数理解析研究所講究録 1468 巻 2006 年 113-136

SOME APPLICATIONS OF KLOOSTERMANIA

PHILIPPE MICHEL

1. INTRODUCTION

Given $a, b, c \geqslant 1$ such that $(ab, c) = 1$, Kloosterman sums are a special kind of algebraic exponential sums given by the following expression

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

These sums were introduced by Kloosterman in 1926 [Klo26] on the occasion of the so-called *Kloosterman refinement* in the circle method to give an asymptotic expression of the number of representations, $r_{abcd}(n)$, of a large integer n by a diagonal[1] quaternary definite quadratic form

$$ax_1^2 + bx_2^2 + cx_3^2 + dx_4^2 = n.$$

He also provided a non trivial bound for these sums which enabled him to show that the Hasse principle holds for such quadratic forms. This was the first indication that Kloosterman sums have to do which modular forms as $r_{abcd}(n)$ is the n-th Fourier coefficient of a theta series of weight two. A more direct connection between Kloosterman sums and modular forms came with Petersson's formula which express Fourier coefficients of modular forms in terms of Kloosterman sums.

The word *Kloostermania* of the title was invented by M. Huxley in the 80's to highlight a series of striking developments that took place at that time in analytic number theory and which build further of on the modular nature of Kloosterman sums. The starting point was Kuznetzov's extension of Petersson's formula to the full automorphic spectrum for the modular group $SL_2(\mathbf{Z})$: his formula enabled him to use spectral theory of the modular surface to solve essentially Linnik's conjecture on the existence of cancellation of sums of Kloosterman sums over the modulus [Kuz80]; but Kloostermania really lifted-off with the landmark paper of Deshouillers/Iwaniec [DI82] who generalized Kuznetzov's formula to arbitrary congruence sugroups. These formulae were them used it to derive bounds of sums of exponential sums and provided a powerful new tool to analytic number theory. Amongst the many applications that followed from them, arguably the most striking are the works of Bombieri/Frieldander/Iwaniec and Fouvry who obtained improvments over the Bombieri/Vinogradov theorem which go far beyond the possibilities of the Generalized Riemann hypothesis [BFI86, BFI89, Fou84].

[1]later Kloosterman extended this result to general definite quaternary quadratic forms

Besides their modular nature, Kloosterman sums, as a special type of algebraic exponential sums, also enjoy rather deep algebraic/geometric properties (the simplest being Weil's bound) as follows from the work of Deligne and Katz in ℓ-adic cohomology. In this paper, we review both of these aspects of Kloosterman sums and present several recent applications which make use of their arithmetic/geometrical and spectral properties.

Acknowledgments. The present survey essentially follows the lectures that I gave at the RIMS on the occasion of the workshop "Automorphic forms and automorphic L-functions" and I would like to thanks the organizers H. Saito and M. Furusawa for their kind invitation and for the excellent working conditions.

2. Kloosterman sums from the algebraic viewpoint

Let us write once again the definition of 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 various elementary properties:

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

ie. $Kl(a,b;c)\in\mathbf{R}$.

There is another simple property which follows for the Chinese Remainder Theorem, namely the:

Twisted Multiplicativity Property: if c_1, c_2 are coprime integers and $c = c_1c_2$, 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.

In the prime modulus case, the first non trivial bound for Kloosterman sums was given by Kloosterman himself: by computing (using purely elementary methods) the fourth moment of Kloosterman sums, he obtained

$$\sum_{a\neq 0(\text{mod } p)} |Kl(a,b;p)|^4 \leqslant 16p^3$$

so that

$$|Kl(a,b;p)| \leqslant 2p^{3/4}.$$

This bound was already sufficient to resolve the problem of representing an integer by a quaternary quadratic form.

In the 50's, as a consequence of his resolution of the Riemann Hypothesis for curves over functions fields, A. Weil [Wei48] established a stronger (probably optimal) bound

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

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 $\alpha_p\beta_p = p$.

It is then natural to raise the question of the optimality of Weil's bound. For this 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 generally, for c a squarefree integer and $(ab,c) = 1$, the *angle* $\theta_{c,ab} \in [0,\pi[$, of $Kl(a,b;c)$ is defined by

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

The Twisted Multiplicativity is then expressed by the formula

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

3. Horizontal and Vertical distribution laws

The angle of a Kloosterman sum $\theta_{c,a}$ depends on two parameters: the **modulus** c (which in our case will be squarefree with a fixed number of prime factors) and the **argument** a (which is an integer coprime with c): following the terminology introduced by N. Katz, we call

- **Horizontal**, a law which describes the asymptotic distribution of the family of angles
$$\{\theta_{c,a}\}_{\substack{1\leqslant a\leqslant c\\(a,c)=1}},$$
for $c \to +\infty$ amongst the squarefree integers with a fixed number of prime factors.
- **Vertical**, a law which describes the asymptotic distribution of the family of angles
$$\{\theta_{c,a}\}_{\substack{1\leqslant c\leqslant C\\(a,c)=1}},$$
for a fixed non-zero integer a, $C \to +\infty$ and c ranging amongst the squarefree integers less than C, coprime with a and with a fixed number of prime factors.

Of the two types of possible laws above, horizontal laws are by far the most mysterious.

4. The Vertical Sato/Tate law (VST)

In the vertical direction, the situation is completely understood thanks to the work of Deligne and Katz [Del74, Del80, Kat88]

VST . *When $p \to +\infty$, 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 $\theta \in [0,\pi]$

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

The proof of Katz's vertical Sato/Tate law is a combination of three key ingredients:

(1) Deligne's Equidistribution Theorem for Frobenius conjugacy classes (a consequence of his fundamental theorem on weights)
(2) 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 $\mathbf{G}_{m,\mathbf{F}_p} = \mathbf{P}^1_{\mathbf{F}_p} - \{0,\infty\}$, irreducible with trivial determinant: $\mathcal{K}l$ "is" a 2-dim irreducible representation

$$\mathcal{K}l : \pi_1^{arith}(\mathbf{G}_m) \to SL_2(E_\lambda).$$

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

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

(3) 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 independent of p)
   - $\mathcal{K}l(\pi_1^{geom}(\mathbf{G}_m)) = \mathcal{K}l(\pi_1^{arith}(\mathbf{G}_m)) = SL_2$.

4.1. 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 irreducible 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

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

- Now, by Katz's determination of the geometric monodromy group of $\mathcal{K}l$, $\mathrm{Sym}^k\mathcal{K}l$ is geometrically irreducible hence $H_c^0 = H_c^2 = 0$, and by Deligne's theorem

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

- Finally, by the Grothendieck/Ogg/Shafarevitch formula, $\dim H_c^1(\mathbf{G}_m|\mathrm{Sym}^k\mathcal{K}l)$ 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.

