HORIZONTAL VS. VERTICAL SATO/TATE

(OR VERTICAL VS. HORIZONTAL SATO/TATE ?)

PHILIPPE MICHEL
UNIV. MONTPELLIER II & INSTITUT UNIVERSITAIRE DE FRANCE

Princeton, Dec. 12, 2003

1. Kloosterman sums

Given 3 integers $a, b, c \geqslant 1$ such that $(ab, c) = 1$

$$Kl(a, b; c) = \sum_{\substack{x(c) \\ (x,c)=1}} e(\frac{ax + b\overline{x}}{c}), \; x\overline{x} \equiv 1(c), \; e(*) = \exp(2\pi i *)$$

Such sums satisfy elementary properties:

$$Kl(a, b; c) = Kl(1, ab; c), \; \overline{Kl(a, b; c)} = Kl(-a, -b; c) = Kl(a, b; c),$$

ie. $Kl(a, b; c) \in \mathbf{R}$. Another elementary property which will be important for the rest of the talk is

Twisted Multiplicativity: for $c = c_1 c_2$, $(c_1, c_2) = 1$, one has

$$Kl(a, b; c) = Kl(a\overline{c}_2, b\overline{c}_2; c_1) Kl(a\overline{c}_1, b\overline{c}_1; c_2).$$

In particular, Twisted Multiplicativity reduces the problem of estimating Kloosterman sums to the case of a prime power modulus and the only non-elementary case is that of a prime modulus. As a consequence of his resolution of RH for curves over functions fields, A. Weil established

$$|Kl(a,b;p)| \leqslant 2p^{1/2}$$

more precisely (when $p > 2$)

$$Kl(a,b;p) = \alpha_{p,ab} + \beta_{p,ab}$$

is the sum of two algebraic integers such that $|\alpha_{p,ab}| = |\beta_{p,ab}| = \sqrt{p}$, and $\alpha_p \beta_p = p$.
One defines the *angle*, $\theta_{p,ab} \in [0,\pi[$, of the Kloosterman sum $Kl(a,b;p)$ by

$$Kl(a,b;p) = 2p^{1/2}\cos(\theta_{p,ab}).$$

More generaly, for c squarefree $(ab,c) = 1$, the angle $\theta_{c,ab} \in [0,\pi[$, of $Kl(a,b;p)$ is given by

$$Kl(a,b;c) := 2^{\omega(c)} c^{1/2} \cos(\theta_{c,ab}).$$

Twisted multiplicativity is expressed as

$$\cos(\theta_{c_1c_2,ab}) = \cos(\theta_{c_1,\overline{c_2}^2 ab}) \cos(\theta_{c_2,\overline{c_1}^2 ab})$$

2. The horizontal Sato/Tate conjecture

The "horizontal" Sato/Tate conjecture was formulated (in Katz's "Sommes d' Exponentielles" lectures -written by G. Laumon-) in analogy with the Sato/Tate conjecture for elliptic curves. This conjecture describe the distribution of the angles $\{\theta_{p,ab}\}$ when a, b are *fixed* and when p vary over the set of primes:

Conjecture HST. *Given $a, b \geqslant 1$; as $P \to +\infty$, the angles $\{\theta_{p,ab}\}_{p\leqslant P}$ become equidistributed relatively to the Sato/Tate measure on $[0, \pi]$,*

$$d\mu_{ST}(\theta) = \frac{2}{\pi}\sin^2(\theta)d\theta$$

ie. for any $\theta \in [0, \pi]$

$$\frac{|\{p \leqslant P,\ 0 \leqslant \theta_{p,ab} \leqslant \theta\}|}{|\{p \leqslant P\}|} \to \mu_{ST}([0, \theta]) = \frac{2}{\pi}\int_0^\theta \sin^2(t)dt,\ P \to +\infty$$

It is remarkable how little we know about this conjecture:
One still does not know the answer to the following simple questions:

Question 1. *Are there infinitely many primes* p *such that*

$$Kl(1,1;p) > 0$$

(resp.

$$Kl(1,1;p) < 0)\ ?$$

Question 2. *Is there an* $\varepsilon > 0$*, such that*

$$|Kl(1,1;p)| \geqslant \varepsilon(\times p^{1/2})$$

(resp.

