Algebra & Number Theory

Volume 18
2024
No. 12

msp

Algebra & Number Theory

msp.org/ant

PRODUCTION

production@msp.org
Silvio Levy, Scientific Editor

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

The subscription price for 2024 is US $525/year for the electronic version, and $770/year (+$65, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Algebra & Number Theory (ISSN 1944-7833 electronic, 1937-0652 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

ANT peer review and production are managed by EditFLOW® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

ALGEBRA AND NUMBER THEORY 18:12 (2024)
https://doi.org/10.2140/ant.2024.18.2133

Moduli of linear slices of high degree smooth hypersurfaces

Anand Patel, Eric Riedl and Dennis Tseng

We study the variation of linear sections of hypersurfaces in $\mathbb{P}^n$. We completely classify all plane curves, necessarily singular, whose line sections do not vary maximally in moduli. In higher dimensions, we prove that the family of hyperplane sections of any smooth degree d hypersurface in $\mathbb{P}^n$ varies maximally for $d \geq n+3$. In the process, we generalize the classical Grauert–Mülich theorem about lines in projective space, both to k-planes in projective space and to free rational curves on arbitrary varieties.

1. Introduction

A fundamental technique for studying a degree d complex hypersurface X in projective space $\mathbb{P}^n$ is to intersect it with hyperplanes. The family of varieties thus obtained can be represented by a map to moduli

$$\phi : \mathbb{P}^{n*} \dashrightarrow \mathbb{P}H^0(\mathscr{O}_{\mathbb{P}^{n-1}}(d)) /\!\!/ \mathrm{SL}_n, \quad [\Lambda] \mapsto [\Lambda \cap X].$$

Basic properties of ϕ are still not understood, even under regularity assumptions on X. Take, for instance, the problem of determining the dimension of its image. If X is assumed to be *general*, then ϕ can directly be shown to have maximal rank, i.e., its image is as large as possible, as done in [van Opstall and Veliche 2007]. However, once we assume X is an *arbitrary* hypersurface, the story becomes more complicated, with several authors studying special cases in the last few decades. Even in the case of a reduced plane curve X, showing maximal variation is not a trivial task. Thirty years ago, while studying PGL_3-orbits of plane curves, Aluffi and Faber [1993, Proposition 4.2] cleverly exploited the classical Plücker formulas to prove that *smooth* plane curves of degree at least 5 always have maximum variation of linear sections. However, if the curve X is singular, then ϕ can fail to have maximal rank, and Aluffi and Faber were not able to completely analyze this case.

Quite generally, if the dimension of the projective automorphism group of X is larger than expected (e.g., if X is a cone), then linear slices must fail to vary maximally in moduli. Outside this class of hypersurfaces, we are unaware of any other examples where ϕ fails to have maximal rank, so we pose the following question:

Question 1.1. If ϕ fails to have maximal rank, must X have a positive-dimensional projective automorphism group?

MSC2020: 14J10, 14J70.
Keywords: hypersurfaces, linear sections, plane curves, stability, Grauert–Mülich.

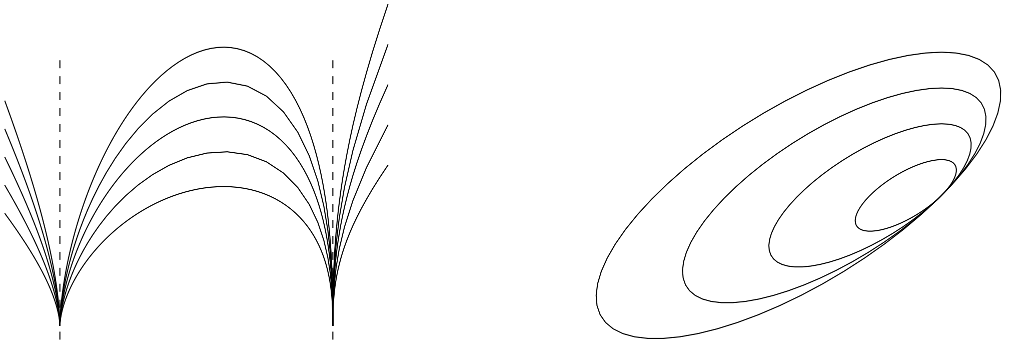

Figure 1. Singular curves for which ϕ fails to have maximal rank. Left: union of orbits under $\mathbb{G}_m$ action. Right: union of orbits under $\mathbb{G}_a$ action (quadritangent conics).

We are concerned exclusively with the case where X is a hypersurface, although one can ask similar questions for other subvarieties of $\mathbb{P}^n$, such as in [Mckernan 1991]. Our first result is to answer Question 1.1 affirmatively when $X \subset \mathbb{P}^2$ is a plane curve:

Theorem 1.2. *If $X \subset \mathbb{P}^2_{\mathbb{C}}$ is an arbitrary plane curve and if ϕ fails to have maximal rank, then X has infinitely many projective automorphisms.*

Given Theorem 1.2, we see that of all curves where ϕ fails to have maximal rank have stabilizer containing $\mathbb{G}_m$ or $\mathbb{G}_a$, where typical examples are depicted in Figure 1.

The map ϕ is even more difficult to understand for larger-dimensional hypersurfaces — we restrict our attention primarily to smooth hypersurfaces. Beauville [1990] investigated the case where ϕ is a *constant map* and classified the smooth hypersurfaces X for which the family of hyperplane sections has constant moduli. This phenomenon happens only for very special hypersurfaces in positive characteristic. In contrast, we prove:

Theorem 1.3. *If $X \subset \mathbb{P}^n_{\mathbb{C}}$ is a smooth hypersurface of degree $d \geq n+3$, then ϕ has maximal rank.*

We can also intersect a hypersurface X with k-planes for smaller k, obtaining natural analogues

$$\phi_k : \mathbb{G}(k,n) \dashrightarrow \mathbb{P}H^0(\mathscr{O}_{\mathbb{P}^k}(d)) /\!\!/ \mathrm{SL}_{k+1}, \quad [\Lambda] \mapsto [\Lambda \cap X],$$

and ask similar questions about ϕ_k. Harris, Mazur, and Pandharipande [Harris et al. 1998], and then later Starr [2006], studied the situation where ϕ_k is expected to be dominant, relating the problem of establishing dominance to the question of unirationality of low degree hypersurfaces. When ϕ_k is expected to be generically finite and dominant, the problem of establishing its degree has also appeared in the literature. In this direction, see [Cadman and Laza 2008; Lee et al. 2020; 2023].

We are able to generalize Theorem 1.3, and prove that ϕ_k has maximal rank under some restrictions on k:

Theorem 1.4. *If $X \subset \mathbb{P}^n_{\mathbb{C}}$ is a smooth hypersurface of degree d, then ϕ_k has maximal rank assuming $d \geq n+3$ and $k \geq \frac{2}{3}n$.*

For $k < \frac{2}{3}n$, we obtain similar statements, but with d forced to be larger (see Theorems 5.8 and 5.9).

Broadening the topic even further, we can intersect X with other types of varieties, for example, rational curves of degree e. In this way, we obtain a map from the variety of degree e rational curves in $\mathbb{P}^n$ to the moduli space of ed points on $\mathbb{P}^1$. Our methods provide results in this context — see Theorems 4.2 and 5.2.

1A. ***Methods.*** The log tangent sheaf $T_{\mathbb{P}^n}(-\log X)$ and the Grauert–Mülich theorem play key roles in our approach. We identify the tangent space of the fiber of ϕ at a point $[\Lambda]$ with sections of the log tangent sheaf $T_{\mathbb{P}^n}(-\log X)$ restricted to Λ. Then, we adapt the argument in the usual Grauert–Mülich theorem [Okonek et al. 1980, Theorem 2.1.4] to produce sections or subsheaves of the log tangent sheaf $T_{\mathbb{P}^n}(-\log X)$. In the plane curve case, this forces X to be an integral curve for a vector field on $\mathbb{P}^2$, leading to the classification in Theorem 4.2. In the higher dimensional case, we appeal to a result of Guenancia regarding the semistability of $T_{\mathbb{P}^n}(-\log X)$, when $(\mathbb{P}^n, X)$ is a log-canonical pair and $d \geq n+2$. In particular, all our results in this case actually hold when $(\mathbb{P}^n, X)$ is a log-canonical pair, not only when X is smooth.

Our methods will produce results in other contexts, for example, if we replace $\mathbb{P}^n$ with a homogeneous space G/P.

2. Preliminaries

We introduce conventions and basic definitions.

2A. ***Notation and conventions.*** We will work over the complex numbers. We identify vector bundles with locally free sheaves throughout, and all our sheaves are coherent. A *subbundle* of a vector bundle V is a locally free subsheaf $W \subset V$ such that V/W is also locally free. For us, a variety is an integral scheme of finite type. If F is a coherent sheaf on a scheme X, we denote by $\mathrm{ev} : H^0(X, F) \otimes \mathscr{O}_X \to F$ the natural evaluation map.

We denote by $\mathrm{Mor}_e(\mathbb{P}^k, \mathbb{P}^n)$ the variety parameterizing morphisms $f : \mathbb{P}^k \to \mathbb{P}^n$ with $f^*\mathscr{O}(1) = \mathscr{O}(e)$. Explicitly, $\mathrm{Mor}_e(\mathbb{P}^k, \mathbb{P}^n)$ is a Zariski open subset of $\mathbb{P}\big(H^0(\mathscr{O}_{\mathbb{P}^k}(e))^{\oplus n+1}\big)$ parameterizing tuples $(A_0, \dots, A_n)$ of homogeneous degree e forms on $\mathbb{P}^k$ which do not vanish simultaneously anywhere on $\mathbb{P}^k$. More generally, $\mathrm{Mor}(X, Y)$ denotes the (not finite-type) scheme parameterizing morphisms from the scheme X to the scheme Y.

Given a torsion-free sheaf E on a projective variety X, we let its slope $\mu(E)$ denote the ratio $\frac{\deg(E)}{\mathrm{rank}(E)}$, where $\deg(E) = \int_X c_1(E)\mathscr{O}_X(1)^{\dim(X)-1}$. We call E semistable (respectively stable) if there is no proper subsheaf F with $\mu(F) > \mu(E)$ (respectively $\mu(F) \geq \mu(E)$). In general, the *Harder–Narasimhan* filtration of E is

$$0 = E_0 \subsetneq E_1 \subsetneq E_2 \subsetneq \cdots \subsetneq E_a = E,$$

where the subquotients $E_1/E_0, E_2/E_1, \dots, E_a/E_{a-1}$ are semistable and have strictly decreasing slopes. Finally, if E, F are two coherent sheaves, then $\mathrm{Hom}(E, F)$ will denote the vector space of global homomorphisms $E \to F$ while $\underline{\mathrm{Hom}}(E, F)$ will denote the sheaf of local homomorphisms.

2B. ***The map to moduli* Φ.** Suppose $X \in \mathbb{P}^n$ is a degree d hypersurface. After fixing integers $e \geq 1$, $k \leq n-1$, we get the induced *map to moduli*

$$\Phi : \mathrm{Mor}_e(\mathbb{P}^k, \mathbb{P}^n) \dashrightarrow \mathbb{P}H^0(\mathbb{P}^k, \mathscr{O}_{\mathbb{P}^k}(de)), \quad \iota \mapsto [\iota^{-1}(X)].$$

We say that Φ has maximal rank if the dimension of its image is

$$\max\bigl\{\dim(\mathrm{Mor}_e(\mathbb{P}^k, \mathbb{P}^n)), \mathbb{P}H^0(\mathbb{P}^k, \mathscr{O}_{\mathbb{P}^k}(de))\bigr\}.$$

Equivalently, since we are working over $\mathbb{C}$, the derivative of Φ at a general point has maximum rank.

Though our methods give results for all e, k, we are primarily interested in the cases where $e = 1$ or $k = 1$. Therefore, we have only stated our results in these cases. In the case $e = 1$, Φ having maximal rank is equivalent to the map

$$\mathbb{G}(k, n) \dashrightarrow \mathbb{P}H^0(\mathbb{P}^k, \mathscr{O}_{\mathbb{P}^k}(d)) /\!/ \mathrm{SL}_k, \quad [\Lambda] \mapsto [\Lambda \cap X],$$

having maximal rank, assuming the general k-plane slice of X has no infinitesimal automorphisms. Whenever our results apply, this condition will always be satisfied.

2C. ***Log tangent sheaves.*** We now introduce the main tool of the paper. We suspect Lemma 2.2 is well-known to experts but include a proof for want of a suitable reference. Everything in this section should work for a reduced divisor in an arbitrary smooth ambient variety, but we will focus on the case that the ambient variety is projective space.

Let $D \subset \mathbb{P}^n$ be a reduced hypersurface. Viewing D as a divisor in the smooth ambient variety $\mathbb{P}^n$, we get the log tangent sheaf $T_{\mathbb{P}^n}(-\log D)$, which sits inside the exact sequence

$$0 \to T_{\mathbb{P}^n}(-\log D) \to T_{\mathbb{P}^n} \to \mathscr{O}_D(D) \to \mathscr{O}_{D_{\mathrm{sing}}}(D) \to 0,$$

where D_{sing} is the singular subscheme cut out of $\mathbb{P}^n$ by the equation for D and its partials. In terms of background, we only assume what is covered in [Liao 2013, 2.1.2], but see [Saito 1980] for the original reference. One can check that $T_{\mathbb{P}^n}(-\log D)$ is a vector bundle when D is smooth using local coordinates; in general $T_{\mathbb{P}^n}(-\log D)$ is a reflexive sheaf.

Remark 2.1. Informally, local sections of $T_{\mathbb{P}^n}(-\log D)$ represent local vector fields which are tangent to D. This can be seen explicitly by noting that the map $T_{\mathbb{P}^n} \to \mathscr{O}_D(D)$ in the exact sequence above is given by $\theta \mapsto \theta(f)$, where θ is a vector field and f is the (local) equation for D. If we identify $\mathscr{O}_D(D)$ with $N_{D/\mathbb{P}^n}$, the map $T_{\mathbb{P}^n} \to \mathscr{O}_D(D)$ is also $T_{\mathbb{P}^n} \to T_{\mathbb{P}^n}|_D \to N_{D/\mathbb{P}^n}$.

Let $\mathbb{P}^k \xrightarrow{\iota} \mathbb{P}^n$ be a map defined by degree e homogeneous forms, and suppose $Z \subset \mathbb{P}^n$ is a subscheme. We say an infinitesimal deformation $\iota_\epsilon : \mathbb{P}^k \times \operatorname{Spec} \mathbb{C}[\epsilon]/(\epsilon^2) \to \mathbb{P}^n$ *preserves* $\iota^{-1}(Z)$ if $\iota_\epsilon^{-1}(Z) \subset \mathbb{P}^k \times \operatorname{Spec} \mathbb{C}[\epsilon]/(\epsilon^2)$ is the trivial deformation $\iota^{-1}(Z) \times \operatorname{Spec} \mathbb{C}[\epsilon]/(\epsilon^2)$. The point of this section is to prove the following lemma.

Lemma 2.2. *Let $\mathbb{P}^k \xrightarrow{\iota} \mathbb{P}^n$ be a map defined by degree e homogeneous forms whose image is not contained in D. Global sections of $\iota^* T_{\mathbb{P}^n}(-\log D)$ correspond to deformations of the map ι preserving the hypersurface $\iota^{-1}(D)$.*

Proof. First, sections of $\iota^* T_{\mathbb{P}^n}$ correspond to deformations of ι. More explicitly, ι is defined by an $(n+1)$-tuple of degree e forms in $k+1$ variables $A_0(s_0, \dots, s_k), \dots, A_n(s_0, \dots, s_k)$ sending $[s_0 : \cdots : s_k]$ to $[A_0(s_0, \dots, s_k) : \cdots : A_n(s_0, \dots, s_k)]$.

A deformation ι_ϵ of ι is given by another $(n+1)$-tuple of degree e forms in $k+1$ variables $B_0(s_0, \dots, s_k)$, $\dots, B_n(s_0, \dots, s_k)$. Explicitly, as a map from $\operatorname{Spec}(\mathbb{C}[\epsilon]/(\epsilon^2)) \times \mathbb{P}^k \to \mathbb{P}^n$ this is given in coordinates by ϵ, $[s_0, \dots, s_k]$ mapping to $[A_0(s_0, \dots, s_k) + \epsilon B_0(s_0, \dots, s_k) : \cdots : A_n(s_0, \dots, s_k) + \epsilon B_n(s_0, \dots, s_k)]$. The vector space of deformations is given by the quotient space of $(n+1)$-tuples of degree e forms $(B_0, \dots, B_n)$ modulo the 1-dimensional vector space generated by $(A_0, \dots, A_n)$.

Let D be defined by $F = 0$ where F is a reduced homogeneous form in $n+1$ variables. If we pull back the form F under the deformed map ι_ϵ, we obtain

$$\begin{aligned} &F(A_0(s_0, \dots, s_k) + \epsilon B_0(s_0, \dots, s_k), \dots, A_n(s_0, \dots, s_k) + \epsilon B_n(s_0, \dots, s_k)) \\ &= F(A_0(s_0, \dots, s_k), \dots, A_n(s_0, \dots, s_k)) + \epsilon \sum_{i=0}^{n} B_i(s_0, \dots, s_k) \cdot \partial_i F(A_0(s_0, \dots, s_k), \dots, A_n(s_0, \dots, s_k)). \end{aligned}$$

Therefore, deformations ι_ϵ that preserve $\iota^{-1}(D)$ correspond to choices of $B_0, \dots, B_n$ such that

$$\sum_{i=0}^{n} B_i(s_0, \dots, s_k) \cdot \partial_i F(A_0(s_0, \dots, s_k), \dots, A_n(s_0, \dots, s_k)) \tag{2-1}$$

is a scalar multiple of F.

Now, we wish to realize this latter condition as producing sections of the pulled back log tangent sheaf. First, the sections of the pulled back tangent sheaf $\iota^* T_{\mathbb{P}^n}$ can be computed via the Euler sequence

$$0 \to \mathcal{O}_{\mathbb{P}^k} \to \mathcal{O}_{\mathbb{P}^k}(e)^{n+1} \to \iota^* T_{\mathbb{P}^n} \to 0$$

to be the quotient space of $(n+1)$-tuples of linear forms $(B_0, \dots, B_n)$ modulo the 1-dimensional vector space generated by $(A_0, \dots, A_n)$.

The restricted vector field corresponding to $(B_0, \dots, B_n)$ is $\sum_{i=0}^{n} B_i \frac{\partial}{\partial x_i}$. Recall that $T_{\mathbb{P}^n}(-\log D)$ is the kernel of the map $T_{\mathbb{P}^n} \to \mathcal{O}_D(D)$ sending a vector field $\theta := \sum_{i=0}^{n} B_i \frac{\partial}{\partial x_i}$ to $\theta(F)$.

In other words, the defining equation is

$$\sum_{i=0}^{n} B_i \frac{\partial}{\partial x_i} F \equiv 0 \pmod{F}.$$

Pulling back this under ι yields exactly (2-1). □

3. Grauert–Mülich

The goal of this section is to generalize the classical Grauert–Mülich theorem [Okonek et al. 1980, Theorem 2.1.4] in two directions:

Proposition 3.1. *Let $Z \subset \mathbb{P}^N$ be a smooth projective variety and $f : \mathbb{P}^1 \to Z$ be a general map in an open subset $\mathcal{M}$ of $\operatorname{Mor}(\mathbb{P}^1, Z)$ such that $\mathbb{P}^1 \times \mathcal{M} \to Z$ has connected fibers. Suppose f^*T_Z is globally generated.*

Let E be a torsion free sheaf on Z and write f^*E as $\bigoplus_{i=1}^{b} \mathcal{O}(a_i)$ with $a_1 \geq \cdots \geq a_b$. If $a_j > a_{j+1} + 1$ for some j, then E has a subsheaf of rank j and degree $\frac{1}{\deg(f)} \sum_{i=1}^{j} a_i$. In particular, if E is semistable, then the bundle f^*E can be written as $\bigoplus_i \mathcal{O}(a_i)$ with $|a_i - a_{i+1}| \leq 1$.

For applications to slicing by k-planes, we will use Proposition 3.3.

Definition 3.2. Given a torsion free sheaf E on a smooth projective variety, let $\mu^{\max}(E)$ denote the maximum slope of a nontrivial subsheaf of E. It is also the slope of the first subsheaf appearing in its Harder–Narasimhan filtration. Similarly let $\mu^{\min}(E)$ denote the minimum slope of a nontrivial quotient of E. It is also the slope of the quotient of the last two subsheaves appearing in its Harder–Narasimhan filtration.

Proposition 3.3. *Let E be a torsion free sheaf on $\mathbb{P}^n$. Let Λ be a k-plane in $\mathbb{P}^n$, general with respect to E. Let $S \subsetneq E|_\Lambda$ be a sheaf appearing in the Harder–Narasimhan filtration of $E|_\Lambda$ and suppose*

$$\mu^{\min}(S) - \frac{1}{k} > \mu^{\max}(E|_\Lambda / S).$$

Then E is not semistable.

The proofs of Propositions 3.1 and 3.3 are very similar in spirit to standard proofs of Grauert–Mülich, such as the one found in [Okonek et al. 1980]. The argument relies crucially on the following lemma.

Lemma 3.4 (descent lemma from [Okonek et al. 1980, Lemma 2.1.2]). *Let Y and Z be smooth varieties and $\pi : Y \to Z$ be a surjective smooth morphism with connected fibers. Let E be a vector bundle on Z such that π^*E has a vector subbundle S with quotient vector bundle Q. If*

$$\operatorname{Hom}(T_{Y/Z}, \underline{\operatorname{Hom}}(S, Q)) = 0,$$

then S is the pullback of a subbundle of E on Z.

The key technical lemma of this section is Lemma 3.6, whose proof will use the following simple fact.

Lemma 3.5. *Let Y be a variety and E and F be two sheaves on Y. Suppose every semistable subquotient in the Harder–Narasimhan filtration of E has greater slope than every semistable subquotient of F, i.e., that $\mu^{\min}(E) > \mu^{\max}(F)$. Then, $\operatorname{Hom}(E, F) = 0$.*

Proof. Let $0 = E_0 \subset E_1 \subset \cdots \subset E_a = E$ be the Harder–Narasimhan filtration for E and $0 = F_0 \subset F_1 \subset \cdots \subset F_b = F$ be the Harder–Narasimhan filtration for F. Consider a map $\phi : E \to F$. We show $\phi = 0$.

First, ϕ induces a map $E_1 \to F_b/F_{b-1}$, which is zero since the source is semistable and has slope greater than the target, which is also semistable. Therefore, ϕ induces a map $E_1 \to F_{b-1}/F_{b-2}$, which again is zero for the same reason. Continuing this, we find the map $E_1 \to F$ is zero.

Then, we consider the induced map $E_2/E_1 \to F_b/F_{b-1}$ and repeat the argument above to find $E_2/E_1 \to F$ must be the zero map. Continuing this for E_3/E_2 and so on shows that the map ϕ is zero. □

Lemma 3.6. *Let Z be a smooth projective variety and let $\mathcal{U} \to \mathcal{M}$ be a smooth family of projective varieties with a smooth surjective map $\pi : \mathcal{U} \to Z$ having connected fibers. Let E be a torsion free sheaf on Z and let $\mathcal{U}_p$ be a general fiber of $\mathcal{U} \to \mathcal{M}$. Let S be a subsheaf of $\pi^* E|_{\mathcal{U}_p}$ appearing in the Harder–Narasimhan filtration of $\pi^* E|_{\mathcal{U}_p}$ such that*

$$\mu^{\min}(S) + \mu^{\min}(T_{\mathcal{U}/Z}|_{\mathcal{U}_p}) > \mu^{\max}(\pi^* E|_{\mathcal{U}_p}/S).$$

Then there is a subsheaf $\widetilde{S}$ on Z of E such that $\pi^*\widetilde{S}|_{\mathcal{U}_p}$ agrees with S on the locus where S is a vector bundle.

Proof. By [Shatz 1977, Lemmas 5 and 7], we can replace $\mathcal{M}$ by a dense open subset so that the members of the Harder–Narasimham filtration of $\pi^* E|_{\mathcal{U}_p}$ extend to a family over $\mathcal{U}$. Namely, there exists a sequence of subsheaves $0 = S_0 \subset S_1 \subset \cdots \subset S_a = \pi^* E$ that restrict to the Harder–Narasimhan filtration of $E|_{\mathcal{U}_p}$ for all $p \in \mathcal{M}$, so in particular $S = S_i|_{\mathcal{U}_p}$ for some i. If E and all S_j's are locally free, then we can immediately apply Lemma 3.4 to conclude. The next paragraphs deal with the possibility that E or the S_j's are not locally free by passing to a general curve in $\mathcal{U}_p$.

We have an open subset $\mathcal{U}^0 \subset \mathcal{U}$ whose complement has codimension at least 2 and consists of the points over which S_i and $\pi^* E/S_i$ are both vector bundles. The image of $\mathcal{U}^0$ in Z is an open subset Z^0 (by flatness of π) whose complement must also have codimension at least 2.

Now, we can apply Lemma 3.4 in the case $Y = \mathcal{U}^0$ and $Z = Z^0$. In order to do so, we must show that

$$\operatorname{Hom}\bigl(T_{\mathcal{U}^0/Z^0}, \underline{\operatorname{Hom}}(S_i|_{\mathcal{U}^0}, (\pi^* E/S_i)|_{\mathcal{U}^0})\bigr) = 0. \tag{3-1}$$

For this, observe that because all sheaves appearing in (3-1) are locally free, it suffices to show the lack of homomorphisms when we restrict to a general fiber $\mathcal{U}_p$. Then, we use the same idea and restrict to a general complete intersection curve $C \subset \mathcal{U}_p$ of sufficiently high degree. Restricting the Harder–Narasimhan filtration of $\pi^* E|_{\mathcal{U}_p}$ to C results in a sequence of vector subbundles because each semistable subquotient on $\mathcal{U}_p$ is in particular torsion-free, so the Harder–Narasimhan filtration is a sequence of vector subbundles away from a set of codimension at least 2 in $\mathcal{U}_p$. Since C can be chosen to avoid this set, restricting a sequence of subbundles yields a sequence of subbundles. By [Mehta and Ramanathan 1982], this sequence of sub vector bundles on C is the Harder–Narasimhan filtration of $\pi^* E|_C$. To show (3-1), it therefore suffices to show

$$\operatorname{Hom}\bigl(T_{\mathcal{U}/Z}|_C, \underline{\operatorname{Hom}}(S|_C, (\pi^* E|_{\mathcal{U}_p}/S)|_C)\bigr) = \operatorname{Hom}\bigl(T_{\mathcal{U}/Z}|_C \otimes S|_C, (\pi^* E|_{\mathcal{U}_p}/S)|_C\bigr) = 0.$$

We conclude by applying Lemma 3.5, keeping in mind that $S|_C$ is part of the Harder–Narasimhan filtration of $\pi^* E|_C$, and that the following slope equalities hold:

$$\begin{gathered}
\mu^{\max}((\pi^* E|_{\mathcal{U}_p}/S)|_C) = \deg(C)\mu^{\max}(\pi^* E|_{\mathcal{U}_p}/S),\\
\mu^{\min}(S|_C) = \deg(C)\mu^{\min}(S),\\
\mu^{\min}(T_{\mathcal{U}/Z}|_C) = \deg(C)\mu^{\min}(T_{\mathcal{U}/Z}),\\
\mu^{\min}(T_{\mathcal{U}/Z}|_C \otimes S|_C) = \mu^{\min}(S|_C) + \mu^{\min}(T_{\mathcal{U}/Z}|_C),
\end{gathered}$$

where $\deg(C)$ can be defined using any projective embedding of $\mathcal{U}_p$. □

In order for us to apply Lemma 3.6, it is necessary to understand the sheaf $T_{\mathcal{U}/Z}|_{\mathcal{U}_p}$. Lemmas 3.8 and 3.9 identify the sheaf in two common situations.

Definition 3.7. Let Y be a variety and V be a globally generated vector bundle on Y. Then, the *Lazarsfeld–Mukai* bundle of V is the kernel of the evaluation map $\mathcal{O}_Y \otimes H^0(V) \to V$.

Lemma 3.8. *Let Z be a smooth projective variety and $\mathcal{M}$ be a smooth open subset of a component of the Hilbert scheme of varieties on Z. Let $\mathcal{U}$ be the universal family, and suppose that the natural map $\pi : \mathcal{U} \to Z$ is smooth. Let $\mathcal{U}_p$ be a general fiber of $\mathcal{U} \to \mathcal{M}$. Then $T_{\mathcal{U}/Z}|_{\mathcal{U}_p}$ is the Lazarsfeld–Mukai bundle for the normal bundle $N_{\mathcal{U}_p/Z}$, defined by the short exact sequence*

$$0 \to T_{\mathcal{U}/Z}|_{\mathcal{U}_p} \to H^0(N_{\mathcal{U}_p/Z}) \otimes \mathcal{O}_{\mathcal{U}_p} \to N_{\mathcal{U}_p/Z} \to 0.$$

Proof. First we compare the normal sheaf of $\mathcal{U}$ in $\mathcal{M} \times Z$ to the normal sheaf $N_{\mathcal{U}_p/Z}$. We have the diagram

$$\begin{tikzcd}
 & 0 & 0 & & \\
 & \mathcal{O}_{\mathcal{U}_p}^N \arrow[u] \arrow[r, "="] & \mathcal{O}_{\mathcal{U}_p}^N \arrow[u] & & \\
0 \arrow[r] & T_{\mathcal{U}}|_{\mathcal{U}_p} \arrow[u] \arrow[r] & T_{\mathcal{M}\times Z}|_{\mathcal{U}_p} \arrow[u] \arrow[r] & N_{\mathcal{U}/\mathcal{M}\times Z}|_{\mathcal{U}_p} \arrow[r] & 0 \\
0 \arrow[r] & T_{\mathcal{U}_p} \arrow[u] \arrow[r] & T_Z|_{\mathcal{U}_p} \arrow[u] \arrow[r] & N_{\mathcal{U}_p/Z} \arrow[u, "\cong"'] \arrow[r] & 0 \\
 & 0 \arrow[u] & 0 \arrow[u] & &
\end{tikzcd}$$

In this diagram, we have written $H^0(N_{\mathcal{U}_p/Z}) \otimes \mathcal{O}_{\mathcal{U}_p}$ as $\mathcal{O}_{\mathcal{U}_p}^N$, where $N = h^0(N_{\mathcal{U}_p/Z})$. We see that $N_{\mathcal{U}/\mathcal{M}\times Z}|_{\mathcal{U}_p}$ is isomorphic to $N_{\mathcal{U}_p/Z}$ by the eight lemma. Next we relate $N_{\mathcal{U}/\mathcal{M}\times Z}|_{\mathcal{U}_p}$ to the Lazarsfeld–Mukai bundle. Consider the diagram, where the lower right entry is computed by the eight lemma,

$$\begin{tikzcd}
 & 0 & 0 & & \\
 & T_Z|_{\mathcal{U}_p} \arrow[u] \arrow[r, "="] & T_Z|_{\mathcal{U}_p} \arrow[u] & & \\
0 \arrow[r] & T_{\mathcal{U}}|_{\mathcal{U}_p} \arrow[u] \arrow[r] & T_{\mathcal{M}\times Z}|_{\mathcal{U}_p} \arrow[u] \arrow[r] & N_{\mathcal{U}/\mathcal{M}\times Z}|_{\mathcal{U}_p} \arrow[r] & 0 \\
0 \arrow[r] & T_{\mathcal{U}/Z}|_{\mathcal{U}_p} \arrow[u] \arrow[r] & \pi^* T_{\mathcal{M}}|_{\mathcal{U}_p} \arrow[u] \arrow[r] & N_{\mathcal{U}/\mathcal{M}\times Z}|_{\mathcal{U}_p} \arrow[u, "\cong"'] \arrow[r] & 0 \\
 & 0 \arrow[u] & 0 \arrow[u] & &
\end{tikzcd}$$

Then since $\pi^* T_{\mathcal{M}}$ is constant on $\mathcal{U}_p$ and $N_{\mathcal{U}/\mathcal{M}\times Z}|_{\mathcal{U}_p} \cong N_{\mathcal{U}_p/Z}$, we see that the last row becomes

$$0 \to T_{\mathcal{U}/Z}|_{\mathcal{U}_p} \to H^0(N_{\mathcal{U}_p/Z}) \otimes \mathcal{O} \to N_{\mathcal{U}_p/Z} \to 0.$$

The result follows. □

Lemma 3.9. *Let Y and Z be smooth projective schemes and $\mathcal{M}$ be an open subset of* $\operatorname{Mor}(Y, Z)$*. Let $\pi : Y \times \mathcal{M} \to Z$ be the universal map. For $f : Y \to Z$ in $\mathcal{M}$, suppose $f^* T_Z$ is globally generated. Then, the restriction $T_{Y\times\mathcal{M}/Z}|_{Y\times\{[f]\}}$ is an extension of T_Y by the Lazersfeld–Mukai bundle of $f^* T_Z$.*

Proof. We have the relative tangent sequence

$$0 \to T_{Y\times\mathcal{M}/Z}|_{Y\times\{[f]\}} \to T_{Y\times\mathcal{M}}|_{Y\times\{[f]\}} \to f^* T_Z \to 0.$$

We have the natural decomposition $T_{Y\times\mathcal{M}/Z}|_{Y\times\{[f]\}} \cong H^0(f^* T_Z) \otimes \mathcal{O} \oplus T_Y$, with respect to which the natural map $T_{Y\times\mathcal{M}/Z}|_{Y\times\{[f]\}} \to f^* T_Z$ is $\mathrm{ev} + df$. Consider the following commutative diagram, where K is the Lazarsfeld–Mukai bundle of $f^* T_Z$:

$$\begin{tikzcd}
& 0 & 0 & 0 & \\
0 \arrow[r] & T_Y \arrow[u] \arrow[r, equal] & T_Y \arrow[u] \arrow[r] & 0 \arrow[u] \arrow[r] & 0 \\
0 \arrow[r] & T_{Y\times\mathcal{M}/Z}|_{Y\times\{[f]\}} \arrow[u] \arrow[r] & H^0(f^* T_Z) \otimes \mathcal{O} \oplus T_Y \arrow[u] \arrow[r, "\mathrm{ev}+df"] & f^* T_Z \arrow[u] \arrow[r] & 0 \\
0 \arrow[r] & K \arrow[u] \arrow[r] & H^0(f^* T_Z) \otimes \mathcal{O} \arrow[u] \arrow[r, "\mathrm{ev}"] & f^* T_Z \arrow[u, equal] \arrow[r] & 0 \\
& 0 \arrow[u] & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd}$$

The rows and columns are exact and the left column gives $T_{Y\times\mathcal{M}/Z}|_{Y\times\{[f]\}}$ as an extension of K by T_Y. □

Lemma 3.10. *The Lazarsfeld–Mukai bundle of any globally generated vector bundle on $\mathbb{P}^1$ is a direct sum of $\mathcal{O}(-1)$'s.*

Proof. Taking Lazarsfeld–Mukai bundles behaves well with respect to direct sum, so it remains to show the result for line bundles $\mathcal{O}(a)$ with $a \geq 0$. The Lazarsfeld–Mukai bundle M satisfies

$$0 \to M \to \mathcal{O} \otimes H^0(\mathcal{O}(a)) \to \mathcal{O}(a) \to 0.$$

It follows that M has rank a, degree $-a$ and no global sections, so that $M = \mathcal{O}(-1)^a$. The result follows. □

Proof of Proposition 3.1. We apply Lemma 3.6 to our situation, where $\mathcal{M}$ is an open subset of $\operatorname{Mor}(\mathbb{P}^1, Z)$ containing $[f]$ and $\mathcal{U} = \mathbb{P}^1 \times \mathcal{M}$. Then, applying Lemma 3.9 shows $T_{\mathbb{P}^1\times\mathcal{M}/Z}|_{\mathbb{P}^1\times\{[f]\}}$ an extension of $T_{\mathbb{P}^1}$ by the Lazersfeld–Mukai bundle of $f^* T_Z$. By Lemma 3.10, the Lazersfeld–Mukai bundle of $f^* T_Z$

is a sum of $\mathscr{O}(-1)$'s, so $T_{\mathbb{P}^1\times\mathcal{M}/Z}|_{\mathbb{P}^1\times\{[f]\}}$ is an extension of $\mathscr{O}(2)$ by a direct sum of $\mathscr{O}(-1)$'s implying $\mu^{\min}(T_{\mathbb{P}^1\times\mathcal{M}/Z}|_{\mathbb{P}^1\times\{[f]\}})\geq -1$.

Suppose f^*E splits as $\bigoplus_i \mathscr{O}(a_i)$ with $a_1\geq\cdots\geq a_r$ and $a_j\leq a_{j+1}-2$. Letting $S=\bigoplus_{i\leq j}\mathscr{O}(a_i)$, we find

$$a_j+(-1)>a_{j+1},$$
$$\mu^{\min}(S)+\mu^{\min}(T_{\mathbb{P}^1\times\mathcal{M}/Z}|_{\mathbb{P}^1\times\{[f]\}})>\mu^{\max}((f^*E)/S),$$

Therefore, we can apply Lemma 3.6 and conclude. □

Proof of Proposition 3.3. This follows from Lemma 3.6 with $Z=\mathbb{P}^n$, $\mathcal{M}=\mathbb{G}(k,n)$ and $\mathcal{U}$ the universal k-plane. The only thing to check is $\mu^{\min}(T_{\mathcal{U}/\mathbb{P}^n}|_\Lambda)=-\frac{1}{k}$. By Lemma 3.8, $T_{\mathcal{U}/\mathbb{P}^n}|_\Lambda$ lies in the sequence

$$0\to T_{\mathcal{U}/\mathbb{P}^n}|_\Lambda\to H^0(\mathscr{O}_\Lambda(1)^{n-k})\otimes\mathscr{O}_\Lambda\to\mathscr{O}_\Lambda(1)^{n-k}\to 0,$$

and so is isomorphic to $\Omega_\Lambda(1)^{n-k}$ by the Euler sequence. Since $\Omega_\Lambda(1)$ is semistable with slope $-\frac{1}{k}$ [Okonek et al. 1980, Theorem 1.3.2], the result follows. □

4. Plane curves

We now apply the results from the previous section to analyze the map to moduli Φ introduced in Section 2B in the case of plane curves. Throughout this section, C in $\mathbb{P}^2$ denotes a reduced plane curve. (In the nonreduced case, we simply pass to the reduction and apply the results of this section.) Our main results in this section are stated below.

Theorem 4.1. *Let C be a reduced plane curve of degree d. Then, the map*

$$\Phi:\mathrm{Mor}_e(\mathbb{P}^1,\mathbb{P}^2)\dashrightarrow\mathbb{P}(H^0(\mathscr{O}_{\mathbb{P}^1}(ed))),\quad [\iota]\mapsto[\iota^{-1}(C)],$$

has maximal rank if C has finite stabilizer under the action of PGL_3.

In fact, we can classify all cases in Theorem 4.1 where Φ does not have maximal rank.

Theorem 4.2. *We get a complete classification of cases when Φ in Theorem 4.1 does not have maximal rank*:

(1) $d\geq 5$: *C is a union of orbits under an action of $\mathbb{G}_m$ or $\mathbb{G}_a$ on $\mathbb{P}^2$.*

(2) $d=4$:

  (a) $e=1$ *and C is the union of four concurrent lines.*

  (b) $e\geq 2$ *and C is a union of orbits under an action of $\mathbb{G}_m$ or $\mathbb{G}_a$ on $\mathbb{P}^2$.*

(3) $d=3$: $e\geq 2$ *and C is union of concurrent lines.*

Before giving the proofs of these theorems, we need the following two propositions.

Proposition 4.3. *If C is a reduced plane curve and $T_{\mathbb{P}^2}(-\log C)$ admits a nontrivial homomorphism from $\mathscr{O}_{\mathbb{P}^2}(1)$, then C is a union of concurrent lines.*

Proof. First, a nontrivial map from $\mathscr{O}_{\mathbb{P}^2}(1) \to T_{\mathbb{P}^2}(-\log C)$ induces a nontrivial map $\mathscr{O}_{\mathbb{P}^2}(1) \to T_{\mathbb{P}^2}$. Consider the Euler sequence

$$0 \to \mathscr{O}_{\mathbb{P}^2} \to \mathscr{O}_{\mathbb{P}^2}(1)^3 \to T_{\mathbb{P}^2} \to 0.$$

Applying $\underline{\mathrm{Hom}}(\mathscr{O}_{\mathbb{P}^2}(1), \cdot\,)$ to the Euler sequence, we find

$$\underline{\mathrm{Hom}}(\mathscr{O}_{\mathbb{P}^2}(1), T_{\mathbb{P}^2}) \cong \underline{\mathrm{Hom}}(\mathscr{O}_{\mathbb{P}^2}(1), \mathscr{O}_{\mathbb{P}^2}(1)^3)$$

and that the composite map $\mathscr{O}_{\mathbb{P}^2}(1) \to T_{\mathbb{P}^2}(-\log C) \to T_{\mathbb{P}^2}$ lifts to a map $\mathscr{O}_{\mathbb{P}^2}(1) \to \mathscr{O}_{\mathbb{P}^2}(1)^3$.

After a change of coordinates, we can assume the map $\mathscr{O}_{\mathbb{P}^2}(1) \to \mathscr{O}_{\mathbb{P}^2}(1)^3$ is inclusion into the first factor. The map $\mathscr{O}_{\mathbb{P}^2}(1)^3 \to T_{\mathbb{P}^2}$ sends a tuple of linear forms (L_1, L_2, L_3) to $\bigl(L_1\frac{\partial}{\partial_X}, L_2\frac{\partial}{\partial_Y}, L_3\frac{\partial}{\partial_Z}\bigr)$, so we conclude that $L\frac{\partial}{\partial_X}$ is a section of $T_{\mathbb{P}^2}(-\log C)$ for all linear forms L.

By Remark 2.1, we see that away from the point $[1:0:0]$, the tangent vector $\frac{\partial}{\partial_X}$ is in the tangent space of C for every point of C. Restricting to the affine chart $\{Z \neq 0\}$ with coordinates (x, y) and dehomogenizing, this means C restricts to a union of lines parallel to the x-axis. Since these lines and the line at infinity are precisely the lines passing through $[1 : 0 : 0]$, we conclude C is a union of concurrent lines. □

Proposition 4.4. *If C is a reduced plane curve and $T_{\mathbb{P}^2}(-\log C)$ has a section, then C is equivalent to a union of orbits under one of the two actions by $\mathbb{G}_m$ and $\mathbb{G}_a$ as follows:*

$$\mathbb{G}_m \to \mathrm{GL}_3, \quad t \mapsto \begin{pmatrix} t^a & 0 & 0 \\ 0 & t^b & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad a, b \in \mathbb{N},$$

$$\mathbb{G}_a \to \mathrm{GL}_3, \quad t \mapsto \exp\left(t \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} 1 & t & \frac{1}{2}t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{pmatrix}.$$

Explicitly, there are two cases:

(1) *C is projectively equivalent to a union of curves of the form $X^pY^q = cZ^{p+q}$, $c \in \mathbb{C}^\times$, and possibly a subset of the three coordinate lines.*

(2) *C is projectively equivalent to a union of members of the family $\{XZ - Y^2 + cZ^2 \mid c \in \mathbb{C}\}$ of conics quadritangent to $\{XZ - Y^2 = 0\}$ at $[0 : 0 : 1]$, and possibly the line $\{Z = 0\}$.*

Proof. Let s be a section of $T_{\mathbb{P}^2}(-\log C)$. Then, s is also a section of $T_{\mathbb{P}^2}$ and can be written as $L_X\frac{\partial}{\partial_X} + L_Y\frac{\partial}{\partial_Y} + L_Z\frac{\partial}{\partial_Z}$ where L_X, L_Y, L_Z are homogenous linear forms in X, Y and Z.

Let C_0 be a component of C and let $p \in C$ be a smooth point of C_0. We lift $p \in \mathbb{P}^2$ to a point $\tilde{p} \in \mathbb{C}^3 \setminus \{0\}$. Then, C_0 contains the projection under $\mathbb{C}^3 \setminus \{0\} \to \mathbb{P}^2$ of the integral curve $\widetilde{C}$ through $\tilde{p}$ which is the solution to the matrix differential equation

$$\frac{d}{dt}\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} = A\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix}, \quad \begin{pmatrix} X(0) \\ Y(0) \\ Z(0) \end{pmatrix} = \tilde{p}. \tag{4-1}$$

Here A is the 3×3 matrix with complex entries such that

$$A\begin{pmatrix}X\\Y\\Z\end{pmatrix}=\begin{pmatrix}L_X(X,Y,Z)\\L_Y(X,Y,Z)\\L_Z(X,Y,Z)\end{pmatrix}.$$

If the projection of $\widetilde{C}$ to $\mathbb{P}^2$ is not a single point, then the image is dense in C_0. Therefore, C must be (the closure of) a finite union of projections of integral curves in $\mathbb{C}^3\setminus\{0\}$ and 1-dimensional components of the zero locus of s.

After a linear change of coordinates, we can assume that A is in Jordan block form. We keep this choice of coordinates from now on. We let $\tilde{p}=(c_1,c_2,c_3)\in\mathbb{C}^3$ denote a lift of a point on C_0 (to be determined separately in each case) and we let $P(X,Y,Z)$ be a homogenous polynomial defining C_0.

Case 1: A is diagonal. We will show that the first case of Proposition 4.4 happens, so we can assume that $c_1,c_2,c_3\neq 0$ or else $\tilde{p}$ is contained in a coordinate line. Let $\lambda_1,\lambda_2,\lambda_3$ be the eigenvalues of A. Then, the solution to (4-1) is $(X(t)\;\;Y(t)\;\;Z(t))^T=(e^{\lambda_1 t}c_1\;\;e^{\lambda_2 t}c_2\;\;e^{\lambda_3 t}c_3)^T$, where $(\,\cdot\,)^T$ denotes the transpose.

The defining equation $P(X,Y,Z)$ of C_0 is a homogenous polynomial of minimal degree satisfying

$$P(c_1e^{\lambda_1 t},c_2e^{\lambda_2 t},c_3e^{\lambda_3 t})=0. \tag{4-2}$$

We can choose a new grading on $\mathbb{C}[X,Y,Z]$ by the complex numbers $\mathbb{C}$ where the monomial $X^aY^bZ^c$ has the grade $a\lambda_1+b\lambda_2+c\lambda_3$. Let P_ω be the homogenous component of P with grade $\omega\in\mathbb{C}$. By linear independence of characters, the elements in $\{e^{\omega t}\mid\omega\in\mathbb{C}\}$ are linearly independent, and hence $P_\omega(c_1e^{\lambda_1 t},c_2e^{\lambda_2 t},c_3e^{\lambda_3 t})=0$. Therefore, P divides P_ω for all $\omega\in\mathbb{C}$ so P_ω can be nonzero for only one value of ω.

We cannot have $\lambda_1,\lambda_2,\lambda_3$ all equal or else s would be a multiple of $X\frac{\partial}{\partial_X}+Y\frac{\partial}{\partial_Y}+Z\frac{\partial}{\partial_Z}$ which induces the zero vector field on $\mathbb{P}^2$. The monomials $X^aY^bZ^c$ that can appear in P with nonzero coefficients must be the solution to the two linear equations

$$a+b+c=\deg(C_0), \tag{4-3}$$

$$\lambda_1a+\lambda_2b+\lambda_3c=\omega \tag{4-4}$$

for some fixed ω. The solution set to (4-3) is some 1-dimensional complex line ℓ in $\mathbb{C}^3$ and we are interested in the integer solutions $\ell\cap\mathbb{Z}^3$. If $\ell\cap\mathbb{Z}^3$ is empty or a single point, then P is a monomial, hence degree 1 by irreducibility. So the only remaining case is if $\ell\cap\mathbb{Z}^3$ is a 1-dimensional lattice, which can be written in the form $\{(a_0,b_0,c_0)+m(a_1,b_1,c_1)\mid m\in\mathbb{Z}\}$.

Thus, we know that the monomials $X^aY^bZ^c$ that can appear with nonnegative coefficients in P must be in $S=\{(a_0,b_0,c_0)+m(a_1,b_1,c_1)\mid m\in\mathbb{Z}\}\cap\mathbb{Z}^3_{\geq 0}$. If S contains exactly one element, then P is degree one by irreducibility. If S contains exactly two elements, then P is a binomial and must then be of the form $X^aY^b+kZ^{a+b}$ for some $k\neq 0$, because P is irreducible. Finally, one can check S cannot contain three or more elements assuming P is irreducible.

Case 2: A **has exactly two Jordan blocks** Let λ_1 be the eigenvalue of the 2×2 block and λ_2 be the eigenvalue of the 1×1 block. We will show that the first case of Proposition 4.4 happens. We can assume C_0 is not contained in a coordinate line, and therefore assume $\tilde{p}$ is such that $c_2, c_3 \neq 0$. Then, a solution to (4-1) is $(X(t)\ \ Y(t)\ \ Z(t))^T = (e^{\lambda_1 t}c_1 + c_2 t e^{\lambda_1 t}\ \ e^{\lambda_1 t}c_2\ \ e^{\lambda_2 t}c_3)^T$ for $\tilde{p} = (c_1, c_2, c_3)$.

This means $P(e^{\lambda_1 t}c_1 + c_2 t e^{\lambda_1 t}, e^{\lambda_1 t}c_2, e^{\lambda_2 t}c_3) = 0$. Dividing by $e^{\deg(P)\lambda_1 t}$ and letting $\lambda = \lambda_2 - \lambda_1$, we find

$$P(c_1 + c_2 t, c_2, e^{\lambda t}c_3) = 0.$$

Reparameterizing t by $t - \frac{c_1}{c_2}$, we can assume $c_1 = 0$. We claim now that the map $\mathbb{C}[X, Y, Z] \to \mathbb{C}[\![t]\!]$ sending $P(X, Y, Z)$ to $P(c_2 t, c_2, e^{\lambda t}c_3)$ is an injection because $c_2 t, c_2, e^{\lambda t}c_3$ are algebraically independent. The latter claim follows from the fact that the functions $\{t^m e^{\omega t} \mid m \in \mathbb{Z}_{\geq 0},\ \omega \in \mathbb{C}\}$ are linearly independent. Therefore, $P = 0$, i.e., C_0 must be contained in either the $\{Y = 0\}$ or $\{Z = 0\}$ coordinate lines, establishing this case.

Case 3: A **has exactly one Jordan block** Let λ be the unique eigenvalue of A. Subtracting a diagonal matrix from A is equivalent to subtracting the Euler vector field $X\frac{\partial}{\partial X} + Y\frac{\partial}{\partial Y} + Z\frac{\partial}{\partial Z}$ from the vector field s, so we can assume $\lambda = 0$. Then, a solution to (4-1) is $(X(t)\ \ Y(t)\ \ Z(t))^T = \big(c_1 + c_2 t + c_3\frac{1}{2}t^2\ \ c_2 + c_3 t\ \ c_3\big)^T$ for $\tilde{p} = (c_1, c_2, c_3)$. We will show that the second case of Proposition 4.4 happens, so we can assume that $c_3 \neq 0$ or else C_0 is contained in $\{Z = 0\}$.

We know that $P\big(c_1 + c_2 t + c_3\frac{1}{2}t^2, c_2 + c_3 t, c_3\big) = 0$. We change coordinates on t. Letting $t \mapsto t - \frac{c_2}{c_3}$ yields

$$P\Big(c_1 + \frac{1}{2}\frac{c_2^2}{c_3} + c_3\frac{1}{2}t^2, c_3 t, c_3\Big) = 0.$$

Dividing out by a power of c_3 and replacing c_1 with another constant c_1', we find

$$P\Big(c_1' + \frac{1}{2}t^2, t, 1\Big) = 0.$$

As t varies, the curve $\big(c_1' + \frac{1}{2}t^2, t, 1\big)$ parameterizes the conic $XZ - \frac{1}{2}Y^2 - c_1' Z^2$ in $\mathbb{P}^2$, settling this case. □

Proofs of Theorems 4.1 and 4.2. We will prove Theorem 4.2 which implies Theorem 4.1. Let $f : \mathbb{P}^1 \to \mathbb{P}^2$ be a general map of degree e. The log tangent sheaf $T_{\mathbb{P}^2}(-\log C)$ is a vector bundle since it is a reflexive sheaf on a surface. Pulling back $T_{\mathbb{P}^2}(-\log C)$ to $\mathbb{P}^1$ yields a rank-2 vector bundle E of degree $(3-d)e$. We split our analysis into cases.

Case: $d \geq 5$ **or** $d = 4$ **and** $e \geq 2$**.** If Φ is not of maximal rank, then $E \cong \mathscr{O}(a) \oplus \mathscr{O}(b)$ where $a \geq 0$. Since the total degree of E is at most -2, we get $a - b \geq 2$ and we can apply Proposition 3.1 to find a line subbundle of $T_{\mathbb{P}^2}(-\log C)$ of nonnegative degree. This means $T_{\mathbb{P}^2}(-\log C)$ has a section and we conclude by Proposition 4.4.

Case: $d = 4$ **and** $e = 1$**.** If Φ is not of maximal rank, then $h^0(E) \geq 2$. This means $E \cong \mathscr{O}(a) \oplus \mathscr{O}(b)$ where $a \geq 1$. In this case $a - b \geq 3$, so we can apply Proposition 3.1 to find a line subbundle $\mathscr{O}(a)$ of $T_{\mathbb{P}^2}(-\log C)$. Applying Proposition 4.3, we are done.

Case: $d=3$ **and** $e\geq 2$**.** In this case, $\deg(E)=0$ and a dimension count shows that Φ is not of maximal rank whenever $h^0(E)\geq 3$. Hence, we can apply Proposition 3.1 to find a line subbundle of $T_{\mathbb{P}^2}(-\log C)$ of positive degree, so we again conclude using Proposition 4.3.

Case: $d=3$ **and** $e=1$**.** We can find a line ℓ meeting C in three distinct points. This means Φ is automatically surjective, so it is of maximal rank.

Case: $d=2$**.** In this case, $\deg(E)=e$ and Φ is not of maximal rank if and only if $E\cong O(a)\oplus O(b)$ where $a\geq e+2$ and $b\leq -2$. Applying Proposition 3.1, we find a line subbundle of $T_{\mathbb{P}^2}(-\log C)$ of degree at least $\lceil\frac{e+2}{e}\rceil=2$. However, there are no nontrivial maps $\mathscr{O}(2)\to T_{\mathbb{P}^2}$, showing Φ must have maximal rank. □

5. Hyperplane sections

We let X be a smooth degree d hypersurface in $\mathbb{P}^n$. Using the notation from Section 2B, our objective is to prove that Φ has maximal rank when $k=n-1$ and $e=1$. Unlike the plane curve case, we are unable to obtain a complete classification statement like Theorem 4.2. However, we are able to prove that if d is larger than $n+1$, the hyperplane sections of X vary maximally in moduli. We prove Theorem 1.4 and some generalizations, captured below in Theorems 5.8 and 5.9.

Our results all rely on a stability result from Guenancia [2016]. The following version comes from Guenancia's Theorem A by observing that the canonical bundle of a degree d hypersurface in $\mathbb{P}^n$ is ample when $d\geq n+2$.

Theorem 5.1 [Guenancia 2016, Theorem A]. *If X is a smooth hypersurface of degree $d\geq n+2$, then $T_{\mathbb{P}^n}(-\log X)$ is semistable.*

Using Theorem 5.1, the basic strategy is to understand how large the degree d can be such that the restriction of $T_{\mathbb{P}^n}(-\log X)$ to the curve or k-plane can have a section. We use results from Section 3 to do this.

Theorem 5.2. *If X in $\mathbb{P}^n$ is a smooth hypersurface of degree d, then the space of degree e rational curve sections of X vary maximally in modulus when $d>\frac{n(n-1)}{2e}+n+1$.*

Proof. Consider the bundle $T_{\mathbb{P}^n}(-\log X)$. By Theorem 5.1, this bundle is semistable. For d larger than $n+1$, we see that a section of this bundle would give a destabilizing subsheaf, so we know that $T_{\mathbb{P}^n}(-\log X)$ has no sections.

Let $\mathcal{M}=\mathrm{Mor}_e(\mathbb{P}^1,\mathbb{P}^n)$ be the space of parameterized degree e rational curves in $\mathbb{P}^n$. Given a choice of F with $X=V(F)$, there is a natural map $\Phi:\mathcal{M}\to H^0(\mathscr{O}_{\mathbb{P}^1}(ed))$ sending a map $f:\mathbb{P}^1\to\mathbb{P}^n$ to the pullback $f^*F\in H^0(\mathbb{P}^1,\mathscr{O}_{\mathbb{P}^1}(de))$. We know by Lemma 2.2 that the tangent space to the fiber of Φ at a given map $f:\mathbb{P}^1\to\mathbb{P}^n$ is simply $H^0(f^*T_{\mathbb{P}^n}(-\log X))$. To show that Φ is generically finite, we need only show that $h^0(f^*T_{\mathbb{P}^n}(-\log X))=0$.

By Proposition 3.1, we see that $f^*T_{\mathbb{P}^n}(-\log X)$ is a direct sum of line bundles $\bigoplus\mathscr{O}(a_i)$ with consecutive a_i differing by at most 1. Thus, any such bundle on $\mathbb{P}^1$ that has a section will have degree larger than that of the bundle $\mathscr{O}\oplus\mathscr{O}(-1)\oplus\cdots\oplus\mathscr{O}(-n+1)$. From this it follows that any semistable bundle E on $\mathbb{P}^n$ such that f^*E has a section for a general map $f:\mathbb{P}^1\to\mathbb{P}^n$ will have degree at least $-\frac{n(n-1)}{2}$. Thus, if

$\deg f^*T_{\mathbb{P}^n}(-\log X) < -\frac{n(n-1)}{2}$ then $h^0(f^*T_{\mathbb{P}^n}(-\log X)) = 0$. Since $\deg f^*T_{\mathbb{P}^n}(-\log X) = e(n+1-d)$, the result follows. □

We now consider k-plane sections of smooth hypersurfaces. By Proposition 3.3, we need to understand torsion free sheaves on $\mathbb{P}^k$ whose Harder–Narasimhan filtration has subquotients whose slopes do not decrease too quickly, namely $\mu_1 > \mu_2 > \cdots > \mu_a$ with $\mu_i - \mu_{i+1} \leq \frac{1}{k}$ for all i. Understanding the possible slopes that may appear in the Harder–Narasimhan filtration is an interesting combinatorial problem, which we describe below.

Definition 5.3. Let a sequence $(d_1, r_1), (d_2, r_2), \ldots, (d_a, r_a)$ in $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{>0}$ be *k-admissible* if $d_1 \leq 0$ and $0 \leq \frac{d_{i+1}}{r_{i+1}} - \frac{d_i}{r_i} \leq \frac{1}{k}$ for each i. Let $A_{k,n}$ denote the set of k-admissible sequences with $\sum_i r_i = n$ (where a is arbitrary).

Definition 5.4. Define $B_k(n)$ to be $\max\bigl\{\sum_{i=1}^a d_i \mid (d_1, r_1), \ldots, (d_a, r_a) \text{ in } A_{k,n}\bigr\}$.

Lemma 5.5. *If E is a semistable sheaf on $\mathbb{P}^n$ of rank n such that its restriction to a general k-plane has a section, then* $\deg E \geq -B_k(n)$.

Proof. Let Λ be a general k-plane and $0 = E_0 \subset E_1 \subset \cdots \subset E_a = E|_\Lambda$ be the Harder–Narasimhan filtration of $E|_\Lambda$. Let $-d_i$ be the degree of E_i/E_{i-1} and r_i be the rank of E_i/E_{i-1}. Since $E|_\Lambda$ has a section, we see that $d_1 \leq 0$. Since E is semistable, by Proposition 3.3 it follows that the sequence $(-d_1, r_1), \ldots, (-d_a, r_a)$ will be k-admissible. The result follows. □

We can compute a bound for when k-plane sections of a degree d hypersurface in $\mathbb{P}^n$ will vary maximally in moduli in terms of $B_k(n)$.

Theorem 5.6. *Let X be a smooth, degree d hypersurface in $\mathbb{P}^n$ with $d > B_k(n) + n + 1$. Then,*

$$\Phi : \mathrm{Mor}_1(\mathbb{P}^k, \mathbb{P}^n) \dashrightarrow \mathbb{P}H^0(\mathbb{P}^k, \mathscr{O}_{\mathbb{P}^k}(d)), \quad \iota \mapsto [\iota^{-1}(X)],$$

is of maximal rank.

Proof. By Theorem 5.1, $T_{\mathbb{P}^n}(-\log X)$ will be semistable. By Proposition 3.3, $T_{\mathbb{P}^n}(-\log X)|_\Lambda$ will have Harder–Narasimhan filtration as described in the statement of the theorem. Given a hypersurface X together with a choice of defining equation f, we get a map $\phi : \mathrm{Mor}_1(\mathbb{P}^k, \mathbb{P}^n) \to H^0(\mathscr{O}_{\mathbb{P}^k}(d))$ sending a k-plane to the pull-back of f by the k-plane. We wish to show that ϕ is generically finite.

To get a contradiction, suppose ϕ has only positive-dimensional fibers. By Lemma 2.2, the tangent space to a fiber of ϕ at a general point Λ is $H^0(T_{\mathbb{P}^n}(-\log X)|_\Lambda)$, so we know that $T_{\mathbb{P}^n}(-\log X)|_\Lambda$ has a global section. Thus, by Theorem 5.1 and Lemma 5.5, the degree of $T_{\mathbb{P}^n}(-\log X)$ will be at least $-B_k(n)$. It follows that

$$n + 1 - d \geq -B_k(n).$$

This is impossible given the assumptions in the statement of the theorem. □

Then, Theorem 1.4 follows from the following result on $B_k(n)$.

Proposition 5.7. *If $k \geq \frac{2n}{3}$, then $B_k(n) = 1$.*

Proof. Let $(d_1, r_1), \dots, (d_a, r_a)$ be an admissible sequence of total degree $B_k(n)$. Without loss of generality, we may assume $d_1 = 0$, $d_i > 0$ for $i > 1$. Then it follows that $r_2 \geq k$, since $\frac{d_2}{r_2} \leq \frac{1}{k}$. Since $\frac{d_3}{r_3} \leq \frac{2}{k}$, we see that $r_3 \geq \frac{k}{2}$, provided that there are at least three terms in the sequence. However, in this case, $r_1 + r_2 + r_3 \geq 1 + k + \frac{k}{2} = 1 + \frac{3k}{2} > n$, which is impossible. Thus, $a \leq 2$.

Next, we observe that $d_2 \leq 1$, since if $d_2 \geq 2$, then $r_2 \geq d_2 k \geq 2k > n$, a contradiction. It follows that the sum of the d_i is at most 1, and since we know that 1 is achievable with the admissible sequence $(0, n-k)$, $(1, k)$, the result follows. □

We defer more detailed analysis of $B_k(n)$ to the Appendix. From the results in the Appendix and Theorem 5.6 we get the following results.

Theorem 5.8. *If $X \subset \mathbb{P}^n$ is a smooth hypersurface of degree d with $d > 4\big(\frac{n^2}{k^{3/2}} + k^{3/2}\big)$, then the map*

$$\Phi : \mathrm{Mor}_1(\mathbb{P}^k, \mathbb{P}^n) \dashrightarrow \mathbb{P}H^0(\mathbb{P}^k, \mathscr{O}_{\mathbb{P}^k}(d)), \quad \iota \mapsto [\iota^{-1}(X)],$$

is of maximal rank.

Proof. This follows from Proposition A.1, where it is shown $B_k(n) \leq 3\big(\frac{n^2}{k^{3/2}} + k^{3/2}\big)$. To finish, one checks that $\frac{n^2}{k^{3/2}} + k^{3/2} \geq 2n \geq n + 2$. This follows from the AM-GM equality and the fact $n \geq 2$. □

In Theorem 5.8, we prioritized giving a clean statement and proof over giving an optimal constant. Still, one can wonder what the optimal constant by computing $B_k(n)$ for small k and all n. In this case, Corollary A.5 gives the following result:

Theorem 5.9. *If $k \leq 5$, then there is a linear function $\ell(n)$ and an integer C_k such that $|B_k(n) - \frac{n^2}{C_k}| \leq \ell(n)$. Here, $C_2 = 3$, $C_3 = 7$, $C_4 = 11$, $C_5 = 19$. In particular, the map*

$$\Phi : \mathrm{Mor}_1(\mathbb{P}^k, \mathbb{P}^n) \dashrightarrow \mathbb{P}H^0(\mathbb{P}^k, \mathscr{O}_{\mathbb{P}^k}(d)), \quad \iota \mapsto [\iota^{-1}(X)],$$

is of maximal rank if $X \subset \mathbb{P}^n$ is smooth and has degree $d \geq C_k n^2 + \ell(n) + n + 2$.

We expect Theorem 5.9 to hold for all values of k, but we can only check a finite number of cases with a computer. Roughly up to $k = 100$ is what is reasonable with our methods.

Given Theorem 5.9, one can ask how fast C_k grows with k. We trivially know $C_k = O(k^2)$ by relaxing the condition that the d_i are integers in the definition of an admissible sequence to compute $B_k(n)$ (in which case we let all the r_i be equal to 1). We also get $C_k = \Omega(k^{3/2})$ from Proposition A.1. From experimental evidence, we think that the actual answer is strictly between $k^{3/2}$ and k^2 but closer to $k^{3/2}$.

Appendix: Bounds and computations for $B_k(n)$

We will bound $B_k(n)$ for all k, n in Proposition A.1. We also compute $B_k(n)$ for k small and arbitrary n, and give some conjectures about $B_k(n)$ in general.

To give an idea of how the function $B_k(n)$ behaves, we note the results in the Appendix can show $B_5(39) = 39$, corresponding to the admissible sequence

$$(0, 1), (1, 5), (1, 3), (1, 2), (2, 3), (4, 5), (1, 1), (6, 5), (4, 3), (3, 2), (5, 3), (9, 5), (2, 1).$$

There are a couple of features of this admissible sequence we believe hold in general that we will only prove in special cases. First, this admissible sequence can be generated greedily, where we use greed to maximize the ratio $\frac{d}{r}$ of the last piece of the sequence. Second, the admissible sequence is essentially periodic in that the $(1, 1), (6, 5), (4, 3), (3, 2), (5, 3), (9, 5)$ is obtained from $(0, 1), (1, 5), (1, 3), (1, 2), (2, 3), (4, 5)$ by replacing each (d, r) with $(d + r, r)$. We give a finite criterion that can be applied to show both the greedy property and the periodicity in Lemma A.4.

We expect there are many other interesting patterns that can be found. For example, the segment $(0, 1), (1, 5), (1, 3), (1, 2), (2, 3), (4, 5), (1, 1)$ of the admissible sequence above is preserved under reversing the order and replacing each (d, r) with $(r - d, r)$. This pattern continues to hold for larger k and suggests that these optimal admissible sequences can also be generated greedily backwards as well as forwards.

Proposition A.1. *We have* $B_k(n) \leq 3\big(\frac{n^2}{k^{3/2}} + k^{3/2}\big)$

Proof. Let $(d_1, r_1), \dots, (d_a, r_a)$ be an admissible sequence with $\sum_i d_i = B_k(n)$. Let $\mu_i = \frac{d_i}{r_i}$. Let $n(j)$ be the sum of the r_i such that $\mu_i \in [j - 1, j)$.

Since the μ_i contributing to $n(j)$ are all less than j, we observe that

$$B_k(n) \leq \sum_{j=1}^{\infty} j n(j).$$

Thus, understanding the $n(j)$ allows us to bound $B_k(n)$. The sum of all of the $n(j)$ is n. Let J be the last nonzero $n(j)$, so $B_k(n) \leq \sum_{j=1}^{J} j n(j)$.

Let $n_k^{\min}$ be a positive number that is at most $n(j)$ for any $j < J$. Then we obtain an upper bound for $B_k(n)$

$$\begin{aligned} B_k(n) &\leq \sum_{i=1}^{J} j n(j) \leq n_k^{\min} + 2n_k^{\min} + \dots + \left\lceil \frac{n}{n_k^{\min}} \right\rceil n_k^{\min} \\ &= n_k^{\min} \frac{\lceil \frac{n}{n_k^{\min}} \rceil \big(1 + \lceil \frac{n}{n_k^{\min}} \rceil\big)}{2} \\ &\leq n_k^{\min} \frac{\big(\frac{n}{n_k^{\min}} + 1\big)\big(\frac{n}{n_k^{\min}} + 2\big)}{2}. \end{aligned}$$

Thus, it remains to give a bound for $n_k^{\min}$. Fix $j < J$ and let $(d_{i_j+1}, r_{i_j+1}), \dots, (d_{i_j+c(j)}, r_{i_j+c(j)})$ be the part of the admissible sequence with slopes $\frac{d_{i_j+1}}{r_{i_j+1}}, \dots, \frac{d_{i_j+c(j)}}{r_{i_j+c(j)}}$ in $[j - 1, j)$. By definition,

$$r_{i_j+1} + \dots + r_{i_j+c(j)} = n(j).$$

First, we show $c(j) \geq k$: Note that $\frac{d_{i_j+1}}{r_{i_j+1}} < (j - 1) + \frac{1}{k}$. If $j = 1$, then this is true because $\frac{d_{i_j+1}}{r_{i_j+1}} \leq 0$ by definition. If $j > 1$, then this is true because $\frac{d_{i_j+1}}{r_{i_j+1}} \leq \frac{d_{i_j}}{r_{i_j}} + \frac{1}{k} < (j - 1) + \frac{1}{k}$.

Since

$$\frac{d_{i_j+1}}{r_{i_j+1}} < (j - 1) + \frac{1}{k},$$

we then see that

$$\begin{aligned}\frac{d_{i_j+2}}{r_{i_j+2}} &\leq \frac{d_{i_j+1}}{r_{i_j+1}} + \frac{1}{k} < (j-1) + \frac{2}{k}\\ \vdots \quad & \quad\ \vdots\\ \frac{d_{i_j+k}}{r_{i_j+k}} &\leq \frac{d_{i_j+k-1}}{r_{i_j+k-1}} + \frac{1}{k} < j,\end{aligned}$$

so $c(j) \geq k$.

Now, we want to bound $r_{i_j+1} + \cdots + r_{i_j+c(j)} = n(j)$. In the multiset $\{r_{i_j+1}, \ldots, r_{i_j+c(j)}\}$, we know that there is at most element that is equal to 1, fewer than two elements that are equal to 2, fewer than three elements that are equal to 3 and so on. Therefore, if m is the largest integer such that $1+(1+\cdots+m-1) = 1+\frac{m(m-1)}{2}$ is at most k, then $\frac{(m-1)^2}{2} < \frac{m(m_1)}{2} + 1 \leq k$, so $m \leq \sqrt{2k}+1$. Thus,

$$\begin{aligned}n(j) = r_{i_j+1} + \cdots + r_{i_j+c(j)} &\geq 1 + (2\cdot 1 + 3\cdot 2 + \cdots m\cdot(m-1))\\ &= 1 + 2\left(\binom{2}{2} + \binom{3}{2} + \cdots + \binom{m}{2}\right)\\ &= 1 + \frac{(m+1)m(m-1)}{3}\\ &\geq \frac{(\sqrt{2k}+2)(\sqrt{2k}+1)\sqrt{2k}}{3} > \frac{2\sqrt{2}}{3}k^{3/2}.\end{aligned}$$

Thus, choosing $n_k^{\min}$ to be $\frac{2\sqrt{2}}{3}k^{3/2}$ suffices. Plugging into our earlier bound, we get an upper bound for $B_k(n)$ as

$$n_k^{\min}\frac{\left(\frac{n}{n_k^{\min}}+1\right)\left(\frac{n}{n_k^{\min}}+2\right)}{2} = \frac{n^2}{2n_k^{\min}} + \frac{3n}{2} + n_k^{\min} = \frac{n^2}{k^{3/2}}\frac{9}{4\sqrt{2}} + \frac{3n}{2} + \frac{2\sqrt{2}}{3}k^{3/2},$$

which is at most $2\left(\frac{n^2}{k^{3/2}} + n + k^{3/2}\right)$. Applying the AM-GM inequality yields

$$\frac{n^2}{k^{3/2}} + k^{3/2} \geq 2n,$$

yielding the claimed bound. □

We now move on to computing exact values of $B_k(n)$ for small k. Our strategy is a recursive algorithm that requires some conditions to be met, and we suspect that these conditions are always met. In the course of our proof, we will use the three quantities $\mu^{\max}(n)$, $B_k^{\mathrm{upper}}(n)$ and $B_k^{\mathrm{lower}}(n)$. We define $\mu^{\max}(n)$ by

$$\mu^{\max}(n) := \max\left\{\frac{r_a}{d_a}\;\middle|\;(d_1,r_1),\ldots,(d_a,r_a) \text{ in } A_n\right\}.$$

Lemma A.2. *We can compute $\mu^{\max}(n)$ inductively by $\mu^{\max}(1) = 0$ and*

$$\mu^{\max}(n) = \max\left\{\frac{\lfloor(\mu^{\max}(i)+\frac{1}{k})(n-i)\rfloor}{n-i}\;\middle|\;0 < i < n\right\}.$$

Proof. Let

$$\mu_n^{\max'} = \max\left\{ \frac{\lfloor (\mu^{\max}(i) + \frac{1}{k})(n-i) \rfloor}{n-i} \;\middle|\; 0 < i < n \right\}.$$

We use induction. The base case $n = 1$ is vacuous, so assume $n > 1$. First, we show that $\mu_n^{\max\,\prime} \leq \mu_n^{\max}$. Given any $0 < i < n$, $\frac{\lfloor (\mu_i^{\max\,\prime} + \frac{1}{k})(n-i) \rfloor}{n-i}$ is a slope achieved by taking an admissible sequence $(d_1, r_1), \dots, (d_a, r_a)$ in A_i and appending $\big(\lfloor (\mu_i^{\max\,\prime} + \frac{1}{k})(n-i) \rfloor, n-i\big)$. So by definition $\mu_n^{\max\,\prime} \leq \mu_n^{\max}$.

Now we show $\mu_n^{\max\,\prime} \geq \mu_n^{\max}$. Let $(d_1, r_1), \dots, (d_a, r_a)$ be an admissible sequence in A_n achieving $\frac{d_a}{r_a} = \mu_n^{\max}$. If $a = 1$, then $d_a = 0$ so $\mu_n^{\max} = 0$ while $\mu_n^{\max\,\prime}$ is by definition nonnegative. Otherwise, let $i = r_1 + \dots + r_{a-1}$, so $\frac{d_{a-1}}{r_{a-1}} \leq \mu_i^{\max} = \mu_i^{\max\,\prime}$ by definition and the assumption hypothesis. Then, $r_a = n - i$ and

$$\frac{d_a}{r_a} \leq \mu_i^{\max\,\prime} + \frac{1}{k},$$

so $d_a \leq \lfloor (\mu_i^{\max\,\prime} + \frac{1}{k})(n-i) \rfloor$. Then, by definition $\mu_n^{\max\,\prime} \geq \mu_n^{\max}$. Therefore, we are done and $\mu_n^{\max\,\prime} = \mu_n^{\max}$ for all n. □

We define $B_k^{\text{upper}}(n)$ recursively by $B_k^{\text{upper}}(1) = 0$ and

$$B_k^{\text{upper}}(n) = \max\left\{ B_k^{\text{upper}}(i) + \left\lfloor \left(\mu_i^{\max} + \frac{1}{k}\right)(n-i) \right\rfloor \;\middle|\; 0 < i < n \right\}.$$

For B_k^{lower}, we let $B_k^{\text{lower}}(1) = 0$ and let $i(n)$ be the smallest i that maximizes $\frac{\lfloor (\mu^{\max}(i) + \frac{1}{k})(n-i) \rfloor}{n-i}$. Then define B_k^{lower} inductively by

$$B_k^{\text{lower}}(n) = B_k^{\text{lower}}(i(n)) + \left\lfloor \left(\mu^{\max}(i(n)) + \frac{1}{k}\right)(n - i(n)) \right\rfloor.$$

We now show that $B_k(n)$ is bounded by $B_k^{\text{upper}}(n)$ and $B_k^{\text{lower}}(n)$.

Lemma A.3. *We have $B_k^{\text{lower}}(n) \leq B_k(n) \leq B_k^{\text{upper}}(n)$.*

Proof. First we show $B_k^{\text{lower}}(n) \leq B_k(n)$ by induction. To do this, we show by induction that $B_k^{\text{lower}}(n)$ is always achieved by an admissible sequence $(d_1, r_1), \dots, (d_a, r_a)$ with $\frac{d_a}{r_a} = \mu^{\max}(n)$ and $d_1 + \dots + d_a = B_k^{\text{lower}}(n)$. The base case $n = 1$ vacuous, so we assume $n > 1$. Let i be the minimal index maximizing

$$\frac{\lfloor (\mu_i^{\max} + \frac{1}{k})(n-i) \rfloor}{n-i}.$$

By the induction assumption, there is an admissible sequence $(d_1, r_1), \dots, (d_{a-1}, r_{a-1})$ achieving $\frac{d_{a-1}}{r_{a-1}} = \mu_i^{\max}$ and $d_1 + \dots + d_{a-1} = B_k^{\text{lower}}(i)$. By appending $\big(\lfloor (\mu_i^{\max} + \frac{1}{k})(n-i) \rfloor, n-i\big)$ to the sequence we get an admissible sequence $(d_1, r_1), \dots, (d_a, r_a)$ with $\frac{d_a}{r_a} = \mu_n^{\max}$ and $d_1 + \dots + d_a = B_k^{\text{lower}}(n)$.

Finally, we show $B_k^{\text{upper}}(n) \geq B_k(n)$ by induction. The base case $n = 1$ is vacuous, so assume $n > 1$. Let $(d_1, r_1), \dots, (d_a, r_a)$ be an admissible sequence in A_n achieving $d_1 + \dots + d_a = B_k(n)$. If $a = 1$, then $B_k(n) = 0$ and $B_k^{\text{upper}}(n)$ is always nonnegative by definition. If $a > 1$, then let $i = r_1 + \dots + r_{a-1}$ so

$(d_1, r_1), \dots, (d_{a-1}, r_{a-1})$ is an admissible sequence in A_i. By the inductive hypothesis, $d_1 + \cdots + d_{a-1} \le B_k^{\text{upper}}(i)$. We have $r_a = n - i$ and the maximum d_a can be is $\lfloor (\mu^{\max}(i) + \frac{1}{k})(n-i) \rfloor$. Therefore,

$$d_1 + \cdots + d_a \le B_k^{\text{upper}}(i) + \left\lfloor \left(\mu^{\max}(i) + \frac{1}{k} \right)(n-i) \right\rfloor \le B_k^{\text{upper}}(n),$$

finishing the proof. □

From experimental evidence, we suspect $B_k^{\text{lower}}(n)$ and $B_k^{\text{upper}}(n)$ always coincide, which would give a recursive algorithm for $B_k(n)$. However, to give results for small values of k and all n, we want to have a finite criterion that can be verified by a computer. We believe admissible sequences achieving $B_k(n)$ will always following a periodic structure in n with k fixed reflected in Lemma A.4 below.

Lemma A.4. *Suppose $\mu^{\max}(i_0) = \frac{k-1}{k}$ for some i_0. Then $\mu^{\max}(n) = \mu^{\max}(n - i_0) + 1$ all $n \ge i_0$. If in addition $B_k^{\text{upper}}(i) = B_k^{\text{lower}}(i)$ for each $i \le 3i_0$, then $B_k^{\text{lower}}(n) = B_k(n) = B_k^{\text{upper}}(n)$ for all n.*

Using Lemma A.4, one can show $B_k(n + i_0) = B_k(n) + n + B_k(i_0)$. Iterating this shows

$$B_k(n + N i_0) = B_k(n) + nN + N B_k(i_0) + \frac{N(N-1)i_0}{2}$$

for $1 < n \le i_0$ and $N \ge 0$. In particular, $B_k(n) = \Theta\big(\frac{1}{i_0} N^2\big)$.

Proof. First note that if $\mu^{\max}(i) = m + \frac{k-1}{k}$ for m an integer, then

$$\mu^{\max}(i+1) = \max_j \left\{ \frac{\lfloor (\mu^{\max}(j) + \frac{1}{k})(i+1-j) \rfloor}{i+1-j} \;\middle|\; 0 < j \le i \right\} \tag{A-1}$$

$$= \left\lfloor m + \frac{k-1}{k} + \frac{1}{k} \right\rfloor = m + 1. \tag{A-2}$$

In particular, there is a unique i_0 for which $\mu^{\max}(i_0) = \frac{k-1}{k}$ and $\mu^{\max}(i_0 + 1) = 1 = 1 + \mu^{\max}(1)$.

We will now show $\mu^{\max}(n) = \mu^{\max}(n - i_0) + 1$ for all $n \ge i_0$ using induction on n. For the case $n = i_0 + 1$, $\mu^{\max}(i_0 + 1) = 1$ from above.

Now suppose $n > i_0 + 1$. We first note that

$$\mu^{\max}\left(i_0 \left\lfloor \frac{n-1}{i_0} \right\rfloor \right) = \left(\left\lfloor \frac{n-1}{i_0} \right\rfloor - 1 \right) + \frac{k-1}{k}$$

is $\frac{k-1}{k}$ by induction.

Now, we claim that $\mu^{\max}(i) < \mu^{\max}\big(i_0 \lfloor \frac{n-1}{i_0} \rfloor\big)$ for all $i < i_0 \lfloor \frac{n-1}{i_0} \rfloor$. Since $\mu^{\max}(j)$ is weakly increasing in j, the point is to prove they are not equal. If $\mu^{\max}(i)$ was equal to $\mu^{\max}\big(i_0 \lfloor \frac{n-1}{i_0} \rfloor\big)$, then $\mu^{\max}(i+1) = \mu^{\max}(i)$, which contradicts (A-1).

By Lemma A.2, $\mu^{\max}(n)$ will be determined by the i between 0 and n such that $\frac{\lfloor (\mu^{\max}(i) + \frac{1}{k})(n-i) \rfloor}{n-i}$ is maximized. Let $i(n)$ be this i. We claim $i(n) \ge i_0 \lfloor \frac{n-1}{i_0} \rfloor$. To get a contradiction, suppose that $i(n) < i_0 \lfloor \frac{n-1}{i_0} \rfloor$.

Since we have shown above that $\mu^{\max}(i(n)) < \mu^{\max}\big(i_0\lfloor\frac{n-1}{i_0}\rfloor\big)$,

$$\mu^{\max}(i(n)) + \frac{1}{k} < \mu^{\max}\left(i_0\left\lfloor\frac{n-1}{i_0}\right\rfloor\right) + \frac{1}{k} = \left\lfloor\frac{n-1}{i_0}\right\rfloor,$$

contradicting $i(n) < i_0\lfloor\frac{n-1}{i_0}\rfloor$.

Since $i(n) \geq i_0\lfloor\frac{n}{i_0}\rfloor$,

$$\begin{aligned}\mu^{\max}(n) &= \frac{\big\lfloor\big(\mu^{\max}(i(n)) + \frac{1}{k}\big)(n - i(n))\big\rfloor}{n - i(n)}\\ &= \frac{\big\lfloor\big(\mu^{\max}(i(n) - i_0) + 1 + \frac{1}{k}\big)(n - i(n))\big\rfloor}{n - i(n)}\\ &= \frac{\big\lfloor\big(\mu^{\max}(i(n) - i_0) + \frac{1}{k}\big)((n - i_0) - (i(n) - i_0))\big\rfloor}{(n - i_0) - (i(n) - i_0)} + 1\\ &= \max_j\left\{\frac{\big\lfloor\big(\mu^{\max}(j) + \frac{1}{k}\big)(n - i_0 - j)\big\rfloor}{n - i_0 - j} \,\middle|\, 0 < j \leq n - i_0\right\} + 1\\ &= \mu^{\max}(n - i_0) + 1,\end{aligned}$$

where the second and fourth line are by induction and the fifth line is by definition. This concludes our induction for $\mu^{\max}$. From our proof, we also see that

$$i(n) - i_0 = i(n - i_0). \tag{A-3}$$

Next, we want to show the statement regarding $B_k(n)$. It suffices to show that $B_k^{\text{lower}}(n) = B_k^{\text{upper}}(n)$ for all n. We will show this by induction and can assume $n > 3i_0$ and $B_k^{\text{lower}}(i) = B_k^{\text{upper}}(i)$ for all $0 < i < n$. As before, let $i(n)$ be the minimum i that maximizes $\frac{\lfloor(\mu^{\max}(i)+\frac{1}{k})(n-i)\rfloor}{n-i}$. By definition, we want to show

$$B_k(i(n)) + \left\lfloor\left(\mu^{\max}(i(n)) + \frac{1}{k}\right)(n - i(n))\right\rfloor = \max\left\{B_k(i) + \left\lfloor\left(\mu^{\max}(i) + \frac{1}{k}\right)(n - i)\right\rfloor \,\middle|\, 0 < i < n\right\}. \tag{A-4}$$

The inequality $\leq$ is clear as the left side is one of the terms on the right side. Let i' be an index maximizing the right side. We want to show that $i' > i_0$. If $i' \leq i_0$, then $B_k^{\text{upper}}(i') + \big\lfloor\big(\mu^{\max}(i') + \frac{1}{k}\big)(n - i')\big\rfloor$ is less than n as $\mu_{i'}^{\max} + \frac{1}{k} \leq 1$ and $B^{\text{upper}}(i') \leq \mu^{\max}(i')i' < i'$. We also crudely bound $B_k^{\text{lower}}(n)$ from below.

To do so, we first bound $B_k^{\text{lower}}(3i_0)$ by $3i_0$. From the statement of Lemma A.4 regarding $\mu^{\max}$, we know $\mu^{\max}(j) \geq m$ for all $i > m \cdot i_0$. Then,

$$B_k^{\text{lower}}(n) \geq 0 \cdot i_0 + 1 \cdot i_0 + 2 \cdot i_0 + 3(n - 3i_0).$$

Since $n > 3i_0$,

$$3i_0 + 3(n - 3i_0) = 2(n - 3i_0) + n > n,$$

yielding a contradiction.

Since $i' > i_0$, the right side of (A-4) is

$$\begin{aligned}
&\max\Big\{B_k(i)+\Big\lfloor\Big(\mu^{\max}(i)+\frac{1}{k}\Big)(n-i)\Big\rfloor \;\Big|\; 0<i<n\Big\}\\
&\qquad=\max\Big\{B_k(i)+\Big\lfloor\Big(\mu^{\max}(i)+\frac{1}{k}\Big)(n-i)\Big\rfloor \;\Big|\; i_0<i<n\Big\}\\
&\qquad=\max\Big\{B_k(i+i_0)+\Big\lfloor\Big(\mu^{\max}(i+i_0)+\frac{1}{k}\Big)(n-i-i_0)\Big\rfloor \;\Big|\; 0<i<n-i_0\Big\}\\
&\qquad=\max\Big\{B_k(i)+i+B_k(i_0)+\Big\lfloor\Big(\mu^{\max}(i)+1+\frac{1}{k}\Big)(n-i-i_0)\Big\rfloor \;\Big|\; 0<i<n-i_0\Big\}\\
&\qquad=\max\Big\{B_k(i)+\Big\lfloor\Big(\mu^{\max}(i)+\frac{1}{k}\Big)(n-i-i_0)\Big\rfloor \;\Big|\; 0<i<n-i_0\Big\}+n-i_0+B_k(i_0).
\end{aligned}$$

But

$$\max\Big\{B_k(i)+\Big\lfloor\Big(\mu^{\max}(i)+\frac{1}{k}\Big)(n-i-i_0)\Big\rfloor \;\Big|\; 0<i<n-i_0\Big\}=B(n-i_0)$$

by induction. Looking at the left side of (A-4), we get

$$\begin{aligned}
&B_k(i(n))+\Big\lfloor\Big(\mu^{\max}(i(n))+\frac{1}{k}\Big)(n-i(n))\Big\rfloor\\
&\qquad=B_k(i(n)-i_0)+(i(n)-i_0)+B_k(i_0)+\Big\lfloor\Big(\mu^{\max}(i(n)-i_0+i_0)+\frac{1}{k}\Big)(n-i(n))\Big\rfloor\\
&\qquad=B_k(i(n)-i_0)+(i(n)-i_0)+B_k(i_0)+\Big\lfloor\Big(\mu^{\max}(i(n)-i_0)+1+\frac{1}{k}\Big)(n-i(n))\Big\rfloor\\
&\qquad=B_k(i(n)-i_0)+\Big\lfloor\Big(\mu^{\max}(i(n)-i_0)+\frac{1}{k}\Big)(n-i_0-(i(n)-i_0)\Big\rfloor+n-i_0+B_k(i_0)\\
&\qquad=B_k(n-i_0)+B_k(i_0)+n-i_0,
\end{aligned}$$

where the last line is by (A-3). Therefore, both sides of (A-4) are equal, which is what we wanted. □

We can verify the conditions of Lemma A.4 using a Python program for small k. For example, the answer for $k = 2, 3, 4, 5$ are given below.