4.2. **A variant of VST for composite moduli.** When the modulus $c = p_1p_2\dots p_k$ is composite with say $k \geqslant 1$ prime factors, an analog of the vertical Sato/Tate law quickly follows from the twisted multiplicativity,

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

and from the chinese remainder theorem:

for $k \geqslant 1$, let $\mu_{ST}^{(k)}$ denote the measure on $[0,\pi]$ given by the direct image of $\mu_{ST}^{\otimes k}$ of the map $[0,\pi]^k \to [0,\pi]$ given by

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

VST(k). *As $c \to +\infty$ amongst the squarefree integers having k prime factors none of which is small (for example $p|c \Rightarrow p \geqslant c^{1/2k}$) the angles $\{\theta_{c,a}\}_{\substack{1\leqslant a\leqslant c\\(a,c)=1}}$ are equidistributed on $[0,\pi]$ relatively to the measure $\mu_{ST}^{(k)}$.*

5. The Horizontal Sato/Tate conjecture

The **horizontal** analog of the Vertical Sato/Tate law was conjectured by Katz (before his proof of the VST) as a close analog of the Sato/Tate conjecture for elliptic curves.

Conjecture HST. *Given $a \geqslant 1$; as $P \to +\infty$, the angles $\{\theta_{p,a}\}_{\substack{p\leqslant P\\(a,p)=1}}$ 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 $\theta \in [0,\pi]$

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

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*

(1) *infinitely many primes p for which $Kl(1,1;p) > 0$?*
(2) *infinitely many primes p for which $Kl(1,1;p) < 0$?*

Question 2. *Is there an $\varepsilon > 0$ for which*

(1) *there are infinitely many primes p for which $|Kl(1,1;p)| \leqslant (1-\varepsilon)p^{1/2}$?*
(2) *there are infinitely many primes p for which $|Kl(1,1;p)| \geqslant \varepsilon p^{1/2}$?*
(3) *or even, there are infinitely many primes such that $|Kl(1,1;p)| \geqslant \varepsilon$?*

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

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

Apparently, at that moment, no one has a clue on how to get some control on these Euler products (like analytic continuation in a non-obvious region). Of course 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 multiple of 2)...

... but in fact some numerical computations of A. Booker show that it is very unlikely to be the case [Boo00].

Still, there are some reasons to believe in the Horizontal Sato/Tate conjecture: the first one is the validity of Katz's Vertical Sato/Tate law. Another reason is that numerical computations of Kloosterman sum show very good agreement with HST.

But probably the best reason to believe in HST is the fact that for other exponential sums the Horizontal Sato/Tate Law has been proven !

(1) Heath-Brown/Patterson [HBP79] 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.

(2) Duke/Friedlander/Iwaniec [DFI95] 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 classical Sato/Tate conjecture for elliptic curves, the case of Salié sum would correspond to the case of CM elliptic curves (but the proof of equidistribution of Salié sums is much harder).

The proofs of both cases above make use of

- 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).

6. Horizontal Sato/Tate for composite moduli

Interestingly, some variants of the vertical Sato/Tate law can be used to provide results in the *horizontal* direction. For this one has to mollify the problem by allowing composite moduli.

Let $c = p_1 \dots p_k$ be a squarefree integer with a fixed number $k \geqslant 2$ factors; by twisted multiplicativity the angles 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 expect the primes to behave independently of each other, it is reasonable to make the following

Conjecture HST(k). *As $C \to +\infty$, the family of angles $\{\theta_{c,1}\}_{c\leqslant C}$ are equidistributed on $[0, \pi]$ relatively to the Sato/Tate measure of order k, $\mu^{(k)}_{ST}$, where the moduli c range over the squarefree integers $\leqslant C$, with k prime factors none of which is small (for example $p|c \Rightarrow p \geqslant c^{1/2k}$).*

Observe that conjecture **HST(k)** 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 amongst Kloosterman sums with composite moduli. As we shall see, due to the extra flexibility allowed by multiple prime factors in the moduli, both questions can be answered (affirmatively) if k is sufficiently large.

6.1. Three variants of the Vertical Sato/Tate law.

Theorem 6.1. *There exists infinitely many pairs of distinct primes (p, q) such that $|Kl(1,1;pq)| \geqslant \frac{4}{25}\sqrt{pq}$. More precisely for X large enough*

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

In particular (take $P = X^{1/2}$)

$$\sum_{\substack{c \leqslant X\\ \omega(c)=2}} \mu^2(c)\frac{|Kl(1,1;c)|}{\sqrt{c}} \gg \frac{X}{\log^2 X}$$

(here the implied constant could be evaluated exactly).

In particular, this first result shows that Weil's bound is indeed optimal in the horizontal aspect for moduli with 2 prime factors [Mic95]! The next result shows that at the expense of allowing an extra prime factor in the moduli, one can improve the second inequality by a factor $\log X$ [FM02]

Theorem 6.2.

$$\sum_{\substack{c \leqslant X\\ \omega(c)=3}} \mu^2(c)\frac{|Kl(1,1;c)|}{\sqrt{c}} \geqslant 0.078\frac{X}{\log X}$$

Finally we state a mean square estimate on Kloosterman sums with arbitrary moduli, gaining an extra $\exp((\log\log X)^{5/17})$ factor [FM03]

Theorem 6.3. *For $X \to +\infty$, one has*

$$X\frac{\exp((\log\log X)^{5/17})}{\log X} \ll \sum_{c \leqslant X} \frac{|Kl(1,1;c)|^2}{c} \ll X(\log\log X)^3$$

Observe that the last upper bound is also non-trivial (the trivial upper bound being $\ll X(\log X)^3$); on the other hand, a natural probabilistic model for Kloosterman sums predicts that

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

for some constant $b > 0$ that can be explicitly computed.

The proof of these three results are based on common principles:

For some $k \geqslant 2$, let C_k be the set of squarefree integers c less that X with k prime factors. We want to show that for many (ie. a positive proportion of) $c \in C_k$ the Kloosterman sum is large.