$$|Kl(1,1;p)| \leqslant (1-\varepsilon)p^{1/2}\ )$$

for infinitely many primes p *?*

One could answer such questions if one had some non-trivial analytic information on the Euler products

$$L(Kl_1, s) = \prod_{p>2} (1 - \frac{Kl(1,1;p)}{p^s} + \frac{p}{p^{2s}})^{-1} = \prod_{p>2} (1 - \frac{\alpha_{p,1}}{p^s})^{-1} (1 - \frac{\beta_{p,1}}{p^s})^{-1}$$

and

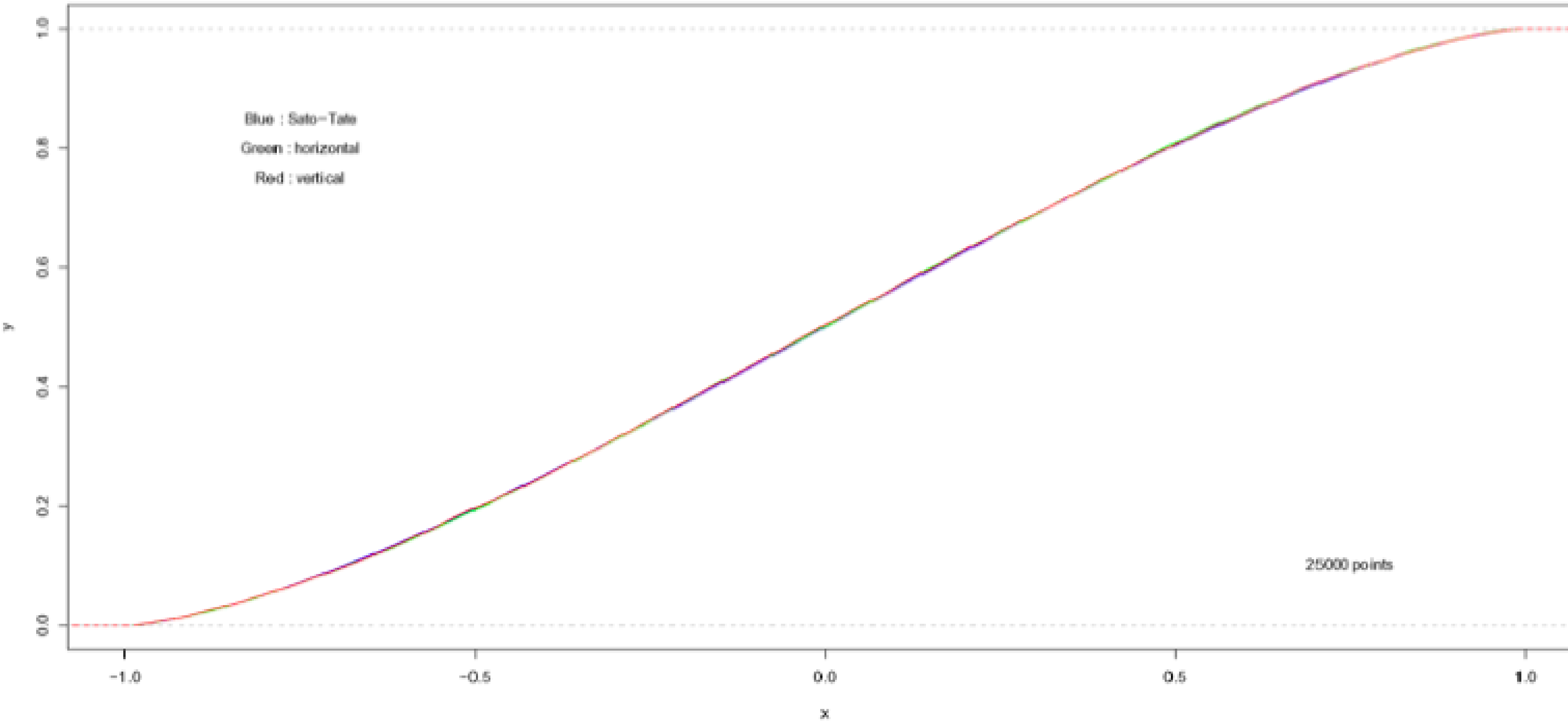
$$L(\mathrm{sym}^2, Kl_1, s) = \prod_{p>2} (1 - \frac{\alpha_{p,1}^2}{p^s})^{-1} (1 - \frac{\beta_{p,1}^2}{p^s})^{-1} (1 - \frac{\alpha_{p,1}\beta_{p,1}}{p^s})^{-1}$$

Unfortunately at that moment, I have no clue on how to get some control on these Euler products (like analytic continuation in a non-obvious region).

