A nonabelian circle method

Victor Wang
(joint work with Nuno Arala, Jayce Getz, Jiaqi Hou,
Chun-Hsien Hsu, and Huajie Li; NSF RTG DMS-2231514)

IST Austria and IoM Academia Sinica

Aarhus Automorphic Forms Conference, August 2025

This project has received funding from the European Union's Horizon 2020 research and innovation program

under the Marie Skłodowska-Curie Grant Agreement No. 101034413

Some matrix equations

Let $M_d(R)$ be the set of $d \times d$ matrices with entries in R.

- $XY = YX$, where $X, Y \in M_d(\mathbb{Z})$. This is interesting for $d \geq 2$ (nonabelian). The number of pairs with entries in $[-T, T]$ as $T \to \infty$ is studied in [Browning–Sawin–W. 2024, Mudgal 2024, Chapman–Mudgal 2025].
- $X^d = A$, where $X \in M_d(\mathbb{Z})$. Again, something interesting happens for $d \geq 2$. For typical A, such as if $\det(A) \neq \square^d$, this has no solutions. How about special A? For scalar $A = kI_d$ with $k \in \mathbb{Z}$, this has $\sim c_k T^{d(d-1)/2}$ solutions as $T \to \infty$ [Eskin–Mozes–Shah 1996].[1]

[1] One can also study other asymptotic aspects of the point count, such as Cesàro convergence over k when $T \asymp |k|^{1/d}$. See results on the Linnik problem in [Einsiedler–Lindenstrauss–Michel–Venkatesh 2011].

Some matrix equations

Let $M_d(R)$ be the set of $d \times d$ matrices with entries in R.

- $XY = YX$, where $X, Y \in M_d(\mathbb{Z})$. This is interesting for $d \geq 2$ (nonabelian). The number of pairs with entries in $[-T, T]$ as $T \to \infty$ is studied in [Browning–Sawin–W. 2024, Mudgal 2024, Chapman–Mudgal 2025].
- $X^d = A$, where $X \in M_d(\mathbb{Z})$. Again, something interesting happens for $d \geq 2$. For typical A, such as if $\det(A) \neq \square^d$, this has no solutions. How about special A? For scalar $A = kI_d$ with $k \in \mathbb{Z}$, this has $\sim c_k T^{d(d-1)/2}$ solutions as $T \to \infty$ [Eskin–Mozes–Shah 1996].[1]
- This talk will concentrate on nonabelian sums of n squares, especially the *best error term as* $n \to \infty$ (Weyl sums).

[1] One can also study other asymptotic aspects of the point count, such as Cesàro convergence over k when $T \asymp |k|^{1/d}$. See results on the Linnik problem in [Einsiedler–Lindenstrauss–Michel–Venkatesh 2011].

Theorem (Arala–Getz–Hou–Hsu–Li–W. 2024)

Let $D/\mathbb{Q}$ be a quaternion algebra ramified at $S \supseteq \{2, \infty\}$. Fix a maximal order $\mathcal{O}_D \subset D$ and a function $w \in C_c^\infty(D^n \otimes \mathbb{R})$, where $n \geq 8$. Then for $\upsilon_1, \dots, \upsilon_n \in \{\pm 1\}$ and $T \geq 1$,

$$\sum_{x \in \mathcal{O}_D^n \colon P(x) = 0} w(x/T) = c_{P,w} T^{4n-8} + O_{w,\epsilon}(T^{3n+\epsilon}),$$

where $P(x) := \upsilon_1 x_1^2 + \cdots + \upsilon_n x_n^2$. (Asymptotic for $n \geq 9$.)

Theorem (Arala–Getz–Hou–Hsu–Li–W. 2024)

Let $D/\mathbb{Q}$ be a quaternion algebra ramified at $S \supseteq \{2, \infty\}$. Fix a maximal order $\mathcal{O}_D \subset D$ and a function $w \in C_c^\infty(D^n \otimes \mathbb{R})$, where $n \geq 8$. Then for $\upsilon_1, \dots, \upsilon_n \in \{\pm 1\}$ and $T \geq 1$,

$$\sum_{x \in \mathcal{O}_D^n :\, P(x)=0} w(x/T) = c_{P,w} T^{4n-8} + O_{w,\epsilon}(T^{3n+\epsilon}),$$

where $P(x) := \upsilon_1 x_1^2 + \cdots + \upsilon_n x_n^2$. (Asymptotic for $n \geq 9$.)

Previously an asymptotic was available for $n \geq 17$, thanks to Myerson's 2018 strengthening of Birch's 1962 classical result.

Remark

The solutions to the aforementioned equation $X^d = I_d$ for $X \in M_d(\mathbb{Z})$ break up into finitely many $\mathrm{GL}_d(\mathbb{Z})$-conjugation orbits. But the equations $XY = YX$ and $P(x) = 0$ seem to lack such a nearly-transitive group action.

For concreteness, one could take $D = \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k$ to be Hamilton's classical quaternions, with

$$i^2 = j^2 = k^2 = ijk = -1 \quad \text{(Broome Bridge, Dublin, 1843)},$$

and $\mathcal{O}_D = \mathbb{Z}\frac{1+i+j+k}{2} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ the Hurwitz quaternions (1919).

For concreteness, one could take $D = \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k$ to be Hamilton's classical quaternions, with

$$i^2 = j^2 = k^2 = ijk = -1 \quad \text{(Broome Bridge, Dublin, 1843)},$$

and $\mathcal{O}_D = \mathbb{Z}\frac{1+i+j+k}{2} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ the Hurwitz quaternions (1919). Issues with zerodivisors currently prevent us from taking $D = M_2(\mathbb{Q})$ and $\mathcal{O}_D = M_2(\mathbb{Z})$ in the previous theorem. However, we have the following level-of-distribution result:

Theorem (Arala–W. 2025+)

Let $d \in \{2, 3\}$ and $w \in C_c^\infty(M_d(\mathbb{R})^n)$. If $b, r \in M_d(\mathbb{Z})$ and $T \asymp |r| > 0$ with $|\det(r)|$ prime and $|\det(r)| \asymp |r|^d$, then

$$\sum_{\substack{x \in M_d(\mathbb{Z})^n \\ x_1^2 + \cdots + x_n^2 - b \in rM_d(\mathbb{Z})}} w(x/T) = \frac{c_w T^{d^2 n}}{|\det(r)|^d} + O_{w,\epsilon}(T^{(d^2 - \frac{d}{2})n + \epsilon}).$$

Asymptotic for $n \geq 2d + 1$. Previously $n > d^2(d^2+1)/(2d-2)$ available by [Birch 1962, Yamagishi 2023]?

Rough idea of the algebraic circle method

Let A be a free $\mathbb{Z}$-module of finite rank. Fix a $\mathbb{Z}$-bilinear[2] map $\mu\colon A \times A \to A$, a $\mathbb{Z}$-linear map $\operatorname{tr}\colon A \to \mathbb{Z}$, and a vector norm $|\cdot|\colon A \otimes \mathbb{R} \to \mathbb{R}_{\geq 0}$. Let $e(t) := e^{2\pi i t}$ for $t \in \mathbb{R}$. If $x \in A$, then

$$\begin{aligned} \mathbf{1}_{x=0} &= \int_{(A\otimes\mathbb{R})/A} e(\theta_1 x_1 + \cdots + \theta_{\operatorname{rank} A} x_{\operatorname{rank} A})\, d\theta \\ &= \int_{(A\otimes\mathbb{R})/A} e(\operatorname{tr}(\theta x))\, d\theta, \end{aligned}$$

provided that the pairing $\operatorname{tr} \circ \mu\colon A \times A \to \mathbb{Z}$ is perfect.[3]

[2] Thus A is *distributive*, but not necessarily commutative or associative.

[3] For non-degenerate $\operatorname{tr} \circ \mu$, the second $(A \otimes \mathbb{R})/A$ needs adjustment.