- **Step 1.**— For some decomposition $k = k_1 + k_2$ with $k_1, k_2 \geqslant 1$, pick $\Omega_1 \subset \mathcal{C}_{k_1}$ and $\Omega_2 \subset \mathcal{C}_{k_2}$, two (big enough) subsets of squarefree integers with k_1 and k_2 prime factors respectively, such that

  $$\Omega_1.\Omega_2 := \{c_1.c_2,\ c_1 \in \Omega_1,\ c_2 \in \Omega_2\} \subset \mathcal{C}_k.$$

  Here big enough means in particular that

  $$|\Omega_1.\Omega_2| \gg |\mathcal{C}_k| \gg_k \frac{X}{\log X}$$

- **Step 2.**— Prove that, for $X \to +\infty$, the two families of angles of Kloosterman sums

  $$\{\theta_{c_1,\overline{c}_2^2},\ (c_1, c_2) \in \Omega_1 \times \Omega_2\}$$
  $$\{\theta_{c_2,\overline{c}_1^2},\ (c_1, c_2) \in \Omega_1 \times \Omega_2\}$$

  become equidistributed on $[0, \pi]$ relatively to the respective measures $\mu_{ST}^{(k_1)}$ and $\mu_{ST}^{(k_2)}$. In particular, this implies that there is $\alpha > 0$ and $\delta > 0$ such that

  $$\left|\{(c_1, c_2) \in \Omega_1 \times \Omega_2,\ s.t.\ |\cos\theta_{c_1,\overline{c}_2^2}| \geqslant \alpha\}\right| \geqslant (1/2 + \delta)\left|\Omega_1.\Omega_2\right|$$
  $$\left|\{(c_1, c_2) \in \Omega_1 \times \Omega_2,\ s.t.\ |\cos\theta_{c_2,\overline{c}_1^2}| \geqslant \alpha\}\right| \geqslant (1/2 + \delta)\left|\Omega_1.\Omega_2\right|$$

- **Step 3.**—Use with the following simple probability argument

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

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

along with the twisted multiplicativity property

$$\cos\theta_{c_1,\overline{c}_2^2}\cos\theta_{c_2,\overline{c}_1^2} = \cos\theta_{c_1c_2,1} = \frac{Kl(1, 1; c_1c_2)}{2^{k_1+k_2}\sqrt{c_1c_2}}$$

to conclude that

$$\left|\left\{c_1c_2 \in \Omega_1.\Omega_2,\ s.t.\ \frac{|Kl(1, 1; c_1c_2)|}{2^{k_1+k_2}\sqrt{c_1c_2}} \geqslant \alpha^2\right\}\right| \geqslant 2\delta\left|\Omega_1.\Omega_2\right| \gg \left|\mathcal{C}_k\right| \gg \frac{X}{\log X}.$$

6.2. **Some variants of the Vertical Sato/Tate laws.** Clearly performing Step 2, ie. the equidistribution of the angles

$$\{\theta_{c_1,\overline{c}_2^2},\ (c_1, c_2) \in \Omega_1 \times \Omega_2\},$$

amounts to proving some variant of Katz's vertical Sato/Tate laws (possibly extended to composite moduli). In this variants, the set of arguments of the angle of Kloosterman sums considered is not merely the set of integers contained in the interval $[1, c_1]$ but a more general set of integers. By standard techniques from analytic number theory is indeed possible to pass to such smaller sets.

In practice the set of moduli Ω_1 and of arguments Ω_2 are contained in two intervals $[1, X_1]$ and $[1, X_2]$ respectively, were X_1, X_2 are two positive powers of X with $X_1X_2 = X$ and one has

$$|\Omega_1| \gg X_1/\log X_1, \ |\Omega_2| \gg X_2/\log X_2.$$

— The easiest case is when the set of argument is somewhat greater that the set of moduli, say

$$X_2 \gg X_1 \log^{100} X.$$

In that case, one case use large sieve techniques (like the Barban/Davenport/Halberstam inequality) to show that on average over the moduli $c_1 \in \Omega_1$, the set of arguments $\overline{c}_2^2 (\mathrm{mod}\ c_1)$, with $c_2 \in \Omega_2$ are very well distributed amongst the congruence classes of the c_1. Since the congruence classes are essentially well covered one can now invoke the vertical Sato/Tate law VST or its variant VST(k_1) to finish the proof of equidistribution.

— A harder case is when $X_2 \leqslant X_1$, in that case not all congruence classes $\mathrm{mod}\ c_1$ are covered and one has to prove that the Vertical Sato/Tate law still holds for restricted subsets:

For example assuming that $c_1 = p$ is a prime, by Weyl's criterion, one needs to prove that for any $k \geqslant 1$

$$\frac{1}{|\Omega_2|} \sum_{c_2 \in \Omega_2} sym_k(\theta_{p,\overline{c}_2^2}) \rightarrow 0,$$

which by a slight abuse of notation is rewritten as

$$\frac{1}{|\Omega_2|} \sum_{c_2 \in \Omega_2} \mathrm{tr}(\mathrm{Frob}_{\overline{c}_2^2} | \mathrm{Sym}_k \mathcal{K}l) \rightarrow 0.$$

Here we have identified $\mathbf{F}_p^\times$ 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 analytic number theory). Of course, in order to prove such estimate we cannot apply directly the methods from algebraic geometry but after several (more or less complicated) transformations, this becomes possible.

The simplest example of a VST over a restricted subset is the case of a short interval:

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

To obtain this result, one express the characteristic function of the integers in the interval $[1, p^{1/2+\varepsilon}]$ in term of the additive characters mod p. Then the Polya/Vinogradov completion method reduce the estimate of the corresponding Weyl sums to prove that

$$\sum_{a \in \mathbf{F}_p^\times} \mathrm{tr}(\mathrm{Frob}_a | \mathrm{Sym}_k \mathcal{K}l) \psi(a) \ll_k p^{1/2}$$

uniformly for ψ ranging over the additive characters of $\mathbf{F}_p$. In particular such bounds lead to the consideration of another family of 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 (easy) algebro-geometric input here, is the (simple) fact that a geometrically irreducible sheaf of rank >1 (ie. $\mathrm{Sym}^k\mathcal{K}l$) remains geometrically irreducible if it is twisted by any sheaf of rank 1 (ie. $\mathcal{L}_\psi$).

Another example is that of the set of 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. After these transformations are performed one need to use the previous variant and another form of the VST

Proposition. *For any $\varepsilon>0$, as $p\to+\infty$, and for any $b\in\mathbf{F}_p-\{0,1\}$, the set of pairs Kloosterman angles $\{(\theta_{p,a},\theta_{p,ba})\}_{1\leqslant a\leqslant p^{1/2+\varepsilon}}$ becomes equidistributed on $[0,\pi]\times[0,\pi]$ for the product of Sato/Tate measures.*

To prove the last proposition, the new sheave to be considered is the Rankin/Selberg sheaf

$$\mathrm{Sym}^k\mathcal{K}l\otimes\mathrm{Sym}^k[b]^*\mathcal{K}l$$

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

Proposition. *If $b\neq1$, the geometric monodromy group of $\mathcal{K}l\oplus[b]^*\mathcal{K}l$ 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 $[b]^*\mathcal{K}l$.

