical central axis of $E_{\mathcal{A}}$, as depicted in Figure 4. Note that the identity graph automorphism is induced from the identity automorphism of the cluster algebra, while the graph automorphism f_0 is induced from the cluster automorphism $(f_+f_-)^3$ by a direct computation. Therefore the second claim is true and we have $\mathrm{Aut}(E_{\mathcal{A}})\cong\mathrm{Aut}(\mathcal{A})\cong\mathbb{D}_6$.

Example 3.7. It is known from a result in [Fomin and Zelevinsky 2003b] that the cluster algebras of type B_n and type C_n have the same exchange graph. Based on a seed Σ_0, the exchange graph of a cluster algebra $\mathcal{A}$ of type B_3 or type C_3 is depicted in Figure 5. For the cluster algebra of type B_3, the quiver of the initial seed Σ_0 is

$$1\longrightarrow 2\xleftarrow{(2,1)}3.$$

For the cluster algebra of type C_3, the quiver of the initial seed Σ_0 is

$$frm[o]--\longrightarrow 2\xleftarrow{(1,2)}3.$$

Let σ be an automorphism of $E_{\mathcal{A}}$. As shown by Example 4 in [Chang and Zhu 2016a], we have $\mathrm{Aut}(\mathcal{A})\cong\mathbb{D}_4$ and $\{\Sigma_0,\Sigma_1,\Sigma_2,\Sigma_3,\Sigma_4,\Sigma_5,\Sigma_6,\Sigma_7\}$ are all the seeds whose quivers are isomorphic to Q. Similarly, to get $\mathrm{Aut}(E_{\mathcal{A}})\cong\mathrm{Aut}(\mathcal{A})$, we prove the two claims stated in Example 3.6. For the first claim, we only need to prove that $\sigma(\Sigma_0)\neq O_i(i=1,2,3,4)$ in Figure 5, and this can be obtained by the fact that these seeds have different combinatorial numbers of layers of geodesic loops:

$$\begin{aligned} N(\ell_{\Sigma_0}^{0})&=\{4,5,6\}, & N(\ell_{\Sigma_0}^{1})&=\{4,5,6\};\\ & & N(\ell_{O_1}^{0})&=\{5,6,6\};\\ N(\ell_{O_2}^{0})&=\{4,5,6\}, & N(\ell_{O_2}^{1})&=\{5,6,6\};\\ N(\ell_{O_3}^{0})&=\{4,5,6\}, & N(\ell_{O_3}^{1})&=\{5,6,6\};\\ & & N(\ell_{O_4}^{0})&=\{5,6,6\}. \end{aligned}$$

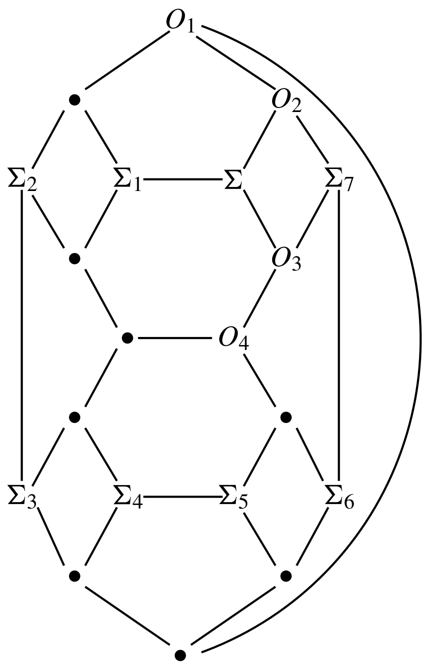


Figure 5. The exchange graph of a cluster algebra of type B_3 or type C_3.

For the second claim, we may also assume $\sigma(\Sigma_0) = \Sigma_0$. Since $N(\ell^0_{\Sigma_0}) = \{4, 5, 6\}$, there are neither rotation symmetries nor reflection symmetries of $E_{\mathcal{A}}$ at Σ_0. So σ must be the identity automorphism of $E_{\mathcal{A}}$, which is induced from the identity automorphism of the cluster algebra. Noticing that there are eight elements in $\mathrm{Aut}(\mathcal{A}) \cong \mathbb{D}_4$, where each one corresponds a graph automorphism which maps Σ_0 to $\Sigma_i,\ 0 \le i \le 7$.

Example 3.8. For cluster algebras of type F_4, let the quiver Q of a seed Σ be

$$1 \longrightarrow 2 \xleftarrow{(2,1)} 3 \longrightarrow 4.$$

Then $\mathrm{Aut}(\mathcal{A}) \cong \mathbb{D}_7$ [Chang and Zhu 2016a]. The variables x_1, x_2, x_3 and the corresponding full subquiver of Q form a seed Σ_1 of type B_3, while x_2, x_3, x_4 and the corresponding full subquiver of Q form a seed Σ_2 of type C_3. By pinning down Σ, rotating the graph $E_{\mathcal{A}}$ induces an automorphism σ of $E_{\mathcal{A}}$, which exchanges the graph $E_{\mathcal{A}_{\Sigma_1}}$ and the graph $E_{\mathcal{A}_{\Sigma_2}}$. However σ does not induce a cluster automorphism of $\mathcal{A}$, and $\mathrm{Aut}(\mathcal{A}) \cong \mathbb{D}_7 \subsetneq \mathbb{D}_7 \rtimes \mathbb{Z}_2 \cong \mathrm{Aut}(E_{\mathcal{A}})$.

Proposition 3.9. *Let Q be a connected quiver with three vertices which is of finite type. Let $\Sigma = (\boldsymbol{x}, Q)$ and $\Sigma' = (\boldsymbol{x}', Q')$ be two seeds (not necessarily mutation* *equivalent to each other). If there is an isomorphism $\sigma : E_{\mathcal{A}} \to E_{\mathcal{A}'}$ such that $\sigma(\Sigma) = \Sigma'$, then $\Sigma' = (\boldsymbol{x}', Q')$ is a finite type seed with Q' connected, and*

(1) *if Σ is of type A_3, then $Q' \cong Q$ (or Q^{op});*

(2) *if Σ is of type B_3 and Σ' is not of type C_3, then $Q' \cong Q$ (or Q^{op});*

(3) *if Σ is of type C_3 and Σ' is not of type B_3, then $Q' \cong Q$ (or Q^{op}).*

Proof. Clearly, since $E'_{\mathcal{A}} \cong E_{\mathcal{A}}$ is of finite type, Q' is a Dynkin type quiver with three vertices. If Q is of type A_3, then by Example 3.6,

$$N(\ell^0_\Sigma) = \{4, 5, 5\} \text{ or } \{5, 5, 5\}.$$

If Q is of type B_3 (or C_3), then from Example 3.7,

$$N(\ell^0_\Sigma) = \{4, 5, 6\} \text{ or } \{5, 6, 6\}.$$

If Q' is a union of a quiver of type A_2 and a point, then from Example 3.3,

$$N(\ell^0_\Sigma) = \{4, 4, 5\}.$$

If Q' is a union of a quiver of type B_2 (or C_2) and a point, then from Example 3.3,

$$N(\ell^0_\Sigma) = \{4, 4, 6\}.$$

If Q' is a union of a quiver of type G_2 and a point, then from Example 3.3,

$$N(\ell^0_\Sigma) = \{4, 4, 8\}.$$

Thus we get the proof by Remark 3.2. □

Example 3.10. Let Q be the quiver in Figure 6; we say it is of type $\tilde{A}_2$. Then it is not hard to see that if a quiver in the mutation class of Q is not isomorphic to Q, then it must be isomorphic to the quiver Q' in Figure 6. Let $\mathcal{A}$ be a cluster algebra with an initial seed

$$\Sigma = (\{x_1, x_2, x_3\}, Q),$$

as in the above examples. To show that $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$, we only need to notice that

$$N(\ell^0_\Sigma) = \{5, 5, \infty\},$$
$$N(\ell^0_{\Sigma'}) = \{5, 5, 5\},$$

where Σ' is a seed of $\mathcal{A}$ with quiver isomorphic to Q'. In fact, from [Assem et al. 2012, Section 3.3],

$$\mathrm{Aut}(\mathcal{A}) = \langle r_1, r_2 \mid r_1 r_2 = r_2 r_1, {r_1}^2 = r_2 \rangle \rtimes \langle \sigma \mid \sigma^2 = 1 \rangle \cong \mathbb{H}_{2,1} \rtimes \mathbb{Z}_2,$$

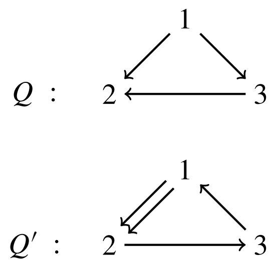


Figure 6. Quivers of type $\tilde{A}_2$.

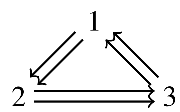


Figure 7. Quiver of type T_3.

where

$$r_1 : \begin{cases} x_1 \mapsto x_3, \\ x_2 \mapsto \mu_1(x_1), \\ x_3 \mapsto x_2, \end{cases} \tag{5}$$

$$r_2 : \begin{cases} x_1 \mapsto x_2, \\ x_2 \mapsto \mu_3\mu_1(x_3), \\ x_3 \mapsto \mu_1(x_1), \end{cases} \tag{6}$$

$$\sigma : \begin{cases} x_1 \mapsto x_2, \\ x_2 \mapsto x_1, \\ x_3 \mapsto x_3. \end{cases} \tag{7}$$

Thus $\mathrm{Aut}(E_{\mathcal{A}}) \cong \mathbb{H}_{2,1} \rtimes \mathbb{Z}_2$.

Example 3.11. Let $\mathcal{A}$ be a cluster algebra from a once punctured torus, which we call a cluster algebra of type T_3; then it is of finite mutation type with quiver always isomorphic to the quiver in Figure 7. By Lemma 2.13, we have $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$.

Corollary 3.12. *Let $\mathcal{A}$ and $\mathcal{A}'$ be two cluster algebras of finite type, or of skew-symmetric finite mutation type, with rank equal to* 2 *or* 3*. Let $\Sigma = (\boldsymbol{x}, B)$ and $\Sigma' = (\boldsymbol{x}', B')$ be two seeds of $\mathcal{A}$ and $\mathcal{A}'$, respectively. If $N(\ell^k_{\Sigma}) = N(\ell^k_{\Sigma'})$ for any $k \in \mathbb{Z}_{\geq 0}$, then there exists an isomorphism $\Phi : E_{\mathcal{A}} \to E_{\mathcal{A}'}$ such that $\Phi(\boldsymbol{x}) = \boldsymbol{x}'$.*

Proof. This follows from Examples 3.6, 3.7, 3.10 and 3.11. □

We expect the result in the corollary to be true for any finite type cluster algebras and finite mutation type cluster algebras. This means that for any seed Σ, the set $N(\ell_\Sigma^k)$ characterizes the exchange graph.

Lemma 3.13. *Let Q be a connected skew-symmetric quiver of finite mutation type.*

(1) *If there are 3 vertices in Q, then Q is one of the following types*:

  (a) *A_3 type.*
  (b) *$\tilde{A}_2$ type.*
  (c) *T_3 type.*

(2) *If there are at least four vertices in Q, then any full subquiver of Q with three vertices is of type A_3 or of type $\tilde{A}_2$.*

Proof. (1) From the classification of cluster algebras of finite mutation type, Q must be block-decomposable, so the proof is a straightforward check by gluing the blocks in Figure 2.

(2) We only need to notice that a quiver of type T_3 is obtained by gluing two blocks of type II in Figure 2, and thus one cannot further glue it with a block to obtain a connected quiver of finite mutation type. □

It is clear that if for any quiver in the mutation equivalent class of Q the number of arrows between any two vertices is at most 2, then Q is of finite mutation type. The above lemma shows that the inverse statement is also true for the cases when there are at least three vertices; that is, we have the following corollary, which has been stated in [Derksen and Owen 2008, Corollary 8].

Corollary 3.14. *A connected quiver Q with at least three vertices is of finite mutation type if and only if for any quiver in its mutation class the number of arrows between any two vertices is at most* 2.

Proposition 3.15. *Let Q be a connected skew-symmetric quiver with three vertices which is of finite mutation type. Let $\Sigma = (\boldsymbol{x}, Q)$ and $\Sigma' = (\boldsymbol{x}', Q')$ be two seeds. If there is an isomorphism $\sigma : E_{\mathcal{A}} \to E_{\mathcal{A}'}$ such that $\sigma(\Sigma) = \Sigma'$, then $Q' \cong Q$ or $Q' \cong Q^{\mathrm{op}}$.*

Proof. Similar to Proposition 3.9, this follows from Lemma 3.13 and Examples 3.11, 3.6 and 3.10. □

3C. *General cases.*

Theorem 3.16. *Let $\mathcal{A}$ be a cluster algebra of finite type. Assuming that it is not of type F_4, let $\Sigma = (\boldsymbol{x}, Q)$ be a labeled seed of $\mathcal{A}$, where Q is a connected quiver with at least three vertices. Then we have* $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$.

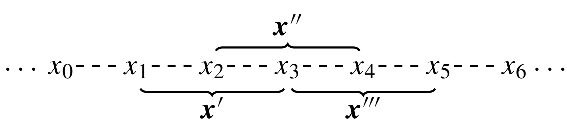
Proof. We need to show that $\mathrm{Aut}(E_{\mathcal{A}}) \subseteq \mathrm{Aut}(\mathcal{A})$. Let Φ be any automorphism of $E_{\mathcal{A}}$. Then it induces an automorphism ϕ of the complex Δ, in particular, which gives a permutation on the cluster variable set $\mathscr{X}$. Let $\boldsymbol{x}' \subseteq \boldsymbol{x}$ be a 3-dimensional complex, and let $Q(\boldsymbol{x}')$ be the full subquiver of $Q(\boldsymbol{x})$ with vertices indexed by the variables in $\boldsymbol{x}'$. Let $\mathcal{A}'$ be the cluster algebra defined by the seed $\Sigma' = (\boldsymbol{x}', Q(\boldsymbol{x}'))$.

Notice that since ϕ is an automorphism of a complex, it maps a simplex to a simplex, and thus induces a bijection from $\mathscr{X}_{\boldsymbol{x}'} = \{\alpha \in \mathscr{X} - (\boldsymbol{x} \setminus \boldsymbol{x}') : \boldsymbol{x} \setminus \boldsymbol{x}' \cup \{\alpha\} \in \Delta\}$ to $\mathscr{X}_{\phi(\boldsymbol{x}')} = \{\alpha \in \mathscr{X} - \phi(\boldsymbol{x} \setminus \boldsymbol{x}') : \phi(\boldsymbol{x} \setminus \boldsymbol{x}') \cup \{\alpha\} \in \Delta\}$, and also induces an isomorphism $\phi_{\boldsymbol{x}'}$ from the link $\Delta_{\boldsymbol{x}'}$ to the link $\Delta_{\phi(\boldsymbol{x}')}$. Moreover, the duality of the isomorphism $\phi_{\boldsymbol{x}'}$ gives an isomorphism between the dual graphs of the complexes; that is, we have an isomorphism

$$\Phi_{\boldsymbol{x}'} : \Gamma_{\boldsymbol{x}'} \to \Gamma_{\phi(\boldsymbol{x}')}. \tag{8}$$

Let $\overline{\Sigma}' = (\phi(\boldsymbol{x}'), Q(\phi(\boldsymbol{x}')))$ be a seed, where $Q(\phi(\boldsymbol{x}'))$ is the full subquiver of $Q(\Phi(\boldsymbol{x}))$ whose vertices are those labeled by elements in $\phi(\boldsymbol{x}')$. Let $\overline{\mathcal{A}}'$ be the cluster algebra defined by $\overline{\Sigma}'$. As shown in the beginning of Section 3A, there are isomorphisms $\Gamma_{\boldsymbol{x}'} \cong E_{\mathcal{A}'}$ and $\Gamma_{\phi(\boldsymbol{x}')} \cong E_{\overline{\mathcal{A}}'}$. Combining these with the isomorphism (8), we have $E_{\mathcal{A}'} \cong E_{\overline{\mathcal{A}}'}$. Since $\mathcal{A}$ is not of type F_4, if $Q(\boldsymbol{x}')$ is of type B_3 (resp. type C_3), then $Q(\phi(\boldsymbol{x}'))$ is not of type C_3 (resp. type B_3). Thus by Proposition 3.9, $Q(\boldsymbol{x}')) \cong Q(\phi(\boldsymbol{x}'))$ or $Q(\boldsymbol{x}') \cong Q(\phi(\boldsymbol{x}'))^{\mathrm{op}}$.

Let $\boldsymbol{x}' = \{x_1, x_2, x_3\} \subseteq \boldsymbol{x}$ and $\boldsymbol{x}'' = \{x_2, x_3, x_4\} \subseteq \boldsymbol{x}$ be two 3-dimensional complexes with exactly two common elements. By the above discussion, we have $Q(\boldsymbol{x}')) \cong Q(\phi(\boldsymbol{x}'))$ or $Q(\boldsymbol{x}') \cong Q(\phi(\boldsymbol{x}'))^{\mathrm{op}}$, and $Q(\boldsymbol{x}'')) \cong Q(\phi(\boldsymbol{x}''))$ or $Q(\boldsymbol{x}'') \cong Q(\phi(\boldsymbol{x}''))^{\mathrm{op}}$. Now assume $b_{x_2x_3} \neq 0$, that is, there exists at least one arrow in $Q(\boldsymbol{x})$ between the vertices labeled by x_2 and x_3, then simultaneously we have $Q(\boldsymbol{x}')) \cong Q(\phi(\boldsymbol{x}'))$ and $Q(\boldsymbol{x}'')) \cong Q(\phi(\boldsymbol{x}''))$, or $Q(\boldsymbol{x}') \cong Q(\phi(\boldsymbol{x}'))^{\mathrm{op}}$ and $Q(\boldsymbol{x}'') \cong Q(\phi(\boldsymbol{x}''))^{\mathrm{op}}$. Finally, due to the arbitrariness of the choice of $\boldsymbol{x}'$ and the connectedness of the quiver, one may show that $Q(\Phi(\boldsymbol{x})) \cong Q$ or $Q(\Phi(\boldsymbol{x})) \cong Q^{\mathrm{op}}$. See the inductive process in the following picture:



Therefore $\Phi : E_{\mathcal{A}} \to E_{\mathcal{A}}$ induces a cluster automorphism of $\mathcal{A}$ by Lemma 2.13. Thus $\mathrm{Aut}(E_{\mathcal{A}}) \subseteq \mathrm{Aut}(\mathcal{A})$ and we have $\mathrm{Aut}(E_{\mathcal{A}}) \cong \mathrm{Aut}(\mathcal{A})$. □

Remark 3.17. By combining the above theorem, Table 1 in [Assem et al. 2012] and Theorem 3.5 in [Chang and Zhu 2016a], we may compute the automorphism groups of the exchange graphs of cluster algebras of finite type; see Table 1. The cases of rank 2 and type F_4 are computed in Examples 3.3 and 3.8, respectively.

Theorem 3.18. *Let $\mathcal{A}$ be a connected skew-symmetric cluster algebra of finite mutation type*; *then* $\mathrm{Aut}(\mathcal{A}) \cong \mathrm{Aut}(E_{\mathcal{A}})$.

Proof. If $\mathcal{A}$ is of finite type of rank 2, that is, of type A_2, then the result follows from Example 3.3. If $\mathcal{A}$ is of infinite type of rank 2, the result follows from Example 3.4. When the rank of $\mathcal{A}$ is at least 3, the proof is similar to the proof of Theorem 3.16 by using the connectedness of the cluster algebra and Proposition 3.15. □

Corollary 3.19. *Let $\mathcal{A}$ be a connected cluster algebra of finite type or of skew-symmetric finite mutation type*, *then an automorphism of $E_{\mathcal{A}}$ is determined by the image of any fixed seed Σ and the images of the seeds adjacent to Σ. More precisely*, *let $\Sigma = (\boldsymbol{x}, B)$ be a seed on $E_{\mathcal{A}}$*, *then an automorphism $\Phi : E_{\mathcal{A}} \to E_{\mathcal{A}}$ is determined by a pair (Σ', ϕ)*, *where $\Sigma' = (\boldsymbol{x}', B')$ is a seed on $E_{\mathcal{A}}$ and $\phi : \boldsymbol{x} \to \boldsymbol{x}'$ is a bijection such that $\Phi(\Sigma) = \Sigma'$ and $\Phi(\mu_x(\boldsymbol{x})) = \mu_{\phi(x)}(\boldsymbol{x}')$ for any $x \in \boldsymbol{x}$.*

Proof. If $\mathcal{A}$ is of finite type of rank 2 and of type F_4, then the conclusion is clear. Otherwise, note that a cluster automorphism is determined by such a pair (Σ', ϕ); thus the proof follows from Theorems 3.16 and 3.18. □

Acknowledgements

The authors thank the reviewer for the careful reading and many valuable suggestions. In particular, the reference [Derksen and Owen 2008] was pointed out by the reviewer.

References

[Assem et al. 2012] I. Assem, R. Schiffler, and V. Shramchenko, "Cluster automorphisms", *Proc. Lond. Math. Soc.* (3) **104**:6 (2012), 1271–1302. MR Zbl

[Cerulli Irelli et al. 2013] G. Cerulli Irelli, B. Keller, D. Labardini-Fragoso, and P.-G. Plamondon, "Linear independence of cluster monomials for skew-symmetric cluster algebras", *Compos. Math.* **149**:10 (2013), 1753–1764. MR Zbl

[Chang and Zhu 2016a] W. Chang and B. Zhu, "Cluster automorphism groups of cluster algebras of finite type", *J. Algebra* **447** (2016), 490–515. MR Zbl

[Chang and Zhu 2016b] W. Chang and B. Zhu, "Cluster automorphism groups of cluster algebras with coefficients", *Sci. China Math.* **59**:10 (2016), 1919–1936. MR Zbl

[Chang and Zhu 2016c] W. Chang and B. Zhu, "On rooted cluster morphisms and cluster structures in 2-Calabi–Yau triangulated categories", *J. Algebra* **458** (2016), 387–421. MR Zbl

[Derksen and Owen 2008] H. Derksen and T. Owen, "New graphs of finite mutation type", *Electron. J. Combin.* **15**:1 (2008), art. id. 139. MR Zbl

[Felikson et al. 2012a] A. Felikson, M. Shapiro, and P. Tumarkin, "Cluster algebras of finite mutation type via unfoldings", *Int. Math. Res. Not.* **2012**:8 (2012), 1768–1804. MR Zbl

[Felikson et al. 2012b] A. Felikson, M. Shapiro, and P. Tumarkin, "Skew-symmetric cluster algebras of finite mutation type", *J. Eur. Math. Soc.* **14**:4 (2012), 1135–1180. MR Zbl

[Fomin and Zelevinsky 2002] S. Fomin and A. Zelevinsky, "Cluster algebras, I: Foundations", *J. Amer. Math. Soc.* **15**:2 (2002), 497–529. MR Zbl

[Fomin and Zelevinsky 2003a] S. Fomin and A. Zelevinsky, "Cluster algebras, II: Finite type classification", *Invent. Math.* **154**:1 (2003), 63–121. MR Zbl

[Fomin and Zelevinsky 2003b] S. Fomin and A. Zelevinsky, "Y-systems and generalized associahedra", *Ann. of Math.* (2) **158**:3 (2003), 977–1018. MR Zbl

[Fomin and Zelevinsky 2007] S. Fomin and A. Zelevinsky, "Cluster algebras, IV: Coefficients", *Compos. Math.* **143**:1 (2007), 112–164. MR Zbl

[Fomin et al. 2008] S. Fomin, M. Shapiro, and D. Thurston, "Cluster algebras and triangulated surfaces, I: Cluster complexes", *Acta Math.* **201**:1 (2008), 83–146. MR Zbl

[Gekhtman et al. 2008] M. Gekhtman, M. Shapiro, and A. Vainshtein, "On the properties of the exchange graph of a cluster algebra", *Math. Res. Lett.* **15**:2 (2008), 321–330. MR Zbl

[Keller 2012] B. Keller, "Cluster algebras and derived categories", pp. 123–183 in *Derived categories in algebraic geometry* (Tokyo, 2011), edited by Y. Kawamata, Eur. Math. Soc., Zürich, 2012. MR Zbl

[Saleh 2014] I. Saleh, "Exchange maps of cluster algebras", *Int. Electron. J. Algebra* **16** (2014), 1–15. MR Zbl

Received June 20, 2019. Revised January 27, 2020.

WEN CHANG
Current address:
SCHOOL OF MATHEMATICS AND INFORMATION SCIENCE
SHAANXI NORMAL UNIVERSITY
XIAN
CHINA
changwen161@163.com
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING CHINA

BIN ZHU
DEPARTMENT OF MATHEMATICAL SCIENCES
TSINGHUA UNIVERSITY
BEIJING
CHINA
zhu-b@tsinghua.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.303

GEOMETRIC MICROLOCAL ANALYSIS IN DENJOY–CARLEMAN CLASSES

STEFAN FÜRDÖS

A systematic geometric theory for the ultradifferentiable (nonquasianalytic and quasianalytic) wavefront set similar to the well-known theory in the classic smooth and analytic setting is developed. In particular an analogue of Bony's theorem and the invariance of the ultradifferentiable wavefront set under diffeomorphisms of the same regularity is proven using a theorem of Dynkin about the almost-analytic extension of ultradifferentiable functions. Furthermore, we prove a microlocal elliptic regularity theorem for operators defined on ultradifferentiable vector bundles. As an application, we show that Holmgren's theorem and several generalizations hold for operators with quasianalytic coefficients.

1. Introduction

The aim of this work is to establish a geometric theory for the wavefront set in ultradifferentiable classes introduced by Hörmander [1971a] analogous to the one for the classical wavefront set. There are a number of recent works dealing with this question; see, e.g., [Adwan and Hoepfner 2015; Berhanu and Hailu 2017; Hoepfner and Medrado 2018]. In this paper we present a unified approach to the problem, which also allows us to treat quasianalytic classes, which the methods introduced up to now were not able to cover. We note that the geometric theory of the ultradifferentiable wavefront set developed here has numerous possible applications, including for example to problems studied by Baouendi and Métivier [1982], Berhanu, Cordaro and Hounie [Berhanu et al. 2008] or Castellanos, Cordaro and Petronilho [Castellanos et al. 2013].

Regarding questions of the regularity of solutions of PDEs, the wavefront set is a crucial notion introduced by Sato [1970] in the analytic category and by

The author was supported by the Austrian Science Fund FWF, international cooperation project Nr. I 3472 and by the Czech Science Foundation GACR grant 17-19437S.
MSC2010: primary 26E10, 35A18; secondary 35A02, 35A30.
Keywords: ultradifferentiable wavefront set, elliptic regularity, uniqueness theorems, Denjoy–Carleman classes, Bony's theorem, quasianalytic uniqueness theorems.

Hörmander [1971b] in the smooth case. Their refinement of the singular support simplifies, for example, the proof of the classical elliptic regularity theorem considerably.

One of the basic features of both the smooth and analytic wavefront sets is that they are invariant under smooth and real-analytic changes of coordinates, respectively. Hence it is possible to define the smooth (or analytic) wavefront set of a distribution given on a smooth (or analytic) manifold. This is mainly due to the fact the smooth (resp. analytic) wavefront set can either be described by the Fourier transform (Hörmander's approach), boundary values of almost analytic (resp. holomorphic) functions (Sato's definition) or by the FBI transform (due to Bros and Iagolnitzer [1975]). The proof of the equivalence of these descriptions in the analytic category is due to Bony [1977].

Various other notions of wavefront sets associated to microlocalizable structures have since then been introduced; e.g., for Sobolev spaces, see, e.g., Lerner [2010]. In this paper we are interested in ultradifferentiable classes, that is, spaces of smooth functions which include strictly all real analytic functions. The most well known example of such classes are the Gevrey classes; see, e.g., [Rodino 1993].

Generally, spaces of ultradifferentiable functions are defined by putting growth conditions either on the derivatives or the Fourier transform of its elements. One family of ultradifferentiable classes, which includes the Gevrey classes, is the category of Denjoy–Carleman classes. The elements of a Denjoy–Carleman class satisfy generalized Cauchy estimates of the form

$$|\partial^\alpha f(x)| \leq C h^{|\alpha|} M_{|\alpha|}$$

on compact sets, where C and h are constants independent of α and $\mathcal{M} = (M_j)_j$ is a sequence of positive real numbers, the weight sequence associated to the Denjoy–Carleman class. Such classes of smooth functions were first investigated by Borel and Hadamard, but were named after Denjoy and Carleman who characterized independently the quasianalyticity of such a class using its weight sequence; see the survey [Thilliez 2008].

There is a rich literature concerning the Denjoy–Carleman classes and their properties. It turns out that conditions on the weight sequence translate to stability conditions of the associated class. For example, if $\mathcal{M}$ is a regular weight sequence in the sense of Dynkin [1976], then it is known that the Denjoy–Carleman class is closed under composition and solving ordinary differential equations and that the implicit function theorem holds in the class; see, e.g., [Bierstone and Milman 2004]. Hence it makes sense in this situation to consider manifolds of Denjoy–Carleman type. In Section 2 we give a brief introduction in the modern theory of Denjoy–Carleman classes and include a survey of the statements from the literature that are needed later on for the convenience of the reader. We note also that using the theory it is straightforward to generalize Nagano's theorem [1966] to orbits of quasianalytic vector fields.

There have been several attempts to define wavefront sets with respect to Denjoy–Carleman classes; see, e.g., [Komatsu 1991] and [Chung and Kim 1997]. The definition in the latter uses the FBI transform and also allows us to define $\mathrm{WF}_{\mathcal{M}}\, u$ for ultradistributions u in the nonquasianalytic case; see, e.g., [Adwan and Hoepfner 2010]. But if we restrict ourselves to distributions, the most wide-reaching definition of an ultradifferentiable wavefront set both with respect to the conditions imposed on the weight sequence and scope of achieved results was given by Hörmander [1971a] utilizing the Fourier transform. Due to the relatively weak conditions that he imposed on the weight sequence, Hörmander was only able to define the ultradifferentiable wavefront set $\mathrm{WF}_{\mathcal{M}}\, u$ of distributions u on real-analytic manifolds but not distributions defined on general ultradifferentiable manifolds. Hörmander's results are reviewed in Section 3.

The main result we need in order to proceed is a theorem of Dynkin [1976]. He showed that for regular weight sequences each function in a regular Denjoy–Carleman class has an almost-analytic extension, whose $\bar{\partial}$-derivative satisfies near $\operatorname{Im} z = 0$ a certain exponential decrease in terms of the weight sequence. We apply this result and several statements of Hörmander [1983] in Section 4 to prove that the Denjoy–Carleman wavefront set can be characterized by such $\mathcal{M}$-almost-analytic extensions. Using this characterization it is possible to modify Hörmander's proof of the invariance of the wavefront set in the real-analytic case to show that in our situation the ultradifferentiable wavefront set for distributions on Denjoy–Carleman manifolds can be well defined.

In Section 5 we show that $\mathrm{WF}_{\mathcal{M}}\, u$ can be characterized by the generalized FBI transform introduced by Berhanu and Hounie [2012]. This shows, in the case of distributions, the equivalence of the wavefront set introduced by Kim and Chung with the definition of Hörmander and generalizes, in that situation, results of Berhanu and Hailu [2017] and Hoepfner and Medrado [2018], especially to quasianalytic classes.

We may note that if we combine our methods with the arguments in [Hoepfner and Medrado 2018] then it is possible to generalize the above results to ultradistributions, in particular the invariance of the wavefront set on ultradifferentiable manifolds. In fact, it should be possible to give variants for ultradistributions of most of the statements that are proven in this paper. However, in order to give a unified presentation, especially regarding the assumptions on the weight sequence $\mathcal{M}$, we consider here only distributions.

As mentioned in the beginning, one of the fundamental results regarding the classical wavefront set is the elliptic regularity theorem which states in its microlocal form that we have for all partial differential operators P with smooth coefficients that $\mathrm{WF}\, u \subseteq \mathrm{WF}\, Pu \cup \mathrm{Char}\, P$, where $\mathrm{Char}\, P$ is the set of characteristic points of P, for all distributions u. Similarly Hörmander proved that $\mathrm{WF}_{\mathcal{M}}\, u \subseteq \mathrm{WF}_{\mathcal{M}}\, u \cup \mathrm{Char}\, P$ holds for operators with real-analytic coefficients. However, recently several authors, e.g., Albanese, Jornet and Oliaro[Albanese et al. 2010] and Pilipović, Teofanov and Tomić [Pilipović et al. 2018], have used the pattern of Hörmander's proof to show this inclusion for ultradifferentiable wavefront sets and operators with ultradifferentiable coefficients for variously defined ultradifferentiable classes.

Arguing similarly, we prove in Section 6 that, if $\mathcal{M}$ is a regular weight sequence that satisfies an additional condition, which is usually referred to in the literature as *moderate growth*, (see, e.g., [Thilliez 2008]), then $\mathrm{WF}_{\mathcal{M}}\, u \subseteq \mathrm{WF}_{\mathcal{M}}\, Pu \cup \mathrm{Char}\, P$ for operators P with coefficients in the Denjoy–Carleman class associated to $\mathcal{M}$. In fact, we show this inclusion for operators with ultradifferentiable coefficients acting on distributional sections of ultradifferentiable vector bundles.

Following the approach given separately by Kawai [Sato et al. 1973] and Hörmander [1971a] in the analytic case, we use the elliptic regularity theorem in Section 7 to prove a generalization of Holmgren's uniqueness theorem to operators with coefficients in quasianalytic Denjoy–Carleman classes. Finally we give quasianalytic versions of the generalizations of the analytic Holmgren's theorem due to Bony [1976], Hörmander [1993], Sjöstrand [1982] and Zachmanoglou [1972].

2. Denjoy–Carleman classes

Throughout this article, Ω denotes an open subset of $\mathbb{R}^n$. A *weight sequence* is a sequence of positive real numbers $(M_j)_{j\in\mathbb{N}_0}$ such that

$$M_0 = 1,$$
$$M_j^2 \le M_{j-1}M_{j+1}, \qquad j \in \mathbb{N}.$$

Definition 2.1. Let $\mathcal{M} = (M_j)_j$ be a weight sequence. We say that a smooth function $f \in \mathcal{E}(\Omega)$ is *ultradifferentiable of class* $\{\mathcal{M}\}$ if and only if for every compact set $K \Subset \Omega$ there exist constants C and h such that for all multi-indices $\alpha \in \mathbb{N}_0^n$,

$$|D^\alpha f(x)| \le Ch^{|\alpha|}M_{|\alpha|}, \qquad x \in K. \tag{2-1}$$

We denote the space of ultradifferentiable functions of class $\{M\}$ on Ω as $\mathcal{E}_{\mathcal{M}}(\Omega)$. Note that $\mathcal{E}_{\mathcal{M}}(\Omega)$ is always a subalgebra of $\mathcal{E}(\Omega)$ [Komatsu 1973].

Example 2.2. For any $s \ge 0$ consider the sequence $\mathcal{M}^s = ((k!)^{s+1})_k$. The space of ultradifferentiable functions associated to $\mathcal{M}^s$ is the well-known space of Gevrey functions $\mathcal{G}^{s+1} = \mathcal{E}_{\mathcal{M}^s}$ of order $s+1$; see, e.g., [Rodino 1993]. If $s=0$ then $\mathcal{G}^1 = \mathcal{E}_{\mathcal{M}^0} = \mathcal{O}$ is the space of real-analytic functions.

Remark 2.3. It is easy to see that $\mathcal{E}_{\mathcal{M}}(\Omega)$ is an infinite-dimensional vector space, since it contains all polynomials. In fact $\mathcal{E}_{\mathcal{M}}(\Omega)$ is a complete locally convex space; see, e.g., [Komatsu 1973]. The topology on $\mathcal{E}_{\mathcal{M}}(\Omega)$ is defined as follows. If $K \Subset \Omega$ is a compact set such that $K = \overline{K^\circ}$ then we define for $f \in \mathcal{E}(K)$,

$$\|f\|_K^h := \sup_{\substack{x\in K\\ \alpha\in\mathbb{N}_0^n}} \left|\frac{D^\alpha f(x)}{h^{|\alpha|}M_{|\alpha|}}\right|$$

and set

$$\mathcal{E}_{\mathcal{M}}^h(K) := \{f \in \mathcal{E}(K) \mid \|f\|_K^h < \infty\}.$$

It is easy to see that $\mathcal{E}_{\mathcal{M}}^h(K)$ is a Banach space. Moreover, $\mathcal{E}_{\mathcal{M}}^h(K) \subsetneq \mathcal{E}_{\mathcal{M}}^k(K)$ for $h < k$ and the inclusion mapping $\iota_h^k : \mathcal{E}_{\mathcal{M}}^h(K) \to \mathcal{E}_{\mathcal{M}}^k(K)$ is compact. Hence the space

$$\mathcal{E}_{\mathcal{M}}(K) := \{f \in \mathcal{E}(K) \mid \text{there exists } h > 0 \text{ such that } \|f\|_K^h < \infty\} = \varinjlim_h \mathcal{E}_{\mathcal{M}}^h(K)$$

is an (LB)-space. We can now write

$$\mathcal{E}_{\mathcal{M}}(\Omega) = \varprojlim_K \mathcal{E}_{\mathcal{M}}(K)$$

as a projective limit. For more details on the topological structure of $\mathcal{E}_{\mathcal{M}}(\Omega)$, see [Komatsu 1973].

We also call $\mathcal{E}_{\mathcal{M}}(\Omega)$ the *Denjoy–Carleman class on Ω associated to the weight sequence $\mathcal{M}$*.

If $\mathcal{M}$ and $\mathcal{N}$ are two weight sequences then

$$\mathcal{M} \preccurlyeq \mathcal{N} :\Longleftrightarrow \sup_{k\in\mathbb{N}_0}\left(\frac{M_k}{N_k}\right)^{1/k} < \infty$$

defines a reflexive and transitive relation on the space of weight sequences. Furthermore it induces an equivalence relation by setting

$$\mathcal{M} \approx \mathcal{N} :\Longleftrightarrow \mathcal{M} \preccurlyeq \mathcal{N} \text{ and } \mathcal{N} \preccurlyeq \mathcal{M}.$$

It holds that $\mathcal{E}_{\mathcal{M}} \subseteq \mathcal{E}_{\mathcal{N}}$ if and only if $\mathcal{M} \preccurlyeq \mathcal{N}$ and $\mathcal{E}_{\mathcal{M}} = \mathcal{E}_{\mathcal{N}}$ if and only if $\mathcal{M} \approx \mathcal{N}$; see [Mandelbrojt 1952] and also [Rainer and Schindl 2014; Thilliez 2008]. For example, if $r < s$ then $\mathcal{G}^{r+1} \subsetneq \mathcal{G}^{s+1}$.

The weight function $\omega_{\mathcal{M}}$ (see [Mandelbrojt 1952; Komatsu 1973]) associated to the weight sequence $\mathcal{M}$ is defined by

$$\omega_{\mathcal{M}}(t) := \sup_{j\in\mathbb{N}_0} \log \frac{t^j}{M_j}, \quad t > 0,$$
$$\omega_{\mathcal{M}}(0) := 0.$$

We note that $\omega_{\mathcal{M}}$ is a continuous increasing function on $[0, \infty)$ and vanishes on the interval [0, 1], and $\omega_{\mathcal{M}} \circ \exp$ is convex. In particular $\omega_{\mathcal{M}}(t)$ increases faster than $\log t^p$ for any $p > 0$ as t tends to infinity. It is possible to extract the weight sequence from the weight function, i.e.,

$$M_k = \sup_t \frac{t^k}{e^{\omega_{\mathcal{M}}(t)}};$$

see [Mandelbrojt 1952; Komatsu 1973].

If f and g are two continuous functions defined on $[0, \infty)$ then we write $f \sim g$ if and only if $f(t) = O(g(t))$ and $g(t) = O(f(t))$ for $t \to \infty$. It can be shown that the weight function ω_s for the Gevrey space $\mathcal{G}^{s+1}$ satisfies

$$\omega_s(t) \sim t^{1/(s+1)}.$$

Sometimes the classes $\mathcal{E}_{\mathcal{M}}$ are defined using the sequence $m_k = M_k/(k!)$ instead of $(M_k)_k$ and (2-1) is replaced by

$$|D^\alpha f(x)| \leq Ch^{|\alpha|}|\alpha|!m_{|\alpha|}.$$

Infrequently the sequences $\mu_k = M_{k+1}/M_k$ or $L_k = M_k^{1/k}$ are also used, with an accordingly modified version of (2-1); see also Remark 3.3. The main reason for the different ways of defining the Denjoy–Carleman classes is the following. In order to show that these classes satisfy certain properties, like the inverse function theorem, one has to put certain conditions on the defining data of the spaces, i.e., the weight sequence; see, e.g., [Rainer and Schindl 2016]. Often these conditions are easier to write down in terms of these other sequences instead of using $(M_j)_j$. In the following our point of view is that the sequences $(M_k)_k$, $(m_k)_k$, $(\mu_k)_k$ and $(L_k)_k$ are all associated to the weight sequence $\mathcal{M}$. We are going to use especially the two sequences $(m_j)_j$ and $(M_j)_j$ indiscriminately.

We may note that sometimes ultradifferentiable functions associated to the weight sequence $\mathcal{M}$ are defined as smooth functions satisfying (2-1) for all $h > 0$ on each compact K; see, e.g., [Ehrenpreis 1970]. One says then that f is ultradifferentiable of class $(\mathcal{M})$ and the corresponding space is the Beurling class associated to $\mathcal{M}$. On the other hand $\mathcal{E}_{\mathcal{M}}$ is then usually called the Roumieu class associated to $\mathcal{M}$; see, e.g., [Komatsu 1973; Rainer and Schindl 2016].

From now on we shall put certain conditions on the weight sequences under consideration.

Definition 2.4. We say that a weight sequence $\mathcal{M}$ is *regular* if and only if it satisfies the following conditions, with $k \in \mathbb{N}$:

(M1) $$m_0 = m_1 = 1.$$

(M2) $$\sup_k \sqrt[k]{\frac{m_{k+1}}{m_k}} < \infty.$$

(M3) $$m_k^2 \le m_{k-1} m_{k+1}.$$

(M4) $$\lim_{k\to\infty} \sqrt[k]{m_k} = \infty.$$

The last condition just means that the space $\mathcal{O}$ of all real-analytic functions is strictly contained in $\mathcal{E}_{\mathcal{M}}$ whereas the first is a useful normalization condition that will help simplify certain computations. It is obvious that if we replace in (M1) the number 1 with some other positive real number we would not change the resulting space $\mathcal{E}_{\mathcal{M}}$.

If $\mathcal{M}$ is a regular weight sequence, then it is well known that the associated Denjoy–Carleman class satisfies certain stability properties; see, e.g., [Bierstone and Milman 2004; Rainer and Schindl 2016]. For example $\mathcal{E}_{\mathcal{M}}$ is *closed under differentiation*, i.e., if $f \in \mathcal{E}_{\mathcal{M}}(\Omega)$ then $D^\alpha f \in \mathcal{E}_{\mathcal{M}}(\Omega)$ for all $\alpha \in \mathbb{N}_0^n$.

Remark 2.5. The fact that $\mathcal{E}_{\mathcal{M}}(\Omega)$ is closed under differentiation implies immediately another stability condition, namely *closedness under division by a coordinate* (see [Bierstone and Milman 2004]):

Suppose that $f \in \mathcal{E}_{\mathcal{M}}(\Omega)$ and $f(x_1, \dots, x_{j-1}, a, x_{j+1}, \dots, x_n) = 0$ for some fixed $a \in \mathbb{R}$ and all $x_k,\ k \neq j$, with the property $(x_1, \dots, x_{j-1}, a, x_{j+1}, \dots, x_n) \in \Omega$. Then we apply the fundamental theorem of calculus to the function

$$f_j : t \longmapsto f(x_1, \dots, x_{j-1}, t(x_j - a) + a, x_{j+1}, \dots, x_n)$$

and obtain

$$\begin{aligned} f(x) &= \int_0^1 \frac{\partial f_j}{\partial t}(t)\, dt \\ &= (x_j - a) \int_0^1 \frac{\partial f}{\partial x_j}(x_1, \dots, x_{j-1}, t(x_j - a) + a, x_{j+1}, \dots, x_n)\, dt \\ &= (x_j - a) g(x). \end{aligned}$$

It is easy to see that $g \in \mathcal{E}_{\mathcal{M}}(\Omega)$ using $\partial f / \partial x_j \in \mathcal{E}_{\mathcal{M}}(\Omega)$.

For the proof of the properties above, only (M2) was used. If we apply also (M3) then it is possible to show that $\mathcal{E}_{\mathcal{M}}(\Omega)$ is *inverse closed*, i.e., if $f \in \mathcal{E}_{\mathcal{M}}(\Omega)$ does not vanish at any point of Ω then

$$\frac{1}{f} \in \mathcal{E}_{\mathcal{M}}(\Omega);$$

see [Rainer and Schindl 2016].

In fact, if $\mathcal{M}$ is a regular weight sequence then the associated Denjoy–Carleman class satisfies also the following stability properties.

Theorem 2.6. *Let $\mathcal{M}$ be a regular weight sequence and $\Omega_1 \subseteq \mathbb{R}^m$ and $\Omega_2 \subseteq \mathbb{R}^n$ be open sets. Then the following holds*:

(1) *The class $\mathcal{E}_\mathcal{M}$ is* ***closed under composition*** *(see [Roumieu 1962] and also [Bierstone and Milman 2004]), i.e., let $F\colon \Omega_1 \to \Omega_2$ be an $\mathcal{E}_\mathcal{M}$-mapping, that is, each component F_j of F is ultradifferentiable of class $\{\mathcal{M}\}$ in Ω_1, and $g \in \mathcal{E}_\mathcal{M}(\Omega_2)$. Then also $g \circ F \in \mathcal{E}_\mathcal{M}(\Omega_1)$.*

(2) *The* ***inverse function theorem*** *holds in the Denjoy–Carleman class $\mathcal{E}_\mathcal{M}$ (see [Komatsu 1979]): Let $F : \Omega_1 \to \Omega_2$ be an $\mathcal{E}_\mathcal{M}$-mapping and $p_0 \in \Omega_1$ such that the Jacobian $F'(p_0)$ is invertible. Then there exist neighborhoods U of p_0 in Ω_1 and V of $q_0 = F(x_0)$ in Ω_2 and an $\mathcal{E}_\mathcal{M}$-mapping $G : V \to U$ such that $G(q_0) = p_0$ and $F \circ G = \mathrm{id}_V$.*

(3) *The* ***implicit function theorem*** *is valid in $\mathcal{E}_\mathcal{M}$ (see [Komatsu 1979]): Let $F : \mathbb{R}^{n+d} \supseteq \Omega \to \mathbb{R}^d$ be an $\mathcal{E}_\mathcal{M}$-mapping and $(x_0, y_0) \in \Omega$ such that $F(x_0, y_0) = 0$ and $\partial F/\partial y(x_0, y_0)$ is invertible. Then there exist open sets $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^d$ with $(x_0, y_0) \in U \times V \subseteq \Omega$ and an $\mathcal{E}_\mathcal{M}$-mapping $G : U \to V$ such that $G(x_0) = y_0$ and $F(x, G(x)) = 0$ for all $x \in V$.*

Furthermore it is true that $\mathcal{E}_\mathcal{M}(\Omega)$ is *closed under solving ODEs*; to be more specific, the following theorem holds.

Theorem 2.7 (Yamanaka [1991]; see also Komatsu [1980]). *Let $\mathcal{M}$ be a regular weight sequence, $0 \in I \subseteq \mathbb{R}$ an open interval, $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^d$ be open and $F \in \mathcal{E}_\mathcal{M}(I \times U \times V)$.*

Then the initial value problem

$$\begin{aligned} x'(t) &= F(t, x(t), \lambda), && t \in I,\ \lambda \in V, \\ x(0) &= x_0, && x_0 \in U, \end{aligned}$$

has locally a unique solution x that is ultradifferentiable near 0.

More precisely, there is an open set $\Omega \subseteq I \times U \times V$ that contains the point $(0, x_0, \lambda)$ and an $\mathcal{E}_\mathcal{M}$-mapping $x = x(t, y, \lambda) : \Omega \to U$ such that the function $t \mapsto x(t, y_0, \lambda_0)$ is the solution of the initial value problem

$$\begin{aligned} x'(t) &= F(t, x(t), \lambda_0), \\ x(0) &= y_0. \end{aligned}$$

For any regular weight sequence $\mathcal{M}$ we can define the associated weight by

$$h_\mathcal{M}(t) = \inf_k t^k m_k \quad \text{if } t > 0 \text{ and } h_\mathcal{M}(0) = 0. \tag{2-2}$$

As above, we have that

$$m_k = \sup_t \frac{h_\mathcal{M}(t)}{t^k}.$$

In order to describe the connection between the weight and the weight function associated to a regular weight sequence we set

$$\tilde{\omega}_{\mathcal{M}}(t) := \sup_{j\in\mathbb{N}_0} \log \frac{t^j}{m_j},$$
$$\tilde{h}_{\mathcal{M}}(t) = \inf_k t^k M_k,$$

for $t > 0$ and $\tilde{\omega}_{\mathcal{M}}(0) = \tilde{h}_{\mathcal{M}}(0) = 0$.

Lemma 2.8. *If $\mathcal{M}$ is a regular weight sequence then*

$$\begin{aligned} h_{\mathcal{M}}(t) &= e^{-\tilde{\omega}_{\mathcal{M}}(1/t)}, \\ \tilde{h}_{\mathcal{M}}(t) &= e^{-\omega_{\mathcal{M}}(1/t)}. \end{aligned} \tag{2-3}$$

Proof. We prove only the equality for $h_{\mathcal{M}}$. Of course, the verification of the other equation is completely analogous. If $t > 0$ is chosen arbitrarily we have by the monotonicity of the exponential function that

$$\exp\left(\tilde{\omega}_{\mathcal{M}}\left(\frac{1}{t}\right)\right) = \exp\left(\sup_k \log \frac{1}{m_k t^k}\right) = \sup_k \frac{1}{m_k t^k} = \frac{1}{\inf_k m_k t^k} = \frac{1}{h_{\mathcal{M}}(t)}. \qquad \square$$

We obtain that $h_{\mathcal{M}}$ is continuous with values in $[0, 1]$, equals 1 on $[1, \infty)$ and goes more rapidly to 0 than t^p for any $p > 0$ for $t \to 0$. Although the weight function is the prevalent concept, the weight has been used, e.g., by Dynkin [1976] and Thilliez [2003].

Example 2.9. If $\mathcal{M} = \mathcal{M}^s$ is the Gevrey sequence of order s then we know already that the associated weight function satisfies $\omega_s(t) \sim t^{1/(1+s)}$. Hence (2-3) shows for $s > 0$ that if we set

$$f_s(t) = e^{-1/t^s}$$

then there are constants C_1, C_2, Q_1 and $Q_2 > 0$ such that

$$C_1 f_s(Q_1 t) \le h_s(t) \le C_2 f_s(Q_2 t)$$

for $t > 0$.

It is well known (see, e.g., [Mather 1971] or [Melin and Sjöstrand 1975]) that a function f is smooth on Ω if and only if there is an almost-analytic extension F of f, i.e., there exists a smooth function F on some open set $\tilde{\Omega} \subseteq \mathbb{C}^n$ with $\tilde{\Omega} \cap \mathbb{R}^n = \Omega$ such that

$$\bar{\partial}_j F = \frac{\partial}{\partial \bar{z}_j} F = \frac{1}{2}\left(\frac{\partial}{\partial x_j} + i \frac{\partial}{\partial y_j}\right) F$$

is flat on Ω and $F|_\Omega = f$. The idea is now that if f is ultradifferentiable then one should find an extension F of f such that the regularity of f is translated in a certain uniform decrease of $\bar{\partial}_j F$ near Ω (see [Dynkin 1993]). Such extensions were constructed, e.g., by Petzsche and Vogt [1984] and Adwan and Hoepfner [2010] under relative restrictive conditions on the weight sequence. The most general result in this regard though was given by Dynkin [1976].

Theorem 2.10. *Let $\mathcal{M}$ be a regular weight sequence and $K \Subset \mathbb{R}^n$ be a compact convex set with $K = \overline{K^\circ}$. Then $f \in \mathcal{E}_\mathcal{M}(K)$ if and only if there exists a test function $F \in \mathcal{D}(\mathbb{C}^n)$ with $F|_K = f$ and if there are constants $C, Q > 0$ such that*

$$|\bar{\partial}_j F(z)| \leq C h_\mathcal{M}(Q d_K(z)), \tag{2-4}$$

where $1 \leq j \leq n$ and d_K is the distance function with respect to K on $\mathbb{C}^n \setminus K$.

We shall note that Dynkin used the function $h_1(t) = \inf_{k\in\mathbb{N}} m_k t^{k-1}$ instead of the weight $h_\mathcal{M}$.[1] But we observe that

$$h_\mathcal{M}(t) = \inf_{k\in\mathbb{N}_0} m_k t^k \leq t \inf_{k\in\mathbb{N}} m_k t^{k-1} = t h_1(t) \leq Ct \inf_{k\in\mathbb{N}} m_{k-1} t^{k-1} = Ct h_\mathcal{M}(t),$$

where we used (M2). Since $h_\mathcal{M}$ is rapidly decreasing for $t \to 0$ we can interchange these two functions in the formulation of Theorem 2.10. In fact, Dynkin's proof gives immediately the following result.

Corollary 2.11. *Let $\mathcal{M}$ be a regular weight sequence, $p \in \Omega$ and $f \in \mathcal{D}'(\Omega)$. If f is ultradifferentiable of class $\{\mathcal{M}\}$ near p, i.e., there exists a compact neighborhood K of p such that $f|_K \in \mathcal{E}_\mathcal{M}(K)$, then there are an open neighborhood $W \subseteq \Omega$ of p, a constant $\rho > 0$ and a function $F \in \mathcal{E}(W + iB(0,\rho))$ such that $F|_W = f|_W$ and*

$$|\bar{\partial}_j F(x+iy)| \leq C h_\mathcal{M}(Q|y|) \tag{2-5}$$

for some positive constants C, Q and all $1 \leq j \leq n$ and $x + iy \in W + iB(0,\rho)$.

One of the main questions in the study of ultradifferentiable functions is if the class under consideration behaves more like the ring of real-analytic functions or the ring of smooth functions. E.g., does the class contain flat functions, that means nonzero elements whose Taylor series at some point vanishes? That leads to following definition.

Definition 2.12. Let $E \subseteq \mathcal{E}(\Omega)$ be a subalgebra. We say that E is quasianalytic if and only if for $f \in E$ the fact that $D^\alpha f(p) = 0$ for some $p \in \Omega$ and all $\alpha \in \mathbb{N}_0^n$ implies that $f \equiv 0$ in the connected component of Ω that contains p.

In the case of Denjoy–Carleman classes quasianalyticity is characterized by the following theorem.

[1] h_1 is in fact the weight associated to the shifted sequence $(m_{k+1})_k$.

Theorem 2.13 [Denjoy 1921; Carleman 1923a; 1923b]. *The space $\mathcal{E}_{\mathcal{M}}(\Omega)$ is quasianalytic if and only if*

$$\sum_{k=1}^{\infty} \frac{M_{k-1}}{M_k} = \infty. \tag{2-6}$$

We say that a weight sequence is quasianalytic if and only if it satisfies (2-6) and nonquasianalytic otherwise.

Example 2.14. Let $\sigma > 0$ be a parameter. We define a family $\mathcal{N}^\sigma$ of regular weight sequences by $N_0 = N_1 = 1$ and

$$N_k^\sigma = k!(\log(k+e))^{\sigma k}$$

for $k \geq 2$. The weight sequence $\mathcal{N}^\sigma$ is quasianalytic if and only if $0 < \sigma \leq 1$; see [Thilliez 2008].

Remark 2.15. Obviously $\mathcal{D}_{\mathcal{M}}(\Omega) = \mathcal{D}(\Omega) \cap \mathcal{E}_{\mathcal{M}}(\Omega)$ is nontrivial if and only if $\mathcal{E}_{\mathcal{M}}(\Omega)$ is nonquasianalytic; see, e.g., [Rudin 1966]. It is well known that the sequences $\mathcal{M}^s$ are nonquasianalytic if and only if $s > 0$. In fact there is a nonquasianalytic regular weight sequence $\tilde{\mathcal{M}}$ such that $\tilde{\mathcal{M}} \preccurlyeq \mathcal{M}^s$ for all $s > 0$; see [Rainer and Schindl 2014, page 125]. Hence

$$\mathcal{O} \subsetneq \mathcal{E}_{\tilde{\mathcal{M}}} \subsetneq \bigcap_{s>0} \mathcal{G}^{s+1}.$$

Using Theorem 2.6 we produce the following definition:

Definition 2.16. Let M be a smooth manifold and $\mathcal{M}$ a regular weight sequence. We say that M is an ultradifferentiable manifold of class $\{\mathcal{M}\}$ if and only if there is an atlas $\mathcal{A}$ of M that consists of charts such that

$$\varphi' \circ \varphi^{-1} \in \mathcal{E}_{\mathcal{M}}$$

for all $\varphi, \varphi' \in \mathcal{A}$.

A mapping $F \colon M \to N$ between two manifolds of class $\{\mathcal{M}\}$ is ultradifferentiable of class $\{\mathcal{M}\}$ if and only if $\psi \circ F \circ \varphi^{-1} \in \mathcal{E}_{\mathcal{M}}$ for any charts φ and ψ of M and N, respectively. We can now consider the category of ultradifferentiable manifolds of class $\{\mathcal{M}\}$. We denote by

$$\mathfrak{X}_{\mathcal{M}}(M) = \mathcal{E}_{\mathcal{M}}(M, TM)$$

the Lie algebra of ultradifferentiable vector fields on M. Note that, if $\mathcal{M}$ is a regular weight sequence, an integral curve of an ultradifferentiable vector field of class $\{\mathcal{M}\}$ is an $\mathcal{E}_{\mathcal{M}}$-curve by Theorem 2.7.

These considerations allow us to state a quasianalytic version of Nagano's theorem [1966].

Theorem 2.17. *Let U be an open neighborhood of $p_0 \in \mathbb{R}^n$ and $\mathcal{M}$ a quasianalytic regular weight sequence. Furthermore let $\mathfrak{g}$ be a Lie subalgebra of $\mathfrak{X}_{\mathcal{M}}(U)$ that is also an $\mathcal{E}_{\mathcal{M}}$-module, i.e., if $X \in \mathfrak{g}$ and $f \in \mathcal{E}_{\mathcal{M}}(U)$ then $fX \in \mathfrak{g}$.*

There exists an ultradifferentiable submanifold W of class $\{\mathcal{M}\}$ in U, such that

$$T_pW = \mathfrak{g}(p) \quad \textit{for all } p \in W. \tag{2-7}$$

Moreover, the germ of W at p_0 is uniquely defined by this property.

The proof of Theorem 2.17 is the same as in the analytic version; see, e.g., Baouendi, Ebenfelt and Rothschild [Baouendi et al. 1999]. We call the uniquely defined germ $\gamma_{p_0}(\mathfrak{g})$ of the manifold constructed in Theorem 2.17 the local Nagano leaf of $\mathfrak{g}$ at p_0. From now on all Lie algebras of ultradifferentiable vector fields that are considered are assumed to be also $\mathcal{E}_{\mathcal{M}}$-modules.

Following Nagano [1966] (see also [Baouendi et al. 1999]), we can also give a global version of Theorem 2.17.

Theorem 2.18. *Let $\mathcal{M}$ be a quasianalytic regular weight sequence. If $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{X}_{\mathcal{M}}(\Omega)$ then $\mathfrak{g}$ admits a foliation of Ω, that is a partition of Ω by maximal integral manifolds.*

Before we close this section we need to introduce another condition for weight sequences. Let $\mathcal{M}$ be a weight sequence. We say that $\mathcal{M}$ is of *moderate growth* if and only if there are constants C and ρ such that

$$M_{j+k} \leq C\rho^{j+k} M_j M_k \tag{M2$'$}$$

for all $(j,k) \in \mathbb{N}_0^2$. Both the Gevrey sequences $\mathcal{M}^s$ and the sequences $\mathcal{N}^\sigma$ from Example 2.14 satisfy (M2$'$) for all s and σ, respectively.

For a discussion of this condition, see, e.g., [Komatsu 1973]. Here we only mention two facts. First, for any weight sequence $\mathcal{M}$, if (M2$'$) holds then (M2) is also satisfied. Furthermore, if $\mathcal{M}$ satisfies (M2$'$) then there is some $s > 0$ such that $\mathcal{E}_{\mathcal{M}} \subseteq \mathcal{G}^{1+s}$; see, e.g., [Thilliez 2003]. On the other hand consider the regular weight sequence $\mathcal{L}$ given by $L_0 = L_1 = 1$ and $L_k = k!2^{k^2}$ for $k \geq 2$. Then $\mathcal{G}^{1+s} \subsetneq \mathcal{E}_{\mathcal{L}}$ for all $s \geq 0$ and therefore $\mathcal{L}$ cannot satisfy (M2$'$).

3. The ultradifferentiable wavefront set

In this and the following two sections we always assume that $\mathcal{M}$ is a regular weight sequence.

Hörmander [1971a] proved the following local characterization of $\mathcal{E}_{\mathcal{M}}$ via the Fourier transform:

Proposition 3.1. *Let $u \in \mathcal{D}'(\Omega)$ and $p_0 \in \Omega$. Then u is ultradifferentiable of class $\{\mathcal{M}\}$ near p_0 if and only if there are an open neighborhood V of p_0, a* *bounded sequence* $(u_N)_N \subseteq \mathcal{E}'(U)$ *such that* $u|_V = (u_N)|_V$ *and some constant* $Q > 0$ *so that*

$$\sup_{\substack{\xi\in\mathbb{R}^n\\ N\in\mathbb{N}_0}} \frac{|\xi|^N|\hat{u}_N(\xi)|}{Q^N M_N} < \infty.$$

Hörmander then used this characterization to define the ultradifferentiable wavefront set:

Definition 3.2. Let $u \in \mathcal{D}'(\Omega)$ and $(x_0, \xi_0) \in T^*\Omega \setminus \{0\}$. We say that u is *microlocally ultradifferentiable of class* $\{\mathcal{M}\}$ at (x_0, ξ_0) if and only if there are a neighborhood V of x_0, a bounded sequence $(u_N)_N \subseteq \mathcal{E}'(\Omega)$ with $u_{N|V} \equiv u_{|V}$ and a conic neighborhood Γ of ξ_0 such that $u_N|_V \equiv u|_V$, where $V \in \mathcal{U}(x_0)$, and a conic neighborhood Γ of ξ_0 such that for some constant $Q > 0$

$$\sup_{\substack{\xi\in\Gamma\\ N\in\mathbb{N}_0}} \frac{|\xi|^N|\hat{u}_N|}{Q^N M_N} < \infty. \tag{3-1}$$

The ultradifferentiable wavefront set $\mathrm{WF}_{\mathcal{M}}\, u$ is then defined as

$$\mathrm{WF}_{\mathcal{M}}\, u := \bigl\{(x,\xi) \in T^*\Omega \setminus \{0\} \mid \\ u \text{ is not microlocally ultradifferentiable of class } \{\mathcal{M}\} \text{ at } (x,\xi)\bigr\}.$$

Remark 3.3. Hörmander [1971a] defined $\mathrm{WF}_{\mathcal{M}}$ for weight sequences that satisfy weaker conditions than those we imposed in Definition 2.4. He required, as we have done, (M2) and that $\mathcal{O} \subseteq \mathcal{E}_{\mathcal{M}}$, but (M3) is replaced by the monotonic growth of the sequence

$$L_N = (M_N)^{1/N}. \tag{3-2}$$

This condition still implies that $\mathcal{E}_{\mathcal{M}}$ is an algebra but gives only that $\mathcal{E}_{\mathcal{M}}$ is closed under composition with analytic mappings.

More precisely, in terms of the sequence $(L_N)_N$, the conditions that Hörmander imposed take the following form. First, $N \le L_N$ and $L_{N+1} \le C L_N$ for all N and a constant $C > 0$ independent of N. Furthermore, as mentioned before, the sequence $(L_N)_N$ is also assumed to be increasing.

Note that Hörmander's classes might not even be defined by weight sequences in the sense of Section 2. Hence Hörmander [1983] was able to define $\mathrm{WF}_{\mathcal{M}}\, u$ for distributions u on real analytic manifolds but not on arbitrary ultradifferentiable manifolds of class $\{\mathcal{M}\}$; note that the implicit function theorem may not hold in an arbitrary ultradifferentiable class defined by weight sequences obeying his conditions. Similarly he proved that

$$\mathrm{WF}_{\mathcal{M}}\, u \subseteq \mathrm{WF}_{\mathcal{M}}\, Pu \cup \mathrm{Char}\, P$$

for linear partial differential operators P with analytic coefficients but not for operators whose coefficients might be only of class $\{\mathcal{M}\}$.

As mentioned before it is possible to modify the arguments of Hörmander in the case of regular weight sequences to show that the above inclusion holds for partial differential operators with ultradifferentiable coefficients as long as $\mathcal{M}$ is regular and of moderate growth. Similarly we are able to define $\operatorname{WF}_{\mathcal{M}} u$ for distributions defined on manifolds of class $\{\mathcal{M}\}$ (for regular $\mathcal{M}$), in this instance using Dynkin's almost-analytic extension of ultradifferentiable functions (i.e., Corollary 2.11).

However, since regular weight sequences also fulfill the conditions of Hörmander we can use all of his results on $\operatorname{WF}_{\mathcal{M}}$. Indeed, in terms of L_N, we have that (M4) implies that $k \leq \gamma L_k$ for all $k \in \mathbb{N}_0$ and a constant $\gamma > 0$ independent of k by Sterling's formula whereas (M2) is equivalent to the existence of a constant $A > 0$ such that $L_k \leq A L_{k-1}$. We note that the last estimate implies $L_N \leq A^N$ for $N \in \mathbb{N}_0$ since $L_1 = 1$. On the other hand, it is well known that if $(M_N)_N$ satisfies (M3) then $(L_N)_N$ is an increasing sequence; see, e.g., [Mandelbrojt 1952].

The following result shows we may choose the distributions u_N in Definition 3.2 in a special manner.

Proposition 3.4 [Hörmander 1983, Lemma 8.4.4]. *Let $u \in \mathcal{D}'(\Omega)$ and let $K \subset \Omega$ be compact, let $F \subseteq \mathbb{R}^n$ be a closed cone such that $\operatorname{WF}_{\mathcal{M}} u \cap (K \times F) = \varnothing$. If $\chi_N \in \mathcal{D}(K)$ and for all α*

$$|D^{\alpha+\beta}\chi_N| \leq C_\alpha h_\alpha^{|\beta|} M_N^{|\beta|/N}, \quad |\beta| \leq N,$$

for some constants $C_\alpha, h_\alpha > 0$ then it follows that $\chi_N u$ is bounded in $\mathcal{E}'^S$ if u is of order S in a neighborhood of K, and further

$$|\widehat{\chi_N u}(\xi)| \leq C \frac{Q^N M_N}{|\xi|^N}, \quad N \in \mathbb{N},\ \xi \in F,$$

for some constants $C, Q > 0$.

We summarize the basic properties of $\operatorname{WF}_{\mathcal{M}}$ according to Hörmander [1983].

Here $\operatorname{sing\,supp} u \subseteq \Omega$ is defined to be the complement of the largest open subset $V \subseteq \Omega$ with $u|_V \in \mathcal{E}_{\mathcal{M}}(V)$.

Theorem 3.5 [Hörmander 1983, Theorems 8.4.5–8.4.7]. *Let $u \in \mathcal{D}'(\Omega)$ and $\mathcal{M}, \mathcal{N}$ be two weight sequences. Then*:

(1) $\operatorname{WF}_{\mathcal{M}} u$ *is a closed conic subset of* $\Omega \times \mathbb{R}^n \setminus \{0\}$.

(2) *The projection of* $\operatorname{WF}_{\mathcal{M}} u$ *in* Ω *is*

$$\pi_1(\operatorname{WF}_{\mathcal{M}} u) = \operatorname{sing\,supp}_{\mathcal{M}} u.$$

(3) $\operatorname{WF} u \subseteq \operatorname{WF}_{\mathcal{N}} u \subseteq \operatorname{WF}_{\mathcal{M}} u$ *if* $\mathcal{M} \preccurlyeq \mathcal{N}$.

(4) *If* $P = \sum p_\alpha D^\alpha$ *is a partial differential operator with ultradifferentiable coefficients of class* $\{\mathcal{M}\}$ *then* $\mathrm{WF}_{\mathcal{M}}\, Pu \subseteq \mathrm{WF}_{\mathcal{M}}\, u$.

Additionally we note that $\mathrm{WF}_{\mathcal{M}}\, u$ satisfies the following *microlocal reflection property*:

$$(x, \xi) \notin \mathrm{WF}_{\mathcal{M}}\, u \Longleftrightarrow (x, -\xi) \notin \mathrm{WF}_{\mathcal{M}}\, \bar{u}. \tag{3-3}$$

In particular, if u is a real-valued distribution, that is, $\bar{u} = u$, then $\mathrm{WF}_{\mathcal{M}}\, u|_x := \{\xi \in \mathbb{R}^n \mid (x, \xi) \in \mathrm{WF}_{\mathcal{M}}\, u\}$ is symmetric at the origin.

Example 3.6. It is easy to see that $\mathrm{WF}_{\mathcal{M}}\, \delta_p = \{p\} \times \mathbb{R}^n \backslash \{0\}$ for any regular weight sequence $\mathcal{M}$.

Remark 3.7. The complicated form of Definition 3.2 compared with the definition of the smooth wavefront set stems from the fact that quasianalytic weight sequences are allowed. Thus in general there may not be any nontrivial test functions of class $\{\mathcal{M}\}$. However if $\mathcal{D}_{\mathcal{M}} \neq \{0\}$ then we can choose in Definition 3.2 the constant sequence $u_N = \varphi u$ for some $\varphi \in \mathcal{D}_{\mathcal{M}}(\Omega)$ with $\varphi(x_0) = 1$, and (3-1) is equivalent to the existence of constants $C, Q > 0$ such that

$$|\widehat{\varphi u}(\xi)| \leq C \inf_N Q^N M_N |\xi|^{-N} \quad \text{for all } \xi \in \Gamma;$$

thus (2-3) implies

$$|\widehat{\varphi u}(\xi)| \leq C \tilde{h}_{\mathcal{M}}\left(\frac{Q}{|\xi|}\right) \leq C \exp\left(-\omega_{\mathcal{M}}\left(\frac{|\xi|}{Q}\right)\right).$$

We conclude (see, e.g., [Rodino 1993] in the case of Gevrey-classes) that for nonquasianalytic weight sequences $\mathcal{M}$, (3-1) is equivalent to

$$\sup_{\xi \in \Gamma} e^{\omega_{\mathcal{M}}(Q|\xi|)} |\widehat{\varphi u}(\xi)| < \infty \quad \text{for some } Q > 0.$$

Proposition 3.1 is then only a restatement of the well-known fact that for nonquasianalytic weight sequences we have that $\varphi \in \mathcal{D}_{\mathcal{M}}$ if and only if $\hat{\varphi} \leq C e^{-\omega_{\mathcal{M}}(Q|\xi|)}$ for some constants C, Q. Therefore it is possible to define ultradifferentiable classes using appropriately defined weight functions instead of weight sequences; see, e.g., in a somehow generalized setting, [Björck 1966]. However, this approach leads only to nonquasianalytic spaces. This restriction was removed by Braun, Meise and Taylor [Braun et al. 1990], who reformulated the defining estimates of these classes to allow also quasianalytic classes. A wavefront set relative to these classes was introduced in [Albanese et al. 2010]; see Section 6. The complicated connection between the classes defined by weight sequences and those given by weight functions was investigated in Bonet, Meise and Melikhov [Bonet et al. 2007]. Recently a new approach to define spaces of ultradifferentiable functions was introduced in [Rainer and Schindl 2014], which encompasses the classes given by weight sequences and weight functions; see also [Rainer and Schindl 2016].

4. Invariance of the wavefront set under ultradifferentiable mappings

Our aim in this section is to develop, using the almost-analytic extension of functions in $\mathcal{E}_{\mathcal{M}}$ given by Dynkin, a geometric description of $\mathrm{WF}_{\mathcal{M}}$ similar to the one that was presented, e.g., by Liess [1999, Section 4], for the smooth wavefront set.

We need to fix some notation: If $\Gamma \subseteq \mathbb{R}^d$ is a cone and $r > 0$ then

$$\Gamma_r := \{y \in \Gamma \mid |y| < r\}.$$

If $\Gamma' \subseteq \Gamma$ is also a cone we write $\Gamma' \Subset \Gamma$ if and only if $(\Gamma' \cap S^{d-1}) \Subset (\Gamma \cap S^{d-1})$.

Analogous to Liess [Liess 1999, Section 2.1] in the smooth category we say that, if $\mathcal{M}$ is a weight sequence, a function $F \in \mathcal{E}(\Omega \times U \times \Gamma_r)$, $U \subseteq \mathbb{R}^d$ open, is *$\mathcal{M}$-almost analytic* in the variables $(x, y) \in U \times \Gamma_r$ with parameter $x' \in \Omega$ if and only if for all $K \Subset \Omega$, $L \Subset U$ and cones $\Gamma' \Subset \Gamma$ there are constants $C, Q > 0$ such that for some r' we have

$$\left|\frac{\partial F}{\partial \bar{z}_j}(x', x, y)\right| \le C h_{\mathcal{M}}(Q|y|), \quad (x', x, y) \in K \times L \times \Gamma'_{r'},\ j = 1, \dots, d, \tag{4-1}$$

where $\partial/\partial \bar{z}_j = \frac{1}{2}(\partial_{x_j} + i\partial_{y_j})$ and $h_{\mathcal{M}}$ is the weight associated to the regular weight sequence $\mathcal{M}$ as defined by (2-2).

We may also say generally that a function $g \in \mathcal{C}(\Omega \times U \times \Gamma_r)$ is of *slow growth* in $y \in \Gamma_r$ if for all $K \Subset \Omega$, $L \Subset U$ and $\Gamma' \Subset \Gamma$ there are constants $c, k > 0$ such that

$$|g(x', x, y)| \le c|y|^{-k}, \quad (x', x, y) \in K \times L \times \Gamma'_r. \tag{4-2}$$

The next theorem is a generalization of [Hörmander 1983, Theorem 4.4.8]; see [Adwan and Hoepfner 2015].

Theorem 4.1. *Let $F \in \mathcal{E}(\Omega \times U \times \Gamma_r)$ be $\mathcal{M}$-almost analytic in the variables $(x, y) \in U \times \Gamma_r$ and of slow growth in the variable $y \in \Gamma_r$. Then the distributional limit u of the sequence $u_\varepsilon = F(\,.\,,\,.\,,\varepsilon) \in \mathcal{E}(\Omega \times U)$ exists. We say that $u = b_\Gamma(F) \in \mathcal{D}'(\Omega \times U)$ is the boundary value of F. Furthermore, we have*

$$\mathrm{WF}_{\mathcal{M}}\, u \subseteq (\Omega \times U) \times (\mathbb{R}^n \times \Gamma^\circ),$$

where $\Gamma^\circ = \{\eta \in \mathbb{R}^d \mid \langle y, \eta \rangle \ge 0 \text{ for all } y \in \Gamma\}$ is the dual cone of Γ in $\mathbb{R}^d$.

Proof. Let $\varphi \in \mathcal{D}(\Omega \times U)$ and $Y_0 \in \Gamma_\delta$. Then there are $K \Subset \Omega$ and $L \Subset U$ such that $\operatorname{supp} \varphi \subseteq K \times L$, and constants $c, k > 0$ exist such that (4-2) holds. We set

$$\Phi_\kappa(x', x, y) = \sum_{|\alpha| \le \kappa} \partial_x^\alpha \varphi(x', x) \frac{(iy)^\alpha}{\alpha!}$$

for $\kappa \ge k$. Obviously $F \cdot \Phi_\kappa$ can be extended to a smooth function on $\mathbb{R}^n \times \mathbb{R}^d \times \Gamma_\delta$

that vanishes outside $K \times L \times \Gamma_\delta$. We consider the function

$$u_\varepsilon : \mathbb{R}^2 \ni (\sigma, \tau) \longmapsto F(x', \tilde{x} + \sigma Y_0, \varepsilon + \tau Y_0)\Phi_\kappa(x', \sigma Y_0, \tau Y_0),$$

where $x' \in \mathbb{R}^n$, $\tilde{x} \in Y_0^\perp = \{z \in \mathbb{R}^d \mid \langle z, Y_0\rangle = 0\}$. If $a < b$ are chosen such that $\varphi(x', \tilde{x} + \sigma Y_0) = 0$ for all $x' \in \mathbb{R}^n$, $\tilde{x} \in Y_0^\perp$ and $\sigma \le a$ or $\sigma \ge b$ then $u_\varepsilon(\sigma, \tau) = 0$ for all $\tau \in [0, 1]$. If $R = [a, b] \times [0, 1]$ then Stokes' theorem states that

$$\int_{\partial R} u_\varepsilon \, d\zeta = \int_R \frac{\partial u_\varepsilon}{\partial \bar{\zeta}} \, d\bar{\zeta} \wedge d\zeta, \tag{4-3}$$

where we have set $\zeta = \sigma + i\tau$.

A simple computation gives

$$2i \frac{\partial}{\partial \bar{\zeta}} \big(\Phi_\kappa(x', \tilde{x} + \sigma Y_0, \tau Y_0)\big) = (\kappa + 1)\tau^\kappa \sum_{|\alpha| = \kappa+1} \partial_x^\alpha \varphi(x', \tilde{x} + \sigma Y_0) \frac{(iY_0)^\alpha}{\alpha!}.$$

Hence formula (4-3) means in detail that

$$\begin{aligned}
&\int_a^b F(x', \sigma Y_0, \varepsilon)\varphi(x', \sigma Y_0)\, d\sigma \\
&\quad = \int_a^b F(x', \sigma Y_0, \varepsilon + Y_0)\Phi_\kappa(x', \sigma Y_0, Y_0)\, d\sigma \\
&\qquad + 2i \int_a^b \int_0^1 \langle \bar{\partial} F(x', \sigma Y_0, \varepsilon + \tau Y_0), Y_0\rangle \Phi_\kappa(x', \sigma Y_0, \tau Y_0)\, d\tau d\sigma \\
&\qquad + (\kappa + 1) \int_a^b \int_0^1 F(x', \sigma Y_0, \varepsilon + \tau Y_0)\tau^\kappa \sum_{|\alpha| = \kappa+1} \frac{\partial_x^\alpha \varphi}{\beta!}\, d\tau d\sigma
\end{aligned}$$

and thus integrating over $\Omega \times Y_0^\perp$ yields

$$\begin{aligned}
&\int_{\Omega \times U} F(x', x, \varepsilon)\varphi(x', x)\, d\lambda(x', x) \\
&\quad = \int_{\Omega \times U} F(x', x, \varepsilon + Y_0)\Phi_\kappa(x', x, Y_0)\, d\lambda(x', x) \\
&\qquad + 2i \int_{\Omega \times U} \int_0^1 \langle \bar{\partial} F(x', x, \varepsilon + \tau Y_0), Y_0\rangle \Phi_\kappa(x', x, \tau Y_0)\, d\tau d\lambda(x', x) \\
&\qquad + (\kappa + 1) \int_{\Omega \times U} \int_0^1 F(x', x, \varepsilon + \tau Y_0)\tau^\kappa \sum_{|\alpha| = \kappa+1} \partial_x^\alpha \varphi(x', x) \frac{(iY_0)^\alpha}{\alpha!} d\lambda(x', x).
\end{aligned} \tag{4-4}$$

Since by assumption $|\tau^\kappa F(x', x, \varepsilon + \tau Y_0)| \le c$ for some constant c and $\bar{\partial}_j F$ decreases rapidly for $\Gamma_r \ni y \to 0$ (see the remarks after Lemma 2.8) the bounded convergence theorem implies that the right-hand side converges for $\varepsilon \to 0$. Hence we define

$$\begin{aligned}
\text{(4-5)}\quad \langle u,\varphi\rangle :=&\\
&\int_{\Omega\times U} F(x',x,Y_0)\Phi_\kappa(x',x,Y_0)\,d\lambda(x',x)\\
&\quad + 2i\int_{\Omega\times U}\int_0^1 \langle\bar\partial F(x',x,\tau Y_0),Y_0\rangle\Phi_\kappa(x',x,\tau Y_0)\,d\tau d\lambda(x',x)\\
&\quad + (\kappa+1)\int_{\Omega\times U}\int_0^1 F(x',x,\tau Y_0)\tau^\kappa \sum_{|\alpha|=\kappa+1}\partial_x^\alpha\varphi(x',x)\frac{(iY_0)^\alpha}{\alpha!}\,d\tau d\lambda(x',x).
\end{aligned}$$

Since there is a constant $\tilde C$ only depending on F and $K\times L$ such that

$$|\langle u,\varphi\rangle|\le \tilde C \sup_{(x',x)\in K\times L}\biggl(\sum_{|\beta|\le\kappa+1}|\partial_x^\beta\varphi(x',x)|\biggr),$$

we deduce that the linear form u on $\mathcal{D}(\Omega\times U)$ given by (4-5) is a distribution.

Now, let $p_0\in\Omega\times U$ and $\omega_2\times V_2\Subset\omega_1\times V_1\Subset\Omega\times U$ be two open neighborhoods of p_0. Using [Hörmander 1983, Theorem 1.4.2] we can choose a sequence $(\varphi_\kappa)_\kappa\subset\mathcal{D}(\omega_1\times V_1)$ such that $\varphi_\kappa|_{\omega_2\times V_2}\equiv 1$ and for all $\gamma\in\mathbb{N}_0^{n+d}$ we have that

$$\text{(4-6)}\qquad |D^{\gamma+\beta}\varphi_\kappa|\le (C_\gamma(\kappa+1))^{|\beta|},\quad |\beta|\le\kappa+1,$$

for a constant $C_\gamma\ge 1$ independent of κ. As before we set for each κ

$$\Phi_\kappa(x',x,y)=\sum_{|\alpha|\le\kappa}\partial_x^\alpha\varphi_\kappa(x',x)\frac{(iy)^\alpha}{\alpha!}.$$

We aim to estimate $\widehat{\varphi_\kappa u}$. In order to do so, let $(\xi,\eta)\in\mathbb{R}^n\times\mathbb{R}^d$ and notice that (4-5) implies for $\kappa\ge k$,

$$\begin{aligned}
\widehat{\varphi_\kappa u}(\xi,\eta) &= \langle u, e^{-i\langle\,.\,,(\xi,\eta)\rangle}\varphi_\kappa\rangle\\
&= \int_{\Omega\times U} F(x',x,Y_0)e^{-i(x'\xi+(x+iY_0)\eta)}\Phi_\kappa(x',x,Y_0)\,d\lambda(x',x)\\
&\quad + 2i\int_{\Omega\times U}\int_0^1\langle\bar\partial F(x',x,\tau Y_0),Y_0\rangle e^{-i(x'\xi+(x+i\tau Y_0)\eta)}\\
&\qquad\qquad \times\Phi_\kappa(x',x,\tau Y_0)\,d\tau d\lambda(x',x)\\
&\quad + (\kappa+1)\int_{\Omega\times U}\int_0^1 F(x',x,\tau Y_0)e^{-i(x'\xi+(x+i\tau Y_0)\eta)}\tau^\kappa\\
&\qquad\qquad \times\sum_{|\alpha|=\kappa+1}\partial_x^\alpha\varphi(x',x)\frac{(iY_0)^\alpha}{\alpha!}\,d\tau d\lambda(x',x)
\end{aligned}$$

for some fixed, but arbitrary $Y_0\in\Gamma_r$ (note that k depends on u, $\omega_1\times V_1$ and Y_0).

Condition (4-6) gives the following estimate for $0 \le \mu \le \kappa + 1$:

$$\left| \sum_{|\alpha|=\mu} \partial_x^\alpha \varphi_\kappa(x', x) \frac{(iY)^\alpha}{\alpha!} \right| \le C_0^\mu (\kappa+1)^\mu \sum_{|\alpha|=\mu} \frac{|Y^\alpha|}{\alpha!} = C_0^\mu (\kappa+1)^\mu \frac{|Y|_1^\mu}{\mu!},$$

where $|Y|_1 = \sum_j |Y_j|$ for $Y = (Y_1, \dots, Y_d) \in \mathbb{R}^d$. Hence we have

$$|\Phi_\kappa(x', x, \tau Y_0)| \le C_1^{\kappa+1},$$

$$\left| (\kappa+1) \sum_{|\alpha|=\kappa+1} \partial_x^\alpha \varphi_\kappa(x', x) \frac{(iY_0)^\alpha}{\alpha!} \right| \le C_1^{\kappa+1},$$

for $C_1 = 2e^{C_0|Y_0|_1}$ and $\tau \in [0, 1]$. We obtain

$$\begin{aligned} |\widehat{\varphi_\kappa u}(\xi, \eta)| & \\ &\le C_1^{\kappa+1} e^{\eta Y_0} + 2C_1^{\kappa+1} C \int_0^1 h_{\mathcal{M}}(Q\tau|Y_0|) e^{\tau\eta Y_0}\, d\tau + C_1^{\kappa+1} \int_0^1 \tau^{\kappa-k} e^{\tau\eta Y_0}\, d\tau \\ &\le C_2 Q_1^\kappa \left(e^{\eta Y_0} + m_{\kappa-k} \int_0^1 \tau^{\kappa-k} e^{\eta Y_0} \right) \\ &= C_2 Q_1^\kappa \big(e^{\eta Y_0} + m_\kappa (\kappa-k)! (-Y_0\eta)^{k-\kappa-1} \big) \end{aligned}$$

for some constants C_2, Q_1 and $Y_0\eta < 0$. If we set $\tilde{Y}_0 = (0, Y_0) \in \mathbb{R}^n \times \mathbb{R}^d$ then obviously

$$\langle \tilde{Y}_0, (\xi, \eta) \rangle = \langle Y_0, \eta \rangle.$$

Therefore we have for $\kappa \ge k$ and $\zeta = (\xi, \eta)$ that

$$|\widehat{\varphi_\kappa u}(\zeta)| = C_3 Q_1^\kappa \big(e^{\tilde{Y}_0\zeta} + m_{\kappa-k} (\kappa-k)! (-\tilde{Y}_0\zeta)^{k-\kappa-1} \big)$$

and $\tilde{Y}_0\zeta < 0$.

Now for any $\zeta_0 \in \mathbb{R}^{n+d}$ with $\langle \tilde{Y}_0, \zeta_0 \rangle < 0$ we can choose an open cone $V \subseteq \mathbb{R}^{n+d}$ such that $\zeta_0 \in V$ and for some constant $c > 0$ we have $\langle \tilde{Y}_0, \zeta \rangle < -c|\zeta|$ if $\zeta \in V$. Furthermore we set $u_\kappa = \varphi_{k+\kappa-1} u$. Clearly the sequence $(u_\kappa)_\kappa$ is bounded in $\mathcal{E}'(\Omega \times U)$ and $u_\kappa|_{\omega_2 \times V_2} \equiv u|_{\omega_2 \times V_2}$. Also using the inequality $e^{-c|\zeta|} \le \kappa!(c|\zeta|)^{-\kappa}$ we conclude

$$|\hat{u}_\kappa(\zeta)| = C_3 Q_1^\kappa \big(\kappa!(c|\zeta|)^{-\kappa} + m_{\kappa-1}(\kappa-1)!(c|\zeta|)^{-\kappa} \big) \le C_3 Q_2^\kappa m_\kappa \kappa! |\zeta|^{-\kappa}, \quad \zeta \in V.$$

Hence $(p_0, \zeta_0) \notin \mathrm{WF}_{\mathcal{M}}\, u$ and therefore

$$\mathrm{WF}_{\mathcal{M}}\, u \subseteq (\Omega \times U) \times (\mathbb{R}^n \times \Gamma^\circ) \setminus \{(0, 0)\}. \qquad \square$$

It is clear that the proof would only require $F \in \mathcal{C}^1$. From now the constants used in the proofs will be generic, i.e., they may change from line to line.