Rough idea of the algebraic circle method

Let A be a free $\mathbb{Z}$-module of finite rank. Fix a $\mathbb{Z}$-bilinear[2] map $\mu\colon A \times A \to A$, a $\mathbb{Z}$-linear map $\operatorname{tr}\colon A \to \mathbb{Z}$, and a vector norm $|\cdot|\colon A \otimes \mathbb{R} \to \mathbb{R}_{\geq 0}$. Let $e(t) := e^{2\pi i t}$ for $t \in \mathbb{R}$. If $x \in A$, then

$$\begin{aligned}\mathbf{1}_{x=0} &= \int_{(A\otimes\mathbb{R})/A} e(\theta_1 x_1 + \cdots + \theta_{\operatorname{rank} A} x_{\operatorname{rank} A})\, d\theta \\ &= \int_{(A\otimes\mathbb{R})/A} e(\operatorname{tr}(\theta x))\, d\theta,\end{aligned}$$

provided that the pairing $\operatorname{tr} \circ \mu\colon A \times A \to \mathbb{Z}$ is perfect.[3]

Proposition (Algebraic Dirichlet-type covering)

Let $\theta \in A \otimes \mathbb{R}$ and $Q \geq 1$. Then there exists $(a, r) \in A^2$ such that $0 \neq |r| \ll Q$ and $|\theta r - a| \ll 1/Q$.

[2] Thus A is *distributive*, but not necessarily commutative or associative.

[3] For non-degenerate $\operatorname{tr} \circ \mu$, the second $(A \otimes \mathbb{R})/A$ needs adjustment.

Quadratic Weyl sums over $\mathbb{Z}$ (classical)

Fix $w \in C_c^\infty(\mathbb{R})$. Let $a, r \in \mathbb{Z} \setminus \{0\}$ with $\gcd(a, r) = 1$. Let $1 \leq T \leq |r|$. Writing $e(\theta) := e^{2\pi i\theta}$ for $\theta \in \mathbb{R}$, we have

$$\Sigma_T(a/r) := \sum_{x \in \mathbb{Z}} w(x/T)e(ax^2/r) = \sum_{c \in \mathbb{Z}} I_r(c)S_{a,r}(c)$$

by Poisson summation, where $I_r(c) = \int_{\mathbb{R}} w(x/T)e(-cx/r)\,dx$ and $S_{a,r}(c) = \frac{1}{r}\sum_{x \in \mathbb{Z}/r\mathbb{Z}} e(\frac{ax^2+cx}{r})$.

- Integration by parts: $I_r(c) \ll_A \frac{T}{|Tc/r|^A}$ for all $A > 0$.
- Squaring and differencing: $S_{a,r}(c) \ll \frac{1}{|r|^{1/2}}$ (Gauss).

Quadratic Weyl sums over $\mathbb{Z}$ (classical)

Fix $w \in C_c^\infty(\mathbb{R})$. Let $a, r \in \mathbb{Z} \setminus \{0\}$ with $\gcd(a, r) = 1$. Let $1 \le T \le |r|$. Writing $e(\theta) := e^{2\pi i\theta}$ for $\theta \in \mathbb{R}$, we have

$$\Sigma_T(a/r) := \sum_{x\in\mathbb{Z}} w(x/T)e(ax^2/r) = \sum_{c\in\mathbb{Z}} I_r(c)S_{a,r}(c)$$

by Poisson summation, where $I_r(c) = \int_{\mathbb{R}} w(x/T)e(-cx/r)\,dx$ and $S_{a,r}(c) = \frac{1}{r}\sum_{x\in\mathbb{Z}/r\mathbb{Z}} e(\frac{ax^2+cx}{r})$.

- Integration by parts: $I_r(c) \ll_A \frac{T}{|Tc/r|^A}$ for all $A > 0$.
- Squaring and differencing: $S_{a,r}(c) \ll \frac{1}{|r|^{1/2}}$ (Gauss).

Thus $\Sigma_T(a/r) \ll_A \frac{T}{|r|^{1/2}} \sum_{c\in\mathbb{Z}} \min(1, |Tc/r|^{-A}) \ll_A \frac{T}{|r|^{1/2}}|r/T|$, essentially coming from $|c| \le |r/T|$. So: $\Sigma_T(a/r) \ll |r|^{1/2}$. (I think this is essentially sharp for all $T \ge r^{1/2+\epsilon}$, if $a = 1$.)

- This is square-root cancellation if $T \asymp |r|$.
- The sums $\Sigma_T(\theta)$, for $\theta \in \mathbb{R}/\mathbb{Z}$, appear in problems such as counting integer solutions to quadratic equations.

Quadratic Weyl sums over $\mathbb{Z}[i]$ (classical)

Fix $w \in C_c^\infty(\mathbb{C})$. Let $a, r \in \mathbb{Z}[i] \setminus \{0\}$ with $|\gcd(a, r)| = 1$. Let $1 \leq T \leq |r|$. Directly adapting to $\mathbb{Z}[i]$ the slide for $\mathbb{Z}$ gives

$$\Sigma_T(a/r) := \sum_{x \in \mathbb{Z}[i]} w(x/T) e(\mathrm{tr}(ar^{-1}x^2)) \ll |\mathbb{Z}[i]/r\mathbb{Z}[i]|^{1/2} = |r|.$$

Again, this is square-root cancellation over x if $T \asymp |r|$.

Quadratic Weyl sums over $\mathbb{Z}[i]$ (classical)

Fix $w \in C_c^\infty(\mathbb{C})$. Let $a, r \in \mathbb{Z}[i] \setminus \{0\}$ with $|\gcd(a, r)| = 1$. Let $1 \leq T \leq |r|$. Directly adapting to $\mathbb{Z}[i]$ the slide for $\mathbb{Z}$ gives

$$\Sigma_T(a/r) := \sum_{x \in \mathbb{Z}[i]} w(x/T) e(\mathrm{tr}(ar^{-1}x^2)) \ll |\mathbb{Z}[i]/r\mathbb{Z}[i]|^{1/2} = |r|.$$

Again, this is square-root cancellation over x if $T \asymp |r|$.

- Key to this generalization is that $r \in \mathit{Center}(\mathbb{Z}[i])$, so that $e(\mathrm{tr}(ar^{-1}x^2))$ depends only on x mod $r \in \mathbb{Z}[i]/r\mathbb{Z}[i]$.
- However, as an exponential sum over $\mathbb{Z}^2$, the quantity $\Sigma_T(a/r)$ has modulus $|r|^2$, rather than $|r|$. Thus, by packaging $\mathbb{Z}^2$ into $\mathbb{Z}[i]$, we are able to get square-root cancellation over x much shorter than the modulus.
- The sums $\Sigma_T(\theta)$ appear when counting solutions to quadratic equations in $\mathbb{Z}[i]$, or equivalently, to a *pair* of quadratic equations (the Weil restriction) in $\mathbb{Z}$.

Weil restriction in analytic number theory

- For the general story over rings of integers $\mathcal{O}_K$ of global fields K, see e.g. [Skinner 1997, Browning–Vishe 2014].
- Difficult variants of the packaging idea include *generalized quadratic forms* over K, which involve conjugated variables $\sigma(x)$, and are not just quadratic forms over the number field [Browning–Pierce–Schindler 2022].

Weil restriction in analytic number theory

- For the general story over rings of integers $\mathcal{O}_K$ of global fields K, see e.g. [Skinner 1997, Browning–Vishe 2014].
- Difficult variants of the packaging idea include *generalized quadratic forms* over K, which involve conjugated variables $\sigma(x)$, and are not just quadratic forms over the number field [Browning–Pierce–Schindler 2022].

Related examples of Weil restriction or similar packaging:

- Solving $x^3 = y^2 + 2$ (Fermat; Euler using $\mathbb{Z}[\sqrt{-2}]$).
- Skolem's method for Thue equations like $x^3 + 2y^3 = k$.
- Prime values of *restricted norm forms* like $x^2 + y^4$ or $x^3 + 2y^3$; see e.g. [Friedlander–Iwaniec 1998, Heath-Brown 2001, . . . , Maynard 2020, Green–Sawhney 2024].
- Linear spaces on hypersurfaces [Brandes 2014].
- Counting $\mathbb{F}_q[t][s]/s^{m+1}$- and $\mathbb{F}_q[t][s,r]/(s^{m+1}, r^2)$-points on hypersurfaces to study singularities on the moduli spaces of curves thereof [Glas–Hase-Liu 2024].