More elaborated transformation can be used, yielding to other geometric statements: for example, at some point one use the following proposition

Proposition. *Given $p>2$, $m_1,m_2\in\mathbf{F}_p^\times$, and let U be the open subset $U:=\mathbf{G}_{m,\mathbf{F}_p}-\{m_1,m_2\}$ Consider the four morphisms $f_1,\ f_2,\ f_3,\ f_4:U\longrightarrow\mathbf{G}_{m,\mathbf{F}_p}$ given by*

$$f_1(T)=\big(m_1(m_1-T)\big)^{-2},\quad f_2(T)=\big(T(m_1-T)\big)^{-2}$$
$$f_3(T)=\big(m_2(m_2-T)\big)^{-2},\quad f_4(T)=\big(T(m_2-T)\big)^{-2},$$

then the geometric monodromy group of the Sheaf

$$f_1^*\mathcal{K}l\oplus f_2^*\mathcal{K}l\oplus f_3^*\mathcal{K}l\oplus f_4^*\mathcal{K}l$$

is as big as possible, ie. equals

$$SL_2\times SL_2\times SL_2\times SL_2.$$

Almost 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 — mostly by Katz —) such example include higher dimensional Kloosterman sums

$$Kl_n(a;p) = \sum_{\substack{x_1x_2\ldots x_n=a\\ x_i\in\mathbf{F}_p^\times}} e(\frac{x_1+x_2+\cdots+x_n}{p}),$$

more general hypergeometric sum or exponential sums obtained by geometric Fourier transforms

$$S_{x^3+x}(a;p) := \sum_{x\in\mathbf{F}_p} e(a\frac{x^3+x}{p})$$

or

$$S_f(a;p) = \sum_{x\in\mathbf{F}_p} e(a\frac{f(x)}{p})$$

where $f(X)\in\mathbf{Z}[X]$ is an irreducible polynomial of degree n with maximal Galois group (ie $\simeq\Sigma_n$).

7. Spectral theory of Kloosterman sums

For $q\geqslant 1$ et $k\geqslant 2$ even, let $S_k(q)$ be the space of holomorphic cusp forms of weight k and level q: $f:\mathbf{H}\to\mathbf{C}$,

$$f(\frac{az+b}{cz+d}) = (cz+d)^k f(z),\ \begin{pmatrix} a & b\\ c & d\end{pmatrix}\in\Gamma_0(q)\subset SL_2(\mathbf{Z}),\ (c\equiv 0(q))$$

The Petersson inner product is given by

$$\langle f,f\rangle := \int_{\Gamma_0(q)\backslash\mathbf{H}} y^k|f(z)|^2\frac{dxdy}{y^2} < +\infty,$$

Let

$$f(z+1) = f(z) = \sum_{n\geqslant 1}\rho_f(n)n^{\frac{k-1}{2}}e(nz)$$

be the Fourier expansion of $f\in S_k(q)$. For $\mathcal{B}_k(q)$ an orthonormal basis of $S_k(q)$, Petterson's formula is

$$\frac{\Gamma(k-1)}{(4\pi)^{k-1}}\sum_{f\in\mathcal{B}_k(q)}\overline{\rho_f}(m)\rho_f(n) = \delta_{m,n} - 2\pi i^k\sum_{c\equiv 0(q)}\frac{Kl(m,n;c)}{c}J_{k-1}(\frac{4\pi\sqrt{mn}}{c}).$$

In particular, Weil's bound for Kloosterman sums implies that for $\sigma > 1/2$

$$\frac{\Gamma(k-1)}{(4\pi)^{k-1}}\sum_{f\in\mathcal{B}_k(q)}\overline{\rho_f}(m)\rho_f(n) = \delta_{m,n} + O((m,n,q)^{1/2}(\frac{\sqrt{mn}}{q})^{\sigma}),$$

hence taking $m = n$ above and $\sigma = 1/2 + \varepsilon$ one gets

$$|\rho_f(n)| \ll_f n^{1/4+\varepsilon}$$

which is a non trivial (the trivial bound being $|\rho_f(n)| \ll_f n^{1/2+\varepsilon}$); of course much better bound are now available thanks to Deligne but it is nevertheless interesting to note that any non-trivial estimate for Kloosterman sum is already sufficient to bound the Fourier coefficients non-trivially: for instance Kloosterman's original bound gives $|\rho_f(n)| \ll_f n^{3/8+\varepsilon}$.

For us the most useful version of this formula is Kloosterman's generalization: consider the L^2-space of functions on the (punctured) modular curve $\Gamma_0(q)\backslash\mathbf{H}$ equipped with the Petersson inner product,

$$\langle f, g\rangle = \int_{\Gamma_0(q)\backslash\mathbf{H}} \overline{f}(z)g(z)\frac{dxdy}{y^2}.$$

Then $L^2(\Gamma_0(q)\backslash\mathbf{H})$ has is decomposed spectrally by eigenfunction of the hyperbolic Laplace operator (Maass forms, which we may also choose to be Hecke-eigenforms)

$$\Delta = -y^2(\frac{\partial^2}{\partial^2 x} + \frac{\partial^2}{\partial^2 y})$$

$$L^2(\Gamma_0(q)\backslash\mathbf{H}) = \mathbf{C} \oplus \bigoplus_{j\geqslant 1} \mathbf{C}.u_j \oplus \bigoplus_{\mathfrak{a}} \frac{1}{4\pi}\int_{\mathbf{R}} \mathbf{C}.E(., 1/2 + it)dt.$$

If u_j is such a Maass (cusp) form satisfying say

$$\Delta.u_j = (\frac{1}{4} + t_j^2)u_j,$$

its Fourier expansion is given by

$$u_j(z) = \sum_{n\neq 0} \rho_j(n)W_{0,it_j}(4\pi|n|y)e(nx).$$

The Petersson/Kuznetzov formula relates sums of Kloosterman sums to Fourier coefficients of modular forms:

Given $\varphi \in C_c^\infty((0,\infty))$ a test function, and $m, n \geqslant 1$ one has

$$\frac{1}{4}\sum_{c\equiv 0\,(q)} \frac{Kl(m,n;c)}{c}\varphi\left(\frac{4\pi\sqrt{mn}}{c}\right) = \sum_{k\equiv 0\,(2)} \Gamma(k-1)\tilde{\varphi}(k-1)\sum_{f\in\mathcal{B}_k(q)} \overline{\rho_f}(m)\rho_f(n)$$
$$+\sum_{j\geqslant 1}\frac{\hat{\varphi}(t_j)}{\cosh(\pi t_j)}\overline{\rho_j}(m)\rho_j(n) + \frac{1}{4\pi}\sum_{\mathfrak{a}}\int_{-\infty}^{+\infty}\frac{\hat{\varphi}(t)}{\cosh(\pi t)}\overline{\rho_{\mathfrak{a}}}(m,t)\rho_{\mathfrak{a}}(n,t)\,dt.$$

There are variants of this formula for forms of weight one and forms of half integral weight and as E. Lapid explained in his lectures all these formulae are better understood in the context of the relative trace formula. Note that in this later context the ultimate goal is to use the connection

$$\text{Spectral Data } (= \text{Fourier coefficient of automorphic forms}) \\ \Leftrightarrow \text{ Geometric Data } (= \text{Sums of Kloosterman sums})$$

to compare Spectral Data associated to very different groups. In analytic number theory (so far) this connection is used, yet in both directions, but only at the level of a single group. This is what analytic number theorists call (after Martin Huxley) **Kloostermania.** In the sequel we will describe some recent applications of Kloostermania and in particular make some links with the first part of this survey.