Remark 4.2. If $F \in \mathcal{E}(\Omega \times U \times V)$ is $\mathcal{M}$-almost analytic with respect to the variables $(x, y) \in U \times V$ we will often write $F(x', x+iy)$ or $F(x', z, \bar{z})$ and consider F as a smooth function on $\Omega \times (U + iV)$. If $\Omega = \varnothing$ then we just say that F is $\mathcal{M}$-almost analytic.

Even though in the remainder of this paper we shall only use the assertion of Lemma 4.5 in the special case $\Omega = \varnothing$ (i.e., without parameters), we have decided to include the general statement because we think it is of independent interest. We also have an application for the parameter version of the theorem in our paper [Fürdös 2019].

Example 4.3. Consider the holomorphic function $F(z) = \frac{1}{z}$ on $\mathbb{C}\backslash\{0\}$. It is well known that the boundary values of F onto the real line from above and beneath, commonly denoted by

$$\frac{1}{x+i0} = b_+ F = \lim_{y\to 0+} \frac{1}{x+iy},$$

$$\frac{1}{x-i0} = b_+ F = \lim_{y\to 0+} \frac{1}{x-iy},$$

satisfy the jump relations (see, e.g., [Duistermaat and Kolk 2010]); in particular,

$$2i\delta = \frac{1}{x-i0} - \frac{1}{x+i0}.$$

We have that both $1/(x+i0)$ and $1/(x-i0)$ are real-analytic outside the origin. Hence the application of Theorem 4.1 together with the jump relations imply that

$$\mathrm{WF}_{\mathcal{M}}\left(\frac{1}{x \pm i0}\right) = \{0\} \times \mathbb{R}_{\pm}.$$

There is a partial converse to the last theorem.

Theorem 4.4. *Let $\Gamma \subseteq \mathbb{R}^n$ be an open convex cone and $u \in \mathcal{D}'(\Omega)$ with $\mathrm{WF}_{\mathcal{M}}\, u \subseteq \Omega \times \Gamma^\circ$. If $V \Subset \Omega$ and Γ' is an open convex cone with $\overline{\Gamma}' \subseteq \Gamma \cup \{0\}$ then there is an $\mathcal{M}$-almost analytic function F on $V + i\Gamma'_r$ of slow growth for some $r > 0$ such that $u|_V = b_{\Gamma'}(F)$*

Proof. By [Hörmander 1983, Theorem 8.4.15] we have that u can be written on a bounded neighborhood U of V as a sum of a function $f \in \mathcal{E}_{\mathcal{M}}(U)$ and the boundary value of a holomorphic function of slow growth on $U + i\Gamma'_r$ for some r. To obtain the assertion use Corollary 2.11 to extend f almost-analytically on V. □

In order to proceed we need a further refinement of a result of Hörmander.

Lemma 4.5. *Let $\Gamma_j \subseteq \mathbb{R}^n\backslash\{0\}$, $j = 1, \dots, N$, be closed cones such that $\bigcup_j \Gamma_j = \mathbb{R}^n\backslash\{0\}$ and $V \Subset \Omega$ is convex. Any $u \in \mathcal{D}'(\Omega)$ can be written on V as a linear* *combination* $u|_V = \sum_j u_j$ *of distributions* $u_j \in \mathcal{D}'(V)$ *that satisfy*

$$\mathrm{WF}_{\mathcal{M}}\, u_j \subseteq \mathrm{WF}_{\mathcal{M}}\, u \cap (V \times \Gamma_j)$$

Proof. Set $v = \varphi u$ where $\varphi \in \mathcal{D}(\Omega)$ such that $\varphi \equiv 1$ on V. The existence of $v_j \in \mathcal{S}'(\mathbb{R}^n)$ such that

$$\mathrm{WF}_{\mathcal{M}}\, v_j \subseteq \mathrm{WF}_{\mathcal{M}}\, v \cap (\mathbb{R}^n \times \Gamma_j)$$

is given in [Hörmander 1983, Corollary 8.4.13]. Set $u_j = (v_j)|_V$. □

Combining Theorem 4.4 with Lemma 4.5 we obtain:

Corollary 4.6. *Let* $u \in \mathcal{D}'(\Omega)$ *and* $(x_0, \xi_0) \in \Omega \times \mathbb{R}^n \backslash \{0\}$. *Then* $(x_0, \xi_0) \notin \mathrm{WF}_{\mathcal{M}}\, u$ *if and only if there are a neighborhood* U *of* x_0, *open convex cones* $\Gamma_1, \dots, \Gamma_N$ *with the properties* $\xi_0 \Gamma_j < 0$, $j = 1, \dots N$ *and* $\Gamma_j \cap \Gamma_k = \varnothing$ *for* $j \neq k$, *and* $\mathcal{M}$*-almost analytic functions* h_j *on* $U + i\Gamma_{r_j}$, $r_j > 0$, *of slow growth such that*

$$u|_U = \sum_{j=1}^{N} b_{\Gamma_j}(h_j).$$

Proof. Without loss of generality, assume that $\mathrm{WF}_{\mathcal{M}}\, u \neq \varnothing$. If $(x_0, \xi_0) \notin \mathrm{WF}_{\mathcal{M}}\, u$ one can find closed cones $V_1, \dots, V_N$ with nonempty interior and $V_j \cap V_k$ has measure zero for $j \neq k$ such that ξ_0 is contained in the interior of V_1 and $V_1 \cap \mathrm{WF}_{\mathcal{M}}\, u = \varnothing$ whereas $\xi_0 \notin V_j$ are acute cones and $\mathrm{WF}_{\mathcal{M}}\, u \cap V_j \neq \varnothing$ for $j = 2, \dots, N$. By Lemma 4.5 we can write u on an open neighborhood U of x_0 as a sum $u = u_1 + \sum_{j=2}^{N} u_j$ with u_1 being an ultradifferentiable function defined on U and $u_j \in \mathcal{D}'(U)$ such that $\mathrm{WF}_{\mathcal{M}}\, u_j \subseteq \mathrm{WF}_{\mathcal{M}}\, u \cap V_j$, $j = 2, \dots, N$. The cones $V_2, \dots, V_N$ are the dual cones of open convex cones $\Gamma_2, \dots, \Gamma_N$, i.e., $\Gamma_j^\circ = V_j$. We can choose cones $\Gamma'_j \Subset \Gamma_j$ and using Theorem 4.4 we find $\mathcal{M}$-almost analytic functions h_j on $U + i\Gamma'_{j,r}$ of slow growth such that $u_j = b_{\Gamma'_j}(h_j)$. It remains to note that $\xi_0 y < 0$ for all $y \in \Gamma'_j$, $j = 2, \dots, N$. □

Let $\Omega_1 \subseteq \mathbb{R}^m$ and $\Omega_2 \subseteq \mathbb{R}^n$ be open. If $F : \Omega_1 \to \Omega_2$ is a $\mathcal{E}_{\mathcal{M}}$-mapping then we denote as in [Hörmander 1983, page 263] the set of normals by

$$N_F = \{(F(x), \eta) \in \Omega_2 \times \mathbb{R}^n : \ DF(x)\eta = 0\},$$

where DF denotes the transpose of the Jacobian of F. The following is a generalization of [Hörmander 1983, Theorem 8.5.1].

Theorem 4.7. *For any* $u \in \mathcal{D}'(\Omega_2)$ *with* $N_F \cap \mathrm{WF}_{\mathcal{M}}\, u = \varnothing$ *we obtain that the pull-back* $F^*u \in \mathcal{D}'(\Omega_1)$ *is well defined and*

$$\mathrm{WF}_{\mathcal{M}}(F^*u) \subseteq F^*(\mathrm{WF}_{\mathcal{M}}\, u). \tag{4-7}$$

Proof. The first part of the statement is [Hörmander 1983, Theorem 8.2.4]. For the proof of the second part of the theorem assume first that there is an open convex cone Γ such that u is the boundary value of an $\mathcal{M}$-almost analytic function Φ on $\Omega_2 + i\Gamma_r$ of slow growth. Hence $\mathrm{WF}_{\mathcal{M}}\, u \subseteq \Omega_2 \times \Gamma^\circ$. If $x_0 \in \Omega_1$ and $DF(x_0)\eta \neq 0$ for $\eta \in \Gamma^\circ \setminus \{0\}$ then $DF(x_0)\Gamma^\circ$ is a closed convex cone. We claim that

$$\mathrm{WF}_{\mathcal{M}}(F^* u)|_{x_0} \subseteq \{(x_0, DF(x_0)\eta) : \eta \in \Gamma^\circ \setminus \{0\}\}.$$

We adapt as usual the argument of Hörmander [1983]. We can write (see [Hörmander 1983, page 296])

$$DF(x_0)\Gamma^\circ = \{\xi \in \mathbb{R}^n \mid \langle h, \xi\rangle \geq 0,\ F'(x_0)h \in \Gamma\}.$$

If $\tilde{F}$ denotes an $\mathcal{M}$-almost analytic extension of F onto $X_0 + i\mathbb{R}^n$, where the neighborhood X_0 of x_0 is convex and relatively compact in Ω_1, which exists due to Theorem 2.10, then Taylor's formula implies that

$$\operatorname{Im} \tilde{F}(x + i\varepsilon h) \in \Gamma, \quad x \in X_0,$$

for $F'(x_0)h \in \Gamma$ if X_0 and $\varepsilon > 0$ are small.

Recalling (4-4) we see that the map

$$\mathbb{R}_{\geq 0} \times (\Gamma \cup \{0\}) \ni (\varepsilon, y) \longmapsto \tilde{\Phi}(\varepsilon, y) := \Phi(\tilde{F}(\,.\, + i\varepsilon h) + iy) \in \mathcal{D}'(X_0)$$

is continuous. If $\varepsilon \to 0$ then $\tilde{\Phi}(\varepsilon, y) \to \tilde{\Phi}(0, y) = \Phi(\tilde{F}(\,.\, + 0i) + iy)$ in $\mathcal{D}'$ and if now $y \to 0$ we have by definition $\tilde{\Phi}(0, y) \to F^* u$. On the other hand if first $y \to 0$ then

$$\tilde{\Phi}(\varepsilon, y) \to \tilde{\Phi}(\varepsilon, 0) = \Phi(\tilde{F}(\,.\, + i\varepsilon h)).$$

Hence by continuity

$$F^* u = \lim_{\varepsilon \to 0} \Phi(\tilde{F}(\,.\, + i\varepsilon h))$$

in $\mathcal{D}'(X_0)$ and by the proof of Theorem 4.1,

$$\mathrm{WF}_{\mathcal{M}}\, F^* u|_{x_0} \subseteq \{(x_0, \xi) \mid \langle h, \xi\rangle \geq 0\}.$$

The claim follows.

Now suppose that $(F(x_0), \eta_0) \notin \mathrm{WF}_{\mathcal{M}}\, u$. By Corollary 4.6 we can write a general distribution u on some neighborhood U_0 of $F(x_0)$ as $\sum_{j=1}^N u_j$ where the distributions u_j, $j = 1, \dots, N$, are the boundary values of some $\mathcal{M}$-almost analytic functions Φ_j on $U_0 + i\Gamma_j$, where the Γ_j are some open convex cones such that $\eta_0 \Gamma_j < 0$ for all $j = 1, \dots, N$. By assumption $DF(x)\eta \neq 0$ when $(F(x), \eta) \in \mathrm{WF}_{\mathcal{M}}\, u$ for $x \in F^{-1}(U_0)$. Hence we can assume that $DF(x)\eta \neq 0$ for $\eta \in \Gamma_j^\circ$ for all $j = 1, \dots, N$ and $x \in F^{-1}(U_0)$ since in the proof of Corollary 4.6 the cones Γ_j, $j = 1, \dots, N$, can be chosen such that $\Gamma^\circ \cap S^{n-1}$ and $\Gamma_j^\circ \cap \mathrm{WF}_{\mathcal{M}}\, u|_x \neq \varnothing$

for $x \in U_0$. By the arguments above we have for a small neighborhood V of x_0 that

$$F^*u|_V = \sum_{j=1}^{N} F^*u_j|_V$$

and $\mathrm{WF}_{\mathcal{M}}(F^*u_j)|_{x_0} \subseteq \{(x_0, DF(x_0)\eta) \mid \eta \in \Gamma_j^\circ \setminus \{0\}\}$ for all j. However, since $\eta_0\Gamma_j < 0$ it follows that $(x_0, DF(x_0)\eta_0) \notin \mathrm{WF}_{\mathcal{M}}(F^*u_j)$ and so $(x_0, DF(x_0)\eta_0) \notin \mathrm{WF}_{\mathcal{M}}(F^*u)$. $\square$

Remark 4.8. If F is an $\mathcal{E}_{\mathcal{M}}$-diffeomorphism we obtain from Theorem 4.7 that

$$\mathrm{WF}_{\mathcal{M}}(F^*u) = F^*(\mathrm{WF}_{\mathcal{M}}\, u).$$

Hence if M is an $\mathcal{E}_{\mathcal{M}}$-manifold and $u \in \mathcal{D}'(M)$ we can define $\mathrm{WF}_{\mathcal{M}}\, u$ invariantly as a subset of $T^*M\backslash\{0\}$. More precisely, there is a subset K_u of T^*M such that the diagram

$$\begin{array}{ccc} & K_u & \\ \swarrow & & \searrow \\ T^*\varphi(U\cap V) \supseteq \mathrm{WF}_{\mathcal{M}}\, v_1 & \xrightarrow{\ \rho^*\ } & \mathrm{WF}_{\mathcal{M}}\, v_2 \subseteq T^*\psi(U\cap V) \end{array}$$

commutes for any two charts φ and ψ of M on $U \subseteq M$ and $V \subseteq M$, respectively. We have set $\rho = \psi \circ \varphi^{-1}$, $v_1 = \varphi^* u \in \mathcal{D}'(\varphi(U\cap V))$ and $v_2 = \psi^* u \in \mathcal{D}'(\psi(U\cap V))$. It follows that $K_u \subseteq T^*M\setminus\{0\}$ has to be closed and fiberwise conic. We set $\mathrm{WF}_{\mathcal{M}}\, u := K_u$.

Analogously we define the wavefront set of a distribution $u \in \mathcal{D}'(M, E)$ with values in an ultradifferentiable vector bundle locally by setting

$$\mathrm{WF}_{\mathcal{M}}\, u|_V = \bigcup_{j=1}^{\nu} u_j,$$

where $V \subseteq M$ is an open subset such that there is a local basis $\omega^1, \dots, \omega^\nu$ of $\mathcal{E}_{\mathcal{M}}(V, E)$ and $u_j \in \mathcal{D}'(V)$ are distributions on V such that

$$u|_V = \sum_{j=1}^{\nu} u_j \omega^j.$$

We close this section by observing that Theorem 4.7 allows us to strengthen a uniqueness result of Boman [1995]:

Theorem 4.9. *Let $\mathcal{M}$ be a quasianalytic weight sequence and $S \subseteq \mathbb{R}^n$ an $\mathcal{E}_{\mathcal{M}}$-submanifold. If u is a distribution defined on a neighborhood of S such that*

$$\mathrm{WF}_{\mathcal{M}}\, u \cap N^*S = \varnothing$$

and

$$\partial^\alpha u|_S = 0 \quad \text{for all } \alpha \in \mathbb{N}_0^n,$$

then u vanishes on some neighborhood of S.

Indeed, locally S is diffeomorphic to

$$S' = \{(x', x'') \in \mathbb{R}^{m+d} \mid x'' = 0\} \subseteq \mathbb{R}^n$$

and the assumptions of the theorem translate to the corresponding conditions for the pullback $w = F^*u$ where $F : \mathbb{R}^n \to \mathbb{R}^n$ is the local $\mathcal{E}_\mathcal{M}$-diffeomorphism that maps S' to S. Then the proof of Theorem 1 in [Boman 1995] gives $w = 0$ in a neighborhood of S'.

5. A generalized version of Bony's Theorem

We have seen that for a distribution u the wavefront set $\mathrm{WF}_\mathcal{M}\, u$ can be described either using the Fourier transform or by its $\mathcal{M}$-almost analytic extensions. A similar fact is true for the analytic wavefront set using holomorphic extensions. The latter was the original approach of Sato [1970]. However, Bros and Iagolnitzer [1975] used the classical FBI transform to describe the set of microlocal analytic singularities. It was Bony [1977] who proved that all three methods actually describe the same set. In the ultradifferentiable setting Chung and Kim [1997] (see also Kim, Chung and Kim [Kim et al. 2001]) used the FBI transform to define an ultradifferentiable singular spectrum for Fourier hyperfunctions. However, they did not mention how this singular spectrum in the case of distributions may be related to $\mathrm{WF}_\mathcal{M}$ as defined by Hörmander. Our next aim is to show an ultradifferentiable version of Bony's theorem. We will work in the generalized setting of [Berhanu and Hounie 2012]. We shall note that recently Berhanu and Hailu [2017] showed that the Gevrey classes can be characterized by this generalized FBI transform and Hoepfner and Medrado [2018] also proved a characterization of the ultradifferentiable wavefront set of ultradistributions for a certain class of nonquasianalytic weight sequences.

Let p be a real, homogeneous, positive, elliptic polynomial of degree $2k$, $k \in \mathbb{N}$, on $\mathbb{R}^n$, i.e.,

$$p(x) = \sum_{\alpha = 2k} a_\alpha x^\alpha, \qquad a_\alpha \in \mathbb{R},$$

and let there be constants $c, C > 0$ such that

$$c|x|^{2k} \le p(x) \le C|x|^{2k}, \qquad x \in \mathbb{R}^n.$$

Let

$$c_p^{-1} = \int e^{-p(x)} dx.$$

As in [Berhanu and Hounie 2012, Section 4] we consider the generalized FBI transform with generating function e^{-p} of a distribution of compact support $u \in \mathcal{E}'(\mathbb{R}^n)$, i.e.,

$$\mathfrak{F}u(t, \xi) = c_p \langle u(x), e^{i\xi(t-x)} e^{-|\xi| p(t-x)} \rangle.$$

The inversion formula is

$$u = \lim_{\varepsilon\to\infty} \int_{\mathbb{R}^n\times\mathbb{R}^n} e^{i\xi(x-t)} e^{-\varepsilon|\xi|^2} \mathfrak{F}u(t,\xi)|\xi|^{n/(2k)}\,dt\,d\xi, \tag{5-1}$$

where of course the distributional limit is meant.

Theorem 5.1. *Let $u\in\mathcal{D}'(\Omega)$ and*

$$(x_0,\xi_0)\in T^*\Omega\setminus\{0\}.$$

Then $(x_0,\xi_0)\notin \mathrm{WF}_{\mathcal{M}}\,u$ if and only if there is a test function $\psi\in\mathcal{D}(\Omega)$ with $\psi|_U\equiv 1$ for some neighborhood U of x_0 such that

$$\sup_{(t,\xi)\in V\times\Gamma} e^{\omega_{\mathcal{M}}(\gamma|\xi|)}|\mathfrak{F}(\psi u)(t,\xi)|<\infty \tag{5-2}$$

for some conic neighborhood $V\times\Gamma$ of (x_0,ξ_0) and some constant $\gamma>0$.

Proof. We modify the proof of [Berhanu and Hounie 2012, Theorem 4.2]; see also the proof of [Hoepfner and Medrado 2018, Theorem 5.2].

First, assume that $(x_0,\xi_0)\notin\mathrm{WF}_{\mathcal{M}}\,u$. By Corollary 4.6 we know that for some neighborhood U of x_0,

$$u|_U=\sum_{j=1}^{N} b_{\Gamma^j}(F_j),$$

where F_j are $\mathcal{M}$-almost analytic on $U\times\Gamma^j_{r_j}$ for cones Γ^j with $\xi_0\Gamma^j<0$. Hence it suffices to prove the necessity of (5-2) for $u=b_\Gamma(F)$ being the boundary value of an $\mathcal{M}$-almost analytic function F on $U\times\Gamma_d$ where Γ is a cone with the property that $\xi_0\Gamma<0$. Without loss of generality, $x_0=0$, and let $r>0$ be such that $B_{2r}(0)\Subset U$ and $\psi\in\mathcal{D}(B_{2r}(0))$ be such that $\psi|_{B_r(0)}\equiv 1$. Furthermore we choose $v\in\Gamma_d$ and set

$$Q(t,\xi,x)=i\xi(t-x)-|\xi|p(t-x).$$

Then

$$\mathfrak{F}(\psi u)(t,\xi)=\lim_{\tau\to 0+}\int_{B_{2r}(0)} e^{Q(t,\xi,x+i\tau v)}\psi(x)F(x+i\tau v)\,dx.$$

As in the proof of Theorem 4.2 in [Berhanu and Hounie 2012] we put $z=x+iy$, $\psi(z)=\psi(x)$ and

$$D_\tau:=\{x+i\sigma v\in\mathbb{C}^n \mid x\in B_{2r}=B_{2r}(0),\ \tau\le\sigma\le\lambda\}$$

for some $\lambda>0$ to be determined later and consider the n-form

$$e^{Q(t,\xi,z)}\psi(z)F(z)\,dz_1\wedge\cdots\wedge dz_n.$$

Since $\psi \in \mathcal{D}(B_{2r}(0))$, Stokes' theorem implies

$$\begin{aligned}(5\text{-}3)\quad \int_{B_{2r}} e^{Q(t,\xi,x+i\tau v)}\psi(x)F(x+i\tau v)\,dx &= \int_{B_{2r}} e^{Q(t,\xi,x+i\lambda v)}\psi(x)F(x+i\lambda v)\,dx \\ &\quad + \sum_{j=1}^{n}\int_{D_\tau} e^{Q(t,\xi,z)}\frac{\partial}{\partial \bar z_j}(\psi(z)F(z))\,d\bar z_j\wedge dz_1\wedge\cdots\wedge dz_n \\ &= \int_{B_{2r}} e^{Q(t,\xi,x+i\lambda v)}\psi(x)F(x+i\lambda v)\,dx \\ &\quad + \sum_{j=1}^{n}\int_{B_{2r}}\int_{\tau}^{\lambda} e^{Q(t,\xi,x+i\sigma v)}\frac{\partial\psi}{\partial\bar z_j}(x+i\sigma v)F(x+i\sigma v)\,d\sigma\,dx \\ &\quad + \sum_{j=1}^{n}\int_{B_{2r}}\int_{\tau}^{\lambda} e^{Q(t,\xi,x+i\sigma v)}\psi(x+i\sigma v)\frac{\partial F}{\partial\bar z_j}(x+i\sigma v)\,d\sigma\,dx.\end{aligned}$$

We need to estimate the integrals on the right-hand side of (5-3). Using the arguments of [Berhanu and Hounie 2012] we see that there is an open cone Γ containing ξ_0 and a bounded neighborhood V of 0 such that the first two integrals can be estimated by $Ce^{-\gamma|\xi|}$ where $C,\gamma>0$ are constants, as long as $\xi\in\Gamma$ and $t\in V$. Since (M4) implies that $\omega_{\mathcal{M}}(t)=O(t)$ for $t\to\infty$; see, e.g., [Komatsu 1973] or [Bonet et al. 2007], we obtain that both integrals can in fact be bounded by $Ce^{-\omega_{\mathcal{M}}(\gamma|\xi|)}$ if $(t,\xi)\in V\times\Gamma$.

In order to estimate the third integral in (5-3) we recall from [Berhanu and Hounie 2012] that for λ small enough there is a constant $c_0>0$ such that

$$\operatorname{Re} Q(t,\xi,x+i\lambda v)\le -\frac{c_0}{2}\lambda|v||\xi|$$

if $\xi\in\Gamma,\ x\in B_{2r}$ and $t\in V$. Hence we have for a generic constant $C_3>0$ and all $k\in\mathbb{N}_0$ that

$$\begin{aligned}&\left|\sum_{j=1}^{n}\int_{B_{2r}}\int_{\tau}^{\lambda} e^{Q(t,\xi,x+i\sigma v)}\psi(x)\frac{\partial F}{\partial\bar z_j}(x+i\sigma v)\,d\sigma\,dx\right| \\ &\qquad\le C_3\int_0^\infty e^{-c_0\sigma|v||\xi|}h_{\mathcal{M}}(\rho\sigma|v|)\,d\sigma\le C_3\int_0^\infty e^{-c_0\sigma|v||\xi|}\rho^k\sigma^k|v|^k m_k\,d\sigma \\ &\qquad= C_3\rho_1^k M_k|\xi|^{-k}.\end{aligned}$$

Lemma 2.8 gives

$$\left|\sum_{j=1}^{n}\int_{B_{2r}}\int_{\tau}^{\lambda} e^{Q(t,\xi,x+i\sigma v)}\psi(x)\frac{\partial F}{\partial\bar z_j}(x+i\sigma v)\,d\sigma\,dx\right|\le C_3\tilde h_{\mathcal{M}}(\rho_1|\xi|^{-1})\le C_3 e^{-\omega_{\mathcal{M}}(\rho_1|\xi|)}.$$

In view of (5-3) we have shown that for $\xi\in\Gamma$ and t in a small enough neighborhood of 0 there are constants $C,\gamma>0$ such that

$$\left|\int_{B_{2r}} e^{Q(t,\xi,x+i\tau v)}\psi(x)F(x+i\tau v)\,dx\right|\le Ce^{-\omega_{\mathcal{M}}(\gamma|\xi|)}.$$

Note that in the estimate the constants C and γ depend on λ but not on $\tau<\lambda$. Thus (5-2) is proven.

On the other hand, assume that (5-2) holds for a point (x_0,ξ_0), i.e., that there are a neighborhood V of x_0, an open cone $\Gamma\subseteq\mathbb{R}^n$ containing ξ_0 and constants $C,\gamma>0$ such that

$$|\mathfrak{F}(\psi u)(x,\xi)|\le Ce^{-\omega_{\mathcal{M}}(\gamma|\xi|)},\quad x\in V,\ \xi\in\Gamma, \tag{5-4}$$

for some test function $\psi\in\mathcal{D}(\Omega)$ that is 1 near x_0. We may assume that $x_0=0$. We have to prove that $(0,\xi_0)\notin \mathrm{WF}_{\mathcal{M}}\,u$ or, equivalently, $(0,\xi_0)\notin\mathrm{WF}_{\mathcal{M}}\,v$ where $v=\psi u$. We invoke the inversion formula (5-1) for the FBI transform

$$v=\lim_{\varepsilon\to\infty}\int_{\mathbb{R}^n\times\mathbb{R}^n} e^{i\xi(x-t)}e^{-\varepsilon|\xi|^2}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dtd\xi$$

and split the occurring integral into two parts

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} e^{i\xi(x-t)}e^{-\varepsilon|\xi|^2}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dtd\xi=I_1^\varepsilon(x)+I_2^\varepsilon(x) \tag{5-5}$$

where

$$I_1^\varepsilon(x)=\int_{\mathbb{R}^n}\int_{|t|\le a} e^{i\xi(x-t)}e^{-\varepsilon|\xi|^2}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dtd\xi$$
$$I_2^\varepsilon(x)=\int_{\mathbb{R}^n}\int_{a\le|t|} e^{i\xi(x-t)}e^{-\varepsilon|\xi|^2}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dtd\xi$$

for a constant a to be determined. Following Berhanu and Hounie [2012] we see that for any choice of a the second integral converges to a holomorphic function in a neighborhood of the origin for $\varepsilon\to 0$.

It remains to look at I_1^ε. Suppose that a is small enough such that $B_a(0)\subseteq V$. Let $\mathcal{C}_j$, $1\le j\le N$, be open, acute cones such that

$$\mathbb{R}^n=\bigcup_{j=1}^N\overline{\mathcal{C}}_j$$

and the intersection $\overline{\mathcal{C}}_j\cap\overline{\mathcal{C}}_k$ has measure zero for $j\ne k$. Furthermore, let $\xi_0\in\mathcal{C}_1$, $\mathcal{C}_1\subseteq\Gamma$ and $\xi_0\notin\overline{\mathcal{C}}_j$ for $j\ne 1$. In particular that means that (5-4) holds on $B_a(0)\times\mathcal{C}_1$, i.e.,

$$|\mathfrak{F}(\psi u)(x,\xi)|\le Ce^{-\omega_{\mathcal{M}}(\gamma|\xi|)},\quad x\in B_a(0),\ \xi\in\mathcal{C}_1. \tag{5-6}$$

Furthermore, arguing as in [Berhanu and Hounie 2012], we can choose open cones Γ_j with the property that $\xi_0\Gamma_j < 0$ for $j = 2, \dots, N$ such that the functions

$$f_j(x+iy) = \int_{\mathcal{C}_j}\int_{B_a(0)} e^{i\xi(x+iy-t)}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dt d\xi$$

are holomorphic on $\mathbb{R}^m \times i\Gamma_j$.

In the remaining case we have to modify the arguments in [Berhanu and Hounie 2012] a little bit. We set

$$f_1^\varepsilon(x) = \int_{\mathcal{C}_1}\int_{B_a(0)} e^{i\xi(x-t)-\varepsilon|\xi|^2}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dt d\xi$$

and

$$f_1(x) = \int_{\mathcal{C}_1}\int_{B_a(0)} e^{i\xi(x-t)}\mathfrak{F}v(t,\xi)|\xi|^{n/(2k)}\,dt d\xi.$$

The functions f_1^ε, $\varepsilon > 0$, extend to entire functions whereas f_1 is smooth due to (5-6) since $e^{-\omega_{\mathcal{M}}}$ is rapidly decreasing (see the remark after the proof of (2-3)). This decrease also shows that f_1^ε converges uniformly to f_1 in a neighborhood of 0 since

$$\begin{aligned}|f_1(x) - f_1^\varepsilon(x)| &\le \int_{\mathcal{C}_1}\int_{B_a(0)} |\mathfrak{F}v(t,\xi)||\xi|^{n/(2k)}|1-e^{-\varepsilon|\xi|^2}|\,dt d\xi \\ &\le C\int_{\mathcal{C}_1} |\xi|^{n/(2k)} e^{-\omega_{\mathcal{M}}(\gamma|\xi|)}|1-e^{-\varepsilon|\xi|^2}|\,d\xi\end{aligned}$$

and the last integral converges to 0 by the monotone convergence theorem.

In fact $f_1 \in \mathcal{E}_{\mathcal{M}}$ because

$$\begin{aligned}|D^\alpha f_1(x)| &\le \int_{\mathcal{C}_1} |\xi|^{n/(2k)}|\xi^\alpha \mathfrak{F}v(t,\xi)|\,dt d\xi \\ &\le C\int_{\mathcal{C}_1} |\xi|^{n/(2k)+|\alpha|} e^{-\omega_{\mathcal{M}}(\gamma|\xi|)}\,d\xi \\ &\le C\gamma^{|\alpha|}M_{|\alpha|},\end{aligned}$$

where in the last step (2-3) and (M2) are used.

So we have showed that on an open neighborhood U of the origin and some open cones Γ_j, $j = 2, \dots, N$, which satisfy $\xi_0\Gamma_j < 0$ we can write

$$v|_U = v_0 + \sum_{j=2}^{N} b_{\Gamma_j} f_j$$

with $v_0 \in \mathcal{E}_{\mathcal{M}}(U)$ and f_j holomorphic on $U + i\Gamma_j$ for $j = 2, \dots, N$. Hence $(0, \xi_0) \notin \mathrm{WF}_{\mathcal{M}}\, v$. $\square$

We summarize our results regarding the description of $\mathrm{WF}_{\mathcal{M}}\, u$ in order to obtain the generalized Bony's theorem alluded to in the beginning of this section (see [Hoepfner and Medrado 2018]).

Theorem 5.2. *Let $u \in \mathcal{D}'(\Omega)$. For $(x_0, \xi_0) \in T^*\Omega\backslash\{0\}$ the following statements are equivalent*:

(1) $(x_0, \xi_0) \notin \mathrm{WF}_{\mathcal{M}}\, u$.

(2) *There are $U \in \mathcal{U}(x_0)$, open convex cones $\Gamma^j \subseteq \mathbb{R}^n$ with $\xi_0 \Gamma^j < 0$ and $\mathcal{M}$-almost analytic functions F_j of slow growth in $U \times \Gamma^j_{\rho_j}$, $\rho_j > 0$ and $j = 1, \dots, N$ for some $N \in \mathbb{N}$ such that*

$$u|_U = \sum_{j=1}^{N} b_{\Gamma^j} F_j.$$

(3) *There are $\varphi \in \mathcal{D}(\Omega)$ with $\varphi \equiv 1$ near x_0, with a neighborhood V of x_0, and an open cone Γ containing ξ_0 such that (5-2) holds.*

We can also give a local version of Theorem 5.2.

Corollary 5.3. *Let $u \in \mathcal{D}'(\Omega)$ and $p \in \Omega$. Then the following are equivalent*:

(1) *The distribution u is of class $\{\mathcal{M}\}$ near p.*

(2) *There are a bounded sequence $(u_N)_N \subseteq \mathcal{E}'(\Omega)$ and an open neighborhood $V \subseteq \Omega$ of p such that $u_N|_V = u|_V$ for all $N \in \mathbb{N}_0$ and (3-1) holds for $\Gamma = \mathbb{R}^n$ and some constant $Q > 0$.*

(3) *There exist an open neighborhood $W \subseteq \Omega$ of p, $r > 0$ and a smooth function F on $W + iB(0, r)$ such that $F|_W = u|_W$ and (2-5) holds for some constants $C, Q > 0$.*

(4) *There are a test function $\psi \in D(\Omega)$ with $\psi \equiv 1$ near p, a neighborhood V of p and constants $C, \gamma > 0$ such that*

$$\sup_{(t,\xi)\in V\times\mathbb{R}^n} e^{\omega_{\mathcal{M}}(\gamma|\xi|)} |\mathfrak{F}(\psi u)(t, \xi)| < \infty.$$

Proof. The equivalence of (1) and (2) is just Proposition 3.1, whereas Corollary 2.11 shows that (1) implies (3). For the fact that (4) implies (1) we note that by Theorem 5.1 we have $(p, \xi) \notin \mathrm{WF}_{\mathcal{M}}\, u$ for all $\xi \in \mathbb{R}^n\backslash\{0\}$. Therefore u has to be ultradifferentiable of class $\{\mathcal{M}\}$ near p. Now we show that (4) follows from (3): Suppose that $u \in \mathcal{E}_{\mathcal{M}}(V)$ on a neighborhood of p and let $F \in \mathcal{E}(W + i\mathbb{R}^n)$ be an $\mathcal{M}$-almost analytic extension of u on a relatively compact neighborhood $W \Subset V$ of p. We choose $\varphi \in \mathcal{D}(W)$, $0 \leq \varphi \leq 1$, and $\varphi \equiv 1$ near p. We consider the map

$$\theta :\ y \longmapsto \theta(y) = y - is\varphi(y)\frac{\xi}{|\xi|}$$

for some $1 > s > 0$ to be determined. Finally let $\psi \in \mathcal{D}(V)$ such that $\psi \equiv 1$ on W. As in the proof of Theorem 5.1 we set $\psi(z) = \psi(x)$ for $z = x + iy \in \mathbb{C}^n$. We put $v = \psi F$ and consider the n-form

$$e^{Q(t,\xi,z)} v(z)\, dz_1 \wedge \cdots \wedge dz_n$$

on

$$D_s = \left\{ x + i\sigma\varphi(x)\frac{\xi}{|\xi|} \in \mathbb{C}^n \;\middle|\; 0 < \sigma < s,\ x \in \operatorname{supp} v \right\}.$$

Stokes' theorem gives us

$$\mathfrak{F}v(t,\xi) = c_p \int_{\theta(\mathbb{R}^n)} e^{Q(t,\xi,z)} v(z,\bar{z})\, dz_1 \wedge \cdots \wedge dz_n + c_p \sum_{j=1}^n \int_{D_s} e^{Q(t,\xi,z)} \frac{\partial v}{\partial \bar{z}_j}(z,\bar{z})\, d\bar{z}_j \wedge dz_1 \wedge \cdots \wedge dz_n.$$

The second integral above is estimated in the same way as the last integral in (5-3). On the other hand the first integral on the right-hand side equals

$$G(t,\xi) = c_p \int_{\mathbb{R}^n} e^{Q(t,\xi,\theta(y))} v(\theta(y)) \det\theta'(y)\, dy.$$

We note that

$$\operatorname{Re} Q(t,\xi,\theta(y)) \le -s\varphi(y)|\xi|\big(1 + O(s\varphi(y))\big) - c_0|t-y|^{2k}$$

and hence

$$\begin{aligned} |G(t,\xi)| &\le C \int_{B_\delta(p)} e^{\operatorname{Re} Q(t,\xi,\theta(y))}\, dy + C \int_{\substack{\mathbb{R}^n \setminus B_\delta(p)\\ y \in \operatorname{supp}(v\circ\theta)}} e^{\operatorname{Re} Q(t,\xi,\theta(y))}\, dy \\ &= I_1(t,\xi) + I_2(t,\xi), \end{aligned}$$

where $B_\delta(p) \subseteq \{x \in \mathbb{R}^n \mid \varphi(x) = 1\}$, can be estimated as follows; see [Berhanu et al. 2008]. Set $s = \delta/4$. We obtain

$$I_1(t,\xi) \le Ce^{-c|\xi|}$$

for all $\xi \in \mathbb{R}^n$ if t is in some bounded neighborhood of p. Furthermore

$$I_2(t,x) \le C \int_{\substack{\mathbb{R}^n \setminus B_r(p)\\ y \in \operatorname{supp}(u\circ\theta)}} e^{-|\xi||t-y|^{2k}}\, dy \le Ce^{-(\delta/2)^{2k}|\xi|}$$

for all ξ and $|t-p| \le \frac{\delta}{2}$.

Hence we have showed that there are constants $c, C > 0$ such that

$$|\mathfrak{F}u(t,\xi)| \le Ce^{-\omega_{\mathcal{M}}(c|\xi|)}$$

for all $\xi \in \mathbb{R}^n$ and t in a bounded neighborhood of p. □

6. Elliptic regularity

As mentioned in the introduction, Albanese, Jornet and Oliaro [2010] used the pattern of the proof of [Hörmander 1983, Theorem 8.6.1] (see Remark 3.3) to prove elliptic regularity for operators with coefficients that are all in the same ultradifferentiable class defined by a weight function; see Remark 3.7. Similarly Hörmander's methods were applied by Pilipović, Teofanov and Tomić [2016; 2018] for certain classes that are defined by more degenerate sequences.

It should be noted that the assumptions Albanese, Jornet and Oliaro put on the weight functions guarantee that the associated class is closed under composition and the inverse function theorem holds. So it would be a reasonable conjecture that the regularity of the defining weight sequence is necessary for elliptic regularity to hold in the category of Denjoy–Carleman classes. But there are weight functions obeying these conditions such that the associated function class cannot be described by regular weight sequences and on the other hand there are regular Denjoy–Carleman classes that cannot be defined by such weight functions; see [Bonet et al. 2007]. It turns out, however, that the regularity of the weight sequence is not enough for the proof of the elliptic regularity theorem, we also have to assume that (M2′) holds. In that case the main result of [Bonet et al. 2007] implies that the Denjoy–Carleman class can be described by a weight function that satisfies the conditions of [Albanese et al. 2010]. Hence, we could use their elliptic regularity theorem, but we would have to show that their definition of the ultradifferentiable wavefront set coincides with the definition of Hörmander. Instead we give here a proof in full detail partially in preparation for the forthcoming paper by Fürdös, Nenning, Rainer and Schindl [Fürdös et al. 2020], where we deal with the problem in the far more general setting of the ultradifferentiable classes introduced in [Rainer and Schindl 2014]; see Remark 3.7.

Furthermore, we show here that Hörmander's proof can be modified in a way to investigate the regularity of solutions of a determined system of linear partial differential equations

$$\begin{aligned} P_{11}u_1 + \cdots + P_{1\nu}u_\nu &= f_1 \\ \vdots \qquad\qquad & \quad \vdots \\ P_{\nu 1}u_1 + \cdots + P_{\nu\nu}u_\nu &= f_\nu \end{aligned}$$

where the P_{jk}, $1 \leq j, k \leq \nu$, are partial differential operators with $\mathcal{E}_{\mathcal{M}}$-coefficients.

More precisely, using the geometric theory for the ultradifferentiable wavefront set developed in Section 4, we can work in the following setting (see [Hörmander 1983, Chapter 6] or [Chazarain and Piriou 1982]).

Let M be an ultradifferentiable manifold of class $\{\mathcal{M}\}$ and E and F two vector bundles of class $\{\mathcal{M}\}$ on M with the same fiber dimension ν. An ultradifferentiable partial differential operator $P:\mathcal{E}_{\mathcal{M}}(M,E)\to\mathcal{E}_{\mathcal{M}}(M,F)$ of class $\{\mathcal{M}\}$ is given locally by

$$Pu=\begin{pmatrix}P_{11}&\cdots&P_{1\nu}\\ \vdots&\ddots&\vdots\\ P_{\nu 1}&\cdots&P_{\nu\nu}\end{pmatrix}\begin{pmatrix}u_1\\ \vdots\\ u_\nu\end{pmatrix}, \tag{6-1}$$

where the P_{jk} are linear partial differential operators with ultradifferentiable coefficients defined in suitable chart neighborhoods. If

$$Q(x,D)=\sum_{|\alpha|\le m} q_\alpha(x)D^\alpha$$

is a differential operator of order $\le m$ on some open set $\Omega\subseteq\mathbb{R}^n$ then the principal symbol q is defined to be

$$q(x,\xi)=\sum_{|\alpha|=m} q_\alpha(x)\xi^\alpha.$$

Hence the order of P is of order $\le m$ if and only if no operator P_{jk} on any chart neighborhood is of order higher than m and P is of order m if the operator is not of order $\le m-1$. The principal symbol p of P is an ultradifferentiable mapping defined on T^*M with values in the space of fiber-linear maps from E to F that is homogenous of degree m in the fibers of T^*M. It is given locally by

$$p(x,\xi)=\begin{pmatrix}p_{11}(x,\xi)&\cdots&p_{1\nu}(x,\xi)\\ \vdots&\ddots&\vdots\\ p_{\nu 1}(x,\xi)&\cdots&p_{\nu\nu}(x,\xi)\end{pmatrix}, \tag{6-2}$$

where p_{jk} is the principal symbol of the operator P_{jk}. See [Chazarain and Piriou 1982] for more details. We say that P is not characteristic (or noncharacteristic) at a point $(x,\xi)\in T^*M\setminus\{0\}$ if $p(x,\xi)$ is an invertible linear mapping. We define the set of all characteristic points

$$\operatorname{Char} P=\{(x,\xi)\in T^*M\setminus\{0\} : P \text{ is characteristic at } (x,\xi)\}.$$

Theorem 6.1. *Suppose that $\mathcal{M}$ is a regular weight sequence that satisfies also* (M2′). *Let M be an $\mathcal{E}_{\mathcal{M}}$-manifold and E, F two ultradifferentiable vector bundles on M of the same fiber dimension. If $P(x,D)$ is a differential operator between E and F with $\mathcal{E}_{\mathcal{M}}$-coefficients and p its principal symbol, then*

$$\mathrm{WF}_{\mathcal{M}}\, u\subseteq \mathrm{WF}_{\mathcal{M}}(Pu)\cup\operatorname{Char} P,\qquad u\in\mathcal{D}'(M,E). \tag{6-3}$$

Proof. We write $f=Pu$. Since the problem is local we work on some chart neighborhood Ω such that in suitable trivializations of E and F we may write $u=(u_1,\dots,u_\nu)\in\mathcal{D}'(\Omega,\mathbb{C}^\nu)$, $f=(f_1,\dots,f_\nu)\in\mathcal{D}'(\Omega,\mathbb{C}^\nu)$, and P and its principal symbol p are of the forms (6-1) and (6-2), respectively. In particular, P is of order m on Ω.

We have to prove that if $(x_0,\xi_0)\notin \mathrm{WF}_{\mathcal{M}}\, f\cup \operatorname{Char} P$ then $(x_0,\xi_0)\notin \mathrm{WF}_{\mathcal{M}}\, u$. Assuming this we find that there has to be a compact neighborhood K of x_0 and a closed conic neighborhood V of ξ_0 in $\mathbb{R}^n\backslash\{0\}$ satisfying

$$\det p(x,\xi)\neq 0,\quad (x,\xi)\in K\times V, \tag{6-4}$$

$$(K\times V)\cap \mathrm{WF}_{\mathcal{M}}(Pu)_j=\varnothing,\quad j=1,\dots,\nu. \tag{6-5}$$

We consider the formal adjoint $Q=P^t$ of P with respect to the pairing

$$\langle f,g\rangle=\sum_{\tau=1}^{\nu}\int f_\tau(x)g_\tau(x)\,dx,\quad f,g\in\mathcal{D}(\Omega,\mathbb{C}^\nu).$$

If $P=(P_{jk})_{jk}$ then $Q=(Q_{jk})_{jk}=(P^t_{kj})_{jk}$ where P^t_{jk} denotes the formal adjoint of the scalar operator $P_{jk}(x,D)=\sum p^\alpha_{jk}(x)D^\alpha$, i.e., for $v\in\mathcal{E}(\Omega)$,

$$P^t_{jk}(x,D)v=\sum_{|\alpha|\le m}(-D)^\alpha(p^\alpha_{jk}(x)v(x)).$$

Let $(\lambda_N)_N\subseteq\mathcal{D}(K)$ be a sequence of test functions satisfying $\lambda_N|_U\equiv 1$ on a fixed neighborhood U of x_0 for all N and for all $\alpha\in\mathbb{N}_0^n$ there are constants $C_\alpha,h_\alpha>0$ such that

$$|D^{\alpha+\beta}\lambda_N|\le C_\alpha(h_\alpha N)^{|\beta|},\quad |\beta|\le N. \tag{6-6}$$

If $u=(u^1,\dots,u^\nu)\in\mathcal{D}'(\Omega,\mathbb{C}^\nu)$, then the sequence $u^\tau_N=\lambda_{2N}u^\tau$ is bounded in $\mathcal{E}'$ and each of these distributions is equal to u^τ in U for all τ. Hence we have to prove that $(u^\tau_N)_N$ satisfies (3-1), i.e.,

$$\sup_{\substack{\xi\in V\\ N\in\mathbb{N}_0}}\frac{|\xi|^N|\hat{u}^\tau_N|}{Q^N M_N}<\infty$$

for a constant $Q>0$ independent of N.

In order to do so, set $\Lambda^\tau_N=\lambda_N e_\tau\in\mathcal{D}'(\Omega,\mathbb{C}^\nu)$ and observe

$$\hat{u}^\tau_N(\xi)=\langle u^\tau,e^{-i\langle\,\cdot\,,\xi\rangle}\lambda_{2N}\rangle=\langle u,e^{-i\langle\,\cdot\,,\xi\rangle}\Lambda^\tau_{2N}\rangle.$$

Following the argument in the proof of [Hörmander 1983, Theorem 8.6.1] we want to solve the equation $Qg^\tau=e^{-ix\xi}\Lambda^\tau_{2N}$. We make the ansatz

$$g^\tau=e^{-ix\xi}B(x,\xi)w^\tau,$$

where $B(x,\xi)$ is the inverse matrix of the transpose of $p(x,\xi)$, which exists if $(x,\xi)\in K\times V$ and is homogeneous of degree $-m$ in ξ; note that the principal symbol of $Q = P^t$ is $B^{-1}(x, -\xi)$. Using this we conclude that w has to satisfy

$$w^\tau - Rw^\tau = \Lambda_{2N}^\tau. \tag{6-7}$$

Here $R = R_1 + \cdots + R_m$ with $R_j|\xi|^j$ being (matrix) differential operators of order $\le j$ with coefficients in $\mathcal{E}_\mathcal{M}$ that are homogeneous of degree 0 in ξ if $x \in K$ and $\xi \in V$.

A formal solution of (6-7) would be

$$w^\tau = \sum_{k=0}^\infty R^k \Lambda_{2N}^\tau.$$

However, this sum may not converge and even if it would converge, in the estimates we want to obtain we are not allowed to consider derivatives of arbitrary high order. Hence we set

$$w_N^\tau := \sum_{j_1+\cdots+j_k \le N-m} R_{j_1} \cdots R_{j_k} \Lambda_{2N}^\tau$$

and compute

$$w_N^\tau - Rw_N^\tau = \Lambda_{2N}^\tau - \sum_{\sum_{s=1}^k j_s > N-m \ge \sum_{s=2}^k j_s} R_{j_1} \cdots R_{j_k} \Lambda_{2N}^\tau = \Lambda_{2N}^\tau - \rho_N^\tau.$$

Equivalently, we have

$$Q(e^{-ix\xi} B(x,\xi) w_N^\tau) = e^{-ix\xi}(\Lambda_{2N}^\tau(x) - \rho_N^\tau(x,\xi)).$$

We obtain now

$$\begin{aligned} \hat{u}_N^\tau(\xi) &= \langle u, e^{-i\langle\,.\,,\xi\rangle} \Lambda_{2N}^\tau \rangle \\ &= \langle u, Q(e^{-i\langle\,.\,,\xi\rangle} B(\,.\,,\xi) w_N^\tau)\rangle + \langle u, e^{-i\langle\,.\,,\xi\rangle} \rho_N^\tau(\,.\,,\xi)\rangle \\ &= \langle f, e^{-i\langle\,.\,,\xi\rangle} B(\,.\,,\xi) w_N^\tau\rangle + \langle u, e^{-i\langle\,.\,,\xi\rangle} \rho_N^\tau(\,.\,,\xi)\rangle \end{aligned} \tag{6-8}$$

and continue by estimating the right-hand side of (6-8). For this purpose we need the following lemma.

Lemma 6.2. *There exist constants C and h depending only on R and the constants appearing in (6-6) such that, if $j = j_1 + \cdots + j_k$ and $j + |\beta| \le 2N$, we have*

$$|D^\beta(R_{j_1} \cdots R_{j_k} \Lambda_{2N}^\tau)_\sigma| \le Ch^N M_N^{(j+|\beta|)/N} |\xi|^{-j}, \quad \xi \in V, \sigma = 1, \ldots, \nu. \tag{6-9}$$

Proof. Since both sides of (6-9) are homogeneous of degree $-j$ in $\xi \in V$ it suffices to prove the lemma for $|\xi| = 1$. Moreover we can write

$$(R_{j_1} \cdots R_{j_k} \Lambda_{2N}^\tau)_\sigma = \tilde{R}_\sigma^\tau \lambda_{2N}, \quad \sigma = 1, \ldots, \nu,$$

with $\tilde{R}_\sigma^\tau$ being a certain linear combination of products of components of the operators R_{j_s}. Especially the coefficients of $\tilde{R}_\sigma^\tau$ are all of class $\{\mathcal{M}\}$ on a common neighborhood of K and since there are only finitely many of them we may assume that they all can be considered as elements of $\mathcal{E}^q_{\mathcal{M}}(K)$ for some $q > 0$. We denote the set of the coefficients of the operators $\tilde{R}^\tau_{2N}\sigma$ by $\mathcal{R}$. Recall that (M4) implies that there is a constant $\delta > 0$ such that $N < \delta \sqrt[N]{M_N}$. Hence (6-6) implies that for all $\alpha \in \mathbb{N}_0^n$ we have

$$|D^{\alpha+\beta}\lambda_{2N}| \le C_\alpha h_\alpha^{|\beta|}(2N)^{|\beta|} \le C_\alpha(2h_\alpha\delta)^{|\beta|}M_N^{|\beta|/N} \tag{6-10}$$

for $|\beta| \le 2N$.

Considering all these arguments, the proof of the lemma is a consequence of the following result. □

Lemma 6.3. *Let $K \subseteq \Omega$ be compact, $(\lambda_N)_N \subseteq \mathcal{D}(K)$ be a sequence satisfying (6-10), $q \ge 1$ and $a_1, \dots, a_{j-1} \in \mathcal{R} \cup \{1\}$. Then there are constants $C, h > 0$ independent of N such that for $j \le 2N$ we have*

$$|D_{i_1}(a_1 D_{i_2}(a_2 \cdots D_{i_{j-1}}(a_{j-1}D_{i_j}\lambda_{2N})\cdots))| \le Ch^j M_N^{j/N}. \tag{6-11}$$

Proof. We begin by noting that (M3) implies that $m_j m_{k-j} \le m_k$ for all $j \le k \in \mathbb{N}$; see [Komatsu 1973]. Furthermore we can assume that there is a constant $C_1 > 1$ such that for all $k \le j-1$,

$$|D^\alpha a_k| \le C_1 q^{|\alpha|}M_{|\alpha|}$$

on K. Obviously the expression $D_{i_1}a_1D_{i_2}a_2\cdots D_{i_{j-1}}a_{j-1}D_{i_j}\lambda_{2N}$ can be written as a sum of terms of the form $(D^{\alpha_1}a_1)\cdots(D^{\alpha_{j-1}}a_{j-1})D^{\alpha_j}\lambda_{2N}$ where $|\alpha_1|+\cdots+|\alpha_j| = j$.

We set $h \ge C_1\max(q, h_0)$. If there are $C_{k_1,\dots,k_j}$ terms with $|\alpha_1| = k_1, \dots, |\alpha_j| = k_j$ then we have the following estimate on K:

$$\begin{aligned}
|D_{i_1}a_1D_{i_2}a_2\cdots D_{i_{j-1}}a_{j-1}D_{i_j}\lambda_{2N}|&\\
&\le C\sum q^{j-k_j}C_1^{j-1}C_{k_1,\dots,k_j}m_{k_1}\cdots m_{k_{j-1}}k_1!\cdots k_{j-1}!h_0^{k_j}M_N^{k_j/N}\\
&\le Ch^j\sum m_{j-k_j}C_{k_1,\dots,k_j}k_1!\cdots k_{j-1}!M_N^{k_j/N}\\
&\le Ch^j\sum C_{k_1,\dots,k_j}\frac{k_1!\cdots k_{j-1}!}{(j-k_j)!}M_{j-k_j}M_N^{k_j/N}.
\end{aligned}$$

Since $j - k_j \le 2N$, we observe that (M2′) implies that there are two indices $\sigma_1, \sigma_2 \le N$, $\sigma_1 + \sigma_2 = k - k_j$, such that $M_{k-k_j} \le C\rho^{j-k_j}M_{\sigma_1}M_{\sigma_2}$ for some constants C, ρ that are independent of j and N. Now we have

$$M_{j-k_j}M_N^{k_j/N} = C\rho^{j-k_j}M_{\sigma_1}M_{\sigma_2}M_N^{k_j/N} \le C\rho^{j-k_j}M_N^{(\sigma_1+\sigma_2)/N}M_N^{k_j/N} = C\rho^{j-k_j}M_N^{j/N}.$$

since $\sqrt[N]{M_N}$ is increasing. As noted in [Albanese et al. 2010] it is possible to estimate

$$\frac{k_1!\cdots k_{j-1}!}{(j-k_j)!} = \frac{k_1!\cdots k_{j-1}!k_j!j!}{(j-k_j)!k_j!j!} \le 2^j\frac{k_1!\cdots k_j!}{j!},$$

and also (see [Hörmander 1983, page 308])

$$\sum C_{k_1,\dots,k_j} k_1!\cdots k_j! = (2j-1)!!.$$

Since $(2j-1)!!/(j!2^j) \le 1$ we obtain

$$\begin{aligned}|D_{i_1}a_1 D_{i_2}a_2\cdots D_{i_{j-1}}a_{j-1}D_{i_j}\lambda_{2N}| &\le C(4\rho h)^j \frac{(2k-1)!!}{j!2^j} M_N^{j/N}\\ &\le C(4\rho h)^j M_N^{j/N}.\end{aligned}$$

□

In order to estimate $\hat{u}_N^\tau$, we note that due to the boundedness of the sequence $(u_N^\tau)_N \subseteq \mathcal{E}'$ the Banach and Steinhaus theorem implies that there are constants μ and c such that

$$|\hat{u}_N^\tau| \le c(1+|\xi|)^\mu$$

for all N and therefore if $|\xi| \le \sqrt[N]{M_N}$ then

$$|\xi|^N |\hat{u}_N^\tau| \le C M_N^{(N+\mu)/N} \le C\delta^{\mu N} M_N, \tag{6-12}$$

since (M2) implies that there is a constant $\delta > 0$ such that $\sqrt[N]{M_N} \le \delta \sqrt[N-1]{M_{N-1}}$.

Hence it suffices to estimate the terms on the right-hand side of (6-8) for $\xi \in V$, $|\xi| > \sqrt[N]{M_N}$. We begin with the second term.

As in the scalar case there are constants μ and $C > 0$ that only depend on u and K such that for all $\psi \in \mathcal{D}(\Omega, \mathbb{C}^\nu)$ with $\operatorname{supp}\psi \subseteq K$,

$$|\langle u, \psi\rangle| \le C \sum_{|\alpha|\le\mu} \sup_K |D^\alpha \psi|.$$

Note that $\operatorname{supp}_x \rho_N^\tau(\,.\,,\xi) \subseteq K$ for all $\xi \in V$ and $N \in \mathbb{N}$. Thence

$$\begin{aligned}|\langle u, e^{-i\langle\,.\,,\xi\rangle}\rho_N^\tau(\,.\,,\xi)\rangle| &\le C \sum_{|\alpha|\le\mu}\sum_{\beta\le\alpha} |\xi|^{|\alpha|-|\beta|} \sup_{x\in K} |D_x^\beta \rho_N^\tau(x,\xi)|\\ &\le C \sum_{|\alpha|\le\mu} |\xi|^{\mu-|\alpha|} \sup_{x\in K} |D_x^\alpha \rho_N^\tau(x,\xi)|\end{aligned}$$

for $\xi \in V$, $|\xi| \ge 1$ and $N \in \mathbb{N}$. There are at most 2^N terms of the form $R_{j_1}\cdots R_{j_k}\Lambda_{2N}^\tau$ in ρ_N^τ and each term can be estimated by (6-9) by setting $N \ge j > N-m$ and hence

$$|D_x^\alpha \rho_N^\tau(x,\xi)| \le Ch^N 2^N |\xi|^{m-N} M_N^{(N+|\alpha|)/N}$$

for $x \in K$ and $\xi \in V$, $|\xi| > 1$. Applying (M2) therefore gives

$$\begin{aligned}|\langle u, e^{-i\langle\,.\,,\xi\rangle}\rho_N^\tau(\,.\,,\xi)\rangle| &\le Ch^N 2^N |\xi|^{\mu+m-N} M_N^{(N+\mu)/N}\\ &\le Ch^N |\xi|^{\mu+m-N} M_N.\end{aligned} \tag{6-13}$$

The first term in (6-8) is more difficult to estimate. To begin with, observe that Lemma 6.2 gives

$$\begin{aligned}|D^\beta w_N^\tau(x,\xi)| &\le Ch^N \sum_{j=0}^{N-m} M_N^{(j+|\beta|)/N}|\xi|^{-j}\\ &\le Ch^N M_N^{|\beta|/N}\sum_{j=0}^{N-m} M_N^{(j-j)/N}\\ &\le Ch^N M_N^{|\beta|/N}(N-m)\\ &\le Ch^N M_N^{|\beta|/N}\end{aligned}$$

for $N>m,\ |\beta|\le N$ and $\xi\in V,\ |\xi|>\sqrt[N]{M_N}$. Recall that for $N\le m$ we have set $w_N^\tau=\Lambda_{2N}^\tau=\lambda_{2N}^\tau e_\tau$. Hence by the above and (6-10) it follows that

(6-14) $$|D^\beta w_N^\tau(x,\xi)|\le Ch^N M_N^{|\beta|/N}$$

for all $N\in\mathbb{N},\ |\beta|\le N$ and $\xi\in V,\ |\xi|>\sqrt[N]{M_N}$.

On the other hand, since the components of $B(x,\xi)$ are ultradifferentiable of class $\{\mathcal{M}\}$ and homogeneous in $\xi\in V$ of degree $-m$ we note that it is possible to show, as above, using an analogue to Lemma 6.2, the estimate

(6-15) $$|D_x^\beta(w_N^\tau(x,\xi)|\xi|^m B(x,\xi))|\le Ch^N M_N^{|\beta|/N}\qquad |\beta|\le N,\ \xi\in V,\ |\xi|>\sqrt[N]{M_N}.$$

In order to finish the proof of Theorem 6.1 we need an additional lemma.

Lemma 6.4. *Let $f\in\mathcal{D}'(\Omega)$, K be a compact subset of Ω and $V\subseteq\mathbb{R}^n\backslash\{0\}$ be a closed cone such that*

$$\mathrm{WF}_{\mathcal{M}} f\cap(K\times V)=\varnothing.$$

Furthermore let $w_N\in(\mathcal{E}(\Omega\times V)$ be such that $\operatorname{supp} w_N(\,.\,,\xi)\subseteq K$ *for all $\xi\in V$ and (6-14) holds.*

If μ denotes the order of f in a neighborhood of K then

(6-16) $$|\widehat{w_N f}(\xi)|=|\langle w_N(\,.\,,\xi)f,e^{-i\langle\,.\,,\xi\rangle}\rangle|\le Ch^N|\xi|^{\mu+n-N}M_{N-\mu-n},$$

for $N>\mu+n$ and $\xi\in\Gamma,\ |\xi|>\sqrt[N]{M_N}$.

Proof. By Proposition 3.4 we can find a sequence $(f_N)_N$ that is bounded in $\mathcal{E}'^\mu$ and equal to f in some neighborhood of K and such that

(6-17) $$|\hat f_N(\eta)|\le CQ^N M_N/|\eta|^N,\qquad \eta\in W,$$

where W is a conic neighborhood of Γ. Then $w_N f=w_N f_{N'}$ for $N'=N-\mu-n$.

If we denote the partial Fourier transform of $w_N(x,\xi)$ by

$$\hat w_N(\eta,\xi)=\int_\Omega e^{-ix\eta}w_N(x,\xi)\,dx$$

then obviously (6-14) is equivalent to

$$|\eta^\beta \hat{w}_N(\eta,\xi)| \le Ch^N M_N^{|\beta|/N}$$

for $|\beta|\le N,\ \xi\in V,\ |\xi| > \sqrt[N]{M_N}$ and $\eta\in\mathbb{R}^n$. Since $|\eta|\le\sqrt{n}\max|\eta_j|$ we conclude that

$$|\eta|^\ell |\hat{w}_N(\eta,\xi)| \le Ch^N M_N^{\ell/N} \tag{6-18}$$

for $\ell\le N,\ \eta\in\mathbb{R}^n$ and $\xi\in V,\ |\xi| > \sqrt[N]{M_N}$. Hence we obtain

$$\begin{aligned}(|\eta| + M_N^{1/N})^N |\hat{w}_N(\eta,\xi)| &= \sum_{k=0}^{N}\binom{N}{k} M_N^{k/N}|\eta|^{N-k}|\hat{w}_N(\eta,\xi)| \\ &\le Ch^N \sum_{k=0}^{N}\binom{N}{k} M_N^{k/N} M_N^{(N-k)/N} \\ &\le Ch^N M_N\end{aligned} \tag{6-19}$$

if $\eta\in\mathbb{R}^n,\ \xi\in V$ and $|\xi| > \sqrt[N]{M_N}$. Like Hörmander [1983] and Albanese, Jornet and Oliaro [2010], we consider

$$\begin{aligned}\widehat{w_N f}(\xi) &= \frac{1}{(2\pi)^n}\int \hat{w}_N(\eta,\xi)\hat{f}_{N'}(\xi-\eta)\,d\eta \\ &= \frac{1}{(2\pi)^n}\int_{|\eta|<c|\xi|} \hat{w}_N(\eta,\xi)\hat{f}_{N'}(\xi-\eta)\,d\eta \\ &\qquad + \frac{1}{(2\pi)^n}\int_{|\eta|>c|\xi|} \hat{w}_N(\eta,\xi)\hat{f}_{N'}(\xi-\eta)\,d\eta\end{aligned}$$

for some $0<c<1$. The boundedness of the sequence $(f_N)_N$ in $\mathcal{E}'^\mu$ implies as before that

$$|\hat{f}_N(\xi)| \le C(1+|\xi|)^\mu.$$

Hence we conclude that

$$\begin{aligned}(2\pi)^n|\widehat{w_N f}(\xi)| &\le \|\hat{w}_N(\,.\,,\xi)\|_{L^1} \sup_{|\xi-\eta|<c|\xi|}|\hat{f}_{N'}(\eta)| \\ &\qquad + C\int_{|\eta|>c|\xi|} |\hat{w}_N(\eta,\xi)|(1+c^{-1})^\mu(1+|\eta|)^\mu\,d\eta\end{aligned}$$

since $|\eta|\ge c|\xi|$ gives

$$|\xi-\eta| \le (1+c^{-1})|\eta|.$$

On the other hand, there is a constant $0<c<1$ such that $\eta\in W$ when $\xi\in V$ and $|\xi-\eta|\le c|\xi|$. Then $|\eta|\ge(1-c)|\xi|$ and we can replace the supremum above by $\sup_{\eta\in W}|\hat{f}_{N'}(\eta)|$. Furthermore, by (6-19),

$$\begin{aligned}\|\hat{w}_N(\,.\,,\xi)\|_{L_1} &= \int_{\mathbb{R}^n}|\hat{w}_N(\eta,\xi)|\,d\eta \le Ch^N M_N\int_{\mathbb{R}^n}(|\eta|+\sqrt[N]{M_N})^N\,d\eta\\ &\le Ch^N M_N\int_{\sqrt[N]{M_N}}^{\infty}s^{-N'-1}\,ds\\ &\le Ch^N M_N\frac{M_N^{N'/N}}{N'}\le Ch^N M_N^{\mu+n}.\end{aligned}$$

Thence it follows for $\xi\in V,\ |\xi|>\sqrt[N]{M_N}$, that

$$\begin{aligned}|\widehat{w_N f}(\xi)| &\le C_1(1-c)^{-N'}\|\hat{w}_N(\,.\,,\xi)\|_{L^1}|\xi|^{-N'}\sup_{\eta\in W}|\hat{f}_{N'}(\eta)||\eta|^{N'}\\ &\qquad + C_2(1+c^{-1})^{N'+\mu}\int_{|\eta|>c|\xi|}(1+|\eta|)^{\mu}|\hat{w}_N(\eta,\xi)|\,d\eta\\ &\le C_1h^N M_N^{(n+\mu)/N}Q^{N'}M_{N'}|\xi|^{-N'}+C_2\tilde{h}^N M_N\int_{|\eta|>c|\xi|}|\eta|^{-N'-n}\,d\eta\\ &\le Ch^N M_{N'}|\xi|^{-N'},\end{aligned}$$

where we have also used (M2), (6-17) and (6-18). □

Due to (6-15) we can replace w_N in (6-16) with $(w_N^\tau|\xi|^m B)_\sigma,\ \sigma=1,\dots,\nu$, and obtain

$$|\langle f, e^{-i\langle\,.\,,\xi\rangle}B(\,.\,,\xi)w_N^\tau\rangle|\le Ch^N|\xi|^{\mu+n-N}M_{N-\mu-n} \tag{6-20}$$

for $\xi\in V,\ |\xi|>\sqrt[N]{M_N}$.

We consider now the sequence $(v_N^\tau)_N=(u^\tau_{N+m+n+\mu})_N$. If $\xi\in V,\ |\xi|\le(M_{N''})^{1/N''}$, $N''=N+\mu+n+m$, then by (6-12)

$$|\xi|^N|\hat{v}_N^\tau|\le C\delta^N M_{N''}\le C\delta_1^N M_N\quad\text{for some }\delta_1>0.$$

On the other hand (6-8), (6-13) and (6-20) give

$$|\xi|^N|\hat{v}_N^\tau(\xi)|\le C_1h_1^N M_{N+m}|\xi|^{-m}+C_2h_2^N M_{N+\mu+m+n}|\xi|^{-n}\le Ch^N M_N$$

for $\xi\in V,\ |\xi|>(M_{N''})^{1/N''}$.

Thus we have shown for all $\tau=1,\dots,\nu$ that the bounded sequence $(v_N^\tau)_N\subseteq\mathcal{E}'(\Omega)$ satisfies

$$\sup_{\substack{\xi\in V\\ N\in\mathbb{N}}}\frac{|\xi|^N|v_N^\tau(\xi)|}{Q^N M_N}<\infty$$

for some $Q>0$. Clearly $u^\tau|_U\equiv(v_N^\tau)|_U$ and hence

$$(x_0,\xi_0)\notin\mathrm{WF}_{\mathcal{M}}\,u^\tau$$

for all $\tau=1,\dots,\nu$. □

For elliptic operators, i.e., operators P with $\operatorname{Char} P = \varnothing$, the following holds obviously.

Corollary 6.5. *If P is an elliptic operator with ultradifferentiable coefficients of class $\{\mathcal{M}\}$ and $u \in \mathcal{D}'$ then*

$$\mathrm{WF}_{\mathcal{M}}\, Pu = \mathrm{WF}_{\mathcal{M}}\, u.$$

7. Uniqueness theorems

Hörmander [1971a] and Kawai (see [Sato et al. 1973]) independently noticed that results like Theorem 6.1 in the analytic category can be used to prove Holgrem's uniqueness theorem [1901]. We show here that Theorem 6.1 can also be used to give a quasianalytic version of Holmgren's uniqueness theorem. We follow mainly the presentation of [Hörmander 1983].

First recall [Hörmander 1993, Theorem 6.1]:

Proposition 7.1. *Let $I \subseteq \mathbb{R}$ be an interval and $x_0 \in \partial \operatorname{supp} u$, then $(x_0, \pm 1) \in \mathrm{WF}_{\mathcal{M}}\, u$ for any quasianalytic regular weight sequence $\mathcal{M}$.*

As noted in [Hörmander 1993], Proposition 7.1 immediately generalizes to a result for several variables (see [Hörmander 1983, Theorem 8.5.6], and see [Kim et al. 2001] for a similar result):

Theorem 7.2. *Let $\mathcal{M}$ be a quasianalytic regular weight sequence, $u \in \mathcal{D}'(\Omega)$, $x_0 \in \operatorname{supp} u$ and $f : \Omega \to \mathbb{R}$ be a function of class $\{\mathcal{M}\}$ with the properties*

$$df(x_0) \neq 0, \qquad f(x) \leq f(x_0) \quad \text{if } x_0 \neq x \in \operatorname{supp} u.$$

Then we have

$$(x_0, \pm df(x_0)) \in \mathrm{WF}_{\mathcal{M}}\, u.$$

Proof. If we replace f by $f(x) - |x - x_0|^2$ we see that we may assume that $f(x) < f(x_0)$ for $x_0 \neq x \in \operatorname{supp} u$. Furthermore, since $df(x_0) \neq 0$ we can assume that $x_0 = 0$ and $f(x) = x_n$. Next we choose a neighborhood U of 0 in $\mathbb{R}^{n-1}$ so that $U \times \{0\} \Subset \Omega$. By assumption $\operatorname{supp} u \cap (\bar{U} \times \{0\}) = \{0\}$. Hence there is an open interval $I \subseteq \mathbb{R}$ with $0 \in I$ such that

$$U \times I \Subset \Omega \quad \text{and} \quad \operatorname{supp} u \cap (\partial U \times I) = \varnothing. \tag{7-1}$$

If A is an entire analytic function in the variables $x' = (x_1, \dots, x_{n-1})$ then we consider the distribution $U_A \in \mathcal{D}'(I)$ given by $\langle U_A, \psi\rangle = \langle u_A \otimes \psi\rangle$. Note U_A is well defined due to (7-1). By [Hörmander 1983, Theorem 8.5.4'] we have that

$$\mathrm{WF}_{\mathcal{M}}(U_A) \subseteq \big\{(x_n, \xi_n) \in I \times \mathbb{R}\backslash\{0\} \mid \text{there exists } x' \in U : (x', x_n, 0, \xi_n) \in \mathrm{WF}_{\mathcal{M}}\, u\big\}.$$

Note that (x', x_n) above must be close to 0 for x_n small.

Assume, e.g., that $(0, e_n) \notin \mathrm{WF}_{\mathcal{M}}\, u$, $e_n = (0, \dots, 0, 1)$. Then I can be chosen so small that $(x, e_n) \notin \mathrm{WF}_{\mathcal{M}}\, u$ for $x \in U \times I$. We conclude that $(x_n, 1) \notin \mathrm{WF}_{\mathcal{M}}\, U_A$ if $x_n \in I$. Proposition 7.1 implies that $U_A = 0$ on I since $U_A = 0$ on $I \cap \{x_n > 0\}$. That means actually that

$$\langle u|_{U\times I}, A \otimes \varphi\rangle = 0$$

for all $\varphi \in \mathcal{D}(I)$. Since A was chosen arbitrarily from a dense subset of $\mathcal{E}(\mathbb{R}^{n-1})$ it follows that $u = 0$ on $U \times I$. □

For the rest of this section $\mathcal{M}$ is going to be a quasianalytic regular weight sequence that satisfies (M2′).

In order to give Theorem 7.2 a more invariant form we need to recall some facts from [Hörmander 1983].

Definition 7.3. Let F be a closed subset of a $\mathcal{C}^2$ manifold X. The *exterior normal set* $N_e(F) \subseteq T^*X\backslash\{0\}$ is defined as the set of all points (x_0, ξ_0) such that $x_0 \in F$ and there exists a real-valued function $f \in \mathcal{C}^2(X)$ with $df(x_0) = \xi_0 \neq 0$ and $f(x) \leq f(x_0)$ when $x \in F$.

In fact, following the remarks in [Hörmander 1983, page 300] we observe that it would be sufficient for f to be defined locally around x_0. Furthermore f could then also be chosen to be real-analytic in a chart neighborhood near x_0. If g is $\mathcal{C}^1$ near a point $\tilde{x} \in F$ and $dg(\tilde{x}) = \tilde{\xi} \neq 0$ then $(\tilde{x}, \tilde{\xi}) \in \overline{N_e(F)} \subseteq T^*X\backslash\{0\}$. It is clear that if $(x_0, \xi_0) \in N_e(F)$ then $x_0 \in \partial F$. In fact, if $\pi : T^*\Omega \to \Omega$ is the canonical projection then $\pi(N_e(F))$ is dense in ∂F; see [Hörmander 1983, Proposition 8.5.8.]. The *interior normal set* $N_i(F) \subseteq T^*X\backslash\{0\}$ consists of all points (x_0, ξ_0) with $(x_0, -\xi_0) \in N_e(F)$. The *normal set* of F is defined as

$$N(F) = N_e(F) \cup N_i(F) \subseteq T^*X\backslash\{0\}.$$

In this notation Theorem 7.2 takes the following form.

Theorem 7.4. *Let $u \in \mathcal{D}'(\Omega)$. Then*

$$\overline{N(\operatorname{supp} u)} \subseteq \mathrm{WF}_{\mathcal{M}}\, u.$$

Theorem 7.4 combined with Theorem 6.1 gives:

Theorem 7.5. *Let P be a partial differential operator with $\mathcal{E}_{\mathcal{M}}$-coefficients and $u \in \mathcal{D}'(\Omega)$ be a solution of $Pu = 0$. Then*

$$\overline{N(\operatorname{supp} u)} \subseteq \operatorname{Char} P,$$

i.e., the principal symbol p_m of P must vanish on $N(\operatorname{supp} u)$.

In fact, we can now derive the *quasianalytic Holmgren uniqueness theorem.* We recall that a $\mathcal{C}^1$-hypersurface M is characteristic at a point x with respect to a partial differential operator P, if and only if for a defining function φ of M near x we have that $(x, d\varphi(x)) \in \operatorname{Char} P$.

Corollary 7.6. *Let P be a partial differential operator with $\mathcal{E}_{\mathcal{M}}$-coefficients. If X is a $\mathcal{C}^1$-hypersurface in Ω that is noncharacteristic at x_0 and $u \in \mathcal{D}'(\Omega)$ is a solution of $Pu = 0$ that vanishes on one side of X near x_0 then $u \equiv 0$ in a full neighborhood of x_0.*

In fact, (see [Zachmanoglou 1969]) it is possible to reformulate Corollary 7.6.

Corollary 7.7. *Let P be a differential operator with coefficients in $\mathcal{E}_{\mathcal{M}}(\Omega)$. Furthermore let $F \in \mathcal{E}_{\mathcal{M}}(\mathbb{R}^n)$ be a real-valued function of the form*

$$F(x) = f(x') - x_n, \quad x' = (x_1, \dots, x_{n-1}),$$

where $f \in \mathcal{E}_{\mathcal{M}}(\mathbb{R}^{n-1})$ and suppose that the level hypersurfaces of F are nowhere characteristic with respect to P in Ω. Set also $\Omega_c = \{x \in \Omega \mid F(x) < c\}$ for $c \in \mathbb{R}$. If $u \in \mathcal{D}'(\Omega)$ is a solution of $P(x, D)u = 0$ and there is $c \in \mathbb{R}$ such that $\Omega_c \cap \operatorname{supp} u$ is relatively compact in Ω, then $u = 0$ in Ω_c.

Proof. We set for $c \in \mathbb{R}$,

$$\omega_c = \{x \in \Omega \mid F(x) = c\}.$$

Note that for each $c \in \mathbb{R}$ the set ω_c is not relatively compact in Ω. Therefore Ω_c is not relatively compact in Ω either for any c since $\partial\Omega_c = \omega_c$.

By assumption there is a $c \in \mathbb{R}$ such that $K = \operatorname{supp} u \cap \overline{\Omega}_c$ is compact in Ω. In particular, K is bounded in Ω. Hence there has to be $\tilde{c} < c$ such that

$$K \subseteq \{x \in \Omega \mid \tilde{c} \le F(x) \le c\}.$$

Let $c_1 < c$ be the greatest real number such that the inclusion above holds for $\tilde{c} = c_1$. Since K is compact there is a point $p \in \partial K$ such that $F(p) = c_1$. It follows that $p \in \partial \operatorname{supp} u \cap \omega_{c_1}$. Thus we can apply Corollary 7.6 because ω_{c_1} is nowhere characteristic for P. Hence u vanishes in a full neighborhood of p. This contradicts the choice of c_1. We conclude that u has to vanish on Ω_c. □

Note that Hörmander [1963] used the analytic version of Corollary 7.7 to prove Holmgren's uniqueness theorem.

Remark 7.8. We have formulated our results for scalar operators on open sets of $\mathbb{R}^n$ but they remain of course valid on ultradifferentiable manifolds of class $\{\mathcal{M}\}$. Actually, all the conclusions in this section hold even for determined systems of operators and vector-valued distributions. Indeed, we have only to verify that Theorem 7.2 holds also for distributions with values in $\mathbb{C}^\nu$, but this is trivial: If $f(x) \le f(x_0)$ for $x \in \operatorname{supp} u$ then $f(x) \le f(x_0)$ for all $x \in \operatorname{supp} u_j$ and any $1 \le j \le n$, since $\operatorname{supp} u = \bigcup_{j=1}^{\nu} \operatorname{supp} u_j$. Hence Theorem 7.2 implies

$$(x_0, \pm df(x_0)) \in \bigcap_{j=1}^{\nu} \mathrm{WF}_{\mathcal{M}}\, u_j \subseteq \mathrm{WF}_{\mathcal{M}}\, u.$$

Following an idea of Bony [1969; 1976], it is possible to generalize the results above. For the formulation we need some additional notation. Consider a smooth real-valued function p on $T^*\Omega$. The *Hamiltonian vector field* H_p of p is defined by

$$H_p = \sum_{j=1}^{n} \left(\frac{\partial p}{\partial \xi_j} \frac{\partial}{\partial x_j} - \frac{\partial p}{\partial x_j} \frac{\partial}{\partial \xi_j} \right).$$

An integral curve of H_p, i.e., a solution of the Hamilton–Jacobi equations

$$\frac{dx_j}{dt} = \frac{\partial p}{\partial \xi_j}(x, \xi),$$
$$\frac{d\xi_j}{dt} = -\frac{\partial p}{\partial x_j}(x, \xi),$$

$j = 1, \dots, n$, is called a *bicharacteristic* if p vanishes on it. If q is another smooth real-valued function on $T^*\Omega$ then the *Poisson bracket* is defined by $\{p, q\} := H_p(q)$ or, in coordinates,

$$\{p, q\} = \sum_{j=1}^{n} \left(\frac{\partial p}{\partial \xi_j} \frac{\partial q}{\partial x_j} - \frac{\partial p}{\partial x_j} \frac{\partial q}{\partial \xi_j} \right).$$

See [Grigis and Sjöstrand 1994] or [Hörmander 1983] for more details.

We continue by recalling a result of Sjöstrand [1982] (see also [Hörmander 1983]).

Theorem 7.9. *Let F be a closed subset of Ω and suppose that $p \in \mathcal{E}(T^*\Omega \setminus \{0\})$ is real-valued and vanishes on $N_e(F)$. If $(x_0, \xi_0) \in N_e(F)$ then the bicharacteristic $t \mapsto (x(t), \xi(t))$ with $(x(0), \xi(0)) = (x_0, \xi_0)$ stays for $|t|$ small in $N_e(F)$.*

The analogous statement is of course also true for $N_i(F)$ replacing $N_e(F)$.

Corollary 7.10 [Bony 1976]. *Let F be a closed subset of Ω and set*

$$\mathcal{N}_F := \{p \in \mathcal{E}(T^*\Omega\backslash\{0\}) \mid p \equiv 0 \text{ on } N(F)\}.$$

Then $\mathcal{N}_F$ is an ideal in $\mathcal{E}(T^*\Omega\backslash\{0\})$ that is closed under Poisson brackets.

We obtain the quasianalytic version of a result of Bony [1969; 1976].

Theorem 7.11. *Let P a differential operator with $\mathcal{E}_{\mathcal{M}}$-coefficients on Ω and Π the Poisson algebra that is generated by all functions $f \in \mathcal{E}(T^*\Omega\backslash\{0\})$ that vanish on* Char P.

If $u \in \mathcal{D}'(\Omega)$ is a solution of the homogeneous equation $Pu = 0$ then all functions in Π have to vanish on $N(\operatorname{supp} u)$.

Corollary 7.12. *If the elements of Π have no common zeros and u vanishes in a neighborhood of a point $p_0 \in \Omega$ then u must vanish in the connected component of Ω that contains p_0.*

We continue by taking a closer look at Theorem 7.9. Let $\pi : T^*\Omega \to \Omega$ be the canonical projection and $(x_0, \xi_0) \in T^*\Omega\backslash\{0\}$. If q is a smooth function on $T^*\Omega\backslash\{0\}$ that vanishes on $N(F)$, $F \subseteq \Omega$ closed, and $\lambda(t)$ is the bicharacteristic through (x_0, ξ_0) then we conclude that the bicharacteristic curve $\gamma(t) = \pi \circ \lambda$ must stay in ∂F for small t in view of the remarks before Theorem 7.4.

Now suppose that Q is a real vector field on Ω and q its symbol. If we denote by γ the integral curve of Q through x_0 and by λ the bicharacteristic of q through (x_0, ξ_0) where (x_0, ξ_0) then it is trivial that $\gamma = \pi \circ \lambda$.

Definition 7.13. We say that a partial differential operator P on Ω with $\mathcal{E}_{\mathcal{M}}$-coefficients is $\mathcal{M}$-admissible if and only if there are ultradifferentiable real-valued vector fields $Q_1, \dots, Q_d$ with symbols $q_1, \dots, q_d$ such that each q_j vanishes on $\operatorname{Char} P$.

Following the approach of Sjöstrand [1982] we can generalize results of Zachmanoglou [1972] (see also [Bony 1976]) to the quasianalytic setting.

Proposition 7.14. *Let P be an $\mathcal{M}$-admissible operator. If $\mathcal{L} = \mathcal{L}(Q_1, \dots, Q_d)$ is the Lie algebra generated by the vector fields Q_j, $j = 1, \dots, d$, $\varphi \in \mathcal{C}^1(\Omega, \mathbb{R})$ near a point $x_0 \in \Omega$ such that $(x_0, \varphi'(x_0)) \in \operatorname{Char} P$ and $u \in \mathcal{D}'(\Omega)$ is a solution of $Pu = 0$ such that near x_0 we have $x_0 \in \operatorname{supp} u \subseteq \{\varphi \geq 0\}$. Then each $Q \in \mathcal{L}$ is tangent to $\{\varphi = 0\}$ at x_0 and the local Nagano leaf $\gamma_{x_0}(\mathcal{L})$ is contained in $\operatorname{supp} u$.*

Proof. By assumption all $Q_1, \dots, Q_d$ are tangent to $\{\varphi = 0\}$ at x_0 and hence also all $Q \in \mathcal{L}$. From the remarks before Definition 7.13 and Theorem 7.4 we see that all integral curves of the vector fields in $\mathcal{L}$ must be contained in $\partial \operatorname{supp} u$ for a small neighborhood of x_0. Inspecting the construction of the representative of the local Nagano leaf in the proof of Theorem 2.17 we see that $\gamma_{x_0}(\mathcal{L}) \subseteq \operatorname{supp} u$ near x_0. □

In fact, we have the following global theorem (for the analytic case see [Zachmanoglou 1972; Bony 1976, Theorem 2.4.]).

Theorem 7.15. *Let P an $\mathcal{M}$-admissable differential operator. If $u \in \mathcal{D}'(\Omega)$ is a solution of $Pu = 0$ and $p_0 \notin \operatorname{supp} u$ then every integral curve of the vector fields $Q_1, \dots, Q_d$ through p_0 stays in $\Omega \setminus \operatorname{supp} u$.*

Proof. Let $\Gamma = \Gamma_{p_0}(\mathcal{L})$ be the global Nagano leaf of $\mathcal{L} = \mathcal{L}(Q_1, \dots, Q_d)$ through p_0 and suppose that $\partial \operatorname{supp} u \cap \Gamma \neq \varnothing$. Then there has to be a point $q_0 \in \Gamma \cap \partial \operatorname{supp} u$

such that for all neighborhoods $V \subseteq \Omega$ of x_0 we have

$$(\Gamma \cap V) \cap (\Omega \setminus \operatorname{supp} u) \neq \varnothing.$$

Let V be small enough such that $\Gamma \cap V$ is the representative of the local Nagano leaf of $\mathcal{L}$ at q_0 constructed in the proof of Theorem 2.17. Then

$$\Gamma \setminus \operatorname{supp} u \cap V \neq \varnothing.$$

Thence there is a vector field $X \in \mathcal{L}$ such that if $\gamma(t) = \exp tX$ is the integral curve of X through q_0 then $\gamma(0) = q_0$ and $\gamma(1) = q_1 \in V \setminus \operatorname{supp} u$. Possibly shrinking V and applying an ultradifferentiable coordinate change in V we may assume that $q_0 = 0,\ q_1 = (0, \dots, 0, 1)$ and

$$X = \frac{\partial}{\partial x_n}.$$

We note that in these new coordinates the assumption on P can be stated in the following way. Let $\xi \in \mathbb{R}^n$ with $\xi_n \neq 0$ then $p_m(x, \xi) \neq 0$ for all $x \in V$. There is also a neighborhood $V_1 \subseteq V$ of q_1 such that u vanishes on V_1.

We adapt the proof of [Zachmanoglou 1969, Theorem 1]. Let $r > 0$ and $\delta > 0$ be small enough so that

$$U = \{x \in \mathbb{R}^n \mid |x'| < r,\ -\delta < x_n < 1\}$$

is contained in V and

$$\{x \in \mathbb{R}^n \mid |x'| < r,\ x_n = 1\} \subseteq V_1.$$

We consider the real-analytic function

$$F(x) = (1 + \delta)\frac{|x'|^2}{r^2} - \delta - x_n.$$

The normals of the level hypersurfaces of F are always nonzero in the direction of the n-th unit vector. It follows that the level hypersurfaces are everywhere noncharacteristic with respect to P in V. Set

$$U_1 = \left\{x \in U : \ F(x) < -\frac{\delta}{2}\right\}$$

and note that if $x \in U_1$ then $x_n > -\delta/2$. It is easy to see that $U_1 \cap \operatorname{supp} u$ is relatively compact in U. We conclude that $u = 0$ in U_1 by Corollary 7.7. That is a contradiction to the assumption $q_0 \in \partial \operatorname{supp} u$. □

If $Q_1, \dots, Q_d$ are real-valued vector fields with $\mathcal{E}_{\mathcal{M}}$-coefficients, the operators

$$P_0 = Q_1 + i Q_2, \qquad P_k = \sum_{j=1}^{d} Q_j^{2k}, \quad k \in \mathbb{N},$$

are $\mathcal{M}$-admissible.