Quadratic Weyl sums over $\mathbb{Z}\langle i, j\rangle$ $(i^2 = j^2 = -1)$

Let $\mathbb{L} = \mathbb{Z}\langle i, j\rangle = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$. Fix $w \in C_c^\infty(\mathbb{L} \otimes \mathbb{R})$. Given $x = x_1 + x_2 i + x_3 j + x_4 k$, let $x^\dagger := x_1 - x_2 i - x_3 j - x_4 k$, $\operatorname{trd}(x) := x^\dagger + x = 2x_1$, and $\operatorname{nrd}(x) := x^\dagger x = x_1^2 + \cdots + x_4^2$.

Theorem (Arala–Getz–Hou–Hsu–Li–W. 2024)

Let $a, r \in \mathbb{L} \setminus \{0\}$ *with* $\gcd_{\mathbb{Z}}(ar^\dagger, \operatorname{nrd}(r)) \asymp \gcd_{\mathbb{Z}}(r)$,[a] *where* gcd *is computed in* $\mathbb{Z}^5$ *and in* $\mathbb{Z}^4$*, respectively. If* $T \asymp |r|$*, then*

$$\Sigma_T(ar^{-1}) := \sum_{x \in \mathbb{L}} w(x/T) e(\operatorname{trd}(ar^{-1}x^2)) \ll_\epsilon T^{3+\epsilon}.$$

[a] For example, take $\operatorname{nrd}(r)$ square-free and $\gcd(\operatorname{nrd}(a), \operatorname{nrd}(r)) = 1$.

Quadratic Weyl sums over $\mathbb{Z}\langle i, j\rangle$ ($i^2 = j^2 = -1$)

Let $\mathbb{L} = \mathbb{Z}\langle i, j\rangle = \mathbb{Z} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$. Fix $w \in C_c^\infty(\mathbb{L} \otimes \mathbb{R})$. Given $x = x_1 + x_2 i + x_3 j + x_4 k$, let $x^\dagger := x_1 - x_2 i - x_3 j - x_4 k$, $\operatorname{trd}(x) := x^\dagger + x = 2x_1$, and $\operatorname{nrd}(x) := x^\dagger x = x_1^2 + \cdots + x_4^2$.

Theorem (Arala–Getz–Hou–Hsu–Li–W. 2024)

Let $a, r \in \mathbb{L} \setminus \{0\}$ with $\gcd_{\mathbb{Z}}(ar^\dagger, \operatorname{nrd}(r)) \asymp \gcd_{\mathbb{Z}}(r)$,[a] *where* gcd *is computed in $\mathbb{Z}^5$ and in $\mathbb{Z}^4$, respectively. If $T \asymp |r|$, then*

$$\Sigma_T(ar^{-1}) := \sum_{x \in \mathbb{L}} w(x/T) e(\operatorname{trd}(ar^{-1}x^2)) \ll_\epsilon T^{3+\epsilon}.$$

[a] For example, take $\operatorname{nrd}(r)$ square-free and $\gcd(\operatorname{nrd}(a), \operatorname{nrd}(r)) = 1$.

Are there near-equality cases? If $\operatorname{trd}(ar^{-1}) \in \mathbb{Z}$, or equivalently $\operatorname{trd}(ar^\dagger) \equiv 0 \bmod \operatorname{nrd}(r)$, then the contribution to $\Sigma_T(ar^{-1})$ from $\operatorname{trd}(x) = 0$ is of size T^3 (no oscillation).

Theorem (Arala–Getz–Hou–Hsu–Li–W. 2024)

Let $a, r \in \mathbb{L} \setminus \{0\}$ *with* $\gcd_{\mathbb{Z}}(ar^\dagger, \mathrm{nrd}(r)) \asymp \gcd_{\mathbb{Z}}(r)$. *If* $T \asymp |r|$, *then*

$$\Sigma_T(ar^{-1}) := \sum_{x \in \mathbb{L}} w(x/T) e(\mathrm{trd}(ar^{-1}x^2)) \ll_\epsilon T^{3+\epsilon}.$$

The proof uses Fourier analysis, Cartan decomposition, matrix identities, Gauss sums, and the geometry of numbers. Crucially, the vector $ar^{-1} \in \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k$ is rather special:

$$ar^{-1} = \frac{ar^\dagger}{\mathrm{nrd}(r)} = \frac{b_1 + b_2 i + b_3 j + b_4 k}{\mathrm{nrd}(r)},$$

say, where $(b_1, b_2, b_3, b_4) \in \mathbb{Z}^4$ satisfies

$$b_1^2 + \cdots + b_4^2 = \mathrm{nrd}(ar^\dagger) \equiv 0 \bmod \mathrm{nrd}(r).$$

Contrast with the classical 4-dimensional circle method, whose fractions have smaller denominator but lack algebraic structure.

Poisson summation

Let $a, r \in \mathbb{L} \setminus \{0\}$ and $T \asymp |r|$. Then

$$\Sigma_T(ar^{-1}) := \sum_{x \in \mathbb{L}} w(x/T) e(\mathrm{trd}(ar^{-1}x^2)) = \sum_{c \in \mathbb{L}} I_r(c) S_{a,r}(c)$$

by Poisson summation in $(\mathbb{Z}^4/\mathrm{nrd}(r)\mathbb{Z}^4) \times \mathbb{R}^4$, where

$$I_r(c) = \int_{\mathbb{L}\otimes\mathbb{R}} w(x/T) e(-\frac{\mathrm{trd}(cx)}{\mathrm{nrd}(r)})\, dx,$$

$$S_{a,r}(c) = \frac{1}{\mathrm{nrd}(r)^4} \sum_{x \in \mathbb{L}/\mathrm{nrd}(r)\mathbb{L}} e(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{\mathrm{nrd}(r)}).$$

Since $r \notin \mathit{Center}(\mathbb{L})$, the sum in $S_{a,r}(c)$ is genuinely over $\mathbb{L}/\mathrm{nrd}(r)\mathbb{L}$, not just over $\mathbb{L}/r\mathbb{L}$.

Poisson summation

Let $a, r \in \mathbb{L} \setminus \{0\}$ and $T \asymp |r|$. Then

$$\Sigma_T(ar^{-1}) := \sum_{x \in \mathbb{L}} w(x/T) e(\mathrm{trd}(ar^{-1}x^2)) = \sum_{c \in \mathbb{L}} I_r(c) S_{a,r}(c)$$

by Poisson summation in $(\mathbb{Z}^4/\mathrm{nrd}(r)\mathbb{Z}^4) \times \mathbb{R}^4$, where

$$I_r(c) = \int_{\mathbb{L}\otimes\mathbb{R}} w(x/T) e(-\frac{\mathrm{trd}(cx)}{\mathrm{nrd}(r)})\, dx,$$

$$S_{a,r}(c) = \frac{1}{\mathrm{nrd}(r)^4} \sum_{x \in \mathbb{L}/\mathrm{nrd}(r)\mathbb{L}} e(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{\mathrm{nrd}(r)}).$$

Since $r \notin \mathit{Center}(\mathbb{L})$, the sum in $S_{a,r}(c)$ is genuinely over $\mathbb{L}/\mathrm{nrd}(r)\mathbb{L}$, not just over $\mathbb{L}/r\mathbb{L}$. Integration by parts gives $I_r(c) \ll_A \frac{T^4}{|Tc/\mathrm{nrd}(r)|^A}$ for all $A > 0$. Thus we may pretend that

$$|c| \leq \mathrm{nrd}(r)/T.$$

But $\mathrm{nrd}(r)/T \asymp |r|^2/T \asymp T$. Should we give up?

Local estimates and vanishing phenomena

Proposition (Arala–Getz–Hou–Hsu–Li–W. 2024)

Assume $\gcd_{\mathbb{Z}}(r) = 1$ *and* $2 \nmid N = \mathrm{nrd}(r)$. *Let* $c \in \mathbb{L}$.

1. *If* $S_{a,r}(c) \neq 0$, *there exists* $c_0 = c_0(a, r) \in \mathbb{L}$ *such that* $cr \in c_0 r\mathbb{Z} + N\mathbb{L}$. *(We can take* $c_0 = ar^\dagger b_0$ *for any sufficiently generic* $b_0 = b_0(r) \in \mathbb{L}$*.)*

Local estimates and vanishing phenomena

Proposition (Arala–Getz–Hou–Hsu–Li–W. 2024)

Assume $\gcd_{\mathbb{Z}}(r) = 1$ *and* $2 \nmid N = \mathrm{nrd}(r)$. *Let* $c \in \mathbb{L}$.