The most obvious application of Kloosterman sums is as above to provide estimate for Fourier coefficients of automorphic form (and more generally) to give estimates for the spectral parameters of the associated automorphic representations. As above, Weil's bound can be used along with Kuznetzov's formula to prove that for a Hecke eigenform $g(z)$

- $|\lambda_g(p^\alpha)| \leqslant 2p^{\alpha/4}$, and
- $|\Im m t_g| \leqslant 1/4$ if f is a Maass form with Laplace eigenvalue

$$\lambda_g = (1/2 + it_g)(1/2 - it_g).$$

The last bound is due to Selberg and was also obtained by Gelbart/Jacquet as a consequence of the existence of the adjoint square lift. Of course due to the recent progress on the functioriality conjecture by Kim and Shahidi [KS02, Kim03], one can do much better: with $1/4$ replaced by $7/64$ (Kim/Sarnak).

However Kuznetzov's formula is powerful to control *linear combinations* of Fourier coefficients (rather than individual ones), ie. sums of the form

$$\sum_{|t_j|\leqslant T} \frac{1}{\cosh(\pi t_j)} |\sum_{n\leqslant N} a_n \rho_j(n)|^2;$$

such sums are expressed in terms of Kloosterman sums

$$\sum_c \sum_{m,n} \overline{a_m} a_n \frac{Kl(m,n;c)}{c} \varphi_T\Big(\frac{\sqrt{mn}}{c}\Big),$$

and depending on the case, one may either use Weil's bound or use the shape of Kloosterman sums to factor that expression further as

$$\sum_c \frac{1}{c} \sum_{\substack{a(c)\\(a,c)=1}} \Big(\sum_m \overline{a_m} e\Big(\frac{am}{c}\Big)\Big)\Big(\sum_n a_n e\Big(\frac{\overline{a}n}{c}\Big)\Big).$$

Combining this with the classical large sieve inequality one gets a general large sieve inequality for linear combination of Fourier coefficients of modular forms (which for some applications is stronger than the Ramanujan/Petersson conjecture)

$$\sum_{|t_j|\leqslant T}\frac{1}{\cosh(\pi t_j)}|\sum_{n\leqslant N}a_n\rho_j(n)|^2\ll(T^2+\frac{N}{q})\sum_{n\leqslant N}|a_n|^2.$$

8. Dimension of the space of Modular forms of weight 1

Let $q>1$ et $\chi:(\mathbf{Z}/q\mathbf{Z})^\times\to\mathbf{C}$ an odd Dirichlet character, we denote by $S_1(q,\chi)$ the space of holomorphic forms of weight 1 level q and nebentypus χ: ie. of the $f(z)$ such that

$$f(\frac{az+b}{cz+d})=\chi(d)(cz+d)f(z),\ \begin{pmatrix}a&b\\c&d\end{pmatrix}\in\Gamma_0(q).$$

Contrarily to the case of forms of weight $k\geqslant 2$, there is no way to compute explicitly the dimension of this space.

In fact, there should be only very few of such forms: Serre conjectured that $\dim S_1(q,\chi)\ll_\varepsilon q^{1/2+\varepsilon}$. In this direction, the trace formula would give (at best) the upper bound $\dim S_1(q,\chi)\ll q/\log q$ but the first really non-trivial result is due to W. Duke [Duk95] who proved that

$$\dim S_1(q,\chi)\ll_\varepsilon q^{1-1/12+\varepsilon}.$$

Duke's method was geometric in nature, in [MV02] the spectral theory of automorphic forms has been used to give a further improvement

Theorem 8.1.

$$\dim S_1(q,\chi)\ll_\varepsilon q^{1-1/7+\varepsilon}.$$

In particular since χ is of order divisible by 60, one even has

$$\dim S_1(q)\ll_\varepsilon q^{1-1/7+\varepsilon}.$$

One of the key point of the proof, already used by Duke (and proved by Deligne/Serre [DS74]), is that any such Hecke-eigenform, f say, is associated with a complex two dimensional complex Galois representation ρ_f, such that the Hecke-eigenvalues of f at a prime $p\not| q$ equals

$$\lambda_f(p)=\mathrm{tr}(\rho_f\mathrm{Frob}_p),\ \chi(p)=\det(\rho_f\mathrm{Frob}_p).$$

The outcome is that (since the -finite- images of 2-dim. complex Galois representation are classified) there exists complex numbers $a_2,\ a_8,\ a_{12}$ of modulus at most 1 (depending only on χ) such that for any $f\in S_1(q,\chi)$, one has the linear relation

$$a_{12}\rho_f(p^{12})+a_8\rho_f(p^8)+a_2\rho_f(p^2)=\rho_f(1).$$

Hence

$$\sum_{f\in\mathcal{B}_1(q,\chi)} |\rho_f(1)|^2|\{p\leqslant N,\ (p,q)=1\}|^2$$

$$= \sum_{f\in\mathcal{B}_1(q,\chi)} |\sum_{p\leqslant N} a_{12}\rho_f(p^{12})+a_8\rho_f(p^8)+a_2\rho_f(p^2)|^2;$$

There is of course no Petersson trace formula for $S_1(q,\chi)$ enabling to estimate the above sum but we observe that if f belong to $S_1(q,\chi)$ then $y^{1/2}f(z)$ is Maass form of weight 1 and eigenvalue $1/4$ for the Laplace operator of weight 1

$$\Delta_1 = -y^2(\frac{\partial^2}{\partial^2 x}+\frac{\partial^2}{\partial^2 y})+iy\frac{\partial}{\partial x}.$$

Hence, we may embed the basis $\mathcal{B}_1(q,\chi)$ in an orthonormal basis of Maass form of weight 1. Then by positivity, one can bound the sum above by the corresponding full spectral sum (including all Maass form and Eisenstein series). This is a priori wasteful but not too much thanks to Kuznetzov formula and Weil's bound for the corresponding Kloosterman sums.

This way we obtain

$$\sum_{f\in\mathcal{B}_1(q,\chi)} |\rho_f(1)|^2(\frac{N}{\log N})^2 \ll \frac{N}{\log N}+\frac{N^6}{q}(\frac{N}{\log N})^2$$

we conclude by choosing N optimally ($N=q^{1/7}$) and by observing that

$$\sum_{f\in\mathcal{B}_1(q,\chi)} |\rho_f(1)|^2 \gg \dim S_1(q,\chi)q^{\varepsilon-1}.$$

Remark that in the proof above, we have used the linear relation (valid a priori only for Maass forms of Galois type)

$$a_{12}\rho_f(p^{12})+a_8\rho_f(p^8)+a_2\rho_f(p^2)=\rho_f(1),$$

to (coarsely) detect (and isolate) automorphic Maass forms of weight 1 of Galois type from all other Maass forms, thus the above estimate can be interpreted as a very coarse and very primitive manifestation of Langlands "Beyond endoscopy" program.