For our last result we need to recall the notion of finite type which was introduced by Hörmander [1967]. We say a collection of smooth real vector fields $X_1, \dots, X_d$ on Ω is of finite type (of length at most r) if at any point $p \in \Omega$ the tangent space $T_p\Omega$ is generated by $X_j(p)$ and some iterated commutators $[X_{i_1}, [X_{i_2}, [\dots, [X_{i_{q-1}}, X_{i_q}]\dots]]](p)$, where $q \leq r$.

A straightforward application of Theorem 7.15 gives the following corollary.

Corollary 7.16. *Let Ω be connected and assume the real vector fields $X_1, \dots, X_d$ are of class $\{\mathcal{M}\}$ and of finite type and let $u \in \mathcal{D}'(\Omega)$ be a solution of $P_k u = 0$. If u vanishes on an open subset of Ω then $u \equiv 0$ in Ω.*

References

[Adwan and Hoepfner 2010] Z. Adwan and G. Hoepfner, "Approximate solutions and micro-regularity in the Denjoy–Carleman classes", *J. Differential Equations* **249**:9 (2010), 2269–2286. MR Zbl

[Adwan and Hoepfner 2015] Z. Adwan and G. Hoepfner, "Denjoy–Carleman classes: boundary values, approximate solutions and applications", *J. Geom. Anal.* **25**:3 (2015), 1720–1743. MR Zbl

[Albanese et al. 2010] A. A. Albanese, D. Jornet, and A. Oliaro, "Quasianalytic wave front sets for solutions of linear partial differential operators", *Integral Equations Operator Theory* **66**:2 (2010), 153–181. MR Zbl

[Baouendi and Métivier 1982] M. S. Baouendi and G. Métivier, "Analytic vectors of hypoelliptic operators of principal type", *Amer. J. Math.* **104**:2 (1982), 287–319. MR Zbl

[Baouendi et al. 1999] M. S. Baouendi, P. Ebenfelt, and L. P. Rothschild, *Real submanifolds in complex space and their mappings*, Princeton Math. Series **47**, Princeton Univ. Press, 1999. MR Zbl

[Berhanu and Hailu 2017] S. Berhanu and A. Hailu, "Characterization of Gevrey regularity by a class of FBI transforms", pp. 451–482 in *Recent applications of harmonic analysis to function spaces, differential equations, and data science*, edited by I. Pesenson et al., Birkhäuser, Cham, 2017. MR Zbl

[Berhanu and Hounie 2012] S. Berhanu and J. Hounie, "A class of FBI transforms", *Comm. Partial Differential Equations* **37**:1 (2012), 38–57. MR Zbl

[Berhanu et al. 2008] S. Berhanu, P. D. Cordaro, and J. Hounie, *An introduction to involutive structures*, New Math. Monogr. **6**, Cambridge Univ. Press, 2008. MR Zbl

[Bierstone and Milman 2004] E. Bierstone and P. D. Milman, "Resolution of singularities in Denjoy–Carleman classes", *Selecta Math.* (*N.S.*) **10**:1 (2004), 1–28. MR Zbl

[Björck 1966] G. Björck, "Linear partial differential operators and generalized distributions", *Ark. Mat.* **6** (1966), 351–407. MR Zbl

[Boman 1995] J. Boman, "Microlocal quasianalyticity for distributions and ultradistributions", *Publ. Res. Inst. Math. Sci.* **31**:6 (1995), 1079–1095. MR Zbl

[Bonet et al. 2007] J. Bonet, R. Meise, and S. N. Melikhov, "A comparison of two different ways to define classes of ultradifferentiable functions", *Bull. Belg. Math. Soc. Simon Stevin* **14**:3 (2007), 425–444. MR Zbl

[Bony 1969] J.-M. Bony, "Une extension du théorème de Holmgren sur l'unicité du problème de Cauchy", *C. R. Acad. Sci. Paris Sér. A-B* **268** (1969), A1103–A1106. MR Zbl

[Bony 1976] J. M. Bony, "Extensions du théorème de Holmgren", exposé 17 in *Séminaire Goulaouic–Schwartz,* 1975/1976*: Équations aux dérivées partielles et analyse fonctionnelle*, Centre Math. École Polytech., Paris, 1976. MR Zbl

[Bony 1977] J. M. Bony, "Équivalence des diverses notions de spectre singulier analytique", exposé 3 in *Séminaire Goulaouic–Schwartz,* 1976/1977*: Équations aux dérivées partielles et analyse fonctionnelle*, Centre Math. École Polytech., Paris, 1977. MR Zbl

[Braun et al. 1990] R. W. Braun, R. Meise, and B. A. Taylor, "Ultradifferentiable functions and Fourier analysis", *Results Math.* **17**:3-4 (1990), 206–237. MR Zbl

[Bros and Iagolnitzer 1975] J. Bros and D. Iagolnitzer, "Tuboïdes et structure analytique des distributions, II: Support essentiel et structure analytique des distributions", exposé 18 in *Séminaire Goulaouic–Lions–Schwartz,* 1974/1975*: Équations aux dérivées partielles linéaires et non linéaires*, Centre Math. École Polytech., Paris, 1975. MR Zbl

[Carleman 1923a] T. Carleman, "Sur les fonctions indéfiniment dérivables", *C. R. Acad. Sci. Paris* **177** (1923), 422–424. Zbl

[Carleman 1923b] T. Carleman, "Sur les fonctions quasi-analytiques", pp. 181–196 in *Conférences faites au cinquième congrés des mathematiciens Scandinaves* (Helsinki, 1922), Akadem. Buchh., Helsinki, 1923. Zbl

[Castellanos et al. 2013] J. E. Castellanos, P. D. Cordaro, and G. Petronilho, "Gevrey vectors in involutive tube structures and Gevrey regularity for the solutions of certain classes of semilinear systems", *J. Anal. Math.* **119** (2013), 333–364. MR Zbl

[Chazarain and Piriou 1982] J. Chazarain and A. Piriou, *Introduction to the theory of linear partial differential equations*, Stud. Math. Appl. **14**, North-Holland, Amsterdam, 1982. MR Zbl

[Chung and Kim 1997] S.-Y. Chung and D. Kim, "A quasianalytic singular spectrum with respect to the Denjoy–Carleman class", *Nagoya Math. J.* **148** (1997), 137–149. MR Zbl

[Denjoy 1921] A. Denjoy, "Sur les fonctions quasi-analytiques de variable réele", *C. R. Acad. Sci. Paris* **173** (1921), 1329–1331. Zbl

[Duistermaat and Kolk 2010] J. J. Duistermaat and J. A. C. Kolk, *Distributions: theory and applications*, Birkhäuser, Boston, 2010. MR Zbl

[Dynkin 1976] E. M. Dynkin, "Pseudoanalytic continuation of smooth functions: uniform scale", pp. 40–73 in *Mathematical programming and related questions* (Drohobych, Ukraine, 1974), edited by B. S. Mitjagin and B. P. Titarenko, Central. Ékonom.-Mat. Inst. Akad. Nauk SSSR, Moscow, 1976. In Russian; translated in *Amer. Math. Soc. Transl. (2)* **115** (1980), 33–58. MR Zbl

[Dynkin 1993] E. M. Dynkin, "The pseudoanalytic extension", *J. Anal. Math.* **60** (1993), 45–70. MR Zbl

[Ehrenpreis 1970] L. Ehrenpreis, *Fourier analysis in several complex variables*, Pure Appl. Math. **17**, Wiley, New York, 1970. MR Zbl

[Fürdös 2019] S. Fürdös, "Ultradifferentiable CR manifolds", *J. Geom. Anal.* (online publication April 2019).

[Fürdös et al. 2020] S. Fürdös, D. N. Nenning, A. Rainer, and G. Schindl, "Almost analytic extensions of ultradifferentiable functions with applications to microlocal analysis", *J. Math. Anal. Appl.* **481**:1 (2020), art. id. 123451. MR Zbl

[Grigis and Sjöstrand 1994] A. Grigis and J. Sjöstrand, *Microlocal analysis for differential operators: an introduction*, Lond. Math. Soc. Lect. Note Ser. **196**, Cambridge Univ. Press, 1994. MR Zbl

[Hoepfner and Medrado 2018] G. Hoepfner and R. Medrado, "The FBI transforms and their use in microlocal analysis", *J. Funct. Anal.* **275**:5 (2018), 1208–1258. MR Zbl

[Holmgren 1901] E. Holmgren, "Über Systeme von linearen partiellen Differentialgleichungen", *Öfv. Kongl. Vetenskaps-Akad. För.* **58** (1901), 91–103. Zbl

[Hörmander 1963] L. Hörmander, *Linear partial differential operators*, Grundlehren der Math. Wissenschaften **116**, Springer, 1963. MR Zbl

[Hörmander 1967] L. Hörmander, "Hypoelliptic second order differential equations", *Acta Math.* **119** (1967), 147–171. MR Zbl

[Hörmander 1971a] L. Hörmander, "Uniqueness theorems and wave front sets for solutions of linear differential equations with analytic coefficients", *Comm. Pure Appl. Math.* **24** (1971), 671–704. MR Zbl

[Hörmander 1971b] L. Hörmander, "Linear differential operators", pp. 121–133 in *Actes du Congrès International des Mathématiciens, I* (Nice, 1970), edited by M. Berger et al., Gauthier-Villars, Paris, 1971. MR Zbl

[Hörmander 1983] L. Hörmander, *The analysis of linear partial differential operators, I: Distribution theory and Fourier analysis*, Grundlehren der Math. Wissenschaften **256**, Springer, 1983. MR Zbl

[Hörmander 1993] L. Hörmander, "Remarks on Holmgren's uniqueness theorem", *Ann. Inst. Fourier* (*Grenoble*) **43**:5 (1993), 1223–1251. MR Zbl

[Kim et al. 2001] J. G. Kim, S.-Y. Chung, and D. Kim, "Microlocal analysis in the Denjoy–Carleman class", *J. Korean Math. Soc.* **38**:3 (2001), 561–575. MR Zbl

[Komatsu 1973] H. Komatsu, "Ultradistributions, I: Structure theorems and a characterization", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **20** (1973), 25–105. MR Zbl

[Komatsu 1979] H. Komatsu, "The implicit function theorem for ultradifferentiable mappings", *Proc. Japan Acad. Ser. A Math. Sci.* **55**:3 (1979), 69–72. MR Zbl

[Komatsu 1980] H. Komatsu, "Ultradifferentiability of solutions of ordinary differential equations", *Proc. Japan Acad. Ser. A Math. Sci.* **56**:4 (1980), 137–142. MR Zbl

[Komatsu 1991] H. Komatsu, "Microlocal analysis in Gevrey classes and in complex domains", pp. 161–236 in *Microlocal analysis and applications* (Montecatini Terme, Italy, 1989), edited by L. Cattabriga and L. Rodino, Lecture Notes in Math. **1495**, Springer, 1991. MR Zbl

[Lerner 2010] N. Lerner, *Metrics on the phase space and non-selfadjoint pseudo-differential operators*, Pseudo-Diff. Oper. Theory Appl. **3**, Birkhäuser, Basel, 2010. MR Zbl

[Liess 1999] O. Liess, "Carleman regularization in the $\mathscr{C}^\infty$-category", *Ann. Univ. Ferrara Sez. VII* (*N.S.*) **45**:1 (1999), 213–240. MR Zbl

[Mandelbrojt 1952] S. Mandelbrojt, *Séries adhérentes, régularisation des suites, applications*, Gauthier-Villars, Paris, 1952. MR Zbl

[Mather 1971] J. N. Mather, "On Nirenberg's proof of Malgrange's preparation theorem", pp. 116–120 in *Proceedings of Liverpool Singularities Symposium, I* (Liverpool, 1969/1970), edited by C. T. C. Wall, Lecture Notes in Math. **192**, Springer, 1971. MR Zbl

[Melin and Sjöstrand 1975] A. Melin and J. Sjöstrand, "Fourier integral operators with complex-valued phase functions", pp. 120–223 in *Fourier integral operators and partial differential equations* (Nice, 1974), edited by J. Chazarain, Lecture Notes in Math. **459**, Springer, 1975. MR Zbl

[Nagano 1966] T. Nagano, "Linear differential systems with singularities and an application to transitive Lie algebras", *J. Math. Soc. Japan* **18** (1966), 398–404. MR Zbl

[Petzsche and Vogt 1984] H.-J. Petzsche and D. Vogt, "Almost analytic extension of ultradifferentiable functions and the boundary values of holomorphic functions", *Math. Ann.* **267**:1 (1984), 17–35. MR Zbl

[Pilipović et al. 2016] S. Pilipović, N. Teofanov, and F. Tomić, "Beyond Gevrey regularity", *J. Pseudo-Differ. Oper. Appl.* **7**:1 (2016), 113–140. MR Zbl

[Pilipović et al. 2018] S. Pilipović, N. Teofanov, and F. Tomić, "Beyond Gevrey regularity: superposition and propagation of singularities", *Filomat* **32**:8 (2018), 2763–2782. MR

[Rainer and Schindl 2014] A. Rainer and G. Schindl, "Composition in ultradifferentiable classes", *Studia Math.* **224**:2 (2014), 97–131. MR Zbl

[Rainer and Schindl 2016] A. Rainer and G. Schindl, "Equivalence of stability properties for ultradifferentiable function classes", *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat.* **110**:1 (2016), 17–32. MR Zbl

[Rodino 1993] L. Rodino, *Linear partial differential operators in Gevrey spaces*, World Sci., River Edge, NJ, 1993. MR Zbl

[Roumieu 1962] C. Roumieu, "Ultra-distributions définies sur R^n et sur certaines classes de variétés différentiables", *J. Anal. Math.* **10** (1962), 153–192. MR Zbl

[Rudin 1966] W. Rudin, *Real and complex analysis*, McGraw-Hill, New York, 1966. MR Zbl

[Sato 1970] M. Sato, "Hyperfunctions and partial differential equations", pp. 91–94 in *Proceedings of the International Conference on Functional Analysis and Related Topics* (Tokyo, 1969), Tokyo Univ. Press, 1970. Zbl

[Sato et al. 1973] M. Sato, T. Kawai, and M. Kashiwara, "Microfunctions and pseudo-differential equations", pp. 265–529 in *Hyperfunctions and pseudo-differential equations* (Katata, Japan, 1971), edited by H. Komatsu, Lecture Notes in Math. **287**, Springer, 1973. MR Zbl

[Sjöstrand 1982] J. Sjöstrand, "Singularités analytiques microlocales", pp. 1–166 in Astérisque **95**, Soc. Math. France, Paris, 1982. MR Zbl

[Thilliez 2003] V. Thilliez, "Division by flat ultradifferentiable functions and sectorial extensions", *Results Math.* **44**:1-2 (2003), 169–188. MR Zbl

[Thilliez 2008] V. Thilliez, "On quasianalytic local rings", *Expo. Math.* **26**:1 (2008), 1–23. MR Zbl

[Yamanaka 1991] T. Yamanaka, "On ODEs in the ultradifferentiable class", *Nonlinear Anal.* **17**:7 (1991), 599–611. MR Zbl

[Zachmanoglou 1969] E. C. Zachmanoglou, "An application of Holmgren's theorem and convexity with respect to differential operators with flat characteristic cones", *Trans. Amer. Math. Soc.* **140** (1969), 109–115. MR Zbl

[Zachmanoglou 1972] E. C. Zachmanoglou, "Foliations and the propagation of zeroes of solutions of partial differential equations", *Bull. Amer. Math. Soc.* **78** (1972), 835–839. MR Zbl

Received June 5, 2019. Revised April 5, 2020.

STEFAN FÜRDÖS
FACULTY OF MATHEMATICS, UNIVERSITY OF VIENNA
UNIVERSITY OF VIENNA
VIENNA
AUSTRIA
stefan.fuerdoes@univie.ac.at

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.353

AFFINE STRUCTURES ON LIE GROUPOIDS

HONGLEI LANG, ZHANGJU LIU AND YUNHE SHENG

We study affine structures on a Lie groupoid, including affine k-vector fields, k-forms and (p, q)-tensors. We show that the space of affine structures is a 2-vector space over the space of multiplicative structures. Moreover, the space of affine multivector fields with the Schouten bracket and the space of affine vector-valued forms with the Frölicher–Nijenhuis bracket are graded strict Lie 2-algebras, and affine (1, 1)-tensors constitute a strict monoidal category. Such higher structures can be seen as the categorification of multiplicative structures on a Lie groupoid.

1. Introduction

Geometric structures on a Lie groupoid that are compatible with the groupoid multiplication are called multiplicative structures. They have been studied intensively and their infinitesimal correspondings have been developed. See [Iglesias-Ponte et al. 2012; Mackenzie and Xu 2000; Xu 1995] and [Bursztyn and Cabrera 2012; Bursztyn et al. 2009; Crainic et al. 2015] for multiplicative multivector fields and multiplicative forms, respectively, and see [Bursztyn and Drummond 2019] for the theory of multiplicative tensors. Beyond this, there are also multiplicative Dirac structures [Jotz Lean 2019; Ortiz 2013], multiplicative generalized complex structures [Jotz Lean et al. 2016], multiplicative contact and Jacobi structures [Crainic and Salazar 2015; Crainic and Zhu 2007; Iglesias-Ponte and Marrero 2003], multiplicative distributions [Jotz Lean and Ortiz 2014] and multiplicative Manin pairs [Li-Bland and Ševera 2011], etc. We refer to [Kosmann-Schwarzbach 2016] for a survey on this subject.

Our purpose is to study geometric structures that are compatible with the affinoid structure on a Lie groupoid. This is motivated by Weinstein's work [1990], where he studied Poisson manifolds also carrying affinoid structures. An affinoid structure on a space X is a subset of X^4 whose elements are called parallelograms, with axioms modeled on the properties of the quaternary relation $\{(g, h, l, k) : hg^{-1} = kl^{-1}\}$ on

The research is supported by NSFC (11922110, 11901568).
MSC2010: primary 53D17, 53D18; secondary 22A22, 70G45.
Keywords: affine structure, multiplicative structure, 2-vector space, strict monoidal category, Lie 2-algebra.

a group or a groupoid. Groupoids are affinoid spaces, but not every affinoid space arises in this way. Mackenzie [1992; 2000] regarded affinoid structures as a type of double groupoid. He gave the equivalence of affinoid structures, butterfly diagrams and generalized principal bundles and studied their infinitesimal invariants.

The multiplicativity condition for a k-vector field (a k-form) on a Lie groupoid is known as the graph $\{(g, h, gh) : s(g) = t(h)\}$ of the groupoid multiplication being a coisotropic (an isotropic) submanifold of $\mathcal{G} \times \mathcal{G} \times \mathcal{G}$. A Lie groupoid $\mathcal{G}$ carries an affinoid structure with the set of parallelograms given by $\{(g, h, l, hg^{-1}l)\}$ when $hg^{-1}l$ is well defined. So the affine condition is naturally defined to be that the set of parallelograms is a coisotropic or an isotropic submanifold of $\mathcal{G}\times\mathcal{G}\times\mathcal{G}\times\mathcal{G}$ for k-vector fields or k-forms respectively. This gives the notions of affine k-vector fields and affine k-forms on a Lie groupoid. This topic was first studied in [Weinstein 1990]. Then Lu [1990] studied the dressing transformation, Poisson cohomology and also the symplectic groupoids of affine Poisson structures on Lie groups. For more information on affine Poisson structures, see also [Dazord et al. 1991; Dazord and Sondaz 1991; Urbański 1994]. To define affine (p, q)-tensors on a Lie groupoid $\mathcal{G}$, we consider the tangent and cotangent groupoids of $\mathcal{G}$. A (p, q)-tensor on $\mathcal{G}$ can be viewed as a function on the Lie groupoid $\tilde{\mathcal{G}} := \oplus^q T\mathcal{G} \oplus^p T^*\mathcal{G} \rightrightarrows \oplus^q TM \oplus^p A^*$. Then a (p, q)-tensor is said to be affine if it is an affine function (0-form) on the Lie groupoid $\tilde{\mathcal{G}}$. This definition coincides with the previous definitions for affine k-vector fields and affine k-forms.

We shall first make clear the relations between affine and multiplicative structures. For Lie groups, Lu [1990] obtained two multiplicative bivector fields from an affine bivector field by using the right and left translations. We generalize this result to the case of Lie groupoids and obtain two multiplicative k-vector fields (k-forms, (p, q)-tensors) from an affine k-vector field (k-form, (p, q)-tensor). Furthermore, we show that the space of affine structures is a 2-vector space over the vector space of multiplicative structures. Thus affine structures can be viewed as the categorification of multiplicative structures and affine structures define an equivalence relation on multiplicative structures. For some cases, multiplicative structures are functors, as morphisms of Lie groupoids; then affine structures are natural transformations between these multiplicative structures. Moreover, for affine multivector fields, the Schouten bracket gives rise to a graded strict Lie 2-algebra structure on the aforementioned 2-vector space. This recovers the strict Lie 2-algebra structure on 1-vector fields in [Berwick-Evans and Lerman 2016] and is equivalent to the graded Lie 2-algebra in [Bonechi et al. 2018]; see also [Ortiz and Waldron 2019]. We give the geometric support of this graded Lie 2-algebra structure. We also prove that affine vector-valued forms are closed under the Frölicher–Nijenhuis bracket and thus constitute a graded strict Lie 2-algebra. For affine (1, 1)-tensors, the composition of affine (1, 1)-tensors defines a strict monoidal category structure on the aforementioned 2-vector space.

We remark that on Lie groups, affine (p, q)-tensors and multiplicative (p, q)-tensors are the same when $q \neq 0$. In particular, the affine k-forms and multiplicative k-forms are the same. An affine k-vector field differs from a multiplicative k-vector field by an element in $\wedge^k \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of the Lie group. Affine k-vector fields, k-forms and (1, 1)-tensors on pair groupoids are also analyzed in detail.

The organization of this paper is as follows. In Section 2, we recall Lie 2-algebras, monoidal categories, tangent and cotangent Lie groupoids. In Section 3, we introduce the notion of affine k-vector fields and clarify the relation with multiplicative k-vector fields. We show that the space of affine k-vector fields is a 2-vector space. Moreover, affine multivector fields are closed under the Schouten bracket; we thus get a graded strict Lie 2-algebra structure on this 2-vector space. In Section 4, we introduce the notion of affine k-forms and study their properties analogously. In Section 5, affine tensors on a Lie groupoid are introduced. We obtain a graded strict Lie 2-algebra structure on the space of vector-valued forms and a strict monoidal category structure on the space of affine (1, 1)-tensors.

2. Preliminary

2A. *Strict Lie 2-algebras.* Lie 2-algebras are the categorification of Lie algebras, whose underlying spaces are 2-vector spaces. See [Baez and Crans 2004] for more details. Let Vect be the category of vector spaces.

Definition 2.1 [Baez and Crans 2004]. A 2-*vector space* is a category in Vect.

Explicitly, a 2-vector space is a category $V_1 \rightrightarrows V_0$ whose spaces of objects and arrows are both vector spaces, such that the source and target maps $s, t : V_1 \to V_0$, the identity-assigning map $\iota : V_0 \hookrightarrow V_1$, and the composition $\circ : V_1 \times_{V_0} V_1 \to V_1$ are all linear.

A 2-vector space is completely determined by the vector spaces V_0, V_1 with the source, target and the identity-assigning map. Actually, given $f \in V_1$, define its arrow part by $\vec{f} = f - \iota(s(f))$. Then $s(\vec{f}) = 0$ and $t(\vec{f}) = t(f) - s(f)$. So we can identify $f : x \to y$ with the pair $(x, \vec{f})$. With this notation, the composition of $f : x \to y$ and $g : y \to z$ is defined as $g \circ f = (x, \vec{f} + \vec{g})$. Any arrow $(x, \vec{f})$ has an inverse $(x + t(\vec{f}), -\vec{f})$, so a 2-vector space is always a Lie groupoid.

A 2-vector space is equivalent to a 2-term chain complex of vector spaces. On the one hand, given a 2-vector space $V_1 \rightrightarrows V_0$, the corresponding 2-term complex is $t : \ker s \to V_0$. On the other hand, given a chain complex $C_1 \to C_0$, the 2-vector space is $C_0 \oplus C_1 \to C_0$. We refer to [Baez and Crans 2004] for the details.

Definition 2.2 [Baez and Crans 2004]. A *strict Lie* 2-*algebra* is a 2-vector space V together with a skew-symmetric bilinear functor, the bracket, $[\cdot, \cdot] : V \times V \to V$ satisfying the Jacobi identity.

A strict Lie 2-algebra is equivalent to a strict 2-term L_∞-algebra. Namely, a 2-term complex $d : C_1 \to C_0$ with skew-symmetric brackets $[\cdot,\cdot] : C_0 \times C_0 \to C_0$ and $[\cdot,\cdot] : C_0 \times C_1 \to C_1$ satisfying the Jacobi identity and $[da, b] = [a, db]$ and $d[x, a] = [x, da]$ for $x \in C_0$ and $a, b \in C_1$. See [Baez and Crans 2004] for details.

When the spaces of objects and morphisms are graded vector spaces and the Lie bracket is a graded Lie bracket, we call it a *graded strict Lie 2-algebra*. We refer to [Bonechi et al. 2018] for the explicit definition.

2B. *Strict monoidal categories.* A monoidal category is a category $\mathcal{C}$ with a bifunctor $\circ : \mathcal{C}\times\mathcal{C} \to \mathcal{C}$, its product, which is associative up to a natural isomorphism, and with an object which is a left unit and a right unit for the product up to natural isomorphisms. For our purpose, we only consider the category with a product which is strict associative and has a strict two-sided identity object.

Definition 2.3 [Mac Lane 1971]. A *strict monoidal category* $(\mathcal{C}, \circ, e)$ is a category $\mathcal{C}$ with a bifunctor $\circ : \mathcal{C}\times\mathcal{C} \to \mathcal{C}$, which is associative:

$$\circ(\circ \times 1) = \circ(1 \times \circ) : \mathcal{C}\times\mathcal{C}\times\mathcal{C} \to \mathcal{C},$$

and with an object e which is a left and right unit for $\circ$:

$$\circ(e \times 1) = \mathrm{id}_{\mathcal{C}} = \circ(1 \times e) : \mathcal{C} \to \mathcal{C}.$$

The bifunctor $\circ$ here assigns to each pair of objects x, y an object $x \circ y$ and to each pair of arrows $f : x \to x',\ g : y \to y'$ an arrow $f \circ g : x \circ y \to x' \circ y'$. Thus $\circ$ being a bifunctor means that

$$1_x \circ 1_y = 1_{x\circ y}, \quad (f' \circ g') \cdot (f \circ g) = (f' \cdot f) \circ (g' \cdot g),$$

whenever f', f and g', g are composable. Here $\cdot$ is the multiplication in the category $\mathcal{C}$. The associative law and the unit law in the definition hold both for objects and arrows.

2C. *Tangent and cotangent Lie groupoids.* We recall the definition of the tangent and cotangent Lie groupoids of a Lie groupoid.

Denote the source and target maps for a Lie groupoid $\mathcal{G} \rightrightarrows M$ by $s, t : \mathcal{G} \to M$. Two elements $g, h \in \mathcal{G}$ are multiplicable or composable if and only if $s(g) = t(h)$ and their product is written as $g \cdot h$ or simply as gh. Such a pair is called a *multiplicable pair*. We denote the space of multiplicable pairs by $\mathcal{G}^{(2)}$. Let A be the Lie algebroid of $\mathcal{G}$. For $u \in \Gamma(A)$, the right and left translations $\overrightarrow{u}, \overleftarrow{u} \in \mathfrak{X}(\mathcal{G})$ are defined by

$$\overrightarrow{u}(g) = dR_g(u_{t(g)}), \quad \overleftarrow{u}(g) = -dL_g(d\,\mathrm{inv}(u_{s(g)})),$$

where R_g and L_g are the right and left multiplications on $\mathcal{G}$ and $\mathrm{inv} : \mathcal{G} \to \mathcal{G}$ is the inverse map in $\mathcal{G}$.

Throughout this paper, by abuse of notation, we use the same notations s and t

to denote the source and target of any Lie groupoid and we adopt the same notation to denote a map and its tangent map.

Given a Lie groupoid $\mathcal{G} \rightrightarrows M$, its tangent bundle $T\mathcal{G} \rightrightarrows TM$ with the differentials of the structure maps of $\mathcal{G}$ is again a Lie groupoid.

Its cotangent bundle $T^*\mathcal{G}$ is also equipped with a Lie groupoid structure which is over A^*, written as $T^*\mathcal{G} \rightrightarrows A^*$. First we have the inclusion $A^* \hookrightarrow T^*\mathcal{G}$ since $T^*\mathcal{G}|_M \cong T^*M \oplus A^*$. The source and target maps of an element $\xi \in T^*_g\mathcal{G}$ with $g \in \mathcal{G}$ are

$$\langle s(\xi), u\rangle = \langle \xi, \overleftarrow{u}\rangle, \quad \langle t(\xi), v\rangle = \langle \xi, \overrightarrow{v}\rangle \qquad \text{for all } u \in A_{s(g)}, v \in A_{t(g)},$$

So for any $u \in \Gamma(A)$, seen as a function on the base manifold A^*, we get the formulas

$$s^*u = \overleftarrow{u}, \qquad t^*u = \overrightarrow{u}. \tag{1}$$

For a multiplicable pair $(g, h) \in \mathcal{G}^{(2)}$, if $\xi \in T^*_g\mathcal{G}$ and $\eta \in T^*_h\mathcal{G}$ are multiplicable, the product is the element $\xi \cdot \eta \in T^*_{gh}\mathcal{G}$ such that

$$(\xi \cdot \eta)(X \cdot Y) = \xi(X) + \eta(Y) \quad \text{for all } (X, Y) \in T\mathcal{G}^{(2)},$$

where $X \cdot Y \in T_{gh}\mathcal{G}$ is the product of $X \in T_g\mathcal{G}$ and $Y \in T_h\mathcal{G}$ in the Lie groupoid $T\mathcal{G}$. See, for example, [Kosmann-Schwarzbach 2016; Lang and Liu 2018] for more explanation of the cotangent groupoid.

3. Affine k-vector fields on a Lie groupoid

An affinoid structure on a space X is a subset of X^4 whose elements are seen as parallelograms, with axioms modeled on the properties of the quaternary relation $\{(g, h, l, k) \mid hg^{-1} = kl^{-1}\}$ on a group or a groupoid [Weinstein 1990]. In particular, a groupoid has an affinoid structure with parallelograms given by the relation

$$\{(g, h, l, hg^{-1}l) : s(g) = s(h), t(g) = t(l)\}.$$

A k-vector field on a Lie groupoid is called affine when it is compatible with the affinoid structure in the sense that the submanifold of parallelograms is coisotropic. While a k-vector field is multiplicative when the graph $\{(g, h, gh) : s(g) = t(h)\}$ of the multiplication, or space of triangles, is coisotropic.

Let V be a vector space and $\Pi \in \wedge^k V$. A subspace $W \subset V$ is coisotropic with respect to Π if

$$\Pi(\xi_1, \ldots, \xi_k) = 0 \quad \text{for all } \xi_1, \ldots, \xi_k \in W^0,$$

where W^0 is the annihilator space of W, namely,

$$W^0 = \{\xi \in V^* : \xi(w) = 0 \text{ for all } w \in W\}.$$

More generally, for a manifold M and $\Pi \in \mathfrak{X}^k(M)$, a submanifold S of M is coisotropic with respect to Π if T_xS is coisotropic with respect to Π_x for all $x \in S$.

The following definition is motivated by Weinstein's [1990] definition for affine Poisson structures on a Lie groupoid.

Definition 3.1. A k-vector field $\Pi \in \mathfrak{X}^k(\mathcal{G})$ on a Lie groupoid $\mathcal{G}$ is called *affine* if the submanifold

$$S := \{(g, h, l, hg^{-1}l) : s(g) = s(h), t(g) = t(l)\} \subset \mathcal{G} \times \mathcal{G} \times \mathcal{G} \times \mathcal{G}$$

is coisotropic with respect to $\Pi \oplus (-1)^{k+1}\Pi \oplus (-1)^{k+1}\Pi \oplus \Pi$.

Comparatively, a k-vector field on a Lie groupoid $\mathcal{G}$ is multiplicative [Iglesias-Ponte et al. 2012; Mackenzie and Xu 2000; Xu 1995] if it satisfies that the submanifold $\{(g, h, gh) : s(g) = t(h)\} \subset \mathcal{G} \times \mathcal{G} \times \mathcal{G}$ is coisotropic relative to $\Pi \oplus \Pi \oplus (-1)^{k+1}\Pi$.

It is shown in [Chen et al. 2013; Iglesias-Ponte et al. 2012; Weinstein 1990] that a k-vector field $\Pi \in \mathfrak{X}^k(\mathcal{G})$ is multiplicative if and only if it is affine and the base manifold M is coisotropic with respect to Π. We refer the readers to [Chen et al. 2013, Lemma 2.3], where the authors pointed out that some of the conditions listed in [Iglesias-Ponte et al. 2012, Theorem 2.19] are redundant.

As shown in [Iglesias-Ponte et al. 2012] for the $k = 2$ case, for any $\mu \in T^*_{gh}\mathcal{G}$, the covector $(-\mu, L^*_{\mathcal{X}}\mu, R^*_{\mathcal{Y}}\mu, -L^*_{\mathcal{X}}R^*_{\mathcal{Y}}\mu)$ is conormal to S, that is, an element in the annihilator space of TS at $(gh, h, g, s(g))$, where $\mathcal{X}$ and $\mathcal{Y}$ are bisections passing through g and h. Another two classes of vectors conormal to S are $(-t^*\eta, t^*\eta, 0, 0)$ and $(-s^*\xi, s^*\xi, 0, 0)$ for $\eta \in T^*_{t(g)}M$ and $\xi \in T^*_{s(g)}M$. We thus get an explicit description of affine k-vector fields:

Lemma 3.2 [Chen et al. 2013, Lemma 2.3]. *A k-vector field $\Pi \in \mathfrak{X}^k(\mathcal{G})$ on a Lie groupoid $\mathcal{G}$ is affine if and only if the following two conditions hold*:

(i) *For any $(g, h) \in \mathcal{G}^{(2)}$*,

$$\Pi(gh) = L_{\mathcal{X}}\Pi(h) + R_{\mathcal{Y}}\Pi(g) - L_{\mathcal{X}} \circ R_{\mathcal{Y}}(\Pi(s(g))), \tag{2}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are any two local bisections passing through g and h respectively.

(ii) *For any $\xi \in \Omega^1(M)$, $\iota_{t^*\xi}\Pi$ is right-invariant.*

An equivalent description of (2) is that $[\Pi, \overrightarrow{X}]$ is right-invariant for $X \in \Gamma(A)$ [Mackenzie and Xu 2000].

Remark 3.3. We emphasize that our definition of affine multivector fields is different from that in [Iglesias-Ponte et al. 2012; Xu 1995], where they call multivector fields satisfying (2) *affine multivector fields*. We have an extra condition.

On the other hand, we shall see that one affine k-vector field defines two multiplicative k-vector fields.

The restriction on M for a k-vector field $\Pi \in \mathfrak{X}^k(\mathcal{G})$ has $k+1$ components:

$$\Pi|_M \in \Gamma(\wedge^k T\mathcal{G}|_M) \cong \Gamma(\wedge^k(TM \oplus A)) = \Gamma(\wedge^k TM \oplus (\wedge^{k-1}TM \otimes A) \oplus \cdots \oplus \wedge^k A).$$

We denote by π the $\wedge^k A$-component:

$$\pi = \mathrm{pr}_{\wedge^k A} \Pi|_M \in \Gamma(\wedge^k A).$$

So the base manifold M is coisotropic with respect to $\Pi \in \mathfrak{X}^k(\mathcal{G})$ if $\pi = 0$.

Proposition 3.4. *Let Π be a k-vector field on the Lie groupoid $\mathcal{G}$ with $\pi = \mathrm{pr}_{\wedge^k A}\Pi|_M$. Define*

$$\Pi_r = \Pi - \overrightarrow{\pi}, \quad \Pi_l = \Pi - \overleftarrow{\pi}. \tag{3}$$

Then Π is affine if and only if Π_l or Π_r is a multiplicative k-vector field on $\mathcal{G}$. Here the right and left translations are

$$\overrightarrow{\pi}(g) := R_g(\pi_{t(g)}), \quad \overleftarrow{\pi}(g) := -L_g(\mathrm{inv}(\pi_{s(g)})) \qquad \text{for all } g \in \mathcal{G}.$$

Proof. It is known from [Mackenzie and Xu 2000] that a k-vector field Π on $\mathcal{G}$ satisfying $\Pi(gh) = L_{\mathcal{X}}\Pi(h) + R_{\mathcal{Y}}\Pi(g) - L_{\mathcal{X}} \circ R_{\mathcal{Y}}(\Pi(s(g)))$ is equivalent to saying that $[\Pi, \overrightarrow{X}]$ is right-invariant for any $X \in \Gamma(A)$.

For any $X \in \Gamma(A)$, $[\Pi_r, \overrightarrow{X}]$ is right-invariant if and only if $[\Pi, \overrightarrow{X}]$ is right-invariant. So Π satisfies (2) if and only if Π_r satisfies (2). Besides, since $t \circ R_g = t$, it is clear that $\iota_{t^*\xi}\overrightarrow{\pi}$ is right-invariant for $\xi \in \Omega^1(M)$. Also it is obvious that M is coisotropic with respect to Π_r. We conclude that Π_r is multiplicative if and only if Π is affine.

For Π_l, since $[\Pi_l, \overrightarrow{X}] = [\Pi, \overrightarrow{X}]$, $\iota_{t^*\xi}\overleftarrow{\pi} = 0$ and M is coisotropic with respect to Π_l, we obtain the conclusion that Π is affine if and only if Π_l is multiplicative. □

Example 3.5. Multiplicative k-vector fields on $\mathbb{R}^n$ are linear k-vector fields. An affine k-vector field is of the form

$$\sum f^{i_1,\dots,i_k}(x)\frac{\partial}{\partial x^{i_1}} \wedge \cdots \wedge \frac{\partial}{\partial x^{i_k}} + \sum c^{i_1,\dots,i_k}\frac{\partial}{\partial x^{i_1}} \wedge \cdots \wedge \frac{\partial}{\partial x^{i_k}},$$

where $f^{i_1,\dots,i_k}(x)$ is a linear function on $\mathbb{R}^n$ and $c^{i_1,\dots,i_k}$ is a constant. Namely, an affine k-vector field is a sum of a linear k-vector field and a constant k-vector field.

Example 3.6. Multiplicative k-vector fields on the pair groupoid $M \times M \rightrightarrows M$ all have the form $(\Pi, -\Pi)$ for $\Pi \in \mathfrak{X}^k(M)$ and affine k-vector fields on $M \times M$ are of the form (Π, Π') for two k-vector fields $\Pi, \Pi' \in \mathfrak{X}^k(M)$.

Example 3.7. Let $\mathcal{G}$ be a Lie groupoid with Lie algebroid A. For any $\pi \in \Gamma(\wedge^k A)$, the k-vector field $\Pi = \overrightarrow{\pi}$ is affine and the associated two multiplicative k-vector fields are $\Pi_r = 0$ and $\Pi_l = \overrightarrow{\pi} - \overleftarrow{\pi}$.

The space $\mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$ of affine k-vector fields is a vector space with the space $\mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G})$ of multiplicative k-vector fields as a linear subspace.

Theorem 3.8. *We have a 2-vector space*

$$\mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G}),$$

where the groupoid structure is as follows: the source and target maps are given by $s(\Pi) = \Pi_r$ *and* $t(\Pi) = \Pi_l$ *as defined in* (3), *and the multiplication* $*$ *is*

$$\Pi * \Pi' = \Pi + \overleftarrow{\pi'},$$

for a pair Π, Π' *of affine k-vector fields such that* $\Pi_r = \Pi'_l$*. Here* $\pi' = \mathrm{pr}_{\wedge^k A}\Pi'|_M$ *is the* $\wedge^k A$*-component of* $\Pi'|_M$*.*

Proof. We first verify the groupoid structure. It follows from $\Pi_r = \Pi'_l$ that $\Pi - \overrightarrow{\pi} = \Pi' - \overleftarrow{\pi'}$. Then

$$s(\Pi * \Pi') = \Pi + \overleftarrow{\pi'} - \overrightarrow{\pi} - \overrightarrow{\pi'} = \Pi' - \overrightarrow{\pi'} = \Pi'_r = s(\Pi'),$$

and

$$t(\Pi * \Pi') = \Pi + \overleftarrow{\pi'} - \overleftarrow{\pi} - \overleftarrow{\pi'} = \Pi_l = t(\Pi).$$

Here we have used the fact that

$$\mathrm{pr}_{\wedge^k A}\overleftarrow{\pi'}|_M = (-1)^k \mathrm{pr}_{\wedge^k A}\mathrm{inv}(\pi') = \mathrm{pr}_{\wedge^k A}(\pi' - \rho(\pi')) = \pi',$$

where if $\pi' = X_1 \wedge \cdots \wedge X_k$, we have

$$(-1)^k \mathrm{inv}(\pi') = -\mathrm{inv}(X_1) \wedge \cdots \wedge (-\mathrm{inv}(X_k)) = (X_1 - \rho(X_1)) \wedge \cdots (X_k - \rho(X_k)).$$

For the associativity of this multiplication, let Π'' be another affine k-vector field such that $\Pi'_r = \Pi''_l$. We see

$$(\Pi * \Pi') * \Pi'' = \Pi + \overleftarrow{\pi'} + \overleftarrow{\pi''} = \Pi * (\Pi' * \Pi'').$$

Also, it is immediate that all the groupoid structures are linear. This gives a 2-vector space structure on $\mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$. □

The inverse of $\Pi \in \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$ in this 2-vector space is

$$\Pi^{-1} = \Pi - (\overrightarrow{\pi} + \overleftarrow{\pi}), \qquad \pi = \mathrm{pr}_{\wedge^k A}\Pi|_M. \tag{4}$$

Remark 3.9. Lu [1990] considered the case when $\Pi \in \mathfrak{X}^2(\mathcal{G})$ is an affine Poisson vector field on a Lie group. This affine vector field Π^{-1} is also Poisson and is called the *opposite affine Poisson structure* of Π. We see here that it is actually the inverse of Π in the 2-vector space given above.

Corollary 3.10. *The associated* 2*-term chain complex of vector spaces for the* 2*-vector space in the above theorem is*

$$\Gamma(\wedge^k A) \to \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G}), \quad \pi \mapsto \overrightarrow{\pi} - \overleftarrow{\pi}.$$

In addition to this, since affine multivector fields are closed under the Schouten bracket [Iglesias-Ponte et al. 2012], we further obtain a graded strict Lie 2-algebra on this 2-vector space. See [Baez and Crans 2004] for the details of Lie 2-algebras.

Theorem 3.11. *We have a graded strict Lie 2-algebra structure on*

$$\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows \oplus_k \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G}),$$

where the bracket is the Schouten bracket.

Proof. The Schouten bracket defines a graded Lie algebra structure on $\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$. It suffices to check that it is a functor. Let $\Pi_1, \Pi_1' \in \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$ and $\Pi_2, \Pi_2' \in \mathfrak{X}^l_{\mathrm{aff}}(\mathcal{G})$ be two multiplicable pairs, that is $(\Pi_1)_r = (\Pi_1')_l$ and $(\Pi_2)_r = (\Pi_2')_l$. The Schouten bracket $[\cdot\,,\cdot] : \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}) \times \mathfrak{X}^l_{\mathrm{aff}}(\mathcal{G}) \to \mathfrak{X}^{k+l-1}_{\mathrm{aff}}(\mathcal{G})$ being a functor means

$$[(\Pi_1, \Pi_2) * (\Pi_1', \Pi_2')] = [\Pi_1, \Pi_2] * [\Pi_1', \Pi_2']. \tag{5}$$

Actually, by Theorem 3.8, the left-hand side of (5) is equal to

$$[\Pi_1 + \overleftarrow{\pi_1'}, \Pi_2 + \overleftarrow{\pi_2'}] = [\Pi_1, \Pi_2] + [\Pi_1, \overleftarrow{\pi_2'}] + [\overleftarrow{\pi_1'}, \Pi_2] - \overleftarrow{[\pi_1', \pi_2']},$$

where $\pi_1' = \mathrm{pr}_{\wedge^k A} \Pi_1'|_M$ and $\pi_2' = \mathrm{pr}_{\wedge^l A} \Pi_2'|_M$. And the right-hand side of (5) amounts to

$$[\Pi_1, \Pi_2] + \overleftarrow{\mathrm{pr}_{\wedge^{k+l-1} A}[\Pi_1', \Pi_2']|_M}.$$

By straightforward calculation, we have

$$\begin{aligned}
[\Pi_1', \Pi_2'] &= [(\Pi_1')_l, (\Pi_2')_l] + [\overleftarrow{\pi_1'}, (\Pi_2')_l] + [(\Pi_1')_l, \overleftarrow{\pi_2'}] + [\overleftarrow{\pi_1'}, \overleftarrow{\pi_2'}] \\
&= [(\Pi_1')_l, (\Pi_2')_l] + [\overleftarrow{\pi_1'}, (\Pi_2)_r] + [(\Pi_1)_r, \overleftarrow{\pi_2'}] + [\overleftarrow{\pi_1'}, \overleftarrow{\pi_2'}] \\
&= [(\Pi_1')_l, (\Pi_2')_l] + [\overleftarrow{\pi_1'}, \Pi_2] + [\Pi_1, \overleftarrow{\pi_2'}] - \overleftarrow{[\pi_1', \pi_2']},
\end{aligned}$$

where we have used the fact that $[\overleftarrow{X}, \overrightarrow{Y}] = 0$ for any $X, Y \in \Gamma(A)$. Moreover, Π_1 is affine so $[\Pi_1, \overleftarrow{\pi_2'}]$ is left-invariant and so is $[\overleftarrow{\pi_1'}, \Pi_2]$. From this and the fact that $(\Pi_1')_l$ ($(\Pi_2')_l$) has no component in $\wedge^k A$ ($\wedge^l A$), we see

$$\overleftarrow{\mathrm{pr}_{\wedge^{k+l-1} A}[\Pi_1', \Pi_2']|_M} = [\overleftarrow{\pi_1'}, \Pi_2] + [\Pi_1, \overleftarrow{\pi_2'}] - \overleftarrow{[\pi_1', \pi_2']}.$$

Thus we get (5). This finishes the proof. □

Remark 3.12. In [Berwick-Evans and Lerman 2016; Ortiz and Waldron 2019], the authors constructed a strict Lie 2-algebra on the multiplicative 1-vector fields and their natural transformations. They proved the Morita invariance of this construction and obtained a strict Lie 2-algebra structure on the differentiable stack. Actually, this is our case for $k = 1$ when writing the strict Lie 2-algebra as a Lie algebra crossed module. Another remark is that our graded Lie 2-algebra is actually the same as the one in [Bonechi et al. 2018], where they wrote it in the 2-term L_∞-algebra form $\Gamma(\wedge^\bullet A) \to \mathfrak{X}^\bullet_{\mathrm{mult}}(\mathcal{G})$. Moreover, this Lie 2-algebra is Morita invariant and is used to define multivector fields on a differentiable stack in [Bonechi et al. 2018]. Here we see affine multivector fields as the geometric support of this graded Lie 2-algebra structure.

Now we move to consider the infinitesimal of affine k-vector fields.

For an affine k-vector field $\Pi \in \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$, by Lemma 3.2, define $\delta_\Pi f \in \Gamma(\wedge^{k-1} A)$ and $\delta_\Pi X \in \Gamma(\wedge^k A)$ for any $f \in C^\infty(M)$ and $X \in \Gamma(A)$, such that

(6) $$\overrightarrow{\delta_\Pi f} = [\Pi, t^* f], \quad \overrightarrow{\delta_\Pi X} = [\Pi, \overrightarrow{X}].$$

Recall that a *k-differential* [Iglesias-Ponte et al. 2012] on a Lie algebroid A is a pair of maps

$$\delta_0 : C^\infty(M) \to \Gamma(\wedge^{k-1} A), \quad \delta_1 : \Gamma(A) \to \Gamma(\wedge^k A),$$

satisfying

$$\delta_0(fg) = \delta_0(f)g + f\delta_0(g), \quad \delta_1(fX) = \delta_0(f)X + f\delta_1(X)$$

for all $f, g \in C^\infty(M)$, $X \in \Gamma(A)$, and

$$\delta_1[X, Y] = [\delta_1(X), Y] + [X, \delta_1(Y)], \quad X, Y \in \Gamma(A).$$

Denote by $\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$ (resp. $\oplus_k \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G})$) and $\oplus_k \mathcal{A}_k$ the spaces of affine (resp. multiplicative) vector fields on $\mathcal{G}$ and k-differentials on A. It is straightforward to check that they are graded Lie algebras with the Schouten bracket and the commutator Lie bracket.

With these notions, by (6), we have a map

(7) $$\delta : \oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}) \to \oplus_k \mathcal{A}_k, \quad \Pi \mapsto \delta_\Pi.$$

The universal lifting theorem says that

$$\delta|_{\oplus_k \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G})} : \oplus_k \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G}) \to \oplus_k \mathcal{A}_k$$

is an isomorphism of graded Lie algebras when $\mathcal{G}$ is s-connected and s-simply connected [Iglesias-Ponte et al. 2012].

As a direct consequence of Proposition 3.4, we have the following isomorphism of graded Lie algebras.

Proposition 3.13. *We have an isomorphism*

$$\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}) \to \oplus_k \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G}) \triangleright (\oplus_k \Gamma(\wedge^k A)), \quad \Pi \mapsto (\Pi - \overrightarrow{\pi}, \pi), \quad \pi = \mathrm{pr}_{\wedge^k A} \Pi|_M,$$

of graded Lie algebras, where the brackets on $\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$ and $\oplus_k \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G})$ are the Schouten bracket, the bracket on $\oplus_k \Gamma(\wedge^k A)$ is the graded Lie bracket induced by the Lie bracket on A, and the mixed bracket is

$$[\Gamma, \pi] = \delta_\Gamma(\pi) \in \Gamma(\wedge^{k+l-1} A), \quad \Gamma \in \mathfrak{X}^k_{\mathrm{mult}}(\mathcal{G}), \pi \in \Gamma(\wedge^l A).$$

Proof. By Proposition 3.4, this map is an isomorphism of graded vector spaces whose inverse is $(\Gamma, \pi) \mapsto \Gamma + \vec{\pi}$. Identifying an element $\pi \in \Gamma(\wedge^k A)$ with the affine k-vector field $\vec{\pi}$, we see that the Lie bracket on the right-hand side is actually induced by the Schouten bracket on the left-hand side under the isomorphism. Hence the right-hand side is a graded Lie algebra and this map is an isomorphism of graded Lie algebras.

We could also check directly that this map is a morphism of graded Lie algebras. The key point is

(8) $$\operatorname{pr}_{\wedge^{k+l-1}A}[\Pi, \Pi']|_M = \delta_\Pi(\pi') - (-1)^{(k-1)(l-1)}\delta_{\Pi'}(\pi) - [\pi, \pi'], \quad \Pi \in \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}), \Pi' \in \mathfrak{X}^l_{\mathrm{aff}}(\mathcal{G}).$$

The proof of this is similar to the proof of Theorem 3.11 for the right translation. □

The map δ defined in (7) is not a bijection on $\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$. In fact, for $\Pi \in \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$, we have

(9) $$\delta_{\Pi - \overrightarrow{\pi}} = \delta_\Pi - [\pi, \cdot], \quad \delta_{\Pi - \overleftarrow{\pi}} = \delta_\Pi, \quad \pi = \operatorname{pr}_{\wedge^k A}\Pi|_M.$$

Proposition 3.13 together with the universal lifting theorem for multiplicative multivector fields tells us that the kernel of the map δ is $\oplus_k \Gamma(\wedge^k A)$ and we obtain the universal lifting theorem for affine multivector fields.

Theorem 3.14. *Let $\mathcal{G}$ be an s-simply connected and s-connected Lie groupoid with Lie algebroid A. We have a graded Lie algebra isomorphism*

$$\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G}) \cong \oplus_k \mathcal{A}_k \triangleright (\oplus_k \Gamma(\wedge^k A)), \quad \Pi \mapsto (\delta_\Pi - [\pi, \cdot], \pi), \quad \pi = \operatorname{pr}_{\wedge^k A}\Pi|_M,$$

where the brackets on $\oplus_k \mathfrak{X}^k_{\mathrm{aff}}(\mathcal{G})$ and $\oplus_k \mathcal{A}_k$ are the Schouten bracket and the commutator bracket, the bracket on $\oplus_k \Gamma(\wedge^k A)$ is the graded Lie bracket induced by the Lie bracket on A and the mixed bracket is

$$[\delta, \pi] = \delta(\pi) \in \Gamma(\wedge^{k+l-1} A), \quad \delta \in \mathcal{A}_k, \pi \in \Gamma(\wedge^l A).$$

Here δ acts on π as a degree $k-1$ derivation.

Proof. This follows from (9), Proposition 3.13 and the universal lifting theorem for multiplicative multivector fields [Iglesias-Ponte et al. 2012]. □

Next, we consider the case when an affine bivector field Π on a Lie groupoid $\mathcal{G}$ is also Poisson. We shall generalize Lu's results for Lie groups [1990].

Proposition 3.15. *Let Π be an affine bivector field on a Lie groupoid $\mathcal{G}$, and Π_r, Π_l be the multiplicative bivector fields given by (3). Then*

(i) *Π_r (resp. Π_l) is Poisson if and only if $[\Pi, \Pi]$ is right (resp. left)-invariant;*

(ii) *if Π_r is Poisson, then Π is Poisson if and only if $2\delta_{\Pi_r}\pi + [\pi, \pi] = 0$, where $\pi = \operatorname{pr}_{\wedge^2 A}\Pi|_M$.*

Proof. For (i), direct calculation shows that

$$[\Pi_r, \Pi_r] = [\Pi - \overrightarrow{\pi}, \Pi - \overrightarrow{\pi}] = [\Pi, \Pi] - 2[\Pi, \overrightarrow{\pi}] + [\overrightarrow{\pi}, \overrightarrow{\pi}]$$
$$= [\Pi, \Pi] - 2\overrightarrow{\delta_\Pi \pi} + \overrightarrow{[\pi, \pi]}.$$

Therefore if Π_r is Poisson, $[\Pi, \Pi]$ is right-invariant. Conversely, by (8), we have $\mathrm{pr}_{\wedge^3 A}[\Pi, \Pi]|_M = 2\delta_\Pi \pi - [\pi, \pi]$. If $[\Pi, \Pi]$ is right-invariant, we must have

$$[\Pi, \Pi] = \overrightarrow{2\delta_\Pi \pi - [\pi, \pi]}$$

and hence $[\Pi_r, \Pi_r] = 0$.

For (ii), following from

$$[\Pi, \Pi] = [\Pi_r + \overrightarrow{\pi}, \Pi_r + \overrightarrow{\pi}] = [\Pi_r, \Pi_r] + 2\overrightarrow{\delta_{\Pi_r} \pi} + \overrightarrow{[\pi, \pi]},$$

we get the result. □

As a corollary, if an affine bivector field Π is Poisson, the associated two multiplicative vector fields Π_r and Π_l are also Poisson.

Corollary 3.16. *Let Π be an affine Poisson structure on a Lie groupoid $\mathcal{G}$ with $\pi = \mathrm{pr}_{\wedge^2 A}\Pi|_M$. Then its inverse as introduced in (4),*

$$\Pi^{-1} = \Pi - (\overrightarrow{\pi} + \overleftarrow{\pi}),$$

is also an affine Poisson structure on $\mathcal{G}$.

Proof. Π^{-1} is obviously affine. To see that it is Poisson, we have

$$[\Pi^{-1}, \Pi^{-1}] = [\Pi - (\overrightarrow{\pi} + \overleftarrow{\pi}), \Pi - (\overrightarrow{\pi} + \overleftarrow{\pi})]$$
$$= -2[\Pi, \overrightarrow{\pi}] - 2[\Pi, \overleftarrow{\pi}] + [\overleftarrow{\pi}, \overleftarrow{\pi}] + [\overrightarrow{\pi}, \overrightarrow{\pi}] = [\Pi_r, \Pi_r] + [\Pi_l, \Pi_l].$$

Therefore, by Proposition 3.15, Π^{-1} is Poisson. □

Example 3.17. Let $\mathcal{G}$ be a Lie groupoid with Lie algebroid A. For $\pi \in \Gamma(\wedge^2 A)$, the bivector field $\Pi = \overrightarrow{\pi}$ is affine. It is Poisson if and only if π satisfies the classical Yang–Baxter equation $[\pi, \pi] = 0$. Moreover, we have $\Pi_r = 0$, $\Pi_l = \overrightarrow{\pi} - \overleftarrow{\pi}$ and $\Pi^{-1} = -\overleftarrow{\pi}$.

Besides, given any $\gamma \in \Gamma(\wedge^2 A)$, define $\Pi = \overrightarrow{\pi} + \overleftarrow{\gamma}$. Then we get $\Pi_r = \overleftarrow{\gamma} - \overrightarrow{\gamma}$ and $\Pi_l = \overrightarrow{\pi} - \overleftarrow{\pi}$ and $\Pi^{-1} = -\overleftarrow{\pi} - \overrightarrow{\gamma}$. Furthermore, direct calculation shows that Π is Poisson if and only if

$$\overrightarrow{[\pi, \pi]} = \overleftarrow{[\gamma, \gamma]},$$

which implies that $[\pi, \pi] = [\gamma, \gamma] \in \wedge^3 \ker\rho$ and both of them are Ad-invariant.

Affine Poisson structures give rise to a natural equivalence relation between multiplicative Poisson structures on a Lie groupoid (Poisson groupoids), which further give an equivalence relation on Lie bialgebroids.

4. Affine k-forms on a Lie groupoid

A k-form $\Theta \in \Omega^k(\mathcal{G})$ on a Lie groupoid $\mathcal{G}$ is multiplicative if the graph of multiplication $\{(g, h, gh) : s(g) = t(h)\}$, or space of triangles, is an isotropic submanifold of $\mathcal{G} \times \mathcal{G} \times \mathcal{G}$ with respect to $\Theta \oplus \Theta \oplus -\Theta$. Algebraically, a k-form Θ on $\mathcal{G} \rightrightarrows M$ is multiplicative [Bursztyn and Cabrera 2012; Bursztyn et al. 2009; Crainic et al. 2015] if it satisfies

$$m^*\Theta = \mathrm{pr}_1^*\Theta + \mathrm{pr}_2^*\Theta,$$

where m and $\mathrm{pr}_1, \mathrm{pr}_2 : \mathcal{G}^{(2)} \to \mathcal{G}$ are the groupoid multiplication and the projections to the first and second components, respectively.

One consequence of the multiplicativity condition is that Θ is isotropic on M, namely, $\iota^*\Theta = 0$, where $\iota : M \hookrightarrow \mathcal{G}$ is the natural inclusion. In other words, the restriction of Θ on M has no component in $\wedge^k T^*M$. Relaxing this condition, we shall get the notion of affine k-forms.

The restriction of a k-form $\Theta \in \Omega^k(\mathcal{G})$ on M has $k+1$ components:

$$\begin{aligned}\Theta|_M \in \Gamma(\wedge^k T^*\mathcal{G}|_M) &= \Gamma(\wedge^k(A^* \oplus T^*M)) \\ &= \Gamma(\wedge^k A^* \oplus (\wedge^{k-1} A^* \otimes T^*M) \oplus \cdots \oplus \wedge^k T^*M).\end{aligned}$$

Denote by θ the $\wedge^k T^*M$-component: $\theta = \mathrm{pr}_{\wedge^k T^*M} \Theta|_M$. In other words, $\theta = \iota^*\Theta$ for $\iota : M \hookrightarrow \mathcal{G}$.

Definition 4.1. A k-form $\Theta \in \Omega^k(\mathcal{G})$ on a Lie groupoid $\mathcal{G}$ is *affine* if it satisfies

$$m^*\Theta = \mathrm{pr}_1^*\Theta + \mathrm{pr}_2^*\Theta - {\mathrm{pr}_1}^* s^*\theta, \tag{10}$$

where $\theta := \mathrm{pr}_{\wedge^k T^*M} \Theta|_M$.

Since $s \circ \mathrm{pr}_1 = t \circ \mathrm{pr}_2 : \mathcal{G}^{(2)} \to \mathcal{G}$, the affine condition has another expression:

$$m^*\Theta = \mathrm{pr}_1^*\Theta + \mathrm{pr}_2^*\Theta - {\mathrm{pr}_2}^* t^*\theta. \tag{11}$$

One direct consequence of the definition is that the de Rham differential of an affine k-form on $\mathcal{G}$ is an affine $(k+1)$-form.

Unlike the multiplicative case, it is not obvious from (10) that a k-form is affine if the submanifold of parallelograms is isotropic in $\mathcal{G} \times \mathcal{G} \times \mathcal{G} \times \mathcal{G}$.

Proposition 4.2. *A k-form Θ on $\mathcal{G}$ is affine if and only if the space of parallelograms*

$$\Gamma = \{(g, h, l, hg^{-1}l) : s(g) = s(h), t(g) = t(l)\}$$

is an isotropic submanifold of $\mathcal{G} \times \mathcal{G} \times \mathcal{G} \times \mathcal{G}$ with respect to $\Theta \oplus -\Theta \oplus -\Theta \oplus \Theta$, that is,

$$\iota^*(\mathrm{pr}_1^*\Theta - \mathrm{pr}_2^*\Theta - \mathrm{pr}_3^*\Theta + \mathrm{pr}_4^*\Theta) = 0, \tag{12}$$

where $\mathrm{pr}_i : \mathcal{G} \times \mathcal{G} \times \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ is the projection to the i-th component and $\iota : \Gamma \hookrightarrow \mathcal{G} \times \mathcal{G} \times \mathcal{G} \times \mathcal{G}$ is the inclusion.

Proof. The tangent space of Γ at $(g, h, l, hg^{-1}l)$ consists of 4-tuples

$$(X_g, Y_h, Z_l, Y_h \cdot \mathrm{inv}(X)_{g^{-1}} \cdot Z_l)$$

of tangent vectors, where $Y_h \cdot \mathrm{inv}(X)_{g^{-1}} \cdot Z_l$ means the multiplication of three tangent vectors in $T\mathcal{G}$. Applying (12) to k such vectors, we have

$$\Theta(Y_h^1 \cdot \mathrm{inv}(X)_{g^{-1}}^1 \cdot Z_l^1, \dots, Y_h^k \cdot \mathrm{inv}(X)_{g^{-1}}^k \cdot Z_l^k) = -\Theta(X_g^1, \dots, X_g^k) + \Theta(Y_h^1, \dots, Y_h^k) + \Theta(Z_l^1, \dots, Z_l^k), \tag{13}$$

where $\mathrm{inv}(X)_{g^{-1}}^i := \mathrm{inv}(X_g^i)$. In particular, we choose $(h, l) \in \mathcal{G}^{(2)}$ and $g = 1_{t(l)} = 1_{s(h)}$. Moreover, each (X_g^i, Y_h^i, Z_l^i) is chosen to satisfy $t(Z_l^i) = s(Y_h^i) = X_g^i$. Then the equation becomes

$$\Theta(Y_h^1 \cdot Z_l^1, \dots, Y_h^k \cdot Z_l^k) = -\Theta(s(Y_h^1), \dots, s(Y_h^k)) + \Theta(Y_h^1, \dots, Y_h^k) + \Theta(Z_l^1, \dots, Z_l^k). \tag{14}$$

This is exactly (10).

Conversely, if Θ is affine, by setting $l = h^{-1}$ and $Z_l^i = \mathrm{inv}(Y)_l^i$ in (14), we get

$$\Theta(\mathrm{inv}(Y)_l^1, \dots, \mathrm{inv}(Y)_l^k) = \Theta(s(Y_h^1), \dots, s(Y_h^k)) - \Theta(Y_h^1, \dots, Y_h^k) + \Theta(t(Y_h^1), \dots, t(Y_h^k)). \tag{15}$$

Applying (14) twice to the left-hand side of (13), we obtain

$$\begin{aligned}
&\Theta(Y_h^1 \cdot \mathrm{inv}(X)_{g^{-1}}^1 \cdot Z_l^1, \dots, Y_h^k \cdot \mathrm{inv}(X)_{g^{-1}}^k \cdot Z_l^k)\\
&= \Theta(Y_h^1, \dots, Y_h^k) + \Theta(\mathrm{inv}(X)_{g^{-1}}^1 \cdot Z_l^1, \dots, \mathrm{inv}(X)_{g^{-1}}^k \cdot Z_l^k) - \Theta(s(X_g^1), \dots, s(X_g^k))\\
&= \Theta(Y_h^1, \dots, Y_h^k) + \Theta(\mathrm{inv}(X)_{g^{-1}}^1, \dots, \mathrm{inv}(X)_{g^{-1}}^k) + \Theta(Z_l^1, \dots, Z_l^k)\\
&\qquad - \Theta(t(X_g^1), \dots, t(X_g^k)) - \Theta(s(X_g^1), \dots, s(X_g^k))\\
&= \Theta(Y_h^1, \dots, Y_h^k) - \Theta(X_g^1, \dots, X_g^k) + \Theta(Z_l^1, \dots, Z_l^k),
\end{aligned}$$

where we have used (15) in the last step. Hence, we get (12). □

Regarding the relation between multiplicative and affine k-forms, we have already seen that a multiplicative k-form is an affine k-form which is isotropic on M. On the other hand, an affine k-form is associated with two multiplicative k-forms.

Proposition 4.3. *Let $\Theta \in \Omega^k(\mathcal{G})$ be a k-form on $\mathcal{G}$ with $\theta = \mathrm{pr}_{\wedge^k T^*M} \Theta|_M \in \Omega^k(M)$. Define two k-forms on $\mathcal{G}$:*

$$\Theta_l := \Theta - s^*\theta, \qquad \Theta_r := \Theta - t^*\theta. \tag{16}$$

Then Θ is affine if and only if Θ_l (resp. Θ_r) is a multiplicative k-form.

Proof. By straightforward calculation, we have

$$m^*\Theta_l = m^*\Theta - m^*s^*\theta, \quad \mathrm{pr}_1^*\Theta_l = \mathrm{pr}_1^*\Theta - \mathrm{pr}_1^*s^*\theta, \quad \mathrm{pr}_2^*\Theta_l = \mathrm{pr}_2^*\Theta - \mathrm{pr}_2^*s^*\theta.$$

Following from $s \circ \mathrm{pr}_2 = s \circ m : \mathcal{G}^{(2)} \to \mathcal{G}$, we see that the equation $m^*\Theta_l = \mathrm{pr}_1^*\Theta_l + \mathrm{pr}_2^*\Theta_l$ holds if and only if (10) holds. Similarly, noticing that $t \circ \mathrm{pr}_1 = t \circ m$, we get that Θ_r is multiplicative if and only if Θ satisfies (11), that is, Θ is affine. □

Denote by $\Omega^k_{\mathrm{aff}}(\mathcal{G})$ and $\Omega^k_{\mathrm{mult}}(\mathcal{G})$ the spaces of affine and multiplicative k-forms, respectively. It is immediate that $\Omega^k_{\mathrm{aff}}(\mathcal{G})$ is a vector space with $\Omega^k_{\mathrm{mult}}(\mathcal{G})$ being a linear subspace.

Theorem 4.4. *We have a* 2*-vector space*

$$\Omega^k_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows \Omega^k_{\mathrm{mult}}(\mathcal{G}),$$

where the groupoid structure is given as follows: *the source and target maps are*

$$s(\Theta) = \Theta_r, \quad t(\Theta) = \Theta_l \qquad \text{for all } \Theta \in \Omega^k_{\mathrm{aff}}(\mathcal{G}),$$

where Θ_r and Θ_l are defined in (16), *and the multiplication is*

$$\Theta * \Theta' = \Theta + s^*\theta', \quad \theta' = \mathrm{pr}_{\wedge^k T^*M}\Theta'|_M$$

for a pair $\Theta, \Theta' \in \Omega^k_{\mathrm{aff}}(\mathcal{G})$ such that $\Theta_r = \Theta'_l$.

Proof. The proof is similar to that for Theorem 3.8. □

Corollary 4.5. *The* 2*-term chain complex of vector spaces associated to the above* 2*-vector space $\Omega^k_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows \Omega^k_{\mathrm{mult}}(\mathcal{G})$ is*

$$\Omega^k(M) \to \Omega^k_{\mathrm{mult}}(\mathcal{G}), \quad \theta \mapsto t^*\theta - s^*\theta.$$

It is seen from the definition that the affine and multiplicative forms are closed under the de Rham differential. So we get two subcomplexes of the de Rham complex on $\mathcal{G}$:

$$(\Omega^\bullet_{\mathrm{mult}}(\mathcal{G}), d) \subset (\Omega^\bullet_{\mathrm{aff}}(\mathcal{G}), d) \subset (\Omega^\bullet(\mathcal{G}), d).$$

Proposition 4.6. *The map*

$$\Phi : \Omega^\bullet_{\mathrm{aff}}(\mathcal{G}) \to \Omega^\bullet_{\mathrm{mult}}(\mathcal{G}) \oplus \Omega^\bullet(M), \quad \Theta \mapsto (\Theta - t^*\theta, \theta), \quad \theta = \mathrm{pr}_{\wedge^\bullet T^*M}\Theta|_M,$$

is an isomorphism of cochain complexes, where the differentials are the de Rham differential. Thus we get an isomorphism on the cohomology

$$H^\bullet_{\mathrm{aff}}(\mathcal{G}) \cong H^\bullet_{\mathrm{mult}}(\mathcal{G}) \oplus H^\bullet(M).$$

Proof. The inverse of Φ can be defined by $(\Lambda, \lambda) \mapsto \Lambda + t^*\lambda$ for any $\Lambda \in \Omega^k_{\mathrm{mult}}(\mathcal{G})$ and $\lambda \in \Omega^k(M)$. So Φ is an isomorphism. Next, we check that it is a cochain map, namely, $d\circ\Phi=\Phi\circ d$. Since $\theta=\iota^*\Theta$ for $\iota: M\hookrightarrow\mathcal{G}$, we have $d\theta=\mathrm{pr}_{\wedge^{k+1}T^*M}d\Theta$. Then we have

$$d\circ\Phi(\Theta)=(d\Theta-t^*d\theta, d\theta)=\Phi\circ d(\Theta).$$

Thus it induces an isomorphism on the cohomology. □

Now we discuss the infinitesimal of affine k-forms. It is known from [Bursztyn and Cabrera 2012; Crainic et al. 2015] that there is a one-to-one correspondence between multiplicative k-forms on $\mathcal{G}$ and IM k-forms on its Lie algebroid A when $\mathcal{G}$ is s-connected and s-simply connected.

A pair (μ,ν) of bundle maps

$$\mu: A\to\wedge^{k-1}T^*M,\quad \nu: A\to\wedge^kT^*M,\quad k\geq 1,$$

is called an *IM k-form* on A if

$$\begin{aligned}\iota_{\rho(X)}\mu(Y)&=-\iota_{\rho(Y)}\mu(X),\\ \mu([X,Y])&=\mathcal{L}_{\rho(X)}\mu(Y)-\iota_{\rho(Y)}d\mu(X)-\iota_{\rho(Y)}\nu(X),\\ \nu([X,Y])&=\mathcal{L}_{\rho(X)}\nu(Y)-\iota_{\rho(Y)}d\nu(X)\quad\text{for all } X,Y\in\Gamma(A).\end{aligned}$$

The pair (μ,ν) determines a linear k-form on the vector bundle A. These conditions are described in such a way that the induced map $\oplus^k_A TA\to\mathbb{R}$ is a Lie algebroid morphism with the tangent Lie algebroid structure on $\oplus^k_A TA\to\oplus^k TM$ and the trivial Lie algebroid structure on $\mathbb{R}\to\{*\}$. See [Bursztyn and Cabrera 2012] for details.

By Proposition 4.6, the infinitesimals of affine k-forms on a Lie groupoid $\mathcal{G}$ are clear.

Proposition 4.7. *If $\mathcal{G}$ is an s-connected and s-simply connected Lie groupoid, there is a one-to-one correspondence between affine k-forms Θ on $\mathcal{G}$ and triples (μ,ν,θ) of IM k-forms (μ,ν) and $\theta\in\Omega^k(M)$. That is*

$$\Omega^k_{\mathrm{aff}}(\mathcal{G})\cong\Omega^k_{\mathrm{mult}}(\mathcal{G})\oplus\Omega^k(M)\cong\Omega^k_{\mathrm{IM}}(A)\oplus\Omega^k(M),$$

$$\Theta\mapsto(\Theta_r:=\Theta-t^*\theta,\theta:=\mathrm{pr}_{\wedge^kT^*M}\Theta|_M)\mapsto(\mu,\nu,\theta).$$

Example 4.8. For a Lie groupoid $\mathcal{G}\rightrightarrows M$, given any $\theta\in\Omega^k(M)$, then $s^*\theta$ and $t^*\theta$ are affine k-forms on $\mathcal{G}$ and $s^*\theta-t^*\theta$ is a multiplicative k-form on $\mathcal{G}$.

Example 4.9. On a Lie group G, affine k-forms are multiplicative k-forms. They are nonzero only when k is 0 and 1. This is because for any $k\geq 2$,

$$\begin{aligned}\Theta((X_1,0,X_3,\dots,X_k)\cdot(0,Y_2,Y_3,\dots,Y_k))\\ &=\Theta(R_{h_1}X_1, L_{g_2}Y_2, X_3\cdot Y_3,\dots,X_k\cdot Y_k)\\ &=\Theta(X_1,0,X_3,\dots,X_k)+\Theta(0,Y_2,\dots,Y_k)=0\end{aligned}$$

for $X_i\in T_{g_i}G,\ Y_j\in T_{h_j}G$. So multiplicative 1-forms on a Lie group are always closed.

On the abelian group $\mathbb{R}^n$, multiplicative 1-forms are constant 1-forms. They have the form $\Theta = \sum_i c_i dx^i$, where c_i is a constant.

Example 4.10. For the pair Lie groupoid $M \times M \rightrightarrows M$, multiplicative k-forms all have the form $\mathrm{pr}_1^*\alpha - \mathrm{pr}_2^*\alpha$ for $\alpha \in \Omega^k(M)$, where $\mathrm{pr}_i : M \times M \to M$ is the projection to the i-th component. Affine k-forms are of the form $\mathrm{pr}_1^*\alpha + \mathrm{pr}_2^*\beta$ for any two k-forms $\alpha, \beta \in \Omega^k(M)$.

5. Affine tensors on a Lie groupoid

5A. ***Definition of affine tensors.*** A tensor on a Lie groupoid is said to be affine if it is affine as a function on a more complicated Lie groupoid. This is motivated by the notion of multiplicative tensors on a Lie groupoid introduced in [Bursztyn and Drummond 2019].

Affine functions on a Lie groupoid $\mathcal{G} \rightrightarrows M$ are naturally defined as affine 0-forms on $\mathcal{G}$.

Definition 5.1. A function $F \in C^\infty(\mathcal{G})$ is *affine* if it satisfies

$$F(gh) = F(g) + F(h) - F(s(g)) \quad \text{for all } (g,h) \in \mathcal{G}^{(2)}, \tag{17}$$

or $F(gh) = F(g) + F(h) - F(t(h))$ as $s(g) = t(h)$.

In particular, if an affine function F satisfies $F|_M = 0$, it is called a multiplicative function. By the definition, the space of affine functions is a vector space with the space of multiplicative functions as a subspace.

As a corollary of Proposition 4.3 for 0-forms, we have:

Lemma 5.2. *Let $F \in C^\infty(\mathcal{G})$ be a function on a Lie groupoid $\mathcal{G} \rightrightarrows M$ and $f = \iota^* F \in C^\infty(M)$ be the restriction of F on M, where $\iota : M \hookrightarrow \mathcal{G}$ is the natural inclusion. Define*

$$F_l = F - s^* f, \qquad F_r = F - t^* f.$$

Then F is affine if and only if F_l or F_r is a multiplicative function on $\mathcal{G}$.

Example 5.3. For any function $f \in C^\infty(M)$, we see that $s^*f - t^*f \in C^\infty(\mathcal{G})$ is a multiplicative function on $\mathcal{G}$ and s^*f and t^*f are affine functions on $\mathcal{G}$.

Consider the Lie groupoid

$$\tilde{\mathcal{G}} : \oplus^q T\mathcal{G} \oplus^p T^*\mathcal{G} \rightrightarrows \oplus^q TM \oplus^p A^*.$$

A (p,q)-tensor $F \in \Gamma(\wedge^p T\mathcal{G} \otimes \wedge^q T^*\mathcal{G})$ on $\mathcal{G}$ can be viewed as a function on $\tilde{\mathcal{G}}$.

Definition 5.4. A (p,q)-tensor $F \in \Gamma(\wedge^p T\mathcal{G} \otimes \wedge^q T^*\mathcal{G})$ on a Lie groupoid $\mathcal{G}$ is called *affine* if it is an affine function on $\tilde{\mathcal{G}}$.

The following proposition ensures the consistence of this definition with Definitions 3.1 and 4.1 for the cases of affine k-vector fields and affine k-forms.

Proposition 5.5. (i) *An affine $(p,0)$-tensor is an affine p-vector field as defined in Definition 3.1.*

(ii) *An affine $(0,q)$-tensor is an affine q-form as defined in Definition 4.1.*

Proof. Let $F\in\Gamma(\wedge^p T\mathcal{G})$ be an affine $(p,0)$-tensor. Namely,

$$F(\xi^1_g\cdot\eta^1_h,\dots,\xi^p_g\cdot\eta^p_h)=F(\xi^1_g,\dots,\xi^p_g)+F(\eta^1_h,\dots,\eta^p_h)-F(s(\xi^1_g),\dots,s(\xi^p_g)),$$

for $\xi^i_g\in T^*_g\mathcal{G}$ and $\eta^i_h\in T^*_h\mathcal{G}$ such that $s(\xi^i_g)=t(\eta^i_h)$, where s,t are the source and target maps in the Lie groupoid $T^*\mathcal{G}\rightrightarrows A^*$.

Let f equal $\mathrm{pr}_{\wedge^p A}F|_M$, the projection of F restricting on M to $\wedge^k A$. We claim that F is an affine $(p,0)$-tensor if and only if $F-\overrightarrow{f}$ is a multiplicative $(p,0)$-tensor, that is,

$$(F-\overrightarrow{f})(\xi^1_g\cdot\eta^1_h,\dots,\xi^p_g\cdot\eta^p_h)=(F-\overrightarrow{f})(\xi^1_g,\dots,\xi^p_g)+(F-\overrightarrow{f})(\eta^1_h,\dots,\eta^p_h).$$

By (1), we have $t^*f=\overrightarrow{f}$, where t is the target map in $T^*\mathcal{G}\rightrightarrows A^*$. The assertion holds by Lemma 5.2.

By [Iglesias-Ponte et al. 2012, Proposition 2.7], $F-\overrightarrow{f}$ is a multiplicative function if and only if it is a multiplicative p-vector field on $\mathcal{G}$, which is further equivalent to F being an affine p-vector field by Proposition 3.4. So F is an affine $(p,0)$-tensor if and only if F is an affine p-vector field.

If $F\in\Gamma(\wedge^q T^*\mathcal{G})$ is an affine $(0,q)$-tensor, then

$$F(X\cdot Y)=F(X)+F(Y)-F(s(X)),$$

where $X\in\wedge^q T_g\mathcal{G},\ Y\in\wedge^q T_h\mathcal{G},\ (g,h)\in\mathcal{G}^{(2)}$ and $s(X)=t(Y)$, which implies that

$$m^*F=\mathrm{pr}_1^*F+\mathrm{pr}_2^*F-\mathrm{pr}_1^*s^*F.$$

So F is an affine q-form as defined in Definition 4.1. □

Regarding the relation between affine and multiplicative (p,q)-tensors, we also have the assertion as for affine k-vector fields and affine k-forms.

Let $f\in\Gamma(\wedge^p A\otimes\wedge^q T^*M)$. View it as a function on the base manifold of the Lie groupoid

$$\tilde{\mathcal{G}}:\oplus^q T\mathcal{G}\oplus^p T^*\mathcal{G}\rightrightarrows\oplus^q TM\oplus^p A^*.$$

By Example 5.3, $s^*_{\tilde{\mathcal{G}}}f$ and $t^*_{\tilde{\mathcal{G}}}f$ are affine functions on $\tilde{\mathcal{G}}$ and hence affine (p,q)-tensors on $\mathcal{G}$, where $s_{\tilde{\mathcal{G}}}$ and $t_{\tilde{\mathcal{G}}}$ are the source and target maps of the Lie groupoid $\tilde{\mathcal{G}}$.

Lemma 5.6. *Let $f\in\Gamma(\wedge^p A\otimes\wedge^q T^*M)$. We denote $\overleftarrow{f}:=s^*_{\tilde{\mathcal{G}}}f$ and $\overrightarrow{f}:=t^*_{\tilde{\mathcal{G}}}f$, where $\overleftarrow{f},\overrightarrow{f}\in\Gamma(\wedge^p T\mathcal{G}\otimes\wedge^q T^*\mathcal{G})$. We have*

$$\overleftarrow{f}(X_1,\dots,X_q)=\overleftarrow{f(sX_1,\dots,sX_q)},\qquad \overrightarrow{f}(X_1,\dots,X_q)=\overrightarrow{f(tX_1,\dots,tX_q)},\tag{18}$$

for $X_i\in T\mathcal{G}$.

Proof. This follows from the definition and (1). □

If assuming $f = u \otimes \beta$ for $u \in \Gamma(\wedge^p A)$ and $\beta \in \Omega^q(M)$, we get

$$\overleftarrow{f} = \overleftarrow{u} \otimes s^*\beta, \quad \overrightarrow{f} = \overrightarrow{u} \otimes t^*\beta.$$

The following result is a direct consequence of Lemma 5.2.

Proposition 5.7. *Let $F \in \Gamma(\wedge^p T\mathcal{G} \otimes \wedge^q T^*\mathcal{G})$ and $f = \mathrm{pr}_{\wedge^p A \otimes \wedge^q T^*M} F|_M$. Define*

$$F_l = F - \overleftarrow{f}, \quad F_r = F - \overrightarrow{f},$$

where $\overleftarrow{f}$ and $\overrightarrow{f}$ are defined in (18). Then F is an affine (p, q)-tensor on $\mathcal{G}$ if and only if F_l or F_r is a multiplicative (p, q)-tensor.

Denote by $T^{p,q}_{\mathrm{aff}}(\mathcal{G})$ and $T^{p,q}_{\mathrm{mult}}(\mathcal{G})$ the spaces of affine and multiplicative (p, q)-tensors on $\mathcal{G}$, respectively. It is immediate that $T^{p,q}_{\mathrm{aff}}(\mathcal{G})$ is a vector space with $T^{p,q}_{\mathrm{mult}}(\mathcal{G})$ being a linear subspace.

Theorem 5.8. *With the above notation, we have a 2-vector space*

$$T^{p,q}_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows T^{p,q}_{\mathrm{mult}}(\mathcal{G}),$$

where the source and target maps of the groupoid structure are

$$s(F) = F_r, \quad t(F) = F_l \qquad \text{for all } F \in T^{p,q}_{\mathrm{aff}}(\mathcal{G}),$$

and for a pair $F_1, F_2 \in T^{p,q}_{\mathrm{aff}}(\mathcal{G})$ such that $(F_1)_r = (F_2)_l$, the multiplication is

$$F_1 * F_2 = F_1 + \overleftarrow{f_2}, \quad f_2 = \mathrm{pr}_{\wedge^p A \otimes \wedge^q T^*M} F_2|_M.$$

Proof. The proof is similar to that for Theorem 3.8. □

Corollary 5.9. *The 2-term complex of vector spaces of the above 2-vector space is*

$$\Gamma(\wedge^p A \otimes \wedge^q T^*M) \to T^{p,q}_{\mathrm{mult}}(\mathcal{G}), \quad f \mapsto \overrightarrow{f} - \overleftarrow{f}.$$

An *IM (p, q)-tensor* ([Bursztyn and Drummond 2019]) on a Lie algebroid A is a triple (D, l, r), where

$$l : A \to \wedge^p A \otimes \wedge^{q-1} T^*M$$

and

$$r : T^*M \to \wedge^{p-1} A \otimes \wedge^q T^*M$$

are bundle maps covering the identity, and

$$D : \Gamma(A) \to \Gamma(\wedge^p A \otimes \wedge^q T^*M)$$

is an $\mathbb{R}$-linear map satisfying

$$D(fX) = f D(X) + df \wedge l(X) - X \wedge r(df), \quad f \in C^\infty(M),\ X \in \Gamma(A).$$

The following equations hold:

$$\begin{aligned} D[X,Y] &= X\cdot D(Y) - Y\cdot D(X),\\ l[X,Y] &= X\cdot l(Y) - \iota_{\rho(Y)} D(X),\\ r(\mathcal{L}_{\rho(X)}\alpha) &= X\cdot r(\alpha) - \iota_{\rho^*(\alpha)} D(X),\\ \iota_{\rho(X)} l(Y) &= -\iota_{\rho(Y)} l(X),\\ \iota_{\rho^*(\alpha)} r(\beta) &= -\iota_{\rho^*(\beta)} r(\alpha),\\ \iota_{\rho(X)} r(\alpha) &= \iota_{\rho^*(\alpha)} l(X), \end{aligned}$$

for $X, Y \in \Gamma(A)$ and $\alpha, \beta \in \Omega^1(M)$. Here $\cdot$ denotes the action of $\Gamma(A)$ on

$$\Gamma(\wedge^p A \otimes \wedge^q T^* M)$$

by

$$X\cdot(Z\otimes\gamma) = [X,Z]\otimes\gamma + Z\otimes\mathcal{L}_{\rho(X)}\gamma, \quad \gamma\in\Omega^q(M),\ Z\in\Gamma(\wedge^p A).$$

Denote by $T^{p,q}_{\mathrm{IM}}A$ the space of IM (p,q)-tensors on A.

The universal lifting theorem for multiplicative (p,q)-tensors is given in [Bursztyn and Drummond 2019] as follows: If $\mathcal{G}$ is an s-simply connected and s-connected Lie groupoid, then there is a one-to-one correspondence between multiplicative (p,q)-tensors on $\mathcal{G}$ and IM (p,q)-tensors on the Lie algebroid A of $\mathcal{G}$.

Based on this result and Proposition 5.7, we have the universal lifting theorem for affine (p,q)-tensors.

Proposition 5.10. *If $\mathcal{G}$ is an s-simply connected and s-connected Lie groupoid, then we have the following isomorphisms of vector spaces*:

$$T^{p,q}_{\mathrm{aff}}(\mathcal{G}) \cong T^{p,q}_{\mathrm{mult}}(\mathcal{G}) \oplus \Gamma(\wedge^p A\otimes\wedge^q T^*M) \cong T^{p,q}_{\mathrm{IM}}A \oplus \Gamma(\wedge^p A\otimes\wedge^q T^*M),$$
$$F \mapsto (F - \overrightarrow{f}, f) \mapsto (D, l, r, f),$$

where $f = \mathrm{pr}_{\wedge^p A\otimes\wedge^q T^*M} F|_M$ and $\overrightarrow{f}$ is defined in (18).

5B. ***The Frölicher–Nijenhuis bracket on affine vector-valued forms.*** A vector-valued form on a manifold M is an element in $\Omega^\bullet(M, TM) = \Gamma(TM\otimes\wedge^\bullet T^*M)$. So a vector-valued q-form on M is actually a $(1,q)$-tensor. The space of vector-valued forms relative to the Frölicher–Nijenhuis bracket is a graded Lie algebra [Frölicher and Nijenhuis 1956].