1. *If* $S_{a,r}(c) \neq 0$, *there exists* $c_0 = c_0(a, r) \in \mathbb{L}$ *such that* $cr \in c_0 r\mathbb{Z} + N\mathbb{L}$. *(We can take* $c_0 = ar^\dagger b_0$ *for any sufficiently generic* $b_0 = b_0(r) \in \mathbb{L}$.*)*
2. *Assume* $\gcd_{\mathbb{Z}}(ar^\dagger, N) = 1$. *Let* $K \geq 1$ *be the largest divisor of* N *such that* $(c - c^\dagger)r \in K\mathbb{L}$. *Then*

$$S_{a,r}(c) \ll \frac{K^{1/2}}{N^{3/2}}.$$

Thus the sum $S_{a,r}(c)$ is controlled by lattices of the form

$$\Lambda(K, r, c_0) := \{c \in \mathbb{L} : (c - c^\dagger)r \in K\mathbb{L},\ cr \in c_0 r\mathbb{Z} + N\mathbb{L}\}.$$

Proof of proposition (non-vanishing constraint)

Assume $\gcd_{\mathbb{Z}}(r) = 1$ and $2 \nmid N = \mathrm{nrd}(r)$. By definition,

$$S_{a,r}(c) = \frac{1}{N^4} \sum_{x \in \mathbb{L}/N\mathbb{L}} e\left(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{N}\right),$$

Since $r \notin Center(\mathbb{L})$, the sum over x is usually not $r\mathbb{L}$-periodic. To quantify the failure of periodicity, we replace x with $x + ry$ and average over $y \in \mathbb{L}$, getting

$$\begin{aligned} S_{a,r}(c) &= \frac{1}{N^8} \sum_{x,y \in \mathbb{L}/N\mathbb{L}} e\left(\frac{\mathrm{trd}(ar^\dagger x(x+ry) + c(x+ry))}{N}\right) \\ &= \frac{1}{N^4} \sum_{x \in \mathbb{L}/N\mathbb{L}\colon ar^\dagger xr + cr \equiv 0} e\left(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{N}\right). \end{aligned}$$

Proof of proposition (non-vanishing constraint)

Assume $\gcd_{\mathbb{Z}}(r) = 1$ and $2 \nmid N = \mathrm{nrd}(r)$. By definition,

$$S_{a,r}(c) = \frac{1}{N^4} \sum_{x \in \mathbb{L}/N\mathbb{L}} e(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{N}),$$

Since $r \notin Center(\mathbb{L})$, the sum over x is usually not $r\mathbb{L}$-periodic. To quantify the failure of periodicity, we replace x with $x + ry$ and average over $y \in \mathbb{L}$, getting

$$\begin{aligned} S_{a,r}(c) &= \frac{1}{N^8} \sum_{x,y \in \mathbb{L}/N\mathbb{L}} e(\frac{\mathrm{trd}(ar^\dagger x(x+ry) + c(x+ry))}{N}) \\ &= \frac{1}{N^4} \sum_{x \in \mathbb{L}/N\mathbb{L}:\, ar^\dagger xr + cr \equiv 0} e(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{N}). \end{aligned}$$

Some pair of $\mathbb{Z}$-module isomorphisms $\mathbb{L}/N\mathbb{L} \to M_2(\mathbb{Z}/N\mathbb{Z})$ sends the map $x \mapsto r^\dagger x r$ to $m \mapsto \left[\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right] m \left[\begin{smallmatrix} 1 & 0 \\ 0 & N \end{smallmatrix}\right] \equiv \left[\begin{smallmatrix} 0 & 0 \\ m_{21} & 0 \end{smallmatrix}\right]$. Thus $x \mapsto r^\dagger x r$ has image $r^\dagger b_0 r\mathbb{Z}$ mod $N\mathbb{L}$ for some $b_0 \in \mathbb{L}$...

Proof of proposition (further cancellation)

Assume $S_{a,r}(c) \neq 0$. From the previous slide, we have

$$S_{a,r}(c) = \frac{1}{N^4} \sum_{x \in \mathbb{L}/N\mathbb{L}\colon ar^\dagger xr + cr \equiv 0} e\Big(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{N}\Big),$$

which vanishes (empty sum) unless $cr \in ar^\dagger b_0 r\mathbb{Z} + N\mathbb{L}$. So

$$\#\{x \in \mathbb{L}/N\mathbb{L} : \ ar^\dagger xr + cr \equiv 0\} = \# \ker(x \mapsto ar^\dagger xr).$$

If $\gcd_{\mathbb{Z}}(ar^\dagger, N) = 1$, then $ar^\dagger$ and $r^\dagger$ lie in the same Cartan decomposition class modulo N, so

$$\# \ker(x \mapsto ar^\dagger xr) = \# \ker(x \mapsto r^\dagger xr) = \frac{N^4}{\# \operatorname{im}(x \mapsto r^\dagger xr)} = N^3.$$

Proof of proposition (further cancellation)

Assume $S_{a,r}(c) \neq 0$. From the previous slide, we have

$$S_{a,r}(c) = \frac{1}{N^4} \sum_{x \in \mathbb{L}/N\mathbb{L}\colon ar^\dagger xr + cr \equiv 0} e\Big(\frac{\mathrm{trd}(ar^\dagger x^2 + cx)}{N}\Big),$$

which vanishes (empty sum) unless $cr \in ar^\dagger b_0 r\mathbb{Z} + N\mathbb{L}$. So

$$\#\{x \in \mathbb{L}/N\mathbb{L} : \ ar^\dagger xr + cr \equiv 0\} = \#\ker(x \mapsto ar^\dagger xr).$$

If $\gcd_{\mathbb{Z}}(ar^\dagger, N) = 1$, then $ar^\dagger$ and $r^\dagger$ lie in the same Cartan decomposition class modulo N, so

$$\#\ker(x \mapsto ar^\dagger xr) = \#\ker(x \mapsto r^\dagger xr) = \frac{N^4}{\#\operatorname{im}(x \mapsto r^\dagger xr)} = N^3.$$

To improve on the triangle inequality $|S_{a,r}(c)| \leq \frac{1}{N}$, replace x with $x + k$ and average over $k \in \mathbb{Z}$. If $K = \gcd(\mathrm{trd}(ar^\dagger), N)$,

$$S_{a,r}(c) \ll \frac{\mathbf{1}_{\exists x \in \mathbb{L},\ ar^\dagger xr + cr \in N\mathbb{L},\ \mathrm{trd}(2ar^\dagger x + c) \in K\mathbb{Z}}}{(N/K)^{1/2} N} \quad \text{(Gauss)}.$$

Proof of proposition (lattice simplification)

From the previous slide, if $K = \gcd(\operatorname{trd}(ar^\dagger), N)$,

$$S_{a,r}(c) \ll \frac{\mathbf{1}_{\exists x \in \mathbb{L},\, ar^\dagger x r + cr \in N\mathbb{L},\, \operatorname{trd}(2ar^\dagger x + c) \in K\mathbb{Z}}}{(N/K)^{1/2} N} \quad \text{(Gauss)}.$$

Since $ar^\dagger + ra^\dagger = \operatorname{trd}(ar^\dagger) \in K\mathbb{L}$, we find, on replacing $ar^\dagger$ with $-ra^\dagger$ in the conditions above, that

$$cr \in ra^\dagger x r + K\mathbb{L}, \quad \operatorname{trd}(c) \in \operatorname{trd}(2ra^\dagger x) + K\mathbb{Z}.$$

Proof of proposition (lattice simplification)

From the previous slide, if $K = \gcd(\operatorname{trd}(ar^\dagger), N)$,

$$S_{a,r}(c) \ll \frac{\mathbf{1}_{\exists x \in \mathbb{L},\ ar^\dagger xr + cr \in N\mathbb{L},\ \operatorname{trd}(2ar^\dagger x + c) \in K\mathbb{Z}}}{(N/K)^{1/2} N} \quad \text{(Gauss)}.$$