Corollary A.5. *We have the following closed-form expressions for $B_k(n)$ for $k = 2, 3, 4, 5$. For $k = 2$ and $n \geq 0$,*

$$B_2(3n+1)=\frac{3n^2+n}{2},\quad B_2(3n+2)=\frac{3n^2+3n}{2},\quad B_2(3n+3)=\frac{3n^2+5n+2}{2}.$$

For $k = 3$,

$$\begin{aligned}
&B_3(7n+1)=\frac{7n^2+n}{2}, && B_3(7n+2)=\frac{7n^2+3n}{2}, && B_3(7n+3)=\frac{7n^2+5n}{2},\\
&B_3(7n+4)=\frac{7n^2+7n+2}{2}, && B_3(7n+5)=\frac{7n^2+9n+2}{2}, && B_3(7n+6)=\frac{7n^2+11n+4}{2},\\
&B_3(7n+7)=\frac{7n^2+13n+6}{2}.
\end{aligned}$$

For $k = 4$,

$$B_4(11n+1) = \frac{11n^2+n}{2}, \qquad B_4(11n+2) = \frac{11n^2+3n}{2}, \qquad B_4(11n+3) = \frac{11n^2+5n}{2},$$

$$B_4(11n+4) = \frac{11n^2+7n}{2}, \qquad B_4(11n+5) = \frac{11n^2+9n+2}{2}, \qquad B_4(11n+6) = \frac{11n^2+11n+2}{2},$$

$$B_4(11n+7) = \frac{11n^2+13n+4}{2}, \qquad B_4(11n+8) = \frac{11n^2+15n+4}{2}, \qquad B_4(11n+9) = \frac{11n^2+17n+6}{2},$$

$$B_4(11n+10) = \frac{11n^2+19n+8}{2}, \qquad B_4(11n+11) = \frac{11n^2+21n+10}{2}.$$

For $k = 5$,

$$B_5(19n+1) = \frac{19n^2+n}{2}, \qquad B_5(19n+2) = \frac{19n^2+3n}{2}, \qquad B_5(19n+3) = \frac{19n^2+5n}{2},$$

$$B_5(19n+4) = \frac{19n^2+7n}{2}, \qquad B_5(19n+5) = \frac{19n^2+9n}{2}, \qquad B_5(19n+6) = \frac{19n^2+11n+2}{2},$$

$$B_5(19n+7) = \frac{19n^2+13n+2}{2}, \qquad B_5(19n+8) = \frac{19n^2+15n+2}{2}, \qquad B_5(19n+9) = \frac{19n^2+17n+4}{2},$$

$$B_5(19n+10) = \frac{19n^2+19n+4}{2}, \qquad B_5(19n+11) = \frac{19n^2+21n+6}{2}, \qquad B_5(19n+12) = \frac{19n^2+23n+6}{2},$$

$$B_5(19n+13) = \frac{19n^2+25n+8}{2}, \qquad B_5(19n+14) = \frac{19n^2+27n+10}{2}, \qquad B_5(19n+15) = \frac{19n^2+29n+10}{2},$$

$$B_5(19n+16) = \frac{19n^2+31n+12}{2}, \qquad B_5(19n+17) = \frac{19n^2+33n+14}{2}, \qquad B_5(19n+18) = \frac{19n^2+35n+16}{2},$$

$$B_5(19n+19) = \frac{19n^2+37n+18}{2}.$$

Acknowledgements

The authors would like to thank Paolo Aluffi, Izzet Coskun, Carel Faber and Joe Harris for helpful conversations, and the anonymous referee for helpful comments.

References

[Aluffi and Faber 1993] P. Aluffi and C. Faber, "Linear orbits of smooth plane curves", *J. Algebraic Geom.* **2**:1 (1993), 155–184. MR Zbl

[Beauville 1990] A. Beauville, "Sur les hypersurfaces dont les sections hyperplanes sont à module constant", pp. 121–133 in *The Grothendieck Festschrift*, *I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, MA, 1990. MR Zbl

[Cadman and Laza 2008] C. Cadman and R. Laza, "Counting the hyperplane sections with fixed invariants of a plane quintic: three approaches to a classical enumerative problem", *Adv. Geom.* **8**:4 (2008), 531–549. MR Zbl

[Guenancia 2016] H. Guenancia, "Semistability of the tangent sheaf of singular varieties", *Algebr. Geom.* **3**:5 (2016), 508–542. MR Zbl

[Harris et al. 1998] J. Harris, B. Mazur, and R. Pandharipande, "Hypersurfaces of low degree", *Duke Math. J.* **95**:1 (1998), 125–160. MR Zbl

[Lee et al. 2020] M. Lee, A. Patel, H. Spink, and D. Tseng, "Orbits in $(\mathbb{P}^r)^n$ and equivariant quantum cohomology", *Adv. Math.* **362** (2020), art. id. 106951. MR Zbl

[Lee et al. 2023] M. Lee, A. Patel, and D. Tseng, "Equivariant degenerations of plane curve orbits", *Trans. Amer. Math. Soc.* **376**:10 (2023), 6799–6843. MR Zbl

[Liao 2013] X. Liao, *Chern classes of sheaves of logarithmic vector fields for free divisors*, Ph.D. thesis, Florida State University, 2013, available at https://www.proquest.com/docview/1468444341.

[Mckernan 1991] J. Mckernan, *On the hyperplane sections of a variety in projective space*, Ph.D. thesis, Harvard University, 1991, available at https://www.proquest.com/docview/303940435.

[Mehta and Ramanathan 1982] V. B. Mehta and A. Ramanathan, "Semistable sheaves on projective varieties and their restriction to curves", *Math. Ann.* **258**:3 (1982), 213–224. MR Zbl

[Okonek et al. 1980] C. Okonek, M. Schneider, and H. Spindler, *Vector bundles on complex projective spaces*, Progr. Math. **3**, Birkhäuser, Boston, MA, 1980. MR Zbl

[van Opstall and Veliche 2007] M. A. van Opstall and R. Veliche, "Variation of hyperplane sections", pp. 255–260 in *Algebra, geometry and their interactions*, edited by A. Corso et al., Contemp. Math. **448**, Amer. Math. Soc., Providence, RI, 2007. MR Zbl

[Saito 1980] K. Saito, "Theory of logarithmic differential forms and logarithmic vector fields", *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **27**:2 (1980), 265–291. MR Zbl

[Shatz 1977] S. S. Shatz, "The decomposition and specialization of algebraic families of vector bundles", *Compos. Math.* **35**:2 (1977), 163–187. MR Zbl

[Starr 2006] J. M. Starr, "Fano varieties and linear sections of hypersurfaces", preprint, 2006. arXiv math/0607133

Communicated by Gavril Farkas
Received 2020-09-20 Revised 2023-05-31 Accepted 2023-10-12

anand.patel@okstate.edu
Department of Mathematics, Oklahoma State University, Stillwater, OK, United States

eriedl@nd.edu
Department of Mathematics, University of Notre Dame, Notre Dame, IN, United States

dennisctseng@gmail.com
Department of Mathematics, University of Harvard, Cambridge, MA, United States

ALGEBRA AND NUMBER THEORY 18:12 (2024)
https://doi.org/10.2140/ant.2024.18.2157

Separating G_2-invariants of several octonions

Artem Lopatin and Alexandr N. Zubkov

We describe separating G_2-invariants of several copies of the algebra of octonions over an algebraically closed field of characteristic two. We also obtain a minimal separating and a minimal generating set for G_2-invariants of several copies of the algebra of octonions in case of a field of odd characteristic.

1. Introduction

All vector spaces and algebras are considered over an algebraically closed field $\mathbb{F}$ of arbitrary characteristic $p = \operatorname{char} \mathbb{F} \geq 0$.

We continue the study of the invariants of the diagonal action of the exceptional simple group G_2 on the space of several octonions, over a field of positive characteristic. Over the field of complex numbers, this was done in [20]. This result has been generalized to an arbitrary infinite field of odd characteristic in [23], using a much finer technique of modules with good filtration, together with some results from the theory of groups with triality.

Unfortunately, the technique of modules with good filtration no longer works over a field of characteristic two and the complete description of the generating invariants in this case seems to be an extremely difficult problem. Thus, it makes sense to describe separating invariants, since they satisfy the most important property of ordinary invariants to separate closed orbits in the Zariski topology. The latter problem is usually more accessible and it does not require extremely technical methods. We describe the separating invariants over an algebraically closed field of characteristic two, using a detailed description of the subalgebras of the octonion algebra (up to the action of G_2) and the Hilbert–Mumford criterion (the "if" part; see Section 3B).

The article is organized as follows. In Sections 2A and 2B we define the octonion algebra $\boldsymbol{O}$, the group G_2 and the algebra of G_2-invariants $\mathbb{F}[\boldsymbol{O}^n]^{G_2}$ of n copies of the algebra of octonions $\boldsymbol{O}$. We use notation from [23]. Generators and relations between generators for $\mathbb{F}[\boldsymbol{O}^n]^{G_2}$ were described by Schwarz [20] over $\mathbb{F} = \mathbb{C}$. Zubkov and Shestakov described generators for $\mathbb{F}[\boldsymbol{O}^n]^{G_2}$ over an arbitrary field with $\operatorname{char} \mathbb{F} \neq 2$ (see Section 2D), but generators for the algebra $\mathbb{F}[\boldsymbol{O}^n]^{G_2}$ are still not known in case $p = 2$. The invariants for the action of F_4 on several copies of the split Albert algebra were studied in [10]. Our results are formulated in Section 2E. In Section 3 some definitions and notation are given. In Section 4

The work was supported by UAEU grant G00003324 and partially supported in accordance with the state task of the IM SB RAS, project FWNF-2022-0003.
MSC2020: 14L30, 17A75, 17D05, 20F29, 20G05.
Keywords: polynomial invariants, exceptional groups, separating set, generating set, matrix invariants, positive characteristic.

we describe a minimal generating and a minimal separating set for $\mathbb{F}[\boldsymbol{O}^n]^{G_2}$ in case $p \neq 2$. In Section 5 a minimal generating set is constructed for the subalgebra $T_n \subset \mathbb{F}[\boldsymbol{O}^n]^{G_2}$ of trace invariants in case $p = 2$. In Section 6 subalgebras of $\boldsymbol{O}$ of dimension ≤ 3 are described modulo G_2-action in case $p = 2$. This result is applied in Section 7 to obtain our main result which is the description of a separating set for $\mathbb{F}[\boldsymbol{O}^n]^{G_2}$ in case $p = 2$.

2. Invariants of octonions

2A. ***Octonions.*** The *octonion algebra* $\boldsymbol{O} = \boldsymbol{O}(\mathbb{F})$, also known as the *split Cayley algebra*, is the vector space of all matrices

$$a = \begin{pmatrix} \alpha & \boldsymbol{u} \\ \boldsymbol{v} & \beta \end{pmatrix} \quad \text{with } \alpha, \beta \in \mathbb{F} \text{ and } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^3,$$

endowed with the multiplication

$$aa' = \begin{pmatrix} \alpha\alpha' + \boldsymbol{u}\cdot\boldsymbol{v}' & \alpha\boldsymbol{u}' + \beta'\boldsymbol{u} - \boldsymbol{v}\times\boldsymbol{v}' \\ \alpha'\boldsymbol{v} + \beta\boldsymbol{v}' + \boldsymbol{u}\times\boldsymbol{u}' & \beta\beta' + \boldsymbol{v}\cdot\boldsymbol{u}' \end{pmatrix}, \quad \text{where } a' = \begin{pmatrix} \alpha' & \boldsymbol{u}' \\ \boldsymbol{v}' & \beta' \end{pmatrix},$$

$\boldsymbol{u}\cdot\boldsymbol{v} = u_1v_1 + u_2v_2 + u_3v_3$ and $\boldsymbol{u}\times\boldsymbol{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)$. For short, define $\boldsymbol{c}_1 = (1,0,0)$, $\boldsymbol{c}_2 = (0,1,0)$, $\boldsymbol{c}_3 = (0,0,1)$, $\boldsymbol{0} = (0,0,0)$ from $\mathbb{F}^3$. Consider the following basis of $\boldsymbol{O}$:

$$e_1 = \begin{pmatrix} 1 & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{0} & 1 \end{pmatrix}, \quad \boldsymbol{u}_i = \begin{pmatrix} 0 & \boldsymbol{c}_i \\ \boldsymbol{0} & 0 \end{pmatrix}, \quad \boldsymbol{v}_i = \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{c}_i & 0 \end{pmatrix}$$

for $i = 1, 2, 3$. The unity of $\boldsymbol{O}$ is denoted by $1_{\boldsymbol{O}} = e_1 + e_2$. We identify octonions

$$\alpha 1_{\boldsymbol{O}}, \quad \begin{pmatrix} 0 & \boldsymbol{u} \\ \boldsymbol{0} & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{v} & 0 \end{pmatrix}$$

with $\alpha \in \mathbb{F}$, $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^3$, respectively. Similarly to $\boldsymbol{O}(\mathbb{F})$ we define the algebra of octonions $\boldsymbol{O}(\mathcal{A})$ over any commutative associative $\mathbb{F}$-algebra $\mathcal{A}$.

The algebra $\boldsymbol{O}$ has a linear involution

$$\bar{a} = \begin{pmatrix} \beta & -\boldsymbol{u} \\ -\boldsymbol{v} & \alpha \end{pmatrix}, \quad \text{satisfying } \overline{aa'} = \bar{a}'\bar{a},$$

a *norm* $n(a) = a\bar{a} = \alpha\beta - \boldsymbol{u}\cdot\boldsymbol{v}$, and a nondegenerate symmetric bilinear *form*

$$q(a, a') = n(a + a') - n(a) - n(a') = \alpha\beta' + \alpha'\beta - \boldsymbol{u}\cdot\boldsymbol{v}' - \boldsymbol{u}'\cdot\boldsymbol{v}.$$

Define the linear function *trace* by $\mathrm{tr}(a) = a + \bar{a} = \alpha + \beta$. The subspace $\{a \in \boldsymbol{O} \mid \mathrm{tr}(a) = 0\}$ of traceless octonions is denoted by $\boldsymbol{O}_0$. Notice that

$$\mathrm{tr}(aa') = \mathrm{tr}(a'a) \quad \text{and} \quad n(aa') = n(a)n(a'). \tag{2-1}$$

The following quadratic equation holds:

$$a^2 - \mathrm{tr}(a)a + n(a) = 0. \tag{2-2}$$

Since

$$n(a+a') = n(a) + n(a') - \mathrm{tr}(aa') + \mathrm{tr}(a)\,\mathrm{tr}(a'), \tag{2-3}$$

the linearization of (2-2) implies

$$aa' + a'a - \mathrm{tr}(a)a' - \mathrm{tr}(a')a - \mathrm{tr}(aa') + \mathrm{tr}(a)\,\mathrm{tr}(a') = 0. \tag{2-4}$$

The algebra $\boldsymbol{O}$ is a simple *alternative* algebra, i.e., the following identities hold for $a, b \in \boldsymbol{O}$:

$$a(ab) = (aa)b, \quad (ba)a = b(aa). \tag{2-5}$$

The linearization implies that

$$a(a'b) + a'(ab) = (aa' + a'a)b, \quad (ba)a' + (ba')a = b(aa' + a'a). \tag{2-6}$$

The trace is associative, i.e., for all $a, b, c \in \boldsymbol{O}$ we have

$$\mathrm{tr}((ab)c) = \mathrm{tr}(a(bc)). \tag{2-7}$$

Note that

$$2n(a) = -\mathrm{tr}(a^2) + \mathrm{tr}^2(a) \quad \text{for each } a \in \boldsymbol{O}. \tag{2-8}$$

More details on $\boldsymbol{O}$ can be found in Sections 1 and 3 of [23].

2B. ***The group G_2.*** The group $G_2 = G_2(\mathbb{F})$ is known to be the group $\mathrm{Aut}(\boldsymbol{O})$ of all automorphisms of the algebra $\boldsymbol{O}$. The group G_2 contains a Zariski closed subgroup $\mathrm{SL}_3 = \mathrm{SL}_3(\mathbb{F})$. Namely, every $g \in \mathrm{SL}_3$ defines the following automorphism of $\boldsymbol{O}$:

$$a \to \begin{pmatrix} \alpha & \boldsymbol{u}g \\ \boldsymbol{v}g^{-T} & \beta \end{pmatrix},$$

where g^{-T} stands for $(g^{-1})^T$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^3$ are considered as row vectors. In what follows SL_3 is regarded as this subgroup of G_2. For every $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{O}$ define $\delta_1(\boldsymbol{u}), \delta_2(\boldsymbol{v})$ from $\mathrm{Aut}(\boldsymbol{O})$ as

$$\delta_1(\boldsymbol{u})(a') = \begin{pmatrix} \alpha' - \boldsymbol{u}\cdot\boldsymbol{v}' & (\alpha' - \beta' - \boldsymbol{u}\cdot\boldsymbol{v}')\boldsymbol{u} + \boldsymbol{u}' \\ \boldsymbol{v}' - \boldsymbol{u}' \times \boldsymbol{u} & \beta' + \boldsymbol{u}\cdot\boldsymbol{v}' \end{pmatrix}, \quad \delta_2(\boldsymbol{v})(a') = \begin{pmatrix} \alpha' + \boldsymbol{u}'\cdot\boldsymbol{v} & \boldsymbol{u}' + \boldsymbol{v}' \times \boldsymbol{v} \\ (-\alpha' + \beta' - \boldsymbol{u}'\cdot\boldsymbol{v})\boldsymbol{v} + \boldsymbol{v}' & \beta' - \boldsymbol{u}'\cdot\boldsymbol{v} \end{pmatrix}.$$

The group G_2 is generated by SL_3 and $\delta_1(t\boldsymbol{u}_i), \delta_2(t\boldsymbol{v}_i)$ for all $t \in \mathbb{F}$ and $i = 1, 2, 3$ (for example, see Section 3 of [23]). By straightforward calculations we can see that

$$\hbar : \boldsymbol{O} \to \boldsymbol{O}, \quad \text{defined by } a \to \begin{pmatrix} \beta & -\boldsymbol{v} \\ -\boldsymbol{u} & \alpha \end{pmatrix}, \tag{2-9}$$

belongs to G_2 (see also the proof of Lemma 1 of [23]).

The action of G_2 on $\boldsymbol{O}$ satisfies the properties

$$\overline{ga} = g\bar{a}, \quad \mathrm{tr}(ga) = \mathrm{tr}(a), \quad n(ga) = n(a), \quad q(ga, ga') = q(a, a').$$

Thus, $\boldsymbol{O}_0$ is a G_2-submodule of $\boldsymbol{O}$.

Consider the diagonal action of G_2 on the vector space $\boldsymbol{O}^n = \boldsymbol{O} \oplus \cdots \oplus \boldsymbol{O}$ (n copies), that is, $g(a_1, \dots, a_n) = (ga_1, \dots, ga_n)$ for all $g \in G_2$ and $a_1, \dots, a_n \in \boldsymbol{O}$. The coordinate ring of the affine variety $\boldsymbol{O}^n$ is the polynomial $\mathbb{F}$-algebra $K_n = \mathbb{F}[\boldsymbol{O}^n] = \mathbb{F}[z_{ij} \mid 1 \le i \le n,\ 1 \le j \le 8]$, where $z_{ij} : \boldsymbol{O}^n \to \mathbb{F}$ is defined by $(a_1, \dots, a_n) \to \alpha_{ij}$ for

$$a_i = \begin{pmatrix} \alpha_{i1} & (\alpha_{i2}, \alpha_{i3}, \alpha_{i4}) \\ (\alpha_{i5}, \alpha_{i6}, \alpha_{i7}) & \alpha_{i8} \end{pmatrix} \in \boldsymbol{O}. \tag{2-10}$$

The action of $\mathrm{GL}(\boldsymbol{O})$ on $\boldsymbol{O}$ induces the action on K_n by $(gf)(\underline{a}) = f(g^{-1}\underline{a})$ for all $g \in \mathrm{GL}(\boldsymbol{O})$, $f \in K_n$, $\underline{a} \in \boldsymbol{O}^n$.

To explicitly describe the action of G_2 on K_n consider the *generic octonions*

$$Z_i = \begin{pmatrix} z_{i1} & (z_{i2}, z_{i3}, z_{i4}) \\ (z_{i5}, z_{i6}, z_{i7}) & z_{i8} \end{pmatrix} \in \boldsymbol{O}(K_n)$$

for $1 \le i \le n$. Given $g \in G_2$, denote by $g \bullet Z_i$ the octonion

$$\begin{pmatrix} gz_{i1} & (gz_{i2}, gz_{i3}, gz_{i4}) \\ (gz_{i5}, gz_{i6}, gz_{i7}) & gz_{i8} \end{pmatrix} \in \boldsymbol{O}(K_n).$$

For any commutative algebra $\mathcal{A}$, the action of G_2 on $\boldsymbol{O}$ extends for $\boldsymbol{O}(\mathcal{A})$ by $\mathcal{A}$-linearity. In particular, G_2 acts on $\boldsymbol{O}(K_n)$. It is easy to see that

$$g \bullet Z_i = g^{-1} Z_i, \tag{2-11}$$

where $g^{-1}Z_i$ stands for the action of g^{-1} on the octonion $Z_i \in \boldsymbol{O}(K_n)$.

The algebra of *G_2-invariants of several octonions* (*octonion G_2-invariants*, for short) is

$$K_n^{G_2} = \mathbb{F}[\boldsymbol{O}^n]^{G_2} = \{f \in \mathbb{F}[\boldsymbol{O}^n] \mid gf = f \text{ for all } g \in G_2\}.$$

In other words,

$$K_n^{G_2} = \{f \in \mathbb{F}[\boldsymbol{O}^n] \mid f(g\underline{a}) = f(\underline{a}) \text{ for all } g \in G_2,\ \underline{a} \in \boldsymbol{O}^n\}.$$

Similarly we can define $\mathbb{F}[\boldsymbol{O}_0^n]^{G_2}$, since $\boldsymbol{O}_0 \subset \boldsymbol{O}$ is invariant with respect to G_2-action. Namely, the coordinate ring of the affine variety $\boldsymbol{O}_0^n$ is $K_{0,n} = \mathbb{F}[\boldsymbol{O}_0^n] = \mathbb{F}[z_{ij} \mid 1 \le i \le n,\ 1 \le j \le 7]$. The *generic traceless octonions* are

$$X_i = \begin{pmatrix} z_{i1} & (z_{i2}, z_{i3}, z_{i4}) \\ (z_{i5}, z_{i6}, z_{i7}) & -z_{i1} \end{pmatrix}.$$

The analogue of formula (2-11) also holds for the generic traceless octonions, namely, $g \bullet X_i = g^{-1}X_i$ for all $g \in G_2$ and $1 \le i \le n$. The algebra of *G_2-invariants of several traceless octonions* is $K_{0,n}^{G_2}$.

2C. ***Separating sets.*** Consider a finite-dimensional vector space $\mathcal{V}$ and a linear group $G < \mathrm{GL}(\mathcal{V})$. In 2002, Derksen and Kemper [2] introduced the notion of separating invariants as a weaker concept than generating invariants. Given a subset S of $\mathbb{F}[\mathcal{V}]^G$ and u, v of $\mathcal{V}$, we write $S(u) \neq S(v)$ if there exists an invariant $f \in S$ with $f(u) \neq f(v)$. In this case we say that *u, v are separated by S*. If $u, v \in \mathcal{V}$

are separated by $\mathbb{F}[\mathcal{V}]^G$, then we say that they *are separated*. A subset $S \subset \mathbb{F}[\mathcal{V}]^G$ of the invariant ring is called *separating* if for any u, v from $\mathcal{V}$ that are separated we have that they are separated by S. It is well-known that there always exists a finite separating set (see [2, Theorem 2.3.15]). We say that a separating set is minimal if it is minimal w.r.t. inclusion. Obviously, any generating set is also separating. Minimal separating sets and upper bounds on degrees of elements of a separating set for different actions were constructed in [1; 3; 4; 5; 7; 11; 12; 14; 16; 21].

2D. ***Known results.*** Denote by $\mathrm{alg}_{\mathbb{F}}\{Z\}_n$ the nonassociative $\mathbb{F}$-algebra generated by the generic octonions $Z_1, \dots, Z_n$ and $1_{\boldsymbol{O}}$. Any product of the generic octonions is called a word of $\mathrm{alg}_{\mathbb{F}}\{Z\}_n$. The unit $1_{\boldsymbol{O}} \in \mathrm{alg}_{\mathbb{F}}\{Z\}_n$ is called the empty word. For every $A, B \in \mathrm{alg}_{\mathbb{F}}\{Z\}_n$ we have

$$\mathrm{tr}(gA) = \mathrm{tr}(A), \quad n(gA) = n(A), \quad g(AB) = (gA)(gB). \tag{2-12}$$

Lemma 2.1. (a) *The trace of any (nonassociative) product of $X_1, \dots, X_n$ and $n(X_i)$ belongs to $K_{0,n}^{G_2}$.*

(b) *The trace of any (nonassociative) product of $Z_1, \dots, Z_n$ and $n(Z_i)$ belongs to $K_n^{G_2}$.*

(c) *The trace of any (nonassociative) product of $Z_1, \dots, Z_n, \overline{Z_1}, \dots, \overline{Z_n}$ belongs to $K_n^{G_2}$.*

Proof. Let $w = w(Z_1, \dots, Z_n)$ be some (nonassociative) product of $Z_1, \dots, Z_n$. Given $g \in G_2$, equalities (2-11), (2-12) imply that

$$g\,\mathrm{tr}(w) = \mathrm{tr}(w(g \bullet Z_1, \dots, g \bullet Z_n)) = \mathrm{tr}(w(g^{-1}Z_1, \dots, g^{-1}Z_n)) = \mathrm{tr}(g^{-1}w) = \mathrm{tr}(w).$$

The case of $n(Z_i)$ is considered similarly. Part (b) is proven. The proof of part (a) is the same. Part (c) follows from part (b) and formulas

$$\mathrm{tr}(\bar{a}) = \mathrm{tr}(a), \quad n(\bar{a}) = n(a), \quad \mathrm{tr}(\bar{a}b) = \mathrm{tr}(a)\,\mathrm{tr}(b) - \mathrm{tr}(ab)$$

for all $a, b \in \boldsymbol{O}$. □

In case $\mathbb{F} = \mathbb{Q}$ for every $A_1, \dots, A_4 \in \mathrm{alg}_{\mathbb{F}}\{Z\}_n$ denote by $Q'(A_1, A_2, A_3, A_4)$ the complete skew symmetrization of $\mathrm{tr}\big(((A_1A_2)A_3)A_4\big)$ with respect to its arguments, i.e.,

$$Q'(A_1, A_2, A_3, A_4) = \frac{1}{24}\sum_{\sigma \in \mathcal{S}_4} (-1)^{\sigma}\,\mathrm{tr}\big(((A_{\sigma(1)}A_{\sigma(2)})A_{\sigma(3)})A_{\sigma(4)}\big).$$

In [23] it was shown that all coefficients of $Q'(X_1, X_2, X_3, X_4)$ belong to $\mathbb{Z}\big[\frac{1}{2}\big]$. Lemma 4.1 (see below) implies that all coefficients of $Q'(Z_1, Z_2, Z_3, Z_4)$ also belong to $\mathbb{Z}\big[\frac{1}{2}\big]$. Thus $Q'(A_1, A_2, A_3, A_4)$ is well-defined over an arbitrary field of odd characteristic.

In case $\mathrm{char}\,\mathbb{F} \neq 2$,

- the algebra of invariants $K_{0,n}^{G_2}$ is generated by $\mathrm{tr}(X_iX_j)$, $\mathrm{tr}((X_iX_j)X_k)$, $Q'(X_i, X_j, X_k, X_l)$;
- the algebra of invariants $K_n^{G_2}$ is generated by $\mathrm{tr}(Z_i)$, $\mathrm{tr}(Z_iZ_j)$, $\mathrm{tr}((Z_iZ_j)Z_k)$, $Q'(Z_i, Z_j, Z_k, Z_l)$

for all $1 \le i, j, k, l \le n$ (see [23, Corollary 9 and Section 1]).

2E. ***New results.*** Denote by $S_{0,n}$ the set

$$\{n(X_i) \mid 1 \le i \le n\} \cup \bigl\{\mathrm{tr}\bigl(\bigl(\cdots((X_{i_1}X_{i_2})X_{i_3})\cdots\bigr)X_{i_k}\bigr) \mid 1 \le i_1 < \cdots < i_k \le n,\ k > 1\bigr\}$$

and by S_n the set

$$\{n(Z_i) \mid 1 \le i \le n\} \cup \bigl\{\mathrm{tr}\bigl(\bigl(\cdots((Z_{i_1}Z_{i_2})Z_{i_3})\cdots\bigr)Z_{i_k}\bigr) \mid 1 \le i_1 < \cdots < i_k \le n,\ k > 0\bigr\}.$$

Given $1 \le k \le n$, denote by $S_{0,n}^{(k)}$ and $S_n^{(k)}$ the subset of $S_{0,n}$ and S_n (respectively) of elements of degree less or equal to k.

In case $\mathrm{char}\,\mathbb{F} = 2$ generators for the algebras $K_{0,n}^{G_2}$ and $K_n^{G_2}$ are not known. We introduce the algebra of *trace G_2-invariants of octonions $T_n \subset K_n^{G_2}$*, i.e., the algebra T_n is generated by $n(Z_1), \dots, n(Z_n)$ and the traces of all (nonassociative) products of $Z_1, \dots, Z_n$ (see Lemma 2.1). In case $\mathrm{char}\,\mathbb{F} \neq 2$ we obviously have that $T_n = K_n^{G_2}$. We obtain the following results:

- $S_n^{(4)}$ is a minimal (w.r.t. inclusion) generating set for $K_n^{G_2}$ in case $\mathrm{char}\,\mathbb{F} \neq 2$ (see Proposition 4.3).
- $S_n^{(4)}$ is a minimal (w.r.t. inclusion) separating set for $K_n^{G_2}$ in case $\mathrm{char}\,\mathbb{F} \neq 2$ (see Proposition 4.5).
- T_n is minimally generated by S_n in case $\mathrm{char}\,\mathbb{F} = 2$ (see Theorem 5.2).
- $S_n^{(8)}$ is a separating set for $K_n^{G_2}$ in case $\mathrm{char}\,\mathbb{F} = 2$ (see Theorem 7.11).

3. Auxiliaries

3A. ***Indecomposable invariants.*** Denote by $\mathbb{F}\{\mathbb{X}\}_n$ the free nonassociative and noncommutative unital $\mathbb{F}$-algebra with free generators $x_1, \dots, x_n$, which are called letters. A word w is a nonempty product of letters. The number of letters in w is the degree $\deg(w)$ of w. The degree of w in x_i is denoted by $\deg_{x_i}(w)$ and the total degree of w is denoted by $\deg(w)$. The multidegree of a word w is $\mathrm{mdeg}(w) = (\deg_{x_1}(w), \dots, \deg_{x_n}(w))$. A word w with $\deg_{x_i}(w) \le 1$ for all i is called multilinear. An element $f = \sum_i \alpha_i w_i$ of $\mathbb{F}\{\mathbb{X}\}_n$, where $\alpha_i \in \mathbb{F}$ and w_i is a word, is $\mathbb{N}$-homogeneous ($\mathbb{N}^n$-homogeneous, respectively) if there exists d ($\Delta \in \mathbb{N}^n$, respectively) such that $\deg(w_i) = d$ ($\mathrm{mdeg}(w_i) = \Delta$, respectively) for all i, where $\mathbb{N}$ stands for nonnegative integers. Define homomorphisms of $\mathbb{F}$-algebras $\phi_0 : \mathbb{F}\{\mathbb{X}\}_n \to \mathrm{alg}_{\mathbb{F}}\{X\}_n$ and $\phi : \mathbb{F}\{\mathbb{X}\}_n \to \mathrm{alg}_{\mathbb{F}}\{Z\}_n$ by $x_i \to X_i$ and $x_i \to Z_i$ (respectively) for all i. In other words, for $f = f(x_1, \dots, x_n) \in \mathbb{F}\{\mathbb{X}\}_n$ we have $\phi(f) = f(Z_1, \dots, Z_n) \in \mathrm{alg}_{\mathbb{F}}\{Z\}_n$. We write $x_{i_1} \circ \cdots \circ x_{i_k}$ for some nonassociative product of $x_{i_1}, \dots, x_{i_k}$. Similar notation we use for nonassociative products in $\mathrm{alg}_{\mathbb{F}}\{Z\}_n$.

For $f \in K_n$ denote by $\deg(f)$ its degree and by $\mathrm{mdeg}(f)$ its multidegree, i.e., $\mathrm{mdeg}(f) = (t_1, \dots, t_n)$, where t_i is the total degree of the polynomial f in z_{ij}, $1 \le j \le 8$, and $\deg(f) = t_1 + \cdots + t_n$. For $f \in K_{0,n}$ the degree and multidegree are defined as above. It is well-known that the algebras $K_{0,n}^{G_2}$ and $K_n^{G_2}$ have $\mathbb{N}$-gradings by degrees and $\mathbb{N}^n$-gradings by multidegrees.

Consider an $\mathbb{N}^n$-graded unital (possibly, nonassociative) algebra $\mathcal{A}$ with the component of degree zero equal to $\mathbb{F}$. Denote by $\mathcal{A}^+$ the subalgebra generated by homogeneous elements of positive degree. A set $\{a_i\} \subseteq \mathcal{A}$ is a minimal (by inclusion) $\mathbb{N}^n$-homogeneous generating set (m.h.g.s.) of $\mathcal{A}$ as a unital algebra if and only if the a_i's are $\mathbb{N}^n$-homogeneous and $\{\overline{a_i}\} \cup \{1\}$ is a basis of the vector space $\overline{\mathcal{A}} = \mathcal{A}/(\mathcal{A}^+)^2$.

We say that an element $a \in \mathcal{A}$ is *decomposable* and we write $a \equiv 0$ if $a \in (\mathcal{A}^+)^2$. In other words, a decomposable element is equal to a polynomial in elements of strictly less degree. Therefore, the largest degree of indecomposable elements of $\mathcal{A}$ is equal to the least upper bound for the degrees of elements of a m.h.g.s. for $\mathcal{A}$.

3B. ***One-parameter subgroups of G_2.*** Consider a finite-dimensional vector space $\mathcal{V}$ and a linear (closed) group $G < \mathrm{GL}(\mathcal{V})$. For a point $v \in \mathcal{V}$ and for a one-parameter subgroup $\theta : \mathbb{F}^\times \to G$ we have $\theta(t)v = \sum_{i \in I(v)} t^i v^{(i)}$ for all $t \in \mathbb{F}^\times$, where $I(v) = \{i \in \mathbb{Z} \mid v^{(i)} \neq 0\}$. Following [13] we say that $\lim_{t\to 0} \theta(t)v$ exists if and only if $I(v)$ consists of nonnegative integers. Then $\lim_{t\to 0} \theta(t)v = 0$ if and only if $I(v)$ consists of positive integers only, otherwise $\lim_{t\to 0} \theta(t)v = v^{(0)}$. It is clear that if $\lim_{t\to 0} \theta(t)v$ exists, then it is contained in $\overline{Gv}$. Indeed, if f is a polynomial function on $\mathcal{V}$, that vanishes on the G-orbit of v, then $h(t) = f(\theta(t)v)$ is a polynomial in t, such that $h(t) = 0$ for any $t \neq 0$. Since $\mathbb{F}$ is infinite, $h(t)$ is identically zero, that is, $h(0) = f(v^{(0)}) = 0$.

Given $\underline{\lambda} \in \mathbb{Z}^3$ with $\lambda_1 + \lambda_2 + \lambda_3 = 0$, the *standard* one-parameter subgroup $\theta_{\underline{\lambda}}$ of G_2 is defined by

$$\theta_{\underline{\lambda}}(t)e_i = e_i, \quad \theta_{\underline{\lambda}}(t)\boldsymbol{u}_j = t^{\lambda_j}\boldsymbol{u}_j, \quad \theta_{\underline{\lambda}}(t)\boldsymbol{v}_j = t^{-\lambda_j}\boldsymbol{v}_j,$$

for all $i = 1, 2$ and $1 \le j \le 3$.

4. Minimal generating and separating sets

In this section we write $\mathrm{tr}(i_1, \dots, i_k)$ for $\mathrm{tr}((\cdots((Z_{i_1}Z_{i_2})Z_{i_3})\cdots)Z_{i_k})$, where $1 \le i_1, \dots, i_k \le n$. The following lemma can be proven by straightforward calculations.

Lemma 4.1. *Assume that* $\mathrm{char}\,\mathbb{F} \neq 2$. *Then*

$$\begin{aligned} Q'(Z_1, Z_2, Z_3, Z_4) &\\ = \mathrm{tr}(1234) + \tfrac{1}{2}\big(&-\mathrm{tr}(1)\,\mathrm{tr}(2)\,\mathrm{tr}(3)\,\mathrm{tr}(4) - \mathrm{tr}(1)\,\mathrm{tr}(234) - \mathrm{tr}(2)\,\mathrm{tr}(134) - \mathrm{tr}(3)\,\mathrm{tr}(124) \\ &- \mathrm{tr}(4)\,\mathrm{tr}(123) - \mathrm{tr}(12)\,\mathrm{tr}(34) + \mathrm{tr}(13)\,\mathrm{tr}(24) - \mathrm{tr}(14)\,\mathrm{tr}(23)\,\mathrm{tr}(1)\,\mathrm{tr}(2)\,\mathrm{tr}(34) \\ &+ \mathrm{tr}(1)\,\mathrm{tr}(4)\,\mathrm{tr}(23) + \mathrm{tr}(2)\,\mathrm{tr}(3)\,\mathrm{tr}(14) + \mathrm{tr}(3)\,\mathrm{tr}(4)\,\mathrm{tr}(12)\big). \end{aligned}$$

Recall that the definition of T_n was given in Section 2E.

Lemma 4.2. *Let $w \in \mathbb{F}\{\mathbb{X}\}_n$ be a word.*

1. *If w is not multilinear and* $\deg(w) > 2$, *then* $\mathrm{tr}(w(Z_1, \dots, Z_n)) \equiv 0$ *in T_n.*
2. *If w is multilinear and w is a product of letters $x_{i_1}, \dots, x_{i_k}$ for* $1 \le i_1 < \cdots < i_k \le n$, *then*

$$\mathrm{tr}(w(Z_1, \dots, Z_n)) \equiv \pm\mathrm{tr}(i_1, \dots, i_k) \quad \text{in } T_n.$$

3. *For all* $1 \le i_1 < \cdots < i_k \le n$ *with* $k \ge 3$ *and every permutation $\sigma \in \mathcal{S}_k$ we have*

$$\mathrm{tr}(i_{\sigma(1)}, \dots, i_{\sigma(k)}) \equiv (-1)^\sigma\,\mathrm{tr}(i_1, \dots, i_k) \quad \text{in } T_n.$$

Proof. Combining (2-4) and (2-6) we obtain that

$$a(a'b)+a'(ab)=(\operatorname{tr}(a)a'+\operatorname{tr}(a')a+\operatorname{tr}(aa')-\operatorname{tr}(a)\operatorname{tr}(a'))b$$

for all $a, a', b \in \boldsymbol{O}$. Since $\mathbb{F}$ is infinite, the same equality holds for the generic octonions. We multiply it from the left and from the right by the generic octonions and then apply the trace. Since the trace is a linear function, we obtain that

$$\operatorname{tr}\big(C_1\circ\cdots\circ C_r\circ(A(A'B))\circ C_{r+1}\circ\cdots\circ C_s\big)\equiv-\operatorname{tr}\big(C_1\circ\cdots\circ C_r\circ(A'(AB))\circ C_{r+1}\circ\cdots\circ C_s\big) \quad (4\text{-}1)$$

for all products of the generic octonions A, A', B, $C_1, \dots, C_s$ with $0\le r\le s$ and $s\ge 0$. Similarly, we obtain that

$$\operatorname{tr}\big(C_1\circ\cdots\circ C_r\circ((BA)A')\circ C_{r+1}\circ\cdots\circ C_s\big)\equiv-\operatorname{tr}\big(C_1\circ\cdots\circ C_r\circ((BA')A)\circ C_{r+1}\circ\cdots\circ C_s\big). \quad (4\text{-}2)$$

In the same manner as above, (2-2) and (2-4) imply that

$$\operatorname{tr}(C_1\circ\cdots\circ C_r\circ(A^2)\circ C_{r+1}\circ\cdots\circ C_s)\equiv 0, \quad (4\text{-}3)$$

$$\operatorname{tr}(C_1\circ\cdots\circ C_r\circ(AA')\circ C_{r+1}\circ\cdots\circ C_s)\equiv-\operatorname{tr}(C_1\circ\cdots\circ C_r\circ(A'A)\circ C_{r+1}\circ\cdots\circ C_s), \quad (4\text{-}4)$$

where in both cases $0\le r\le s$ and $s>0$. We claim that

If $W=Z_{i_1}\circ\cdots\circ Z_{i_k}$ is a product of generic octonions where $1\le i_1,\dots,i_k\le n$,
then $\operatorname{tr}(W)\equiv\pm\operatorname{tr}(i_{\sigma(1)},\dots,i_{\sigma(k)})$ for some $\sigma\in\mathcal{S}_k$. (4-5)

Assume that claim (4-5) does not hold. Then there exists $\tau\in\mathcal{S}_k$ and the maximal $2\le r<k$ such that some product $W'=Z_{i_{\tau(1)}}\circ\cdots\circ Z_{i_{\tau(k)}}$ satisfies $\operatorname{tr}(W)\equiv\pm\operatorname{tr}(W')$ and

$$W'=C_1\circ\cdots\circ(U(V_1V_2))\circ\cdots\circ C_s \quad\text{or}\quad W'=C_1\circ\cdots\circ(VU)\circ\cdots\circ C_s,$$

where

- $U=\big(\cdots((Z_{j_1}Z_{j_2})Z_{j_3})\cdots\big)Z_{j_r}$ for some $1\le j_1,\dots,j_r\le n$,
- V, V_1, V_2 are some products of generic octonions,
- $C_1,\dots,C_s$ are generic octonions with $s\ge 0$.

By (2-1) and (4-4), we can assume that $W'=C_1\circ\cdots\circ(U(V_1V_2))\circ\cdots\circ C_s$. Consequently, applying equivalence (4-1) and equivalence (2-1) or (4-4), we obtain that

$$\operatorname{tr}\big(C_1\circ\cdots\circ(U(V_1V_2))\circ\cdots\circ C_s\big)\equiv-\operatorname{tr}\big(C_1\circ\cdots\circ(V_1(UV_2))\circ\cdots\circ C_s\big)\equiv\pm\operatorname{tr}\big(C_1\circ\cdots\circ((UV_2)V_1)\circ\cdots\circ C_s\big).$$

If V_2 is a product of more than one generic octonions, then $V_2=V_2'V_2''$ for some products V_2', V_2'' of generic octonions and we repeat the reasoning for $C_1\circ\cdots\circ(U(V_2'V_2''))\circ\cdots\circ C_s$, and so on. Finally, we can assume that $V_2=Z_j$ for some j; a contradiction to the maximality of r.

Equivalences (4-2) and (4-4) imply that part 3 is valid for $1\le i_1,\dots,i_k\le n$, where $k\ge 3$. This fact together with claim (4-5) imply part 2. Similarly, this fact together with claim (4-5) and formula (4-3) imply part 1. □

Proposition 4.3. *In case* char $\mathbb{F} \neq 2$ *the algebra of invariants* $K_n^{G_2}$ *is minimally generated by* $S_n^{(4)}$.

Proof. The description of generators for $K_n^{G_2}$ from [23] (see Section 2D for the details) together with Lemmas 4.1, 4.2 and formula (2-8) imply that the set $S_n^{(4)}$ generates the algebra $K_n^{G_2}$. By Corollary 1 of [23] and formula (2-8), the invariants

$$\mathrm{tr}(i), \quad n(Z_i), \quad \mathrm{tr}(12), \quad \mathrm{tr}(13), \quad \mathrm{tr}(23), \quad \mathrm{tr}(123),$$

where $1 \leq i \leq 3$, are algebraically independent over $\mathbb{F}$. Thus the required statement is proven for $n \leq 3$.

Assume $n \geq 4$. Thus $S_n^{(4)} \backslash \{f\}$ is not a generating set for any $f \in S_n^{(4)}$ with $\deg(f) \neq 4$.

Assume that $S_n^{(4)} \backslash \{\mathrm{tr}(1234)\}$ is a generating set. Then tr(1234) is a linear combination of tr(12) tr(34), tr(13) tr(24), tr(14) tr(23) and products containing $\mathrm{tr}(i)$ for some $1 \leq i \leq 4$. Considering substitutions

$$Z_1 \to \boldsymbol{v}_1, \quad Z_2 \to \boldsymbol{v}_2, \quad Z_3 \to \boldsymbol{v}_3, \quad Z_4 \to e_1 - e_2$$

and using equalities $\mathrm{tr}(((\boldsymbol{v}_1\boldsymbol{v}_2)\boldsymbol{v}_3)(e_1 - e_2)) = -1$ and $\mathrm{tr}(\boldsymbol{v}_i(e_1 - e_2)) = 0$ for $1 \leq i \leq 3$, we obtain a contradiction. The proposition is proven. □

Remark 4.4. 1. By (2-8), in the formulation of Proposition 4.3 we can replace $n(Z_i)$ by $\mathrm{tr}(Z_i^2)$ for all $1 \leq i \leq n$.

2. It easily follows from the proof of Proposition 4.3 (see also Section 1 of [23]) that $K_{0,n}^{G_2}$ is minimally generated by $S_{0,n}^{(4)}$ when char $\mathbb{F} \neq 2$.

Proposition 4.5. *Assume* char $\mathbb{F} \neq 2$. *Then* $S_{0,n}^{(4)}$ *and* $S_n^{(4)}$ *are minimal separating sets for* $K_{0,n}^{G_2}$ *and* $K_n^{G_2}$ (*respectively*) *for all* $n > 0$.

Proof. By Proposition 4.3 and Remark 4.4, the sets $S_{0,n}^{(4)}$ and $S_n^{(4)}$ are separating for $K_{0,n}^{G_2}$ and $K_n^{G_2}$ (respectively). For $a = 0$, $b = \boldsymbol{u}_1 + \boldsymbol{v}_1$ we have $\mathrm{tr}(a) = n(a) = \mathrm{tr}(b) = 0$, but $n(b) = -1$. For $a = 0$, $b = e_1$ we have $\mathrm{tr}(a) = n(a) = n(b) = 0$, but $\mathrm{tr}(b) = 1$. Hence, S_1 is a minimal separating set for $K_1^{G_2}$. Claims 1, 2, 3 (see below) imply that $S_{0,n}^{(4)}$ is a *minimal* separating set for $K_{0,n}^{G_2}$. Therefore, $S_n^{(4)}$ is also a minimal separating set for $K_n^{G_2}$.

Claim 1. Let $n = 2$. Then $S_{0,2} \backslash \{\mathrm{tr}(X_1X_2)\}$ is not separating $K_{0,2}^{G_2}$.

To prove this claim consider $\underline{a} = (0, 0)$ and $\underline{b} = (\boldsymbol{u}_1, \boldsymbol{v}_1)$ from $\boldsymbol{O}_0^2$. Then $\mathrm{tr}(a_1a_2) \neq \mathrm{tr}(b_1b_2)$.

Claim 2. Let $n = 3$. Then $S_{0,3} \backslash \{\mathrm{tr}((X_1X_2)X_3)\}$ is not separating for $K_{0,3}^{G_2}$.

To prove this claim we consider $\underline{a} = (0, 0, 0)$ and $\underline{b} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$ from $\boldsymbol{O}_0^3$. Then $\mathrm{tr}(a_ia_j) = \mathrm{tr}(b_ib_j) = 0$ for all $1 \leq i < j \leq 3$, but $\mathrm{tr}(a_1a_2a_3) \neq \mathrm{tr}(b_1b_2b_3)$.

Claim 3. Let $n = 4$. Then $S_{0,4} \backslash \big\{\mathrm{tr}\big(((X_1X_2)X_3)X_4\big)\big\}$ is not separating for $K_{0,4}^{G_2}$.

To prove this claim we consider $\underline{a} = (\boldsymbol{u}_1, \boldsymbol{v}_1, c, \boldsymbol{u}_2)$ and $\underline{b} = (\boldsymbol{u}_1, \boldsymbol{v}_1, c, -\boldsymbol{v}_2)$ from $\boldsymbol{O}_0^4$, where $c = e_1 + \boldsymbol{u}_2 - \boldsymbol{v}_2 - e_2$. Then $\mathrm{tr}(a_ia_4) = \mathrm{tr}(b_ib_4)$ for $1 \leq i \leq 3$ and $\mathrm{tr}((a_ia_j)a_4) = \mathrm{tr}((b_ib_j)b_4)$ for $1 \leq i < j \leq 3$, but $\mathrm{tr}\big(((a_1a_2)a_3)a_4\big) = 0$ and $\mathrm{tr}\big(((b_1b_2)b_3)b_4\big) = -1$. □

5. Trace invariants

The group $\mathrm{GL}_2 = \mathrm{GL}_2(\mathbb{F})$ acts on $M_2^n = M_2(\mathbb{F})^{\oplus n}$ diagonally by conjugation. The coordinate ring $\mathbb{F}[M_2^n] = \mathbb{F}[z_{i1}, z_{i2}, z_{i5}, z_{i8} \mid 1 \le i \le n]$ is also a GL_2-module, where the *generic matrices* are

$$\widehat{Z}_i = \begin{pmatrix} z_{i1} & z_{i2} \\ z_{i5} & z_{i8} \end{pmatrix}.$$

We consider $\mathbb{F}[M_2^n]$ as a subalgebra of K_n. In [6] it was shown that

$$\{\det(\widehat{Z}_i) \mid 1 \le i \le n\} \cup \{\mathrm{tr}(\widehat{Z}_{i_1} \cdots \widehat{Z}_{i_k}) \mid 1 \le i_1 < \cdots < i_k \le n,\ k > 0\}, \tag{5-1}$$

is a minimal generating set for $\mathbb{F}[M_2^n]^{\mathrm{GL}_2}$, where $k \le 3$ in case $\mathrm{char}\,\mathbb{F} \ne 2$. In particular, all elements from set (5-1) are indecomposable. A minimal separating set for $\mathbb{F}[M_2^n]^{\mathrm{GL}_2}$ was obtained in [11].

Define a surjective homomorphism of $\mathbb{F}$-algebras $\Psi : K_n \to \mathbb{F}[M_2^n]$ as follows: $z_{i3} \to 0$, $z_{i4} \to 0$, $z_{i6} \to 0$, $z_{i7} \to 0$ for all i. We can naturally extend Ψ to the linear map $\widehat{\Psi} : \boldsymbol{O}(K_n) \to \boldsymbol{O}(\mathbb{F}[M_2^n])$ by

$$\widehat{\Psi}\begin{pmatrix} f_1 & (f_2, f_3, f_4) \\ (f_5, f_6, f_7) & f_8 \end{pmatrix} = \begin{pmatrix} \Psi(f_1) & (\Psi(f_2), \Psi(f_3), \Psi(f_4)) \\ (\Psi(f_5), \Psi(f_6), \Psi(f_7)) & \Psi(f_8) \end{pmatrix}$$

for $f_1, \dots, f_8 \in K_n$.

For an associative commutative $\mathbb{F}$-algebra $\mathcal{A}$ define a map $\mathcal{F} : M_2(\mathcal{A}) \to \boldsymbol{O}(\mathcal{A})$ by

$$\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \to \begin{pmatrix} a_1 & (a_2, 0, 0) \\ (a_3, 0, 0) & a_4 \end{pmatrix}$$

for $a_1, \dots, a_4 \in \mathcal{A}$. It is easy to see that $\mathcal{F}$ is an injective homomorphism of algebras preserving the trace, since $(a, 0, 0) \times (b, 0, 0) = 0$ for all $a, b \in \mathcal{A}$. In what follows, we consider $\mathcal{A} = K_n$. Since the homomorphism $\widehat{\Psi}$ commutes with the trace and the norm, we obtain the following lemma.

Lemma 5.1. *For all* $1 \le i, i_1, \dots, i_k \le n$ *we have*

(a) $\widehat{\Psi}\big(\cdots((Z_{i_1} Z_{i_2}) Z_{i_3}) \cdots\big) Z_{i_k}) = \mathcal{F}(\widehat{Z}_{i_1} \cdots \widehat{Z}_{i_k})$;

(b) $\Psi\big(\mathrm{tr}\big(\big(\cdots((Z_{i_1} Z_{i_2}) Z_{i_3}) \cdots\big) Z_{i_k}\big)\big) = \mathrm{tr}(\widehat{Z}_{i_1} \cdots \widehat{Z}_{i_k})$;

(c) $\Psi(n(Z_i)) = \det(\widehat{Z}_i)$.

Lemmas 2.1, 5.1 and the description of generators of $\mathbb{F}[M_2^n]^{\mathrm{GL}_2}$ imply that

$$\mathbb{F}[M_2^n]^{\mathrm{GL}_2} \subset \Psi(K_n^{G_2}), \tag{5-2}$$

where we have equality in case $\mathrm{char}\,\mathbb{F} \ne 2$.

Theorem 5.2. *In case* $\mathrm{char}\,\mathbb{F} = 2$ *the algebra of trace* G_2*-invariants* T_n *is minimally generated by* S_n.

Proof. By Lemma 4.2 and formula (2-8), the algebra T_n is generated by S_n. To show that S_n is a minimal generating set, it is enough to prove that every element $f \in S_n$ is indecomposable in T_n. Assume the contrary. If $f = \mathrm{tr}\big(\big(\cdots((Z_{i_1} Z_{i_2}) Z_{i_3}) \cdots\big) Z_{i_k}\big)$ from S_n were decomposable in T_n, then by parts (b), (c) of Lemma 5.1, $\Psi(f) = \mathrm{tr}(\widehat{Z}_{i_1} \cdots \widehat{Z}_{i_k})$ would be decomposable in $\mathbb{F}[M_2^n]^{\mathrm{GL}_2}$; a contradiction. Similarly, if $f = n(Z_i)$ were decomposable in T_n, then $\Psi(f) = \det(\widehat{Z}_i)$ would be decomposable in $\mathbb{F}[M_2^n]^{\mathrm{GL}_2}$; a contradiction. □

6. Subalgebras of $\boldsymbol{O}$ of low dimension

The group G_2 acts naturally on the set of subalgebras of $\boldsymbol{O}$. For a subalgebra $\mathcal{A}$ of $\boldsymbol{O}$ we denote by $[\mathcal{A}]$ the G_2-orbit of $\mathcal{A}$ and we say that $[\mathcal{A}]$ is the equivalence class of $\mathcal{A}$. Obviously, all algebras in $[\mathcal{A}]$ are isomorphic to $\mathcal{A}$. Denote by $\Omega(\boldsymbol{O})$ the set of G_2-orbits (i.e., equivalence classes) in the set of subalgebras of $\boldsymbol{O}$. Since all algebras from a given equivalence class $\mathfrak{A} \in \Omega(\boldsymbol{O})$ have the same dimension, we call it the *dimension* of $\mathfrak{A}$. A set of (linearly independent) octonions is said to be a *basis* of $\mathfrak{A}$, provided they form a basis of an algebra from $\mathfrak{A}$. An equivalence class $\mathfrak{A} \in \Omega(\boldsymbol{O})$ is called *closed* if there exists a subalgebra $\mathcal{A}$ of $\boldsymbol{O}$ with $[\mathcal{A}] = \mathfrak{A}$ and there is an $\mathbb{F}$-basis $a_1, \dots, a_n$ of $\mathcal{A}$ such that the G_2-orbit of $(a_1, \dots, a_n)$ is closed in $\boldsymbol{O}^n$. More details on the definition of a closed equivalence class can be found in Remark 7.3 (see below). Denote by

$$\mathbb{M} = \begin{pmatrix} * & (*,0,0) \\ (*,0,0) & * \end{pmatrix} \quad \text{and} \quad \mathbb{S} = \begin{pmatrix} * & (*,*,0) \\ (*,0,*) & * \end{pmatrix},$$

the subalgebra of *quaternions* and *sextonions* of $\boldsymbol{O}$, respectively, where the term *sextonions* was introduced in [15]. Note that $\mathcal{F} : M_2(\mathbb{F}) \to \mathbb{M}$ is an isomorphism of $\mathbb{F}$-algebras (see Section 5 for the details).

The main result of this section is the following statement.

Proposition 6.1. *Assume* $\operatorname{char} \mathbb{F} = 2$ *and an equivalence class* $\mathfrak{A} \in \Omega(\boldsymbol{O})$ *has dimension* $d \leq 3$. *Then one of the following sets is a basis for* $\mathfrak{A}$:

$d = 1$: $\{1_{\boldsymbol{O}}\}$, $\{\boldsymbol{u}_1\}$, $\{e_1\}$;

$d = 2$: $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1\}$, $\{\boldsymbol{u}_1, \boldsymbol{v}_2\}$, $\{e_1, \boldsymbol{u}_1\}$, $\{e_1, \boldsymbol{v}_1\}$, $\{e_1, e_2\}$;

$d = 3$: $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, \boldsymbol{v}_2\}$, $\{e_1, e_2, \boldsymbol{u}_1\}$, $\{e_1, \boldsymbol{u}_1, \boldsymbol{v}_2\}$, $\{\boldsymbol{u}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$.

We do not require for a subalgebra of $\boldsymbol{O}$ to be unital. The proof of Proposition 6.1 will be given in a series of propositions and lemmas, which are interesting on their own.

Proposition 6.2 [17, Proposition 3.3]. *For each* $a \in \boldsymbol{O}$ *there exists* $g \in G_2$ *such that* ga *is a canonical octonion of one of the following types*:

(D) $\left(\begin{smallmatrix} \alpha_1 & \mathbf{0} \\ \mathbf{0} & \alpha_8 \end{smallmatrix}\right)$,

(K$_1$) $\left(\begin{smallmatrix} \alpha_1 & (1,0,0) \\ \mathbf{0} & \alpha_1 \end{smallmatrix}\right)$,

for some $\alpha_1, \alpha_8 \in \mathbb{F}$. *These canonical octonions are unique modulo permutation* $\alpha_1 \leftrightarrow \alpha_8$ *for type* (D).

Proposition 6.3 [17, Theorem 4.4]. *Assume* $\operatorname{char} \mathbb{F} = 2$. *For each* $(a, b) \in \boldsymbol{O}_0^2$ *there exists* $g \in G_2$ *such that* $g(a, b)$ *is a pair of one of the following types*:

(EE) $(\alpha_1 1_{\boldsymbol{O}}, \beta_1 1_{\boldsymbol{O}})$,

(EK$_1$) $\left(\alpha_1 1_{\boldsymbol{O}}, \left(\begin{smallmatrix} \beta_1 & (1,0,0) \\ \mathbf{0} & \beta_1 \end{smallmatrix}\right)\right)$, (K$_1$E) $\bigl(\bigl(\begin{smallmatrix}\alpha_1 & (1,0,0)\\ \mathbf{0} & \alpha_1\end{smallmatrix}\bigr), \beta_1 1_{\boldsymbol{O}}\bigr)$,

(K$_1$L$_1$) $\bigl(\bigl(\begin{smallmatrix}\alpha_1 & (1,0,0)\\ \mathbf{0} & \alpha_1\end{smallmatrix}\bigr), \bigl(\begin{smallmatrix}\beta_1 & (\beta_2,0,0)\\ \mathbf{0} & \beta_1\end{smallmatrix}\bigr)\bigr)$ *with* $\beta_2 \neq 0$,

(K$_1$L$_1^\top$) $\bigl(\bigl(\begin{smallmatrix}\alpha_1 & (1,0,0)\\ \mathbf{0} & \alpha_1\end{smallmatrix}\bigr), \bigl(\begin{smallmatrix}\beta_1 & \mathbf{0}\\ (\beta_5,0,0) & \beta_1\end{smallmatrix}\bigr)\bigr)$ *with* $\beta_5 \neq 0$,

(K$_1$M$_1$) $\bigl(\bigl(\begin{smallmatrix}\alpha_1 & (1,0,0)\\ \mathbf{0} & \alpha_1\end{smallmatrix}\bigr), \bigl(\begin{smallmatrix}\beta_1 & (0,1,0)\\ \mathbf{0} & \beta_1\end{smallmatrix}\bigr)\bigr)$,

(K$_1$M$_1^\top$) $\bigl(\bigl(\begin{smallmatrix}\alpha_1 & (1,0,0)\\ \mathbf{0} & \alpha_1\end{smallmatrix}\bigr), \bigl(\begin{smallmatrix}\beta_1 & \mathbf{0}\\ (0,1,0) & \beta_1\end{smallmatrix}\bigr)\bigr)$,

where $\alpha_1, \beta_1, \beta_2, \beta_5 \in \mathbb{F}$.

Remark 6.4 [17, Lemma 3.2]. Assume $a = \bigl(\begin{smallmatrix}\alpha & \boldsymbol{u}\\ \boldsymbol{v} & \beta\end{smallmatrix}\bigr) \in \boldsymbol{O}$. Then:

(a) If $\boldsymbol{u} \neq 0$, then there exists $g \in \mathrm{SL}_3$ such that $ga = \bigl(\begin{smallmatrix}\alpha & (1,0,0)\\ \boldsymbol{v}' & \beta\end{smallmatrix}\bigr)$, where $\boldsymbol{v}' = (*, 0, 0)$ or $\boldsymbol{v}' = (0, 1, 0)$.

(b) If $\boldsymbol{v} \neq 0$, then there exists $g \in \mathrm{SL}_3$ such that $ga = \bigl(\begin{smallmatrix}\alpha & \boldsymbol{u}'\\ (1,0,0) & \beta\end{smallmatrix}\bigr)$, where $\boldsymbol{u}' = (*, 0, 0)$ or $\boldsymbol{u}' = (0, 1, 0)$.

(c) There exist $g, g', g'' \in \mathrm{SL}_3$ such that

$$g(\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_3) = (\boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_3, -\boldsymbol{v}_2),$$
$$g'(\boldsymbol{u}_2, \boldsymbol{v}_2, \boldsymbol{u}_1, \boldsymbol{v}_3) = (\boldsymbol{u}_2, \boldsymbol{v}_2, \boldsymbol{u}_3, -\boldsymbol{v}_1),$$
$$g''(\boldsymbol{u}_3, \boldsymbol{v}_3, \boldsymbol{u}_1, \boldsymbol{v}_2) = (\boldsymbol{u}_3, \boldsymbol{v}_3, \boldsymbol{u}_2, -\boldsymbol{v}_1).$$

(d) If $\boldsymbol{u} = (\gamma_1, \gamma_2, \gamma_3)$ with $\gamma_2 \neq 0$ or $\gamma_3 \neq 0$ and $\boldsymbol{v} = (\delta, 0, 0)$, then there exists $g \in \mathrm{SL}_3$ such that $ga = \bigl(\begin{smallmatrix}\alpha & (\gamma_1,1,0)\\ (\delta,0,0) & \beta\end{smallmatrix}\bigr)$ and $g(\boldsymbol{u}_1, \boldsymbol{v}_1) = (\boldsymbol{u}_1, \boldsymbol{v}_1)$.

(e) If $\boldsymbol{v} = (\gamma_1, \gamma_2, \gamma_3)$ with $\gamma_2 \neq 0$ or $\gamma_3 \neq 0$ and $\boldsymbol{u} = (\delta, 0, 0)$, then there exists $g \in \mathrm{SL}_3$ such that $ga = \bigl(\begin{smallmatrix}\alpha & (\delta,0,0)\\ (\gamma_1,1,0) & \beta\end{smallmatrix}\bigr)$ and $g(\boldsymbol{u}_1, \boldsymbol{v}_1) = (\boldsymbol{u}_1, \boldsymbol{v}_1)$.

The following lemma is an immediate consequence of the Cayley–Dickson doubling process (see also Section 2.1 of [22]). Its analogue over a finite field is part (ii) of Lemma 3.3 from [9].

Lemma 6.5. *Every automorphism of the* $\mathbb{F}$*-algebra* $\mathbb{M}$ *can be extended to an automorphism of the algebra* $\boldsymbol{O}$.

Lemma 6.6. *If* $\mathcal{A} \subset \boldsymbol{O}$ *is a nonzero subalgebra, then there exists* $g \in G_2$ *such that* $1_{\boldsymbol{O}} \in g\mathcal{A}$ *or* $\boldsymbol{u}_1 \in g\mathcal{A}$ *or* $e_1 \in g\mathcal{A}$. *In particular, if* $\operatorname{char}\mathbb{F} = 2$ *and* $\mathcal{A} \not\subset \boldsymbol{O}_0$ *is a nonzero subalgebra of* $\boldsymbol{O}$, *then there exists* $g \in G_2$ *such that* $e_1 \in g\mathcal{A}$.

Proof. This follows from Proposition 6.2, the known corresponding statement for the algebra $\mathbb{M} \simeq M_2(\mathbb{F})$ and Lemma 6.5. □

6A. ***The case of traceless subalgebra.*** In this section we assume that $\operatorname{char}\mathbb{F} = 2$ and $\mathcal{A} \subset \boldsymbol{O}$ is a subalgebra of traceless octonions, that is, $\mathcal{A} \subset \boldsymbol{O}_0$.

Remark 6.7. If $a = \bigl(\begin{smallmatrix}\alpha & \boldsymbol{u}\\ \boldsymbol{v} & \beta\end{smallmatrix}\bigr) \in \mathcal{A}$ is *triangular* (i.e., $\boldsymbol{u} = \mathbf{0}$ or $\boldsymbol{v} = \mathbf{0}$), and $\alpha \neq 0$ or $\beta \neq 0$, then $1_{\boldsymbol{O}} \in \mathcal{A}$.

Proof. Since $\alpha = \beta$ is nonzero, considering $a^2 = \alpha^2 1_{\boldsymbol{O}}$ completes the proof. □

Lemma 6.8. *If* $\dim\mathcal{A} \geq 2$, *then there exists* $g \in G_2$ *such that one of the following possibilities holds*:

(a) $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1\} \subset g\mathcal{A}$;

(b) $\{\boldsymbol{u}_1, \boldsymbol{v}_2\} \subset g\mathcal{A}$ *and* $1_{\boldsymbol{O}} \notin g\mathcal{A}$.

Proof. By Lemma 6.6, we assume that one of the following alternatives holds:

1. $1_{\boldsymbol{O}} \in \mathcal{A}$. There exists $a \in \mathcal{A}$ such that $\{1_{\boldsymbol{O}}, a\}$ are linearly independent. Since $G_2 1_{\boldsymbol{O}} = 1_{\boldsymbol{O}}$, by Proposition 6.2, we can assume that $a = \alpha e_1 + \beta e_2$ or $a = \alpha 1_{\boldsymbol{O}} + \boldsymbol{u}_1$ for some $\alpha, \beta \in \mathbb{F}$. In the first case we have $\alpha = \beta$ and $\{1_{\boldsymbol{O}}, a\}$ are linearly dependent; a contradiction. In the second case we obtain that $\boldsymbol{u}_1 = a - \alpha 1_{\boldsymbol{O}}$ lies in $\mathcal{A}$.

2. $\boldsymbol{u}_1 \in \mathcal{A}$ and $1_{\boldsymbol{O}} \notin \mathcal{A}$. There exists $b \in \mathcal{A}$ such that $\{\boldsymbol{u}_1, b\}$ are linearly independent. Consider $g \in G_2$ such that $g(\boldsymbol{u}_1, b) = (a', b')$ is one of the pairs from Proposition 6.3. Since $\mathrm{tr}(a') = n(a') = 0$ and $a' \neq 0$, one easily sees that $a' = \boldsymbol{u}_1$. By Remark 6.7 and the fact that $1_{\boldsymbol{O}} \notin \mathcal{A}$ one sees that both diagonal entries of b' are equal to zero. Using the fact that $\{\boldsymbol{u}_1, b'\}$ are linearly independent, we obtain that the pair $(\boldsymbol{u}_1, b')$ has one of the following types:

($\mathrm{K_1L_1^\top}$) $b' = \beta \boldsymbol{v}_1$, where $\beta \in \mathbb{F} \setminus \{0\}$. Since $\boldsymbol{u}_1 b' = \beta e_1$ and $\mathrm{tr}(e_1) \neq 0$, we obtain a contradiction.

($\mathrm{K_1M_1}$) $b' = \boldsymbol{u}_2$. Since $\boldsymbol{u}_1 b' = \boldsymbol{v}_3$, acting by a suitable element of SL_3 and using part (c) of Remark 6.4, we obtain case (b).

($\mathrm{K_1M_1^\top}$) $b' = \boldsymbol{v}_2$, i.e., we have case (b). □

Lemma 6.9. *If* $\dim \mathcal{A} \geq 3$, *then there exists* $g \in G_2$ *such that one of the following possibilities holds*:

(a) $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, \boldsymbol{v}_2\} \subset g\mathcal{A}$;

(b) $\{\boldsymbol{u}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\} \subset g\mathcal{A}$ *and* $1_{\boldsymbol{O}} \notin g\mathcal{A}$.

Proof. By Lemma 6.8, one can assume that one of the following possibilities holds:

1. $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1\} \subset \mathcal{A}$. There exists $b \in \mathcal{A}$ such that $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, b\}$ are linearly independent. Consider $g \in G_2$ such that $g(\boldsymbol{u}_1, b) = (a', b')$ is one of the pairs from Proposition 6.3. Since $\mathrm{tr}(a') = n(a') = 0$ and $a' \neq 0$, one easily sees that $a' = \boldsymbol{u}_1$. Let β' be the diagonal element of b'. Since $G_2 1_{\boldsymbol{O}} = 1_{\boldsymbol{O}}$, taking $b'' = b' - \beta' 1_{\boldsymbol{O}}$ instead of b', we can assume that $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, b''\} \subset \mathcal{A}$ are linearly independent and $(\boldsymbol{u}_1, b'')$ has one of types from Proposition 6.3, where the diagonal elements of b'' are zeros. Consider the possible types for $(\boldsymbol{u}_1, b'')$:

($\mathrm{K_1L_1^\top}$) $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, \beta \boldsymbol{v}_1\} \subset \mathcal{A}$ for some nonzero $\beta \in \mathbb{F}$. Since $\boldsymbol{u}_1 \boldsymbol{v}_1 = e_1$, we obtain a contradiction.

($\mathrm{K_1M_1}$) $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, \boldsymbol{u}_2\} \subset \mathcal{A}$. Since $\boldsymbol{u}_1 \boldsymbol{u}_2 = \boldsymbol{v}_3$, acting by a suitable element of SL_3 from part (c) of Remark 6.4 we obtain case (a).

($\mathrm{K_1M_1^\top}$) $\{1_{\boldsymbol{O}}, \boldsymbol{u}_1, \boldsymbol{v}_2\} \subset \mathcal{A}$, i.e., we have case (a).

2. $\{\boldsymbol{u}_1, \boldsymbol{v}_2\} \subset \mathcal{A}$ and $1_{\boldsymbol{O}} \notin \mathcal{A}$. Consider $b \in \mathcal{A}$ such that $\{\boldsymbol{u}_1, \boldsymbol{v}_2, b\}$ are linearly independent. One can assume that $b = \left(\begin{smallmatrix} \beta_1 & (0,\beta_3,\beta_4) \\ (\beta_5,0,\beta_7) & \beta_1 \end{smallmatrix}\right)$ for some $\beta_i \in \mathbb{F}$. Since

$$\boldsymbol{u}_1 b = \begin{pmatrix} \beta_5 & (\beta_1, 0, 0) \\ (0, -\beta_4, \beta_3) & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{v}_2 b = \begin{pmatrix} 0 & (-\beta_7, 0, \beta_5) \\ (0, \beta_1, 0) & \beta_3 \end{pmatrix},$$

we have $\beta_3 = \beta_5 = 0$. The equality $b^2 = (\beta_1^2 + \beta_4\beta_7)1_{\boldsymbol{O}}$ implies that $\{\boldsymbol{u}_1, \boldsymbol{v}_2, b\} \subset \mathcal{A}$, where the element $b = \beta_1 1_{\boldsymbol{O}} + \beta_4\boldsymbol{u}_3 + \beta_7\boldsymbol{v}_3$ is nonzero and $\beta_1^2 = \beta_4\beta_7$.

Let $\beta_1 = 0$. Then $\boldsymbol{u}_3$ lies in $\mathcal{A}$ or case (b) holds. If $\boldsymbol{u}_3 \in \mathcal{A}$, then $\{\boldsymbol{u}_1, \boldsymbol{v}_2, \boldsymbol{u}_3\} \subset \mathcal{A}$; thus, $\{\boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_3\} \subset \hbar\,\mathcal{A}$ and part (c) of Remark 6.4 implies that case (b) holds.

Let $\beta_1 \neq 0$. Then $\beta_4, \beta_7 \neq 0$ and for $g = \delta_1(0, 0, \beta_1/\beta_7)$ from G_2 we have

$$g(\boldsymbol{u}_1, \boldsymbol{v}_2, b) = \left(\boldsymbol{u}_1 + \frac{\beta_1}{\beta_7}\boldsymbol{v}_2,\, \boldsymbol{v}_2,\, \beta_7\boldsymbol{u}_3\right).$$

Therefore, $\{\boldsymbol{u}_1, \boldsymbol{v}_2, \boldsymbol{u}_3\} \subset g\mathcal{A}$ and case (b) holds (see above). □

6B. *The case of nontraceless subalgebra.* In this section we assume that $\operatorname{char}\mathbb{F} = 2$ and $\mathcal{A} \not\subset \boldsymbol{O}_0$ is a subalgebra of $\boldsymbol{O}$.

Lemma 6.10. *If* $\dim\mathcal{A} \geq 2$, *then there exists* $g \in G_2$ *such that one of the following possibilities holds*:

(a) $\{e_1, \boldsymbol{u}_1\} \subset g\mathcal{A}$;

(b) $\{e_1, \boldsymbol{v}_1\} \subset g\mathcal{A}$;

(c) $\{e_1, e_2\} \subset g\mathcal{A}$.

Proof. By Lemma 6.6 we can assume that $e_1 \in \mathcal{A}$. There exists $b \in \mathcal{A}$ such that $\{e_1, b\}$ are linearly independent. One can also assume that $b = \left(\begin{smallmatrix} 0 & \boldsymbol{u} \\ \boldsymbol{v} & \beta \end{smallmatrix}\right)$ for some $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^3$ and $\beta \in \mathbb{F}$.

Assume $\boldsymbol{u} \neq \boldsymbol{0}$. Since $e_1 b = \left(\begin{smallmatrix} 0 & \boldsymbol{u} \\ \boldsymbol{0} & 0 \end{smallmatrix}\right)$, by part (a) of Remark 6.4 there exists $g \in \mathrm{SL}_3$ such that $g(e_1, e_1 b) = (e_1, \boldsymbol{u}_1)$, i.e., the case (a) holds.

Assume $\boldsymbol{v} \neq \boldsymbol{0}$. Since $b\,e_1 = \left(\begin{smallmatrix} 0 & \boldsymbol{0} \\ \boldsymbol{v} & 0 \end{smallmatrix}\right)$, by part (b) of Remark 6.4 there exists $g \in \mathrm{SL}_3$ such that $g(e_1, b\,e_1) = (e_1, \boldsymbol{v}_1)$, i.e., the case (b) holds.

In case $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{0}$ we have $\beta \neq 0$, i.e., the case (c) holds. □

Lemma 6.11. *If* $\dim\mathcal{A} \geq 3$, *then there exists* $g \in G_2$ *such that one of the following possibilities holds*:

(a) $\{e_1, e_2, \boldsymbol{u}_1\} \subset g\mathcal{A}$;

(b) $\{e_1, \boldsymbol{u}_1, \boldsymbol{v}_2\} \subset g\mathcal{A}$.

Before the proof of this lemma we formulate the following remark.

Remark 6.12. (a) $\{e_1, e_2, \boldsymbol{u}_1\} \subset \mathcal{A}$ if and only if $\{e_1, e_2, \boldsymbol{v}_1\} \subset \hbar\,\mathcal{A}$.

(b) $\{e_1, \boldsymbol{u}_1, \boldsymbol{v}_2\} \subset \mathcal{A}$ if and only if $\{e_1, \boldsymbol{u}_2, \boldsymbol{v}_1\} \subset g\mathcal{A}$ for some $g \in G_2$ (see part (c) of Remark 6.4).

Proof of Lemma 6.11. By Lemma 6.10, we assume that one of the following possibilities holds:

1. $\{e_1, \boldsymbol{u}_1\} \subset \mathcal{A}$. There exists $b \in \mathcal{A}$ such that $\{e_1, \boldsymbol{u}_1, b\}$ are linearly independent. We can assume that $b = \left(\begin{smallmatrix} 0 & \boldsymbol{u} \\ \boldsymbol{v} & \beta \end{smallmatrix}\right)$ for some $\boldsymbol{u} = (0, *, *) \in \mathbb{F}^3$, $\boldsymbol{v} = (\gamma_1, \gamma_2, \gamma_3) \in \mathbb{F}^3$ and $\beta \in \mathbb{F}$.

Assume $\boldsymbol{u} \neq \boldsymbol{0}$. Since $e_1 b = \left(\begin{smallmatrix} 0 & \boldsymbol{u} \\ \boldsymbol{0} & 0 \end{smallmatrix}\right)$, by part (d) of Remark 6.4 there exists $g \in \mathrm{SL}_3$ such that $g(e_1, \boldsymbol{u}_1, e_1 b) = (e_1, \boldsymbol{u}_1, \boldsymbol{u}_2)$. By part (c) of Remark 6.4 there exists $g' \in \mathrm{SL}_3$ such that $g'(e_1, \boldsymbol{u}_1, e_1 b) = (e_1, \boldsymbol{u}_1, \boldsymbol{u}_3)$. The equality $\boldsymbol{u}_1\boldsymbol{u}_3 = -\boldsymbol{v}_2$ implies that the case (b) holds.

Assume $\boldsymbol{u}=\boldsymbol{0}$. Note that $b\,e_1=\left(\begin{smallmatrix}0&\boldsymbol{0}\\ \boldsymbol{v}&0\end{smallmatrix}\right)$ lies in $\mathcal{A}$.

Let $\gamma_1\neq 0$. The equality $b\,\boldsymbol{u}_1=\gamma_1 e_2$ implies that the case (a) holds.

Otherwise, $\gamma_1=0$. In case $\gamma_2\neq 0$ or $\gamma_3\neq 0$, by part (e) of Remark 6.4 there exists $g\in\mathrm{SL}_3$ such that $g(e_1,\boldsymbol{u}_1,b\,e_1)=(e_1,\boldsymbol{u}_1,\boldsymbol{v}_2)$, that is, the case (b) holds. If $\gamma_2=\gamma_3=0$, then $\beta\neq 0$ and $e_2\in\mathcal{A}$, that is, the case (a) holds.

2. The case $\{e_1,\boldsymbol{v}_1\}\subset\mathcal{A}$ is similar to case 1.

3. $\{e_1,e_2\}\subset\mathcal{A}$. There exists $b=\left(\begin{smallmatrix}\alpha&\boldsymbol{u}\\ \boldsymbol{v}&\beta\end{smallmatrix}\right)$ in $\mathcal{A}$ such that $\{e_1,e_2,b\}$ are linearly independent. We can assume that $\alpha=\beta=0$.

Assume $\boldsymbol{u}\neq\boldsymbol{0}$. Since $e_1b=\left(\begin{smallmatrix}0&\boldsymbol{u}\\ \boldsymbol{0}&0\end{smallmatrix}\right)$, by part (a) of Remark 6.4 there exists $g\in\mathrm{SL}_3$ such that $g(e_1,e_2,e_1b)=(e_1,e_2,\boldsymbol{u}_1)$, equivalently, the case (a) holds.

Otherwise, $\boldsymbol{v}\neq\boldsymbol{0}$. Since $b=\left(\begin{smallmatrix}0&\boldsymbol{0}\\ \boldsymbol{v}&0\end{smallmatrix}\right)$ by part (b) of Remark 6.4 there exists $g\in\mathrm{SL}_3$ such that $g(e_1,e_2,b)=(e_1,e_2,\boldsymbol{v}_1)$. By part (a) of Remark 6.12 we obtain that case (a) holds. □

6C. ***Proof of Proposition 6.1.*** Assume $\mathfrak{A}=[\mathcal{A}]$ for some subalgebra $\mathcal{A}$ of $\boldsymbol{O}$. Lemmas 6.6, 6.8, 6.9, 6.10, 6.11 imply that there exist $g\in G_2$ such that $g\mathcal{A}$ contains one of the sets from the formulation of Proposition 6.1. Since the $\mathbb{F}$-span of each of these sets is an algebra, the proof is completed.

7. Separating invariants in case char $\mathbb{F}=2$

In this section we assume that char $\mathbb{F}=2$. We introduce some notation for $\underline{a}\in\boldsymbol{O}^n$:

- the *rank* $\mathrm{rk}(\underline{a})$ is the dimension of the subspace of $\boldsymbol{O}$ spanned by $a_1,\dots,a_n$;
- $\mathrm{alg}(\underline{a})$ is the $\mathbb{F}$-algebra (in general, nonunital) generated by $a_1,\dots,a_n$.

Obviously, $\mathrm{rk}(g\underline{a})=\mathrm{rk}(\underline{a})$ for every $g\in G_2$. The following remark is well-known (for example, see Corollary 2.3.6 of [2]).

Remark 7.1. Assume $\underline{a}\in\boldsymbol{O}^n$. Then there exists a unique closed G_2-orbit $O=O_{\underline{a}}$ in the closure of $G_2\underline{a}$, and $O_{\underline{a}}$ is the only closed orbit in the fiber

$$\{\underline{c}\in\boldsymbol{O}^n\mid f(\underline{a})=f(\underline{c})\text{ for all }f\in K_n^{G_2}\}.$$

In particular, $f(\underline{a})=f(\underline{c})$ for every $f\in K_n^{G_2}$ and $\underline{c}\in O_a$.

Observe that the group GL_n acts naturally on $\boldsymbol{O}^n$ on the right as follows: for any $A=(\alpha_{ij})\in\mathrm{GL}_n$ and $\underline{a}\in\boldsymbol{O}^n$ we set

$$(\underline{a}A)_i=\sum_{1\le k\le n}\alpha_{ki}a_k\quad\text{for }1\le i\le n.$$

This action commutes with the left G_2-action.

Lemma 7.2. *Given $\underline{a},\underline{b}\in\boldsymbol{O}^n$, define $\underline{a}'=\underline{a}A$ and $\underline{b}'=\underline{b}A$ for some $A\in\mathrm{GL}_n$. Then*:

(a) *$G_2\underline{a}=G_2\underline{b}$ if and only if $G_2\underline{a}'=G_2\underline{b}'$.*

(b) *Given some $d \geq 2$, we have that $\underline{a}$ and $\underline{b}$ are not separated by $S_n^{(d)}$ if and only if $\underline{a}'$ and $\underline{b}'$ are not separated by $S_n^{(d)}$.*

(c) *$G_2\underline{a}$ is closed if and only if $G_2\underline{a}'$ is closed.*

Proof. Since A is invertible, for each part of this lemma it is sufficient to prove the "only if" implication.

(a) For each $g \in G_2$ the equality $g\underline{a} = \underline{b}$ implies $g\underline{a}' = \underline{b}'$, hence our claim follows.

(b) Assume that $\underline{a}$ and $\underline{b}$ are not separated by $S_n^{(d)}$, i.e., $f(\underline{a}) = f(\underline{b})$ for all $f \in S_n^{(d)}$. The linearity of the trace together with Lemma 4.2 and formulas (2-3), (2-8) imply that $h(\underline{a}') = h(\underline{b}')$ for all $h \in S_n^{(d)}$.

(c) The right action by A on $\boldsymbol{O}^n$ gives a homeomorphism of $\boldsymbol{O}^n$ with respect to the Zariski topology. Hence it sends closed subsets to closed subsets. Moreover, it maps G_2-orbits to G_2-orbits. □

The following remark is a consequence of part (c) of Lemma 7.2.

Remark 7.3. An equivalence class $\mathfrak{A} \in \Omega(\boldsymbol{O})$ is closed if and only if for every subalgebra $\mathcal{A}$ of $\boldsymbol{O}$ with $[\mathcal{A}] = \mathfrak{A}$ we have that if $\mathcal{A}$ is the $\mathbb{F}$-span of some $a_1, \dots, a_n$, then the G_2-orbit of $(a_1, \dots, a_n)$ is closed in $\boldsymbol{O}^n$.

Proposition 7.4. *The set $S_m^{(8)} \subset K_m^{G_2}$ is separating for every $m > 0$ if and only if $S_n^{(8)}$ separates different closed G_2-orbits of $\underline{a} = (a_1, \dots, a_l, 0, \dots, 0) \in \boldsymbol{O}^n$ and $\underline{b} \in \boldsymbol{O}^n$ for all $n > 0$, where*

- *$a_1, \dots, a_l$ is a basis of some subalgebra $\mathcal{A}$ of $\boldsymbol{O}$,*
- *$b_1, \dots, b_n$ of $\boldsymbol{O}$ linearly generate some subalgebra $\mathcal{B}$ of $\boldsymbol{O}$,*
- *$\dim \mathcal{A} \geq \dim \mathcal{B}$.*

Proof. We only have to prove the "if" part of the statement. Assume that $\underline{a}, \underline{b} \in \boldsymbol{O}^n$ are not separated by $S_n^{(8)}$ for some $n > 0$. To obtain the required, we will show that $G_2\underline{a} = G_2\underline{b}$.

By Remark 7.1 we can assume that $G_2\underline{a}$ and $G_2\underline{b}$ are closed.

Claim 1. Given an $\mathbb{F}$-basis $a_1', \dots, a_l'$ of $\mathbb{F}$-span of $a_1, \dots, a_n$, without loss of generality, we can assume that $\underline{a} = (a_1', \dots, a_l', 0, \dots, 0) \in \boldsymbol{O}^n$.

To prove claim 1, we consider $A \in \mathrm{GL}_n$ such that $\underline{a}A = (a_1', \dots, a_l', 0, \dots, 0)$. Parts (a), (b), (c) of Lemma 7.2 imply that we can consider $\underline{a}A, \underline{b}A$ instead of $\underline{a}, \underline{b}$ and claim 1 is proven.

Denote by $\mathcal{A}$ the algebra generated by $a_1, \dots, a_n$ and by $\mathcal{B}$ the algebra generated by $b_1, \dots, b_n$. Without loss of generality we can assume that $\dim \mathcal{A} \geq \dim \mathcal{B}$.

Claim 2. Without loss of generality, we can assume that $\mathbb{F}$-span of $a_1, \dots, a_n$ is $\mathcal{A}$ and $\mathbb{F}$-span of $b_1, \dots, b_n$ is $\mathcal{B}$.