It would be the case if $L(Kl_1, s + 1/2)$ were the Hecke L-function of an automorphic form (probably a weight 0 Maass form of level a multiplie of 2) but in fact numerical computations of A. Booker show that it is very unlikely to be the case...

But still, there are reasons to believe in the Horizontal Sato/Tate conjecture:

—- Numerical computations of Kloosterman sum show very good agreement with HST.


Blue : Sato−Tate
Green : horizontal
Red : vertical
25000 points
1.0
0.8
0.6
0.4
0.2
0.0
y
−1.0
−0.5
0.0
0.5
1.0
x

0.004
0.002
0.000
-0.002
0
5000
10000
15000
20000
25000
Red : Horizontal vs. Sato-Tate
Green : Vertical vs. Sato-Tate

—- Other cases of Horizontal Sato/Tate (for other types of exponential sums) have been established

- Heath-Brown/Patterson established the (uniform) equidistribution in $[0, 2\pi]$ of the angles of cubic Gauss sums

$$G((\frac{-}{\pi})_3) = \sqrt{N_{K/\mathbf{Q}}(\pi)}e^{i\theta_\pi}$$

  (associated to the cubic residue symbol of $K = \mathbf{Q}(\sqrt{-3})$) for (split) prime moduli π.
- Duke/Friedlander/Iwaniec established the equidistribution of angles of Salié sums

$$S(1,1;p) = \sum_{\substack{x(p)\\(x,p)=1}} (\frac{x}{p})e(\frac{x+\overline{x}}{p}) =: 2\sqrt{p}\cos(\theta^S_{p,1})$$

  for the uniform measure on $[0, \pi]$. In the analogy with the Sato/Tate conjecture for elliptic curves, the case of Salié sum correspond to the case of CM elliptic curves (but the proof of equidistribution is much harder).

In fact, both cases above combine :

- Sieve techniques (to detect prime moduli amongst arbitrary moduli)
- The analytic theory of automorphic forms (to show that the Weyl sums corresponding to these equidistribution problems are small).

— The third main reason to believe in the HST conjecture is the Vertical Sato/Tate law proved by Katz:

Katz's vertical Sato/Tate law. *When $p \longrightarrow +\infty$, the angles $\{\theta_{p,a}\}_{1 \leqslant a \leqslant p-1}$ become equidistributed relatively to the Sato/Tate measure on $[0, \pi]$,*

$$d\mu_{ST}(\theta) = \frac{2}{\pi} \sin^2(\theta) d\theta$$

ie. for any $\theta \in [0, \pi]$

$$\frac{|\{1 \leqslant a \leqslant p-1,\ 0 \leqslant \theta_{p,a} \leqslant \theta\}|}{p-1} \longrightarrow \mu_{ST}([0, \theta]) = \frac{2}{\pi} \int_0^\theta \sin^2(t) dt,\ P \longrightarrow +\infty$$

3. THE VERTICAL SATO/TATE LAW

The proof of Katz's vertical Sato/Tate law is a combination of three key ingredients:
— Deligne's Equidistribution Theorem for Frobenius conjugacy classes (a consequence of his fundamental theorem on weights)
— Katz's construction of the "Kloosterman sheaf": for $p > 2$ and $\ell \neq p$, there exists an ℓ-adic sheaf $\mathcal{K}l$