Since $ar^\dagger + ra^\dagger = \operatorname{trd}(ar^\dagger) \in K\mathbb{L}$, we find, on replacing $ar^\dagger$ with $-ra^\dagger$ in the conditions above, that

$$cr \in ra^\dagger xr + K\mathbb{L}, \quad \operatorname{trd}(c) \in \operatorname{trd}(2ra^\dagger x) + K\mathbb{Z}.$$

Right-multiplying the latter by r, we get

$$(c + c^\dagger)r \equiv (2ra^\dagger x + 2x^\dagger ar^\dagger)r \equiv 2ra^\dagger xr \bmod K\mathbb{L}.$$

The rightmost term is $\equiv 2cr \bmod K\mathbb{L}$, so we conclude that

$$(c^\dagger - c)r \equiv 0 \bmod K\mathbb{L}.$$

Thus if $c_0 = c_0(a, r) := ar^\dagger b_0$, then c lies in the lattice

$$\Lambda(K, r, c_0) := \{c \in \mathbb{L} : (c - c^\dagger)r \in K\mathbb{L},\ cr \in c_0 r\mathbb{Z} + N\mathbb{L}\},$$

since $ar^\dagger xr + cr \in N\mathbb{L} \Rightarrow cr \in ar^\dagger b_0 r\mathbb{Z} + N\mathbb{L}$ (from earlier).

Applying the proposition

Earlier we showed something like

$$\Sigma_T(ar^{-1}) := \sum_{x\in\mathbb{L}} w(x/T)e(\mathrm{trd}(ar^{-1}x^2)) \ll \sum_{|c|\leq N/T} T^4|S_{a,r}(c)|.$$

The sum $S_{a,r}(c)$ is controlled by lattices of the form

$$\Lambda(K,r,c_0) := \{c \in \mathbb{L} : (c-c^\dagger)r \in K\mathbb{L},\ cr \in c_0 r\mathbb{Z} + N\mathbb{L}\},$$

for some $c_0 = c_0(a,r)$. Specifically, by the proposition,

$$\Sigma_T(ar^{-1}) \ll T^4 \sum_{\substack{K|N \\ c\in\Lambda(K,r,c_0) \\ |c|\leq N/T}} \frac{K^{1/2}}{N^{3/2}}.$$

Applying the proposition

Earlier we showed something like

$$\Sigma_T(ar^{-1}) := \sum_{x \in \mathbb{L}} w(x/T) e(\mathrm{trd}(ar^{-1}x^2)) \ll \sum_{|c| \leq N/T} T^4 |S_{a,r}(c)|.$$

The sum $S_{a,r}(c)$ is controlled by lattices of the form

$$\Lambda(K, r, c_0) := \{c \in \mathbb{L} : (c - c^\dagger)r \in K\mathbb{L},\ cr \in c_0 r\mathbb{Z} + N\mathbb{L}\},$$

for some $c_0 = c_0(a, r)$. Specifically, by the proposition,

$$\Sigma_T(ar^{-1}) \ll T^4 \sum_{\substack{K|N \\ c \in \Lambda(K,r,c_0) \\ |c| \leq N/T}} \frac{K^{1/2}}{N^{3/2}}.$$

It remains to analyze the lattice $\Lambda(K, r, c_0)$ for each $K \mid N$.

Geometry of numbers

Consider the lattices

$$\Lambda(K, r, c_0) := \{c \in \mathbb{L} : (c - c^\dagger)r \in K\mathbb{L},\ cr \in c_0 r\mathbb{Z} + N\mathbb{L}\}.$$

Lemma (Arala–Getz–Hou–Hsu–Li–W. 2024)

Suppose $K \mid N = \mathrm{nrd}(r)$, where $r \in \mathbb{L}$ is a primitive vector. Let $c_0 \in \mathbb{L}$. Then for all $B > 0$, we have

$$\#(\Lambda(K, r, c_0) \cap [-B, B]^4) \ll 1 + B + \frac{B^2}{K^{1/2}} + \frac{B^3}{(KN)^{1/2}} + \frac{B^4}{KN}.$$

Geometry of numbers

Consider the lattices

$$\Lambda(K, r, c_0) := \{c \in \mathbb{L} : (c - c^\dagger)r \in K\mathbb{L},\ cr \in c_0 r\mathbb{Z} + N\mathbb{L}\}.$$

Lemma (Arala–Getz–Hou–Hsu–Li–W. 2024)

Suppose $K \mid N = \mathrm{nrd}(r)$, where $r \in \mathbb{L}$ is a primitive vector. Let $c_0 \in \mathbb{L}$. Then for all $B > 0$, we have

$$\#(\Lambda(K, r, c_0) \cap [-B, B]^4) \ll 1 + B + \frac{B^2}{K^{1/2}} + \frac{B^3}{(KN)^{1/2}} + \frac{B^4}{KN}.$$

Proof strategy.

It suffices to lower-bound partial products of successive minima λ_i [Schmidt 1968]. Use lower bounds $\lambda_1 \gg 1$, $\lambda_2 \gg K^{1/2}$, and $\lambda_1\lambda_2\lambda_3\lambda_4 \asymp |\mathbb{L}/\Lambda| = \frac{N^4}{|\Lambda/N\mathbb{L}|} \gg KN$, combined with the upper bound $\lambda_4 \ll (KN)^{1/2}$ to lower-bound $\lambda_1\lambda_2\lambda_3$. □

Earlier we showed something like

$$\Sigma_T(ar^{-1}) := \sum_{x\in\mathbb{L}} w(x/T)e(\mathrm{trd}(ar^{-1}x^2)) \ll T^4 \sum_{\substack{K|N \\ c\in\Lambda(K,r,c_0) \\ |c|\leq N/T}} \frac{K^{1/2}}{N^{3/2}}.$$

By the lemma,

$$\Sigma_T(ar^{-1}) \ll T^4 \sum_{K|N} \frac{K^{1/2}}{N^{3/2}}(\frac{N}{T} + \frac{(N/T)^2}{K^{1/2}} + \frac{(N/T)^3}{(KN)^{1/2}} + \frac{(N/T)^4}{KN}).$$

Earlier we showed something like

$$\Sigma_T(ar^{-1}) := \sum_{x\in\mathbb{L}} w(x/T)e(\mathrm{trd}(ar^{-1}x^2)) \ll T^4 \sum_{\substack{K|N \\ c\in\Lambda(K,r,c_0) \\ |c|\leq N/T}} \frac{K^{1/2}}{N^{3/2}}.$$

By the lemma,

$$\Sigma_T(ar^{-1}) \ll T^4 \sum_{K|N} \frac{K^{1/2}}{N^{3/2}}(\frac{N}{T}+\frac{(N/T)^2}{K^{1/2}}+\frac{(N/T)^3}{(KN)^{1/2}}+\frac{(N/T)^4}{KN}).$$

Summing over K using the divisor bound gives

$$\Sigma_T(ar^{-1}) \ll_\epsilon T^4\frac{N^\epsilon}{N^{3/2}}(N^{1/2}\frac{N}{T}+(N/T)^2+\frac{(N/T)^3}{N^{1/2}}+\frac{(N/T)^4}{N}).$$

Since $N = \mathrm{nrd}(r) \asymp |r|^2 \asymp T^2$, it follows that

$$\Sigma_T(ar^{-1}) \ll_\epsilon T^{3+\epsilon}.$$

(The proof when the quantity $\gcd_{\mathbb{Z}}(r)$ is large, rather than 1, is more technical but still doable.)

Theorem (Arala–Getz–Hou–Hsu–Li–W. 2024)

Let $D/\mathbb{Q}$ be a quaternion algebra ramified at $S \supseteq \{2, \infty\}$. Fix a maximal order $\mathcal{O}_D \subset D$ and a function $w \in C_c^\infty(D^n \otimes \mathbb{R})$, where $n \geq 8$. Then for $\upsilon_1, \dots, \upsilon_n \in \{\pm 1\}$ and $T \geq 1$,

$$\sum_{x \in \mathcal{O}_D^n \colon P(x)=0} w(x/T) = c_{P,w} T^{4n-8} + O_{w,\epsilon}(T^{3n+\epsilon}),$$

where $P(x) := \upsilon_1 x_1^2 + \cdots + \upsilon_n x_n^2$. (Asymptotic for $n \geq 9$.)