In [Bursztyn and Drummond 2013], the authors proved that the multiplicative vector-valued forms on a Lie groupoid are closed under the Frölicher–Nijenhuis bracket. Thus they form a graded Lie algebra. We shall prove that affine vector-valued forms are also closed under the Frölicher–Nijenhuis bracket. Moreover, the space of affine vector-valued forms is a graded strict Lie 2-algebra over the space of multiplicative vector-valued forms.

One formula for the Frölicher–Nijenhuis bracket

$$[\,\cdot\,,\cdot\,]_{\mathrm{FN}} : \Omega^k(M, TM) \times \Omega^q(M, TM) \to \Omega^{k+q}(M, TM)$$

is as follows:

$$\begin{aligned}[X \otimes \phi, Y \otimes \psi]_{\mathrm{FN}} = [X, Y] \otimes \phi \wedge \psi + Y \otimes \phi \wedge \mathcal{L}_X \psi - X \otimes \mathcal{L}_Y \phi \wedge \psi \\ + (-1)^k (Y \otimes d\phi \wedge \iota_X \psi + X \otimes \iota_Y \phi \wedge d\psi),\end{aligned} \tag{19}$$

where $X, Y \in \mathfrak{X}(M)$, $\phi \in \Omega^k(M)$ and $\psi \in \Omega^q(M)$, and the bracket $[\,\cdot\,,\cdot\,]$ and d on the right-hand side are the Schouten bracket and the de Rham differential. When $k = q = 0$, this bracket agrees with the Schouten bracket on vector fields. We refer to [Bursztyn and Drummond 2013; 2019] for an intrinsic definition of this bracket.

Now we discuss the vector-valued forms on a Lie groupoid $\mathcal{G}$. Recall that $T^{1,q}_{\mathrm{aff}}(\mathcal{G})$ and $T^{1,q}_{\mathrm{mult}}(\mathcal{G})$ are spaces of affine and multiplicative $(1, q)$-tensors, respectively.

Theorem 5.11. *Let $F \in T^{1,k}_{\mathrm{aff}}(\mathcal{G})$ and $N \in T^{1,q}_{\mathrm{aff}}(\mathcal{G})$ be affine tensors. Then $[F, N]_{\mathrm{FN}}$ is an affine $(1, k+q)$-tensor on $\mathcal{G}$.*

Proof. By Proposition 5.7, from F and N, we get two multiplicative tensors:

$$F_r = F - \overrightarrow{f}, \quad N_r = N - \overrightarrow{n},$$

where $f = \mathrm{pr}_{A \otimes \wedge^k T^*M} F|_M$ and $n = \mathrm{pr}_{A \otimes \wedge^q T^*M} N|_M$ and $\vec{f}$ and $\vec{n}$ are defined in (18). Based on this, we have

$$\begin{aligned}[F, N]_{\mathrm{FN}} &= [F_r + \overrightarrow{f}, N_r + \overrightarrow{n}]_{\mathrm{FN}} \\ &= [F_r, N_r]_{\mathrm{FN}} + [F_r, \overrightarrow{n}]_{\mathrm{FN}} + [\overrightarrow{f}, N_r]_{\mathrm{FN}} + [\overrightarrow{f}, \overrightarrow{n}]_{\mathrm{FN}}.\end{aligned} \tag{20}$$

By [Bursztyn and Drummond 2013, Theorem 4.3], we have $[F_r, N_r]_{\mathrm{FN}} \in T^{1,k+p}_{\mathrm{mult}}(\mathcal{G})$, a multiplicative $(1, k+p)$-tensor. By [Bursztyn and Drummond 2019, Lemma 5.3], we have

$$[\overrightarrow{f}, N_r]_{\mathrm{FN}} = \overrightarrow{D_N(f)},$$

where $D_N : \Gamma(A \otimes \wedge^k T^*M) \to \Gamma(A \otimes \wedge^{k+q} T^*M)$ is determined by the IM $(1, q)$-tensor (D, l, r) of the multiplicative $(1, q)$-tensor N_r. We refer to [Bursztyn and Drummond 2019] for details. Now it suffices to check that

$$[\overrightarrow{f}, \overrightarrow{n}]_{\mathrm{FN}} = \overrightarrow{s}, \tag{21}$$

for some $s \in \Gamma(A \otimes \wedge^{k+q} T^*M)$. Assume $f = u \otimes \alpha$, $n = v \otimes \beta$ for $u, v \in \Gamma(A)$ and $\alpha \in \Omega^k(M)$, $\beta \in \Omega^q(M)$. Then by (19),

$$\begin{aligned}[\overrightarrow{f}, \overrightarrow{n}]_{\mathrm{FN}} &= [\overrightarrow{u} \otimes t^*\alpha, \overrightarrow{v} \otimes t^*\beta] \\ &= \overrightarrow{[u, v]} \otimes t^*(\alpha \wedge \beta) + \overrightarrow{v} \otimes t^*(\alpha \wedge \mathcal{L}_{\rho(u)}\beta) - \overrightarrow{u} \otimes t^*(\mathcal{L}_{\rho(v)}\alpha \wedge \beta) \\ &\quad + (-1)^k (\overrightarrow{v} \otimes t^*(d\alpha \wedge \iota_{\rho(u)}\beta) + \overrightarrow{u} \otimes t^*(\iota_{\rho(v)}\alpha \wedge d\beta)),\end{aligned}$$

where we have used the relations $\iota_{\overrightarrow{u}} t^*\beta = t^*\iota_{\rho(u)}\beta$ and $d\circ t^* = t^*\circ d$. Write

$$s = [u, v]\otimes\alpha\wedge\beta + v\otimes\alpha\wedge\mathcal{L}_{\rho(u)}\beta - u\otimes\mathcal{L}_{\rho(v)}\alpha\wedge\beta + (-1)^k(v\otimes d\alpha\wedge\iota_{\rho(u)}\beta + u\otimes\iota_{\rho(v)}\alpha\wedge d\beta).$$

We get (21). Following (20), we have shown that

$$[F, N]_{\mathrm{FN}} = [F_r, N_r]_{\mathrm{FN}} - (-1)^{kq}\overrightarrow{D_F(n)} + \overrightarrow{D_N(f)} + \overrightarrow{s},$$

where $D_F : \Gamma(A\otimes\wedge^q T^*M) \to \Gamma(A\otimes\wedge^{k+q}T^*M)$ is determined by the IM $(1, k)$-tensor (D', l', r') of the multiplicative $(1, k)$-tensor F_r. Thus $[F, N]_{\mathrm{FN}}$ is affine. □

Proposition 5.12. *We have a graded strict Lie 2-algebra structure on*

$$\oplus_k T^{1,k}_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows \oplus_k T^{1,k}_{\mathrm{mult}}(\mathcal{G}),$$

where the bracket is the Frölicher–Nijenhuis bracket.

Proof. By Theorems 5.8 and 5.11, we only need to show that the Frölicher–Nijenhuis bracket is a functor. This is similar to the proof in Theorem 3.11. We omit the detail. □

5C. *The strict monoidal category of affine* (1, 1)*-tensors.* Another important case is affine (1, 1)-tensors, which can be used to define affine Nijenhuis operators on a Lie groupoid. On the space of affine (1, 1)-tensors, in addition to the 2-vector space structure proposed in Theorem 5.8, we shall also construct a strict monoidal category structure in this subsection.

A (1, 1)-tensor on a Lie groupoid $\mathcal{G}$ is *multiplicative* if the induced bundle map $T\mathcal{G} \to T\mathcal{G}$ is a Lie groupoid morphism [Laurent-Gengoux et al. 2009], which amounts to saying that the corresponding function on $T^*\mathcal{G}\oplus T\mathcal{G} \rightrightarrows A^*\oplus TM$ is multiplicative by [Bursztyn and Drummond 2019, Proposition 3.9]. Thus the composition of two multiplicative (1, 1)-tensors is still a multiplicative (1, 1)-tensor. We shall show that the composition of two affine (1, 1)-tensors is also an affine (1, 1)-tensor.

By (17), a (1, 1)-tensor $N \in \Gamma(T\mathcal{G}\otimes T^*\mathcal{G})$ is affine if it satisfies

$$N(X\cdot Y, \xi\cdot\eta) = N(X, \xi) + N(Y, \eta) - n(s_{T\mathcal{G}}X, s_{T^*\mathcal{G}}\xi), \quad n = \mathrm{pr}_{A\otimes T^*M} N|_M,$$

where $(X, Y) \in T\mathcal{G}^{(2)}$ and $(\xi, \eta) \in T^*\mathcal{G}^{(2)}$ are multiplicable pairs in $T\mathcal{G}$ and $T^*\mathcal{G}$ covering the same pair $(g, h) \in \mathcal{G}^{(2)}$. Here $s_{T\mathcal{G}}$ and $s_{T^*\mathcal{G}}$ are the source maps of the tangent and cotangent groupoids, respectively.

A multiplicative (1, 1)-tensor N corresponds to a Lie groupoid morphism

$$(N, n_{TM}) : T\mathcal{G} \to T\mathcal{G},$$

where $n_{TM} : TM \to TM$ is the map on the base manifold. Since N preserves the

s-fibers, it induces a bundle map $n_A : A \to A$. Then we have $N|_M = n_{TM} + n_A$. For an affine (1, 1)-tensor N, from the difference of affine and multiplicative (1, 1)-tensors, we have that the restriction of N on M is

$$N|_M = \begin{pmatrix} n_{TM} & 0 \\ n & n_A \end{pmatrix} : TM \oplus A \to TM \oplus A,$$

where $n_{TM} : TM \to TM$ and $n_A : A \to A$ and $n = \mathrm{pr}_{A\otimes T^*M} N|_M$.

Lemma 5.13. *The composition $N \circ N'$ of two affine $(1, 1)$-tensors N and N' is still an affine $(1, 1)$-tensor with*

$$(N \circ N')_l = N_l \circ N'_l, \quad (N \circ N')_r = N_r \circ N'_r \quad \textit{for all } N, N' \in T^{1,1}_{\mathrm{aff}}(\mathcal{G}).$$

Moreover, the $A \otimes T^*M$-component of $N \circ N'|_M$ is

$$\mathrm{pr}_{A\otimes T^*M} N \circ N'|_M = n_A \circ n' + n \circ n'_{TM} + n \circ \rho \circ n',$$

where

$$N|_M = \begin{pmatrix} n_{TM} & 0 \\ n & n_A \end{pmatrix}, \quad N'|_M = \begin{pmatrix} n'_{TM} & 0 \\ n' & n'_A \end{pmatrix} : TM \oplus A \to TM \oplus A$$

are the decompositions of N and N' restricting on M and $\rho : A \to TM$ is the anchor map.

Proof. Write $N = N_r + \overrightarrow{n}$ and $N' = N'_r + \overrightarrow{n'}$, where n and n' are the $A \otimes T^*M$-components of $N|_M$ and $N'|_M$, respectively, and $\overrightarrow{n}$, $\overrightarrow{n'}$ are defined in (18). Then

$$N \circ N' = N_r \circ N'_r + N_r \circ \overrightarrow{n'} + \overrightarrow{n} \circ N'_r + \overrightarrow{n} \circ \overrightarrow{n'}.$$

By Proposition 5.7, $(N_r, n_{TM}), (N'_r, n'_{TM}) : T\mathcal{G} \to T\mathcal{G}$ are morphisms of Lie groupoids. So we have the formulas

$$N_r(\overrightarrow{u}) = \overrightarrow{n_A(u)} \quad \text{for all } u \in \Gamma(A)$$

and $t \circ N'_r = n'_{TM} \circ t$. Applying this to $X \in T_g\mathcal{G}$, we get

$$\begin{aligned} N_r \circ \overrightarrow{n'}(X) &= N_r \overrightarrow{n'(tX)} = \overrightarrow{n_A(n'(tX))} = \overrightarrow{n_A \circ n'}(X), \\ \overrightarrow{n} \circ N'_r(X) &= \overrightarrow{n(tN'_r(X))} = \overrightarrow{n(n'_{TM}(tX))} = \overrightarrow{n \circ n'_{TM}}(X), \\ \overrightarrow{n} \circ \overrightarrow{n'}(X) &= \overrightarrow{n}\, \overrightarrow{n'(tX)} = \overrightarrow{n(\rho(n'(tX)))} = \overrightarrow{n \circ \rho \circ n'}(X). \end{aligned}$$

Thus we proved that

$$N \circ N' = N_r \circ N'_r + \overrightarrow{n_A \circ n' + n \circ n'_{TM} + n \circ \rho \circ n'}.$$

Notice that the composition $N_r \circ N'_r$ of two multiplicative (1, 1)-tensors is still multiplicative. Then by Proposition 5.7, we obtain that $N \circ N'$ is an affine $(1, 1)$-tensor with the properties as desired. □

Actually, the 2-vector space $T^{1,1}_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows T^{1,1}_{\mathrm{mult}}(\mathcal{G})$ from Theorem 5.8 with the composition is a strict monoidal category.

Theorem 5.14. *We have a strict monoidal category structure on the* 2*-vector space*

$$T^{1,1}_{\mathrm{aff}}(\mathcal{G}) \rightrightarrows T^{1,1}_{\mathrm{mult}}(\mathcal{G}),$$

with the product being the composition of two affine $(1, 1)$*-tensors and the unit given by the identity* $I : T\mathcal{G} \to T\mathcal{G}$.

Proof. It is obvious that the identity $I : T\mathcal{G} \to T\mathcal{G}$, as a multiplicative $(1, 1)$-tensor, is a left and right unit for the composition.

It suffices to verify that the composition $\circ : T^{1,1}_{\mathrm{aff}}(\mathcal{G}) \times T^{1,1}_{\mathrm{aff}}(\mathcal{G}) \to T^{1,1}_{\mathrm{aff}}(\mathcal{G})$ is a bifunctor. Let N_1, N_2, N_3, N_4 be four affine $(1, 1)$-tensors such that $(N_1)_r = (N_3)_l$ and $(N_2)_r = (N_4)_l$. That is,

$$N_1 - \overrightarrow{n_1} = N_3 - \overleftarrow{n_3} \quad \text{and} \quad N_2 - \overrightarrow{n_2} = N_4 - \overleftarrow{n_4}.$$

By Lemma 5.13, we see $(N_1 \circ N_2)_r = (N_3 \circ N_4)_l$. Then we prove

$$(N_1 * N_3) \circ (N_2 * N_4) = (N_1 \circ N_2) * (N_3 \circ N_4). \tag{22}$$

The left-hand side of (22) is equal to

$$(N_1 + \overleftarrow{n_3}) \circ (N_2 + \overleftarrow{n_4}) = N_1 \circ N_2 + N_1 \circ \overleftarrow{n_4} + \overleftarrow{n_3} \circ N_2 + \overleftarrow{n_3} \circ \overleftarrow{n_4},$$

and by Lemma 5.13, the right-hand side of (22) amounts to

$$N_1 \circ N_2 + \overleftarrow{(n_{3A} \circ n_4 + n_3 \circ n_{4TM} + n_3 \circ \rho \circ n_4)}.$$

By the same calculation in Lemma 5.13 for the left translation instead of the right translation, we get

$$N_1 \circ \overleftarrow{n_4} = (N_3 - \overleftarrow{n_3} + \overrightarrow{n_1}) \circ \overleftarrow{n_4} = \overleftarrow{n_{3A} \circ n_4} + \overrightarrow{n_1} \circ \overleftarrow{n_4} = \overleftarrow{n_{3A} \circ n_4},$$

which follows from

$$\overrightarrow{n_1} \circ \overleftarrow{n_4}(X) = \overrightarrow{n_1}(\overleftarrow{n_4(sX)}) = \overrightarrow{n_1(t\overleftarrow{n_4(sX)})} = 0, \quad X \in T_g\mathcal{G}.$$

Similarly, we have

$$\overleftarrow{n_3} \circ N_2 = \overleftarrow{n_3 \circ n_{4TM}}.$$

Observe that

$$\overleftarrow{n_3} \circ \overleftarrow{n_4} = \overleftarrow{n_3 \circ \rho \circ n_4}$$

for the same reason as for the right translation proved in Lemma 5.13. This proves (22). □

Remark 5.15. This strict monoidal category from affine (1, 1)-tensors is related to the 2-vector spaces constructed from affine 2-vector fields and 2-forms as in Theorems 3.8 and 4.4 if we take into consideration the generalized tangent bundle $T\mathcal{G} \oplus T^*\mathcal{G} \rightrightarrows TM \oplus A^*$. Identify an affine 2-vector field $\Pi \in \mathfrak{X}^2(\mathcal{G})$ with a matrix

$$\begin{pmatrix} I & \Pi \\ 0 & I \end{pmatrix} : T\mathcal{G} \oplus T^*\mathcal{G} \to T\mathcal{G} \oplus T^*\mathcal{G}.$$

Then the addition in the vector space $\mathfrak{X}^2_{\mathrm{aff}}(\mathcal{G})$ is actually the composition of two affine 2-vector fields as matrices. Likewise, viewing an affine 2-form $\Theta \in \Omega^2(\mathcal{G})$ as a matrix

$$\begin{pmatrix} I & 0 \\ \Theta & I \end{pmatrix} : T\mathcal{G} \oplus T^*\mathcal{G} \to T\mathcal{G} \oplus T^*\mathcal{G},$$

we get that the addition in $\Omega^2_{\mathrm{aff}}(\mathcal{G})$ is the composition of two affine 2-forms as matrices.

Remark 5.16. As a multiplicative (1, 1)-tensor can be characterized as a Lie groupoid morphism from $T\mathcal{G}$ to $T\mathcal{G}$, a multiplicative p-vector field defines a morphism of Lie groupoids from $\oplus^{p-1} T^*\mathcal{G}$ to $T\mathcal{G}$ and a multiplicative p-form defines a morphism of Lie groupoids from $\oplus^{p-1} T\mathcal{G}$ to $T^*\mathcal{G}$ [Bursztyn and Drummond 2019]. In these cases, viewing multiplicative structures as functors, we see that affine structures are actually natural transformations between these multiplicative structures.

Example 5.17. One example of multiplicative (1, 1)-tensors on a Lie groupoid $\mathcal{G}$ is $\overrightarrow{n} - \overleftarrow{n}$ for any $n \in \Gamma(A \otimes T^*M)$. And $\overrightarrow{n}$ and $\overleftarrow{n}$ are both affine (1, 1)-tensors on $\mathcal{G}$.

Example 5.18. An affine (1, 1)-tensor on a Lie group G is a multiplicative (1, 1)-tensor on G, which is a G-equivariant linear map from $\mathfrak{g} := \mathrm{Lie}(G)$ to $\mathfrak{g}$. Namely,

$$T^{1,1}_{\mathrm{aff}}(G) = \{N \in \mathrm{End}(\mathfrak{g}) \mid N(\mathrm{Ad}_g u) = \mathrm{Ad}_g N(u),\ g \in G, u \in \mathfrak{g}\}.$$

The product of two affine (1, 1)-tensors in the strict monoidal category structure is the composition of linear maps and the groupoid multiplication is trivial, i.e., $N * N = N$, meaning that an affine (1, 1)-tensor can only multiply itself, which results in itself.

Example 5.19. For the pair groupoid $\mathcal{G} = M \times M \rightrightarrows M$, a multiplicative (1, 1)-tensor on $\mathcal{G}$ is always of the form $\overrightarrow{N} - \overleftarrow{N}$ for a bundle map $N : TM \to TM$, where by definition,

$$(\overrightarrow{N} - \overleftarrow{N})(X, Y) = \overrightarrow{N(X)} - \overleftarrow{N(Y)} = (N(X), 0) + (0, N(Y)) = (N(X), N(Y)),$$

for all $(X, Y) \in T_xM \times T_yM$. In fact, for any $u \in \mathfrak{X}(M)$, we get $\overrightarrow{u}(x, y) = \frac{d}{dt}|_{t=0}(\phi^u_t(x), x)(x, y) = (u, 0)$, where $\phi^u_t(x)$ is a flow of u such that $\phi^u_0(x) = x$.

And $\overleftarrow{u}(x,y) = -\frac{d}{dt}|_{t=0}(x,y)(y,\phi_t^u(y)) = -(0,u)$, where the minus sign comes from the convention that $\overleftarrow{u}(\cdot) = -L_{(\cdot)}\mathrm{inv}(u)$.

By the relation between affine and multiplicative (1, 1)-tensors, an affine (1, 1)-tensor is of the form $\overrightarrow{N} + \overleftarrow{N'}$ for two bundle maps $N, N' \in \mathrm{End}(TM)$. For simplicity, we write an affine (1, 1)-tensor as (N, N'). The product in the strict monoidal category structure is the composition of two (1, 1)-tensors:

$$(N_1, N_2) \circ (N_3, N_4) = (N_1 \circ N_3, -N_2 \circ N_4) \quad \text{for all } N_i \in \mathrm{End}(TM),$$

which follows from

$$\begin{aligned}(\overrightarrow{N_1} + \overleftarrow{N_2}) \circ (\overrightarrow{N_3} + \overleftarrow{N_4})(X,Y) &= (\overrightarrow{N_1} + \overleftarrow{N_2})(N_3(X), -N_4(Y)) \\ &= (N_1 \circ N_3(X), N_2 \circ N_4(Y)).\end{aligned}$$

For the groupoid multiplication, two affine (1, 1)-tensors (N_1, N_2) and (N_3, N_4) are multiplicable if and only if $N_2 = -N_3$ and the multiplication is

$$(N_1, N_2) * (N_3, N_4) = \overrightarrow{N_1} + \overleftarrow{N_2} + \overleftarrow{N_3} + \overleftarrow{N_4} = \overrightarrow{N_1} + \overleftarrow{N_4} = (N_1, N_4).$$

The next example we are interested in is the direct sum of the pair groupoid and a Lie group: $M \times M \times G \rightrightarrows M$. The following proposition tells us that this case is just the direct sum of Examples 5.18 and 5.19. So the strict monoidal category structure for this case is also clear.

Proposition 5.20. *An affine* (1, 1)*-tensor N on $M \times M \times G \rightrightarrows M$ is of the form*

$$N(X,Y,g,u) = (N_1(X), N_2(Y), g, L(u)) \quad \textit{for all } N_1, N_2 \in \mathrm{End}(TM),\ L \in \mathrm{End}(\mathfrak{g})^G,$$

for $X \in T_xM$, $Y \in T_yM$, $g \in G$ and $u \in \mathfrak{g}$. It is multiplicative if and only if $N_1 = N_2$.

Proof. A multiplicable (1, 1)-tensor field on $\mathcal{G}$ is a bundle map

$$N = (N_1, N_2, N_3) : TM \times TM \times TG \to TM \times TM \times TG$$

over the base manifold $M \times M \times G$ and also a groupoid morphism over TM to itself. We claim that a multiplicative (1, 1)-tensor field can only be of the form

$$N(X,Y,g,u) = (N(X), N(Y), g, L(u)) \quad \text{for all } X \in T_xM,\ Y \in T_yM,\ g \in G,\ u \in \mathfrak{g},$$

for some $N \in \mathrm{End}(TM)$ and a G-equivariant linear map $L \in \mathrm{End}(\mathfrak{g})^G$. In fact, since N preserves the s and t-fibers, the first and second components N_1, N_2 have to be the same. Then N being a groupoid morphism requires that

$$N_3(X, Y, g, u) + \mathrm{Ad}_g N_3(Y, Z, h, v) = N_3(X, Z, gh, u + \mathrm{Ad}_g v).$$

Since N_3 is a bundle map, it follows that $N_3(X, Y, g, u)$ is independent of X and Y and it is determined by a G-equivariant linear map $L \in \mathrm{End}(\mathfrak{g})$. We thus easily get that an affine (1, 1)-tensor field N on $M \times M \times G \rightrightarrows M$ is of the form

$$N(X, Y, g, u) = (N_1(X), N_2(Y), g, L(u)) \quad \text{for all } N_1, N_2 \in \mathrm{End}(TM),\ L \in \mathrm{End}(\mathfrak{g}).$$

Hence the strict monoidal structure for this case is clear. □

At the end of this section, we provide a class of affine (1, 1)-tensors coming from the composition of affine 2-vector fields and affine 2-forms.

Proposition 5.21. *The composition $\Pi \circ \Theta : T\mathcal{G} \to T\mathcal{G}$ of an affine 2-vector field $\Pi \in \mathfrak{X}^2(\mathcal{G})$ and an affine 2-form $\Theta \in \Omega^2(\mathcal{G})$ is an affine (1, 1)-tensor with*

$$\mathrm{pr}_{A\otimes T^*M}(\Pi \circ \Theta)|_M = \pi_{A^*} \circ \theta + \pi \circ \theta_{TM} + \pi \circ \rho^* \circ \theta,$$

where $\pi \in \Gamma(\wedge^2 A)$, $\pi_{A^*} \in \Gamma(A \otimes TM)$ and $\theta_{TM} \in \Gamma(T^*M \otimes A^*)$ and $\theta \in \Omega^2(M)$ are the corresponding components of Π and Θ restricting on M, respectively.

Moreover, the associated two multiplicative (1, 1)-tensors are

$$(\Pi \circ \Theta)_l = \Pi_l \circ \Theta_l, \quad (\Pi \circ \Theta)_r = \Pi_r \circ \Theta_r.$$

Proof. Denote $\Pi = \Pi_r + \overrightarrow{\pi}$ and $\Theta = \Theta_r + t^*\theta$, where Π_r and Θ_r are the associated multiplicative 2-vector field and 2-form. Then

$$\Pi \circ \Theta = \Pi_r \circ \Theta_r + \overrightarrow{\pi} \circ \Theta_r + \Pi_r \circ t^*\theta + \overrightarrow{\pi} \circ t^*\theta. \tag{23}$$

Acting on $X \in T_g\mathcal{G}$ and pairing with $\xi \in T_g^*\mathcal{G}$, we find

$$\langle \overrightarrow{\pi} \circ \Theta_r(X), \xi\rangle = \langle \pi, t\Theta_r(X) \wedge t\xi\rangle = \langle \pi, \theta_{TM}(tX) \wedge t\xi\rangle = \langle \overrightarrow{\pi \circ \theta_{TM}(tX)}, \xi\rangle,$$

where we have used (1) and the fact that (Θ_r, θ_{TM}) is a Lie groupoid morphism from $T\mathcal{G}$ to $T^*\mathcal{G}$. This implies that

$$\overrightarrow{\pi} \circ \Theta_r = \overrightarrow{\pi \circ \theta_{TM}}.$$

On the other hand, using (1) again and the fact that $(\Pi_r, \pi_{A^*}) : T^*\mathcal{G} \to T\mathcal{G}$ is a Lie groupoid morphism, we have

$$\langle \Pi_r \circ t^*\theta(X), \xi\rangle = -\langle \theta, tX \wedge t\Pi_r(\xi)\rangle = -\langle \theta(tX), \pi_{A^*}t(\xi)\rangle = \langle \overrightarrow{\pi_{A^*} \circ \theta(tX)}, \xi\rangle.$$

Thus,

$$\Pi_r \circ t^*\theta = \overrightarrow{\pi_{A^*} \circ \theta}.$$

At the end, observe that $t\overrightarrow{\pi}(\xi) = \rho\pi(t\xi)$ since

$$\langle t\overrightarrow{\pi}(\xi), \alpha\rangle = \langle \pi(t\xi), t^*\alpha\rangle = \langle \pi(t\xi), \rho^*\alpha\rangle \quad \text{for all } \alpha \in \Omega^1(M).$$

We obtain

$$\langle \overrightarrow{\pi} \circ t^*\theta(X), \xi\rangle = -\langle \theta(tX), t\overrightarrow{\pi}(\xi)\rangle = -\langle \theta(tX), \rho\pi(t\xi)\rangle = \langle \overrightarrow{\pi \circ \rho^* \circ \theta(tX)}, \xi\rangle.$$

Hence,

$$\overrightarrow{\pi} \circ t^*\theta = \overrightarrow{\pi \circ \rho^* \circ \theta}.$$

Coupled with (23), we get

$$\Pi \circ \Theta = \Pi_r \circ \Theta_r + \overrightarrow{(\pi \circ \theta_{TM} + \pi_{A^*} \circ \theta + \pi \circ \rho^* \circ \theta)}.$$

Since $\Pi_r \circ \Theta_r : T\mathcal{G} \to T\mathcal{G}$ is a Lie groupoid morphism and thus gives a multiplicative $(1, 1)$-tensor, we have proved that $\Pi \circ \Theta$ is an affine $(1, 1)$-tensor. The other assertions are also clear. □

Acknowledgments

We would like to thank the referee for very helpful suggestions and comments.

References

[Baez and Crans 2004] J. C. Baez and A. S. Crans, "Higher-dimensional algebra, VI: Lie 2-algebras", *Theory Appl. Categ.* **12** (2004), 492–538. MR Zbl

[Berwick-Evans and Lerman 2016] D. Berwick-Evans and E. Lerman, "Lie 2-algebras of vector fields", preprint, 2016. arXiv

[Bonechi et al. 2018] F. Bonechi, N. Ciccoli, C. Laurent-Gengoux, and P. Xu, "Shifted Poisson structures on differentiable stacks", preprint, 2018. arXiv

[Bursztyn and Cabrera 2012] H. Bursztyn and A. Cabrera, "Multiplicative forms at the infinitesimal level", *Math. Ann.* **353**:3 (2012), 663–705. MR Zbl

[Bursztyn and Drummond 2013] H. Bursztyn and T. Drummond, "Lie groupoids and the Frölicher–Nijenhuis bracket", *Bull. Braz. Math. Soc.* (*N.S.*) **44**:4 (2013), 709–730. MR Zbl

[Bursztyn and Drummond 2019] H. Bursztyn and T. Drummond, "Lie theory of multiplicative tensors", *Math. Ann.* **375**:3-4 (2019), 1489–1554. MR Zbl

[Bursztyn et al. 2009] H. Bursztyn, A. Cabrera, and C. Ortiz, "Linear and multiplicative 2-forms", *Lett. Math. Phys.* **90**:1-3 (2009), 59–83. MR Zbl

[Chen et al. 2013] Z. Chen, M. Stiénon, and P. Xu, "Poisson 2-groups", *J. Differential Geom.* **94**:2 (2013), 209–240. MR Zbl

[Crainic and Salazar 2015] M. Crainic and M. A. Salazar, "Jacobi structures and Spencer operators", *J. Math. Pures Appl.* (9) **103**:2 (2015), 504–521. MR Zbl

[Crainic and Zhu 2007] M. Crainic and C. Zhu, "Integrability of Jacobi and Poisson structures", *Ann. Inst. Fourier* (*Grenoble*) **57**:4 (2007), 1181–1216. MR Zbl

[Crainic et al. 2015] M. Crainic, M. A. Salazar, and I. Struchiner, "Multiplicative forms and Spencer operators", *Math. Z.* **279**:3-4 (2015), 939–979. MR Zbl

[Dazord and Sondaz 1991] P. Dazord and D. Sondaz, "Groupes de Poisson affines", pp. 99–128 in *Symplectic geometry, groupoids, and integrable systems* (Berkeley, 1989), edited by P. Dazord and A. Weinstein, Math. Sci. Res. Inst. Publ. **20**, Springer, 1991. MR Zbl

[Dazord et al. 1991] P. Dazord, J.-H. Lu, D. Sondaz, and A. Weinstein, "Affinoïdes de Poisson", *C. R. Acad. Sci. Paris Sér. I Math.* **312**:7 (1991), 523–527. MR Zbl

[Frölicher and Nijenhuis 1956] A. Frölicher and A. Nijenhuis, "Theory of vector-valued differential forms, I: Derivations of the graded ring of differential forms", *Indag. Math.* **18** (1956), 338–359. MR Zbl

[Iglesias-Ponte and Marrero 2003] D. Iglesias-Ponte and J. C. Marrero, "Jacobi groupoids and generalized Lie bialgebroids", *J. Geom. Phys.* **48**:2-3 (2003), 385–425. MR Zbl

[Iglesias-Ponte et al. 2012] D. Iglesias-Ponte, C. Laurent-Gengoux, and P. Xu, "Universal lifting theorem and quasi-Poisson groupoids", *J. Eur. Math. Soc.* **14**:3 (2012), 681–731. MR Zbl

[Jotz Lean 2019] M. Jotz Lean, "Dirac groupoids and Dirac bialgebroids", *J. Symplectic Geom.* **17**:1 (2019), 179–238. MR Zbl

[Jotz Lean and Ortiz 2014] M. Jotz Lean and C. Ortiz, "Foliated groupoids and infinitesimal ideal systems", *Indag. Math.* (*N.S.*) **25**:5 (2014), 1019–1053. MR Zbl

[Jotz Lean et al. 2016] M. Jotz Lean, M. Stiénon, and P. Xu, "Glanon groupoids", *Math. Ann.* **364**:1-2 (2016), 485–518. MR Zbl

[Kosmann-Schwarzbach 2016] Y. Kosmann-Schwarzbach, "Multiplicativity, from Lie groups to generalized geometry", pp. 131–166 in *Geometry of jets and fields* (Będlewo, Poland, 2015), edited by K. Grabowska et al., Banach Center Publ. **110**, Polish Acad. Sci. Inst. Math., Warsaw, 2016. MR Zbl

[Lang and Liu 2018] H. Lang and Z. Liu, "Coadjoint orbits of Lie groupoids", *J. Geom. Phys.* **129** (2018), 217–232. MR Zbl

[Laurent-Gengoux et al. 2009] C. Laurent-Gengoux, M. Stiénon, and P. Xu, "Integration of holomorphic Lie algebroids", *Math. Ann.* **345**:4 (2009), 895–923. MR Zbl

[Li-Bland and Ševera 2011] D. Li-Bland and P. Ševera, "Quasi-Hamiltonian groupoids and multiplicative Manin pairs", *Int. Math. Res. Not.* **2011**:10 (2011), 2295–2350. MR Zbl

[Lu 1990] J.-H. Lu, *Multiplicative and affine Poisson structures on Lie groups*, Ph.D. thesis, University of California, Berkeley, 1990, Available at https://tinyurl.com/jhuthesis.

[Mac Lane 1971] S. Mac Lane, *Categories for the working mathematician*, Grad. Texts in Math. **5**, Springer, 1971. MR Zbl

[Mackenzie 1992] K. C. H. Mackenzie, "Double Lie algebroids and second-order geometry, I", *Adv. Math.* **94**:2 (1992), 180–239. MR Zbl

[Mackenzie 2000] K. C. H. Mackenzie, "Affinoid structures and connections", pp. 175–186 in *Poisson geometry* (Warsaw, 1998), edited by J. Grabowski and P. Urbański, Banach Center Publ. **51**, Polish Acad. Sci. Inst. Math., Warsaw, 2000. MR Zbl

[Mackenzie and Xu 2000] K. C. H. Mackenzie and P. Xu, "Integration of Lie bialgebroids", *Topology* **39**:3 (2000), 445–467. MR Zbl

[Ortiz 2013] C. Ortiz, "Multiplicative Dirac structures", *Pacific J. Math.* **266**:2 (2013), 329–365. MR Zbl

[Ortiz and Waldron 2019] C. Ortiz and J. Waldron, "On the Lie 2-algebra of sections of an $\mathscr{LA}$-groupoid", *J. Geom. Phys.* **145** (2019), art. id. 103474. MR Zbl

[Urbański 1994] P. Urbański, "Affine Poisson structures in analytical mechanics", pp. 123–129 in *Quantization and infinite-dimensional systems* (Bialowieza, Poland, 1993), edited by J.-P. Antoine et al., Plenum, New York, 1994. MR Zbl

[Weinstein 1990] A. Weinstein, "Affine Poisson structures", *Int. J. Math.* **1**:3 (1990), 343–360. MR Zbl

[Xu 1995] P. Xu, "On Poisson groupoids", *Int. J. Math.* **6**:1 (1995), 101–124. MR Zbl

Received March 18, 2019. Revised February 21, 2020.

HONGLEI LANG
DEPARTMENT OF APPLIED MATHEMATICS
CHINA AGRICULTURAL UNIVERSITY
BEIJING
CHINA
hllang@cau.edu.cn

ZHANGJU LIU
SCHOOL OF MATHEMATICAL SCIENCES
PEKING UNIVERSITY
BEIJING
CHINA
liuzj@pku.edu.cn

YUNHE SHENG
DEPARTMENT OF MATHEMATICS
JILIN UNIVERSITY
CHANGCHUN
CHINA
shengyh@jlu.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.383

STRONG NEGATIVE TYPE IN SPHERES

RUSSELL LYONS

It is known that spheres have negative type, but only subsets with at most one pair of antipodal points have strict negative type. These are conditions on the (angular) distances within any finite subset of points. We show that subsets with at most one pair of antipodal points have strong negative type, a condition on every probability distribution of points. This implies that the function of expected distances to points determines uniquely the probability measure on such a set. It also implies that the distance covariance test for stochastic independence, introduced by Székely, Rizzo and Bakirov, is consistent against all alternatives in such sets. Similarly, it allows tests of goodness of fit, equality of distributions, and hierarchical clustering with angular distances. We prove this by showing an analogue of the Cramér–Wold theorem.

1. Introduction

We introduce the topic by borrowing from [Lyons 2014].

Let (X, d) be a metric space. One says that (X, d) has *negative type* if for all $n \geq 1$ and all lists of n red points x_i and n blue points x'_i in X, the sum $2\sum_{i,j} d(x_i, x'_j)$ of the distances between the $2n^2$ ordered pairs of points of opposite color is at least the sum $\sum_{i,j}(d(x_i, x_j) + d(x'_i, x'_j))$ of the distances between the $2n^2$ ordered pairs of points of the same color. It is not obvious that euclidean space has this property, but it is well known. By considering repetitions of x_i and taking limits, we arrive at a superficially more general property: For all $n \geq 1$, $x_1, \ldots, x_n \in X$, and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ with $\sum_{i=1}^n \alpha_i = 0$, we have

$$\sum_{i,j \leq n} \alpha_i \alpha_j d(x_i, x_j) \leq 0. \tag{1-1}$$

We say that (X, d) has *strict negative type* if, for every n and all n-tuples of distinct points $x_1, \ldots, x_n$, equality holds in (1-1) only when $\alpha_i = 0$ for all i. Again, euclidean

Research partially supported by NSF grant DMS-1612363.
MSC2010: primary 44A12, 45Q05, 51K99, 51M10; secondary 62G20, 62H15, 62H20, 62H30.
Keywords: Cramér–Wold, hemispheres, expected distances, distance covariance, equality of distributions, goodness of fit, hierarchical clustering.

spaces have strict negative type. A simple example of a metric space of nonstrict negative type is ℓ^1 on a 2-point space, i.e., $\mathbb{R}^2$ with the ℓ^1-metric.

We say that a (Borel) probability measure μ on X has *finite first moment* if $\int d(o, x)\,\mathrm{d}\mu(x) < \infty$ for some (hence all) $o \in X$; we write $P_1(X, d)$ for the set of such probability measures. Suppose that $\mu_1, \mu_2 \in P_1(X, d)$. By approximating μ_i by probability measures of finite support, we obtain a yet more general property, namely, that when X has negative type,

$$\int d(x_1, x_2)\,\mathrm{d}(\mu_1 - \mu_2)^2(x_1, x_2) \le 0. \tag{1-2}$$

We say that (X, d) has *strong negative type* if it has negative type and equality holds in (1-2) only when $\mu_1 = \mu_2$. See [Lyons 2018] for an example of a (countable) metric space of strict but not strong negative type. The notion of strong negative type was first defined by Zinger, Kakosyan and Klebanov [1992]. Lyons [2013] used it to show that a metric space X has strong negative type if and only if the theory of distance covariance holds in X just as in euclidean spaces, as introduced by Székely, Rizzo and Bakirov [2007]. Similarly, it allows tests of goodness of fit, equality of distributions, and hierarchical clustering with angular distances: see the review in [Székely and Rizzo 2017]. Lyons [2013] noted that if (X, d) has negative type, then (X, d^r) has strong negative type when $0 < r < 1$.

Define

$$a_\mu(x) := \int d(x, x')\,\mathrm{d}\mu(x')$$

for $x \in X$ and $\mu \in P_1(X, d)$. Lyons [2013] remarked that if (X, d) has negative type, then the map $\alpha\colon \mu \mapsto a_\mu$ is injective on $\mu \in P_1(X)$ if and only if X has strong negative type. (There are also metric spaces not of negative type for which α is injective.)

A list of metric spaces of negative type appears as Theorem 3.6 of [Meckes 2013]. All euclidean spaces have strong negative type; see [Lyons 2013] for a discussion of various proofs.

That real and complex hyperbolic spaces $\mathbb{H}^n$ have negative type was shown in [Gangolli 1967, Section 4], and was made explicit in [Faraut and Harzallah 1974, Corollary 7.4]; that they have strict negative type was shown by Hjorth, Kokkendorff and Markvorsen [Hjorth et al. 2002]. Lyons [2014] showed that real hyperbolic spaces have strong negative type. The remaining constant-curvature, simply connected spaces are spheres.

Let S^n denote the unit-radius sphere centered at the origin of $\mathbb{R}^{n+1}$. Although spheres have negative type (in their intrinsic metric), not even circles have strict negative type. For example, in S^1, take two red points $\{(1, 0), (-1, 0)\}$ and two blue points $\{(0, 1), (0, -1)\}$. Nevertheless, antipodal symmetry is the only obstruction to strict negative type: the main result, Theorem 9.1, of [Hjorth et al. 1998] is that a subset of a sphere has strict negative type if and only if that subset contains at most one pair of antipodal points. We strengthen this to strong negative type:

Theorem 1.1. *If $B \subset S^n$ contains at most one pair of antipodal points, then B has strong negative type.*

We begin by proving a special case:

Theorem 1.2. *If $H \subset S^n$ is an open hemisphere, then H has strong negative type.*

We may parametrize open hemispheres as

$$H_t := \{x \in S^n : t \cdot x > 0\}$$

for $t \in S^n$. A crucial ingredient in the proof of Theorem 1.2 is an analogue of the Cramér–Wold theorem:

Theorem 1.3. *Let H be an open hemisphere in an n-dimensional sphere, S^n. For a finite signed measure μ on H and $t \in S^n$, define $b_\mu(t) := \mu(H \cap H_t)$. The map $\mu \mapsto b_\mu$ is injective. Moreover, if D is a dense subset of S^n, then $\mu \mapsto b_\mu \restriction D$ is injective.*

Let $R\colon x \mapsto -x$ be the reflection in the origin. If K is a Borel subset of S^n such that K and its image under R partition S^n and such that the interior of K is a hemisphere, then call K a *partitioning hemisphere*. Given a probability measure μ on S^n, let μ_R denote the maximal measure that is invariant under R and such that $\mu_R \le \mu$. Note that if μ is a probability measure on S^n that is invariant under R, then $\mu(K) = 1/2$ for every partitioning hemisphere, K. Therefore, for every probability measure μ on S^n with $\mu_R \ne 0$, there is a probability measure $\nu \ne \mu$ such that $\mu(K) = \nu(K)$ for every partitioning hemisphere, K. Moreover, if $\mu_R \ne 0$ but $\mu_R(A) = 0$ for every $(n-1)$-dimensional great sphere A in S^n, then there is another probability measure $\nu \ne \mu$ such that $\mu(H) = \nu(H)$ for every open hemisphere, H. We extend Theorem 1.3 to show the converse, which we use to prove Theorem 1.1:

Theorem 1.4. *Let μ be a probability measure on S^n such that $\mu_R = 0$. If ν is a probability on S^n such that $\mu(K) = \nu(K)$ for every partitioning hemisphere, K, then $\nu = \mu$. Similarly, if ν is a probability on S^n such that $\mu(H_t) = \nu(H_t)$ for every t belonging to a dense subset D of S^n, then $\nu = \mu$.*

The last assertion of Theorem 1.4 is essentially known: see, e.g., Lemmas 2.3 and 2.4 of [Rubin 1999].

We also have the following fact:

Proposition 1.5. *If μ is a probability measure on S^n such that $\mu(K) = 1/2$ for every partitioning hemisphere, K, then μ is R-invariant. Similarly, if μ is a probability measure on S^n such that $\mu(H_t) = \mu(H_{-t})$ for a dense set of t, then μ is R-invariant.*

The last assertion is again essentially known: see [Schneider 1970, Korollar 3.2]. The work of [Rubin 1999] and [Schneider 1970], as well as other authors who study related questions, uses spherical harmonics. This is a powerful tool that leads to more general results, although those extensions do not seem relevant to negative type. We give elementary proofs that rely only on the Cramér–Wold theorem for euclidean spaces:

Theorem 1.6. *If μ is a complex Borel measure on $\mathbb{R}^n$ such that $\mu(H) = 0$ for every open halfspace in $\mathbb{R}^n$, then $\mu = 0$.*

Proof. For $t \in S^{n-1}$, define μ_t on $\mathbb{R}$ by

$$\mu_t(-\infty, a) := \mu\{x \in \mathbb{R}^n : \ t \cdot x < a\} \quad (a \in \mathbb{R}).$$

Then $\mu_t = 0$, whence its Fourier transform $\widehat{\mu}_t$ satisfies $\widehat{\mu}_t(b) = 0$ for all $b \in \mathbb{R}$. Because $\widehat{\mu}_t(b) = \widehat{\mu}(bt)$, it follows that the Fourier transform of μ also vanishes, whence so does μ. □

Even this theorem can be proved without Fourier analysis — see [Walther 1997] or [Lyons and Zumbrun 2018].

2. Proofs

Proof of Theorem 1.3. By the bounded convergence theorem, $b_\mu \restriction D$ determines $b_\mu(t)$ for all t such that $\mu(\partial(H \cap H_t)) = 0$, and therefore $b_\mu \restriction D$ determines all of b_μ by continuity from below: for every $t \in S^n$, there are $s_k \in S^n$ such that $\mu(\partial(H \cap H_{s_k})) = 0$ and $H \cap H_{s_k}$ increase to $H \cap H_t$.

We may take H to be the upper open hemisphere, $\{(t_1, t_2, \dots, t_{n+1}) \in S^n : t_{n+1} > 0\}$. Define $\phi \colon H \to \mathbb{R}^{n+1}$ by

$$\phi(t_1, \dots, t_{n+1}) := (t_1/t_{n+1}, \dots, t_n/t_{n+1}, 1).$$

Then ϕ is a homeomorphism from H to the affine hyperplane

$$H' := \{(t_1, t_2, \dots, t_{n+1}) \in \mathbb{R}^{n+1} : \ t_{n+1} = 1\},$$

namely, $\phi(t)$ is the intersection of H' with the line through the origin and t. Furthermore, ϕ maps $H \cap H_t$ to an open halfspace in H' and every open halfspace in H' is the image under ϕ of some $H \cap H_t$. Therefore, b_μ determines the measures of all open halfspaces with respect to the pushforward $\phi_*\mu$ on H'. The classical theorem of Cramér and Wold applied to H' shows that this determines $\phi_*\mu$, which in turn determines μ. □

Proof of Theorem 1.2. Write σ for the volume measure on S^n normalized to have mass π. Then for all $x_1, x_2 \in S^n$, we have

$$d(x_1, x_2) = \int |\mathbf{1}_{H_t}(x_1) - \mathbf{1}_{H_t}(x_2)|^2 \, \mathrm{d}\sigma(t).$$

This well-known fact is easy to see: By rotation-invariance of σ, the right-hand side depends only on $d(x_1, x_2)$. By considering three points on a great circle, we find that the dependence is linear. Finally, by taking antipodal points, we verify that the constant of linearity is 1.

Therefore, if μ_1 and μ_2 are probabilities on S^n, we may write

$$\int d(x_1, x_2)\, \mathrm{d}(\mu_1 - \mu_2)^2(x_1, x_2) = \iint |\mathbf{1}_{H_t}(x_1) - \mathbf{1}_{H_t}(x_2)|^2\, \mathrm{d}(\mu_1 - \mu_2)^2(x_1, x_2)\, \mathrm{d}\sigma(t).$$

Expanding the square in the integrand and using the facts that

$$\int \mathbf{1}_{H_t}(x)\, \mathrm{d}\nu^2(x, y) = \nu(H_t)\nu(S^n)$$

and

$$\int \mathbf{1}_{H_t}(x)\mathbf{1}_{H_t}(y)\, \mathrm{d}\nu^2(x, y) = \nu(H_t)^2$$

for any finite signed measure, ν, we obtain that

$$\int d(x_1, x_2)\, \mathrm{d}(\mu_1 - \mu_2)^2(x_1, x_2) = -2 \int (\mu_1(H_t) - \mu_2(H_t))^2\, \mathrm{d}\sigma(t).$$

It is evident from this that (S^n, d) has negative type. In order to prove (H, d) has strong negative type, it suffices to show that if μ_1 and μ_2 are concentrated on H and satisfy $\mu_1(H_t) = \mu_2(H_t)$ for σ-a.e. t, then $\mu_1 = \mu_2$. But this is immediate from Theorem 1.3. □

Given any signed measure θ, define the antisymmetric measure $\bar{\theta} := \theta - R_*\theta$, where $R_*\theta$ is the pushforward of θ by R. For positive θ with $\theta_R = 0$, we have $\theta = \bar{\theta}^+$, the positive part of $\bar{\theta}$. For positive θ without assuming that $\theta_R = 0$, we have

$$2\,\theta(S^n) \geq |\bar{\theta}|(S^n), \text{ with equality if and only if } \theta_R = 0. \tag{2-1}$$

Lemma 2.1. *Let μ and ν be probability measures on S^n. If $\mu(K) = \nu(K)$ for every partitioning hemisphere, K, then $\bar{\mu} = \bar{\nu}$. Similarly, if $\mu(H_t) = \nu(H_t)$ for every $t \in D$, where D is a dense subset of S^n, then $\bar{\mu} = \bar{\nu}$.*

Proof. We claim that there is an $(n-1)$-dimensional great sphere A in S^n with $\mu(A) = \nu(A) = 0$. To see this, we build A inductively by dimension. First, because only countably many points have positive mass, there is a pair A_0 of antipodal points with $\mu(A_0) = \nu(A_0) = 0$. Second, all uncountably many 1-dimensional great spheres in S^n that contain A_0 have pairwise intersections exactly A_0, whence there is a 1-dimensional great sphere $A_1 \supset A_0$ with $\mu(A_1) = \nu(A_1) = 0$. We may continue this procedure recursively, finding a k-dimensional great sphere $A_k \supset A_{k-1}$ for $1 \leq k \leq n-1$ with $\mu(A_k) = \nu(A_k) = 0$. Finally, take $A := A_{n-1}$.

Let H be one of the two open hemispheres comprising $S^n \setminus A$. Note that $\mu(H) = \nu(H)$ under either assumption (in the second case, we use a continuity argument like that at the start of the proof of Theorem 1.3).

Let K be a partitioning hemisphere. Because $\mu(A) = 0$ and $\mu(S^n) = 1$, we have

$$\begin{aligned}\bar{\mu}(H \cap K) &= \mu(H \cap K) + \mu(H \cap RK) - \mu(H \cap RK) - \mu(RH \cap RK)\\ &= \mu(H) - \mu(RK) = \mu(H) + \mu(K) - 1.\end{aligned}$$

A similar equation holds for ν. Hence, the assumption that $\mu(K) = \nu(K)$ for every partitioning hemisphere, K, yields

$$\bar{\mu}(H \cap K) = \bar{\nu}(H \cap K)$$

for every such K.

Now every set $H \cap H_t$ is of the form $H \cap K$ for some partitioning hemisphere, K. It follows that

$$\bar{\mu}(H \cap H_t) = \bar{\nu}(H \cap H_t)$$

for every t, whence by Theorem 1.3, it follows that $\bar{\mu} = \bar{\nu}$.

We now prove the second assertion of the lemma. Note that in the preceding proof, we did not use the full strength of the assumption that $\mu(K) = \nu(K)$ for every partitioning hemisphere, K, but only that $\mu(K_t) = \nu(K_t)$ for partitioning hemispheres K_t satisfying $K_t \cap H = H_t \cap H$ for $t \in D$; we may also require that $K_t \cap RH = \overline{H_t} \cap RH$. Let u be such that $H_u = H$, and let s_k be on the geodesic segment from u to t with $s_k \notin \{u, t\}$, $s_k \to t$, and $\mu(\partial H_{s_k}) = \nu(\partial H_{s_k}) = 0$. (Such s_k exist because $\mu(A) = \nu(A) = 0$.) Let $t_{k,j} \in D$ converge to s_k as $j \to \infty$. By the bounded convergence theorem, $\lim_{j\to\infty} \mu(H_{t_{k,j}}) = \mu(H_{s_k})$ and similarly for ν, whence $\mu(H_{s_k}) = \nu(H_{s_k})$. In addition, we have $\lim_{k\to\infty} \mu(H_{s_k}) = \mu(K_t)$ and similarly for ν. Hence, $\mu(K_t) = \nu(K_t)$ for every $t \in D$, whence $\bar{\mu} = \bar{\nu}$. □

Proof of Theorem 1.4. By Lemma 2.1, either assumption implies that $\bar{\mu} = \bar{\nu}$. We may conclude from (2-1) that $2 = 2\,\nu(S^n) \geq |\bar{\nu}|(S^n) = |\bar{\mu}|(S^n) = 2\,\mu(S^n) = 2$, and hence, again from (2-1), that $\nu_R = 0$. Since also $\mu_R = 0$, we obtain the desired conclusion, $\mu = \bar{\mu}^+ = \bar{\nu}^+ = \nu$. □

Proof of Theorem 1.1. If B contains no antipodal points, then every μ concentrated on B has $\mu_R = 0$, whence the proof that B has strong negative type is exactly as for Theorem 1.2, using Theorem 1.4 in place of Theorem 1.3.

If B contains one antipodal pair, $\{x, Rx\}$, then it still suffices to show that for probabilities μ and ν concentrated on B, the assumption $\mu(H_t) = \nu(H_t)$ for a dense set of t implies $\mu = \nu$. By Lemma 2.1, such an assumption yields $\bar{\mu} = \bar{\nu}$. Because $\bar{\mu} = \overline{\mu - \mu_R}$ and $\mu - \mu_R$ is a positive measure with $(\mu - \mu_R)_R = 0$, and similarly for ν, we obtain $\overline{\mu - \mu_R} = \overline{\nu - \nu_R}$ and $\mu - \mu_R = \overline{\mu - \mu_R}^+ = \overline{\nu - \nu_R}^+ = \nu - \nu_R$.

Therefore, $\mu_R(H_t) = \nu_R(H_t)$ for a dense set of t. Because μ_R and ν_R are supported by $\{x, Rx\}$, it follows that $\mu_R = \nu_R$, and so $\mu = \nu$, as desired. □

Proof of Proposition 1.5. For both assertions, we may apply Lemma 2.1 to the pair of measures μ and $R_*\mu$, getting $\bar{\mu} = -\bar{\mu}$, whence $\bar{\mu} = 0$. Thus, $\mu = R_*\mu$, as desired. □

3. Acknowledgement

I thank Marcos Matabuena for asking me about strong negative type for the angular metric on compositional data, i.e., on the probability simplex.

References

[Faraut and Harzallah 1974] J. Faraut and K. Harzallah, "Distances hilbertiennes invariantes sur un espace homogène", *Ann. Inst. Fourier* (*Grenoble*) **24**:3 (1974), xiv, 171–217. MR Zbl

[Gangolli 1967] R. Gangolli, "Positive definite kernels on homogeneous spaces and certain stochastic processes related to Lévy's Brownian motion of several parameters", *Ann. Inst. H. Poincaré Sect. B* (*N.S.*) **3** (1967), 121–226. MR Zbl

[Hjorth et al. 1998] P. Hjorth, P. Lisoněk, S. Markvorsen, and C. Thomassen, "Finite metric spaces of strictly negative type", *Linear Algebra Appl.* **270** (1998), 255–273. MR Zbl

[Hjorth et al. 2002] P. G. Hjorth, S. L. Kokkendorff, and S. Markvorsen, "Hyperbolic spaces are of strictly negative type", *Proc. Amer. Math. Soc.* **130**:1 (2002), 175–181. MR Zbl

[Lyons 2013] R. Lyons, "Distance covariance in metric spaces", *Ann. Probab.* **41**:5 (2013), 3284–3305. MR Zbl

[Lyons 2014] R. Lyons, "Hyperbolic space has strong negative type", *Illinois J. Math.* **58**:4 (2014), 1009–1013. MR Zbl

[Lyons 2018] R. Lyons, "Errata to "Distance covariance in metric spaces" [MR3127883]", *Ann. Probab.* **46**:4 (2018), 2400–2405. MR Zbl

[Lyons and Zumbrun 2018] R. Lyons and K. Zumbrun, "A calculus proof of the Cramér–Wold theorem", *Proc. Amer. Math. Soc.* **146**:3 (2018), 1331–1334. MR Zbl

[Meckes 2013] M. W. Meckes, "Positive definite metric spaces", *Positivity* **17**:3 (2013), 733–757. MR Zbl

[Rubin 1999] B. Rubin, "Inversion and characterization of the hemispherical transform", *J. Anal. Math.* **77** (1999), 105–128. MR Zbl

[Schneider 1970] R. Schneider, "Über eine Integralgleichung in der Theorie der konvexen Körper", *Math. Nachr.* **44** (1970), 55–75. MR Zbl

[Székely and Rizzo 2017] G. J. Székely and M. L. Rizzo, "The energy of data", *Annual Review of Statistics and Its Application* **4**:1 (2017), 447–479.

[Székely et al. 2007] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, "Measuring and testing dependence by correlation of distances", *Ann. Statist.* **35**:6 (2007), 2769–2794. MR Zbl

[Walther 1997] G. Walther, "On a conjecture concerning a theorem of Cramér and Wold", *J. Multivariate Anal.* **63**:2 (1997), 313–319. Addendum, *J. Multivariate Anal.* **67**(2) (1998), 431. MR Zbl

[Zinger et al. 1992] A. A. Zinger, A. V. Kakosyan, and L. B. Klebanov, "A characterization of distributions by mean values of statistics and certain probabilistic metrics", *J. Soviet Math.* **59**:4 (1992), 914–920. Stability problems for stochastic models. Translated from *Problemy Ustoichivosti Stokhasticheskikh Modelei, Trudy Seminara*, 1989, pp. 47–55. MR Zbl

Received May 16, 2019. Revised April 4, 2020.

RUSSELL LYONS
DEPARTMENT OF MATHEMATICS
INDIANA UNIV, BLOOMINGTON
BLOOMINGTON, IN
UNITED STATES
rdlyons@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.391

EXCEPTIONAL GROUPS OF RELATIVE RANK ONE AND GALOIS INVOLUTIONS OF TITS QUADRANGLES

BERNHARD MÜHLHERR AND RICHARD M. WEISS

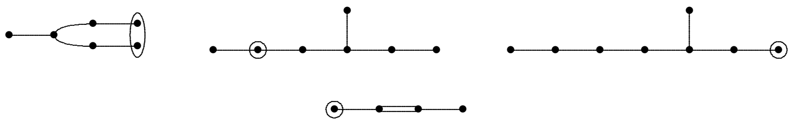
We show that every Moufang set associated with one of the Tits indices ${}^2E^{29}_{6,1}$, $E^{48}_{7,1}$, $E^{91}_{8,1}$ or $F^{21}_{4,1}$ in arbitrary characteristic can be obtained as the fixed point building of a Galois involution acting on a Tits quadrangle parametrized by a quadrangular algebra. This result is used to calculate an explicit formula for the structure map of an arbitrary Moufang set in this class.

MSC2010: 20E42, 51E12, 51E24.
Keywords: building, Moufang set, Tits polygon, exceptional group.

1. Introduction

Let $G(K)$ be the group of rational points of an absolutely simple algebraic group G defined over K and of relative rank 1. Let X be the set of parabolic subgroups P of G defined over K and for each $P \in X$, let U_P be the group of K-rational points of the unipotent radical of P. For each $P \in X$, the group U_P acts sharply transitively on $X \backslash \{P\}$. Thus the triple $(G(K), X, \{U_P\}_{P \in X})$ is a *Moufang set* as defined in Definition 4.1.

Let $\mathcal{T}_1$ denote the following set of Tits indices:



where the circled vertex in the last case corresponds to a short root. These are the Tits indices called ${}^2E_{6,1}^{29}$, $E_{7,1}^{48}$, $E_{8,1}^{91}$ and $F_{4,1}^{21}$ in [Tits 1966a]. The groups $G(K)$ that correspond to the Tits indices in $\mathcal{T}_1$ are precisely the groups of exceptional absolute type and relative rank 1 such that for each parabolic subgroup P defined over K, the unipotent radical of P is nonabelian but has a center of dimension greater than 1. We denote the class of Moufang sets associated with these groups by $\mathcal{M}$.

In this paper, we use the theory of *Tits polygons* introduced in [Mühlherr and Weiss ≥ 2020] to investigate the structure of the Moufang sets in $\mathcal{M}$. A Tits polygon is a bipartite graph Γ such that for each vertex v, the set of vertices adjacent to v is endowed with an "opposition relation" subject to certain axioms; see Section 3 for details. Moufang polygons (which were classified in [Tits and Weiss 2002]) are precisely the Tits polygons in which the opposition relations are all trivial.

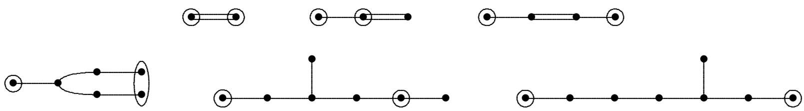
Let T be a spherical Tits index of relative rank 2. The *relative Coxeter group* associated with T (as defined in [Mühlherr et al. 2015, 20.32 and 20.34]; cf. [Tits 1966a, 2.5.2]) is a dihedral group of order $2n$ for $n = 3, 4, 6$ or 8. We call n the *polygon type* of T.

We say that a Tits index T is *linked* to a building Δ if the *absolute type of* T is also the Coxeter diagram of Δ.

Suppose that Δ is a spherical building of type (W, S) and that T is a Tits index of relative rank 2 linked to Δ. Let V be the set of residues of Δ whose type is the complement in S of either one of the two circles in T and let Γ be the bipartite graph with vertex set V, where two residues are adjacent whenever they have a nonempty intersection. If Δ is assumed to be Moufang as defined in Definition 6.5 (which is always the case if Δ is the spherical building associated to the group of rational points of an absolutely simple algebraic group of relative rank at least 2), then the graph Γ is endowed canonically with the structure of a Tits n-gon $X_{\Delta,T}$, where n is the polygon type of T.

We say that a building is *exceptional* if it is the spherical building associated to the group of rational points of an algebraic group of exceptional absolute type and relative rank at least 2. Let $\mathcal{E}$ be the set of pairs (Δ, T), where Δ is an exceptional building and T is a Tits index of relative rank 2 linked to Δ. An *exceptional Tits polygon* is a Tits polygon isomorphic to $X_{\Delta,T}$ for some pair (Δ, T) in $\mathcal{E}$. For each $n = 3$, 4, 6 and 8, let $\mathcal{E}_n$ be the subset consisting of all (Δ, T) in $\mathcal{E}$ such that n is the polygon type of T.

We have $\mathcal{E}_4 = \{(\Delta, T) \in \mathcal{E} \mid T \in \mathcal{T}_2\}$, where $\mathcal{T}_2$ denotes the following set of Tits indices:



Note that the Tits index in $\mathcal{T}_2$ of absolute type F_4 does not appear in [Tits 1966a]; it is, however, a Tits index as defined in [Mühlherr et al. 2015, 20.1]. Let

$$\mathcal{Q} = \{X_{\Delta,T} \mid (\Delta, T) \in \mathcal{E}_4\}.$$

Quadrangular algebras are a class of algebras that arose in the classification of Moufang quadrangles in [Tits and Weiss 2002]. They were classified in [Mühlherr and Weiss 2019; Weiss 2006b]. In [Mühlherr and Weiss 2020], we showed that the Tits quadrangles in $\mathcal{Q}$ are classified by quadrangular algebras. More precisely, each Tits quadrangle in $\mathcal{Q}$ can be coordinatized by commutator relations parameterized by the corresponding quadrangular algebra and theses parametrizations give rise to a one-to-one correspondence between the Tits quadrangles in $\mathcal{Q}$ (up to isomorphism) and nondegenerate quadrangular algebras whose associated quadratic form is of dimension at least 5 (up to isotopy).

In Theorem 8.14, the main result of this paper, we show that every Moufang set in the class $\mathcal{M}$ defined above arises as the fixed point building of a Galois involution acting on a Tits quadrangle in $\mathcal{Q}$. We then use the corresponding quadrangular algebra to calculate explicitly the structure of a root group $\mathcal{U}$ and the *structure map* τ (as defined in Definition 10.8) of the fixed point building of a Galois involution acting on a Tits quadrangle in $\mathcal{Q}$ (modulo Conjecture 19.14). The product of these calculations can be found in Notation 9.18, (9.21) and Theorem 19.7.

Our result and our calculations are valid over arbitrary fields. In particular, we do not make any restriction on the characteristic of K.

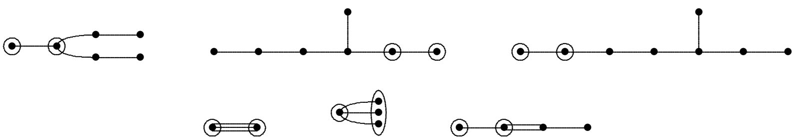
In [Boelaert et al. 2019, 2.3.5 and 6.5.3], the structure maps for the Moufang sets in $\mathcal{M}$ were described in terms of the structurable division algebras attached to forms of tensor products of two composition algebras. The theory of structurable algebras requires, however, that $\text{char}(K) \neq 2$ or 3. In [De Medts and Van Maldeghem 2010, 2.1; Callens and De Medts 2014, 5.3], the structure maps for the Moufang sets associated with the Tits indices $F_{4,1}^{21}$ and ${}^2E_{6,1}^{29}$ over an arbitrary field are given in terms of an octonion division algebras by exploiting a connection to the corresponding projective plane, but this method does not extend to the cases $E_{7,1}^{48}$ and $E_{8,1}^{91}$.

Let $\hat{\mathcal{T}}_1$ denote the following set of Tits indices:

These are the Tits indices called ${}^2E_{6,1}^{35}$, $E_{7,1}^{66}$, $E_{8,1}^{133}$ and ${}^{3,6}D_{4,1}^{9}$ in [Tits 1966a]. Let $\hat{\mathcal{M}}$ denote the class of Moufang sets associated with the Tits indices in $\hat{\mathcal{T}}_1$. These are the Moufang sets arising from an exceptional group of relative rank 1 such that the unipotent radical of a parabolic subgroup defined over K has a 1-dimensional center. We have

$$\{(\Delta, T) \in \mathcal{E} \mid T \in \hat{\mathcal{T}}_2\} \subset \mathcal{E}_6,$$

where $\hat{\mathcal{T}}_2$ denotes the following set of Tits indices:

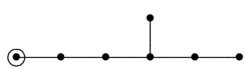


and the circled vertices in the last case corresponds to long roots. Note that the Tits index in $\hat{\mathcal{T}}_2$ of absolute type F_4 does not appear in [Tits 1966a]; it is, however, a Tits index as defined in [Mühlherr et al. 2015, 20.1]. See also [Mühlherr and Weiss $\geq$ 2020, 2.4.6]. Let

$$\mathcal{H} = \{X_{\Delta,T} \in \mathcal{E}_6 \mid T \in \hat{\mathcal{T}}_2\}.$$

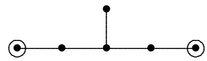
In [Mühlherr and Weiss $\geq$ 2020, 3.8.13], we showed that every Moufang set in $\hat{\mathcal{M}}$ arises as the fixed point building of a Galois involution acting on a Tits hexagon in $\mathcal{H}$. As in the previous case, we then calculated $\mathcal{U}$ and τ for an arbitrary fixed point building of this kind. In this case, the role of quadrangular algebras was played by *cubic norm structures*; see [Mühlherr and Weiss $\geq$ 2020, 3.4.7 and 3.7.1]. The class $\hat{\mathcal{M}}$ is also treated in [Boelaert et al. 2019], but again with the restriction that the characteristic of K is not 2 or 3.

The only exceptional Tits index of relative rank 1 in [Tits 1966a] that is not in $\mathcal{T}_1$ or $\hat{\mathcal{T}}_1$ is $E_{7,1}^{78}$:

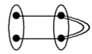


The Moufang sets associated with this index have abelian root groups. They are classified by Albert division algebras and the structure map of one of them is the simply the inverse map of the corresponding Albert algebra. See also Remark 6.14.

We mention, too, that the set $\mathcal{E}_3$ consists of the pairs (Δ, T) in $\mathcal{E}$ such that T is the Tits index



and the set $\mathcal{E}_8$ consists of the pairs (Δ, T) in $\mathcal{E}$ such that T is



(a Tits index in the sense of [Mühlherr et al. 2015, 20.1]).

Organization 1.1. This paper is organized as follows. In Sections 2–5 we introduce quadrangular algebras, Tits polygons, Moufang sets and Tits indices and describe the construction of the Tits polygon $X_{\Delta,T}$. In Section 6 we assemble material about descent, Galois groups and the construction of Moufang sets as fixed point buildings. In Section 7 we describe the connection between the exceptional quadrangles and quadrangular algebras and we prove our main theorem about Moufang sets in Theorem 8.14. This is the result that says that every Moufang set in $\mathcal{M}$ can be found as the fixed point building of a Galois involution of an exceptional Tits quadrangle.

In Section 9 we introduce an arbitrary Galois involution ω of an exceptional Tits quadrangle Δ and choose suitable coordinates for Δ. In Section 10 we introduce the structure map of the Moufang set $\Delta^{\langle\omega\rangle}$ associated with ω. In Sections 11–19 we calculate. These calculations result in the formula for τ in Proposition 19.6 and Theorem 19.7; see also Conjecture 19.14 and Proposition 19.15.

Conventions 1.2. Let G be a group. As in [Tits and Weiss 2002], we set $a^b = b^{-1}ab$ and

$$[a, b] = a^{-1}b^{-1}ab$$

for all $a, b \in G$. We compose permutations from left to right. Other functions will be written on the left and composed from right to left.

2. Quadrangular algebras

The notion of an anisotropic quadrangular algebra was introduced in [Weiss 2006b] and the following more general notion was introduced in [Mühlherr and Weiss 2019].

Definition 2.1. A *quadrangular algebra* is an ordered set

$$\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta),$$

where K is a field, L is a vector space over K, q is a nondegenerate quadratic form on L, f is the bilinear form associated with q, ε is an element of L such that $q(\varepsilon) = 1$, $\mathcal{X}$ is a nontrivial vector space over K, $(a, v) \mapsto a \cdot v$ is a map from $\mathcal{X} \times L$

to $\mathcal{X}$ (usually denoted simply by juxtaposition), h is a map from $\mathcal{X} \times \mathcal{X}$ to L and θ a map from $\mathcal{X} \times L$ to L satisfying the following axioms:

(A1) The map $\cdot$ is bilinear (over K).

(A2) $a \cdot \varepsilon = a$ for all $a \in \mathcal{X}$.

(A3) $(av)\bar{v} = q(v)a$ for all $a \in \mathcal{X}$ and all $v \in L$, where $\bar{v} = f(v, \varepsilon)\varepsilon - v$.

(B1) h is bilinear (over K).

(B2) $h(a, bv) = h(b, av) + f(h(a, b), \varepsilon)v$ for all $a, b \in \mathcal{X}$ and all $v \in L$.

(B3) $f(h(av, b), \varepsilon) = f(h(a, b), v)$ for all $a, b \in \mathcal{X}$ and all $v \in L$.

(C1) For each $a \in \mathcal{X}$, the map $v \mapsto \theta(a, v)$ is linear (over K).

(C2) $\theta(ta, v) = t^2\theta(a, v)$ for all $t \in K$, all $a \in \mathcal{X}$ and all $v \in L$.

(C3) There exists a function g from $\mathcal{X} \times \mathcal{X}$ to K such that

$$\theta(a+b, v) = \theta(a, v) + \theta(b, v) + h(a, bv) - g(a, b)v$$

for all $a, b \in \mathcal{X}$ and all $v \in L$.

(C4) There exists a function ϕ from $\mathcal{X} \times L$ to K such that

$$\theta(av, w) = q(v)\overline{\theta(a, \bar{w})} - f(w, \bar{v})\overline{\theta(a, v)} + f(\theta(a, v), \bar{w})\bar{v} + \phi(a, v)w$$

for all $a \in \mathcal{X}$ and $v, w \in L$, where $\bar{u}$ for $u \in L$ is as in (A3).

(D1) Let $\pi(a) = \theta(a, \varepsilon)$ for all $a \in \mathcal{X}$. Then

$$a\theta(a, v) = (a\pi(a))v$$

for all $a \in \mathcal{X}$ and all $v \in L$.

Notation 2.2. Let $\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta)$ be a quadrangular algebra. We say that Ξ is a quadrangular algebra *over* K if we want to specify the field K. The map h is *nondegenerate* if for each nonzero $a \in \mathcal{X}$ there exists $b \in \mathcal{X}$ such that $h(a, b) \neq 0$. We say that Ξ is δ-*standard* for some $\delta \in L$ if $f(\pi(a), \delta) = 0$ for all $a \in \mathcal{X}$ and either $\operatorname{char}(K) \neq 2$ and $\delta = \frac{1}{2}\varepsilon$ or $\operatorname{char}(K) = 2$, $f(\varepsilon, \delta) = 1$ and $q(\delta) \neq 0$. We say that Ξ is *anisotropic* if the quadratic form q is anisotropic *and*

$$\pi(a) = t\varepsilon \text{ for some } t \in K \text{ only if } a = 0,$$

where π is as in (D1). If either of these conditions fails to hold, we say that Ξ is *isotropic*.

The classification of δ-standard nondegenerate quadrangular algebras is given in [Mühlherr and Weiss 2019]. This classification extended the classification of anisotropic quadrangular algebras given in [Weiss 2006b]. We will say more about this in Remark 7.10

Notation 2.3. Let $\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta)$ be a quadrangular algebra. We set

$$Q(a) = f(\pi(a), \varepsilon)$$

for all $a \in \mathcal{X}$.

Almost all of the results in [Weiss 2006b, Chapters 3 and 4] (where it is assumed that Ξ is anisotropic) remain valid under the assumptions that Ξ is δ-standard for some $\delta \in L$ and $|K| > 2$. This is shown in [Mühlherr and Weiss 2019, Section 7, see, in particular, Conclusion 7.5]; We note that the hypothesis $|K| > 2$ is required for the result [Weiss 2006b, 3.22].

We will assume that the reader has access to [Weiss 2006b, Chapters 3 and 4] and will frequently cite results from these two chapters (keeping the caveats in [Mühlherr and Weiss 2019, Conclusion 7.5] in mind). Note that the map $v \mapsto \bar{v}$ in (A3) is called σ in [Weiss 2006b]. We refer to reader also to the remarks (i)–(viii) on page 7 of [Weiss 2006b], especially the remark (iv).

If $a \in \mathcal{X}$ and $u, v \in L$, we will write auv in place of $(au)v$. Since we are not assuming the existence of a multiplication on L, this convention should not cause any confusion.

By [Weiss 2006b, 4.1 and 4.5(iii)], the functions ϕ in (C4) and Q in Notation 2.2 are both identically zero if $\mathrm{char}(K) \neq 2$. We draw the reader's attention also to [Weiss 2006b, 1.4 and 3.6–3.8]. We will make particularly frequent use of these identities.

We have assembled a few additional identities, which are not in [Weiss 2006b], in the Appendix; see, in particular, Remark A.5.

We close this section with a definition which will not be used until Theorem 7.4.

Notation 2.4. Let $\mathcal{R}_\Xi$ denote the set $\mathcal{X} \times K$ endowed with the multiplication given by

$$(a, t) \cdot (b, r) = \big(a + b, t + r + g(b, a)\big)$$

for all $(a, t), (b, r)$. By [Weiss 2006b, 4.4], g is bilinear. It follows that $\mathcal{R}_\Xi$ is a group in which

$$(a, t)^{-1} = \big(-a, -t + g(a, a)\big)$$

for all (a, t).

3. Tits polygons

The notion of a Tits polygon was introduced in [Mühlherr and Weiss $\geq$ 2020]. In this section we give the definition and assemble just a few of their basic properties.

Definition 3.1. A *dewolla* is a triple

$$X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v \in V}),$$

where:

(i) Γ is a bipartite graph with vertex set V and $|\Gamma_v| \geq 3$ for each $v \in V$, where Γ_v denotes the set of vertices adjacent to v.

(ii) For each $v \in V$, $\equiv_v$ is an antireflexive symmetric relation on Γ_v. We say that vertices $u, w \in V$ are *opposite at* v if $u, w \in \Gamma_v$ and $u \equiv_v w$ and a path $(w_0, w_1, \ldots, w_m)$ in Γ is called *straight* if w_{i-1} and w_{i+1} are opposite at w_i for all $i \in [1, m-1]$.

(iii) $\mathcal{A}$ is a set of connected subgraphs γ containing $2n$ vertices for some $n \geq 3$ such that for each vertex v on γ, Γ_v contains exactly two vertices in γ and these two vertices are opposite at v.

The parameter n is called the *level* of X. The automorphism group $\mathrm{Aut}(X)$ is the subgroup of $\mathrm{Aut}(\Gamma)$ consisting of all $g \in \mathrm{Aut}(\Gamma)$ such that $\gamma^g \in \mathcal{A}$ for all $\gamma \in \mathcal{A}$ and for all $u, v, w \in V$ such that u and w are opposite at v, u^g and w^g are opposite at v^g. A *root* of X is a straight path of length n.

Definition 3.2. A *Tits n-gon* is a dewolla

$$X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v \in V})$$

of level n for some $n \geq 3$ such that Γ is connected and the following axioms hold:

(i) For all $v \in V$ and all $u, w \in \Gamma_v$, there exists $z \in \Gamma_v$ that is opposite both u and w at v.

(ii) For each straight path $\delta = (w_0, \ldots, w_k)$ of length $k \leq n-1$, δ is the unique straight path of length at most k from w_0 to w_k.

(iii) For $G = \mathrm{Aut}(X)$ and for each root $\alpha = (w_0, \ldots, w_n)$ of X, the group U_α acts transitively on the set of vertices opposite w_{n-1} at w_n, where U_α is the pointwise stabilizer of

$$\Gamma_{w_1} \cup \Gamma_{w_2} \cup \cdots \cup \Gamma_{w_{n-1}}$$

in $\mathrm{Aut}(X)$. The group U_α is called the *root group* associated with the root α.

A *Tits polygon* is a Tits n-gon for some $n \geq 3$. A Tits n-gon is called a *Tits triangle* if $n = 3$, a *Tits quadrangle* if $n = 4$, etc.

Definition 3.3. A *Moufang n-gon* is a Tits n-gon $X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v \in V})$ for some $n \geq 3$ in which for all vertices v, the relation $\equiv_v$ is trivial, i.e., every two vertices adjacent to v are opposite at v. In this case, $\mathcal{A}$ is the set of all circuits of Γ of length $2n$, so X is uniquely determined by Γ alone (and it is usual to refer to the Moufang n-gon as Γ rather than X). See [Mühlherr and Weiss $\geq$ 2020, 1.2.2 and 1.2.3]. Moufang polygons were classified in [Tits and Weiss 2002, 17.1–17.8].

Notation 3.4. Let $X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v\in V})$ be a Tits n-gon. A *coordinate system* for X is a pair $(\gamma, i \mapsto w_i)$ where γ is an element of $\mathcal{A}$ and $i \mapsto w_i$ is a surjection from $\mathbb{Z}$ to the vertex set of γ such that w_{i-1} is adjacent to w_i for each i. For each coordinate system $(\gamma, i \mapsto w_i)$, we denote by U_i the root group associated with the root $(w_i, w_{i+1}, \dots, w_{i+n})$ for each $i \in \mathbb{Z}$ and call the map $i \mapsto U_i$ the associated *root group labeling*. Thus $w_i = w_j$ and $U_i = U_j$ whenever i and j have the same image in $\mathbb{Z}_{2n}$. If $i \mapsto w_i$ is a coordinate system, then so is $i \mapsto w_{n+1-i}$. These two coordinate systems are called *opposite*.

For the rest of this section, we fix a Tits n-gon $X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v\in V})$ and a coordinate system $(\gamma, i \mapsto w_i)$ of X and let $i \mapsto U_i$ be the corresponding root group labeling.

Proposition 3.5. *The following hold*:

(i) $[U_i, U_j] \subset U_{[i+1,j-1]}$ *for all* i, j *such that* $i < j < i+n$, *where*

$$U_{[k,m]} = \begin{cases} U_k U_{k+1} \cdots U_m & \text{if } k \le m, \text{ and} \\ 1 & \text{otherwise.} \end{cases}$$

In particular, $[U_i, U_{i+1}] = 1$ *for all* i.

(ii) *The product map* $U_1 \times U_2 \times \cdots \times U_n \to U_{[1,n]}$ *is bijective.*

Proof. This holds by [Mühlherr and Weiss ≥ 2020, 1.3.38]. □

Proposition 3.6. *For each* i, U_i *acts sharply transitively on the set of vertices that are opposite* w_{i+1} *at* w_i *and on the set of vertices that are opposite* w_{i+n-1} *at* w_{i+n}.

Proof. This holds by [Mühlherr and Weiss ≥ 2020, 1.3.25]. □

Notation 3.7. Let

$$U_i^\sharp = \{a \in U_i \mid w_{i+n+1}^a \text{ is opposite } w_{i+n+1} \text{ at } w_{i+n}\}$$

for each i. By [Mühlherr and Weiss ≥ 2020, 1.4.8], we have

$$U_i^\sharp = \{a \in U_i \mid w_{i-1}^a \text{ is opposite } w_{i-1} \text{ at } w_i\}$$

for each i.

Proposition 3.8. *For each* $i \in \mathbb{Z}$, *there exist unique maps* κ_γ *and* λ_γ *from* $U_i^\sharp$ *to* $U_{i+n}^\sharp$ *such that for each* $a \in U_i^\sharp$, *the product*

$$\mu_\gamma(a) := \kappa_\gamma(a) \cdot a \cdot \lambda_\gamma(a) \tag{3.9}$$

interchanges the vertices w_{i+n-1} *and* w_{i+n+1}. *For each* $a \in U_i^\sharp$, *the element* $\mu_\gamma(a)$ *fixes the vertices* w_i *and* w_{i+n} *and interchanges the vertices* w_{i+j} *and* w_{i-j} *for all* $j \in \mathbb{Z}$ *and*

$$U_k^{\mu_\gamma(a)} = U_{2i+n-k} \tag{3.10}$$

for all $k \in \mathbb{Z}$.

Proof. This holds by [Mühlherr and Weiss ≥ 2020, 1.4.4]. □

4. Moufang sets

The notion of a Moufang set was introduced in [Tits 1992].

Definition 4.1. A *Moufang set* is a triple $M = (G, X, \{U_x\}_{x\in X})$, where X is a set such that $|X| \geq 3$, G is a group acting on X and $\{U_x\}_{x\in X}$ is a conjugacy class of subgroups of G such that for each $x \in X$, U_x fixes x and acts sharply transitively on $X\backslash\{x\}$. The groups U_x for $x \in X$ are called the *root groups* of the Moufang set. We call X the *underlying set* of M. The action of G on X is not necessarily faithful; see Definition 4.8 below. If in an example X and the root groups of M are clear from the context, we sometimes refer to G itself as a Moufang set.

Example 4.2. The group $G = \mathrm{PGL}_2(D)$ is a Moufang set for each skew field D. In this example, the underlying set is the projective line $\mathrm{PG}(D)$, i.e., the set of 1-dimensional subspaces of a 2-dimensional right vector space V over D. Choose a basis v_1, v_2 of V and use it to identify V with $D \oplus D$ and G with the group of 2×2 invertible matrices over D acting on V on the left. Then $U_{\langle v_1\rangle}$ is the subgroup of G consisting of all matrices with 1s on the diagonal and 0 below the diagonal.

Example 4.3. Let Δ be a Moufang quadrangle or hexagon. If ρ is a polarity of Δ (i.e., a non-type-preserving involution), then the set of edges of Δ fixed by ρ has the structure of a Moufang set. See [De Medts et al. 2017; Tits 1966b] for explicit examples.

Example 4.4. Let Γ be a Moufang polygon, let $D = \mathrm{Aut}(\Gamma)$ and let v be a vertex of Γ. Let X be the set Γ_v of vertices adjacent to v, let G be subgroup of $\mathrm{Sym}(X)$ induced by the stabilizer D_v and for each $x \in X$, let U_x be the subgroup of G induced by the root group U_α, where α is a path $(z_0, z_1, \ldots, z_n)$ of length n in Γ such that $z_0 = v$ and $z_1 = x$. By [Tits and Weiss 2002, 4.6 and 4.8], the triple $(G, X, \{U_x\}_{x\in X})$ is a Moufang set.

Example 4.5. Let $G(K)$ be the group of rational points of an absolutely simple algebraic group G defined over K and of relative rank 1. Let X be the set of parabolic subgroups of G defined over K and for each P in X, let U_P be the group of K-rational points of the unipotent radical of P. As already observed in the introduction, the triple $(G(K), X, \{U_P\}_{P\in X})$ is a Moufang set.

A fifth source of Moufang sets will be described in Theorem 6.11(v) below. See [De Medts and Segev 2009] for more information about Moufang sets.

Remark 4.6. A building of rank 1 is simply a set of chambers without any additional structure. See, for example, [Weiss 2003, 1.4 and 7.1]. The buildings of rank 1 that arise "in nature", however, always arise in conjunction with a permutation group with respect to which the building has the structure of a Moufang set.

Definition 4.7. Let $M = (G, X, \{U_x\}_{x\in X})$ and $M' = (G', X, \{U'_x\}_{x\in X'})$ be two Moufang sets. We call a bijection π from X to X' such that $\pi U_x \pi^{-1} = U'_{\pi(x)}$ for all $x \in X$ an *equivalence map* from M to M'. An equivalence map that carries G to G' is called an *isomorphism*. We say that M and M' are *equivalent* if there exists an equivalence map from M to M'.

Definition 4.8. Let $M = (G, X, \{U_x\}_{x\in X})$ be a Moufang set. We say that M is faithful if G acts faithfully on X. Let $G^\dagger = \langle U_x \mid x \in X\rangle$. Then $G^\dagger$ acts transitively on X, hence $\{U_x\}_{x\in X}$ is a single conjugacy class of $G^\dagger$ and thus $(G^\dagger, X, \{U_x\}_{x\in X})$ is also a Moufang set.

Observation 4.9. Every Moufang set M is equivalent to a unique faithful Moufang set $M_0 = (G, X, \{U_x\}_{x\in X})$ such that $G = G^\dagger$.

5. Tits indices

In this section, we explain how Tits polygons arise "in nature."

Notation 5.1. Let Π be a Coxeter diagram with vertex set S and let (W, S) denote the corresponding Coxeter system. For each subset $J \subset S$, we denote by W_J the subgroup of W generated by J. A subset J is *spherical* if W_J is finite. If J is a spherical subset of S, we denote by w_J the longest element in the Coxeter system (W_J, J). For each spherical subset J of S, $s \mapsto w_J s w_J$ is an automorphism of the subdiagram Π_J spanned by J; see, for example, [Mühlherr et al. 2015, 19.6].

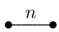
Definition 5.2. A *Tits index* is a triple

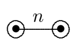
$$T = (\Pi, \Theta, A),$$

where Π is a Coxeter diagram with vertex set S, Θ is a subgroup of $\mathrm{Aut}(\Pi)$ and A is a Θ-invariant subset of S such that for each $s \in S\backslash A$, the subset $J_s := A \cup s^\Theta$ is spherical, and A is stabilized by the automorphism $s \mapsto w_{J_s} s w_{J_s}$ of Π_{J_s}, where s^Θ is the Θ-orbit containing s. The Coxeter diagram Π is called the *absolute type* of T and $|S|$ is called the *absolute rank* of T. For each $s \in S\backslash A$, we denote by $\tilde{s}$ the product $w_A w_{J_s}$. There is one element $\tilde{s}$ for each Θ-orbit disjoint from A. Let $\tilde{S}$ denote the set of all these elements and let $\widetilde{W} = \langle \tilde{S}\rangle$. By [Mühlherr et al. 2015, 20.4, 20.32 and 20.34], $(\widetilde{W}, \tilde{S})$ is also a Coxeter system. It is called the *relative Coxeter system* of T and $|\tilde{S}|$ is called the *relative rank* of T. The *relative type* $\widetilde{\Pi}$ of T is the Coxeter diagram corresponding to its relative Coxeter system.