- $\mathcal{K}l$ has rank 2, is lisse on $\mathbf{G}_{m,\mathbf{F}_p} = \mathbf{P}^1_{\mathbf{F}_p} - \{0, \infty\}$, irreductible with trivial determinant: $\mathcal{K}l$ "is" a 2-dim irreductible representation

$$\mathcal{K}l : \pi_1^{arith}(G_m) \longrightarrow SL_2(E_\lambda).$$

- $\mathcal{K}l$ is pure of weight 0 and for $a \in \mathbf{G}_m(\mathbf{F}_p) = \mathbf{F}_p^\times$,

$$\mathrm{tr}(\mathrm{Frob}_a|\mathcal{K}l) = \frac{Kl(1, a; p)}{\sqrt{p}}.$$

— Katz computed the ramification at 0 and ∞ of $\mathcal{K}l$, enabling him to show that the geometric monodromy group of $\mathcal{K}l$ is as big as possible

- $\mathcal{K}l$ has unipotent ramification at 0 and is totally wild at ∞ with swan conductor equal to 1 (in particular this is independant of p)
- $\mathcal{K}l(\pi_1^{geom}(\mathbf{G}_m)) = \mathcal{K}l(\pi_1^{arith}(\mathbf{G}_m)) = SL_2.$

3.1. The proof of the Vertical Sato/Tate law. One embed $\overline{\mathbf{Q}_\ell}$ into $\mathbf{C}$ and one choose K a maximal compact subgroup of $SL_2(\mathbf{C})$ ($K = SU(2)$ say). $\mathcal{K}l$ being pure of weight 0 with image contained in $SL_2(\mathbf{C})$, the Frobenius conjugacy classes $\{\mathcal{K}l(\mathrm{Frob}_a)\}_{a\in\mathbf{F}_p^\times}$ define conjugacy classes into $K^\natural$. The latter is identified with $[0,\pi]$ and the direct image of the Haar measure is μ_{ST}. By Weyl equidistribution criterion and Peter/Weyl theorem (and the unitary trick) is is then sufficient to show that for any non-trivial irreductible representation of SL_2, (sym_k say) the corresponding Weyl sum is small

$$\frac{1}{p-1}\sum_{a=1}^{p-1}\mathrm{tr}(\mathrm{Frob}_a|\mathrm{sym}^k\mathcal{K}l) = \frac{1}{p-1}\sum_{a=1}^{p-1} sym_k(\theta_{p,a}) \to 0,$$

as $p\to+\infty$ (here $sym_k(\theta) = \frac{\sin((k+1)\theta)}{\sin(\theta)}$).

By the Lefschetz trace formula, the latter sum equals

$$\frac{1}{p-1}\big[\mathrm{tr}(\mathrm{Frob}_p|H^0_c(\mathbf{G}_m|\mathrm{sym}^k\mathcal{K}l)) - \mathrm{tr}(\mathrm{Frob}_p|H^1_c(\mathbf{G}_m|\mathrm{sym}^k\mathcal{K}l)) \\ + \mathrm{tr}(\mathrm{Frob}_p|H^2_c(\mathbf{G}_m|\mathrm{sym}^k\mathcal{K}l))\big].$$

Now, by Katz's determination of the geometric monodromy group of $\mathcal{K}l$, $\mathrm{sym}^k\mathcal{K}l$ is geometrically irreductible hence $H^0_c = H^2_c = 0$, and by Deligne's theorem

$$|\mathrm{tr}(\mathrm{Frob}_p|H^1_c(\mathbf{G}_m|\mathrm{sym}^k\mathcal{K}l))| \leqslant \dim H^1_c(\mathbf{G}_m|\mathrm{sym}^k\mathcal{K}l)p^{1/2}.$$

Finally, by the Grothendieck/Ogg/Shafarevitch formula, $\dim H^1_c(\mathbf{G}_m|\mathrm{sym}^k\mathcal{K}l))$ can be estimated only in terms of the ramification of $\mathrm{sym}^k\mathcal{K}l$ and shown to be bounded in terms of $\dim \mathrm{sym}^k = k+1$ but independently of p.

