Nonabelian equidistribution via sum-product phenomena

A nonabelian circle method for sums of squares

Victor Wang

Institute of Mathematics, Academia Sinica

Joint work with N. Arala, J. R. Getz, J. Hou, C.-H. Hsu, and H. Li

The 2026 Qiuzhen International Mathematics Conference
August 2026

Supported in part by NSF RTG DMS-2231514.
Supported by NSTC Project Grant 114-2115-M-001-010-MY2.
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101034413.

Background: cancellation questions

For an entire, self-dual, analytically normalized geometric $L(s, V) = \sum_{n\geq 1} a_V(n)n^{-s}$ of bounded degree and root number $+1$, $a_V(p) \ll 1$ for primes p (Weil, Deligne):

GRH (elementary)

$$\sum_{p\leq x} a_V(p) \ll_{V,\varepsilon} x^{1/2+\varepsilon}.$$

Square-root cancellation over primes.

Subconvexity (AFE form)

$$\sum_{n\geq 1} \frac{a_V(n)}{\sqrt{n}} W_V\left(\frac{n}{\sqrt{C(V)}}\right) \ll C(V)^{1/4-\delta}.$$

W_V smooth, $\delta > 0$; conductor $C(V)$, AFE length $\sqrt{C(V)}$.

Weyl sums

$$\sum_{m\leq T} \mathrm{e}(\alpha P(m)).$$

Polynomial phases: $\alpha \in \mathbb{R}$, $P \in \mathbb{Z}[X]$, and $\mathrm{e}(t) = \mathrm{e}^{2\pi i t}$.

Today we work in the third setting; GRH and subconvexity are context, not inputs.

Some matrix equations

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

- $XY = YX$, where $X, Y \in M_d(\mathbb{Z})$, $d \geq 2$. Counting with entries in $[-T, T]$ as $T \to \infty$ in [Browning–Sawin–W. 2024, Mudgal 2024, Chapman–Mudgal 2025].
- $X^d = A$, where $X \in M_d(\mathbb{Z})$, $d \geq 2$. If $\det(A) \neq \square^d$, no solutions. If $A = kI_d$ with $t^d - k \in \mathbb{Z}[t]$ irreducible, this has $\sim c_k T^{d(d-1)/2}$ solutions [Eskin–Mozes–Shah 1996]. See also [Habegger–Ostafe–Shparlinski 2024].
- This talk will concentrate on nonabelian sums of n squares, especially the most basic measure of equidistribution as $n \to \infty$ (Weyl sums). In many cases the pointwise Weyl bounds will do better than mean-value alternatives.

Sums of squares in a division algebra

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 $\upsilon_1, \ldots, \upsilon_n \in \{\pm 1\}$ and $T \geq 1$,

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

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

If $k \neq 0$, solutions $X \in M_d(\mathbb{Z})$ to $X^d = kI_d$ lie in 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.

From quaternions to matrices

Example: Hamilton quaternions $D = \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k$,

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

and $\mathcal{O}_D = \mathbb{Z}\frac{1+i+j+k}{2} + \mathbb{Z}i + \mathbb{Z}j + \mathbb{Z}k$ (Hurwitz, 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. 2026+)

Let $d \geq 2$ 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

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

Asymptotic for $n \geq 2d+1$. Previously, $n > 4d + 4 + \frac{4}{d-1}$ sufficed, by [Myerson 2018] along fibers $x_1^2 + \cdots + x_n^2 = b + ry$.

Rough idea of the algebraic circle method

Let A be a free $\mathbb{Z}$-module of finite rank. Fix a $\mathbb{Z}$-bilinear 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.

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 $|\theta r - a| \ll 1/Q$.

In the circle method, we plug in polynomial sequences for x, leading to Weyl sums over x (next slide) for each $\theta \approx a/r$.

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).

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 $|c| \leq |r/T|$. So: $\Sigma_T(a/r) \ll |r|^{1/2}$.

- This is square-root cancellation 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 $r\mathbb{Z}[i] = \mathbb{Z}[i]r$, so that $e(\mathrm{tr}(ar^{-1}x^2))$ depends only on $x \bmod r \in \mathbb{Z}[i]/r\mathbb{Z}[i]$.

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 $\gcd_{\mathbb{Z}}(ar^\dagger, \operatorname{nrd}(r)) \asymp \gcd_{\mathbb{Z}}(r)$, where $\gcd$ is computed in $\mathbb{Z}^5$ and in $\mathbb{Z}^4$, respectively. If $T \asymp |r|$, then

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

Example: take $\operatorname{nrd}(r)$ square-free and $\gcd(\operatorname{nrd}(a), \operatorname{nrd}(r)) = 1$.

Are there near-equality cases? If $\operatorname{trd}(ar^{-1}) \in \mathbb{Z}$, then $\Sigma_T(ar^{-1})$ has a non-oscillatory contribution of size T^3 from $\operatorname{trd}(x) = 0$.

Why the quaternion estimate works

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*

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

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.