Let us prove claim 2. It is an easy exercise in linear algebra to show that there exists $A \in \mathrm{GL}_n$ such that $\underline{a}A = (a_1', \dots, a_l', 0, \dots, 0)$ and $\underline{b}A = (0, \dots, 0, b_d', \dots, b_t', 0, \dots, 0)$, where $a_1', \dots, a_l'$ is a basis for $\mathbb{F}$-span of $a_1, \dots, a_n$ and $b_d', \dots, b_t'$ is a basis for $b_1, \dots, b_n$. Similarly to claim 1, without loss of generality, we can take $\underline{a}A, \underline{b}A$ instead of $\underline{a}, \underline{b}$, that is, we assume that

$$\underline{a} = (a_1, \dots, a_l, 0, \dots, 0) \quad \text{and} \quad \underline{b} = (0, \dots, 0, b_d, \dots, b_t, 0, \dots, 0),$$

where $l \leq 8 = \dim \boldsymbol{O}$ and $t - d + 1 \leq 8$. There exist words $v_1, \dots, v_r$ of $\mathbb{F}\{\mathbb{X}\}_n$ such that the $\mathbb{F}$-span of the set $a_1, \dots, a_n, v_1(\underline{a}), \dots, v_r(\underline{a})$ is $\mathcal{A}$. Similarly, there exist words $w_1, \dots, w_s$ of $\mathbb{F}\{\mathbb{X}\}_n$ such that the $\mathbb{F}$-span of the set $b_1, \dots, b_n, w_1(\underline{b}), \dots, w_s(\underline{b})$ is $\mathcal{B}$.

Since the map $\boldsymbol{O}^n \to \boldsymbol{O}^{r+s}$ given by $\underline{x} \to (v_1(\underline{x}), \dots, v_r(\underline{x}), w_1(\underline{x}), \dots, w_s(\underline{x}))$ is a morphism of affine algebraic varieties, the G_2-orbits of

$$\underline{c}_1 = (a_1, \dots, a_n, v_1(\underline{a}), \dots, v_r(\underline{a}), w_1(\underline{a}), \dots, w_s(\underline{a})),$$
$$\underline{c}_2 = (b_1, \dots, b_n, v_1(\underline{b}), \dots, v_r(\underline{b}), w_1(\underline{b}), \dots, w_s(\underline{b}))$$

are closed. Obviously, $G_2\underline{a} = G_2\underline{b}$ if and only if $G_2\underline{c}_1 = G_2\underline{c}_2$. By Lemma 4.2 and formula (2-8), for any $f \in S^{(8)}_{n+r+s}$ we have that $f(\underline{c}_1)$ is a nonassociative polynomial in $\mathrm{tr}\big((\dots(a_{i_1}a_{i_2})\dots)a_{i_k}\big)$ and $n(a_i)$ for $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq i \leq n$. But this trace is zero in case $k > 8$ by the construction of $\underline{a}$. The same fact holds also for $f(\underline{c}_2)$. Thus, $\underline{a}$ and $\underline{b}$ are not separated by $S^{(8)}_n$ if and only if $\underline{c}_1$ and $\underline{c}_2$ are not separated by $S^{(8)}_{n+r+s}$. Therefore, we can consider $\underline{c}_1, \underline{c}_2$ instead of $\underline{a}, \underline{b}$ and claim 2 is proven.

Since claims 1 and 2 imply that $\underline{a}, \underline{b}$ satisfy conditions from the formulation of the lemma, we obtain that $G_2\underline{a} = G_2\underline{b}$. □

Lemma 7.5. 1. *For every $a \in \mathbb{M}$ with* $\mathrm{tr}(a) = 1$ *and* $n(a) = 0$ *there exists g from the stabilizer* $\mathrm{Stab}_{G_2}(\mathbb{M}) = \{g \in G_2 \mid g\,\mathbb{M} \subset \mathbb{M}\}$ *such that* $ga = e_1$.

2. *For every $a \in \mathbb{M}$ with* $\mathrm{tr}(a) = 0$ *and* $n(a) = 1$ *there exists* $g \in \mathrm{Stab}_{G_2}(\mathbb{M})$ *such that* $ga \in \{1_{\boldsymbol{O}}, 1_{\boldsymbol{O}} + \boldsymbol{u}_1\}$.

3. *Given nonzero $\gamma \in \mathbb{F}$, there exists* $\xi_\gamma \in \mathrm{Stab}_{G_2}(\mathbb{M})$ *such that for every* $\alpha_1, \dots, \alpha_4 \in \mathbb{F}$ *we have*

$$\xi_\gamma \begin{pmatrix} \alpha_1 & (\alpha_2, 0, 0) \\ (\alpha_3, 0, 0) & \alpha_4 \end{pmatrix} = \begin{pmatrix} \alpha_1 & (\gamma\alpha_2, 0, 0) \\ (\gamma^{-1}\alpha_3, 0, 0) & \alpha_4 \end{pmatrix}.$$

4. *Assume that* $\underline{a} = (e_1, e_2)$ *and* $\underline{b} \in \mathbb{M}^2$ *satisfy* $S^{(2)}_n(\underline{a}) = S^{(2)}_n(\underline{b})$. *Then there exists* $g \in \mathrm{Stab}_{G_2}(\mathbb{M})$ *such that* $gb_1 = e_1$ *and* $gb_2 \in \{e_2, e_2 + \boldsymbol{u}_1, e_2 + \boldsymbol{v}_1\}$.

5. *If* $b \in \mathbb{M}$ *satisfies* $\mathrm{tr}(b) = n(b) = \mathrm{tr}(e_1 b) = 0$, *then* $b \in \mathbb{F}\boldsymbol{u}_1$ *or* $b \in \mathbb{F}\boldsymbol{v}_1$.

Proof. **1.** For $A = \mathcal{F}^{-1}(a)$ we have $\mathrm{tr}(A) = 1$ and $\det(A) = 0$. Hence there exists $g \in \mathrm{GL}_2$ such that $g^{-1}Ag = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and Lemma 6.5 completes the proof.

2. For $A = \mathcal{F}^{-1}(a)$ we have $\mathrm{tr}(A) = 0$ and $\det(A) = 1$. Hence there exists $g \in \mathrm{GL}_2$ such that $g^{-1}Ag = \left(\begin{smallmatrix} \lambda & 1 \\ 0 & \lambda \end{smallmatrix}\right)$ or $g^{-1}Ag = \left(\begin{smallmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{smallmatrix}\right)$ for some $\lambda, \lambda_1, \lambda_2$ and Lemma 6.5 completes the proof.

3. Given $g = \left(\begin{smallmatrix} 1 & 0 \\ 0 & \gamma \end{smallmatrix}\right) \in \mathrm{GL}_2$, we have

$$g^{-1} \begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix} g = \begin{pmatrix} \alpha_1 & \gamma\alpha_2 \\ \gamma^{-1}\alpha_3 & \alpha_4 \end{pmatrix}.$$

Lemma 6.5 concludes the proof.

4. By part 1 we assume that $b_1 = e_1$. Define $\mathcal{F}^{-1}(b_2) = B_2 = \left(\begin{smallmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{smallmatrix}\right)$. Since $0 = \mathrm{tr}(a_1a_2) = \mathrm{tr}(b_1b_2)$, we obtain $\beta_1 = 0$. The equalities $\mathrm{tr}(b_2) = 1$ and $n(b_2) = 0$ imply $\beta_4 = 1$ and $\beta_2\beta_3 = 0$. Part 3 concludes the proof.

5. Define $\mathcal{F}^{-1}(b) = B = \left(\begin{smallmatrix}\beta_1 & \beta_2\\ \beta_3 & \beta_4\end{smallmatrix}\right)$. Since $0 = \mathrm{tr}(e_1 b) = \beta_1$, the equalities $\mathrm{tr}(b) = n(b) = 0$ conclude the proof. □

Lemma 7.6. *Assume that* $\underline{a} = (a_1, 0, \dots, 0) \in \mathbb{M}^n$ *and* $\underline{b} \in \mathbb{M}^n$ *are not separated by* $S_n^{(2)}$, *where* $a_1 \in \{1_{\boldsymbol{O}}, e_1\}$ *and* $\dim(\mathrm{alg}(\underline{b})) \le 1$. *Then* $G_2\,\underline{a} = G_2\,\underline{b}$.

Proof. **1.** Let $a_1 = 1_{\boldsymbol{O}}$. Since $\mathrm{tr}(b_1) = 0$ and $n(b_1) = 1$, by part 2 of Lemma 7.5 we can assume that $b_1 = 1_{\boldsymbol{O}}$ or $b_1 = 1_{\boldsymbol{O}} + \boldsymbol{u}_1$.

In the first case, $\dim(\mathrm{alg}(\underline{b})) \le 1$ implies $\underline{b} = (1_{\boldsymbol{O}}, \beta_2 1_{\boldsymbol{O}}, \dots, \beta_n 1_{\boldsymbol{O}})$ for some $\beta_2, \dots, \beta_n \in \mathbb{F}$. Since $0 = n(b_i) = \beta_i^2$ for all $1 < i \le n$, we have $\underline{a} = \underline{b}$.

In the second case we have that b_1 and $b_1^2 = 1_{\boldsymbol{O}}$ are linearly independent; a contradiction.

2. Let $a_1 = e_1$. Since $\mathrm{tr}(b_1) = 1$ and $n(b_1) = 0$, by part 1 of Lemma 7.5 we can assume that $b_1 = e_1$. Then the condition $\dim(\mathrm{alg}(\underline{b})) \le 1$ implies that $\underline{b} = (e_1, \beta_2 e_1, \dots, \beta_n e_1)$ for some $\beta_2, \dots, \beta_n \in \mathbb{F}$. For each $1 < i \le n$ we have $0 = \mathrm{tr}(a_1 0) = \mathrm{tr}(b_1 b_i) = \beta_i$. Therefore, $\underline{a} = \underline{b}$. □

Lemma 7.7. *Assume that* $\underline{a} = (e_1, e_2, 0, \dots, 0) \in \mathbb{M}^n$ *and* $\underline{b} \in \mathbb{M}^n$ *are not separated by* $S_n^{(2)}$ *and* $\dim(\mathrm{alg}(\underline{b})) \le 2$. *Then* $G_2\,\underline{a} = G_2\,\underline{b}$.

Proof. By part 4 of Lemma 7.5 we can assume that $b_1 = e_1$ and $b_2 \in \{e_2, e_2 + \boldsymbol{u}_1, e_2 + \boldsymbol{v}_1\}$.

Let $b_2 = e_2$. For $3 \le i \le n$ part 5 of Lemma 7.5 implies that $b_i \in \mathbb{F}\boldsymbol{u}_1$ or $b_i \in \mathbb{F}\boldsymbol{v}_1$, since $\mathrm{tr}(b_i) = n(b_i) = \mathrm{tr}(b_1 b_i) = 0$. It follows from $\dim(\mathrm{alg}(\underline{b})) \le 2$ that $b_i = 0$ for all $3 \le i \le n$. Therefore, $\underline{a} = \underline{b}$.

In case $b_2 = e_2 + \boldsymbol{u}_1$ we consider $b_1 b_2 = \boldsymbol{u}_1$ and obtain that $\{e_1, \boldsymbol{u}_1, e_2\} \subset \mathrm{alg}(\underline{b})$; a contradiction.

In case $b_2 = e_2 + \boldsymbol{v}_1$ we consider $b_2 b_1 = \boldsymbol{v}_1$ and obtain that $\{e_1, \boldsymbol{v}_1, e_2\} \subset \mathrm{alg}(\underline{b})$; a contradiction. □

Lemma 7.8. *If* $\underline{a} = (e_1, e_2, \boldsymbol{u}_1, \boldsymbol{v}_1, 0, \dots, 0) \in \mathbb{M}^n$ *and* $\underline{b} \in \mathbb{M}^n$ *are not separated by* $S_n^{(3)}$, *then* $G_2\,\underline{a} = G_2\,\underline{b}$.

Proof. By part 4 of Lemma 7.5 we can assume that $b_1 = e_1$ and $b_2 \in \{e_2, e_2 + \boldsymbol{u}_1, e_2 + \boldsymbol{v}_1\}$. Assume $3 \le i \le n$. We have $\mathrm{tr}(b_i) = n(b_i) = \mathrm{tr}(b_1 b_i) = 0$, since $\mathrm{tr}(a_1 a_3) = \mathrm{tr}(a_1 a_4) = 0$. Thus part 5 of Lemma 7.5 implies that $b_i = \beta_i \boldsymbol{u}_1$ or $b_i = \beta_i \boldsymbol{v}_1$ for some $\beta_i \in \mathbb{F}$. Since $\mathrm{tr}(b_3 b_4) = \mathrm{tr}(a_3 a_4) = 1$, we obtain that $\beta_3 \beta_4 = 1$ and either $b_3 = \beta_3 \boldsymbol{u}_1$, $b_4 = \beta_4 \boldsymbol{v}_1$ or $b_3 = \beta_3 \boldsymbol{v}_1$, $b_4 = \beta_4 \boldsymbol{u}_1$ for some nonzero $\beta_3, \beta_4 \in \mathbb{F}$ with $\beta_3\beta_4 = 1$. Hence equalities $\mathrm{tr}(b_3 b_2) = \mathrm{tr}(b_4 b_2) = 0$ imply that $b_2 = e_2$.

1. Let $b_3 = \beta_3 \boldsymbol{u}_1$, $b_4 = \beta_3^{-1} \boldsymbol{v}_1$. By part 3 of Lemma 7.5 we can assume that $\beta_3 = 1$.

Consider $5 \le i \le n$. If $b_i = \beta_i \boldsymbol{u}_1$, then $0 = \mathrm{tr}(a_4 a_i) = \mathrm{tr}(b_4 b_i) = \beta_i$. If $b_i = \beta_i \boldsymbol{v}_1$, then $0 = \mathrm{tr}(a_3 a_i) = \mathrm{tr}(b_3 b_i) = \beta_i$. Therefore, $\underline{a} = \underline{b}$.

2. If $b_3 = \beta_3 \boldsymbol{v}_1$, $b_4 = \beta_3^{-1} \boldsymbol{u}_1$, we have $0 = \mathrm{tr}((b_1 b_3) b_4) = \mathrm{tr}((a_1 a_3) a_4) = \mathrm{tr}(\boldsymbol{u}_1 \boldsymbol{v}_1) = 1$; a contradiction. □

Lemma 7.9. *If* $\underline{a} = (e_1, e_2, \boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2, \boldsymbol{u}_3, \boldsymbol{v}_3, 0, \dots, 0) \in \boldsymbol{O}^n$ *and* $\underline{b} \in \boldsymbol{O}^n$ *are not separated by* $S_n^{(3)}$, *then* $G_2\,\underline{a} = G_2\,\underline{b}$.

Proof. Given $c_1, \dots, c_8$, denote by $M_{c_1,\dots,c_8}$ the Gram matrix $(\mathrm{tr}(c_i c_j))_{1 \le i,j \le 8}$. Since the trace is a bilinear nondegenerate form on $\boldsymbol{O}$ and $a_1, \dots, a_8$ are linearly independent, we obtain that $\det(G_{a_1,\dots,a_8}) = \det(G_{b_1,\dots,b_8})$ is nonzero. Hence, $b_1, \dots, b_8$ are also linearly independent. In particular, $\mathbb{F}$-span of $b_1, \dots, b_8$ is $\boldsymbol{O}$.

For every $1 \le i \le 8$ and $8 < j \le n$ we have that $\mathrm{tr}(a_i a_j) = \mathrm{tr}(b_i b_j)$ is zero. Therefore, $\mathrm{tr}(b b_j) = 0$ for all $b \in \boldsymbol{O}$. Since tr is nondegenerate on $\boldsymbol{O}$, we obtain $\underline{b} = (b_1, \dots, b_8, 0, \dots, 0)$.

For every $1 \le i, j \le 8$ there exists $1 \le k_{ij} \le 8$ and $\eta_{ij} \in \mathbb{F}$ such that $a_i a_j = \eta_{ij} a_{k_{ij}}$. Therefore, for each $1 \le l \le 8$ we have that $\mathrm{tr}((a_i a_j - \eta_{ij} a_{k_{ij}}) a_l) = \mathrm{tr}((b_i b_j - \eta_{ij} b_{k_{ij}}) b_l)$ is zero. Hence, $b_i b_j = \eta_{ij} b_{k_{ij}}$. Consider a linear map $f : \boldsymbol{O} \to \boldsymbol{O}$ defined on the basis of $\boldsymbol{O}$ by $f(a_i) = b_i$ for all $1 \le i \le 8$. Since the multiplication table for $a_1, \dots, a_8$ is the same as for $b_1, \dots, b_8$, we can see that $f \in G_2$. The required is proven. □

The following statement is a corollary of Proposition 6.1.

Corollary 7.10. *Assume* $\mathrm{char}\,\mathbb{F} = 2$ *and a closed equivalence class* $\mathfrak{A} \in \Omega(\boldsymbol{O})$ *has the dimension* $d \le 3$. *Then one of the following sets is a basis for* $\mathfrak{A}$:

$d = 1$: $\{1_{\boldsymbol{O}}\}$, $\{e_1\}$;

$d = 2$: $\{e_1, e_2\}$.

Proof. We need to show that any basis $\{a_1, \dots, a_n\}$ from Proposition 6.1, different from the above bases, generates nonclosed equivalence class. For each $\underline{a} = (a_1, \dots, a_n)$ the arguments are the same: we find an element $\underline{a}'$ in the closure of $G_2\underline{a}$ such that $\mathrm{rk}(\underline{a}') < \mathrm{rk}(\underline{a})$, which obviously implies that $G_2\underline{a}$ is not closed.

- If $\underline{a} = (\boldsymbol{u}_1) \in \boldsymbol{O}^1$, then for the standard one-parameter subgroup $\theta_{\underline{\lambda}}$ with $\underline{\lambda} = (1, -1, 0)$ the element $\underline{a}' = \lim_{t\to 0} \theta_{\underline{\lambda}}(t)\underline{a} = (0)$ lies in $\overline{G_2\underline{a}}$ (see Section 3B for more details).
- If $\underline{a} = (1_{\boldsymbol{O}}, \boldsymbol{u}_1)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (1_{\boldsymbol{O}}, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (\boldsymbol{u}_1, \boldsymbol{v}_2)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (0, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (e_1, \boldsymbol{u}_1)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (e_1, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (e_1, \boldsymbol{v}_1)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(-1,1,0)}(t)\underline{a} = (e_1, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (1_{\boldsymbol{O}}, \boldsymbol{u}_1, \boldsymbol{v}_2)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (1_{\boldsymbol{O}}, 0, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (e_1, e_2, \boldsymbol{u}_1)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (e_1, e_2, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (e_1, \boldsymbol{u}_1, \boldsymbol{v}_2)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (e_1, 0, 0)$ lies in $\overline{G_2\underline{a}}$.
- If $\underline{a} = (\boldsymbol{u}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$, then $\underline{a}' = \lim_{t\to 0} \theta_{(1,-1,0)}(t)\underline{a} = (0, 0, \boldsymbol{v}_3)$ lies in $\overline{G_2\underline{a}}$. □

Theorem 7.11. *The set* $S_n^{(8)}$ *is a separating set for* $K_n^{G_2}$ *in case* $\mathrm{char}\,\mathbb{F} = 2$.

Proof. We will apply Proposition 7.4 to obtain the required statement. Assume that G_2-orbits of $\underline{a} = (a_1, \dots, a_l, 0, \dots, 0) \in \boldsymbol{O}^n$, $\underline{b} \in \boldsymbol{O}^n$ are closed, $a_1, \dots, a_l$ is a basis of some subalgebra $\mathcal{A}$ of $\boldsymbol{O}$, octonions $b_1, \dots, b_n$ linearly generate some subalgebra $\mathcal{B}$ of $\boldsymbol{O}$, and $\dim \mathcal{A} \ge \dim \mathcal{B}$. We assume that $\underline{a}$ and $\underline{b}$ are not separated by $S_n^{(8)}$.

Let $\dim \mathcal{A} = 8$. We may choose that $\underline{a} = (e_1, e_2, \boldsymbol{u}_1, \boldsymbol{v}_1, \boldsymbol{u}_2, \boldsymbol{v}_2, \boldsymbol{u}_3, \boldsymbol{v}_3, 0, \dots, 0)$ by Lemma 7.2. Then Lemma 7.9 implies that $G_2\,\underline{a} = G_2\,\underline{b}$.

Let $\dim \mathcal{A} < 8$. Then $\mathcal{A}$ lies in a maximal proper subalgebra of $\boldsymbol{O}$. By Theorem 5 of [19], the algebra of sextonions $\mathbb{S}$ is the unique maximal proper subalgebra of $\boldsymbol{O}$ modulo G_2-action (see also Remark 7.12

below). So $\mathcal{A} \subset \mathbb{S}$, that is, for all $1 \le i \le l$ we have

$$a_i = \begin{pmatrix} \alpha_{i1} & (\alpha_{i2}, \alpha_{i3}, 0) \\ (\alpha_{i4}, 0, \alpha_{i5}) & \alpha_{i6} \end{pmatrix}$$

for some $\alpha_{ij} \in \mathbb{F}$. Similarly, we can assume that $\mathcal{B} \subset \mathbb{S}$.

Since for the standard one-parameter subgroup $\theta_{\underline{\lambda}}$ with $\underline{\lambda} = (0, 1, -1)$ we have

$$\theta_{\underline{\lambda}}(t) a_i = \begin{pmatrix} \alpha_{i1} & (\alpha_{i2}, t\alpha_{i3}, 0) \\ (\alpha_{i4}, 0, t\alpha_{i5}) & \alpha_{i6} \end{pmatrix},$$

the limit $\underline{a}' = \lim_{t\to 0} \theta_{\underline{\lambda}}(t)\underline{a}$ exists (see Section 3B for more details). Obviously, $\underline{a}' = (a_1', \dots, a_l', 0, \dots, 0)$ lies in $\mathbb{M}^n$. The orbit $G_2\underline{a}$ is closed, therefore, $\underline{a}' \in G_2\,\underline{a}$. Replacing $\underline{a}$ by $\underline{a}'$ we may assume that $\underline{a} \subset \mathbb{M}^n$. Therefore, $\mathcal{A} \subset \mathbb{M}$. In the same manner we can assume that $\mathcal{B} \subset \mathbb{M}$.

In case $\dim \mathcal{A} = 4$ by Lemma 7.2 we may choose that $\underline{a} = (e_1, e_2, \boldsymbol{u}_1, \boldsymbol{v}_1, 0, \dots, 0)$ and Lemma 7.8 implies that $G_2\,\underline{a} = G_2\,\underline{b}$.

Let $\dim \mathcal{A} \le 3$. By Corollary 7.10 and Lemma 7.2 we can assume that $\underline{a}$ is one of the next elements: $(1_{\boldsymbol{O}})$, (e_1), (e_1, e_2). If $\underline{a} = (1_{\boldsymbol{O}})$ or $\underline{a} = (e_1)$, then Lemma 7.6 implies that $G_2\,\underline{a} = G_2\,\underline{b}$. If $\underline{a} = (e_1, e_2)$, then Lemma 7.7 implies that $G_2\,\underline{a} = G_2\,\underline{b}$.

Finally, by Proposition 7.4 the set $S_n^{(8)}$ is separating for $K_n^{G_2}$. □

Remark 7.12. In the proof of Theorem 5 of [19], which claims that $\mathbb{S}$ is the unique maximal proper subalgebra of $\boldsymbol{O}$ modulo G_2-action, there is a small error, but this does not interfere with the case of an algebraically closed field. See [8; 18] for more details.

Remark 7.13. It follows from Theorem 7.11 that the set $S_{0,n}^{(8)}$ is a separating set for $K_{0,n}^{G_2}$ in case $\operatorname{char} \mathbb{F} = 2$.

Acknowledgement

We are very grateful to the anonymous referees for helpful comments.

References

[1] F. B. Cavalcante and A. Lopatin, “Separating invariants of three nilpotent 3 × 3 matrices”, *Linear Algebra Appl.* **607** (2020), 9–28. MR Zbl

[2] H. Derksen and G. Kemper, *Computational invariant theory*, Invariant Theory Algebr. Transform. Groups **1**, Springer, 2002. MR Zbl

[3] H. Derksen and V. Makam, “Algorithms for orbit closure separation for invariants and semi-invariants of matrices”, *Algebra Number Theory* **14**:10 (2020), 2791–2813. MR Zbl

[4] M. Domokos, “Degree bound for separating invariants of abelian groups”, *Proc. Amer. Math. Soc.* **145**:9 (2017), 3695–3708. MR Zbl

[5] M. Domokos, “Characteristic free description of semi-invariants of 2 × 2 matrices”, *J. Pure Appl. Algebra* **224**:5 (2020), art. id. 106220. Addendum in **224**:6 (2020), art. id. 106270. MR Zbl

[6] M. Domokos, S. G. Kuzmin, and A. N. Zubkov, “Rings of matrix invariants in positive characteristic”, *J. Pure Appl. Algebra* **176**:1 (2002), 61–80. MR Zbl

[7] E. Dufresne, J. Elmer, and M. Sezer, "Separating invariants for arbitrary linear actions of the additive group", *Manuscripta Math.* **143**:1-2 (2014), 207–219. MR Zbl

[8] S. M. Gagola, III, "Maximal subalgebras of the octonions", *J. Pure Appl. Algebra* **217**:1 (2013), 20–21. MR Zbl

[9] A. N. Grishkov, M. d. L. M. Giuliani, and A. V. Zavarnitsine, "Classification of subalgebras of the Cayley algebra over a finite field", *J. Algebra Appl.* **9**:5 (2010), 791–808. MR Zbl

[10] A. V. Iltyakov and I. P. Shestakov, "On invariants of F_4 and the center of the Albert algebra", *J. Algebra* **179**:3 (1996), 838–851. MR Zbl

[11] I. Kaygorodov, A. Lopatin, and Y. Popov, "Separating invariants for 2 × 2 matrices", *Linear Algebra Appl.* **559** (2018), 114–124. MR Zbl

[12] G. Kemper, A. Lopatin, and F. Reimers, "Separating invariants over finite fields", *J. Pure Appl. Algebra* **226**:4 (2022), art. id. 106904. MR Zbl

[13] G. R. Kempf, "Instability in invariant theory", *Ann. of Math.* (2) **108**:2 (1978), 299–316. MR Zbl

[14] M. Kohls and M. Sezer, "Separating invariants for the Klein four group and cyclic groups", *Int. J. Math.* **24**:6 (2013), art. id. 1350046. MR Zbl

[15] J. M. Landsberg and L. Manivel, "The sextonions and $E_{7\frac{1}{2}}$", *Adv. Math.* **201**:1 (2006), 143–179. MR Zbl

[16] A. Lopatin and F. Reimers, "Separating invariants for multisymmetric polynomials", *Proc. Amer. Math. Soc.* **149**:2 (2021), 497–508. MR Zbl

[17] A. Lopatin and A. N. Zubkov, "Classification of G_2-orbits for pairs of octonions", preprint, 2022. arXiv 2208.08122

[18] H. P. Petersson, "Maximal subalgebras of octonions", *J. Pure Appl. Algebra* **217**:9 (2013), 1700–1701. MR Zbl

[19] M. L. Racine, "On maximal subalgebras", *J. Algebra* **30** (1974), 155–180. MR Zbl

[20] G. W. Schwarz, "Invariant theory of G_2 and Spin_7", *Comment. Math. Helv.* **63**:4 (1988), 624–663. MR Zbl

[21] R. J. Sousa Ferreira and A. Lopatin, "Minimal generating and separating sets for $O(3)$-invariants of several matrices", *Oper. Matrices* **17**:3 (2023), 639–651. MR Zbl

[22] T. A. Springer and F. D. Veldkamp, *Octonions, Jordan algebras and exceptional groups*, Springer, 2000. MR Zbl

[23] A. N. Zubkov and I. P. Shestakov, "Invariants of G_2 and Spin(7) in positive characteristic", *Transform. Groups* **23**:2 (2018), 555–588. MR Zbl

Communicated by Gavril Farkas
Received 2022-08-29 Revised 2023-08-19 Accepted 2023-12-18

dr.artem.lopatin@gmail.com
*Departamento de Matemática,
Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil*

a.zubkov@yahoo.com
*Department of Mathematical Sciences, United Arab Emirates University,
Abu Dhabi, United Arab Emirates*

Sobolev Institute of Mathematics, Omsk Branch, Omsk, Russia

ALGEBRA AND NUMBER THEORY 18:12 (2024)
https://doi.org/10.2140/ant.2024.18.2179

Scattering diagrams for generalized cluster algebras

Lang Mou

We construct scattering diagrams for Chekhov–Shapiro generalized cluster algebras where exchange polynomials are factorized into binomials, generalizing the cluster scattering diagrams of Gross, Hacking, Keel and Kontsevich. They turn out to be natural objects arising in Fock and Goncharov's cluster duality. Analogous features and structures (such as positivity and the cluster complex structure) in the ordinary case also appear in the generalized situation. With the help of these scattering diagrams, we show that generalized cluster variables are theta functions and hence have certain positivity property with respect to the coefficients in the binomial factors.

1. Introduction

We study generalized cluster algebras in the sense of [Chekhov and Shapiro 2014]. These algebras are generalizations of the (ordinary) cluster algebras introduced by Fomin and Zelevinsky [2002], allowing more general exchange polynomials (as opposed to only binomials) in mutations.

We will see that generalized cluster algebras cannot only be studied in a similar way as cluster algebras [Fomin and Zelevinsky 2002; 2003; 2007; Berenstein et al. 2005], but that they also naturally appear in the context of the *cluster duality* proposed by Fock and Goncharov [2009]. A modified version of Fock and Goncharov's cluster duality was formulated and proved by Gross, Hacking, Keel and Kontsevich [Gross et al. 2018]. In this paper, we extend the scheme therein to study generalized cluster algebras.

Generalized cluster algebras come in a family containing ordinary cluster algebras. Each algebra in this family can be viewed as (a subalgebra of) the algebra of regular functions of a generalized $\mathcal{A}$-cluster variety. The (generalized version of) cluster duality says this family is in a sense dual to another family of generalized $\mathcal{X}$-cluster varieties. In this paper, we demonstrate this duality by reconstructing a family

MSC2020: 13F60.
Keywords: generalized cluster algebra, scattering diagram, cluster variety.

of generalized cluster algebras with principal coefficients $\mathscr{A}^{\text{prin}}$ from a general fiber of the corresponding dual family of $\mathcal{X}$-cluster varieties.

In the ordinary case, the reconstruction is done through a *cluster scattering diagram*, the main technical tool developed in [Gross et al. 2018], which is a mathematical structure associated to the dual $\mathcal{X}$-cluster variety. For our purpose of studying generalized cluster algebras, we construct *generalized cluster scattering diagrams*. This is done by allowing more general wall-crossing functions on the initial incoming walls. It turns out that many features (such as the positivity property of wall-crossings and the cluster complex structure) in the ordinary case still hold in the generalized situation; see Theorems 6.31 and 7.10.

Using the techniques of scattering diagrams (and related objects such as broken lines) transplanted from [Gross et al. 2018], we are able to prove that generalized cluster monomials are theta functions. As a result, they have certain positivity property coming from that of the scattering diagram. We remark that this positivity is with respect to the coefficients appearing in the binomial factors of exchange polynomials, thus weaker than a conjectural positivity of Chekhov and Shapiro (Conjecture 8.13) with respect to the coefficients of exchange polynomials themselves; See Theorem 8.12 and Section 8.5 for the precise statements.

We next describe the contents of the paper in more detail.

1.1. ***Generalized cluster algebras.*** Our way of generalizing cluster algebras is slightly different from [Chekhov and Shapiro 2014], in the way we deal with coefficients. In a sense, one can go from one formulation to the other, in particular when the coefficients are evaluated in some algebraically closed field; see Sections 3.2, 3.5 and also 8.5. We replace Fomin and Zelevinsky's binomial exchange relation

$$x_k'x_k = p_k^+\prod_{i=1}^n x_i^{[b_{ik}]_+} + p_k^-\prod_{i=1}^n x_i^{[-b_{ik}]_+}$$

with a more general polynomial exchange relation

$$x_k'x_k = \prod_{j=1}^{r_k}\Bigl(p_{k,j}^+\prod_{i=1}^n x_i^{[b_{ik}/r_k]_+} + p_{k,j}^-\prod_{i=1}^n x_i^{[-b_{ik}/r_k]_+}\Bigr).$$

We require the coefficients $p_{k,j}^{\pm}$ (in some semifield $(\mathbb{P},\oplus,\cdot)$) to satisfy the normalized condition $p_{k,j}^+\oplus p_{k,j}^- = 1$. The normalization makes mutations deterministic and a particular choice of coefficients named *principal coefficients* (as in [Fomin and Zelevinsky 2007]) available in the generalized situation.

It turns out many algebraic and combinatorial features of cluster algebras are also inherited by generalized cluster algebras. The same finite type classification as for cluster algebras [Fomin and Zelevinsky 2003] and the generalized Laurent phenomenon have already been obtained in [Chekhov and Shapiro 2014]. We show that the dependence on coefficients in the generalized case behaves very much like the ordinary case [Fomin and Zelevinsky 2007]. In particular, a generalized version of the separation formula, Theorem 3.20, is made available through an analogous notion of principal coefficients. The well-known sign coherence of c-vectors (see Section 3.3) is also extended to the generalized case in Proposition 3.17. We note that there is a rather different version of normalized generalized cluster algebras with a certain reciprocal restriction in [Nakanishi 2015] where some results on the structures of seeds parallel to [Fomin and Zelevinsky 2007] were also established.

Another remarkable feature of an ordinary cluster algebra is the positivity of its cluster variables, i.e., they are all Laurent polynomials in the initial variables x_i and coefficients $p_i^{\pm}$ with nonnegative integer coefficients. This was proved by Lee and Schiffler [2015] for skew-symmetric types and by Gross, Hacking, Keel, and Kontsevich [Gross et al. 2018] for the more general skew-symmetrizable types. We extend the positivity to our generalized case (see Theorem 3.8), showing that the Laurent expression of any cluster variable in x_i and $p_{k,j}^{\pm}$ has nonnegative integer coefficients. We note that the positivity obtained here is (in the reciprocal case) a weak form of a positivity conjecture of Chekhov and Shapiro (which we reformulated in Conjecture 8.13); see Remark 3.9 and Section 8.5.

1.2. ***Generalized cluster varieties.*** Let L be a lattice of finite rank. Fix an algebraically closed field $\Bbbk$ of characteristic zero. The (ordinary) cluster varieties of [Fock and Goncharov 2009] are schemes of the form

$$V = \bigcup_{s} T_{L,s}$$

where each $T_{L,s}$ is a copy of the torus $L \otimes \Bbbk^*$ and they are glued together via birational maps called *cluster mutations*. Here s runs over a set of *seeds* (a seed roughly being a labeled basis of L) iteratively generated by mutations. A cluster mutation is give by the birational map

$$\mu_{(m,n)} : T_L \dashrightarrow T_L, \quad \mu_{(m,n)}^*(z^{\ell}) = z^{\ell}(1+z^m)^{\langle \ell,n\rangle}, \quad \ell \in L^*,$$

for a pair of vectors $(n, m) \in L \times L^*$, where $\langle \cdot, \cdot \rangle$ denotes the natural paring between L^* and L. It has a natural dual by switching the roles of m and n, $\mu_{(-n,m)} : T_{L^*} \to T_{L^*}$. Gluing T_{L^*} via these maps gives the *dual cluster variety* $V^{\vee} := \bigcup_s T_{L^*,s}$.

Depending on the types of seeds and mutations chosen, one obtains either Fock–Goncharov $\mathcal{A}$-cluster varieties or $\mathcal{X}$-cluster varieties, which are dual constructions as above. A cluster algebra $\mathscr{A}$ can be embedded into the upper cluster algebra $\overline{\mathscr{A}}$, defined to be the algebra of regular functions on the corresponding $\mathcal{A}$-variety, while the dual $\mathcal{X}$-variety encodes the so-called Y-variables; see Section 4.

One can actually encode coefficients in each cluster mutation, the above construction thus leading to families of cluster varieties. They mutate along with seeds under certain rules. In the $\mathcal{A}$-case, they mutate as Y-variables (see [Fomin and Zelevinsky 2007; Fock and Goncharov 2009]). In the $\mathcal{X}$-case, the mutation rule of the coefficients has been worked out by Bossinger, Frías-Medina, Magee and Nájera Chávez [Bossinger et al. 2020].

We extend these dynamics of coefficients to the generalized situation for both the $\mathcal{A}$- and $\mathcal{X}$-cases. We define a *generalized cluster mutation* as

$$\mu^*(z^{\ell}) = z^{\ell} \prod_{j=1}^{r} (t_j^- + t_j^+ z^m)^{\langle \ell,n\rangle},$$

which depends on some coefficients $t_j^{\pm}$ in $\Bbbk^*$; see Section 4. Thus an ordinary cluster mutation can be viewed as a specialization of a generalized one. Generalized cluster varieties are then defined by gluing tori via the generalized mutations. We obtain two families of generalized cluster varieties

$$\pi_{\mathcal{A}} : \mathcal{A} \to \operatorname{Spec}(R), \quad \pi_{\mathcal{X}} : \mathcal{X} \to \operatorname{Spec}(R),$$

where the coefficients vary in some torus $\operatorname{Spec}(R) = (\mathbb{G}_m)^d$.

One key observation of Gross, Hacking and Keel [Gross et al. 2015] is that a cluster variety can be investigated through its toric models, and mutations between seeds are basically switching between neighboring toric models. A toric model is a construction of a log Calabi–Yau variety by blowing up a hypersurface on the toric boundary of some toric variety. In the cluster situation, the toric variety depends on the choice of a seed $\boldsymbol{s}$ which also tells us where on the toric boundary to blow up. The resulting log Calabi–Yau variety is shown in [Gross et al. 2015] (under certain nice conditions) to be isomorphic to the corresponding cluster variety up to codimension two subsets. We extend this description to the generalized case, for both $\mathcal{A}$- and $\mathcal{X}$-type varieties; see Theorem 5.4 and Section 5.3.

1.3. ***Scattering diagrams.*** Cluster scattering diagrams are the main technical tool in [Gross et al. 2018]. They have their origin in [Kontsevich and Soibelman 2006; Gross and Siebert 2011] in mirror symmetry. Roughly speaking, in the cluster case, a scattering diagram is a collection of walls in a real vector space with attached *wall-crossing functions* (some of them giving information on mutations). Similar to a cluster algebra which starts with one cluster with information to perform mutations in n directions iteratively, its scattering diagram can be constructed by initially setting up n incoming walls and letting them propagate, generating only outgoing walls.

To get a generalized cluster scattering diagram, we replace ordinary wall-crossings (which correspond to ordinary cluster mutations) on the initial incoming walls by the generalized ones of the form

$$f = \prod_{j=1}^{r} (1 + t_j z^m),$$

where the t_j are treated as formal parameters. Given a seed $\boldsymbol{s}$ (in the generalized sense), the associated data of incoming walls uniquely determines a consistent scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$, which we call *the generalized cluster scattering diagram*.

We show that the behavior of $\mathfrak{D}_{\boldsymbol{s}}$ under mutations is analogous to that of the ordinary case, in a way it is canonically associated to a mutation equivalence class of seeds. This is called the *mutation invariance* in [Gross et al. 2018, Theorem 1.24]. See Theorem 6.27 for the precise description of the following theorem.

Theorem 1.1 (Theorem 6.27). *There is a piecewise linear operation T_k such that $T_k(\mathfrak{D}_{\boldsymbol{s}})$ is equivalent to $\mathfrak{D}_{\mu_k(\boldsymbol{s})}$, where $\mu_k(\boldsymbol{s})$ denotes the mutation in direction k of the seed $\boldsymbol{s}$.*

In analogy with the ordinary case, the mutation invariance leads to the *cluster complex structure* of $\mathfrak{D}_{\boldsymbol{s}}$.

Theorem 1.2 (Theorem 7.10). *There is the cluster cone complex $\Delta_{\boldsymbol{s}}^{+}$ such that $\mathfrak{D}_{\boldsymbol{s}}$ is a union of codimension one cones of $\Delta_{\boldsymbol{s}}^{+}$ (with explicit attached wall-crossing functions) and walls outside $\Delta_{\boldsymbol{s}}^{+}$.*

We observe in Lemma 6.19 that $\mathfrak{D}_{\boldsymbol{s}}$ is equivalent to the *tropical vertex scattering diagram* $\mathfrak{D}_{(X_\Sigma, H)}$ of [Argüz and Gross 2022] associated to the corresponding $\mathcal{X}$-type toric model associated to $\boldsymbol{s}$. It is shown in [Argüz and Gross 2022, Theorem 6.1] that $\mathfrak{D}_{(X_\Sigma, H)}$ is further equivalent (after a certain piecewise linear operation) to the *canonical scattering diagram* $\mathfrak{D}_{(X,D)}$ (see [Gross and Siebert 2022; Argüz and Gross 2022, Section 2]) of the log Calabi–Yau pair (X, D) from the toric model. We thus see that $\mathfrak{D}_{\boldsymbol{s}}$

is canonically associated to the corresponding $\mathcal{X}$-cluster variety, with a different seed $\boldsymbol{s}'$ simply giving another representative $\mathfrak{D}_{\boldsymbol{s}'}$.

1.4. ***Cluster dualities.*** The cluster duality of Fock and Goncharov predicts that, in the ordinary case, the varieties $\mathcal{A}$ and $\mathcal{X}$ (see Section 4 for our convention as we do not need the Langlands dual data) are dual in the sense that the upper cluster algebra $\overline{\mathscr{A}}$ has a basis parametrized by the tropical set $\mathcal{X}^{\mathrm{trop}}(\mathbb{Z})$ (see [Gross et al. 2018, Section 2] for a definition) and vice versa. A modified version of this duality (and when it is true) is the main subject of study of [Gross et al. 2018].

The strategy of [Gross et al. 2018] to get the desired basis is as follows. First the tropical set $\mathcal{X}^{\mathrm{trop}}(\mathbb{Z})$ (resp. $\mathcal{X}^{\mathrm{trop}}(\mathbb{R})$) can be identified with the cocharacter lattice M (resp. $M_{\mathbb{R}} := M \otimes \mathbb{R}$) of a chosen seed torus $T_{M,\boldsymbol{s}} = M \otimes \Bbbk^*$ contained in the variety $\mathcal{X}$. By the mutation invariance, the ordinary cluster scattering diagram $\mathfrak{D}_{\boldsymbol{s}}^{\mathrm{ord}}$ (see Section 6.3) naturally lives in $\mathcal{X}^{\mathrm{trop}}(\mathbb{R})$. Denote by Δ^+ the set of integral points inside the cluster complex (which is again a canonical subset of $\mathcal{X}^{\mathrm{trop}}(\mathbb{Z})$ by mutation invariance).

For any integral point $m \in \mathcal{X}^{\mathrm{trop}}(\mathbb{Z})$, using the scattering diagram $\mathfrak{D}_{\boldsymbol{s}}^{\mathrm{ord}}$, one can construct the *theta function* ϑ_m, which in general is only a formal power series in a completion $\widehat{\Bbbk[M]}_{\boldsymbol{s}}$ which depends on $\boldsymbol{s}$. However, it is shown in [Gross et al. 2018, Theorem 4.9] that for $m \in \Delta^+$, ϑ_m is indeed a Laurent polynomial in $\Bbbk[M]$ and corresponds to a cluster monomial. Furthermore, there is a canonically defined (i.e., independent of $\boldsymbol{s}$) subset Θ of $\mathcal{X}^{\mathrm{trop}}(\mathbb{Z})$ containing Δ^+ such that for any $m \in \Theta$, ϑ_m is a Laurent polynomial on every seed torus. It is also shown in [Gross et al. 2018] that the vector space

$$\mathrm{mid}(\mathcal{A}) := \bigoplus_{m \in \Theta} \vartheta_m$$

has an associative algebra structure whose structure constants are defined through *broken lines*. This algebra $\mathrm{mid}(\mathcal{A})$ can be embedded in $\overline{\mathscr{A}}$ so that for $m \in \Delta^+$, ϑ_m is sent to the corresponding cluster monomial. While we do not know in general when $\mathrm{mid}(\mathcal{A})$ equals $\overline{\mathscr{A}}$ (see [Gross et al. 2018, Theorem 0.3]), we do have a basis of $\mathrm{mid}(\mathcal{A})$ parametrized by the subset Θ. Strictly speaking, this process is done through the principal coefficients case.

Our insight is that it is natural to consider the above cluster duality for generalized cluster varieties. In the principal coefficients case, we take a general fiber $\mathcal{X}_\lambda^{\mathrm{prin}} := \pi_{\mathcal{X}}^{-1}(\lambda)$ of the family

$$\pi_{\mathcal{X}} : \mathcal{X}^{\mathrm{prin}} \to \mathrm{Spec}(R).$$

The generalized cluster scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$ then lives in the tropical set $(\mathcal{X}_\lambda^{\mathrm{prin}})^{\mathrm{trop}}(\mathbb{R})$ which is identified with $M_{\mathbb{R}}$ given a chosen seed $\boldsymbol{s}$. Towards a generalized version of the cluster duality, we show:

Theorem 1.3 (Theorem 8.12). *For any $m \in \Delta_{\boldsymbol{s}}^+$, the theta function ϑ_m constructed from the generalized cluster scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$ corresponds to the cluster monomial of the generalized cluster algebra $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$ whose g-vector is m. Moreover, it is a Laurent polynomial in the initial cluster variables x_i and coefficients $p_{i,j}$ with nonnegative integer coefficients.*

It follows from the above theorem that the family

$$\pi_{\mathcal{A}} : \mathcal{A}^{\text{prin}} \to \operatorname{Spec}(R)$$

can be reconstructed from a general fiber $\mathcal{X}_{\lambda}^{\text{prin}}$ (through any of its toric models); see Proposition 8.3.

In principle, in the generalized case, one could consider the subset Θ as in [Gross et al. 2018] and the corresponding generalized middle cluster algebra $\operatorname{mid}(\mathcal{A}^{\text{prin}})$. This would lead to a formulation of generalized cluster duality similar to the ordinary case in [Gross et al. 2018, Theorem 0.3]. Then the usual problem on when the full Fock–Goncharov conjecture is true remains and can be discussed as in [Gross et al. 2018, Section 8].

1.5. ***Relations to other works.*** There are examples of generalized cluster scattering diagrams from representation theory, where they are realized as Bridgeland's stability scattering diagrams [2017] for quivers (with loops) with potentials; see [Labardini-Fragoso and Mou 2024] for such examples arising from surfaces with orbifold points.

In rank two, the scattering diagram $\mathfrak{D}_s$ has already appeared in [Gross et al. 2010; Gross and Pandharipande 2010] from origins other than cluster algebras. There the wall-crossing functions are shown to encode relative Gromov–Witten invariants on certain log Calabi–Yau surfaces. Some conjectural wall-crossing functions in [Gross and Pandharipande 2010] were later verified in [Reineke and Weist 2013] using techniques from quiver representations; see Example 6.22.

The recent paper of Cheung, Kelley and Musiker [Cheung et al. 2023] (outlined in [Cheung et al. 2021]) and some part of Kelley's PhD thesis [2021] have significant overlaps with this paper and the author's PhD thesis [Mou 2020, Chapter 8]. We in the following highlight the differences and relationships concerning scattering diagrams.

In [Mou 2020, Chapter 8], a class of generalized cluster scattering diagrams were constructed and properties including mutation invariance and cluster complex structure were proved. In that work, a palindromic and monic restriction (termed *reciprocal* in [Chekhov and Shapiro 2014]) on the coefficients was imposed. Such a scattering diagram can be obtained from applying to $\mathfrak{D}_s$ of the current paper an evaluation λ such that the initial wall-crossings are specialized to reciprocal polynomials, i.e., of the form

$$f = 1 + a_1 z^{w} + \cdots + a_{r-1} z^{(r-1)w} + z^{rw},$$

where $r \in \mathbb{Z}_{\geq 0}$, $w \in M$, and $a_k = a_{r-k}$ in some ground field $\Bbbk$; see Section 6.4. Scattering diagrams almost identical to these (with the reciprocal restriction) were later considered by Cheung, Kelley and Musiker [Cheung et al. 2021], with more details provided in [Kelley 2021]. The authors treat the coefficients a_i as formal variables. They also outlined the construction of theta functions, following [Gross et al. 2018].

The current paper aims to fill in gaps and missing details in [Mou 2020], enhance the setup therein to include more general coefficients, and discuss the positivity of generalized cluster algebras using scattering diagram techniques. Shortly after this paper was posted on the arXiv, [Cheung et al. 2023] appeared on the arXiv, completing the program [Cheung et al. 2021]. Despite many similarities between the current paper and [Cheung et al. 2023], our approaches of treating coefficients differ somewhat. In [Cheung et al. 2023], the $\boldsymbol{y}$-variables in a generalized seed and the coefficients $\boldsymbol{a} = (a_i)$ in a generalized exchange polynomial are treated separately. The coefficients $\boldsymbol{a}$ are assumed to be reciprocal and remain unchanged under mutations. In contrast, we view the coefficients $\boldsymbol{a}$ as deriving from the $\boldsymbol{y}$-variables (denoted as $\boldsymbol{p}$ in our notation) by factorizing an exchange polynomial into binomials, with each binomial governed by one coefficient in the style of Fomin and Zelevinsky. This approach allows us to realize more general exchange polynomials (beyond just reciprocal ones), at least for an algebraically closed ground field, by specialization from principal coefficients (see Sections 3.5 and 8.5). This setup also enables us to formulate and prove a form of positivity for generalized cluster algebras, a topic not extensively discussed in [Cheung et al. 2023].

2. Preliminaries

In this section, we review some preliminaries commonly used in the theory of cluster algebras [Fomin and Zelevinsky 2002].

2.1. *Semifields.*

Definition 2.1. A semifield $(\mathbb{P}, \oplus, \cdot)$ is a torsion free (multiplicative) abelian group $\mathbb{P}$ with a binary operation *addition* $\oplus$ which is commutative, associative and distributive.

We denote by $\mathbb{ZP}$ the group ring of $\mathbb{P}$ and by $\mathbb{NP} \subset \mathbb{ZP}$ the subset of linear combinations with coefficients in $\mathbb{N}$. Denote by $\mathbb{QP}$ the field of fractions of $\mathbb{ZP}$.

For an element $p \in \mathbb{P}$, we define in $\mathbb{P}$ two elements:

$$p^+ := \frac{p}{p \oplus 1} \quad \text{and} \quad p^- := \frac{1}{p \oplus 1}.$$

Definition 2.2. Let I be a finite set. We define $\mathrm{Trop}(s_i \mid i \in I)$ to be the (multiplicative) abelian group with free generators s_i indexed by I, with the operation addition $\oplus$:

$$\prod_{i\in I} s_i^{a_i} \oplus \prod_{i\in I} s_i^{b_i} := \prod_{i\in I} s_i^{\min\{a_i, b_i\}}.$$

It is clear that $\mathrm{Trop}(s_i \mid i \in I)$ is a semifield. Such a semifield is called a *tropical semifield.*

For $n \in \mathbb{Z}$, we write $[n]_+ := \max\{n, 0\}$. The elements $s^{\pm}$ for

$$s = \prod_{i\in I} s_i^{a_i} \in \mathrm{Trop}(s_i \mid i \in I)$$

has the following simple expressions:

$$s^+ = \prod_{i\in I} s_i^{[a_i]_+} \quad \text{and} \quad s^- = \prod_{i\in I} s_i^{[-a_i]_+}.$$

Definition 2.3. Denote by $\mathbb{Q}_{\mathrm{sf}}(u_1, \dots, u_l)$ the set of all rational functions in l independent variables which can be written as subtraction-free rational expressions in $u_1, \dots, u_l$. Then the set $\mathbb{Q}_{\mathrm{sf}}(u_1, \dots, u_l)$ is a semifield with respect to the usual addition and multiplication.

We call such $\mathbb{Q}_{\mathrm{sf}}(u_1, \dots, u_l)$ a *universal semifield* since for another arbitrary semifield $\mathbb{P}$ and its elements $p_1, \dots, p_l$, there exists a unique map of semifields from $\mathbb{Q}_{\mathrm{sf}}(u_1, \dots, u_l)$ to $\mathbb{P}$ sending u_i to p_i; see [Berenstein et al. 1996, Lemma 2.1.6].

2.2. *Mutations of matrices.*

Definition 2.4. A matrix $B \in \mathrm{Mat}_{n\times n}(\mathbb{Z})$ is called (left) *skew-symmetrizable* if there exists a diagonal matrix $D = \mathrm{diag}(d_i \mid 1 \le i \le n)$ with $d_i \in \mathbb{Z}_{>0}$ such that

$$DB + (DB)^T = 0.$$

Such a matrix D is called a (left) *skew-symmetrizer* of B.

Definition 2.5 [Fomin and Zelevinsky 2002, Definition 4.2]. Let $B = (b_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{Z})$ be a skew-symmetrizable matrix. For $k = 1, \dots, n$, we define $\mu_k(B) = (b'_{ij}) \in \mathrm{Mat}_{n\times n}(\mathbb{Z})$ the *mutation of B in direction k* by setting

(1) $b'_{ik} = -b_{ik}$ and $b'_{kj} = -b_{kj}$ for $1 \le i, j \le n$;

(2) for $i \ne k$ and $j \ne k$,

$$b'_{ij} = \begin{cases} b_{ij} + b_{ik}b_{kj} & \text{if } b_{ik} > 0 \text{ and } b_{jk} < 0; \\ b_{ij} - b_{ik}b_{kj} & \text{if } b_{ik} < 0 \text{ and } b_{jk} > 0; \\ b_{ij} & \text{otherwise.} \end{cases}$$

It is clear that the matrix $\mu_k(B)$ is again skew-symmetrizable with the same set of skew-symmetrizers of B. One can easily check that a mutation is involutive in the same direction, i.e., $\mu_k \circ \mu_k(B) = B$.

3. Generalized cluster algebras

3.1. *Generalized cluster algebras.*

Cluster algebras were originally invented by Fomin and Zelevinsky [2002], which we later refer to as *ordinary cluster algebras*. A generalization of cluster algebras has been provided by Chekhov and Shapiro [2014]. Our definition of generalized cluster algebras below may be considered as a special case (of a slight generalization) of theirs but with a normalization analogous to the one in [Fomin and Zelevinsky 2002, Definition 5.3] for ordinary cluster algebras. The relation and difference will be explained in Section 3.2.

We follow the pattern of [Fomin and Zelevinsky 2007] to define generalized cluster algebras. Most of the key notions here are the generalized versions of their correspondents in the ordinary case.

Definition 3.1. A (*generalized*) *labeled seed* Σ of rank $n \in \mathbb{N}$ is a triple $(\boldsymbol{x}, \boldsymbol{p}, B)$, where:

- $\boldsymbol{p} = (\boldsymbol{p}_1, \dots, \boldsymbol{p}_n)$ is an n-tuple of collections of elements, where each $\boldsymbol{p}_i = (p_{i,1}, \dots, p_{i,r_i})$ is a r_i-tuple of elements in a semifield $(\mathbb{P}, \oplus, \cdot)$ for some positive integer r_i.
- $\boldsymbol{x} = \{x_1, \dots, x_n\}$ is a collection of algebraically independent rational functions of n variables over $\mathbb{QP}$. In other words, the $x_1, \dots x_n$ are elements in some field of rational functions $\mathcal{F} = \mathbb{QP}(u_1, \dots, u_n)$ such that $\mathcal{F} = \mathbb{QP}(x_1, \dots, x_n)$.

- $B \in \mathrm{Mat}_{n\times n}(\mathbb{Z})$ is skew-symmetrizable such that for any $i = 1, \dots, n$, its i-th column is divisible by r_i. The diagonal matrix $D = \mathrm{diag}(r_i)$ is not necessarily a skew-symmetrizer of B.

For convenience, let I be the index set $\{1, \dots, n\}$. For an arbitrary positive integer k, we use the interval $[1, k]$ to represent the set $\{1, \dots, k\}$. We will often call a labeled seed simply a *seed* if there is no confusion.

Associated to a labeled seed $\Sigma = (\boldsymbol{x}, \boldsymbol{p}, B)$, for each $i \in I$, there is the *exchange polynomial*

$$\theta_i(u, v) = \theta[\boldsymbol{p}_i](u, v) := \prod_{l=1}^{r_i} (p_{i,l}^{+} u + p_{i,l}^{-} v) \in \mathbb{ZP}[u, v].$$

Write $\beta_{ij} = b_{ij}/r_j \in \mathbb{Z}$. We put

$$u_{j;+} := \prod_{i:b_{ij}>0} x_i^{\beta_{ij}}, \quad u_{j;-} := \prod_{i:b_{ij}<0} x_i^{-\beta_{ij}}$$

$$p_{i;+} := \prod_{l=1}^{r_i} p_{i,l}^{+}, \qquad p_{i;-} := \prod_{l=1}^{r_i} p_{i,l}^{-} \in \mathbb{P}.$$

Note that all the above notions are with respect to Σ.

Definition 3.2. For any $k \in I$, we define the *mutation of a seed* $\Sigma = (\boldsymbol{x}, \boldsymbol{p}, B)$ *in direction* k as a new labeled seed $\mu_k(\boldsymbol{x}, \boldsymbol{p}, B) := ((x_i'), (\boldsymbol{p}_i'), B')$, where $\boldsymbol{p}_i' = (p_{i,j}' \mid j \in [1, r_i])$ in the following way:

(1) $$B' = \mu_k(B);$$

(2) $$p_{k,j}' = p_{k,j}^{-1} \quad \text{for } j \in [1, r_k];$$

(3) $$\text{for } i \neq k,\ j \in [1, r_i], \qquad p_{i,j}' = \begin{cases} p_{i,j} \cdot (p_{k;-})^{\beta_{ki}} & \text{if } \beta_{ik} > 0, \\ p_{i,j} \cdot (p_{k;+})^{\beta_{ki}} & \text{if } \beta_{ik} \leq 0, \end{cases}$$

or equivalently

$$\text{for } i \neq k, \quad p_{i,j}' = p_{i,j} \Big(\prod_{l=1}^{r_k} (1 \oplus p_{k,l}^{\mathrm{sgn}(\beta_{ik})}) \Big)^{-\beta_{ki}};$$

(4) $$x_i' = \begin{cases} x_i & \text{if } i \neq k, \\ x_k^{-1} \theta[\boldsymbol{p}_k](u_{k;+}, u_{k;-}) & \text{if } i = k. \end{cases}$$

Lemma 3.3. *The mutation μ_k is involutive, i.e., $\mu_k \circ \mu_k(\Sigma) = \Sigma$.*

Proof. We check that μ_k is involutive on each component of a seed. We denote

$$\mu_k \circ \mu_k(\Sigma) = ((x_i''), (p_{i,j}'' \mid j \in [1, r_i])_{i \in I}, B'').$$

For this seed, we simply denote the relevant objects appearing in Definition 3.2 by adding a double prime to the old symbols, while for $\mu_k(\Sigma)$, we add a single prime.

(1) First of all, the matrix mutation μ_k is an involution, as shown in [Fomin and Zelevinsky 2002].

(2) We have for $j \in [1, r_k]$,

$$p_{k,j}'' = (p_{k,j}')^{-1} = p_{k,j}.$$

(3) For $i \neq k$, we have for $j \in [1, r_i]$,

$$\begin{aligned} p''_{i,j} &= \begin{cases} p'_{i,j} \cdot (p'_{k;-})^{\beta'_{ki}} & \text{if } \beta'_{ik} > 0, \\ p'_{i,j} \cdot (p'_{k;+})^{\beta'_{ki}} & \text{if } \beta'_{ik} \leq 0 \end{cases} \\ &= \begin{cases} p_{i,j} \cdot (p_{k;+})^{\beta_{ki}} \cdot (p'_{k;-})^{-\beta_{ki}} & \text{if } \beta_{ik} < 0, \\ p_{i,j} \cdot (p_{k;-})^{\beta_{ki}} \cdot (p'_{k;+})^{-\beta_{ki}} & \text{if } \beta_{ik} \geq 0 \end{cases} \\ &= p_{i,j}. \end{aligned}$$

The last equality is because $p'_{k;+} = p_{k;-}$ and $p'_{k;-} = p_{k;+}$.

(4) Finally for the $\boldsymbol{x}$ part, we have

$$\begin{aligned} x''_i &= \begin{cases} x'_i & \text{if } i \neq k, \\ (x'_k)^{-1}\theta[\boldsymbol{p}'_k](u'_{k;+}, u'_{k;-}) & \text{if } i = k \end{cases} \\ &= \begin{cases} x_i & \text{if } i \neq k, \\ x_k \cdot \theta[\boldsymbol{p}_k](u_{k;+}, u_{k;-})^{-1}\theta[\boldsymbol{p}'_k](u'_{k;+}, u'_{k;-}) & \text{if } i = k \end{cases} \\ &= x_i. \end{aligned}$$

The last equality is because that $\theta[\boldsymbol{p}'_k](u, v) = \theta[\boldsymbol{p}_k](v, u)$ and $u'_{k;\pm} = u_{k;\mp}$.

So overall we have proven that $\mu_k \circ \mu_k(\Sigma) = \Sigma$. □

Fix a positive integer n. We consider the (nonoriented) *n-regular tree* $\mathbb{T}_n$ whose edges are labeled by the numbers $1, \dots, n$ as in [Fomin and Zelevinsky 2002]. Lemma 3.3 makes the following definition well-defined.

Definition 3.4. A (*generalized*) *cluster pattern* is an assignment of a labeled seed $\Sigma_t = (\boldsymbol{x}_t, \boldsymbol{p}_t, B^t)$ to every vertex $t \in \mathbb{T}_n$, such that for any k-labeled edge with endpoints t and t', the seed $\Sigma_{t'}$ is the mutation in direction k of Σ_t, i.e., $\Sigma_{t'} = \mu_k(\Sigma_t)$.

The elements in Σ_t are written as follows:

$$\boldsymbol{x}_t = (x_{i;t} \mid i \in I), \quad \boldsymbol{p}_{i;t} = (p_{i,j;t} \mid j \in [1, r_i]), \quad B^t = (b^t_{ij}).$$

The part $\boldsymbol{x}$ of a labeled seed is called a (labeled) *cluster*, elements x_i are called *cluster variables*, elements $p_{i,j}$ are called *coefficients* and B is called *exchange matrix*.

Two seeds are *mutation-equivalent* if one is obtained from the other by a sequence of mutations. If a seed Σ appears in a cluster pattern, then by definition any seeds mutation-equivalent to Σ must also appear. On the other hand, assigning a seed of rank n to any vertex of $\mathbb{T}_n$ would uniquely determine a cluster pattern.

By definition, all cluster variables live in some ambient field $\mathcal{F}$ of rational functions of n variables. One may identify $\mathcal{F}$ with $\mathbb{QP}(x_1, \cdots, x_n)$ where $(x_1, \dots, x_n)$ is a cluster.

Definition 3.5. Given a generalized cluster pattern, the (*generalized*) *cluster algebra* $\mathscr{A}$ is defined to be the $\mathbb{ZP}$-subalgebra of the ambient field $\mathcal{F}$ generated by all cluster variables $x_{i;t}$ in all seeds that appear in the cluster pattern. Since a cluster pattern is determined by any of its seed, we denote $\mathscr{A} = \mathscr{A}(\Sigma)$ where $\Sigma = (\boldsymbol{x}, \boldsymbol{p}, B)$ is any seed in this cluster pattern.

Remark 3.6. In the case where $r_i = 1$ for any $i \in I$, all the above generalized notions recover the original versions of [Fomin and Zelevinsky 2007].

For convenience, one can specify one vertex $t_0 \in \mathbb{T}_n$ to be *initial*, thus the associated seed being called the *initial seed* with the *initial* cluster, cluster variables, coefficients and exchange matrix. All other seeds are obtained by applying mutations iteratively to the initial one. For the following two theorems, we denote by $(x_1, \dots, x_n)$ the initial cluster.

Theorem 3.7 (generalized Laurent phenomenon, cf. [Fomin and Zelevinsky 2002] and [Chekhov and Shapiro 2014]). *Let $(\boldsymbol{x}, \boldsymbol{p}, B)$ be a generalized labeled seed. Then any cluster variable of $\mathscr{A}(\boldsymbol{x}, \boldsymbol{p}, B)$ is a Laurent polynomial over $\mathbb{ZP}$ in the initial cluster variables x_i, i.e., an element in $\mathbb{ZP}[x_1^{\pm}, \dots, x_n^{\pm}]$.*

Proof. The generalized Laurent phenomenon was already obtained in [Chekhov and Shapiro 2014, Theorem 2.5]. Since our setting is slightly different, we give a proof for completeness.

The proof completely follows from the discussion in [Fomin and Zelevinsky 2002, Section 3]. The generalized Laurent property is a direct corollary of [loc. cit., Theorem 3.2]. One only needs to check the following hypothesis required by [loc. cit., Theorem 3.2]: for any subgraph

$$t_0 \xrightarrow{\;i\;} t_1 \xrightarrow{\;j\;} t_2 \xrightarrow{\;i\;} t_3$$

in the tree $\mathbb{T}_n$, if we define the following three *exchange polynomial* in n variables $x_1, \dots, x_n$ by writing

$$P(\boldsymbol{x}_{t_0}) = \theta[\boldsymbol{p}_{i;t_0}](u^{t_0}_{i;+}, u^{t_0}_{i;-}), \quad Q(\boldsymbol{x}_{t_1}) = \theta[\boldsymbol{p}_{j;t_1}](u^{t_1}_{j;+}, u^{t_1}_{j;-}), \quad R(\boldsymbol{x}_{t_2}) = \theta[\boldsymbol{p}_{i;t_2}](u^{t_2}_{i;+}, u^{t_2}_{i;-}),$$

then they satisfy $R = C \cdot (P|_{x_j \leftarrow Q_0/x_j})$, where $Q_0 = Q|_{x_i=0}$ for some $C \in \mathbb{NP}[x_1, \dots x_n]$.

Notice that since $t_0 \xrightarrow{\;i\;} t_1$, we have

$$P = \prod_{l=1}^{r_i} \Big(p^{+}_{i,l;t_1} \prod_k x_k^{[\beta^{t_1}_{ki}]_+} + p^{-}_{i,l;t_1} \prod_k x_k^{[-\beta^{t_1}_{ki}]_+} \Big).$$

When $\beta^{t_1}_{ij} = 0$, $\beta^{t_0}_{ji} = -\beta^{t_1}_{ij} = 0$. So x_j does not appear in P, implying $P = P|_{x_j \leftarrow Q_0/x_j}$. In this case, we have for any $l \in [1, r_i]$

$$p_{i,l;t_0} = p^{-1}_{i,l;t_2}, \quad \beta^{t_0}_{li} = -\beta^{t_2}_{li}.$$

Thus we have $R = P$.

When $\beta^{t_1}_{ij} < 0$ (implying $\beta^{t_1}_{ji} > 0$), then

$$Q_0/x_j = p_{j;+;t_1} x_j^{-1} \prod_k x_k^{[b^{t_1}_{kj}]_+}$$

and

$$P|_{x_j \leftarrow Q_0/x_j} = \prod_{l=1}^{r_i} \Big(p^{+}_{i,l;t_1} p^{\beta^{t_1}_{ji}}_{j;+;t_1} x_j^{-\beta^{t_1}_{ji}} \prod_{k \neq j} x_k^{[\beta^{t_1}_{ki}]_+ + \beta^{t_1}_{ji} \cdot [b^{t_1}_{kj}]_+} + p^{-}_{i,l;t_1} \prod_k x_k^{[-\beta^{t_1}_{ki}]_+} \Big)$$

We take the ratio of the two monomials in each factor of the above product to obtain monomials

$$p_{i,l;t_1} \cdot p_{j;+;t_1}^{\beta_{ji}^{t_1}} \cdot \prod_k x_k^{\beta_{ki}^{t_2}}.$$

We get exactly the same monomials if we do the same for R. So R and $P|_{x_j \leftarrow Q_0/x_j}$ only differ by a monomial factor in $\mathbb{NP}[x_1, \dots, x_n]$. The case when $\beta_{ij}^{t_1} > 0$ is analogous. Hence the hypothesis is checked and the Laurent phenomenon follows from [Fomin and Zelevinsky 2002, Theorem 3.2]. □

The following Theorem 3.8 is a generalization of the well-known positivity for ordinary cluster algebras. In the case of ordinary cluster algebras, the positivity was originally conjectured by Fomin and Zelevinsky [2002]. It has been proved by Lee and Schiffler [2015] when the exchange matrix B is skew-symmetric and by Gross, Hacking, Keel, and Kontsevich [Gross et al. 2018] when B is more generally skew-symmetrizable.

Theorem 3.8 (positivity). *In a generalized cluster algebra, each of the coefficients in the Laurent polynomial corresponding to any cluster variable from Theorem 3.7 is a nonnegative integer linear combination of elements in $\mathbb{P}$. In other words, any cluster variable is an element in $\mathbb{NP}[x_1^{\pm}, \dots, x_n^{\pm}]$.*

Proof. By the separation formula Theorem 3.20 and Remark 3.21, we only need to show the positivity in the principal coefficients case (to be defined in Definition 3.13). In this case, we prove the positivity in Theorem 8.12. □

Remark 3.9. Chekhov and Shapiro [2014, Conjecture 5.1] conjectured a positivity for generalized cluster algebras under a reciprocal condition; see also the formulation in Conjecture 8.13. In the reciprocal case, this positivity implies Theorem 3.8. We do not know how to show this stronger positivity in general; see the discussion in Section 8.5.

3.2. ***Relation to Chekhov and Shapiro's definition.*** Chekhov and Shapiro [2014] defined generalized cluster algebras by considering more general exchange polynomials. Precisely, a labeled seed in that setting is a triple $(\boldsymbol{x}, (\overline{\alpha_i} \mid i \in I), B)$, where $\boldsymbol{x}$ and B are the same as in Definition 3.1 and $\bar{\alpha}_i = (\alpha_{i,k} \in \mathbb{P} \mid 0 \le k \le r_i)$ for $i \in I$. Here we only take $\mathbb{P}$ as a multiplicative abelian group. The coefficients $\alpha_{i,k}$ are responsible for expressing the exchange polynomial defined as

$$\theta_i(u, v) := \sum_{k=0}^{r_i} \alpha_{i,k} u^{r_i - k} v^k \in \mathbb{ZP}[u, v].$$

The mutation $(\boldsymbol{x}', (\bar{\alpha}_i'), B') = \mu_k(\boldsymbol{x}, (\overline{\alpha_i}), B)$ is defined in the following way:

(1) $B' = \mu_k(B)$.

(2) $\alpha'_{k,j} = \alpha_{k,r_k - j}$ and for $i \neq k$, the coefficients satisfy

$$\alpha'_{i,j}/\alpha'_{i,0} = \begin{cases} \alpha_{k,0}^{j\beta_{ki}} \cdot \alpha_{i,j}/\alpha_{i,0} & \text{if } \beta_{ik} > 0 \\ \alpha_{k,r_k}^{j\beta_{ki}} \cdot \alpha_{i,j}/\alpha_{i,0} & \text{if } \beta_{ik} \le 0. \end{cases}$$

(3) $x'_i = x_i$ for $i \neq k$ and

$$x'_k x_k = \theta_i(u_{k;+}, u_{k;-}).$$

Remark 3.10. In this setting, it does no harm to allow the coefficients $\alpha_{i,k}$ to be elements of $\mathbb{Z}\mathbb{P}$, as long as the above rule (2) is satisfied. For example, one may check that the Laurent phenomenon still holds for cluster variables.

Now assume the multiplicative abelian group $\mathbb{P}$ has an addition $\oplus$ so that it is a semifield. In our setting the exchange polynomials are given by $\theta[\boldsymbol{p}_i](u, v)$, thus the coefficients $\alpha_{i,j}$ corresponding to the coefficients of $\theta[\boldsymbol{p}_i](u, v)$ when expanded as polynomial of u and v. Under Definition 3.2, the polynomials $\theta[\boldsymbol{p}_i](u, v)$ mutate in a way satisfying the rule (2) above. In fact, when talking about coefficients $\alpha_{i,j}/\alpha_{i,0}$, we can normalize our polynomial

$$\tilde{\theta}[\boldsymbol{p}_i](u, v) = \prod_{j\in[1,r_i]} (p_{i,j}u + v).$$

So when expanded as a sum of monomials in u and v, the coefficients of $\tilde{\theta}[\boldsymbol{p}_i]$ are $\prod_{j\in J} p_{i,j}$ for a subset $J \subset [1, r_i]$. According to the mutation formula in Definition 3.2, under μ_k, we have

$$\prod_{j\in J} p'_{i,j} = p_{k;\pm}^{|J|\beta_{ki}} \prod_{j\in J} p_{i,j},$$

which satisfies the rule (2). So our definition of generalized cluster algebras can be viewed as a special case of [Chekhov and Shapiro 2014] if we ease the condition $\alpha_{i,k} \in \mathbb{P}$ into $\alpha_{i,k} \in \mathbb{Z}\mathbb{P}$.

We note that the above rule (2) in [Chekhov and Shapiro 2014] is not enough to uniquely determine the coefficients $(\bar{\alpha}'_i)$ after mutation, whereas the mutation formula in Definition 3.2 is deterministic because of the normalization condition $p^+_{i,j} \oplus p^-_{i,j} = 1$.

One advantage of our definition is the availability of principal coefficients analogous to [Fomin and Zelevinsky 2007, Definition 3.1], to be discussed in the next section.

3.3. ***Principal coefficients.*** As in [Fomin and Zelevinsky 2007] for ordinary cluster algebras, we have the notion of principal coefficients for generalized cluster algebras.

Definition 3.11. We say a generalized cluster algebra $\mathscr{A}$ is of geometric type if $\mathbb{P}$ is a tropical semifield

$$\operatorname{Trop}(s_a \mid a \in I'),$$

where I' is a finite index set.

Proposition 3.12. *Let $\mathscr{A}$ be a generalized cluster algebra of geometric type. For each seed Σ in $\mathscr{A}$'s cluster pattern and $i \in I$, we introduce matrices*

$$C^{(i)} = C^{(i)}_{\Sigma} = (c^{(i)}_{aj}) \in \operatorname{Mat}_{|I'|\times r_i}(\mathbb{Z})$$

to encode the coefficients $p_{i,j}$ by columns of $C^{(i)}$:

$$p_{i,j} = \prod_{a\in I'} s_a^{c^{(i)}_{aj}} \in \mathbb{P}.$$

Denote by $(\bar{c}_{aj}^{(i)})$ *the matrices corresponding to the seed* $\mu_k(\Sigma)$ *for some* $k \in I$. *Then we have*

$$\bar{c}_{aj}^{(i)} = \begin{cases} -c_{aj}^{(i)} & \textit{if } i = k; \\ c_{aj}^{(i)} + \Big(\sum_{j=1}^{r_k} [-c_{aj}^{(k)}]_+\Big)\beta_{ki} & \textit{if } i \neq k \textit{ and } \beta_{ik} > 0; \\ c_{aj}^{(i)} + \Big(\sum_{j=1}^{r_k} [c_{aj}^{(k)}]_+\Big)\beta_{ki} & \textit{if } i \neq k \textit{ and } \beta_{ik} \leq 0. \end{cases}$$

Proof. In the tropical semifield $\operatorname{Trop}(s_a \mid a \in I')$, we have the expressions

$$p_{i,j}^{+} = \prod_{a \in I'} s_a^{[c_{aj}^{(i)}]_+} \quad \text{and} \quad p_{i,j}^{-} = \prod_{a \in I'} s_a^{[-c_{aj}^{(i)}]_+}.$$

Then the result follows by spelling out the mutation formula of coefficients ((3) of Definition 3.2). □

The matrices and their dynamics in Proposition 3.12 have led to a remarkable combinatorial phenomenon in cluster theory known as *the sign coherence of c-vectors*. We shall explain it below.

Definition 3.13. We say a generalized cluster algebra $\mathscr{A}$ has *principal coefficients* at a vertex $t_0 \in \mathbb{T}_n$ if $\mathbb{P}$ is the tropical semifield

$$\operatorname{Trop}(\boldsymbol{p}) := \operatorname{Trop}(p_{i,j} \mid i \in I,\ j \in [1, r_i]),$$

and the seed Σ_{t_0} has coefficients $\boldsymbol{p}_i = (p_{i,1}, \dots p_{i,r_i})$. In this case, the cluster algebra, denoted as $\mathscr{A}^{\text{prin}}(\Sigma_{t_0})$, is by definition a subalgebra of

$$\mathbb{Z}[x_{i;t_0}^{\pm}, p_{i,j}^{\pm} \mid i \in I,\ j \in [1, r_i]].$$

In the case of principal coefficients, the columns of the matrices $C_{\Sigma_t}^{(i)}$ are called *generalized c-vectors*. At the initial seed $\Sigma = \Sigma_{t_0}$ with principal coefficients, each $C_{\Sigma}^{(i)}$ is an identity matrix I_{r_i} extended (vertically) by zeros.

Theorem 3.14 (sign coherence of generalized c-vectors). *In the principal coefficients case, for any* $t \in \mathbb{T}_n$, *for any* $i \in I$ *and any* $j \in [1, r_i]$, *the entries of the* j*-th column of* $C_{\Sigma_t}^{(i)}$ *are either all nonnegative or all nonpositive.*

When $r_i = 1$ for each $i \in I$, i.e., in the case of ordinary cluster algebras, each $C^{(i)} = C_{\Sigma_t}^{(i)}$ is just a column vector with n entries, altogether forming a matrix $C = (C^{(1)}, \dots, C^{(n)})$. They are the so-called *C-matrices* in [Fomin and Zelevinsky 2007] whose columns are *c-vectors*. In this case, Theorem 3.14 then says that each column of any C is either nonnegative or nonpositive. This is well-known in the theory of cluster algebras as *the sign coherence of c-vectors*, which has already been proved by Derksen, Weyman and Zelevinsky [Derksen et al. 2010] for skew-symmetric exchange matrices and by Gross, Hacking, Keel and Kontsevich [Gross et al. 2018] for skew-symmetrizable ones. We will see in Proposition 3.17 that Theorem 3.14 follows from the already established sign coherence of c-vectors.

We set the index set

$$I' = \bigsqcup_{i \in I} I_i', \quad I_i' := \{(i, j) \mid j \in [1, r_i]\}.$$

Lemma 3.15. *Let $\Sigma = \Sigma_{t_0}$ be a seed with principal coefficients. We have the following properties for the matrices $C^{(i)}_{\Sigma_t}$ for any seed Σ_t, $t \in \mathbb{T}_n$.*

(1) *Let $i, k \in I$ such that $k \neq i$. Then for any $a, a' \in I'_k$ and any $1 \leq j, j' \leq r_i$, we have*

$$c^{(i)}_{a,j} = c^{(i)}_{a',j'}.$$

(2) *Let $i \in I$. We have*

$$c^{(i)}_{(i,1),1} = c^{(i)}_{(i,2),2} = \cdots = c^{(i)}_{(i,r_i),r_i} = c \pm 1$$

and

$$c^{(i)}_{(i,k),j} = c \quad \text{for } k \neq j$$

for some integer c.

Proof. We prove this lemma by induction on the distance from t to t_0 in $\mathbb{T}_n$. The base case is for $C^{(i)}_{\Sigma}$ where the entries in (1) are all zeroes and the ones in (2) are 1 when $k = j$ and 0 otherwise. Then the properties stated in the lemma are preserved under the mutation formula given in Proposition 3.12. □

Let $\overline{\mathbb{P}}$ be the tropical semifield $\operatorname{Trop}(\bar{p}_i \mid i \in I)$. There is a group homomorphism

$$\pi : \mathbb{P} \to \overline{\mathbb{P}}, \quad p_{i,j} \mapsto \bar{p}_i.$$

For $t \in \mathbb{T}_n$, denote the image of $p_{i,j;t}$ in $\overline{\mathbb{P}}$ by $\bar{p}_{i;t}$ (which is independent of j by Lemma 3.15) and the image of $\prod_{j=1}^{r_i} p_{i,j;t}$ in $\overline{\mathbb{P}}$ by $p_{i;t} = \bar{p}^{r_i}_{i;t}$.

Lemma 3.16. *The elements $p_{i;t}$ behave in the following way under the mutation μ_k. If $t' \overset{k}{\text{---}} t$ and we write $p'_i = p_{i;t'}$ and $p_i = p_{i;t}$, then we have*

$$p'_i = \begin{cases} p_i^{-1} & \text{if } i = k; \\ p_i \cdot (p_k^-)^{b_{ki}} & \text{if } i \neq k \text{ and } \beta_{ik} > 0; \\ p_i \cdot (p_k^+)^{b_{ki}} & \text{if } i \neq k \text{ and } \beta_{ik} \leq 0. \end{cases}$$

So they behave under mutations in the same way as $p_{i,1;t}$ in the case where $r_i = 1$, $i \in I$, i.e., the case of ordinary cluster algebras.

Proof. By the generalized mutation formula of coefficients, we have

$$\prod_{j=1}^{r_i} p'_{i,j} = \begin{cases} \displaystyle\prod_{j=1}^{r_i} p^{-1}_{i,j} & \text{if } i = k; \\ \displaystyle\prod_{j=1}^{r_i} p_{i,j} \cdot \Big(\prod_{j=1}^{r_k} p^-_{k,j}\Big)^{b_{ki}} & \text{if } i \neq k \text{ and } \beta_{ik} > 0; \\ \displaystyle\prod_{j=1}^{r_i} p_{i,j} \cdot \Big(\prod_{j=1}^{r_k} p^+_{k,j}\Big)^{b_{ki}} & \text{if } i \neq k \text{ and } \beta_{ik} \leq 0. \end{cases}$$

By the matrix description of the elements $p_{k,j}$ in Lemma 3.15, we have that

$$\prod_{j=1}^{r_k} p^{\pm}_{k,j} = \Big(\prod_{j=1}^{r_k} p_{k,j}\Big)^{\pm} \in \mathbb{P}, \quad \pi\Big(\prod_{j=1}^{r_k} p^{\pm}_{k,j}\Big) = p^{\pm}_k \in \overline{\mathbb{P}}.$$

The result then follows. □

Proposition 3.17. *The sign coherence of c-vectors implies the sign coherence of generalized c-vectors.*

Proof. In the case where all $r_i = 1$, the sign coherence then says each column of the matrix $C = (C^{(1)}, \dots, C^{(n)})$ is either nonnegative or nonpositive.

On the other hand, by Lemma 3.16, the elements p_i behave under mutations in the exact same way as the coefficients in seeds when all $r_i = 1$ (thus we only have one p_i for each i). Thus the column $C^{(i)}$ of Σ_t serves as the coordinates of $p_{i;t}$ in terms of the initial coefficients p_i. Then the sign coherence tells that one of p_i^+ and p_i^- is 1. It then follows from Lemma 3.15 that the corresponding $p_{i,j}^+$ or $p_{i,j}^-$ for each $j \in [1, r_i]$ is also 1, hence the generalized sign coherence. □

The following lemma will be useful later.

Lemma 3.18. *In the principal coefficient case, for any $t \in \mathbb{T}_n$, the set of coefficients in seed Σ_t*

$$\{p_{i,j;t} \mid i \in I,\ j \in [1, r_i]\}$$

form a $\mathbb{Z}$-basis of $\mathbb{P} \cong \mathbb{Z}^d$ where $d = \sum_{i \in I} r_i$.

Proof. This follows directly from the mutation formula Proposition 3.12 and Lemma 3.15. □

3.4. ***Separation formula.*** In this section, we describe the separation formula for generalized cluster variables, which can be derived in the exact same way as [Fomin and Zelevinsky 2007, Theorem 3.7], so we omit the proof.

Definition 3.19. Let $\mathscr{A}^{\mathrm{prin}}(\Sigma_{t_0})$ be a generalized cluster algebra with principal coefficients at $\Sigma_{t_0} = (\boldsymbol{x}, \boldsymbol{p}, B)$. We define the rational function

$$X_{l;t} \in \mathbb{Q}_{\mathrm{sf}}(\boldsymbol{x}, \boldsymbol{p})$$

corresponding to the subtraction-free rational expression of the cluster variable $x_{l;t}$ by iterating exchange relations. Here $(\boldsymbol{x}, \boldsymbol{p})$ denote the set of all variables in $\boldsymbol{x}$ and $\boldsymbol{p}$.

Define the rational function

$$F_{l;t}(\boldsymbol{p}) = X_{l;t}((1, \dots, 1), \boldsymbol{p}) \in \mathbb{Q}_{\mathrm{sf}}(\boldsymbol{p}).$$

In general, for a subtraction free expression F in $\mathbb{Q}_{\mathrm{sf}}(x_1, \dots, x_n)$ and an arbitrary semifield $\mathbb{P}$, we use the notation

$$F\ |_{\mathbb{P}}\ (y_1, \dots y_n) \in \mathbb{P}$$

for the evaluation at $x_i = y_i$. This evaluation is well-defined (i.e., independent of the expression used) because of the universal property of the semifield $\mathbb{Q}_{\mathrm{sf}}(x_1, \dots, x_n)$; see Section 2.1.

Theorem 3.20 (cf. [Fomin and Zelevinsky 2007, Proposition 3.6, Theorem 3.7]).

(1) *We have*

$$X_{l;t} \in \mathbb{Z}[x_i^{\pm}; p_{i,j} \mid i \in I,\ j \in [1, r_i]], \quad F_{l;t} \in \mathbb{Z}[p_{i,j} \mid i \in I,\ j \in [1, r_i]].$$

(2) *Let $\mathscr{A}$ be a generalized cluster algebra over an arbitrary semifield $\mathbb{P}'$, with an initial seed $\Sigma_{t_0} = (\boldsymbol{x}, \boldsymbol{p}, B)$. Then the cluster variables in $\mathscr{A}$ can be expressed in the initial cluster as*

$$x_{l;t} = \frac{X_{l;t} \mid_{\mathcal{F}} (\boldsymbol{x}, \boldsymbol{p})}{F_{l;t} \mid_{\mathbb{P}'} (\boldsymbol{p})},$$

where $\mathcal{F}$ is the ambient field for $\mathscr{A}$.

Remark 3.21. Suppose the positivity for $x_{l;t}$ in the principal coefficients case (where we denote the semifield by $\mathbb{P}$) has been established. This means $X_{l;t}$ has a subtraction free expression as a Laurent polynomial (i.e., whose coefficients are in $\mathbb{NP}$). Then the evaluation $X_{l;t} \mid_{\mathcal{F}} (\boldsymbol{x}, \boldsymbol{p})$ also has positive coefficients in $\mathbb{NP}'$, while $F_{l;t} \mid_{\mathbb{P}'} (\boldsymbol{p})$ is an element in $\mathbb{P}'$. Thus it follows by Theorem 3.20 that $x_{l;t}$ has positive coefficients in the case of arbitrary $\mathbb{P}'$.

3.5. *Cluster algebras with specialized coefficients.* We fix a field $\Bbbk$ of characteristic 0 and consider the case of geometric coefficients. In this case, the generalized cluster algebra $\mathscr{A}(\Sigma)$ for $\Sigma = (\boldsymbol{x}, \boldsymbol{p}, B)$ can be viewed as a subring of $\Bbbk\mathbb{P}[x_1^{\pm}, \dots, x_n^{\pm}]$ where $\Bbbk\mathbb{P}$ is the group algebra of $\mathbb{P}$ over $\Bbbk$.

Let $\lambda : \mathbb{P} \to \Bbbk^*$ be a group homomorphism (which we will later refer to as an *evaluation*). It extends to a $\Bbbk$-algebra homomorphism

$$\lambda : \Bbbk\mathbb{P}[x_1^{\pm}, \dots, x_n^{\pm}] \to \Bbbk[x_1^{\pm}, \dots, x_n^{\pm}].$$

We denote the image of $\mathscr{A}(\Sigma) \otimes \Bbbk$ by $\mathscr{A}(\Sigma, \lambda)$. So we have a family of $\Bbbk$-algebras parametrized by $(\Bbbk^*)^l$ if the free abelian group $\mathbb{P}$ is of rank l. Each $\mathscr{A}(\Sigma, \lambda)$ is in fact the $\Bbbk$-subalgebra generated by cluster variables (with coefficients specialized by λ) within $\Bbbk[x_1^{\pm}, \dots, x_n^{\pm}]$. These are what we call (generalized) *cluster algebras with specialized coefficients*.