3.2. Some vertical variants. Note that one can identify $\mathbf{F}_p^\times$ with the interval of integers $[1, p-1]$ (which is odd from the view point of algebraic geometry but perfectly natural for the view point of number theory); one may then ask whether the Vertical Equidistribution law continue to hold for smaller (but sufficiently large and regular) subsets of integers in $[1, p-1]$. This is indeed the case and for example one has a Sato/Tate law for rather short intervals

Theorem. *Given any $\varepsilon > 0$, as $p \to +\infty$, the set of Kloosterman angles $\{\theta_{p,a}\}_{1 \leqslant a \leqslant p^{1/2+\varepsilon}}$ becomes equidistributed for the Sato/Tate measure.*

This variant is consequence of the Polya/Vinogradov completion method and of the algebro-geometric arguments given above. In fact the method lead to the consideration of new sheaves: the twisted sheaves

$$\mathrm{sym}^k \mathcal{K}l \otimes \mathcal{L}_\psi$$

where $\mathcal{L}_\psi$ ranges over the rank one sheaves on $\mathbf{A}^1_{\mathbf{F}_p}$ associated with the characters ψ of $(\mathbf{F}_p, +)$. The new algebro-geometric input, here is the (simple) fact that a geometrically irreductible sheaf of rank > 1 (like $\mathrm{sym}^k \mathcal{K}l$) twisted by a sheaf of rank 1 remains geometrically irreductible.

One also has a vertical Sato/Tate law over the primes less than p

Theorem. *As $p \to +\infty$, the set of Kloosterman angles $\{\theta_{p,q}\}_{\substack{1\leqslant q\leqslant p-1 \\ q \ prime}}$ becomes equidistributed for the Sato/Tate measure.*

The second variant is more involved and require more sophisticated transformations coming from sieve methods.
Again after these transformations, new sheaves need to be considered: namely the (Rankin/Selberg) sheaves

$$\mathrm{sym}^k \mathcal{K}l \otimes \mathrm{sym}^k [a]^* \mathcal{K}l$$

where $a \in \mathbf{F}_p^\times - \{1\}$ and $[a] : x \to ax$ denote the (non-trivial) translation on $\mathbf{G}_m$. At the end, the main geometrical result needed to establish this variant is an independance statement for Kloosterman sheaves

Proposition. *If $a \neq 1$, the geometric monodromy group of $\mathcal{K}l \oplus [a]^*\mathcal{K}l$ is as big as possible: ie. equals $SL_2 \times SL_2$.*

The latter proposition follows from the Goursat/Kolchin/Ribet criterion which is verified either by using the Rankin/Selberg method or by comparing the monodromies at ∞ of $\mathcal{K}l$ and $[a]^*\mathcal{K}l$.

4. Horizontal vs. Vertical Sato/Tate

Interestingly, the above variants of the vertical Sato/Tate law can be used to provide results in the *horizontal* direction. However for this we have to mollify the problem by allowing non-prime moduli.

Let $c = p_1 \dots p_k$ be a squarefree integer with a fixed number $k \geqslant 2$ factors by twisted multiplicativity the angle of the Kloosterman sum $Kl(1,1;c)$ satisfies

$$\cos(\theta_{c,1}) = \frac{Kl(1,1;c)}{2^k\sqrt{c}} = \prod_{p|c} \cos(\theta_{p,\overline{(c/p)}^2})$$

then since we do not expect the primes to see each other, it is reasonable to make the following

Conjecture HST(k). *As c ranges over the squarefree integers having k prime factors none of which is small (for example $p|c \Rightarrow p \geqslant c^{1/2k}$) the angles $\{\theta_{c,1}\}_c$ are equidistributed on $[0,\pi]$ relatively to the measure $\mu_{ST}^{(k)}$ given by the direct image of $\mu_{ST}^{\otimes^k}$ of the map on $[0,\pi]^k$ given by*

$$(\theta_1, \dots, \theta_k) \to \arccos(\cos(\theta_1)\dots\cos(\theta_k)).$$

Observe that this conjecture is not implied by Conjecture HST and at the present time seems as intractable as the original one. On the other hand, one can ask the same basic questions about the size and the existence of sign changes of Kloosterman sums with composite moduli. Due to the extra flexibility in the allowed moduli, both questions can be answered (affirmatively) if k is sufficiently large.