Estimates like $\Sigma_T(ar^{-1}) \ll_\epsilon T^{3+\epsilon}$ are half the proof. To obtain the main term $c_{P,w} T^{4n-8}$, we need to estimate sums roughly of the shape (built out of the sums $S_{a,r}(0)$)

$$\frac{T^{4n}}{(T^2)^4} \sum_{0 \neq r \ll T} \sum_{\substack{a \in \mathbb{L}/r\mathbb{L} \\ \gcd_{\mathbb{Z}}(ar^\dagger, N) \asymp \gcd_{\mathbb{Z}}(r)}} \frac{1}{N^{4n}} \sum_{x \in (\mathbb{L}/N\mathbb{L})^n} e\left(\frac{\operatorname{trd}(ar^\dagger P(x))}{N}\right)$$

where $N = \operatorname{nrd}(r)$.

To obtain the main term $c_{P,w}T^{4n-8}$, we need to estimate sums roughly of the shape (built out of the sums $S_{a,r}(0)$)

$$\frac{T^{4n}}{(T^2)^4}\sum_{0\neq r\ll T}\sum_{\substack{a\in\mathbb{L}/r\mathbb{L}\\ \gcd_{\mathbb{Z}}(ar^\dagger,N)\asymp\gcd_{\mathbb{Z}}(r)}}\frac{1}{N^{4n}}\sum_{x\in(\mathbb{L}/N\mathbb{L})^n}e(\frac{\mathrm{trd}(ar^\dagger P(x))}{N})$$

where $N=\mathrm{nrd}(r)$.

- This is done by spectrally expanding the sum over r in terms of suitably (maximally) invariant automorphic representations on $(\mathbb{L}\otimes\mathbf{A}_{\mathbb{Q}})^\times$; the invariance is maximal because there is no cx term in $S_{a,r}(c)$ for $c=0$.

To obtain the main term $c_{P,w}T^{4n-8}$, we need to estimate sums roughly of the shape (built out of the sums $S_{a,r}(0)$)

$$\frac{T^{4n}}{(T^2)^4}\sum_{0\neq r\ll T}\sum_{\substack{a\in\mathbb{L}/r\mathbb{L}\\ \gcd_{\mathbb{Z}}(ar^\dagger,N)\asymp\gcd_{\mathbb{Z}}(r)}}\frac{1}{N^{4n}}\sum_{x\in(\mathbb{L}/N\mathbb{L})^n}e\left(\frac{\mathrm{trd}(ar^\dagger P(x))}{N}\right)$$

where $N=\mathrm{nrd}(r)$.

- This is done by spectrally expanding the sum over r in terms of suitably (maximally) invariant automorphic representations on $(\mathbb{L}\otimes\mathbf{A}_{\mathbb{Q}})^\times$; the invariance is maximal because there is no cx term in $S_{a,r}(c)$ for $c=0$.
- The trivial representation leads to an Euler product resembling $\zeta_{\mathbb{L}\otimes\mathbb{Q}}(s+\frac{3}{2})$ (with a simple pole at $s=0$), whereas the nontrivial representations are put into the error term using the Jacquet–Langlands correspondence (including the fact that the trivial representation for a ramified local quaternion algebra corresponds to the Steinberg representation on GL_2).

What about matrices?

- The story for $P(x) = 0$ in $M_2(\mathbb{Z})$ is likely quite different than that for $\mathbb{L}$.
- However, a Duke–Friedlander–Iwaniec type delta symbol expansion[4] may well allow one to count solutions to $\det(P(x)) = 0$; note that $\mathrm{nrd}(P(x)) = 0 \Leftrightarrow P(x) = 0$ in a division algebra, but not in a split matrix algebra.
- If so, that might involve $\zeta_{M_2(\mathbb{Q})}(s + \frac{3}{2}) = \zeta(s+2)\zeta(s+1)$ (with simple poles at $s = -1, 0$). The additional pole may lead to a main term of size T^{4n-6} rather than T^{4n-8}.[5]

[4] like what we used in [Arala–Getz–Hou–Hsu–Li–W. 2024] for technical convenience, although the present slides are written more classically

[5] To explain where this comes from would require reworking our previous discussion to account for differences between circle method and delta method setups (the latter involves a difference of two un-sieved divisor problems, thus requiring additional cancellation of poles).

What about bigger algebras? Let $d \in \{2,3\}$, $r \in M_d(\mathbb{Z})$, and $T \asymp |r| > 0$, with $|\det(r)|$ prime and $|\det(r)| \asymp |r|^d$.

Theorem (Arala–W. 2025+)

Let $w \in C_c^\infty(M_d(\mathbb{R}))$. If $a \in M_d(\mathbb{Z}) \setminus M_d(\mathbb{Z})r$, then

$$\Sigma_T(ar^{-1}) := \sum_{x \in M_d(\mathbb{Z})} w(x/T) e(\operatorname{tr}(ar^{-1}x^2)) \ll_\epsilon T^{d^2 - \frac{d}{2} + \epsilon}.$$

Averaging over $a \in M_d(\mathbb{Z})/M_d(\mathbb{Z})r$ ("polygon method") gives:

Theorem (Arala–W. 2025+)

Let $w \in C_c^\infty(M_d(\mathbb{R})^n)$. If $b \in M_d(\mathbb{Z})$, then

$$\sum_{\substack{x \in M_d(\mathbb{Z})^n \\ x_1^2 + \cdots + x_n^2 - b \in rM_d(\mathbb{Z})}} w(x/T) = \frac{c_w T^{d^2 n}}{|\det(r)|^d} + O_{w,\epsilon}(T^{(d^2 - \frac{d}{2})n + \epsilon}).$$

Generalizing from $d = 2$

Let $N = \det(r) \asymp |r|^d \asymp T^d$. We have something like

$$\Sigma_T(ar^{-1}) := \sum_{x \in M_d(\mathbb{Z})} w(x/T) e(\operatorname{tr}(ar^{-1}x^2)) \ll \sum_{|c| \leq N/T} T^{d^2} |S_{a,r}(c)|$$

by Poisson summation in $M_d(\mathbb{Z}/N\mathbb{Z}) \times M_d(\mathbb{R})$, where

$$S_{a,r}(c) = \frac{1}{N^{d^2}} \sum_{x \in M_d(\mathbb{Z}/N\mathbb{Z})} e(\frac{\operatorname{tr}(a \operatorname{adj}(r) x^2 + cx)}{N}),$$

where $\operatorname{adj}(r) r = N$. Averaging over shifts $x \mapsto x + ry$ gives

$$S_{a,r}(c) = \frac{1}{N^{d^2}} \sum_{x \in M_d(\mathbb{Z}/N\mathbb{Z}) \colon a \operatorname{adj}(r) x r + cr \equiv 0} e(\frac{\operatorname{tr}(a \operatorname{adj}(r) x^2 + cx)}{N}).$$

By Cartan decomposition, $\# \operatorname{im}(x \mapsto \operatorname{adj}(r) x r) = N^{d-1}$.