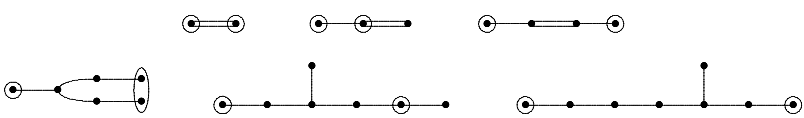
Definition 5.3. A Tits index (Π, Θ, A) is *split* if Θ is trivial and $A = \varnothing$.

Basic Construction 5.4. Let $T = (\Pi, \Theta, A)$ be a Tits index as in Definition 5.2 whose absolute type Π is an irreducible spherical Coxeter diagram and whose relative rank is 2. Thus $\widetilde{W}$ is a dihedral group, where $\widetilde{W}$ is as in Definition 5.2. Let $n = |\widetilde{W}|/2$ and let $J_i = A \cup s_i^\Theta$ for $i = 1$ and 2, where s_1 and s_2 are representatives of the two Θ-orbits disjoint from A. By [Mühlherr et al. 2015, 20.39–20.40], $n \geq 3$. Let Δ be a building of type Π. We assume that Δ is Moufang as defined in [Mühlherr and Weiss 2017, 2.7]. (This is automatic if the rank of T is greater than 2.) Let V_i be the set of J_i-residues of Δ for $i = 1$ and 2 and let Γ be the bipartite graph whose vertex set is $V := V_1 \cup V_2$, where a residue in V_1 is adjacent to a residue in V_2 whenever their intersection is an A-residue of Δ. We declare that two residues $u = R_1$ and $w = R_2$ in V are opposite at a residue $v = R_0$ in V (and we write $u \equiv_v w$) if u and w are adjacent to v and $R_0 \cap R_1$ and $R_0 \cap R_2$ are opposite residues of R_0. Let $\mathcal{A}$ be the set of $2n$ circuits γ in Γ such that for some apartment Σ of Δ, every vertex of γ contains chambers of Σ and for every vertex v of γ, the two vertices of γ adjacent to v are opposite at v. By [Mühlherr and Weiss $\geq$ 2020, 1.2.12 and 1.2.28(i)], the triple $(\Gamma, \mathcal{A}, \{\equiv_v\}_{v \in V})$ is a Tits polygon. We denote this Tits polygon by $X_{\Delta,T}$ and we call the Tits polygons that arise in this way Tits polygons *of index type*.

Remark 5.5. Let Δ be a Moufang n-gon for some $n \geq 3$. In other words, Δ is a building of type  satisfying the Moufang condition defined in [Weiss 2003, 11.2]. By [Mühlherr and Weiss $\geq$ 2020, 1.2.2], Δ is naturally endowed with the structure of a Tits n-gon and, in fact, we have $\Delta = X_{\Delta,T}$ as defined in Basic Construction 5.4, where T is the Tits index



Remark 5.6. Tits indices are often represented figures drawn according to the conventions described in [Mühlherr et al. 2015, 34.2]. Drawn according to these conventions, the Tits indices of relative type B_2 (which is the same as relative type C_2) whose absolute type is an exceptional Coxeter diagram are



The indices of absolute type C_2 and C_3 are included here because of the existence of the exceptional Moufang quadrangles (which are buildings of type C_2 whose automorphism groups are forms of groups of type E_6, E_7, E_8 and F_4) and because of the existence of a family of buildings of type C_3 associated with a form of E_7. These are the buildings that appear in Definition 7.1(i) and (ii) below.

The following result should have been included in [Mühlherr and Weiss $\geq$ 2020].

Theorem 5.7. *Let Δ, $T = (\Pi, \Theta, A)$ and $X_{\Delta,T} = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v \in V})$ be as in Basic Construction 5.4. Then every automorphism of X has a unique extension to an automorphism of Δ.*

Proof. Let $X = X_{\Delta,T}$ and let J_1, J_2 and R be as in Basic Construction 5.4. Thus $A = J_1 \cap J_2$ and R is an A-residue. Let G be stabilizer of the set $\{J_1, J_2\}$ in $\mathrm{Aut}(\Delta)$, let $G_0 = \mathrm{Aut}(X)$, let $G^\dagger$ denote the subgroup of G generated by all the root groups of Δ, let $G_0^\dagger$ denote the subgroup of G_0 generated by all the root groups of X as defined in Definition 3.2(iii) and let χ denote the restriction map from G to G_0. By [Mühlherr and Weiss $\geq$ 2020, 1.2.10], χ is injective. It will suffice, therefore, to show that χ is surjective.

Let $\mathcal{P}_i$ be the set stabilizers in $G^\dagger$ of all the J_i-residues of Δ for $i = 1$ and 2. Let N denote stabilizer of the set $\{\mathcal{P}_1, \mathcal{P}_2\}$ in $\mathrm{Aut}(G^\dagger)$. Since $G^\dagger$ is normal in $\mathrm{Aut}(\Delta)$, there is a canonical map from $\mathrm{Aut}(\Delta)$ to $\mathrm{Aut}(G^\dagger)$. By [Tits 1974, 5.8–5.10, 6.6, 6.13, 8.6, 9.3 and 10.4] (i.e., the solution to "Problem B" as described in [Tits 1974, Introduction]), this map is an isomorphism. Its restriction to G is an isomorphism from G to N. By [Mühlherr and Weiss $\geq$ 2020, 1.5.3], $\chi(G^\dagger) = G_0^\dagger$. The group $G_0^\dagger$ is normal in G_0 and G_0 stabilizes the subset $\chi(\{\mathcal{P}_1, \mathcal{P}_2\})$ of conjugacy classes of $G_0^\dagger$. It follows that $G_0 = E \cdot \chi(G)$, where $E = C_{G_0}(G_0^\dagger)$.

Let D denote the stabilizer of the A-residue R in $G^\dagger$, let Y denote the set of A-residues that are fixed by D and let M denote the stabilizer of Y in $G^\dagger$. Then M acts transitively on Y (since $G^\dagger$ acts transitively on the set of A-residues of Δ) and $D \subset M$. By [Weiss 2003, 11.16], it follows that M is the stabilizer in $G^\dagger$ of a residue R' of Δ containing R. Suppose $|Y| > 1$. Then $R \neq R'$. Hence we can choose j-adjacent chambers $c \in R$ and $d \in R' \backslash R$ for some $j \notin A$. Let Σ be an apartment containing c and d and let α be the unique root of Σ containing c but not d. Choose $g \in U_\alpha^*$ and let R_0 be the A-residue containing d. Then $g \in D$. By [Weiss 2003, 11.4], d and d^g are unequal but j-adjacent. Thus by [Weiss 2003, 7.21], $R_0^g \neq R_0$. Thus the residues R_0 and R_0^g both lie in R' and hence are in the same M-orbit as R. Thus R_0 and R_0^g both lie in Y. This contradicts the fact that D acts trivially on Y. We conclude that $|Y| = 1$. Thus $\chi(D)$ fixes only one edge of Γ. We have $\chi(D) = \chi(D)^g$ for all $g \in E$. Since $E \subset G_0$, the elements of E map edges to edges. It follows that E fixes the unique edge fixed by $\chi(D)$. Since $G_0^\dagger$ acts transitively on the edge set Γ (by [Mühlherr and Weiss $\geq$ 2020, 1.3.6 and 1.3.40]), it follows that E fixes every edge of Γ. Hence $E = 1$. Thus $G_0 = \chi(G)$. □

6. Descent

In this section we assemble the results in [Mühlherr et al. 2015] on descent in buildings that we will require.

Definition 6.1. Let Δ be a building and let G_0 be a subgroup of $\mathrm{Aut}(\Delta)$. A G_0-*residue* is a residue of Δ stabilized by G_0. A G_0-*chamber* is a G_0-residue which is minimal with respect to inclusion. A G_0-*panel* is a G_0-residue P such that for some G_0-chamber C, P is minimal in the set of all G_0-residues containing C properly.

Definition 6.2. Let Δ and G_0 be as in Definition 6.1. The group G_0 is *anisotropic* if Δ itself is the unique G_0-chamber and *isotropic* if this is not the case. Thus G_0 is isotropic if and only if there exist G_0-panels (equivalently, if there exist G_0-residues other than Δ itself).

Notation 6.3. Let Δ be a building and let G_0 be an isotropic subgroup of $\mathrm{Aut}(\Delta)$. We denote by Δ^{G_0} the graph with vertex set the set of all G_0-chambers, where two G_0-chambers are joined by an edge of Δ^{G_0} if and only if there is a G_0-panel containing them both.

Definition 6.4. Let Δ be a building. A *descent group* of Δ is an isotropic subgroup G_0 of $\mathrm{Aut}(\Delta)$ such that each G_0-panel contains at least three G_0-chambers.

Definition 6.5. Let Δ be a building (assumed to be thick). As in [Weiss 2003, 11.7] and [Mühlherr and Weiss 2017, 2.7], we say that Δ is *Moufang* if Δ is spherical, irreducible of rank at least 2 and for each root α of Δ, the root group U_α acts transitively on the set of apartments of Δ containing α. The spherical building associated with the group of rational points of an absolutely simple algebraic group of relative rank at least 2 is always Moufang, but there are a few other families as well. For a summary of the classification of Moufang buildings, see [Weiss 2009, 3.14–3.15].

Remark 6.6. A building of rank 1 is simply a set with no further structure. Thus the notion of a Moufang set is an extension of the notion of a Moufang building to buildings rank 1. (Buildings of rank 1 are automatically spherical and irreducible.)

Theorem 6.7. *Let Δ be a spherical building which is Moufang and let G_0 be a Galois subgroup of* $\mathrm{Aut}(\Delta)$ *as defined in [Mühlherr and Weiss 2017,* 4.5*]. Suppose that G_0 acts with finite orbits on the chamber set of Δ. Then G_0 is a descent group of Δ.*

Proof. This holds by [Mühlherr and Weiss 2017, 12.2(ii)]. □

Definition 6.8. Let Δ be as in Theorem 6.11. A *Galois involution* of Δ is an element ω of $\mathrm{Aut}(\Delta)$ of order 2 such that $\langle\omega\rangle$ is a Galois group as defined in [Mühlherr and Weiss 2017, 4.5].

Proposition 6.9. *Suppose that R is a residue of a Moufang building Δ. Let Σ be an apartment containing chambers of R and let U_R denote the subgroup generated by the root groups U_α for all roots α of Σ containing $R\cap\Sigma$. Then U_R is independent of the choice of Σ.*

Proof. This holds by [Mühlherr et al. 2015, 24.17]. □

Definition 6.10. The group U_R in Proposition 6.9 is called the *unipotent radical* of the residue R.

The following is a special case of the main results of [Mühlherr et al. 2015, Part 3].

Theorem 6.11. *Let G_0 be a descent group of a spherical building Δ. Let Π be the Coxeter diagram of Δ, let S denote the vertex set of Π and let Θ denote the subgroup of* $\mathrm{Aut}(\Pi)$ *induced by G_0. Then the following hold*:

(i) *The graph Δ^{G_0} is a building with respect to a canonical coloring of its edges.*

(ii) *All G_0-chambers are residues of Δ of the same type $A \subset S$, the set A is Θ-invariant and the rank k of Δ^{G_0} is the number of Θ-orbits in S disjoint from A.*

(iii) *The triple $T := (\Pi, \Theta, A)$ is a Tits index and Δ^{G_0} is a building of type $\widetilde{\Pi}$, where $\widetilde{\Pi}$ is the relative Coxeter diagram of T.*

(iv) *If Δ is Moufang and $k \geq 2$, then Δ^{G_0} is also Moufang.*

(v) *Suppose that Δ is Moufang and that $k = 1$ and let X denote the set of all G_0-chambers. Let G be the subgroup of* $\mathrm{Sym}(X)$ *induced by $C_{\mathrm{Aut}(\Delta)}(G_0)$ and for each $R \in X$, let $\widetilde{U}_R$ denote the subgroup of* $\mathrm{Sym}(X)$ *induced by $C_{U_R}(G_0)$, where U_R is as in Proposition 6.9. Then*

$$(G, X, \{\widetilde{U}_R\}_{R\in X})$$

is a Moufang set as defined in Definition 4.1.

Proof. Assertions (i) and (ii) hold by [Mühlherr et al. 2015, 22.20(v) and (viii)], assertion (iii) holds by [Mühlherr et al. 2015, 22.20(iv) and (viii)] and the remaining two assertions hold by [Mühlherr et al. 2015, 24.31]. □

Definition 6.12. Let G_0 and Δ be as in Theorem 6.11. We refer to the triple T in Theorem 6.11(iii) as the *Tits index of* G_0. (In fact, the Tits index of a descent group G_0 is defined also when Δ is not assumed to be spherical; see [Mühlherr et al. 2015, 22.20 and 22.22].)

Definition 6.13. A *fixed point building* is a building of the form Δ^{G_0} for some pair (Δ, G_0) as in Theorem 6.11. If the rank of Δ^{G_0} is 1 and Δ is Moufang, we interpret Δ^{G_0} to mean the Moufang set described in Theorem 6.11(v).

Remark 6.14. Apart from the Moufang sets described in Examples 4.2 and 4.3 and the Moufang sets described in Example 4.4 in the case that Γ is an indifferent quadrangle as defined in [Tits and Weiss 2002, 16.4], all other Moufang sets known to us (including those described in Example 4.5) arise (up to equivalence as defined in Definition 4.7) as the fixed point buildings Δ^{G_0} for some Moufang building Δ and some descent group with finite orbits G_0 whose Tits index is of relative rank 1.

7. The exceptional Tits quadrangles

In the following definition, we refer the reader to Remark 5.5 for (i) and we make use of the notation for spherical buildings described in [Weiss 2009, 30.15] (corresponding to the cases (vi), (xii) and (xiii) of [Weiss 2009, 30.14]) in (ii), (iii) and (iv).

Definition 7.1. We will say that a Tits quadrangle is *exceptional* if it is isomorphic to $X_{\Delta,T}$, where

(i) Δ is a Moufang quadrangle of type E_6, E_7, E_8 or F_4 as defined in [Tits and Weiss 2002, 16.6 and 16.7] and T is the split Tits index of absolute type C_2, or

(ii) Δ is the building $C_3^{\mathcal{I}}(C, K, \sigma)$ for some octonion division algebra C with center K and standard involution σ and T is the Tits index of absolute type C_3 displayed in Remark 5.6, or

(iii) Δ is the building $E_\ell(K)$ for some field K and $\ell = 6$, 7 or 8 and T is the Tits index of absolute type E_ℓ displayed in Remark 5.6, or

(iv) Δ is the building $F_4(C, K)$ for some composition division algebra (C, K) and T is the Tits index of absolute type F_4 displayed in Remark 5.6.

We will call a building *exceptional* if it is isomorphic to one of the buildings in (i)–(iv).

Notation 7.2. Let the pair (Δ, T) be as Definition 7.1 and let Π denote the Coxeter diagram associated with Δ. Thus $\Pi = C_2$, C_3, E_ℓ, F_4 in cases (i), (ii), (iii), (iv), respectively. If $\Pi = F_4$ but C/K is not a field extension in characteristic 2, we number the vertices of Π from left to right with the integers 1, 2, 3, 4 and choose the map Typ defined in [Mühlherr and Weiss $\geq$ 2020, 1.2.13] so that the $\{2, 3, 4\}$-residues are isomorphic to $\mathcal{B}_3^{\mathcal{Q}}(K, C, q)$, where q is the norm of the composition algebra (C, K). If $\Pi = C_2$ and X is a Moufang quadrangle of type E_6, E_7 or E_8, we number the vertices of Π from left to right with the integers 1, 2 and choose Typ so that a root group U_α is nonabelian if and only if the root α starts and ends at a vertex of type 2. Note that if $\Pi = F_4$ and C/K is a field extension in characteristic 2, then (C, K) is not necessarily an invariant of X and our choice of Typ is not uniquely determined. If $\Pi = C_2$ and X is a Moufang quadrangle of type F_4 or $\Pi = F_4$ and C/K is a field extension in characteristic 2 or $\Pi = E_6$, we choose Typ (between the two possible choices) arbitrarily. In the remaining cases, Π has no nontrivial symmetry and there is only one choice for Typ.

Remark 7.3. Since the Tits indices in Remark 5.6 are all invariant under $\mathrm{Aut}(\Pi)$, the Tits quadrangles in Definition 7.1 do not depend on the choice of Typ in Notation 7.2; see [Mühlherr and Weiss $\geq$ 2020, 1.2.13].

The exceptional Tits quadrangles were investigated in [Mühlherr and Weiss 2020]. Here is a reformulation of some of the results in that paper; we will cite others in Proposition 7.12 below.

Theorem 7.4. *Let X be an exceptional Tits quadrangle, let $(\gamma, i \mapsto w_i)$ be a coordinate system of X and let $i \mapsto U_i$ be the corresponding root group labeling. Let K be the field that appears in [Tits and Weiss 2002,* 16.6 *or* 16.7*] in the case that X is as in Definition 7.1*(i); *in the other cases, let K be as in Definition 7.1. If X is in case Definition 7.1*(iii) *or* (iv), *suppose that $|K| > 4$. Then the following hold*:

(i) *After replacing $(\gamma, i \mapsto w_i)$ by its opposite if necessary, there exists a δ-standard quadrangular algebra*

$$\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta)$$

over K and isomorphisms x_i from the group $\mathcal{R}_\Xi$ defined in Notation 2.4 to U_i for $i = 1$ and 3 and from the additive group of L to U_i for $i = 2$ and 4 such that

$$\begin{aligned} [x_1(a,t), x_4(v)^{-1}] &= x_2(\theta(a,v)+tv)x_3(av, tq(v)+\phi(a,v)), \\ [x_1(a,t), x_3(b,s)^{-1}] &= x_2(h(a,b)), \quad \textit{and} \\ [x_2(u), x_4(v)^{-1}] &= x_3(0, f(u,v)) \end{aligned} \tag{7.5}$$

for all $(a,t), (b,s) \in \mathcal{R}$ and all $u, v \in L$, where ϕ is as in (C4).

(ii) *If X is a Moufang polygon of type E_ℓ for $\ell = 6, 7$ or 8, then q is a quadratic form of type E_ℓ as defined in [Weiss 2006b,* 2.13*].*

(iii) *If X is a Moufang polygon of type F_4, then q is a quadratic form of type F_4 as defined in [Weiss 2006b,* 2.15*].*

(iv) *If X and C are as in Definition 7.1*(ii), *then q is the reduced norm of C.*

(v) *If X and ℓ are as in Definition 7.1*(iii), *then q is a split quadratic form of dimension $4 + 2^{\ell-5}$.*

(vi) *If X and (C, K) are as in Definition 7.1*(iv), *then q is the orthogonal sum of the reduced norm of (C, K) and a hyperbolic quadratic form of dimension* 4.

(vii) *X satisfies the hypothesis [Mühlherr and Weiss 2020,* 8.2(b)*].*

Proof. This holds by [Mühlherr and Weiss 2020, 7.4, 8.2, 10.4]. □

Remark 7.6. The quadratic forms in Theorem 7.4(ii)–(vi) are all of dimension at least 5.

Remark 7.7. If X and ℓ are as in Definition 7.1(iii), let (C, K) denote the split composition algebra of dimension $2^{\ell-5}$. Thus $q = q_C$ if q is as in Theorem 7.4(v) or (vi), where q_C is as defined in [Mühlherr and Weiss 2020, 7.3].

Remark 7.8. The quadratic forms in Theorem 7.4(ii)–(iv) are anisotropic. Over a finite field, there are no anisotropic quadratic forms of dimension greater than 2; see, for example, [Tits and Weiss 2002, 34.3]. By Remark 7.6, therefore, the field K must be infinite in cases Definition 7.1(i) and (ii).

Remark 7.9. The quadrangular algebra Ξ in Theorem 7.4 is anisotropic when X is as in Definition 7.1(i) but not when X is as in one of the other cases.

Remark 7.10. The quadrangular algebras that appear in Theorem 7.4 are, up to isotopy as defined in [Mühlherr and Weiss 2019, 5.3], the only δ-standard nondegenerate quadrangular algebras (as defined in Notation 2.2) satisfying $|K| > 4$ and $\dim_K L > 4$. This holds by [Weiss 2006b, 6.42 and 7.57] in the anisotropic case and by [Mühlherr and Weiss 2019, 8.16, 9.8 and 10.16] in the isotropic case.

Hypothesis 7.11. For the rest of this section, let X, $(\gamma, i \mapsto w_i)$, $i \mapsto U_i$, Ξ and $x_1, \dots, x_4$ be as in Theorem 7.4.

Proposition 7.12. *Let $Y_i = \{x_i(0, t) \mid t \in K\}$ for all odd i and let β_i denote the root $(w_i, w_{i+1}, \dots, w_{i+4})$ for all i. There exists a Tits quadrangle $\hat{X} = (\hat{\Gamma}, \hat{\mathcal{A}}, \{\hat{\equiv}_v\}_{v \in \hat{V}})$ such that the following hold*:

(i) *$\hat{\Gamma}$ is a subgraph of Γ containing γ and $\gamma \in \hat{\mathcal{A}}$.*

(ii) *For all even i, $\hat{\Gamma}$ is normalized by U_i and the natural embedding of U_i into* $\mathrm{Aut}(\hat{X})$ *is an injective isomorphism from U_i to the root group $\hat{U}_i$ of $\hat{X}$ corresponding to the root β_i.*

(iii) *For all odd i, $\hat{\Gamma}$ is normalized by Y_i and the natural embedding of Y_i into* $\mathrm{Aut}(\hat{X})$ *is an injective isomorphism from Y_i to the root group $\hat{U}_i$ of $\hat{X}$ corresponding to the root β_i.*

(iv) *The isomorphisms in* (ii) *and* (iii) *extend to an isomorphism from the subgroup $Y_1U_2Y_3U_4$ of* $\mathrm{Aut}(X)$ *to the subgroup $\hat{U}_{[1,4]} := \hat{U}_1\hat{U}_2\hat{U}_3\hat{U}_4$ of* $\mathrm{Aut}(\hat{X})$.

Proof. This holds by [Mühlherr and Weiss 2020, 5.1 and 10.4(iii)]. □

Proposition 7.13. *Let G_γ denote the pointwise stabilizer of γ in $G :=$* $\mathrm{Aut}(X)$ *and let δ be as in Notation 2.2 and Theorem 7.4. Then the following hold*:

(i) *Let $\varphi \in G_\gamma$. Then there exists σ-linear automorphisms ξ and ψ of L, respectively $\mathcal{X}$, for some $\sigma \in$* $\mathrm{Aut}(K)$ *and an element $\eta \in K^*$ such that the following hold*:

 (a) $q(\xi(v)) = q(\lambda)q(v)^\sigma$, *where* $\lambda := \xi(\varepsilon)$,

 (b) $\psi(av)\lambda = \psi(a)\xi(v)$,

 (c) $h(\psi(a), \psi(b)\lambda) = \eta\xi(h(a, b))$ *and*

 (d) $\theta(\psi(a), \xi(v)) = \eta\xi(\theta(a, v)) + M(a)\xi(v)$, *where*

$$M(a) = \begin{cases} 0 & \textit{if } \mathrm{char}(K) \neq 2, \\ f\big(\theta(\psi(a), \lambda), \xi(\delta)\big)q(\lambda)^{-1} & \textit{if } \mathrm{char}(K) = 2 \end{cases}$$

for all $a, b \in \mathcal{X}$ *and all* $v \in L$ *and*

$$
\begin{aligned}
x_1(a,t)^{\varphi} &= x_1(\psi(a), \eta t^{\sigma} + M(a)),\\
x_2(v)^{\varphi} &= x_2(\eta\xi(v)),\\
x_3(a,t)^{\varphi} &= x_3\big(\psi(a)\lambda, (\eta t^{\sigma} + M(a))q(\lambda) + \phi(\psi(a),\lambda)\big),\\
x_4(v)^{\varphi} &= x_4(\xi(v))
\end{aligned}
\tag{7.14}
$$

for all $(a, t) \in S$ *and all* $v \in L$. *If* $\dim_K L > 1$, *then* η *is uniquely determined by* ψ *and* ξ.

(ii) *Suppose that* $\dim_K L > 4$. *Let* $\sigma \in \mathrm{Aut}(K)$, *let* ξ *and* ψ *be* σ*-linear automorphisms of* L, *respectively* $\mathcal{X}$, *and let* η *be an element of* K^* *such that* (a)–(d) *hold. Then there exists* $\varphi \in G_\gamma$ *satisfying the identities* (7.14).

Proof. This holds by [Mühlherr and Weiss 2020, 13.1]. □

Notation 7.15. By [Mühlherr and Weiss 2020, 6.4(i) and (ii)], $x_1(0, 1) \in U_1^\sharp$ and $x_4(\varepsilon) \in U_4^\sharp$, where ε is as in Definition 2.1 and $U_1^\sharp$ and $U_4^\sharp$ as in Notation 3.7; see also [Mühlherr and Weiss 2020, 5.1(iii) and 10.4(iii)]. We can thus set $m_1 = \mu_\gamma(x_1(0, 1))$ and $m_4 = \mu_\gamma(x_4(\varepsilon))$, where μ_γ is as in (3.9).

Proposition 7.16. *Let* m_1 *and* m_4 *be as in Notation 7.15. Then the following hold*:

(i) $\big(x_1(a,t)x_2(u)x_3(b,s)\big)^{m_4} = x_1(b,s)x_2(-h(b,a) - \bar{u})x_3(-a,t),$ *and*

(ii) $\big(x_2(u)x_3(b,s)x_4(v)\big)^{m_1} = x_2(v)x_3(b, s - f(u,v))x_4(-u)$

for all $(a, t), (b, s) \in \mathcal{R}$ *and* $u, v \in L$.

Proof. Choose $(a, t), (b, s) \in \mathcal{R}$ and $u, v \in L$. By [Mühlherr and Weiss 2020, 6.4(iv), 8.19, 8.25 and the first display in 8.59], we have $x_1(a,t)^{m_4} = x_3(-a,t)$ and $x_2(u)^{m_4} = x_2(-\bar{u})$ and $x_3(b,s)^{m_4} = x_1(b,s)$ and by [loc. cit., 6.4(v) and 8.31], we have $x_2(u)^{m_1} = x_4(-u)$, $x_3(a,t)^{m_1} = x_3(a,t)$ and $x_4(v)^{m_1} = x_2(v)$. The claims hold, therefore, by Conventions 1.2 and (7.5). □

8. Galois groups

The main result of this section is Theorem 8.14, in which we show that the Moufang sets described in Example 4.5 associated with one of the Tits indices displayed in Hypothesis 8.4 can all be obtained as the fixed point building of a Galois involution acting on one of the Tits quadrangles described in Notation 7.2.

Let Δ, T, $X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v\in V})$,

$$\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta)$$

and $x_1, \dots, x_4$ be as in Theorem 7.4. Thus the pair (Δ, X) is as in one of the four cases of Definition 7.1. Let (C, K) be as in Definition 7.1(ii) or (iv) if X is in one of these two cases and let (C, K) be as in Remark 7.7 if X is in the case Definition 7.1(iii).

Proposition 8.1. *If K is separably closed, then X is as in Definition 7.1(iii) or* (iv) *and one of the following holds*:

(i) *(C, K) is split and $\Delta = F_4(K, K)$, $E_6(K)$, $E_7(K)$ or $E_8(K)$.*

(ii) *C/K is an inseparable field extension*, $\mathrm{char}(K) = 2$, *$C^2 \subset K$ and $\Delta = F_4(C, K)$.*

Proof. A quadratic form of type E_6, E_7, E_8 or F_4 has a subform that is similar to the norm of a separable quadratic extension. Since the only composition algebras over a separably closed fields are those that appear in (i) and (ii), the claims hold. □

Notation 8.2. Let $\hat{X}$ be as in Proposition 7.12, let $\hat{Q}$ be the orthogonal sum of q with a hyperbolic quadratic form of dimension 4, let $\hat{V} = K^4 \oplus L$ be the underlying vector space. Let $\hat{\Delta} = \mathcal{B}(\hat{Q})$ as defined in [Mühlherr et al. 2015, 35.5] if $\hat{Q}$ is not hyperbolic and let $\hat{\Delta} = \mathcal{D}(\hat{Q})$ as defined in [loc. cit., 35.9] if $\hat{Q}$ is hyperbolic. Thus $\hat{\Delta}$ is a building of type $\hat{\Pi} = B_{m+2}$ or D_{m+2}, where m is the Witt index of q. By [loc. cit., 35.10], we can identify $\mathrm{Aut}(\hat{\Delta})$ with $\mathrm{P\Gamma O}(\hat{Q})$. Let $\hat{T}$ be the Tits index $(\hat{\Pi}, \hat{A}, \hat{\Theta})$, where $\hat{A}$ is the unique subset of the vertex set of $\hat{\Pi}$ that spans a subdiagram of type B_m or D_m and $\hat{\Theta} = 1$. By [Mühlherr and Weiss 2020, 6.1, 6.3 and 6.10], $\hat{X}$ is isomorphic to the Tits quadrangle $X_{\hat{\Delta},\hat{T}}$. By Theorem 5.7, therefore, we can identify $\mathrm{Aut}(\hat{X})$ with $\mathrm{Aut}(\hat{\Delta})$. Let ψ denote the map from $\mathrm{Aut}(\hat{X})$ to $\mathrm{Aut}(K)$ such that $\psi(g) = \rho$ whenever g lifts to a ρ-linear element of $\Gamma\mathrm{O}(\hat{Q})$.

Proposition 8.3. *Let ψ be as in Notation 8.2, let G, G_γ and $\sigma_\varphi = \sigma \in \mathrm{Aut}(K)$ for each $\varphi \in G_\gamma$ be as Proposition 7.13. Then the following hold*:

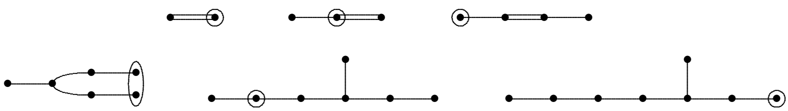
(i) *$\psi(\varphi) = \sigma_\varphi$ for all $\varphi \in G_\gamma$.*

(ii) *A subgroup G_0 of G is Galois as defined in [Mühlherr and Weiss 2017,* 4.5*] if and only if the restriction of ψ to G_0 is injective.*

Proof. In [Mühlherr et al. 2015, 3.20], a description of the Moufang quadrangle $B_2^{\mathcal{I}}(K, L, q)$ and its root groups associated to a fixed apartment is given. If we assume in this description that q is merely nondegenerate rather than anisotropic, the calculations in [loc. cit., 3.20] yield a description of $\hat{X}$, a coordinate system $(\gamma, i \mapsto w_i)$ (where γ is the circuit called Σ in [loc. cit., 3.20]), root groups $U_1, \ldots, U_4$ and isomorphisms $x_1, \ldots, x_4$ (called $\dot{\delta}$, $\dot{\beta}$, $\dot{\gamma}$, $\dot{\alpha}$ in [loc. cit., 3.20]). In this description, the vertices of $\hat{X}$ are the 1- and 2-dimensional totally isotropic subspaces of the quadratic space $(K, \hat{V}, \hat{Q})$ described in Notation 8.2 and two 2-dimensional totally isotropic subspaces W_1, W_2 of $\hat{V}$ that intersect in a 1-dimensional subspace Y are opposite at Y if and only if the subspace spanned by W_1 and W_2 is not totally isotropic. Given this description of the isomorphisms $x_1, \dots, x_4$, we can observe that (i) holds.

Let $g \in G$. The root groups of $\hat{\Delta}$ are all in the kernel of ψ. Let W be a 3-dimensional totally isotropic subgroup of V, let e_1, e_2, e_3 be a basis of W, and let h be an element in the subgroup of $\mathrm{Aut}(\hat{X})$ generated by all its root groups such that the product gh fixes the three 1-dimensional subspaces spanned by e_1, e_2 and e_3. Let $x_{\alpha_1}, x_{\alpha_2}, x_{\alpha_3}$ be as in [Mühlherr and Weiss 2016, §7] with W in place of V. The projective plane associated with W is a residue of $\hat{\Delta}$ which we denote by R and there exist $\lambda_1, \lambda_2, \lambda_3 \in E^*$ such that $x_{\alpha_i}(t)^{gh} = x_{\alpha_i}(\lambda_i t^{\sigma_g})$ for $i = 1, 2$ and 3 and all $t \in E$. By [Mühlherr et al. 2015, 29.15], it follows that ψ is a Galois map of $\hat{\Delta}$ as defined in [loc. cit., 29.25]. Thus (ii) holds. □

Hypothesis 8.4. Let G_0 be an arbitrary Galois group of Δ (as defined in [Mühlherr and Weiss 2017, 4.5]) that acts on the chamber set of Δ with finite orbits. By Theorem 6.7, G_0 is a descent group of Δ. Let $T_0 = (\Pi, A_0, \Theta_0)$ be its Tits index and let S be the vertex set of Π. The Coxeter diagram Π is as in Notation 7.2. If $\Pi = C_2$ or F_4, we assume that the type function of Δ is chosen as described in Notation 7.2. We also suppose that T_0 is one of the diagrams



Remark 8.5. The group G_0 acts on $X = X_{\Delta,T}$. Let R_1 and R_1' be two A_0-residues stabilized by G_0. By [Mühlherr et al. 2015, 22.10(i)], R_1 and R_1' are opposite in Δ. By [Mühlherr et al. 2015, 24.21], $\mathrm{Aut}(\Delta)$ acts transitively on ordered pairs of opposite A_0-residues. Replacing G_0 by a conjugate subgroup of $\mathrm{Aut}(\Delta)$, we can assume from now on that R_1 and R_1' are the A_0-residues corresponding to the vertices w_4 and w_0 of Γ.

Proposition 8.6. *The group G_0 stabilizes $\hat{X}$, where $\hat{X}$ is as in Proposition 7.12.*

Proof. Since G_0 fixes w_0 and w_4, it normalizes $E := \langle Z(U_{w_0}), Z(U_{w_4})\rangle$, where U_{w_i} denotes the maximal unipotent subgroup of the stabilizer of the residue w_i in $\mathrm{Aut}(\Delta)$. By (7.5) and [Mühlherr and Weiss ≥ 2020, 1.2.25], we have $Z(U_{w_4}) = Y_3U_4Y_5$ and $Z(U_{w_0}) = Y_7U_0Y_1$. By Proposition 7.12(ii)–(iii), E stabilizes $\hat{X}$. By [Mühlherr and Weiss ≥ 2020, 1.3.4], the group $\langle U_0, U_4\rangle$ acts transitively on $\hat{\Gamma}_{w_4} = \Gamma_{w_4}$ and by [Mühlherr and Weiss 2020, 8.2(b) and 10.4(ii)], the group $\langle Y_1, Y_5\rangle$ acts transitively on $\hat{\Gamma}_{w_5}$. By [Mühlherr and Weiss ≥ 2020, 1.2.21], $\hat{\Gamma}$ is connected. It follows that E acts transitively on the set of edges of $\hat{X}$. Since G_0 normalizes E and fixes the edge $\{w_4, w_5\}$, we conclude that G_0 stabilizes $\hat{X}$. □

Remark 8.7. We have $\hat{X} = X_{\hat{\Delta},\hat{T}}$, where $\hat{\Delta}$ and $\hat{T}$ are as in Notation 8.2. By Theorem 5.7 and Proposition 8.6, G_0 acts on $\hat{\Delta}$. Let $\hat{G}_0$ denote its image in $\mathrm{Aut}(\hat{\Delta})$.

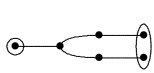
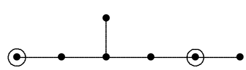
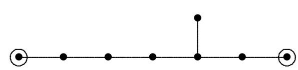
We now want to apply [Mühlherr et al. 2015, 35.13] with $\hat{\Delta}$ in place of Δ and $\hat{G}_0$ in place of Γ. By Proposition 8.3(ii), $\hat{G}_0$ is a Galois group of $\hat{\Delta}$. It is not necessarily finite. In the proof of [loc. cit., 35.13], however, the finiteness of the Galois group is used only in the proof of [loc. cit., 2.41] which is, in turn, a corollary of [loc. cit., 2.40] and thus of [Springer 1998, 11.1.6]. For this result to apply, it suffices to know that $\hat{G}_0$ acts with finite orbits on the chamber set of Δ. This follows from our assumption in Hypothesis 8.4 that G_0 acts with finite orbits on the chamber set of Δ.

Notation 8.8. For each subgroup G_1 of G_0, G_1 is also a Galois group of Δ and hence, by Theorem 6.7, has a Tits index as defined in Definition 6.12. We denote this Tits index by T_{G_1}.

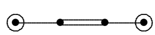
Proposition 8.9. *Suppose that K is separably closed. Then there exists a subgroup G_1 of index 2 in G_0 such that the Tits index T_{G_1} defined in Notation 8.8 has relative rank at least 2.*

Proof. Let ψ be as in Notation 8.2 and let $\hat{G}_0$ be as in Remark 8.7. By [Mühlherr et al. 2015, 35.13] and Remark 8.7, $\hat{\Delta}^{\hat{G}_0} \cong B_1^{\mathcal{Q}}(F, M, Q_0)$ for some anisotropic quadratic space (F, M, Q_0), where $\psi(G_0) = \mathrm{Gal}(K/F)$. Since the bilinear form associated with the quadratic form $\hat{Q}$ in Notation 8.2 is not identically zero, the same holds for Q_0. Since K is separably closed, there exists a subfield E of K over which Q_0 is not anisotropic such that E/F is a separable quadratic extension. Let $G_1 = \psi^{-1}(\mathrm{Gal}(K/E))$ and let $\hat{G}_1$ be the image of G_1 in $\hat{G}_0$. By [Mühlherr et al. 2015, 35.13], we have $\hat{\Delta}^{\hat{G}_1} \cong B_m^{\mathcal{Q}}(E, M_1, Q_1)$ for some $m \geq 1$, where $M_1 = M \otimes_F E$ and Q_1 is the anisotropic part of $Q \otimes_F E$. Thus $\dim_E Q_1 < \dim_F Q_0$. By [Mühlherr et al. 2015, 35.13(i)–(ii)], it follows that $m > 1$. Thus $\hat{G}_1$ acts isotropically (as defined in Definition 6.2) on the residue $\hat{R}_1$ of $\hat{\Delta}$ that corresponds to the vertex w_4. Let R_1 be as in Remark 8.5. If $\Pi \neq E_7$, then $R_1 = \hat{R}_1$ and hence G_1 acts isotropically on R_1. If $\Pi \neq E_7$, then $R_1 = \hat{R}_0 \times \hat{R}_1$, where $\hat{R}_0$ is a building of type A_1. Since $\hat{G}_1$ acts isotropically on $\hat{R}_1$, it follows by [Mühlherr et al. 2015, 21.37(iii)] that G_1 acts isotropically on R_1 also in this case. Thus the relative rank of T_{G_1} is at least 2. □

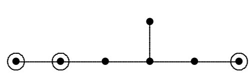
Proposition 8.10. *Let G_1 be as in Proposition 8.9 and let $T_1 := T_{G_1}$ be as in Notation 8.8. Then T_1 is one of the Tits indices*

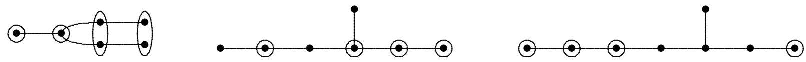
or



of relative rank 2 or the Tits index



of relative rank 3 *or one of the Tits indices*



of relative rank 4 *or* T_1 *is split as defined in Definition 5.3.*

Proof. The Tits index T_1 is as in Hypothesis 8.4 with at least one more circle. The only such Tits indices are those indicated. □

Proposition 8.11. *Suppose K is separably closed, let $\hat{G}_1$ be as in Proposition 8.9 and let $\Delta_1 = \Delta^{\hat{G}_1}$. Then the following hold:*

(i) *If the relative rank of T_1 is* 2, *then Δ_1 is an exceptional Moufang quadrangle.*

(ii) *If the relative rank of T_1 is* 3, *then $\Delta_1 \cong C_3^{\mathcal{I}}(C, K, \sigma)$ for some octonion division algebra (C, K) with standard involution σ.*

(iii) *If the relative rank of T_1 is* 4 *and $\Pi = E_\ell$ for $\ell = 6, 7$ or* 8, *then $\Delta_1 \cong F_4(D, E)$ for some composition division algebra (D, E) of dimension $2^{\ell-5}$ with nonzero trace.*

(iv) *If T_1 is split and $\Pi = F_4$, then either Δ_1 is the split building of type F_4 over E or $\Delta_1 \cong F_4(D, E)$ for some inseparable field extension D/E such that* $\mathrm{char}(E) = 2$ *and $D^2 \subset E$.*

(v) *If T_1 is split and $\Pi = E_6$, E_7 or E_8, then Δ_1 is the split building of type Π over E.*

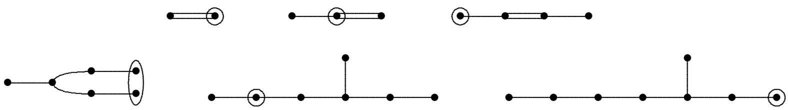
Proof. The assertion (i) holds by [Tits and Weiss 2002, 42.6], (ii) holds by [Tits 1974, 9.1–9.2], (iii) and (iv) hold by [Tits 1974, 10.2–10.3] and (v) holds by [Tits 1974, 6.13]. □

Proposition 8.12. *Suppose K is separably closed and Δ_1 is as in Proposition 8.11. Then there is a Galois group $\overline{G}$ of Δ_1 of order* 2 *such the fixed point building $\Delta_1^{\overline{G}}$ is isomorphic to Δ^G.*

Proof. Let $\overline{G}$ denote the image of G in $\mathrm{Aut}(\Delta_1)$. Then $\overline{G}$ is a Galois group and hence, by Theorem 6.7, $\overline{G}$ is a descent group of Δ_1. The claim holds, therefore, by [Mühlherr et al. 2015, 22.47(i)]. □

Remark 8.13. Suppose that K is separably closed and let Δ_1 and $\overline{G}$ be as in Proposition 8.12 and let Π_1 be the Coxeter diagram of Δ_1. If $\Pi_1 = C_2$ or F_4, we assume that the type function of Δ_1 is chosen as in Notation 7.2. Let $\overline{T} = (\Pi_1, \overline{A}, \overline{\Theta})$ be the Tits index of $\overline{G}$ as defined in Definition 6.12. Then $\overline{T}$ is the unique Tits index of absolute type Π_1 in Hypothesis 8.4. To see this we observe that in each case the A-residues of Δ stabilized by G_0 are precisely the $\overline{A}$-residues of Δ_1 stabilized by $\overline{G}$.

Theorem 8.14. *Let M be an arbitrary Moufang set associated with one of the Tits indices ${}^2E_{6,1}^{29}$, $E_{7,1}^{48}$, $E_{8,1}^{91}$ or $F_{4,1}^{21}$ as described in [Tits 1966a]. Then there exists an exceptional building Δ as defined in Definition 7.1 of type Π with type function chosen as in Notation 7.2 and a Galois group G_0 of order 2 of Δ whose Tits index is the unique Tits index of absolute type Π among the following*:



such that Δ^{G_0} is isomorphic to M.

Proof. The Tits indices ${}^2E_{6,1}^{29}$, $E_{7,1}^{48}$, $E_{8,1}^{91}$ or $F_{4,1}^{21}$ are precisely the indices that appear in Hypothesis 8.4. Thus the claims are simply restatements of Proposition 8.12 and Remark 8.13. □

Remark 8.15. Note that Proposition 8.12 and Remark 8.13 are actually slightly more general than Theorem 8.14 since they include the case Proposition 8.1(ii) which is not considered in [Tits 1966a].

9. The Galois involution ω

In light of Theorem 8.14, we now make once and for all the following assumptions.

Hypothesis 9.1. We assume that Δ, T, $X = (\Gamma, \mathcal{A}, \{\equiv_v\}_{v\in V})$, $(\gamma, i \mapsto w_i)$, $i \mapsto U_i$, $x_1, \ldots, x_4$ and

$$\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta)$$

are as in Notation 7.2 and Theorem 7.4. Let $v \mapsto \bar{v}$, ϕ, g and π be as in (A3), (C3), (C4) and (D1), let Q be as in Notation 2.3, let G and G_γ be as in Proposition 7.13 and let G_0 and T_0 be as in Hypothesis 8.4. We assume that $|G_0| = 2$ and let ω be the nontrivial element in G_0. Thus ω is a Galois involution of Δ as defined in Definition 6.8. Since the relative rank of T_0 is 1, the fixed point building Δ^{G_0} is the Moufang set described in Theorem 6.11(v). We denote this Moufang set by

$$M(\Delta, \omega).$$

Its underlying set (as defined in Definition 4.1) is the set of vertices of Γ fixed by ω. We also assume that $|K| > 4$.

Proposition 9.2. *The coordinate system $(\gamma, i \mapsto w_i)$ of X can be chosen so that $\omega = m_4\varphi$ for some $\varphi \in G_\gamma$, where m_4 is as in Notation 7.15.*

Proof. The fixed points of ω are all vertices of Γ at even distance from w_4. Let v, v_1 be two of them. By [Mühlherr et al. 2015, 25.2; Mühlherr and Weiss ≥ 2020, 1.2.12], u and u^ω are opposite at v for all $u \in \Gamma_v$. By [Mühlherr et al. 2015, 22.10(i)],

v and v_1 are opposite residues of Δ. Therefore there exist $u \in \Gamma$ and $u_1 \in \Gamma_{v_1}$ such that the edges $\{u, v\}$ and $\{u_1, v_1\}$ are opposite residues of Δ. By [Mühlherr and Weiss $\geq$ 2020, 1.2.19], it follows that there exists a straight 4-path from u to v. We conclude that $(\gamma, i \mapsto w_i)$ can be chosen so that w_0 and w_4 are fixed by ω and ω interchanges w_3 and w_5. By Definition 3.2(ii), it follows that ω maps w_i to w_{-i} for all i. By Proposition 3.8, m_4 also maps w_i to w_{-i} for all i. Thus $\omega = m_4\varphi$ for some $\varphi \in G_\gamma$. □

Notation 9.3. Let φ be as in Proposition 9.2 and let $\xi \in \mathrm{Aut}_K(L)$, $\psi \in \mathrm{Aut}_K(\mathcal{X})$, $\sigma \in \mathrm{Aut}(K)$, $\eta \in K^*$, $\lambda \in L$ and $M\colon \mathcal{X} \to K$ be as in Proposition 7.13 applied to φ. Note that since ξ is a bijection, it follows from Proposition 7.13(a) that $q(\lambda) \neq 0$.

Proposition 9.4. $\quad f(\xi(u), \xi(v)) = q(\lambda) f(u, v)^\sigma$ *for all* $u, v \in L$.

Proof. This is a consequence of Proposition 7.13(a). □

Proposition 9.5. *Let* ω, λ, σ, ξ *and* M *be as in* Notation 9.3. *Then the following hold*:

(i) $\psi^2(a)\lambda = -a$ *for all* $a \in \mathcal{X}$.

(ii) σ *has order* 2.

(iii) $\eta^{\sigma+1} = \beta$, *where* $\beta^{-1} = q(\lambda)$.

(iv) $M(\psi(a)) = \eta M(a)^\sigma + \phi(a, \lambda^{-1})$ *for all* $a \in \mathcal{X}$.

(v) $\xi(\overline{\xi(\bar{u})}) = q(\lambda)u$ *for all* $u \in L$.

Proof. By Proposition 7.16(i), we have $x_3(a,t)^{m_4} = x_1(a,t)$ and $x_1(a,t)^{m_4} = x_3(-a,t)$ for all $(a,t) \in \mathcal{R}$. Thus

$$\begin{aligned}(9.6)\quad x_3(a,t) &= x_3(a,t)^{\omega^2}\\ &= x_3\big(-\psi^2(a)\lambda,\ \big(\eta(\eta t^\sigma + M(a))^\sigma + M(\psi(a))\big)q(\lambda) + \phi(\psi^2(a), \lambda)\big)\end{aligned}$$

for all $(a,t) \in \mathcal{R}$ by (7.14). Comparing first coordinates, we obtain (i). Since ψ is σ-linear, it follows that $\sigma^2 = 1$. By Proposition 8.3 and Hypothesis 9.1, on the other hand, σ is nontrivial. Thus (ii) holds. Comparing second coordinates, we have

$$\eta^{\sigma+1}t + \eta M(a)^\sigma + M(\psi(a)) = \beta\big(\phi(\psi^2(a), \lambda) + t\big)$$

for all $(a,t) \in \mathcal{R}$, where β is as in (iii). Setting $a = 0$ and $t = 1$ in this equation, we obtain (iii) and thus

$$(9.7)\qquad \eta M(a)^\sigma + M(\psi(a)) = \beta\phi(\psi^2(a), \lambda).$$

By Proposition 7.13(d) and [Weiss 2006b, 4.5(iii)], M and ϕ are both identically zero if $\mathrm{char}(K) \neq 2$. If $\mathrm{char}(K) = 2$, then by [Mühlherr and Weiss 2019, (7.10)] and (i), we have

$$(9.8)\quad \phi(\psi^2(a), \lambda) = \phi(-a\lambda^{-1}, \lambda) = \phi(a\bar{\lambda}, \lambda)\beta^2 = \phi(a, \bar{\lambda})\beta = \phi(a, \lambda^{-1})q(\lambda)$$

for all $a \in \mathfrak{X}$. By (9.7), therefore, (iv) holds. By [Weiss 2006b, 6.4(iv)], we have $x_2(u)^{m_4} = x_2(-\bar{u})$ for all $u \in L$ and by Proposition 7.13(i), ξ is σ-linear. Thus

$$x_2(u) = x_2(u)^{\omega^2} = x_2\big(\eta^{\sigma+1}\xi(\overline{\xi(\bar{u})})\big) \tag{9.9}$$

for all u by (7.14). By (iii), therefore, (v) holds. □

Proposition 9.10. *Let $\zeta(u) = \eta\xi(\bar{u})$ for all $u \in L$. Then ζ is a σ-linear automorphism of L of order 2 and $q(\zeta(u)) = \eta^{1-\sigma}q(u)^\sigma$ for all $u \in L$.*

Proof. Since ξ is σ-linear, so is ζ. By Proposition 9.5(ii), (iii) and (v), ζ is of order 2. We have $q(\bar{u}) = q(u)$ for all $u \in L$. Thus the second claim holds by Propositions 7.13(a) and 9.5(iii). □

Proposition 9.11. *Let $v \in L$ and let ζ be as in Proposition 9.10. Then v is fixed by ζ if and only if $v = u + \zeta(u)$ for some $u \in L$.*

Proof. Since ζ is of order 2, $v = u + \zeta(u)$ implies $\zeta(v) = v$. Suppose, conversely, that $\zeta(v) = v$. Since σ is of order 2, we can choose $\delta_0 \in K$ such that $\delta_0 + \delta_0^\sigma = 1$. Then $u + \zeta(u) = (\delta_0 + \delta_0^\sigma)v = v$ for $u = \delta_0 v$. □

Remark 9.12. Let $F = K^\sigma := \{t \in K \mid t^\sigma = t\}$ and let $L_\zeta = \{v \in L \mid \zeta(v) = -v\}$. Then L_ζ is a vector space over F and by [Mühlherr et al. 2015, 2.40] applied to the group generated by $v \mapsto -\zeta(v)$, $\dim_F L_\zeta = \dim_K L$.

Remark 9.13. By Proposition 9.10, $q(u) \in \eta F$ for all $u \in L_\zeta$, where F and L_ζ are as in Remark 9.12. Thus $\eta^{-1}q$ is a quadratic form on L_ζ as a vector space over F.

Proposition 9.14. *For each $b \in \mathfrak{X}$, there exists $u \in L$ such that*

$$u + \zeta(u) = h(\psi(b), b),$$

where ζ is as in Proposition 9.10.

Proof. Let $b \in \mathfrak{X}$. Then

$$\zeta(h(\psi(b), b)) = \eta\xi(\overline{h(\psi(b), b)}) = -\eta\xi\big(h(b, \psi(b))\big) = -h(\psi(b), \psi^2(b)\lambda)$$

by Proposition 7.13(c). By Proposition 9.5(i), it follows that $h(\psi(b), b)$ is fixed by ζ. The claim holds, therefore, by Proposition 9.11. □

Proposition 9.15. *Let $(a, t), (b, s) \in \mathcal{R}$ and $u \in L$. Then $x_1(a, t)x_2(u)x_3(b, s)$ is centralized by ω if and only if $a = \psi(b)$, $t = \eta s^\sigma + M(b)$ and $u + \zeta(u) = h(a, b)$.*

Proof. By Proposition 7.16(i), we have

$$\begin{aligned}\big(x_1(a, t)x_2(u)x_3(b, s)\big)^{m_4} &= x_1(b, s)x_2(-\bar{u} - h(b, a))x_3(-a, t)\\ &= x_1(b, s)x_2(\overline{h(a, b) - u})x_3(-a, t).\end{aligned}$$

By (7.14) and Propositions 9.2 and 9.10, therefore,

$$\big(x_1(a,t)x_2(u)x_3(b,s)\big)^{\omega} = x_1\big(\psi(b),\ \eta s^{\sigma} + M(b)\big)x_2\big(\zeta(h(a,b)-u)\big) \cdot x_3\big(-\psi(a)\lambda,\ (\eta t^{\sigma}+M(a))q(\lambda) + \phi(\psi(a),\lambda)\big).$$

Thus $x_1(a,t)x_2(u)x_3(b,s)$ is centralized by ω if and only if

$$\psi(b) = a, \quad t = \eta s^{\sigma} + M(b) \quad \text{and} \quad u + \zeta(u) = h(a,b) \tag{9.16}$$

as well as

$$-\psi(a)\lambda = b \quad \text{and} \quad s = (\eta t^{\sigma} + M(a))q(\lambda) + \phi(\psi(a),\lambda). \tag{9.17}$$

By Proposition 9.5(i)–(iv) and [Mühlherr and Weiss 2019, (7.10)], (9.16) implies that $-\psi(a)\lambda = b$ and

$$\begin{aligned}(\eta t^{\sigma} + M(a))q(\lambda) &= \big(\eta^{\sigma+1}s + \eta M(b)^{\sigma} + M(\psi(b))\big)q(\lambda)\\ &= s + \phi(b,\lambda^{-1})q(\lambda)\\ &= s + \phi(\psi(a)\lambda,\bar{\lambda})q(\lambda)^{-1}\\ &= s + \phi(\psi(a),\lambda).\end{aligned}$$

Therefore (9.16) implies (9.17). □

Notation 9.18. Let

$$\mathcal{U} = \big\{[b,s,u] \mid (b,s,u) \in \mathcal{R} \times L \text{ and } u + \zeta(u) = h(\psi(b),b)\big\},$$

where

$$[b,s,u] = x_1\big(\psi(b),\eta s^{\sigma} + M(b)\big)x_2(u)x_3(b,s).$$

Thus $\mathcal{U} = C_{U_{[1,3]}}(\omega)$ by Proposition 9.15.

Remark 9.19. By Definition 6.10 and [Mühlherr and Weiss ≥ 2020, 1.2.25], $U_{[1,3]}$ is the unipotent radical of the residue of Δ corresponding to w_4. Hence the group $\mathcal{U}$ is the root group U_{w_4} of the Moufang set $M(\Delta,\omega)$ defined in Hypothesis 9.1.

Remark 9.20. Let $[b_1,u_1,s_1], [b_2,u_2,s_2] \in \mathcal{U}$. Then

$$[b_1,s_1,u_1]\cdot[b_2,s_2,u_2] = x_1\big(\psi(b_1),\ \eta s_1^{\sigma} + M(b_1)\big)x_1\big(\psi(b_2),\ \eta s_2^{\sigma} + M(b_2)\big) \cdot x_2\big(u_1+u_2+h(\psi(b_2),b_1)\big)x_3(b_1,s_1)x_3(b_2,s_2)$$

by (7.5). By Notation 2.4, we have

$$\begin{aligned}&x_1\big(\psi(b_1),\ \eta s_1^{\sigma} + M(b_1)\big)x_1\big(\psi(b_2),\ \eta s_2^{\sigma} + M(b_2)\big)\\ &\qquad = x_1\big(\psi(b_1+b_2),\ \eta(s_1+s_2)^{\sigma} + M(b_1) + M(b_2) + g(\psi(b_2),\psi(b_1))\big)\end{aligned}$$

and

$$x_3(b_1,s_1)x_3(b_2,s_2) = x_3\big(b_1+b_2,\ s_1+s_2+g(b_2,b_1)\big).$$

Since $\mathcal{U} = C_{U_{[1,3]}}$ is closed under multiplication, we conclude by Notation 9.18 that

$$(9.21)\quad [b_1, s_1, u_1] \cdot [b_2, s_2, u_2] \\ = \big[b_1 + b_2,\ s_1 + s_2 + g(b_2, b_1),\ u_1 + u_2 + h(\psi(b_2), b_1)\big]$$

and

$$(9.22)\qquad M(b_1 + b_2) + \eta g(b_2, b_1)^\sigma = M(b_1) + M(b_2) + g(\psi(b_2), \psi(b_1)).$$

for all $[b_1, s_1, u_1], [b_2, s_2, u_2] \in \mathcal{U}$. It follows from (9.21) that

$$(9.23)\qquad [b, s, u]^{-1} = \big[-b,\ -s + g(b, b),\ -u + h(\psi(b), b)\big]$$

for all $[b, s, u] \in \mathcal{U}$.

10. The structure map τ

The goal of this section is to introduce the structure map of the Moufang set $M(\Delta, \omega)$ defined in Hypothesis 9.1. See Notation 10.7 and Definition 10.8.

Proposition 10.1. *Let m_1 and m_4 be as in Notation 7.15, let ω and φ be as in Proposition 9.2, let $r_0 = (m_1 m_4)^2$ and let $h_0 \in G$. Then*

$$[\omega, r_0 h_0] = 1$$

if and only if

$$h_0^{m_4}\varphi = \varphi^{r_0} h_0.$$

Proof. By [Mühlherr and Weiss ≥ 2020, 1.4.15], $r_0 = m_4 m_1 m_4 m_1$ and hence $m_4 r_0 = m_4 m_1 m_4 m_1 m_4 = r_0 m_4$. Thus $[\omega, r_0 h_0] = 1$ if and only if

$$r_0 m_4 \cdot h_0^{m_4}\varphi = r_0 h_0 \cdot m_4\varphi \\ = m_4\varphi \cdot r_0 h_0 = m_4 r_0 \cdot \varphi^{r_0} h_0 = r_0 m_4 \cdot \varphi^{r_0} h_0.$$

□

Proposition 10.2. *Let $\omega = m_4\varphi$ be as in Proposition 9.2, let ξ, ψ, σ, η and λ be as Notation 9.3 and let r_0 be as Proposition 10.1. Then conjugation by φ^{r_0} induces the maps*

$$x_2(u) \mapsto x_2(\eta^\sigma \xi(u))$$

and

$$x_3(b, s) \mapsto x_3\big(\eta^\sigma \psi(b)\lambda,\ \eta^\sigma s^\sigma + \eta^{\sigma-1} M(b) + \eta^{2\sigma} \phi(\psi(b), \lambda)\big)$$

on U_2 and U_3.

Proof. By (7.14) and Proposition 7.16(ii), conjugation by φ^{m_1} induces the maps

$$x_2(u) \mapsto x_2(\xi(u))$$
$$x_3(b,s) \mapsto x_3\big(\psi(b)\lambda,\ (\eta s^\sigma + M(b))q(\lambda) + \phi(\psi(b),\lambda)\big)$$
$$x_4(v) \mapsto x_4(\eta\xi(v))$$

on U_2, U_3 and U_4. We apply now the notation in [Tits and Weiss 2002, 5.10]. Conjugating both sides of the identity

$$[x_1(a,t), x_4(\varepsilon)^{-1}]_3 = x_3(a,t) \tag{10.3}$$

by $h_1 := \varphi^{m_1}$, we thus obtain

$$\begin{aligned} x_3\big(\psi(a)\lambda,\ (\eta t^\sigma + M(a))q(\lambda) + \phi(\psi(a),\lambda)\big) &= [x_1(a_1,t_1), x_4(\eta\lambda)^{-1}]_3 \\ &= x_3\big(a_1\cdot\eta\lambda,\ t_1 q(\eta\lambda) + \phi(a_1,\eta\lambda)\big), \end{aligned}$$

where $x_1(a_1,t_1) = x_1(a,t)^{h_1}$. By Proposition 9.5(iii), $\eta^{\sigma+1} = q(\lambda)^{-1}$. Solving for a_1 and then for t_1, we conclude that h_1 induces the map

$$x_1(a,t) \mapsto x_1\big(\eta^{-1}\psi(a),\ \eta^{-1}t^\sigma + \eta^{-2}M(a)\big)$$

on U_1. By Proposition 7.16(i), it follows that conjugation by $\varphi^{m_1m_4}$ induces the maps

$$x_1(a,t) \mapsto x_1\big(\psi(a)\lambda,\ (\eta t^\sigma + M(a))q(\lambda) + \phi(\psi(a),\lambda)\big)$$
$$x_2(u) \mapsto x_2(\overline{\xi(\bar{u})})$$
$$x_3(b,s) \mapsto x_3\big(\eta^{-1}\psi(b),\ \eta^{-1}s^\sigma + \eta^{-2}M(b)\big)$$

on U_1, U_2 and U_3. Conjugating both sides of the identity

$$[x_1(0,1), x_4(v)^{-1}]_2 = x_2(v)$$

by $h_2 := \varphi^{m_1m_4}$, we obtain

$$[x_1(0,\eta q(\lambda)), x_4(v_1)^{-1}]_2 = x_2(\overline{\xi(\bar{v})}),$$

where $x_4(v_1) = x_4(v)^{h_2}$. It follows that $\varphi^{m_1m_4}$ induces the map

$$x_4(v) \mapsto x_4(\eta^\sigma\overline{\xi(\bar{v})})$$

on U_4. By Proposition 7.16(i), therefore, conjugation by $\varphi^{m_1m_4m_1}$ induces the maps

$$x_2(u) \mapsto x_2(\eta^\sigma\overline{\xi(\bar{u})})$$
$$x_3(b,s) \mapsto x_3\big(\eta^{-1}\psi(b),\ \eta^{-1}s^\sigma + \eta^{-2}M(b)\big)$$
$$x_4(v) \mapsto x_4(\overline{\xi(\bar{v})})$$

on U_2, U_3 and U_4. By [Mühlherr and Weiss 2019, (7.10)], we have $\phi(bv, \bar{v}) = \phi(b, v)q(v)$ for all $(b, v) \in \mathcal{R}$. Using (10.3) again, we can thus conclude that conjugation by $\varphi^{m_1 m_4 m_1}$ induces the map

$$x_1(a, t) \mapsto x_1\big(\eta^\sigma \psi(a)\lambda,\ \eta^\sigma t^\sigma + \eta^{\sigma-1} M(a) + \eta^{2\sigma}\phi(\psi(a), \lambda)\big)$$

on U_1. The claims follow now by one more application of Proposition 7.16(i). □

Notation 10.4. Let η be as in Notation 9.3, let $\psi_0(a) = \eta a$ for all $a \in \mathcal{X}$, let $\xi_0(v) = \eta^{-1} v$ for all $v \in L$, let σ_0 be the identity automorphism of K and let $\eta_0 = \eta^2$. Then ψ_0, ξ_0, σ_0 and η_0 fulfill the identities Proposition 7.13(a)–(d) with $\lambda = \eta^{-1}$; note that by [Weiss 2006b, 4.1], $M(a) = 0$ in (d). Let h_0 be the element of G_γ obtained by applying Proposition 7.13(ii) to ψ_0, ξ_0, σ_0 and η_0. By [Weiss 2006b, (4.13)], $\phi(a, \varepsilon) = 0$ for all $a \in \mathcal{X}$. Thus conjugation by h_0 induces the maps

$$\begin{aligned} x_1(a, t) &\mapsto x_1(\eta a, \eta^2 t) && \text{on } U_1, \\ x_2(v) &\mapsto x_2(\eta v) && \text{on } U_2, \quad \text{and} \\ x_3(b, s) &\mapsto x_3(b, s) && \text{on } U_3. \end{aligned}$$

Let $\nu = r_0 h_0$, where r_0 is as in Proposition 10.1.

Proposition 10.5. *Let h_0 and η be as in Notation 10.4. Then conjugation by $h_0^{m_4}$ induces the maps*

$$x_2(u) \mapsto x_2(\eta u) \quad \textit{and} \quad x_3(b, s) \mapsto x_3(\eta b, \eta^2 s)$$

on U_2 and U_3.

Proof. This holds by Proposition 7.16(i). □

Proposition 10.6. $[\omega, \nu] = 1$, *where ω is as in Proposition 9.2 and ν is as in Notation 10.4.*

Proof. Note that by Proposition 7.13(d), $M(tb) = t^{2\sigma} M(b)$ for all $b \in \mathcal{X}$ and all $t \in K$. By (7.14), Propositions 10.2, 10.5 and Notation 10.4, therefore, conjugation by $h_0^{m_4}\varphi$ and conjugation by $\varphi^{r_0} h_0$ both induce the maps

$$\begin{aligned} x_2(u) &\mapsto x_2(\eta^{\sigma+1}\xi(u)) && \text{on } U_2, \quad \text{and} \\ x_3(b, s) &\mapsto x_3\big(\eta^\sigma \psi(b)\lambda,\ \eta^\sigma s^\sigma + \eta^{\sigma-1} M(b) + \eta^{2\sigma}\phi(\psi(b), \lambda)\big) && \text{on } U_3. \end{aligned}$$

By [Mühlherr and Weiss ≥ 2020, 1.4.17(ii)], it follows that $h_0^{m_4}\varphi = \varphi^{r_0} h_0$. By Proposition 10.1, therefore, $[\omega, \nu] = 1$. □

Notation 10.7. Let $\mathcal{U}$ be as in Notation 9.18 and let $\nu = (m_1 m_4)^2 h_0$ be as in Notation 10.4. Thus ν interchanges w_4 and w_0 and by Proposition 10.6, ν commutes with ω. By [Mühlherr et al. 2015, 24.10(i)], the proper residues fixed by ω are pairwise opposite in Δ and by [Mühlherr et al. 2015, 24.21], $\mathcal{U}$ acts sharply transitively on the set of vertices fixed by ω that are distinct from w_4. Choose $g \in \mathcal{U}^*$. Then w_0^g is opposite both w_4 and w_0. Hence $w_0^{g\nu}$ is also opposite both w_4 and w_0. Thus

$$w_0^{g\nu} = w_0^{\tau(g)}$$

for a unique element $\tau(g)$ of $\mathcal{U}^*$. Since ν stabilizes the set $w_0^{\mathcal{U}^*}$, the map τ is a permutation of $\mathcal{U}^*$.

Definition 10.8. The map τ defined in Notation 10.7 is the map called f in [Thompson 1972, (3.1)], which Thompson called the *structure equation* of the Moufang set he was investigating. For this reason, we call the permutation τ the *structure map* of the Moufang set $M(\Delta, \omega)$. By [De Medts and Weiss 2006, 3.1], the Moufang set $M(\Delta, \omega)$ is uniquely determined by the pair $(\mathcal{U}, \tau)$.

11. The subgraph Λ

We have described the group $\mathcal{U}$ in Notation 9.18 and (9.21) and defined τ in Notation 10.7. Our goal in the remainder of this paper is to determine τ explicitly. In this section we begin to investigate how the product $m_1 m_4$ operates on Γ.

Notation 11.1. We continue with all the notation and assumptions in Hypothesis 9.1. Let $\mathcal{R}_\Xi$ as in Notation 2.4. We set $\mathcal{R} = \mathcal{R}_\Xi$ and

$$\mathcal{R}^\times = \{(a, t) \in \mathcal{R} \mid q(\pi(a) + t\varepsilon) \neq 0\}.$$

Observations 11.2. By Conventions 1.2 and (7.5),

$$x_4(v)x_1(a, t) = x_1(a, t)x_2(\theta(a, v) + tv)x_3\big(av, tq(v) + \phi(a, v)\big)x_4(v)$$

for all $(a, t) \in \mathcal{R}$ and all $v \in L$. Replacing v by $-v$ in this identity and applying (C1) and [Weiss 2006b, 4.14], we thus conclude that

$$x_1(a, t)x_4(v) = x_4(v)x_1(a, t)x_2(-\theta(a, v) - tv)x_3\big(-av, tq(v) + \phi(a, v)\big)$$

for all $(a, t) \in \mathcal{R}$ and all $v \in L$. By Conventions 1.2 and (7.5), we also have

$$x_3(b, s)x_1(a, t) = x_1(a, t)x_3(b, s)x_2(h(a, b))$$

and

$$x_1(a, t)x_3(b, s) = x_3(b, s)x_1(a, t)x_2(-h(a, b))$$

for all $(a, t), (b, s) \in \mathcal{R}$ and

$$x_4(v)x_2(u) = x_2(u)x_4(v)x_3(0, f(u, v))$$

and

$$x_2(u)x_4(v) = x_4(v)x_2(u)x_3(0, -f(u, v))$$

for all $u, v \in L$.

Using these observations, we now calculate that

$$(11.3)\quad \begin{aligned} &x_1(a,t)x_2(u)x_3(b,s)x_4(v)\\ &= x_4(v)x_1(a,t)x_2(-\theta(a,v)-tv+u)\\ &\qquad\cdot x_3\big(b-av,\ tq(v)+\phi(a,v)+s-f(u,v)-g(b,av)\big)\end{aligned}$$

and

$$(11.4)\quad \begin{aligned} &x_4(v)x_1(a,t)x_2(u)x_3(b,s)\\ &= x_1(a,t)x_2(\theta(a,v)+tv+u)\\ &\qquad\cdot x_3\big(b+av,\ tq(v)+\phi(a,v)+f(u,v)+s+g(b,av)\big)x_4(v)\end{aligned}$$

for all (a,t), $(b,s)\in\mathcal{R}$ and for all $u,v\in L$.

Notation 11.5. Let Γ and $\mathcal{A}$ be as in Hypothesis 9.1, let Λ be the subgraph of Γ defined in [Mühlherr and Weiss $\geq$ 2020, 1.5.4] and let $U_+ = U_{[1,4]}$ as defined in Proposition 3.5(i). Thus Λ is the subgraph consisting of all the vertices and edges of Γ that lie on an element of $\mathcal{A}$ containing the vertices w_4 and w_5.

Let Ψ be the graph whose vertex set is the disjoint union of the set $\{w_4, w_5\}$, the set of right cosets in U_+ of the subgroups U_4, $U_{[3,4]}$ and $U_{[2,4]}$ and the set of right cosets in U_+ of the subgroups U_1, $U_{[1,2]}$ and $U_{[1,3]}$ and whose edge set consists of the unordered pairs $\{w_4, w_5\}$ as well as $\{U_1g, U_4g\}$,

$$\begin{aligned} &\{w_4, U_{[1,3]}g\},\quad \{U_{[1,3]}g, U_{[1,2]}g\},\quad \{U_{[1,2]}g, U_1g\}\\ \text{and}\quad &\{w_5, U_{[2,4]}g\},\quad \{U_{[2,4]}g, U_{[3,4]}g\},\quad \{U_{[3,4]}g, U_4g\}\end{aligned}$$

for all $g\in U_+$. By [Mühlherr and Weiss $\geq$ 2020, 1.5.8], there is a unique isomorphism from Ψ to Λ that sends w_4 to w_4, w_5 to w_5, w_i^g to $U_{[1,i]}g$ and w_{5+i}^g to $U_{[1+i,4]}g$ for $i=1,2,3$ and for all $g\in U_+$. We identify Λ with Ψ via this isomorphism. Note that with this identification, the set $w_0^{\mathcal{U}}$ that appears in Notation 10.7 is now $\{U_4g \mid g\in\mathcal{U}\}$.

Proposition 11.6. *Let $g\in U_+$. The following hold*:

(i) *For all $(b,s)\in\mathcal{R}$ and all $u,v\in L$, the vertex*

$$U_1x_2(u)x_3(b,s)x_4(v)$$

of Λ is adjacent to the vertex U_4g if and only if U_4g equals

$$U_4x_1(a,t)x_2\big(-\theta(a,v)-tv+u\big)x_3\big(b-av,\ tq(v)+\phi(a,v)+s-f(u,v)-g(b,av)\big)$$

for some $(a,t)\in\mathcal{R}$.

(ii) *For all (a,t), $(b,s)\in\mathcal{R}$ and all $u\in L$, the vertex*

$$U_4x_1(a,t)x_2(u)x_3(b,s)$$

of Λ is adjacent to the vertex U_1g if and only if U_1g equals

$$U_1x_2\big(\theta(a,v)+tv+u\big)x_3\big(b+av,\ tq(v)+\phi(a,v)+f(u,v)+s+g(b,av)\big)x_4(v)$$

for some $v\in L$.

Proof. Let $e\in U_+$. The vertices of Λ adjacent to U_1e are $U_{[1,2]}e$ and $U_4x_1(a,t)e$ for all $(a,t)\in\mathcal{R}$ and the vertices of Λ adjacent to U_4e are $U_{[3,4]}e$ and $U_1x_4(v)e$ for all $v\in L$. The two claims hold, therefore, by (11.3) and (11.4). □

Proposition 11.7. *Let z be a vertex of Γ. There exists a root (v_0,v_1,v_2,v_3,v_4) such that $v_0=w_4$, $v_1=w_5$ and $v_4=z$ if and only if z is a vertex of Λ of the form U_4g for some $g\in U_{[1,3]}$.*

Proof. It suffices to observe that if $z=U_4g$ for some $g\in U_{[1,3]}$, then

$$\big(w_4,w_5,U_{[2,4]}g,U_{[3,4]}g,U_4g\big)$$

is a root. □

Proposition 11.8. *Let $\mathcal{R}^\times$ be as in Notation 11.1, let*

$$\mathcal{O}=\big\{U_4x_1(a,t)x_2(u)x_3(b,s)\mid(a,t),(b,s)\in\mathcal{R},\ u\in L\big\},$$
$$\mathcal{B}=\big\{U_4x_1(a,t)x_2(u)x_3(b,s)\in\mathcal{O}\mid(a,t)\in\mathcal{R}^\times\big\},$$

and let $p=U_4x_1(a,t)x_2(u)x_3(b,s)\in\mathcal{O}$. Then $p^{m_1m_4}\in\mathcal{O}$ if and only if $p\in\mathcal{B}$.

Proof. Let α denote the sequence of vertices

$$\big(U_{[2,4]},\ w_5,\ U_{[2,4]}x_1(a,t),\ U_{[3,4]}x_1(a,t)x_2(u),\ U_4x_1(a,t)x_2(u)x_3(b,s)\big).$$

The sequence α is a root (i.e., a straight path of length 4 in Γ) if and only if the vertices $U_{[2,4]}$ and $U_{[2,4]}x_1(a,t)$ are opposite at w_5. By [Mühlherr and Weiss 2020, 9.12], $U_{[2,4]}$ and $U_{[2,4]}x_1(a,t)$ are opposite at w_5 if and only if $p\in\mathcal{B}$. Thus the sequence $\alpha^{m_1m_4}$ (which starts at w_4 and ends at $p^{m_1m_4}$) is a root if and only if $p\in\mathcal{B}$. Every root starting at w_4 ends at a vertex in $\mathcal{O}$. Thus if $p\in\mathcal{B}$, then $p^{m_1m_4}\in\mathcal{O}$. By [Mühlherr and Weiss $\geq$ 2020, 1.2.28(ii) and 1.3.18], every path of length at most 4 from w_4 to a vertex in $\mathcal{O}$ is a root. Thus if $p^{m_1m_4}\in\mathcal{O}$, then $p\in\mathcal{B}$. □

12. The action of m_1m_4 on the set $\mathcal{B}$

The main result of this section is Proposition 12.15.

Proposition 12.1. *Let $(a,t)\in\mathcal{R}^\times$, $u\in L$ and $(b,s)\in\mathcal{R}$. Then the image of the vertex*

$$p:=U_4x_1(a,t)x_2(u)x_3(b,s)$$

of Λ under m_1 is the vertex

$$U_4x_1(a_0,t_0)x_2(u_0)x_3(b_0,s_0),$$

where

(i) $a_0 = q(\pi(a)+t\varepsilon)^{-1}a\big(\pi(a)+(Q(a)-t)\varepsilon\big)$,

(ii) $t_0 = q(\pi(a)+t\varepsilon)^{-1}(Q(a)-t)$,

(iii) $u_0 = q(\pi(a)+t\varepsilon)^{-1}\theta(a,u)+t_0u$,

(iv) $b_0 = b+a_0u$, *and*

(v) $s_0 = t_0q(u)+\phi(a_0,u)+g(b,a_0u)+s$.

Proof. By Proposition 11.8, there exist elements $(a_0,t_0),(b_0,s_0)\in\mathcal{R}$ and $u_0\in L$ such that $p^{m_1} = U_4x_1(a_0,t_0)x_2(u_0)x_3(b_0,s_0)$. By Proposition 11.6(ii), p is adjacent to the vertex

$$q_v := U_1x_2(\hat{z}_v)x_3(b+av,\hat{s}_v)x_4(v)$$

for all $v\in L$, where

$$\hat{z}_v = \theta(a,v)+tv+u \quad\text{and}\quad \hat{s}_v = tq(v)+\phi(a,v)+f(u,v)+s+g(b,av).$$

By Proposition 7.16(ii), the identity $f(\theta(a,v),v) = Q(a)q(v)$ and the fact that Q is identically zero if $\operatorname{char}(K)\neq 2$ (by [Weiss 2006b, 4.1(i)]), the element m_1 maps q_v to

$$U_1x_2(v)x_3(b+av,\hat{t}_v)x_4(-\hat{z}_v)$$

for all $v\in L$, where

$$\hat{t}_v = -tq(v)+\phi(a,v)+s+g(b,av)+Q(a)q(v).$$

Since p and q_v are adjacent for all v, so are their images under m_1. Therefore, by Proposition 11.6(i),

$$u_0 = \theta(a_0,\hat{z}_v)+t_0\hat{z}_v+v \tag{12.2}$$

as well as

$$b_0 = b+av+a_0\hat{z}_v \tag{12.3}$$

and

$$s_0 = t_0q(\hat{z}_v)+\phi(a_0,\hat{z}_v)+\hat{t}_v+f(\hat{z}_v,v)+g(b+av,\ a_0\hat{z}_v) \tag{12.4}$$

for all $v\in L$. Setting $v=0$ in (12.3) and (12.4), we conclude that (iv) and (v) hold. Setting $v=\varepsilon$ in (12.3), it follows from (iv) that

$$a+a_0(\pi(a)+t\varepsilon)=0.$$

Thus (i) holds.

We have $q(\pi(a)+(Q(a)-t)\varepsilon)=q(\pi(a)+t\varepsilon)$ and hence

$$\begin{aligned}\theta(a_0,\hat{z}_v)&=q(\pi(a)+t\varepsilon)^{-2}\theta\big(a\pi(a)+(Q(a)-t)a,\ \hat{z}_v\big)\\&=q(\pi(a)+t\varepsilon)^{-1}\theta(a,\hat{z}_v)\end{aligned}$$

for all $v\in L$ by (i) and Proposition A.4(i). By (12.2), therefore,

$$u_0=q(\pi(a)+t\varepsilon)^{-1}\theta(a,\hat{z}_v)+t_0\hat{z}_v+v \tag{12.5}$$

for all $v\in L$. Setting $v=0$ in (12.5), we conclude that (iii) holds. Setting $v=\varepsilon$ in (12.5), we obtain

$$u_0=q(\pi(a)+t\varepsilon)^{-1}\theta(a,\pi(a)+t\varepsilon+u)+t_0(\pi(a)+t\varepsilon+u)+\varepsilon.$$

By [Weiss 2006b, 4.21], we have $\theta(a,\pi(a))=Q(a)\pi(a)-q(\pi(a))\varepsilon$. By (iii), therefore,

$$\begin{aligned}-q(\pi(a)+t\varepsilon)t_0(\pi(a)+t\varepsilon)&=\theta(a,\pi(a))+t\pi(a)+q(\pi(a)+t\varepsilon)\varepsilon\\&=(t+Q(a))(\pi(a)+t\varepsilon).\end{aligned}$$

Since $\pi(a)+t\varepsilon\neq 0$, we conclude that (ii) holds. □

Corollary 12.6. *Let $(a,t)\in\mathcal{R}^\times$, $u\in L$ and $(b,s)\in\mathcal{R}$. Then the image of the vertex*

$$p:=U_4x_1(a,t)x_2(u)x_3(b,s)$$

of Λ under m_1m_4 is the vertex

$$U_4x_1(a_1,t_1)x_2(u_1)x_3(b_1,s_1),$$

where

(i) $a_1=b+a_0u$,

(ii) $t_1=-q(\pi(a)+t\varepsilon)^{-1}(t+Q(a))q(u)+\phi(a_0,u)+g(b,a_0u)+s$,

(iii) $u_1=-h(b+a_0u,a_0)+q(\pi(a)+t\varepsilon)^{-1}\big((t+Q(a))\bar{u}-\overline{\theta(a,u)}\big)$,

(iv) $b_1=-a_0$,

(v) $s_1=-q(\pi(a)+t\varepsilon)^{-1}(t+Q(a))$,

and $a_0=q(\pi(a)+t\varepsilon)^{-1}a\big(\pi(a)+(Q(a)-t)\varepsilon\big)$.

Proof. First apply Proposition 12.1 and then apply Proposition 7.16(i). □

Notation 12.7. Let (a,t), $(b,s)\in\mathcal{R}$ and $u\in L$. We set

$$P=\pi(a)+r\varepsilon\quad\text{and}\quad\rho=q(P),\quad\text{where } r=Q(a)-t.$$

Note that $\overline{P}=-(\pi(a)+t\varepsilon)$ and hence

$$\rho=q(\pi(a)+r\varepsilon)=q(\pi(a)+t\varepsilon). \tag{12.8}$$

We then set

$$v = \theta(a, u) + ru. \tag{12.9}$$

Thus $av = aPu$ by (D1) and

$$\begin{aligned} \theta(a, v) &= \theta(a, \theta(a, u)) + r\theta(a, u) \\ &= -q(\pi(a))u + (Q(a) + r)\theta(a, u) \\ &= -q(\pi(a))u - t\theta(a, u) \end{aligned} \tag{12.10}$$

by (12.9) and [Weiss 2006b, 4.1(i) and 4.21].

Proposition 12.11. *Let (a, t), u, P be as in Notation 12.7. Then*

$$h(aP, au) = \big(Q(a)t + 2q(\pi(a))\big)u + (Q(a) - 2t)\theta(a, u).$$

Proof. We have

$$\begin{aligned} h(aP, au) &= h(a, aPu) + f(h(aP, a), \varepsilon)u \\ &= h(a, av) + f(h(a, a), P)u, \end{aligned}$$

where v is as in Notation 12.7. Thus

$$\begin{aligned} h(aP, au) &= 2\big(\theta(a, v) + f(\pi(a), P)u\big) \\ &= 2\big(-q(\pi(a))u - t\theta(a, u) + 2q(\pi(a))u\big) \\ &= 2\big(q(\pi(a))u - t\theta(a, u)\big) \end{aligned}$$

by [Weiss 2006b, 4.1(i) and (iii) and 4.5(i)] if $\operatorname{char}(K) \neq 2$ and

$$\begin{aligned} h(aP, au) &= Q(a)(v + Q(a)u) \\ &= Q(a)(\theta(a, u) + tu) \end{aligned}$$

by [Weiss 2006b, 3.15 and 3.16] if $\operatorname{char}(K) = 2$. □

Proposition 12.12. *Let (a, t), u, P, v be as in Notation 12.7. Then*

$$h(aP, av) = \rho\big(Q(a)u + 2\theta(a, u)\big) = \rho h(a, au).$$

Proof. Replacing u by v in Proposition 12.11, we obtain the first equality. The second equality follows by [Weiss 2006b, 3.15, 3.16 and 4.5(i)]. □

Lemma 12.13. *Let (a, t), u, b, u_1 be as in Corollary 12.6. Then*

$$u_1 = \rho^{-1}\big(\overline{h(aP, b)} + t\bar{u} + \overline{\theta(a, u)}\big),$$

where P is as in Notation 12.7.

Proof. By Corollary 12.6(iii), we have

$$\bar{u}_1 = \rho^{-2}h(aP, \rho b + av) + \rho^{-1}\big((t + Q(a))u - \theta(a, u)\big),$$

where v is as in Notation 12.7. The claim holds, therefore, by Proposition 12.12. □

Lemma 12.14. $\phi(aP, u) = \phi(a, v) = \rho\phi(a, u).$

Proof. This holds by Proposition A.4(ii) and the last display on page 119 of [Tits and Weiss 2002]. □

Proposition 12.15. *Let* (a, t), u, (b, s), (a_1, t_1), u_1, (b_1, s_1) *be as in Corollary 12.6 and let* $r = Q(a) - t$, $P = \pi(a) + r\varepsilon$, $\rho = q(P)$ *and* $v = \theta(a, u) + ru$ *as in Notation 12.7. Then the following hold*:

(i) $a_1 = \rho^{-1}(\rho b + av)$.

(ii) $t_1 = \rho^{-1}\big(rq(u) + \phi(a, u) + g(b, av) + \rho s\big)$.

(iii) $\bar{u}_1 = \rho^{-1}\big(h(aP, b) + tu + \theta(a, u)\big)$.

(iv) $b_1 = -\rho^{-1}aP$.

(v) $s_1 = \rho^{-1}r$.

Proof. This holds by Lemma 12.13 and Lemma 12.14. □

13. Some more identities

In this section, we suppose that (a, t), u, (b, s), r, P, ρ, v are as in Notation 12.7 and Proposition 12.15 and assemble a number of identities we will need in the next few sections.

Proposition 13.1. $\theta(a, v) + tv = -\rho u.$

Proof. This holds by (12.9) and (12.10). □

Proposition 13.2. $q(h(b, av)) = \rho q(h(a, bu)) = \rho q(h(b, au)).$

Proof. By Proposition A.4(iii), we have

$$\rho h(a, bu) = \theta(a, h(aP, bu)) + rh(aP, bu).$$

We have $\rho = q(P)$. Hence

$$\begin{aligned}\rho^2 q(h(a, bu)) &= q\big(\theta(a, h(aP, bu))\big) + rf\big(\theta(a, h(aP, bu)),\ h(aP, bu)\big) \\ &\qquad + r^2 q(h(aP, bu)) \\ &= q(\pi(a))q(h(aP, bu)) + rQ(a)q(h(aP, bu)) + r^2 q(h(aP, bu)) \\ &= \rho q(h(aP, bu))\end{aligned}$$

by [Weiss 2006b, 4.9(i) and 4.22]. Thus $q(h(aP, bu)) = \rho q(h(a, bu))$. By (D1), therefore, $q(h(b, av)) = q(h(b, aPu) = q(h(aP, bu)) = \rho q(h(a, bu))$. The other equality holds by Proposition A.7. □

Proposition 13.3. $$\phi(a, P) = 0.$$

Proof. By Proposition A.4(ii) and [Weiss 2006b, (4.14)], $\phi(a, P) = \phi(aP, \varepsilon)$ and by [Weiss 2006b, 4.5(iii) and (4.13)], $\phi(aP, \varepsilon) = 0$. □

Proposition 13.4. $$q(v) = \rho q(u).$$

Proof. This holds by Proposition A.3(i). □

Proposition 13.5. $$Q(au) = Q(a)q(u).$$

Proof. By [Weiss 2006b, 4.1(i)], Q is identically zero if $\mathrm{char}(K) \neq 2$. The claim holds, therefore, by [Weiss 2006b, 3.21]. □

Proposition 13.6. $$Q(av) = \rho Q(a)q(u).$$

Proof. This holds by Propositions 13.4 and 13.5. □

Lemma 13.7.

$$f(\theta(a, v), w)v - f(w, v)\theta(a, v) = \rho\big(f(\theta(a, u), w)u - f(w, u)\theta(a, u)\big) \quad \textit{for all } w \in L.$$

Proof. This holds by (12.9) and (12.10). □

Lemma 13.8. $\theta(av, w) = \rho\theta(au, w)$ *for all* $w \in L$.