9. The Shifted Convolution Problem

Given f and g two modular forms, the **Shifted Convolution Problem (SCP)** consist in evaluating the sum

$$S(f,g,;h)=\sum_{m\pm n=h}\rho_f(m)\rho_g(n)F(\frac{m}{M},\frac{n}{N})$$

where $h\neq 0$, $F(x,y)$ is a test function and M,N are parameters going to $+\infty$.

Note that for $h=0$ the sum becomes

$$\sum_m \rho_f(m)\rho_g(m)F(\frac{m}{M},\frac{m}{N})$$

and so is closely related to the Rankin/Selberg L-function of the pair (f,g).

Observe that if one takes for f and g, the theta series associated to two definite positive binary quadratic forms $Q(x,y)$, $R(x,y)$, and one choose $\pm=+$ and for F the constant function 1, the SCP becomes the problem of evaluating the number of the representations of h by the quaternary quadratic form $Q(x,y)+R(x',y')$ which was Kloosterman's original motivation.

Kloosterman original way for solving this instance of the SCP, was to use the circle method: since

$$\int_{[0,1]} e(h\alpha)d\alpha=\delta_{h=0},$$

one has

$$S(f,g,h)=\int_{[0,1]} S(f,g;\alpha)e(-h\alpha)d\alpha$$

$$S(f,g;\alpha)=\sum_m \rho_f(m)e(m\alpha)\sum_n \rho_g(n)e(\pm n\alpha)F(\frac{m}{M},\frac{n}{N}).$$

The essence of the circle method consists in replacing (possibly up to a good error term) the above integral over $[0,1]$ by a discrete weighted average of the function $S(f,g;\alpha)e(-h\alpha)$ over points α ranging over a subset of the set of rational numbers in $[0,1]$, with denominator bounded by some parameter C, $\{\frac{a}{c},\ (a,c)=1,\ c\leqslant C\}\subset[0,1]$. In the specific case of the SCP, one can then use the modular properties of f and g to evaluate the corresponding sums $S(f,g;\frac{a}{c})$ which makes eventually Kloosterman sums appear. In Kloosterman's original context, this approximation step is called Kloosterman's refinement but nowadays more flexible treatments are available (the δ-symbol method of Friedlander/Iwaniec or Jutila's method of overleaping intervals [DFI94a, Jut97])

The outcome is a sum for the following shape

$$\frac{1}{L}\sum_c w(\frac{c}{C})\sum_{\substack{a(c)\\(a,c)=1}} e(h\frac{a}{c})\sum_m \rho_f(m)e(m\frac{a}{c})\sum_n \rho_g(n)e(n\frac{a}{c})G(m,n,h;c)$$

where w and G are compactly supported test functions and L is total weight of the summation

$$L=\sum_c w(\frac{c}{C})\varphi(c).$$

Now the modularity of f and g is exploited by means of the Voronoi summation formula which has the roughly shape

$$\sum_m \rho_f(m)e(m\frac{a}{c})G(m)=\frac{1}{c}\sum_m \rho_f(m)e(-m\frac{\overline{a}}{c})\widehat{G}(m)$$

where G is a test function and $\widehat{G}$ is a Bessel transform of G (which depends of the infinity type of g).

Applying such formula to f and g, and summing over $a(\mathrm{mod}\ c)$ Kloosterman sums appear since the sum becomes

$$\frac{1}{L}\sum_c w(\frac{c}{C})\sum_m \rho_f(m)\sum_n \rho_g(n)\frac{Kl(m\pm n,h;c)}{c}\widehat{G}(m,n,h;c)$$

with possibly an extra principal main term (if f and g are not cuspidal). Now applying Weil's bound one deduce a non trivial upper bound for (the error term of) the Shifted Convolution Sum. This is essentially the path Kloosterman followed to evaluate the number of representations of a integer by a quaternary quadratic form.

9.1. **Back to Spectral theory.** Nowadays, one can get estimates which are stronger than the ones provided by Weil's bound. Indeed pursuing the transformation further, one can apply Kuznetzov trace formula but **backwards** (from Kloosterman sum to Fourier coefficients of automorphic forms) getting a

$$\frac{1}{L}\sum_j \frac{1}{\cosh(\pi t_j)}\overline{\rho}_j(h)\Big(\sum_{|h'|\leqslant H'} a_{f,g}(h')\rho_j(h')\Big)\widetilde{G}(h',h;t_j) + \text{Holomorphic} + \text{Eisenstein}$$

where $a_{f,g}(h')$ is of the shape

$$a_{f,g}(h') = \sum_{\substack{m\pm n=h'\\ |m|,|n|\leqslant H'}} \rho_f(m)\rho_g(n).$$

Note that the coefficient $a_{f,g}(h')$ is again a Shifted convolution Sum; however, in practice the range of the variable H' is quite different from the initial range (shorter) and moreover the $a_{f,g}(h')$ need only to be bounded on average over h' (for instance by using the Large Sieve Inequality for Fourier coefficients of automorphic forms), so the argument is not circular. At this point we may take advantage of the fact that much better bounds are available for the Fourier coefficients $\rho_j(h)$ than the ones provided through Weil's bound:

$$|\rho_j(h')| \ll_\varepsilon |\rho_j(1)|n^{7/64+\varepsilon}.$$

There is a further advantage of this approach, as on several occasions the SCP occurs with a extra averaging over the h variable: one may need to evaluate

$$\sum_{h\leqslant H}\chi(h)S(f,g;h)$$

where $\chi(h)$ are oscillating complex numbers. On these occasions, even the bound provided by the Ramanujan/Petersson conjecture for individual h may not be sufficient to bound the averaged shifted convolution sum properly; however going back to the previous expression our averaged sum becomes

$$\frac{1}{L}\sum_j \frac{1}{\cosh(\pi t_j)}\big(\sum_h \chi(h)\overline{\rho}_j(h)\big)\big(\sum_{h'} a_{f,g}(h')\rho_j(h')\big)\widetilde{G}(h',h;t_j)+$$
$$\text{Holomorphic contribution}+\text{ Eisenstein contribution.}$$

then one can hope to get extra cancellation by estimating non-trivially the sum

$$\sum_h \chi(h)\overline{\rho}_j(h)$$

for each j: interesting example of such coefficients are given by Dirichlet characters or Fourier coefficients of modular forms.

10. Application to the Subconvexity problem

One of the most important recent application of the SCP is the resolution of the Subconvexity problem for GL_2-automorphic L-functions: here we state the problem only for the conductor aspect.