Generalizing from $d = 2$ (further cancellation)

Assume $S_{a,r}(c) \neq 0$. From the previous slide, we have

$$S_{a,r}(c) = \frac{1}{N^{d^2}} \sum_{x \in M_d(\mathbb{Z}/N\mathbb{Z}):\, a\,\mathrm{adj}(r)xr+cr\equiv 0} e\left(\frac{\mathrm{tr}(a\,\mathrm{adj}(r)x^2 + cx)}{N}\right),$$

which vanishes unless $cr \in a\,\mathrm{adj}(r)M_d(\mathbb{Z})r + NM_d(\mathbb{Z})$. So

$$\#\{x \in M_d(\mathbb{Z}/N\mathbb{Z}) :\ a\,\mathrm{adj}(r)xr+cr \equiv 0\} = \#\ker(x \mapsto a\,\mathrm{adj}(r)xr).$$

But $0 \neq \mathrm{rank}(a\,\mathrm{adj}(r) \bmod N) \leq \mathrm{rank}(\mathrm{adj}(r) \bmod N) = 1$, so $a\,\mathrm{adj}(r)$ and $\mathrm{adj}(r)$ lie in the same Cartan decomposition class modulo N, so

$$\#\ker(x \mapsto a\,\mathrm{adj}(r)xr) = \#\ker(x \mapsto \mathrm{adj}(r)xr) = \frac{N^{d^2}}{N^{d-1}}.$$

Generalizing from $d = 2$ (further cancellation)

Assume $S_{a,r}(c) \neq 0$. From the previous slide, we have

$$S_{a,r}(c) = \frac{1}{N^{d^2}} \sum_{x \in M_d(\mathbb{Z}/N\mathbb{Z}) :\, a\,\mathrm{adj}(r)xr + cr \equiv 0} e\big(\frac{\mathrm{tr}(a\,\mathrm{adj}(r)x^2 + cx)}{N}\big),$$

which vanishes unless $cr \in a\,\mathrm{adj}(r)M_d(\mathbb{Z})r + NM_d(\mathbb{Z})$. So

$\#\{x \in M_d(\mathbb{Z}/N\mathbb{Z}) :\ a\,\mathrm{adj}(r)xr + cr \equiv 0\} = \#\ker(x \mapsto a\,\mathrm{adj}(r)xr)$.

But $0 \neq \mathrm{rank}(a\,\mathrm{adj}(r) \bmod N) \leq \mathrm{rank}(\mathrm{adj}(r) \bmod N) = 1$, so $a\,\mathrm{adj}(r)$ and $\mathrm{adj}(r)$ lie in the same Cartan decomposition class modulo N, so

$$\#\ker(x \mapsto a\,\mathrm{adj}(r)xr) = \#\ker(x \mapsto \mathrm{adj}(r)xr) = \frac{N^{d^2}}{N^{d-1}}.$$

Average over $x + \mathbb{Z}$. If $K = \gcd(\mathrm{tr}(a\,\mathrm{adj}(r)), N)$,

$$S_{a,r}(c) \ll \frac{\mathbf{1}_{\exists x \in M_d(\mathbb{Z}),\ a\,\mathrm{adj}(r)xr + cr \in NM_d(\mathbb{Z}),\ \mathrm{tr}(2a\,\mathrm{adj}(r)x + c) \in K\mathbb{Z}}}{(N/K)^{1/2} N^{d-1}} \quad \text{(Gauss)}.$$

Geometry of numbers

For each $K \mid N$, we have a lattice

$$\Lambda_{a,r}(K) := \{c \in M_d(\mathbb{Z}) : \exists x \in M_d(\mathbb{Z}),\ a\,\mathrm{adj}(r)xr + cr \in NM_d(\mathbb{Z}), \\ \mathrm{tr}(2a\,\mathrm{adj}(r)x + c) \in K\mathbb{Z}\}.$$

It can be shown that

$$\mathrm{adj}(r)(2c - \mathrm{tr}(c)) \equiv 0 \bmod KM_d(\mathbb{Z})$$

but this seems to be less useful than it was for $d = 2$. We have many successive minima to deal with, since $\mathrm{rank}\,\Lambda_{a,r}(K) = d^2$.

Geometry of numbers

For each $K \mid N$, we have a lattice

$$\Lambda_{a,r}(K) := \{c \in M_d(\mathbb{Z}) : \exists x \in M_d(\mathbb{Z}),\ a\,\mathrm{adj}(r)xr + cr \in NM_d(\mathbb{Z}), \\ \mathrm{tr}(2a\,\mathrm{adj}(r)x + c) \in K\mathbb{Z}\}.$$

It can be shown that

$$\mathrm{adj}(r)(2c - \mathrm{tr}(c)) \equiv 0 \bmod KM_d(\mathbb{Z})$$

but this seems to be less useful than it was for $d = 2$. We have many successive minima to deal with, since $\operatorname{rank} \Lambda_{a,r}(K) = d^2$. We will use Mahler's transference theorem

$$\lambda_i(\Lambda^*_{a,r}(K))\lambda_{d^2-i+1}(\Lambda_{a,r}(K)) \asymp_d 1,$$

which is like applying Poisson summation (again! but we took absolute values after the first Poisson, so this is not circular).

The dual lattice

By definition, $\Lambda^* = \{f \in M_d(\mathbb{Q}) : \mathrm{tr}(fc) \in \mathbb{Z}\ \forall c \in \Lambda\}$ and

$$\Lambda_{a,r}(K) := \{c \in M_d(\mathbb{Z}) : \exists x \in M_d(\mathbb{Z}),\ a\,\mathrm{adj}(r)xr + cr \in NM_d(\mathbb{Z}), \\ \mathrm{tr}(2a\,\mathrm{adj}(r)x + c) \in K\mathbb{Z}\}.$$

Parameterizing $c = y\,\mathrm{adj}(r) - a\,\mathrm{adj}(r)x$ with $x, y \in M_d(\mathbb{Z})$, we see that the mod-K hyperplane $K \mid \mathrm{tr}(a\,\mathrm{adj}(r)x + y\,\mathrm{adj}(r))$ cuts out $\Lambda_{a,r}(K)$.

The dual lattice

By definition, $\Lambda^* = \{f \in M_d(\mathbb{Q}) : \operatorname{tr}(fc) \in \mathbb{Z}\ \forall c \in \Lambda\}$ and

$$\Lambda_{a,r}(K) := \{c \in M_d(\mathbb{Z}) : \exists x \in M_d(\mathbb{Z}),\ a\,\operatorname{adj}(r)xr + cr \in NM_d(\mathbb{Z}), \\ \operatorname{tr}(2a\,\operatorname{adj}(r)x + c) \in K\mathbb{Z}\}.$$

Parameterizing $c = y\,\operatorname{adj}(r) - a\,\operatorname{adj}(r)x$ with $x, y \in M_d(\mathbb{Z})$, we see that the mod-K hyperplane $K \mid \operatorname{tr}(a\,\operatorname{adj}(r)x + y\,\operatorname{adj}(r))$ cuts out $\Lambda_{a,r}(K)$. We may decouple this from the mod-1 hyperplane $\operatorname{tr}(fc) \in \mathbb{Z}$; by duality, the mod-1 hyperplane contains the mod-K hyperplane if and only if

$$M_d(\mathbb{Z})^2 + (-fa\,\operatorname{adj}(r), \operatorname{adj}(r)f)\mathbb{Z} \subseteq M_d(\mathbb{Z})^2 + (\tfrac{a\,\operatorname{adj}(r)}{K}, \tfrac{\operatorname{adj}(r)}{K})\mathbb{Z}.$$

In particular, this implies $\delta := Nf \in rM_d(\mathbb{Z}) + \frac{N}{K}\mathbb{Z} \subseteq M_d(\mathbb{Z})$. It follows upon writing $f = \delta/N$ that

$$N\Lambda^*_{a,r}(K) = \{\delta \in M_d(\mathbb{Z}) : \exists \mu \in \mathbb{Z},\ (\delta + \tfrac{N}{K}\mu)a\,\operatorname{adj}(r) \in NM_d(\mathbb{Z}), \\ \operatorname{adj}(r)(\delta - \tfrac{N}{K}\mu) \in NM_d(\mathbb{Z})\}.$$

Eigenvalue repulsion argument

Lemma

If $\delta \in M_d(\mathbb{Z})$ *and* $\gcd(2\mu, N) = 1$ *with* $0 < |\delta| \leq \epsilon |r|$ *and*

$$(\delta + \mu) a \operatorname{adj}(r), \operatorname{adj}(r)(\delta - \mu) \in N M_d(\mathbb{Z}),$$

then $|\mu| \geq \epsilon N^{1/2}$. *(Prime* $N = |\det(r)| \asymp |r|^d$.*)*

Proof.

We have $\operatorname{rank}(\delta \pm \mu \bmod N) \leq d - 1$, so for some $z \in \mathbb{Z}$

$$\det(t - \delta) \equiv (t - \mu)(t + \mu)(t - z) \bmod N.$$

Eigenvalue repulsion argument

Lemma

If $\delta \in M_d(\mathbb{Z})$ *and* $\gcd(2\mu, N) = 1$ *with* $0 < |\delta| \leq \epsilon |r|$ *and*

$$(\delta + \mu) a \operatorname{adj}(r), \operatorname{adj}(r)(\delta - \mu) \in N M_d(\mathbb{Z}),$$

then $|\mu| \geq \epsilon N^{1/2}$*. (Prime* $N = |\det(r)| \asymp |r|^d$*.)*

Proof.