Proof. By (C4), Lemma 12.14, Proposition 13.4 and Lemma 13.7, we have

$$\begin{aligned}\overline{\theta(av, w)} &= \rho\big(q(u)\theta(a, \bar{w}) + \phi(a, u)\bar{w}\big) + f(\theta(a, v), \bar{w})v - f(v, \bar{w})\theta(a, v)\\ &= \rho\big(q(u)\theta(a, \bar{w}) + \phi(a, u)\bar{w} + f(\theta(a, u), \bar{w})u - f(v, \bar{w})\theta(a, u)\big)\\ &= \rho\overline{\theta(au, w)}\end{aligned}$$

for all $w \in L$. □

Proposition 13.9. *Let* $\omega = f\big(h(b, av), \pi(au)\big) + f(h(b, av), \varepsilon)\phi(a, u)$. *Then*

$$\omega = q(u)\big(f\big(h(a, b), \theta(a, v)\big) + Q(a)f(h(a, b), v)\big).$$

Proof. By [Weiss 2006b, 3.6] and Proposition A.1, we have

$$\begin{aligned}f\big(h(b, av), \pi(au)\big) &= -f\big(h(av, b), \overline{\pi(au)}\big)\\ &= -f\Big(h(av, b),\ q(u)\pi(a) - f(u, \varepsilon)\theta(a, u)\\ &\qquad\qquad + f(\theta(a, u), \varepsilon)v + \phi(a, u)\varepsilon\Big)\end{aligned}$$

and thus

$$\begin{aligned}\omega &= -q(u)f(h(av, b), \pi(a)) + f(u, \varepsilon)f(h(av, b), \theta(a, u))\\ &\qquad - f(\theta(a, u), \varepsilon)f(h(av, b), u)\\ &= -q(u)f(h(av\pi(a), b), \varepsilon) + f(u, \varepsilon)f(h(av\theta(a, u), b), \varepsilon)\\ &\qquad - f(\theta(a, u), \varepsilon)f(h(avu, b), \varepsilon).\end{aligned} \tag{13.10}$$

By (A3), (D1) and [Weiss 2006b, 3.8 and 4.9(i)], we have

$$
\begin{aligned}
(13.11)\qquad av\pi(a) &= a\big(f(v,\varepsilon)\varepsilon - \bar{v}\big)\pi(a)\\
&= f(v,\varepsilon)a\pi(a) - a\bar{v}\pi(a)\\
&= f(v,\varepsilon)a\pi(a) + a\overline{\pi(a)}v - f(\pi(a),v)a\\
&= f(v,\varepsilon)a\pi(a) - a\pi(a)v + Q(a)av - f(\pi(a),v)a\\
&= f(v,\varepsilon)a\pi(a) - a\theta(a,v) + Q(a)av - f(\pi(a),v)a
\end{aligned}
$$

and

$$
\begin{aligned}
(13.12)\qquad avu &= -av\bar{u} + f(u,\varepsilon)av\\
&= -q(u)aP + f(u,\varepsilon)av
\end{aligned}
$$

as well as

$$
\begin{aligned}
av\theta(a,u) &= f\big(\theta(a,u),\varepsilon\big)av - av\overline{\theta(a,u)}\\
&= f\big(\theta(a,u),\varepsilon\big)av + a\theta(a,u)\bar{v} - f\big(\theta(a,u),v\big)a\\
&= f\big(\theta(a,u),\varepsilon\big)av + a\theta(a,u)\big(\overline{\theta(a,u)} + r\bar{u}\big) - f\big(\theta(a,u),v\big)a\\
&= f\big(\theta(a,u),\varepsilon\big)av + q(\theta(a,u))a + ra\theta(a,u)\bar{u} - f\big(\theta(a,u),v\big)a\\
&= f\big(\theta(a,u),\varepsilon\big)av + q(\pi(a))q(u)a + rq(u)a\pi(a) - f\big(\theta(a,u),v\big)a.
\end{aligned}
$$

Since

$$
\begin{aligned}
f\big(\theta(a,u),v\big) &= f\big(\theta(a,u),\theta(a,u)\big) + rf\big(\theta(a,u),u\big)\\
&= 2q(\theta(a,u)) + Q(a)q(u)r = q(u)\big(2q(\pi(a)) + Q(a)\big),
\end{aligned}
$$

it follows that

$$
(13.13)\qquad av\theta(a,u) = f\big(\theta(a,u),\varepsilon\big)av - q(\pi(a))q(u)a - ra\overline{\pi(a)}.
$$

By (13.10)–(13.13) and some calculation, we conclude that

$$
\omega = q(u)\Big(f\big(h(a\theta(a,v),b),\varepsilon\big) + Q(a)f(h(av,b),\varepsilon)\Big).
$$

The claim follows now by (B3). □

Lemma 13.14. $b\theta(a,v)\bar{v} = \rho b\theta(a,u)\bar{u}.$

Proof. We have

$$
\begin{aligned}
&b\theta(a,v)\bar{v}\\
&\quad= -b\big(q(\pi(a))u + t\theta(a,u)\big)\big(\overline{\theta(a,u)} + r\bar{u}\big)\\
&\quad= -\big(q(u)rq(\pi(a)) + tq(\theta(a,u))\big)b - q(\pi(a))bu\overline{\theta(a,u)} - rtb\theta(a,u)\bar{u}\\
&\quad= Q(a)q(u)q(\pi(a))b + \big(q(\pi(a)) - rt\big)b\theta(a,u)\bar{u} - q(\pi(a))f(\theta(a,u),u)b\\
&\quad= \big(q(\pi(a)) - rt\big)b\theta(a,u)\bar{u} = \rho\theta(a,u)\bar{u}
\end{aligned}
$$

by (12.9) and (12.10). □

Lemma 13.15. *The following holds*:

$$avh(av,b)=\rho\big(q(u)ah(a,b)+q(u)b\pi(a)-b\theta(a,u)\bar{u}\big).$$

Proof. By [Weiss 2006b, 3.22], we have

$$\begin{aligned}avh(av,b)&=-a\overline{h(av,b)}\bar{v}+f(\overline{h(av,b)},v)a\\&=ah(b,av)\bar{v}-f(h(b,av),v)a\\&=ah(a,bv)\bar{v}+f(h(b,a),\varepsilon)av\bar{v}-f(h(b,av\bar{v}),\varepsilon)a\\&=ah(a,bv)\bar{v}\\&=ah(a,b)v\bar{v}+\big(b\pi(a)v-b\theta(a,v)\big)\bar{v}\\&=q(v)ah(a,b)+q(v)b\pi(a)-b\theta(a,v)\bar{v}.\end{aligned}$$

The claim holds, therefore, by Proposition 13.4 and Lemma 13.14. □

Lemma 13.16. $f\big(\theta(a,u),h(aP,b)\big)=f\big(h(a,b),\theta(a,v)\big).$

Proof. By [Weiss 2006b, 4.20], we have

$$\begin{aligned}aP\theta(a,u)&=a\pi(a)\theta(a,u)+ra\theta(a,u)\\&=-a\overline{\pi(a)}\theta(a,u)-ta\theta(a,u)\\&=-q(\pi(a))au-ta\theta(a,u)=a\theta(a,v).\end{aligned}$$

The claim follows by (B3). □

14. The form Θ

In this section, we prove the following result:

Proposition 14.1. *Let* (a,t), u, (b,s), (a_1,t_1), ρ, r, v *be as in Proposition 12.15 and let*

$$\Theta(a,t,u,b,s)=\rho q(\pi(a_1)+t_1\varepsilon). \tag{14.2}$$

Then

$$\begin{aligned}\Theta(a,t,u,b,s)=\rho q(\pi(b)+s\varepsilon)&+q(h(a,bu))+q(u)^2\\&-q(u)f(h(a,b),u)+\big(sQ(a)+tQ(b)\big)q(u)\\&+f(u,\varepsilon)f\big(\theta(a,u),\pi(b)\big)-f(\theta(a,u),\varepsilon)f(u,\pi(b))\\&-f\big(h(a,b),\overline{\theta(b,\bar{v})}+sv\big)\\&-q(u)f\big(\pi(a),\pi(b)\big)-2stq(u).\end{aligned} \tag{14.3}$$

Proof. Let

$$z = \pi(\rho b + av) + \rho\big(rq(u) + \phi(a,u) + g(b,av) + \rho s\big)\varepsilon.$$

Thus

$$\rho^3\Theta(a,t,u,b,s) = \rho^4 q(\pi(a_1) + t_1\varepsilon) = q(z)$$

by Proposition 12.15. By (C3) and Lemma 13.8, we have

$$\begin{aligned} z &= \rho^2\pi(b) + \rho h(b,av) + \pi(av) + \rho\big(\phi(a,u) + \rho s + rq(u)\big)\varepsilon \\ &= \rho\big(\rho\pi(b) + h(b,av) + \pi(au) + \big(\phi(a,u) + \rho s + rq(u)\big)\varepsilon\big). \end{aligned}$$

By Proposition A.3(ii), we have

$$q\big(\pi(au) + rq(u) + \phi(a,u)\varepsilon\big) = \rho q(u)^2.$$

By [Weiss 2006b, 4.5(iii)], Proposition 13.2 and Proposition 13.6, it follows that

$$\begin{aligned} \rho^{-2}q(z) = {} & \rho^2 q(\pi(b)) + \rho q(h(b,au)) + \rho q(u)^2 + \rho^2 s^2 \\ & + \rho s Q(a)q(u) + 2\rho rsq(u) \\ & + \rho^2 Q(b)s + \rho f\big(\pi(b), h(b,av)\big) \\ & + \rho f\big(\pi(b), \pi(au)\big) + \rho Q(b)\big(\phi(a,u) + rq(u)\big) \\ & + f\big(h(b,av), \pi(au)\big) + f\big(h(b,av), \varepsilon\big)\phi(a,u) \\ & + \big(\rho s + rq(u)\big) f\big(h(b,av), \varepsilon\big). \end{aligned} \tag{14.4}$$

By Propositions 13.1 and 13.9, we have

$$\begin{aligned} & f\big(h(b,av), \pi(au)\big) + f\big(h(b,av), \varepsilon\big)\phi(a,u) + rq(u) f\big(h(b,av), \varepsilon\big) \\ & \qquad = q(u)\big(f\big(h(a,b), \theta(a,v)\big) + Q(a) f(h(a,b), v) - rf(h(a,b), v)\big) \\ & \qquad = -\rho q(u) f\big(h(a,b), u\big). \end{aligned}$$

Replacing $2\rho rsq(u)$ by $-2\rho stq(u)$ in (14.4), we thus conclude that

$$\begin{aligned} \Theta = \rho^{-3}q(z) = {} & \rho q(\pi(b) + s\varepsilon) + q(h(a,bu)) + rq(u)Q(b) + Q(a)q(u)s \\ & - 2stq(u) + f\big(\pi(b), h(b,av)\big) + sf\big(h(b,av), \varepsilon\big) \\ & + f\big(\pi(b), \pi(au)\big) + \phi(a,u)Q(b) \\ & + q(u)^2 - q(u) f\big(h(a,b), u\big). \end{aligned} \tag{14.5}$$

By (B3), (D1) and [Weiss 2006b, 3.6 and 3.7], we have

$$\begin{aligned} f\big(\pi(b), h(b,av)\big) + sf\big(h(b,av), \varepsilon\big) &= f\big(h(b(\pi(b) + s\varepsilon)\bar{v}, a), \varepsilon\big) \\ &= -f\big(h(a,b), \overline{\theta(b,\bar{v})} + sv\big) \end{aligned}$$

and by (C4), we have

$$\begin{aligned} &f\bigl(\pi(b), \pi(au)\bigr) + Q(b)\phi(a,u)\\ &= f\bigl(\overline{\pi(b)},\ q(u)\pi(a) - f(u,\varepsilon)\theta(a,u) + f(\theta(a,u),\varepsilon)u\bigr)\\ &= -q(u)f(\pi(a),\pi(b)) + q(u)Q(a)Q(b) + f(u,\varepsilon)f(\pi(b),\theta(a,u))\\ &\qquad\qquad\qquad\qquad - f(\theta(a,u),\varepsilon)f(\pi(b),u). \end{aligned}$$

By (14.5), therefore, $\Theta(a,t,u,b,s)$ is as in (14.3). □

Remark 14.6. Note that a_1 and t_1 are only defined under the assumption that ρ is nonzero. In particular, we cannot conclude from (14.2) that $\rho = 0$ implies that the expression $\Theta(a,t,u,b,s)$ in (14.3) is zero (and, in fact, this is clearly not true).

15. The smallest F_4-cases

We interrupt our calculations to make a few remarks about the case that $\Delta = F_4(C,K)$ and either $\mathrm{char}(K) \neq 2$ and $C = K$ or $\mathrm{char}(K) = 2$ and C/K is a field extension such that $C^2 \subset K$. By Theorem 7.4(vi), $\dim_K L = 5$ if $\mathrm{char}(K) \neq 2$ and f is degenerate if $\mathrm{char}(K) = 2$. If $\mathrm{char}(K) \neq 2$, then by [Mühlherr and Weiss 2019, 5.10], Ξ is isotopic to the quadrangular algebra $\mathcal{Q}_4(C,K)$ defined in [loc. cit, 4.12] and hence by [loc. cit, 4.13 and 10.5], $q(\pi(a)) = 0$ and $a\pi(a) = 0$ for all $a \in \mathcal{X}$. (If the composition of q and π is identically zero in one isotope of Ξ, then by [Mühlherr and Weiss 2019, 5.3; Weiss 2006b, 8.7], it is identically zero in all isotopes of Ξ.) By [Mühlherr and Weiss 2019, 8.1(i), 8.5 and 9.2], we have $Q(a) = 0$ and $f(h(a,b),u) = 0$ for all $a, b \in \mathcal{X}$ and all $u \in L$ if $\mathrm{char}(K) = 2$.

Proposition 15.1. *Suppose that* $\mathrm{char}(K) \neq 2$ *and* $a\pi(a) = 0$ *for all* $a \in \mathcal{X}$ *and let* $\Theta = \Theta(a,t,u,b,s)$ *be as in* (14.3). *Then*

$$\Theta = \bigl(st + \tfrac{1}{2} f(h(a,b),u) - q(u)\bigr)^2$$

for all $(a,t) \in \mathcal{R}^\times$, $u \in L$ *and* $(b,s) \in \mathcal{R}$.

Proof. Choose $(a,t) \in \mathcal{R}^\times$, $u \in L$ and $(b,s) \in \mathcal{R}$. By (14.2), we have $\Theta = (tt_1)^2$, where

$$t_1 = t^{-1}\bigl(st - g(b,au) - q(u)\bigr).$$

The claim holds, therefore, since

$$\begin{aligned} g(b,au) &= \tfrac{1}{2} f\bigl(h(b,au),\varepsilon\bigr) && \text{by [Weiss 2006b, 4.1(i) and (4.3)]}\\ &= \tfrac{1}{2} f\bigl(\overline{h(b,au)},\bar{\varepsilon}\bigr) && \text{by [Weiss 2006b, 1.4]}\\ &= -\tfrac{1}{2} f\bigl(h(au,b),\varepsilon\bigr) && \text{by [Weiss 2006b, 3.6]}\\ &= -\tfrac{1}{2} f\bigl(h(a,b),u\bigr) && \text{by (B3).} \end{aligned}$$

□

Remark 15.2. Suppose that $\mathrm{char}(K) \neq 2$ and that $a\pi(a) = 0$ for all $a \in \mathcal{X}$. By (A3), we have $q(\pi(a)) = 0$ for all $a \in \mathcal{X}$. By (B3), we have $f\big(\pi(b), h(b, av)\big) = f(\big(h(b\pi(b), av), \varepsilon\big) = 0$. Similarly, we have

$$f\big(\pi(au), h(au, b)\big) = f\big(h(au\pi(au), b), \varepsilon\big) = 0$$

and $f\big(\pi(b), h(au, b)\big) = -f\big(h(au, b\pi(b), \varepsilon\big) = 0$. By [Weiss 2006b, 4.1(i) and 4.3], we have

$$g(au, b) = \tfrac{1}{2} f\big(h(a, b), u\big)$$

and by Proposition A.7, $q(h(au, b)) = q(h(a, bu))$. Applying (C3) to $\pi(au + b)$ and then applying q to both sides of this equation, we find thus that

$$q\big(h(a, bu)\big) + f\big(\pi(au), \pi(b)\big) = \tfrac{1}{4} f\big(h(a, b), u\big)^2.$$

With these observations, Proposition 15.1 can be confirmed directly from (14.5).

Proposition 15.3. *Suppose that $f(h(a, b), X) = 0$ and $Q(a) = 0$ for all $a, b \in \mathcal{X}$ and that* $\mathrm{char}(K) = 2$ *and let $\Theta = \Theta(a, t, u, b, s)$ be as in (14.3). Then*

$$\Theta = q(\pi(a) + t\varepsilon)q(\pi(b) + s\varepsilon) + q(h(a, bu)) + f\big(\pi(au), \pi(b)\big) + q(u)^2$$

for all $(a, t) \in \mathcal{R}^\times$, $u \in L$ and $(b, s) \in \mathcal{R}$.

Proof. This holds by (14.5). □

16. The element $(m_1 m_4)^2$

Let $(a, t), (a_1, t_1) \in \mathcal{R}^\times$, $u \in L$ and $(b, s) \in \mathcal{R}$ be as in Proposition 12.15 and let $\Theta = \Theta(a, t, u, b, s)$ be as in Proposition 14.1.

Notation 16.1. Suppose that $\Theta \neq 0$. Thus by (14.2), $q(\pi(a_1) + t_1\varepsilon) \neq 0$. By Proposition 12.15 applied to the image of the vertex p under $m_1 m_4$, it follows that there exist $(a_2, t_2), (b_2, s_2) \in \mathcal{R}$ and $u_2 \in L$ such that

$$U_4 x_1(a_2, t_2) x_2(u_2) x_3(b_2, s_2)$$

is the image of the vertex

$$p := U_4 x_1(a, t) x_2(u) x_3(b, s)$$

of Λ under $(m_1 m_4)^2$.

The expressions we obtain for $a_2, \ldots, s_2$ by applying Proposition 12.15 have $\rho^{-1}\Theta$ in the denominator and, in various places, negative powers of ρ appear in the numerator. Our goal for the rest of this section is to obtain expressions for s_2 and b_2 in which all the negative powers of ρ have been eliminated. In the following two sections, we do the same for u_2.

Notation 16.2. Let $r = Q(a) - t$, $P = \pi(a) + r\varepsilon$, $\rho = q(P)$ and $v = \theta(a, u) + ru$ as in Notation 12.7.

Proposition 16.3. *Let s_2 be as in Notation 16.1. Then*

$$s_2 = \frac{tq(u) - \rho s + g(av, b) + \phi(a, u) + \rho Q(b)}{\Theta}.$$

Proof. By Proposition 12.15 and (14.2), we have $s_2 = z/\Theta$, where

$$z = -rq(u) + \phi(a, u) - g(b, av) - \rho s + \rho^{-1} Q(\rho b + av).$$

If $\mathrm{char}(K) \neq 2$, then ϕ and Q are identically zero and by [Weiss 2006b, 4.4 and 4.5(ii)], g is alternating, so the claim holds. We can assume, therefore, that $\mathrm{char}(K) = 2$. Then

$$\begin{aligned} \rho^{-1} Q(\rho b + av) &= \rho Q(b) + \rho^{-1} Q(av) + f(h(b, av), \varepsilon) \\ &= \rho Q(b) + Q(a)q(u) + f(h(b, av), \varepsilon) \end{aligned}$$

by (C3) and Proposition 13.6. By Notation 2.2 and [Weiss 2006b, 4.3], we have

$$\begin{aligned} g(b, av) + f(h(b, av), \varepsilon) &= f(h(b, av), \delta + \varepsilon) \\ &= f(h(b, av), \bar{\delta}) \\ &= f(h(av, b), \delta) = g(av, b), \end{aligned}$$

where δ is as in Theorem 7.4(i). Thus the claim holds also in this case. □

Proposition 16.4. *Let b_2 be as in Notation 16.1. Then*

$$b_2 = \frac{y}{\Theta},$$

where

$$y = q(u)a\big(u - h(a, b)\big) + (\rho b + av)\big(\overline{\pi(b)} + s\varepsilon\big) - b\bar{u}\theta(a, u) - bh(av, b) - tq(u)b$$

Proof. By Proposition 12.15 and (14.2), we have

$$b_2 = \frac{-(\rho b + av)z}{\rho^2 \Theta}, \tag{16.5}$$

where

$$z = \pi(\rho b + av) + \big(Q(\rho b + av) - \rho r q(u) + \rho\phi(a, u) - \rho g(b, av) - \rho^2 s\big)\varepsilon.$$

By (C3), we have

$$\pi(\rho b + av) = \rho^2 \pi(b) + \pi(av) + \rho\big(h(b, av) - g(b, av)\varepsilon\big).$$

If $\mathrm{char}(K) \neq 2$, then Q is identically zero and

$$2g(b, av) = f(h(b, av), \varepsilon)$$

by [Weiss 2006b, 4.1(iii) and 4.3]. If $\mathrm{char}(K)=2$, then

$$Q(\rho b+av)=\rho^2 Q(b)+Q(av)+\rho f(h(b,av),\varepsilon).$$

Thus

$$\begin{aligned}\pi(\rho b+av)&+Q(\rho b+av)\varepsilon-\rho g(b,av)\varepsilon\\ &=\rho^2\pi(b)+\pi(av)+\rho h(b,av)+\big(\rho^2 Q(b)+Q(av)-\rho f(h(b,av),\varepsilon)\big)\varepsilon\\ &=\rho^2\pi(b)+\pi(av)-\rho\overline{h(b,av)}+\big(\rho^2 Q(b)+Q(av)\big)\varepsilon\\ &=\rho^2\pi(b)+\pi(av)+\rho h(av,b)+\big(\rho^2 Q(b)+Q(av)\big)\varepsilon\end{aligned}$$

in all characteristics. By Proposition 13.6, we have

$$Q(av)-\rho rq(u)=\rho(Q(a)-r)q(u)=\rho tq(u).$$

Thus

$$z=\rho^2\pi(b)+\pi(av)+\rho h(av,b)+\big(\rho^2 Q(b)+\rho tq(u)+\rho\phi(a,u)-\rho^2 s\big)\varepsilon. \tag{16.6}$$

Hence

$$\frac{\rho bz}{\rho^2\Theta}=\frac{bz_1}{\Theta}+\frac{b\pi(av)}{\rho\Theta}, \tag{16.7}$$

where

$$z_1=\rho\pi(b)+h(av,b)+\big(\rho Q(b)+tq(u)+\phi(a,u)-\rho s\big)\varepsilon.$$

By Lemma 12.14, and Propositions 13.4 and A.2, we have

$$\begin{aligned}av\pi(av)&=q(v)a\theta(a,v)+\phi(a,v)av\\ &=\rho\big(q(u)a\theta(a,v)+\phi(a,u)av\big)\end{aligned}$$

and by Proposition 13.1, we have

$$a\theta(a,v)+tav=-\rho au.$$

By (16.6), therefore,

$$\frac{avz}{\rho^2\Theta}=\frac{av\pi(b)+(Q(b)-s)av-q(u)au}{\Theta}+\frac{avh(av,b)}{\rho\Theta}. \tag{16.8}$$

By Lemma 13.8, we have $b\pi(av)=\rho b\pi(au)$. By Lemma 13.15, and Equations (16.5), (16.7) and (16.8), it follows that

$$b_2=-\frac{y_1}{\Theta},$$

where

$$y_1 = \rho b\pi(b) + bh(av, b) + \big(\rho Q(b) + tq(u) + \phi(a, u) - \rho s\big)b + b\pi(au) \\ + av\pi(b) + (Q(b) - s)av - q(u)au + q(u)ah(a, b) \\ + q(u)b\pi(a) - b\theta(a, u)\bar{u}.$$

Next note that

$$b\pi(au) + \phi(a, u)b = q(u)b\overline{\pi(a)} - f(u, \varepsilon)b\overline{\theta(a, u)} + f\big(\theta(a, u), \varepsilon\big)b\bar{u}$$

by Proposition A.1 and

$$-f(u, \varepsilon)b\overline{\theta(a, u)} + f\big(\theta(a, u), \varepsilon\big)b\bar{u} - b\theta(a, u)\bar{u} \\ = -f(u, \varepsilon)b\overline{\theta(a, u)} + b\overline{\theta(a, u)}\bar{u} \\ = -b\overline{\theta(a, u)}u.$$

Hence

$$b\pi(au) + \phi(a, u)b - b\theta(a, u)\bar{u} + q(u)b\pi(a) = q(u)Q(a)b - b\overline{\theta(a, u)}u \\ = b\bar{u}\theta(a, u)$$

by [Weiss 2006b, 4.9(i)]. Therefore

$$b_2 = -\frac{y_2}{\Theta},$$

where

$$y_2 = q(u)a\big(h(a, b) - u\big) + av\pi(b) + (Q(b) - s)av + b\bar{u}\theta(a, u) \\ + \rho b\pi(b) + bh(av, b) + \big(\rho(Q(b) - s) + tq(u)\big)b$$

Since

$$b\pi(b) + Q(b)b - sb = -b(\overline{\pi(b)} + s\varepsilon)$$

and

$$av\pi(b) + Q(b)av - sav = -av(\overline{\pi(b)} + s\varepsilon),$$

the claim holds. □

17. The element u_2, part I

Our next goal is to prove Proposition 18.8. In this section, we begin the proof. We continue to assume that (a, t), (b, s), (a_1, t_1), u, $r = Q(a) - t$, $P = \pi(a) + r\varepsilon$, $\rho = q(P)$ and $v = \theta(a, u) + ru$ are as in Proposition 12.15, that $\Theta = \Theta(a, t, u, b, s)$ is as in Proposition 14.1, that $\Theta \neq 0$ and that u_2 is as in Notation 16.1.

Proposition 17.1. *The following holds*:

$$\begin{aligned}h\big(aP, bh(av,b)\big) = \rho\big(h(b, ah(au,b)) + f(h(a,b),\varepsilon)h(au,b)\big)\\ + f\big(h(a,b),v\big)h(a\pi(a),b)\\ + \big(f\big(h(a,b),\theta(a,v)\big) + Q(a)f(h(a,b),v)\big)h(a,b).\end{aligned}$$

Proof. We have

$$\begin{aligned}&h(aP, bh(av,b))\\ &\qquad = h(b, aPh(av,b)) + f(h(aP,b),\varepsilon)h(av,b)\\ &\qquad = h\big(b, a\theta(a, h(av,b))\big) + rh\big(b, ah(av,b)\big) + f\big(h(a,b),P\big)h(av,b)\end{aligned}$$

and by Proposition A.6(iii), we have

$$\begin{aligned}h\big(b, a\theta(a, h(av,b))\big) = -f\big(h(a,b),\varepsilon\big)h\big(b, a\theta(a,v)\big) - f\big(h(b,a),\pi(a)\big)h(b,av)\\ + Q(a)h\big(b, a\overline{h(av,b)}\big) + h\big(b, a\overline{h(a\theta(a,v),b)}\big)\\ + f\big(h(a,b),v\big)h\big(b, a\pi(a)\big).\end{aligned}$$

Thus by (12.9) and (12.10), we have

$$\begin{aligned}&h\big(b, a\theta(a, h(av,b))\big)\\ &= q(\pi(a))f\big(h(a,b),\varepsilon\big)h(b,au) + tf\big(h(a,b),\varepsilon\big)h\big(b, a\theta(a,u)\big)\\ &\qquad - f\big(h(b,a),\pi(a)\big)h\big(b, a\theta(a,u)\big) - rf\big(h(b,a),\pi(a)\big)h(b,au)\\ &\qquad + Q(a)h\big(b, a\overline{h(a\theta(a,u),b)}\big) + Q(a)rh\big(b, a\overline{h(au,b)}\big)\\ &\quad - q(\pi(a))h\big(b, a\overline{h(au,b)}\big) - th\big(b, a\overline{h(a\theta(a,u),b)}\big) + f\big(h(a,b),v\big)h\big(b, a\pi(a)\big)\end{aligned}$$

as well as

$$rh\big(b, ah(av,b)\big) = rh\big(b, ah(a\theta(a,u),b)\big) + r^2h\big(b, ah(au,b)\big)$$

and

$$\begin{aligned}&f\big(h(a,b),P\big)h(av,b)\\ &\quad = N\big(h(a\theta(a,u),b) + rh(au,b)\big)\\ &\quad = N\big(h(b, a\theta(a,u)) + f\big(h(a\theta(a,u),b),\varepsilon\big)\varepsilon + rh(b,au) + rf(h(au,b),\varepsilon)\varepsilon\big)\\ &\quad = N\big(h(b, a\theta(a,u)) + f(h(a,b),v)\varepsilon + rh(b,au)\big),\end{aligned}$$

where

$$N := f\big(h(a,b),\pi(a)\big) + rf\big(h(a,b),\varepsilon\big).$$

Next we observe that

$$\begin{aligned}Q(a)h\big(b, a\overline{h(a\theta(a,u),b)}\big) - th\big(b, a\overline{h(a\theta(a,u),b)}\big) + rh\big(b, ah(a\theta(a,u),b)\big)\\ = rf\big(h(a\theta(a,u),b),\varepsilon\big)h(b,a)\\ = rf\big(h(a,b),\theta(a,u)\big)h(b,a)\\ = rf\big(h(a,b),\theta(a,u)\big)\big(h(a,b) - f(h(a,b),\varepsilon)\varepsilon\big)\end{aligned}$$

and

$$\begin{aligned}\big(q(\pi(a))f(h(a,b),\varepsilon) - rf\big(h(b,a),\pi(a)\big) + Nr\big)h(b,au)\\ = \big(q(\pi(a))f(h(a,b),\varepsilon) + rf\big(h(a,b),\overline{\pi(a)}\big) + Nr\big)h(b,au)\\ = \big(q(\pi(a)) + rQ(a) + r^2\big)f\big(h(a,b),\varepsilon\big)h(b,au)\\ = \rho f(h(a,b),\varepsilon)h(b,au)\\ = \rho f(h(a,b),\varepsilon)\big(h(au,b) - f(h(au,b),\varepsilon)\varepsilon\big)\end{aligned}$$

as well as

$$\begin{aligned}&Q(a)rh\big(b, a\overline{h(au,b)}\big) - q(\pi(a))h\big(b, a\overline{h(au,b)}\big) + r^2h\big(b, ah(au,b)\big)\\ &= \big(Q(a)r + q(\pi(a)) + r^2\big)h\big(b, ah(au,b)\big)\\ &\qquad + \big(Q(a)r - q(\pi(a))\big)f\big(h(au,b),\varepsilon\big)h(b,a)\\ &= \rho h\big(b, ah(au,b)\big) + \big(Q(a)r - q(\pi(a))\big)f\big(h(a,b),u\big)h(a,b)\\ &\qquad - \big(Q(a)r - q(\pi(a))\big)f\big(h(a,b),u\big)f\big(h(a,b),\varepsilon\big)\varepsilon.\end{aligned}$$

We also have

$$\begin{aligned}\big(tf\big(h(a,b),\varepsilon\big) - f\big(h(b,a),\pi(a)\big) + N\big)h\big(b, a\theta(a,u)\big)\\ = \big(tf\big(h(a,b),\varepsilon\big) + f\big(h(a,b),\overline{\pi(a)}\big) + N\big)h\big(b, a\theta(a,u)\big)\\ = \big(t + r + Q(a)\big)f\big(h(a,b),\varepsilon\big)h\big(b, a\theta(a,u)\big) = 0\end{aligned}$$

and

$$\begin{aligned}f\big(h(a,b),v\big)h\big(b, a\pi(a)\big) + f\big(h(a,b),v\big)f\big(h(a,b),\pi(a)\big)\varepsilon\\ = f\big(h(a,b),v\big)\big(h\big(b,a\pi(a)\big) - f\big(h(b,a\pi(a)),\varepsilon\big)\varepsilon\big)\\ = f\big(h(a,b),v\big)h\big(a\pi(a),b\big).\end{aligned}$$

Finally, we observe that

$$rv - r\theta(a,u) + \big(Q(a)r + q(\pi(a))\big)u = \rho u$$

by (12.9) and (12.10) and hence

$$rf\big(h(a,b),v\big)f\big(h(a,b),\varepsilon\big)-rf\big(h(a,b),\theta(a,u)\big)f\big(h(a,b),\varepsilon\big)\\ +\big(Q(a)r+q(\pi(a))\big)f\big(h(a,b),u\big)f\big(h(a,b),\varepsilon\big)\\ =\rho f\big(h(a,b),u\big)f\big(h(a,b),\varepsilon\big).$$

Assembling all these calculations, we obtain the desired formula. □

Proposition 17.2. *The following holds*:

$$h\big(aP,bh(b,av)\big)=\rho\big(h\big(b,ah(b,au)\big)+f\big(h(a,b),\varepsilon\big)h(b,au)\big).$$

Proof. We have

$$\begin{aligned}h\big(aP,bh(b,av)\big)&=h\big(aP,bh(av,b)\big)+f\big(h(b,av),\varepsilon\big)h(aP,b)\\&=h\big(aP,bh(av,b)\big)-f\big(h(a,b),v\big)h(aP,b)\\&=h\big(aP,bh(av,b)\big)-f\big(h(a,b),v\big)\big(h(a\pi(a),b)+rh(a,b)\big)\end{aligned}$$

and thus

$$h\big(aP,bh(b,av)\big)=\rho\big(h\big(b,ah(au,b)\big)+f\big(h(a,b),\varepsilon\big)h(au,b)\big)\\ +\big(f\big(h(a,b),\theta(a,v)\big)+tf\big(h(a,b),v\big)\big)h(a,b).$$

by Proposition 17.1. By Proposition 13.1, we have

$$f\big(h(a,b),\theta(a,v)\big)+tf\big(h(a,b),v\big)=-\rho f\big(h(a,b),u\big).$$

Hence

$$\begin{aligned}&h\big(aP,bh(b,av)\big)\\&\quad=\rho\big(h\big(b,ah(au,b)\big)+f\big(h(a,b),\varepsilon\big)h(au,b)-f\big(h(a,b),u\big)h(a,b)\big)\\&\quad=\rho\big(h\big(b,ah(au,b)\big)+f\big(h(a,b),\varepsilon\big)h(b,au)-f\big(h(au,b),\varepsilon\big)h(b,a)\big)\\&\quad=\rho\big(f\big(h(a,b),\varepsilon\big)h(b,au)-h\big(b,a\overline{h(au,b)}\big)\big)\end{aligned}$$

and so the claim holds. □

Proposition 17.3. *The following holds*:

$$h\big(aP,b\bar{u}\theta(a,u)\big)=-q(\pi(a))q(u)h(a,b)-f\big(h(a,b),\theta(a,v)\big)u\\ -tq(u)h\big(a\pi(a),b\big)+f\big(h(a,b),v\big)\theta(a,u).$$

Proof. We have

$$\begin{aligned}
h\big(aP, b\bar{u}\theta(a,u)\big) &= h\big(b\bar{u}, aP\theta(a,u)\big) + f\big(h(aP,b\bar{u}),\varepsilon\big)\theta(a,u)\\
&= h\big(b\bar{u}, a\big(\theta(a,\theta(a,u)) + r\theta(a,u)\big)\big) + f\big(h(a,b),v\big)\theta(a,u)\\
&= h\big(b\bar{u}, a\big((Q(a)+r)\theta(a,u) - q(\pi(a))u\big)\big) + f\big(h(a,b),v\big)\theta(a,u)\\
&= (Q(a)+r)h\big(a\pi(a), b\bar{u}u\big) + (Q(a)+r)f\big(h(b\bar{u},a\pi(a)),\varepsilon\big)u\\
&\qquad - q(\pi(a))h(a,b\bar{u}u) - q(\pi(a))f(h(b\bar{u},a),\varepsilon)u\\
&\qquad + f\big(h(a,b),v\big)\theta(a,u)\\
&= -tq(u)h\big(a\pi(a),b\big) + tf\big(h(a\pi(a),b),u\big)u\\
&\qquad - q(u)q(\pi(a))h(a,b) + q(\pi(a))f(h(a,b),u)u\\
&\qquad + f\big(h(a,b),v\big)\theta(a,u)\\
&= -tq(u)h\big(a\pi(a),b\big) - f\big(h(a,b),\theta(a,v)\big)u\\
&\qquad - q(u)q(\pi(a))h(a,b) + f\big(h(a,b),v\big)\theta(a,u).
\end{aligned}$$

□

Proposition 17.4. *The following holds*:

$$h\big(aP, ah(a,b)\big) = \big(Q(a)+2t\big)h(a\pi(a),b) + \big(Q(a)r + 2q(\pi(a))\big)h(a,b).$$

Proof. Set $u = h(a,b)$ in Proposition 12.11 and then apply Proposition A.6(i). □

Proposition 17.5. *The following holds*:

$$h\big(aP, bh(b,aP)u\big) = \rho h\big(b\bar{u}, ah(a,b)\big).$$

Proof. We have

$$\begin{aligned}
h(av, b\bar{u}) = h(aPu, b\bar{u}) &= h(b, aPu\bar{u}) + f\big(h(aPu,b),\varepsilon\big)\bar{u}\\
&= q(u)h(b,aP) + f\big(h(a,b),v\big)\bar{u}
\end{aligned}$$

and thus

$$q(u)h(b,aP) = h(av,b\bar{u}) - f\big(h(a,b),v\big)\bar{u}.$$

Hence

$$\begin{aligned}
q(u)h\big(aP, bh(b,aP)u\big) &= h\big(aP, bh(av,b\bar{u})u\big) - q(u)f\big(h(a,b),v\big)h(aP,b)\\
&= -h\big(aP, b\bar{u}\overline{h(av,b\bar{u})}\big) + f\big(h(av,b\bar{u}),\bar{u}\big)h(aP,b)\\
&\qquad - q(u)f\big(h(a,b),v\big)h(aP,b).
\end{aligned}$$

Since

$$f\big(h(av,b\bar{u}),\bar{u}\big) = f\big(h(av,b\bar{u}\cdot u),\varepsilon\big) = q(u)f\big(h(a,b),v\big),$$

it follows that

$$q(u)h\big(aP, bh(b,aP)u\big) = h\big(aP, b\bar{u}h(b\bar{u},av)\big).$$

Replacing b by $b\bar{u}$ in Proposition 17.2, we thus obtain

$$\begin{aligned} q(u)h\big(aP, bh(b,aP)u\big) &= \rho\big(h\big(b\bar{u}, ah(b\bar{u}, au)\big) + f\big(h(a, b\bar{u}), \varepsilon\big)h(b\bar{u}, au)\big) \\ &= \rho\big(h\big(b\bar{u}, ah(b\bar{u}, au)\big) + f\big(h(a,b), u\big)h(b\bar{u}, au)\big). \end{aligned}$$

Since

$$\begin{aligned} h(b\bar{u}, au) &= h(a, b\bar{u}u) + f\big(h(b\bar{u}, a), \varepsilon\big)u \\ &= q(u)h(a,b) - f\big(h(a,b), u\big)u, \end{aligned}$$

it follows that

$$h\big(b\bar{u}, ah(b\bar{u}, au)\big) = q(u)h\big(b\bar{u}, ah(a,b)\big) - f\big(h(a,b), u\big)h(b\bar{u}, au).$$

Hence

$$q(u)h\big(aP, bh(b,aP)u\big) = \rho q(u)h\big(b\bar{u}, ah(a,b)\big).$$

Thus the claim holds for all u such that $q(u) \neq 0$. Since every element of L can be written as a sum $u_1 + u_2$ with $q(u_1)$ and $q(u_2)$ nonzero, it follows that the claim holds in general. □

Proposition 17.6. *The following holds*:

$$\begin{aligned} h\big(aP, av\overline{\pi(b)}\big) = \rho\Big(h\big(a, a\overline{\theta(b,\bar{u})}\big) + h\big(b, ah(b, au)\big) - h\big(b\bar{u}, ah(a,b)\big) \\ + f\big(h(a,b), \varepsilon\big)h(b, au)\Big). \end{aligned}$$

Proof. We have

$$\begin{aligned} &h\big(aP, av\overline{\pi(b)}\big) \\ &\qquad = h\big(aP, aPu\overline{\pi(b)}\big) \\ &\qquad = -h\big(aP, aP\bar{u}\overline{\pi(b)}\big) + f(u,\varepsilon)h\big(aP, aP\overline{\pi(b)}\big) \\ &\qquad = h\big(aP, aP\pi(b)u\big) - f\big(u, \overline{\pi(b)}\big)h(aP, aP) + f(u,\varepsilon)h\big(aP, aP\overline{\pi(b)}\big) \\ &\qquad = h\big(aP, aP\pi(b)u\big) + f\big(u, \pi(b)\big)h(aP, aP) - f(u,\varepsilon)h\big(aP, aP\pi(b)\big) \end{aligned}$$

By [Weiss 2006b, 3.22], it follows that

$$\begin{aligned} h\big(aP, av\overline{\pi(b)}\big) = h\big(aP, aP\theta(b,u)\big) + h\big(aP, bh(b, av)\big) - h\big(aP, bh(b, aP)u\big) \\ + f\big(u, \pi(b)\big)h(aP, aP) - f(u,\varepsilon)h\big(aP, aP\pi(b)\big). \end{aligned}$$

Let $N = \theta(b,u) - f(u,\varepsilon)\pi(b) + f(u,\pi(b))\varepsilon$, so that

$$h(aP, aPN) = h\big(aP, aP\theta(b,u)\big) + f\big(u, \pi(b)\big)h(aP, aP) - f(u,\varepsilon)h\big(aP, aP\pi(b)\big).$$

Since

$$\begin{aligned} N &= -\theta(b,\bar u) + f\big(u,\pi(b)\big)\varepsilon \\ &= -\theta(b,\bar u) + f\big(\bar u,\overline{\pi(b)}\big)\varepsilon \\ &= -\theta(b,\bar u) - f\big(\bar u,\pi(b)\big)\varepsilon + Q(b)f(u,\varepsilon)\varepsilon \\ &= -\theta(b,\bar u) + f\big(\theta(b,\bar u),\varepsilon\big)\varepsilon \quad \text{by [Weiss 2006b, 4.9(iii)]} \\ &= \overline{\theta(b,\bar u)}, \end{aligned}$$

it follows by Proposition 12.12 that

$$h(aP,aPN) = \rho h\big(a,a\overline{\theta(b,\bar u)}\big).$$

The claim holds, therefore, by Proposition 17.2 and Proposition 17.5. □

Proposition 17.7. *The following holds*:

$$h\big(b,ah(au,b)\big) - h\big(b,ah(b,au)\big) + f\big(h(a,b),\varepsilon\big)h(au,b) \\ - f\big(h(a,b),\varepsilon\big)h(b,au,b) = f\big(h(a,b),u\big)h(a,b).$$

Proof. The expression on the left-hand side equals

$$\begin{aligned} & h\big(b,ah(au,b)+a\overline{h(au,b)}\big) + f\big(h(a,b),\varepsilon\big)\big(h(au,b)+\overline{h(au,b)}\big) \\ &\qquad = f\big(h(au,b),\varepsilon\big)h(b,a) + f\big(h(a,b),\varepsilon\big)f\big(h(au,b),\varepsilon\big)\varepsilon \\ &\qquad = f\big(h(a,b),u\big)\big(f\big(h(a,b),\varepsilon\big)\varepsilon - \overline{h(a,b)}\big) \\ &\qquad = f\big(h(a,b),u)\big)h(a,b). \end{aligned}$$

□

Proposition 17.8. *We have* $h(b_1,b_2) = \frac{1}{\Theta}(\omega+\rho^{-1}\xi)$, *where*

$$\omega = h\big(b\bar u,ah(a,b)\big) - h\big(a,a\overline{\theta(b,\bar u)}\big) + f\big(h(a,b),u\big)h(a,b) \\ - s\big(2\theta(a,u)+Q(a)u+h(aP,b)\big) - h\big(aP,b\overline{\pi(b)}\big)$$

and

$$\begin{aligned} \xi = {} & \Big(q(u)q(\pi(a)) - q(u)r^2 + f\big(h(a,b),\theta(a,v)\big) + Q(a)f\big(h(a,b),v\big)\Big)h(a,b) \\ & + \big(q(u)Q(a)+2q(u)t+f\big(h(a,b),v\big)\big)h\big(a\pi(a),b\big) \\ & + \big(q(u)Q(a)t - 2q(u)q(\pi(a)) - f\big(h(a,b),\theta(a,v)\big)\big)u \\ & + \big(q(u)Q(a)+2tq(u)+f\big(h(a,b),v\big)\big)\theta(a,u). \end{aligned}$$

Proof. We have

$$h(b_1,b_2) = -\frac{1}{\Theta}\rho^{-1}h(aP,y),$$

where y is as in Proposition 16.4. The claim holds, therefore, by Propositions 12.11, 12.12, 17.1, 17.3, 17.4, 17.6, 17.7 and some calculation. □

18. The element u_2, part II

We continue with all the notation and assumptions of the previous two sections.

Proposition 18.1. *We have* $\rho t_1\bar{u}_1 = \alpha + \rho^{-1}\beta$, *where*

$$\begin{aligned} &\alpha = s\big(h(aP,b)+tu+\theta(a,u)\big),\\ \text{and}\quad &\beta = \big(rq(u)+\phi(a,u)+g(b,av)\big)\big(h(aP,b)+tu+\theta(a,u)\big). \end{aligned}$$

Proof. This holds by Proposition 12.15(ii) and (iii). □

Lemma 18.2. *The following holds*:

$$\begin{aligned} \overline{\theta(av,u_1)} = \big(\phi(a,u)t + Q(a)tq(u) + q(\pi(a))q(u) + f\big(h(a,b),\theta(a,v)\big)\big)u\\ + \big(\phi(a,u) - q(u)t - f(h(a,b),v)\big)\theta(a,u)\\ + \big(\phi(a,u) - q(u)r\big)h\big(a\pi(a),b\big)\\ + \big(\phi(a,u)r - q(u)r^2 + \rho q(u)\big)h(a,b). \end{aligned}$$

Proof. By (C4), and Lemmas 12.13 and 13.8, we have

$$\begin{aligned} \overline{\theta(av,u_1)} &= \rho\overline{\theta(au,u_1)}\\ &= \overline{\theta\big(au, \overline{h(aP,b)} + t\bar{u} + \overline{\theta(a,u)}\big)}\\ &= q(u)\theta\big(a, h(aP,b)+tu+\theta(a,u)\big)\\ &\quad - f\big(u, h(aP,b)+tu+\theta(a,u)\big)\theta(a,u)\\ &\quad + f\big(\theta(a,u), h(aP,b)+tu+\theta(a,u)\big)u\\ &\quad + \phi(a,u)\big(h(aP,b)+tu+\theta(a,u)\big). \end{aligned}$$

By Proposition A.4(iii),

$$\theta\big(a,h(aP,b)\big) = \rho h(a,b) - rh(aP,b).$$

By Lemma 13.16 and [Weiss 2006b, 4.9(i), 4.21 and 4.22], therefore, the claim follows. □

Proposition 18.3. *The following hold*:

(i) $h(b,av\bar{u}) = q(u)h(b,aP)$.

(ii) $h(b,av\overline{\theta(a,u)}) = q(u)\big(\big(q(\pi(a))+Q(a)r\big)h(b,a) - rh\big(b,a\pi(a)\big)\big)$.

(iii) $avh(b,aP) = \rho ah(a,b)\bar{u} - f\big(h(a,b),v\big)aP$.

(iv) $h\big(b,avh(b,aP)\big) = \rho h\big(b,ah(a,b)\bar{u}\big) - f\big(h(a,b),v\big)h(b,aP)$.

Proof. Since $av\bar{u} = aPu\bar{u} = q(u)aP$, (i) holds. By [Weiss 2006b, 4.9(i) and 4.20], we have

$$\begin{aligned} av\overline{\theta(a,u)} &= aPu\overline{\theta(a,u)} \\ &= -aP\theta(a,u)\bar{u} + f(\theta(a,u),u)aP \\ &= a(\overline{\pi(a)} + t\varepsilon)\theta(a,u)\bar{u} + f(\theta(a,u),u)aP \\ &= q(\pi(a))au\bar{u} + ta\pi(a)u\bar{u} + Q(a)q(u)aP \\ &= q(u)\big(q(\pi(a))a + ta\pi(a) + Q(a)a(\pi(a)+r\varepsilon)\big). \end{aligned}$$

Thus (ii) holds.

Next note that

$$\begin{aligned} avh(b,aP) &= aPuh(b,aP) \\ &= -aP\overline{h(b,aP)}\bar{u} + f\big(\overline{h(b,aP)},u\big)aP \\ &= aPh(aP,b)\bar{u} - f\big(h(aP,b),u\big)aP \\ &= aPh(aP,b)\bar{u} - f\big(h(a,b),v\big)aP. \end{aligned} \tag{18.4}$$

By Proposition A.4(i) and two applications of [Weiss 2006b, 3.22], we have

$$\begin{aligned} aPh(aP,b)\bar{u} &= aPh(aP,b\bar{u}) + b\theta(aP,\bar{u}) - b\pi(aP)\bar{u} \\ &= aPh(aP,b\bar{u}) + \rho\big(b\theta(a,\bar{u}) - b\pi(a)\bar{u}\big) \\ &= aPh(aP,b\bar{u}) + \rho\big(ah(a,b)\bar{u} - ah(a,b\bar{u})\big) \end{aligned} \tag{18.5}$$

and by Proposition A.4(iii), we have

$$\begin{aligned} aPh(aP,b\bar{u}) &= a(\pi(a)+r\varepsilon)h(aP,b\bar{u}) \\ &= a\theta\big(a,h(aP,b\bar{u})\big) + rah(aP,b\bar{u}) \\ &= \rho ah(a,b\bar{u}). \end{aligned}$$

By (18.4) and (18.5), therefore,

$$avh(b,aP) = \rho ah(a,b)\bar{u} - f\big(h(a,b),v\big)aP.$$

Thus (iii) holds. The assertion (iv) follows immediately. □

Corollary 18.6. *The following holds*:

$$\begin{aligned} \rho\overline{h\big(b,avu_1\big)} = \rho h\big(ah(a,b)\bar{u},b\big) + \big((r-t)q(u) - f(h(a,b),v)\big)h(a\pi(a),b) \\ + \big(r^2q(u) - q(\pi(a))q(u) - rf(h(a,b),v)\big)h(a,b). \end{aligned}$$

Proof. This holds by Lemma 12.13 and Proposition 18.3. □

Proposition 18.7. *We have $\rho\overline{\theta(a_1,u_1)} = \nu + \rho^{-1}\zeta$, where*

$$\nu = -\overline{\theta\big(b,h(b,aP)\big)} + t\overline{\theta(b,\bar{u})} + \theta\big(b,\overline{\theta(a,u)}\big) + q(u)h(a,b) + h\big(ah(a,b)\bar{u},b\big)$$

and

$$\begin{aligned}\zeta = \big(\phi(a,u)t + Q(a)tq(u) + q(\pi(a))q(u) + f\big(h(a,b),\theta(a,v)\big)\big)u \\ + \big(\phi(a,u) - q(u)t - f(h(a,b),v)\big)\theta(a,u) \\ + \big(r\phi(a,u) - q(\pi(a))q(u) - rf(h(a,b),v)\big)h(a,b) \\ + \big(\phi(a,u) - q(u)t - f(h(a,b),v)\big)h(a\pi(a),b) \\ - g(b,av)\big(h(aP,b) + tu + \theta(a,u)\big).\end{aligned}$$

Proof. This follows from (C3), Proposition 12.15(i) and (iii), Lemma 18.2 and Corollary 18.6. □

Proposition 18.8. *Let u_2 be as in Notation 16.1. Then*

$$u_2 = \frac{\delta}{\Theta},$$

where

$$\begin{aligned}\delta = t\overline{\theta(b,\bar{u})} - q(u)u - h\big(a, a\overline{\theta(b,\bar{u})}\big) + \overline{\theta(b,\overline{\theta(a,u)})} + q(u)h(a,b) \\ + 2h\big(b\bar{u}, ah(a,b)\big) + 2q(h(a,b))u - sv\end{aligned}$$

Proof. By Proposition 12.15 and (14.2), we have

$$\begin{aligned}u_2 &= \overline{h(-b_2,b_1)} + \frac{\rho}{\Theta}\big(t_1\bar{u}_1 + \overline{\theta(a_1,u_1)}\big) \\ &= h(b_1,b_2) + \frac{\rho}{\Theta}\big(t_1\bar{u}_1 + \overline{\theta(a_1,u_1)}\big).\end{aligned}$$

The sum of the elements ξ in Proposition 17.8, β in Proposition 18.1 and ζ in Proposition 18.7 is

$$\iota := f\big(h(a,b),\theta(a,v) + tv\big)h(a,b) - \rho q(u)u.$$

By Proposition 13.1, we have

$$\rho^{-1}\iota = -f\big(h(a,b),u\big)h(a,b) - q(u)u$$

and thus the sum of $\rho^{-1}\iota$ with the elements ω in Proposition 17.8, α in Proposition 18.1 and ν in Proposition 18.7 is

$$\begin{aligned}t\overline{\theta(b,\bar{u})} - \overline{\theta(b,h(b,aP))} - q(u)u - h\big(a, a\overline{\theta(b,\bar{u})}\big) \\ + \overline{\theta(b,\overline{\theta(a,u)})} + q(u)h(a,b) + h\big(ah(a,b)\bar{u}, b\big) \\ + h\big(b\bar{u}, ah(a,b)\big) - sv - h\big(aP, b\overline{\pi(b)}\big).\end{aligned}$$

By Proposition A.6(i), we have

$$\begin{aligned}\overline{\theta\big(b,h(b,aP)\big)} &= -\overline{h(b\pi(b),aP)} + Q(b)\overline{h(b,aP)} \\ &= h\big(aP, b\pi(b)\big) + Q(b)h(aP,b) = -h\big(aP, b\overline{\pi(b)}\big)\end{aligned}$$

and by (B2) and [Weiss 2006b, 3.6], we have

$$\begin{aligned} h\big(ah(a,b)\bar{u}, b\big) &= -\overline{h\big(b, ah(a,b)\bar{u}\big)} \\ &= -\overline{h\big(ah(a,b), b\bar{u}\big)} - f\big(h(b, ah(a,b)), \varepsilon\big)u \\ &= h\big(b\bar{u}, ah(a,b)\big) + f\big(h(a,b), h(a,b)\big)u \\ &= h\big(b\bar{u}, ah(a,b)\big) + 2q(h(a,b))u. \end{aligned}$$

Hence the claim holds. □

19. A formula for τ

We are now in a position to determine a formula for the structure map of the Moufang set $M(\Delta, \omega)$ defined in Definition 10.8 (modulo Conjecture 19.14). Our main result is Theorem 19.7.

We continue with all the notation and assumptions in Hypothesis 9.1. Let ω be as in Proposition 9.2, let ξ, ψ, σ, η and λ be as in Notation 9.3, let ν be as in Notation 10.4 and let τ be as in Notation 10.7.

Proposition 19.1. $$\varepsilon = \eta^{\sigma+1}\xi(\bar{\lambda}).$$

Proof. By Proposition 9.5(v), $\xi(\bar{\lambda}) = \xi(\overline{\xi(\varepsilon)}) = \xi(\overline{\xi(\bar{\varepsilon})}) = q(\lambda)\varepsilon$. Hence by Proposition 9.5(iii), the claim holds. □

Proposition 19.2. *Let* $x_1(a,t)x_2(u)x_3(b,s) \in C_{U_{[1,3]}}(\omega)$. *Then*

(i) $\pi(a) = \eta^{\sigma+2}\xi(\theta(b,\bar{\lambda})) + M(b)\varepsilon$.

(ii) $\eta^{\sigma+1} f\big(\xi(\theta(b,\bar{\lambda})), \varepsilon\big) = Q(b)^\sigma$.

(iii) $q(\pi(a) + t\varepsilon) = \eta^2 q(\pi(b) + s\varepsilon)^\sigma$.

(iv) $Q(a) = \eta Q(b)^\sigma$.

Proof. By Proposition 9.15, we have $a = \psi(b)$ and

$$q(\pi(a) + t\varepsilon) = q\big(\pi(\psi(b)) + (\eta s^\sigma + M(b))\varepsilon\big) \tag{19.3}$$

and by Propositions 7.13(d) and 19.1, we have

$$\begin{aligned} \pi(\psi(b)) &= \theta(\psi(b), \varepsilon) \\ &= \eta^{\sigma+1}\theta(\psi(b), \xi(\bar{\lambda})) \\ &= \eta^{\sigma+2}\xi(\theta(b,\bar{\lambda})) + M(b)\varepsilon. \end{aligned}$$

Thus (i) holds. Since $M(b) = 0$ if $\mathrm{char}(K) \neq 2$, it follows that

$$\begin{aligned} &q\big(\pi(\psi(b)) + (\eta s^\sigma + M(b))\varepsilon\big) \\ &\quad = q\big(\eta^{\sigma+2}\xi(\theta(b,\bar{\lambda})) + \eta s^\sigma \varepsilon\big) \\ &\quad = \eta^{2\sigma+4} q(\xi(\theta(b,\bar{\lambda}))) + \eta^{\sigma+3} s^\sigma f\big(\xi(\theta(b,\bar{\lambda})), \varepsilon\big) + \eta^2 s^{2\sigma}. \end{aligned} \tag{19.4}$$

By Propositions 9.4, 9.5(iii) and 19.1, we have

$$\begin{aligned} f\big(\xi(\theta(b,\bar\lambda)),\varepsilon\big) &= \eta^{\sigma+1} f\big(\xi(\theta(b,\bar\lambda)),\xi(\bar\lambda)\big) \\ &= f(\theta(b,\bar\lambda),\bar\lambda)^\sigma. \end{aligned}$$

By [Weiss 2006b, 4.9], we have $f(\theta(b,\bar\lambda),\bar\lambda) = Q(b)q(\lambda)$. Therefore (ii) holds and thus

$$\eta^{\sigma+3} s^\sigma f\big(\xi(\theta(b,\bar\lambda)),\varepsilon\big) = \eta^2 s^\sigma Q(b)^\sigma. \tag{19.5}$$

By Propositions 7.13(i), 9.5(iii) and [Weiss 2006b, 4.22], we have

$$\begin{aligned} \eta^{2\sigma+4} q\big(\xi(\theta(b,\bar\lambda))\big) &= \eta^{2\sigma+4} q(\lambda) q(\theta(b,\bar\lambda))^\sigma \\ &= \eta^{2\sigma+4} q(\lambda)^{\sigma+1} q(\pi(b))^\sigma = \eta^2 q(\pi(b))^\sigma. \end{aligned}$$

By (19.3), (19.4) and (19.5), therefore, (iii) holds. By (i), we have

$$Q(a) = f(\pi(a),\varepsilon) = \eta^{\sigma+2} f\big(\xi(\theta(b,\bar\lambda)),\varepsilon\big).$$

By (ii), therefore, (iv) holds. □

Proposition 19.6. *Let Θ be as in (14.3), let ψ, η, σ and M be as in Notation 9.3, let ζ be as in Proposition 9.10 and let $\mathcal{U}$ be as in Notation 9.18. Let*

$$\Theta_0(b,s,u) = \Theta\big(\psi(b),\ \eta s^\sigma + M(b),\ u,\ b,\ s\big)$$

for all $[b,s,u] \in \mathcal{U}$. Then

$$\begin{aligned} \Theta_0(b,s,u) = \eta^2 q\big(\pi(b)+s\varepsilon\big)^{\sigma+1} &+ q\big(h(\psi(b),bu)\big) - q(u)^2 - q(u) f(u,\zeta(u)) \\ &- q(u) f\big(\pi(\psi(b)),\pi(b)\big) + f\big(\theta(b,\bar u),\overline{\theta(\psi(b),u)}\big) \\ &+ f(u,\varepsilon) f\big(\theta(\psi(b),u),\ \pi(b)\big) - f(\pi(b),u) f\big(\theta(\psi(b),u),\ \varepsilon\big) \\ &- f\big(\zeta(u),\overline{\theta(b,\overline{\theta(\psi(b),u)})} + s\theta(\psi(b),u)\big) \\ &+ \big(\eta(Q(b)+s)^\sigma + M(b)\big) f\big(\zeta(u),\overline{\theta(b,\bar u)} + su\big) \end{aligned}$$

for all $[b,s,u] \in \mathcal{U}$.

Proof. Recall that the map M in Proposition 7.13(d) is identically zero when $\operatorname{char}(K) \neq 2$ and that by Notation 9.18, $h(\psi(b),b) = \zeta(u)+u$ for all $[b,s,u] \in \mathcal{U}$. The claim holds, therefore, by Proposition 19.2 and a bit of calculation. □

Here, at last, is our formula:

Theorem 19.7. *Let ω be as in Proposition 9.2, let ψ, η, σ and M be as in Notation 9.3, let ζ be as in Proposition 9.10, let $\mathcal{U}$ be as in Notation 9.18 and let τ be as in Notation 10.7. Let $[b,s,u] \in \mathcal{U}^*$ and let $\Theta_0 = \Theta_0(b,s,u)$ be as in Proposition 19.6. Suppose that*

$$q(\pi(b)+s\varepsilon)\Theta_0(b,s,u) \neq 0. \tag{19.8}$$

Then

$$\tau([b,s,u]) = \Big[\frac{\hat{b}}{\Theta_0}, \frac{\hat{s}}{\Theta_0}, \frac{\hat{u}}{\Theta_0}\Big], \tag{19.9}$$

where

$$\begin{aligned}\hat{s} = \big(\eta s^\sigma + M(b)\big)q(u) + \eta^2 q(\pi(b)+s\varepsilon)^\sigma(Q(b)-s) + \phi(\psi(b),u)\\ + g\big(\psi(b)\theta(\psi(b),u),b\big) + \big(\eta(Q(b)-s)^\sigma + M(b)\big)g(\psi(b)u,b),\end{aligned}$$

$$\begin{aligned}\hat{b} = -q(u)\psi(b)\zeta(u) + \eta^2 q\big(\pi(b)+s\varepsilon\big)^\sigma b\big(\overline{\pi(b)}+s\varepsilon\big)\\ + \big(\eta(Q(b)-s)^\sigma + M(b)\big)\big(\psi(b)u(\overline{\pi(b)}+s\varepsilon) - bh(\psi(b)u,b)\big)\\ + \psi(b)\theta(\psi(b),u)\big(\overline{\pi(b)}+s\varepsilon\big) - bh\big(\psi(b)\theta(\psi(b),u),b\big)\\ - b\bar{u}\theta(\psi(b),u) - \big(\eta s^\sigma + M(b)\big)q(u)b.\end{aligned}$$

and

$$\begin{aligned}\hat{u} = \eta\big(\eta s^\sigma + M(b)\big)\overline{\theta(b,\bar{u})}\\ - \eta h\big(\psi(b),\psi(b)\overline{\theta(b,\bar{u})}\big) + \eta\overline{\theta\big(b,\overline{\theta(\psi(b),u)}\big)}\\ + 3\eta q(u)\zeta(u) + 2\eta^{2-\sigma}q(u)^\sigma u\\ + 2\eta h\big(b\bar{u},\psi(b)\zeta(u)\big)\\ - \eta s\theta(\psi(b),u) + \eta\big(\eta(Q(b)+s)^\sigma + M(b)\big)su.\end{aligned}$$

Proof. By Proposition 19.2(iii), $q(\pi(a)+t\varepsilon) \neq 0$, where $a = \psi(b)$ and $t = \eta s^\sigma + M(b)$ and by Notation 9.18, $h(\psi(b),b) = \zeta(u)+u$ for all $[b,s,u] \in \mathcal{U}$. By Proposition 9.10, we have $q(\zeta(u)) = \eta^{1-\sigma}q(u)^\sigma$. The claim holds, therefore, by Notation 10.4, 16.1, and Propositions 16.3, 16.4, 18.8. □

Proposition 19.10. *Let $\mathcal{U}$ be as in Notation 9.18 and Θ_0 as in Proposition 19.6. Let $(b,s,u) \in \mathcal{U}$ and suppose that either $q(\pi(b)+s\varepsilon) \neq 0$ or $(b,s) = (0,0)$ but $u \neq 0$. Then $\Theta_0(b,s,u) \neq 0$.*

Proof. Let $\nu = (m_1m_4)^2h_0$ be as in Notation 10.4, let

$$p = U_4x_1(a,t)x_2(u)x_3(b,s),$$

where $a = \psi(b)$ and $t = \eta s^\sigma + M(b)$, and let $\mathcal{O}$ and $\mathcal{B}$ be as in Proposition 11.8. By Notation 10.7,

$$p^\nu = U_4\tau(b,s,u).$$

Since h_0 stabilizes the set $U_4\mathcal{U}$, we conclude that

$$p^{(m_1m_4)^2} \in U_4\mathcal{U} \subset \mathcal{O}. \tag{19.11}$$

Suppose now that $q(\pi(b)+s\varepsilon)\neq 0$. By Proposition 19.2(iii), $q(\pi(a)+t\varepsilon)\neq 0$. By Proposition 11.8, therefore, $p^{m_1m_4}\in\mathcal{O}$. By (19.11) and a second application of Proposition 11.8, it follows that $p^{m_1m_4}\in\mathcal{B}$. Thus $q(\pi(a_1)+t_1\varepsilon)\neq 0$, where a_1 and t_1 are as in Proposition 12.15. By (14.2), therefore, $\Theta_0(b,s,u)=\Theta(a,t,u,b,s)\neq 0$.

Suppose, instead, that $(b,s)=(0,0)$ but $u\neq 0$. Thus $p=U_4x_2(u)$ and

$$\alpha := \big(U_4x_2(u), U_{[3,4]}x_2(u), U_{[2,4]}, U_{[3,4]}, U_4\big)$$

is a path of length 4 from p to U_4. By Theorem 6.11(v), the pair (p,U_4) is in the same G-orbit as the pair (w_4,U_4), where G is as in Hypothesis 9.1. Since

$$(w_4, w_5, U_{[2,4]}, U_{[3,4]}, U_4)$$

is a straight 4-path (i.e., a root) from w_4 to U_4, it follows that there exists a root from (w_4,U_4). By [Mühlherr and Weiss $\geq$ 2020, 1.2.28(ii) and 1.3.18], every path of length at most 4 from w_4 to U_4 is a root. Hence α is a root. In particular, α is straight and thus the vertices $U_{[3,4]}x_2(u)$ and $U_{[3,4]}$ are opposite at $U_{[2,4]}$. By [Mühlherr and Weiss 2020, 6.4(ii)], therefore, $q(u)\neq 0$. By (14.3), we have $\Theta(0,0,u,0,0)=q(u)^2\neq 0$ and hence $\Theta_0(b,s,u)\neq 0$ also in this case. □

Corollary 19.12. *Let F be as in Remark 9.12, let $\mathcal{V}=\{[b,s,u]\in\mathcal{U}\mid b=0\}$ and let*

$$q_0(0,s,u)=q(u)-\eta s^{\sigma+1}$$

for all $(0,s,u)\in\mathcal{V}$. Then q_0 is an anisotropic quadratic form on $\mathcal{V}$ as a vector space over F.

Proof. By Proposition 9.15, $[0,s,u]\in\mathcal{U}$ for $s\in K$ and $u\in L$ if and only if $\zeta(u)=-u$. Thus if $[0,s,u]\in\mathcal{V}$, then $f(u,\zeta(u))=-2q(u)$. By Remark 9.13, q_0 is a quadratic form on $\mathcal{V}$ as a vector space over F and by Proposition 19.6,

$$\Theta_0(0,s,u)=(q(u)-\eta s^{\sigma+1})^2$$

for all $(0,s,u)\in\mathcal{V}$. By Proposition 19.10, therefore, q_0 is anisotropic. □

Corollary 19.13. *The field K is infinite.*

Proof. By Remarks 7.6 and 9.12, $\dim_F q_0\geq 7$ and K/F is a quadratic extension. Over a finite field, there are no anisotropic quadratic forms of dimension greater than 2; see, for example, [Tits and Weiss 2002, 34.3]. By Corollary 19.12, therefore, F is infinite. Hence K is infinite. □

Conjecture 19.14. We conjecture that $\Theta_0(b,s,t)\neq 0$ for all $[b,s,u]\in\mathcal{U}$ such that $(b,s)\neq 0$ but $q(\pi(b)+s\varepsilon)=0$. This holds vacuously if the quadrangular algebra Ξ is anisotropic; see Notation 2.2. By Proposition 19.10, this conjecture would imply that Θ_0 is anisotropic.

Proposition 19.15. *Suppose that Conjecture 19.14 holds. Then the hypothesis (19.8) in Theorem 19.7 is superfluous.*

Proof. Exactly as at the end of [Mühlherr and Weiss ≥ 2020, Section 3.7], we observe that the function τ defined in Notation 10.7 is a regular map from $\mathcal{U}^*$ to itself. Since Θ_0 is anisotropic, we can let $\hat{\tau}$ denote the map from $\mathcal{U}^*$ to $U_{[1,3]}$ defined by the expression on the right-hand side of (19.9). The map $\hat{\tau}$ is also regular and by Theorem 19.7, τ and $\hat{\tau}$ agree on a Zariski dense subset of $\mathcal{U}^*$. By Corollary 19.13, K is infinite. It follows that $\tau = \hat{\tau}$. □

Conjecture 19.16. Suppose only that ω is as Proposition 9.2. We conjecture, but with less confidence, that the hypothesis that $G_0 := \langle\omega\rangle$ is as in Hypothesis 8.4 is, in fact, equivalent with the assumption that the restriction of Θ to $C_{U_{[1,4]}}(\omega)$ is anisotropic.

Conjecture 19.17. We conjecture that the form Θ_0 is, up to similarity, an invariant of the Moufang set $M(\Delta, \omega)$.

Appendix

In this appendix we assemble a few elementary properties of quadrangular algebras that cannot be found in [Weiss 2006b].

Let $\Xi = (K, L, q, f, \varepsilon, \mathcal{X}, \cdot, h, \theta)$ be a quadrangular algebra as defined in Definition 2.1. Let $v \mapsto \bar{v}$, ϕ and π be as in (A1), (C4) and (D1).

Proposition A.1. *Let $a \in \mathcal{X}$ and $u \in L$. Then*

$$\pi(au) = q(u)\overline{\pi(a)} - f(u, \varepsilon)\overline{\theta(a, u)} + f(\theta(a, u), \varepsilon)\bar{u} + \phi(a, u)\varepsilon.$$

Proof. Set $w = \varepsilon$ in (C4). □

Proposition A.2. *For all $a \in \mathcal{X}$ and all $u \in L$,*

$$au\pi(au) = q(u)a\theta(a, u) + \phi(a, u)au.$$

Proof. This holds by [Weiss 2006a, 3.18]. (In [Weiss 2006a], it is assumed that the quadrangular algebra is anisotropic, but given [Mühlherr and Weiss 2019, Conclusion 7.5], it is straightforward to check that the proof remains valid in the present context.) □

Proposition A.3. *Let $a \in \mathcal{X}$, $t \in K$ and $u \in L$. Then the following hold*:

(i) $$q(\theta(a, u) + tu) = q(\pi(a) + t\varepsilon)q(u).$$

(ii) $$q\big(\pi(au) + tq(u) + \phi(a, u)\big) = q(\pi(a) + t\varepsilon)q(u)^2.$$

Proof. These assertions hold by [Weiss 2009, 21.10]. (In [Weiss 2009], it is assumed that the quadrangular algebra is anisotropic, but given [Mühlherr and Weiss 2019, Conclusion 7.5], it is straightforward to check that the proof remains valid in the present context.) □

Proposition A.4. *Let $a \in \mathcal{X}$, $r \in K$ and $u \in L$ and let $P = \pi(a) + r\varepsilon$. Then the following hold*:

(i) $\theta(aP, u) = q(P)\theta(a, u)$.

(ii) $\phi\big(a, \theta(a, u) + ru\big) = \phi(aP, u)$.

(iii) $q(P)h(a, b) = \theta\big(a, h(aP, b)\big) + rh(aP, b)$.

Proof. These assertions hold by [Tits and Weiss 2002, 13.67(i)–(iii)]. (Given [Mühlherr and Weiss 2019, Conclusion 7.5], it is straightforward to check that the proofs in [Tits and Weiss 2002] remain valid in the present context.) □

Remark A.5. The notion of a quadrangular algebra had not yet been formulated when [Tits and Weiss 2002] was written. The axioms defining a quadrangular algebra, however, can all be found in [Tits and Weiss 2002, Chapter 13] with the same notion as in Definition 2.1 with one small exception: If we call the function g introduced in [Tits and Weiss 2002, 13.26] for the moment g' and let g be as in (C3), then $g'(a, b) = g(b, a)$ for all $a, b \in \mathcal{X}$. See, in particular, [Tits and Weiss 2002, 13.37] and the remark (viii) on page 7 of [Weiss 2006b].

Proposition A.6. *The following hold for all $a, b \in \mathcal{X}$ and all $u \in L$, where Q is as in Notation 2.3.*

(i) $\theta(a, h(a, b)) = Q(a)h(a, b) - h(a\pi(a), b)$.

(ii) $\theta\big(a, h(a\pi(a), b)\big) = q(\pi(a))h(a, b)$.

(iii) $$\theta(a, h(au, b)) = -f(h(a, b), \varepsilon)\theta(a, u) - f\big(h(b, a), \pi(a)\big)u + Q(a)\overline{h(au, b)} + \overline{h(a\theta(a, u), b)} + f(h(a, b), u)\pi(a).$$

Proof. The assertions (i) and (ii) hold by (e) and (f) at the bottom of page 120 of [Tits and Weiss 2002] (in the proof of [Tits and Weiss 2002, 13.67](iii)). Choose $a, b \in \mathcal{X}$ and $u \in L$. First note that by [Weiss 2006b, 4.9(i)],

$$\begin{aligned} f\big(\theta(a, u), h(b, au)\big) &= f\big(h(b, au\overline{\theta(a, u)}), \varepsilon\big) \\ &= -f\big(h(b, a\theta(a, u)\bar{u}), \varepsilon\big) + f\big(h(b, a), \varepsilon\big)f\big(\theta(a, u), u\big) \\ &= -f\big(h(b, a\pi(a)u\bar{u}), \varepsilon\big) + Q(a)q(u)f\big(h(b, a), \varepsilon\big) \\ &= -q(u)f\big(h(b, a\pi(a)), \varepsilon\big) + Q(a)q(u)f\big(h(b, a), \varepsilon\big) \\ &= q(u)f\big(h(b, a), -\overline{\pi(a)} + Q(a)\varepsilon\big) \\ &= q(u)f\big(h(b, a), \pi(a)\big). \end{aligned}$$

By (C4), therefore, we have

$$\begin{aligned}\overline{\theta\big(au, h(au,b)\big)} &= q(u)\theta\big(a, \overline{h(au,b)}\big) - f\big(h(au,b), \bar{u}\big)\theta(a,u) \\ &\qquad + f\big(\theta(a,u), \overline{h(au,b)}\big)u + \phi(a,u)\overline{h(au,b)} \\ &= q(u)\theta\big(a, \overline{h(au,b)}\big) - f\big(h(au\bar{u},b), \varepsilon\big)\theta(a,u) \\ &\qquad - f\big(\theta(a,u), h(b,au)\big)u + \phi(a,u)\overline{h(au,b)} \\ &= q(u)f\big(h(au,b),\varepsilon\big)\pi(a) - q(u)\theta(a,h(au,b)) \\ &\qquad - q(u)f\big(h(a,b),\varepsilon\big)\theta(a,u) - q(u)f\big(h(b,a),\pi(a)\big)u \\ &\qquad + \phi(a,u)\overline{h(au,b)}.\end{aligned}$$

By Propositions 13.5, A.2 and (i), on the other hand, we have

$$\begin{aligned}\overline{\theta\big(au, h(au,b)\big)} &= q(u)Q(a)\overline{h(au,b)} - \overline{h(au\pi(au),b)} \\ &= q(u)Q(a)\overline{h(au,b)} - q(u)\overline{h(a\theta(a,u),b)} + \phi(a,u)\overline{h(au,b)}.\end{aligned}$$

We set these two expressions equal and delete the term $\phi(a,u)\overline{h(au,b)}$ that appears on both sides. All the remaining terms have $q(u)$ as a factor. If we assume that $q(u) \neq 0$ and delete all these factors, we obtain (iii). Thus (iii) holds under the assumption that $q(u) \neq 0$. Since every term in (iii) is linear in u and every element of L can be written as a sum $u_1 + u_2$ with $q(u_1)$ and $q(u_2)$ nonzero, we conclude that (iii) holds in general. (This is the same argument we used at the end of the proof of Proposition 17.5.) □

We leave it to the reader to confirm that if we set $u = \varepsilon$ and $u = \pi(a)$ in Proposition A.6(iii), we obtain Proposition A.6(i) and (ii), respectively.

Proposition A.7. *Let $a, b \in \mathcal{X}$ and $u \in L$. Then $q(h(a,bu)) = q(h(b,au))$.*

Proof. By (B2), we have

$$q(h(b,au)) = q(h(a,bu)) + f(h(b,a),\varepsilon)f(h(a,bu),u) + f(h(b,a),\varepsilon)^2 q(u).$$

By (A3) and [Weiss 2006b, 3.7],

$$f(h(a,bu),u) = f(h(a,bu\bar{u}),\varepsilon) = f(h(a,b),\varepsilon)q(u)$$

and by [Weiss 2006b, 3.6],

$$f(h(a,b),\varepsilon) = f(\overline{h(a,b)},\bar{\varepsilon}) = -f(h(b,a),\varepsilon).$$

Hence $f(h(b,a),\varepsilon)f(h(a,bu),u) + f(h(b,a),\varepsilon)^2 q(u) = 0$. □

Acknowledgments

The work of Mühlherr was partially supported by a grant from the DFG and the work of Weiss was partially supported by a collaboration grant from the Simons Foundation. Significant portions of this research were carried out when both authors were guests of the National Center for Theoretical Sciences in Taipei and, subsequently, when Weiss was a guest of the Korea Institute for Advanced Study.

References

[Boelaert et al. 2019] L. Boelaert, T. De Medts, and A. Stavrova, *Moufang sets and structurable division algebras*, Mem. Amer. Math. Soc. **1245**, Amer. Math. Soc., Providence, RI, 2019. MR Zbl

[Callens and De Medts 2014] E. Callens and T. De Medts, "Moufang sets arising from polarities of Moufang planes over octonion division algebras", *Manuscripta Math.* **143**:1-2 (2014), 171–189. MR Zbl

[De Medts and Segev 2009] T. De Medts and Y. Segev, "A course on Moufang sets", *Innov. Incidence Geom.* **9** (2009), 79–122. MR Zbl

[De Medts and Van Maldeghem 2010] T. De Medts and H. Van Maldeghem, "Moufang sets of type F_4", *Math. Z.* **265**:3 (2010), 511–527. MR Zbl

[De Medts and Weiss 2006] T. De Medts and R. M. Weiss, "Moufang sets and Jordan division algebras", *Math. Ann.* **335**:2 (2006), 415–433. MR Zbl

[De Medts et al. 2017] T. De Medts, Y. Segev, and R. M. Weiss, "Tits endomorphisms and buildings of type F_4", *Ann. Inst. Fourier* (*Grenoble*) **67**:6 (2017), 2349–2421. MR Zbl

[Mühlherr and Weiss 2016] B. Mühlherr and R. M. Weiss, "Galois involutions and exceptional buildings", *Enseign. Math.* **62**:1-2 (2016), 207–260. MR Zbl

[Mühlherr and Weiss 2017] B. Mühlherr and R. M. Weiss, "Rhizospheres in spherical buildings", *Math. Ann.* **369**:1-2 (2017), 839–868. MR Zbl

[Mühlherr and Weiss 2019] B. Mühlherr and R. M. Weiss, "Isotropic quadrangular algebras", *J. Math. Soc. Japan* **71**:4 (2019), 1321–1380. MR Zbl

[Mühlherr and Weiss 2020] B. Mühlherr and R. M. Weiss, "The exceptional Tits quadrangles", *Transform. Groups* (online publication May 2020).

[Mühlherr and Weiss ≥ 2020] B. Mühlherr and R. M. Weiss, "Tits polygons", To appear in *Mem. Amer. Math. Soc.*

[Mühlherr et al. 2015] B. Mühlherr, H. P. Petersson, and R. M. Weiss, *Descent in buildings*, Ann. of Math. Stud. **190**, Princeton Univ. Press, 2015. MR Zbl

[Springer 1998] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progr. Math. **9**, Birkhäuser, Boston, 1998. MR Zbl

[Thompson 1972] J. G. Thompson, "Toward a characterization of $E_2^*(q)$, II", *J. Algebra* **20** (1972), 610–621. MR Zbl

[Tits 1966a] J. Tits, "Classification of algebraic semisimple groups", pp. 33–62 in *Algebraic groups and discontinuous subgroups* (Boulder, CO, 1965), edited by A. Borel and G. D. Mostow, Proc. Symp. Pure Math. **9**, Amer. Math. Soc., Providence, RI, 1966. MR Zbl

[Tits 1966b] J. Tits, "Les groupes simples de Suzuki et de Ree", exposé 210 in *Séminaire Bourbaki,* 1960/61, W. A. Benjamin, Amsterdam, 1966. Reprinted as pp. 65–82 in *Séminaire Bourbaki* **6**, Soc. Math. France, Paris, 1995. MR

[Tits 1974] J. Tits, *Buildings of spherical type and finite BN-pairs*, Lecture Notes in Math. **386**, Springer, 1974. MR Zbl

[Tits 1992] J. Tits, "Twin buildings and groups of Kac–Moody type", pp. 249–286 in *Groups, combinatorics and geometry* (Durham, UK, 1990), edited by M. Liebeck and J. Saxl, Lond. Math. Soc. Lect. Note Ser. **165**, Cambridge Univ. Press, 1992. MR Zbl

[Tits and Weiss 2002] J. Tits and R. M. Weiss, *Moufang polygons*, Springer, 2002. MR Zbl

[Weiss 2003] R. M. Weiss, *The structure of spherical buildings*, Princeton Univ. Press, 2003. MR Zbl

[Weiss 2006a] R. M. Weiss, "Moufang quadrangles of type E_6 and E_7", *J. Reine Angew. Math.* **590** (2006), 189–226. MR Zbl

[Weiss 2006b] R. M. Weiss, *Quadrangular algebras*, Math. Notes **46**, Princeton Univ. Press, 2006. MR Zbl

[Weiss 2009] R. M. Weiss, *The structure of affine buildings*, Ann. of Math. Stud. **168**, Princeton Univ. Press, 2009. MR Zbl

Received September 27, 2019. Revised April 9, 2020.

BERNHARD MÜHLHERR
MATHEMATISCHES INSTITUT
UNIVERSITÄT GIESSEN
GERMANY

bernhard.m.muehlherr@math.uni-giessen.de

RICHARD M. WEISS
MATHEMATICS
TUFTS UNIVERSITY
MEDFORD, MA
UNITED STATES

rweiss@tufts.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.455

GLOBALLY ANALYTIC PRINCIPAL SERIES REPRESENTATION AND LANGLANDS BASE CHANGE

JISHNU RAY

S. Orlik and M. Strauch have studied locally analytic principal series representation for general p-adic reductive groups generalizing an earlier work of P. Schneider for GL(2) and related the condition of irreducibility of such locally analytic representation with that of a suitable Verma module. We take the case of GL(n) and study the globally analytic principal series representation under the action of the pro-p Iwahori subgroup of $GL(n, \mathbb{Z}_p)$, following the notion of globally analytic representations introduced by M. Emerton. Furthermore, we relate the condition of irreducibility of our globally analytic principal series to that of a Verma module. Finally, using the Steinberg tensor product theorem, we construct the Langlands base change of our globally analytic principal series to a finite unramified extension of $\mathbb{Q}_p$.

1. Introduction

In this paper we construct a globally analytic (also called rigid analytic) principal series representation of the pro-p Iwahori subgroups of $\mathrm{GL}_n(\mathbb{Z}_p)$ and determine when it is irreducible. Furthermore, we construct base change of our rigid analytic representation to a finite unramified extension of $\mathbb{Q}_p$. This extends earlier works of Robert [1984; 1985] for SL_2 and Clozel [2018] for GL_2.

Denote by G the pro-p Iwahori subgroup of $\mathrm{GL}_n(\mathbb{Z}_p)$ (the group of matrices in $\mathrm{GL}_n(\mathbb{Z}_p)$ that are lower unipotent modulo $p\mathbb{Z}_p$) and by B the subgroup of matrices in $\mathrm{GL}_n(\mathbb{Z}_p)$ which are lower triangular modulo $p\mathbb{Z}_p$. Let P_0 and T_0 be the set of upper triangular and diagonal matrices in B, respectively. Let $Q_0 = P_0 \cap G$. Let P^+ be the Borel subgroup of upper triangular matrices in $\mathrm{GL}_n(\mathbb{Z}_p)$, and let W be the Weyl group (isomorphic to the permutation group S_n) of $\mathrm{GL}_n(\mathbb{Q}_p)$ with respect to its maximal torus. Define

$$P_w^+ := B \cap w P^+ w^{-1}, \quad w \in W.$$

MSC2010: primary 20G25, 22E50; secondary 20G05.
Keywords: rigid analytic representations, Tate algebra, affinoids, rigid analytic geometry, p-adic Langlands, p-adic representations, p-adic Lie groups.

Let K be a finite extension of $\mathbb{Q}_p$, let $\chi : T_0 \to K^\times$ be a locally analytic character with $\chi(t_1, \dots, t_n) = \chi_1(t_1) \cdots \chi_n(t_n)$, and $\chi_i(t) = t^{c_i} = e^{c_i \log(t)}$, where $c_i \in K$, for t sufficiently close to 1 in $\mathbb{Z}_p$ and where e is the exponential function.

Throughout this article, we use the following definitions.

(1) Let $\mathcal{A}_{\mathrm{loc}}(B, K)$ be the space of locally analytic functions on B with values in K. These are functions from B to K such that for any $x \in B$ we can find a ball $B_{r_x}(x)$ of radius r_x (depending on x) around x such that f can be written as a power series with coefficients in K which converges on $B_{r_x}(x)$ [Schneider 2011].

(2) Let $\mathrm{ind}_{P_0}^{B}(\chi)_{\mathrm{loc}} = \{f \in \mathcal{A}_{\mathrm{loc}}(B, K) : f(gb) = \chi(b^{-1}) f(g),\ b \in P_0,\ g \in B\}$.

(3) Let $\mathrm{ind}_{Q_0}^{G}(\chi)_{\mathrm{loc}} = \{f \in \mathcal{A}_{\mathrm{loc}}(G, K) : f(gb) = \chi(b^{-1}) f(g),\ b \in Q_0,\ g \in G\}$. Note that $B = UP_0$ and $G = UQ_0$, where U is the lower unipotent subgroup of B (or G). Therefore, we can obviously see that, as a vector space,

$$\mathrm{ind}_{P_0}^{B}(\chi)_{\mathrm{loc}} \cong \mathcal{A}_{\mathrm{loc}}(U, K) \cong \mathrm{ind}_{Q_0}^{G}(\chi)_{\mathrm{loc}}.$$

(4) Let d be the dimension of the p-adic Lie group U. Let $\mathcal{A}(U, K)$ be the space of globally analytic functions inside $\mathcal{A}_{\mathrm{loc}}(B, K)$. Any element $f \in \mathcal{A}(U, K)$ is of the form $f = \sum_{\nu \in \mathbb{N}^d} c_\nu a^\nu$ with $\lim_{|\nu| \to \infty} |c_\nu| = 0$. The space $\mathcal{A}(U, K)$ is a K-Banach space with the sup norm on f defined by

$$|f| = \sup |c_\nu|$$

(see [Bosch 2014, Chapter 2]). This is also known as the Tate algebra of globally analytic (sometimes called "rigid analytic") functions on U (see [Bosch 2014]). By (2) above, the vector space of globally analytic functions inside $\mathrm{ind}_{P_0}^{B}(\chi)_{\mathrm{loc}}$ (or $\mathrm{ind}_{Q_0}^{G}(\chi)_{\mathrm{loc}}$) is isomorphic to $\mathcal{A}(U, K)$.