We point out that an ordinary cluster algebra with trivial coefficients (i.e., when $\mathbb{P}$ is trivial) is actually a generalized cluster algebra with specialized coefficients. Suppose B is a skew-symmetrizable matrix and let r_i be the gcd of the i-th column (if that column is nonzero). Let $\mathscr{A}^{\mathrm{prin}}(\Sigma)$ be the generalized cluster algebra with principal coefficients where Σ has exchange matrix B. Choose a group homomorphism $\lambda : \mathrm{Trop}(\boldsymbol{p}) \to \Bbbk^*$ such that the specialized exchange polynomials equals the usual cluster exchange binomial, i.e.,

$$\prod_{j=1}^{r_i} (\lambda(p_{i,j})u + v) = u^{r_i} + v^{r_i}.$$

Of course such λ always exists assuming $\Bbbk$ is algebraically closed. Then it is easy to check that every generalized mutation becomes an ordinary mutation: if $t \overset{k}{\frac{\quad}{}} t'$,

$$x_{k;t'} = x_{k;t}^{-1}\Big(\prod_{i \in I} x_i^{[b_{ik}^t]_+} + \prod_{i \in I} x_i^{[-b_{ik}^t]_+}\Big).$$

Thus the algebra $\mathscr{A}^{\mathrm{prin}}(\Sigma, \lambda)$ has the exact same cluster variables as the ordinary cluster algebra with trivial coefficients, and can thus be viewed as an ordinary cluster algebra.

t	B^t	$p_{1,1;t}$	$p_{2,1;t}$	$p_{2,2;t}$	$x_{1;t}$	$x_{2;t}$
0	$\left[\begin{smallmatrix}0&-2\\1&0\end{smallmatrix}\right]$	t_{11}	t_{21}	t_{22}	A_1	A_2
1	$\left[\begin{smallmatrix}0&2\\-1&0\end{smallmatrix}\right]$	t_{11}^{-1}	t_{21}	t_{22}	$A_1^{-1}(1+t_{11}A_2)$	A_2
2	$\left[\begin{smallmatrix}0&-2\\1&0\end{smallmatrix}\right]$	t_{11}^{-1}	t_{21}^{-1}	t_{22}^{-1}	$A_1^{-1}(1+t_{11}A_2)$	$A_2^{-1}\big(1+t_{21}A_1^{-1}(1+t_{11}A_2)\big)$ $\times\big(1+t_{22}A_1^{-1}(1+t_{11}A_2)\big)$
3	$\left[\begin{smallmatrix}0&2\\-1&0\end{smallmatrix}\right]$	t_{11}	$t_{11}^{-1}t_{21}^{-1}$	$t_{11}^{-1}t_{22}^{-1}$	$A_1A_2^{-1}\big((1+t_{21}A_1^{-1})(1+t_{22}A_1^{-1})$ $+t_{11}t_{21}t_{22}A_1^{-2}A_2\big)$	$A_2^{-1}\big(1+t_{21}A_1^{-1}(1+t_{11}A_2)\big)$ $\times\big(1+t_{22}A_1^{-1}(1+t_{11}A_2)\big)$
4	$\left[\begin{smallmatrix}0&-2\\1&0\end{smallmatrix}\right]$	$t_{11}^{-1}t_{21}^{-1}t_{22}^{-1}$	$t_{11}t_{21}$	$t_{11}t_{22}$	$A_1A_2^{-1}\big((1+t_{21}A_1^{-1})(1+t_{22}A_1^{-1})$ $+t_{11}t_{21}t_{22}A_1^{-2}A_2\big)$	$A_2^{-1}(t_{21}+A_1)(t_{22}+A_1)$
5	$\left[\begin{smallmatrix}0&2\\-1&0\end{smallmatrix}\right]$	$t_{11}t_{21}t_{22}$	t_{22}^{-1}	t_{21}^{-1}	A_1	$A_2^{-1}(t_{21}+A_1)(t_{22}+A_1)$
6	$\left[\begin{smallmatrix}0&-2\\1&0\end{smallmatrix}\right]$	t_{11}	t_{22}	t_{21}	A_1	A_2

Table 1. Labeled seeds of $\mathscr{A}^{\text{prin}}$.

3.6. ***An example in type B_2 with principal coefficients.*** We consider $\mathscr{A}^{\text{prin}}(\boldsymbol{x}, \boldsymbol{p}, B)$ with principal coefficients for $B = \left[\begin{smallmatrix}0&-2\\1&0\end{smallmatrix}\right]$ which is of type B_2 in the finite type classification [Fomin and Zelevinsky 2003; Chekhov and Shapiro 2014, Theorem 2.7]. We write $x_{i;t_0} = A_i$, and $p_{i,j;t_0} = t_{ij}$. For the subgraph

$$t_0 \overset{1}{\text{—}} t_1 \overset{2}{\text{—}} t_2 \overset{1}{\text{—}} t_3 \overset{2}{\text{—}} t_4 \overset{1}{\text{—}} t_5 \overset{2}{\text{—}} t_6$$

of $\mathbb{T}_2$, we have the associated labeled seeds calculated in Table 1

We note that the Σ_{t_6} is not exactly the same as the Σ_{t_0} but up to a switch of $p_{2,1}$ and $p_{2,2}$.

3.7. ***Generalized Y-seeds.*** We define generalized Y-seeds (with coefficients) and their mutations. The formulation to including coefficients in Y-seeds comes from [Bossinger et al. 2020]. The following definition is a generalization of [Bossinger et al. 2020, Definition 2.15], which is an enhancement of a Y-seed of [Fomin and Zelevinsky 2007].

Definition 3.22. A *generalized labeled Y-seed* (with coefficients) Δ is a triple $(\boldsymbol{y}, \boldsymbol{q}, B)$, where

- $\boldsymbol{q} = (\boldsymbol{q}_1, \dots, \boldsymbol{q}_n)$ is an n-tuple of r_i-tuples $\boldsymbol{q}_i = (q_{i,1}, \dots q_{i,r_i})$ of elements in a semifield $\mathbb{P}$ for positive integers r_i, $1 \le i \le n$;
- $\boldsymbol{y} = \{y_1, \dots, y_n\}$ is a collection of elements in some universal semifield $\mathbb{QP}_{\text{sf}}(u_1, \dots, u_l)$;
- B is a left skew-symmetrizable integer matrix such that the i-th column is divisible by r_i for every i.

Definition 3.23. For $k \in \{1, \dots, n\}$, we define the *mutation* of a Y-seed $(\boldsymbol{y}, \boldsymbol{q}, B)$ in direction k as a new Y-seed $\mu_k(\boldsymbol{y}, \boldsymbol{q}, B) := ((y'_i), (\boldsymbol{q}'_i), B')$ in the following way:

$$B' = \mu_k(B); \tag{3-1}$$

$$q'_{k,j} = q_{k,j}^{-1} \quad \text{for } j \in [1, r_k];$$

$$\text{for } i \neq k,\ j \in [1, r_i], \quad q'_{i,j} = \begin{cases} q_{i,j} \cdot \Bigl(\prod_{l=1}^{r_k} q_{k,l}^{-}\Bigr)^{-\beta_{ik}} & \text{if } -\beta_{ki} > 0, \\ q_{i,j} \cdot \Bigl(\prod_{l=1}^{r_k} q_{k,l}^{+}\Bigr)^{-\beta_{ik}} & \text{if } -\beta_{ki} \leq 0, \end{cases} \tag{3-2}$$

or equivalently

$$\text{for } i \neq k, \quad q'_{i,j} = q_{i,j} \prod_{l=1}^{r_k} (1 \oplus q_{k,l}^{\mathrm{sgn}(-\beta_{ki})})^{\beta_{ik}};$$

$$y'_i = \begin{cases} y_i \prod_{l=1}^{r_k} (q_{k,l}^{\mathrm{sgn}(\beta_{ik})} y_k^{\mathrm{sgn}(\beta_{ik})} + q_{k,l}^{\mathrm{sgn}(-\beta_{ik})})^{\beta_{ik}} & \text{if } i \neq k, \\ y_k^{-1} & \text{if } i = k. \end{cases} \tag{3-3}$$

As in Lemma 3.3, it is straightforward to check that the mutation μ_k on a generalized Y-seed is involutive in the same direction.

Definition 3.24. A *generalized Y-pattern* is an association

$$t \mapsto \Delta_t = (\boldsymbol{y}_t, \boldsymbol{q}_t, B^t)$$

to every vertex $t \in \mathbb{T}_n$ a generalized labeled Y-seed Δ_t such that if t and t' are connected by an edge labeled by $k \in I$, then we have

$$\Delta_{t'} = \mu_k(\Delta_t).$$

Definition 3.25. We say that a generalized Y-pattern has principal coefficients at a vertex $t_0 \in \mathbb{T}_n$ if $\mathbb{P}$ is the tropical semifield

$$\mathrm{Trop}(q_{i,j;t_0} \mid i \in I,\ j \in [1, r_i]).$$

Given a Y-pattern, the elements $y_{i;t}$ for $t \in \mathbb{T}_n$ are called *Y-variables*.

Remark 3.26. In the case that for any $i \in I$,

$$q_{i,1} = q_{i,2} = \cdots = q_{i,r_i},$$

a generalized Y-seed with coefficients as in Definition 3.22 becomes a labeled Y-seed with coefficients in [Bossinger et al. 2020]. In this case, the mutation formula of Y-variables is independent of the choice r_i. So we get back to the nongeneralized version by letting the coefficients $q_{i,j}$, $j \in [1, r_i]$, equal. While in the cluster case, one recovers the nongeneralized seed mutation by choosing $r_i = 1$. This asymmetry suggests that our generalization is a natural one.

To the best knowledge of the author, the generalized version of Y-patterns has not been considered in the literature. It is interesting to see if these generalized patterns appear naturally anywhere.

4. Generalized cluster varieties

Cluster varieties were introduced by Fock and Goncharov [2009], giving a geometric view to cluster algebras (of geometric types). We follow [Gross et al. 2015] to define relevant notions such as fixed data and seeds. However, in order to deal with generalized coefficients, some new gadgets are needed.

Definition 4.1. We recall the *fixed data* Γ from [Gross et al. 2015]. The fixed data Γ consists of

- a lattice N of finite rank with a skew-symmetric bilinear form $\omega : N \times N \to \mathbb{Q}$;
- an *unfrozen sublattice* $N_{\mathrm{uf}} \subset N$, a saturated sublattice of N;
- an index set $I = \{1, \dots, \operatorname{rank} N\}$ and a subset $I_{\mathrm{uf}} = \{1, \dots, \operatorname{rank} N_{\mathrm{uf}}\}$;
- positive integers d_i for $i \in I$ with greatest common divisor 1;
- a sublattice $N^\circ \subset N$ of finite index such that $\omega(N_{\mathrm{uf}}, N^\circ) \subset \mathbb{Z}$, $\omega(N, N_{\mathrm{uf}} \cap N^\circ) \subset \mathbb{Z}$;
- $M = \operatorname{Hom}(N, \mathbb{Z})$, $M^\circ = \operatorname{Hom}(N^\circ, \mathbb{Z})$;

4.1. *Generalized $\mathcal{A}$-cluster variety.*

Definition 4.2. Given fixed data Γ, an $\mathcal{A}$-*seed with* (*generalized*) *coefficients* is a pair $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{p})$ consisting of a *seed* $\boldsymbol{e} = (e_i)_{i \in I}$ which is a labeled collection of elements in N and a labeled collection of tuples of coefficients $\boldsymbol{p} = (\boldsymbol{p}_i)_{i \in I_{\mathrm{uf}}}$, where $\boldsymbol{p}_i = (p_{i,j})_{j \in [1, r_j]}$ and $p_{i,j}$ belongs to some tropical semifield $\mathbb{P}$ such that

(1) $\{e_i \mid i \in I\}$ is a basis for N;

(2) $\{e_i \mid i \in I_{\mathrm{uf}}\}$ is a basis for N_{uf};

(3) $\{d_i e_i \mid i \in I\}$ is a basis for N°;

(4) for $i \in I_{\mathrm{uf}}$, the elements $w_i := \omega(-, d_i e_i)/r_i$ belong to M.

For such a seed $\boldsymbol{s}$, we define two matrices $B = B(\boldsymbol{s}) = (b_{ij})$ and $\tilde{B} = \tilde{B}(\boldsymbol{s}) = (\beta_{ij})$ by setting

$$b_{ij} := \omega(e_i, d_j e_j) \quad \text{and} \quad \beta_{ij} := \langle e_i, w_j \rangle = b_{ij}/r_j.$$

Definition 4.3. Given $\boldsymbol{s}$ an $\mathcal{A}$-seed with coefficients, for $k \in I_{\mathrm{uf}}$, we define the *mutation in direction* k, $\mu_k(\boldsymbol{s}) = (\boldsymbol{e}', \boldsymbol{p}')$ by

$$e'_i = \begin{cases} -e_k & \text{if } i = k, \\ e_i + [\langle e_i, -r_k w_k \rangle]_+ e_k & \text{if } i \neq k; \end{cases}$$

and

$$p'_{k,j} = p_{k,j}^{-1} \quad \text{for } j \in [1, r_k];$$

$$\text{for } i \neq k,\ j \in [1, r_i], \quad p'_{i,j} = \begin{cases} p_{i,j} \cdot (p_{k;-})^{\beta_{ki}} & \text{if } \beta_{ik} > 0, \\ p_{i,j} \cdot (p_{k;+})^{\beta_{ki}} & \text{if } \beta_{ik} \leq 0. \end{cases}$$

Remark 4.4. If we write $w_i' = \omega\big(-, \frac{d_i}{r_i} e_i'\big)$ as the mutations of w_i, then they are given by

$$w_i' = \begin{cases} -w_k, & \text{if } i = k; \\ w_i + [\langle r_k e_k, w_i\rangle]_+ w_k, & \text{if } i \neq k. \end{cases}$$

Denote the dual basis of (e_i) by (e_i^*) and the dual of $(e_i') = \mu_k(\boldsymbol{e})$ by $(e_i^{\prime,*})$. We have

$$e_i^{\prime,*} = \begin{cases} -e_k^* + \sum\limits_j [\langle e_j, -r_k w_k\rangle]_+ e_j^* & \text{if } i = k; \\ e_i^* & \text{if } i \neq k. \end{cases}$$

If there is no confusion, we will call an $\mathcal{A}$-seed with coefficients simply a seed.

Let $R = \Bbbk\mathbb{P}$, the group algebra of $\mathbb{P}$ over the ground field $\Bbbk$. To any $\mathcal{A}$-seed $\boldsymbol{s}$, we associate a copy of the R-torus $T_{N,\boldsymbol{s}}(R) := \operatorname{Spec}(\Bbbk[M] \otimes_{\Bbbk} R)$.

Definition 4.5. To the mutation μ_k from $\boldsymbol{s}$ to $\mu_k(\boldsymbol{s})$, there is an associated birational morphism (over R)

$$\mu_k : T_{N,\boldsymbol{s}}(R) \dashrightarrow T_{N,\mu_k(\boldsymbol{s})}(R), \quad \mu_k^*(z^m) = z^m f_k^{-\langle e_k, m\rangle},$$

where

$$f_k := \prod_{j=1}^{r_k} (p_{k,j}^- + p_{k,j}^+ z^{w_k}) \in R[M].$$

We call this birational transformation the *$\mathcal{A}$-cluster mutation* associated to the mutation μ_k of seeds.

Definition 4.6. We define the *oriented rooted tree* $\mathfrak{T}_n$ (where $n = |I_{\mathrm{uf}}|$) as in [Gross et al. 2015]. It is the infinite tree generated from a root v_0 such that

(1) v_0 has outgoing edges labeled by $I_{\mathrm{uf}} = \{1, \dots, n\}$;

(2) any other vertex has one unique incoming edge, and outgoing edges labeled by I_{uf}.

Let $v_0 \in \mathfrak{T}_n$ be the root. Then for any other vertex $v \in \mathfrak{T}_n$, there is a unique oriented path from v_0 to v. We associate a seed $\boldsymbol{s}$ to the root v_0, the unique path from v to v_0 determines a seed $\boldsymbol{s}_v$ by applying the mutations in directions of the labelings in the path to the initial seed $\boldsymbol{s}$. Therefore we have an association $v \mapsto \boldsymbol{s}_v$ for $v \in \mathfrak{T}_n \setminus \{v_0\}$ and $v_0 \mapsto \boldsymbol{s}$ such that for an edge $v \xrightarrow{k} v'$ in $\mathfrak{T}_n$, then

$$\boldsymbol{s}_{v'} = \mu_k(\boldsymbol{s}_v).$$

Suppose the unique path from v_0 to v walks through edges labeled by $k_1, k_2, \dots, k_l$. There is then the birational map

$$\mu_{v_0,v} := \mu_{k_l} \circ \cdots \circ \mu_{k_2} \circ \mu_{k_1} : T_{N,\boldsymbol{s}}(R) \dashrightarrow T_{N,\boldsymbol{s}_v}(R).$$

For arbitrary two vertices v and v' in $\mathfrak{T}_n$, there is the birational map

$$\mu_{v,v'} := \mu_{v_0,v'} \circ \mu_{v_0,v}^{-1} : T_{N,\boldsymbol{s}_v}(R) \dashrightarrow T_{N,\boldsymbol{s}_{v'}}(R).$$

These birational maps surely satisfy the cocycle condition. We use the following lemma to glue $T_{N,\boldsymbol{s}_v}$ together.

Lemma 4.7 [Bossinger et al. 2020, Lemma 3.10; Gross et al. 2015, Proposition 2.4]. *Let $\mathcal{I}$ be a set and $\{S_i \mid i \in I\}$ be a collection of integral separated schemes of finite type over a locally Noetherian ring R, with birational maps (of R-schemes) $f_{ij} : S_i \dashrightarrow S_j$ for all i, j, verifying the cocycle condition $f_{jk} \circ f_{ij} = f_{ik}$ as rational maps and such that f_{ii} is the identity map. Let $U_{ij} \subset S_i$ be the largest open subscheme such that $f_{ij} : U_{ij} \to f_{ij}(U_{ij})$ is an isomorphism. Then there is an R-scheme*

$$S = \bigcup_{i \in \mathcal{I}} S_i$$

obtained by gluing the S_i along the open sets U_{ij} via the maps f_{ij}.

Definition 4.8. Let Γ be fixed data and $\boldsymbol{s}$ be an $\mathcal{A}$-seed with coefficients. We apply Lemma 4.7 to glue together the collection of tori indexed by $\mathfrak{T}_n$ to get the *generalized $\mathcal{A}$-cluster variety* associated to $\boldsymbol{s}$ (as an R-scheme)

$$\mathcal{A}_{\boldsymbol{s}} = \mathcal{A}_{\Gamma,\boldsymbol{s}} := \bigcup_{v \in \mathfrak{T}} T_{N,\boldsymbol{s}_v}(R).$$

We now explain how to obtain a generalized cluster pattern from $\mathcal{A}_{\boldsymbol{s}}$, justifying the name generalized $\mathcal{A}$-cluster variety. We assume $N_{\mathrm{uf}} = N$, thus $I_{\mathrm{uf}} = I$.[1]

Recall we have the association $v \mapsto \boldsymbol{s}_v = \mu_{v_0,v}(\boldsymbol{s})$ for $v \in \mathfrak{T}_n$. We write $\boldsymbol{s}_v = (\boldsymbol{e}_v, \boldsymbol{p}_v)$ where $\boldsymbol{e}_v = (e_{i;v} \mid i \in I)$, $\boldsymbol{p}_v = (\boldsymbol{p}_{i;v} \mid i \in I)$ and $\boldsymbol{p}_{i;v} = (p_{i,j;v} \mid j \in [1, r_i])$.

Sending v_0 to any vertex t_0 in the n-regular tree $\mathbb{T}_n$ gives a unique surjective map

$$\pi : \mathfrak{T}_n \to \mathbb{T}_n, \quad v_0 \mapsto t_0$$

such that the labeling on edges is preserved.

For any seed $v \in \mathfrak{T}_n$, there is the corresponding labeled seed with coefficients (in the sense of Definition 3.1)

$$\Sigma_v = \Sigma(\boldsymbol{s}_v) := (\boldsymbol{x}_v, \boldsymbol{p}_v, B^v),$$

where

$$x_{i,v} := \mu^*_{v_0,v}(z^{e^*_{i;v}}) \in \mathbb{QP}(x_1, \dots, x_n), \quad b^v_{ij} := \omega(e_{i;v}, d_j e_{j;v}),$$

where $x_i = x_{i,v_0}$.

Lemma 4.9. *If two vertices v and v' vertices of $\mathfrak{T}_n$ descend to the same vertex in $\mathbb{T}_n$, i.e., $\pi(v) = \pi(v')$, then their corresponding labeled seeds with coefficients are identical, i.e., $\Sigma_v = \Sigma_{v'}$.*

Proof. Suppose the unique path in $\mathfrak{T}_n$ from v_0 to v goes through edges labeled by $k_1, \dots, k_l$ in order. We show in the following by induction that

$$\mu_{k_l} \circ \cdots \circ \mu_{k_1}(\Sigma_{v_0}) = \Sigma_v,$$

where the operation μ_k is the mutation in direction k of labeled seeds with coefficients in the sense of Definition 3.2.

[1] This is because we do not define cluster patterns with frozen directions. This can be done by making mutations only available at a subset of a given cluster, leaving the rest variables frozen. However, one can always treat the frozen variables as making up coefficients in a cluster pattern.

Let $v_1 \xrightarrow{k} v_2$ be in $\mathfrak{T}_n$. Then one checks $B^{v_2} = \mu_k(B^{v_1})$ using the fact that $\boldsymbol{e}_{v_2} = \mu_k(\boldsymbol{e}_{v_1})$, which is standard from [Gross et al. 2015]. The coefficients parts $\boldsymbol{p}_{v_1}$ and $\boldsymbol{p}_{v_2}$ are related by the mutation μ_k by definition. So we only need to check that $\boldsymbol{x}_{v_1}$ and $\boldsymbol{x}_{v_2}$ are also related by μ_k.

Note that $\mu^*_{v_0,v_2} = \mu^*_{v_0,v_1} \circ \mu^*_k$. So we have for $i \neq k$

$$\begin{aligned}
x_{i;v_2} &= \mu^*_{v_0,v_1}(\mu^*_k(z^{e^*_{i;v_2}})) = \mu^*_{v_0,v_1}(z^{e^*_{i;v_1}}) = x_{i;v_1},\\
x_{k;v_2} &= \mu^*_{v_0,v_1}(\mu^*_k(z^{e^*_{k;v_2}}))\\
&= \mu^*_{v_0,v_1}\Big(z^{-e^*_{k;v_1}+\sum[-b^{v_1}_{ik}]_+e^*_{i;v_1}} \prod_{j=1}^{r_k}(p^-_{k,j;v_1} + p^+_{k,j;v_1}z^{w_{k;v_1}})\Big)\\
&= \mu^*_{v_0,v_1}\Big(z^{-e^*_{k;v_1}} \prod_{j=1}^{r_k}(p^-_{k,j;v_1}z^{w^-_{k;v_1}} + p^+_{k,j;v_1}z^{w^+_{k;v_1}})\Big)\\
&= \mu^*_{v_0,v_1}(z^{-e^*_{k;v_1}}) \prod_{j=1}^{r_k}\big(p^-_{k,j;v_1}\mu^*_{v_0,v_1}(z^{w^-_{k;v_1}}) + p^+_{k,j;v_1}\mu^*_{v_0,v_1}(z^{w^+_{k;v_1}})\big)\\
&= x^{-1}_{k;v_1} \prod_{j=1}^{r_k}\Big(p^-_{k,j;v_1}\prod_{i\in I}x^{[-\beta_{ik}]_+}_{i;v_1} + p^+_{k,j;v_1}\prod_{i\in I}x^{[\beta_{ik}]_+}_{i;v_1}\Big).
\end{aligned}$$

The only unexplained notation in the above equations is that for any $w = \sum_{i\in I} a_i e^*_i \in M$, we write

$$w^- := \sum_{i\in I}[-a_i]_+e^*_i \quad \text{and} \quad w^+ := \sum_{i\in I}[a_i]_+e^*_i.$$

Now we have checked that $\mu_k(\Sigma_{v_1}) = \Sigma_{v_2}$. By induction on the distance from v to the root v_0, we conclude that $\mu_{k_l} \circ \cdots \circ \mu_{k_1}(\Sigma_{v_0}) = \Sigma_v$ for any $v \in \mathfrak{T}_n$. Since μ_k is involutive, we can reduce the sequence $(k_1, \cdots, k_l)$ by deleting pairs of consecutive identical indices until there is none. So Σ_v only depends on the reduced sequence of edge labels from v_0 to v. Now notice that two vertices v and v' in $\mathfrak{T}_n$ have the same projection t in $\mathbb{T}_n$ if and only if they have the same reduced sequence of edge labels from v_0, meaning the same labeled seed with coefficients $\Sigma_t := \Sigma_v = \Sigma_{v'}$. □

Proposition 4.10. *According to the above lemma, we have that the labeled seeds Σ_v and $\Sigma_{v'}$ are equal if $\pi(v) = \pi(v') = t \in \mathbb{T}_n$. So we can denote them all by Σ_t. The association $t \mapsto \Sigma_t$ for every $t \in \mathbb{T}_n$ is a cluster pattern.*

Proof. Suppose the unique path from t_0 to some $t \in \mathbb{T}_n$ walks through edges in order of $k_1, \dots, k_l$. Then already in the proof of the above lemma, we have

$$\Sigma_t = \mu_{k_l} \circ \cdots \circ \mu_{k_1}(\Sigma_{t_0}).$$

This association by definition gives a cluster pattern. □

Definition 4.11. The (generalized) *upper cluster algebra* $\overline{\mathscr{A}}(\boldsymbol{s})$ (of an $\mathcal{A}$-seed $\boldsymbol{s}$ with coefficients) is defined to be the R-algebra

$$H^0(\mathcal{A}_{\boldsymbol{s}}, \mathcal{O}_{\mathcal{A}_{\boldsymbol{s}}}) = \bigcap_{v\in\mathfrak{T}_n} H^0(T_{N,s_v}(R), \mathcal{O}_{T_{N,s_v}(R)}),$$

the ring of regular functions on the (generalized) $\mathcal{A}$-cluster variety $\mathcal{A}_{\boldsymbol{s}}$.

By definition the upper cluster algebra is the algebra of all Laurent polynomials that remains Laurent polynomials after an arbitrary sequence of mutations. It follows from the Laurent phenomenon that all cluster variables are elements in the upper cluster algebra, thus the inclusion

$$\mathscr{A}(\boldsymbol{s}) \subset \overline{\mathscr{A}}(\boldsymbol{s}),$$

where the former denotes the subalgebra generated by cluster variables, i.e., the cluster algebra (over R).

The notion of principal coefficients can be easily translated into the current setting.

Definition 4.12. An $\mathcal{A}$-seed $\boldsymbol{s}$ is said to have *principal coefficients* if the associated labeled seed $\Sigma(\boldsymbol{s})$ has principal coefficients.

The associated cluster pattern with $t_0 \mapsto \Sigma(\boldsymbol{s})$, $t \mapsto \Sigma(\boldsymbol{s}_v)$ (where $t = \pi(v)$) then has principal coefficients at t_0. In this case, we denote the corresponding cluster variety by $\mathcal{A}_{\boldsymbol{s}}^{\mathrm{prin}}$.

4.2. ***Generalized $\mathcal{X}$-cluster variety.*** Given fixed data Γ as in the last section, we define the notion of (generalized) $\mathcal{X}$-seeds with coefficients.

Definition 4.13. An $\mathcal{X}$-seed with (generalized) coefficients $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{q})$ is the same as an $\mathcal{A}$-seed. We use the symbol $\boldsymbol{q}$ instead of $\boldsymbol{p}$ to stress that it is an $\mathcal{X}$-seed.

What distinguish $\mathcal{X}$-seeds with $\mathcal{A}$-seeds is the mutation.

Definition 4.14. Given an $\mathcal{X}$-seed $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{q})$, we define the *mutation in direction k*, $\mu_k(\boldsymbol{s}) = (\boldsymbol{e}', \boldsymbol{q}')$ by

$$e_i' = \begin{cases} -e_k & \text{if } i = k, \\ e_i + [\langle e_i, -r_k w_k\rangle]_+ e_k & \text{if } i \neq k; \end{cases}$$

and

$$q_{k,j}' = q_{k,j}^{-1} \quad \text{for } j \in [1, r_k];$$

$$\text{for } i \neq k,\ j \in [1, r_i], \quad q_{i,j}' = \begin{cases} q_{i,j} \cdot (q_{k;-})^{-\beta_{ik}} & \text{if } -\beta_{ki} > 0, \\ q_{i,j} \cdot (q_{k;+})^{-\beta_{ik}} & \text{if } -\beta_{ki} \leq 0, \end{cases}$$

So the pure seed part $\boldsymbol{e}$ behaves in the same way under mutation as in an $\mathcal{A}$-seed while the coefficients part $\boldsymbol{q}$ mutates differently, but same as the coefficients in a labeled Y-seed. Roughly, if in $\mathcal{A}$-seeds, the matrix B governs the mutation of coefficients, then in $\mathcal{X}$-seeds, $-B^T$ does the job.

Definition 4.15. Let $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{q})$ be an $\mathcal{X}$-seed with coefficients. Then there is the associated *$\mathcal{X}$-cluster mutation*

$$\mu_k : T_M(R) \dashrightarrow T_M(R), \quad \mu_k^*(z^n) = z^n \cdot \Big(\prod_{l=1}^{r_k} (q_{k,l}^- + q_{k,l}^+ z^{e_k}) \Big)^{-\langle n, -w_k\rangle},$$

where $T_M(R)$ is the R-torus $\operatorname{Spec}(\Bbbk[N] \otimes R)$.

Definition 4.16. Let $\boldsymbol{s}$ be an $\mathcal{X}$-seed for Γ. Then there is a unique association $v \mapsto \boldsymbol{s}_v$ for every $v \in \mathfrak{T}_n$ such that $v_0 \mapsto \boldsymbol{s}$ and adjacent associated seeds are related by mutations of $\mathcal{X}$-seeds in corresponding directions. Define the *generalized $\mathcal{X}$-cluster variety* associated to $\boldsymbol{s}$ to be the R-scheme

$$\mathcal{X}_{\boldsymbol{s}} = \mathcal{X}_{\Gamma,\boldsymbol{s}} := \bigcup_{v \in \mathfrak{T}_n} T_{M,\boldsymbol{s}_v}(R)$$

obtained by gluing $T_{M,\boldsymbol{s}_v}(R)$ via $\mathcal{X}$-cluster mutations using Lemma 4.7.

Write $\boldsymbol{s}_v = ((e_{i;v}), (\boldsymbol{q}_{i;v}))$. Let us keep track of the monomials $z^{e_{i;v}} \in \Bbbk[N]$ (instead of $z^{e^*_{i;v}}$ in the $\mathcal{A}$-case). We define

$$y_{i;v} := \mu^*_{v,v_0}(z^{e_{i;v}}) \in \operatorname{Frac}(\Bbbk[N] \otimes R).$$

It turns out that these $y_{i;v}$ are the Y-variables of the Y-pattern induced by the $\mathcal{X}$-seed $\boldsymbol{s}$ described as follows. We take $\boldsymbol{s}$ as the initial seed. Analogous to the $\mathcal{A}$-situation, any vertex $v \in \mathfrak{T}_n$ descends to a vertex $t = \pi(v) \in \mathbb{T}_n$.

Proposition 4.17. *For $v \in \mathfrak{T}_n$, define the generalized labeled Y-seed $\Delta_v = ((y_{i;v}), (\boldsymbol{q}_{i;v}), B^v)$. Then we have $\Delta_v = \Delta_{v'}$ if $\pi(v) = \pi(v') = t \in \mathbb{T}_n$. Then the association $t \mapsto \Delta_t$ for $t \in \mathbb{T}_n$ is a generalized Y-pattern with coefficients where $\Delta_t := \Delta_v$ for any v such that $t = \pi(v)$.*

Proof. We first note that the Y-variables $y_{i;v}$ live in the universal semifield $\mathbb{QP}_{\mathrm{sf}}(y_1, \ldots, y_n)$ where $y_i = z^{e_i}$ are the initial Y-variables. The proof is completely analogous to Proposition 4.10. We leave the details to the reader. □

4.3. ***Special coefficients.*** By construction, given an $\mathcal{A}$-seed (resp. $\mathcal{X}$-seed) $\boldsymbol{s}$, there is the flat family

$$\pi_{\mathcal{A}} : \mathcal{A}_{\boldsymbol{s}} \to \operatorname{Spec} R \quad (\text{resp. } \pi_{\mathcal{X}} : \mathcal{X}_{\boldsymbol{s}} \to \operatorname{Spec} R).$$

Let λ be a $\Bbbk$-point of $\operatorname{Spec} R$. Then the special fiber $\pi^{-1}(\lambda)$ is a $\Bbbk$-scheme and can be viewed as a *generalized cluster variety with special coefficients*, denoted by $\mathcal{A}_{\boldsymbol{s},\lambda}$ (resp. $\mathcal{X}_{\boldsymbol{s},\lambda}$). They are also glued together by tori via birational morphisms (namely the $\mathcal{A}$- or $\mathcal{X}$-mutations specialized at λ)

$$\mathcal{A}_{\boldsymbol{s},\lambda} = \bigcup_{v \in \mathfrak{T}_n} T_{N,v}, \quad \mathcal{X}_{\boldsymbol{s},\lambda} = \bigcup_{v \in \mathfrak{T}_n} T_{M,v}.$$

The $\mathcal{A}$-type varieties (resp. $\mathcal{X}$-type varieties) lead to cluster patterns (resp. Y-patterns) with specialized coefficients. We have as before in the $\mathcal{A}$-case the inclusion of algebras

$$\mathscr{A}(\boldsymbol{s}, \lambda) \subset \overline{\mathscr{A}}(\boldsymbol{s}, \lambda) := H^0(\mathcal{A}_{\boldsymbol{s},\lambda}, \mathscr{O}_{\mathcal{A}_{\boldsymbol{s},\lambda}}).$$

4.4. ***Cluster duality.*** The cluster duality of Fock and Goncharov predicts, in the ordinary case, that the varieties $\mathcal{A}_{\boldsymbol{s}}$ and $\mathcal{X}_{\boldsymbol{s}}$ are dual in the sense that the upper cluster algebra $\overline{\mathscr{A}}(\boldsymbol{s})$ has a basis parametrized by the tropical set $\mathcal{X}^{\mathrm{trop}}(\mathbb{Z})$ (and vice versa). Note here $\boldsymbol{s}$ is viewed as a seed without coefficients so we do not need to distinguish between $\mathcal{A}$- and $\mathcal{X}$-seeds. Strictly speaking, this statement is not true as in some cases $\mathcal{X}_{\boldsymbol{s}}$ may have too few regular functions [Gross et al. 2015]. This duality (named the

Fock–Goncharov full conjecture) is the main subject of study (on a precise modified formulation and when it is true) in [Gross et al. 2018].

Our point of view of is that it is more natural to include generalized cluster varieties in cluster dualities, which we will demonstrate in the principal coefficients case. We denote the $\mathcal{X}$-cluster variety with principal coefficients by $\mathcal{X}_{\boldsymbol{s}}^{\mathrm{prin}}$, where the coefficient group is the tropical semifield

$$\mathbb{P} = \operatorname{Trop}(q_{ij} \mid i \in I,\ j \in [1, r_i]).$$

The scheme $\mathcal{X}_{\boldsymbol{s}}^{\mathrm{prin}}$ is over $\operatorname{Spec}(R)$ where $R = \Bbbk\mathbb{P}$. There are evaluations λ sending q_{ij} to $\lambda_{ij} \in \Bbbk^*$. Each λ specifies an $\mathcal{X}$-cluster variety with special coefficients as in the following diagram:

$$\begin{tikzcd}
\mathcal{X}_{\boldsymbol{s},\lambda}^{\mathrm{prin}} \arrow[r, hook] \arrow[d] & \mathcal{X}_{\boldsymbol{s}}^{\mathrm{prin}} \arrow[d, "\pi_{\mathcal{X}}"] \\
\operatorname{Spec}(\Bbbk) \arrow[r, hook, "\lambda"] & \operatorname{Spec}(R)
\end{tikzcd}$$

With a general choice coefficients, $\mathcal{X}_{\boldsymbol{s},\lambda}^{\mathrm{prin}}$ should be considered mirror dual to the family

$$\pi_{\mathcal{A}} : \mathcal{A}_{\boldsymbol{s}}^{\mathrm{prin}} \to \operatorname{Spec}(R),$$

where $\boldsymbol{s}$ is viewed as an $\mathcal{A}$-seed with coefficients. We shall not fully justify this statement in this paper, but instead will show that the family $\pi_{\mathcal{A}} : \mathcal{A}_{\boldsymbol{s}}^{\mathrm{prin}} \to \operatorname{Spec}(R)$ (as well as the generalized cluster algebra $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$) can be reconstructed from $\mathcal{X}_{\boldsymbol{s},\lambda}^{\mathrm{prin}}$, through a consistent wall-crossing structure (or scattering diagram) $\mathfrak{D}_{\boldsymbol{s}}$ associated to $\mathcal{X}_{\boldsymbol{s},\lambda}^{\mathrm{prin}}$; see Section 8.

5. Toric models and mutations

This section is a generalization of [Gross et al. 2015, Section 3] aiming for generalized cluster varieties. A *log Calabi–Yau pair* (X, D) is a smooth projective variety X (over an algebraically closed field $\Bbbk$) with a reduced simple normal crossing divisor D such that $K_X + D = 0$ where K_X is the canonical divisor of X. A *log Calabi–Yau variety* U is the interior of a log Calabi–Yau pair (X, D), i.e., $U = X \setminus D$. Described in [Gross et al. 2015], particularly relevant in cluster theory are log Calabi–Yau pairs (X, D) obtained from a blow-up $\pi : X \to X_\Sigma$ where X_Σ is the toric variety associated to a fan Σ in $\mathbb{R}^n$. The blow-up is along a hypersurface in the toric boundary of X_Σ, and D is given by the strict transform of the toric boundary. We will see that both generalized $\mathcal{X}$- and $\mathcal{A}$-varieties can be realized as log Calabi–Yau varieties obtained this way (up to codimension two subsets).

5.1. ***Toric models.*** Fix a lattice $N \cong \mathbb{Z}^n$ and let M be its dual. Suppose for $i \in I = [1, l]$ we have pairs of vectors $(e_i, w_i) \in N \times M$ such that $\langle e_i, w_i\rangle = 0$. We assume that all nonzero e_i are primitive, but some of them may equal. For each i, we fix a positive integer r_i. We also take functions (elements in $\Bbbk[M]$)

$$f_i = a_{i,0} + a_{i,1} z^{w_i} + \cdots + a_{i,r_i} z^{r_i w_i}$$

with nonzero $a_{i,0}$ and a_{i,r_i}.

We construct in below a log Calabi–Yau variety U_Λ using the data

$$\Lambda := ((e_i)_{i\in I}, (w_i)_{i\in I}, (f_i)_{i\in I}).$$

The following construction is what we mean by a *toric model* for U_Λ and we call such Λ a *toric model data*.

Construction 5.1 (cf. [Gross et al. 2015, Construction 3.4]). Given the data Λ, consider the fan

$$\Sigma = \Sigma_\Lambda := \{\mathbb{R}_{\geq 0} e_i \mid i \in I\} \cup \{0\}$$

in $N_\mathbb{R}$. Let X_Σ be the toric variety defined by Σ, and D_i be the irreducible toric boundary divisor corresponding to $\mathbb{R}_{\geq 0} e_i$. Note that since $\langle e_i, w_i\rangle = 0$, z^{w_i} does not vanish on D_i. Let Z_i be the zero locus of f_i on D_i, i.e., the closed subscheme $\overline{V}(f_i) \cap D_i$, which is a hypersurface. Blow up X_Σ along $\bigcup_{i=1}^{l} Z_i$ to obtain

$$\pi : \widetilde{X}_\Sigma \to X_\Sigma.$$

Let $\widetilde{D}_i$ be the strict transform of D_i in $\widetilde{X}_\Sigma$. Then the open subscheme $U_\Lambda := \widetilde{X}_\Sigma \setminus \bigcup_i \widetilde{D}_i$ is a log Calabi–Yau variety.

Definition 5.2. For $k \in I$, we say a toric model data Λ *k-mutable* if the pairs (e_i, w_i) satisfy the condition

$$\langle e_i, w_k\rangle = 0 \Longrightarrow \langle e_k, w_i\rangle = 0$$

for any $i \in I$.

We define mutations of a k-mutable toric model data.

Definition 5.3. Let Λ be a k-mutable toric model data and $\Lambda' = ((e_i'), (w_i'), (f_i'))$ be another set of data. Write $\beta_{ij} = \langle e_i, w_j\rangle$. We write $\Lambda' = \mu_k(\Lambda)$ (or say they are *μ_k-equivalent*) if they satisfy the following conditions:

- $e_k' = -e_k$ and $w_k' = -w_k'$;
- if $i \neq k$ and $\beta_{ik} \geq 0$, $e_i' = e_i$ and $w_i' = w_i$;
- if $i \neq k$ and $\beta_{ik} \leq 0$, $e_i' = e_i - \langle e_i, r_k w_k\rangle e_k$ and $w_i' = w_i + \langle e_k, w_i\rangle r_k w_k$;

and if writing $f_i' = a_{i,0}' + a_{i,1}' z^{w_i'} + \cdots + a_{i,r_i}' z^{r_i w_i'}$,

- $a_{k,j}' = a_{k,r_k-j}$ for $j \in [1, r_k]$;
- for $i \neq k$, $j \in [1, r_i]$,

$$a_{i,j}'/a_{i,0}' = \begin{cases} (a_{k,0})^{j\beta_{ki}} \cdot a_{i,j}/a_{i,0} & \text{if } \beta_{ik} > 0, \\ (a_{k,r_k})^{j\beta_{ki}} \cdot a_{i,j}/a_{i,0} & \text{if } \beta_{ik} \leq 0. \end{cases} \tag{5-1}$$

We note that the mutation μ_k is not deterministic for the (f_i) part, and is not involutive for the $((e_i), (w_i))$ part.

Applying Construction 5.1 to $\Lambda' = \mu_k(\Lambda)$, we obtain another log Calabi–Yau variety $U_{\Lambda'}$. Note that both U_Λ and $U_{\Lambda'}$ contain the torus T_N. Consider the birational morphism

$$\mu_k : T_N \dashrightarrow T_N, \quad \mu_k^*(z^m) = z^m \cdot f_k^{-\langle m, e_k\rangle}.$$

The following theorem is a generalization of the results in [Gross et al. 2015, Section 3].

Theorem 5.4. *The birational morphism μ_k extends to an isomorphism $\mu_k : U_\Lambda \to U_{\Lambda'}$ outside codimension two subsets if* $\dim \overline{V}(f_k) \cap Z_i < \dim Z_i$ *whenever* $\langle e_i, w_k\rangle = 0$ *for* $i \in I$.

Proof. We first make up some auxiliary varieties. Let $\Sigma^+ = \Sigma \cup \{\mathbb{R}_{\geq 0} e_k'\}$ and $\Sigma^- = \Sigma' \cup \{\mathbb{R}_{\geq 0} e_k\}$. We can blow up X_{Σ^+} (resp. X_{Σ^-}) in the same way as we do so for X_σ (resp. $X_{\Sigma'}$) to obtain $\widetilde{X}_+$ (resp. $\widetilde{X}_-$). Removing the strict transforms of the toric boundaries, we can still get U_Λ and $U_{\Lambda'}$. Following Lemma 3.6 in [Gross et al. 2015], we show that μ_k extends to an isomorphism (outside codimension two subsets) between $\widetilde{X}_+$ and $\widetilde{X}_-$, mapping the toric boundary of one to that of the other.

Suppose we only blow up X_{Σ^+} along Z_k and X_{Σ^-} along Z_k'. Then the blow-up $\widetilde{X}_+$ has a covering of open subsets

$$\widetilde{X}_+ = \widetilde{\mathbb{P}}_+ \cup \Big(\bigcup_{i \neq k} U_i \Big) \tag{5-2}$$

where $\widetilde{\mathbb{P}}_+$ is the blow-up along Z_k of the toric variety of the fan $\{\mathbb{R}_{\geq 0} e_k', \mathbb{R}_{\geq 0} e_k\}$ and U_i is the standard open toric chart corresponding to the ray $\mathbb{R}_{\geq 0} e_i$. Replacing U_i with $U_i \setminus \overline{V}(f_k)$ for $i \neq k$, (5-2) is still a covering but up to codimension two (with $\overline{V}(f_k) \cap D_i$ missing). More precisely, f_k is a regular function on U_i if $\langle w_k, e_i\rangle \geq 0$. In this case, $\overline{V}(f_k) \cap D_i$ is just the zero locus of the restriction of f_k on D_i, i.e., $V(f_k) \cap D_i$. As z^{w_k} vanishes on Z_i when $\langle w_k, e_i\rangle > 0$, $\overline{V}(f_k) \cap D_i = \varnothing$ since f_k has nonzero constant term. When $\langle w_k, e_i\rangle < 0$, then $\overline{V}(f_k) \cap D_i = V(z^{-r_k w_k} f_k) \cap D_i$ where $z^{-r_k w_k} f_k = f_k'$ is a regular function on U_i. So $\overline{V}(f_k) \cap D_i$ is still empty since f_k' has nonzero constant. Therefore we only fail to cover $\overline{V}(f_k) \cap D_i$ when $\langle w_k, e_i\rangle = 0$, which is a codimension two subset.

By Lemma 3.2 of [Gross et al. 2015], μ_k extends to a regular isomorphism from $\widetilde{\mathbb{P}}_+$ to $\widetilde{\mathbb{P}}_-$. Here $\widetilde{\mathbb{P}}_-$ is the blow-up along Z_k' of the toric variety defined by the fan $\{\mathbb{R}_{\geq 0} e_k', \mathbb{R}_{\geq 0} e_k\}$. We check that μ_k also extends to a regular isomorphism from $U_i \setminus \overline{V}(f_k)$ to $U_i' \setminus \overline{V}(f_k')$. Note that these are affine schemes so we check that μ_k^* extends to an isomorphism between their rings of regular functions. There are two cases.

(1) If $\langle e_i, w_k\rangle \geq 0$, then $e_i' = e_i$. Note that f_k is a regular function on U_i as well as on U_i'. Thus we have

$$U_i \setminus \overline{V}(f_k) = U_i \setminus V(f_k) \quad \text{and} \quad U_i' \setminus \overline{V}(f_k') = U_i' \setminus V(f_k).$$

For $\langle m, e_i\rangle \geq 0$, z^m defines a regular function on U_i' and

$$\mu_k^*(z^m) = z^m f_k^{-\langle m, e_k\rangle}$$

is a regular function on $U_i \setminus V(f_k)$.

(2) If $\langle e_i, w_k\rangle < 0$, then $e_i' = e_i - \langle e_i, r_k w_k\rangle e_k$. Instead of f_k, the function $f_k' = z^{-r_k w_k} f_k$ is a regular function on U_i and $\overline{V}(f_k) = V(f_k')$. For $\langle m, e_i'\rangle \geq 0$ and z^m a regular function on U_i', we have

$$\mu_k^*(z^m) = z^m f_k^{-\langle m, e_k\rangle} = z^{m - r_k w_k \langle m, e_k\rangle}(f_k')^{-\langle m, e_k\rangle}.$$

We check that $\langle m - r_k w_k\langle m, e_k\rangle, e_i\rangle = \langle m - r_k w_k \langle m, e_k\rangle, e_i' + \langle e_i, r_k w_k\rangle e_k\rangle = \langle m, e_i'\rangle > 0$. Thus $\mu_k^*(z^m)$ is a regular function on $U_i \setminus \overline{V}(f_k) = U_i \setminus V(f_k')$.

Therefore μ_k^* is a morphism between regular functions. In all the cases above, one checks that sending z^m to $z^m f_k^{\langle m, e_k\rangle}$ is the inverse of μ_k^*. Summarizing, we have so far proven that there is an isomorphism

$$\mu_k : U_+ := \widetilde{\mathbb{P}}_+ \cup \Big(\bigcup_{i \neq k} U_i \setminus \overline{V}(f_k)\Big) \to U_- := \widetilde{\mathbb{P}}_- \cup \Big(\bigcup_{i \neq k} U_i' \setminus \overline{V}(f_k')\Big)$$

extending the birational morphism μ_k between tori.

Now we analyze the impact of blowing up the hypersurfaces Z_i (and Z_i') for $i \neq k$. When $\langle w_k, e_i\rangle \neq 0$, as discussed $D_i \cap \overline{V}(f_k) = \varnothing$, so $Z_i \subset D_i$ is contained in U_+. Since $\langle w_k', e_i'\rangle = -\langle w_k, e_i\rangle \neq 0$, the same is true for Z_i', i.e., $Z_i' \subset U_-$. We would like to show that $\mu_k(Z_i) = Z_i'$ when $\langle w_k, e_i\rangle \neq 0$. There are two cases.

(1) Suppose $\langle w_k, e_i\rangle > 0$. In this case, $e_i' = e_i$ and $w_i' = w_i$. By definition $Z_i' = D_i' \cap V(f_i') = V(z^{m_0}) \cap V(f_i') \subset U_i'$ for some m_0 such that $\langle m_0, e_i'\rangle = 1$. Now we have $\mu_k^*(z^{m_0}) = z^{m_0} f_k^{-\langle m_0, e_k\rangle}$ and

$$\mu_k^*(f_i') = a_{i,0}' + a_{i,1}' z^{w_i} f_k^{-\langle w_i, e_k\rangle} + \cdots + a_{i,r_i}' z^{r_i w_i} f_k^{-\langle r_i w_i, e_k\rangle}.$$

Note that f_k is invertible on $U_i \setminus \overline{V}(f_k)$ and restricts to constant p_{k0} on D_i. So $V(\mu_k^*(z^{m_0}))$ is just the divisor D_i and

$$\mu_k^*(f_i')|_{D_i} = a_{i,0}' + a_{i,1}' a_{k,0}^{-\beta_{ki}} z^{w_i} + \cdots a_{i,r_i}' a_{k,0}^{-r_i \beta_{ki}} z^{r_i w_i} = \lambda \cdot f_i|_{D_i}.$$

for some nonzero $\lambda \in \Bbbk$ by the μ_k-equivalence assumption on Λ and Λ'. Therefore $\mu_k(Z_i) = Z_i'$.

(2) Suppose $\langle w_k, e_i\rangle < 0$. In this case we have $e_i' = v_i - \langle r_k w_k, e_i\rangle e_k$ and $w_i' = w_i + \langle w_i, e_k\rangle r_k w_k$. Still $Z_i' = V(z^{m_0}) \cap V(f_i')$. Now instead of f_k, the function $f_k' = z^{-r_k w_k} f_k$ is a regular function on U_i and restricts to constant a_{k,r_k} on D_i. First, $\mu_k^*(z^{m_0}) = z^{m_0 - \langle m_0, e_k\rangle r_k w_k}(f_k')^{\langle m_0, e_k\rangle}$. Since f_k' is invertible on $U_i \setminus V(f_k)$, $V(\mu_k^*(z^{m_0})) = D_i$ as $\langle m_0 + \langle m_0, e_k\rangle r_k w_k, e_i\rangle = 1$. Secondly we have

$$\begin{aligned}\mu_k^*(f_i') &= a_{i,0}' + a_{i,1}' z^{w_i'} f_k^{-\langle w_i, e_k\rangle} + \cdots + a_{i,r_i}' z^{r_i w_i'} f_k^{-\langle r_i w_i, e_k\rangle} \\ &= a_{i,0}' + a_{i,1}' z^{w_i' - \langle w_i, e_k\rangle r_k w_k}(f_k')^{-\langle w_i, e_k\rangle} + \cdots + a_{i,r_i}' z^{r_i w_i' - \langle r_i w_i, e_k\rangle r_k w_k}(f_k')^{-\langle r_i w_i, e_k\rangle}.\end{aligned}$$

Hence

$$\mu_k^*(f_i') \mid_{D_i} = a_{i,0}' + a_{i,1}' a_{k,r_k}^{-\beta_{ki}} z^{w_i} + \cdots + a_{i,r_i}' a_{k,r_k}^{-r_i \beta_{ki}} z^{r_i w_i} = \lambda \cdot f_i \mid_{D_i}$$

for some nonzero $\lambda \in \Bbbk$ again by the μ_k-equivalence assumption. Therefore in this case we also have $\mu_k(Z_i) = Z_i'$.

Finally we consider the case $\langle w_k, e_i\rangle = 0$. The argument we need is exactly the same as in the last paragraph of the proof in [Gross et al. 2015]. By the assumption $\langle w_k, e_i\rangle = 0 \Longrightarrow \langle w_i, e_k\rangle = 0$, so we have

$$\mu_k^*(f_i') = f_i,$$

and thus $\mu_k(Z_i) = Z_i'$. The problem is that Z_i may not be fully contained in $D_i \setminus V(f_k)$, with $V(f_k) \cap Z_i$ missing. If $V(f_k) \cap Z_i$ contains a irreducible component of Z_i, then U_Λ would contain the corresponding exceptional divisor while blowing up in U_+ does not. However the isomorphism $\mu_k : U_+ \to U_-$ need not extend as isomorphism across this exceptional divisor. Now we need the further hypothesis $\dim V(f_k) \cap Z_i < \dim Z_i$ so that the missing part in the blow-up center is of at least codimension three in U_i. After blowing up the corresponding locus in U_+ and U_+, we have the diagram

$$\begin{tikzcd}
\widetilde{U}_+ \arrow[r, "\mu_k"] \arrow[d, "\pi"] & \widetilde{U}_- \arrow[d, "\pi"] \\
U_+ \arrow[r, "\mu_k"] & U_-
\end{tikzcd}$$

where vertical arrows are blow-ups and horizontal arrows are genuine isomorphisms. Removing the strict transform of the toric boundary, we have immersions

$$\widetilde{U}_+ \setminus \widetilde{D} \subset U_\Lambda \quad \text{and} \quad \widetilde{U}_- \setminus \widetilde{D} \subset U_{\Lambda'}$$

missing codimension two loci. Summarizing, the birational map μ_k can be extended to an isomorphism $\mu_k : U_\Lambda \dashrightarrow U_{\Lambda'}$ outside sets of codimension two. □

A sufficient condition for the assumption in Theorem 5.4 to hold is

$$\forall \langle e_i, w_k\rangle = 0, \quad \dim \overline{V}(f_k) \cap Z_i < \dim Z_i.$$

Definition 5.5 (cf. [Berenstein et al. 2005, Definition 1.4]). A toric model data $\Lambda = ((e_i), (w_i), (f_i))$ is said to be *coprime* if the functions f_i are pairwise coprime as elements in the ring $\Bbbk[M]$.

Corollary 5.6. *The result in Theorem 5.4 holds if Λ is coprime.*

Proof. Note that $Z_i = \overline{V}(f_i) \cap D_i$. If needed, multiply some monomial z^m to f_i so that $\tilde{f}_i = z^m f_i$ is a regular function on D_i. Do the same to f_k to get $\tilde{f}_k$. By the coprime condition on Λ, $\tilde{f}_i$ and $\tilde{f}_k$ are still coprime, so we have

$$\dim V(\tilde{f}_k) \cap V(\tilde{f}_i) < \dim V(\tilde{f}_i),$$

where the above subschemes are taken inside D_i. □

The following is an easy-to-check condition on Λ for the coprimeness to hold.

Lemma 5.7. *If the vectors w_i are linear independent, then Λ is coprime.*

5.2. ***The upper bound.*** Suppose we are given the data $\Lambda = ((e_i), (w_i), (f_i))$. Assume that Λ is i-mutable for any $i \in I$. For $i \in I$, let $T_N^{(i)}$ be a copy of the torus T_N. Then we have birational maps for each $i \in I$,

$$\mu_i : T_N \dashrightarrow T_N^{(i)}, \quad \mu_i^*(z^m) = z^m f_i^{-\langle e_i, m\rangle}.$$

We glue the $|I| + 1$ tori along the maps μ_i to obtain a scheme X_Λ.

In previous section, we know that not only the torus T_N, U_Λ also contains the torus $T_N^{(i)}$, that is, we have the following diagram for every $i \in I$:

$$\begin{tikzcd}
T_N \arrow[r, dashed, "\mu_i"] \arrow[d, hook] & T_N^{(i)} \arrow[dl, hook] \\
U_\Sigma &
\end{tikzcd}$$

These diagrams determine a unique morphism $\psi : X_\Lambda \to U_\Lambda$.

Lemma 5.8 [Gross et al. 2015, Lemma 3.5]. *The morphism $\psi : X_\Lambda \to U_\Lambda$ satisfies the following properties*:

(1) *If* $\dim Z_i \cap Z_j < \dim Z_i$ *for all $i \neq j$, then ψ is an isomorphism outside a set of codimension at least two.*

(2) *If $Z_i \cap Z_j = \varnothing$ for all $i \neq j$, then ψ is an open immersion. In particular, in this case, X_Λ is separated.*

In the $\mathcal{A}$-cluster case to be explained later, the variety X_Λ may be named *the upper bound* according to [Fomin and Zelevinsky 2007].

5.3. ***Toric models for cluster varieties.*** In this section, we realize generalize cluster varieties as log Calabi–Yau varieties utilizing Construction 5.1.

5.3.1. ***$\mathcal{A}$-cluster cases.*** Suppose we have fixed data Γ and an $\mathcal{A}$-seed with coefficients $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{p})$. We further choose an evaluation $\lambda : \mathbb{P} \to \Bbbk^*$. This amounts to pick a $\Bbbk$-point of $\operatorname{Spec}(\Bbbk\mathbb{P})$. These lead to the generalized $\mathcal{A}$-cluster variety $\mathcal{A}_{\boldsymbol{s},\lambda}$ with special coefficients.

Meanwhile consider the toric model data

$$\Lambda(\boldsymbol{s}, \lambda) := ((e_i)_{i \in I_{\mathrm{uf}}}, (w_i)_{i \in I_{\mathrm{uf}}}, (f_i)_{i \in I_{\mathrm{uf}}})$$

defined as follows. The vectors $(e_i)_{i \in I_{\mathrm{uf}}}$ are taken from the seed $\boldsymbol{s}$. Recall that we have the exchange matrix $B = (b_{ij})$ where $b_{ij} := \omega(e_i, d_j e_j)$. Write $\beta_{ij} = b_{ij}/r_j$. Note that $\{e_i \mid i \in I\}$ form a basis of the lattice N and we denote by e_i^* the dual basis of M. Then define

$$w_i := \omega(-, d_i e_i / r_i) = \sum_{j \in I} \beta_{ij} e_i^* \in M, \quad f_i := \lambda(\theta[\boldsymbol{p}_i](z^{w_i}, 1)) \in \Bbbk[M].$$

Then Construction 5.1 applies to the toric model data $\Lambda(\boldsymbol{s}, \lambda)$, and thus there is the associated log Calabi–Yau variety $U_{\Lambda(\boldsymbol{s},\lambda)}$. Recall that we also have the scheme $X_{\Lambda(\boldsymbol{s},\lambda)}$ obtained by gluing $n + 1$ copies of the torus T_N as in Section 5.2. We call $X_{\Lambda(\boldsymbol{s},\lambda)}$ *the upper bound* for $(\boldsymbol{s}, \lambda)$, which by definition is an open subscheme of $\mathcal{A}_{\boldsymbol{s},\lambda}$.

The following lemma is easy to verify by direct computations.

Lemma 5.9. *We have $\mu_k(\Lambda(\boldsymbol{s},\lambda)) = \Lambda(\mu_k(\boldsymbol{s}),\lambda)$ in the sense of Definition 5.3. The latter μ_k is the mutation of an $\mathcal{A}$-seed with coefficients.*

Proposition 5.10. (1) *The morphism $\psi : X_{\Lambda(\boldsymbol{s},\lambda)} \to U_{\Lambda(\boldsymbol{s},\lambda)}$ is an open immersion with image an open subset whose complement has codimension at least two.*

(2) *The birational map $\mu_k : U_{\Lambda(\boldsymbol{s},\lambda)} \dashrightarrow U_{\Lambda(\mu_k(\boldsymbol{s}),\lambda)}$ is an isomorphism outside codimension two in each of the listed situations*:

A. *The functions f_i have general coefficients.*

B. *The seed $\boldsymbol{s}$ is mutation equivalent to one with principal coefficients, and $\lambda \in (\Bbbk^*)^{|I'|}$ is general enough.*

Proof. (1) follows from Lemma 5.8, part (2) — as we only need to check the hypothesis $Z_i \cap Z_j = \varnothing$ for all $i \neq j$. In fact, in $\mathcal{A}$-cluster case, since $e_i \neq e_j$, we have $T_{N/\langle e_i\rangle} \cap T_{N/\langle e_j\rangle} = \varnothing$ for all $i \neq j$, where $T_{N/\langle e_i\rangle}$ is viewed as the dense torus contained in the divisor D_i. As Z_i is a closed subset of $T_{N/\langle e_i\rangle}$, the hypothesis holds.

(2) follows from Theorem 5.4. We need to check that whenever $\langle e_i, u_k\rangle = 0$,

$$\dim \overline{V}(f_k) \cap \overline{V}(f_i) \cap D_i < \dim \overline{V}(f_i) \cap D_i.$$

A sufficient condition is the functions f_i being coprime. Note that for $i \in I$,

$$f_i = \prod_{j=1}^{r_i} (\lambda(p_{i,j}^+)z^{w_i} + \lambda(p_{i,j}^-)).$$

When these f_i have general coefficients (case A), they are coprime. In case B, one may replace f_i by

$$\tilde{f}_i = \prod_{j=1}^{r_i} (\lambda(p_{i,j})z^{w_i} + 1).$$

Since the elements $p_{i,j}$ for $i \in I$ and $j \in [1, r_i]$ form a $\mathbb{Z}$ basis in $\mathbb{P}$ (by Lemma 3.18) when $\boldsymbol{s}$ is mutation equivalent to one with principal coefficients, these $\tilde{f}_i$ are coprime as long as λ is general. □

Remark 5.11. Suppose we are in the situation of case B of Proposition 5.10(2). Then we have isomorphisms of the rings of regular functions

$$\Bbbk[X_{\Lambda(\boldsymbol{s},\lambda)}] \cong \Bbbk[U_{\Lambda(\boldsymbol{s},\lambda)}] \cong \Bbbk[U_{\Lambda(\mu_k(\boldsymbol{s}),\lambda)}].$$

The equality then extends to any seed $\boldsymbol{s}_v$ that is mutation equivalent to $\boldsymbol{s}$. It then follows that they are all isomorphic to the upper cluster algebra

$$\mathscr{A}(\boldsymbol{s},\lambda) = \Bbbk[\mathcal{A}_{\boldsymbol{s},\lambda}].$$

The cluster variables in seed $\boldsymbol{s}$ are $x_{i,\boldsymbol{s}} := z^{e_i^*}$. Each $x_{i,\boldsymbol{s}}$ extends to a regular function on the toric variety X_Σ corresponding to the toric model data $\Lambda(\boldsymbol{s}, \lambda)$. Then $x_{i,\boldsymbol{s}}$ pulls back to the blow-up $\widetilde{X}_\Sigma$ and restricts to a regular function on the open subvariety $U_{\Lambda(\boldsymbol{s},\lambda)}$. It follows from (2) of Proposition 5.10 that $x_{i,\boldsymbol{s}}$ is also a regular function on $X_{\Lambda(\boldsymbol{s}_v,\lambda)}$ and in particular is a Laurent polynomial if restricted to $T_{N,\boldsymbol{s}_v}$. This explains the generalized Laurent phenomenon Theorem 3.7, which was observed in [Gross et al. 2015] for the ordinary case.

5.3.2. ***$\mathcal{X}$-cluster cases.*** Suppose we have fixed data Γ and an $\mathcal{X}$-seed with coefficients $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{q})$. Let us make the assumption that for any $j \in I_{\mathrm{uf}}$,

$$r_j = \gcd(b_{ij}, i \in I).$$

This is equivalent to say that each w_j for $j \in I_{\mathrm{uf}}$ is primitive as an element of $M = \mathrm{Hom}(N, \mathbb{Z})$. Switching the roles of (e_i) and (w_i), we obtain the toric model data

$$\Omega(\boldsymbol{s}, \lambda) = ((-w_i), (e_i), (g_i))$$

for M instead of N, where

$$g_i := \lambda(\theta[\boldsymbol{q}_i](z^{e_i}, 1)) \in \Bbbk[N]$$

with some chosen evaluation λ. Since the matrix B is skew-symmetrizable, $\Omega(\boldsymbol{s}, \lambda)$ is k-mutable for any $k \in I_{\mathrm{uf}}$.

Lemma 5.12. *The assumption that* $r_j = \gcd(b_{ij}, i \in I)$ *is invariant under mutations.*

Proof. This is because if the j-th column of B is divisible by r_j then the same is true for the matrix $\mu_k(B) = (b'_{ij})$. Thus we have

$$\gcd(b_{ij}, i \in I) = \gcd(b'_{ij}, i \in I)$$

as μ_k is involutive on B. □

The above lemma shows that we have well-defined data $\Omega(\mu_k(\boldsymbol{s}), \lambda)$.

Lemma 5.13. *We have* $\mu_k(\Omega(\boldsymbol{s}, \lambda)) = \Omega(\mu_k(\boldsymbol{s}), \lambda)$, *where the later* μ_k *is the mutation for an* $\mathcal{X}$*-seed with coefficients.*

Proof. This lemma is analogous to Lemma 5.9 and is also easy to check. However, to show that the carefully chosen signs and conventions are the correct ones, we record some details here.

In the notations of Definition 5.3, for the data $\Omega(\boldsymbol{s}, \lambda)$, we take $e_i = -w_i$ and $w_i = e_i$. So after the mutation μ_k in sense of Definition 5.3, for $i \neq k$

$$(-w_i)' = \begin{cases} -w_i - \langle (-w_i), r_k e_k \rangle (-w_k) & \text{if } \langle -w_i, e_k \rangle \leq 0, \\ -w_i & \text{if } \langle -w_i, e_k \rangle > 0. \end{cases}$$

Note that the two conditions are equivalent to $\beta_{ik} \leq 0$ and $\beta_{ik} > 0$ respectively. And in these two cases, we have

$$(-w_i)' = -w_i - \langle e_k, w_i \rangle r_k w_k \quad \text{and} \quad -w_i$$

respectively. This is exactly $-w_i'$ for $w_i' = \omega(-, d_i e_i'/r_i)$ from the seed $\mu_k(\boldsymbol{s})$. Similarly, one checks that the $\boldsymbol{e}$ part is also compatible with mutations.

As for coefficients, for the data $\Omega(\mu_k(\boldsymbol{s}), \lambda)$, we have

$$g_i'(u, v) = \lambda(\theta[\boldsymbol{q}_i'](u, v).$$

Here $q_{i,j}'$ is obtain from $\mathcal{X}$-type mutations for coefficients (see Definition 4.14) which coincides with Definition 5.3. □

Recall that $X_{\Omega(\boldsymbol{s},\lambda)}$ is the upper bound for $\Omega(\boldsymbol{s}, \lambda)$ as defined in Section 5.2.

Proposition 5.14. *For the $\mathcal{X}$-type constructions,*

(1) *the morphism $\psi : X_{\Omega(\boldsymbol{s},\lambda)} \to U_{\Omega(\boldsymbol{s},\lambda)}$ is an open immersion with image being an open subset whose complement has codimension at least two*;

(2) *the birational map $\mu_k : U_{\Omega(\boldsymbol{s},\lambda)} \dashrightarrow U_{\Omega(\mu_k(\boldsymbol{s}),\lambda)}$ is an isomorphism outside codimension two subsets.*

Proof. The proof of (1) is completely analogous to that of (1) of Proposition 5.10. For (2), it follows from that for any $\mathcal{X}$-seed $\boldsymbol{s}$, the data $\Omega(\boldsymbol{s}, \lambda)$ is always coprime by Lemma 5.7 as the vectors e_i form a basis of N. □

6. Scattering diagrams

This section deals with scattering diagrams. Our main objects of study *generalized cluster scattering diagrams* will be defined in Section 6.2.

6.1. ***The tropical vertex.*** We start with a more general setup of scattering diagrams as in [Argüz and Gross 2022, Section 5.1.1]. Let N be a lattice of finite rank, $M = \operatorname{Hom}_{\mathbb{Z}}(N, \mathbb{Z})$ and $M_{\mathbb{R}} = M \otimes_{\mathbb{Z}} \mathbb{R}$. Let P be a monoid with a monoid map $r : P \to M$. Denote by $P^\times$ the groups of units of P and let $\mathfrak{m}_P = P \setminus P^\times$. An ideal of the monoid P induces a monomial ideal of the ring $\Bbbk[P]$, where $\Bbbk$ is a ground field. So we use the same letter to denote both. For any monomial ideal $I \subset \Bbbk[P]$, define the ring

$$R_I := \Bbbk[P]/I.$$

Denote by $\widehat{\Bbbk[P]}$ the completion of $\Bbbk[P]/\mathfrak{m}_P^n$ for $n \in \mathbb{N}$.

For I such that its radical $\sqrt{I}$ is equal to $\mathfrak{m}_P$ (e.g., $I = \mathfrak{m}_P^n$ for some $n \in \mathbb{N}$), define the *module of log derivations* $\Theta(R_I) := R_I \otimes_{\mathbb{Z}} N$ as follows.

If we write the element $z^p \otimes n$ as $z^p \partial_n$ for $p \in P$ and $n \in N$, then it acts on R_I by

$$z^p \partial_n (z^{p'}) = \langle n, r(p') \rangle z^{p+p'}, \quad p' \in P.$$

Then the submodule $\mathfrak{m}_P\Theta(R_I)$ is a Lie algebra with the commutator bracket

$$[z^{p_1}\partial_{n_1}, z^{p_2}\partial_{n_2}] = z^{p_1+p_2}\partial_{\langle r(p_2),n_1\rangle n_2-\langle r(p_1),n_2\rangle n_1}.$$

Taking exponential of elements in this Lie algebra, we get group elements in $\operatorname{Aut}(R_I)$. There is a nilpotent Lie subalgebra of $\mathfrak{m}_P\Theta(R_I)$ defined by

$$\mathfrak{v}_I := \bigoplus_{m\in P\setminus I, r(m)\neq 0} z^m(\Bbbk\otimes r(m)^\perp).$$

Since it is nilpotent, this Lie subalgebra (as a set) is in bijection with the corresponding algebraic group $\mathbb{V}_I := \exp(\mathfrak{v}_I)\subset \operatorname{Aut}(R_I)$. Taking the projective limit with respect to the ideals $\mathfrak{m}_P^n$ for $n\in\mathbb{N}$, we get a pro-unipotent group $\widehat{\mathbb{V}}$, which is in bijection with the pro-nilpotent Lie algebra $\hat{\mathfrak{v}} := \varprojlim \mathfrak{v}_{\mathfrak{m}_P^n}$. The group $\widehat{\mathbb{V}}$ is called the *higher-dimensional tropical vertex group*, acts by automorphisms on $\widehat{\Bbbk[P]}$. We also denote (without completion)

$$\mathfrak{v} := \bigoplus_{m\in P,\, r(m)\neq 0} z^m(\Bbbk\otimes r(m)^\perp).$$

Definition 6.1. A *scattering diagram* in $M_\mathbb{R}$ over R_I is a finite set $\mathfrak{D}$ of *walls* where each *wall* $(\mathfrak{d}, f_\mathfrak{d})$ is a rational polyhedral cone $\mathfrak{d}\subset M_\mathbb{R}$ of codimension one along with an attached element called *wall-crossing function*

$$f_\mathfrak{d} = \sum_{\substack{m\in P\setminus I\\ r(m)\in\Lambda_\mathfrak{d}}} c_m z^m \in R_I,$$

where $\Lambda_\mathfrak{d}\subset M$ is the integral tangent space of any point in $\mathfrak{d}$, i.e., $\Lambda_\mathfrak{d} = M\cap\mathbb{R}\langle\mathfrak{d}\rangle$. We require that $f_\mathfrak{d}\equiv 1 \bmod \mathfrak{m}_P$.

Remark 6.2. Upon choosing a generator n_0 of $\Lambda_\mathfrak{d}^\perp\cap N$, the wall-crossing function $f_\mathfrak{d}$ induces an element in $\mathbb{V}_I\subset\operatorname{Aut}(R_I)$ by the action

$$z^p\mapsto z^p f_\mathfrak{d}^{\langle r(p),n_0\rangle}.$$

So this wall-crossing automorphism depends on how one crosses the wall. One may view that this wall-crossing automorphism depends on the direction in which one transversally crosses the wall. With n_0 chosen, such an automorphism can be equivalently represented by the corresponding Lie algebra element $\log(f_\mathfrak{d})\partial_{n_0}\in\mathfrak{v}_I$.

Let $\operatorname{Supp}(\mathfrak{D})$ be the union of all walls in $\mathfrak{D}$. Let $\operatorname{Sing}(\mathfrak{D})$ be the union of at least codimension two intersections of every pair of walls and the boundary of every wall. Let $\gamma : [0, 1]\to M_\mathbb{R}$ be a piecewise smooth proper map such that the end points $\gamma(0)$ and $\gamma(1)$ avoid $\operatorname{Supp}(\mathfrak{D})$ and whose image is disjoint from $\operatorname{Sing}(\mathfrak{D})$. We also assume that γ meets walls transversally.

Suppose that γ crosses walls $\mathfrak{d}_1, \dots, \mathfrak{d}_s$ in $\mathfrak{D}$ at times

$$0 < t_1\leq t_2\leq\cdots\leq t_s < 1.$$

These numbers t_i are obtained by considering the finite set $\gamma^{-1}(\mathrm{Supp}(\mathfrak{D})) \subset [0, 1]$ as γ is proper. It is possible that $t_i = t_j$ as walls may overlap. Suppose γ crosses a wall $(\mathfrak{d}, f_{\mathfrak{d}})$ at time t. Denote by $\xi_{\gamma,\mathfrak{d}}$ the element in $\mathbb{V}_I$ with the action

$$z^p \mapsto z^p f_{\mathfrak{d}}^{\langle r(p), n_0\rangle}, \quad p \in P \setminus I$$

where n_0 is chosen such that $\langle n_0, \gamma'(t_i)\rangle > 0$.

Definition 6.3. We define the *path-ordered product* of γ in $\mathfrak{D}$ to be the element

$$\mathfrak{p}_{\gamma,\mathfrak{D}} := \xi_{\gamma,\mathfrak{d}_s}\xi_{\gamma,\mathfrak{d}_{s-1}} \cdots \xi_{\gamma,\mathfrak{d}_1} \in \mathbb{V}_I.$$

Definition 6.4. A scattering diagram $\mathfrak{D}$ over R_I is *consistent* if the path-ordered product $\mathfrak{p}_{\gamma,\mathfrak{D}}$ only depends on the endpoints $\gamma(0)$ and $\gamma(1)$ for any path $\gamma : [0, 1] \to M_{\mathbb{R}}$ for which $\mathfrak{p}_{\gamma,\mathfrak{D}}$ is well-defined.

Recall that we have the completed algebra $\widehat{\Bbbk[P]} := \varprojlim R_{\mathfrak{m}_P^k}$. For an element $f \in \widehat{\Bbbk[P]}$, denote by $f^{<k}$ its projection in $R_{\mathfrak{m}_P^k}$.

Definition 6.5. A *scattering diagram* in $M_{\mathbb{R}}$ over $\widehat{\Bbbk[P]}$ is a (possibly infinite) set $\mathfrak{D}$ of walls $(\mathfrak{d}, f_{\mathfrak{d}})$ with $\mathfrak{d}$ a rational polyhedral cone of codimension one and the wall-crossing function

$$f_{\mathfrak{d}} = \sum_{\substack{m \in P \\ r(m) \in \Lambda_{\mathfrak{d}}}} c_m z^m \in \widehat{\Bbbk[P]},$$

such that modulo the ideal $\mathfrak{m}_P^n$, the collection $\mathfrak{D}^{<n} := \{(\mathfrak{d}, f_{\mathfrak{d}}^{<n})\}$ is a scattering diagram over $R_{\mathfrak{m}_P^n}$. A scattering diagram $\mathfrak{D}$ is consistent if $\mathfrak{D}^{<n}$ is consistent for any $n \in \mathbb{N}$.

The path-ordered product for $\mathfrak{D}$ over $\widehat{\Bbbk[P]}$ is defined through the projective limit of path-ordered products for $\mathfrak{D}^{<n}$:

$$\mathfrak{p}_{\gamma,\mathfrak{D}} := \varprojlim \mathfrak{p}_{\gamma,\mathfrak{D}^{<n}} \in \widehat{\mathbb{V}} \subset \mathrm{Aut}(\widehat{\Bbbk[P]}).$$

Definition 6.6. We say two scattering diagrams $\mathfrak{D}$ and $\mathfrak{D}'$ (over the same algebra) are *equivalent* if for any γ, we have $\mathfrak{p}_{\gamma,\mathfrak{D}} = \mathfrak{p}_{\gamma,\mathfrak{D}'}$ whenever both path-ordered products are well-defined.

Definition 6.7. We say a wall $\mathfrak{d}$ has *direction* m_0 for some $m_0 \in M$ if the attached wall-crossing function $f_{\mathfrak{d}}$ only contains monomials z^p such that $r(p) = -km_0$ for some $k \in \mathbb{N}$. A wall $(\mathfrak{d}, f_{\mathfrak{d}})$ with direction m_0 is called *incoming* if $\mathfrak{d} = \mathfrak{d} - \mathbb{R}_{\geq 0} m_0$.

Next we explain how to assign a scattering diagram to an $\mathcal{X}$-type toric model. We are actually in a particular situation within the more general framework of [Argüz and Gross 2022], which works for any log Calabi–Yau variety obtained from blowing-up a toric variety along hypersurfaces in the toric boundary.

Let $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{q})$ be an $\mathcal{X}$-seed with principal coefficients for some fixed data Γ. We assume that $N_{\mathrm{uf}} = N$ to avoid frozen directions. As usual, write $\boldsymbol{e} = (e_i)$. We assume the condition that $r_j = \gcd(b_{ij} \mid i \in I)$ for any $j \in I$. This assumption implies any $w_i := \frac{d_i}{r_i}\omega(-, e_i) \in M$ is primitive. Recall that we have used the fan

$$\Sigma_0 := \{0\} \cup \{-\mathbb{R}_{\geq 0} w_i\}$$

to describe the toric model of $U = U_{\Omega(s,\lambda)}$. The functions (in the data $\Omega(s, \lambda)$ to define U) are then

$$g_i = \prod_{j=1}^{r_i} (1 + \lambda_{ij} z^{e_i}) \in \Bbbk[N].$$

We pick a complete fan Σ in $M_\mathbb{R}$ containing Σ_0. For example, we may take a refinement of (the cone complex induced by) the hyperplane arrangement $\{e_i^\perp \mid i \in I\}$. Let X_Σ be the corresponding (complete) toric variety, with D_i being the boundary toric divisor corresponding to the ray $-\mathbb{R}_{\geq 0} w_i$. Let $H = \bigcup_i H_i$, where

$$H_i = \bigcup_{j\in[1,r_i]} H_{ij} := \bigcup_{j\in[1,r_i]} \overline{V}(1 + \lambda_{ij} z^{e_i}) \cap D_i$$

which is a union of disjoint hypersurfaces in D_i (as the coefficients $\lambda_{ij} \in \Bbbk^*$ are general). These hypersurfaces are exactly where we blow up X_Σ to obtain the log Calabi–Yau variety $U_{\Omega(s,\lambda)}$.

Take the monoid

$$P := M \oplus \prod_{i\in I} \mathbb{N}^{r_i},$$

with the natural projection $r : P \to M$. We write multiplicatively $t_{i,1}, t_{i,2}, \ldots t_{i,r_i}$ for the generators of $\mathbb{N}^{r_i}$. For each ray $\rho_i := -\mathbb{R}_{\geq 0} w_i$ and H_{ij}, there is a finite scattering diagram $\mathfrak{D}_{ij}$ called a *widget* from a certain *tropical hypersurface* [Argüz and Gross 2022, Definition 5.3 and Section 5.1.3]. In our case, they are given by:

Lemma 6.8. *The widget $\mathfrak{D}_{ij}$ consists of all codimension one cones of the fan Σ contained in the hyperplane $e_i^\perp$ containing ρ_i, with the same wall-crossing function $(1 + t_{i,j} z^{w_i})$. In other words, we have*

$$\mathfrak{D}_{ij} = \{(\sigma, 1 + t_{i,j} z^{w_i}) \mid \sigma \in \Sigma,\ \dim \sigma = n - 1,\ \sigma \subset e_i^\perp,\ \rho_i \subset \sigma\}.$$

Proof. By definition [Argüz and Gross 2022, Definition 5.3 and Section 5.1.3], $\mathfrak{D}_{ij}$ consists of walls $(\sigma, (1+t_{i,j} z^{w_i})^{\omega_\sigma})$ where σ runs through all codimension one cones in Σ containing ρ_i and $\omega_\sigma = H_{ij} \cdot D_\sigma$ is the intersection number computed in the divisor D_i. Here D_σ is the one-dimensional toric stratum in D_i corresponding to σ. Note that if $e_i \notin \sigma^\perp$, then z^{e_i} or z^{-e_i} vanishes along D_σ. So $H_{ij} = \overline{V}(1 + \lambda_{ij} z^{e_i})$ does not intersect D_σ and thus $\omega_\sigma = 0$. If $\sigma \subset e_i^\perp$, as e_i is primitive, the intersection is at the point $z^{e_i} = -1/\lambda_{ij}$, where z^{e_i} can be viewed as the coordinate on D_σ. Thus the multiplicity ω_σ is 1. □

Note that by Definition 6.7 every wall in $\mathfrak{D}_{ij}$ is incoming since $-w_i$ is contained in every σ.

Theorem 6.9 [Argüz and Gross 2022, Theorem 5.6 and Section 5.1.3]. *Consider the scattering diagram (with only incoming walls)*

$$\mathfrak{D}_{(X_\Sigma,H),\mathrm{in}} := \bigcup_{i\in I} \bigcup_{j\in[1,r_i]} \mathfrak{D}_{ij}.$$

There exists a unique (up to equivalence) consistent scattering diagram $\mathfrak{D}_{(X_\Sigma,H)}$ over $\Bbbk\widehat{[P]}$ containing $\mathfrak{D}_{(X_\Sigma,H),\mathrm{in}}$ such that $\mathfrak{D}_{(X_\Sigma,H)} \setminus \mathfrak{D}_{(X_\Sigma,H),\mathrm{in}}$ consists only of nonincoming walls.

6.2. ***Generalized cluster scattering diagrams.*** Instead of applying Theorem 6.9 to (X_Σ, H), there is another way to obtain the same scattering diagram by generalizing the construction of cluster scattering diagrams in [Gross et al. 2018].

Given fixed data Γ and an $\mathcal{A}$-seed $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{p})$ with principal coefficients, we are going to define the *generalized cluster scattering diagram* $\mathfrak{D}_{\boldsymbol{s}}$.

Recall that we have the semifield $\mathbb{P} = \mathrm{Trop}(\boldsymbol{p})$, isomorphic to $\prod_{i\in I}\mathbb{Z}^{r_i}$ as an abelian group. Let $P = P_{\boldsymbol{s}}$ as before be $M \oplus \prod_{i\in I}\mathbb{N}^{r_i}$, but regarded as a submonoid of $M \oplus \mathbb{P}$ generated by M and $\boldsymbol{p}$. There is a submonoid $P^\oplus = P^\oplus_{\boldsymbol{s}} \subset P$ generated by elements

$$\{(w_i, p_{i,j}) \mid i \in I,\ j \in [1, r_i]\}.$$

One could take the completion of $P^\oplus$ with respect to the ideal $P^+ := P^\oplus \setminus \{0\}$, giving that $\widehat{\Bbbk[P^\oplus]} \subset \widehat{\Bbbk[P]}$. In N, there is a submonoid $N^\oplus_{\boldsymbol{s}} = N^\oplus$ generated by $\{e_i \mid i \in I\}$. Denote $N^+ = N^\oplus \setminus \{0\}$. We also consider the monoid map

$$\pi : P^\oplus \to N^\oplus, \quad (w_i, p_{i,j}) \mapsto e_i.$$

Let $n = \sum_{i\in I}\alpha_i e_i \in N$. Define

$$\bar{n} := \sum_{i\in I}\alpha_i \cdot \frac{d_i}{r_i} e_i \in N_{\mathbb{R}}.$$

These $\bar{n}$ form a sublattice $\overline{N}$ of $N_{\mathbb{R}}$ isomorphic to N. We have the similar notion $\overline{N}^+$, the monoid generated by $\bar{e}_i$.

There is a subspace $\mathfrak{g}$ of the tropical vertex lie algebra $\mathfrak{v}$ defined as

$$\mathfrak{g} = \mathfrak{g}_s := \bigoplus_{n\in N^+}\mathfrak{g}_n, \quad \mathfrak{g}_n := \bigoplus_{\substack{\pi(p)=n\\ p\in P^+}} z^p \cdot (\Bbbk \otimes \bar{n}).$$

Lemma 6.10. *The subspace* $\mathfrak{g}$ *is an* N^+*-graded Lie subalgebra of* $\mathfrak{v}$.

Proof. For any $n = \sum_{i\in I}\alpha_i e_i \in N^+$, consider the elements

$$\prod_{i,j} p_{i,j}^{c_{i,j}} \cdot z^{p^*(n)}$$

such that $\sum_{j\in[1,r_i]} c_{i,j} = \alpha_i$ and

$$p^*(n) := \omega(-, \bar{n}) = \sum_{i\in I}\alpha_i\omega(-, d_i e_i/r_i) = \sum_{i\in I}\alpha_i w_i.$$

Those elements form a basis of the vector space $\mathfrak{g}_n$. We check that for two such elements

$$\begin{aligned}[p_1 z^{p^*(n_1)}\partial_{\bar{n}_1}, p_2 z^{p^*(n_2)}\partial_{\bar{n}_2}] &= p_1 p_2 \cdot z^{p^*(n_1+n_2)}\partial_{\omega(\bar{n}_1,\bar{n}_2)\bar{n}_2 - \omega(\bar{n}_2,\bar{n}_1)\bar{n}_1}\\ &= \omega(\bar{n}_1, \bar{n}_2) p_1 p_2 \cdot z^{p^*(n_1+n_2)}\partial_{\bar{n}_1+\bar{n}_2} \in \mathfrak{g}_{n_1+n_2}.\end{aligned}$$ □

Remark 6.11. One may also view the above Lie algebra $\mathfrak{g}$ as being $\overline{N}^+$-graded where both $\overline{N}$ and N are sublattices of $N_{\mathbb{R}}$. When later considering a scattering diagram $\mathfrak{D}$ over an $\overline{N}^+$-graded Lie algebra $\mathfrak{g}$ (instead of N^+-graded), the walls live in $M_{\mathbb{R}}$ with integral normal vectors in $\overline{N}^+$.

Consider the ideals $(N^+)^k \subset N^+$ for $k \geq 1$. These correspond to the monomial ideals $(P^+)^k$. Then we have quotient Lie algebras (and their corresponding groups $G^{<k}$)

$$\mathfrak{g}^{<k} := \mathfrak{g}/\mathfrak{g}_{(N^+)^k} = \bigoplus_{n \in N^+ \setminus (N^+)^k} \mathfrak{g}_n,$$

and their projective limits

$$\hat{\mathfrak{g}} = \prod_{n \in N^+} \mathfrak{g}_n \quad \text{and} \quad G := \exp(\hat{\mathfrak{g}}).$$

The group $G^{<k}$ acts on $\Bbbk[P^\oplus]/(P^+)^k$ by automorphisms as in Remark 6.2.