4.1. The size of Kloosterman sums.

Theorem (M.). *There exists infinitely many pairs of distincts primes (p,q) such that $|Kl(1,1;pq)|\geqslant\frac{4}{25}\sqrt{pq}$. More precisely for P large enough*

$$|\{(p,q),\ p\neq q,\ p,q\leqslant P,\ |Kl(1,1;pq)|\geqslant\frac{2}{25}\sqrt{pq}\}|\gg P^2/\log^2 P$$

The argument of the proof combines a (simple) probability argument with a vertical Sato/Tate law over primes:

Lemma 1. *Given* (Ω, μ) *a probability space and* $\Omega_1, \Omega_2 \subset \Omega$ *such that* $\mu(\Omega_1) + \mu(\Omega_2) > 1$ *then*

$$\mu(\Omega_1 \cap \Omega_2) \geqslant \mu(\Omega_1) + \mu(\Omega_2) - 1 > 0$$

Theorem. *As* $P \to +\infty$*, the sets*

$$\{\theta_{p,\overline{q}^2},\ p \neq q,\ 3 \leqslant p, q \leqslant P, \},\ \{\theta_{q,\overline{p}^2},\ p \neq q,\ 3 \leqslant p, q \leqslant P, \}$$

are equidistributed relatively to the Sato/tate measure.

One then choose

$$\Omega_1 = \{(p,q),\ p \neq q,\ p, q \leqslant P,\ |\cos(\theta_{p,\overline{q}^2})| \geqslant 2/5\},$$

and

$$\Omega_2 = \{(p,q),\ p \neq q,\ p, q \leqslant P,\ |\cos(\theta_{q,\overline{p}^2})| \geqslant 2/5\}$$

and use

$$\mu_{ST}(\{\theta,\ |\cos\theta| > 2/5\}) > 1/2$$

and the fact that for (p,q) in $\Omega_1 \cap \Omega_2 \neq \emptyset$

$$|\cos(\theta_{pq,1})| = |\cos(\theta_{p,\overline{q}^2})||\cos(\theta_{q,\overline{p}^2})| \geqslant 4/25.$$

One can elaborate further on this ideas and complement them with other techniques from analytic number theory (the Large Sieve, the Barban Davenport/Halberstam Theorem etc...) as well as with some other sophisticated variants of the horizontal Sato/tate laws (some of these variants were provided by Katz for quite different purposes) to obtain other results in the same direction: we state two further results which will be useful in the rest of this talk

Theorem (Fouvry/M.). *For $u_0 = 23.9$, one has for X large enough*

$$\sum_{\substack{c \leqslant X \\ p|c \Rightarrow p \geqslant X^{1/u_0}}} \mu^2(2c) \frac{|Kl(1,1;c)|}{\sqrt{c}} \geqslant 0.166 \frac{X}{\log X}$$

Theorem (Fouvry/M.). *For some absolute (explicit) constant $0 < a(< 1)$ one has*

$$X \frac{\exp((\log\log X)^a)}{\log X} \ll \sum_{c \leqslant X} \frac{|Kl(1,1;c)|^2}{c} \ll X(\log\log X)^{4-1}$$

Observe that these bounds are non-trivial (the trivial upper bound being $\ll X(\log X)^{4-1}$ and a being positive); on the other hand, the probabilistic model for Kloosterman sums predicts that

$$\sum_{c \leqslant X} \frac{|Kl(1,1;c)|^2}{c} \simeq bX.$$

All what has been said so far can be generalized to wide classes of families of algebraic exponential sums (essentially families for which the geometric monodromy group has been computed - by Katz- like hypergeometric Kloosterman sums or exponential sums obtained by geometric Fourier transforms: see [GKM,ESDE]).