Given π an automorphic representation of $GL_n(\mathbf{A}_\mathbf{Q})$ of conductor q, the **convexity bound** is the following

$$\text{for } \Re es = 1/2,\ L(\pi, s) \ll_{\varepsilon,\pi_\infty,s} q^{1/4+\varepsilon}.$$

The subconvexity problem (**ScP**) consists in replacing the exponent $1/4$ by one strictly smaller. This problem as many applications, in particular to Linnik's equidistribution problems on modular or Shimura curves [Duk88, Mic04].

The SCP has been the key for the resolution of many instance of the ScP (notably through the work of Duke/Friedlander/Iwaniec [DFI93, DFI94b, DFI01, DFI02]). Recently, in joint work with G. Harcos, we could solve the problem for Rankin/Selberg L-functions in great generality [HM04]:

Theorem 10.1. *Let g be a fixed cusp form and f be a cusp form of level q (say). Then, if the product of the nebentypus $\chi_f\chi_g$ is not trivial, one has*

$$L(f\otimes g, s) \ll q^{1/2-1/2700}.$$

This result is obtained by solving an averaged version of the SCP. It has for application the equidistribution of short orbits of Heegner point of Shimura curves associated to indefinite quaternion algebras over $\mathbf{Q}$.

Remark that in the modern approaches to the SCP, Kloosterman sums acts merely as a catalyst : either Kuznetzov trace formula, or some transformations make them appear but a backwards application of the Kuznetzov formula make them disappear afterwards. This suggest that Kloosterman sums could be avoided for some problems. This is indeed, the case of the ScP: recently A. Venkatesh, [Ven05] found a different approach to the ScP which builds only on the realization of the L-function in terms of a (square of a) period. In particular his approach avoids entirely the use of Fourier coefficients and Kloosterman sums. In fact, his method, which is based on the ergodic properties of Hecke operators, works smoothly over an arbitrary number field ! Amongst other cases, Venkatesh could solve the ScP for

- the standard L-function of a $GL_2(\mathbf{A}_F)$-automorphic representation with trivial nebentypus (a large conductor).
- the Rankin/Selberg L-function of a pair of $GL_2(\mathbf{A}_F)$-automorphic representations with trivial nebentypus, one being fixed, the other having large conductor.

More recently, A. Venkatesh and I were able to remove the assumptions on the nebentypus above: the proof has several difference with Venkatesh's one (in particular, it does not use -directly- ergodicity of Hecke operators); in fact the principal extra ingredient was inspired by the resolution of the ScP for Rankin/Selberg L-functions over $\mathbf{Q}$ discussed above [MV05]. A consequence of these results is the resolution of the ScP for modular Artin L-functions over a number field and to the equidistribution of Galois orbits of special points on quaternionic Shimura varieties over totally real fields (which is meaningful in the context of the Andre/Oort conjectures).

11. Algebraic and modular aspect of Kloosterman sums combined

We conclude this survey by two further application which combine both of the modular and the algebraic aspects of Kloosterman sums discussed so far.

11.1. Number variance on the modular surface and the error term in Weyl's law. The main motivation of Selberg for developing the trace formula was for proving the existence of cusp forms. In particular for the full modular curve he obtained the Weyl law: for $T \geqslant 1$

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

here MainTerm is well understood and asymptotic to

$$\mathrm{MainTerm} \simeq \frac{vol(SL_2(\mathbf{Z})\backslash\mathbf{H}}{4\pi}T^2;$$

on the other hand $S(T)$ is rather small error term ($= O(\frac{T}{\log T})$) but its asymptotic properties are not so well understood. Selberg also established a lower bound for the variance of this error term showing that this error term is often not too small:

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

Until recently there has been no progress at all concerning the evaluation of this variance. However last year X. Li and P. Sarnak [LS05], by using the Petersson/Kuznetzov formula – instead of Selberg's trace formula –and the lower bound (along with other analytic techniques)

$$X\frac{\exp((\log\log X)^{5/17})}{\log X} \ll \sum_{c\leqslant X}\frac{|Kl(1,1;c)|^2}{c}$$

were able to make the first (modest but meaningful) improvement over Selberg's lower bound namely:

$$\int_T^{2T}|S(t)|^2dt \gg \frac{T^2}{\log^2 T}\exp((\log\log T)^{5/17}).$$

In fact, they provided further refinements of this estimate. These refinement can be interpreted in term of Quantum Chaos: while it is expected that the local scaled spacing distributions of the Laplace eigenvalues of a generic hyperbolic surface should be Gaussian, it is expected that for arithmetic surfaces this distribution should be Poissonian. The results of Li and Sarnak then show that the distribution in the case of the modular surface is definitely not Gaussian and give the first hint for a Poissonian behavior: as we have seen, this is the consequence of the arithmetical structure of the Kloosterman sums associated to congruence subgroup of $SL_2(\mathbf{Z})$.

11.2. **On the sign of Kloosterman sums.** For this concluding application we return to the question raised in section 5 on the existence of sign changes of Kloosterman sums. Here we shall use Kuznetzov's formula in the opposite directon (from automorphic forms to Kloosterman sums). Indeed by using his formula (and Roelcke lower bound on the first eigenvalue λ_1 of $SL_2(\mathbf{Z})\backslash\mathbf{H}$), Kuznetzov proved the bound

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

which is clearly non trivial by comparison with the bound provided by Weil's estimate and gave the first non-trivial result towards Linnik's conjecture (that the sum above should be $\ll X^{1/2+\varepsilon}$); however until recently (as was remarked by Serre) this was not evident whether this estimate accounted for the existence of sign changes amongst the $Kl(1,1;c)$ or for an extraordinary uniform smallness of Kloosterman sums. However thanks to the lower bounds of section 6.1 one has

Proposition. *There are infinitely many integers c such that*

$$Kl(1,1;c) > 0, \quad \textit{(resp. } Kl(1,1;c) < 0\textit{)}.$$

In fact the number such integers which are less than X is at least $\gg X/\log X$.