For $n_0 \in N^+$ primitive, we define as in [Gross et al. 2018] a Lie algebra (and its corresponding pro-unipotent group)

$$\mathfrak{g}^{\parallel}_{n_0} := \bigoplus_{k>0} \mathfrak{g}_{k \cdot n_0} \subset \mathfrak{g} \quad \text{and} \quad G^{\parallel}_{n_0} := \exp(\hat{\mathfrak{g}}^{\parallel}_{n_0}) \subset G.$$

There is a general framework for scattering diagrams over an N^+-graded Lie algebra (as opposed to the tropical vertex case); see [Kontsevich and Soibelman 2014, Section 2.1; Gross et al. 2018, Section 1.1]. In this case, one could make use of an existence-and-uniqueness theorem of [Kontsevich and Soibelman 2014] (see also [Gross et al. 2018, Theorem 1.21]) to obtain a consistent scattering diagram with certain prescribed *incoming data*. The cluster scattering diagram of [Gross et al. 2018] can be defined this way, which we will extend to the generalized case in Definition 6.17.

Definition 6.12. A *wall* in $M_\mathbb{R}$ (for N^+ and an N^+-graded Lie algebra $\mathfrak{g}$) is a pair $(\mathfrak{d}, g_\mathfrak{d})$ such that

(1) $g_\mathfrak{d}$ belongs to $G^{\parallel}_{n_0}$ for some primitive $n_0 \in N^+$;

(2) $\mathfrak{d} \subset n_0^\perp \subset M_\mathbb{R}$ is a codimension one convex rational polyhedral cone.

Remark 6.13. The above definition works for general N^+-graded Lie algebras. In the case that $\mathfrak{g}$ is a Lie subalgebra of the tropical vertex Lie algebra $\mathfrak{v}$, the group $G^{\parallel}_{n_0}$ is embedded in $\mathrm{Aut}(\widehat{\Bbbk[P^\oplus]})$. Then the wall-crossing element $g_\mathfrak{d}$ can be equivalently represented by a function $f_\mathfrak{d} \in \widehat{\Bbbk[P^\oplus]}$.

Now every wall has a direction $-p^*(n_0) \in M$ in the sense of Definition 6.7. We call a wall $(\mathfrak{d}, g_\mathfrak{d})$ with direction m_0 *incoming* if $\mathfrak{d} = \mathfrak{d} - \mathbb{R}_{\geq 0} m_0$ and *nonincoming* (or *outgoing*) otherwise.

Definition 6.14. A *scattering diagram* over an N^+-graded algebra $\mathfrak{g}$ in $M_\mathbb{R}$ is a collection of walls such that for every degree $k > 0$, there are only a finite number of $(\mathfrak{d}, g_\mathfrak{d}) \in \mathfrak{D}$ with the image of $g_\mathfrak{d}$ in $G^{<k}$ not being identity.

The path-ordered product of a path $\gamma : [0, 1] \to M_\mathbb{R}$ for a scattering diagram $\mathfrak{D}$ over $\mathfrak{g}$ can be defined similarly as in Definition 6.3. We note that when γ crosses a wall $(\mathfrak{d}, g_\mathfrak{d})$ at time t, then the element $\xi_{\gamma,\mathfrak{d}}$ also depends on $\gamma'(t)$:

$$\xi_{\gamma,\mathfrak{d}} = \begin{cases} g_\mathfrak{d} & \text{if } \langle n_0, \gamma'(t) \rangle > 0, \\ g_\mathfrak{d}^{-1} & \text{if } \langle n_0, \gamma'(t) \rangle < 0. \end{cases}$$

The consistency for these scattering diagrams is defined using path-ordered products in the same way as Definition 6.3.

Theorem 6.15 [Kontsevich and Soibelman 2014, Proposition 2.1.12; Gross et al. 2018, Theorem 1.21]. *Let $\mathfrak{D}_{\mathrm{in}}$ be a scattering diagram over $\mathfrak{g}$ consisting only of incoming walls. Then there exists a unique* (*up to equivalence*) *consistent scattering diagram $\mathfrak{D}$ containing $\mathfrak{D}_{\mathrm{in}}$ such that $\mathfrak{D} \setminus \mathfrak{D}_{\mathrm{in}}$ consists only of outgoing walls.*

Now we get back to the cluster situation. Suppose given fixed data Γ and $\boldsymbol{s}$ an $\mathcal{A}$-seed with principal coefficients. Unlike the previous section, here we do not assume the maximality of the positive integers r_i, i.e., r_i needs not to be $\gcd(b_{ki} \mid k \in I)$.

We calculate in the following how the group $G_{n_0}^{\parallel}$ is embedded in $\mathrm{Aut}(\widehat{\Bbbk[P^{\oplus}]})$. Suppose $n_0 = \sum_{i \in I} \alpha_i e_i$, a primitive element in N^+. Consider any element

$$x = \sum_{k>0} \sum_{\substack{p \in \mathbb{P}^{\oplus} \\ \pi(p)=kn_0}} c_p \cdot p \cdot z^{kp^*(n_0)} \partial_{k\bar{n}_0} \in \hat{\mathfrak{g}}_{n_0}^{\parallel}, \quad c_p \in \Bbbk.$$

For nonzero $n \in N_{\mathbb{Q}}$, denote by $\mathrm{ind}(n)$ the largest number in $\mathbb{Q}_{\geq 0}$ such that $n/\mathrm{ind}(n) \in N$. Thus $n/\mathrm{ind}(n)$ is primitive in N.

Lemma 6.16. *The group element* $\exp(x) \in G_{n_0}^{\parallel}$ *acts on* $\widehat{\Bbbk[P^{\oplus}]}$ *as an automorphism by*

$$z^m \mapsto z^m \exp\Bigl(\sum_{k>0} \sum_{\substack{p \in \mathbb{P}^{\oplus} \\ \pi(p)=kn_0}} \mathrm{ind}(\bar{n}_0) k c_p \cdot p \cdot z^{kp^*(n_0)}\Bigr)^{\langle r(m), \bar{n}_0/\mathrm{ind}(\bar{n}_0)\rangle}, \quad m \in P^{\oplus}.$$

Proof. This follows by rewriting x as

$$x = \Bigl(\sum_{k>0} \sum_{\substack{p \in \mathbb{P}^{\oplus} \\ \pi(p)=kn_0}} \mathrm{ind}(\bar{n}_0) k c_p \cdot p \cdot z^{kp^*(n_0)}\Bigr) \partial_{\bar{n}_0/\mathrm{ind}(\bar{n}_0)}. \qquad \square$$

Due to Lemma 6.16, any $\exp(x) \in G_{n_0}^{\parallel}$ can be represented by a function f as in Lemma 6.16 such that the action of $\exp(x)$ sends z^m to $z^m f^{\langle r(m), \bar{n}_0/\mathrm{ind}(\bar{n}_0)\rangle}$.

Given $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{p})$, for each $i \in I$, consider the hyperplane $e_i^{\perp}$ with the attached wall-crossing function

$$f_i = \prod_{j=1}^{r_i} (1 + p_{i,j} z^{w_i}) \in \Bbbk[P^{\oplus}].$$

As discussed, the function f_i represents an element in $G_{e_i}^{\parallel}$.

Definition 6.17. Let $\mathfrak{D}_{\boldsymbol{s},\mathrm{in}}$ be the scattering diagram over $\mathfrak{g}$ in $M_{\mathbb{R}}$ consisting only of the incoming walls of the form $\mathfrak{d}_i := (e_i^{\perp}, f_i)$, i.e.,

$$\mathfrak{D}_{\boldsymbol{s},\mathrm{in}} := \{(e_i^{\perp}, f_i) \mid i \in I\}.$$

We define *the generalized cluster scattering* $\mathfrak{D}_{\boldsymbol{s}}$ to be the unique (up to equivalence) consistent scattering diagram associated to $\mathfrak{D}_{\boldsymbol{s},\mathrm{in}}$ guaranteed by Theorem 6.15.

Remark 6.18. One may tend to think of $\mathfrak{D}_s$ as a scattering diagram over $\widehat{\Bbbk[P^\oplus]}$ or over $\widehat{\Bbbk[P]}$ (as $\mathfrak{g}$ is a Lie subalgebra of $\mathfrak{v}$) in Definition 6.5. However there is one subtle issue. Suppose that there is a wall $(\mathfrak{d} \subset n_0^\perp, f_\mathfrak{d})$ in $\mathfrak{D}_s$ for some $n_0 \in N^+$ primitive. Then the wall-crossing action is given by

$$\xi_{f_\mathfrak{d}}(z^p) = z^p f_\mathfrak{d}^{\langle \bar{n}_0/\mathrm{ind}(\bar{n}_0), r(p)\rangle}.$$

Since in general $\bar{n}_0$ may not be proportional to n_0, the cone $\mathfrak{d}$ may not be contained in $\bar{n}_0^\perp$. In this case, the wall $(\mathfrak{d}, f_\mathfrak{d})$ does not qualify as a wall in Definition 6.5. This issue can be resolved in the following two ways (so that one can view $\mathfrak{D}_s$ as a scattering diagram of Definition 6.5).

(1) We could regard $\mathfrak{g}$ as graded by $\overline{N}^+ \subset N_\mathbb{Q}$ (rather than N^+-graded) and modify Definition 6.12 (the definition of a wall $(\mathfrak{d}, g_\mathfrak{d})$) so that $\mathfrak{d}$ is a codimension one cone in some hyperplane $n_0^\perp$ for $n_0 \in \overline{N}^+$ and $g_\mathfrak{d}$ belongs to $G_{n_0}^\parallel$.

(2) Another way to resolve the issue is to consider the dual $\eta^* : M_\mathbb{R} \to M_\mathbb{R}$ of the linear map

$$\eta : N_\mathbb{R} \to N_\mathbb{R}, \quad n \mapsto \bar{n}.$$

We then apply $(\eta^*)^{-1}$ to every wall $(\mathfrak{d}, f_\mathfrak{d})$ to get the collection

$$(\eta^*)^{-1}(\mathfrak{D}_s) := \{((\eta^*)^{-1}(\mathfrak{d}), f_\mathfrak{d}) \mid (\mathfrak{d}, f_\mathfrak{d}) \in \mathfrak{D}_s\}$$

Then the cone $(\eta^*)^{-1}(\mathfrak{d})$ is indeed contained in $\bar{n}_0^\perp$. So this collection of walls is a scattering diagram in Definition 6.5.

From now on, to avoid any further confusion, the notation $\mathfrak{D}_s$ is reserved for the consistent scattering diagram $(\eta^*)^{-1}(\mathfrak{D}_s)$ over $\widehat{\Bbbk[P^\oplus]}$.

Lemma 6.19. *Let s be a seed with principal coefficients for some generalized fixed data Γ (viewed of both $\mathcal{A}$- and $\mathcal{X}$-type) with the condition that for each $i \in I$, the element*

$$w_i = \omega(-, d_i e_i / r_i)$$

is primitive in M. In this case, we have defined both scattering diagrams $\mathfrak{D}_{(X_\Sigma, H)}$ (with a chosen general evaluation λ) and $\mathfrak{D}_s$. Identify the parameters $t_{i,j}$ with $p_{i,j}$. Then $\mathfrak{D}_{(X_\Sigma, H)}$ and $\mathfrak{D}_s$ are equivalent as scattering diagrams over $\widehat{\Bbbk[P]}$.

Proof. We require ω_i to be primitive so that $\mathfrak{D}_{(X_\Sigma, H)}$ is defined. According to Remark 6.18, $\mathfrak{D}_s$ is viewed as a scattering diagram over $\widehat{\Bbbk[P]}$ in the same $M_\mathbb{R}$ as $\mathfrak{D}_{(X_\Sigma, H)}$ so it is legitimate to compare them. Let $\widetilde{\mathfrak{D}}$ be the consistent scattering diagram over $\mathfrak{g}$ obtained using the initial data $\mathfrak{D}_{(X_\Sigma, H),\mathrm{in}}$. Notice that the walls in $\mathfrak{D}_{(X_\Sigma, H),\mathrm{in}}$ are parts of the hyperplanes $e_i^\perp$. We then subdivide the walls in $\mathfrak{D}_{s,\mathrm{in}}$ so that $\mathfrak{D}_{(X_\Sigma, H),\mathrm{in}}$ becomes the subset of incoming walls. Thus $\widetilde{\mathfrak{D}}$ is equivalent to $\mathfrak{D}_s$ by Theorem 6.15.

On the other hand, $\widetilde{\mathfrak{D}}$ is also a scattering diagram over $\widehat{\Bbbk[P]}$. By Theorem 6.9, It is also equivalent to $\mathfrak{D}_{(X_\Sigma, H)}$ since they have the same incoming walls. Therefore we have $\mathfrak{D}_{(X_\Sigma, H)} \cong \widetilde{\mathfrak{D}} \cong \mathfrak{D}_s$. □

6.3. ***The cluster scattering diagrams of GHKK.*** The ordinary cluster scattering diagram $\mathfrak{D}_s^{\mathrm{ord}}$ corresponds to the case where $r_i = 1$ for each $i \in I$. Thus there is only one parameter $p_i := p_{i,1}$ for each $i \in I$. The lattice $\overline{N}$ is generated by $\bar{e}_i = d_i e_i$. The initial incoming walls are then

$$\{(e_i^{\perp}, 1 + p_i z^{w_i}) \mid i \in I\},$$

where $w_i = \omega(-, \bar{e}_i) \in M$.

This scattering diagram is closely related to the *cluster scattering diagram* $\mathfrak{D}_s^{\mathrm{GHKK}}$ of Gross, Hacking, Keel and Kontsevich [Gross et al. 2018, Theorem 1.12]. We explain the difference and relation here. The scattering diagram $\mathfrak{D}_s^{\mathrm{GHKK}}$ is actually defined for $\overline{N}$ and $\overline{M} := \mathrm{Hom}(\overline{N}, \mathbb{Z})$ (in the ordinary case equal to N° and M° respectively). Under the injectivity assumption [Gross et al. 2018, Section 1.1], the incoming walls are

$$\{(e_i^{\perp}, 1 + z^{\omega(e_i,-)}) \mid i \in I\},$$

where $\omega(e_i, -)$ is in M°. The injectivity assumption means that $\omega(e_i, -)$ generate a strict convex cone. If this is not the case, we may extend M° to $M^{\circ} \oplus \mathbb{P}$ (identified with $M^{\circ} \oplus N$ in [Gross et al. 2018]) and let incoming walls be

$$\{(e_i^{\perp}, 1 + p_i z^{\omega(e_i,-)}) \mid i \in I\}.$$

It lives in $(M^{\circ} \oplus N) \otimes \mathbb{R}$, or in $M^{\circ} \otimes \mathbb{R}$ if regarding p_i as formal parameters as we do. Then $\mathfrak{D}_s^{\mathrm{GHKK}}$ is defined to be the unique consistent scattering diagram over $\widehat{\Bbbk[P]}$ with only these incoming walls, where $P \subset M^{\circ} \oplus N$ is a submonoid contained in a strictly convex cone and containing the cone generated by $(p_i, \omega(e_i, -))$. The Lie algebra $\mathfrak{g}$, however, is naturally graded by N^{+} (generated by e_i's), not $\overline{N}^{+}$ (generated by $\bar{e}_i$'s). Thus if one uses Theorem 6.15 to define $\mathfrak{D}_s^{\mathrm{GHKK}}$, the same rescaling issue in Remark 6.18 still exists and can be resolved in a similar way. In [Gross et al. 2018], $\mathfrak{D}_s^{\mathrm{GHKK}}$ is regarded as living in $M_{\mathbb{R}}^{\circ}$ with the integral normal vectors of walls being in N°.

The structures of $\mathfrak{D}_s^{\mathrm{GHKK}}$ and $\mathfrak{D}_s^{\mathrm{ord}}$ are very much alike. For example, they both admit cluster complex structures; see [Gross et al. 2018, Theorem 2.13] and Theorem 7.10. It turns out that in the convention of [Fomin and Zelevinsky 2007] (e.g., the definition of g-vectors), $\mathfrak{D}_s^{\mathrm{GHKK}}$ corresponds to the cluster algebra associated to $-B^T$ while $\mathfrak{D}_s^{\mathrm{ord}}$ corresponds to the one associated to B, where $B = (b_{ij})$ with $b_{ij} = \omega(e_i, \bar{e}_j)$.

6.4. ***Scattering diagrams with special coefficients.*** Just as specializing a cluster algebra $\mathscr{A}$ at some evaluation $\lambda : \mathbb{P} \to \Bbbk^*$, one can do the same to $\mathfrak{D}_s$, obtaining a consistent scattering diagram with special coefficients.

We consider another monoid $Q = M \oplus \prod_{i \in I} \mathbb{N}$ (with t_i being the standard generators of $\prod_{i\in I} \mathbb{N}$). Let $\lambda : \mathbb{P} \to \Bbbk^*$, $p_{i,j} \mapsto \lambda_{i,j}$ be an evaluation. Define the map (abusing the same notation λ)

$$\lambda : \Bbbk[P] \to \Bbbk[Q], \quad z^m \mapsto z^m \text{ for } m \in M, \quad p_{i,j} \mapsto \lambda_{i,j} t_i.$$

Lemma 6.20. *The collection*

$$\lambda(\mathfrak{D}_{\boldsymbol{s}}) := \{(\mathfrak{d}, \lambda(f_{\mathfrak{d}})) \mid (\mathfrak{d}, f_{\mathfrak{d}}) \in \mathfrak{D}_{\boldsymbol{s}}\}$$

obtained by applying the algebra homomorphism λ to every wall-crossing function $f_{\mathfrak{d}}$ is a consistent scattering diagram over $\widehat{\Bbbk[Q]}$.

Proof. The algebra homomorphism λ respects the completions of $\Bbbk[P]$ and $\Bbbk[Q]$. So $\lambda(f_{\mathfrak{d}})$ belongs to $\widehat{\Bbbk[Q]}$. Recall we have the monoid map $r : P \to M$ which forgets the components in $\bigoplus_{i\in I} \mathbb{N}^{r_i}$. We use the same notation $r : Q \to M$ for the analogous map on Q. Then $(\mathfrak{d}, \lambda(f_{\mathfrak{d}}))$ becomes a wall over $\widehat{\Bbbk[Q]}$, and $\lambda(\mathfrak{D}_{\boldsymbol{s}})$ is a scattering diagram over $\widehat{\Bbbk[Q]}$.

The consistency of $\lambda(\mathfrak{D}_{\boldsymbol{s}})$ follows from the consistency of $\mathfrak{D}_{\boldsymbol{s}}$ as λ is an algebra homomorphism. □

We call $\lambda(\mathfrak{D}_{\boldsymbol{s}})$ the (generalized) cluster scattering diagram of $\boldsymbol{s}$ with special coefficients λ. In fact, the ordinary cluster scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$ when $r_i = 1$ can be obtained this way. We denote the ordinary one by $\mathfrak{D}_{\boldsymbol{s}}^{\mathrm{ord}}$. Its incoming walls are

$$(e_i^{\perp}, 1 + p_i z^{\omega(-,d_i e_i)}).$$

If there exist coefficients $\lambda_{ij} \in \Bbbk^*$ such that

$$\prod_{j=1}^{r_i} (1 + \lambda_{ij} t_i z^{w_i}) = 1 + t_i^{r_i} z^{r_i w_i} = 1 + t_i^{r_i} z^{\omega(-,d_i e_i)},$$

then we can apply the corresponding morphism $\lambda : \Bbbk[P] \to \Bbbk[Q]$ to $\mathfrak{D}_{\boldsymbol{s}}$ so that

$$\lambda(\mathfrak{D}_{\boldsymbol{s}}) \cong \mathfrak{D}_{\boldsymbol{s}}^{\mathrm{ord}}$$

as they have the exact same set of incoming walls. Here $t_i^{r_i}$ is identified with p_i. The existence of such an evaluation λ amounts to find the r_i roots of the polynomial $1 + x^{r_i}$ in $\Bbbk$, which is always possible if $\Bbbk$ is algebraically closed.

6.5. *Examples.*

We illustrate some examples of generalized cluster scattering diagrams in this section.

Example 6.21. Consider the fixed data Γ consisting of

- the lattice $N = \mathbb{Z}^2$ with the standard basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$, and the skew-symmetric form ω be determined by $\omega(e_1, e_2) = -1$;
- $N_{\mathrm{uf}} = N$;
- the rank $r = 2$ and $I = I_{\mathrm{uf}} = \{1, 2\}$;
- positive integers $d_1 = 1$ and $d_2 = 2$;
- the sublattice N° generated by e_1 and $2e_2$;
- $M = \mathrm{Hom}(N, \mathbb{Z})$, $M^\circ = \mathrm{Hom}(N^\circ, \mathbb{Z})$.

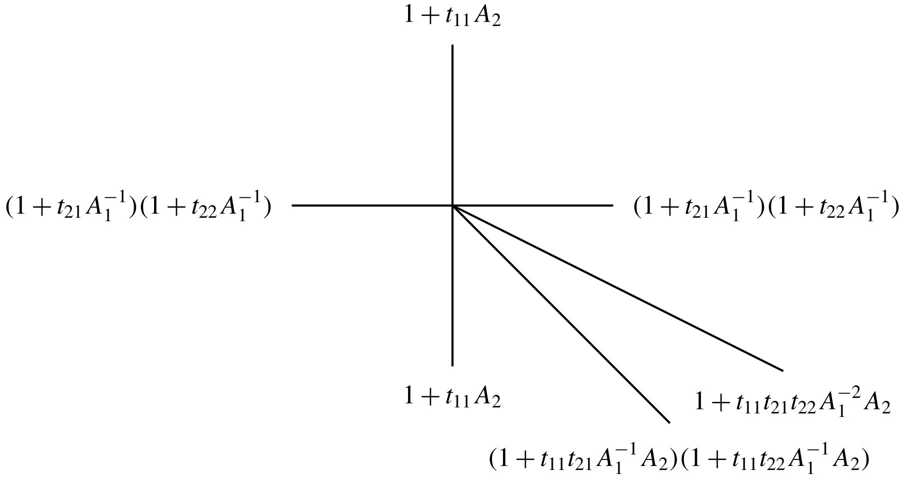
$1+t_{11}A_2$
$(1+t_{21}A_1^{-1})(1+t_{22}A_1^{-1})$
$(1+t_{21}A_1^{-1})(1+t_{22}A_1^{-1})$
$1+t_{11}A_2$
$1+t_{11}t_{21}t_{22}A_1^{-2}A_2$
$(1+t_{11}t_{21}A_1^{-1}A_2)(1+t_{11}t_{22}A_1^{-1}A_2)$

Figure 1. The generalized cluster scattering diagram for Example 6.21.

Let $\boldsymbol{s}$ be a seed consisting of $\boldsymbol{e}=(e_1,e_2)$ and $\boldsymbol{p}_1=(t_{11})$, $\boldsymbol{p}_2=(t_{21},t_{22})$. We have matrices

$$B=\begin{pmatrix}0&-2\\1&0\end{pmatrix}\quad\text{and}\quad\beta=\begin{pmatrix}0&-1\\1&0\end{pmatrix}.$$

In this case we have $\bar{e}_i=d_ie_i/r_i=e_i$. So $\overline{N}=N$ and we shall not worry about the rescaling issue. Then $w_1=e_2^*$ and $w_2=-e_1^*$. We write $A_i=z^{e_i^*}$ for $i=1,2$. The coefficients group is $\mathbb{P}=\mathbb{Z}^3$ with generators $\{t_{11},t_{21},t_{22}\}$. The initial incoming scattering diagram is

$$\mathfrak{D}_{\boldsymbol{s},\mathrm{in}}=\{(e_1^\perp,1+t_{11}A_2),(e_2^\perp,(1+t_{21}A_1^{-1})(1+t_{22}A_1^{-1}))\}.$$

The resulting generalized cluster scattering diagram is

$$\mathfrak{D}_{\boldsymbol{s}}=\mathfrak{D}_{\boldsymbol{s},\mathrm{in}}\cup\{(\mathbb{R}_{>0}(1,-1),f_{(1,-1)}),(\mathbb{R}_{>0}(2,-1),f_{(2,-1)})\},$$

where

$$f_{(1,-1)}=(1+t_{11}t_{21}A_1^{-1}A_2)(1+t_{11}t_{22}A_1^{-1}A_2)\quad\text{and}\quad f_{(2,-1)}=1+t_{11}t_{21}t_{22}A_1^{-2}A_2.$$

The scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$ is depicted in Figure 1.

Example 6.22. Consider the fixed data Γ consisting of

- the lattice $N=\mathbb{Z}^2$ with the standard basis $e_1=(1,0)$ and $e_2=(0,1)$, and the skew-symmetric form ω be determined by $\omega(e_1,e_2)=-1$;
- $N_{\mathrm{uf}}=N$;
- the rank $r=2$ and $I=I_{\mathrm{uf}}=\{1,2\}$;
- positive integers $\lambda_1=1$ and $\lambda_2=1$;
- the sublattice N° generated by e_1 and e_2;
- $M=\mathrm{Hom}(N,\mathbb{Z})$, $M^\circ=\mathrm{Hom}(N^\circ,\mathbb{Z})$.

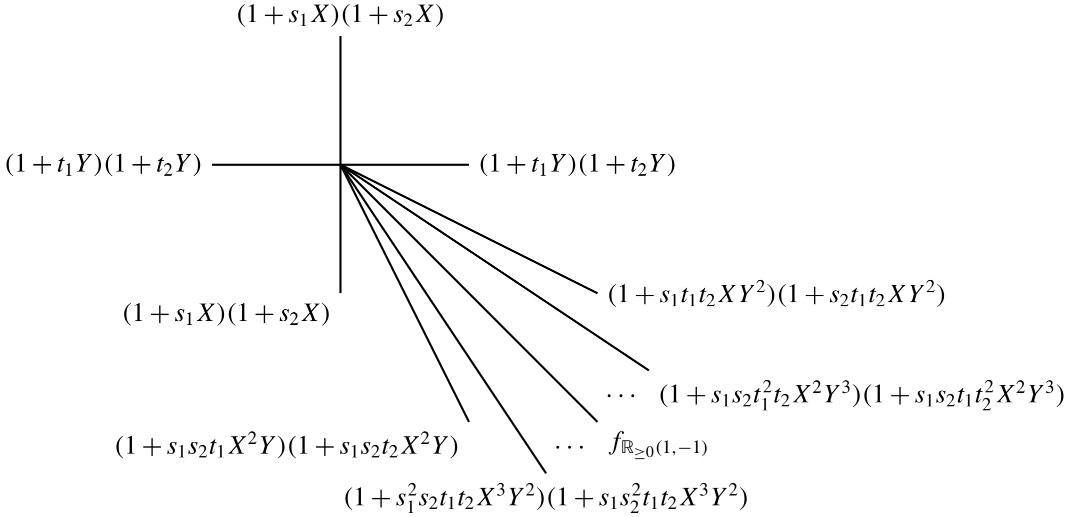
$(1+s_1X)(1+s_2X)$
$(1+t_1Y)(1+t_2Y)$
$(1+t_1Y)(1+t_2Y)$
$(1+s_1X)(1+s_2X)$
$(1+s_1t_1t_2XY^2)(1+s_2t_1t_2XY^2)$
$\cdots$ $(1+s_1s_2t_1^2t_2X^2Y^3)(1+s_1s_2t_1t_2^2X^2Y^3)$
$(1+s_1s_2t_1X^2Y)(1+s_1s_2t_2X^2Y)$
$\cdots$ $f_{\mathbb{R}_{\geq 0}(1,-1)}$
$(1+s_1^2s_2t_1t_2X^3Y^2)(1+s_1s_2^2t_1t_2X^3Y^2)$

Figure 2. The generalized cluster scattering diagram for Example 6.22.

The seed is given by $\boldsymbol{e}=(e_1,e_2)$ and $\boldsymbol{p}_1=(s_1,s_2)$, $\boldsymbol{p}_2=(t_1,t_2)$. The corresponding $\mathfrak{D}_{\boldsymbol{s}}$ is depicted in Figure 2. We write $X=z^{e_2^*}$ and $Y=z^{-e_1^*}$. The five rays depicted in the fourth quadrant are in the directions $(2,-1)$, $(3,-2)$, $(1,-1)$, $(2,-3)$ and $(1,-2)$ in clockwise order. In fact, in the fourth quadrant there are additional nontrivial walls whose underlying cones are $\mathbb{R}_{\geq 0}(n,-(n+1))$ and $\mathbb{R}_{\geq 0}(n+1,-n)$ for each positive integer $n\geq 3$ (which we omit in the figure below). The wall-crossing function, for example for $\mathbb{R}_{\geq 0}(2k,-(2k+1))$ for $k\in\mathbb{Z}_{>0}$, is

$$f_{\mathbb{R}_{\geq 0}(2k,-(2k+1))}=(1+s_1^{k+1}s_2^kt_1^kt_2^kX^{2k+1}Y^{2k})(1+s_1^ks_2^{k+1}t_1^kt_2^kX^{2k+1}Y^{2k}),$$

which can be obtained using Theorem 7.10.

The wall-crossing function attached to the ray $\mathbb{R}_{\geq 0}(1,-1)$

$$f_{\mathbb{R}_{\geq 0}(1,-1)}=\frac{(1+s_1t_1XY)(1+s_1t_2XY)(1+s_2t_1XY)(1+s_2t_2XY)}{(1-s_1s_2t_1t_2X^2Y^2)^4}$$

is much more difficult to calculate. This was explicitly obtained by Reineke and Weist [2013] by relating the wall-crossing functions to quiver representations.

6.6. ***Mutation invariance of*** $\mathfrak{D}_{\boldsymbol{s}}$**.** A first step to investigate the structure of $\mathfrak{D}_{\boldsymbol{s}}$ is through a comparison with $\mathfrak{D}_{\mu_k(\boldsymbol{s})}$. For the ordinary case, this is called the *mutation invariance* in [Gross et al. 2018]. In the generalized situation, we show an analogous mutation invariance still holds. One just needs to take care of the generalized coefficients $p_{i,j}$.

Notice that the definition of $\mathfrak{D}_{\boldsymbol{s}}$ does not involve the semifield structure of $\mathbb{P}$. So one can view that the coefficients part $\boldsymbol{p}$ actually provides a $\mathbb{Z}$-basis of the multiplicative abelian group $\mathbb{P}$ (grouped and labeled in a certain way). Thus even though $\mu_k(\boldsymbol{s})$ no longer has principal coefficients in $\mathbb{P}$, $\mathfrak{D}_{\mu_k(\boldsymbol{s})}$ is still defined. To stress that the coefficients are no longer semifield elements, we use $t_{i,j}$ instead of $p_{i,j}$.

Now $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{t})$ consists of $\boldsymbol{e}$ a labeled basis of N and tuples of coefficients $\boldsymbol{t} = (\boldsymbol{t}_i)$.

Definition 6.23. Define the mutation $\mu_k^+(\boldsymbol{s}) = (\boldsymbol{e}', \boldsymbol{t}')$ such that $\boldsymbol{e}' = \mu_k(\boldsymbol{e})$ as before and for the coefficients,

$$t'_{i,j} = \begin{cases} t_{k,j}^{-1} & \text{if } i = k, \\ t_{i,j} \cdot \prod_{l=1}^{r_k} t_{k,l}^{[\beta_{ki}]_+} & \text{if } i \neq k. \end{cases}$$

Remark 6.24. Note that this mutation does not depend on any semifield structure on $\mathbb{P}$. So it is different from the μ_k from Definition 4.3 for mutations of many steps. For this reason, we call $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{p})$ a seed with coefficients (avoiding the type $\mathcal{A}$- or $\mathcal{X}$-) and use the new symbol μ_k^+ for mutations in this context (as we will see in Section 7.1 the meaning of the sign $+$).

Definition 6.25. We set

$$\mathcal{H}_{k,+} := \{m \in M_\mathbb{R} \mid \langle e_k, m\rangle \geq 0\}, \quad \mathcal{H}_{k,-} := \{m \in M_\mathbb{R} \mid \langle e_k, m\rangle \leq 0\}.$$

For $k \in I$, define the piecewise linear transformation $T_k : M_\mathbb{R} \to M_\mathbb{R}$ by

$$T_k(m) := \begin{cases} m + \langle e_k, m\rangle r_k w_k, & m \in \mathcal{H}_{k,+}, \\ m, & m \in H_{k,-}. \end{cases}$$

One sees that in the two half spaces, the map T_k is actually the restriction of two linear maps $T_{k,+}$ and $T_{k,-}$ respectively. The map T_k is with respect to the seed $\boldsymbol{s}$ and thus sometimes will be denoted as $T_k^{\boldsymbol{s}}$. The vector $r_k w_k$ can also be expressed as $r_k w_k = \omega(-, d_k e_k) = \sum_{i=1}^n b_{ik} e_i^*$. One checks that

$$T_{k,+}(w_i) = w_i + \beta_{ki} r_k w_k.$$

Recall we have the projection $r : M \oplus \mathbb{P} \to M$. The transformation T_k can be lifted to $M \oplus \mathbb{P}$ by

$$\widetilde{T}_k(m, p) := \begin{cases} (m + \langle e_k, m\rangle r_k w_k,\ p \cdot t_k^{\langle e_k, m\rangle}), & m \in \mathcal{H}_{k,+}, \\ (m, p), & m \in \mathcal{H}_{k,-}, \end{cases}$$

where $t_k = \prod_{l=1}^{r_k} t_{k,l}$. Note that $\widetilde{T}_k$ on its domain of linearity is the restriction of two linear transformations $\widetilde{T}_{k,\varepsilon}$ respectively.

Construction 6.26. We define the scattering diagram $T_k(\mathfrak{D}_{\boldsymbol{s}})$ as in [Gross et al. 2018, Definition 1.22] (but taking care of the parameters $t_{i,j}$ here) in the following steps.

(1) Replace each wall in $\mathfrak{D}_{\boldsymbol{s}}$ not fully contained in $e_k^\perp$ if necessary by splitting it into two new walls

$$(\mathfrak{d} \cap \mathcal{H}_{k,+}, f_\mathfrak{d}) \quad \text{and} \quad (\mathfrak{d} \cap \mathcal{H}_{k,-}, f_\mathfrak{d}).$$

Regard this new collection of walls as the current representative of $\mathfrak{D}_{\boldsymbol{s}}$.

(2) For a wall $(\mathfrak{d}, f_\mathfrak{d})$ contained in $\mathcal{H}_{k,\varepsilon}$, define the wall $T_{k,\varepsilon}(\mathfrak{d}, f_\mathfrak{d}) = (T_{k,\varepsilon}(\mathfrak{d}), \widetilde{T}_{k,\varepsilon}(f_\mathfrak{d}))$ where the new wall-crossing function $\widetilde{T}_{k,\varepsilon}(f_\mathfrak{d})$ is the one obtained from $f_\mathfrak{d}$ by replacing each monomial of the form

$$pz^m \quad \text{by} \quad \widetilde{T}_{k,\varepsilon}(pz^m),$$

where the later is the monomial corresponding to $\widetilde{T}_{k,\varepsilon}(m, p) \in M \oplus \mathbb{P}$. For example, we have

$$\widetilde{T}_{k,+}(t_{i,j}z^{w_i}) = t_{i,j}t_k^{\beta_{ki}}z^{w_i+\beta_{ki}r_kw_k}, \quad \text{while} \quad \widetilde{T}_{k,-}(t_{i,j}z^{w_i}) = t_{i,j}z^{w_i}.$$

We call these walls uniformly by $T_k(\mathfrak{d}, f_\mathfrak{d})$ no matter which half they belong to. We stress that the sign ε in $T_{k,\varepsilon}$ is determined by which half space the wall $\mathfrak{d}$ lies in.

(3) Consider the collection of walls

$$T_k(\mathfrak{D}_s) := \left\{ T_k(\mathfrak{d}, f_\mathfrak{d}) \,\middle|\, (\mathfrak{d}, f_\mathfrak{d}) \in \mathfrak{D}(s) \setminus \left(e_k^\perp, \prod_{j=1}^{r_k} (1 + t_{k,j}z^{w_k}) \right) \right\} \cup \left\{ \left(e_k^\perp, \prod_{j=1}^{r_k} (1 + t_{k,j}^{-1}z^{-w_k}) \right) \right\}.$$

Denote the monoid $(P')^\oplus := P^\oplus_{\mu_k^+(s)} \subset M \oplus \mathbb{P}$. While $\mathfrak{D}_s$ is over $\widehat{\Bbbk[P_s^\oplus]}$, $\mathfrak{D}_{\mu_k^+(s)}$ is over $\widehat{\Bbbk[(P')^\oplus]}$.

Theorem 6.27 (cf. [Gross et al. 2018, Theorem 1.24]). *The set of walls $T_k(\mathfrak{D}_s)$ is indeed a consistent scattering diagram over $\widehat{\Bbbk[(P')^\oplus]}$, and furthermore is equivalent to $\mathfrak{D}_{\mu_k^+(s)}$.*

We find it most natural to understand the mutation invariance by making connection to the *canonical wall structure* (or canonical scattering diagram) [Gross and Siebert 2022] via [Argüz and Gross 2022, Theorem 6.1], where $\mathfrak{D}_s$ can be viewed as associated to the toric model $U_{\Omega(s,\lambda)}$ for general λ. However, as in Section 5.3.2, this would require the condition

$$r_i = \gcd(b_{ij}, i \in I).$$

Fortunately, we can prove the mutation invariance following the same strategy in [Gross et al. 2018] without this condition. The proof occupies the rest of the section.

First define a monoid $\overline{P}$ containing both $P^\oplus$ and $(P')^\oplus$. Let σ be the cone in $(M \oplus \mathbb{P})_\mathbb{R}$ generated by

$$\{(w_i, t_{i,j}) \mid i \in I, j \in [1, r_i]\} \cup \{(-w_k, -t_{k,j}) \mid 1 \le j \le r_k\}.$$

Take $\overline{P} = \sigma \cap (M \oplus \mathbb{P})$ and we tend to talk about scattering diagrams over $\widehat{\Bbbk[\overline{P}]}$. However the ideal $\mathfrak{m}_{\overline{P}}$ misses the elements $(w_k, t_{k,j})$. This means a wall such as

$$(e_k^\perp, (1 + t_{k,j}z^{w_k}))$$

in $\mathfrak{D}_s$ does not qualify as a wall over $\widehat{\Bbbk[\overline{P}]}$. For this reason, we extend the definition of scattering diagram as in [Gross et al. 2018, Definition 1.27] (slightly generalizing the *slab* for our needs).

Define

$$\overline{N}_s^{+,k} := \left\{ \sum_{i \in I} a_i \bar{e}_i \,\middle|\, a_i \in \mathbb{Z}_{\ge 0} \text{ for } i \ne k,\, a_k \in \mathbb{Z}, \text{ and } \sum_{i \in I \setminus \{k\}} a_i > 0 \right\} \subset \overline{N}.$$

Since $\overline{N}_s^{+,k} = \overline{N}^{+,k}_{\mu_k^+(s)}$, we denote them by $\overline{N}^{+,k}$.

Definition 6.28 (cf. [Gross et al. 2018, Definition 1.27]). A *wall* for $\overline{P}$ is a pair $(\mathfrak{d}, f_\mathfrak{d})$ with $\mathfrak{d}$ as before but with primitive normal vector n_0 in $\overline{N}^{+,k}$ and

$$f_\mathfrak{d} = 1 + \sum_{k \ge 1, \pi(t) = kn_0} c_{k,t} \cdot t z^{k\omega(-,n_0)} \equiv 1 \mod \mathfrak{m}_{\overline{P}}.$$

The *slab* for s and $k \in I$ means the pair

$$\mathfrak{d}_k := \Big(e_k^{\perp}, \prod_{j=1}^{r_k}(1+t_{k,j}z^{w_k})\Big).$$

A scattering diagram $\mathfrak{D}$ for $\bar{P}$ is a collection of walls and possibly this single slab, with the condition that for each $k > 0$, $f_{\mathfrak{d}} \equiv 1 \bmod \mathfrak{m}_{\bar{P}}^k$ for all but finitely many walls in $\mathfrak{D}$.

We quote the following very hard theorem from [Gross et al. 2018]. The objects here are understood in our definitions so there are minor differences. However, one can still prove the theorem in the exact same way. So we omit its proof here.

Theorem 6.29 [Gross et al. 2018, Theorem 1.28]. *There exists a unique* (*up to equivalence*) *consistent scattering diagram* $\overline{\mathfrak{D}}_s$ *in the sense of Definition 6.28 such that*

(1) $\overline{\mathfrak{D}}_s \supseteq \mathfrak{D}_{s,\mathrm{in}}$,

(2) $\overline{\mathfrak{D}}_s \setminus \mathfrak{D}_{s,\mathrm{in}}$ *consists only of outgoing walls.*

Furthermore, $\overline{\mathfrak{D}}_s$ *is also a scattering diagram for the* $\bar{N}_s^+$*-graded Lie algebra* $\mathfrak{g}_s$. *As such*, *it is equivalent to* $\mathfrak{D}_s$.

Proof of Theorem 6.27. First we choose a representative for $\mathfrak{D}_s$ given by Theorem 6.29. Now $T_k(\mathfrak{D}_s)$ becomes a scattering diagram in the sense of Definition 6.28 for the seed $s' = \mu_k^+(s)$. This is because

(1) the operation T_k removes the old slab $\mathfrak{d}_k$ and adds the new slab

$$\mathfrak{d}_k' := \Big(e_k^{\perp}, \prod_{j=1}^{r_k}(1+t_{k,j}^{-1}z^{-w_k})\Big);$$

(2) for a wall (contained in either $\mathcal{H}_{k,+}$ or $\mathcal{H}_{k,-}$), $\widetilde{T}_k$ sends a monomial of the form $\prod_{i,j}(t_{i,j}z^{w_i})^{a_{ij}}$ in its wall-crossing function to

$$\prod_{i,j}(t_{i,j}t_k^{\beta_{ki}}z^{w_i+\beta_{ki}r_kw_k})^{a_{ij}} \quad \text{or} \quad \prod_{i,j}(t_{i,j}z^{w_i})^{a_{ij}}.$$

So if $tz^m \in \mathfrak{m}_{\bar{P}}^i$ for some i, so is $\widetilde{T}_k(tz^m)$.

We next show that

(1) $T_k(\mathfrak{D}_s)$ and $\mathfrak{D}_{s'}$ have the same set of slabs and incoming walls;

(2) $T_k(\mathfrak{D}_s)$ is consistent as a scattering diagram with a slab.

Then by the uniqueness statement of Theorem 6.29, $T_k(\mathfrak{D}_s)$ and $\mathfrak{D}_{s'}$ are equivalent.

Statement (1) follows from the same argument in *Step I* of [Gross et al. 2018, Proof of Theorem 1.24].

For (2), we check the consistency of $T_k(\mathfrak{D}_s)$, that is, for any loop γ, $\mathfrak{p}_{\gamma,T_k(\mathfrak{D}_s)} = \mathrm{id}$ whenever it is defined.

If γ is confined in one of the half spaces, the path-ordered product is identity because of the consistency of $\mathfrak{D}_s$. So we assume that γ crosses the slab $\mathfrak{d}'_k$. Split γ into four subpaths γ_1, γ_2, γ_3 and γ_4 such that

(1) γ_1 starts at a point in $\mathcal{H}_{k,-}$ and only crosses the slab $\mathfrak{d}'_k$;

(2) γ_2 is contained entirely in $\mathcal{H}_{k,+}$;

(3) γ_3 only crosses $\mathfrak{d}'_k$ back to $\mathcal{H}_{k,-}$;

(4) γ_4 is contained entirely in $\mathcal{H}_{k,-}$.

Let $\widetilde{T}_{k,+} : \Bbbk[M \oplus \mathbb{P}] \to \Bbbk[M \oplus \mathbb{P}]$ be the algebra automorphism induced by $\widetilde{T}_{k,+}$ (see (2) in the Construction 6.26 the action of $\widetilde{T}_{k,+}$ on monomials). Denote by $\mathfrak{p}_{\mathfrak{d}'_k}$ the wall-crossing automorphism

$$z^m \mapsto z^m \prod_{j=1}^{r_k} (1 + t_{k,j}^{-1} z^{-w_k})^{-\langle e_k, m\rangle}.$$

So we have

$$\mathfrak{p}_{\gamma_1, T_k(\mathfrak{D}_s)} = \mathfrak{p}_{\mathfrak{d}'_k}, \tag{6-1}$$

$$\mathfrak{p}_{\gamma_2, T_k(\mathfrak{D}_s)} = \widetilde{T}_{k,+} \circ \mathfrak{p}_{\gamma_2, \mathfrak{D}_s} \circ \widetilde{T}_{k,+}^{-1}, \tag{6-2}$$

$$\mathfrak{p}_{\gamma_3, T_k(\mathfrak{D}_s)} = \mathfrak{p}_{\mathfrak{d}'_k}^{-1}, \tag{6-3}$$

$$\mathfrak{p}_{\gamma_4, T_k(\mathfrak{D}_s)} = \mathfrak{p}_{\gamma_4, \mathfrak{D}_s}. \tag{6-4}$$

All the above equalities except (6-2) are by definitions. To show (6-2), we see that it suffices to show the case where γ_2 only crosses one wall $\mathfrak{d}$ contained in $n_0^\perp$ with the wall-crossing function $f(m_0)$. We write $\widetilde{T} = \widetilde{T}_{k,+}$ and $T = T_{k,+}$. Then we compute the action of the right-hand side of (6-2) on z^m:

$$z^m \mapsto \widetilde{T}^{-1}(z^m) \mapsto \widetilde{T}^{-1}(z^m) f(z^{m_0})^{\langle T^{-1}(m), n_0\rangle} \mapsto z^m f(\widetilde{T}(z^{m_0}))^{\langle m, (T^{-1})^*(n_0)\rangle}.$$

Note that the wall $\mathfrak{d}$ gets transformed under T_k to be contained in $(T^{-1})^*(n_0)$ with $f(\widetilde{T}(z^{m_0}))$. So the above action is the same as $\mathfrak{p}_{\gamma_2, T_k(\mathfrak{D}_s)}(z^m)$.

To show $\mathfrak{p}_{\gamma, T_k(\mathfrak{D}_s)} = \mathrm{id}$, it suffices to show that

$$\widetilde{T}_{k,+}^{-1} \circ \mathfrak{p}_{\mathfrak{d}'_k} = \mathfrak{p}_{\mathfrak{d}_k}, \tag{6-5}$$

so that $\mathfrak{p}_{\gamma, T_k(\mathfrak{D}_s)} = \mathfrak{p}_{\gamma, T_k(\mathfrak{D}_s)} = \mathrm{id}$.

Letting the left-hand side act on some monomial, we have

$$\begin{aligned}
\widetilde{T}_{k,+}^{-1} \circ \mathfrak{p}_{\mathfrak{d}'_k}(tz^m) &= \widetilde{T}_{k,+}^{-1}\Bigl(tz^m \prod_{j=1}^{r_k} (1 + t_{k,j}^{-1} z^{-w_k})^{-\langle e_k, m\rangle}\Bigr) \\
&= t \cdot t_k^{-\langle e_k, m\rangle} \cdot z^{m - \langle e_k, m\rangle r_k w_k} \prod_{j=1}^{r_k} (1 + t_{k,j}^{-1} z^{-w_k})^{-\langle e_k, m\rangle} \\
&= tz^m \prod_{j=1}^{r_k} (1 + t_{k,j}^{-1} z^{w_k})^{-\langle e_k, m\rangle} \\
&= \mathfrak{p}_{\mathfrak{d}_k}(tz^m).
\end{aligned} \tag{6-6}$$

This finishes the proof. □

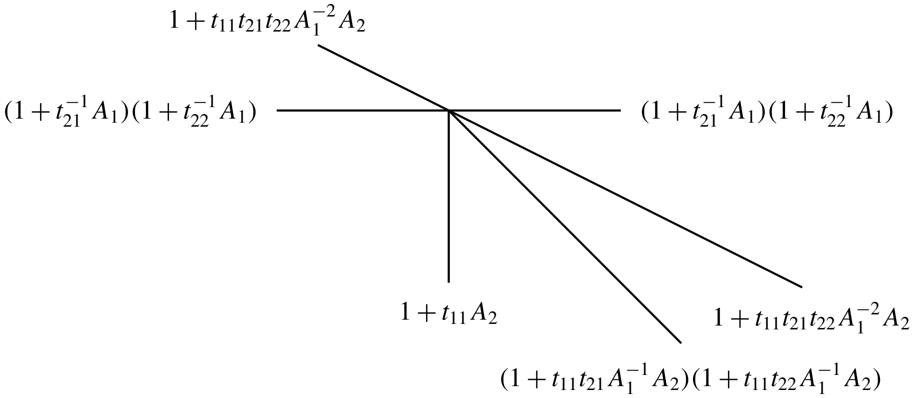
$1+t_{11}t_{21}t_{22}A_1^{-2}A_2$
$(1+t_{21}^{-1}A_1)(1+t_{22}^{-1}A_1)$
$(1+t_{21}^{-1}A_1)(1+t_{22}^{-1}A_1)$
$1+t_{11}A_2$
$1+t_{11}t_{21}t_{22}A_1^{-2}A_2$
$(1+t_{11}t_{21}A_1^{-1}A_2)(1+t_{11}t_{22}A_1^{-1}A_2)$

Figure 3. The generalized cluster scattering diagram for Example 6.30.

Example 6.30. In this example we compute $T_2(\mathfrak{D}_s)$ for the scattering diagram $\mathfrak{D}_s$ in Example 6.21. Recall that the exchange matrix for s is $B = \left(\begin{smallmatrix} 0 & -2 \\ 1 & 0 \end{smallmatrix}\right)$. So we have $T_{2,+}(e_2^*) = e_2^* - 2e_1^*$, which determines the ray $\mathbb{R}_{\geq 0}(e_2^* - 2e_1^*)$ of Figure 3.

6.7. ***Positivity.*** The scattering diagram $\mathfrak{D}_s$ has the following positivity.

Theorem 6.31 (cf. [Gross et al. 2018, Theorem 1.28]). *The scattering diagram $\mathfrak{D}_s$ is equivalent to a scattering diagram all of whose walls $(\mathfrak{d}, f_\mathfrak{d})$ satisfy $f_\mathfrak{d} = (1+tz^m)^c$ for some $m = \omega(-, \bar{n})$, $n \in N^+$, some $t \in \mathbb{P}$ such that $\pi(t) = n$, and c being a positive integer. In other words, if we write $n = \sum_{i\in I} \alpha_i e_i$, then*

(1) *$\mathfrak{d}$ is contained in $\bar{n}^\perp \subset M_\mathbb{R}$ where $\bar{n} = \sum_{i\in I} \alpha_i \frac{d_i}{r_i} e_i$;*

(2) *$m = \sum_{i\in I} \alpha_i w_i = \omega(-, \bar{n})$;*

(3) *if writing $t = \prod_{i,j} t_{i,j}^{\alpha_{i,j}}$, then $\sum_{j=1}^{r_i} \alpha_{ij} = \alpha_i$.*

Proof. This theorem essentially follows from [Gross et al. 2018, Appendix C.3], the proof of the positivity of $\mathfrak{D}_s^{\mathrm{GHKK}}$. We use a representative of $\mathfrak{D}_s$ constructed in the same algorithm used to produce $\mathfrak{D}_s^{\mathrm{GHKK}}$ in the proof of [Gross et al. 2018, Theorem 1.28]. We will construct order by order a sequence of finite scattering diagrams $\mathfrak{D}_1 \subset \mathfrak{D}_2 \subset \cdots$ (over $\widehat{k[P_s^\oplus]}$ or the graded Lie algebra $\mathfrak{g}_s$) such that their union

$$\mathfrak{D} = \bigcup_{k=1}^{\infty} \mathfrak{D}_k$$

is equivalent to $\mathfrak{D}_s$. We then prove inductively that every wall in $\mathfrak{D}_k$ has the positivity property.

Let $\mathfrak{D}_1 = \mathfrak{D}_{s,\mathrm{in}}$. Note that $\mathfrak{D}_1$ is equivalent to $\mathfrak{D}_s$ modulo $(P^+)^2$. Suppose that we have defined up to $\mathfrak{D}_k$ which is equivalent to $\mathfrak{D}$ modulo $(P^+)^{k+1}$, and assume that every wall in $\mathfrak{D}_k$ has wall-crossing function of the form $(1+tz^m)^c$ for some positive integer c. We construct $\mathfrak{D}_{k+1}$ as follows, and show that it is equivalent to $\mathfrak{D}$ modulo $(P^+)^{k+2}$ and furthermore that it still has the same positivity property for its wall-crossing functions.

There is a finite rational polyhedral cone complex that underlies the support of $\mathfrak{D}_k$ (which is true for any scattering diagram with finitely many walls). We call the codimension two cells *joints*. Let $\mathfrak{j}$ be a joint of $\mathfrak{D}_k$. Then by [Gross et al. 2018, Definition-Lemma C.2], it falls into two classes:

(1) *Parallel*, if every wall with the normal vector n containing $\mathfrak{j}$ has $\omega(-, n)$ tangent to $\mathfrak{j}$.

(2) *Perpendicular*, if every wall with the normal vector n containing $\mathfrak{j}$ has $\omega(-, n)$ not tangent to $\mathfrak{j}$.

Let $\gamma_{\mathfrak{j}}$ be a simple loop around $\mathfrak{j}$ small enough so that it only intersects walls containing $\mathfrak{j}$. By our assumption, the path-ordered product $\mathfrak{p}_{\gamma_{\mathfrak{j}},\mathfrak{D}_k}$ is identity modulo $(P^+)^{k+1}$, but modulo $(P^+)^{k+2}$, it can be written as

$$\mathfrak{p}_{\gamma_{\mathfrak{j}},\mathfrak{D}_k} = \exp\Bigl(\sum_{d(t,m)=k+1} c_{t,m} t z^m \partial_{n(t,m)}\Bigr),$$

where $c_{t,m} \in \Bbbk$. Here we define the degree $d(t, m) := k+1$ if $(t, m) \in (P^+)^{k+1} \setminus (P^+)^{k+2}$, and $n(t, m)$ is primitive in N^+ uniquely determined by (t, m).

If $\mathfrak{j}$ is perpendicular, we define a set of walls

$$\mathfrak{D}[\mathfrak{j}] := \{(\mathfrak{j} - \mathbb{R}_{\geq 0} m,\ (1 + t z^m)^{\pm c_{t,m}}) \mid d(t, m) = k+1\},$$

where $\mathfrak{j} - \mathbb{R}_{\geq 0} m$ is of codimension one since m is not tangent to $\mathfrak{j}$. Here the function $(1 + t z^m)^{\pm c_{t,m}}$ makes sense as a power series. The sign ± in the power is chosen so that when $\gamma_{\mathfrak{j}}$ crosses $\mathfrak{j} - \mathbb{R}_{\geq 0} m$, the wall-crossing automorphism is

$$\exp(-c_{t,m} t z^m \partial_{n(t,m)}).$$

In this way, if we add the walls in $\mathfrak{D}[\mathfrak{j}]$ to $\mathfrak{D}_k$, we have the path-ordered product $\mathfrak{p}_{\gamma_{\mathfrak{j}},\mathfrak{D}_k \cup \mathfrak{D}[\mathfrak{j}]} = \mathrm{id}$ modulo $(P^+)^{k+2}$. We then define

$$\mathfrak{D}_{k+1} = \mathfrak{D}_k \cup \bigcup_{\mathfrak{j}} \mathfrak{D}[\mathfrak{j}],$$

where the union is over all perpendicular joints of $\mathfrak{D}_k$.

There are two things we need to show in the induction:

(1) $\mathfrak{D}_{k+1}$ is equivalent to $\mathfrak{D}_s$ modulo $(P^+)^{k+2}$.

(2) All the walls in $\mathfrak{D}_{k+1}$ have wall-crossing functions of the form $(1 + t z^m)^c$ for some positive integer c.

Part (1) follows from the argument in [Gross et al. 2018, Lemma C.6 and Lemma C.7]. This part guarantees that the constructed union $\mathfrak{D}$ is equivalent to $\mathfrak{D}_s$.

Part (2) is about the positivity of wall-crossings. By the construction of $\mathfrak{D}_{k+1}$, we only need to examine the new walls emerging from perpendicular joints of $\mathfrak{D}_k$. Let $\mathfrak{j}$ be a perpendicular joint of $\mathfrak{D}_k$. The integral normal space $\mathfrak{j}^\perp \cap N$ is a rank two saturated sublattice O of N. Locally at $\mathfrak{j}$, $\mathfrak{D}_k \cup \mathfrak{D}[\mathfrak{j}]$ induces a scattering diagram living in $O_{\mathbb{R}}^\vee = M_{\mathbb{R}}/(\Lambda_j \otimes \mathbb{R})$. Precisely, consider the set of walls

$$\mathfrak{D}' = \{((\mathfrak{d} + \Lambda_{\mathfrak{j}} \otimes \mathbb{R})/(\Lambda_{\mathfrak{j}} \otimes \mathbb{R}),\ f_{\mathfrak{d}}) \mid \mathfrak{j} \subset \mathfrak{d},\ (\mathfrak{d}, f_{\mathfrak{d}}) \in \mathfrak{D}_k \cup \mathfrak{D}[\mathfrak{j}]\}.$$

The wall-crossing functions $f_{\mathfrak{d}}$ are all of the form

$$(1+tz^m)^c,$$

$c \in \Bbbk$ ($f_{\mathfrak{d}}$ makes sense as a power series). The wall $\mathfrak{d}$ has some primitive normal vector $o \in O \cap N^+$, and m is proportional to $\omega(-, o)$. We also know since $\mathfrak{j}$ is perpendicular, $\overline{m} \neq 0$ (the image of m under the quotient $M \to O^\vee$) in $O^\vee_{\mathbb{R}}$. And the one-dimensional wall $\bar{\mathfrak{d}} = (\mathfrak{d} + \Lambda_{\mathfrak{j}} \otimes \mathbb{R})/(\Lambda_{\mathfrak{j}} \otimes \mathbb{R})$ is contained in $\mathbb{R}(\overline{m})$, orthogonal to the normal vector o. Then $\mathfrak{D}'$ is a rank two scattering diagram in $O^\vee_{\mathbb{R}}$ over $\widehat{\Bbbk[P^+]}$, with the monoid map from P^+ to $O^\vee$ being $r : P \to M$ postcomposed by the quotient from M to $O^\vee$. It is consistent up to modulo $(P^+)^{k+2}$. Then by [Gross et al. 2018, Proposition C.13], the wall-crossing functions admit the positivity property, i.e., the power c is always a positive integer. This shows the positivity for $\mathfrak{D}_{k+1}$ assuming that of $\mathfrak{D}_k$. Therefore, the union $\mathfrak{D}$ is also positive by induction; hence so is $\mathfrak{D}_{\boldsymbol{s}}$. □

7. The cluster complex structure

In this section, we study the *cluster complex structure* of the scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$, which is a description of parts of the walls of $\mathfrak{D}_{\boldsymbol{s}}$. The construction of such a structure of $\mathfrak{D}_{\boldsymbol{s}}$ is analogous to [Gross et al. 2018, Construction 1.30].

7.1. ***The cluster complex.*** Take a representative for the scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$ with minimal support (which always exists). By Theorem 6.29, one can choose such a representative $\mathfrak{D}_{\boldsymbol{s}}$ so that there are no other walls contained in the initial incoming ones $\mathfrak{d}_i$.

Define

$$\mathcal{C}^+ = \mathcal{C}^+_{\boldsymbol{s}} := \{m \in M_{\mathbb{R}} \mid \langle e_i, m\rangle \geq 0 \ \ \forall i \in I\},$$
$$\mathcal{C}^- = \mathcal{C}^-_{\boldsymbol{s}} := \{m \in M_{\mathbb{R}} \mid \langle e_i, m\rangle \leq 0 \ \ \forall i \in I\}.$$

The closed cones $\mathcal{C}^\pm_{\boldsymbol{s}}$ are closures of connected components of $M_{\mathbb{R}} \setminus \operatorname{Supp}(\mathfrak{D}_{\boldsymbol{s}})$. They are thus called *chambers*. By the mutation invariance Theorem 6.27, we have that the cones

$$T_k^{-1}(\mathcal{C}^\pm_{\mu_k^+(\boldsymbol{s})}) \subset M_{\mathbb{R}} \setminus \operatorname{Supp}(\mathfrak{D}_{\boldsymbol{s}})$$

are also closures of connected components. Applying mutations on seeds provides an iterative way to construct chambers of $M_{\mathbb{R}} \setminus \operatorname{Supp}(\mathfrak{D}_{\boldsymbol{s}})$ as follows.

Note again that the coefficients part of $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{t})$ does not mutate as in Definition 4.3, which requires setting the tropical semifield $\mathbb{P}$ from the initial seed and once for all. Instead, we regard the coefficients part $\boldsymbol{t}$ as in the multiplicative group $\mathbb{P}$ and mutates in the way specified by Section 6.6. In this way, we can apply mutations iteratively on $\boldsymbol{s}$.

Let us consider the rooted tree $\mathfrak{T}_n$ from Definition 4.6. There is an association $v \mapsto \boldsymbol{s}_v$ such that $v_0 \mapsto \boldsymbol{s}$ and adjacent seeds with coefficients are related by the corresponding mutation (in the sense of Section 6.6) of the labeled edges. Once this association is done, we denote the rooted tree by $\mathfrak{T}_{\boldsymbol{s}}$.

Suppose the unique path from v_0 to a vertex v goes through the arrows labeled by $\{k_1, k_2, \dots, k_l\}$. Define the piecewise linear map

$$T_{v_0,v} = T_{k_l} \circ \cdots \circ T_{k_2} \circ T_{k_1} : M_{\mathbb{R}} \to M_{\mathbb{R}}.$$

Since $\mathcal{C}^{\pm}_{\boldsymbol{s}}$ are chambers of the scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$, then again due to the mutation invariance, we have that

$$\mathcal{C}^{\pm}_{v} := T^{-1}_{v_0,v}(\mathcal{C}^{\pm}_{\boldsymbol{s}_v})$$

are chambers of $\mathfrak{D}_{\boldsymbol{s}}$.

Each $\mathcal{C}^{\pm}_{v}$ is a simplicial (rational polyhedral) cone of maximal dimension, as each T_k is a linear isomorphism on its domains of linearity. The intersection $\mathcal{C}^{+}_{\boldsymbol{s}} \cap \mathcal{C}^{+}_{\mu^{+}_{k}(\boldsymbol{s})}$ is their common facet generated by $\{e^*_i \mid i \neq k\}$. Each facet of $\mathcal{C}_v$ is canonically labeled by an index $i \in I$. Inductively, for any two vertices v and v' connected by an arrow labeled by $k \in I$, then $\mathcal{C}^{+}_{v}$ and $\mathcal{C}^{+}_{v'}$ share a common facet labeled by k.

We borrow the following notation from [Gross et al. 2018]: we use the short-hand subscription notation $v \in \boldsymbol{s}$ for an object parametrized by a vertex $v \in \mathfrak{T}_{\boldsymbol{s}}$ with the root v_0 labeled by $\boldsymbol{s}$. This is done to emphasize the dependence on the initial seed $\boldsymbol{s}$.

Definition 7.1. We denote by $\mathcal{C}^{\pm}_{v\in\boldsymbol{s}}$ the chambers $\mathcal{C}^{\pm}_{v}$ of $\subset M_{\mathbb{R}} \setminus \operatorname{Supp}(\mathfrak{D}_{\boldsymbol{s}})$. We write $\Delta^{\pm}_{\boldsymbol{s}}$ for the set of chambers $\mathcal{C}^{\pm}_{v\in\boldsymbol{s}}$ for v running over all vertices of $\mathfrak{T}_{\boldsymbol{s}}$. We call elements in $\Delta^{+}_{\boldsymbol{s}}$ *cluster chambers*.

Remark 7.2. As we have pointed out, $\mathcal{C}^{+}_{v} \cap \mathcal{C}^{+}_{v'}$ is a common facet if v and v' are adjacent in $\mathfrak{T}_{\boldsymbol{s}}$. More generally, by adding all the faces of every $\mathcal{C}^{+}_{v}$ to the set $\Delta^{+}_{\boldsymbol{s}}$, we obtain a collection of cones which form a cone complex, still denoted by $\Delta^{+}_{\boldsymbol{s}}$. For this reason, we call $\Delta^{+}_{\boldsymbol{s}}$ *the cluster* (*cone*) *complex* and $\Delta^{-}_{\boldsymbol{s}}$ *the negative cluster* (*cone*) *complex*.

The simplicial cone $\mathcal{C}^{\pm}_{v\in\boldsymbol{s}}$ is determined by (the generators of) its one-dimensional faces. The cone $\mathcal{C}^{+}_{\boldsymbol{s}_v}$ is generated by the dual vectors $\{e^*_{i;v} \mid i \in I\}$. These are pulled back by $T^{-1}_{v_0,v}$ to be the generators of $\mathcal{C}^{+}_{v\in\boldsymbol{s}}$.

Definition 7.3. We define the *g-vectors* for $v \in \mathfrak{T}_{\boldsymbol{s}}$ as a tuple

$$\boldsymbol{g}_v = (g_{i;v} \mid i \in I), \quad \text{where } g_{i;v} := T^{-1}_{v_0,v}(e^*_{i;v}) \in M.$$

We will use the notation $\boldsymbol{g}_{v\in\boldsymbol{s}}$ to emphasize the initial seed $\boldsymbol{s}$.

Remark 7.4. Denote the dual vectors (in N) of $\boldsymbol{g}_v$ by $\boldsymbol{g}^*_v = (g^*_{i;v} \mid i \in I)$. They are normal vectors of the facets of $\mathcal{C}^{+}_{v}$. Since the walls of $\mathfrak{D}_{\boldsymbol{s}}$ only have normal vectors in $N^{+}_{\boldsymbol{s}}$ or $-N^{+}_{\boldsymbol{s}}$, the vector $g^*_{i;v}$ has a well-defined sign

$$\varepsilon_{i;v} = \operatorname{sgn}(g^*_{i;v}) = \begin{cases} + & \text{if } g^*_{i;v} \in N^{+}_{\boldsymbol{s}}, \\ - & \text{if } g^*_{i;v} \in N^{-}_{\boldsymbol{s}}. \end{cases}$$

We will show later the vectors $\boldsymbol{g}_v$ can be calculated iteratively by a variant of mutations as defined below.

Definition 7.5. Let $\boldsymbol{e} = (e_i \mid i \in I)$ be a seed (without coefficients) for Γ. Define the *signed mutation* $\mu_k^\varepsilon(\boldsymbol{e}) = (e'_i \mid i \in I)$ for $\varepsilon \in \pm$ as follows:

$$e'_i = \begin{cases} -e_k & \text{if } i = k, \\ e_i + [-\varepsilon\omega(e_i, d_k e_k)]_+ e_k & \text{if } i \neq k. \end{cases}$$

So the signed mutation μ_k^+ coincides with our previous Definition 6.23 (ignoring the coefficients part).

On the mutation of the dual of $\boldsymbol{e}$, we use the same notation $\mu_k^\varepsilon(\boldsymbol{e}^*) = (f'_i \mid i \in I)$ where $\boldsymbol{e}^* = (f_i \mid i \in I)$. Then

$$f'_i = \begin{cases} f_i & \text{if } i \neq k \\ -f_k + \sum_{i \in I} [-\varepsilon\omega(e_i, d_k e_k)]_+ f_k & \text{if } i = k. \end{cases}$$

There is another tuple of vectors in M that changes under signed mutations. For a seed $\boldsymbol{s}$, let $\boldsymbol{w} = (w_i \mid i \in I)$, where

$$w_i := \omega\Big(-, \frac{d_k}{r_k} e_k\Big) = \sum_{j \in I} b_{ji} f_i \in M.$$

Let $\boldsymbol{w}' = (w'_i)$ associated to $\mu_k^\varepsilon(\boldsymbol{e})$. Then we have

$$w'_i = \begin{cases} -w_k & \text{if } i = k, \\ w_i + [\varepsilon\omega(e_k, d_k e_i)]_+ w_k & \text{if } i \neq k. \end{cases}$$

We will later denote $\mu_k^\varepsilon(\boldsymbol{w}) = \boldsymbol{w}'$.

There are also signed mutations for coefficients. Recall we have fixed a multiplicative abelian group $\mathbb{P} = \prod_{i \in I} \mathbb{Z}^{r_i}$. The coefficients $\boldsymbol{t} = (t_{i,j} \mid i \in I, j \in [1, r_i])$ are a basis of $\mathbb{P}$.

Definition 7.6. For $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{t})$, a seed $\boldsymbol{e}$ together with coefficients $\boldsymbol{t} = (t_{i,j})$ in $\mathbb{P}$, we define its *signed mutation in direction k*, $\mu_k^\varepsilon(\boldsymbol{e}, (t_{i,j})) = (\boldsymbol{e}', (t'_{i,j}))$ for $\varepsilon \in \pm$ by setting $\boldsymbol{s}' = \mu_k^\varepsilon(\boldsymbol{s})$ and

$$t'_{i,j} = \begin{cases} t_{k,j}^{-1} & \text{if } i = k, \\ t_{i,j} \cdot \prod_{l=1}^{r_k} t_{k,l}^{[\varepsilon\omega(e_k, e_i)]_+} & \text{if } i \neq k. \end{cases}$$

Proposition 7.7 (cf. [Mou 2020, Proposition 4.4.9]). *For every $v \in \mathfrak{T}_{\boldsymbol{s}}$, the dual of g-vectors $\boldsymbol{g}_v^*$ is a seed of N. These seeds and their duals, i.e., the g-vectors, can obtained iteratively as follows*:

(1) $\boldsymbol{g}_{v_0} = \boldsymbol{e}^*$ *and* $\boldsymbol{g}_{v_0}^* = \boldsymbol{e}$.

(2) *For any $v \xrightarrow{k} v'$ in $\mathfrak{T}_{\boldsymbol{s}}$, we have*

$$\boldsymbol{g}_{v'}^* = \mu_k^{\varepsilon_{k;v}}(\boldsymbol{g}_v^*), \quad \boldsymbol{g}_{v'} = \mu_k^{\varepsilon_{k;v}}(\boldsymbol{g}_v).$$

Proof. We prove this proposition by induction on the distance from v to v_0. The base case is when $v = v_0$, in which we have

$$\boldsymbol{g}_{v'}^* = \mu_k^+(\boldsymbol{e}) = \mu_k^+(\boldsymbol{g}_{v_0}^*), \quad \boldsymbol{g}_{v'} = \mu_k^+(\boldsymbol{e}^*) = \mu_k^+(\boldsymbol{g}_{v_0}).$$

Now assuming that $v \neq v_0$ and suppose that the unique path from v_0 to v starts with $v_0 \xrightarrow{i} v_1$ for some $i \in I$. Write $\boldsymbol{s}_1 = \boldsymbol{s}_{v_1} = \mu_i^+(\boldsymbol{s})$. By induction, we assume that the proposition holds for g-vectors with respect to the seed $\boldsymbol{s}_1$:

$$\boldsymbol{g}_{v' \in \boldsymbol{s}_1} = \mu_k^{\varepsilon}(\boldsymbol{g}_{v \in \boldsymbol{s}_1}),$$

where $\varepsilon = \varepsilon_{k;v \in \boldsymbol{s}_1} = \operatorname{sgn}(g^*_{k;v \in \boldsymbol{s}_1})$ with respect to $\boldsymbol{s}_1$. Note that by definition

$$\boldsymbol{g}_{v' \in \boldsymbol{s}} = (T_i^{\boldsymbol{s}})^{-1}(\boldsymbol{g}_{v' \in \boldsymbol{s}_1}), \quad \boldsymbol{g}_{v \in \boldsymbol{s}} = (T_i^{\boldsymbol{s}})^{-1}(\boldsymbol{g}_{v \in \boldsymbol{s}_1}),$$

and we want to prove

$$\boldsymbol{g}_{v' \in \boldsymbol{s}} = \mu_k^{\delta}(\boldsymbol{g}_{v \in \boldsymbol{s}}),$$

where $\delta = \varepsilon_{k;v \in \boldsymbol{s}}$ with respect to $\boldsymbol{s}$.

Then it amounts to show that

$$(T_i^{\boldsymbol{s}})^{-1} \circ \mu_k^{\varepsilon}(\boldsymbol{g}_{v \in \boldsymbol{s}_1}) = \mu_k^{\delta} \circ (T_i^{\boldsymbol{s}})^{-1}(\boldsymbol{g}_{v \in \boldsymbol{s}_1}). \tag{7-1}$$

We split the discussion into the following two cases. The codimension one skeletons of the chambers $\mathcal{C}^+_{v \in \boldsymbol{s}_1}$ and $\mathcal{C}^+_{v' \in \boldsymbol{s}_1}$ are in the essential support of $\mathfrak{D}_{\boldsymbol{s}_1}$. As v and v' are adjacent, these two chambers share a common facet. Therefore they are either separated by the hyperplane $e_i^{\perp}$ or contained in the same half space (since the hyperplane is also in the essential support).

Case 1. The two groups of g-vectors $\boldsymbol{g}_{v \in \boldsymbol{s}_1}$ and $\boldsymbol{g}_{v' \in \boldsymbol{s}_1}$ are separated by $e_i^{\perp}$. In this case, the normal vector $g^*_{k;v \in \boldsymbol{s}_1}$ is in the direction of e_i. The signs δ and ε on the two sides of (7-1) are then different. We assume that $\varepsilon = \operatorname{sgn}(g^*_{k;v \in \boldsymbol{s}_1}) = +$; the other case is analogous. By our assumption, $\boldsymbol{g}^*_{v \in \boldsymbol{s}_1}$ qualifies as a seed of fixed data Γ, thus forming a basis of N, which implies $g^*_{k;v \in \boldsymbol{s}_1} = e_i$. Since $\{\lambda_j g^*_{j;v \in \boldsymbol{s}_1} \mid j \in I\}$ form a basis of the sublattice N°, we have $d_i = d_k$. We note that the map $T_i^{\boldsymbol{s}}$ is actually determined by the vectors e_i and $d_i e_i$. On the left-hand side of (7-1), $T_i^{\boldsymbol{s}}$ is the identity, while on the right-hand side, it is $T_{i,+}^{\boldsymbol{s}}$. So we need to show the equality

$$\mu_k^+(\boldsymbol{g}_{v \in \boldsymbol{s}_1}) = \mu_k^- \circ (T_{i,+}^{\boldsymbol{s}})^{-1}(\boldsymbol{g}_{v \in \boldsymbol{s}_1}).$$

To simplify the notation, we denote $\boldsymbol{g} = \boldsymbol{g}_{v \in \boldsymbol{s}_1}$ and $g_i = g_{i;v \in \boldsymbol{s}_1}$. On the left side of the equality, the tuple $\mu_k^+(\boldsymbol{g}) = (g_i')$ differs with $\boldsymbol{g}$ by only one vector

$$g_k' = -g_k + \sum_{i \in I} [-b_{ik}^v]_+ g_i.$$

On the right-hand side, we first have

$$(T_{i,+}^{\boldsymbol{s}})^{-1}(g_k) = -g_k + \sum_{i \in I} -b_{ik}^v g_i,$$

while other g-vectors remain unchanged under $(T_{i,+}^{\boldsymbol{s}})^{-1}$. It is easy to check that the dual of $(T_{i,+}^{\boldsymbol{s}})^{-1}$ is an automorphism of (N, ω), that is, it is a linear automorphism on N preserving the form ω. Thus we have, if writing $\mu_k^- \circ (T^{s}_{i,+})^{-1}(\boldsymbol{g}_{v\in \boldsymbol{s}_1}) = (g''_i)$,

$$g''_k = -g_k + \sum_{i\in I} -b^v_{ik} g_i + \sum_{i\in I} [b^v_{ik}]_+ g_i = g'_k, \quad \text{and} \quad g''_i = g_i \quad \text{for } i \neq k.$$

This finishes the proof of the desired equality.

Case 2. The g-vectors $\boldsymbol{g}_{v\in\boldsymbol{s}_1}$ and $\boldsymbol{g}_{v'\in\boldsymbol{s}_1}$ are all contained in the same half $\mathcal{H}^{\boldsymbol{s}}_{i,+}$ or $\mathcal{H}^{\boldsymbol{s}}_{i,-}$. Again we need to prove (7-1). We observe that the two signs δ and ε are equal. In fact, the sign ε of $g^*_{k;v\in\boldsymbol{s}_1}$ depends on its coordinates in $e_{j;v_1}$ for $j \neq i$ since $g^*_{k;v\in\boldsymbol{s}_1}$ is not purely proportional to e_i. The same is true for the sign δ which only depends on $g^*_{k;v\in\boldsymbol{s}}$'s coordinates in e_j for $j \neq i$. Since $g^*_{k;v\in\boldsymbol{s}}$ only differ in the direction of e_i, and also because $e_{j;v_1}$ and e_j also differ by multiples of e_i, we conclude that $\varepsilon = \delta$. The equality (7-1) then directly follows from a fact we already mentioned in Case 1 that the dual of $(T^{s}_{i,\varepsilon})^{-1}$ acts as an automorphism on (N, ω). □

A direct corollary of Proposition 7.7 is another description of c-vectors mentioned in Section 3.3. Recall that we have $\pi : \mathbb{P} \to \overline{\mathbb{P}}$, $p_{i,j} \mapsto \bar{p}_i$. We write the group operation in $\mathbb{P}$ and $\overline{\mathbb{P}}$ by addition instead of multiplication.

Corollary 7.8. *We identify the lattice $\overline{N}$ with $\overline{\mathbb{P}}$ by $\bar{e}_i = \frac{d_i}{r_i} e_i \mapsto \bar{p}_i$. Then we have for any $i \in I$ and $v \in \mathfrak{T}_{\boldsymbol{s}}$,*

$$\frac{d_i}{r_i} g^*_{i;v} = \bar{p}_{i;v}, \quad d_i g^*_{i;v} = r_i \bar{p}_{i;v} = p_{i;v}.$$

Proof. For the initial vertex v_0, this is given by the identification $\bar{e}_i \mapsto \bar{p}_i$. The iteration of $g^*_{i;v}$ is provided by signed mutations according to Proposition 7.7. We have if $v \xrightarrow{k} v'$ in $\mathfrak{T}_{\boldsymbol{s}}$,

$$g^*_{i;v'} = \begin{cases} -g^*_{k;v} & \text{if } i = k, \\ g^*_{i;v} + [-\varepsilon b^v_{ik}]_+ g^*_{i;v} & \text{if } i \neq k, \end{cases}$$

where $\varepsilon = \operatorname{sgn}(g^*_{k;v})$. What is implicit is that we have already known that $g^*_{i;k}$ is either nonnegative or nonpositive. On the other hand, the mutation of $p_{i;v}$ is given by

$$p_{i;v'} = \begin{cases} -p_{k;v} & \text{if } i = k, \\ p_{i;v} + b^v_{ki} \cdot p^+_{k;v} & \text{if } i \neq k \text{ and } b_{ik} \leq 0, \\ p_{i;v} + b^v_{ki} \cdot p^-_{k;v} & \text{if } i \neq k \text{ and } b_{ik} > 0. \end{cases}$$

Thus assuming $d_i g^*_{i;v} = p_{i;v}$ for all $i \in I$ would imply $d_i g^*_{i;v'} = p_{i;v'}$ for all $i \in I$ as they have the same mutation formula when $p_{k;v}$ has a well-defined sign. Therefore the result is proved by induction on the distance from v to v_0. □

Lemma 7.9. *The generalized coefficients $p_{i,j;v}$ have the following signed mutation formula. If $v \xrightarrow{k} v'$ in $\mathfrak{T}_{\boldsymbol{s}}$, then*

$$p_{i,j;v'} = \begin{cases} -p_{k,j;v} & \text{if } i = k, \\ p_{i,j;v} + [\varepsilon \beta^v_{ki}]_+ \cdot \sum_{j=1}^{r_k} p_{k,j;v} & \text{if } i \neq k \end{cases}$$

where $\varepsilon = \operatorname{sgn}(g^*_{k;v})$.

Proof. By Corollary 7.8, $p_{i;v}$ is sign coherent because $g^*_{i;v}$ is so. As we have already shown in Proposition 3.17 that the sign coherence of $p_{i;v}$ implies that of $p_{i,j;v}$, the result follows by induction. □

7.2. ***Wall-crossings.*** We next study the wall-crossing functions attached to walls of the cluster chambers. Each cluster chamber $\mathcal{C}^+_{v\in\boldsymbol{s}}$ has exactly n facets $\mathfrak{d}_{i;v\in\boldsymbol{s}}$ naturally indexed by I (a facet has the same index as its normal vector $g^*_{i,v\in\boldsymbol{s}}$). The wall $(\mathfrak{d}_{i;v\in\boldsymbol{s}}, f_{i;v\in\boldsymbol{s}})$ is pulled back by $T^{-1}_{v_0,v}$ from the scattering diagram $\mathfrak{D}_{\boldsymbol{s}_v}$ (with coefficients $\boldsymbol{t}_v$). The wall-crossing function $f_{i;v}$ has the following description. Here we identify the initial coefficients $t_{i,j}$ with $p_{i,j}$, and endow $\mathbb{P}$ the semifield structure $\mathrm{Trop}(\boldsymbol{p})$.

Theorem 7.10. *The scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$ has a representative in its equivalent class such that it is the union of the scattering diagram*

$$\mathfrak{D}(\Delta^+_{\boldsymbol{s}}) := \{(\mathfrak{d}_{i;v}, f_{i;v}) \mid i \in I,\, v \in \mathfrak{T}_{\boldsymbol{s}}\}, \quad \textit{where } f_{i;v} = \prod_{j=1}^{r_i} \Big(1 + p^{\varepsilon_{i;v}}_{i,j;v} \cdot z^{\varepsilon_{i;v} \sum_{j=1}^{n} \beta^v_{ji} g_{j;v}}\Big)$$

and another one whose support is disjoint from $\Delta^+_{\boldsymbol{s}}$.

Proof. We prove this theorem by induction on the distance from v to v_0. We first note that by Lemma 7.9 the coefficients $p_{i,j;v} \in \mathbb{P}$ can be computed iteratively by signed mutations. The vectors

$$w_{i;v} := \sum_{j=1}^{n} \beta^v_{ji} g_{j;v} = \omega\Big(-, \frac{d_i}{r_i} g^*_{i;v}\Big) \in M$$

can also be computed iteratively by signed mutations since the g-vectors do by Proposition 7.7.

Assume that the result is true for the distance between two vertices no greater v_0 and v. Suppose we have that $v \xrightarrow{k} v' \in \mathfrak{T}_{\boldsymbol{s}}$ and that the unique path from v_0 to v_1 starts from $v_0 \xrightarrow{i_0} v_1$.

Let's look at the chambers $\tau := \mathcal{C}^+_{v\in\boldsymbol{s}_1}$ and $\tau' := \mathcal{C}^+_{v'\in\boldsymbol{s}_1}$ in $\mathfrak{D}_{\boldsymbol{s}_1}$. They have g-vectors satisfying

$$\boldsymbol{g}_{v'\in\boldsymbol{s}_1} = \mu^{\varepsilon}_k(\boldsymbol{g}_{v\in\boldsymbol{s}_1}),$$

where $\varepsilon = \varepsilon_{k;v\in\boldsymbol{s}_1} := \mathrm{sgn}(g^*_{k;v\in\boldsymbol{s}_1})$. For the wall-crossing functions, by our assumption, for $i \in I$, we have

$$f_{i;v\in\boldsymbol{s}_1} = \prod_{j=1}^{r_i} (1 + p^{\varepsilon_{i;v\in\boldsymbol{s}_1}}_{i,j;v\in\boldsymbol{s}_1} z^{\varepsilon_{i;v\in\boldsymbol{s}_1} w_{i;v\in\boldsymbol{s}_1}}),$$

$$f_{i;v'\in\boldsymbol{s}_1} = \prod_{j=1}^{r_i} (1 + p^{\varepsilon_{i;v'\in\boldsymbol{s}_1}}_{i,j;v'\in\boldsymbol{s}_1} z^{\varepsilon_{i;v'\in\boldsymbol{s}_1} w_{i;v'\in\boldsymbol{s}_1}}).$$

These two functions are related by the signed mutation μ^{ε}_k. More precisely, we have

$$\mu^{\varepsilon}_k(\boldsymbol{g}^*_{v\in\boldsymbol{s}_1}, \boldsymbol{p}_{v\in\boldsymbol{s}_1}) = (\boldsymbol{g}^*_{v'\in\boldsymbol{s}_1}, \boldsymbol{p}_{v'\in\boldsymbol{s}_1}), \quad \mu^{\varepsilon}_k(\boldsymbol{w}_{v\in\boldsymbol{s}_1}) = \boldsymbol{w}_{v'\in\boldsymbol{s}_1}.$$

We want to pull back the chambers $\mathcal{C}^+_{v\in\boldsymbol{s}_1}$ and $\mathcal{C}^+_{v'\in\boldsymbol{s}_1}$, as well as the wall-crossing functions $f_{i;v\in\boldsymbol{s}_1}$ and $f_{i;v'\in\boldsymbol{s}_1}$ to $\mathfrak{D}_{\boldsymbol{s}}$ via the operation $(T^{\boldsymbol{s}}_{i_0})^{-1}$ to get the chambers $\sigma := \mathcal{C}^+_{v\in\boldsymbol{s}}$, $\sigma' := \mathcal{C}^+_{v'\in\boldsymbol{s}}$ and the wall-crossing functions $f_i := f_{i;v\in\boldsymbol{s}}$ and $f'_i := f_{i;v'\in\boldsymbol{s}}$ by the mutation invariance Theorem 6.27. We want to show that f_i and f'_i are also related by signed mutations. In the following, we calculate f_i and f'_i in detail by applying $\widetilde{T}^{-1}_{i_0}$ to $f_{i;v\in\boldsymbol{s}_1}$ and $f_{i;v'\in\boldsymbol{s}_1}$. This depends on the following two cases as in the proof of Proposition 7.7:

(1) The two chambers τ and τ' are separated by the hyperplane $e_{i_0}^{\perp}$.

(2) They are contained in the same half space $\mathcal{H}_{i_0,+}$ or $\mathcal{H}_{i_0,-}$.