(5) The action of G on the globally analytic vectors $\mathcal{A}(U, K)$ of $\mathrm{ind}_{P_0}^{B}(\chi)_{\mathrm{loc}}$ is given by the left translation $h \cdot f(g) \mapsto f(h^{-1}g),\ h \in G$.

(6) Recall that for any K-Banach space V with norm $|\cdot|$, a representation π of G on V is called a globally analytic representation if the map

$$g \mapsto g \cdot v = \pi(g)v$$

is globally analytic on G for all $v \in V$. Therefore, in coordinates $(x_1, \dots, x_l)$ with $l = \dim(G)$,

$$g \cdot v = \sum_k x^k v_k,$$

where $v_k \in V$ and $|v_k| \to 0$. Here $k = (k_1, \dots, k_l)$ and $x^k = x_1^{k_1} \cdots x_l^{k_l},\ k_i \in \mathbb{N}$ (see [Emerton 2017; Clozel 2018, Section 2]). For a detailed discussion on globally analytic representation, see [Emerton 2017].

(7) Write $\chi = (\chi_1, \dots, \chi_n)$, $\chi_i(1+pu_i) = e^{c_i \log(1+pu_i)}$ for $c_i \in K$, u_i close to 0, $i \in [1, n]$. The exponential is analytic (in K) in the domain $v_p(z) > e/(p-1)$ where $e = e(K)$ is the ramification index of K and v_p is the normalized valuation, $v_p(p) = 1$. Now,

$$v_p(c_i \log(1+pu_i)) = v_p(c_i) + 1 + v_p(u_i).$$

So we say that χ is analytic if and only if $v_p(c_i) > e/(p-1) - 1$ (see (3-10)).

Then, we show in Lemmas 3.2, 3.3 and 3.7 that the action of G on the *globally analytic vectors* $\mathcal{A}(U, K)$ of $\mathrm{ind}_{P_0}^B(\chi)_{\mathrm{loc}}$ is a globally analytic action of G, that is, it gives a globally analytic representation of G, in the sense of Emerton [2017].

Let μ be the linear form from the Lie algebra of the torus T_0 to K given by

$$\mu = (-c_1, \dots, c_n) : \mathrm{Diag}(t_1, \dots, t_n) \mapsto \sum_{i=1}^{n} -c_i t_i,$$

where $t = (t_i) \in \mathrm{Lie}(T_0)$ and $c_i \in K$. For negative root $\alpha = (i, j)$, $i > j$, let $H_{(i,j)} = E_{i,i} - E_{j,j}$ where $E_{i,i}$ is the standard elementary matrix.

Using Theorems 3.8 and 3.9, we will show the following:

Theorem. *Assume $p > n+1$ and χ is analytic. Then the space of globally analytic vectors of* $\mathrm{ind}_{P_0}^B(\chi)_{\mathrm{loc}}$ *is an admissible and globally analytic representation of G. Furthermore, the space of globally analytic vectors of* $\mathrm{ind}_{P_0}^B(\chi)_{\mathrm{loc}}$ *is irreducible if and only if $-\mu(H_\alpha) + i - j \notin \{1, 2, 3, \dots\}$ for all $\alpha = (i, j) \in \Phi^-$.*

Here, the admissibility is in the sense of [Emerton 2017] (see also [Clozel 2018, Section 2.3]).

For global analyticity, we compute explicitly the action of G on the Tate algebra of globally analytic functions of $\mathrm{ind}_{P_0}^B(\chi)_{\mathrm{loc}}$ and show that the action map is a globally analytic function on G seen as a rigid analytic space. For this, we have to do a lot of new technical computations, which were not necessary for the GL(2) case by Clozel [2018]. In particular, see Lemmas 3.4–3.7. For the irreducibility we first use the action of the Lie algebra of G to show that any nonzero closed G-invariant subspace of the globally analytic vectors of $\mathrm{ind}_{P_0}^B(\chi)_{\mathrm{loc}}$ contains the constant function 1. Unlike the GL(2) case, the remaining part of the argument for the proof of irreducibility uses the notion of Verma modules and its condition of irreducibility, a result of Bernstein, Gelfand and Gelfand.

Strikingly, one can check easily that the condition that we obtain for irreducibility of the globally analytic principal series is the same condition as the irreducibility of the locally analytic principal series $\mathrm{ind}_{P_0}^B(\chi)_{\mathrm{loc}}$, deduced by Schneider and Teitelbaum [2002, Theorems 5.4 and 5.6, Corollary 5.7] for GL(2) and Orlik and Strauch for GL(n) [Orlik and Strauch 2010, Theorems 3.4.12 and 4.1.1] although our style of proof is completely different from their proof. Note that in Theorem 4.1.1 of [Orlik and Strauch 2010], there is a typo where the authors quote the result in [Dixmier 1977] of the irreducibility of a Verma module due to Bernstein, Gelfand and Gelfand. For the correct result of irreducibility one should look at [Dixmier 1977, Theorem 7.6.24].

The preliminaries are included in Section 2. The main result is discussed in Section 3A. In Section 3B, we extend these results to the pro-p Iwahori group of $\mathrm{GL}_n(L)$ where L is an unramified finite extension of $\mathbb{Q}_p$. Then, in Theorem 3.19, we use the Steinberg tensor product [1963] to construct base change in the context of Langlands functoriality.

In Section 4, we deal with the globally analytic vectors induced from the Weyl orbit of the upper triangular Borel subgroup of the Iwahori subgroup B, i.e., the globally analytic vectors of $\mathrm{ind}_{P_w^+}^{B}(\chi^w)_{\mathrm{loc}}$, where $\chi_w(h) = \chi(w^{-1}hw)$.

Our work is just the tip of an iceberg in the domain of globally analytic representations and it leads to a plethora of future questions; some of them are discussed at the end of Section 4.

2. Base change maps for analytic functions

We introduce the basic notions of rigid analytic geometry, including a brief discussion on the restriction of scalars. Then we briefly recall (following [Clozel 2018]) the notions of holomorphic and Langlands base change functors from a globally analytic representation over $\mathbb{Q}_p$ to a representation over L. The Langlands base change is related to the "Steinberg tensor product" described at the end of Section 1.1 of [Clozel 2017] for GL(2).

2A. Let L be a finite unramified extension of $\mathbb{Q}_p$ of degree N, with ring of integers $\mathcal{O}_L$. Given a formal scheme $\mathbb{X}_{\mathcal{O}_L}$ over $\mathcal{O}_L$, Bertapelle [2000] constructs a Weil restriction functor which associates to $\mathbb{X}_{\mathcal{O}_L}$ another formal scheme over $\mathbb{Z}_p$. Let $\mathbb{X}_L$ be the rigid analytic space associated to the formal scheme $\mathbb{X}_{\mathcal{O}_L}$. Bertapelle's construction gives a Weil restriction functor (we will call it as restriction of scalars) which associates to $\mathbb{X}_L$ another rigid analytic space $\mathrm{Res}_{L/\mathbb{Q}_p}(\mathbb{X}_L)$ over $\mathbb{Q}_p$. Although we will not recall the construction of this functor for general rigid analytic spaces and formal schemes, we do recall how this functor behaves with respect to affinoid rigid analytic spaces which is what we will need in this article. An interested reader should consult [Bertapelle 2000] for the most general construction of this restriction functor. Let (B^1/L) be the (rigid analytic affinoid) closed unit ball over L with its Tate algebra of analytic functions $\mathcal{T}_L = L\langle x\rangle$ and G_L be a rigid analytic group isomorphic as a rigid analytic space to $(B^1/L)^d$ which is a rigid analytic space with affinoid algebra $\mathcal{A}(G_L) := \widehat{\otimes}^d \mathcal{T}_L := \mathcal{T}_d(L) = L\langle x_1, \ldots, x_d\rangle$, the Tate algebra of analytic functions in d variables with coefficients in L. The restriction of scalars functor [Bertapelle 2000] associates to G_L a rigid analytic space $\mathrm{Res}_{L/\mathbb{Q}_p} G_L$ over $\mathbb{Q}_p$. In general, this functor does not behave trivially, but L being unramified, we obtain

$$\mathrm{Res}_{L/\mathbb{Q}_p}(B^1/L) \cong (B^1/\mathbb{Q}_p)^N,$$

[Clozel 2018, Lemma 1.1] which is canonically obtained by the choice of a basis (e_i) of $\mathcal{O}_L$ over $\mathbb{Z}_p$. This is defined in the following way. For an affinoid $\mathbb{Q}_p$-algebra $\mathcal{B}$ and for $f \in \mathrm{Hom}_L(L\langle x\rangle, \mathcal{B} \otimes_{\mathbb{Q}_p} L)$ with $f(x) = \sum b_i e_i \ \ (b_i \in \mathcal{B})$, we canonically define a function $g \in \mathrm{Hom}_{\mathbb{Q}_p}(\mathbb{Q}_p\langle x_1, \ldots, x_N\rangle, \mathcal{B})$ with $g(x_i) = b_i$ which is given by

$$g(x_1, \ldots, x_N) = f\Big(\sum e_i x_i\Big) \tag{2-1}$$

([Clozel 2018, Section 1.1]; see also [Bertapelle 2000, Proposition 1.8]). Since e_i is integral and $|x_i| \le 1$ it is easy to see that the series on the right converges. As the restriction of scalars is compatible with direct products [Bertapelle 2000, Proposition 1.8], $\mathrm{Res}_{L/\mathbb{Q}_p} G_L = (B^1/\mathbb{Q}_p)^{dN}$. Henceforth, we write $\mathrm{Res}\, G_L$ to denote $\mathrm{Res}_{L/\mathbb{Q}_p} G_L$.

2B. Assume now that $G_L \cong (B^1/L)^d$ is obtained by *extension* of scalars from $\mathbb{Q}_p$. Then, the Tate algebra $\mathcal{A}(G_L)$ is equal to $\mathcal{A}(G_{\mathbb{Q}_p}) \otimes L$. The comultiplication map m^*, defined by a morphism

$$m^* : \mathcal{A}(G_L) \to \mathcal{A}(G_L) \widehat{\otimes} \mathcal{A}(G_L)$$

with image inside the completed tensor product, is obtained by extension of scalars from

$$m_0^* : \mathcal{A}(G_{\mathbb{Q}_p}) \to \mathcal{A}(G_{\mathbb{Q}_p}) \widehat{\otimes} \mathcal{A}(G_{\mathbb{Q}_p}).$$

Note that (2-1) associates to $f \in \mathcal{A}(G_L)$ (with L-coefficients, i.e., in $\mathcal{T}_d(L)$) a function $g \in \mathcal{A}(\mathrm{Res}\, G_L) \otimes L$ (the function g given in (2-1) will have coefficients in L). In particular, we get a map $\mathcal{A}(G_{\mathbb{Q}_p}) \to \mathcal{A}(\mathrm{Res}\, G_L) \otimes L$ in composing with the "tautological map"

$$\mathcal{A}(G_{\mathbb{Q}_p}) \to \mathcal{A}(G_L).$$

This gives us the map

$$b_1 : \mathcal{A}(G_{\mathbb{Q}_p}) \to \mathcal{A}(\mathrm{Res}\, G_L) \otimes L, \tag{2-2}$$

and we call it as a "holomorphic base change" map. The Galois group

$$\Sigma = \mathrm{Gal}(L/\mathbb{Q}_p)$$

naturally acts on L which induces a natural action of Σ on the pro-p Iwahori with coefficients in L. This action is $\mathbb{Q}_p$-linear. Therefore Σ acts naturally on G_L with a $\mathbb{Q}_p$-linear action. Note also that the action of Σ on $\mathrm{Res}\, G_L$ is $\mathbb{Q}_p$-linear. Recall from (2-2) that b_1 sends $\mathcal{A}(G_{\mathbb{Q}_p})$ to the functions that are L-holomorphic (given by power series $\sum a_m \underline{x}^m, \ a_m \in L$, with $\underline{x} = (x_1, \ldots, x_d)$ being the variable).

Let $\sigma \in \Sigma$. Then the action associated to σ sends a power series in $\mathcal{A}(G_{\mathbb{Q}_p})$ to $\left(\sum a_m \underline{x}^m\right)^\sigma := \sum \sigma(a_m)\underline{x}^m$. This gives rise to a map

$$b : \mathcal{A}(G_{\mathbb{Q}_p}) \to \mathcal{A}(\operatorname{Res} G_L) \otimes L, \tag{2-3}$$

$$b(f) = \prod_{\sigma\in\Sigma} b_1(f)^\sigma. \tag{2-4}$$

We now consider all Tate algebras to have coefficients in L and we denote them by $\mathcal{A}_L$. That is, $\mathcal{A}_L(\operatorname{Res} G_L) = \mathcal{A}(\operatorname{Res} G_L) \otimes L$ and $\mathcal{A}_L(G_L) := \mathcal{A}(G_L)$. Clozel [2018, Proposition 1.5] has shown that the map b constructed in (2-3) is actually a tensor product and gives rise to an isomorphism $\mathcal{A}_L(\operatorname{Res} G_L) \cong \widehat{\otimes}_\sigma \mathcal{A}_L(G_L)$.

Fix a finite extension K of $\mathbb{Q}_p$ and an injection $i : L \subset K$. If $\sigma \in \operatorname{Gal}(L/\mathbb{Q}_p)$, we then have the injection

$$i \circ \sigma : L \to K.$$

Denote by V a (globally) analytic representation of $G_{\mathbb{Q}_p}$ on a K- Banach space. Then V naturally extends to an analytic representation of G_L; this is called the *holomorphic base change* of V in [Clozel 2018]. For $\sigma \in \operatorname{Gal}(L/\mathbb{Q}_p)$, write V^σ the representation of G_L associated to $i \circ \sigma$. Then, the *full (Langlands) base change* of V is defined to be the globally analytic representation of $\operatorname{Res}_{L/\mathbb{Q}_p}(G_L)$ on $\widehat{\otimes}_\sigma V^\sigma$ (see [Clozel 2018, Definition 3.2]).

3. Globally analytic principal series for GL(n)

We first recall the notion of locally analytic principal series representation induced from the Borel to the Iwahori subgroup of $\mathrm{GL}(n, \mathbb{Z}_p)$. Then we treat the action of the pro-p Iwahori on the subspace of rigid analytic functions within the locally analytic principal series and show that this action is a globally analytic action (Theorem 3.8). This gives us the globally analytic induced principal series representation under the pro-p Iwahori subgroup G. Furthermore, we treat the condition of irreducibility of the globally analytic principal series by translating an irreducibility condition of a suitable Verma module (Theorem 3.9). Finally in Section 3B we base change our globally analytic representation to a finite unramified extension L of $\mathbb{Q}_p$.

3A. We consider the case of principal series for $\mathrm{GL}_n(\mathbb{Z}_p)$. Denote by G the pro-p Iwahori subgroup of $\mathrm{GL}_n(\mathbb{Z}_p)$, i.e., the group of matrices in $\mathrm{GL}_n(\mathbb{Z}_p)$ that are lower unipotent modulo $p\mathbb{Z}_p$ and by B the subgroup of matrices in $\mathrm{GL}_n(\mathbb{Z}_p)$ which are lower triangular modulo $p\mathbb{Z}_p$. Let $P_0 \supset T_0$ be the set of upper triangular and diagonal matrices in B, and let $\chi : T_0 \to K^\times$ be a locally analytic character with

$$\chi(t_1, \dots, t_n) = \chi_1(t_1) \cdots \chi_n(t_n),$$

and $\chi_i(t) = t^{c_i}$. Here t^{c_i} is the exponential $e^{c_i \log(t)}$ where $c_i \in K$ for t sufficiently close to 1 in $\mathbb{Z}_p$.

We first consider the locally analytic induced representation of B,

$$J_{\text{loc}} = \text{ind}_{P_0}^{B}(\chi)_{\text{loc}} = \{f \in \mathcal{A}_{\text{loc}}(B, K) : f(gb) = \chi(b^{-1})f(g),\ b \in P_0,\ g \in B\},$$

where χ is naturally extended to P_0 and $\mathcal{A}_{\text{loc}}(B, K)$ is the space of locally analytic functions on B. With U the lower unipotent subgroup of B with entries in $\mathbb{Z}_p$ in the lower triangular part, 1 in the diagonal entries and 0 elsewhere, we have the natural decomposition

$$B = U P_0.$$

Since χ is fixed and G is an open normal subgroup of B, the restriction of the functions of J_{loc} to $G \subset B$ is injective [Clozel 2018, Section 3.3]. With $Q_0 = P_0 \cap G$, we deduce that the vector space of J_{loc} is

$$I_{\text{loc}} = \{f \in \mathcal{A}_{\text{loc}}(G, K) : f(gb) = \chi(b^{-1})f(g),\ b \in Q_0,\ g \in G\}. \tag{3-1}$$

With the decomposition $G = U Q_0$, we see that $I_{\text{loc}} \cong \mathcal{A}_{\text{loc}}(\mathbb{Z}_p^{(n(n-1))/2}, K) = \mathcal{A}_{\text{loc}}(U, K)$. Here, $\mathbb{Z}_p$ is seen as the rigid analytic (additive) group $B^1(\mathbb{Z}_p)$. The group G acts on I_{loc} by the left translation

$$h \cdot f(g) \mapsto f(h^{-1}g). \tag{3-2}$$

Let $E_{i,j}$ be the elementary matrices with 1 in the (i, j)-th place and 0 elsewhere. From now on, we assume

$$p > n + 1;$$

then G is p-saturated in the sense of [Lazard 1965, III, 3.2.7.5] and thus, it is the ordered product (as a rigid analytic group) of the following one-parameter subgroups:

(1) First, for $y \in \mathbb{Z}_p$, take the one-parameter lower unipotent matrices by the following lexicographic order: the one-parameter group of matrices $(1 + yE_{i,j})$ comes before the one-parameter group of matrices$(1 + yE_{k,l})$ if and only if $i < k$ or $i = k$ and $j < l$. Note that $(1 + yE_{i,j})$ and $(1 + yE_{k,l})$ are lower unipotent, and hence $i > j$ and $k > l$.

(2) Then, for $t_k \equiv 1[p]$ and $k \in [1, n]$, take the one-parameter diagonal subgroups $(t_k E_{k,k} + \sum_{i=1, i\neq k}^{n} E_{i,i})$ starting from the top left extreme to the low right extreme.

(3) Finally, for $y \in p\mathbb{Z}_p$, take the upper unipotent matrices in the following order: the one-parameter group of matrices $(1 + yE_{i,j})$ comes before the one-parameter group of matrices $(1 + yE_{k,l})$ if and only if $i \geq k$ or $i = k$ and $j > l$. Note that $(1 + yE_{i,j})$ and $(1 + yE_{k,l})$ are upper unipotent, and hence $i < j$ and $k < l$.

That is, for the lower unipotent matrices, we start with the top and left extreme and then fill the lines from the left, going down and for the upper unipotent matrices we start with the low and right extreme and then fill the lines from the right, going up. (See [Lazard 1965, III, 3.3.2] for the rigid analyticity and see Theorem 2.2.1 and Remark 2.2.2 of [Ray 2020] for the order of the product, i.e., an ordered Lazard basis of G, although in [Ray 2020] we have taken G to be upper unipotent matrices modulo p but this does not matter).

Let now

$$\mathcal{A} = \mathcal{A}(U, K) = \mathcal{A}(\mathbb{Z}_p^{(n(n-1))/2}, K)$$

be the subspace of globally analytic functions of

$$I_{\mathrm{loc}} = \mathcal{A}_{\mathrm{loc}}(U, K).$$

Thus $f \in \mathcal{A}$ is a globally analytic function in the variables $a_{i,j}$ on U, that is,

$$f(A) = \sum_{\nu \in \mathbb{N}^d} c_\nu a^\nu$$

such that $c_\nu \in K$ and $|c_\nu| \to 0$ as $|\nu| \to \infty$. Here $d = (n(n-1))/2$, $a = (a_{2,1}, a_{3,1}, a_{3,2}, \ldots, a_{n,n-1}) \in \mathbb{Z}_p^d$ with the lexicographic ordering of $a_{i,j}$ as in (1), $\nu = (\nu_{2,1}, \nu_{3,1}, \ldots, nu_{n,n-1}) \in \mathbb{N}^d$, $a^\nu = a_{2,1}^{\nu_{2,1}} \cdots a_{n,n-1}^{\nu_{n,n-1}}$ and $|\nu| = \nu_{2,1} + \cdots + \nu_{n,n-1}$.

We now seek conditions such that if f is a globally analytic function on G and the action of G is defined as above, then the map

$$h \mapsto h \cdot f(g) = f(h^{-1}g)$$

is globally analytic.

Lemma 3.1. *With the above notation, for $p > n+1$, the action of G on $f \in \mathcal{A}(U, K)$, i.e., the map $h \mapsto h \cdot f$, is a globally analytic function on G if and only if it is so for all one-parameter (rigid analytic) subgroups and the diagonal subgroup of which G is the product.*

Proof. This follows from the same argument as in the discussion after Lemma 3.4 of [Clozel 2018]. □

Thus, our goal is to verify the analyticity of the action of the diagonal subgroup, the one-parameter lower unipotent subgroups and the one-parameter upper unipotent subgroups of G which are treated in Lemmas 3.2, 3.3 and 3.7, respectively.

Let $A = (a_{i,j})_{i,j}$ be any matrix in U (i.e., $a_{i,i} = 1$ and $a_{i,j} = 0$ for $i < j$) and

$$T = \mathrm{Diag}(t_1, \ldots, t_n) = \sum_{k=1}^{n} t_k E_{k,k}$$

be any element in the diagonal $T_0 \cap G$, where $t_k \in 1 + p\mathbb{Z}_p$. Assume $f \in I_{\mathrm{loc}}$, then the action of T on f, given by (3-2), is

$$\begin{aligned} T\cdot f(A) &= f(\mathrm{Diag}(t_1^{-1},\ldots,t_n^{-1})A) \\ &= f\Big(\Big(\sum_{k=1}^{n} t_k^{-1}E_{k,k}\Big)\Big(\sum_{i,j=1}^{n} a_{i,j}E_{i,j}\Big)\Big) \\ &= f\Big(\sum_{j,k=1}^{n} t_k^{-1}a_{k,j}E_{k,j}\Big) \\ &= f\Big(\Big(\sum_{k,j=1}^{n} t_k^{-1}t_j a_{k,j}E_{k,j}\Big)\Big(\sum_{j=1}^{n} t_j^{-1}E_{j,j}\Big)\Big) \\ &= f\Big(\sum_{k,j=1}^{n} t_k^{-1}t_j a_{k,j}E_{k,j}\Big)\chi(t_1,\ldots,t_n) \qquad \text{(from (3-1))}. \end{aligned}$$

Interchanging indices $k\to i$, we obtain

$$\Big(\sum_{i=1}^{n} t_i E_{i,i}\Big)\cdot f\Big(\sum_{i,j=1}^{n} a_{i,j}E_{i,j}\Big) = f\Big(\sum_{i,j=1}^{n} t_i^{-1}t_j a_{i,j}E_{i,j}\Big)\chi(t_1,\ldots,t_n) \tag{3-3}$$

with $a_{i,i}=1,\ a_{i,j}=0$ for $i<j$ and $t_i\equiv 1 \pmod p$.

Taking $f=1$ we see that $\chi(t_1,\ldots,t_n)$ must be an analytic function. By (3-3), for fixed $k\in[1,n]$ considering the action of the matrix $\big(t_kE_{k,k}+\sum_{i=1,i\neq k}^{n}E_{i,i}\big)$ on f we obtain

$$\begin{aligned} &\Big(t_kE_{k,k}+\sum_{i=1,i\neq k}^{n}E_{i,i}\Big)f(A) \\ &\quad= f\Big(\sum_{\substack{u,v\neq k\\ u>v}} a_{u,v}E_{u,v}+a_{k,k}E_{k,k}+\sum_{j=1}^{k-1}t_k^{-1}a_{k,j}E_{k,j}+\sum_{i=k+1}^{n}t_ka_{i,k}E_{i,k}\Big) \\ &\qquad\times\chi(1,\ldots,t_k,\ldots,1) \\ &\quad:= f(\mathcal{C})\chi(1,\ldots,t_k,\ldots,1) \end{aligned} \tag{3-4}$$

where $\mathcal{C}$ is the matrix

$$\Big(\sum_{\substack{u,v\neq k\\ u>v}} a_{u,v}E_{u,v}+a_{k,k}E_{k,k}+\sum_{j=1}^{k-1}t_k^{-1}a_{k,j}E_{k,j}+\sum_{i=k+1}^{n}t_ka_{i,k}E_{i,k}\Big).$$

Assume now that f is globally analytic in the variables $a_{i,j}$ on U, that is,

$$f(A)=\sum_{\nu\in\mathbb{N}^d} c_\nu a^\nu, \tag{3-5}$$

such that $c_\nu \in K$ and $|c_\nu| \to 0$. Then with $t_k = 1 + p\xi_k,\ \xi_k \in \mathbb{Z}_p$,

$$f(\mathcal{C}) = \sum_\nu c_\nu \Bigl(a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}\Bigr)\Bigl(\prod_{j=1}^{k-1} (t_k^{-1} a_{k,j})^{\nu_{k,j}}\Bigr)\Bigl(\prod_{i=k+1}^{n} (t_k a_{i,k})^{\nu_{i,k}}\Bigr) \tag{3-6}$$

$$= \sum_\nu c_\nu \Bigl(a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}\Bigr)\Bigl(\prod_{j=1}^{k-1} (1+p\xi_k)^{-\nu_{k,j}} a_{k,j}^{\nu_{k,j}}\Bigr)\Bigl(\prod_{i=k+1}^{n} (1+p\xi_k)^{\nu_{i,k}} a_{i,k}^{\nu_{i,k}}\Bigr). \tag{3-7}$$

Recall that for $|v| < 1$, $m \in \mathbb{N}$, we have $(1-v)^{-m} = \sum_{q=0}^{\infty} \binom{m+q-1}{q} v^q$. Now, inserting the expressions

$$(1+p\xi_k)^{-\nu_{k,j}} = \sum_{q_{k,j}=0}^{\infty} \binom{\nu_{k,j}+q_{k,j}-1}{q_{k,j}} (-p\xi_k)^{q_{k,j}}$$

and

$$(1+p\xi_k)^{\nu_{i,k}} = \sum_{u_{i,k}=0}^{\nu_{i,k}} \binom{\nu_{i,k}}{u_{i,k}} p^{u_{i,k}} \xi_k^{u_{i,k}}$$

into (3-7) we obtain, with $|q| := q_{k,1} + \cdots + q_{k,k-1}$, $|u| = u_{k+1,k} + \cdots + u_{n,k}$ and $v_{\max} = \prod_{i=k+1}^{n} \nu_{i,k}$,

$$\begin{aligned} f(\mathcal{C}) &= \sum_\nu c_\nu \Bigl(a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}\Bigr)\Bigl(\prod_{j=1}^{k-1}\Bigl(\sum_{q_{k,j}=0}^{\infty} \binom{\nu_{k,j}+q_{k,j}-1}{q_{k,j}} (-p\xi_k)^{q_{k,j}} a_{k,j}^{\nu_{k,j}}\Bigr)\Bigr) \\ &\qquad \times \Bigl(\prod_{i=k+1}^{n} \sum_{u_{i,k}=0}^{\nu_{i,k}} \binom{\nu_{i,k}}{u_{i,k}} p^{u_{i,k}} \xi_k^{u_{i,k}} a_{i,k}^{\nu_{i,k}}\Bigr) \\ &= \sum_\nu c_\nu \Bigl(a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}\Bigr)\Bigl(\sum_{N\geq 0}^{\infty} \xi_k^N \Bigl(\sum_{|q|=N} \prod_{j=1}^{k-1} \binom{\nu_{k,j}+q_{k,j}-1}{q_{k,j}} (-p)^{q_{k,j}} a_{k,j}^{\nu_{k,j}}\Bigr)\Bigr) \\ &\qquad \times \Bigl(\sum_{M=0}^{v_{\max}} \xi_k^M \Bigl(\sum_{|u|=M} \prod_{i=k+1}^{n} \binom{\nu_{i,k}}{u_{i,k}} p^{u_{i,k}} a_{i,k}^{\nu_{i,k}}\Bigr)\Bigr). \end{aligned}$$

Let f_N and g_M be defined by

$$f_N = \Bigl(\sum_{|q|=N} \prod_{j=1}^{k-1} \binom{\nu_{k,j}+q_{k,j}-1}{q_{k,j}} (-p)^{q_{k,j}} a_{k,j}^{\nu_{k,j}}\Bigr), \tag{3-8}$$

$$g_M = \Bigl(\sum_{|u|=M} \prod_{i=k+1}^{n} \binom{\nu_{i,k}}{u_{i,k}} p^{u_{i,k}} a_{i,k}^{\nu_{i,k}}\Bigr). \tag{3-9}$$

Then,

$$f(\mathcal{C}) = \sum_{\nu} c_\nu (a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}) \Big(\sum_{N\geq 0}^{\infty} \xi_k^N f_N\Big)\Big(\sum_{M=0}^{\nu_{\max}} \xi_k^M g_M\Big)$$
$$= \sum_{m\geq 0}^{\infty} \xi_k^m \Big(\sum_{\nu} c_\nu (a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}) \sum_{N+M=m} f_N g_M\Big).$$

Recall from the introduction that any element $f \in \mathcal{A}(U, K)$ is of the form

$$f = \sum_{\nu\in\mathbb{N}^d} c_\nu a^\nu$$

with $\lim_{|\nu|\to\infty} |c_\nu| = 0$. The space $\mathcal{A}(U, K)$ is a K-Banach space with the sup norm on f defined by

$$|f| = \sup|c_\nu|$$

(see [Bosch 2014, Chapter 2]). Recall that for any K-Banach space V with norm $|\cdot|$, a representation π of G on V is called a globally analytic representation if the map

$$g \mapsto g\cdot v = \pi(g)v$$

is globally analytic on G for all $v \in V$. Thus, in coordinates $(x_1, \dots, x_l)$ with $l = \dim(G)$,

$$g\cdot v = \sum_k x^k v_k$$

where $v_k \in V$ and $|v_k| \to 0$. Here $k = (k_1, \dots, k_l)$ and $x^k = x_1^{k_1} \cdots x_l^{k_l}$, $k_i \in \mathbb{N}$ (see [Emerton 2017; Clozel 2018, Section 2]).

Now, with $t_k = 1 + p\xi_k$, $\xi_k \in \mathbb{Z}_p$, in order to show that the action of the one-parameter diagonal subgroup $t_k(E_{k,k}) + \sum_{i\neq k,i=1}^{n} E_{i,i}$ on $f \in \mathcal{A}(U, K)$ is analytic we have to show that the map

$$\mathbb{Z}_p \to \mathcal{A}(U, K),$$
$$\xi_k \mapsto \Big((1 + p\xi_k)E_{k,k} + \sum_{i=1,i\neq k}^{n} E_{i,i}\Big) f = f(\mathcal{C})\chi(1, \dots, 1 + p\xi_k, \dots, 1)$$

is a globally analytic map on $\mathbb{Z}_p$. The norm of the coefficient of ξ_k^m, in (3-10), is

$$\Big|\Big(\sum_{\nu} c_\nu \Big(a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}\Big) \sum_{N+M=m} f_N g_M\Big)\Big|.$$

Notice that, since $N, M \leq m$ and $f_N, g_M \in \mathbb{Z}_p$ from (3-8) and (3-9), the quantity $\big(a_{k,k}^{\nu_{k,k}} \prod_{u,v\neq k,\, u>v} a_{u,v}^{\nu_{u,v}} \sum_{N+M=m} f_N g_M\big)$ has finite sum and product and hence lies in $\mathbb{Z}_p$. Hence,

$$\left|\left(\sum_{\nu} c_\nu \left(a_{k,k}^{\nu_{k,k}} \prod_{\substack{u,v\neq k\\ u>v}} a_{u,v}^{\nu_{u,v}}\right) \sum_{N+M=m} f_N g_M\right)\right| \to 0$$

as $|c_\nu| \to 0$ with $\nu \to \infty$. This gives the analyticity of the action $f \to f(\mathcal{C})$. We treat the analyticity of the character χ in general. Write $\chi = (\chi_1, \ldots, \chi_n)$, $\chi_i(1+pu_i) = e^{c_i \log(1+pu_i)}$ for $c_i \in K$, u_i close to 0, $i \in [1, n]$. The exponential is analytic (in K) in the domain $v_p(z) > e/(p-1)$ where $e = e(K)$ is the verification index and v_p is the normalized valuation, $v_p(p) = 1$. Now,

$$v_p(c_i \log(1+pu_i)) = v_p(c_i) + 1 + v_p(u_i).$$

So we must have

$$v_p(c_i) + 1 > \frac{e}{p-1};$$

that is,

$$v_p(c_i) > \frac{e}{p-1} - 1. \tag{3-10}$$

We say that χ is "analytic" if and only if the c_i's verify the condition (3-10) and in the rest of this text we assume that our character χ is analytic. It is easy to see that if χ is analytic, then $\chi(1, \ldots, 1+p\xi_k, \ldots, 1)$ is an analytic function on ξ_k. The character

$$\begin{aligned}\chi(1,\ldots,1+p\xi_k,\ldots,1) = \chi_k(1+p\xi_k) &= \sum_{n=0}^{\infty} c_n (1+p\xi_k)^n \quad \text{(since } \chi_k \text{ is analytic)}\\ &= \sum_{n=0}^{\infty} c_n \sum_{u=0}^{n} \binom{n}{u} p^u \xi_k^u\\ &= \sum_{u=0}^{\infty} \xi_k^u \left(p^u \sum_{n\geq u}^{\infty} c_n \binom{n}{u}\right).\end{aligned}$$

The norm of the coefficient of ξ_k^u is

$$\left| p^u \sum_{n\geq u}^{\infty} c_n \binom{n}{u} \right|$$

which goes to 0 as $|c_n| \to 0$ with $n \to \infty$. Thus, we have shown:

Lemma 3.2. *Under the hypothesis* (3-10), *for each* $k \in [1, n]$, *the action of the one-parameter diagonal subgroup* $\left(t_k E_{k,k} + \sum_{i=1, i\neq k}^{n} E_{i,i}\right)$ *of* G *on* $\mathcal{A}(\mathbb{Z}_p^{(n(n-1))/2}, K)$ *given by* (3-4) *is an analytic action.*

For $y \in \mathbb{Z}_p$ and $i > j$, with i, j fixed between $1, \dots, n$, the action of the one-parameter (rigid analytic) subgroup $(1 + yE_{i,j})$ on $f(A)$, given by (3-2) is

$$(1 + yE_{i,j})f(A) = f((1 + yE_{i,j})^{-1}A) = f((1 - yE_{i,j})A) = f\left((1 - yE_{i,j})\left(\sum_{\substack{k \geq l \\ k,l \in [1,n]}} a_{k,l}E_{k,l}\right)\right) = f\left(\sum_{\substack{k \geq l \\ k,l \in [1,n]}} a_{k,l}E_{k,l} - \sum_{l=1,\dots,j} ya_{j,l}E_{i,l}\right) := f(\mathcal{B}), \tag{3-11}$$

where $\mathcal{B}$ is the matrix

$$\sum_{k \geq l,\ k,l \in [1,n]} a_{k,l}E_{k,l} - \sum_{l=1}^{j} ya_{j,l}E_{i,l}.$$

One can easily see that the matrix $\mathcal{B} = (b_{u,v})$ is lower unipotent and differs from matrix A only in the first j entries of its i-th row. In particular,

$$b_{i,v} = a_{i,v} - ya_{j,v}$$

for all $v \in [1, j]$, $a_{j,j} = 1$, and all other $b_{u,v}$ are the same as $a_{u,v}$ (recall that A is lower unipotent).

Now, let f be a globally analytic function on U as in (3-5). That is

$$f(a) = \sum_{\nu \in \mathbb{N}^d} c_\nu a^\nu$$

with $a^\nu = a_{2,1}^{\nu_{2,1}} \cdots a_{n,n-1}^{\nu_{n,n-1}}$ and $|c_\nu| \to 0$. Then, we have to show that

$$(1 + yE_{i,j})f = f(\mathcal{B})$$

gives an analytic map

$$\mathbb{Z}_p \to \mathcal{A}(U, K),$$
$$y \to (1 + yE_{i,j})f = f(\mathcal{B}).$$

The power series $f(\mathcal{B})$ is equal to

$$\sum_\nu c_\nu \left(\left(\prod_{\substack{u>v \\ u=i \Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\right)\left(\prod_{k=1}^{j}(a_{i,k} - ya_{j,k})^{\nu_{i,k}}\right)\right).$$

For each $k \in [1, j]$, inserting the expansion

$$(a_{i,k} - ya_{j,k})^{\nu_{i,k}} = \sum_{m_{i,k}=0}^{\nu_{i,k}} \binom{\nu_{i,k}}{m_{i,k}} y^{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}$$

into $f(\mathcal{B})$ we obtain, for $M=\prod_{k=1}^{j}\nu_{i,k},\ |m|=m_{i,1}+\cdots+m_{i,j}$,

$$\begin{aligned}
f(\mathcal{B}) &= \sum_{\nu} c_\nu \Biggl(\Biggl(\prod_{\substack{u>v\\ u=i\Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\Biggr)\Biggl(\prod_{k=1}^{j}\Biggl(\sum_{m_{i,k}=0}^{\nu_{i,k}}\binom{\nu_{i,k}}{m_{i,k}} y^{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}\Biggr)\Biggr)\Biggr)\\
&= \sum_{\nu} c_\nu \Biggl(\Biggl(\prod_{\substack{u>v\\ u=i\Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\Biggr)\Biggl(\sum_{N=0}^{M} y^N\Biggl(\sum_{\substack{|m|=N\\ m_{i,\star}\in[0,\nu_{i,\star}]}}\prod_{k=1}^{j}\binom{\nu_{i,k}}{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}\Biggr)\Biggr)\Biggr)\\
&= \sum_{\nu} c_\nu \Biggl(\sum_{N=0}^{M} y^N\Biggl(\Biggl(\prod_{\substack{u>v\\ u=i\Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\Biggr)\sum_{\substack{|m|=N\\ m_{i,\star}\in[0,\nu_{i,\star}]}}\prod_{k=1}^{j}\binom{\nu_{i,k}}{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}\Biggr)\Biggr)\\
&= \sum_{N=0}^{M} y^N\Biggl(\sum_{\nu} c_\nu \Biggl(\Biggl(\prod_{\substack{u>v\\ u=i\Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\Biggr)\sum_{\substack{|m|=N\\ m_{i,\star}\in[0,\nu_{i,\star}]}}\prod_{k=1}^{j}\binom{\nu_{i,k}}{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}\Biggr)\Biggr)\\
&= \sum_{N=0}^{M} y^N f_N,
\end{aligned}$$

where

$$f_N := \sum_{\nu} c_\nu \Biggl(\Biggl(\prod_{\substack{u>v\\ u=i\Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\Biggr)\sum_{\substack{|m|=N\\ m_{i,\star}\in[0,\nu_{i,\star}]}}\prod_{k=1}^{j}\binom{\nu_{i,k}}{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}\Biggr).$$

Define

$$s(N,\nu) := \Biggl(\Biggl(\prod_{\substack{u>v\\ u=i\Rightarrow v>j}} a_{u,v}^{\nu_{u,v}}\Biggr)\sum_{\substack{|m|=N\\ m_{i,\star}\in[0,\nu_{i,\star}]}}\prod_{k=1}^{j}\binom{\nu_{i,k}}{m_{i,k}}(-a_{j,k})^{m_{i,k}} a_{i,k}^{\nu_{i,k}-m_{i,k}}\Biggr)$$

such that $f_N=\sum_\nu c_\nu s(N,\nu)$. Notice that since $m_{i,k}\le \nu_{i,k}$ for all $k\in[1,j]$, the sum in $s(N,\nu)$ is a finite sum and thus $s(N,\nu)$ lies in $\mathbb{Z}_p$. Therefore, the norm of the coefficient of y^N is $|f_N|=\bigl|\sum_\nu c_\nu s(N,\nu)\bigr|$ which goes to 0 as $|c_\nu|\to 0$ with $|\nu|\to\infty$. This gives the analyticity of the map $y\to(1+E_{i,j})f=f(\mathcal{B})$.

Therefore, we have shown:

Lemma 3.3. *For $y\in\mathbb{Z}_p$ and $i>j$, the action of the lower unipotent (rigid analytic) one-parameter subgroup $(1+yE_{i,j})$ of G on $f\in\mathcal{A}(\mathbb{Z}_p^{(n(n-1))/2},K)$ given by (3-11) is an analytic action.*

It remains to check the analyticity of the action (3-2) by triangular superior matrices of the form $(1+yE_{i,j})$ for $i<j,\ i,j\in[1,n],\ y\in p\mathbb{Z}_p$. Recall that the action of $(1+yE_{i,j})$ on $f\in I_{\mathrm{loc}}$ given by (3-2) is

$$(1+yE_{i,j})f(A)=f((1+yE_{i,j})^{-1}A)=f((1-yE_{i,j})A).$$

Recall the action of Q_0 given by (3-1), that is, $f(gb) = \chi(b^{-1})f(g)$ with $b \in Q_0$. Hence, our objective is to write the matrix $(1 - yE_{i,j})A$ as the product of two matrices X and Z with $X \in U$ and $Z \in Q_0$, that is,

$$(1 - yE_{i,j})A = XZ,$$

where X is a lower unipotent matrix with entries in $\mathbb{Z}_p$ and Z is an upper triangular matrix with diagonal elements in $1 + p\mathbb{Z}_p$ and such that the elements above the diagonal have entries in $p\mathbb{Z}_p$.

Lemma 3.4. *For $i < j$ and $y \in p\mathbb{Z}_p$, there exists a unique matrix decomposition $(1 - yE_{i,j})A = XZ$ with $X = (x_{k,l})_{k,l} \in U$ and $Z = (z_{r,s})_{r,s} \in Q_0$. Also,*

(i) *all the diagonal elements $z_{r,r}$ of Z are of the form*

$$\frac{1 - yh_{r,r}(y, a)}{1 - yg_{r,r}(y, a)},$$

(ii) *all the elements $z_{r,s}$, for $r < s$, of Z are of the form*

$$\frac{yh_{r,s}(y, a)}{1 - yg_{r,s}(y, a)},$$

(iii) *all the elements $x_{k,l}$ with $k > l$ of the lower triangular unipotent matrix X are of the form*

$$\frac{h_{k,l}(y, a)}{1 - yg_{k,l}(y, a)},$$

where $h_{\star,\star}(y, a)$ and $g_{\star,\star}(y, a)$ are polynomial functions with integral coefficients in y and $a_{2,1}, a_{3,1}, a_{3,2}, \ldots, a_{n,n-1}$ (entries of the lower unipotent matrix A).

Proof. We prove the lemma by an easy inductive argument. The base case $n = 2$ is clear from the matrix equation

$$\begin{pmatrix} 1 & -y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ a_{2,1} & 1 \end{pmatrix} = \begin{pmatrix} 1 - ya_{2,1} & -y \\ a_{2,1} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ x_{2,1} & 1 \end{pmatrix} \begin{pmatrix} z_{1,1} & z_{1,2} \\ 0 & z_{2,2} \end{pmatrix}$$
$$= \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{1,1}x_{2,1} & z_{2,2} + z_{1,2}x_{2,1} \end{pmatrix}$$

with $x_{2,1} = a_{2,1}/(1 - ya_{2,1})$, $z_{1,1} = 1 - ya_{2,1}$, $z_{1,2} = -y$, $z_{2,2} = 1/(1 - ya_{2,1})$. Assume, by induction hypothesis that our lemma is true for GL($n - 1$). We show it for GL(n). Let us first suppose that $i > 1$, that is, with some elementary matrix E', where $1 - yE' \in \mathrm{GL}(n - 1)$, we have

$$(1 - yE_{i,j}) = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & 1 - yE' \\ 0 & \end{array}\right).$$

The matrix A, being lower unipotent, can be written in the following block form:

$$A = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline a_{2,1} & \\ \vdots & A' \\ a_{n,1} & \end{array}\right),$$

with $A' \in \mathrm{GL}(n-1)$. Setting $\underline{\mathrm{a}}$ to be the column vector

$$\begin{pmatrix} a_{2,1} \\ a_{3,1} \\ \vdots \\ a_{n,1} \end{pmatrix},$$

we have

$$(1 - yE_{i,j})A = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & 1 - yE' \\ 0 & \end{array}\right) \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline a_{2,1} & \\ \vdots & A' \\ a_{n,1} & \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline (1 - yE')\underline{\mathrm{a}} & (1 - yE')A' \end{array}\right).$$

We want to decompose the above matrix in the form

$$\left(\begin{array}{c|c} 1 & 0 \\ \hline (1 - yE')\underline{\mathrm{a}} & (1 - yE')A' \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline x_{2,1} & \\ \vdots & X' \\ x_{n,1} & \end{array}\right) \left(\begin{array}{c|c} z_{1,1} & z_{1,2} \cdots z_{1,n} \\ \hline 0 & \\ \vdots & Z' \\ 0 & \end{array}\right)$$

with $x_{2,1}, \ldots, x_{n,1} \in \mathbb{Z}_p$, $z_{1,1} \in 1 + p\mathbb{Z}_p$ and $z_{1,2}, \ldots, z_{1,n} \in p\mathbb{Z}_p$. Denote $\underline{\mathrm{z}}$ to be the row vector $[z_{1,2}, \ldots, z_{1,n}]$, $\underline{\mathrm{x}}$ to be the column vector

$$\begin{pmatrix} x_{2,1} \\ x_{3,1} \\ \vdots \\ x_{n,1} \end{pmatrix}.$$

Hence, we want to solve

$$\left(\begin{array}{c|c} 1 & 0 \\ \hline (1 - yE')\underline{\mathrm{a}} & (1 - yE')A' \end{array}\right) = \left(\begin{array}{c|c} z_{1,1} & \underline{\mathrm{z}} \\ \hline z_{1,1}\underline{\mathrm{x}} & \underline{\mathrm{x}} \cdot \underline{\mathrm{z}} + X'Z' \end{array}\right).$$

So we must have

(1) $z_{1,1} = 1$,

(2) $\underline{\mathrm{z}} = 0$,

(3) $z_{1,1}\underline{\mathrm{x}} = \underline{\mathrm{x}} = (1 - yE')\underline{\mathrm{a}}$ (using $z_{1,1} = 1$ from (1)),

(4) $\underline{\mathrm{x}} \cdot \underline{\mathrm{z}} + X'Z' = X'Z' = (1 - yE')A'$ (as $\underline{\mathrm{z}} = 0$ from (2)).

By the induction hypothesis, we can find X' and Z' satisfying (4) with entries as in Lemma 3.4. Also, (3) is of the form

$$\begin{pmatrix} 1 & & & \\ & & -y & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \begin{pmatrix} a_{2,1} \\ a_{3,1} \\ \vdots \\ a_{n,1} \end{pmatrix} = \begin{pmatrix} x_{2,1} \\ x_{3,1} \\ \vdots \\ x_{n,1} \end{pmatrix}$$

Clearly, we can solve $x_{2,1}, \ldots, x_{n,1}$ from the above matrix equation satisfying Lemma 3.4 and in fact the solutions do not have any denominators.

So by induction we are reduced to the case $i = 1$, that is, when

$$(1 - yE_{i,j}) = \left(\begin{array}{c|c} 1 & 0 \cdots -y \cdots 0 \\ \hline 0 & \\ \vdots & 1 \\ 0 & \end{array}\right).$$

Our goal is to solve, for X and Z, the matrix equation

$$(1 - yE_{1,j})A = XZ. \tag{3-12}$$

Expanding right-hand side of (3-12), we obtain

$$\begin{aligned} \mathcal{B} = (b_{u,v})_{u,v} = XZ &= \Bigg(1 + \sum_{\substack{k\in[1,n] \\ l\in[1,k-1]}} x_{k,l} E_{k,l}\Bigg)\Bigg(\sum_{\substack{r\in[1,n] \\ s\in[r,n]}} z_{r,s} E_{r,s}\Bigg) \\ &= \sum_{\substack{r\in[1,n] \\ s\in[r,n]}} z_{r,s} E_{r,s} + \sum_{\substack{k\in[1,n] \\ r\in[1,k-1] \\ s\in[r,n]}} x_{k,r} z_{r,s} E_{k,s}. \end{aligned}$$

Therefore,

$$b_{u,v} = \begin{cases} \displaystyle\sum_{r=1}^{v} x_{u,r} z_{r,v} & \text{if } u > v, \\ \displaystyle z_{u,v} + \sum_{r=1}^{u-1} x_{u,r} z_{r,v} & \text{if } u \le v. \end{cases} \tag{3-13}$$

Recall that our matrix

$$A = \sum_{k\ge l} a_{k,l} E_{k,l}$$

is lower unipotent, that is, $a_{k,k} = 1$ for all $k \in [1, n]$ and $a_{k,l} = 0$ for $k < l$. Expanding the left-hand side of (3-12), we obtain

$$(1-yE_{1,j})A=(1-yE_{1,j})\Big(\sum_{k\geq l}a_{k,l}E_{k,l}\Big)=\sum_{k\geq l}a_{k,l}E_{k,l}-\sum_{l=1}^{j}ya_{j,l}E_{1,l}$$
$$=\sum_{\substack{k\in[1,n]\\ l\in[1,k]\\ k\neq 1}}a_{k,l}E_{k,l}+\sum_{l=2}^{j}(-ya_{j,l})E_{1,l}+(1-ya_{j,1})E_{1,1}.$$

Note that the first row of the matrix $(1-yE_{1,j})A$ is

$$\sum_{l=2}^{j}(-ya_{j,l})E_{1,l}+(1-ya_{j,1})E_{1,1}.$$

From (3-12), the matrices $(1-yE_{1,j})A$ and $\mathcal{B}=(b_{u,v})_{u,v}$ are equal. Thus, equating $b_{u,v}$ from (3-13) with the above expression of the matrix $(1-yE_{1,j})A$, we obtain the following equations (with the convention that $x_{k,l}=0$ for $k\leq l$ and $z_{r,s}=0$ for $r>s$):

(1) For $u\neq 1$ and $u>v$, $b_{u,v}=\sum_{r=1}^{v}x_{u,r}z_{r,v}=a_{u,v}$.

(2) For $u\neq 1$ and $u=v$, $b_{u,v}=z_{u,u}+\sum_{r=1}^{u-1}x_{u,r}z_{r,v}=a_{u,u}=1$.

(3) For $u\neq 1$ and $u<v$, $b_{u,v}=z_{u,v}+\sum_{r=1}^{u-1}x_{u,r}z_{r,v}=a_{u,v}=0$.

(4) For $u=v=1$, $b_{1,1}=z_{1,1}=1-ya_{j,1}$.

(5) For $u=1$ and $u<v$, $b_{1,v}=z_{1,v}=-ya_{j,v}$.

Note that in (5), for $v>j$, $b_{1,v}=-ya_{j,v}=0$ (as A is lower unipotent). Setting $v=1$ in (1), for $u\in[2,n]$, we obtain

$$x_{u,1}=\frac{a_{u,1}}{z_{1,1}}=\frac{a_{u,1}}{1-ya_{j,1}}\qquad(\text{as } z_{1,1}=1-ya_{j,1}\text{ from (4)}).\tag{3-14}$$

Now, let $C=(c_{k,l})_{k,l}=(1-yE_{1,j})A$ and $\mathcal{B}=(b_{u,v})_{u,v}$ as above. We proceed by equating, in stage 1, the first row of the matrix $\mathcal{B}$ with the first row of the matrix C, starting from the leftmost entry (i.e., given by (4) and (5) above) and solve for $z_{\star,\star}$. Then in the next stage (say, stage $1+\frac{1}{2}$) we equate the first column of the matrix $\mathcal{B}$ with the first column of the matrix C starting from the uppermost entry ($b_{2,1}=c_{2,1}$) and solve for $x_{\star,\star}$ (i.e., those given by (3-14)). In stage 2, we do the same with the second row and in the stage $2+\frac{1}{2}$ we equate the second column of the matrix $\mathcal{B}$ with C (given by (1), (2) and (3)) and proceed like this until the last (n-th) stage. Our objective is to solve $x_{\star,\star}$ and $z_{\star,\star}$ while equating the matrix $\mathcal{B}$ with C and to show (i), (ii) and (iii) of Lemma 3.4. We prove this by induction.

Assume, by induction hypothesis, at stages m and $m+\frac{1}{2}$, $1\leq m<n$, that we have found $x_{k,l}$ for $k\in[2,n]$, $l\in[1,m]$, $k>l$ and $z_{r,s}$ for $r\in[1,m]$, $s\in[1,n]$, $r\leq s$, having the forms (i), (ii) and (iii), respectively. Then, at stage $(m+1)$, we have to equate $b_{m+1,v} = c_{m+1,v}$ for $v \in [m+1, n]$. Equating $b_{m+1,m+1} = c_{m+1,m+1}$, we deduce, by (2), that

$$\begin{aligned} z_{m+1,m+1} &= 1 - \sum_{r=1}^{m} x_{m+1,r} z_{r,m+1} \\ &= 1 - \frac{y h_1(y,a)}{1 - y g_1(y,a)} = \frac{1 - y(h_1 + g_1)}{1 - y g_1}, \end{aligned}$$

where the second equality is by induction hypothesis, for some polynomial functions $h_1(y,a)$ and $g_1(y,a)$ with integral coefficients in y and $a_{2,1}, a_{3,1}, a_{3,2}, \ldots, a_{n,n-1}$.

Similarly, equating $b_{m+1,v} = c_{m+1,v}$ for $v \in [m+2, n]$, we obtain, by (3), that

$$\begin{aligned} z_{m+1,v} &= -\sum_{r=1}^{m} x_{m+1,r} z_{r,v} \\ &= \frac{-y h_2(y,a)}{1 - y g_2(y,a)} \qquad \text{(again by induction hypothesis).} \end{aligned}$$

At stage $(m+1) + \frac{1}{2}$ we have to equate $b_{u,m+1} = c_{u,m+1}$ for all $u \in [m+2, n]$. So, by (1), we get

$$\begin{aligned} x_{u,m+1} z_{m+1,m+1} &= a_{u,m+1} - \sum_{r=1}^{m} x_{u,r} z_{r,m+1} \\ &= a_{u,m+1} - \frac{y h_3(y,a)}{1 - y g_3(y,a)} \qquad \text{(again by induction hypothesis)} \\ &= \frac{h_4(y,a)}{1 - y g_3(y,a)} \end{aligned}$$

for some h_4 and g_3 with integral coefficients. Therefore,

$$x_{u,m+1} = \frac{h_4(y,a)}{1 - y g_3(y,a)} \cdot \frac{1}{z_{m+1,m+1}} = \frac{h_4(y,a)}{1 - y g_3(y,a)} \frac{1 - y g_1}{1 - y(h_1 + g_1)} = \frac{h_5(y,a)}{1 - y g_5(y,a)},$$

with polynomials h_5 and g_5 having integral coefficients. This completes our induction argument and finishes the proof of Lemma 3.4. □

Now, let $f \in I_{\mathrm{loc}}$. Then, by Lemma 3.4, and with X and $z_{\star,\star}$ as contained therein, the action of $(1 + y E_{i,j})$ on f is given by

$$(1 + y E_{i,j}) f(A) = f(X) \chi(z_{1,1}^{-1}, \ldots, z_{n,n}^{-1}). \tag{3-15}$$

Recall that for $|v| < 1$, we have

$$(1 - v)^{-m} = \sum_{q=0}^{\infty} \binom{m+q-1}{q} v^q.$$

Assume now that $f \in \mathcal{A}(\mathbb{Z}^{(n(n-1))/2}, K)$ is a globally analytic function. Thus, f is an element in the Tate algebra of U with $\frac{1}{2}n(n-1)$ variables. In order to show that the action of $(1+yE_{i,j})$ on $f \in \mathcal{A}(U, K)$, given by (3-15), is globally analytic we have to show that

$$\prod_{r=1}^{n} \chi_r(z_{r,r}^{-1}) f(X)$$

is a globally analytic function in y.

Lemma 3.5. *If the action $f \to g,\ g(A) = f(X)$, where $A = XZ$, is globally analytic, then $f \to \prod_{r=1}^{n} \chi_r(z_{r,r}^{-1}) g$ is globally analytic.*

Proof. With (3-10), our character χ is analytic. Hence,

$$\begin{aligned}\chi_r(z_{r,r}^{-1}) &= \chi_r\left(\frac{1-yg_{r,r}(y,a)}{1-yh_{r,r}(y,a)}\right) && \text{(from Lemma 3.4)}\\ &= \sum_{n=0}^{\infty} c_n \left(\frac{1-yg_{r,r}(y,a)}{1-yh_{r,r}(y,a)}\right)^n && (\text{for } |c_n| \to 0).\end{aligned}$$

We are reduced to showing that $(y,a) \to (1-yg_{r,r}(y,a))/(1-yh_{r,r}(y,a))$ is analytic in y and this is true because

$$\frac{1-yg_{r,r}(y,a)}{1-yh_{r,r}(y,a)} = (1-yg_{r,r}(y,a))\left(\sum_{n=0}^{\infty}(yh_{r,r}(y,a))^n\right). \qquad \square$$

Therefore, with Lemma 3.5, to prove that the action of $(1+yE_{i,j})$ on $f \in \mathcal{A}(U, K)$, given by (3-15), is globally analytic, we only need to show that the action $f \to g,\ g(A) = f(X)$, where $A = XZ$, is globally analytic.

Lemma 3.6. *The action $f \to g,\ g(A) = f(X)$, where $A = XZ$, is globally analytic.*

Proof. Recall that the lower unipotent matrix X is $((x_{k,l})_{k,l})$ with

$$x_{k,l} = \frac{h_{k,l}(y,a)}{1-yg_{k,l}(y,a)}$$

given by Lemma 3.4. Write

$$x_{k,l} = h_{k,l}(y,a)\sum_{n=0}^{\infty} y^n g_{k,l}(y,a)^n = \sum_{n=0}^{\infty} y^n g_{n,k,l}(y,a).$$

Since f is analytic, $f = \sum_N f_N x^N$ with $N = (N_{k,l}) \in \mathbb{Z}_p^{n(n-1)/2}$, $x^N = \prod_{k>l} x_{k,l}^{N_{k,l}}$. The norm $|f_N| \to 0$ as $N \to \infty$. Then,

$$f(X) = f((x_{k,l})_{k,l}) = \sum_N f_N \prod_{\substack{k,l\\k>l}} \left(\sum_{n=0}^{\infty} y^n g_{n,k,l}(y,a)\right)^{N_{k,l}}.$$

As

$$\left(\sum_{n=0}^{\infty} y^n g_n\right)^M = \sum_{v\ge 0} y^v \sum_{v_1+\cdots+v_M=v} g_{v_1}\cdots g_{v_M},$$

we obtain that

$$f(X) = \sum_{N=(N_{k,l})} f_N \prod_{k,l} \sum_{v\ge 0} y^v \Bigl(\sum_{v_1+\cdots+v_{N_{k,l}}=v} g_{v_1,k,l}\cdots g_{v_{N_{k,l}},k,l}\Bigr).$$

Define $a_{k,l}(v) = \bigl(\sum_{v_1+\cdots+v_{N_{k,l}}=v} g_{v_1,k,l}\cdots g_{v_{N_{k,l}},k,l}\bigr)$; then,

$$\begin{aligned} f(X) &= \sum_{N=(N_{k,l})} f_N \prod_{k,l}\sum_{v\ge 0} y^v a_{k,l}(v) \\ &= \sum_{N=(N_{k,l})} f_N \sum_{v\ge 0}^{\infty} y^v \sum_{\sum v_{k,l=v}} \prod_{k,l} a_{k,l}(v_{k,l}) \\ &= \sum_{N=(N_{k,l})} f_N \sum_{v\ge 0}^{\infty} y^v \sum_{\sum v_{k,l=v}} \prod_{k,l} \sum_{v_1+\cdots+v_{N_{k,l}}=v_{k,l}} g_{v_1,k,l}\cdots g_{v_{N_{k,l}},k,l}. \end{aligned}$$

The coefficient of y^v is

$$\sum_{N=(N_{k,l})} f_N \sum_{\sum v_{k,l=v}} \prod_{k,l} \sum_{v_1+\cdots+v_{N_{k,l}}=v_{k,l}} g_{v_1,k,l}\cdots g_{v_{N_{k,l}},k,l} := \sum_{N=(N_{k,l})} f_N s_N,$$

where

$$s_N := \sum_{\sum v_{k,l=v}} \prod_{k,l} \sum_{v_1+\cdots+v_{N_{k,l}}=v_{k,l}} g_{v_1,k,l}\cdots g_{v_{N_{k,l}},k,l}.$$

Here $N_{k,l}$ is finite and $v_{k,l}\le v$ and hence the sum s_N is a finite sum in $\mathbb{Z}_p$. As, with $N\to\infty$, $|f_N|\to 0$, we obtain that $\bigl|\sum_N f_N s_N\bigr|\to 0$ and this completes the proof. □

This shows the analyticity of the action given by (3-15). So we have shown:

Lemma 3.7. *For $y\in p\mathbb{Z}_p$ and $i<j$, the action of the upper unipotent (rigid analytic) one-parameter subgroup $(1+yE_{i,j})$ of G on $f\in\mathcal{A}(\mathbb{Z}_p^{n(n-1)/2}, K)$, given by (3-15) is an analytic action.*

Note that, by Section 3A, the vector space of locally analytic functions of principal series

$$\operatorname{ind}_{P_0}^B(\chi)_{\mathrm{loc}} = \{f\in\mathcal{A}_{\mathrm{loc}}(B,K) : f(gb)=\chi(b^{-1})f(g),\ b\in P_0,\ g\in B\}$$

is isomorphic to the vector space of the locally analytic functions

$$I_{\mathrm{loc}} \cong \mathcal{A}_{\mathrm{loc}}(\mathbb{Z}_p^{n(n-1)/2}, K).$$

Denote by $\mathrm{ind}_{P_0}^{B}(\chi)$ the space of globally analytic vectors of $\mathrm{ind}_{P_0}^{B}(\chi)_{\mathrm{loc}}$, which is

$$\mathcal{A} := \mathcal{A}(\mathbb{Z}_p^{n(n-1)/2}, K).$$

Also, the representation on $\mathcal{A}$ is admissible: indeed, $\mathcal{A}$ is a subspace of $\mathcal{A}(G)$ defined by the conditions $f(gb) = \chi(b^{-1})f(g)$ (f is then analytic on G since χ is) and this is a closed subspace. Thus by Lemmas 3.1–3.3 and Lemma 3.7 we have shown the following theorem.

Theorem 3.8. *Assume $p > n+1$. Let χ be an analytic character of T_0 (see (3-10)). The action of G on the induced principal series $\mathrm{ind}_{P_0}^{B}(\chi)$ is a globally analytic action. Moreover, the globally analytic representation of G on $\mathrm{ind}_{P_0}^{B}(\chi)$ is admissible in the sense of [Emerton 2017].*

Recall that $\chi = (\chi_1, \dots, \chi_n)$ where $\chi_i(1 + pu_i) = e^{c_i \log(1+pu_i)}$ for $c_i \in K$, with u_i close to 0, $i \in [1, n]$. Also, recall from (3-5) that $f \in \mathcal{A}$ implies that $f(A) = \sum_{\nu \in \mathbb{N}^d} c_\nu a^\nu$ with $|c_\nu| \to 0$ as $|\nu| = \nu_{2,1} + \nu_{3,1} + \cdots + \nu_{n,n-1} \to \infty$.

In the following, we will have conditions on the character χ such that the globally analytic representation of G on $\mathcal{A}$ is irreducible.

Let μ be the linear form from the Lie algebra of the torus T_0 to K given by

$$\mu = (-c_1, \dots, c_n) : \mathrm{Diag}(t_1, \dots, t_n) \mapsto \sum_{i=1}^{n} -c_i t_i,$$

where $t = (t_i) \in \mathrm{Lie}(T_0)$. For negative root $\alpha = (i, j)$, $i > j$, let $H_{(i,j)}$ be the matrix $E_{i,i} - E_{j,j}$ where $E_{i,i}$ is the standard elementary matrix.

Theorem 3.9. *Let the c_i's satisfy (3-10) and $p > n+1$, then the globally analytic representation $\mathcal{A} \cong \mathrm{ind}_{P_0}^{B}(\chi)$ of G is topologically irreducible if and only if $-\mu(H_{\alpha=(i,j)}) + i - j \notin \{1, 2, 3, \dots\}$ for all $\alpha = (i, j) \in \Phi^-$.*

Assume $\mathcal{X} \subset \mathcal{A}$ is a closed nontrivial G-invariant subspace. Let Φ, Φ^-, Φ^+, Π be the roots, negative roots, positive roots and simple roots, respectively, associated to G. Consider $f \in \mathcal{A}$. Then, from (3-5),

$$f = \sum_{\nu \in \mathbb{N}^d} c_\nu a^\nu$$

where $d = n(n-1)/2$, $c_\nu \in K$, $|c_\nu| \to 0$ as $|\nu| := \sum_{\alpha \in \Phi^-} \nu_\alpha \to \infty$.
Here, $\nu = (\nu_\alpha, \alpha \in \Phi^-) \in \mathbb{N}^d$ and $a_\nu = \prod_{\alpha \in \Phi^-} a_\alpha^{\nu_\alpha}$. In some arguments we will have to order the exponents ν_α. We use the following lexicographic order. Let $\alpha = (i, j)$ and $\alpha' = (k, l)$. Then ν_α comes before $\nu_{\alpha'}$ if and only if $i < k$ or $i = k$ and $j < l$, i.e., $\nu = (\nu_{2,1}, \nu_{3,1}, \nu_{3,2}, \dots, \nu_{n,n-1})$ (see also the discussion before (3-5)). For $N \geq 0$, let τ_N be the natural truncation

$$\mathcal{A} \to K[a]_N := \oplus_{|\nu| \leq N} K a^\nu.$$

The latter space is the space of polynomials in several variables with total degree $\le N$. As τ_N is equivariant under the action of the diagonal subgroup of G given by formulas (3-3) and (3-4) and the associated characters of the diagonal torus of G are linearly independent, $\tau_N(\mathcal{X})$ is a direct sum of monomials given by

$$\mathcal{X}_N := \tau_N(\mathcal{X}) = \left\{ \sum_{\nu \in M_N} c_\nu a^\nu \right\},$$

where M_N is the set of exponents of all the elements in $\mathcal{X}_N$. If $N \le N'$ and $\nu \in M_N$, then as $\mathcal{X}_N$ is the image of the degree N truncation operator τ_N on $\mathcal{X}$, the surjectivity

$$K[a]_{N'} \twoheadrightarrow K[a]_N$$

implies that $M_N \subset M_{N'}$. This is because supposing

$$a^\nu \in \mathcal{X}_N \subset K[a]_N$$

(i.e., $\nu \in M_N$), then there exists a pullback $a^\nu + g \in \mathcal{X}_{N'}$ of a under the surjection $\mathcal{X}_{N'} \twoheadrightarrow \mathcal{X}_N$ (this map is a surjection by definition of $\mathcal{X}_N$ and $\mathcal{X}_{N'}$) such that g is a power series with the total degree of all its monomials strictly higher than N but less than or equal to N'. As $a^\nu + g \in \mathcal{X}_{N'}$ we see that $\nu \in M_{N'}$.

This shows that if $N \le N'$, then $M_N \subset M_{N'}$. Conversely, $\nu \in M_{N'}$ and $|\nu| \le N$ implies $\nu \in M_N$. Therefore, the multisets M_N and $M_{N'}$ are compatible and by letting $N \to \infty$ we see that, as $\mathcal{X}$ is closed, there exists M (the exponents of elements of $\mathcal{X}$) such that

(1) $f \in \mathcal{X} \implies c_\nu = 0$ for all $\nu \notin M$ and

(2) if $\nu \in M$, $a^\nu \in \tau_N(\mathcal{X})$ for all $N \ge |\nu|$; thus there exists

$$f := a^\nu + \sum_{|r|>N} c_r a^r \in \mathcal{X},$$

where $r = (r_\alpha, \alpha \in \Phi^-) \in \mathbb{N}^d$, $|c_r| \to 0$.

For $\alpha \in \Phi^-$, let $Y_\alpha \in \mathfrak{g} = \mathrm{Lie}(G)$ be the infinitesimal generator associated to the unipotent subgroup $1 + yE_\alpha$, $y \in \mathbb{Z}_p$, E_α being the standard elementary matrix at α.

Lemma 3.10. *The multi-index* 0 *is in* M*.*

Proof. $M \ne \varnothing$, because if so, then $\mathcal{X} = 0$, which is not true as, by assumption, $\mathcal{X}$ is nontrivial. Now if $\nu = (\nu_\alpha, \alpha \in \Phi^-) \in M$, then by (2) above,

$$f = a^\nu + \sum_{|r|>N} c_r a^r \in \mathcal{X},$$

where $N \ge |\nu|$ and $r \in \mathbb{N}^d$. By (3-11), the action of $Y_\beta = Y_{(i,j)}$ on f (where

$\beta = (i, j) \in \Phi^-$ is fixed) is given by

$$
\begin{aligned}
Y_\beta(f) = \frac{d}{dy}\bigg|_{y=0} \Bigg(\Bigg(\prod_{\substack{\alpha=(u,v)\\ u=i \Rightarrow v>j}} a_\alpha^{\nu_\alpha} \Bigg) \Bigg(\prod_{k=1}^{j} (a_{i,k} + y a_{j,k})^{\nu_{i,k}} \Bigg) \\
+ \sum_{|r|>N} c_r \Bigg(\prod_{\substack{\alpha=(u,v)\\ u=i \Rightarrow v>j}} a_\alpha^{r_\alpha} \Bigg) \Bigg(\prod_{k=1}^{j} (a_{i,k} + y a_{j,k})^{r_{i,k}} \Bigg) \Bigg) \\
= \prod_{\substack{\alpha=(u,v)\\ u=i \Rightarrow v>j}} a_\alpha^{\nu_\alpha} \Bigg(\sum_{l=1}^{j} \nu_{i,l} a_{j,l} a_{i,l}^{\nu_{i,l}-1} \prod_{\substack{k\in[1,j]\\ k\neq l}} a_{i,k}^{\nu_{i,k}} \Bigg) \\
\times \sum_{|r|>N} c_r \prod_{\substack{\alpha=(u,v)\\ u=i \Rightarrow v>j}} a_\alpha^{r_\alpha} \Bigg(\sum_{l=1}^{j} r_{i,l} a_{j,l} a_{i,l}^{r_{i,l}-1} \prod_{\substack{k\in[1,j]\\ k\neq l}} a_{i,k}^{r_{i,k}} \Bigg) \\
= A + \sum_{|r|>N} c_r B.
\end{aligned}
$$

The first term in the right-hand side of the above equation is

$$
A := \prod_{\substack{\alpha=(u,v)\\ u=i \Rightarrow v>j}} a_\alpha^{\nu_\alpha} \Bigg(\sum_{l=1}^{j} \nu_{i,l} a_{j,l} a_{i,l}^{\nu_{i,l}-1} \prod_{\substack{k\in[1,j]\\ k\neq l}} a_{i,k}^{\nu_{i,k}} \Bigg) = \sum_{l=1}^{j} \nu_{i,l} a_{j,l}^{\nu_{j,l}+1} a_{i,l}^{\nu_{i,l}-1} \prod_{\substack{\alpha\neq(i,l)\\ \alpha\neq(j,l)}} a_\alpha^{\nu_\alpha}
$$

and

$$
B = \prod_{\substack{\alpha=(u,v)\\ u=i \Rightarrow v>j}} a_\alpha^{r_\alpha} \Bigg(\sum_{l=1}^{j} r_{i,l} a_{j,l} a_{i,l}^{r_{i,l}-1} \prod_{\substack{k\in[1,j]\\ k\neq l}} a_{i,k}^{r_{i,k}} \Bigg) = \sum_{l=1}^{j} r_{i,l} a_{j,l}^{r_{j,l}+1} a_{i,l}^{r_{i,l}-1} \prod_{\substack{\alpha\neq(i,l)\\ \alpha\neq(j,l)}} a_\alpha^{r_\alpha}.
$$

Notice that the monomials in B have total degree $|r|$ except, when $l = j$, the term $r_{i,j} a_{i,j}^{r_{i,j}-1} \prod_{\alpha\neq(i,j)} a_\alpha^{r_\alpha}$ (note that $a_{j,j} = 1$ by convention) which has total degree $|r| - 1$.

As $Y_{(i,j)}(f) \in \mathcal{X}$, we see that $(\nu_\alpha, \nu_{i,j} - 1, \alpha \in \Phi^-, \alpha \neq (i, j)) \in M$; these are the exponents when we take $l = j$ in A. This shows that if $M \neq \varnothing$, then $0 \in M$ because we can descend the $\nu_{i,j}$'s successively for every negative root (i, j) and this completes the proof of Lemma 3.10. □

Lemma 3.11. *The constants a^0 are in $\mathcal{X}$.*

Proof. Let $T_k \in \mathfrak{g}$ be the infinitesimal generator associated to the diagonal subgroup $\mathrm{Diag}(1, \dots, t_k, \dots, 1)$, where $t_k \in 1 + p\mathbb{Z}_p$ and t_k is at the (k, k)-th place. By

Lemma 3.10, $0 \in M$. This implies there exists c_r for $|r| > 0$ such that

$$f = c_0 + \sum_{|r|>0} c_r a^r \in \mathcal{X}$$

(where $c_0 \neq 0$). By (3-6), from the action of $\mathrm{Diag}(1, \dots, t_k, \dots, 1)$ on f, the function obtained from $T_k(f)$ gives that

$$\sum_{|r|>0} c_r \Big(\sum r_\delta - \sum r_\beta\Big) a^r \in \mathcal{X}, \tag{3-16}$$

where $\sum r_\delta$ is $\sum_{i\in[k+1,n],\,\delta=(i,k)} r_\delta$ and $\sum r_\beta$ is $\sum_{j\in[1,k-1],\,\beta=(k,j)} r_\beta$.

The function obtained from $T_k^{p-1}(f)$ gives that

$$\sum_{|r|>0} c_r \Big(\sum r_\delta - \sum r_\beta\Big)^{p-1} a^r \in \mathcal{X}.$$

This implies that

$$E_k f := c_0 + \sum_{|r|>0} c_r \bigg(1 - \Big(\sum r_\delta - \sum r_\beta\Big)^{p-1}\bigg) a^r \in \mathcal{X}.$$

If $p \mid \sum r_\delta - \sum r_\beta$, then

$$\bigg(1 - \Big(\sum r_\delta - \sum r_\beta\Big)^{p-1}\bigg)^l \to 1 \quad \text{as } l \to \infty.$$

If $p \nmid \big(\sum r_\delta - \sum r_\beta\big)$, then

$$\bigg(1 - \Big(\sum r_\delta - \sum r_\beta\Big)^{p-1}\bigg)^l \to 0 \quad \text{as } l \to \infty.$$

Then

$$A_{k,1} f := c_0 + \sum_{\substack{|r|>0 \\ p \mid (\sum r_\delta - \sum r_\beta)}} c_r a^r \in \mathcal{X}.$$

Similar to (3-16), applying now the transformation T_k on $A_{k,1} f$, dividing by p, and iterating all the above steps, we see that

$$A_{k,2}(f) := c_0 + \sum_{\substack{|r|>0 \\ p^2 \mid (\sum r_\delta - \sum r_\beta)}} c_r a^r \in \mathcal{X}.$$

Repeating this s times, for $s \in \mathbb{N}$, we obtain

$$A_{k,s}(f) := c_0 + \sum_{\substack{|r|>0 \\ p^s \mid (\sum r_\delta - \sum r_\beta)}} c_r a^r \in \mathcal{X}.$$

This implies, for $s \in \mathbb{N}$,

$$\Big(\prod_{k=1}^{n} A_{k,s}\Big)(f) = c_0 + Q_s(f) \in \mathcal{X}, \tag{3-17}$$

where $Q_s(f) = \sum c_r a^r$ where the sum runs over all $r = (r_\alpha, \alpha \in \Phi^-)$ with $|r| > 0$ such that for all $k \in [1, n]$,

$$p^s \;\Big|\; \Big(\sum_{\substack{\delta=(i,k)\\ i\in[k+1,n]}} r_\delta - \sum_{\substack{\beta=(k,j)\\ j\in[1,k-1]}} r_\beta \Big).$$

We need to show that $Q_s(f) \to 0$ as $s \to \infty$, i.e., we have to show that

$$\forall N_\epsilon, \exists S, \text{ such that } \forall s > S, v_p(c_r) > N_\epsilon, \forall r \in \mathbb{N}^d \text{ such that } |r| > 0, \tag{3-18}$$

whenever

$$p^s \;\Big|\; \Big(\sum_{\substack{\delta=(i,k)\\ i\in[k+1,n]}} r_\delta - \sum_{\substack{\beta=(k,j)\\ j\in[1,k-1]}} r_\beta \Big) \qquad \text{for all } k \in [1, n]. \tag{3-19}$$

But as f is globally analytic, $|c_r| \to 0$ as $|r| = \sum_{\alpha\in\Phi^-} r_\alpha \to \infty$, which means that

$$\forall N_\epsilon, \exists S' \text{ such that } v_p(c_r) > N_\epsilon \text{ whenever } |r| > S'. \tag{3-20}$$

Choose an S such that $p^S > S'$.

For $k = 1$, (3-19) implies $p^s \mid r_{2,1} + r_{3,1} + r_{4,1} + \cdots + r_{n,1}$ which means that $r_{2,1}+r_{3,1}+r_{4,1}+\cdots+r_{n,1} \geq p^s > S'$ except when $r_{2,1} = r_{3,1} = r_{4,1} = \cdots = r_{n,1} = 0$. If this happens, then consider (3-19) with $k = 2$, i.e., $p^s \mid r_{3,2}+r_{4,2}+\cdots+r_{n,2}-r_{2,1}$ (where $r_{2,1} = 0$), i.e., $r_{3,2} + r_{4,2} + \cdots + r_{n,2} \geq p^s > S'$ except when $r_{3,2} = r_{4,2} = \cdots = r_{n,2} = 0$. Repeating this process, since we have started with an r such that $|r| > 0$, we see that any r as in (3-19), with $|r| > 0$, satisfies $|r| > S'$ for all $s > S$ and by (3-20) this implies that $v_p(c_r) > N_\epsilon$, which was the desired condition in (3-18). This shows that $Q_s(f) \to 0$ as $s \to \infty$ which gives that c_0 is in $\mathcal{X}$ (see (3-17)). This completes the proof of Lemma 3.11. □

In the following, we complete the proof of Theorem 3.9 which was to find conditions such that the globally analytic representation $\mathcal{A}$ of G is topologically irreducible. It uses an argument concerning Verma modules and the condition of irreducibility of $\mathcal{A}$ comes from a result of Bernstein, Gelfand and Gelfand determining the condition of irreducibility of that Verma module.

Let $\mathfrak{g} = \mathrm{Lie}(G)$, let $\mathfrak{h} = \mathrm{Lie}(T_0)$, and let $\mathfrak{b}$ (resp. $\mathfrak{b}^-$) be the upper (resp. lower) triangular Borel subalgebra containing $\mathfrak{h}$. Let $\mathfrak{u}^- = \mathrm{Lie}(U)$. Therefore $\mathfrak{g} \cong \mathfrak{gl}_n$, the set of all $n \times n$ matrices with coefficients in $\mathbb{Z}_p$. So $\mathfrak{h}$ and $\mathfrak{b}$ are the subalgebras of $\mathfrak{gl}_n$ consisting of diagonal and upper triangular matrices, respectively.

Recall that here c_i's $\in K$ are such that $\chi_i(t) = t^{c_i}$, for $t \to 1$. Let

$$V_{-\mu} := U(\mathfrak{g}) \otimes_{U(\mathfrak{b}^-)} K$$

be the Verma module where $U(\mathfrak{b}^-)$ acts on K via the action of $\mathfrak{b}^- = \mathfrak{u}^- \oplus \mathfrak{h}$, $\mathfrak{u}^-$ acting trivially, and $\mathfrak{h}$ via $-\mu \in \mathfrak{h}^\star = \mathrm{Hom}(\mathfrak{h}, K)$ given by

$$\mu = (-c_1, \dots, -c_n) : \mathrm{Diag}(t_1, \dots, t_n) \mapsto \sum_{i=1}^{n} -c_i t_i, \tag{3-21}$$

where $t = (t_i) \in \mathfrak{h}$, $U(\mathfrak{g})$ is the universal enveloping algebra of $\mathfrak{g}$. (Note that Dixmier [1977, Section 7.1.14] has a different normalization for the Verma module).

Let $\mathcal{A}_{\mathrm{fin}}$ be the set of polynomials within the rigid analytic functions $\mathcal{A}$. For $k \in [1, n]$, let $T_k \in \mathfrak{h}$ be the infinitesimal generator associated to the one parameter diagonal subgroup $\mathrm{Diag}(1, \dots, t_k, \dots, 1)$, $t_k \in 1 + p\mathbb{Z}_p$, where t_k is at the (k, k)-th place and $f = a^r \in \mathcal{A}_{\mathrm{fin}}$. The elements T_k form a basis of $\mathfrak{h}$. By (3-4) and (3-6), the action of $\mathrm{Diag}(1, \dots, t_k, \dots, 1)$ on f is given by

$$\begin{aligned}\mathrm{Diag}(1, \dots, t_k, \dots, 1)(f) \\ = \Bigg(\Bigg(\prod_{\substack{\alpha=(u,v)\\ u,v\neq k}} a_\alpha^{r_\alpha}\Bigg)\Bigg(\prod_{\substack{\delta=(i,k)\\ i\in[k+1,n]}} a_\delta^{r_\delta} t_k^{r_\delta}\Bigg)\Bigg(\prod_{\substack{\beta=(k,j)\\ j\in[1,k-1]}} a_\beta^{r_\beta} t_k^{-r_\beta}\Bigg)\Bigg)(\chi_k(t_k)).\end{aligned}$$

As $\chi_k(t_k) = t_k^{c_k}$, so the action of T_k on f is

$$\begin{aligned} T_k \cdot f &= c_k a^r + \Bigg(\sum_{\substack{\delta=(i,k)\\ i\in[k+1,n]}} r_\delta - \sum_{\substack{\beta=(k,j)\\ j\in[1,k-1]}} r_\beta\Bigg) a^r \\ &= \Bigg(c_k + \sum_{\substack{\delta=(i,k)\\ i\in[k+1,n]}} r_\delta - \sum_{\substack{\beta=(k,j)\\ j\in[1,k-1]}} r_\beta\Bigg) a^r = \Bigg(-\mu - \sum_{i=1}^{d} \alpha_i r_{\alpha_i}\Bigg)(T_k) a^r. \end{aligned}$$

Here the α_i's are the negative roots.

Thus if $H \in \mathfrak{h}$, then

$$H \cdot a^r = \Bigg(-\mu - \sum_{i=1}^{d} \alpha_i r_{\alpha_i}\Bigg)(H) a^r. \tag{3-22}$$

Decomposing

$$\mathcal{A}_{\mathrm{fin}} = \oplus_{\xi \in \mathfrak{h}^\star} \mathcal{A}_{\mathrm{fin}}(\xi)$$

in the form of $\mathfrak{h}$-eigenspaces, we see from (3-22) that the monomials a^r are $\mathfrak{h}$-finite and the dimensions of eigenspaces of $\mathcal{A}_{\mathrm{fin}}$ under $\mathfrak{h}$ are finite: The eigenvectors are of the form $\xi \in -\mu - \sum_{i=1}^{d} \mathbb{N}\alpha_i \in \mathfrak{h}^\star$ and the multiplicity $\mathrm{mult}(\xi) = \dim \mathcal{A}(\xi)$ of ξ is equal to

$$\dim \mathcal{A}(\xi) = \mathrm{mult}(\xi) = \left\{ \text{number of families } (r_{\alpha_i}) \in \mathbb{N}^d \;\middle|\; \xi = -\mu - \sum_{i=1}^{d} r_{\alpha_i}\alpha_i \right\}, \tag{3-23}$$

which is finite.

With $f_0 = 1 \in \mathcal{A}_{\mathrm{fin}}$, $H \cdot f_0 = -\mu(H) f_0$ and the action $\mathfrak{u}^- \cdot f_0$ is equal to 0 because the action of any element of $\mathfrak{u}^-$ on f_0 is given by derivation (see proof of Lemma 3.10). So, the map $u \to u \cdot f_0$ for $u \in \mathfrak{g}$ induces a $\mathfrak{g}$-homomorphism

$$\phi : V_{-\mu} \to \mathcal{A}, \tag{3-24}$$

where $V_{-\mu} := U(\mathfrak{g}) \otimes_{U(\mathfrak{b}^-)} K$.

Moreover, $v \in V_{-\mu}$ implies v is $\mathfrak{h}$-finite (see [Dixmier 1977, Chapter 7]). This gives $\phi(v) \in \mathcal{A}$ is $\mathfrak{h}$-finite which means that $\phi(v) \in \mathcal{A}_{\mathrm{fin}}$. This is because (3-22) gives, by continuity, that $f \in \mathcal{A}$; hence $f = \sum_{r=(r_{\alpha_i})} c_r a^r$ implies

$$H \cdot f = \sum_{r=(r_{\alpha_i})} \Bigl(-\mu - \sum_{i=1}^{d} r_{\alpha_i}\alpha_i \Bigr)(H) c_r a^r.$$

Then $H \cdot f = \lambda f$ implies $\lambda = \bigl(-\mu - \sum_{i=1}^{d} r_{\alpha_i}\alpha_i\bigr)(H)$ if $c_r \neq 0$. Therefore the cardinality of the set $\{c_r \neq 0\}$ is finite and the $\mathfrak{h}$-finite vectors of $\mathcal{A}$ are just $\mathcal{A}_{\mathrm{fin}}$.

The map $\phi : V_{-\mu} \to \mathcal{A}_{\mathrm{fin}}$ in (3-24) is clearly nonzero because the vector $1 \in V_{-\mu}$ goes to f_0.

Lemma 3.12. *If the Verma module $V_{-\mu}$ is irreducible then the globally analytic G-representation $\mathcal{A}$ is irreducible.*

Proof. Suppose the Verma module $V_{-\mu}$ is irreducible. Then the map $\phi : V_{-\mu} \to \mathcal{A}_{\mathrm{fin}}$ is injective. Also by (3-23), under the action of $\mathfrak{h}$, since the eigenvectors of $V_{-\mu}$ and $\mathcal{A}_{\mathrm{fin}}$ and their multiplicities match, that is $\dim \mathcal{A}(\xi) = \dim \mathcal{A}_{\mathrm{fin}}(\xi) = \dim V_{-\mu}(\xi)$, we deduce that ϕ is an isomorphism.

The dimension of $\dim \mathcal{A}(\xi)$ is given by (3-23). On the other hand, using that our Verma module $V_{-\mu}$ is defined by $\mathfrak{b}^-$ and $-\mu$ (rather than $\lambda - \rho^-$ as in Dixmier's parametrization [1977, 7.1.4]), Dixmier's formula [1977, 7.1.6] yields

$$\dim V_{-\mu}(\xi) = \mathrm{mult}(\xi) = \left\{ \text{number of families } (r_{\alpha_i}) \in \mathbb{N}^d \;\middle|\; \xi = \lambda - \rho^- - \sum_{i=1}^{d} r_{\alpha_i}\alpha_i \right\},$$

where $\rho^- = \frac{1}{2}\sum_{\alpha \in \Phi^-} \alpha$ is half the sum of negative roots (since we have used $\mathfrak{b}^-$ to define the Verma module instead of Dixmier's $\mathfrak{b}^+$). We easily see that the above dimension $\dim V_{-\mu}(\xi)$ is equivalent to $\dim \mathcal{A}(\xi)$ (3-23) with $\lambda - \rho^- = -\mu$.