The matrix Weyl estimate

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

Theorem (Arala–W. 2026+)

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

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

Generalizing from $d = 2$

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

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

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

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

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

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

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

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 $\mathrm{rank}\,\Lambda_{a,r}(K) = d^2$. 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 $\Lambda_{a,r}(K)$. We may decouple this from the mod-1 hyperplane $\mathrm{tr}(fc) \in \mathbb{Z}$; by duality, the mod-1 hyperplane contains the mod-K hyperplane if and only if

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

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

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

Scalar (eigenvalue) repulsion for prime $K = N$

Lemma

Fix a small $\varepsilon > 0$. Let $A = \{\lambda \in \mathbb{Z}/N\mathbb{Z} : \exists \delta \in M_d(\mathbb{Z}),\ |\delta| \le \varepsilon |r|^{1-2\varepsilon},\ (\delta + \lambda) a\, \mathrm{adj}(r), \mathrm{adj}(r)(\delta - \lambda) \in N M_d(\mathbb{Z})\}$. Then there exists $c = c(a, r) \in (\mathbb{Z}/N\mathbb{Z})^\times$ such that

$$4c\lambda_1\lambda_2 \in I \bmod N$$

for all $\lambda_1, \lambda_2 \in A$, where $I := \{x \in \mathbb{Z} : |x| \le \varepsilon N\}$.

Proof sketch.

Multiplying gives $\mathrm{adj}(r)(\delta^2 - \lambda^2) a\, \mathrm{adj}(r) \equiv 0 \bmod N^2$, since $(\delta - \lambda)(\delta + \lambda) = \delta^2 - \lambda^2$. But $\gcd(\mathrm{adj}(r) a\, \mathrm{adj}(r), N^2) = N$. Generically project onto vectors u, v with $|u|, |v| \ll |r|^{d-1+\varepsilon}$ to get $(u^t v)\lambda^2 \equiv u^t \delta^2 v \ll |r|^{2d-2\varepsilon} \bmod N^2$. Take $c = \frac{u^t v}{N} \in \mathbb{Z}$, so $c\lambda^2 \in I$. Since A and I are roughly closed under finite addition, dispersion $4\lambda_1\lambda_2 = (\lambda_1 + \lambda_2)^2 - (\lambda_1 - \lambda_2)^2$ wins. $\square$

Schmidt backwards

By sum-product phenomena such as the Glibichuk–Konyagin "8AB Theorem", our lemma implies that the set of integers $\mu \in \mathbb{Z}$ associated to vectors $\delta \in N\Lambda^*_{a,r}(K)$ with $|\delta| \leq \varepsilon|r|^{1-2\varepsilon}$ has cardinality $O(K^{1/2})$. (This is trivial if $K = 1$.)

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

Let $T_\varepsilon := \varepsilon|r|^{1-2\varepsilon} \asymp T^{1-2\varepsilon}$. For any $\mu \in \mathbb{Z}$, we have

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

so $K^{1/2} \gg \#\{\delta \in N\Lambda^*_{a,r}(K) : |\delta| \leq T_\varepsilon\} \gg \frac{T_\varepsilon^{d^2-j}}{(\lambda_1 \cdots \lambda_{d^2-j})(N\Lambda^*_{a,r}(K))} \asymp \frac{(\lambda_{d^2} \cdots \lambda_{j+1})(\Lambda_{a,r}(K))}{(N/T)^{d^2-j+O(\varepsilon)}}$ 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 $(\lambda_1 \cdots \lambda_j)(\Lambda_{a,r}(K)) \gg \frac{K(N/T)^{d^2-d}}{K^{1/2}(N/T)^{d^2-j+\varepsilon}} = K^{1/2}(N/T)^{j-d-\varepsilon}$.

Schmidt forwards

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

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

which is $\ll \frac{(N/T)^{d+\varepsilon}}{K^{1/2}}$. Since $S_{a,r}(c)$ is controlled by lattice conditions $c \in \Lambda_{a,r}(K)$, we have something like

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

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

From Weyl cancellation to equidistribution

Assume $d \geq 2$, $r \in M_d(\mathbb{Z})$, and $T \asymp |r| > 0$, with $|\det(r)| \asymp |r|^d$ prime.

Theorem (Arala–W. 2026+)

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

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

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

Theorem (Arala–W. 2026+)

Let $w \in C_c^\infty(M_d(\mathbb{R})^n)$ and $b \in M_d(\mathbb{Z})$. Then

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

The obstruction becomes the source of cancellation

1. A nonabelian shift turns failed central periodicity into an exact vanishing lattice.
2. A scalar Gauss sum and mod-N^2 sum-product repulsion control complementary sides of that lattice.
3. Mahler transference reunites the two inputs into

$$\Sigma_T(ar^{-1}; w) \ll T^{d^2} N^{-1/2} T^{\varepsilon}.$$

Consequence

For $n \geq 2d+1$, sums of n matrix squares equidistribute modulo $rM_d(\mathbb{Z})$ when $\det r = N > 0$ is prime, $\|r\| \asymp T$, and $N \asymp T^d$.

Some questions

- What about more general nonabelian angles? This is a bit cleaner if we replace M_d with a $d \times d$ division algebra, in which case we (Arala–W. 2026+) believe we can handle more general moduli r, such as relaxing "$\det(r)$ prime" to "$\det(r)$ square-free" or "r is of generic split Cartan form".[1]
- How does this all relate to the incomplete-Eisenstein-series perspective of [Nelson, Leung–Young]?

[1] Since sum-product phenomena are more complicated for composite moduli N, this requires some additional ideas such as vector sieving in lattices, and a simple argument from "un-representation theory" in intervals mod N (cf. Brüdern–Kawada–Wooley).