Case 1. In this case, the normal vector $g^*_{k;v\in \boldsymbol{s}_1}$ is either e_{i_0} or $-e_{i_0}$. Assume it is e_{i_0}; the other case is similar. Then the chamber τ is in $\mathcal{H}_{i_0,+}$ while τ' is in $\mathcal{H}_{i_0,-}$. First of all, we have $f_k = f_k'$ obtained simply by reversing the monomials in $f_{k;v\in \boldsymbol{s}_1} = f_{k;v'\in \boldsymbol{s}_1}$. Since T_{i_0} (as well as $\widetilde{T}_{i_0}$) is identity on $\mathcal{H}_{i_0,-}$, we have for $i \neq k$, $f_i' = f_{i;v'\in \boldsymbol{s}_1}$. Note that for the signs, for $i \in I$,

$$\varepsilon_{i;v'\in \boldsymbol{s}_1} = \varepsilon_{i;v'\in \boldsymbol{s}}$$

unless $g^*_{i;v'\in \boldsymbol{s}_1}$ is proportional to e_{i_0}, which only happens for $g^*_{k;v'\in \boldsymbol{s}_1} = -g^*_{k;v\in \boldsymbol{s}_1}$, where we have

$$\varepsilon_{k;v'\in \boldsymbol{s}_1} = -, \quad \varepsilon_{k;v'\in \boldsymbol{s}} = +.$$

So we conclude for any $i \in I$,

$$f_i' = \prod_{j=1}^{r_i} (1 + (p_{i,j;v'\in \boldsymbol{s}_1} z^{w_{i;v'\in \boldsymbol{s}_1}})^{\varepsilon_{i;v'\in \boldsymbol{s}}}).$$

For $f_{i;v\in \boldsymbol{s}_1}$ and f_i, we first consider the signs $\varepsilon_{i;v\in \boldsymbol{s}_1}$ and $\varepsilon_{i;v\in \boldsymbol{s}}$. Since the dual of $T_{i_0}^{-1}$ on N only shifts in the direction of e_{i_0}, we have for $i \neq k$

$$\varepsilon_{i;v\in \boldsymbol{s}_1} = \varepsilon_{i;v\in \boldsymbol{s}},$$

as the vectors $g^*_{i;v\in \boldsymbol{s}_1}$ and $g^*_{i;v\in \boldsymbol{s}}$ must have the same sign in all the other directions except for e_{i_0}, and the only one proportional to e_{i_0} is $g^*_{k;v\in \boldsymbol{s}_1}$. Thus we have for $i \neq k$,

$$f_i = \prod_{j=1}^{r_i} (1 + \widetilde{T}_{i_0}^{-1} (p_{i,j;v\in \boldsymbol{s}_1} z^{w_{i;v\in \boldsymbol{s}_1}})^{\varepsilon_{i;v\in \boldsymbol{s}}})$$

We want to show that f_i and f_i' are related by the mutation μ_k^{δ}. Precisely, it amounts to show that

$$\mu_k^{\delta}(\widetilde{T}_{i_0}^{-1}(p_{i,j;v\in \boldsymbol{s}_1} z^{w_{i;v\in \boldsymbol{s}_1}} \mid i \in I,\ j \in [1, r_i])) = \mu_k^{\varepsilon}(p_{i,j;v\in \boldsymbol{s}_1} z^{w_{i;v\in \boldsymbol{s}_1}} \mid i \in I,\ j \in [1, r_i]), \tag{7-2}$$

where δ is the sign $\varepsilon_{k;v\in \boldsymbol{s}}$. Here we abuse the notation $\mu_k^{\pm}$ which acts on a tuple of functions, but it should be clear what it means. By our assumption, $\varepsilon = +$ and $\delta = -\varepsilon = -$. Then this follows from the general fact that for any seed $(\boldsymbol{e}, \boldsymbol{t})$ and $k \in I$, we have

$$\mu_k^{-}(\widetilde{T}_k^{-1}(t_{i,j} z^{w_i} \mid i \in I,\ j \in [1, r_i]) = \mu_k^{+}(t_{i,j} z^{w_i} \mid i \in I,\ j \in [1, r_i]).$$

Case 2. Suppose τ and τ' are both contained in the same half space. According to our above discussion, as in the notation of (7-2), it then amounts to check that

$$\widetilde{T}_{i_0}^{-1}(\mu_k^{\varepsilon}(p_{i,j;v\in \boldsymbol{s}_1} z^{w_{i;v\in \boldsymbol{s}_1}} \mid i \in I,\ j \in [1, r_i])) = \mu_k^{\delta}(\widetilde{T}_{i_0}^{-1}(p_{i,j;v\in \boldsymbol{s}_1} z^{w_{i;v\in \boldsymbol{s}_1}} \mid i \in I,\ j \in [1, r_i])),$$

where $\delta = \varepsilon_{k;v\in \boldsymbol{s}}$. As we have discussed in the **Case 2** of the proof of Proposition 7.7, the signs are equal: $\delta = \varepsilon$. Then the rest follows immediately from the fact that the dual of $T_{i_0,\varepsilon}$ acts as an automorphism on the data (N, ω). □

8. Reconstructing $\mathscr{A}^{\mathrm{prin}}$

In this section, we see how to reconstruct the generalized cluster algebra $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$ as well as the variety $\mathcal{A}^{\mathrm{prin}}(\boldsymbol{s})$ from $\mathcal{X}_{\boldsymbol{s},\lambda}$ through $\mathfrak{D}_{\boldsymbol{s}}$.

8.1. ***Reconstructing*** $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$ ***from*** $\mathfrak{D}_{\boldsymbol{s}}$**.** Given fixed data Γ and an $\mathcal{A}$-seed with principal coefficients $\boldsymbol{s} = (\boldsymbol{e}, \boldsymbol{p})$, denote by $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$ the corresponding generalized cluster algebra. Recall that we denote by $x_{i;v}$ the cluster variables associated to the seed $\boldsymbol{s}_v$.

Consider the generalized cluster scattering diagram $\mathfrak{D}_{\boldsymbol{s}}$, whose wall-crossings act on $\widehat{\Bbbk[P]}$ by automorphisms. For two vertices $v, v' \in \mathfrak{T}_{\boldsymbol{s}}$, let γ be a path from the chamber $\mathcal{C}^+_{v\in\boldsymbol{s}}$ to $\mathcal{C}^+_{v'\in\boldsymbol{s}}$ and consider the path-ordered product

$$\mathfrak{p}_{v,v'} = \mathfrak{p}^{\boldsymbol{s}}_{v,v'} := \mathfrak{p}_{\gamma,\mathfrak{D}_{\boldsymbol{s}}} : \widehat{\Bbbk[P]} \to \widehat{\Bbbk[P]}.$$

Since $\mathfrak{D}_{\boldsymbol{s}}$ is consistent and one can always choose some γ contained in the cluster complex, the path-ordered product $\mathfrak{p}_{v,v'}$ can also be viewed as an automorphism of $\mathrm{Frac}(M \oplus \mathbb{P})$.

Proposition 8.1. *Let $\mathcal{C}^+_{v\in\boldsymbol{s}}$ be a cluster chamber and $\boldsymbol{g}_v$ the set of g-vectors. Then for any $i \in I$,*

$$x_{i;v} = \mathfrak{p}_{v,v_0}(z^{g_{i;v}}) \in \mathrm{Frac}(M \oplus \mathbb{P}).$$

Proof. We prove this by induction on the distance from v to v_0 in $\mathfrak{T}_{\boldsymbol{s}}$. Suppose the statement is true for a vertex $v \in \mathfrak{T}_{\boldsymbol{s}}$ and we have $v \xrightarrow{i} v'$ in $\mathfrak{T}_{\boldsymbol{s}}$. Then the chambers $\mathcal{C}^+_v$ and $\mathcal{C}^+_{v'}$ are separated by the wall $\mathfrak{d}_{i;v}$ with the wall-crossing $f_{i;v}$ given in Theorem 7.10. Denote $\varepsilon = \mathrm{sgn}(g^*_{i;v}) \in \{+, -\}$. Then we have

$$\mathfrak{p}_{v',v}(z^{g_{i;v'}}) = z^{g_{i;v'}} \prod_{j=1}^{r_i} \big(1 + p^{\varepsilon}_{i,j;\boldsymbol{s}_v} \cdot z^{\sum_{j=1}^{n} \varepsilon\beta^v_{ji} g_{j;v}}\big)^{-\langle g_{i;v'}, g^*_{i;v}\rangle}.$$

By Proposition 7.7, we have

$$g_{i;v'} = -g_{i;v} + \sum_{j=1}^{n} [-\varepsilon r_i \beta^v_{ji}]_+ g_{j;v}.$$

This leads to

$$\mathfrak{p}_{v',v}(z^{g_{i;v'}}) = z^{-g_{i;v}} \prod_{j=1}^{r_i} \big(z^{\sum_{j\in I}[-\varepsilon\beta^v_{ji}]_+ g_{j;v}} + p^{\varepsilon}_{i,j;\boldsymbol{s}_v} \cdot z^{\sum_{j\in I}[\varepsilon\beta^v_{ji}]_+ g_{j;v}}\big).$$

Note that by sign coherence, $p_{i,j;\boldsymbol{s}_v}$ has the same sign as ε. So the above equation is exactly the exchange relation of cluster variables. Applying the path-ordered product $\mathfrak{p}_{v,v_0}$ on both sides of the above equality finishes the induction. □

By the generalized Laurent phenomenon Theorem 3.7, we know that $x_{i;v}$ actually lives in $\Bbbk[M \oplus \mathbb{P}]$.

Corollary 8.2. *The set of cluster variables of $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$ is in bijection with the set of g-vectors.*

Proof. We send a cluster variable $x_{i;v}$ to the g-vector $g_{i;v}$. To show that $x_{i;v}$ is uniquely determined by $g_{i;v}$, we observe that the formula $\mathfrak{p}_{v,v_0}(z^{g_{i;v}})$ is independent of the choice of v. Suppose there is another chamber $\mathcal{C}^+_{v'\in\boldsymbol{s}}$ such that $g_{i;v}$ is one of the generators. Choose a path γ from $\mathcal{C}^+_{v\in\boldsymbol{s}}$ to $\mathcal{C}^+_{v'\in\boldsymbol{s}}$ close enough to the ray $\mathbb{R}_+ g_{i;v}$ so that it only crosses walls containing $\mathbb{R}_+ g_{i;v}$. The two path-ordered products $\mathfrak{p}_{v,v_0}$ and $\mathfrak{p}_{v',v_0}$ differ by $\mathfrak{p}_\gamma$, which acts on $z^{g_{i;v}}$ by identity. Thus $\mathfrak{p}_{v,v_0}(z^{g_{i;v}}) = \mathfrak{p}_{v',v_0}(z^{g_{i;v}})$. □

8.2. ***Reconstructing $\mathcal{A}_{\boldsymbol{s}}^{\mathrm{prin}}$ from $\mathfrak{D}_{\boldsymbol{s}}$.*** Recall that there is a surjective map from $\mathfrak{T}_{\boldsymbol{s}}$ to $\Delta_{\boldsymbol{s}}^+$ (the set of cluster chambers) sending v to $\mathcal{C}^+_{v\in\boldsymbol{s}}$. For each vertex $v \in \mathfrak{T}_{\boldsymbol{s}}$, we associate a torus $T_{N,v}(R) = T_N(R)$. To a pair of vertices v and v', we associate the birational morphism

$$\mathfrak{q}_{v,v'} = \mathfrak{q}^{\boldsymbol{s}}_{v,v'} : T_{N,v}(R) \dashrightarrow T_{N,v'}(R), \qquad \mathfrak{q}^*_{v,v'} := \mathfrak{p}_{v',v}.$$

Then there is an R-scheme obtained by gluing $T_{N,v}(R)$, $v \in \mathfrak{T}_{\boldsymbol{s}}$ via these birational morphisms

$$\mathcal{A}^{\mathrm{prin}}_{\mathrm{scat},\boldsymbol{s}} := \bigcup_{v\in\mathfrak{T}_{\boldsymbol{s}}} T_{N,v}(R).$$

One can actually relate $\mathcal{A}^{\mathrm{prin}}_{\mathrm{scat},\boldsymbol{s}}$ to the previously defined cluster variety

$$\mathcal{A}^{\mathrm{prin}}_{\boldsymbol{s}} := \bigcup_{v\in\mathfrak{T}_{\boldsymbol{s}}} T_{N,\boldsymbol{s}_v}(R),$$

which is obtained by gluing together the same set of tori via $\mathcal{A}$-cluster mutations.

Recall the piecewise linear map $T_{v_0,v} : M_\mathbb{R} \to M_\mathbb{R}$ that sends the cluster chamber $\mathcal{C}^+_{v\in\boldsymbol{s}}$ to $\mathcal{C}^+_{\boldsymbol{s}_v}$. When restricted to a domain of linearity, $T_{v_0,v}$ becomes a linear automorphism on M. Denote the restriction of $T_{v_0,v}$ on $\mathcal{C}^+_{v\in\boldsymbol{s}}$ by $T_{v_0,v}|_{\mathcal{C}^+_{v\in\boldsymbol{s}}}$. In particular, $T_{v_0,v_0}|_{\mathcal{C}^+_{\boldsymbol{s}}}$ is the identity map. These linear isomorphisms induce isomorphisms (or R-schemes) between tori

$$\psi_{v_0,v} : T_{N,\boldsymbol{s}_v}(R) \to T_{N,v\in\boldsymbol{s}}(R), \qquad \psi^*_{v_0,v}(z^m) = z^{T_{v_0,v}|_{\mathcal{C}^+_{v\in\boldsymbol{s}}}(m)}.$$

Proposition 8.3. *The isomorphisms $\psi_{v_0,v}$ glue to be an isomorphism*

$$\psi_{v_0} : \mathcal{A}^{\mathrm{prin}}_{\boldsymbol{s}} \to \mathcal{A}^{\mathrm{prin}}_{\mathrm{scat},\boldsymbol{s}}.$$

Proof. The morphisms $\mu_{v,v'}$ (resp. $\mathfrak{q}_{v,v'}$) are generated $\mu_{v_0,v}$ (resp. $\mathfrak{q}_{v_0,v}$) for all v in $\mathfrak{T}_{\boldsymbol{s}}$. So the statement is equivalent to the commutativity of the following diagram (for any v).

$$\begin{tikzcd}
T_{N,\boldsymbol{s}} \arrow[r, "\psi_{v_0,v_0}=\mathrm{id}"] \arrow[d, dashed, "\mu_{v_0,v}"] & T_{N,v_0\in\boldsymbol{s}} \arrow[d, dashed, "\mathfrak{q}_{v_0,v}"] \\
T_{N,\boldsymbol{s}_v} \arrow[r, "\psi_{v_0,v}"] & T_{N,v\in\boldsymbol{s}}
\end{tikzcd}$$

To show $\mathfrak{q}_{v_0,v} = \psi_{v_0,v} \circ \mu_{v_0,v}$, we pull back the functions $z^{g_{i;v}}$ (for all $i \in I$) via these birational morphisms. On the left-hand side, we get the cluster variables

$$x_{i;v} = \mathfrak{q}^*_{v_0,v}(z^{g_{i;v}})$$

by Proposition 8.1. On the right-hand side, these $z^{g_{i;v}}$ get pulled back to $z^{e^*_{i;v}}$ by $\psi^*_{v_0,v}$ as $T_{v_0,v}|_{\mathcal{C}^+_{v\in \boldsymbol{s}}}$ sends the chamber $\mathcal{C}^+_{v\in \boldsymbol{s}}$ to the chamber $\mathcal{C}^+_{\boldsymbol{s}_v}$. Then via $\mu^*_{v_0,v}$, we still get cluster variables

$$x_{i;v} = \mu^*_{v_0,v}(z^{e^*_{i;v}}).$$

As $\{g_{i;v} \mid i \in I\}$ form a basis of M, we conclude that $\mathfrak{q}_{v_0,v} = \psi_{v_0,v} \circ \mu_{v_0,v}$, which finishes the proof. □

We next see in a certain sense the variety $\mathcal{A}^{\mathrm{prin}}_{\boldsymbol{s}}$ is independent of $\boldsymbol{s}$. This is a subtle issue as for the cluster algebra $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$, the initial seed $\Sigma(\boldsymbol{s})$ is distinguished from others since it has principal coefficients.

To resolve this, we again treat $\mathbb{P}$ as only a multiplicative abelian group. Consider $\boldsymbol{s}' = \mu^+_k(\boldsymbol{s})$ in the sense of Theorem 6.27. The tree $\mathfrak{T}_{\boldsymbol{s}'}$ is naturally embedded in $\mathfrak{T}_{\boldsymbol{s}}$, along with the association of seeds with coefficients. First of all, it is clear that the inclusion

$$\bigcup_{v\in\mathfrak{T}_{\boldsymbol{s}'}} T_{N,v\in\boldsymbol{s}} \subset \mathcal{A}^{\mathrm{prin}}_{\mathrm{scat},\boldsymbol{s}}$$

is an equality. The gluing maps are given by path-ordered products of $\mathfrak{D}_{\boldsymbol{s}}$.

Consider for $v \in \mathfrak{T}_{\boldsymbol{s}'}$, the isomorphism (of R-schemes)

$$\varphi_v : T_{N,v\in\boldsymbol{s}'} \to T_{N,v\in\boldsymbol{s}}$$

such that $\varphi^*_v : \Bbbk[M \oplus \mathbb{P}] \to \Bbbk[M \oplus \mathbb{P}]$ is given by the linear transformation

$$T_k\,|_{\mathcal{C}^+_{v\in\boldsymbol{s}}} : M \oplus \mathbb{P} \to M \oplus \mathbb{P}.$$

Proposition 8.4. *The maps φ_v for $v \in \mathfrak{T}_{\boldsymbol{s}'}$ glue together to have an isomorphism of $\Bbbk[\mathbb{P}]$-schemes*

$$\varphi : \mathcal{A}^{\mathrm{prin}}_{\mathrm{scat},\boldsymbol{s}'} \to \mathcal{A}^{\mathrm{prin}}_{\mathrm{scat},\boldsymbol{s}}.$$

Proof. Let v and v' be two vertices in $\mathfrak{T}_{\boldsymbol{s}'}$. Since each φ_v is an isomorphism, the statement is equivalent to the commutativity of the following diagram (for any v and v').

$$\begin{tikzcd}
T_{N,v\in\boldsymbol{s}'} \arrow[r, "\varphi_v"] \arrow[d, dashed, "\mathfrak{q}^{\boldsymbol{s}'}_{v,v'}"] & T_{N,v\in\boldsymbol{s}} \arrow[d, dashed, "\mathfrak{q}^{\boldsymbol{s}}_{v,v'}"] \\
T_{N,v'\in\boldsymbol{s}'} \arrow[r, "\varphi_{v'}"] & T_{N,v'\in\boldsymbol{s}}
\end{tikzcd}$$

In terms of algebras, this amounts to showing that

$$T_k\,|_{\mathcal{C}^+_{v\in\boldsymbol{s}}} \circ \mathfrak{p}^{\boldsymbol{s}}_{v,v'} = \mathfrak{p}^{\boldsymbol{s}'}_{v,v'} \circ T_k\,|_{\mathcal{C}^+_{v'\in\boldsymbol{s}}} : \Bbbk[M \oplus \mathbb{P}] \dashrightarrow \Bbbk[M \oplus \mathbb{P}].$$

If the two chambers $\mathcal{C}^+_{v\in\boldsymbol{s}}$ and $\mathcal{C}^+_{v'\in\boldsymbol{s}}$ are on the same side of the hyperplane $e_k^\perp$, the above equality is just (6-2). If they are separated by $e_k^\perp$, it is the same as (6-5) and has been checked in (6-6). □

Combined with Proposition 8.3, we see that the construction $\mathcal{A}_{\boldsymbol{s}}^{\mathrm{prin}}$ is independent of $\boldsymbol{s}$. In terms of the corresponding cluster algebra $\mathscr{A}^{\mathrm{prin}}(\boldsymbol{s})$, once it has principal coefficients on some seed $\boldsymbol{s}$, it can be made to do so at any seed mutation equivalent to $\boldsymbol{s}$.

8.3. ***Broken lines and theta functions.*** This section is a recast of [Gross et al. 2018, Section 3] in the generalized situation. Recall the setting of scattering diagrams in Definition 6.5.

Definition 8.5 (broken line, cf. [Gross et al. 2018, Definition 3.1]). Let $\mathfrak{D}$ be a scattering diagram over $\widehat{\Bbbk[P]}$ with a monoid map $r : P \to M$. Let $p_0 \in P \setminus \ker(r)$ and $Q \in M_{\mathbb{R}} \setminus \mathrm{Supp}(\mathfrak{D})$. A *broken line* for p_0 with endpoint Q is a piecewise linear continuous proper map

$$\gamma : (-\infty, 0] \to M_{\mathbb{R}} \setminus \mathrm{Sing}(\mathfrak{D})$$

with a finite number of domains of linearity $L_1, L_2, \dots, L_k$ (open intervals in $(-\infty, 0]$), where each $L = L_i \subset (-\infty, 0]$ is labeled by a monomial $c_L z^{p_L} \in \Bbbk[P]$ with $p_L \in P$. This data should satisfy:

(1) $\gamma(0) = Q$.

(2) If $L = L_1$ is the first domain of linearity of γ, i.e., $L = (-\infty, t)$ for some $t \leq 0$, then $c_L z^{p_L} = z^{p_0}$.

(3) For $t \in L$ any domain of linearity, $m_L := r(p_L) = -\gamma'(t)$.

(4) For two consecutive domains of linearity $L = (a, t)$ (a can be $-\infty$) and $L' = (t, b)$, the monomial $c_{L'} z^{p_{L'}}$ is a term in the formal power series

$$\mathfrak{p}_{\gamma(t),\mathfrak{D}}(c_L z^{p_L}) = c_L z^{p_L} \prod_{\substack{(\mathfrak{d}, f_{\mathfrak{d}}) \\ \gamma(t) \in \mathfrak{d}}} f_{\mathfrak{d}}^{-\langle n_0, m_L \rangle}.$$

Here $n_0 \in N$ is primitive, serving as a normal vector of every $\mathfrak{d}$ appearing in the product such that $\langle n_0, \gamma'(t) \rangle > 0$. So the power $-\langle n_0, m_L \rangle$ is always a positive integer.

Definition 8.6 (theta function, [Gross et al. 2018, Definition 3.3]). Let $\mathfrak{D}$ be a scattering diagram over $\widehat{\Bbbk[P]}$. Let $p_0 \in P \setminus \ker(r)$ and $Q \in M_{\mathbb{R}} \setminus \mathrm{Supp}(\mathfrak{D})$. For a broken line γ for p_0 with end point Q, define

$$\mathrm{Mono}(\gamma) := c_Q z^{p_Q},$$

where (by abuse of notation) Q stands for the last linear segment of γ. We define the *theta function* for p_0 with endpoint Q as the formal sum

$$\vartheta_{Q,p_0} := \sum_{\gamma} \mathrm{Mono}(\gamma),$$

where the sum is over the set of all broken lines for p_0 with endpoint Q.

For $p_0 = \ker(r)$, we define for any endpoint Q

$$\vartheta_{Q,p_0} = z^{p_0}.$$

We collect some important properties for theta functions from [Gross et al. 2018].

Theorem 8.7. (1) *The theta function ϑ_{Q,p_0} is in $\widehat{\Bbbk[P]}$.*

(2) *Suppose that $\mathfrak{D}$ is consistent. Then for Q, $Q' \in M_{\mathbb{R}} \setminus \operatorname{Supp}(\mathfrak{D})$ whose coordinates are linearly independent over $\mathbb{Q}$, and $p_0 \in P$,*

$$\vartheta_{Q',p_0} = \mathfrak{p}_{\gamma,\mathfrak{D}}(\vartheta_{Q,p_0}),$$

where γ is a path in $\mathfrak{D}$ from Q to Q' such that its path-ordered product is well-defined.

Proof. Part (1) essentially follows from the proof of [Gross et al. 2018, Proposition 3.4]. We are using a different monoid P here, but the same proof still works with $J := \mathfrak{m}_P = P \setminus M$.

Part (2), as pointed out in the proof of [Gross et al. 2018, Theorem 3.5], is again a special case of [Carl et al. 2024, Section 4]. Here the generic condition on the coordinates of Q and Q' is just to make sure that any broken line does not cross any joint of $\mathfrak{D}$. Modulo $\mathfrak{m}_P^k$, the independence of ϑ_{Q,m_0} on Q within one chamber follows from [Carl et al. 2024, Lemma 4.7]. The compatibility between Q and Q' in different chambers follows from [Carl et al. 2024, Lemma 4.9]. See also a more general discussion on the global property of theta functions in [Gross et al. 2022, Section 3.3]. □

In the case of generalized cluster scattering diagrams $\mathfrak{D}_s$ (see Definition 6.17), the monoid P is $M \oplus \bigoplus_{i \in I} \mathbb{N}^{r_i}$ (contained in $M \oplus \mathbb{P}$) with the natural projection r to the direct summand M. We have the following properties of theta functions.

Proposition 8.8 (mutation invariance of broken line, cf. [Gross et al. 2018, Proposition 3.6]). *The piecewise linear transformation $T_k : M_{\mathbb{R}} \to M_{\mathbb{R}}$ (with a lift on $M \oplus \mathbb{P}$) defines a one-to-one correspondence $\gamma \mapsto T_k(\gamma)$ between broken lines for p_0 with endpoint Q for $\mathfrak{D}_s$ and broken lines for $T_k(p_0)$ with endpoint $T_k(Q)$ for $\mathfrak{D}_{\mu_k(s)}$. This correspondence satisfies, depending on whether $Q \in \mathcal{H}_{k,+}$ or $\mathcal{H}_{k,-}$,*

$$\operatorname{Mono}(T_k(\gamma)) = T_{k,\pm}(\operatorname{Mono}(\gamma)),$$

where $T_{k,\pm}$ acts on a monomial as in Theorem 6.27. In particular, we have

$$\vartheta^{\mu_k(s)}_{T_k(Q),T_k(p_0)} = T_{k,\pm}(\vartheta^{s}_{Q,p_0}).$$

Proof. We use $T_k(\gamma)$ to denote the piecewise linear map $T_k \circ \gamma : (-\infty, 0] \to M_{\mathbb{R}}$. Suppose L is a domain of linearity of γ labeled with monomial $c_L z^{p_L}$. If $\gamma(L)$ is contained in one of the half spaces $\mathcal{H}_{k,\pm}$, L is also a domain of linearity for $T_k(\gamma)$. We apply the action of $T_{k,\pm}$ on the monomial $c_L z^{p_L}$ (where the sign is chosen depending on which half space L is in). If $\gamma(L)$ crosses $e_k^{\perp}$, split L into L^+ and L^-, and apply $T_{k,\pm}$ respectively to the monomial $c_L z^{p_L}$. One then needs to check the piecewise linear path $T_k \circ \gamma$ together with the new monomial data we just obtained is a broken line for $T_k(p_0)$ with endpoint $T_k(Q)$ in $\mathfrak{D}_{\mu_k(s)}$ as in [Gross et al. 2018, Proposition 3.6]. The inverse of the operation $\gamma \mapsto T_k \circ \gamma$ is also clear. The rest of the statement follows easily. □

Proposition 8.9 (cf. [Gross et al. 2018, Proposition 3.8]). *Consider the scattering diagram $\mathfrak{D}_s$.*

(1) *Let $Q \in \operatorname{Int}(\mathcal{C}_s^+)$ be an end point, and let $p \in P$ be such that $r(p) \in \mathcal{C}_s^+ \cap M$. Then $\vartheta_{Q,p} = z^p$.*

(2) *Let $\mathcal{C}_v^+ \in \Delta_s^+$ be a cluster chamber for some $v \in \mathfrak{T}_s$, and $Q \in \operatorname{int}(\mathcal{C}_v^+)$ and $m \in \mathcal{C}_v^+ \cap M$. Then $\vartheta_{Q,p} = z^p$ if $r(p) = m$.*

Proof. Part (1) is essentially [Gross et al. 2018, Proposition 3.8], although there the scattering diagram is actually different from $\mathfrak{D}_s$ in terms of wall-crossing functions. However, the bending behavior of a broken line on a wall is totally analogous, so the exact same argument still applies.

Part (2) is the generalized version of [Gross et al. 2018, Corollary 3.9]. By Proposition 8.8, the transformation $T_{v_0,v} : M_\mathbb{R} \to M_\mathbb{R}$ defines a one-to-one correspondence between the broken lines for p with $r(p) \in \mathcal{C}_v^+ \cap M$ and $Q \in \operatorname{int}(\mathcal{C}_v^+)$ in $\mathfrak{D}_s$, and the ones for $T_{v_0,v}(p)$ with $r(T_{v_0,v}(p)) \in \mathcal{C}_{s_v}^+ \cap M$ and $T_{v_0,v}(Q) \in \operatorname{int}(\mathcal{C}_{s_v}^+)$. However the only broken lines of the later is labeled by the final monomial $z^{p'}$ for $p' = T_{v_0,v}(p)$ by part (1). The result follows. □

8.4. *Cluster monomials as theta functions.*

Definition 8.10. Let s be a generalized $\mathcal{A}$-seed with principal coefficients. Then for $v \in \mathfrak{T}_s$, a cluster monomial in this seed is a monomial on the torus $T_{N,v}(R) \subset \mathcal{A}_s^{\text{prin}}$ of the form z^m where m is a nonnegative $\mathbb{N}$-linear combination of $\{e_{i;v}^* \mid i \in I\}$. By the Laurent phenomenon, such a monomial extends to a regular function on the whole cluster variety $\mathcal{A}_s^{\text{prin}}$.

Remark 8.11. One may regard a cluster monomial as a function on the initial torus $T_{N,v_0}(R)$. While being a monomial on the cluster variables $x_{i;v}$, it is also a Laurent polynomial in the initial cluster variables x_i by the Laurent phenomenon.

The following description of cluster monomials is a generalized version of [Gross et al. 2018, Theorem 4.9]. It proves the positivity (see Theorem 3.8) of generalized cluster monomials.

Theorem 8.12. *Let $\mathfrak{D}_s$ be the generalized cluster scattering diagram of a seed s. Let $Q \in \operatorname{int}(\mathcal{C}_s^+)$ a general end point and $m \in \mathcal{C}_v^+ \cap M$ for some $v \in \mathfrak{T}_s$. Then the theta function $\vartheta_{Q,m}$ is an element in $z^m \cdot \mathbb{N}[P]$ which expresses the cluster monomial associated to m of the algebra $\mathscr{A}^{\text{prin}}(s)$ in the initial seed s.*

Proof. We first note that m is regarded as a point in P through the inclusion of M in P. Let Q' be a base point in $\operatorname{int}(\mathcal{C}_v^+)$ and γ be a path going from Q' to Q. By part (2) of Theorem 8.7, we have

$$\vartheta_{m,Q} = \mathfrak{p}_\gamma(\vartheta_{m,Q'}).$$

As a theta function, $\vartheta_{m,Q}$ is a (formal) sum of monomials belonging to $z^m \widehat{\Bbbk[P]}$. By the positivity Theorem 6.31 of $\mathfrak{D}_s$, $\vartheta_{m,Q}$ has positive integer coefficients, thus an element in $z^m \widehat{\mathbb{N}[P]}$. By part (2) of Proposition 8.9, $\vartheta_{m,Q'} = z^m$. We know that the cone $\mathcal{C}_v^+$ has integral generators $\{g_{i;v} \mid i \in I\}$ in M. Thus m is a nonnegative linear combination of these g-vectors.

On the other hand, by Proposition 8.1, we have the following expression of a cluster variable

$$x_{i;v} = \mathfrak{p}_\gamma(z^{g_{i;v}}).$$

It follows immediately that $\vartheta_{m,Q}$ is a monomial of these $x_{i;v}$, thus expressing a cluster monomial. Finally by the generalized Laurent phenomenon Theorem 3.7, we have $\vartheta_{m,Q} \in z^m \cdot \mathbb{N}[P]$. □

Since $\vartheta_{m,Q}$ does not depend on Q as long as it is chosen generally in the positive chamber, we simply write it as ϑ_m. Consider the set of functions

$$\{\vartheta_m \mid m \in \Delta_{\boldsymbol{s}}^+(\mathbb{Z})\},$$

where $\Delta_{\boldsymbol{s}}^+(\mathbb{Z}) = \bigcup_{v\in\mathfrak{T}_s} \mathcal{C}_v^+ \cap M$. These are all cluster monomials. In general, they do not form an R-basis of the cluster algebra $\mathscr{A}^{\text{prin}}(\boldsymbol{s})$ or the upper cluster algebra $\overline{\mathscr{A}}^{\text{prin}}(\boldsymbol{s})$. But one can follow [Gross et al. 2018, Section 7.1] to define the set $\Theta \subset M$ such that for any $m \in \Theta$, ϑ_m is only a sum of monomials from finitely many broken lines. Consider the free R-module

$$\operatorname{mid}(\mathcal{A}_{\boldsymbol{s}}^{\text{prin}}) := \bigoplus_{m\in\Theta} R \cdot \vartheta_m.$$

It is shown in [Gross et al. 2018, Theorem 7.5] that in the ordinary case there are natural inclusions of R-modules

$$\mathscr{A}^{\text{prin}}(\boldsymbol{s}) \subset \operatorname{mid}(\mathcal{A}_{\boldsymbol{s}}^{\text{prin}}) \subset \overline{\mathscr{A}}^{\text{prin}}(\boldsymbol{s})$$

such that for the first inclusion, cluster monomials are sent to the corresponding theta functions, and for the second inclusion, any theta function is sent to the corresponding universal Laurent polynomials on $\mathcal{A}_{\boldsymbol{s}}^{\text{prin}}$ (see [Gross et al. 2018, Proposition 7.1]). We expect that this is also true in the generalized case.

8.5. ***More on positivity.*** Chekhov and Shapiro [2014] proposed a positivity conjecture which is stronger than Theorem 3.8. We formulate a version here.

A generalized cluster algebra in the sense of [Chekhov and Shapiro 2014] (see Section 3.2) is called *reciprocal* if any of its exchange polynomials $\theta_i(u, v)$ is monic and palindromic, i.e., $\theta_i(u, v) = \theta_i(v, u)$ and has leading coefficient 1. In this way, the exchange polynomials do not change under mutations. Note that $\theta_i(u, v)$ can have coefficients in $\mathbb{ZP}$ (rather than just in $\mathbb{P}$) in general.

Conjecture 8.13 (cf. [Chekhov and Shapiro 2014, Conjecture 5.1]). *Any cluster variable of a reciprocal generalized cluster algebra whose exchange polynomials have coefficients in* $\mathbb{P}$ (*or more generally in* $\mathbb{NP}$) *is expressed as a positive Laurent polynomial in the initial cluster*, *i.e.*, *an element in* $\mathbb{NP}[x_1^{\pm}, \dots, x_n^{\pm}]$ *where the* x_i*'s are the initial cluster variables.*

Chekhov and Shapiro [2014, Section 5] pointed out that this conjecture is true for any generalized cluster algebra associated to a surface with arbitrary orbifold points (see also [Banaian and Kelley 2020] for a proof using snake graphs). The rank two case of this conjecture has been resolved in [Rupel 2013].

We consider here a related situation where the reciprocal assumption is not required. Let $\mathbb{P}$ be an abelian group of finite rank. Consider an algebraic closure $\Bbbk = \overline{\mathbb{QP}}$ of the field of rational functions $\mathbb{QP}$. Let $\mathscr{A}^{\text{prin}}(\Sigma)$ be a generalized cluster algebra with principal coefficients as of Definition 3.13. The coefficients group is the tropical semifield $\operatorname{Trop}(\boldsymbol{p})$. Recall that the initial exchange polynomials have the form

$$\theta_i(u, v) = \prod_{j=1}^{r_i} (p_{i,j}u + v).$$

Let $\lambda : \operatorname{Trop}(\boldsymbol{p}) \to \Bbbk^*$ be an evaluation (as in Section 3.5) such that each $\lambda(\theta_i(u, v))$ satisfies:

(A) All its coefficients are in $\mathbb{ZP}$ (in $\mathbb{NP}$ if assuming positivity).

(B) $\lambda\big(\prod_{j=1}^{r_i} p_{i,j}\big)$ is an element in $\mathbb{P}$.

By the mutation formula of coefficients, the exchange polynomials after any steps of mutations still satisfy these two conditions. Therefore the cluster algebra with special coefficients $\mathscr{A}^{\mathrm{prin}}(\Sigma, \lambda)$ can be viewed as a generalized cluster algebra of [Chekhov and Shapiro 2014] (with the coefficients group $\mathbb{P}$). Note that any reciprocal generalized cluster algebra can be obtained this way.

The scattering diagram $\lambda(\mathfrak{D}_s)$ (see Section 6.4) is responsible for $\mathscr{A}^{\mathrm{prin}}(\Sigma, \lambda)$. It is over $\Bbbk\widehat{\big[M \oplus \prod_{i \in I} \mathbb{N}\big]}$ with formal parameters t_i. Note that by the generalized Laurent phenomenon, the cluster variables of $\mathscr{A}^{\mathrm{prin}}(\Sigma, \lambda)$ are all in $\mathbb{ZP}[x_1^{\pm}, \dots, x_n^{\pm}]$.

Theorem 8.14. *Let $\mathscr{A}^{\mathrm{prin}}(\Sigma, \lambda)$ be a generalized cluster algebra as above assuming* (***A***), (***B***), *and that the initial exchange polynomials have coefficients in $\mathbb{NP}$. Let $\mathbf{s}$ be an $\mathcal{A}$-seed such that $\Sigma(\mathbf{s}) = \Sigma$. If there exists a representative of $\lambda(\mathfrak{D}_s)$ such that every wall-crossing function is in $\mathbb{NP}\widehat{\big[M \oplus \prod_{i \in I} \mathbb{N}\big]}$, then any cluster variable is expressed as a positive Laurent polynomial in the initial cluster*, *i.e.*, *an element in $\mathbb{NP}[x_1^{\pm}, \dots, x_n^{\pm}]$.*

Proof. As in Theorem 8.12, the positivity of cluster variables follows from the positivity of the scattering diagram $\lambda(\mathfrak{D}_s)$ since every broken line ends with a monomial with coefficients in $\mathbb{NP} \subset \Bbbk$. Expressing a cluster variable as a theta function for $\lambda(\mathfrak{D}_s)$ $\big($and evaluated at $t_i = 1$ where the t_i's are the standard generators of $\prod_{i \in I} \mathbb{N}\big)$, the result follows. □

If $\mathscr{A}^{\mathrm{prin}}(\Sigma, \lambda)$ is of finite type (i.e. there are only finitely many distinguished cluster variables), then the cluster complex Δ_s^+ is finite and complete in $M_{\mathbb{R}}$ by Corollary 8.2. By Theorem 7.10, we have that $\mathfrak{D}_s = \mathfrak{D}(\Delta_s^+)$ and the wall-crossing function on any facet of any cluster chamber has coefficients in $\mathbb{NP}$ under the evaluation λ if assuming so for the initial ones. Then the positivity follows in this case from Theorem 8.14. It is not hard to check that in Example 6.22 the expansion of the wall-crossing function $f_{\mathbb{R}_{\geq 0}(1,-1)}$ has every coefficient in $\mathbb{N}[s_1 s_2, s_1 + s_2, t_1 t_2, t_1 + t_2]$. By the description in Example 6.22 of all other walls, all wall-crossings functions in this scattering diagram are positive in this sense. This then implies all cluster variables are positive, i.e., have coefficients in $\mathbb{N}[s_1 s_2, s_1 + s_2, t_1 t_2, t_1 + t_2]$.

Acknowledgements

I would like to thank Daniel Labardini-Fragoso for showing me an example of generalized cluster scattering diagram in the beginning; Fang Li and Siyang Liu for helpful discussions and the hospitality during my visit at Zhejiang University. During the preparation of the manuscript, I was supported by the Royal Society through the Newton International Fellowship NIF\R1\201849.

References

[Argüz and Gross 2022] H. Argüz and M. Gross, "The higher-dimensional tropical vertex", *Geom. Topol.* **26**:5 (2022), 2135–2235. MR Zbl

[Banaian and Kelley 2020] E. Banaian and E. Kelley, "Snake graphs from triangulated orbifolds", *SIGMA Symmetry Integrability Geom. Methods Appl.* **16** (2020), art. id. 138. MR Zbl

[Berenstein et al. 1996] A. Berenstein, S. Fomin, and A. Zelevinsky, "Parametrizations of canonical bases and totally positive matrices", *Adv. Math.* **122**:1 (1996), 49–149. MR Zbl

[Berenstein et al. 2005] A. Berenstein, S. Fomin, and A. Zelevinsky, "Cluster algebras, III: Upper bounds and double Bruhat cells", *Duke Math. J.* **126**:1 (2005), 1–52. MR Zbl

[Bossinger et al. 2020] L. Bossinger, B. Frías-Medina, T. Magee, and A. Nájera Chávez, "Toric degenerations of cluster varieties and cluster duality", *Compos. Math.* **156**:10 (2020), 2149–2206. MR Zbl

[Bridgeland 2017] T. Bridgeland, "Scattering diagrams, Hall algebras and stability conditions", *Algebr. Geom.* **4**:5 (2017), 523–561. MR Zbl

[Carl et al. 2024] M. Carl, M. Pumperla, and B. Siebert, "A tropical view on Landau–Ginzburg models", *Acta Math. Sin.* (*Engl. Ser.*) **40**:1 (2024), 329–382. MR Zbl

[Chekhov and Shapiro 2014] L. Chekhov and M. Shapiro, "Teichmüller spaces of Riemann surfaces with orbifold points of arbitrary order and cluster variables", *Int. Math. Res. Not.* **2014**:10 (2014), 2746–2772. MR Zbl

[Cheung et al. 2021] M.-W. Cheung, E. Kelley, and G. Musiker, "Cluster scattering diagrams and theta basis for reciprocal generalized cluster algebras", *Sém. Lothar. Combin.* **85B** (2021), art. id. 86. MR Zbl

[Cheung et al. 2023] M.-W. Cheung, E. Kelley, and G. Musiker, "Cluster scattering diagrams and theta functions for reciprocal generalized cluster algebras", *Ann. Comb.* **27**:3 (2023), 615–691. MR Zbl

[Derksen et al. 2010] H. Derksen, J. Weyman, and A. Zelevinsky, "Quivers with potentials and their representations, II: Applications to cluster algebras", *J. Amer. Math. Soc.* **23**:3 (2010), 749–790. MR Zbl

[Fock and Goncharov 2009] V. V. Fock and A. B. Goncharov, "Cluster ensembles, quantization and the dilogarithm", *Ann. Sci. École Norm. Sup.* (4) **42**:6 (2009), 865–930. 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 2003] S. Fomin and A. Zelevinsky, "Cluster algebras, II: Finite type classification", *Invent. Math.* **154**:1 (2003), 63–121. MR Zbl

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

[Gross and Pandharipande 2010] M. Gross and R. Pandharipande, "Quivers, curves, and the tropical vertex", *Port. Math.* **67**:2 (2010), 211–259. MR Zbl

[Gross and Siebert 2011] M. Gross and B. Siebert, "From real affine geometry to complex geometry", *Ann. of Math.* (2) **174**:3 (2011), 1301–1428. MR Zbl

[Gross and Siebert 2022] M. Gross and B. Siebert, "The canonical wall structure and intrinsic mirror symmetry", *Invent. Math.* **229**:3 (2022), 1101–1202. MR Zbl

[Gross et al. 2010] M. Gross, R. Pandharipande, and B. Siebert, "The tropical vertex", *Duke Math. J.* **153**:2 (2010), 297–362. MR Zbl

[Gross et al. 2015] M. Gross, P. Hacking, and S. Keel, "Birational geometry of cluster algebras", *Algebr. Geom.* **2**:2 (2015), 137–175. MR Zbl

[Gross et al. 2018] M. Gross, P. Hacking, S. Keel, and M. Kontsevich, "Canonical bases for cluster algebras", *J. Amer. Math. Soc.* **31**:2 (2018), 497–608. MR Zbl

[Gross et al. 2022] M. Gross, P. Hacking, and B. Siebert, *Theta functions on varieties with effective anti-canonical class*, Mem. Amer. Math. Soc. **1367**, Amer. Math. Soc., Providence, RI, 2022. MR Zbl

[Kelley 2021] E. Kelley, *Structural properties of reciprocal generalized cluster algebras*, Ph.D. thesis, University of Minnesota, 2021, available at https://www.proquest.com/docview/2590085968.

[Kontsevich and Soibelman 2006] M. Kontsevich and Y. Soibelman, “Affine structures and non-Archimedean analytic spaces”, pp. 321–385 in *The unity of mathematics*, edited by P. Etingof et al., Progr. Math. **244**, Birkhäuser, Boston, MA, 2006. MR Zbl

[Kontsevich and Soibelman 2014] M. Kontsevich and Y. Soibelman, “Wall-crossing structures in Donaldson–Thomas invariants, integrable systems and mirror symmetry”, pp. 197–308 in *Homological mirror symmetry and tropical geometry*, edited by R. Castano-Bernard et al., Lect. Notes Unione Mat. Ital. **15**, Springer, 2014. MR Zbl

[Labardini-Fragoso and Mou 2024] D. Labardini-Fragoso and L. Mou, “Gentle algebras arising from surfaces with orbifold points of order 3, I: Scattering diagrams”, *Algebr. Represent. Theory* **27**:1 (2024), 679–722. MR Zbl

[Lee and Schiffler 2015] K. Lee and R. Schiffler, “Positivity for cluster algebras”, *Ann. of Math.* (2) **182**:1 (2015), 73–125. MR Zbl

[Mou 2020] L. Mou, *Wall-crossing structures in cluster algebras*, Ph.D. thesis, University of California, Davis, 2020, available at https://www.proquest.com/docview/2458758157.

[Nakanishi 2015] T. Nakanishi, “Structure of seeds in generalized cluster algebras”, *Pacific J. Math.* **277**:1 (2015), 201–217. MR Zbl

[Reineke and Weist 2013] M. Reineke and T. Weist, “Refined GW/Kronecker correspondence”, *Math. Ann.* **355**:1 (2013), 17–56. MR Zbl

[Rupel 2013] D. Rupel, “Greedy bases in rank 2 generalized cluster algebras”, preprint, 2013. arXiv 1309.2567

Communicated by Sergey Fomin
Received 2022-11-28 Revised 2023-10-31 Accepted 2024-01-22

langmou@math.uni-koeln.de *Mathematisches Institut, Universität zu Köln, Köln, Germany*

ALGEBRA AND NUMBER THEORY 18:12 (2024)
https://doi.org/10.2140/ant.2024.18.2247

Matrix Kloosterman sums

Márton Erdélyi and Árpád Tóth

We study a family of exponential sums that arises in the study of expanding horospheres on GL_n. We prove an explicit version of general purity and find optimal bounds for these sums.

1. Introduction

1A. ***The subject of the paper.*** We derive nontrivial bounds for certain exponential sums that are natural generalizations of the classical Kloosterman sum to the noncommutative algebra $M_n(\mathbb{F}_q)$ of $n \times n$ matrices over a finite field $\mathbb{F}_q$ with $q = p^f$ elements.

To define these sums let $\mathbb{F}_p$ be the prime field of $\mathbb{F}_q$, and $\overline{\mathbb{F}}$ be an algebraic closure of $\mathbb{F}_q$ so that for $m \geq 1$ $\mathbb{F}_{q^m} \subset \overline{\mathbb{F}}$ is the unique degree-m extension of $\mathbb{F}_q$. Let

$$\varphi_0 : \mathbb{F}_p \to \mathbb{C}^*$$

be the additive character which maps $1 \in \mathbb{F}_p$ to $\zeta = \exp(1/p) = e^{2\pi i/p}$, and fix the additive characters

$$\varphi = \varphi_0 \circ \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p} \quad \text{and} \quad \varphi_m = \varphi_0 \circ \mathrm{Tr}_{\mathbb{F}_{q^m}/\mathbb{F}_p}$$

of $\mathbb{F}_q$ and $\mathbb{F}_{q^m}$.

Let $M_n(\mathbb{F}_{q^m})$ be the algebra of $n \times n$ matrices over $\mathbb{F}_{q^m}$, and $\mathrm{GL}_n(\mathbb{F}_{q^m}) = M_n^*(\mathbb{F}_{q^m}) \subset M_n(\mathbb{F}_{q^m})$ be the general linear group. Let ψ (resp. ψ_m) be the additive character of $M_n(\mathbb{F}_q)$ (resp. $M_n(\mathbb{F}_{q^m})$) defined by

$$\psi = \varphi \circ \mathrm{tr}$$

Erdélyi is supported by NKFIH (National Research, Development and Innovation Office) grants FK127906, FK135885 and the MTA–RI Lendület "Momentum" Analytic Number Theory and Representation Theory Research Group. Tóth is supported by the MTA Rényi Intézet Lendület Automorphic Research Group and by NKFIH grants K119528 and FK135885.
MSC2020: primary 11L05; secondary 11L07, 20G40.
Keywords: Kloosterman sums, estimates for exponential sums.

(resp. $\psi_m = \varphi_m \circ \mathrm{tr}$), where $\mathrm{tr} = \mathrm{tr}_n : M_n(\mathbb{F}_{q^m}) \to \mathbb{F}_{q^m}$ is the matrix trace. For $a \in M_n(\mathbb{F}_{q^m})$ define the sum

$$K_n(a, \mathbb{F}_{q^m}) = \sum_{x \in \mathrm{GL}_n(\mathbb{F}_{q^m})} \psi_m(ax + x^{-1}), \tag{1}$$

generalizing the classical Kloosterman sum [1927]

$$K_1(\alpha, \mathbb{F}_{q^m}) = K(\alpha, \mathbb{F}^*_{q^m}) = \sum_{x \in \mathbb{F}^*_{q^m}} \varphi_m(\alpha x + x^{-1}). \tag{2}$$

When the field in question is clear, or when the arguments used do not depend on it, we will write M_n and $K_n(a)$ for $M_n(\mathbb{F}_q)$ and $K_n(a, \mathbb{F}_q)$.

The interest in these sums arose in connection with a conjecture of Marklof concerning the equidistribution of certain special points associated to expanding horospheres. This conjecture, originally motivated by Marklof's work on Frobenius numbers [2010], was proved by Einsiedler, Moses, Shah and Shapira [2016] using methods of ergodic theory. In the case of SL_2 the connection to classical Kloosterman sums was known already to Marklof (see Section 2.1 of [Einsiedler et al. 2016]) and together with Lee [2018] they proved an effective version of the conjecture for SL_3. This proof strongly hinted that nontrivial bounds of the sums in (1) could yield a proof of Marklof's conjecture with an effective power saving for higher rank situations as well. One of the main purposes of this paper is to provide such bounds; they are formulated in Theorems 1.8 and 1.10. These bounds, together with further extensions in [Erdélyi et al. 2024b], were then recently used by El-Baz, Lee and Strömbergsson for realizing the above goal in [El-Baz et al. 2022].

There is however also intrinsic interest in these sums as natural generalizations[1] of Kloosterman's sum. The relevance and widespread use in analytic number theory of $K_1(\alpha)$ (see, for example, [Heath-Brown 2000]) is immediate from the fact that it is the additive Fourier transform of the function $x \mapsto \varphi(x^{-1})$ on $\mathbb{F}_q^*$ (extended by 0),

$$\varphi(x^{-1}) = \frac{1}{q} \sum_{\alpha \in \mathbb{F}_q} K_1(-\alpha)\varphi(\alpha x),$$

and that suitable estimates are available for $K_1(\alpha)$. In fact one knows [Weil 1948a] (see also [Carlitz 1969]) that if α is not 0, then the associated zeta-function is rational,

$$Z(T) = \exp\biggl(-\sum_{m \geqslant 1} \frac{K_1(\alpha, \mathbb{F}^*_{q^m})}{m} T^m\biggr) = \frac{1}{1 - K_1(\alpha)T + qT^2},$$

from which

$$K_1(\alpha, \mathbb{F}^*_{q^m}) = -(\lambda_1^m + \lambda_2^m)$$

[1] The special case when a is a scalar matrix was first considered by Hodges [1956] and reappeared again in various other contexts. See, for example, [Kim 1998; Fulman 2001; Chae and Kim 2003]. We thank Ofir Gorodetsky for bringing our attention to these earlier works.

for some $\lambda_1, \lambda_2 \in \mathbb{C}$. Weil's proof [1948b] of the Riemann hypothesis over function fields shows, using [1948a], that $|\lambda_i| = \sqrt{q}$ which gives the optimal bound

$$|K_1(\alpha, \mathbb{F}_{q^m}^*)| \leq 2q^{m/2}$$

for α not 0. (The explicit description of the connection between exponential sums of this type and the Riemann hypothesis for curves over function fields goes back to [Hasse 1935].)

There are a number of extensions of these results in the commutative setting especially the so called hyper-Kloosterman sums [Mordell 1963; SGA 4½ 1977; Luo et al. 1995; Kowalski et al. 2017], and both these and the classical Kloosterman sums are ubiquitous in the theory of exponential sums [Katz 1988]. There is also a deep connection between Kloosterman sums and modular forms [Poincaré 1911; Petersson 1932; Linnik 1963; Selberg 1965; Deshouillers and Iwaniec 1982; Goldfeld and Sarnak 1983], and the notion of Kloosterman sum is extended to GL_n [Friedberg 1987; Stevens 1987], as well as to other algebraic groups [Dabrowski 1993; Dabrowski and Reeder 1998], with many applications.

The sums $K_n(a)$ considered in this paper are more natural from a ring-theoretic point of view. If A is a finite-dimensional algebra over a finite field, then by the Wedderburn–Artin theorem the additive Fourier transform of $\psi(x^{-1})$ (extended from A^* to A by 0) leads naturally to the matrix Kloosterman sums of (1). These of course are also related to the group GL_n but at the same time they are very strongly tied to the standard representation of this group. From this ring-theoretic point of view we have again

$$\psi(x^{-1}) = \frac{1}{q^{n^2}} \sum_{a \in M_n(\mathbb{F}_q)} K_n(-a)\psi(ax)$$

in the simple ring of $n \times n$ matrices over a finite field.

The other main goal of the paper is then to generalize the classical results above from the Kloosterman sums K_1 to K_n, especially to understand the associated cohomology. The difficulty of this task stems from the fact that when $K_n(a)$ is viewed as an exponential sum on the affine variety

$$V = \{(x, \Delta) \in M_n(\mathbb{F}_q) \times \mathbb{F}_q : \det(x)\Delta = 1\},$$

the part at infinity of the projective closure of V, defined by the equation $\det x = 0$, is singular. However the sums $K_n(a)$ provide a rare example for which their cohomology and so their zeta function is explicitly expressible in terms of one-dimensional exponential sums and so the weights in the sense of Deligne [1980] can be understood in elementary terms. This realization that exponential sums on algebraic groups can be treated rather explicitly is the other main achievement of the paper. The evaluations for the matrix Kloosterman sums in concrete terms, especially the semisimple case in Theorem 1.1, is reminiscent of Herz's work on Bessel functions of matrix arguments [1955]. This link to transcendental special functions continues a long line of similar connections, for example, those of the Gauss, Jacobi, and Kloosterman sums to the gamma, beta and Bessel functions. As an important by-product the concrete nature of these evaluations lead automatically to the estimates required for the equidistribution problem mentioned above.

1B. ***Statements of the results.*** The statements below refer to a fixed finite field $\mathbb{F}_q$, and so we will write $K_n(a)$ for the sum $K_n(a, \mathbb{F}_q)$ in (1). We start with the following reduction. Let a_1, a_2 be matrices of size $n_1 \times n_1$ and $n_2 \times n_2$, and let $a_1 \oplus a_2$ denote the block matrix $\left(\begin{smallmatrix} a_1 & 0 \\ 0 & a_2 \end{smallmatrix}\right)$ of size $(n_1+n_2) \times (n_1+n_2)$.

Theorem 1.1. (i) *Assume that $a_1 \in M_{n_1}(\mathbb{F}_q)$, $a_2 \in M_{n_2}(\mathbb{F}_q)$ and that their characteristic polynomials $p_{a_i}(\lambda)$ are relatively prime. Then*

$$K_{n_1+n_2}(a_1 \oplus a_2) = q^{n_1 n_2} K_{n_1}(a_1) K_{n_2}(a_2).$$

(ii) *Assume that $a \in M_n(\mathbb{F}_q)$ has characteristic polynomial $p_a(\lambda) = \prod_{i=1}^n (\lambda - \alpha_i)$, with $\alpha_i \in \mathbb{F}_q$ all different. Then*

$$K_n(a) = q^{n(n-1)/2} \prod_{i=1}^{n} K_1(\alpha_i),$$

where $K_1(\alpha_i)$ is as in (2).

Now assume that all roots of the characteristic polynomial of a are in $\mathbb{F}_q$. By the theorem above we may assume that a has a unique eigenvalue α. Our first result in this direction is for nilpotent matrices.

Theorem 1.2. *Assume that $a \in M_n(\mathbb{F}_q)$ is nilpotent. Then*

$$K_n(a) = K_n(0) = (-1)^n q^{n(n-1)/2}. \tag{3}$$

In general we have:

Theorem 1.3. *Assume that $a \in \mathrm{GL}_n(\mathbb{F}_q)$ has a unique eigenvalue $\alpha \neq 0$. Denote by λ the partition of n consisting of the sizes of the blocks in the Jordan normal form of a. There exists a polynomial $P_\lambda(A, G, K) \in \mathbb{Z}[A, G, K]$, that depends only on the block partition λ, such that*

$$K_n(a) = P_\lambda(q, q-1, K_1(\alpha)).$$

Remark 1. While irrelevant for estimates, the separation of q and $q-1$ in the above polynomial is natural from the cohomological point of view, as these correspond to sums over the additive group $\mathbb{A}^1$ and the multiplicative group $\mathbb{G}_m$.

The proof of Theorem 1.3 is constructive and allows one to express the Kloosterman sums $K_n(a)$ recursively as a polynomial in q, $q-1$ and $K_1(\alpha_i)$, where α_i runs through the eigenvalues of a. For example, if I_n denotes the identity matrix of size $n \times n$ then we have:

Theorem 1.4. *Assume that $a = \alpha I_n$, $\alpha \neq 0$. Then*

$$K_n(\alpha I_n) = q^{n-1} K_1(\alpha) K_{n-1}(\alpha I_{n-1}) + q^{2n-2}(q^{n-1} - 1)\, K_{n-2}(\alpha I_{n-2}). \tag{4}$$

The recursion formulas for a general partition are somewhat complicated to state but easy to describe algorithmically. See Section 5B, which also contains further examples. These formulas are based on a parabolic Bruhat decomposition (Section 2B). Using the finer decomposition via a Borel subgroup, one can derive closed form expressions. For example, we have:

Theorem 1.5. *If $\alpha \in \mathbb{F}_q^*$ then*

$$K_n(\alpha I_n) = \sum_{\substack{w \in S_n \\ w^2 = I}} q^{n(n-1)/2+N_w}(q-1)^{e_w} K_1(\alpha)^{f_w},$$

where S_n is the symmetric group and where for an involution $w \in S_n$, f_w (resp. e_w) is the number of fixed points (resp. involution pairs) of w (i.e., $n = 2e + f$) and

$$N_w = \bigl|\{(i, j) \mid 1 \leq i < j \leq n,\ w(j) < w(i) \leq j\}\bigr|.$$

One can similarly express $K_n(a)$ for $a = \alpha I + \sum_{i=1}^{n-1} \boldsymbol{e}_{i,i+1}$, when $\alpha \neq 0$ and where $\boldsymbol{e}_{i,j}$ is the elementary matrix with 1 in the (i, j)-th position and 0 everywhere else; see (58) in Section 5E.

The use of the Borel subgroup Bruhat decomposition is also very suitable for deriving estimates for these generalized Kloosterman sums. As a first step we have the following.

Theorem 1.6. *If a has a unique eigenvalue α then*

$$|K_n(a)| \leq |K_n(\alpha I)| \leq 4q^{(3n^2-\delta(n))/4},$$

where $\delta(n) = 0$ if n is even and $\delta(n) = 1$ if n is odd.

Thus if the characteristic polynomial of a splits over $\mathbb{F}_q$, then the estimates do not require much input from étale cohomology. However to bound the sum $K_n(a)$ in general it seems unavoidable to use cohomological methods. The main input from étale cohomology is Lemma 3.16 from which we can derive the following.

Theorem 1.7. *Let $a \in \mathrm{GL}_n(\mathbb{F}_q)$ be a regular semisimple element (i.e., the characteristic polynomial p_a has no multiple roots over $\overline{\mathbb{F}}$). Then the exponential sum $K_n(a)$ is cohomologically pure — that is, all the cohomology groups are trivial but the middle one and all the weights are n^2.*

In particular for these regular semisimple elements, we have "square-root cancellation"

$$|K_n(a)| \leq 2^n q^{n^2/2}.$$

Remark 2. The conditions on the multiplicities of the roots of p_a can be formulated as polynomial equations with integral coefficients in the variables $a_{i,j}$; thus Theorem 1.7 is a concrete illustration of the theorems on "generic purity" of Katz and Laumon [1985] and Fouvry and Katz [2001, Theorem 1.1].

It is an intriguing question if the set of regular semisimple elements is the actual purity locus. The methods of Theorem 1.3, at least for low ranks, allow one to see that matrices, whose Jordan normal form over the algebraic closure has an eigenvalue with more than one Jordan block, are too large for square root cancellations as can be seen by inspection. Interestingly matrices with only one block for each eigenvalue do exhibit square root cancellation. Our methods do not yield enough information to decide whether these sums are cohomologically pure; see Section 5E.

Remark 3. Will Sawin suggested to us a geometric approach that could shed even more light on these sums. The sum $K_n(a)$ may be viewed as the trace of Frobenius on the stalk at a of a complex of sheaves (in fact, a perverse sheaf). The geometry of this complex is linked to the Springer correspondence, in the sense that the stalk at a should be related to the cohomology of the fixed points of a acting on various flag varieties. This is an interesting and promising approach whose exact shape can be conjectured from the recursion formulas. While the paper was going through publication this suggestion was elaborated in [Erdélyi et al. 2024a], settling the purity question in the previous remark in the positive.

Our main theorem for bounding $K_n(a)$ is as follows.

Theorem 1.8. *For all $a \in \mathrm{GL}_n(\mathbb{F}_q)$ we have*

$$|K_n(a)| \ll_n q^{(3n^2-\delta(n))/4},$$

where $\delta(n) = 0$ if n is even and $\delta(n) = 1$ if n is odd.

Remark 4. It is possible to refine the statement based on the characteristic polynomial of a. If the characteristic polynomial is $p_a(t) = \prod_{i=1}^{r}(t-\alpha_i)^{n_i}$ for some pairwise distinct $\alpha_i \in \overline{\mathbb{F}}$, then

$$|K_n(a)| \ll_n \prod_{i=1}^{r} q^{(3n_i^2-\delta(n_i))/4} \prod_{1\le i<j\le r} q^{n_i n_j}.$$

In the classical picture it is natural to look at the more general expression

$$K_1(\alpha,\beta) = K(\alpha,\beta,\mathbb{F}_q^*) = \sum_{\gamma\in\mathbb{F}_q^*} \varphi(\alpha\gamma+\beta\gamma^{-1}), \tag{5}$$

which, in the case $\beta \neq 0$, immediately reduces to $K_1(\alpha\beta^{-1})$ by the simple identity

$$K_1(\alpha,\beta) = K_1(\alpha\delta,\beta\delta^{-1})$$

valid for any $\delta \in \mathbb{F}_q^*$. However the case $\beta = 0$ is again interesting in its own way as it is the Fourier transform of the characteristic function of the set of invertible elements. While trivial to evaluate it is also unavoidable in the analytic applications.

We will look at these sums for $n\times n$ matrices and so we let

$$K_n(a,b) = K_n(a,b,\mathbb{F}_q) = \sum_{x\in\mathrm{GL}_n(\mathbb{F}_q)} \psi(ax+bx^{-1}). \tag{6}$$

It is easy to see that just like as above, we have $K_n(a,b) = K_n(b,a)$ and for any invertible g, h

$$K_n(gah^{-1}, hbg^{-1}) = K_n(a,b). \tag{7}$$

Therefore if $\det a$ or $\det b$ is nonzero, we may use the results above to obtain the bound $|K_n(a,b)| \ll q^{3n^2/4}$. However unlike in dimension 1, the other cases are not settled completely by Theorem 1.2, as the orbit structure of pairs of matrices (a,b) under the $\mathrm{GL}_n\times\mathrm{GL}_n$-action given in (7) gets more intricate when $n > 1$. We start with the most natural case of $b = 0$.

Theorem 1.9. *Assume that $a \in M_n(\mathbb{F}_q)$ has rank r. Then*

$$K_n(a, 0, \mathbb{F}_q) = (-1)^r q^{-r(r+1)/2} q^{rn} |\mathrm{GL}_{n-r}(\mathbb{F}_q)|.$$

Since for an r-dimensional subspace the Moebius function of the poset of subspaces evaluates to $(-1)^r q^{r(r-1)/2}$ by the q-binomial theorem [Stanley 1986, Formula (1.87)], this result is vaguely reminiscent to the evaluation of the Ramanujan sums

$$\sum_{\substack{1 \le x \le q \\ (x,q)=1}} e^{2\pi i a x/q} = \sum_{d|(a,q)} \mu(d) \frac{q}{d}.$$

One can see that the sums $K_n(a, b)$ can get significantly larger than $q^{3n^2/4}$ since for rank $a = 1$ we have $|K_n(a, 0)| \gg q^{n^2-n}$. The next theorem shows that this is close to the worst case scenario.

Theorem 1.10. *Let $a, b \in M_n(\mathbb{F}_q)$ be singular $n \times n$ matrices such that*

$$r = \mathrm{rk}(b) \ge s = \mathrm{rk}(a).$$

Then

$$|K_n(a, b, \mathbb{F}_q)| \le 2q^{n^2 - rn + r^2 + \binom{\min(r, n-r)}{2}}.$$

Corollary 1.11. *If $a, b \in M_n(\mathbb{F}_q)$ are not both* 0, *we have the general estimate*

$$|K_n(a, b, \mathbb{F}_q)| \le 2q^{n^2-n+1}.$$

Remark 5. Apart from the constant 2 which can probably be replaced by $1 + o(1)$ as $q^n \to \infty$, this bound is sharp, since

$$K_n(\boldsymbol{e}_{1,n}, \boldsymbol{e}_{1,n}) = q^{2n-2} |\mathrm{GL}_{n-2}(\mathbb{F}_q)| + (q-1) q^{n-1} |\mathrm{GL}_{n-1}(\mathbb{F}_q)| \sim q^{n^2-n+1}. \tag{8}$$

(Here again $\boldsymbol{e}_{i,j}$ stands for the elementary matrix with 1 in the (i, j)-th position and 0 everywhere else.)

1C. ***The organization of the paper.*** The paper naturally splits into four parts. The first deals with matrices a all of whose eigenvalues are in the ground field $\mathbb{F}_q$ with the aim of evaluating the generalized Kloosterman sums in terms of classical ones. Later sections deal with the nonsplit case using cohomology, and the entirely different and more combinatorial case of degenerate matrices. These are included because they are needed for the equidistribution problem in [El-Baz et al. 2022]. The final part provides some examples and highlights some open questions. The material bifurcates on several occasions and this may somewhat obscure the insight one gains from the results. Therefore we first highlight the generic case of regular semisimple matrices before further details.

To treat regular semisimple elements in concrete terms we need to assume that they are split over the ground field. The calculation to reduce to block upper diagonal a with the blocks having no common eigenvalue is presented in Section 2A. We show here that in that case the Kloosterman sum $K_n(a)$ factors as a product of Kloosterman sums of smaller ranks corresponding to the diagonal blocks. This is elementary and leads immediately to Theorem 1.1. If one is merely interested in this generic case then one can jump to Section 3F which shows how to circumvent the problem when the eigenvalues are only defined in a field extension. In this regard one is also naturally lead to Conjecture 5.9 which suggests that an evaluation might be possible even in the nonsplit case.

The finer picture of the non semisimple case is both of natural interest and required by the applications. We spend a great deal of the paper on them. After the proof of Theorem 1.1, Section 2 and parts of Section 3 handle this case still under the assumption that the eigenvalues are in the ground field $\mathbb{F}_q$. To reach our goals we will evaluate the subsums of $K_n(a)$ restricted to Bruhat cells in various decompositions.

While the calculations in Section 2 are somewhat lengthy the overall structure is simple. We use Bruhat decomposition with respect to a maximal parabolic fixing a line. In Sections 2B and 2C we introduce the necessary notation for this task. This decomposition is the first step in the plain old Gauss–Jordan elimination and it is natural to expect that this process should lead to an inductive structure for matrix Kloosterman sums. An elementary computation in Section 2C justifies this expectation.

Since nilpotent matrices can be put into Jordan normal form over any field this step immediately shows that Theorem 1.3 holds, at least if the polynomial in the statement is allowed to depend on the characteristic p. It is easy to see that in the simplest case of Theorem 1.4 there is a recursion that works over all primes simultaneously. The bounds in Section 3 are based only on this result and if one is only interested in the estimates the rest of the section can be skipped.

However from the exponential sum point of view independence on the characteristic is of great interest in itself. Therefore it is important that the recursions to lower rank can be done across all finite fields universally for matrices whose eigenvalues are in that field. The rest of Section 2 is then devoted to show this. Section 2D presents the technical core by showing that one may restrict to matrices with entries 1 and 0. These matrices can be lifted to $\mathbb{Z}$ and can be put into Jordan normal form over $\mathbb{Q}$. In effect this establishes that Theorem 1.3 holds for almost all primes simultaneously. The final step in the proof is then a simple criterion for the existence of a Jordan normal form over $\mathbb{Z}$ in Section 2F. While this Jordan normal form reduction can be made explicit, the rather technical calculations are postponed to Section 5B.

The second main part of the paper, Section 3, is about estimates. This again starts with an elementary but structural observation that repeated applications of the reduction in Section 2 is equivalent to the use of the Bruhat decomposition with respect to a Borel subgroup. We are able to refine the usual Bruhat decompositions just enough to establish the first bound in Theorem 1.6.

While the proof of the above estimates are still group-theoretic in nature, in general one needs methods from cohomology. Conversely the cohomological apparatus relies on these more classical arguments. The connection is given by a key lemma, Lemma 3.16, whose simple proof somewhat disguises its importance. This statement allows one to push "trivial cancellations" over a field extension back to the original field.

In Section 3E we review the cohomological apparatus by listing the necessary tools for the linear algebra that follows. Sections 3F and 3G then give the proofs of the main theorems, including the special case that the sums $K_n(a)$ are "pure" when a is a regular semisimple element. These results are based on the fact that the cohomology groups attached to the subsum restricted to a Bruhat cell vanishes in large enough degree.

The third part deals with $K_n(a,b)$ in the degenerate cases. After dealing with the case $b=0$ the general situation is reduced to this special case when $b=0$. Again one needs a double coset decomposition suitable for this task. The machinery here is combinatorial in nature using Gaussian binomials [Stanley 1986].

In the last part we provide examples to illustrate our results. We give explicit evaluations for low ranks. The case $K_2(a,b)$ is explicitly computed for any $a,b\in M_2(\mathbb{F}_q)$ including the nonsplit semisimple case. There are further calculations that illustrate the difficulties to get explicit results for $n\geq 4$. These sections also contain some observations and open questions that are interesting on their own.

Notational conventions. Letters of the Greek alphabet $\alpha,\dots,\lambda,\dots,\xi$ will denote scalars, that is, elements of the field $\mathbb{F}_q$, or its algebraic closure. Lower case letters of the Latin alphabet $a,\dots,g,\dots,x$ will denote matrices, but their entries, while scalars, will usually be denoted with Latin characters $a_{ij},\dots,g_{ij},\dots,x_{ij}$ as well. Upper case letters $A,\dots,G,\dots,X$ will denote sets of matrices, usually but not necessarily subgroups. While these depend on n, that dependence is usually suppressed for better readability. There are a few exceptions that should not cause confusion: a general partition will be denoted λ, I will denote the identity matrix (I_n if the size is not clear from the context), K will denote Kloosterman sums, and on occasion $S_1,S_2,\dots$ will denote some auxiliary sums.

For the cohomological arguments in Sections 3E–3G we have to work with algebraic varieties, so then by $A,\dots,G,\dots,X$ we will also denote the affine algebraic varieties of M_n defined by simple algebraic equations in the matrix entries.

There are two notions of a trace in the paper: $\mathrm{Tr}=\mathrm{Tr}_{\mathbb{F}_{q^m}/\mathbb{F}_q}$ will denote the trace from an overfield to the ground field, while $\mathrm{tr}=\mathrm{tr}_n$ will denote the trace of a matrix of size $n\times n$.

If u is a unipotent matrix that is upper triangular, we will denote $\bar{u}=u-I$ the strictly upper triangular part of u. Similarly, if U is the group of upper triangular unipotent matrices and $V\subset U$, then $\overline{V}$ denotes the set (or variety) $\{\bar{v}\mid v\in V\}$.

An important role is played by the Weyl group $W\simeq N(T)/T$ of GL_n. Here T is the set of diagonal matrices in GL_n, and $N(T)$ is its normalizer. If S_n is the group of permutations on the letters $\{1,\dots,n\}$ then $W\simeq S_n$, and we will choose a specific isomorphism in Section 3A.

Cohomology will mean ℓ-adic cohomology with compact support, with some unspecified $\ell\neq p$.

2. Identities via reduction of rank

2A. ***Splitting to primary components.*** We start with some elementary but important observations. Recall from (5) and (2) that

$$K_n(a,b)=\sum_{x\in\mathrm{GL}_n(\mathbb{F}_q)}\psi(ax+bx^{-1})\quad\text{and}\quad K_n(a)=K_n(a,I).$$

The sum is clearly unchanged if x is replaced by x^{-1}, or if x is replaced by cx for some $c\in\mathrm{GL}_n(\mathbb{F}_q)$. This leads immediately to:

Lemma 2.1. *If $a, b \in M_n(\mathbb{F}_q)$ and $c \in \mathrm{GL}_n(\mathbb{F}_q)$ then we have*

$$K_n(a,b) = K_n(b,a), \quad K_n(ac,b) = K_n(a,cb) \quad \textit{and} \quad K_n(a) = K_n(cac^{-1}).$$

Therefore, $K_n(a)$ only depends on the conjugacy class of a and so using conjugation we can put a into Jordan normal form over $\overline{\mathbb{F}}$. To exploit this we recall the following result of Sylvester [1885] whose proof is given here for the convenience of the reader.

Lemma 2.2. *Let $M_{k,l}$ be the vector space of $k \times l$ matrices over an arbitrary field F. If $a_k \in M_k(F)$ and $a_l \in M_l(F)$ are matrices with no common eigenvalue over the algebraic closure $\overline{F}$ of F then the linear endomorphism of $M_{k,l}$ given by*

$$v \mapsto va_l - a_k v$$

is an isomorphism.

Proof. It suffices to prove that this map is injective. Assume that

$$va_l - a_k v = 0$$

for some v. Then for any polynomial $p(t)$

$$vp(a_l) = p(a_k)v.$$

Let $p_k(t) \in F[t]$ (resp. $p_l(t)$) be the characteristic polynomial of a_k (resp. a_l). Then by the Cayley–Hamilton theorem we have

$$0 = p_k(a_k)v = vp_k(a_l) \quad \text{and} \quad 0 = vp_l(a_l).$$

By our assumption on the eigenvalues we have that $p_k(t)$ and $p_l(t)$ are relatively prime in $F[t]$. Thus there are polynomials $r_k(t)$, $r_l(t)$ such that $p_k(t)r_k(t) + p_l(t)r_l(t) = 1$ which implies $0 = v$; hence our claim. □

To use this observation let $U_{[k,l]}$ be the linear subgroup of GL_n whose set of points for any ring R is

$$U_{[k,l]}(R) = \left\{ \left(\begin{array}{c|c} I_k & v \\ \hline 0 & I_l \end{array} \right) \;\middle|\; v \in M_{k,l}(R) \right\}.$$

The fact that this is a subvariety will play a role in the cohomological arguments, but for now we will only use the particular subgroup $U_{k,l}(\mathbb{F}_q) \subset \mathrm{GL}_n(\mathbb{F}_q)$. As usual since the field $\mathbb{F}_q$ is fixed, for easier reading we will write $K_n(a)$ for $K_n(a, \mathbb{F}_q)$, and G and $U_{k,l}$ for the sets $\mathrm{GL}_n(\mathbb{F}_q)$ and $U_{k,l}(\mathbb{F}_q)$.

Proposition 2.3. *Assume that*

$$a = \left(\begin{array}{c|c} a_k & b \\ \hline 0 & a_l \end{array} \right)$$

with $a_k \in M_k(\mathbb{F}_q)$, $a_l \in M_l(\mathbb{F}_q)$, $b \in M_{k,l}(\mathbb{F}_q)$ for some $k, l \in \mathbb{N}$ such that $k + l = n$ and a_k and a_l have no common eigenvalue in $\overline{\mathbb{F}}$. Then

$$K_n(a) = q^{kl} K_k(a_k) K_l(a_l).$$

Proof. Since tr is invariant under conjugation, we have

$$K_n(a)=\sum_{x\in G}\psi(ax+x^{-1})=\frac{1}{q^{kl}}\sum_{x\in G}\sum_{u\in U_{[k,l]}}\psi(a(u^{-1}xu)+(u^{-1}xu)^{-1})$$
$$=\frac{1}{q^{kl}}\sum_{x\in G}\sum_{u\in U_{[k,l]}}\psi((uau^{-1})x+x^{-1}).$$

Now

$$uau^{-1}=\left(\begin{array}{c|c}a_k & b+va_l-a_kv\\ \hline 0 & a_l\end{array}\right)$$

and so

$$K_n(a)=\frac{1}{q^{kl}}\sum_{x\in G}\psi(ax+x^{-1})\biggl(\sum_{v\in\mathbb{F}_q^{k\times l}}\psi_k((va_l-a_kv)x')\biggr),$$

where x' is the $l\times k$ matrix which we get by deleting the first k rows and last l columns of x and ψ_k is the $k\times k$ matrix trace composed with φ.

From Lemma 2.2 we have

$$\sum_{v\in\mathbb{F}_q^{k\times l}}\psi_k((va_l-a_kv)x')=\sum_{v\in\mathbb{F}_q^{k\times l}}\psi_k(vx')=\begin{cases}0 & \text{if } x'\neq 0,\\ q^{kl} & \text{if } x'=0.\end{cases}$$

This immediately yields

$$K_n(a)=\sum_{\substack{x\in G\\ x'=0}}\psi(ax+x^{-1})=q^{kl}K_k(a_k)K_l(a_l). \qquad \square$$

Proof of Theorem 1.1. The first claim was proved above. For the second using the invariance under conjugation we may assume that $a=\operatorname{diag}(\alpha_1,\dots,\alpha_n)$. By Proposition 2.3 we have

$$K_n(a)=q^{n-1}K_1(\alpha_n)K_{n-1}(a'),$$

where $a'=\operatorname{diag}(\alpha_1,\dots,\alpha_{n-1})$. The result follows by induction. □

2B. ***A parabolic Bruhat decomposition.*** We prepare the proofs of Theorems 1.3 and 1.4. Our goal is to express the Kloosterman sum $K_n(a)$ in terms of sums $K_{n-1}(a')$ and $K_{n-2}(a'')$ where the matrices a' and a'' are derived from a by deleting one or two rows and columns.

When a has a single eigenvalue α, it is conjugate to

$$a=\alpha I+\bar{a}, \tag{9}$$

where $\bar{a}$ is strictly upper triangular. Since $K_n(a)$ is conjugacy invariant, we will assume that a itself is in the above form. Our reductions are then based on a parabolic Bruhat decomposition for GL_n. While it can be deduced from general facts — see [Springer 1998, Theorem 8.3.8; Borel 1991, Proposition IV.14.21(iii)] — it is easier to work them out explicitly for the special case at hand.

Let P be the closed subgroup of GL_n defined by the vanishing of $g_{n,1}, \dots, g_{n,n-1}$. If F is a field, P may be described alternatively as follows. Let $\mathrm{GL}_n(F)$ act on row vectors by multiplication on the right: $v \mapsto vg$. Then $P(F)$ is the stabilizer of the line $\{\lambda\, \boldsymbol{e}_n : \lambda \in F\}$ where $\boldsymbol{e}_n$ is the row vector $(0, \dots, 0, 1)$:

$$P(F) = \{g \in \mathrm{GL}_n(F) \mid \boldsymbol{e}_n g = \lambda \boldsymbol{e}_n,\ \lambda \in F^*\}. \tag{10}$$

Since the arguments in this and the following sections will not be used in the cohomological proofs we will only concentrate on the set $P(\mathbb{F}_q)$.

Then as sets $G = \mathrm{GL}_n(\mathbb{F}_q) = \bigsqcup_{k=1}^{n} P(\mathbb{F}_q) w_{(kn)} P(\mathbb{F}_q)$, where $w_{(kn)}$ is the permutation matrix corresponding to the transposition (kn). To make this a parameterization let

$$U_k = \left\{ I + \sum_{j=k+1}^{n} u_j \boldsymbol{e}_{k,j} \,\middle|\, u_{k,j} \in F \right\} \tag{11}$$

be the set of unipotent matrices with nonzero elements only in the k-th row. (While this is again an algebraic group scheme, this fact will not play any role.)

We will only deal with $F = \mathbb{F}_q$ and from here on we will write P and U_k for $P(\mathbb{F}_q)$ and $U_k(\mathbb{F}_q)$. We then have the following decomposition into disjoint sets.

Lemma 2.4. *Let $X_k = U_k w_{(kn)} P$. The map $U_k \times P \to X_k$, $(u, g) \mapsto u w_{(kn)} g$, is a bijection and $G = \bigsqcup_{k=1}^{n} U_k w_{(kn)} P$.*

Proof. Let x be a matrix in $G = \mathrm{GL}_n(\mathbb{F}_q)$ with rows $\boldsymbol{x}_i$, and write

$$\boldsymbol{e}_n = \sum_{j=1}^{n} u_j \boldsymbol{x}_j,$$

where $\boldsymbol{e}_n = (0, \dots, 0, 1)$. We claim that

$$X_k = \bigl\{x \in G \mid \min\{j \mid u_j \neq 0\} = k\bigr\}.$$

It is clear that $X_k P = X_k$ and that $U_k w \subset X_k$ with $w = w_{(kn)}$. Conversely if we let

$$u = I + \sum_{j=k+1}^{n} (u_j/u_k) \boldsymbol{e}_{k,j}$$

then we have $ux \in wP$.

Finally it is enough to show that if $u_1 w p_1 = u_2 w p_2$, then $u_1 = u_2$. To this effect note that the above implies that $\boldsymbol{e}_n w u_2^{-1} u_1 = \boldsymbol{e}_n p_2 p_1^{-1} w$. However $\boldsymbol{e}_n w = \boldsymbol{e}_k$ and so the k-th rows of u_1 and u_2 are the same, which implies $u_1 = u_2$. □

By the lemma we have

$$K_n(a) = \sum_{g \in \mathrm{GL}_n(\mathbb{F}_q)} \psi(ag + g^{-1}) = \sum_{k=1}^{n} \sum_{x \in X_k} \psi(ax + x^{-1}). \tag{12}$$

When summing over X_k, the case of $k=n$, when $X_n = P$, is trivial. To see this we will give an explicit Levi decomposition of $P = P(\mathbb{F}_q)$. This fixes notation and will also be used in our further calculations on X_k for $k < n$. Note that again we will be working not with the algebraic groups but the fixed finite groups that they give rise to for $\mathbb{F}_q$.

For $h \in \mathrm{GL}_{n-1}(\mathbb{F}_q)$ and $\mu \in \mathrm{GL}_1(\mathbb{F}_q)$ let

$$[h,\mu] = \begin{bmatrix} h & 0 \\ 0 & \mu \end{bmatrix} \in \mathrm{GL}_n(\mathbb{F}_q) \tag{13}$$

and let

$$L = \{[h,\mu] : h \in \mathrm{GL}_{n-1}(\mathbb{F}_q),\ \mu \in \mathrm{GL}_1(\mathbb{F}_q)\} \subset \mathrm{GL}_n(\mathbb{F}_q). \tag{14}$$

Also let

$$V = \left\{ I + \sum_{l=1}^{n-1} v_l \boldsymbol{e}_{l,n} : v_l \in \mathbb{F}_q \right\}. \tag{15}$$

Then

$$P = LV = VL.$$

Proposition 2.5. *If a is as in* (9) *then*

$$\sum_{x\in X_n} \psi(ax + x^{-1}) = q^{n-1} K_1(\alpha) K_{n-1}(a'),$$

where a' is the matrix one gets by deleting the last row and column of a.

Proof. Since

$$\sum_{x\in X_n} \psi(ax+x^{-1}) = \sum_{\substack{g\in L\\ v\in V}} \psi(agv + (gv)^{-1}) = q^{n-1} \sum_{g\in L} \psi(ag+g^{-1}),$$

the claim follows from the description of L in (14) and that

$$\operatorname{tr}(agv + (gv)^{-1}) = \operatorname{tr}(ag + g^{-1}) = \operatorname{tr}(a'h + h^{-1}) + \alpha\lambda + \lambda^{-1}. \qquad \square$$

2C. ***The sum over the nontrivial cells.*** We continue to assume that $a = \alpha I_n + \bar{a}$, with $\bar{a}$ strictly upper triangular, so that a has a unique eigenvalue α. We will show that for $k < n$ the sum

$$\sum_{x\in X_k} \psi(ax+x^{-1})$$

can be expressed as a sum over the subvariety

$$L_k(\alpha) = \{ g \in L \mid g_{k,j} = 0 \text{ for all } j \neq k \text{ and } \alpha g_{k,k} = g_{n,n}^{-1} \}. \tag{16}$$

However we will only work over the set of points in $\mathbb{F}_q$ and write $L_l(\alpha)$ for $L_k(\alpha)(\mathbb{F}_q)$. For $\alpha = 0$ these sets are empty, while for $\alpha \in \mathbb{F}_q^*$ they are subvarieties of L isomorphic to $\mathrm{GL}_{n-2} \times \mathrm{GL}_1 \times \mathbb{A}^{n-2}$ that can be visualized as elements $g \in L$ of the form

$$g = \begin{bmatrix} h_{11} & * & h_{12} & 0 \\ 0 & \mu & 0 & 0 \\ h_{21} & * & h_{22} & 0 \\ 0 & 0 & 0 & 1/(\alpha\mu) \end{bmatrix}. \tag{17}$$

Here the blocks correspond to the partition

$$\{1, \dots, n\} = \{1, \dots, k-1\} \sqcup \{k\} \sqcup \{k+1, \dots, n-1\} \sqcup \{n\}$$

for $1 < k < n-1$, while for $k = 1$ and $n-1$ we have to adapt (17) to 3×3 blocks

$$g = \begin{bmatrix} \mu & 0 & 0 \\ * & h'' & 0 \\ 0 & 0 & 1/(\alpha\mu) \end{bmatrix}, \quad g = \begin{bmatrix} h'' & * & 0 \\ 0 & \mu & 0 \\ 0 & 0 & 1/(\alpha\mu) \end{bmatrix}. \tag{18}$$

This is merely a preliminary step in the reduction to rank $n-2$, but is already quite useful, a fact that we will illustrate by proving Theorems 1.2 and 1.4.

The reduction to the special form in (17) is based on the following calculation.

Proposition 2.6. *For a fixed $g \in L$,*

$$\sum_{v \in V} \psi(\alpha w_{(kn)} g v + (w_{(kn)} g v)^{-1}) = 0$$

unless

$$g_{k,j} = 0 \quad \textit{for all } j \neq k \qquad \textit{and} \qquad \alpha g_{k,k} = g_{n,n}^{-1}.$$

When these conditions hold

$$\sum_{v \in V} \psi(\alpha w_{(kn)} g v + (w_{(kn)} g v)^{-1}) = q^{n-1} \psi(\alpha w_{(kn)} g + (w_{(kn)} g)^{-1}).$$

Proof. Let $\overline{V} = \{\bar{v} = \sum_{l=1}^{n-1} v_l \boldsymbol{e}_{l,n} : v_l \in \mathbb{F}_q\}$. If $v = I + \bar{v} \in V$ then $v^{-1} = I - \bar{v}$. The sum in question then becomes

$$\sum_{v \in V} \psi(\alpha w_{(kn)} g v + (w_{(kn)} g v)^{-1}) = \psi(\alpha w_{(kn)} g + (w_{(kn)} g)^{-1}) \sum_{\bar{v} \in \overline{V}} \psi(\alpha w_{(kn)} g \bar{v} - \bar{v}(w_{(kn)} g)^{-1}),$$

which vanishes unless the linear function

$$v \mapsto \operatorname{tr}(\alpha w_{(kn)} g \bar{v} - \bar{v}(w_{(kn)} g)^{-1}) = \alpha \sum_{l=1}^{n-1} g_{k,l} v_l - g_{n,n}^{-1} v_k$$

is trivial. □

We can now prove the following claim about the sum over X_k.

Proposition 2.7. *Let $a = \alpha I + \bar{a}$, where $\bar{a}$ is strictly upper triangular, $k < n$, X_k is as in Lemma 2.4 and $L_k(\alpha)$ is as in* (16). *Then we have*

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = q^{n-1} \sum_{g \in L_k(\alpha)} \sum_{u \in U_k} \psi(a^u w_{(kn)} g + (w_{(kn)} g)^{-1}),$$

where $a^u = u^{-1} a u$.

Proof. Since $Pu = P$ for any $u \in U_k$, we have

$$X_k = \bigsqcup_{u \in U_k} u w_{(kn)} P = \bigsqcup_{u \in U_k} u w_{(kn)} P u^{-1},$$

and so by $u^{-1} a u = \alpha I + u^{-1} \bar{a} u$ we have

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = \sum_{\substack{g \in L \\ v \in V \\ u \in U_k}} \psi(\alpha w_{(kn)} g v + (w_{(kn)} g v)^{-1}) \psi(u^{-1} \bar{a} u w g v).$$

A direct calculation, based on the fact that the last row of $\bar{a} = a - \alpha I$ is identically 0, then shows that

$$\operatorname{tr}(u^{-1} \bar{a} u w g v) = \operatorname{tr}(u^{-1} \bar{a} u w g) \tag{19}$$

is independent of v. Therefore

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = \sum_{\substack{g \in L \\ u \in U_k}} \psi(u^{-1} \bar{a} u w g) \sum_{v \in V} \psi(\alpha w_{(kn)} g v + (w_{(kn)} g v)^{-1}).$$

The inner sum is identical to the one in Proposition 2.6; thus the proposition is proven. □

Remark 6. We comment briefly on identity (19). The calculations are simplified by using $M_n(\mathbb{F}_q)$, writing $v = I + \bar{v}$, and observing that $\operatorname{tr} x \bar{v} = \sum_{l=1}^{n} x_{n,l} v_l$ which clearly vanishes if the last row of the matrix x is identically 0.