So $V_{-\mu} \cong \mathcal{A}_{\text{fin}}$. Suppose $\mathcal{X}$ is a nonzero closed subspace of $\mathcal{A}$. Then by Lemma 3.11, we have $1 \in \mathcal{X}$. Thus $\mathcal{A}_{\text{fin}} = U(\mathfrak{g}) \cdot 1 \subset \mathcal{X}$. Since $\mathcal{X}$ is closed, $\mathcal{X} = \mathcal{A}$. □

Now we prove the converse of Lemma 3.12.

Recall that a closed subspace of $\mathcal{A}$ is G-invariant if and only if it is invariant by $\mathfrak{g}$ [Clozel 2018, Proposition 2.4]. Moreover it follows from the definition of globally analytic representations (compare [Clozel 2018, Section 2.2]) that the action of $\mathfrak{g}$ on $\mathcal{A}$ is continuous. If $V \subset \mathcal{A}_{\text{fin}}$ is invariant by $\mathfrak{g}$, it follows that its closure $\overline{V}$ is G-invariant.

Recall that $\mathcal{A}_{\text{fin}}$ is the set of $\mathfrak{h}$-finite vectors in $\mathcal{A}$. In particular, if $\mathcal{X} \subset \mathcal{A}$ is closed, the space $\mathcal{X}_{\mathfrak{h}\text{-fin}}$ of $\mathfrak{h}$-finite vectors in $\mathcal{X}$ is $\mathcal{X} \cap \mathcal{A}_{\text{fin}}$.

Lemma 3.13. *Assume $V \subset \mathcal{A}_{\text{fin}}$ is invariant by $\mathfrak{g}$. Then $V = \overline{V} \cap \mathcal{A}_{\text{fin}} = \overline{V}_{\mathfrak{h}\text{-fin}}$.*

Proof. By (3-23), $\mathcal{A}(\xi)$ is the subspace of the Tate algebra spanned by a finite number of monomials a^r. In particular, the obvious projection $p_\xi : \mathcal{A} \to \mathcal{A}(\xi)$ is continuous. Assume $v \in \overline{V} \cap \mathcal{A}_{\text{fin}}$. Thus $v \in \bigoplus_\xi \mathcal{A}(\xi)$ (finite sum of finite-dimensional subspaces) and $v = \lim v_m,\ v_m \in V$. If P is the projection on $\bigoplus_\xi \mathcal{A}(\xi),\ v = Pv = \lim Pv_m$. But $Pv' \in V \cap \bigoplus_\xi \mathcal{A}(\xi)$ for any $v' \in V$. Thus $v \in V$, as a limit in a finite-dimensional space. □

Lemma 3.13 obviously gives the following Corollary.

Corollary 3.14. *Suppose V is a nonzero proper subspace of $\mathcal{A}_{\text{fin}}$ stable by $\mathfrak{g}$. Then $\overline{V}$ is a nonzero proper closed G-invariant subspace of $\mathcal{A}$.*

Lemma 3.15. *If the globally analytic G- representation $\mathcal{A}$ is irreducible then the Verma module $V_{-\mu}$ is irreducible.*

Proof. Let $W \subset \mathcal{A}_{\text{fin}}$ be the image of $V_{-\mu}$ by ϕ. Then $W \neq 0$. If $\mathcal{A}$ is an irreducible G-module, $W = \mathcal{A}_{\text{fin}}$ by Corollary 3.14. Thus we have a surjective map $\phi : V_{-\mu} \to \mathcal{A}_{\text{fin}}$. But, as we noticed, the dimensions of $V_{-\mu}(\xi)$ and of $\mathcal{A}_{\text{fin}}(\xi)$ coincide. This implies that ϕ is an isomorphism. On the other hand (again by the Corollary 3.14), W is irreducible. Thus $V_{-\mu}$ is irreducible. □

Now we determine the condition when the Verma module $V_{-\mu}$ is irreducible. Recall that

$$\mu = (-c_1, \ldots, -c_n) : \mathrm{Diag}(t_1, \ldots, t_n) \mapsto \sum_{i=1}^{n} -c_i t_i,$$

where $t = (t_i) \in \mathfrak{h}$. For negative root $\alpha = (i, j),\ i > j$, let $H_{\alpha=(i,j)}$ be the matrix $E_{i,i} - E_{j,j}$ where $E_{i,i}$ is the standard elementary matrix.

Lemma 3.16. *The Verma module $V_{-\mu}$ is irreducible if and only if*

$$(-\mu)(H_{\alpha=(i,j)}) + i - j \notin \{1, 2, 3, \ldots\}$$

for all $\alpha = (i, j) \in \Phi^-$.

Proof. Let $\rho^- = \frac{1}{2}\sum_{\alpha\in\Phi^-}\alpha$. For $\alpha = (i, j) \in \Phi^-$, $H_\alpha = H_{i+1,i} + \cdots + H_{j,j-1}$ and $\rho^-(H_{k+1,k}) = 1$. This gives that $\rho^-(H_{\alpha=(i,j)}) = i - j$. By Theorem 7.6.24 of [Dixmier 1977], the condition of irreducibility of our $V_{-\mu}$ is $(-\mu + \rho^-)(H_\alpha) \notin \{1, 2, 3, \dots\}$ for all negative roots $\alpha \in \Phi^-$. (This is because Dixmier's $\mathfrak{b}^+$ is our $\mathfrak{b}^-$ and so we have to work with negative roots.) This gives the condition

$$(-\mu)(H_\alpha) + i - j \notin \{1, 2, 3, \dots\}. \qquad \square$$

Lemmas 3.16, 3.15, and 3.12 together prove Theorem 3.9.

3B. With L an unramified finite extension of $\mathbb{Q}_p$, suppose V is a globally analytic representation of $G(\mathbb{Q}_p)$ on a K-Banach space where $L \subset K$. Then Clozel showed the following proposition for holomorphic base change.

Proposition 3.17 (Clozel). *V extends naturally to a globally analytic representation of $G(L)$.*

Proof. See [Clozel 2018, Proposition 3.1]. $\square$

All the arguments of Section 3A extend automatically to the group $G(L)$. As L is unramified, the conditions for the character χ to be analytic, that is, those given by (3-10), remain unchanged. Moreover, note that the representation $\mathcal{A}(B_1^{n(n-1)/2}, K)$ (where now $B_1^{n(n-1)/2}$ is seen as a product of $\frac{1}{2}n(n-1)$ closed rigid balls of radius 1 as an L-analytic space) given by Lemmas 3.2, 3.3 and 3.7 is L-analytic. The restriction of $\mathcal{A}(B_1^{n(n-1)/2}, K)$ to $G(\mathbb{Q}_p)$ is simply the previous representation. Indeed, the representation of $G(L)$ is obtained from the representation of $G(\mathbb{Q}_p)$ by holomorphic base change (see Proposition 3.17). Denote by $I_{\mathbb{Q}_p}(\chi)$ and $I_L(\chi)$, respectively, the two *globally analytic* representations (the character χ is defined by the parameters $(c_1, \dots, c_n)$, we agree to identify the characters for the two fields). Then we have:

Theorem 3.18. *For a given embedding $L \hookrightarrow K$, with μ as in (3-21), if*

$$-\mu(H_\alpha) + i - j \notin \{1, 2, 3, \dots\} \quad \textit{for all } \alpha = (i, j) \in \Phi^-,$$

then $I_L(\chi)$ is an admissible, irreducible (under both $G(L)$ and $G(\mathbb{Q}_p)$) globally analytic representation and it is the holomorphic base change of $I_{\mathbb{Q}_p}(\chi)$.

$I_L(\chi)$ is admissible, as holomorphic base change respects admissibility [Clozel 2018, Proposition 3.1]. With the notation of Section 2B, define the full (Langlands) base change of $I_{\mathbb{Q}_p}$ to be the representation of $\mathrm{Res}_{L/\mathbb{Q}_p} G(\mathbb{Q}_p)$ on $\widehat{\otimes}_\sigma (I_L(\chi))^\sigma := I(\chi \circ N_{L/\mathbb{Q}_p})$, where $N_{L/\mathbb{Q}_p}$ is the norm map from L to $\mathbb{Q}_p$ and $\widehat{\otimes}$ is the completed tensor product (see also [Clozel 2018, Definition 3.8]) and $\sigma \in \mathrm{Gal}(L/\mathbb{Q}_p)$. Note that, for each factor, the embedding $i : L \to K$ must be replaced by $i \circ \sigma$. Finally, we then have:

Theorem 3.19. *Let μ be as in (3-21). Assume $-\mu(H_\alpha)+i-j \notin \{1,2,3,\dots\}$ for all $\alpha=(i,j)\in\Phi^-$. Then the completed tensor product $\widehat{\otimes}_\sigma(I_L(\chi))^\sigma$ is irreducible, and is the representation of $G(L)$ on the space of globally analytic vectors, induced from $\chi\circ N_{L/\mathbb{Q}_p}$.*

Proof. Notice that by assumption, each factor in the completed tensor product is irreducible and admits the same description as in Theorem 3.18. The space of the representation $I(\chi\circ N_{L/\mathbb{Q}_p})$ is $\widehat{\otimes}_\sigma\mathcal{A}(U,K)=\mathcal{A}(\mathrm{Res}_{L/\mathbb{Q}_p}U,K)$, which is a space of globally analytic vectors (by Theorem 3.18) in the locally analytic representation $I_{\mathrm{loc}}(\chi\circ N_{L/\mathbb{Q}_p})$ of $\mathrm{Res}_{L/\mathbb{Q}_p}(G)$. The proof of irreducibility of $\widehat{\otimes}_\sigma(I_L(\chi))^\sigma$ follows from Theorem 3.9 using a natural generalization of the argument in [Clozel 2018]. □

4. Analyticity for the induction from the Weyl orbits of the upper triangular Borel subgroup of B

In this section we treat the global analyticity of the principal series induced from Weyl orbits of the Borel subgroup (Theorem 4.3). Then we base change our globally analytic representation to L.

Denote by $\mathbb{P}$ the Borel subgroup of the upper triangular matrices in $\mathrm{GL}_n(\mathbb{Q}_p)$, by $\mathbb{T}$ the maximal torus of $\mathrm{GL}_n(\mathbb{Q}_p)$, by P^+ the Borel subgroup of the upper triangular matrices in $\mathrm{GL}_n(\mathbb{Z}_p)$, and by W the ordinary Weyl group of $\mathrm{GL}_n(\mathbb{Q}_p)$ with respect to $\mathbb{T}$ which is isomorphic to the group of $n\times n$ permutation matrices. Write $P_w^+=B\cap wP^+w^{-1}$. Here B is the Iwahori subgroup in Section 3A. Denote by $\mathrm{ind}_{\mathbb{P}}^{\mathrm{GL}_n(\mathbb{Q}_p)}(\chi)_{\mathrm{loc}}$ the locally analytic induction, that is,

$$\begin{aligned}\mathrm{ind}_{\mathbb{P}}^{\mathrm{GL}_n(\mathbb{Q}_p)}(\chi)_{\mathrm{loc}}=\{f\in\mathcal{A}_{\mathrm{loc}}(\mathrm{GL}_n(\mathbb{Q}_p),K):f(gb)\\ =\chi(b^{-1})f(g),\ g\in\mathrm{GL}_n(\mathbb{Q}_p),\ b\in\mathbb{P}\}.\end{aligned}$$

The Iwasawa decomposition [Orlik and Strauch 2010, Section 3.2.2] gives

$$\mathrm{ind}_{\mathbb{P}}^{\mathrm{GL}_n(\mathbb{Q}_p)}(\chi)_{\mathrm{loc}}\cong\mathrm{ind}_{P^+}^{\mathrm{GL}_n(\mathbb{Z}_p)}(\chi)_{\mathrm{loc}}$$

as a $\mathrm{GL}_n(\mathbb{Z}_p)$-equivariant topological isomorphism. By the Bruhat–Tits decomposition [Orlik and Strauch 2010, Section 3.2.2; Cartier 1979, Section 3.5],

$$\mathrm{GL}_n(\mathbb{Z}_p)=\bigsqcup_{w\in W}BwP^+,$$

we obtain the decomposition

$$\mathrm{ind}_{P^+}^{\mathrm{GL}_n(\mathbb{Z}_p)}(\chi)_{\mathrm{loc}}\cong\bigoplus_{w\in W}\mathrm{ind}_{P_w^+}^{B}(\chi^w)_{\mathrm{loc}},$$

a B-equivariant decomposition of topological vector spaces, where the action of χ^w is given by $\chi^w(h)=\chi(w^{-1}hw)$. Let $\mathrm{ind}_{P_w^+}^{B}(\chi^w)$ be the space of globally analytic functions of $\mathrm{ind}_{P_w^+}^B(\chi^w)_{\mathrm{loc}}$. Our goal is to show that for all $w\in W$, $\mathrm{ind}_{P_w^+}^B(\chi^w)$ is a globally analytic representation of G. We have already showed, in Section 3, that for $w=\mathrm{Id}$ and χ analytic, the induction $\mathrm{ind}_{P_0}^B(\chi)$ is a globally analytic representation of G. (Note that $B\cap P^+=P_0$.) Recall that U is the lower triangular unipotent subgroup of $\mathrm{GL}_n(\mathbb{Z}_p)$. Consider the decomposition (see Lemma 3.3.2 of [Orlik and Strauch 2010])

$$\begin{aligned} B &= (wUw^{-1}\cap B)(wP^+w^{-1}\cap B)\\ &= (wUw^{-1}\cap B)(P_w^+). \end{aligned}$$

For GL_3, and

$$w=\begin{pmatrix}0&0&1\\1&0&0\\0&1&0\end{pmatrix},$$

the above decomposition is

$$B=\begin{pmatrix}\mathbb{Z}_p^\times & p\mathbb{Z}_p & p\mathbb{Z}_p\\ \mathbb{Z}_p & \mathbb{Z}_p^\times & p\mathbb{Z}_p\\ \mathbb{Z}_p & \mathbb{Z}_p & \mathbb{Z}_p^\times\end{pmatrix}=\begin{pmatrix}1 & p\mathbb{Z}_p & p\mathbb{Z}_p\\ 0 & 1 & 0\\ 0 & \mathbb{Z}_p & 1\end{pmatrix}\begin{pmatrix}\mathbb{Z}_p^\times & 0 & 0\\ \mathbb{Z}_p & \mathbb{Z}_p^\times & p\mathbb{Z}_p\\ \mathbb{Z}_p & 0 & \mathbb{Z}_p^\times\end{pmatrix}.$$

We extend a character χ of $\mathbb{T}\cap\mathrm{GL}_n(\mathbb{Z}_p)$ to a character of P_w^+ by acting trivially on the nondiagonal elements of P_w^+. By definition,

$$\mathrm{ind}_{P_w^+}^B(\chi)_{\mathrm{loc}}=\{f\in\mathcal{A}_{\mathrm{loc}}(B,K):f(gb)=\chi(b^{-1})f(g),\ b\in P_w^+,\ g\in B\}.$$

With the decomposition $B=(wUw^{-1}\cap B)(P_w^+)$, the vector space of locally analytic functions $\mathrm{ind}_{P_w^+}^B(\chi)_{\mathrm{loc}}$ is the same as $\mathcal{A}_{\mathrm{loc}}(wUw^{-1}\cap B,K)$. Let $\mathcal{A}(wUw^{-1}\cap B,K)$ be the subspace of globally analytic functions of $\mathcal{A}_{\mathrm{loc}}(wUw^{-1}\cap B,K)$. With $i\neq j$ fixed, $y\in\mathbb{Z}_p$ if $i>j$ and $y\in p\mathbb{Z}_p$ if $i<j$, recall that the action of the one-parameter subgroup on

$$f\in\mathcal{A}(wUw^{-1}\cap B,K)$$

is given by

$$\begin{aligned} (1+yE_{i,j})f(C) &= f((1+yE_{i,j})^{-1}C) \qquad (\text{with } C\in wUw^{-1}\cap B)\\ &= f((1-yE_{i,j})C)\\ &= f((1-yE_{i,j})wAw^{-1}) \quad (\text{with } C=wAw^{-1} \text{ for } A\in U). \end{aligned} \tag{4-1}$$

Our goal is to show that this action is globally analytic.

Since $w^{-1}\in W$, write w^{-1} in the form of a permutation matrix, i.e.,

$$w^{-1}=\sum_{r=1}^n E_{r,j_r}$$

with $j_r \neq j_s$ for $r \neq s$. Then,

$$w^{-1}(1 - yE_{i,j}) = \left(\sum_{r=1}^{n} E_{r,j_r}\right)(1 - yE_{i,j}) = \left(\sum_{r=1}^{n} E_{r,j_r}\right) - yE_{k,j},$$

where k is such that $j_k = i$. As the inverse of a permutation matrix is its transpose, we obtain

$$w^{-1}(1 - yE_{i,j})w = \left(\left(\sum_{r=1}^{n} E_{r,j_r}\right) - yE_{k,j}\right)\left(\sum_{s=1}^{n} E_{j_s,s}\right) = 1 - yE_{k,l},$$

where l is such that $j_l = j$. So we have deduced that

$$(1 - yE_{i,j})w = w(1 - yE_{k,l}) \qquad (k, l \text{ such that } j_k = i,\ j_l = j). \tag{4-2}$$

Inserting (4-2) into (4-1), we obtain

$$(1 + yE_{i,j})f(C) = f(w(1 - yE_{k,l})Aw^{-1}). \tag{4-3}$$

Now, the globally analytic function f on $w(1 - yE_{k,l})Aw^{-1}$ equals some globally analytic function g on $(1 - yE_{k,l})A$, because the conjugacy action of w on the matrix $(1 - yE_{k,l})A$ is just permuting the entries of $(1 - yE_{k,l})A$. So, (4-3) is

$$\begin{aligned} f(w(1 - yE_{k,l})Aw^{-1}) &= g((1 - yE_{k,l})A) \\ &= (1 + yE_{k,l})g(A) \qquad (\text{recall } A \in U) \end{aligned}$$

and we know from Lemmas 3.3 and 3.7 that the action of $(1 + yE_{k,l})$ on $g(A)$ is globally analytic. Thus, we have shown that:

Lemma 4.1. *The action of the lower and the upper unipotent one-parameter subgroups of G of the form $(1+yE_{i,j})$ on $f \in \mathcal{A}(wUw^{-1} \cap B, K)$ is a globally analytic action.*

A similar argument also shows that the action of the diagonal subgroup of G on $\mathcal{A}(wUw^{-1} \cap B, K)$ is globally analytic. More precisely, we write

$$w^{-1}\,\mathrm{Diag}(t_1, \ldots, t_n)w = \mathrm{Diag}(t'_1, \ldots, t'_n)$$

with $(t'_1, \ldots, t'_n)$ a permutation of $(t_1, \ldots, t_n)$. Then, with $C \in wUw^{-1} \cap B$,

$$\begin{aligned} \mathrm{Diag}(t_1^{-1}, \ldots, t_n^{-1})f(C) &= f(\mathrm{Diag}(t_1, \ldots, t_n)wAw^{-1}) \qquad (C = wAw^{-1}) \\ &= f(w[\mathrm{Diag}(t'_1, \ldots, t'_n)]Aw^{-1}) \\ &= g(\mathrm{Diag}(t'_1, \ldots, t'_n)A) \qquad (\text{for some analytic } g) \\ &= \mathrm{Diag}(t_1^{-1}, \ldots, t_n^{-1})g(A) \end{aligned}$$

and by Lemmas 3.2 and 3.1, the action of the diagonal subgroup of G on $g(A)$ is a globally analytic action. Therefore, we have shown:

Lemma 4.2. *The action of the diagonal subgroup of G on $\mathcal{A}(wUw^{-1} \cap B, K)$ is globally analytic.*

Recall that the vector space $\mathcal{A}(wUw^{-1} \cap B, K)$ is isomorphic to $\mathrm{ind}_{P_w^+}^{B}(\chi^w)$. Thus, Lemmas 4.1 and 4.2 together give:

Theorem 4.3. *Assume $p > n+1$. Then, for all $w \in W$, the action of the pro-p Iwahori group G on $\mathrm{ind}_{P_w^+}^{B}(\chi^w)$ is globally analytic.*

Following the notation of Section 3B, we fix L a finite unramified extension of $\mathbb{Q}_p$ inside K. For each $w \in W$, consider the globally analytic admissible representation $I_{w,\mathbb{Q}_p}(\chi) := \mathcal{A}(wUw^{-1} \cap B, K)$ of $G(\mathbb{Q}_p)$. By Section 2B, $\mathcal{A}(wUw^{-1} \cap B, K)$ extends naturally to a globally analytic admissible representation of $G(L)$ called the "holomorphic base change" which we denote by $I_{w,L}(\chi)$. With the notation of Section 2B, define the full Langlands base change to be the representation of $\mathrm{Res}_{L/\mathbb{Q}_p} G(\mathbb{Q}_p)$ on $\oplus_{w\in W}(\widehat{\otimes}_\sigma I_{w,L}(\chi)^\sigma)$ (see [Clozel 2018, Section 3.5]). Finally, as in Theorem 3.19, we will then have:

Theorem 4.4. *The Langlands base change*

$$\oplus_{w\in W}(\widehat{\otimes}_\sigma I_{w,L}(\chi^w)^\sigma)$$

is a globally analytic admissible representation of $G(L)$.

In conclusion, for $p > n+1$, we have shown that for all $w \in W$, $\mathrm{ind}_{P_w^+}^{B}(\chi^w)$ is a globally analytic representation of the pro-p Iwahori G under the analyticity assumption on the character χ. Furthermore we have treated the case of irreducibility of the principal series when $w = \mathrm{Id}$. We hope that it is possible to adapt and generalize the argument of our irreducibility proof to treat the case when $w \neq \mathrm{Id}$. Also it is an interesting future project to determine the globally analytic vectors of more general p-adic representations of $\mathrm{GL}(2, \mathbb{Q}_p)$, for example the "trianguline" representation of [2008] (see also [Colmez 2014]), which corresponds to a quotient of principal series. Also one can explore the connection with the globally analytic vectors of p-adic representations (under the pro-p Iwahori or a suitable rigid analytic subgroup of GL(2)) and (φ, Γ)-modules [Colmez 2010], similar to the existing correspondence for locally analytic representations [Colmez and Dospinescu 2014, Section VI.3].

There are other interesting questions that our work leads to. The most interesting of them is to show Schur's lemma for globally analytic representations. Schur's lemma for locally analytic representations is known by the works of Gabriel Dospinescu and Benjamin Schraen. For our case of topologically irreducible globally analytic principal series, Schur's lemma easily follows from our proof of irreducibility and Proposition 7.1.8(iv) of [Dixmier 1977]. The interesting question is to show Schur's lemma for general globally analytic representations.

Acknowledgements

The author is grateful to Laurent Clozel and Peter Schneider for several discussions relating to this article. The author thanks the editor Don Blasius and the anonymous referee for the careful review, which eliminated several errors and typos.

References

[Bertapelle 2000] A. Bertapelle, "Formal Néron models and Weil restriction", *Math. Ann.* **316**:3 (2000), 437–463. MR Zbl

[Bosch 2014] S. Bosch, *Lectures on formal and rigid geometry*, Lecture Notes in Math. **2105**, Springer, 2014. MR Zbl

[Cartier 1979] P. Cartier, "Representations of p-adic groups: a survey", pp. 111–155 in *Automorphic forms, representations and L-functions, I* (Corvallis, OR, 1977), edited by A. Borel and W. Casselman, Proc. Sympos. Pure Math. **33**, Amer. Math. Soc., Providence, RI, 1979. MR Zbl

[Clozel 2017] L. Clozel, "Globally analytic p-adic representations of the pro-p-Iwahori subgroup of GL(2) and base change, I: Iwasawa algebras and a base change map", *Bull. Iranian Math. Soc.* **43**:4 (2017), 55–76. MR Zbl

[Clozel 2018] L. Clozel, "Globally analytic p-adic representations of the pro-p Iwahori subgroup of GL(2) and base change, II: A Steinberg tensor product theorem", pp. 1–33 in *Cohomology of arithmetic groups* (Bonn, Germany, 2016), edited by J. W. Cogdell et al., Springer Proc. Math. Stat. **245**, Springer, 2018. MR Zbl

[Colmez 2008] P. Colmez, "Représentations triangulines de dimension 2", pp. 213–258 in *Représentations p-adiques de groupes p-adiques, I: Représentations galoisiennes et* (φ, Γ)*-modules*, edited by L. Berger et al., Astérisque **319**, Soc. Math. France, Paris, 2008. MR Zbl

[Colmez 2010] P. Colmez, "Représentations de $\mathrm{GL}_2(\mathbb{Q}_p)$ et (φ, Γ)-modules", pp. 281–509 in *Représentations p-adiques de groupes p-adiques, II: Représentations de* $\mathrm{GL}_2(\mathbb{Q}_p)$ *et* (φ, Γ)*-modules*, edited by L. Berger et al., Astérisque **330**, Soc. Math. France, Paris, 2010. MR Zbl

[Colmez 2014] P. Colmez, "La série principale unitaire de $\mathrm{GL}_2(\mathbb{Q}_p)$: vecteurs localement analytiques", pp. 286–358 in *Automorphic forms and Galois representations, I* (Durham, UK, 2011), edited by F. Diamond et al., Lond. Math. Soc. Lect. Note Ser. **414**, Cambridge Univ. Press, 2014. MR Zbl

[Colmez and Dospinescu 2014] P. Colmez and G. Dospinescu, "Complétés universels de représentations de $\mathrm{GL}_2(\mathbb{Q}_p)$", *Algebra Number Theory* **8**:6 (2014), 1447–1519. MR Zbl

[Dixmier 1977] J. Dixmier, *Enveloping algebras*, North-Holland Math. Library **14**, North-Holland, Amsterdam, 1977. MR Zbl

[Emerton 2017] M. Emerton, *Locally analytic vectors in representations of locally p-adic analytic groups*, Mem. Amer. Math. Soc. **1175**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Lazard 1965] M. Lazard, "Groupes analytiques p-adiques", *Inst. Hautes Études Sci. Publ. Math.* **26** (1965), 389–603. MR Zbl

[Orlik and Strauch 2010] S. Orlik and M. Strauch, "On the irreducibility of locally analytic principal series representations", *Represent. Theory* **14** (2010), 713–746. MR Zbl

[Ray 2020] J. Ray, "Explicit presentation of an Iwasawa algebra: the case of pro-p Iwahori subgroup of $\mathrm{SL}_n(\mathbb{Z}_p)$", *Forum Math.* **32**:2 (2020), 319–338. MR Zbl

[Robert 1984] A. Robert, "Représentations p-adiques irréductibles de sous-groupes ouverts de $\mathrm{SL}_2(\mathbf{Z}_p)$", *C. R. Acad. Sci. Paris Sér. I Math.* **298**:11 (1984), 237–240. MR

[Robert 1985] A. Robert, “Représentations p-adiques de dimension infinie de sous-groupes ouverts de $SL_2(\mathbf{Q}_p)$”, exposé 23, pp. 1–14 in *Groupe d’étude d’analyse ultramétrique. 12e année: 1984/85. Fasc. 2*, edited by Y. Amice et al., Secrétariat Math., Paris, 1985. MR

[Schneider 2011] P. Schneider, *p-adic Lie groups*, Grundlehren der Math. Wissenschaften **344**, Springer, 2011. MR Zbl

[Schneider and Teitelbaum 2002] P. Schneider and J. Teitelbaum, “Locally analytic distributions and p-adic representation theory, with applications to GL_2”, *J. Amer. Math. Soc.* **15**:2 (2002), 443–468. MR Zbl

[Steinberg 1963] R. Steinberg, “Representations of algebraic groups”, *Nagoya Math. J.* **22** (1963), 33–56. MR Zbl

Received September 22, 2018. Revised February 27, 2020.

JISHNU RAY
DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER BC
CANADA
jishnuray1992@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 307, No. 2, 2020
https://doi.org/10.2140/pjm.2020.307.491

ZEROS OF p-ADIC HYPERGEOMETRIC FUNCTIONS, p-ADIC ANALOGUES OF KUMMER'S AND PFAFF'S IDENTITIES

NEELAM SAIKIA

We classify all the zeros and nonzero values of a family of hypergeometric series in the p-adic setting. These values of hypergeometric series in the p-adic setting lead to transformations of hypergeometric series in the p-adic setting which can be described as p-adic analogues of Kummer's and Pfaff's linear transformations on classical hypergeometric series. We also evaluate certain summation identities for hypergeometric series in the p-adic setting as well as Gaussian hypergeometric series.

1. Introduction and statement of results

The main goal of this paper is to study zeros of hypergeometric series in the p-adic setting introduced by D. McCarthy [2012a; 2013]. We also establish analogues of classical hypergeometric series transformations, particularly very special cases of Kummer's and Pfaff's linear transformations, for hypergeometric series in the p-adic setting. These types of questions were posed by McCarthy [2013]. We now begin with the definition of classical hypergeometric series. For a complex number a and a nonnegative integer k the rising factorial denoted by $(a)_k$ is defined by $(a)_k := a(a+1)(a+2)\cdots(a+k-1)$ for $k > 0$ and $(a)_0 := 1$. Then for $a_i, b_i, \lambda \in \mathbb{C}$ with $b_i \notin \{\dots, -3, -2, -1, 0\}$, the classical hypergeometric series ${}_{r+1}F_r$ is defined by

$$ {}_{r+1}F_r\left(\begin{matrix} a_1, & a_2, & \dots, & a_{r+1} \\ & b_1, & \dots, & b_r \end{matrix} \,\middle|\, \lambda\right) := \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_{r+1})_k}{(b_1)_k \cdots (b_r)_k} \cdot \frac{\lambda^k}{k!}. $$

This series converges for $|\lambda| < 1$. Classical hypergeometric series play important role in different areas of mathematics. For example, they have significant applications in modular forms, elliptic curves, representation theory, differential equations etc. [McCarthy 2012b; 2010; Mortenson 2005]. J. Greene [1987] introduced the notion

The author acknowledges the financial support of the Department of Science and Technology, Government of India for financial support under the INSPIRE Faculty Award.
MSC2010: 11S80, 11T24, 33C20, 33C99, 33E50.
Keywords: character sum, Gauss sums, Jacobi sums, p-adic gamma function.

of hypergeometric series over finite fields which are finite field analogues of classical hypergeometric series. Let p be an odd prime and $\mathbb{F}_p$ denote a finite field with p elements. Let $\widehat{\mathbb{F}_p^\times}$ denote the group of all multiplicative characters of $\mathbb{F}_p^\times$ and $\overline{\chi}$ denote the inverse of a multiplicative character χ. We extend the domain of each $\chi \in \widehat{\mathbb{F}_p^\times}$ to $\mathbb{F}_p$ by simply setting $\chi(0) := 0$ including the trivial character ε. For multiplicative characters χ and ψ of $\mathbb{F}_p$ the Jacobi sum is defined by

$$J(\chi, \psi) := \sum_{y \in \mathbb{F}_p} \chi(y)\psi(1-y), \tag{1-1}$$

and the normalized Jacobi sum known as binomial is defined by

$$\binom{\chi}{\psi} := \frac{\psi(-1)}{p} J(\chi, \overline{\psi}). \tag{1-2}$$

Let n be a nonnegative integer. For multiplicative characters $A_1, A_2, \ldots, A_{n+1}$, and $B_1, B_2, \ldots, B_n$ of $\mathbb{F}_p$ with $t \in \mathbb{F}_p$, Greene [1987] defined the ${}_{n+1}F_n(\cdots)$ hypergeometric function over finite field $\mathbb{F}_p$ by

$${}_{n+1}F_n\left(\begin{matrix} A_1, & A_2, & \ldots, & A_{n+1} \\ & B_1, & \ldots, & B_n \end{matrix} \;\middle|\; t\right) := \frac{p}{p-1} \sum_{\chi \in \widehat{\mathbb{F}_p^\times}} \binom{A_1\chi}{\chi} \cdots \binom{A_{n+1}\chi}{B_n\chi} \chi(t).$$

This function is also known as Gaussian hypergeometric function. These functions were developed to have a parallel study of character sums analogous to special functions. Gaussian hypergeometric functions satisfy many identities which are often analogues of classical hypergeometric series identities. For more details, see [Greene 1987]. Since the entries of the Gaussian hypergeometric function are multiplicative characters so results involving Gaussian hypergeometric functions often be restricted to primes in certain congruence classes for the existence of characters of specific orders, see for example [Evans and Greene 2009; Fuselier 2010; Lennon 2011a; 2011b]. To overcome these limitations, McCarthy [2012a; 2013] defined a function ${}_nG_n[\cdots]$ in terms of quotients of p-adic gamma functions which can be best described as an analogue of classical hypergeometric series in the p-adic setting. Let $\mathbb{Z}_p$ and $\mathbb{Q}_p$ denote the ring of p-adic integers and the field of p-adic numbers, respectively. Let $\Gamma_p(\cdot)$ denote the Morita's p-adic gamma function. Let ω denote the Teichmüller character of $\mathbb{F}_p$, satisfying $\omega(a) \equiv a \pmod{p}$, and $\bar{\omega}$ denote the character inverse of ω. For $x \in \mathbb{Q}$ let $\lfloor x \rfloor$ denote the greatest integer less than or equal to x and $\langle x \rangle$ denote the fractional part of x, satisfying $0 \le \langle x \rangle < 1$. We now recall the McCarthy hypergeometric function ${}_nG_n[\cdots]$ in the p-adic setting.

Definition 1.1 [McCarthy 2013, Definition 5.1]. Let p be an odd prime and $t \in \mathbb{F}_p$. For positive integer n and $1 \le k \le n$, let $a_k, b_k \in \mathbb{Q} \cap \mathbb{Z}_p$. Then the function ${}_nG_n[\cdots]$

is defined as

$$
{}_nG_n\left[\begin{matrix} a_1, & a_2, & \dots, & a_n \\ b_1, & b_2, & \dots, & b_n \end{matrix} \;\middle|\; t\right] := \frac{-1}{p-1}\sum_{a=0}^{p-2}(-1)^{an}\,\overline{\omega}^a(t)
\times \prod_{k=1}^{n}(-p)^{-\lfloor\langle a_k\rangle - a/(p-1)\rfloor - \lfloor\langle -b_k\rangle + a/(p-1)\rfloor}\,\frac{\Gamma_p\big(\big\langle a_k-\frac{a}{p-1}\big\rangle\big)}{\Gamma_p(\langle a_k\rangle)}\,\frac{\Gamma_p\big(\big\langle -b_k+\frac{a}{p-1}\big\rangle\big)}{\Gamma_p(\langle -b_k\rangle)}.
$$

This function is also known as p-adic hypergeometric function. It is clear from the Definition 1.1 that the value of the ${}_nG_n[\cdots]$ function depends only on the fractional part of the parameters a_k and b_k. Therefore, we may assume that $0 \le a_k, b_k < 1$. Gaussian hypergeometric functions satisfy many powerful transformation formulas that are often mirror symmetrical to their classical counterparts, for details see [Greene 1987]. Note that these results can be converted into identities involving ${}_nG_n[\cdots]$ via the transformations [McCarthy 2013, Lemma 3.3; 2012c, Proposition 2.5] between finite field hypergeometric function and p-adic hypergeometric series. However, the new identities involving ${}_nG_n[\cdots]$ will be valid for the primes p, where the original characters existed over $\mathbb{F}_p$. Therefore, it will be interesting to extend such results to almost all primes. Fuselier and McCarthy [2016] established certain transformation identities for p-adic hypergeometric series in full generality. In particular, they proved a transformation result analogous to a Whipple's result for ${}_3F_2$-classical hypergeometric series. These transformations eventually leaded to settle one supercongruence conjecture of Rodriguez-Villegas between a truncated ${}_4F_3$-classical hypergeometric series and the Fourier coefficients of a certain weight four modular form. This is one of the motivation to study transformation formulas with the expectation that transformation formulas will lead to new identities. Let χ_4 be a multiplicative character of $\mathbb{F}_p$ of order 4. Also, let φ be the quadratic character of $\mathbb{F}_p$. Consider the classical hypergeometric series

$$
{}_2F_1\left(\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ & \frac{1}{2} \end{matrix} \;\middle|\; t\right).
$$

Then the finite field analogue of this series can be considered as ${}_2F_1\left(\begin{smallmatrix} \chi_4, & \chi_4^3 \\ & \varphi \end{smallmatrix} \mid t\right)$. Using the transformations [McCarthy 2013, Lemma 3.3; 2012c, Proposition 2.5] the function

$$
{}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \;\middle|\; \frac{1}{t}\right]
$$

can be described as the p-adic analogue of the classical hypergeometric series

$$
{}_2F_1\left(\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ & \frac{1}{2} \end{matrix} \;\middle|\; t\right).
$$

We know that classical hypergeometric series satisfy many powerful identities. For example, Gauss [1812], Kummer [1836], Whipple [Lidl and Niederreiter 1983, p. 54], Saalchütz [Slater 1966, p. 49], Dixon [Slater 1966, p. 51], and Watson [Slater 1966, p. 54] studied special values of classical hypergeometric series. For instance, the following evaluation of classical hypergeometric series in terms of quotients of classical gamma function was due to Gauss [1812]. If $R(c-a-b) > 0$ then

$$(1\text{-}3) \qquad {}_2F_1\left(\begin{matrix} a, & b \\ & c \end{matrix} \,\Big|\, 1\right) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.$$

If we put $a = \frac{1}{4}$, $b = \frac{3}{4}$ and $c = 1 + \frac{1}{2}$ in (1-3) then we have

$$(1\text{-}4) \qquad {}_2F_1\left(\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ & 1+\frac{1}{2} \end{matrix} \,\Big|\, 1\right) = \frac{\Gamma\left(1+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(1+\frac{1}{4}\right)\Gamma\left(\frac{3}{4}\right)}.$$

Also, consider Kummer's theorem [1836]:

$$(1\text{-}5) \qquad {}_2F_1\left(\begin{matrix} a, & b \\ & 1+b-a \end{matrix} \,\Big|\, -1\right) = \frac{\Gamma(1+b-a)\Gamma\left(1+\frac{b}{2}\right)}{\Gamma(1+b)\Gamma\left(1+\frac{b}{2}-a\right)}.$$

Putting $a = \frac{1}{4}$ and $b = \frac{3}{4}$ into (1-5) we have

$$(1\text{-}6) \qquad {}_2F_1\left(\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ & 1+\frac{1}{2} \end{matrix} \,\Big|\, -1\right) = \frac{\Gamma\left(1+\frac{1}{2}\right)\Gamma\left(1+\frac{3}{8}\right)}{\Gamma\left(1+\frac{3}{4}\right)\Gamma\left(1+\frac{1}{8}\right)}.$$

Classical hypergeometric series with dihedral monodromy group can be expressed as elementary functions as their hypergeometric equations can be reformulated to Fuchsian equations with cyclic monodromy groups. For example, two interesting cases that can be expressed as square roots inside powers are:

$$(1\text{-}7) \qquad {}_2F_1\left(\begin{matrix} \frac{a}{2}, & \frac{a+1}{2} \\ & a+1 \end{matrix} \,\Big|\, z\right) = \left(\frac{1+\sqrt{1-z}}{2}\right)^{-a},$$

and

$$(1\text{-}8) \qquad {}_2F_1\left(\begin{matrix} \frac{a}{2}, & \frac{a+1}{2} \\ & \frac{1}{2} \end{matrix} \,\Big|\, z\right) = \frac{(1-\sqrt{z})^{-a} + (1+\sqrt{z})^{-a}}{2}.$$

All these evaluations of Gauss, Kummer etc. motivate us to study the special values of p-adic hypergeometric series ${}_2G_2[\cdots]$. Indeed, we completely determine all the possible zeros and nonzero values of a certain family of ${}_2G_2[\cdots]$. We first discuss a theorem that classify all the zeros and nonzero values of the function ${}_2G_2[\cdots]$. For brevity we write $a \neq \square$ if a is not square in $\mathbb{F}_p$.

Theorem 1.2. *Let $p \geq 3$ be a prime and $t \in \mathbb{F}_p^{\times}$. Then we have the following values.*

(1)

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, 1\right] = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8, \\ -1 & \text{if } p \equiv \pm 3 \pmod 8. \end{cases} \tag{1-9} $$

(2) *Let $t \neq 1$ and $\frac{t-1}{t}$ be a square in $\mathbb{F}_p^\times$ such that $\frac{t-1}{t} = a^2$ for some $a \in \mathbb{F}_p^\times$. Then we have*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, t\right] = \varphi(2) \cdot (\varphi(1+a) + \varphi(1-a)). \tag{1-10} $$

(3) *If $\frac{t-1}{t} \neq \square$ in $\mathbb{F}_p$ then*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, t\right] = 0. \tag{1-11} $$

From Theorem 1.2 we obtain the following corollary which explicitly states whenever the function

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, t\right] $$

is zero and nonzero.

Corollary 1.3. *Let $p \geq 3$ be a prime and $t \in \mathbb{F}_p^\times$. Then*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, t\right] = 0 $$

if $\frac{t-1}{t} = a^2$ for some $a \in \mathbb{F}_p^\times$ with $\varphi(1+a) = -\varphi(1-a)$ or if $\frac{t-1}{t} \neq \square$ in $\mathbb{F}_p^\times$. On the other hand,

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, t\right] \neq 0 $$

if $t = 1$ or if $\frac{t-1}{t} = a^2$ for some $a \in \mathbb{F}_p^\times$ with $\varphi(1+a) \neq -\varphi(1-a)$.

As a consequence of the Theorem 1.2 we state the following corollary.

Corollary 1.4. *Let $p \geq 3$ be a prime and $t \in \mathbb{F}_p^\times$. Then the only possible values the function*

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \,\middle|\, t\right] $$

can take are $0, \pm 1$ *and* ± 2.

Remark 1.5. Note that (1-9) can be described as a p-adic analogue of (1-4). Theorem 1.2 provides a p-adic analogue of (1-6). The value of the function

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,-1\right] $$

completely depends on the prime as if $p \equiv \pm 3 \pmod 8$ then it will be equal to zero. However, (1-6) cannot be equal to zero. Theorem 1.2 can also be described as p-adic analogue of (1-7) and (1-8) for $a = \frac{1}{2}$.

Another purpose of this paper is to establish p-adic analogous of the Kummer's linear transformation [Bailey 1935, p. 4 Equation (1)]:

$$ {}_2F_1\left(\begin{matrix}a, & b\\ & c\end{matrix}\;\middle|\, z\right) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}\cdot {}_2F_1\left(\begin{matrix}a, & b\\ & 1+a+b-c\end{matrix}\;\middle|\, 1-z\right) + \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-z)^{c-a-b}\cdot {}_2F_1\left(\begin{matrix}c-a, & c-b\\ & 1+c-a-b\end{matrix}\;\middle|\, 1-z\right). \tag{1-12} $$

The next theorem provides transformations of p-adic hypergeometric series which can be described as p-adic analogue of a particular case of Kummer's linear transformation (1-12). This theorem is obtained as a consequence of Theorem 1.2.

Theorem 1.6. *Let $p \geq 3$ be a prime and $x \in \mathbb{F}_p$ be such that $x \neq 0, 1$. Then we have the following*:

(1) *If x and $1-x$ are not squares in $\mathbb{F}_p$ then*

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{x}\right] = {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{1-x}\right]. \tag{1-13} $$

(2) *If $x = b^2$ for some $b \in \mathbb{F}_p$ and $1-x$ is not a square in $\mathbb{F}_p$ then*

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{x}\right] = {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{1-x}\right] - \varphi(2)(\varphi(1+b)+\varphi(1-b)). \tag{1-14} $$

(3) *If x, and $1-x$ are both squares such that $1-x = a^2$ and $x = b^2$ for some $a, b \in \mathbb{F}_p$ then*

$$ (\varphi(1+b)+\varphi(1-b)){}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{x}\right] = (\varphi(1+a)+\varphi(1-a)){}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{1-x}\right]. \tag{1-15} $$

(4) *If $1-x = a^2$ for some $a \in \mathbb{F}_p$ and x is not a square in $\mathbb{F}_p$ then*

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{x}\right] = {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\,\frac{1}{1-x}\right] + \varphi(2)(\varphi(1+a)+\varphi(1-a)). \tag{1-16} $$

Note that the finite field analogue of Theorem 1.6 involving characters of order 4 follows from Greene's evaluation [1987, Theorem 4.4(i)] and if we use this result of Greene along with the relations [McCarthy 2013, Lemma 3.3; 2012c, Proposition 2.5] then we also obtain a similar transformation for the p-adic hypergeometric series

$$ {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \;\middle|\; t\right] $$

under the condition that $p \equiv 1 \pmod 4$. However, Theorem 1.6 has no congruence condition on primes.

Fuselier and McCarthy [2016] evaluated certain summation identities for p-adic hypergeometric series. This motivates us to study summation identities of p-adic hypergeometric series.

Theorem 1.7. *Let $p \geq 3$ be a prime. Let $x \in \mathbb{F}_p^{\times}$. Then we have the following.*

(1)

$$ \sum_{t\in\mathbb{F}_p^{\times}} \varphi(t(t-1))\, {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \;\middle|\; t\right] = -1 - p\cdot\varphi(2). \tag{1-17} $$

(2) *If $x \neq 1$ and $1-x$ is not a square in $\mathbb{F}_p$ then we have*

$$ \sum_{t\in\mathbb{F}_p^{\times}} \varphi(t(t-1))\, {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \;\middle|\; \frac{t}{x}\right] = -1. \tag{1-18} $$

(3) *If $x \neq 1$ and $1-x = a^2$ for some $a \in \mathbb{F}_p$ then*

$$ \sum_{t\in\mathbb{F}_p^{\times}} \varphi(t(t-1))\, {}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & 0 \end{matrix} \;\middle|\; \frac{t}{x}\right] = -1 - p\varphi(2)(\varphi(1+a)+\varphi(1-a)). \tag{1-19} $$

The following theorem gives a summation identity of Gaussian hypergeometric functions involving characters of order 4.

Theorem 1.8. *Let $p \geq 3$ be a prime such that $p \equiv 1 \pmod 4$. Let $x \in \mathbb{F}_p^{\times}$ and χ_4 be a multiplicative character of $\mathbb{F}_p$ of order 4. Then we have the following.*

(1)

$$ \sum_{t\in\mathbb{F}_p^{\times}} \varphi(1-t)\, {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \;\middle|\; t\right) = \frac{1}{p} + \varphi(2). \tag{1-20} $$

(2) *If $x \neq 1$ and $1-x$ is not a square in $\mathbb{F}_p$ then we have*

$$ \sum_{t\in\mathbb{F}_p^{\times}} \varphi(x-t)\, {}_2F_1\left(\begin{matrix} \chi_4, & \chi_4^3 \\ & \varepsilon \end{matrix} \;\middle|\; t\right) = \frac{\varphi(x)}{p}. \tag{1-21} $$

(3) *If* $x \neq 1$ *and* $1-x=a^2$ *for some* $a \in \mathbb{F}_p^\times$ *then*

$$\sum_{t\in\mathbb{F}_p^\times} \varphi(x-t)\,{}_2F_1\left(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\middle|\; t\right) = \frac{\varphi(x)}{p} + \varphi(2x)(\varphi(1+a)+\varphi(1-a)). \tag{1-22}$$

Apart from Kummer's transformation there are other interesting transformation formulas exist in the literature. For example, Euler [Slater 1966, p. 10], Whipple [1925], and Dixon [1903] studied transformation properties of classical hypergeometric series. In this paper, we are interested in the Pfaff's transformation [Slater 1966, p. 31]

$${}_2F_1\left(\begin{matrix}a, & b\\ & c\end{matrix}\;\middle|\; z\right) = (1-z)^{-a}\,{}_2F_1\left(\begin{matrix}a, & c-b\\ & c\end{matrix}\;\middle|\; \tfrac{z}{z-1}\right). \tag{1-23}$$

In particular, if $a=\frac{1}{4}$, $b=\frac{3}{4}$, and $c=\frac{1}{2}$ then the above result gives

$${}_2F_1\left(\begin{matrix}\frac14, & \frac34\\ & \frac12\end{matrix}\;\middle|\; z\right) = (1-z)^{-1/4}\,{}_2F_1\left(\begin{matrix}\frac14, & \frac{-1}{4}\\ & \frac12\end{matrix}\;\middle|\; \tfrac{z}{z-1}\right). \tag{1-24}$$

We know that p-adic analogue of

$${}_2F_1\left(\begin{matrix}\frac14, & \frac{-1}{4}\\ & \frac12\end{matrix}\;\middle|\; z\right)$$

can be described as the function

$${}_2G_2\left[\begin{matrix}\frac14, & \frac{-1}{4}\\ 0, & \frac12\end{matrix}\;\middle|\; \tfrac{1}{z}\right] = {}_2G_2\left[\begin{matrix}\frac14, & 1-\frac{1}{4}\\ 0, & \frac12\end{matrix}\;\middle|\; \tfrac{1}{z}\right].$$

Then a p-adic analogue of Pfaff's transformation (1-24) is described in the next theorem.

Theorem 1.9. *Let* $p\geq 3$ *be a prime and* $1\neq x\in\mathbb{F}_p^\times$. *Then we have the following.*

(1) *If* $1-x\neq\square$ *then*

$${}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & \frac12\end{matrix}\;\middle|\; \tfrac{1}{x}\right] = {}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & \frac12\end{matrix}\;\middle|\; \tfrac{x-1}{x}\right]. \tag{1-25}$$

(2) *If* $1-x=a^2$ *for some* $a\in\mathbb{F}_p^\times$ *then*

$$\begin{aligned}\varphi(a)(\varphi(a+1)+\varphi(a-1))\,{}_2G_2&\left[\begin{matrix}\frac14, & \frac34\\ 0, & \frac12\end{matrix}\;\middle|\; \tfrac{1}{x}\right]\\ &= (\varphi(1+a)+\varphi(1-a))\,{}_2G_2\left[\begin{matrix}\frac14, & \frac34\\ 0, & \frac12\end{matrix}\;\middle|\; \tfrac{x-1}{x}\right].\end{aligned} \tag{1-26}$$

The rest of this paper is organized as follows. We introduce some basic definitions in Section 2 including Gauss sum and p-adic gamma function. In Section 2 we state some results including the Hasse–Davenport result, and the Gross–Koblitz formula. We give the proofs of the main theorems in Section 3.

2. Notation and preliminary results

Let $\overline{\mathbb{Q}_p}$ be the algebraic closure of $\mathbb{Q}_p$ and $\mathbb{C}_p$ be the completion of $\overline{\mathbb{Q}_p}$. Since each $\chi \in \widehat{\mathbb{F}_p^\times}$ takes values from μ_{p-1}, the group of $(p-1)$-th roots of unity in $\mathbb{C}^\times$, and $\mathbb{Z}_p^\times$ contains μ_{p-1}, so we may assume that the multiplicative characters of $\mathbb{F}_p^\times$ to be mapped $\chi : \mathbb{F}_p^\times \to \mathbb{Z}_p^\times$. Recall that $\omega : \mathbb{F}_p^\times \to \mathbb{Z}_p^\times$ is the Teichmüller character. Also, $\widehat{\mathbb{F}_p^\times} = \{\omega^j : 0 \le j \le p-2\}$ and $\bar{\omega}$ denotes the inverse of ω.

Preliminary results on multiplicative characters and Gauss sums. The following result gives the orthogonality relation of multiplicative characters.

Lemma 2.1 [Ireland and Rosen 2005, Chapter 8]. *Let p be an odd prime. Then we have*

$$\sum_{\chi \in \widehat{\mathbb{F}_p^\times}} \chi(x) = \begin{cases} p-1 & \text{if } x = 1, \\ 0 & \text{if } x \neq 1. \end{cases} \tag{2-1}$$

Let ζ_p be a fixed primitive p-th root of unity in $\overline{\mathbb{Q}_p}$. For $\chi \in \widehat{\mathbb{F}_p^\times}$, the Gauss sum is defined by

$$g(\chi) := \sum_{x \in \mathbb{F}_p} \chi(x)\, \zeta_p^x.$$

From the definition we can say that $g(\varepsilon) = -1$. For more details on Gauss sums, see [Berndt et al. 1998]. We now introduce some properties of Gauss sums. Let $\delta : \widehat{\mathbb{F}_p^\times} \to \{0, 1\}$ be defined by

$$\delta(\chi) = \begin{cases} 1 & \text{if } \chi = \varepsilon, \\ 0 & \text{if } \chi \neq \varepsilon. \end{cases} \tag{2-2}$$

We start with a result that provides a formula for the multiplicative inverse of Gauss sum.

Lemma 2.2 [Greene 1987, Equation 1.12]. *Let $\chi \in \widehat{\mathbb{F}_p^\times}$. Then*

$$g(\chi)g(\overline{\chi}) = p\chi(-1) - (p-1)\delta(\chi). \tag{2-3}$$

Another important product formula for Gauss sums is the Hasse–Davenport formula.

Theorem 2.3 [Berndt et al. 1998, Hasse–Davenport relation, Theorem 11.3.5]. *Let χ be a character of order m of $\mathbb{F}_p$ for some positive integer m. For a multiplicative character ψ of $\mathbb{F}_p$ we have*

$$\prod_{i=0}^{m-1} g(\psi\chi^i) = g(\psi^m)\psi^{-m}(m)\prod_{i=1}^{m-1} g(\chi^i). \tag{2-4}$$

The following lemma relates Gauss and Jacobi sums.

Lemma 2.4 [Greene 1987, Equation 1.14]. *Let $\chi_1, \chi_2 \in \widehat{\mathbb{F}_p^\times}$. Then*

$$J(\chi_1,\chi_2) = \frac{g(\chi_1)g(\chi_2)}{g(\chi_1\chi_2)} + (p-1)\chi_2(-1)\delta(\chi_1\chi_2). \tag{2-5}$$

Let χ, ψ be multiplicative characters of $\mathbb{F}_p$. Then the following special values of binomials are very useful to prove our main results, for more details we refer [Greene 1987, Equations 2.12, 2.7]:

$$\binom{\chi}{\varepsilon} = \binom{\chi}{\chi} = -\frac{1}{p} + \frac{p-1}{p}\delta(\chi), \tag{2-6}$$

$$\binom{\chi}{\psi} = \binom{\psi\overline{\chi}}{\psi}\psi(-1). \tag{2-7}$$

p-adic preliminaries. We recall the p-adic gamma function, for further details see [Koblitz 1980]. For a positive integer n, the p-adic gamma function $\Gamma_p(n)$ is defined as

$$\Gamma_p(n) := (-1)^n \prod_{0<j<n,\ p\nmid j} j$$

and one can extend it to all $x \in \mathbb{Z}_p$ by setting $\Gamma_p(0) := 1$ and

$$\Gamma_p(x) := \lim_{x_n \to x} \Gamma_p(x_n)$$

for $x \neq 0$, where x_n runs through any sequence of positive integers p-adically approaching x. Two important product formulas of the p-adic gamma function from [Gross and Koblitz 1979] are as follows. If $x \in \mathbb{Z}_p$ then

$$\Gamma_p(x)\Gamma_p(1-x) = (-1)^{a_0(x)}, \tag{2-8}$$

where $a_0(x) \equiv x \pmod p$ such that $a_0(x) \in \{1, 2, \ldots, p\}$. If $m \in \mathbb{Z}^+$, $p\nmid m$ and $x = \frac{r}{p-1}$ with $0 \le r \le p-1$ then

$$\prod_{h=0}^{m-1} \Gamma_p\Big(\frac{x+h}{m}\Big) = \omega(m^{(1-x)(1-p)})\Gamma_p(x)\prod_{h=1}^{m-1} \Gamma_p\Big(\frac{h}{m}\Big). \tag{2-9}$$

Another interesting product formula of p-adic gamma function given in [McCarthy 2013] as a consequence of (2-9) described as follows. Let $t \in \mathbb{Z}^+$ and $p \nmid t$. Then for $0 \le j \le p-2$ we have

$$\omega(t^{-tj})\Gamma_p\left(\left\langle\frac{-tj}{p-1}\right\rangle\right)\prod_{h=1}^{t-1}\Gamma_p\left(\frac{h}{t}\right)=\prod_{h=1}^{t}\Gamma_p\left(\left\langle\frac{h}{t}-\frac{j}{p-1}\right\rangle\right). \tag{2-10}$$

Let $\pi \in \mathbb{C}_p$ be the fixed root of the polynomial $x^{p-1}+p$, which satisfies the congruence condition $\pi \equiv \zeta_p - 1 \pmod{(\zeta_p-1)^2}$. The Gross–Koblitz formula relates the Gauss sum and p-adic gamma function as follows.

Theorem 2.5 Gross–Koblitz formula [1979]. *For $j \in \mathbb{Z}$,*

$$g(\bar{\omega}^j) = -\pi^{(p-1)\langle j/(p-1)\rangle}\Gamma_p\left(\left\langle\frac{j}{p-1}\right\rangle\right).$$

The next two lemmas are helpful in the proof of our main results. These two lemmas are direct applications of the Gross–Koblitz formula.

Lemma 2.6. *For $1 \le j \le p-2$*

$$\Gamma_p\left(\left\langle 1-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{j}{p-1}\right\rangle\right) = -\omega^j(-1). \tag{2-11}$$

Proof. Applying the Gross–Koblitz formula (Theorem 2.5) on the left-hand side of (2-11) and then using (2-3) it is straightforward to verify the lemma. □

Lemma 2.7. *For $1 \le j \le p-2$ we have*

$$\frac{(-p)^{-\lfloor 1/2+j/(p-1)\rfloor}}{\Gamma_p\left(\frac{1}{2}\right)}\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle 1-\frac{j}{p-1}\right\rangle\right) = \frac{1}{p}\sum_{t\in\mathbb{F}_p^\times}\bar{\omega}^j(-t)\varphi(t(t-1)). \tag{2-12}$$

Proof. Let

$$U = \frac{(-p)^{-\lfloor 1/2+j/(p-1)\rfloor}}{\Gamma_p\left(\frac{1}{2}\right)}\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle 1-\frac{j}{p-1}\right\rangle\right).$$

Using the Gross–Koblitz formula (Theorem 2.5), (2-3), and (2-5) we obtain

$$U = \frac{\varphi\omega^j(-1)}{p}J(\varphi\bar{\omega}^j,\varphi) = \frac{1}{p}\sum_{t\in\mathbb{F}_p^\times}\bar{\omega}^j(-t)\varphi(t(t-1)).$$

This completes the proof of the lemma. □

3. Proof of the theorems

We begin with a proposition which explicitly determines the value of a character sum. We use this proposition to prove Theorems 1.2 and 1.7.

Proposition 3.1. *For $x \in \mathbb{F}_p^\times$ we have*

$$\sum_{\chi \in \widehat{\mathbb{F}_p^\times}} g(\varphi\chi^2)g(\varphi\overline{\chi})g(\overline{\chi})\chi\left(\frac{x}{4}\right) = \begin{cases} 0 & \text{if } 1-x \neq \square, \\ p(p-1)\varphi(-2) & \text{if } x = 1, \\ p(p-1)\varphi(-2)(\varphi(1+a)+\varphi(1-a)) & \text{if } x \neq 1 \text{ and } 1-x = a^2. \end{cases}$$

Proof. Let
$$A = \sum_{\chi \in \widehat{\mathbb{F}_p^\times}} g(\varphi\chi^2)g(\varphi\overline{\chi})g(\overline{\chi})\chi\left(\frac{x}{4}\right).$$

Multiplying both numerator and denominator by $g(\varphi\chi)$ we can write

$$A = \sum_{\chi \in \widehat{\mathbb{F}_q^\times}} \frac{g(\varphi\chi^2)g(\overline{\chi})}{g(\varphi\chi)} g(\varphi\chi)g(\varphi\overline{\chi})\chi\left(\frac{x}{4}\right). \tag{3-1}$$

Applying (2-5) we have

$$\frac{g(\varphi\chi^2)g(\overline{\chi})}{g(\varphi\chi)} = J(\varphi\chi^2, \overline{\chi}) - (p-1)\chi(-1)\delta(\varphi\chi). \tag{3-2}$$

Also, applying (2-3) we have

$$g(\varphi\chi)g(\varphi\overline{\chi}) = p\varphi\chi(-1) - (p-1)\delta(\varphi\chi). \tag{3-3}$$

Substituting (3-2) and (3-3) into (3-1) we obtain

$$A = A_1 + A_2 + A_3 + A_4, \tag{3-4}$$

where

$$A_1 = p\varphi(-1) \sum_{\chi \in \widehat{\mathbb{F}_p^\times}} J(\varphi\chi^2, \overline{\chi})\chi\left(\frac{-x}{4}\right), \tag{3-5}$$

$$\begin{aligned} A_2 &= -(p-1) \sum_{\chi \in \widehat{\mathbb{F}_p^\times}} \delta(\varphi\chi)J(\varphi\chi^2, \overline{\chi})\chi\left(\frac{x}{4}\right) \\ &= -(p-1)\varphi(x)J(\varphi, \varphi), \end{aligned} \tag{3-6}$$

$$\begin{aligned} A_3 &= -p(p-1)\varphi(-1) \sum_{\chi \in \widehat{\mathbb{F}_p^\times}} \delta(\varphi\chi)\chi\left(\frac{x}{4}\right) \\ &= -p(p-1)\varphi(-x), \end{aligned} \tag{3-7}$$

$$A_4=(p-1)^2\sum_{\chi\in\widehat{\mathbb{F}_p^\times}}\delta(\varphi\chi)\chi\left(\frac{-x}{4}\right)=(p-1)^2\varphi(-x). \tag{3-8}$$

Using (1-2), and (2-6) in (3-6) we have

$$A_2=-p(p-1)\varphi(-x)\binom{\varphi}{\varphi}=(p-1)\varphi(-x). \tag{3-9}$$

Adding (3-7), (3-8), and (3-9) we obtain

$$A_2+A_3+A_4=0. \tag{3-10}$$

Substituting (3-10) into (3-4) we have

$$A=A_1=p\varphi(-1)\sum_{\chi\in\widehat{\mathbb{F}_p^\times}}J(\varphi\chi^2,\overline{\chi})\chi\left(\frac{-x}{4}\right).$$

(1-2) gives

$$A=p^2\varphi(-1)\sum_{\chi\in\widehat{\mathbb{F}_p^\times}}\binom{\varphi\chi^2}{\chi}\chi\left(\frac{x}{4}\right).$$

If we use (2-7) then we have $\binom{\varphi\chi^2}{\chi}=\chi(-1)\binom{\varphi\overline{\chi}}{\chi}$. This yields

$$A=p^2\varphi(-1)\sum_{\chi\in\widehat{\mathbb{F}_p^\times}}\binom{\varphi\overline{\chi}}{\chi}\chi\left(\frac{-x}{4}\right)=p\varphi(-1)\sum_{\chi\in\widehat{\mathbb{F}_p^\times},\,y\in\mathbb{F}_p}\varphi(y)\overline{\chi}(y)\overline{\chi}(1-y)\chi\left(\frac{x}{4}\right). \tag{3-11}$$

Replacing χ by $\overline{\chi}$ in (3-11) we obtain

$$A=p\varphi(-1)\sum_{y\in\mathbb{F}_p}\varphi(y)\sum_{\chi\in\widehat{\mathbb{F}_p^\times}}\chi\left(\frac{4y(1-y)}{x}\right). \tag{3-12}$$

Using the orthogonality relation (2-1) we can say that second sum present in (3-12) is nonzero if and only if the equation $4y^2-4y+x=0$ admits a solution. We know that $4y^2-4y+x=0$ is solvable if and only if $1-x$ is a square in $\mathbb{F}_p$. Let $1-x=a^2$ for some $a\in\mathbb{F}_p$. Then $\frac{1}{2}(1\pm a)$ are solutions of $4y^2-4y+x=0$. Hence, we obtain

$$A=\begin{cases}p(p-1)\varphi(-2) & \text{if } x=1,\\ p(p-1)\varphi(-2)(\varphi(1+a)+\varphi(1-a)) & \text{if } x\neq 1 \text{ and } 1-x=a^2,\\ 0 & \text{if } 1-x\neq\square.\end{cases}$$

This completes the proof. □

In the next proposition, we again consider the same character sum as considered in Proposition 3.1 and express the sum as a special value of p-adic hypergeometric series. We use this proposition to prove Theorem 1.2.

Proposition 3.2. *For $x \in \mathbb{F}_p^\times$ we have*

$$\sum_{\chi\in\widehat{\mathbb{F}_p^\times}} g(\varphi\chi^2)g(\varphi\overline{\chi})g(\overline{\chi})\chi\left(\frac{x}{4}\right) = p(1-p)\varphi(2)\Gamma_p\left(\tfrac{1}{4}\right)\Gamma_p\left(\tfrac{3}{4}\right){}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\Big|\;\frac{1}{x}\right].$$

Proof. Let

$$A = \sum_{\chi\in\widehat{\mathbb{F}_p^\times}} g(\varphi\chi^2)g(\varphi\overline{\chi})g(\overline{\chi})\chi\left(\frac{x}{4}\right).$$

Since $\widehat{\mathbb{F}_p^\times} = \{\omega^j : 0 \le j \le p-2\}$, replacing χ by ω^j and applying the Gross–Koblitz formula we obtain

$$A = -\sum_{j=0}^{p-2}\omega^j\left(\frac{x}{4}\right)\pi^{(p-1)\ell_j}\Gamma_p\left(\left\langle\frac{1}{2}-\frac{2j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\frac{j}{p-1}\right), \tag{3-13}$$

$$\text{where}\quad \ell_j = \left\langle\frac{1}{2}-\frac{2j}{p-1}\right\rangle + \left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle + \left(\frac{j}{p-1}\right).$$

Applying (2-9) with $x = \langle\frac{1}{2}-\frac{2j}{p-1}\rangle$ and $m=2$ we obtain

$$\Gamma_p\left(\left\langle\frac{1}{2}-\frac{2j}{p-1}\right\rangle\right) = \frac{\Gamma_p\left(\frac{1}{2}\langle\frac{1}{2}-\frac{2j}{p-1}\rangle\right)\Gamma_p\left(\frac{1}{2}\langle\frac{1}{2}-\frac{2j}{p-1}\rangle+\frac{1}{2}\right)}{\Gamma_p\left(\frac{1}{2}\right)\omega\left(2^{(1-p)(1-\langle 1/2-2j/(p-1)\rangle)}\right)}.$$

Taking j in the intervals $\left[0, \lfloor\frac{p-1}{4}\rfloor\right]$, $\left(\lfloor\frac{p-1}{4}\rfloor, \lfloor\frac{3(p-1)}{4}\rfloor\right]$ and $\left(\lfloor\frac{3(p-1)}{4}\rfloor, p-2\right]$ we verify that

$$\Gamma_p\left(\frac{1}{2}\left\langle\frac{1}{2}-\frac{2j}{p-1}\right\rangle\right)\Gamma_p\left(\frac{1}{2}\left\langle\frac{1}{2}-\frac{2j}{p-1}\right\rangle+\frac{1}{2}\right) = \Gamma_p\left(\left\langle\frac{1}{4}-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{3}{4}-\frac{j}{p-1}\right\rangle\right)$$

and $\omega(2^{(1-p)(1-\langle 1/2-2j/(p-1)\rangle)}) = \varphi(2)\,\overline{\omega}^j(4)$. Therefore, we can write

$$\Gamma_p\left(\left\langle\frac{1}{2}-\frac{2j}{p-1}\right\rangle\right) = \frac{\Gamma_p\left(\langle\frac{1}{4}-\frac{j}{p-1}\rangle\right)\Gamma_p\left(\langle\frac{3}{4}-\frac{j}{p-1}\rangle\right)}{\Gamma_p\left(\frac{1}{2}\right)\varphi(2)\overline{\omega}^j(4)}. \tag{3-14}$$

Substituting (3-14) into (3-13) we obtain

$$A = -\frac{\varphi(2)}{\Gamma_p\left(\frac{1}{2}\right)}\sum_{j=0}^{p-2}\omega^j(x)\,\pi^{(p-1)\ell_j}\Gamma_p\left(\left\langle\frac{1}{4}-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{3}{4}-\frac{j}{p-1}\right\rangle\right) \times\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\frac{j}{p-1}\right). \tag{3-15}$$

Now,

$$\ell_j = \left\langle \frac{1}{2} - \frac{2j}{p-1} \right\rangle + \left\langle \frac{1}{2} + \frac{j}{p-1} \right\rangle + \left(\frac{j}{p-1} \right)$$
$$= 1 - \left\lfloor \frac{1}{2} - \frac{2j}{p-1} \right\rfloor - \left\lfloor \frac{1}{2} + \frac{j}{p-1} \right\rfloor.$$

By considering $\lfloor \frac{1}{2} - \frac{2j}{p-1} \rfloor = 2k + s$ for some $k \in \mathbb{Z}$ and $s = 0, 1$ it is straight forward to verify that

$$\left\lfloor \frac{1}{2} - \frac{2j}{p-1} \right\rfloor = \left\lfloor \frac{1}{4} - \frac{j}{p-1} \right\rfloor + \left\lfloor \frac{3}{4} - \frac{j}{p-1} \right\rfloor.$$

This gives

$$\ell_j = 1 - \left\lfloor \frac{1}{4} - \frac{j}{p-1} \right\rfloor - \left\lfloor \frac{3}{4} - \frac{j}{p-1} \right\rfloor - \left\lfloor \frac{1}{2} + \frac{j}{p-1} \right\rfloor. \tag{3-16}$$

Substituting (3-16) into (3-15) we obtain

$$A = p(1-p)\varphi(2)\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big)\,{}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \;\middle|\; \frac{1}{x}\right].$$

This completes the proof. □

Proof of Theorem 1.2. Comparing Propositions 3.1 and 3.2 for $x = 1$ we have

$${}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \;\middle|\; 1\right] = \frac{-\varphi(-1)}{\Gamma_p\big(\frac{1}{4}\big)\Gamma_p\big(\frac{3}{4}\big)}. \tag{3-17}$$

Applying (2-9) $\big($with $x = \frac{1}{2}$ and $m = 2\big)$ we obtain

$$\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big) = \varphi(2)\Gamma_p\big(\tfrac{1}{2}\big)^2 = -\varphi(-2). \tag{3-18}$$

Note that the last equality is obtained by using (2-8) $\big($with $x = \frac{1}{2}\big)$. Substituting (3-18) into (3-17) and using the fact that

$$\varphi(2) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \pmod 8, \\ -1 & \text{if } p \equiv \pm 3 \pmod 8, \end{cases}$$

we prove (1-9). Now, letting $x \neq 0, 1$, and $1 - x = a^2$ for some $a \in \mathbb{F}_p^\times$ and then comparing Propositions 3.1 and 3.2, we obtain

$${}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \;\middle|\; \frac{1}{x}\right] = \frac{-\varphi(-1)}{\Gamma_p\big(\frac{1}{4}\big)\Gamma_p\big(\frac{3}{4}\big)}(\varphi(1+a) + \varphi(1-a)). \tag{3-19}$$

Then substituting (3-18) into (3-19) we have

$${}_2G_2\left[\begin{matrix} \frac{1}{4}, & \frac{3}{4} \\ 0, & \frac{1}{2} \end{matrix} \;\middle|\; \frac{1}{x}\right] = \varphi(2) \cdot (\varphi(1+a) + \varphi(1-a)). \tag{3-20}$$

Replacing x by $\frac{1}{t}$ in (3-20) we obtain (1-10). Finally, if $1-x$ is not a square in $\mathbb{F}_p$ then again, comparing Propositions 3.1 and 3.2 we obtain

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;\frac{1}{x}\right]=0. \tag{3-21}$$

Replacing x by $\frac{1}{t}$ in (3-21) we derive (1-11). This completes the proof of the theorem. □

Proof of Theorem 1.6. If $t\neq 0, 1$ and $1-\frac{1}{t}=\frac{t-1}{t}\neq\square$ then from (1-11) we have

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;t\right]=0. \tag{3-22}$$

Replacing t by $\frac{1}{x}$ in (3-22) we obtain that if $1-x\neq\square$ then

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;\frac{1}{x}\right]=0. \tag{3-23}$$

Similarly, if $t\neq 0, 1$ and $1-\frac{1}{1-t}=\frac{t}{t-1}\neq\square$ then (1-11) gives

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;1-t\right]=0. \tag{3-24}$$

Replacing $1-t$ by $\frac{1}{1-x}$ in (3-24) we can write that if $x\neq\square$ then

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;\frac{1}{1-x}\right]=0. \tag{3-25}$$

Combining (3-23) and (3-25) we obtain (1-13). Now, let $x=b^2$. Putting $1-x=\frac{1}{t}$ we have $1-\frac{1}{t}=b^2$. Applying (1-10) we have

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;t\right]=\varphi(2)\cdot(\varphi(1+b)+\varphi(1-b)).$$

Therefore, if $x=b^2$ then

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;\frac{1}{1-x}\right]=\varphi(2)\cdot(\varphi(1+b)+\varphi(1-b)). \tag{3-26}$$

Combining (3-23) and (3-26) we deduce (1-14). Let $1-x=a^2$. Also, let $x=\frac{1}{t}$. Then $1-\frac{1}{t}=a^2$. Using (1-10) we obtain that

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;t\right]=\varphi(2)\cdot(\varphi(1+a)+\varphi(1-a)).$$

Therefore, if $1-x=a^2$ then

$$ {}_2G_2\left[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & \frac{1}{2}\end{matrix}\;\middle|\;\frac{1}{x}\right]=\varphi(2)\cdot(\varphi(1+a)+\varphi(1-a)). \tag{3-27}$$

Combining (3-26) and (3-27) we derive (1-15). Finally, comparing (3-25) and (3-27) we obtain (1-16). This completes the proof of the theorem. □

Proof of Theorem 1.7. Again, consider the sum

$$A=\sum_{\chi\in\widehat{\mathbb{F}_p^\times}} g(\varphi\chi^2)g(\varphi\overline{\chi})g(\overline{\chi})\chi\left(\frac{x}{4}\right).$$

Then from (3-15), and (3-16) we have

$$A=-\frac{\varphi(2)}{\Gamma_p\left(\frac{1}{2}\right)}\sum_{j=0}^{p-2}\omega^j(x)\pi^{(p-1)\ell_j}\Gamma_p\left(\left\langle\frac{1}{4}-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{3}{4}-\frac{j}{p-1}\right\rangle\right) \times\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\frac{j}{p-1}\right), \tag{3-28}$$

$$\text{where}\quad \ell_j=1-\left\lfloor\tfrac{1}{4}-\tfrac{j}{p-1}\right\rfloor-\left\lfloor\tfrac{3}{4}-\tfrac{j}{p-1}\right\rfloor-\left\lfloor\tfrac{1}{2}+\tfrac{j}{p-1}\right\rfloor.$$

Now, the term for $j=0$ present in (3-28) is equal to $p\varphi(2)\Gamma_p\left(\frac{1}{4}\right)\Gamma_p\left(\frac{3}{4}\right)$. Therefore, we have

$$A=-\frac{\varphi(2)}{\Gamma_p\left(\frac{1}{2}\right)}\sum_{j=1}^{p-2}\omega^j(x)\pi^{(p-1)\ell_j}\Gamma_p\left(\left\langle\frac{1}{4}-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{3}{4}-\frac{j}{p-1}\right\rangle\right) \times\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\frac{j}{p-1}\right)+p\varphi(2)\Gamma_p\left(\tfrac{1}{4}\right)\Gamma_p\left(\tfrac{3}{4}\right). \tag{3-29}$$

Using (2-11) we can write

$$\begin{aligned}A&=p\varphi(2)\Gamma_p\left(\tfrac{1}{4}\right)\Gamma_p\left(\tfrac{3}{4}\right)+\frac{\varphi(2)}{\Gamma_p\left(\frac{1}{2}\right)}\sum_{j=1}^{p-2}\omega^j(-x)\pi^{(p-1)\ell_j}\Gamma_p\left(\left\langle\frac{1}{4}-\frac{j}{p-1}\right\rangle\right)\\&\quad\times\Gamma_p\left(\left\langle\frac{3}{4}-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\left\langle1-\frac{j}{p-1}\right\rangle\right)\Gamma_p\left(\frac{j}{p-1}\right)^2\\&=-p\varphi(2)\sum_{j=1}^{p-2}\omega^j(-x)\frac{(-p)^{(-\lfloor1/2+j/(p-1)\rfloor)}\Gamma_p\left(\left\langle\frac{1}{2}+\frac{j}{p-1}\right\rangle\right)}{\Gamma_p\left(\frac{1}{2}\right)}\Gamma_p\left(\left\langle1-\frac{j}{p-1}\right\rangle\right)\\&\quad\times(-p)^{-\lfloor1/4-j/(p-1)\rfloor-\lfloor3/4-j/(p-1)\rfloor}\Gamma_p\left(\frac{j}{p-1}\right)^2\Gamma_p\left(\left\langle\frac{1}{4}-\frac{j}{p-1}\right\rangle\right)\\&\quad\times\Gamma_p\left(\left\langle\frac{3}{4}-\frac{j}{p-1}\right\rangle\right)+p\varphi(2)\Gamma_p\left(\tfrac{1}{4}\right)\Gamma_p\left(\tfrac{3}{4}\right).\end{aligned}$$

Also, applying (2-12) we obtain

$$A=-\varphi(2)\sum_{t\in\mathbb{F}_p^\times}\varphi(t(t-1))\sum_{j=1}^{p-2}\omega^j\Big(\frac{x}{t}\Big)(-p)^{-\lfloor 1/4-j/(p-1)\rfloor-\lfloor 3/4-j/(p-1)\rfloor} \times\Gamma_p\Big(\Big\langle\frac{1}{4}-\frac{j}{p-1}\Big\rangle\Big)\Gamma_p\Big(\Big\langle\frac{3}{4}-\frac{j}{p-1}\Big\rangle\Big)\Gamma_p\Big(\frac{j}{p-1}\Big)^2 + p\varphi(2)\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big). \tag{3-30}$$

The term under summation for $j=0$ is equal to

$$\begin{aligned}-\varphi(2)\sum_{t\in\mathbb{F}_p^\times}\varphi(t(t-1))\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big)&=-\varphi(-2)\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big)J(\varphi,\varphi)\\&=-p\varphi(2)\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big)\binom{\varphi}{\varphi}\\&=\varphi(2)\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big).\end{aligned}$$

Note that the last equality is obtained by applying (2-6). Using this value in (3-30) we obtain

$$A=(p-1)\varphi(2)\Gamma_p\big(\tfrac{1}{4}\big)\Gamma_p\big(\tfrac{3}{4}\big)\Bigg(1+\sum_{t\in\mathbb{F}_p^\times}\varphi(t(t-1))\,{}_2G_2\Bigg[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix}\;\Bigg|\;\frac{t}{x}\Bigg]\Bigg). \tag{3-31}$$

Now, from Proposition 3.1 comparing the values of A and using (3-18) we deduce (1-17), (1-18), and (1-19). This completes the proof of the theorem. □

Proof of Theorem 1.8. Applying the transformations [McCarthy 2013, Lemma 3.3; 2012c, Proposition 2.5] for $x, t\neq 0$ we obtain

$$\begin{aligned}{}_2G_2\Bigg[\begin{matrix}\frac{1}{4}, & \frac{3}{4}\\ 0, & 0\end{matrix}\;\Bigg|\;\frac{t}{x}\Bigg]&=\binom{\chi_4^3}{\varepsilon}^{-1}{}_2F_1\Bigg(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\Bigg|\;\frac{x}{t}\Bigg)\\&=-p\cdot{}_2F_1\Bigg(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\Bigg|\;\frac{x}{t}\Bigg).\end{aligned} \tag{3-32}$$

Note that we obtain the last equality by using (2-6).

Let $x=1$. Then substituting (3-32) into (1-17) we have

$$-p\sum_{t\in\mathbb{F}_p^\times}\varphi(t(t-1))\,{}_2F_1\Bigg(\begin{matrix}\chi_4, & \chi_4^3\\ & \varepsilon\end{matrix}\;\Bigg|\;\frac{1}{t}\Bigg)=-1-p\varphi(2). \tag{3-33}$$

Replacing t by $1/t$ we derive (1-20). Similarly, if $x\neq 1$ and $1-x$ is not a square then substituting (3-32) into (1-18) and replacing t by x/t we obtain (1-21). Finally, if $x\neq 1$ and $1-x=a^2$ then substituting (3-32) into (1-19) and replacing t by x/t we deduce (1-22). This completes the proof. □

Proof of Theorem 1.9. If $1-x \neq \square$ then applying Propositions 3.1 and 3.2 we have

$$ {}_2G_2\left[\begin{matrix} \frac14, & \frac34 \\ 0, & \frac12 \end{matrix} \;\Big|\; \frac{1}{x}\right] = {}_2G_2\left[\begin{matrix} \frac14, & \frac34 \\ 0, & \frac12 \end{matrix} \;\Big|\; \frac{x-1}{x}\right] = 0. $$

This proves (1-25). Now, let $1-x=a^2$ for some $a\in\mathbb{F}_p^{\times}$. Let a^{-1} denote the inverse of a in $\mathbb{F}_p^{\times}$. Then again applying Propositions 3.1 and 3.2 we have

$$ {}_2G_2\left[\begin{matrix} \frac14, & \frac34 \\ 0, & \frac12 \end{matrix} \;\Big|\; \frac{1}{x}\right] = -\frac{\varphi(-1)}{\Gamma_p\left(\frac14\right)\Gamma_p\left(\frac34\right)}(\varphi(1+a)+\varphi(1-a)), \tag{3-34} $$

and

$$ {}_2G_2\left[\begin{matrix} \frac14, & \frac34 \\ 0, & \frac12 \end{matrix} \;\Big|\; \frac{x-1}{x}\right] = -\frac{\varphi(-1)}{\Gamma_p\left(\frac14\right)\Gamma_p\left(\frac34\right)}(\varphi(1+a^{-1})+\varphi(1-a^{-1})). \tag{3-35} $$

Comparing (3-34) and (3-35) we prove (1-26). This completes the proof. $\square$

Concluding Remarks. Let $\mathbb{F}_q$ be a finite field with q elements, where $q=p^r$. We note that all the transformations and special values of p-adic hypergeometric series that are proved in this paper can also be extended to the q-version of the p-adic hypergeometric series ${}_nG_n[\cdots \mid t]_q$ with $t\in\mathbb{F}_q$ using [McCarthy 2013, Definition 5.1]. We avoid this case here for brevity. We also make the same comment for Gaussian hypergeometric functions over $\mathbb{F}_q$. We believe that using this method we can settle many other transformation formulas for p-adic hypergeometric series that are analogous to classical hypergeometric series transformations. This is considered as the subject of forthcoming work.

Acknowledgement

The author would like to thank the referee for his/her review and highly appreciates the valuable suggestions and comments.

References

[Bailey 1935] W. N. Bailey, *Generalized hypergeometric series*, Cambridge Tracts in Mathematics and Mathematical Physics **32**, Cambridge University Press, London, 1935. JFM

[Berndt et al. 1998] B. C. Berndt, R. J. Evans, and K. S. Williams, *Gauss and Jacobi sums*, Wiley, New York, 1998. MR Zbl

[Dixon 1903] A. C. Dixon, "Summation of a certain series", *Proc. Lond. Math. Soc.* **35** (1903), 284–289. MR Zbl

[Evans and Greene 2009] R. Evans and J. Greene, "Clausen's theorem and hypergeometric functions over finite fields", *Finite Fields Appl.* **15**:1 (2009), 97–109. MR Zbl

[Fuselier 2010] J. G. Fuselier, "Hypergeometric functions over $\mathbb{F}_p$ and relations to elliptic curves and modular forms", *Proc. Amer. Math. Soc.* **138**:1 (2010), 109–123. MR Zbl

[Fuselier and McCarthy 2016] J. G. Fuselier and D. McCarthy, "Hypergeometric type identities in the p-adic setting and modular forms", *Proc. Amer. Math. Soc.* **144**:4 (2016), 1493–1508. MR Zbl

[Gauss 1812] C. F. Gauss, "Disquisitiones generales circa seriem infinitam $1+\frac{\alpha\beta}{1\cdot\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot2\cdot\gamma(\gamma+1)}xx+\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot2\cdot3\cdot\gamma(\gamma+1)(\gamma+2)}x^3+$ etc.", pp. 233–279 in *Commentationes societatis regiae scientiarum Gottingensis recentiores*, vol. II, 1812. Reprinted as pp. 123–163, 207–229 in *Carl Friedrich Gauss Werke*, Band III, Königlichen Gesellschaft der Wissenschaften, Göttingen, 1866.

[Greene 1987] J. Greene, "Hypergeometric functions over finite fields", *Trans. Amer. Math. Soc.* **301**:1 (1987), 77–101. MR Zbl

[Gross and Koblitz 1979] B. H. Gross and N. Koblitz, "Gauss sums and the p-adic Γ-function", *Ann. of Math.* (2) **109**:3 (1979), 569–581. MR Zbl

[Ireland and Rosen 2005] K. Ireland and M. Rosen, *A classical introduction to modern number theory*, International ed., Springer, 2005.

[Koblitz 1980] N. Koblitz, *p-adic analysis: a short course on recent work*, London Mathematical Society Lecture Note Series **46**, Cambridge University Press, 1980. MR Zbl

[Kummer 1836] E. E. Kummer, "Über die hypergeometrische Reihe $1+\frac{\alpha\cdot\beta}{1\cdot\gamma}x+\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot2\cdot\gamma(\gamma+1)}x^2+\frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot2\cdot3\cdot\gamma(\gamma+1)(\gamma+2)}x^3+\cdots$", *J. Reine Angew. Math.* **15** (1836), 39–83. MR Zbl

[Lennon 2011a] C. Lennon, "Gaussian hypergeometric evaluations of traces of Frobenius for elliptic curves", *Proc. Amer. Math. Soc.* **139**:6 (2011), 1931–1938. MR Zbl

[Lennon 2011b] C. Lennon, "Trace formulas for Hecke operators, Gaussian hypergeometric functions, and the modularity of a threefold", *J. Number Theory* **131**:12 (2011), 2320–2351. MR Zbl

[Lidl and Niederreiter 1983] R. Lidl and H. Niederreiter, *Finite fields*, Encyclopedia of Mathematics and its Applications **20**, Addison-Wesley Publishing Company, Reading, MA, 1983. MR Zbl

[McCarthy 2010] D. McCarthy, "${}_3F_2$ hypergeometric series and periods of elliptic curves", *Int. J. Number Theory* **6**:3 (2010), 461–470. MR Zbl

[McCarthy 2012a] D. McCarthy, "Extending Gaussian hypergeometric series to the p-adic setting", *Int. J. Number Theory* **8**:7 (2012), 1581–1612. MR Zbl

[McCarthy 2012b] D. McCarthy, "On a supercongruence conjecture of Rodriguez-Villegas", *Proc. Amer. Math. Soc.* **140**:7 (2012), 2241–2254. MR Zbl

[McCarthy 2012c] D. McCarthy, "Transformations of well-poised hypergeometric functions over finite fields", *Finite Fields Appl.* **18**:6 (2012), 1133–1147. MR Zbl

[McCarthy 2013] D. McCarthy, "The trace of Frobenius of elliptic curves and the p-adic gamma function", *Pacific J. Math.* **261**:1 (2013), 219–236. MR Zbl

[Mortenson 2005] E. Mortenson, "Supercongruences for truncated ${}_{n+1}F_n$ hypergeometric series with applications to certain weight three newforms", *Proc. Amer. Math. Soc.* **133**:2 (2005), 321–330. MR Zbl

[Slater 1966] L. J. Slater, *Generalized hypergeometric functions*, Cambridge University Press, 1966. MR Zbl

[Whipple 1925] F. J. W. Whipple, "On well-poised series, generalized hypergeometric series having parameters in pairs, each pair with the same sum", *Proc. London Math. Soc.* (2) **24**:4 (1925), 247–263. MR Zbl

Received March 19, 2020. Revised May 1, 2020.

NEELAM SAIKIA
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
NORTH GUWAHATI
ASSAM
INDIA
nlmsaikia1@gmail.com

CONTENTS

Volume 307, no. 1 and no. 2

Guidelines for Authors

Authors may submit articles at msp.org/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript 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 307 No. 2 August 2020