The next step consists in limiting the number of allowed prime factors of the moduli c above and thus to prove an Horizontal type result for Kloosterman sums with almost prime moduli : the following theorem solve the first basic question concerning HST(k) [FM02]:

Theorem 11.1. *There exists infinitely many squarefree c (a positive proportion in fact) having at most 23 prime factors such that $Kl(1,1;c) > 0$ (resp. $Kl(1,1;c) < 0$).*

Sketch of Proof. Refining the lower bounds of the first lecture one can prove that $(u_0 = 1/23.9)$

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

Thus it is sufficient to show that

$$|\sum_{\substack{c \leqslant X \\ p|c \Rightarrow p \geqslant X^{1/u_0}}} \mu^2(2c)\frac{Kl(1,1;c)}{\sqrt{c}}| \leqslant 0.1659999\frac{X}{\log X}. \tag{2}$$

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 that when it works, the sieve provides bounds of the shape

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

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

In the present case (to force positivity) we need to sieve the two sequences

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

Morever a necessary condition for the sieve to work is to control such sequences well in arithmetic progressions to large moduli: this lead 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's formula for $\Gamma_0(q)$ and of the 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): in that way we have good control for q up to size $X^{1/2-\varepsilon}$ (this is an analog of the Bombieri/Vinogradov theorem):

Proposition. *For any $\varepsilon > 0$, $Q \leqslant X^{1/2-\varepsilon}$ and any $B > 0$, one has*

$$\sum_{q \leqslant Q} |\sum_{\substack{c \leqslant X \\ c \equiv 0(q)}} \frac{Kl(1,1;c)}{\sqrt{c}}| \ll_{\varepsilon,B} \frac{X}{\log^B X}.$$

Then, applying to this situation, (a variant of) the sieve of Selberg, one deduce that

$$(3)\qquad \sum_{\substack{c\leqslant X\\ p|c\Rightarrow p\geqslant X^{1/u_0}}} \mu^2(2c)\big(2^{\omega}(c)\pm\frac{Kl(1,1;c)}{\sqrt{c}}\big)\leqslant MT(u_0)\frac{X}{\log X}+0.1659999\frac{X}{\log X}$$

where $MT(u_0)\frac{X}{\log X}$ is a main term which can be proven to be equal to

$$MT(u_0)\frac{X}{\log X}=\sum_{\substack{c\leqslant X\\ p|c\Rightarrow p\geqslant X^{1/u_0}}}\mu^2(2c)2^{\omega}(c)\;+\;O(\frac{X}{\log^2 X}).$$

Subtracting this contribution from (3) we obtain the desired estimate (2).

References

[BFI86] E. Bombieri, J. B. Friedlander, and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), no. 3-4, 203–251.

[BFI89] ______, *Primes in arithmetic progressions to large moduli. III*, J. Amer. Math. Soc. **2** (1989), no. 2, 215–224.

[Boo00] Andrew R. Booker, *A test for identifying Fourier coefficients of automorphic forms and application to Kloosterman sums*, Experiment. Math. **9** (2000), no. 4, 571–581.

[Del74] Pierre Deligne, *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math. (1974), no. 43, 273–307.

[Del80] ______, *La conjecture de Weil. II*, Inst. Hautes Études Sci. Publ. Math. (1980), no. 52, 137–252.

[DS74] Pierre Deligne and Jean-Pierre Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. (4) **7** (1974), 507–530 (1975).

[DI82] J.-M. Deshouillers and H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. Math. **70** (1982/83), no. 2, 219–288.

[Duk88] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. Math. **92** (1988), 73–90.

[Duk95] ______, *The dimension of the space of cusp forms of weight one*, Internat. Math. Res. Notices (1995), no. 2, 99–109 (electronic).

[DFI93] W. Duke, J. Friedlander, and H. Iwaniec, *Bounds for automorphic L-functions*, Invent. Math. **112** (1993), no. 1, 1–8.

[DFI94a] ______, *A quadratic divisor problem*, Invent. Math. **115** (1994), no. 2, 209–217.

[DFI94b] ______, *Bounds for automorphic L-functions. II*, Invent. Math. **115** (1994), no. 2, 219–239.

[DFI95] W. Duke, J. B. Friedlander, and H. Iwaniec, *Equidistribution of roots of a quadratic congruence to prime moduli*, Ann. of Math. (2) **141** (1995), no. 2, 423–441.

[DFI01] ______, *Bounds for automorphic L-functions. III*, Invent. Math. **143** (2001), no. 2, 221–248.

[DFI02] ______, *The subconvexity problem for Artin L-functions*, Invent. Math. **149** (2002), no. 3, 489–577.

[Fou84] Étienne Fouvry, *Autour du théorème de Bombieri-Vinogradov*, Acta Math. **152** (1984), no. 3-4, 219–244.

[FM02] Etienne Fouvry and Philippe Michel, *Sur le changement de signe des sommes de Kloosterman*, Preprint (2002).

[FM03] ______, *Sommes de modules de sommes d'exponentielles*, Pacific J. Math. **209** (2003), no. 2, 261–288.

[HM04] G. Harcos and P. Michel, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points II*, Invent. Math. (to appear) (2004).

[HBP79] D. R. Heath-Brown and S. J. Patterson, *The distribution of Kummer sums at prime arguments*, J. Reine Angew. Math. **310** (1979), 111–130.

[Jut97] M. Jutila, *A variant of the circle method*, Sieve methods, exponential sums, and their applications in number theory (Cardiff, 1995), 1997, pp. 245–254.

[Kat88] Nicholas M. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Mathematics Studies, vol. 116, Princeton University Press, Princeton, NJ, 1988.

[Kim03] Henry H. Kim, *Functoriality for the exterior square of* GL_4 *and the symmetric fourth of* GL_2, J. Amer. Math. Soc. **16** (2003), no. 1, 139–183 (electronic). With appendix 1 by Dinakar Ramakrishnan and appendix 2 by Kim and Peter Sarnak.

[KS02] Henry H. Kim and Freydoon Shahidi, *Functorial products for* $\mathrm{GL}_2 \times \mathrm{GL}_3$ *and the symmetric cube for* GL_2, Ann. of Math. (2) **155** (2002), no. 3, 837–893. With an appendix by Colin J. Bushnell and Guy Henniart.

[Klo26] H. Kloosterman, *On the representations of numbers in the forms* $ax^2+by^2+cz^2+dt^2$, Acta Math. **49** (1926), 407–464.

[Kuz80] N. V. Kuznecov, *The Petersson conjecture for cusp forms of weight zero and the Linnik conjecture. Sums of Kloosterman sums*, Mat. Sb. (N.S.) **111(153)** (1980), no. 3, 334–383, 479.

[Mic95] P. Michel, *Autour de la conjecture de Sato-Tate pour les sommes de Kloosterman. I*, Invent. Math. **121** (1995), no. 1, 61–78.

[Mic04] ______, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Ann. of Math. (2) **160** (2004), no. 1, 185–236.

[MV02] Philippe Michel and Akshay Venkatesh, *On the dimension of the space of cusp forms associated to 2-dimensional complex Galois representations*, Int. Math. Res. Not. (2002), no. 38, 2021–2027.

[MV05] P. Michel and A. Venkatesh, *Periods, subconvexity and equidistribution*, In preparation (2005).

[LS05] X. Li and P. Sarnak, *On the number variance for* $SL_2(\mathbf{Z})\backslash\mathbf{H}$, GAFA (to appear) (2005).

[Ven05] A. Venkatesh, *Sparse equidistribution problems, period bounds, and subconvexity.*, Preprint (2005).

[Wei48] André Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U. S. A. **34** (1948), 204–207.