However one may argue alternatively via interpreting these matrices as linear transformations as follows. The group P is the parabolic subgroup fixing the 1-dimensional subspace $M = \{\lambda \boldsymbol{e}_n \mid \lambda \in \mathbb{F}_q\}$ and so its elements also preserve the flag $\{0\} \subset M \subset \mathbb{F}_q^n$. Any element g of P then gives rise to a linear transformation g' of $M' = \mathbb{F}_q^n / M$. The subgroup V itself is characterized by the property that its elements act trivially both on M and on M'. Let $x = u^{-1} \bar{a} u w g$. Since $\boldsymbol{e}_n x = 0$, the linear transformation x also induces a map x' on M' and $\operatorname{tr} x = \operatorname{tr} x'$. Since $\boldsymbol{e}_n x v = 0$ as well, $\operatorname{tr} x v = \operatorname{tr}(xv)' = \operatorname{tr} x' v' = \operatorname{tr} x'$.

In general all our calculations can easily be proved using block matrices, either by hand or by using a symbolic algebra package. Since this gives an easy way to check the validity of these statements, we will present most of the identities in this matrix interpretation.

As a corollary to Proposition 2.7 we immediately have:

Proof of Theorem 1.2. If $\alpha = 0$ then the set $L_k(\alpha)$ is empty, and so for $k < n$

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = 0.$$

Since $K_1(0) = -1$, Proposition 2.5 gives

$$K_n(a) = -q^{n-1} K_{n-1}(a'),$$

where a' is the matrix one gets by deleting the last rows and columns of a. Since by assumption a is upper triangular nilpotent, so is a' and the theorem follows by induction. □

Proof of Theorem 1.4. If $a = \alpha I$, with $\alpha \in \mathbb{F}_q^*$ a scalar matrix, then $\bar{a} = 0$ and $a^u = \alpha I$. If $1 < k < n-1$ and g is as in (17) then g is invertible if and only if $g'' = \left[\begin{smallmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{smallmatrix}\right]$ is invertible, in which case $(g^{-1})'' = (g'')^{-1}$. It follows that for such k

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = q^{2n-3}(q-1) K_{n-2}(\alpha I) q^{n-k},$$

and it is easy to see that this relation holds for $k = 1, n-1$ as well. This gives

$$K_n(\alpha I) = \sum_{k=1}^{n} \sum_{x \in X_k} \psi(\alpha x + x^{-1}) = q^{n-1} K_1(\alpha) K_{n-1}(\alpha I) + q^{2n-3}(q-1) K_{n-2}(\alpha I) \sum_{k=1}^{n-1} q^{n-k}$$

from which the desired formula follows. □

2D. *Reduction to* $\mathbf{GL}_{n-2}$. Assume that $a = \alpha I + \bar{a}$ has a unique eigenvalue $\alpha \neq 0$, and that $\bar{a}$ is strictly upper triangular. Since the results of the previous section take care of the case when $n = 2$ or $a = \alpha I$, we will assume that $n \geq 3$ and that $\bar{a} \neq 0$.

Recall that $\sum_{x \in X_k} \psi(ax + x^{-1}) = q^{n-1} \sum_{u,g} \psi(a^u w_{(kn)} g + (w_{(kn)} g)^{-1})$, the sums over $u \in U_k$ and $g \in L_k(\alpha)$. We will use the fact that as a variety $L_k(\alpha)$ is isomorphic to $\mathrm{GL}_{n-2}(\mathbb{F}_q) \times \mathbb{F}_q^* \times \mathbb{F}_q^{n-2}$ to express $\sum_{x \in X_k} \psi(ax + x^{-1})$ as a linear combination of Kloosterman sums of rank $n-2$.

To state the reduction step we denote by $m''_{\not{k},\not{n}}$ the matrix one gets by deleting the k-th and n-th rows and columns of a matrix m. For us n will be fixed, and the value of k will be clear from the context, in which case we will often simply write m''. For any matrix m let $m_{(k)}$ denote its k-th row, and $m^{(l)}$ denote its l-th column.

Proposition 2.8. *Assume that $n > 2$, $a = \alpha I + \bar{a}$ with $\bar{a}$ strictly upper triangular and let $u = I + \bar{u} \in U_k$. Then we have that*

$$\sum_{g \in L_k(\alpha)} \psi(a^u w_{(kn)} g + (w_{(kn)} g)^{-1}) = 0$$

unless $\bar{u}_{(k)} \bar{a}^{(j)} = \bar{a}_{k,j}$ for $j = k+1, \ldots, n-1$. When this condition holds

$$\sum_{g \in L_k(\alpha)} \psi(a^u w_{(kn)} g + (w_{(kn)} g)^{-1}) = q^{n-1} K_{n-2}(a'' + z) \sum_{\lambda \in \mathbb{F}_q^*} \varphi(\lambda \xi),$$

where $z = (\bar{a}^{(k)} \bar{u}_{(k)})'' \in M_{n-2}$ and $\xi = a_{k,n} - \bar{u}_{(k)} \bar{a}^{(n)}$.

Remark 7. When $k = 1$ or $n - 1$ the perturbation z vanishes for any u.

Proof. A direct calculation shows that

$$a^u = (I - \bar{u})a(I + \bar{u}) = a - \bar{u}\bar{a} + \bar{a}\bar{u}. \tag{20}$$

First assume that $1 < k < n - 1$ and that

$$g = \begin{bmatrix} g_{11} & y_1 & g_{12} & 0 \\ 0 & \mu & 0 & 0 \\ g_{21} & y_2 & g_{22} & 0 \\ 0 & 0 & 0 & 1/(\alpha\mu) \end{bmatrix} \in L_k(\alpha)$$

as in (17), in which case $g'' = \left[\begin{smallmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{smallmatrix}\right]$ is invertible. Let tr_n denote the $n \times n$ matrix trace. Then

$$\mathrm{tr}_n((wg)^{-1}) = \mathrm{tr}_{n-2}((g'')^{-1}).$$

Moreover

$$\mathrm{tr}_n(awg) = \mathrm{tr}_{n-2}(a''g'') + \mu a_{k,n} \quad \text{and} \quad \mathrm{tr}_n(\bar{a}\bar{u}wg) = \mathrm{tr}_{n-2}((a^{(k)}u_{(k)})''g''),$$

where we have used the fact that $\bar{u}$ has only one nonzero row $\bar{u}_{(k)}$.

Finally note that $\mathrm{tr}(a^u wg + (wg)^{-1})$ does not depend on the $(k-1) \times 1$ matrix y_1, and as a function of y_2 only depends on $\mathrm{tr}(\bar{u}\bar{a}wg)$. The function

$$y_2 \mapsto \mathrm{tr}(\bar{u}\bar{a}wg)$$

is constant if and only if

$$\bar{u}_{(k)}\bar{a}^{(j)} = 0$$

for $j = k+1, \dots, n-1$, and if this condition does not hold the sum over $g \in L_k(\alpha)$ vanishes. This proves the claim for $1 < k < n - 1$. The remaining cases are treated similarly. □

We will now specify the result in the case $a = \alpha I + \bar{a}$ is in Jordan normal form. There is a partition λ associated to a, i.e., a sequence of positive integers $n_1 \le n_2 \le \dots \le n_l$, such that $n_1 + n_2 + \dots + n_l = n$. Conversely, given a partition $\lambda = [n_1, \dots, n_l]$ let

$$N_i = n_1 + \dots + n_i \tag{21}$$

so that $1 \le N_1 < N_2 < \dots < N_l = n$, and form

$$\bar{a}(\lambda) = \sum_{j=1}^{n-1} \varepsilon_j(\lambda)\boldsymbol{e}_{j,j+1}, \quad \text{where} \quad \varepsilon_j(\lambda) = \begin{cases} 0 & \text{if } j = N_i, \text{ for some } i, \\ 1 & \text{otherwise.} \end{cases} \tag{22}$$

Any a with a single eigenvalue α is conjugate to one of the matrices $\alpha I + \bar{a}(\lambda)$ and we will assume from now on that a is already in that form. Since scalar matrices were already dealt with, we will also assume that $\lambda \ne [1, 1, \dots, 1]$, which ensures that $\varepsilon_{n-1} = 1$.

Theorem 2.9. *Assume that $n > 2$, $\alpha \neq 0$, $a = \alpha I + \bar{a}(\lambda)$ with $\varepsilon_j = \varepsilon_j(\lambda) \in \{0, 1\}$ as in* (22) *with $\varepsilon_{n-1} = 1$. Then*:

(i) *We have*

$$\sum_{x \in X_{n-1}} \psi(ax + x^{-1}) = -q^{2n-2} K_{n-2}(a''),$$

where $a'' = a''_{\cancel{n-1},\cancel{n}}$ — the matrix obtained by deleting the last two rows and columns of a.

(ii) *If $k \leq n-2$ then*

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = 0$$

unless $k = N_i$ for one of the N_i in (21) *for which $n_i > 1$.*

(iii) *When $k = N_i = n_1 + \cdots + n_i < n-1$ and $n_i > 1$, we have*

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = q^{2n-2} \sum_{\substack{z \in Z \\ \mu \in \mathbb{F}_q^*}} K_{n-2}(a'' + z)\varphi(\mu\xi_l), \tag{23}$$

where $a'' = a''_{\cancel{k},\cancel{n}}$ and

$$Z = \left\{ \sum_{j=i+1}^{l-1} \xi_j \boldsymbol{e}_{k-1,N_j-1} + \xi_l \boldsymbol{e}_{k-1,n-2} \;\middle|\; \xi_j \in \mathbb{F}_q \text{ for } i+1 \leq j \leq l \right\} \subset M_{n-2}.$$

In Z the elementary matrices $\boldsymbol{e}_{*,*}$ are of size $(n-2)\times(n-2)$.

Proof. The statements are easy corollaries of Proposition 2.8 except for the fact in (ii) that n_i must be greater than 1, which is equivalent to $\varepsilon_{k-1} \neq 0$. Since $k < n-1$ and $\varepsilon_{k-1} = 0$ imply that the parameters in Proposition 2.8 are very simple, all $z = 0$, and $\xi = -\bar{u}_{k,n-1}$. Therefore

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = q^{2n-2} \sum_{u} K_{n-2}(a'') \sum_{\mu \in \mathbb{F}_q^*} \varphi(-\mu\bar{u}_{k,n-1})$$

vanishes. □

While the matrices $a'' + z$ are not in Jordan normal form, they are again matrices with a single eigenvalue α. Therefore collecting them according to their conjugacy classes gives a reduction algorithm, in fact, a characteristic p version of Theorem 1.3 (see Proposition 2.13). In the next two sections we will explicate this strategy and prove that the polynomials that arise this way do not depend on p.

2E. ***Jordan normal forms over $\mathbb{Z}$.*** The proof of Theorem 1.3 will be based on proving that the perturbations arising from the reduction to M_{n-2} can be collected into Jordan normal forms that do not depend on the characteristic p. For this we will need some details about Jordan normal forms for integral nilpotent matrices. This of course is a trivial task over $\mathbb{Q}$, but requires a little care when one works over $\mathbb{Z}$.

Assume, for example, that x is an $n \times n$ nilpotent matrix, and $g \in \mathrm{GL}_n(\mathbb{Z})$ is such that gxg^{-1} is in Jordan normal form as in (22). A moment's thought reveals that $\{vx \mid v \in \mathbb{Z}^n\}$, the row space of x, must be a direct summand of $\mathbb{Z}^n$, which we also think of as row vectors. This by itself is not sufficient for the existence of a Jordan normal form over $\mathbb{Z}$ but we have the following.

Theorem 2.10. *Let $\bar{a}$ be an $n \times n$ nilpotent matrix with integral entries. There exist $g \in \mathrm{GL}_n(\mathbb{Z})$ such that $g\bar{a}g^{-1} = \sum_{j=1}^{n-1} \varepsilon_j \boldsymbol{e}_{j,j+1}$, $\varepsilon_j \in \{0, 1\}$ if and only if*

$$\{v\bar{a}^k \mid v \in \mathbb{Z}^n\}$$

is a direct summand of $\mathbb{Z}^n$ (as an abelian group) for any $k \in \mathbb{N}$.

This is equivalent to the conditions that

$$\{\bar{a}^k v^T \mid v \in \mathbb{Z}^n\}$$

is a direct summand of $(\mathbb{Z}^n)^T$ for any k (here $\cdot^T$ is the matrix transpose).

The following examples illustrate the situation.

Example 2.11. Let $x = \left[\begin{smallmatrix} 0 & 2 \\ 0 & 0 \end{smallmatrix}\right]$. Its Jordan normal form over $\mathbb{Q}$ is $y = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. If $g = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$, the equation $gx = yg$ leads to $c = 0$ and $d = 2a$. Therefore the equation $gxg^{-1} = y$ has no solution in $\mathrm{SL}_2(\mathbb{Z})$, or even $\mathrm{SL}_2(\mathbb{Q})$.

Example 2.12. Let

$$x = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with normal form

$$y = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The $\mathbb{Z}$-span of the rows of x is clearly a direct summand. If $gxg^{-1} = y$ then also $gx^2 = y^2 g$, but

$$x^2 = \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad y^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

and so a Jordan normal form over $\mathbb{Z}$ does not exist.

Theorem 2.10 follows along the lines of the standard proofs in the case of a vector space over a field. For completeness we present such a proof below, but for both this proof and the applications of the theorem it is more convenient to work with linear transformations than matrices.

Let R be either $\mathbb{F}_q$ or $\mathbb{Z}$ and V a free R-module of finite rank n. If $A : V \to V$ is an R-homomorphism, then it gives rise to an $R[T]$-module structure on V, where $R[T]$ is the polynomial ring over R, and

$Tv := A(v)$. If needed we will denote these $R[T]$-modules by V_A to distinguish modules corresponding to different transformations.

We will be interested in the situation when A is nilpotent: $A^n = 0$. If $v \in V$ let k be minimal such that $A^k v = 0$ and let

$$\langle v \rangle = Rv + R(Av) + \cdots + R(A^{k-1}v) \tag{24}$$

denote the cyclic submodule generated by v. In this case we will call V cyclic if $V_A = \langle v \rangle$ for some $v \in V_A$. This happens exactly when the $R[T]$ module V_A is isomorphic to $\mathcal{C}_n = R[T]/(T^n)$. If R is a field, then any V_A is a direct sum of cyclic modules, but this fails for $R = \mathbb{Z}$, and in general when $R[T]$ is not a principal ideal domain.

Proof of Theorem 2.10. The theorem is equivalent to the following statement: if $V \simeq \mathbb{Z}^n$, and $A : V \to V$ is a nilpotent homomorphism, then the $\mathbb{Z}[T]$ module V_A is a direct sum of cyclic modules if and only if $A^k(V)$ is a direct summand of V (as an abelian group) for all $k \in \mathbb{N}$.

By the structure theorem for finitely generated abelian groups a subgroup V' of V is a direct summand if and only if for $k \in \mathbb{Z}$, $v \in V$, $kv \in V'$ implies that $v \in V'$. This immediately shows that $\ker A$ is a direct summand of V, and $A(V) \cap \ker A$ is a direct summand of $\ker A$. Let V_0 be a complementary direct summand so that

$$\ker A = V_0 \oplus (A(V) \cap \ker A). \tag{25}$$

Since A is nilpotent, the rank of $A(V)$ is strictly less than that of V. The condition on A descends to $A(V)$, and so by induction we have that $A : A(V) \to A(V)$ has a cyclic basis, i.e., there are $v_1, \ldots, v_l$ such that $A(V) \simeq \langle Av_1 \rangle \oplus \cdots \oplus \langle Av_l \rangle$. If we let d_i be the smallest integer k such that $A^k v_i = 0$, then we have that the set

$$\{A^j v_i \mid i = 1, \ldots, l,\ j = 1, \ldots, d_i - 1\} \tag{26}$$

is a basis of the free abelian group $A(V)$.

Let $v_{l+1}, \ldots, v_r$ be such that $V_0 = \bigoplus_{i=l+1}^r \mathbb{Z}v_i$, where V_0 is as in (25). Extending the notation from above we let $d_i = 1$ for $i = l+1, \ldots, r$.

We claim that

$$V_A \simeq \bigoplus_{j=1}^{r} \langle v_j \rangle.$$

We need to prove that for each $v \in V$, there is a unique choice of $\alpha_{i,j} \in \mathbb{Z}$ such that

$$v = \sum_{i=1}^{r} \sum_{j=0}^{d_i-1} \alpha_{i,j} A^j v_i. \tag{27}$$

To see uniqueness assume that $v = 0$ is expressed this way. Then $Av = 0$ as well, and the linear independence of the set in (26) shows that $\alpha_{i,j} = 0$ for all $i = 1, \ldots, l$ and $j = 0, \ldots, d_i - 2$. Since by (25) we have that $A^{d_i-1}v_i$, for $i = 1, \ldots, l$, and $v_{l+1}, \ldots, v_r$ are linearly independent, this shows that $\alpha_{i,d_i-1} = 0$ as well for all $i = 1, \ldots, r$.

It remains to show that every $v \in V$ can be expressed as an integral linear combination as in (27). By (26) this is clearly true for Av:

$$Av = \sum_{i=1}^{l} \sum_{j=0}^{d_i - 2} \alpha_{i,j} A^i (A v_j)$$

for some $\alpha_{i,j} \in \mathbb{Z}$. Let $v' = \sum_{i=1}^{l} \sum_{j=0}^{d_i-2} \alpha_{i,j} A^i v_j$. Then $v - v' \in \ker A$; thus proving the claim. □

2F. ***The proof of Theorem 1.3.*** The proof of Theorem 1.3 relies on the following two propositions.

Proposition 2.13. *Assume that $n \geq 2$, $\alpha \in \mathbb{F}_q^*$, $a = \alpha I + \bar{a}(\lambda)$ with $\varepsilon_j \in \{0, 1\}$ as in* (22) *with $\varepsilon_{n-1} = 1$. Also assume that $z_1 = \sum_{j=i+1}^{l-1} \xi_j \boldsymbol{e}_{k-1,N_j-1} + \xi_l \boldsymbol{e}_{k-1,n_2}$, and that*

$$z_2 = \sum_{j=i+1}^{l-1} \eta_j \boldsymbol{e}_{k-1,N_j-1} + \eta_l \boldsymbol{e}_{k-1,n_2}, \quad \textit{where } \eta_j = \begin{cases} 0 & \textit{if } \xi_j = 0, \\ 1 & \textit{if } \xi_j \neq 0. \end{cases}$$

Then $a'' + z_1$ and $a'' + z_2$ are conjugate in $\mathrm{GL}_{n-2}(\mathbb{F}_q)$, *and so $K_{n-2}(a'' + z_1) = K_{n-2}(a'' + z_2)$.*

As a corollary of Proposition 2.13 one immediately has that for $k = N_i$ as above

$$\sum_{x \in X_k} \psi(ax + x^{-1}) = q^{2n-2} \sum_{z \in Z_0} (q-1)^{J+1} (K_{n-2}(a'' + z) - K_{n-2}(a'' + z + \boldsymbol{e}_{k-1,n-2})), \tag{28}$$

where $a'' = a''_{\not k,\not\eta}$, and where z runs over

$$Z_0 = \left\{ z = \sum_{j=i+1}^{l-1} \eta_j \boldsymbol{e}_{k-1,N_j-1} \;\middle|\; \eta_j \in \{0,1\} \text{ for } i+1 \leq j \leq l \right\}, \tag{29}$$

with $J = J(z) = |\{j \mid \eta_j = 1\}|$.

This in itself proves a version of Theorem 1.3 valid for almost all primes. The matrices $a'' + z$, $a'' + z + \boldsymbol{e}_{k-1,n-2}$ can obviously be lifted to $\mathbb{Z}$ where they can be put into Jordan normal form after a rational change of basis. This shows there are well-defined partitions $\lambda''(z)$, $\lambda''(z + \boldsymbol{e}_{k,n-2})$ of $n-2$ associated to the partition λ, the value $k = N_i$ and z which are independent of p except for the finitely many primes dividing the determinant of the change of base matrix. The next proposition shows that such exceptions do not arise.

Proposition 2.14. *Let $\bar{a} \in M_n(\mathbb{Z})$ be as in Proposition 2.13 and $z \in Z_0$ be as in* (29). *There exists a unique partition $\lambda'' \vdash n-2$ such that for any prime p and finite field $\mathbb{F}_q$ of characteristic p the partition $\lambda''(z) \vdash n-2$ associated to the nilpotent matrix $\bar{a}'' + z$ equals λ''.*

Similarly there exist a partition $\mu'' \vdash n-2$ such that the block partition of the matrix $\bar{a}'' + z + \boldsymbol{e}_{k-1,n-2}$ is μ''.

Proof of Proposition 2.13. Again we will use the language of linear transformations. Let V be a finite-dimensional vector-space over $\mathbb{F}_q$, and $A : V \to V$ a nilpotent linear transformation such that

$$V_A \simeq \langle v_0 \rangle \oplus \langle v_1 \rangle \oplus \cdots \oplus \langle v_l \rangle.$$

Let $d_i = \dim \langle v_i \rangle$. We are interested in perturbations $A+Z_1$, $A+Z_2$ of A, where Z_1, Z_2 are such that

$$\begin{aligned} Z_1(A^j v_i) = Z_2(A^j v_i) = 0 \quad &\text{for } i = 1, \dots, l,\ j = 0, \dots, d_i - 1, \\ Z_1(A^j v_0) = Z_2(A^j v_0) = 0 \quad &\text{for } j = 0, \dots, n_0 - 2, \end{aligned}$$

and

$$Z_1(A^{d_0-1}v_0) = \sum_{i=1}^{l} \xi_i A^{d_i-1} v_i \quad \text{and} \quad Z_2(A^{d_0-1}v_0) = \sum_{i=1}^{l} \eta_i A^{d_i-1} v_i,$$

where $\eta_i = 0$ or 1 depending on whether $\xi = 0$ or not.

Let $\phi : V_A \to V_A$ be the A-linear ($\phi \circ A = A \circ \phi$) isomorphism for which

$$\phi(v_i) = \begin{cases} \frac{1}{\xi_i} v_i & \text{for } 1 \le i \le l,\ \xi_i \ne 0, \\ v_i & \text{for } 1 \le i \le l,\ \xi_i = 0. \end{cases}$$

Then $\phi(Z_1(v)) = Z_2(\phi(v))$ as well, showing that the modules V_{A+Z_1} and V_{A+Z_2} are isomorphic. □

Proof of Proposition 2.14. The statement is only meaningful if $n > 2$. Lift the matrix $\bar{a}''$ which only has entries 0 and 1 to a matrix $\tilde{a}$ over $\mathbb{Z}$ by lifting $0_{\mathbb{F}_q}$ to $0_{\mathbb{Z}}$ and $1_{\mathbb{F}_q}$ to $1_{\mathbb{Z}}$. Identify $\mathbb{Z}^{n-2}$ with $1 \times (n-2)$ matrices (row vectors) and let $A : \mathbb{Z}^{n-2} \to \mathbb{Z}^{n-2}$ be the linear transformation

$$v \mapsto v\tilde{a}.$$

In a similar fashion we may lift the matrix z or $z + \boldsymbol{e}_{k-1,n-2}$ to a linear transformation $Z : \mathbb{Z}^{n-2} \to \mathbb{Z}^{n-2}$.

In both cases a simple change of the standard generators shows that $(A+Z)(\mathbb{Z}^{n-2}) = A(\mathbb{Z}^{n-2})$ is a direct summand. One also has that $AZ = Z^2 = 0$ and so

$$(A+Z)^k = (A+Z)A^{k-1}.$$

It also follows that $(A+Z)^k(\mathbb{Z}^{n-2})$ is a direct summand and by Theorem 2.10 has a Jordan normal form over $\mathbb{Z}$. □

Proof of Theorem 1.3. Assume that $a = \alpha I + \bar{a}$, where $\bar{a} = \bar{a}(\lambda)$ as in (22). By (12),

$$K_n(a) = \sum_{k=1}^{n} \sum_{x \in X_k} \psi(ax + x^{-1}).$$

It is enough to prove that each of the sums $\sum_{x \in X_k} \psi(ax + x^{-1})$ is expressible as a polynomial in q, $q-1$ and $K = K_1(\alpha)$.This was already established when $\bar{a} = 0$, so we may assume that $\varepsilon_{n-1} \ne 0$.

Proposition 2.5 and (i) of Theorem 2.9 take care of the cells X_n and X_{n-1}. For the other cells we may refer to statements (ii) and (iii) in Theorem 2.9 which reduce the sum down to the case when $k = N_i$ for one of the $N_i = n_1 + \cdots + n_i$, in which case Proposition 2.13 gives (28). Induction on the rank and Proposition 2.14 then shows that the resulting Kloosterman sums of rank $n-1$ and $n-2$ can be expressed as polynomials in q, $q-1$ and K independently of $p = \operatorname{char} \mathbb{F}_q$. □

It is possible to make this recursion more concrete; see Section 5B for examples.

3. Estimates

3A. ***Kloosterman sums over Bruhat cells: reduction to involutions.*** We will move on to setup the technical background for the proof of Theorem 1.6. We will pursue a path which establishes both of the estimates in Theorem 1.6 as well as Theorem 1.5 simultaneously and which will also allow us to analyze these sums cohomologically.

This approach is based on the Bruhat decomposition of the algebraic group GL_n as well as its specialization in the finite group $G = \mathrm{GL}_n(\mathbb{F}_q)$ with respect to the Borel subgroup B_n of upper triangular matrices, with

$$B_n = T_n U_n,$$

where U_n is the algebraic subgroup of upper triangular unipotent matrices, and T_n is the maximal torus (diagonal matrices). Since in this section n is fixed we will simply write B, T and U. Let W be the Weyl group of GL_n. We then have

$$G = \mathrm{GL}_n(\mathbb{F}_q) = \bigsqcup_{w\in W} C_w(\mathbb{F}_q), \quad C_w(\mathbb{F}_q) = U^\flat(\mathbb{F}_q)wB(\mathbb{F}_q), \tag{30}$$

with $U^\flat = \{u \in U \mid w^{-1}uw \in U^T\}$, where U^T is the unipotent subgroup of the opposite Borel subgroup of lower triangular matrices. It is clear from Gaussian elimination that for a field F all $x \in C_w(F)$ have a unique decomposition $x = uwb$ with $u \in U^\flat(F)$ and $b \in B(F)$. This remains true for the algebraic variety GL_n canonically [Springer 1998, Chapter 8]. (Alternatively, given any field F one may work over an algebraic closure of F, say $\bar{F}$, as in [Borel 1991, IV.14.12, Theorem (a)]. It is clear from the uniqueness that once representatives for W as permutation matrices are fixed, the decomposition above is invariant under any Galois automorphism of $\bar{F}$ fixing F. The same argument shows that the map $U^\flat \times B \to C_w$ is an isomorphism of algebraic varieties, defined over F.)

We will work with

$$K_n^{(w)}(a, \mathbb{F}_q) = \sum_{x\in C_w(\mathbb{F}_q)} \psi(ax + x^{-1}). \tag{31}$$

As a first step we will prove that the above sum vanishes unless $w^2 = I$ and analyze the cells C_w to simplify these sums for $w^2 = I$.

The Weyl group, $W = W_n$, is isomorphic to S_n, the permutation group on n letters. For later calculations we make this identification explicit as follows. For a permutation matrix w we associate the permutation π such that $i = \pi(j)$ if $w_{i,j} = 1$. Conversely given π, we let

$$w_\pi = \sum_{j=1}^{n} \boldsymbol{e}_{\pi(j),j} \tag{32}$$

so that $w_{\pi_1} w_{\pi_2} = w_{\pi_1\pi_2}$. To ease reading the arguments that will follow, we overload the notation and write $w(i)$ for $\pi(i)$ if $w = w_\pi$. This convention leads to

$$(gw)_{ij} = g_{i,w(j)} \quad \text{and} \quad (wg)_{ij} = g_{w^{-1}(i),j} \tag{33}$$

for any $g \in G$, which will be frequently used without further mention.

Proposition 3.1. *Let $a \in M_n(\mathbb{F}_q)$ be an upper triangular matrix, such that* $\det a \neq 0$, *and assume $w \in W$ is such that $w^2 \neq I$. Then*

$$K_n^{(w)}(a, \mathbb{F}_q) = 0.$$

Proof. Let i be minimal such that $i \neq w^2(i)$ and let $j = w(i)$. If $k = w(j)$ then $w^2(k) \neq k$ and so $k > i$.

Consider now the one parameter subgroup

$$\{x_{i,j}(s) = I + s\boldsymbol{e}_{i,j} : s \in \mathbb{F}_q\} \subset B.$$

Clearly we have $x_{ij}(s)B(\mathbb{F}_q) = B(\mathbb{F}_q)$ and so

$$K_n^{(w)}(a, \mathbb{F}_q) = \sum_{\substack{u \in U^\flat(\mathbb{F}_q)\\ b \in B(\mathbb{F}_q)}} \psi(auwb + (uwb)^{-1}) = \frac{1}{q}\sum_{s \in \mathbb{F}_q}\sum_{\substack{u \in U^\flat(\mathbb{F}_q)\\ b \in B(\mathbb{F}_q)}} \psi(auwx_{i,j}(s)b + (uwx_{i,j}(s)b)^{-1}).$$

Note that

$$\frac{\partial}{\partial s}\operatorname{Tr}(auwx_{i,j}(s)b) = \operatorname{Tr}(bauw\boldsymbol{e}_{i,j}) = (bauw)_{j,i} = (bau)_{j,j}$$

since $j = w(i)$. By the assumption $\det a \neq 0$ we have $(bau)_{j,j} \neq 0$.

On the other hand

$$\frac{\partial}{\partial s}\operatorname{Tr}(b^{-1}x_{i,j}(-s)w^{-1}u^{-1}) = -\operatorname{Tr}(\boldsymbol{e}_{i,j}w^{-1}u^{-1}b^{-1}) = -(w^{-1}u^{-1}b^{-1})_{j,i} = -(u^{-1}b^{-1})_{w(j),i} = 0$$

since $w(j) > i$. Since $\psi(b^{-1}x_{i,j}(-s)w^{-1}u^{-1})$ is linear in s, it is equal to $\psi(b^{-1}w^{-1}u^{-1})$ for any s.

Writing $\psi(auwx_{i,j}(s)b) = \psi(auwb)\psi(auwe_{i,j}bs)$ and using that ψ is invariant under conjugation we have $\psi(auwe_{i,j}bs) = \psi(bauwe_{i,j}s) = \psi((bau)_{j,j}s)$. This gives

$$\sum_{s \in \mathbb{F}_q} \psi(auwx_{i,j}(s)b + b^{-1}x_{i,j}(-s)w^{-1}u^{-1}) = \psi(auwb + b^{-1}w^{-1}u^{-1})\sum_{s \in \mathbb{F}_q}\psi((bau)_{j,j}s) = 0. \qquad \square$$

We can also prove Theorems 1.1 and 1.2 this way. For example, for nilpotent matrices we may prove that $K_n^{(w)}(a) = 0$, unless $w = I$. First we may assume that a is strictly upper triangular.

Let $j = \max(k \mid w(k) \neq k)$ and $i = w(j) < j$ and consider the one parameter subgroup

$$\{x_{i,j}(s) = I + s\boldsymbol{e}_{i,j} \mid s \in \mathbb{F}_q\} \leq B(\mathbb{F}_q).$$

We have $x_{i,j}(s)B(\mathbb{F}_q) = B(\mathbb{F}_q)$ and so

$$K_n^{(w)}(a) = \sum_{\substack{u \in U_w(\mathbb{F}_q)\\ b \in B(\mathbb{F}_q)}} \psi(auwb + (uwb)^{-1}) = \frac{1}{q}\sum_{s \in \mathbb{F}_q}\sum_{\substack{u \in U_w(\mathbb{F}_q)\\ b \in B(\mathbb{F}_q)}} \psi(auwx_{i,j}(s)b + (uwx_{i,j}(s)b)^{-1}).$$

By the definition of j for any $u \in U_w(\mathbb{F}_q)$ and $b \in B(\mathbb{F}_q)$

$$\frac{\partial}{\partial s}\operatorname{tr}(auwx_{i,j}(s)b) = \operatorname{tr}(\boldsymbol{e}_{i,j}bauw) = 0.$$

On the other hand

$$\frac{\partial}{\partial s}\,\mathrm{tr}((uwx_{i,j}(s)b)^{-1}) = \mathrm{tr}(\boldsymbol{e}_{i,j}w^{-1}u^{-1}b^{-1}) = b_{i,i}^{-1} \neq 0.$$

This gives

$$\sum_{s\in\mathbb{F}_q}\psi(auwx_{i,j}(s)b + b^{-1}x_{i,j}(-s)w^{-1}u^{-1}) = \psi(auwb+b^{-1}w^{-1}u^{-1})\sum_{s\in\mathbb{F}_q}\psi(b_{i,i}^{-1}s) = 0.$$

So

$$K_n(a,\mathbb{F}_q)=K_n^{(I)}(a,\mathbb{F}_q)=\sum_{b\in B(\mathbb{F}_q)}\psi(ab+b^{-1})=\sum_{b\in B(\mathbb{F}_q)}\psi(b^{-1})=|U|\sum_{t\in T(\mathbb{F}_q)}\prod_{i=1}^{n}\varphi(t_{i,i}^{-1})=(-1)^n q^{n(n-1)/2}$$

as $\psi(b^{-1})=\psi(t^{-1})=\prod_{i=1}^{n}\varphi(t_{i,i}^{-1})$ if $b=tu\in TU=B$.

3B. ***Finer decomposition of individual Bruhat cells.*** We will give a decomposition of the algebraic group U of unipotent upper triangular matrices in GL_n. To show that the underlying maps are morphisms we will work over a general commutative ring R. Therefore the letter U will denote the algebraic group itself, and not its set of points $U(\mathbb{F}_q)$. Similarly further subsets of U denoted with various markings will define affine subvarieties of U and not their set of points in $\mathbb{F}_q$.

The motivation for this refinement of the Bruhat decomposition is as follows. The Bruhat cells $C_w(\mathbb{F}_q) = U^\flat(\mathbb{F}_q)wB(\mathbb{F}_q)$ are already of the form $\mathbb{F}_q^d \times (\mathbb{F}_q^*)^e$ when using the entries of the matrices in $U^\flat(\mathbb{F}_q)$, $T(\mathbb{F}_q)$ and $U(\mathbb{F}_q)$ as coordinates, and exponential sums over such spaces have a well-established theory [Adolphson and Sperber 1989; Denef and Loeser 1991]. Also, $\mathrm{tr}(ax+x^{-1})$ as a function on $\mathbb{F}_q^d$ (using the entries as coordinates of x) is degree 1 in each of these variables. However they have to be collected the right way to use this observation. Therefore we will fiber up the cells C_w into finer (affine) subspaces that are more suitable for this purpose. For later arguments that rely on cohomology it will be important that this refinement gives an isomorphism of affine varieties. While for this purpose one could restrict to $\mathbb{F}_q$-algebras, the results are true over an arbitrary commutative ring. Also this refinement has potential applications elsewhere and so we first set it up for a general element $w \in W$ which is assumed to be fixed.

This section mainly consists in developing a nomenclature for these simple but numerous subsets (many of them subgroups). These subsets are defined under various actions of W on the affine variety $\overline{U}$ of strictly upper triangular nilpotent matrices, whose set of points in any ring is

$$\overline{U}(R) = \{y \in M_n(R) \mid I + y \in U(R)\} \tag{34}$$

Clearly $\overline{U}$ is isomorphic to $\mathbb{A}^{n(n-1)/2}$ as an algebraic variety.

The subsets we are going to define depend on w but we will not work with more than one w at a time, so we drop w from the notation. For example, the subgroups

$$U^\flat = \{u \in U \mid w^{-1}uw \in U^T\} \quad \text{and} \quad U^\sharp = \{u \in U \mid w^{-1}uw \in U\}$$

that satisfy $U^\sharp \cap U^\flat = \{I\}$ and $U = U^\sharp U^\flat = U^\flat U^\sharp$ can be defined as

$$U^\flat = I + \overline{U}^\flat \quad \text{and} \quad U^\sharp = I + \overline{U}^\sharp$$

where

$$\overline{U}^\flat = \overline{U} \cap w\overline{U}^T w^{-1} \quad \text{and} \quad \overline{U}^\sharp = \overline{U} \cap w\overline{U}w^{-1}.$$

The further refinement comes from exploiting the action of W, the group of permutation matrices on M_n via left multiplication and so it is tied to the standard representation of GL_n.

Definition 3.2. For a fixed $w \in W$ let

$$\overline{U}_a = \overline{U} \cap w^{-1}\overline{U}, \quad \overline{U}_b = \overline{U} \cap w^{-1}\overline{U}^T \quad \text{and} \quad \overline{U}_o = \overline{U} \cap w^{-1}D, \tag{35}$$

where D is the set of possibly singular diagonal matrices. Using these subvarieties define

$$U_a = I + \overline{U}_a, \quad U_b = I + \overline{U}_b \quad \text{and} \quad U_o = I + \overline{U}_o. \tag{36}$$

Remark 8. The subscripts "a", "b" and "o" denote the fact that these are the elements $\bar{u} \in \overline{U}$ for which the nonzero entries in $w\bar{u}$ are strictly "above", "below" or "on" the diagonal. Clearly, for any ring R, $\overline{U}(R) = \overline{U}_a(R) \oplus \overline{U}_b(R) \oplus \overline{U}_o(R)$.

We will see below that the map $x \mapsto \mathrm{tr}(ax + x^{-1})$ is linear on $\overline{U}_a$ which leads to immediate cancellations. However for this purpose we first need to setup further notation.

Definition 3.3. Let

$$U_a^\sharp = U^\sharp \cap U_a, \quad U_b^\sharp = U^\sharp \cap U_b, \quad U_a^\flat = U^\flat \cap U_a, \quad U_b^\flat = U^\flat \cap U_b.$$

Remark 9. These subvarieties are defined via their nonvanishing entries. In general let

$$\mathcal{I} = \{(i,j) \mid 1 \le i < j \le n\},$$

and for $\mathcal{J} \subset \mathcal{I}$ define the affine variety

$$\overline{U}_{\mathcal{J}} = \{\bar{u} \in \overline{U} \mid \bar{u}_{i,j} \ne 0 \implies (i,j) \in \mathcal{J}\}.$$

Note that $U_{\mathcal{J}} = I + \overline{U}_{\mathcal{J}}$ is an algebraic subgroup of U if and only if $\mathcal{J}$ is transitive in the sense that

$$(i,j), (j,k) \in \mathcal{J} \implies (i,k) \in \mathcal{J}.$$

For example, $U_a = U_{\mathcal{J}_a}$ where $\mathcal{J}_a = \{(i,j) \in \mathcal{I} \mid w(i) < j\}$, which is transitive, and so U_a is a subgroup.

Example 3.4. Consider the case $n = 5$. We indicate below the indices with the different properties for two involutions:

$$w = (14)(25) \text{ gives } \begin{bmatrix} b & b & o & a \\ & b & b & o \\ & & a & a \\ & & & a \\ & & & \end{bmatrix}, \quad \text{while } w = (15)(23) \text{ gives } \begin{bmatrix} b & b & b & o \\ & o & a & a \\ & & a & a \\ & & & a \\ & & & \end{bmatrix}.$$

To state the main result of this section we need to introduce one more piece of notation. Let

$$\overline{U}_{b/o} = \overline{U}_b \oplus \overline{U}_o$$

be the subset of $\overline{U}$ of elements $\bar{u}$ for which $w\bar{u}$ has nonzero elements only below or on the diagonal, and let

$$U^\sharp_{b/o} = I + \overline{U}^\sharp_{b/o} \quad \text{and} \quad U^\flat_{b/o} = I + \overline{U}^\flat_{b/o},$$

where $\overline{U}^\sharp_{b/o} = \overline{U}^\sharp \cap \overline{U}_{b/o}$, and $\overline{U}^\flat_{b/o} = \overline{U}^\flat \cap \overline{U}_{b/o}$.

Proposition 3.5. (i) *U_a, $U^\sharp U_a$ and $U^\flat_{b/o}$ are algebraic subgroups of U and $U^\sharp U_a \cap U^\flat_{b/o} = \{I\}$.*

(ii) *The morphism*

$$U^\sharp_{b/o} \times U_a \times U^\flat_{b/o} \to U, \quad (u_1, y, u_2) \mapsto u_1 y u_2,$$

is an isomorphism.

(iii) *If $w^2 = I$, then $U_o \subset U^\flat$ is a subgroup, and the morphism*

$$U^\sharp_b \times U_a \times U_o \times U^\flat_b \to U, \quad (u_1, y, u_o, u_2) \mapsto u_1 y u_o u_2,$$

is an isomorphism.

There are many alternative versions of the statement in (i) as, for example, $U_{a/o} = U_a U_o = U_o U_a$ is a subgroup of U as well. Also if one is merely interested in a bijection over $\mathbb{F}_q$, the statements in (ii) and (iii) can easily be proved by a counting argument. One may argue similarly to see that the set $U^\sharp_{b/o}$ is a complement of U_a in $U^\sharp U_a$. To prove that these maps are isomorphisms it is possible to adapt the reasoning of Lemma 8.2.2 in [Springer 1998] but given the concrete nature of the statement we give a self-contained proof here, based on the fact that affine varieties and their maps are determined by their functor of points.

Lemma 3.6. *We have the following*:

(i) *If $u \in U(R)$ and $x \in \overline{U}_a(R)$ then $xu \in \overline{U}_a(R)$.*

(ii) *If $u \in U^\sharp(R)$ and $x \in \overline{U}_a(R)$ then $ux \in \overline{U}_a(R)$.*

(iii) *If $u_1 \in U^\sharp(R)$ and $u_2 \in U^\flat(R)$ then*

$$u_1 u_2 + \overline{U}_a(R) = u_1 U_a(R) u_2.$$

Proof. Since $\overline{U}_a(R) = \overline{U}(R) \cap w^{-1}\overline{U}(R)$, the first and second claims are obvious from the fact that $U(R)\overline{U}(R) = \overline{U}(R)U(R) = \overline{U}(R)$, and that for $u \in U^\sharp(R)$, $w^{-1}uw \in U(R)$.

By the first two claims, if $u_1 \in U^\sharp(R)$ and $u_2 \in U^\flat(R)$ then $u_1 \overline{U}_a(R) u_2 = \overline{U}_a(R)$, from which the third claim is obvious. □

Proof of Proposition 3.5. It is enough to prove that for any commutative ring R the sets $U_a(R)$, $U^\sharp(R)U_a(R)$ and $U^\flat_{b/o}(R)$ are subgroups of $U(R)$, and that the maps

$$U^\sharp_{b/o}(R) \times U_a(R) \times U^\flat_{b/o}(R) \to U(R), \quad (u_1, y, u_2) \mapsto u_1 y u_2,$$

and

$$U_b^\sharp(R) \times U_a(R) \times U_o(R) \times U_b^\flat(R) \to U(R), \quad (u_1, y, u_o, u_2) \mapsto u_1 y u_o u_2,$$

are bijections.

Define on $U(R)$ the equivalence relation

$$u_1 \sim_{\overline{U}_a} u_2 \iff u_1 - u_2 \in \overline{U}_a(R). \tag{37}$$

We start with the proof of the claim about the group property of the three sets in (i). By (i) of the previous lemma, if $u_1 \sim_{\overline{U}_a} u_2$, and $u \in U(R)$, then $u_1 u \simeq_{\overline{U}_a} u_2 u$. Therefore $U_a(R)$, the stabilizer of the equivalence class of the identity I, is a subgroup.

By (ii) of the same lemma $U^\sharp(R)$ normalizes $U_a(R)$ and so $U^\sharp(R)U_a(R)$ is a subgroup. Finally,

$$\mathcal{J} = \{(i, j) \mid j \le w(i),\ w(j) \le w(i)\} \quad \text{and} \quad \mathcal{J}' = \mathcal{I} \setminus \mathcal{J} = \{(i, j) \in \mathcal{I} \mid w(i) < j \text{ or } w(i) < w(j)\},$$

are disjoint transitive subsets. This shows that $U_{b/o}^\flat(R) = U_{\mathcal{J}}(R)$ is a subgroup, and since $U^\sharp(R), U_a(R) \subset U_{\mathcal{J}'}$ we have that $U^\sharp(R)U_a(R) \cap U_{b/o}^\flat(R) = \{I\}$.

Now to prove the claim that every $u \in U(R)$ can be represented in a unique way as

$$u = u_1 y u_2, \quad u_1 \in U_{b/o}^\sharp(R), \ \ y \in U_a(R), \ \ u_2 \in U_{b/o}^\flat(R),$$

we will first show that $U(R) = U_{b/o}^\sharp(R)U_a(R)U_{b/o}^\flat(R)$. Let $u \in U(R)$. We know [Springer 1998, Chapter 8, Proposition 8.2.1] that there are $v_1 \in U^\sharp(R)$ and $v_2 \in U^\flat(R)$ such that $u = v_1 v_2$.

Let $u_1 \in U_{b/o}^\sharp(R)$ be the matrix whose entries agree with v_1 for $(i, j) \in \mathcal{I}$ when $w(i) \ge j$ and are 0 otherwise. Then we have that

$$v_1 \in u_1 + \overline{U}_a(R) \quad \text{and so} \quad v_1 U_a(R) = u_1 U_a(R).$$

Similarly let $u_2 \in U_b^\flat(R)$ be such that $v_2 \in u_2 + \overline{U}_a(R)$, so that $U_a(R)u_2 = U_a(R)v_2$. We have that

$$u = v_1 v_2 \in v_1 U_a(R) v_2 = u_1 U_a(R) u_2.$$

Now for the injectivity assume that $u_1, u_1' \in U_{b/o}^\sharp(R)$, $y, y' \in U_a(R)$, and $u_2, u_2' \in U_{b/o}^\flat(R)$ are such that

$$u_1 y u_2 = u_1' y' u_2'.$$

Since $U^\sharp(R)U_a(R) \cap U_{b/o}^\flat(R) = \{I\}$ we have that $u_2 = u_2'$ and so $u_1 y = u_1' y'$. By (iii) of the previous lemma $u_1 \sim_{\overline{U}_a} u_1'$ which can only happen in $U_{b/o}^\sharp(R)$ if $u_1 = u_1'$.

The very last claim about the case $w^2 = I$ is elementary. □

For convenience we will parameterize C_w using the following lemma.

Lemma 3.7. *Assume that $w^2 = I$. Then $U_o = U_o^\flat$ and the morphisms*

(i) $U^\flat \times T \times U \to C_w$, $(v, t, u) \mapsto vwtuv^{-1}$,

(ii) $U_b^\sharp \times \overline{U}_a \times U_o^\flat \times U_b^\flat \to U$, $(u_1, y, u_o, u_2) \mapsto u_1(I + y)^{-1} u_o u_2$,

are isomorphisms.

Proof. The first claim follows from (30) and the fact that the morphism $U^\flat \times T \times U \to U^\flat \times T \times U$, $(v, t, u) \mapsto (v, t, uv^{-1})$ is an isomorphism of varieties. The second claim is merely a restatement of (iii) of Proposition 3.5. □

The advantage of the parameterization in (i) is obvious: if $x = vwtuv^{-1} \in C_w(\mathbb{F}_q)$ then

$$\operatorname{tr}(ax + x^{-1}) = \operatorname{tr}(a^v wtu + (wtu)^{-1}),$$

where $a^v = v^{-1}av = \alpha I + \bar{a}^v$, with $\bar{a}^v = v^{-1}\bar{a}v$ still strictly upper triangular.

3C. *Some trace calculations.* From now on we will specify to the finite field $\mathbb{F}_q$ and so B, U, $\overline{U}$ etc. will mean $B(\mathbb{F}_q)$, $U(\mathbb{F}_q)$, $\overline{U}(\mathbb{F}_q)$ etc. Assume again that a is a matrix with a unique eigenvalue $\alpha \in \mathbb{F}_q^*$:

$$a = \alpha I + \bar{a} \in M_n(\mathbb{F}_q), \quad \text{where } \bar{a} \text{ is strictly upper triangular.} \tag{38}$$

With (ii) of Lemma 3.7 it is now easy to bound $\sum_{x \in C_w} \psi(ax + x^{-1})$. However the cohomological methods used in the proofs of Theorems 1.7 and 1.8 require repeated use of Lemma 3.16, which in turn relies on understanding

$$x \mapsto \operatorname{tr}(ax + x^{-1})$$

itself as a function on C_w. This is achieved in Propositions 3.8 and 3.9.

Proposition 3.8. *Assume that $w \in W$ satisfies $w^2 = I$ and let $\overline{U}_a$, $U_b^\sharp$, $U_b^\flat$, U_o be as in Definitions 3.2 and 3.3 with values in $\mathbb{F}_q$.*

Assume that $a^v = \alpha I + \bar{a}' \in M_n(\mathbb{F}_q)$ with $\alpha \in \mathbb{F}_q^*$, $\bar{a}' \in \overline{U}$, and that $u_1 \in U_b^\sharp$, $u_o \in U_o$ and $u_2 \in U_b^\flat$. If $y \in \overline{U}_a$, let $g(y) = u_1(I + y)^{-1}u_o u_2$. Then the function

$$y \mapsto \operatorname{tr}(a^v wtg(y) + (wtg(y))^{-1})$$

is affine linear on $\overline{U}_a$. This map is nonconstant unless $u_2 = I$.

When the map is constant its value is

$$\operatorname{tr}(a^v wtu_1u_o + (wtu_1u_o)^{-1}).$$

Further reductions are given by:

Proposition 3.9. *Assume $w \in W$ satisfies $w^2 = I$, and that $a' = \alpha I + \bar{a}' \in M_n(\mathbb{F}_q)$ with $\alpha \in \mathbb{F}_q^*$, $\bar{a}' \in \overline{U}$. Let $u_1 \in U_b^\sharp$ and $t = \operatorname{diag}(t_1, \dots, t_n) \in T$. For $u_o = I + \bar{u}_o \in U_o$ the map*

$$u_o \mapsto \operatorname{tr}(a'wtu_1(I + \bar{u}_o) + (wtu_1(I + \bar{u}_o))^{-1})$$

is affine linear on U_0. This map is nonconstant unless $t_{w(i)}^{-1} = \alpha t_i$ for all $i \neq w(i)$. When the map is constant its value is

$$\sum_{i = w(i)} (\alpha t_i + t_i^{-1}) + \operatorname{tr}(\bar{a}'wdu_1),$$

where $d = \sum_{i < w(i)} t_i \boldsymbol{e}_{i,i}$.

The proof of the above facts rely on some simple lemmas that we present first.

Lemma 3.10. *If a' is upper triangular, $y \in \overline{U}_a$, $u_1 \in U^\sharp$ and $u_2 \in U$, then*

$$\operatorname{tr}(a' w t u_1 (I+y)^{-1} u_2) = \operatorname{tr}(a' w t u_1 u_2) \tag{39}$$

is independent of y.

Proof. By Proposition 3.5 $(I+y)^{-1} = I + y'$ for some $y' \in \overline{U}_a$ and so wy' is strictly upper triangular. By the assumptions on a', u_1 and u_2 we have that a', wtu_1w^{-1} and u_2 are upper triangular, so

$$\operatorname{tr}(a' w t u_1 y' u_2) = \operatorname{tr}(a'(wtu_1w^{-1})(wy')u_2) = 0. \qquad \square$$

Lemma 3.11. *For any $u \in U$, $t \in T$*

$$y \mapsto wyu^{-1}t^{-1}$$

is a linear automorphism of $\overline{U}_a$.

Proof. The map $y \mapsto wyu^{-1}t^{-1}$ is clearly linear, and a bijection to its image. Since $w\overline{U}_a = \overline{U}_a$ and $\overline{U}_a ut = \overline{U}_a$ this image is $\overline{U}_a$. □

Proof of Proposition 3.8. First by Lemma 3.10 we have

$$\operatorname{tr}(a^v wtg(y) + (wtg(y))^{-1}) = \operatorname{tr}(a^v wtu_1u_ou_2) + \operatorname{tr}((wtu_1u_ou_2)^{-1}) + \operatorname{tr}((u_2u_o)^{-1}yu_1^{-1}t^{-1}w)$$

showing instantly that the map $y \mapsto \operatorname{tr}(a^v wtg(y) + (wtg(y))^{-1})$ is affine linear. Also

$$\operatorname{tr}((u_2u_o)^{-1}yu_1^{-1}t^{-1}w) = \operatorname{tr}(w(u_2u_o)^{-1}wy'),$$

where $y' = wyu_1^{-1}t^{-1} \in U$, and so by Lemma 3.11 it is enough to show that the map

$$y' \mapsto \operatorname{tr}(w(u_2u_o)^{-1}wy')$$

is nonconstant unless $u_2 = I$.

Assume that $u_2 \neq I$, and let $z = (u_2u_o)^{-1}$. Since $U_b^\flat$ and $U_o^\flat$ are subgroups, z is in $U_b^\flat U_o^\flat$, but not in $U_o^\flat$ and so there exist (i, j) such that

$$i < j, \quad w(i) > j \quad \text{and} \quad z_{i,j} \neq 0.$$

Also, by the fact that $U_b^\flat \subset U^\flat$, we have that $w(i) > w(j)$, and so

$$\boldsymbol{e}_{w(j),w(i)} \in \overline{U}_a.$$

Now let

$$Y_0 = \{y \in \overline{U}_a : y_{w(j),w(i)} = 0\}$$

so that any element $y \in \overline{U}_a$ may be written as $y = y_0 + s\boldsymbol{e}_{w(j),w(i)}$, where $y_0 \in Y_0$. Then

$$\operatorname{tr}(wzw(y_0 + s\boldsymbol{e}_{w(j),w(i)})) = \operatorname{tr}(wzwy_0) + s\ \operatorname{tr}(wzw\boldsymbol{e}_{w(j),w(i)}) = \operatorname{tr}(wzwy_0) + sz_{i,j}.$$

This proves the proposition. □

For the proof of Proposition 3.9 we need an explicit evaluation:

Lemma 3.12. *Assume $w \in W$ satisfies $w^2 = I$, $\alpha \in \mathbb{F}_q^*$, $\bar{a}' \in \bar{U}$ and that $t = \operatorname{diag}(t_1, \dots, t_n) \in T$ satisfies $t_{w(i)}^{-1} = \alpha t_i$ for all $i < w(i)$. Then*

$$\operatorname{tr}(w(\alpha t + t^{-1})) = \sum_{i = w(i)} \alpha t_i + t_i^{-1}.$$

If $u \in U$, then

$$\operatorname{tr}(\bar{a}' w t u) = \operatorname{tr}(\bar{a}' w d u),$$

where $d = \sum_{i < w(i)} t_i \boldsymbol{e}_{i,i}$.

Proof. First observe that

$$wt = \sum_{i = w(i)} t_i \boldsymbol{e}_{i,i} + \sum_{i < w(i)} (t_i \boldsymbol{e}_{w(i),i} + (\alpha t_i)^{-1} \boldsymbol{e}_{i,w(i)}),$$

from which the first claim follows immediately. The second is a slight variant, using that $\bar{a}'_{i,i} = 0$, from the assumption $\bar{a}' \in \bar{U}$. □

Proof of Proposition 3.9. It is easy to check that $U_o = U_o^{\flat} = \left\{ I + \sum_{i < w(i)} s_i \boldsymbol{e}_{i,w(i)} \mid s_i \in \mathbb{F}_q \right\}$ and it is abelian. Using that we have that $(I + \bar{u}_o)^{-1} = I - \bar{u}_o$. Therefore

$$\operatorname{tr}(a' w t u_1 (I + \bar{u}_o) + (w t u_1 (I + \bar{u}_o))^{-1}) = \operatorname{tr}(a' w t u_1 + (w t u_1)^{-1}) + \operatorname{tr}((a' w t u_1 - (w t u_1)^{-1}) \bar{u}_o)$$

is clearly affine linear as a function of $\bar{u}_o$. To see when it is nonconstant write $\bar{u}_o$ as $\sum_{i < w(i)} s_i \boldsymbol{e}_{i,w(i)}$ so that by the conventions in (33)

$$\bar{u}_o w = \sum_{i < w(i)} s_i \boldsymbol{e}_{i,i} \quad \text{and} \quad w \bar{u}_o = \sum_{i < w(i)} s_i \boldsymbol{e}_{w(i),w(i)}.$$

This leads to

$$\operatorname{tr}(a' w t u_1 \bar{u}_o) = \operatorname{tr}((a' w t u_1 w) w \bar{u}_o) = \sum_{i < w(i)} s_i (a' w t u_1 w)_{w(i),w(i)} = \sum_{i < w(i)} s_i \alpha t_i$$

since $w u_1 w \in U$, and $w \operatorname{diag}(t_i) w = \operatorname{diag}(t_{w(i)})$. In a similar manner

$$\operatorname{tr}((w t u_1)^{-1} \bar{u}_o) = \operatorname{tr}((t u_1)^{-1} w \bar{u}_o) = \sum_{i < w(i)} s_i ((t u_1)^{-1})_{w(i),w(i)} = \sum_{i < w(i)} s_i t_{w(i)}^{-1}.$$

Finally, since $u_1 \in U_b^{\sharp} \subset U^{\sharp}$, $u_1^{-1} \in U^{\sharp}$ as well, and for any $u \in U^{\sharp}$ we have

$$u_{i,w(i)} = 0 \quad \text{in the case } i < w(i).$$

Hence for $u = I + \bar{u} \in U^{\sharp}$, $\operatorname{tr}(w t \bar{u}) = 0$, and $\operatorname{tr}(t w \bar{u}) = 0$. This gives

$$\operatorname{tr}(a' w t u_1 + w u_1^{-1} t^{-1}) = \operatorname{tr}(\alpha w t + w t^{-1}) + \operatorname{tr}(\bar{a}' w t u_1).$$

This finishes the proof of the proposition in view of Lemma 3.12. □

3D. ***The proofs of Theorems 1.5 and 1.6.*** We again interpret the notation for all affine varieties as the set of their $\mathbb{F}_q$-rational points, so C_w stands for $C_w(\mathbb{F}_q)$, etc. Recall that

$$K_n^{(w)}(a) = \sum_{x \in C_w} \psi(ax + x^{-1}).$$

Proposition 3.13. *Assume that $a = \alpha I + \bar{a} \in M_n(\mathbb{F}_q)$ with $\alpha \in \mathbb{F}_q^*$, $\bar{a} \in \bar{U}$. Then*

$$K_n^{(w)}(a) = q^{n_a} \sum_{v} \sum_{t, u_o, u_1} \psi(a^v w t u_1 u_o + (w t u_1 u_o)^{-1}),$$

where $v \in U^\flat$, $a^v = v^{-1} a v$, $t \in T$, $u_o \in U_o$, $u_1 \in U_b^\sharp$ and

$$n_a = \dim U_a = |\{\mathcal{J}_a\}| = \bigl|\{(i, j) \mid i, w(i) < j\}\bigr|.$$

Proof. By Lemma 3.7(ii)

$$K_n^{(w)}(a) = \sum_{v} \sum_{t, u_o, u_1, u_2} \sum_{y} \psi(a' w t g(y) + (w t g(y))^{-1}),$$

where $g(y) = u_1 (I + y)^{-1} u_0 u_2$ and the inner sum is

$$\sum_{y} \psi(a' w t g(y) + (w t g(y))^{-1}) = \begin{cases} 0 & \text{if } u_2 \neq I, \\ q^{n_a} \psi(a' w t u_1 u_o + (w t u_1 u_o)^{-1}) & \text{if } u_2 = I, \end{cases}$$

by Proposition 3.8. □

Proposition 3.14. *We have*

$$K_n^{(w)}(a) = q^{n_a + n_o} K_1(\alpha)^f \sum_{v, d, u} \psi(\bar{a}^v w d u),$$

where $v \in U^\flat$, $u \in U_b^\sharp$, $\bar{a}^v = v^{-1} \bar{a} v$, $n_o = e = \bigl|\{i \mid i < w(i)\}\bigr|$ is the number of involution pairs in w, $f = \bigl|\{i \mid w(i) = i\}\bigr|$ is the number of fixed points of w and $d \in D(w) = \bigl\{\sum_{i < w(i)} t_i \boldsymbol{e}_{i,i} \mid t_i \in \mathbb{F}_q^*\bigr\}$.

Proof. Let

$$T(w) = \{t \in T \mid t_{w(i)}^{-1} = \alpha t_i \text{ if } i < w(i)\}.$$

By Proposition 3.9,

$$\sum_{u_o} \psi(a^v w t u_1 u_o + (w t u_1 u_o)^{-1}) = 0$$

unless $t \in T'(w)$, in which case

$$\sum_{u_o} \psi(a^v w t u_1 u_o + (w t u_1 u_o)^{-1}) = q^{n_o} \prod_{i = w(i)} \phi(\alpha t_i + t_i^{-1}) \psi(\bar{a}^v w d u_1),$$

with $d = \sum_{i < w(i)} t_i \boldsymbol{e}_{i,i}$.

The proposition follows after summing over $t \in T(w)$. □

Proof of Theorem 1.5. We will show that for $\alpha \neq 0$ and $w^2 = I$ we have

$$K_n^{(w)}(\alpha I) = q^{n(n-1)/2+N}(q-1)^e K_1(\alpha)^f, \tag{40}$$

where $N = n^\flat_{a/o} = \dim U^\flat_a + \dim U^\flat_o$.

Since $\bar{a} = 0$ now, $\bar{a}^v = v^{-1}\bar{a}v = 0$ as well, and so $\psi(\bar{a}^v w d u_1) = \psi(0) = 1$. To get the exponent of q note that $n_a + n_o + n^\flat + n^\sharp_b = n_a + n_o + n_b + n^\flat_{a/o}$, where these are denoting the dimension of the corresponding subspaces of U. Then we also have $n_a + n_o + n_b = n(n-1)/2$. □

Proof of Theorem 1.6. If $a = \alpha I + \bar{a}$, $\bar{a} \in \overline{U}$, $w^2 = I$, then by Proposition 3.14

$$|K_n^{(w)}(a)| = q^{n_a+n_o}|K_1(\alpha)^f|\left|\sum_{v,d,u}\psi(\bar{a}^v w d u)\right| \leq q^{n_a+n_o}|K_1(\alpha)^f|\sum_{v,d,u}|\psi(\bar{a}^v w d u)| = |K_n^{(w)}(\alpha I)|$$

since $|\psi(\bar{a}^v w d u_1)| = 1$.

Since $w^2 = I$, every element is either fixed by w or is in an involution pair, and so $n = f + 2n_o$. So while f depends on w,

$$K_1(\alpha)^f = K_1(\alpha)^n K_1(\alpha)^{-2n_o} = \operatorname{sign}(K_1(\alpha))^n |K_1(\alpha)|^n K_1(\alpha)^{-2n_o} = \varepsilon\, |K_1(\alpha)|^f,$$

with the sign

$$\varepsilon = (\operatorname{sign}(K_1(\alpha)))^n$$

independent of w. Here we have used the fact that $K_1(\alpha)$ is real. Thus we immediately have that

$$|K_n^{(w)}(\alpha I)| = \varepsilon K_n^{(w)}(\alpha I).$$

Therefore

$$|K_n(a)| \leq \sum_{w^2=I}|K_n^{(w)}(a)| \leq \sum_{w^2=I}|K_n^{(w)}(\alpha I)| = \varepsilon\sum_{w^2=I}K_n^{(w)}(\alpha I) = \varepsilon K_n(\alpha I) = |K_n(\alpha I)|.$$

It remains to prove that $c_n = |K_n(\alpha I)|/q^{(3n^2-\delta(n))/4} \leq 4$. By the recursion formula (4) we have

$$c_{n+1} \leq c_n |K_1(\alpha)| q^{-n/2} q^{(\delta(n+1)-\delta(n)-3)/4} + c_{n-1}.$$

Recall that

$$\delta(n) = \begin{cases} 0 & \text{if } n \text{ is even,} \\ 1 & \text{if } n \text{ is odd.} \end{cases}$$

Thus

$$\delta(n+1) - \delta(n) = \begin{cases} 1 & \text{if } n \text{ is even,} \\ -1 & \text{if } n \text{ is odd.} \end{cases}$$

If $n = 2k$ this gives $c_{2k+1} - c_{2k-1} \leq 2c_{2k}q^{-k}$, and so

$$c_{2k+1} \leq c_1 + 2\sum_{j=1}^{k} c_{2j}q^{-j}.$$

Similarly

$$c_{2k} \leq 2 \sum_{j=1}^{k-1} c_{2j+1} q^{-j}$$

from which the claim follows easily for $q \geq 3$. The case $q = 2$ is easily checked by hand. □

The sum $\sum_{v,d,u} \psi(\bar{a}^v w d u)$ gives rise to some intriguing questions on its own; see the problems mentioned Section 5C.

3E. ***Review of cohomology.*** With the results of the previous section, for a fixed $a \in M_n$ the bounds can be proven over those extensions of $\mathbb{F}_q$ in which a can be conjugated to Jordan normal form. However, to get the general result, we need to understand certain cohomology groups attached to the sum — which are independent of the field extension. In the rest of this section we will consider a subset of the matrix group $X \subset M_n$ defined by algebraic equations of the matrix entries as the corresponding algebraic variety.

We first introduce the notation and the main tools, then prove cohomological versions of Propositions 2.3 and 3.14. This enables us to prove Theorems 1.7 and 1.8.

Let $\ell \neq p$ be a prime, and $\overline{\mathbb{Q}}_\ell$ be an algebraic closure of the field $\mathbb{Q}_\ell$ of ℓ-adic numbers, such that there is a p-th primitive root of unity ζ contained in $\overline{\mathbb{Q}}_\ell$. Fix the field embedding $\iota_0 : \mathbb{Q}(\zeta) \to \mathbb{C}$ which sends ζ to $e(1/p)$ and let $\mathcal{L}_\varphi$ be the Artin–Schreier sheaf on $\mathbb{A}^1 = \mathbb{A}^1_{\mathbb{F}_q}$ corresponding to the additive character φ.

For a quasiprojective scheme $X/\mathbb{F}_q$ and a morphism $f : X \to \mathbb{A}^1$ the Grothendieck trace formula [1965] yields

$$\sum_{x \in X(\mathbb{F}_{q^m})} \varphi(f(x)) = \sum_{i=0}^{2\dim X} (-1)^i \operatorname{Tr}(\mathrm{Frob}_q^m, H_c^i(\overline{X}, f^*\mathcal{L}_\varphi)),$$

where $\overline{X} = X \otimes_{\mathbb{F}_q} \overline{\mathbb{F}}$, H_c^i is the ℓ-adic cohomology group with compact support in degree i. We use the notation $H_c^\bullet$ for the "complex" of cohomologies. These cohomology groups are finite-dimensional $\overline{\mathbb{Q}}_\ell$-vector spaces and $\mathrm{Frob}_q \in \mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F}_q)$ is the geometric Frobenius acting on them. By Deligne's work [1980] (see also [Kiehl and Weissauer 2001; Milne 1980; 2016]) we know that each Frobenius eigenvalue λ_k^i on H_c^i (for $1 \leq k \leq d_i = \dim H_c^i$) is a Weil number of weight j for some $j \leq i$, that is,

$$|\iota(\lambda_k^i)| = q^{j/2} \tag{41}$$

for all embeddings $\iota : \mathbb{Q}(\lambda) \to \mathbb{C}$, and thus

$$\sum_{x \in X(\mathbb{F}_{q^m})} \varphi(f(x)) = \sum_{i=0}^{2\dim X} (-1)^i \sum_{j=1}^{d_i} (\lambda_j^i)^m.$$

To simplify the notation we write $H_c^i(Y, f) = H_c^i(\overline{Y}, (f^*\mathcal{L}_\varphi)|_{\overline{Y}})$ for arbitrary subschemes $Y \leq X$. As a corollary of the above we have that if

$$H_c^i(Y, f) = 0 \quad \text{for } i > d$$

then

$$\left|\sum_{x\in X(\mathbb{F}_{q^m})} \varphi(f(x))\right| \le Cq^{md/2}, \tag{42}$$

with $C=\sum_{i=0}^{d}\dim H_c^i(Y,f)$. We will consider the cohomologies of the sums in the previous section: the sum $K_n(a)$ corresponds to the scheme $X=G=\mathrm{GL}_n$ and the morphism $f=g : x\mapsto \mathrm{tr}(ax+x^{-1})$ (and also the embedding of $\iota_0 : \mathbb{Q}(\zeta)\to\mathbb{C}$).

In the previous sections we derived bounds for the general Kloosterman sums over a finite extension where the eigenvalues of the coefficient matrix are defined. However bounds over an extension field do not imply that the weights are small. Consider, for example, $X=(\mathbb{A}^1\setminus\{1\})\sqcup\mathbb{A}^0$ and the regular function $f : X\to\mathbb{A}^1$ defined by

$$f(x)=\begin{cases} x & \text{if } x\in\mathbb{A}^1\setminus\{1\},\\ 0 & \text{if } x\in\mathbb{A}^0.\end{cases}$$

Then $\sum_{x\in X(\mathbb{F}_{q^m})}\varphi(f(x))=1-\zeta^m$ which vanishes if $p\mid m$ but only in that case.

The reason for this phenomenon is that the Frobenius eigenvalues on different cohomologies differ by a multiple of a root of unity, and thus cancel in some extensions: here $\dim H_c^1(X,f)=\dim H_c^0(X,f)=1$, $\dim H_c^2(X,f)=0$ and the Frobenius eigenvalues are $\lambda_1^1=\zeta$ and $\lambda_1^0=1$.

Thus, to get the general bound, we will prove that the cohomologies $H_c^i(G,g)$ vanish if i is large enough; hence the weights are not too large.

We will use the following properties of $H_c^\bullet$ (for an overview, see, e.g., [Katz 1980] especially Chapters 3.5. and 4.1-3 and [Laumon 2000; Fresán and Jossen 2020]):

Excision. If $f : X\to\mathbb{A}^1$ is a regular function, $Z\to X$ is a closed immersion and $U\to X$ is the complementary open immersion, then there exists a long exact sequence in the form

$$\cdots\to H_c^i(U,f)\to H_c^i(X,f)\to H_c^i(Z,f)\to H_c^{i+1}(U,f)\to\cdots.$$

Künneth formula. If $f_i : X_i\to\mathbb{A}^1$ for $i=1,2$, π_i is the canonical map $X=X_1\times_{\mathrm{Spec}(\mathbb{F}_q)}X_2\to X_i$, and $f_1+f_2:=\pi_1^*f_1+\pi_2^*f_2$, then

$$(f_1^*\mathcal{L}_\varphi)\boxtimes(f_2^*\mathcal{L}_\varphi)\simeq(f_1+f_2)^*\mathcal{L}_\varphi$$

and

$$H_c^\bullet(X,f_1+f_2)\simeq H_c^\bullet(X_1,f_1)\otimes H_c^\bullet(X_2,f_2),$$

that is, for all i

$$H_c^i(X,f_1+f_2)\simeq\bigoplus_{j+k=i}H_c^j(X_1,f_1)\otimes H_c^k(X_2,f_2).$$

We will also need some knowledge of the cohomologies in simple situations. They are listed in the following theorem.

Theorem 3.15 (cohomology of some basic sheaves). (i) *Cohomology of some basic sheaves on* $\mathbb{A}^1$.

(a) $H_c^i(\mathbb{A}^1, \mathrm{id}) = 0$ *for all* i.

(b) *If* $0 : \mathbb{A}^1 \to \mathbb{A}^1$ *is the zero map, then* $\mathcal{L}_0 = f_0^* \mathcal{L}_\varphi$ *is the constant sheaf and*

$$\dim H_c^i(\mathbb{A}^1, 0) = \begin{cases} 1 & \text{if } i = 2, \\ 0 & \text{if } i \neq 2. \end{cases}$$

The Frobenius eigenvalue on H_c^2 *is* q (*which is of weight* 2).

(ii) *Cohomology of some basic sheaves on* $\mathbb{A}^1 \setminus \mathbb{A}^0$.

(a) *We have*

$$\dim H_c^i(\mathbb{A}^1 \setminus \mathbb{A}^0, \mathrm{id}) = \begin{cases} 1 & \text{if } i = 1, \\ 0 & \text{if } i \neq 1. \end{cases}$$

The Frobenius eigenvalue on H_c^1 *is* 1 (*which is of weight* 0).

(b) *We have*

$$\dim H_c^i(\mathbb{A}^1 \setminus \mathbb{A}^0, 0) = \begin{cases} 1 & \text{if } i = 1, 2, \\ 0 & \text{if } i \neq 1, 2. \end{cases}$$

The Frobenius eigenvalue on H_c^2 *is* q (*which is of weight* 2) *and on* H_c^1 *is* 1 (*weight* 0).

(iii) *Cohomology of sheaves corresponding to Kloosterman sums*; *see* [Weil 1948a]. *If* $\alpha \in \mathbb{F}_q^*$ *and* $f_\alpha : \mathbb{G}_m = \mathbb{A}^1 \setminus \mathbb{A}^0 \to \mathbb{A}^1$ *is the morphism which corresponds to the map*

$$f_\alpha(t) = \alpha \cdot t + 1/t$$

then

$$\dim H_c^i(\mathbb{A}^1 \setminus \mathbb{A}^0, f_\alpha) = \begin{cases} 2 & \text{if } i = 1, \\ 0 & \text{if } i \neq 1, \end{cases}$$

and on H_c^1 *both weights are* 1.

The next observation is essential in what follows.

Lemma 3.16. *Let* $f, g : X \to \mathbb{A}^1$ *be regular functions*, $X_0 = f^{-1}(\{0\})$ *and consider* $f \cdot \mathrm{id}_{\mathbb{A}^1} + g : X \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathbb{A}^1 \to \mathbb{A}^1$. *Then*

$$H_c^\bullet(X \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathbb{A}^1, f \cdot \mathrm{id}_{\mathbb{A}^1} + g) \simeq H_c^\bullet(X_0, g) \otimes H_c^\bullet(\mathbb{A}^1, 0);$$

thus

$$H_c^{i+2}(X \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathbb{A}^1, f \cdot \mathrm{id}_{\mathbb{A}^1} + g) \simeq H_c^i(X_0, g) \otimes H_c^2(\mathbb{A}^1, 0) \quad \textit{for all } i. \tag{43}$$

Proof. Let $V = X \setminus X_0$ and consider the morphism $j = \mathrm{id}_V \otimes (\mathrm{id}_{\mathbb{A}^1} - g)/f : V \times \mathbb{A}^1 \to V \times \mathbb{A}^1$. This is clearly an isomorphism and $j \circ (f \cdot \mathrm{id}_{\mathbb{A}^1} + g) = 0_V + \mathrm{id}_{\mathbb{A}^1}$; thus by the Künneth formula, $H_c^\bullet(V \times \overline{\mathbb{A}^1}, j \circ (f \cdot \mathrm{id}_{\mathbb{A}^1} + g)) \equiv 0$.

Let $Z = X_0 \times \mathbb{A}^1$ and $U = X \times \mathbb{A}^1 \setminus Z$. From the previous argument, $H_c^\bullet(U, f \cdot \mathrm{id}_{\mathbb{A}^1} + g) \equiv 0$. By excision $H_c^i(X \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathbb{A}^1, f \cdot \mathrm{id}_{\mathbb{A}^1} + g) \simeq H_c^i(Z, f \cdot \mathrm{id}_{\mathbb{A}^1} + g)$, but $(f \cdot \mathrm{id}_{\mathbb{A}^1})|_Z \equiv 0$; hence $(f \cdot \mathrm{id}_{\mathbb{A}^1} + g)|_Z = g|_Z = g|_{X_0} + 0|_{\mathbb{A}^1}$ and applying the Künneth formula we get the lemma. □

Remark 10. This lemma is the cohomological form of the straightforward computation

$$\sum_{(x,t)\in X(\mathbb{F})\times F} \varphi(tf(x)+g(x)) = \sum_{x\in X(\mathbb{F})} \varphi(g(x)) \sum_{t\in\mathbb{F}} \varphi(tf(x)) = q \sum_{x\in X_0(\mathbb{F})} \varphi(g(x)).$$

A similar argument as in the proof appears in motivic context in [Fresán and Jossen 2020, Lemma 6.5.3 and Remark 6.5.4].

Applying the lemma repeatedly we get:

Corollary 3.17. *Let $\pi_j : \mathbb{A}^m \to \mathbb{A}^1$ be the projection*

$$(x_1, x_2, \dots, x_m) \mapsto x_j.$$

For $f_j, g : X \to \mathbb{A}^1$, $1 \le j \le m$, let $h : X \times_{\operatorname{Spec}(\mathbb{F}_q)} \mathbb{A}^m \to \mathbb{A}^1$ be defined by

$$h = \sum_{j=1}^{m} f_j \cdot \pi_j + g.$$

Consider $X_0 = \{x \in X \mid h(x, \cdot\,) \equiv 0\} \le X$ as a subscheme. Then

$$H_c^\bullet(X \times_{\operatorname{Spec}(\mathbb{F}_q)} \mathbb{A}^m, h) \simeq H_c^\bullet(X_0, g) \otimes \left(\bigotimes_{j=1}^{m} H_c^\bullet(\mathbb{A}^1, 0) \right);$$

thus

$$H_c^{i+2m}(X \times_{\operatorname{Spec}(\mathbb{F}_q)} \mathbb{A}^m, h) \simeq H_c^i(X_0, g) \otimes \left(\bigotimes_{j=1}^{m} H_c^2(\mathbb{A}^1, 0) \right) \quad \textit{for all } i.$$

We will now show the vanishing of cohomologies of high enough degree for the exponential sums that were used in the previous sections.

3F. *The proof of Theorem 1.7.*

We first start with the reduction to Jordan blocks as in Proposition 2.3.

Let $G = \mathrm{GL}_n$, $G_k = \mathrm{GL}_k$, $G_l = \mathrm{GL}_l$ for some $n = k + l$. Let

$$a = \left(\begin{array}{c|c} a_k & b \\ \hline 0 & a_l \end{array} \right) \in \mathbb{A}^{n\times n}$$

be block upper triangular with $a_k \in \mathbb{A}^{k\times k}$, $a_l \in \mathbb{A}^{l\times l}$. Let $g : G \to \mathbb{A}^1$, $x \mapsto \operatorname{tr}(ax + x^{-1})$, and for the diagonal blocks denote by g_k and g_l the morphisms $x_k \mapsto \operatorname{tr}_k(a_k x_k + x_k^{-1})$ and $x_l \mapsto \operatorname{tr}_l(a_l x_l + x_l^{-1})$, respectively. Let $H^\bullet = H_c^\bullet(G, g)$ and similarly $H_k^\bullet = H_c^\bullet(G_k, g_k)$, $H_l^\bullet = H_c^\bullet(G_l, g_l)$.

Proposition 3.18. *If a_k and a_l have no common eigenvalues, then*

$$H^\bullet \simeq H^\bullet(\mathbb{A}^{k\times l}, 0) \otimes H_k^\bullet \otimes H_l^\bullet, \quad \textit{that is,} \quad H^i \simeq H^{2kl}(\mathbb{A}^{k\times l}, 0) \otimes \left(\bigoplus_{j+j'=i-2kl} H_k^j \otimes H_l^{j'} \right),$$

where $H^\bullet(\mathbb{A}^{k\times l}, 0) = \left(\bigotimes_{i=1}^{kl} H_c^\bullet(\mathbb{A}^1, 0)\right)$

Proof. The morphism $U_{[k,l]} \to \mathbb{A}^{kl}$, $u \mapsto (u_{i,j})_{1\le i\le k, 1\le j\le l}$, is an isomorphism, so we apply Corollary 3.17 with $m = kl$ and $X = G$ on $X \times_{\mathrm{Spec}(\mathbb{F}_q)} U_{[k,l]}$ and

$$h : (x, u) \mapsto \mathrm{tr}(a(u^{-1}xu) + (u^{-1}xu)^{-1}).$$

From the proof of Proposition 2.3 it is clear that $h(x, \cdot)$ is cohomologically nontrivial if and only if $x' = 0$, that is, $x \in X_0$ with $X_0 \le X$ the subscheme of "block upper triangular" matrices and by Corollary 3.17

$$H_c^\bullet(X \otimes_{\mathrm{Spec}(\mathbb{F}_q)} U_{[k,l]}, h) \simeq H_c^\bullet(X_0, g) \otimes H^\bullet(\mathbb{A}^{k\times l}, 0) \simeq H_c^\bullet(G_k, g_k) \otimes H_c^\bullet(G_l, g_l) \otimes H^\bullet(\mathbb{A}^{k\times l}, 0),$$

where the second isomorphism is a consequence of a being a block matrix. Thus

$$X_0 \simeq G_k \times_{\mathrm{Spec}(\mathbb{F}_q)} G_l \times_{\mathrm{Spec}(\mathbb{F}_q)} \mathbb{A}^{k\times l}, \quad g|_{X_0} = g_k + g_l + 0_{\mathbb{A}^{k\times l}}$$

and the Künneth formula.

On the other hand $j : X \to X$, $(u, x) \mapsto (u, uxu^{-1})$, is an isomorphism and $j^*h = 0_{U_k} + g$ so

$$H_c^\bullet(X \otimes_{\mathrm{Spec}(\mathbb{F}_q)} U_{[k,l]}, h) \simeq H_c^\bullet(G, g) \otimes \left(\bigotimes_{i=1}^{kl} H_c^\bullet(\mathbb{A}^1, 0)\right);$$

hence the proposition. □

Proof of Theorem 1.7. Since cohomology does not depend on the finite field in question we may assume that a is a diagonal matrix with nonzero and unequal entries α_i on the diagonal. As above let $g : G \to \mathbb{A}^1$, $x \mapsto \mathrm{tr}(ax+x^{-1})$. Also let $H^\bullet_{K(\alpha_i)} = H_c^\bullet(\mathbb{A}^1 \backslash \mathbb{A}^0, f_\alpha)$. Repeated applications of the proposition above gives

$$H_c^\bullet(G, g) = H^{n(n-1)}(\mathbb{A}^{n(n-1)/2}, 0) \otimes \bigotimes_{i=1}^{n} H^\bullet_{K(\alpha_i)}.$$

By Theorem 3.15(i) and the Künneth formula, $\dim H^{n(n-1)}(\mathbb{A}^{n(n-1)/2}, 0) = 1$, with Frobenius eigenvalue $q^{n(n-1)}$. Similarly by Theorem 3.15(iii) $\bigotimes_{i=1}^n H^\bullet_{K(\alpha_i)}$ is concentrated in degree n where it equals $\bigotimes_{i=1}^n H^1_{K(\alpha_i)}$.

The claim now follows from the purity of the classical Kloosterman sums $K_1(\alpha_i)$. □

3G. ***Bounding the weights in the nonsplit case.***

For the proof of Theorem 1.8 we need to bound the degrees of the nontrivial cohomology groups. Recall that for $w^2 = I$ we defined in Theorem 1.5

$$N = N(w) = \bigl|\{(i, j) \mid 1 \le i < j \le n,\ w(j) < w(i) \le j\}\bigr|.$$

Theorem 3.19. *Assume that $a = \alpha I + \bar{a}$, with $\alpha \in \mathbb{F}_q^*$ and $\bar{a}$ nilpotent, and consider the cohomology $H_c^\bullet(C_w, x \mapsto \mathrm{tr}(ax + x^{-1}))$ associated to the exponential sum* (31).

(i) *If $w^2 \ne I$, then $H_c^\bullet(C_w, x \mapsto \mathrm{tr}(ax + x^{-1})) \equiv 0$.*

(ii) *If $w^2 = I$, then $H_c^i(C_w, x \mapsto \mathrm{tr}(ax + x^{-1})) = 0$ for $i > n^2 + 2N(w)$; thus all weights of the sum $\sum_{x\in C_w} \psi(ax + x^{-1})$ are at most $n^2 + 2N(w)$.*

Proof. (i) We may assume that $\bar{a}$ is upper triangular. The case $w^2 \neq I$ follows from the proof of Proposition 3.1. Let the pair (i, j) be chosen as in there: i is minimal such that $i \neq w^2(i)$ and $j = w(i)$. Consider the subvarieties $Y = \{I + s\boldsymbol{e}_{i,j} \mid s \in \mathbb{F}_q\} \simeq \mathbb{A}^1$, $X_1 = U_b$ and $X_2 = \{x \in B \mid x_{ij} = 0\}$. Then we have the decomposition $C_w = X_1 \times Y \times X_2$ by mapping $x_1 \in X_1, x_2 \in X_2, s \in \mathbb{F}_q$ to $g = x_1 w(I + s\boldsymbol{e}_{i,j})x_2$ (this is indeed an isomorphism of algebraic varieties). If $X = X_1 \times X_2$ then the proof of Proposition 3.1 shows that the map $g \operatorname{tr} a + (uwb)^{-1}$ is of the form $f(x)s + g(x)$ with $X_0 = \{x \in X \mid f(x) = 0\}$ empty. Hence Lemma 3.16 implies the vanishing of cohomology.

Assume now that $w^2 = I$. By Lemma 3.7, $C_w \simeq U^\flat \times T \times U_b^\sharp \times \overline{U}_a \times U_o \times U_b^\flat$ and this is an isomorphism of algebraic varieties.

We can apply Corollary 3.17 in the setting of Proposition 3.8. We have $\mathbb{A}^m = \overline{U}_a$ with $m = n_a$, $h : x \mapsto \operatorname{tr}(ax + x^{-1})$ and

$$X_0 = U_b \times T \times U_b^\sharp \times U_o \hookrightarrow X = U^\flat \times T \times U_b^\sharp \times U_o \times U_b^\flat,$$

where the embedding maps the last coordinate to I. We obtain

$$H_c^\bullet(C_w, x \mapsto \operatorname{tr}(ax + x^{-1})) \simeq H_c^\bullet(U^\flat \times T \times U_b^\sharp \times U_o, g_1) \otimes \left(\bigotimes_{i=1}^{n_a} H_c^\bullet(\mathbb{A}^1, 0) \right),$$

where $g_1 = \operatorname{tr}(a^v wtu_1u_o + (wtu_1u_o)^{-1})$ with the notation of Section 3C.

Applying Corollary 3.17 in the setting of Proposition 3.9 with $\mathbb{A}^m \simeq U_o$, $m = n_o$, $h : x \mapsto \operatorname{tr}(ax + x^{-1})$ and

$$X_0 = U^\flat \times T(w) \times U_b^\sharp \hookrightarrow X = U^\flat \times T \times U_b^\sharp,$$

where $T(w)$ is as in the proof of Proposition 3.14, we obtain

$$H_c^\bullet(C_w, x \mapsto \operatorname{tr}(ax + x^{-1})) \simeq H_c^\bullet(U^\flat \times T(w) \times U_b^\sharp, g_2) \otimes \left(\bigotimes_{i=1}^{n_a+n_o} H_c^\bullet(\mathbb{A}^1, 0) \right),$$

where $g_2 = \sum_{i=w(i)} (\alpha t_i + t_i^{-1}) + \operatorname{tr}(\bar{a}^v wdu_1)$.

The Künneth formula yields

$$H_c^\bullet(C_w, x \mapsto \operatorname{tr}(ax + x^{-1}))$$
$$\simeq H_c^\bullet(U^\flat \times U_b^\sharp, g_3) \otimes \left(\bigotimes_{i=1}^{n_a+n_o} H_c^\bullet(\mathbb{A}^1, 0) \right) \otimes \left(\bigotimes_{j=1}^{e} H_c^\bullet(\mathbb{A}^1 \setminus \mathbb{A}^0, 0) \right) \otimes \left(\bigotimes_{k=1}^{f} H_c^\bullet(\mathbb{A}^1 \setminus \mathbb{A}^0, f_\alpha) \right),$$

with $g_3 = \operatorname{tr}(\bar{a}^v wdu_1)$ and $f_\alpha : x \mapsto \alpha x + x^{-1}$.