We have $\operatorname{rank}(\delta \pm \mu \bmod N) \leq d - 1$, so for some $z \in \mathbb{Z}$

$$\det(t - \delta) \equiv (t - \mu)(t + \mu)(t - z) \bmod N.$$

So $\mu^2 \equiv -\operatorname{tr}(\wedge^2 \delta) \ll |\delta|^2 \ll \epsilon^2 N^{2/d} \bmod N$. If $|\mu| \leq \epsilon N^{1/2}$, then $|\mu| = |\operatorname{tr}(\wedge^2 \delta)|^{1/2}$. Now $|\operatorname{adj}(r)(\delta - \mu)| \ll |r|^{d-1} |\delta| \ll \epsilon N$, so $\operatorname{adj}(r)(\delta - \mu) = 0$, whence $\delta = \mu$. But then we find that $2\mu a \operatorname{adj}(r) \in N M_d(\mathbb{Z})$, so $a \in M_d(\mathbb{Z}) r$; a contradiction. □

Schmidt backwards

The lemma implies that the set of integers $\mu \in \mathbb{Z}$ associated to vectors $\delta \in N\Lambda^*_{a,r}(K)$ with $|\delta| \le \epsilon|r|$ is $\ge \epsilon K^{1/2}$-spaced. (This is trivial if $K = 1$.)

$$N\Lambda^*_{a,r}(K) = \{\delta \in M_d(\mathbb{Z}) : \exists \mu \in \mathbb{Z}, (\delta + \tfrac{N}{K}\mu) a \operatorname{adj}(r) \in NM_d(\mathbb{Z}), \\ \operatorname{adj}(r)(\delta - \tfrac{N}{K}\mu) \in NM_d(\mathbb{Z})\}.$$

But for any $\mu \in \mathbb{Z}$, we have

$$\mathcal{C}_\mu := \#\{|\delta| \ll \epsilon|r| : \operatorname{adj}(r)(\delta - \tfrac{N}{K}\mu) \in NM_d(\mathbb{Z})\} \ll \mathcal{C}_0 \ll 1,$$

so $K^{1/2} \gg \#\{\delta \in N\Lambda^*_{a,r}(K) : |\delta| \ll \epsilon|r|\} \gg \frac{|r|^{d^2-j}}{(\lambda_1 \cdots \lambda_{d^2-j})(N\Lambda^*_{a,r}(K))}$ $\asymp \frac{(\lambda_{d^2} \cdots \lambda_{j+1})(\Lambda_{a,r}(K))}{(N/T)^{d^2-j}}$ for all $0 \le j \le d^2$, by Schmidt and Mahler.

Schmidt backwards

The lemma implies that the set of integers $\mu \in \mathbb{Z}$ associated to vectors $\delta \in N\Lambda^*_{a,r}(K)$ with $|\delta| \leq \epsilon|r|$ is $\geq \epsilon K^{1/2}$-spaced. (This is trivial if $K = 1$.)

$$N\Lambda^*_{a,r}(K) = \{\delta \in M_d(\mathbb{Z}) : \exists \mu \in \mathbb{Z}, (\delta + \tfrac{N}{K}\mu) a \operatorname{adj}(r) \in NM_d(\mathbb{Z}), \\ \operatorname{adj}(r)(\delta - \tfrac{N}{K}\mu) \in NM_d(\mathbb{Z})\}.$$

But for any $\mu \in \mathbb{Z}$, we have

$$\mathcal{C}_\mu := \#\{|\delta| \ll \epsilon|r| : \operatorname{adj}(r)(\delta - \tfrac{N}{K}\mu) \in NM_d(\mathbb{Z})\} \ll \mathcal{C}_0 \ll 1,$$

so $K^{1/2} \gg \#\{\delta \in N\Lambda^*_{a,r}(K) : |\delta| \ll \epsilon|r|\} \gg \frac{|r|^{d^2-j}}{(\lambda_1 \cdots \lambda_{d^2-j})(N\Lambda^*_{a,r}(K))}$ $\asymp \frac{(\lambda_{d^2} \cdots \lambda_{j+1})(\Lambda_{a,r}(K))}{(N/T)^{d^2-j}}$ for all $0 \leq j \leq d^2$, by Schmidt and Mahler. But $(\lambda_1 \cdots \lambda_{d^2})(\Lambda_{a,r}(K)) \asymp K(N/T)^{d^2-d}$ (volume calculation), so $(\lambda_1 \cdots \lambda_j)(\Lambda_{a,r}(K)) \gg \frac{K(N/T)^{d^2-d}}{K^{1/2}(N/T)^{d^2-j}} = K^{1/2}(N/T)^{j-d}$.

Schmidt forwards

Since $(\lambda_1 \cdots \lambda_j)(\Lambda_{a,r}(K)) \gg K^{1/2}(N/T)^{j-d}$, Schmidt gives

$$\#\{c \in \Lambda_{a,r}(K) : |c| \leq N/T\} \ll \sum_{0 \leq j \leq d^2} \frac{(N/T)^j}{K^{1/2}(N/T)^{j-d}} \ll \frac{(N/T)^d}{K^{1/2}}.$$

Schmidt forwards

Since $(\lambda_1 \cdots \lambda_j)(\Lambda_{a,r}(K)) \gg K^{1/2}(N/T)^{j-d}$, Schmidt gives

$$\#\{c \in \Lambda_{a,r}(K) : |c| \leq N/T\} \ll \sum_{0 \leq j \leq d^2} \frac{(N/T)^j}{K^{1/2}(N/T)^{j-d}} \ll \frac{(N/T)^d}{K^{1/2}}.$$

Since $S_{a,r}(c)$ is controlled by lattice conditions $c \in \Lambda_{a,r}(K)$, we have something like

$$\begin{aligned}\Sigma_T(ar^{-1}) &:= \sum_{x \in M_d(\mathbb{Z})} w(x/T) e(\mathrm{tr}(ar^{-1}x^2)) \\ &\ll \sum_{|c| \leq N/T} T^{d^2} |S_{a,r}(c)| \\ &\ll \sum_{K|N} \frac{T^{d^2}}{(N/K)^{1/2} N^{d-1}} \frac{(N/T)^d}{K^{1/2}} \ll T^{d^2-d} N^{1/2}.\end{aligned}$$

This is $\ll T^{d^2-\frac{d}{2}}$, since $N = \det(r) \asymp |r|^d \asymp T^d$.

We have proved the following. Let $d \in \{2,3\}$, $r \in M_d(\mathbb{Z})$, and $T \asymp |r| > 0$, with $|\det(r)|$ prime and $|\det(r)| \asymp |r|^d$.

Theorem (Arala–W. 2025+)

Let $w \in C_c^\infty(M_d(\mathbb{R}))$. If $a \in M_d(\mathbb{Z}) \setminus M_d(\mathbb{Z})r$, then

$$\Sigma_T(ar^{-1}) := \sum_{x \in M_d(\mathbb{Z})} w(x/T) e(\mathrm{tr}(ar^{-1}x^2)) \ll_\epsilon T^{d^2 - \frac{d}{2} + \epsilon}.$$

Averaging over $a \in M_d(\mathbb{Z})/M_d(\mathbb{Z})r$ ("polygon method") gives:

Theorem (Arala–W. 2025+)

Let $w \in C_c^\infty(M_d(\mathbb{R})^n)$. If $b \in M_d(\mathbb{Z})$, then

$$\sum_{\substack{x \in M_d(\mathbb{Z})^n \\ x_1^2 + \cdots + x_n^2 - b \in rM_d(\mathbb{Z})}} w(x/T) = \frac{c_w T^{d^2 n}}{|\det(r)|^d} + O_{w,\epsilon}(T^{(d^2 - \frac{d}{2})n + \epsilon}).$$

Some questions

- What about using Weyl differencing instead of Poisson summation? (It looks messy, but maybe. . . ?)
- How does this all relate to the incomplete Eisenstein series perspective of [Nelson, Leung–Young] that we saw yesterday?