What is peculiar to Kloosterman sums is that they arise in other contexts and in particular in the analytic theory of automorphic forms:
both topics are connected via the Petersson/Kuznetsov trace formula which relate sums of Kloosterman sums of moduli divisible by N to Fourier coefficients of automorphic forms on $\Gamma_0(N)$: here φ denote a sufficiently smooth function, $\check{\varphi}$, $\hat{\varphi}$ some Bessel transforms; $B_k(q)$ denotes an orthonormal basis of the space of holomorphic forms of wt. k and level q, the $\{f_j\}_{j\geqslant 0}$ ranges over an orthnormal basis of wt. 0 Maass cusp forms with Laplace eigenvalue $\lambda_j = 1/4 + t_j^2$ and the last portion is over the Eisenstein series

$$\sum_{c\equiv 0(N)} \frac{S(m,n;c)}{c}\varphi(c) = \sum_{k\equiv 0(2)} \check{\varphi}(k) \sum_{f\in B_k(q)} \overline{\rho_f}(m)\rho_f(n)$$
$$+\sum_{j\geqslant 1} \hat{\varphi}(t_j)\overline{\rho_{f_j}}(m)\rho_{f_j}(n) + \frac{1}{4\pi}\sum_{\mathfrak{a}} \int_{-\infty}^{+\infty} \hat{\varphi}(t_j)\overline{\rho_{\mathfrak{a}}, t}(m)\rho_{\mathfrak{a}}, t(n)dt.$$

For at least 20 years, this connection in one of the cornerstone of the modern analytic theory of number and has and is still widely exploited in both directions. We give below two recent exemples.

4.2. From Kloosterman sums to automorphic forms: the error term in the counting of cusp forms. The main motivation by Selberg for developping the trace formula was for proving the existence of cusp forms. In particular for the full modular curve he obtained the Weyl law: for $T \geqslant 1$

$$N(T) := |\{j \geqslant 1, |t_j| \leqslant T\}| = MainTerm(T) + S(T);$$

here $MainTerm(T)$ is well understood and asymptotic to

$$MainTerm(T) \simeq \frac{vol(X_0(1))}{4\pi} T^2;$$

on the other hand $S(T)$ is rather small error term $(= O(T))$ but not so well understood. Selberg also established that this error term is often not too small:

$$\int_T^{2T} |S(t)|^2 dt \gg T^2 / \log^2 T.$$

Very recently, X. Li and P. Sarnak , by using the Petersson/Kuznetzov formula – instead of Selberg's trace formula –and the lower bound (along with other analytic techniques)

$$X\frac{\exp((\log\log X)^a)}{\log X} \ll \sum_{c\leqslant X}\frac{|Kl(1,1;c)|^2}{c}$$

have given the first (modest but meaningfull) improvment over Selberg's lower bound:

$$\int_T^{2T}|S(t)|^2 dt \gg \frac{T^2}{\log^2 T}\exp((\log\log T)^a)$$

4.3. Sign changes of Kloosterman sums. In the other direction (from automorphic forms to Kloosterman sums), on can use the Kuznetzov formula to prove the existence of sign changes for Kloosterman sums. Indeed by using his formula (and Roelcke lower bound on the first eigenvalue λ_1 of $X_0(1)$), Kuznetzov proved the estimate

$$\sum_{c\leqslant X}\frac{Kl(1,1;c)}{c^{1/2}}\ll X^{1/2+1/6+\varepsilon}$$

which is clearly non trivial by comparison with Weil's bound; however until recently (as was remarked by Serre) this was not evident that this estimate proved the existence of sign changes for $Kl(1,1;c)$. However thanks to the lower bound of the previous sections one has

Proposition. *There are infinitely many integers c such that $Kl(1,1;c)>0$ (resp. $Kl(1,1;c)<0$).*

The next step consists in limiting the number of allowed prime factors of c in the above proposition. One has the following result which answer positively to the question of sign changes of Kloosterman sums of prime moduli:

Theorem (Fouvry/M.). *There exists infinitely many squarefree c (a positive proportion in fact) having all their prime factor $\geqslant c^{1/23.9}$ such that $Kl(1,1;c) > 0$ (resp. $Kl(1,1;c) < 0$).*

Sketch of Proof. In view of the lower bound $(u_0 = 1/23.9)$

$$\sum_{\substack{c\leqslant X\\ p|c\Rightarrow p\geqslant X^{1/u_0}}} \mu^2(2c)\frac{|Kl(1,1;c)|}{\sqrt{c}} \geqslant 0.166\frac{X}{\log X}$$

it is sufficient to show that

$$|\sum_{\substack{c\leqslant X\\ p|c\Rightarrow p\geqslant X^{1/u_0}}} \mu^2(2c)\frac{Kl(1,1;c)}{\sqrt{c}}| \leqslant C\frac{X}{\log X}$$

for some $C < 0.166$. These kind of estimates follows from sieve methods; here we use a variant of Selberg's upper bound sieve. Recall that the input in Selberg's sieve is an non-negative arithmetic function $(a_c)_{c\leqslant X}$ say and provide bounds

$$\sum_{\substack{c\leqslant X\\ p|c\Rightarrow p\geqslant X^{1/u}}} a_c \leqslant C(u)\frac{X}{\log X}$$

with $C(u) > 0$ a decreasing function of u.