Now it is enough to observe that the maximal nontrivial cohomology group is of degree

$$2 \dim U^\flat \times U_b^\sharp + 2(n_a + n_o) + 2e + f,$$

where e is the number of involution pairs, and f is the number fixed elements in w. It is clear that $n = 2e + f$ and the calculation in the proof of (40) shows that the rest is equal to $n(n-1) + 2N(w)$. □

To show the vanishing of cohomologies of the Bruhat cells in higher degrees, we need one more combinatorial lemma. For this the Weyl group W is identified with the symmetric group S_n viewed as the group of permutations of the set $\{1, \dots, n\}$ as in (32). In the notation of Remark 9 we let for any $w \in W$

$$\mathcal{J}^{\flat}_{a/o}(w) = \{(i, j) \mid 1 \leq i < j \leq n,\ w(j) < w(i) \leq j\}$$

Recall that $N(w) = |\mathcal{J}^{\flat}_{a/o}(w)| = \dim \overline{U}^{\flat}_{a/o}$.

Lemma 3.20. *Let e be a positive integer such that $e \leq \lfloor n/2 \rfloor$ and $w_e \in S_n$ be the involution for which*

$$w_e(j) = \begin{cases} n - j + 1 & \text{if } j = 1, \dots, e, \\ j & \text{if } j = e + 1, \dots, n - e. \end{cases}$$

We then have the following:

(i) *If $w^2 = I$ and w is a product of e disjoint transpositions then $N(w) \leq N(w_e)$ with equality only if $w = w_e$.*

(ii) $N(w_e) = e(n - e)$. *In particular $N(w)$ is maximal for the long element $w_{\lfloor n/2 \rfloor}$, and we have $N(w_{\lfloor n/2 \rfloor}) = (n^2 - \delta(n))/4$.*

Proof. We proceed by induction on e. Let $k = w(n)$ and first assume that $k > 1$. Let $v = (12 \dots k) \in S_n$, that is,

$$v(j) = \begin{cases} j + 1 & \text{if } j < k, \\ 1 & \text{if } j = k, \\ j & \text{if } j > k, \end{cases}$$

and $w' = v^{-1} w v \in S_n$. We claim that $(i, j) \in \mathcal{J}^{\flat}_{a/o}(w) \Rightarrow (v(i), v(j)) \in \mathcal{J}^{\flat}_{a/o}(w')$; thus $N(w) \leq N(w')$.

To see this, first assume that $\{i, j\} \cap \{k, n\} = \varnothing$. In this case the claim is clear since v respects the ordering of $X = \{1, 2, \dots, n\} \setminus \{k, n\}$ and $w(X) \subseteq X$.

Now the case $(i, k) \in \mathcal{J}^{\flat}_{a/o}(w)$ does not arise since $w(k) = n$. There is a single j such $(k, j) \in \mathcal{J}^{\flat}_{a/o}(w)$, namely $j = n$, but then $(v(k), v(n)) = (1, n) \in \mathcal{J}^{\flat}_{a/o}(w')$. Finally $(i, n) \in \mathcal{J}^{\flat}_{a/o}(w')$ for all i; thus for all $(i, n) \in \mathcal{J}^{\flat}_{a/o}(w)$ we have $(v(i), v(n)) \in \mathcal{J}^{\flat}_{a/o}(w')$.

If $w(n) \neq 1$, then the last case of the above argument shows $N(w)$ is strictly smaller than $N(w')$.

We now move to the induction step. Let w' be as above if $w(n) \neq 1$ and $w' = w$ otherwise. Let also $w'' \in S_{n-2}$ be the element which arises from the permutation w' restricted to $\{2, \dots, n - 1\}$ which we identify with $\{1, \dots, n - 2\}$ using $j \mapsto j - 1$. Then w'' is a product of $e - 1$ transpositions and the induction hypothesis shows that $N(w'')$ is maximal if and only if w'' arises from $w' = w_e$.

To prove the second part note that if $w(n) = 1$, then $(i, n) \in \mathcal{J}^{\flat}_{a/o}(w)$ for all $i < n$. Again let now $w'' \in S_{n-2}$ be the element which we obtain by deleting the first and last rows and columns of w'. Then $N(w'') = N(w) + n - 1$. This time induction again shows that $N(w_e) = (n-1) + (n-3) + \cdots + (n-2e+1)$ from which the statements follow. □

The following lemma enables us to work on the individual groups GL_{n_j}:

Lemma 3.21. *Let $G = \mathrm{GL}_m$ for some m and $g : \mathrm{GL}_m(\mathbb{F}_q) \to \mathbb{A}^1$, $x \mapsto \mathrm{tr}(ax + x^{-1})$, where $a = \alpha I + \bar{a}$, with $\bar{a}$ nilpotent. Then $H^i(\mathrm{GL}_m, g)$ vanishes if $i > m^2 + (m^2 - \delta(m))/2$.*

Proof. We may assume that $\bar{a}$ is upper triangular. For an involution $w^2 = I$ again let $e = e(w)$ be the number of involution pairs in w. By Theorem 3.19 and Lemma 3.20

$$H_c^i(C_w, g) = \begin{cases} 0 & \text{for any } i, \text{ if } w^2 \neq I, \\ 0 & \text{for } i > m^2 + 2e(m-e), \text{ if } w \neq I,\, w^2 = I, \\ 0 & \text{for } i \neq m^2, \text{ if } w = I. \end{cases} \tag{44}$$

Let l be the standard length function on $W = S_m$ [Borel 1991, 21.21] and consider

$$Y_l = \bigsqcup_{l(w)=l} C_w.$$

We also let $Y_0 = B$ corresponding to the unit element of W.

If $l(w) = l$ then C_w is an open subscheme of

$$X_l = \bigsqcup_{l(w)\leq l} C_w = Y_l \sqcup X_{l-1}.$$

Clearly for any w we have $H_c^i(C_w, g) = 0$ if $i > m^2 + (m^2 - \delta(m))/2$ and so this remain true for Y_l:

$$H_c^i(Y_l, g) = 0 \quad \text{if } i > m^2 + (m^2 - \delta(m))/2.$$

We will now apply induction in the excision long exact sequence on the disjoint union $X_l = Y_l \sqcup X_{l-1}$:

$$\cdots \to H_c^i(X_{l-1}, f) \to H_c^i(X_l, f) \to H_c^i(Y_l, f) \to H_c^{i+1}(X_{l-1}, f) \to \cdots .$$

For $l = 0$ we have that $X_0 = Y_0$ and so

$$H_c^i(X_0, g) = 0$$

already for $i > m^2$. Assume now that $H_c^i(X_{l-1}, g) = 0$ if $i > m^2 + (m^2 - \delta(m))/2$. From the excision long exact sequence we then also have that $H_c^i(X_l, g) = 0$ for $i > m^2 + (m^2 - \delta(m))/2$. □

Proof of Theorem 1.8. Fix $a \in M_n$. As the weights do not change after base change, we might work over a sufficiently large finite extension of $\mathbb{F}_q$, say $\mathbb{F}_{q^{m(a)}}$, over which a is conjugate to a block diagonal matrix, where the blocks on the diagonal are square matrices $a_j \in M_{n_j}$ in Jordan normal form with a unique eigenvalue α_j.

Then by Proposition 3.18 we have

$$H_c^\bullet(G, g) \simeq \left(\bigotimes_{j=1}^{r} H_c^\bullet(\mathrm{GL}_{n_j}, g_j) \right) \otimes \left(\bigotimes_{1\leq i<j\leq r} \bigotimes_{k=1}^{n_i n_j} H_c^\bullet(\mathbb{A}^1, 0) \right),$$

where $g_j : \mathrm{GL}_{n_j}(\mathbb{F}_q) \to \mathbb{A}^1$, $x \mapsto \mathrm{tr}(a_j x + x^{-1})$.

We apply the lemma with $m = n_i$ and $g = g_i$ for $1 \leq i \leq r$, respectively. By (42) and Lemma 3.21 we have that $K_n(a) \ll q^d$, with

$$d = \sum_{i=1}^{r} (3n_i^2 - \delta(n_i))/4 + \sum_{1 \leq i < j \leq r} n_i n_j.$$

To conclude the proof note that

$$\max\biggl(\sum_{i=1}^{r} (3n_i^2 - \delta(n_i))/4 + \sum_{1 \leq i < j \leq r} n_i n_j \biggm| r, n_i \in \mathbb{N} : \sum_{i=1}^{r} n_i = n \biggr) = (3n^2 - \delta(n))/4. \qquad \square$$

4. Degenerate cases

4A. ***Preliminary observations on $K_n(a, b)$.*** The results in this section are combinatorial in nature and will not require cohomology. Therefore from now on we work solely over $\mathbb{F}_q$ and $M_n = M_n(\mathbb{F}_q)$.

Let a and b singular $n \times n$ matrices such that

$$r = \operatorname{rk}(b) \geq s = \operatorname{rk}(a). \tag{45}$$

This section contains some elementary observations about the generalized Kloosterman sums

$$K_n(a, b) = \sum_{x \in \mathrm{GL}_n(\mathbb{F}_q)} \psi(ax + bx^{-1}).$$

First, we clearly have

$$K_n(a, b) = K_n(c_1 a c_2^{-1}, c_2 b c_1^{-1}) \quad \text{for any } c_1, c_2 \in \mathrm{GL}_n(\mathbb{F}_q). \tag{46}$$

In this, and the following sections, when we write a block matrix $a = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right) \in \mathbb{F}_q^{n \times n}$, we always mean the blocks to correspond to the partition $\{1, 2, \dots, n\} = \{1, \dots, r\} \sqcup \{r + 1, \dots, n\}$, r as in (45). For example, by (46) we may assume that $b = E_r$, where

$$E_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \tag{47}$$

is a standard idempotent, but an exact description of the equivalence classes is a delicate question. However all we need is a reasonable set of representatives for the action $(a, b) \mapsto (c_1 a c_2^{-1}, c_2 b c_1^{-1})$ that are suitable for handling the Kloosterman sums. This is most conveniently achieved via a parabolic Bruhat decomposition of $G = \mathrm{GL}_n(\mathbb{F}_q)$ with respect to the subgroups $P_r = P_r(\mathbb{F}_q)$ consisting of elements that when acting on row vectors map the subspace $V_r = \langle \boldsymbol{e}_{r+1}, \dots, \boldsymbol{e}_n \rangle$ to itself. In block matrix notation we have

$$P_r = P_r(\mathbb{F}_q) = \left\{ g \in \mathrm{GL}_n(\mathbb{F}_q) \;\middle|\; g = \begin{pmatrix} g_{11} & g_{12} \\ 0 & g_{22} \end{pmatrix} \right\}.$$

Let $Q_r = P_r^T = \bigl\{ g \in \mathrm{GL}_n(\mathbb{F}_q) \mid g = \left(\begin{smallmatrix} g_{11} & 0 \\ g_{21} & g_{22} \end{smallmatrix}\right) \bigr\}$ be the stabilizer of the columns space $\langle \boldsymbol{e}_1^T, \dots, \boldsymbol{e}_r^T \rangle$, the subspace of linear functionals vanishing on V_r. Then we have the following Bruhat decomposition for the group $G = \mathrm{GL}_n(\mathbb{F}_q)$.

Proposition 4.1. *Let $G = \mathrm{GL}_n(\mathbb{F}_q)$ with P_r, Q_r as above. Then*

$$G = \bigcup_{w \in W_P \backslash W / W_P} Q_r w P_r,$$

where $W_P = W \cap P_r = W \cap Q_r$ and $W_P \backslash W / W_P$ denotes the set of double cosets.

Proof. From $G = \bigcup_{w \in W} BwB$ it is clear that $G = \bigcup_{w \in W} P_{n-r} w P_r$. Let $w_l = \sum_{i=1}^{n} \boldsymbol{e}_{i,n-i}$ be the matrix that corresponds to the longest element $(1,n)(2,n-1)\cdots \in W = S_n$. Then $w_l P_{n-r} w_l = Q_r$, from which $G = w_l G = \bigcup_{w \in W} Q_r w P_r$. It is obvious that if $w' = w_1 w w_2$ with $w_1, w_2 \in W_P$, then $Q_r w' P_r = Q_r w P_r$. □

The following lemma is well known; see, for example, Section 1.3 in [James and Kerber 1981].

Lemma 4.2. *Let $W = S_{1,2,\dots,n}$, $W_P = S_{1,2,\dots,r} \times S_{r+1,r+2,\dots,n}$ and $m = \min(r, n-r)$. A set of double coset representatives of $W_P \setminus W / W_P$ is*

$$\overline{W}_r = \{w_k \mid k = 1, \dots, m\},$$

where for $k \le m$, w_k is the permutation matrix

$$w_k = \sum_{i=1}^{k} \boldsymbol{e}_{i,i+r} + \sum_{i=k+1}^{r} \boldsymbol{e}_{i,i} + \sum_{i=r+k+1}^{n} \boldsymbol{e}_{i,i}, \tag{48}$$

which we also identify with $w_k = (1, r+1)(2, r+2)\cdots(k, r+k) \in W = S_n$.

Recall that we have $E_r = \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

Proposition 4.3. *Let a and b as in* (45). *Then there exist matrices d and $w \in \overline{W}_r$ such that*

$$K_n(a,b) = K_n(E_r \cdot d, E_r \cdot w).$$

Proof. First we can write $a = c_1 E_s d_1$ and $b = d_2 E_r c_2$ for some $c_i, d_j \in \mathrm{GL}_n$, and thus $K_n(a,b) = K_n(E_r d_0, E_r c_0)$, where $d_0 = E_s d_1 d_2$, $c_0 = c_2 c_1$. Here we have used that $E_r E_s = E_s$, since $r \ge s$.

Now we have that $c_0 = qwp$, where $q \in Q_r$, $p \in P_r$ are as in Proposition 4.1, and $w \in \overline{W}_r$ as in Lemma 4.2. Let

$$U_r = \left\{ g \in \mathrm{GL}_n(\mathbb{F}_q) \;\middle|\; g = \begin{pmatrix} I_r & g_{12} \\ 0 & I_{n-r} \end{pmatrix} \right\} \tag{49}$$

be the unipotent radical of P_r, and

$$L_r = P_r \cap Q_r = \left\{ g \in \mathrm{GL}_n(\mathbb{F}_q) \;\middle|\; g = \begin{pmatrix} g_{11} & 0 \\ 0 & g_{22} \end{pmatrix} \right\},$$

so that $P_r = L_r U_r$ and $Q_r = L_r U_r^T$.

We have $L_r = H_1 H_2$, where

$$H_1 = \left\{ g \in \mathrm{GL}_n(\mathbb{F}_q) \;\middle|\; g = \begin{pmatrix} g_{11} & 0 \\ 0 & I_{n-r} \end{pmatrix} \right\}, \quad H_2 = \left\{ g \in \mathrm{GL}_n(\mathbb{F}_q) \;\middle|\; g = \begin{pmatrix} I_r & 0 \\ 0 & g_{22} \end{pmatrix} \right\}. \tag{50}$$

For $g \in L_r$,

$$E_r g = g E_r,$$

and for $u \in U_r$ we have

$$u E_r = E_r \quad \text{and} \quad E_r u^T = E_r.$$

Therefore writing $q = g_1 u_1^T$, $p = g_2 u_2$ we have

$$K_n(E_r d_0, E_r g_1 u_1^T w g_2 u_2) = K_n(g_2 u_2 E_r d_0, g_1 E_r u_1^T w) = K_n(E_r d, w),$$

where $d = g_2 d_0 g_1$. □

4B. ***The proof of Theorem 1.9 and a preliminary bound.*** We will give a proof of Theorem 1.9 on $K_n(a, 0)$ with a of rank r, namely

$$K(a, 0) = K_n(E_r, 0) = (-1)^r q^{-\binom{r+1}{2}} q^{rn} |\mathrm{GL}_{n-r}(\mathbb{F}_q)|. \tag{51}$$

While this evaluation is trivial, for singular a, b it is the basis of our bounds for the general Kloosterman sums given in Proposition 4.4 below.

Proof of Theorem 1.9. We have that

$$K_n(E_r, 0) = \sum_{x \in \mathrm{GL}_n(\mathbb{F}_q)} \psi(E_r x) = \frac{1}{q^{r(n-r)}} \sum_{u \in U_r} \sum_{x \in \mathrm{GL}_n(\mathbb{F}_q)} \psi(E_r u x),$$

where $U_r = \left\{u \in \mathrm{GL}_n(\mathbb{F}_q) \mid u = \left(\begin{smallmatrix} I_r & u_{12} \\ 0 & I_{n-r} \end{smallmatrix}\right)\right\}$ as in (49). Let $x = \left(\begin{smallmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{smallmatrix}\right)$. It is clear that when summing over u first,

$$\sum_{u \in U_r} \psi(E_r u x) = \begin{cases} 0 & \text{if } x_{21} \neq 0, \\ q^{r(n-r)} & \text{if } x_{21} = 0. \end{cases}$$

Therefore

$$K_n(E_r, 0) = q^{r(n-r)} |\mathrm{GL}_{n-r}(\mathbb{F}_q)| \, K_r(I_r, 0),$$

which leads immediately to the claim, in view of Theorem 1.2. □

Proposition 4.4. *Let $w = w_k \in \overline{W}_r$ be as in Lemma 4.2, and $d \in M_n$. Then*

$$|K_n(E_r d, E_r w)| \leq \sum_{j=0}^{k} q^{j(n-r-j)-\binom{j}{2}} \frac{|\mathrm{GL}_{n-r-j}(\mathbb{F}_q)|}{|\mathrm{GL}_{n-r}(\mathbb{F}_q)|} R_w(j),$$

where $R_w(j) = \left|\left\{x \in \mathrm{GL}_n(\mathbb{F}_q) \mid x E_r w = \left(\begin{smallmatrix} y_{11} & y_{12} \\ 0 & y_{22} \end{smallmatrix}\right), \ \mathrm{rk}(y_{22}) = j\right\}\right|$.

Proof. First swap the parameters

$$K_n(E_r d, E_r w) = K_n(E_r w, E_r d)$$

and then use the action of U_r as above to get

$$\begin{aligned}K_n(E_rw, E_rd) &= \sum_{x\in\mathrm{GL}_n(\mathbb{F}_q)} \psi(E_rwx + E_rdx^{-1})\\ &= \frac{1}{q^{r(n-r)}}\sum_{u\in U_r}\sum_{x\in\mathrm{GL}_n(\mathbb{F}_q)} \psi(E_rwux + E_rdx^{-1}u^{-1})\\ &= \frac{1}{q^{r(n-r)}}\sum_{x\in\mathrm{GL}_n(\mathbb{F}_q)}\sum_{u\in U_r} \psi(E_rwux + E_rdx^{-1}).\end{aligned}$$

This shows that

$$K_n(E_rw, E_rd) = \sum_{x\in\mathcal{R}_w} \psi(E_rwx + E_rdx^{-1}),$$

where $\mathcal{R}_w = \bigl\{x \in \mathrm{GL}_n(\mathbb{F}_q) \mid xE_rw = \bigl(\begin{smallmatrix} y_{11} & y_{12}\\ 0 & y_{22}\end{smallmatrix}\bigr) \text{ for some } y_{11}, y_{12}, y_{22}\bigr\}$.

Let

$$\mathcal{R}_w(j) = \left\{x \in \mathrm{GL}_n(\mathbb{F}_q) \;\middle|\; xE_rw = \begin{pmatrix} y_{11} & y_{12}\\ 0 & y_{22}\end{pmatrix},\ \mathrm{rk}(y_{22}) = j\right\} \tag{52}$$

so that $\mathcal{R}_w = \bigsqcup_j \mathcal{R}_w(j)$. Since in xE_rw the last $n-r-k$ columns are 0, $\mathcal{R}_w(j)$ is empty if $j > k$ and so

$$K_n(E_rw, E_rd) = \sum_{j=1}^{k}\sum_{x\in\mathcal{R}_w(j)} \psi(E_rwx + E_rdx^{-1}).$$

Clearly if $x \in \mathcal{R}_w(j)$ then $gx \in \mathcal{R}_w(j)$ for any $g \in P_r$. Therefore let

$$H_2 = \left\{g \in L_r \;\middle|\; g = \begin{pmatrix} I_r & 0\\ 0 & h\end{pmatrix},\ h \in \mathrm{GL}_{n-r}(\mathbb{F}_q)\right\}$$

as in (50) and note that for $g \in H_2$, $x \in \mathcal{R}_w(j)$

$$\mathrm{tr}(E_rwgx) = \mathrm{tr}^{(r)}(y_{11}) + \mathrm{tr}^{(n-r)}(y_{22}h) \quad\text{and}\quad \mathrm{tr}(E_rd(gx)^{-1}) = \mathrm{tr}(E_rdx^{-1}),$$

where $\mathrm{tr}^{(j)}$ is the $j\times j$ matrix trace and y_{11}, y_{22} are as in (52). This immediately implies that for $x \in \mathcal{R}_w(j)$

$$\sum_{g\in H_2} \psi(E_rwgx + E_rd(gx)^{-1}) = K_{n-r}(E_j, 0)\varphi(\mathrm{tr}^{(r)}(y_{11}) + \mathrm{tr}(E_rdx^{-1})),$$

and this gives

$$\begin{aligned}\sum_{x\in\mathcal{R}_w(j)} \psi(E_rwx + E_rdx^{-1}) &= \frac{1}{|\mathrm{GL}_{n-r}(\mathbb{F}_q)|}\sum_{g\in H_2}\sum_{x\in\mathcal{R}_w(j)} \psi(E_rwxg + E_rd(xg)^{-1})\\ &= \sum_{x\in\mathcal{R}_w(j)} \varphi(\mathrm{tr}^{(r)}(y_{11}) + \mathrm{tr}(E_rdx^{-1}))\frac{K_{n-r}(E_j, 0)}{|\mathrm{GL}_{n-r}(\mathbb{F}_q)|}.\end{aligned}$$

The proposition follows from trivially estimating the last sum using $|\varphi(\,\cdot\,)| \le 1$ and the evaluation (51). □

4C. ***The proof of Theorem 1.10.*** We restate the theorem and its corollary. We need to prove that if a and b are singular $n \times n$ matrices such that $s = \mathrm{rk}(a) \leq r = \mathrm{rk}(b) < n$ and $m = \min(r, n-r)$, then

(i) $K_n(a,b) \leq 2q^{n^2-rn+r^2+\binom{m}{2}}$,

(ii) if a, b are not both 0 then $K_n(a,b) \leq 2q^{n^2-n+1}$, and

(iii) this bound is sharp, since

$$K_n(\boldsymbol{e}_{1,n}, \boldsymbol{e}_{1,n}) = q^{2n-2}|\mathrm{GL}_{n-2}(\mathbb{F}_q)| + (q-1)q^{n-1}|\mathrm{GL}_{n-1}(\mathbb{F}_q)| \sim q^{n^2-n+1}.$$

Here (ii) is an obvious corollary of (i), and we start with the proof of that claim (Theorem 1.10). By Proposition 4.4 this will require some estimates for the number $R_w(j) = |\mathcal{R}_w(j)|$.

Lemma 4.5. *Let $w = w_k \in \overline{W}_r$ and $\mathcal{R}_w(j)$ be as in* (52). *Then*

$$\mathcal{R}_w(j) = P_r w_j P_s,$$

where $P_r = L_rU_r$ as in (49), *and $P_s = H_1'H_2U_r$ with H_1, H_2 as in* (50) *and $H_1' = H_1 \cap w_kH_1w_k$.*

Proof. First, if $y \in \mathcal{R}_w(j)E_rw$ then $gy \in \mathcal{R}_w(j)E_rw$ for any $g \in P_r$. This immediately shows that $P_r\mathcal{R}_w(j) = \mathcal{R}_w(j)$.

On the other hand if $g = h_1h_2u$ with $h_1 \in H_1'$, $h_2 \in H_2$ and $u \in U_r$ then

$$xgE_rw = xh_1E_rw = xE_rh_1w = (xE_rw)wh_1w = yh_1'$$

for some $h_1' \in H_1$ and $y = \left(\begin{smallmatrix} y_{11} & y_{12} \\ 0 & y_{22} \end{smallmatrix}\right)$, $\mathrm{rk}(y_{22}) = j$ as in (52), which shows that $\mathcal{R}_w(j)P_s = \mathcal{R}_w(j)$ as well.

The fact that there is a unique orbit represented by w_j is a direct calculation based on the definition of $\mathcal{R}_w(j)$ which implies that x_{21} has rank j, and the last $r-k$ columns of x_{21} are identically 0. □

Lemma 4.6. *In the notation above*

$$R_w(j) = |\mathcal{R}_w(j)| = c_{n-r,k}(j)c_{r,r-j}(r-j)q^{r(j+n-r)}|\mathrm{GL}_{n-r}(\mathbb{F}_q)|,$$

where $c_{k,l}(j) = \left|\{x \in \mathbb{F}_q^{k \times l} \mid \mathrm{rk}(x) = j\}\right|$.

Proof. Assume that $x = \left(\begin{smallmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{smallmatrix}\right) \in \mathcal{R}_w(j)$, and that $x' = \left(\begin{smallmatrix} x_{11} & x_{12}' \\ x_{21} & x_{22}' \end{smallmatrix}\right) \in \mathrm{GL}_n(\mathbb{F}_q)$. Then $x' \in \mathcal{R}_w(j)$ as well, and there is $u \in U_r$, $h \in H_2$, such that $x' = xuh$.

It follows that $R_w(j) = q^{r(n-r)}|\mathrm{GL}_n(\mathbb{F}_q)| \cdot |\mathcal{R}_w'(j)|$, where $\mathcal{R}_w'(j)$ consists of those $n \times r$ matrices $\left\{\left(\begin{smallmatrix} x_{11} \\ x_{21} \end{smallmatrix}\right)\right\}$, which have rank r, and for which the $(n-r) \times r$ matrix x_{21} is such that it has rank j and its last $r-k$ columns are identically 0.

The number of choices for x_{21} for $x = \left(\begin{smallmatrix} x_{11} \\ x_{21} \end{smallmatrix}\right) \in \mathcal{R}_w'(j)$ is $c_{n-r,k}(j)$.

From the transitivity $\mathcal{R}_w(j) = P_r\mathcal{R}_w(j)$ in Lemma 4.5 for each x_{21} there are the same number of possible x_{11}. For x_{21} the matrix with a I_j in the top left corner and zeros everywhere else, it is readily seen that the number of possible x_{11}'s is $q^{rj}c_{r,r-j}(r-j)$ which proves the claim. □

In view of the description in Lemma 4.5 it may seem that Proposition 4.4 is wasteful and a more exact evaluation is possible. While it is true that one can push this approach to get more precise information, we will see below that there are cases when the estimates are of the right order of magnitude. Still we will use some enumerative combinatorics, but merely for getting a good constant to match the q-power in the estimate that arises from Proposition 4.4. To do this it is convenient to use the Gaussian binomial coefficients (q binomials) [Cameron 2017]. For $k \in \mathbb{N}$ let

$$[k]_q = \frac{q^k-1}{q-1}, \quad [k]_q! = \prod_{j=1}^{k}[j]_q, \quad \text{and} \quad \binom{k}{l}_q = \frac{[k]_q!}{[l]_q! \cdot [k-l]_q!}.$$

With this notation we have $|\mathrm{GL}_k(\mathbb{F}_q)| = (q-1)^k q^{\binom{k}{2}}[k]_q!$ and the number of matrices of fixed size and rank ([Landsberg 1893, Formula $(\mathcal{B})$]; see also [Morrison 2006, Section 1.7])

$$c_{k,l}(j) = \bigl|\{x \in \mathbb{F}_q^{k\times l} \mid \mathrm{rk}(x) = j\}\bigr| = (q-1)^j q^{\binom{j}{2}} \frac{[k]_q! \cdot [l]_q!}{[k-j]_q! \cdot [l-j]_q! \cdot [j]_q!}. \tag{53}$$

We may therefore rephrase Lemma 4.6 as

$$R_w(j) = |\mathcal{R}_w(j)| = (q-1)^n q^{\binom{n}{2}+j^2} \frac{[k]_q! \cdot [r]_q! \cdot [n-r]_q!^2}{[k-j]_q! \cdot [n-r-j]_q! \cdot [j]_q!^2}. \tag{54}$$

(Here $w = w_k \in \overline{W}_r$ and $\mathcal{R}_w(j)$ is as in (52).)

Proof of Theorem 1.10. By Lemma 4.6, (53) and (54) we have that the summands in Proposition 4.4 are equal to

$$(q-1)^{n-j} q^{\binom{n}{2}+j^2} \frac{[k]_q! \cdot [r]_q! \cdot [n-r]_q!}{[k-j]_q! \cdot [j]_q!^2},$$

and thus

$$|K_n(E_r d, E_r w)| \le (q-1)^{n-r} q^{\binom{n}{2}}[n-r]_q! \sum_{j=0}^{k} q^{j^2}\binom{k}{j}_q \prod_{j'=j+1}^{r}(q^{j'}-1).$$

Using the trivial identity $q^{j'}-1 \le q^{j'}$ we have for the inner sum

$$\sum_{j=0}^{k} q^{j^2}\binom{k}{j}_q \prod_{j'=j+1}^{r}(q^{j'}-1) < \sum_{j=0}^{k} q^{j^2}\binom{k}{j}_q q^{\binom{r+1}{2}-\binom{j+1}{2}} = q^{\binom{r+1}{2}} \sum_{j=0}^{k} q^{\binom{j}{2}}\binom{k}{j}_q = 2q^{\binom{r+1}{2}} \prod_{j=1}^{k-1}(1+q^j)$$

by the q-binomial theorem [Stanley 1986, Formula (1.87)]. From this,

$$\begin{aligned}|K_n(E_r d, E_r w)| &\le 2q^{\binom{n}{2}+\binom{r+1}{2}} \prod_{j=1}^{k-1}(q^{2j}-1) \prod_{j'=k}^{n-r}(q^{j'}-1) \\ &< 2q^{\binom{n}{2}+\binom{r+1}{2}+2\binom{k}{2}+k(n-r-k+1)+\binom{n-r-k+1}{2}} = 2q^{n^2-rn+r^2+\binom{k}{2}}.\end{aligned}$$

Recall that $m = \min(r, n-r)$, and thus by Lemma 4.2 we have $1 \le k \le m$, so

$$|K_n(E_r d, E_r w)| \le 2q^{n^2-rn+r^2+\binom{m}{2}}. \qquad \square$$

Finally we prove the claim about $K_n(\boldsymbol{e}_{1,n}, \boldsymbol{e}_{1,n})$. Using the Bruhat decomposition with the maximal parabolic subgroup P as in (10) we get

$$K_n(\boldsymbol{e}_{1,n}, \boldsymbol{e}_{1,n}) = \sum_{k=1}^{n} \sum_{u \in U_k} \sum_{p \in P} \varphi((uw_{(kn)}p)_{n,1} + (uw_{(kn)}p)^{-1}_{n,1}).$$

Write $p = \left(\begin{smallmatrix} h & v \\ 0 & \lambda \end{smallmatrix}\right)$ with $h \in \mathrm{GL}_{n-1}(\mathbb{F}_q)$, $v \in \mathbb{F}_q^{(n-1)\times 1}$ and $\lambda \in \mathbb{F}_q^*$. Then

$$(uw_{(kn)}p)_{n,1} = \begin{cases} h_{k1} & \text{if } k < n, \\ 0 & \text{if } k = n, \end{cases} \qquad (uw_{(kn)}p)^{-1}_{n,1} = \begin{cases} \lambda^{-1} & \text{if } k = 1, \\ 0 & \text{if } k > 1, \end{cases}$$

and thus

$$\sum_{g \in U_k w_{(kn)} P} \varphi((x)_{n,1} + (x)^{-1}_{n,1}) = \begin{cases} -q^{n-1}|U_1|\,K_{n-1}(\boldsymbol{e}_{1,1}, 0) & \text{if } k = 1, \\ (q-1)q^{n-1}|U_k|\,K_{n-1}(\boldsymbol{e}_{1,k}, 0) & \text{if } 1 < k < n, \\ |P| & \text{if } k = n. \end{cases}$$

Since $K_{n-1}(\boldsymbol{e}_{1,k}, 0) = K_{n-1}(\boldsymbol{e}_{1,1}, 0) = -q^{n-2}|\mathrm{GL}_{n-2}(\mathbb{F}_q)|$ by Theorem 1.9, we get

$$\begin{aligned} K_n(\boldsymbol{e}_{1,n}, \boldsymbol{e}_{1,n}) &= -q^{2n-3}|\mathrm{GL}_{n-2}(\mathbb{F}_q)|\left(-q^{n-1} + (q-1)\sum_{k=2}^{n-1} q^{n-k}\right) + (q-1)q^{n-1}|\mathrm{GL}_{n-1}(\mathbb{F}_q)| \\ &= q^{2n-2}|\mathrm{GL}_{n-2}(\mathbb{F}_q)| + (q-1)q^{n-1}|\mathrm{GL}_{n-1}(\mathbb{F}_q)| \sim q^{n^2-n+1}. \end{aligned}$$

5. Examples

5A. ***Kloosterman sums of 2×2 matrices.*** Let $a \in M_2(\mathbb{F}_q)$. Since the Kloosterman sum is invariant under conjugation, $K_2(a) = K_2(gag^{-1})$ for any $g \in \mathrm{GL}_2(\mathbb{F}_q)$, we may assume that a is in Frobenius normal form, $a = \left(\begin{smallmatrix} 0 & 1 \\ -d & t \end{smallmatrix}\right)$, where $t = \mathrm{tr}(a)$, $d = \det a$. The Kloosterman sum is then

$$K_2(a) = \sum_{x_{11}x_{22} - x_{12}x_{21} \neq 0} \varphi(-dx_{12} + tx_{22})\varphi(x_{21} + (x_{11} + x_{22})/(x_{11}x_{22} - x_{12}x_{21})).$$

From this presentation it is not at all clear that this sum should behave differently depending on whether $t^2 - 4d$ is or is not a nonzero square or 0, showing that brute force calculations without using the finer group structure are unlikely to highlight any of the features of these sums.

For $n = 2$ the maximal parabolic of Section 2B is also the Borel subgroup, and the two approaches given earlier are the same. They lead in an elementary way to the following evaluations:

a	$\left(\begin{smallmatrix} \alpha & 0 \\ 0 & \beta \end{smallmatrix}\right)$, $\alpha \neq \beta$	$\left(\begin{smallmatrix} \alpha & 0 \\ 0 & \alpha \end{smallmatrix}\right)$, $\alpha \neq 0$	$\left(\begin{smallmatrix} \alpha & 1 \\ 0 & \alpha \end{smallmatrix}\right)$, $\alpha \neq 0$	$\left(\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right)$	$\left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$
$K_2(a)$	$qK_1(\alpha)K_1(\beta)$	$q^3 - q^2 + K_1(\alpha)^2 q$	$-q^2 + K_1(\alpha)^2 q$	q	q

The nonsplit case, $a = \left(\begin{smallmatrix} \alpha & \delta\beta \\ \beta & \alpha \end{smallmatrix}\right)$, with $\beta \neq 0$ and $\delta \notin (\mathbb{F}_q^*)^2$, can also be evaluated explicitly, although this requires some effort. We have

$$K_2(a) = -qK_1\big(\alpha + \beta\sqrt{\delta}, \mathbb{F}_{q^2}^*\big).$$

See the proof of Proposition 5.10 below.

It is also possible to deal directly with the more general sum $K_2(a,b)$. Of course if either $\mathrm{rk}(a)$ or $\mathrm{rk}(b)$ is 2, say $\mathrm{rk}(b)=2$, this leads to the previous evaluation by

$$K_2(a,b)=K_2(ab^{-1},I_2).$$

If one of them, but not both are 0, say $b=0$, and $\mathrm{rk}(a)=1$, then

$$K_2(a,0)=-q(q-1).$$

If both of a and b have rank 1, then we may assume that $a=e_1=\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$, and that b is one of $b_1=\left(\begin{smallmatrix}\alpha&0\\0&0\end{smallmatrix}\right)$ for which

$$K_2(e_1,b_1)=K_1(\alpha)q(q-1),$$

or $b_2=\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)$ for which

$$K_2(e_1,b_2)=-q(q-1),$$

or $b_3=\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$ for which

$$K_2(e_1,b_3)=q^3-q^2+q.$$

Finally one trivially has that $K_2(0,0)=(q^2-1)(q^2-q)$.

5B. ***The recursion in closed form.*** We will describe an algorithm for calculating the polynomials that express $K_n(a)$ when $a=\alpha I_n+\bar{a}(\lambda)$ for some fixed $\alpha\neq 0$ and some partition $\lambda=[n_1,\dots,n_l]$. As usual the fact that $n_1+\cdots+n_l=n$ is denoted by $\lambda\vdash n$. We will rely on the notation of Sections 2B and 2C, where we made the assumption that $n_i\le n_{i+1}$. If an element n_i repeats k times we will write $[\dots,n_i^k,\dots]$ instead of $[\dots,n_i,\dots,n_i,\dots]$, so, for example, we will write $[1^n]$ for the partition that corresponds to the matrix αI.

Also we will denote the polynomials by K_λ as well, instead of the notation P_λ in Theorem 1.3. However we will still write K for $K_{[1]}=K_{[1]}(\alpha)=K_1(\alpha)$, so, for example, Theorem 1.4 can be stated as

$$K_{[1^n]}=q^{n-1}KK_{[1^{n-1}]}+q^{2n-2}(q^{n-1}-1)K_{[1^{n-2}]}.$$

It is clear from the proof of Theorem 5.1 below that if $n_{l-1}<n_l$, the recursion is particularly simple, and has only two terms corresponding to the sums over the cells X_{n-1} and X_n. Therefore in the case $n_{l-1}<n_l$

$$K_\lambda=q^{n-1}KK_{\lambda'}-q^{2n-2}K_{\lambda''},$$

where

$$\lambda'=[n_1,\dots,n_{l-1},n_l-1]\quad\text{and}\quad\lambda''=[n_1,\dots,n_{l-1},n_l-2]$$

possibly reordered into a monotonic sequence, if $n_{l-1}=n_l-1$. If $n_l=2$, then the entries corresponding to $n_l-2=0$ are simply deleted. For example, for $\lambda=[1,2]$ this gives

$$K_{[1,2]}=q^2KK_{[1^2]}-q^4K.$$

The situation is more interesting when the last entry is repeated. This can be handled by the following explicit recursion, which gives an alternative proof of Theorem 1.3.

Theorem 5.1 (recursion algorithm). *Assume that* $\lambda = [n_1^{k_1}, \dots, n_{l-1}^{k_{l-1}}, n_l^{k_l}]$, *with* $k_l > 1$. *Then*

$$K_\lambda = q^{n-1} K K_{\lambda'} - q^{2n-2} K_{\lambda''} - (q^{k_l - 1} - 1) q^{2n-2} (K_{\lambda''} - K_{\lambda'''}),$$

where

$$\begin{aligned}\lambda' &= [n_1^{k_1}, \dots, n_{l-1}^{k_{l-1}}, n_l - 1, n_l^{k_l - 1}],\\ \lambda'' &= [n_1^{k_1}, \dots, n_{l-1}^{k_{l-1}}, n_l - 2, n_l^{k_l - 1}],\\ \lambda''' &= [n_1^{k_1}, \dots, n_{l-1}^{k_{l-1}}, (n_l - 1)^2, n_l^{k_l - 2}],\end{aligned}$$

reordered into a monotonic sequence, *if needed.*

Note that $\lambda' \vdash n-1$ and λ'', $\lambda''' \vdash n-2$. It is essential that the "reduced" partitions λ', λ'', λ''' are put into the canonical nondecreasing form we are using. This process is somewhat inconvenient to express in notation, but easy to do so in practice. For example, if $\lambda = [1, 2, 3^2]$ then $\lambda' = [1, 2^2, 3]$, $\lambda'' = [1^2, 2, 3]$ and $\lambda''' = [1, 2^3]$, while for $\lambda = [1, 2, 4^2]$, $\lambda' = [1, 2, 3, 4]$, $\lambda'' = [1, 2^2, 4]$ and $\lambda''' = [1, 2, 3^2]$, etc.

The algorithm can be used to express $K_n(a)$ for any a with a split characteristic polynomial when n is small, by first using Theorem 1.1. For example, for $n = 3$, $a = \alpha I + \bar{a}(\lambda)$, $\alpha \neq 0$, it gives

λ	$[1^3]$	$[1,2]$	$[3]$
K_λ	$q^3K^3 + (q^5 + 2q^4)(q-1)K$	$q^3K^3 + q^4(q-2)K$	$q^3K^3 - 2q^4K$

.

The first nontrivial example when λ''' appears is $[2^2] \vdash 4$, for which

$$K_{[2^2]} = q^3 K K_{[1,2]} - q^6 K_{[2]} - q^6(q-1)(K_{[2]} - K_{[1^2]}) = q^6 K^4 + q^7(q-3)K^2 + q^8(q^2 - q + 1).$$

To see a more intricate situation we illustrate the algorithm for $n = 6$ and $\lambda = [1, 2, 3]$, when we have

$$K_{[1,2,3]} = q^5 K K_{[1,2^2]} - q^{10} K_{[1^2,2]}, \qquad K_{[1,2^2]} = q^4 K_{[1^2,2]} - 2q^8 K_{[1,2]}$$

and so on.

There are many families where the recursion may be stated in simple terms, for example, if there is only one block, $\lambda = [n]$, when we have

$$K_{[n]} = K q^{n-1} K_{[n-1]} - q^{2n-2} K_{[n-2]}.$$

From this one can get a closed formula for $K_{[n]}$; see Section 5E below.

Theorem 5.1 is an easy corollary of the following two propositions. As usual we assume $\alpha \neq 0$, and $a = \alpha I + \bar{a}(\lambda)$ with $\varepsilon_j \in \{0, 1\}$ as in (22) with $\varepsilon_{n-1} = 1$ and use a'' to denote $a''_{\not k, \not n}$.

Proposition 5.2. *Let* $Z = \bigl\{z = \sum_{j=i+1}^{l-1} \xi_j \boldsymbol{e}_{k-1, N_j - 1} \mid \xi_j \in \mathbb{F}_q\bigr\}$. *In the above notation*,

(i) *$a'' + z$ and a'' are conjugate over $\mathbb{F}_q$*,

(ii) *$a'' + z + \boldsymbol{e}_{k-1,n-2}$ and $a'' + \boldsymbol{e}_{k-1,n-2}$ are conjugate over $\mathbb{F}_q$.*

As an immediate corollary of (28) we get that $\sum_{x\in X_k}\psi(ax+x^{-1})=0$, unless $k=N_i$ for some $k<l$ such that $n_k=n_l$ and then

$$\sum_{x\in X_k}\psi(ax+x^{-1})=q^{2n+k-l-3}(q-1)(K_{n-2}(a'')-K_{n-2}(a''+\boldsymbol{e}_{k-1,n-2})).$$

The second step in the proof is the following:

Proposition 5.3. (i) *If $n_i<n_l$ then $a''+\boldsymbol{e}_{k-1,n-2}$ and a'' are conjugate over $\mathbb{F}_q$.*

(ii) *If $n_i=n_l$, $a''+\boldsymbol{e}_{k-1,n-2}$ is conjugate to a''', where a''' is built from the partition λ''' as in* (22).

While not needed, we remark that the proposition remains true over $\mathbb{Z}$.

The proof of the claims in the propositions will use the linear transformation interpretation from Section 2F. We start with an easy observation.

Lemma 5.4. *Let $V_A=\langle v_0\rangle\oplus\cdots\oplus\langle v_l\rangle\simeq\mathcal{C}_{n_0}\oplus\cdots\oplus\mathcal{C}_{n_l}$. If $v\in V_A$ is such that $A^kv=0$ for $k<n_0$, then $V_A=\langle v_0+v\rangle\oplus\cdots\oplus\langle v_l\rangle$ as well.*

Proof. One easily checks that $\langle v_0+v\rangle\cap(\langle v_1\rangle\oplus\cdots\oplus\langle v_l\rangle)=\{0\}$ and that $\langle v_0+v\rangle+(\langle v_1\rangle\oplus\cdots\oplus\langle v_l\rangle)=V_A$. □

Remark 11. It follows that there is an isomorphism $\phi:V_A\to V_A$ which is trivial on $\langle v_1\rangle\oplus\cdots\oplus\langle v_l\rangle$ and extends $v_0\mapsto v_0+v$. Clearly, it satisfies $A\phi=\phi A$.

The question for us is to determine how a module structure given by $A:V\to V$ changes if A is perturbed by another map $Z:V\to V$. For example, Propositions 5.2 and Proposition 5.3(i) are easy consequences of the following lemma.

Lemma 5.5. *Assume that $V_A\simeq\langle v_0\rangle\oplus\langle v_1\rangle\oplus\cdots\oplus\langle v_l\rangle\simeq\mathcal{C}_{n_0}\oplus\mathcal{C}_{n_1}\oplus\cdots\oplus\mathcal{C}_{n_l}$, that $Z:V\to V$ is such that for $i=1,\dots,l$*

$$Z(A^jv_i)=0\quad \textit{for } j=0,\dots,n_i-1$$

and that for $i=0$ we have

$$Z(A^jv_0)=0\quad \textit{for } j=0,\dots,n_0-2,$$
$$Z(A^{n_0-1}v_0)=\sum_{i=1}^{l}\xi_iA^{n_i-1}v_i.$$

If $n_0<n_i$ for all $i=1,\dots,l$ then $V_{A+Z}\simeq V_A$ as $\mathbb{F}_q[T]$-modules.

Proof. Let $v=\sum_{i=1}^{l}\xi_iA^{n_i-n_0}v_i$. Clearly $A^{n_0}v=0$, and so by Lemma 5.4 there is an A-linear isomorphism ϕ, for which $\phi(v_0)=v_0-v$. One easily checks that $Z\circ\phi=0$ and so ϕ provides the claimed isomorphism. □

Lemma 5.6. *Assume that $V_A \simeq \langle v_0 \rangle \oplus \langle v_1 \rangle \simeq \mathcal{C}_m \oplus \mathcal{C}_m$ and that $Z : V \to V$ is such that*

$$\begin{aligned} Z(A^j v_1) &= 0 \quad \textit{for } j = 0, \dots, m-1, \\ Z(A^j v_0) &= 0 \quad \textit{for } j = 0, \dots, m-2, \\ Z(A^{m-1} v_0) &= \xi A^{m-1} v_1 \quad \textit{for some } \xi \in \mathbb{F}_q^*. \end{aligned}$$

Then $V_{A+Z} \simeq \mathcal{C}_{m-1} \oplus \mathcal{C}_{m+1}$ as $\mathbb{F}_q[T]$-modules.

Proof. It is easy to see that $(A+Z)^j v_0 = A^j v_0$, for $j = 0, \dots, m-1$ and that $(A+Z)^m v_0 = \xi A^{m-1} v_1$. Moreover if we replace v_1 by $v_1' = v_1 - \frac{1}{\xi} v_0$, then $(A+Z)^{m-1} v_1' = 0$. □

Proofs of Propositions 5.2 and 5.3. The two lemmas above give exactly this. □

5C. *Examples for the Kloosterman sums over Borel Bruhat cells.* Let B be the standard Borel subgroup of invertible upper triangular matrices, $w \in W$ an element of the Weyl group, and $C_w = BwB$. We will consider here the sums

$$K_n^{(w)}(a) = \sum_{x \in C_w} \psi(ax + x^{-1}),$$

where $a = \alpha I + \bar{a}$, with $\bar{a} \in \bar{U}$, where $\bar{U}$ is the set of strictly upper triangular matrices. We will first comment on the nature of these sums and then derive some of the properties that will be used below in the section on purity.

As in the proof of Theorem 1.3 in Section 2D one can show that these sums satisfy a recursion that connects them to similar sums of rank $n-1$ or $n-2$, depending on whether $w(n) = n$ or $w(n) < n$. To see this, recall from Proposition 3.14 that

$$K_n^{(w)}(a) = q^{n_a + n_o} K_1(\alpha)^f S_n^{(w)}(\bar{a}), \tag{55}$$

where n_o is the number of involution pairs in w, f is the number of fixed points of w and where the auxiliary sum is given by

$$S_n^{(w)}(\bar{a}) = \sum_{v,d,u_1} \psi(\bar{a}^v w d u),$$

with $v \in U^\flat$, $u \in U_b^\sharp$, $d \in D(w) = \left\{ \sum_{i < w(i)} t_i \boldsymbol{e}_{i,i} \mid t_i \in \mathbb{F}_q^* \right\}$ and $\bar{a}^v = v^{-1} \bar{a} v$. The recursion then proceeds on the sum $S_n^{(w)}(a)$. For example, when $w(n) = k < n$ we again let $m'' = m''_{\not k, \not n}$ denote the matrix one gets by deleting the k-th and n-th rows and columns of an $n \times n$ matrix m. To describe the set of perturbations $Z \subset M_{n-2}$ that arise in the reduction, we let $\bar{a}_{(i)}$ and $\bar{a}^{(i)}$ denote the i-th row, respectively the i-th column, of $\bar{a}$ and consider the set

$$Y = \{ y \in \mathbb{F}_q^n \mid y_i = 0 \text{ for } i \le k \text{ and } y\bar{a} = \bar{a}_{(k)} \}.$$

Using this notation we have the following proposition whose proof goes along the lines of Proposition 2.8 and is omitted here.

Proposition 5.7. *The following equation holds*:

$$S_n^{(w)}(\bar{a}) = q^{n-k-1} \sum_{y\in Y} S_{n-2}^{(w'')}(\bar{a}'' + (\bar{a}^{(k)}y)'') \sum_{t\in\mathbb{F}_q^*} \varphi(ty\bar{a}^{(n)}). \tag{56}$$

The set Y may be empty, and then so is the set of perturbations which arise from the collection of rank-1 matrices

$$Z = \{\bar{a}^{(k)}y \in M_n \mid y \in Y\},$$

in which case the sum is interpreted as 0.

What makes the sums $S_n^{(w)}(\bar{a})$ harder to deal with is that as functions of a they are no longer invariant under conjugation by GL_n. They are invariant under conjugation by elements of B in virtue of (55) and the fact that

$$K_n^{(w)}(a) = K_n^{(w)}(b^{-1}ab)$$

for any $b \in B$, since $b^{-1}C_w b = C_w$. However the B-orbits in the set $\overline{U}$ of strictly upper triangular matrices under the adjoint action are not well understood. It is easy to see that one can no longer straighten out partitions into a nondecreasing order, or even assume that an orbit is represented by a matrix in Jordan normal form. For example, $\{t\boldsymbol{e}_{1,3} \mid t \neq 0\}$ is one of the orbits in the 3×3 case. When $n = 6$, there is even a one-parameter family of orbits, found by Kashin [1990]. In general a full description of the orbits is hard even in low ranks [Bürgstein and Hesselink 1987; Hille and Röhrle 1997].

Returning to the sums $K_n^{(w)}(a)$, it is still quite likely that these can be expressed as polynomials in q and K, independently of the characteristic p since the perturbations arising in the reduction calculations are of a very special nature. In what follows we will provide some low-rank examples, when the independence is easy to establish directly. We will do this by stating certain special cases when the above reduction is sufficient, most importantly the case when $w = (ij)$. There are a number of other special cases when the reduction for $S_n^{(w)}(a)$ can be treated in a simple manner; for example, when $w(n) = 1$ or $n - 1$.

Again we assume that $a = \alpha I + \bar{a}$ is of the form as in (9), $\bar{a} = \sum_{j=1}^{n-1} \varepsilon_j \boldsymbol{e}_{j,j+1}$, and $\lambda \vdash n$ is the partition corresponding to a. Define

$$K_\lambda^{(w)}(\alpha) = \sum_{x\in C_w} \psi(ax + x^{-1}).$$

Theorem 5.8. *Let $i < j$ and $(ij) \in W$ and $\alpha \neq 0$. Then*

$$K_\lambda^{(ij)}(\alpha) = q^{n(n-1)/2} K^{n-2} \cdot \begin{cases} (q-1)q & \text{if } j = i+1 \text{ and } \varepsilon_i = 0, \\ -q & \text{if } j = i+1 \text{ and } \varepsilon_i = 1, \\ (q-1)q^{j-i-d} & \text{if } j > i+1 \text{ and } \varepsilon_i = \varepsilon_{j-1} = 0, \end{cases}$$

where $d = |\{k \in \mathbb{N} \mid i < k < j-1 \text{ and } \varepsilon_k \neq 0\}|$.

If $j > i+1$ and either ε_i or $\varepsilon_{j-1} \neq 0$ then $K_\lambda^{(ij)}(\alpha) = 0$.

This allows us to compute the Bruhat cell polynomials for $n \leq 3$; see Tables 1 and 2.

λ	$w =$ (12)	I
[1, 1]	$q^2(q-1)$	qK^2
[2]	$-q^2$	qK^2

Table 1. The $n = 2$ case.

λ	$w =$ (13)	(12)	(23)	I
[1, 1, 1]	$q^5(q-1)K$	$q^4(q-1)K$	$q^4(q-1)K$	q^3K^3
[2, 1]	0	$q^4(q-1)K$	$-q^4K$	q^3K^3
[1, 2]	0	$-q^4K$	$q^4(q-1)K$	q^3K^3
[3]	0	$-q^4K$	$-q^4K$	q^3K^3

Table 2. The $n = 3$ case.

With a little more work one can calculate all the Bruhat cell polynomials for $n = 4$. We summarize the result in Tables 3 and 4.

The polynomials are not merely a permutation for different rearrangements of a partition; see, for example, the case $\lambda = [1, 2, 1]$ and $w = (14)(23)$.

We finish this section by giving the cell polynomials for the full block ($\lambda = [n]$) case for general n. Let $w = (i_1, j_1)(i_2, j_2) \cdots (i_r, j_r) \in W$ such that $i_k < j_k$ for any k. Then

$$K^w_{[n]}(\alpha) = \begin{cases} (-1)^r K^{n-2r} q^{n(n-1)/2+r} & \text{if } j_k - i_k = 1 \text{ for any } k, \\ 0 & \text{otherwise.} \end{cases}$$

λ	$w =$ (14)(23)	(13)(24)	(12)(34)	(14)	(13)
[1, 1, 1, 1]	$q^{10}(q-1)^2$	$q^9(q-1)^2$	$q^8(q-1)^2$	$q^9(q-1)K^2$	$q^8(q-1)K^2$
[2, 1, 1]	0	0	$-q^8(q-1)$	0	0
[1, 2, 1]	$-q^9(q-1)$	0	$q^8(q-1)^2$	$q^8(q-1)K^2$	0
[1, 1, 2]	0	0	$-q^8(q-1)$	0	$q^8(q-1)K^2$
[3, 1]	0	0	$-q^8(q-1)$	0	0
[2, 2]	0	$q^9(q-1)$	q^8	0	0
[1, 3]	0	0	$-q^8(q-1)$	0	0
[4]	0	0	q^8	0	0

Table 3. The $n = 4$ case (continued below).

λ	$w =$ (24)	(12)	(23)	(34)	I
$[1, 1, 1, 1]$	$q^8(q-1)K^2$	$q^7(q-1)K^2$	$q^7(q-1)K^2$	$q^7(q-1)K^2$	q^6K^4
$[2, 1, 1]$	$q^8(q-1)K^2$	$-q^7K^2$	$q^7(q-1)K^2$	$q^7(q-1)K^2$	q^6K^4
$[1, 2, 1]$	0	$q^7(q-1)K^2$	$-q^7K^2$	$q^7(q-1)K^2$	q^6K^4
$[1, 1, 2]$	0	$q^7(q-1)K^2$	$q^7(q-1)K^2$	$-q^7K^2$	q^6K^4
$[3, 1]$	0	$-q^7K^2$	$-q^7K^2$	$q^7(q-1)K^2$	q^6K^4
$[2, 2]$	0	$-q^7K^2$	$q^7(q-1)K^2$	$-q^7K^2$	q^6K^4
$[1, 3]$	0	$q^7(q-1)K^2$	$-q^7K^2$	$-q^7K^2$	q^6K^4
$[4]$	0	$-q^7K^2$	$-q^7K^2$	$-q^7K^2$	q^6K^4

Table 4. The $n = 4$ case (continued).

Since this is merely for illustration we only give a sketch of the argument. For those $w \in W$ such that $K_n^w(\alpha) = 0$ one can find $(i, j) \in I$ such that $\operatorname{tr}(v^{-1}avwdu)$ is nonconstant and linear in $v_{i,j}$.

5D. ***Regular semisimple matrices.*** We have seen that for an $n \times n$ matrix a, whose characteristic polynomial P_a have no multiple roots, the cohomology associated to the Kloosterman sum $K_n(a)$ is pure. Assume now that this characteristic polynomial P_a is irreducible over $\mathbb{F}_q$. Let $\alpha \in \mathbb{F}_{q^n}$ be an eigenvalue of a, $P_a(\alpha) = 0$. The argument in Section 3E shows that over $\mathbb{F}_{q^n}$

$$H^\bullet = H_c^\bullet(\mathrm{GL}_n, x \mapsto \operatorname{tr}(ax + x^{-1})) = \Bigl(\bigotimes_{i=1}^{n(n-1)/2} H_c^\bullet(\mathbb{A}^1, 0)\Bigr) \otimes \Bigl(\bigotimes_{j=1}^{n} H_c^\bullet(\mathbb{A}^1 \setminus \mathbb{A}^0, f_\alpha)\Bigr),$$

where for $x \in (\mathbb{A}^1 \setminus \mathbb{A}^0)(\mathbb{F}_{q^n})$, $f_\alpha(x) = \alpha x + x^{-1}$, that corresponds to the scalar Kloosterman sum $K_1(\alpha, \mathbb{F}_{q^n}) = \lambda_1 + \lambda_2$. Here λ_1 and $\lambda_2 = \bar{\lambda}_1$ are the $\mathbb{F}_{q^n}$-Frobenius eigenvalues on $H_c^\bullet(\mathbb{A}^1 \setminus \mathbb{A}^0, f_\alpha)$.

Then, clearly, over $\mathbb{F}_{q^n}$ the Frobenius eigenvalues on the 2^n-dimensional space $\bigotimes_{j=1}^n H_c^1(\mathbb{A}^1 \setminus \mathbb{A}^0, f_\alpha)$ are of the form $\bigl(\prod_{i\in I} \lambda_1\bigr)\bigl(\prod_{i\notin I} \lambda_2\bigr) = \lambda_1^{|I|}\lambda_2^{n-|I|}$, where $I \subset \{1, \dots, n\}$—therefore each of $\lambda_1^{|I|}\lambda_2^{n-|I|}$ has multiplicity $\binom{n}{j}$. If we fix some n-th roots of the λ_i, say $\eta_i^n = \lambda_i$, then we have that the Frobenius eigenvalues on H^{n^2} are of the form $\zeta_I q^{n(n-1)/2}\eta_1^{|I|}\eta_2^{n-|I|}$ where again $I \subset \{1, \dots, n\}$, and the ζ_I are n-th roots of unity, $\zeta_I^n = 1$ for all I. It is natural to make the following conjecture.[2]

Conjecture 5.9. If p is large enough, and the characteristic polynomial P_a is irreducible over $\mathbb{F}_q$ then

$$K_n(a, \mathbb{F}_q) = (-1)^{n+1} q^{n(n-1)/2} K_1(\alpha, \mathbb{F}_{q^n}).$$

The conjecture would follow if, for $I = \varnothing$ and $\{1, \dots, n\}$, the $\mathbb{F}_q$-Frobenius eigenvalues were $(-1)^{n+1}q^{n(n-1)/2}\lambda_1$, $(-1)^{n+1}q^{n(n-1)/2}\lambda_2$ and the others canceled after summing. For example, when

[2] While this paper was in print, Elad Zelingher [2023] announced a proof of this conjecture.

$n = 3$, this can happen if the eigenvalues μ_i for $1 \leq i \leq 8$ are the eight summands in the expansion of the product $(\eta_1 + \eta_2)(\omega\eta_1 + \omega^2\eta_2)(\omega^2\eta_1 + \omega\eta_2)q^3$ for $\omega^3 = 1$; this leads to

$$K_n(a, \mathbb{F}_q) = \sum_{i=1}^{8} \mu_i = q^3(\lambda_1 + \lambda_2) = (-1)^4 q^3 K_1(\alpha, \mathbb{F}_{q^3}),$$

exactly as desired.

The conjecture is partly based on the observation that if we let $K = \mathbb{F}_q[a] \subset M_n$, then K is a field naturally isomorphic to $\mathbb{F}_{q^n}$. K acts on M_n by left multiplication and it is easy to describe the K-algebra that arises for any n. We will use this below to handle the case $n = 2$, but such elementary methods get cumbersome and are unlikely to give a proof, or even offer any insight already for $n = 3$.

We now give a few numerical examples for $M_n(\mathbb{F}_{p^n})$ for small n and p checked with computer algebra systems pari/gp and Sage.

(i) Let $n = 3$, $p = 5$ and $\alpha \in \mathbb{F}_{125}$ be one of the roots of $x^3 + x^2 + 1$. Then $K_1(\alpha, \mathbb{F}_{125}) = \big(3 + \sqrt{5}\big)/2$. On the other hand if

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & -1 \end{bmatrix}$$

then $K_3(A, \mathbb{F}_5) = 327.2542$ which agrees with $125K_1(\alpha, \mathbb{F}_{125})$.

(ii) Let $n = 4$, $p = 3$ and $\alpha \in \mathbb{F}_{81}$ be a root of $x^4 + 2x^3 + 2$, and

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Then $K_4(A, \mathbb{F}_3) = 11664$ which agrees with $-729K_1(\alpha, \mathbb{F}_{81})$.

(iii) Let $n = 3$, $p \equiv 1$ (3) and

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \mu & 0 & 0 \end{bmatrix},$$

where $\mu \in \mathbb{F}_p^* \setminus (\mathbb{F}_p^*)^3$. If $\alpha \in \mathbb{F}_{p^3}$ is such that $\alpha^3 = \mu$ then one can check that

$$K(\alpha) = \sum e(3\mu c + (3a^2 - 3\mu cb)/\Delta(a, b, c)),$$

where $e(x) = e^{2\pi i x}$, $\Delta(a, b, c) = a^3 - 3\mu cba + (\mu b^3 + \mu^2 c^3)$ and where the sum is over $(a, b, c) \in \mathbb{F}_p^3 \setminus \{(0, 0, 0)\}$.

A direct calculation using Bruhat decomposition shows that the conjecture in this case is equivalent to

$$K(\alpha) = K_{13}(A) + (1 + (\mu/p))q,$$

where (μ/p) is the Legendre symbol and where

$$K_{13}(A)=\sum_{t_1,t_2,t_3\in\mathbb{F}_p^*} e(\mu t_1^2t_3^2+t_1^2t_3+\mu t_3+1/t_2-1/(\mu t_1 t_3^2)).$$

Up to about $p\le 200$, this can be checked fast even on a personal computer. For example, the order of 2 mod 199 is 99, and we have that

$$K\big(\sqrt[3]{2},\mathbb{F}_{199^3}\big)=3869.8269,$$

while for

$$A=\begin{pmatrix}0&1&0\\0&0&1\\2&0&0\end{pmatrix}$$

we have

$$K_{13}(A)=4267.8269$$

with a difference of 398, which shows that $K_3(A,\mathbb{F}_{199})=K\big(\sqrt[3]{2},\mathbb{F}_{199^3}\big)$. On the other hand 3 mod 199 is a primitive root, and so for

$$A=\begin{pmatrix}0&1&0\\0&0&1\\3&0&0\end{pmatrix}$$

we have

$$K\big(\sqrt[3]{3},\mathbb{F}_{199^3}\big)=K_{13}(A)=-2875.1994.$$

We also checked all $p=3k+1\le 200$, for which 2 is not a cube mod p and found that $K\big(\sqrt[3]{2},\mathbb{F}_{p^3}\big)=K_3(A,\mathbb{F}_p)$ holds for all of them.

Proposition 5.10. *Assume that q is odd and let $\alpha\in\mathbb{F}_{q^2}\setminus\mathbb{F}_q$ and $a=\big(\begin{smallmatrix}0&1\\-N(\alpha)&-\mathrm{Tr}(\alpha)\end{smallmatrix}\big)$, where N and Tr are the norm and trace of the field extension $\mathbb{F}_{q^2}/\mathbb{F}_q$. Then*

$$K_2(a,M_2(\mathbb{F}_q))=-qK_1(\alpha,\mathbb{F}_{q^2}^*).$$

Remark 12. The cohomology complex $H_c^\bullet$ corresponding to the sum $K_2(a,M_2(\mathbb{F}_q))$ satisfies

$$\dim H_c^i=\begin{cases}0 & \text{if } i\ne 4,\\ 4 & \text{if } i=4,\end{cases}$$

and the Frobenius eigenvalues μ_i on H_c^4 satisfy $\mu_1^2=q^2\lambda_1^2$, $\mu_2^2=q^4$, $\mu_3^2=q^4$, and $\mu_4^2=q^2\lambda_2^2$ where $\lambda_2=\bar\lambda_1$ are the eigenvalues corresponding to $K_1(\alpha,\mathbb{F}_{q^2}^*)$. Apart from permutations the proposition determines the sign of the square roots, we have $\mu_1=-q\lambda_1$, $\mu_2=q^2$, $\mu_3=-q^2$ and $\mu_4=-q\lambda_2$. Thus $K_2(a,M_2(\mathbb{F}_{q^m}))=\sum_{i=1}^4\mu_i^m$. Here again we have cancellation: $\mu_2^m+\mu_3^m=0\iff 2\nmid m$.

Proof of Proposition 5.10. As above let $K=\mathbb{F}_q[a]$ be the subring of M_2 generated by a. K is isomorphic to $\mathbb{F}_{q^2}$ by the assumption on α. Also, the vector space $\mathbb{F}_q^2$ as an $\mathbb{F}_q[a]$-module is isomorphic to the $\mathbb{F}_q$ vector space $\mathbb{F}_{q^2}$, with a acting via multiplication by α. For the moment denote this action by L_α, $L_\alpha:\beta\mapsto\alpha\beta$.

Let $\mathbb{F}_{q^2}\langle\tau\rangle$ be the noncommutative ring of twisted polynomials, $\sum_i \xi_i \tau^i$, subject to $\tau\xi = F(\xi)\tau$, where $F(\xi) = \xi^q$ is the Frobenius automorphism of $\mathbb{F}_{q^2}/\mathbb{F}_q$.

There is an obvious map from $\mathcal{M}_2 = \mathbb{F}_{q^2}\langle\tau\rangle/(\tau^2 - 1)$ to $M_2(\mathbb{F}_q)$, sending $\xi_1 + \xi_2\tau$ to the $\mathbb{F}_q$-linear transformation $L_{\xi_0} + L_{\xi_1}F$. It is not difficult to see that this linear map is injective, and so by dimension count, an isomorphism. This identifies M_2 with $\mathcal{M}_2$, and it is easy to check that under this identification $\psi(L_{\xi_0} + L_{\xi_1}F) = \varphi_2(\xi_0)$, where $\varphi_2 = \varphi \circ \operatorname{Tr}_{\mathbb{F}_{q^2}/\mathbb{F}_q}$. It follows that

$$\sum_{x\in M_2(\mathbb{F}_q)^*} \psi(ax + x^{-1}) = \sum_{\xi_0+\xi_1\tau\in\mathcal{M}_2^*} \varphi_2(\alpha\xi_0 + (\xi_0+\xi_1\tau)^{-1}).$$

An easy calculations shows that $(1+\xi\tau) \in \mathcal{M}_2^*$ exactly when $N(\xi) \neq 1$, and then $(1+\xi\tau)^{-1} = \frac{1}{1-N(\xi)}(1-\xi\tau)$. One also has that $(\xi\tau)^{-1} = F(\xi^{-1})\tau$ and so

$$\mathcal{M}_2^* = \{\xi_1\tau \mid \xi_1 \in \mathbb{F}_{q^2}^*\} \cup \{\xi_0(1+\xi_1\tau) \mid \xi_0 \in \mathbb{F}_{q^2}^*,\ \xi_1 \in \mathbb{F}_{q^2},\ N(\xi_1) \neq 1\}.$$

Therefore

$$\sum_{x\in \mathrm{GL}_2(\mathbb{F}_q)} \psi(ax + x^{-1}) = q^2 - 1 + \sum_{\substack{\xi_0\in\mathbb{F}_{q^2}^* \\ \xi_1\in\mathbb{F}_{q^2} \\ N(\xi_1)\neq 1}} \varphi_2\big(\alpha\xi_0 + (1-N(\xi_1))^{-1}\xi_0^{-1}\big).$$

Now the norm map N is a surjective homomorphism from $\mathbb{F}_{q^2}^* \to \mathbb{F}_q^*$, with a kernel of size $q+1$ and so for $\gamma \in \mathbb{F}_q$

$$\big|\{\xi \in \mathbb{F}_{q^2} \mid (1-N(\xi))^{-1} = \gamma\}\big| = \begin{cases} 0 & \text{if } \gamma = 0, \\ 1 & \text{if } \gamma = 1, \\ q+1 & \text{if } \gamma \neq 0, 1. \end{cases}$$

This gives

$$\sum_{x\in \mathrm{GL}_2(\mathbb{F}_q)} \psi(ax + x^{-1}) = q^2 - 1 + (q+1) \sum_{\substack{\xi_0\in\mathbb{F}_{q^2}^* \\ \gamma\in\mathbb{F}_q^*}} \varphi_2(\alpha\xi_0 + \gamma\xi_0^{-1}) - q \sum_{\xi_0\in\mathbb{F}_{q^2}^*} \varphi_2(\alpha\xi_0 + \xi_0^{-1}).$$

Finally

$$\sum_{\substack{\xi_0\in\mathbb{F}_{q^2}^* \\ \gamma\in\mathbb{F}_q^*}} \varphi_2(\alpha\xi_0 + \gamma\xi_0^{-1}) = \sum_{\substack{\xi_0\in\mathbb{F}_{q^2}^* \\ \gamma\in\mathbb{F}_q^*}} \varphi_2(\alpha\xi_0)\phi(\gamma \operatorname{Tr}\xi_0^{-1}) = -\sum_{\substack{\xi_0\in\mathbb{F}_{q^2}^* \\ \operatorname{Tr}\xi_0^{-1}=0}} \varphi_2(\alpha\xi_0) = -(q-1). \qquad \square$$

5E. ***The purity locus.*** We have seen that for a regular semisimple element $a \in M_n(\mathbb{F}_q)$ the Kloosterman sum $K_n(a)$ is pure. The tables above already suggest that for a matrix a with more than one Jordan block for an eigenvalue, that sum cannot be pure. This can be seen without reference to cohomology. To see this assume that a has a single eigenvalue α. Recall that

$$K_n(a) = P(A, G, K) = \sum_{2f\leq n} c_f(q, q-1)K^{n-2f},$$

where c_f are polynomials. We give the A and G weight 1, and K weight $\frac{1}{2}$, so the polynomial P has a weighted degree, which determines the order of magnitude (in q) of its value. Now $f = 1$ corresponds to simple transpositions, and by Theorem 5.8 one sees that these sums are too large in magnitude to be pure if not all ε_i are 0.

It is an intriguing question what happens for $K_{[n]}$ when a has only one Jordan block. The recursion formula gives

$$K_{[n]} = q^{n-1} K K_{[n-1]} - q^{2n-2} K_{[n-2]},$$

where $K = K_1(\alpha)$. Let $k_n = q^{-n(n-1)/2} K_{[n]}$, so that we have

$$k_n = K k_{n-1} - q k_{n-2}.$$

It follows that there exist c_1, c_2 such that

$$k_n = c_1 \lambda_1^n + c_2 \lambda_2^n,$$

where λ_1, λ_2 are the roots of $X^2 - KX + q$. These are exactly the eigenvalues of Frobenius acting on the cohomology of the Kloosterman sheaf. Using that $K = \lambda_1 + \lambda_2$, and that $K_{[2]} = -q^2 + K_1^2 q$ we get that $k_1 = \lambda_1 + \lambda_2$, and that $k_2 = k_1^2 - q = \lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2$. Therefore $c_1 = \frac{\lambda_1}{\lambda_1 - \lambda_2}$, $c_2 = -\frac{\lambda_2}{\lambda_1 - \lambda_2}$ and

$$k_n = \frac{\lambda_1^{n+1} - \lambda_2^{n+1}}{\lambda_1 - \lambda_2} = \sum_{j=0}^{n} \lambda_1^j \lambda_2^{n-j}. \tag{57}$$

This evaluation has an interesting interpretation. Let $X^2 - K_1 X + q = (X - \lambda_1)(X - \lambda_2)$, with $\lambda_{1,2} = q^{1/2} e^{\pm i\theta}$, so that $K_1 = 2q^{1/2}\cos\theta$. We have that

$$k_n = \frac{\lambda_1^{n+1} - \lambda_2^{n+1}}{\lambda_1 - \lambda_2} = q^{n/2} \frac{e^{(n+1)\theta} - e^{-(n+1)\theta}}{e^{\theta} - e^{-\theta}} = \frac{\sin(n+1)\theta}{\sin\theta}.$$

Therefore

$$K_{[n]} = q^{n(n-1)/2} U_n(\cos\theta), \tag{58}$$

where U_n is the Chebyshev polynomial of the second kind. The Sato–Tate distribution of the angles of $K_1(\alpha)$ over the valuations of a global field is then equivalent to nontrivial cancellation in the sums

$$\sum_{N(v) \le x} K_n(a, \mathbb{F}_v)/N(v)^{n(n-1)/2},$$

where $a = \alpha I + \sum_{i=1}^{n-1} e_{i,i+1}$.

Getting back to the question of purity these sums are pure from a numerical point of view, but this in itself does not rule out a cohomology with a nilpotent Frobenius action.

For example, in the case $n = 2$ it is easy to see that the cohomologies corresponding to the Bruhat cells are as follows.

On $C_I = B$ the trace of $\alpha x + x^{-1}$ can be written as a product of two Kloosterman sums over the diagonal elements; thus we have

$$H^\bullet_{C_I} = H^\bullet(C_I, x \mapsto \mathrm{tr}(ax + x^{-1})) = H^\bullet_c(\mathbb{A}^1 - \mathbb{A}^0, f_\alpha) \otimes H^\bullet_c(\mathbb{A}^1 - \mathbb{A}^0, f_\alpha) \otimes H^\bullet_c(\mathbb{A}^1, 0).$$

That implies $\dim H^i_{C_I} = 0$ unless $i = 4$ and $\dim H^4_{C_I} = 4$.

On the nontrivial cell $C_w = UwB$ we have seen that the sum (and the cohomology) cancels on the subvariety $\alpha \det b \neq 1$ and on the rest we have

$$H^\bullet_{C_w} = H^\bullet(C_w, x \mapsto \mathrm{tr}(ax + x^{-1})) = H^\bullet_c(\mathbb{A}^1 - \mathbb{A}^0, \mathrm{id}) \otimes H^\bullet_c(\mathbb{A}^1, 0) \otimes H^\bullet_c(\mathbb{A}^1, 0).$$

That implies $\dim H^i_{C_w} = 0$ unless $i = 5$ and $\dim H^5_{C_w} = 1$.

Thus the long exact sequence of the excision ($C_w = G \setminus C_I$) gives

$$0 \to H^4_G \to H^4_{C_I} \to H^5_{C_w} \to H^5_G \to 0.$$

Either $\dim H^4_G = 4$ and $\dim H^5_G = 1$ or $\dim H^4_G = 3$ and $\dim H^5_G = 0$ seems to be possible.

The same problem exists for higher-degree cases.

Acknowledgements

We thank Will Sawin and Min Lee for suggesting the problem of matrix Kloosterman sums. We thank the anonymous referees whose reports helped in improving the paper. We also thank Sawin and Gergely Zábrádi for helpful conversations and the Rényi Institute of Mathematics for ongoing support.

References

[Adolphson and Sperber 1989] A. Adolphson and S. Sperber, "Exponential sums and Newton polyhedra: cohomology and estimates", *Ann. of Math.* (2) **130**:2 (1989), 367–406. MR Zbl

[Borel 1991] A. Borel, *Linear algebraic groups*, 2nd ed., Grad. Texts in Math. **126**, Springer, 1991. MR Zbl

[Bürgstein and Hesselink 1987] H. Bürgstein and W. H. Hesselink, "Algorithmic orbit classification for some Borel group actions", *Compos. Math.* **61**:1 (1987), 3–41. MR Zbl

[Cameron 2017] P. J. Cameron, *Notes on counting: an introduction to enumerative combinatorics*, Austral. Math. Soc. Lect. Ser. **26**, Cambridge Univ. Press, 2017. MR Zbl

[Carlitz 1969] L. Carlitz, "Kloosterman sums and finite field extensions", *Acta Arith.* **16** (1969), 179–193. MR Zbl

[Chae and Kim 2003] H.-j. Chae and D. S. Kim, "*L* functions of some exponential sums of finite classical groups", *Math. Ann.* **326**:3 (2003), 479–487. MR Zbl

[Dabrowski 1993] R. Dabrowski, "Kloosterman sums for Chevalley groups", *Trans. Amer. Math. Soc.* **337**:2 (1993), 757–769. MR Zbl

[Dabrowski and Reeder 1998] R. Dąbrowski and M. Reeder, "Kloosterman sets in reductive groups", *J. Number Theory* **73**:2 (1998), 228–255. MR Zbl

[Deligne 1980] P. Deligne, "La conjecture de Weil, II", *Inst. Hautes Études Sci. Publ. Math.* **52** (1980), 137–252. MR Zbl

[Denef and Loeser 1991] J. Denef and F. Loeser, "Weights of exponential sums, intersection cohomology, and Newton polyhedra", *Invent. Math.* **106**:2 (1991), 275–294. MR Zbl

[Deshouillers and Iwaniec 1982] J.-M. Deshouillers and H. Iwaniec, "Kloosterman sums and Fourier coefficients of cusp forms", *Invent. Math.* **70**:2 (1982), 219–288. MR Zbl

[Einsiedler et al. 2016] M. Einsiedler, S. Mozes, N. Shah, and U. Shapira, "Equidistribution of primitive rational points on expanding horospheres", *Compos. Math.* **152**:4 (2016), 667–692. MR Zbl

[El-Baz et al. 2022] D. El-Baz, M. Lee, and A. Strömbergsson, "Effective equidistribution of primitive rational points on expanding horospheres", preprint, 2022. arXiv 2212.07408

[Erdélyi et al. 2024a] M. Erdélyi, W. Sawin, and Á. Tóth, "The purity locus of matrix Kloosterman sums", *Trans. Amer. Math. Soc.* **377**:6 (2024), 4117–4132. MR Zbl

[Erdélyi et al. 2024b] M. Erdélyi, Á. Tóth, and G. Zábrádi, "Matrix Kloosterman sums modulo prime powers", *Math. Z.* **306**:4 (2024), art. id. 68. MR Zbl

[Fouvry and Katz 2001] É. Fouvry and N. Katz, "A general stratification theorem for exponential sums, and applications", *J. Reine Angew. Math.* **540** (2001), 115–166. MR Zbl

[Fresán and Jossen 2020] J. Fresán and P. Jossen, "Exponential motives", preprint, 2020, available at https://tinyurl.com/expmot.

[Friedberg 1987] S. Friedberg, "Poincaré series for GL(n): Fourier expansion, Kloosterman sums, and algebreo-geometric estimates", *Math. Z.* **196**:2 (1987), 165–188. MR Zbl

[Fulman 2001] J. Fulman, "A new bound for Kloosterman sums", preprint, 2001. arXiv math/0105172

[Goldfeld and Sarnak 1983] D. Goldfeld and P. Sarnak, "Sums of Kloosterman sums", *Invent. Math.* **71**:2 (1983), 243–250. MR Zbl

[Grothendieck 1965] A. Grothendieck, "Formule de Lefschetz et rationalité des fonctions L", exposé 279 in *Séminaire Bourbaki*, 1964/1965, Benjamin, Amsterdam, 1965. Reprinted as pp. 41–55 in *Séminaire Bourbaki* **9**, Soc. Math. France, Paris, 1995. MR Zbl

[Hasse 1935] H. Hasse, "Theorie der relativ-zyklischen algebraischen Funktionenkörper, insbesondere bei endlichem Konstantenkörper", *J. Reine Angew. Math.* **172** (1935), 37–54. MR Zbl

[Heath-Brown 2000] D. R. Heath-Brown, "Arithmetic applications of Kloosterman sums", *Nieuw Arch. Wiskd.* (5) **1**:4 (2000), 380–384. MR Zbl

[Herz 1955] C. S. Herz, "Bessel functions of matrix argument", *Ann. of Math.* (2) **61** (1955), 474–523. MR Zbl

[Hille and Röhrle 1997] L. Hille and G. Röhrle, "On parabolic subgroups of classical groups with a finite number of orbits on the unipotent radial", *C. R. Acad. Sci. Paris Sér. I Math.* **325**:5 (1997), 465–470. MR Zbl

[Hodges 1956] J. H. Hodges, "Weighted partitions for general matrices over a finite field", *Duke Math. J.* **23** (1956), 545–552. MR Zbl

[James and Kerber 1981] G. James and A. Kerber, *The representation theory of the symmetric group*, Encycl. Math. Appl. **16**, Addison-Wesley, Reading, MA, 1981. MR Zbl

[Kashin 1990] V. V. Kashin, "Orbits of an adjoint and co-adjoint action of Borel subgroups of a semisimple algebraic group", pp. 141–158 in *Problems in group theory and homological algebra*, edited by A. L. Onishchik, Yaroslavl Gos. Univ., 1990. In Russian. MR Zbl

[Katz 1980] N. M. Katz, *Sommes exponentielles*, Astérisque **79**, Soc. Math. France, Paris, 1980. MR Zbl

[Katz 1988] N. M. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Ann. of Math. Stud. **116**, Princeton Univ. Press, 1988. MR Zbl

[Katz and Laumon 1985] N. M. Katz and G. Laumon, "Transformation de Fourier et majoration de sommes exponentielles", *Inst. Hautes Études Sci. Publ. Math.* **62** (1985), 145–202. MR Zbl

[Kiehl and Weissauer 2001] R. Kiehl and R. Weissauer, *Weil conjectures, perverse sheaves and l'adic Fourier transform*, Ergebnisse der Math. (3) **42**, Springer, 2001. MR Zbl

[Kim 1998] D. S. Kim, "Gauss sums for symplectic groups over a finite field", *Monatsh. Math.* **126**:1 (1998), 55–71. MR Zbl

[Kloosterman 1927] H. D. Kloosterman, "On the representation of numbers in the form $ax^2+by^2+cz^2+dt^2$", *Acta Math.* **49**:3-4 (1927), 407–464. MR Zbl

[Kowalski et al. 2017] E. Kowalski, P. Michel, and W. Sawin, "Bilinear forms with Kloosterman sums and applications", *Ann. of Math.* (2) **186**:2 (2017), 413–500. MR Zbl

[Landsberg 1893] G. Landsberg, "Ueber eine Anzahlbestimmung und eine damit zusammenhängende Reihe", *J. Reine Angew. Math.* **111** (1893), 87–88. MR Zbl

[Laumon 2000] G. Laumon, "Exponential sums and l-adic cohomology: a survey", *Israel J. Math.* **120** (2000), 225–257. MR Zbl

[Lee and Marklof 2018] M. Lee and J. Marklof, "Effective equidistribution of rational points on expanding horospheres", *Int. Math. Res. Not.* **2018**:21 (2018), 6581–6610. MR Zbl

[Linnik 1963] J. V. Linnik, "Additive problems and eigenvalues of the modular operators", pp. 270–284 in *Proceedings of the International Congress of Mathematicians* (Stockholm, 1962), edited by V. Stenström, Inst. Mittag-Leffler, Djursholm, Sweden, 1963. MR Zbl

[Luo et al. 1995] W. Luo, Z. Rudnick, and P. Sarnak, "On Selberg's eigenvalue conjecture", *Geom. Funct. Anal.* **5**:2 (1995), 387–401. MR Zbl

[Marklof 2010] J. Marklof, "The asymptotic distribution of Frobenius numbers", *Invent. Math.* **181**:1 (2010), 179–207. MR Zbl

[Milne 1980] J. S. Milne, *Étale cohomology*, Princeton Math. Ser. **33**, Princeton Univ. Press, 1980. MR Zbl

[Milne 2016] J. S. Milne, "The Riemann hypothesis over finite fields from Weil to the present day", pp. 487–565 in *The legacy of Bernhard Riemann after one hundred and fifty years*, *II*, edited by L. Ji et al., Adv. Lect. Math. **35**, Int. Press, Somerville, MA, 2016. MR Zbl

[Mordell 1963] L. J. Mordell, "On a special polynomial congruence and exponential sum", pp. 29–32 in *Calcutta Math. Soc. Golden Jubilee Commemoration*, *I* (Calcutta, 1958-1959), Calcutta Math. Soc., 1963. MR Zbl

[Morrison 2006] K. E. Morrison, "Integer sequences and matrices over finite fields", *J. Integer Seq.* **9**:2 (2006), art. id. 06.2.1. MR Zbl

[Petersson 1932] H. Petersson, "Über die Entwicklungskoeffizienten der automorphen Formen", *Acta Math.* **58**:1 (1932), 169–215. MR Zbl

[Poincaré 1911] H. Poincaré, "Fonctions modulaires et fonctions fuchsiennes", *Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys.* (3) **3** (1911), 125–149. MR Zbl

[Selberg 1965] A. Selberg, "On the estimation of Fourier coefficients of modular forms", pp. 1–15 in *Theory of numbers* (Pasadena, CA, 1963), edited by A. L. Whiteman, Proc. Sympos. Pure Math. **8**, Amer. Math. Soc., Providence, RI, 1965. MR Zbl

[SGA $4\frac{1}{2}$ 1977] P. Deligne, "Applications de la formule des traces aux sommes trigonométriques", pp. 168–232 in *Cohomologie étale* (Séminaire de Géométrie Algébrique du Bois Marie), edited by P. Deligne, Lecture Notes in Math. **569**, Springer, 1977. MR Zbl

[Springer 1998] T. A. Springer, *Linear algebraic groups*, 2nd ed., Progr. Math. **9**, Birkhäuser, Boston, MA, 1998. MR Zbl

[Stanley 1986] R. P. Stanley, *Enumerative combinatorics, I*, Wadsworth & Brooks, Monterey, CA, 1986. MR Zbl

[Stevens 1987] G. Stevens, "Poincaré series on $\mathrm{GL}(r)$ and Kloostermann [sic] sums", *Math. Ann.* **277**:1 (1987), 25–51. MR Zbl

[Sylvester 1885] J. J. Sylvester, "Sur l'équation en matrices $px = xq$", *C. R. Acad. Sci. Paris* **99**:2 (1885), 67–71. Zbl JFM

[Weil 1948a] A. Weil, "On some exponential sums", *Proc. Nat. Acad. Sci. U.S.A.* **34** (1948), 204–207. MR Zbl

[Weil 1948b] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Publ. Math. Inst. Univ. Strasbourg **7**, Hermann & Cie, Paris, 1948. MR Zbl

[Zelingher 2023] E. Zelingher, "On matrix Kloosterman sums and Hall–Littlewood polynomials", preprint, 2023. arXiv 2312.13121

Communicated by Roger Heath-Brown
Received 2023-03-17 Revised 2023-10-30 Accepted 2023-12-18

merdelyi@math.bme.hu Department of Algebra and Geometry,
Budapest University of Technology and Economics, Budapest, Hungary

arpad.toth@ttk.elte.hu Department of Analysis, Eötvös Loránd University, Budapest, Hungary

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the Submission page at the ANT website.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles in *ANT* are usually in English, but articles written in other languages are welcome.

Length There is no a priori limit on the length of an *ANT* article, but *ANT* considers long articles only if the significance-to-length ratio is appropriate. Very long manuscripts might be more suitable elsewhere as a memoir instead of a journal article.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and subject classifications for the article, and, for each author, postal address, affiliation (if appropriate), and email address.

Format. Authors are encouraged to use LaTeX but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages (Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc.) allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with details about how your graphics were generated.

White space. 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.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Algebra & Number Theory

Volume 18 No. 12 2024