In the present case (to force positivity) we consider the two sequence

$$a_c^{\pm} = 2^{\omega(c)} \pm \frac{Kl(1,1;c)}{\sqrt{c}}$$

However a necessary condition to sieve is to control such sequence well in arithmetic progressions to large moduli, which lead us to have good bounds for the sums

$$\sum_{c \equiv 0(q)} \frac{Kl(1,1;c)}{\sqrt{c}}.$$

such bounds can be obtained by means of Kuznetzov formula for $\Gamma_0(q)$ and large sieve inequality (for Maass forms) together with the Luo/Rudnick/Sarnak lower bound for λ_1 (any bound strictly better that Selberg's $\lambda_1 > 3/16$ would be sufficient) gives us good control for q up to size $X^{1/2-\varepsilon}$ (this is an analog of the Bombieri/Vinogradov theorem)... $\square$

5. Appendum: improvements and extensions

Other type of sums can be dealed with by the above method: for example, recently Livné/Patterson established an analog of Kuznetzov formula in the metaplectic setting enabling them to evaluate non-trivially the first moment of cubic Kloosterman sums. Extension of the results above, for such sums and others, are currently being investigated by B. Louvel in his Göttingen PhD thesis.

Other refinment on the sieve side are also possible: by using iteration of Selberg's sieve to form a lower bound sieve, C. Sivak improved $u_0 = 23.9$ to $u_0 = 22.9$; moreover by using an effectivization of Bombieri's asymptotic sieve, she arrived at $u_0 = 20.9$.

Finally, it is likely that most of the progress in reducing the number of prime factor of c will come from an improvement in the constant bound

$$\sum_{\substack{c \leqslant X \\ p|c \Rightarrow p \geqslant X^{1/u_0}}} \mu^2(2c) \frac{|Kl(1,1;c)|}{\sqrt{c}} \geqslant 0.166 \frac{X}{\log X}.$$

In such lower bound the contribution of the c with two primes factors is treated trivially (lower bounded by 0 !).

In some previous work Fouvry and I (building on earlier work of Friedlander/Iwaniec) developped a general series of transformation to detect cancellation in sums of the forms

$$\sum_{\substack{p^\eta \leqslant q \leqslant 2p^\eta \\ q \text{ prime}}} f(q;p)$$

where $f(a;p)$ is a function on $\mathbf{F}_p$ (bounded by 1 say) and for all $\eta \in [6/7, 1]$. However the price to pay on the geometric side is enormous: for one needs very strong estimate for two dimensionnal sums of the form

$$\sum_{x,y\in\mathbf{F}_p} f(x(y+b_1))f(x(y+b_2))f(x(y+b_3))f(x(y+b_1'))f(x(y+b_2'))f(x(y+b_3')) \ll p$$

uniformly for $(b_1, b_2, b_3, b_1', b_2', b_3')$ outside a codimension 1 subvariety of $\mathbf{A}^6_{\mathbf{F}_p}$. If f is the function given as the Frobenius trace of some ℓ-adic sheaf, this means that one has to show that

$$H^4_c = H^3_c = 0,$$

(usually it is not difficult to show that $H^4_c = 0$).

When f corresponds to some sheaves of rank 1, we could establish by averaging over one of the variable getting a one variable sum of exponential sums and by using the methods of [ESDE]. For the present case of Kloosterman sums, one needs to establish such estimate for the functions

$$f_k(a) = \mathrm{tr}(\mathrm{Frob}_{\overline{a}^2}|\mathrm{sym}^k\mathcal{K}l)$$

which is a much harder problem. this question is currently investigated by C. Sivak in Orsay.