

SOME APPLICATIONS OF SMOOTH BILINEAR FORMS WITH KLOOSTERMAN SUMS

VALENTIN BLOMER, ÉTIENNE FOUVRY, EMMANUEL KOWALSKI, PHILIPPE MICHEL,
AND DJORDJE MILIĆEVIĆ

ABSTRACT. We revisit a recent bound of I. Shparlinski and T. P. Zhang on bilinear forms with Kloosterman sums, and prove an extension for correlation sums of Kloosterman sums against Fourier coefficients of modular forms. We use these bounds to improve on earlier results on sums of Kloosterman sums along the primes and on the error term of the fourth moment of Dirichlet L -functions.

1. STATEMENT OF RESULTS

1.1. Preliminaries. This note is motivated by a recent result of I. E. Shparlinski and T. P. Zhang [7] concerning bilinear forms with Kloosterman sums. Given a prime q and $m \in \mathbf{F}_q$, let

$$\mathrm{Kl}_2(m; q) := \frac{1}{\sqrt{q}} \sum_{\substack{x \in \mathbf{F}_q^\times \\ xy=1}} e_q(y + mx)$$

denote the normalized Kloosterman sum, where $e_q(x) = \exp(2\pi ix/q)$. Shparlinski and Zhang ([7, Theorem 3.1]) proved the following theorem.

Theorem 1.1 (Shparlinski–Zhang). *Let q be a prime number and let $\mathcal{M}, \mathcal{N} \subset [1, q-1]$ be intervals of lengths $M, N \geq 1$. Then we have*

$$(1.1) \quad \sum_{m \in \mathcal{M}} \sum_{n \in \mathcal{N}} \mathrm{Kl}_2(mn; q) \ll_\varepsilon q^\varepsilon \left(q^{1/2} + \frac{MN}{q^{1/2}} \right)$$

for any $\varepsilon > 0$, where the implied constant depends only on ε .

In light of the Weil bound for Kloosterman sums $|\mathrm{Kl}_2(m; q)| \leq 2$, the estimate (1.1) is non-trivial as long as MN is a bit larger than $q^{1/2}$. On the other hand, if M or N is close to q , other methods (e.g. the completion method) become more efficient. In particular, the restriction that M and N are $\leq q$ is not really restrictive for applications.

The aim of this paper is two-fold. On the one hand, we put Theorem 1.1 into a slightly more general context in Propositions 1.2 and 1.3; viewing it as a correlation estimate for Kloosterman sums and a divisor function (which itself is a Fourier coefficient of an Eisenstein series), it turns out to be a consequence of a version of the Voronoi summation formula. On the other hand, we give two applications of independent interest to the fourth moment of Dirichlet L -functions in Theorem 1.5

2010 *Mathematics Subject Classification.* 11M06, 11F11, 11L05, 11L40, 11F72, 11T23.

Key words and phrases. L -functions, modular forms, shifted convolution sums, Kloosterman sums, incomplete exponential sums.

V. B. was partially supported by the Volkswagen Foundation. É. F. thanks ETH Zürich and EPF Lausanne for financial support. Ph. M. was partially supported by the SNF (grant 200021-137488) and the ERC (Advanced Research Grant 228304). V. B., Ph. M. and E. K. were also partially supported by a DFG-SNF lead agency program grant (grant 200021L_153647). D. M. was partially supported by the NSF (Grant DMS-1503629) and ARC (through Grant DP130100674).

and sums of Kloosterman sums over primes in Theorem 1.8; these applications are discussed in Subsection 1.3.

1.2. Variations on a theme. Our first result is a smoothed version of the bound (1.1). To state it, we use the following class of smoothing functions. For a modulus $q \geq 1$ and a parameter $Q \geq 1$, we will consider functions satisfying the following conditions:

$$(1.2) \quad \begin{aligned} &W: [0, +\infty[\rightarrow \mathbf{C} \text{ is smooth, } \text{Supp}(W) \subset [1/2, 2], \\ &W^{(j)}(x) \ll_{j,\varepsilon} (q^\varepsilon Q)^j \text{ for any } x \geq 0, j \geq 0 \text{ and } \varepsilon > 0. \end{aligned}$$

Proposition 1.2. *Let q be a prime number and let $Q \geq 1$ be a real number. Let W_1, W_2 be functions satisfying (1.2). For any $M, N \geq 1$ and any integer a coprime with q , we have*

$$(1.3) \quad \sum_{m,n} W_1\left(\frac{m}{M}\right) W_2\left(\frac{n}{N}\right) \text{Kl}_2(amn; q) \ll_\varepsilon (qQ)^\varepsilon Q^2 \left(q^{1/2} + \frac{MN}{q^{1/2}}\right).$$

Furthermore, if W_3 also satisfies (1.2), then for any $Y \geq 1$, we have

$$(1.4) \quad \sum_{m,n} W_1\left(\frac{m}{M}\right) W_2\left(\frac{n}{N}\right) W_3\left(\frac{mn}{Y}\right) \text{Kl}_2(amn; q) \ll_\varepsilon (qQ)^\varepsilon Q^2 \left(q^{1/2} + \frac{MN}{q^{1/2}}\right).$$

In both cases, the implied constant depends only on ε .

The inequalities (1.3) and (1.4) could be easily deduced from the result of Shparlinski and Zhang by summation by parts with respect to the variables m and n . In §2, we will give an alternative proof based on [4, Prop. 2.2]. The Q -dependence in Proposition 1.2 is presented in a compact form well suited for our applications but it is not fully optimized otherwise (in particular, for Theorem 1.5 we will be using $Q = q^\varepsilon$); our proof actually yields a better Q -dependence in some other ranges.

We can view the bounds (1.3) and (1.4) essentially as sums over a single variable weighted by the divisor function d . The advantage of our proof of Proposition 1.2 is that it provides naturally an automorphic generalization, where the divisor function is replaced with Fourier coefficients of modular forms.

Proposition 1.3. *Let $(\lambda_f(n))_{n \geq 1}$ be the Hecke eigenvalues of a holomorphic cuspidal Hecke eigenform f of level 1, normalized so that $|\lambda_f(n)| \leq d(n)$. Let q be a prime number, and let W be a function satisfying (1.2) with $Q = 1$. Let a be an integer coprime to q . For any $N \geq 1$ and any $\varepsilon > 0$, we have*

$$(1.5) \quad \sum_{n \geq 1} \lambda_f(n) \text{Kl}_2(an; q) W\left(\frac{n}{N}\right) \ll_{\varepsilon,f} (qN)^\varepsilon \left(q^{1/2} + \frac{N}{q^{1/2}}\right)$$

where the implied constant depends only on f and ε .

Remark 1.4. This is by no means the most general statement that may be proved along these lines.

As pointed out in [7], the estimates (1.1) and (1.3) are significant improvements of the bound

$$(1.6) \quad \sum_{m,n} W_1\left(\frac{m}{M}\right) W_2\left(\frac{n}{N}\right) \text{Kl}_2(amn; q) \ll_{\varepsilon,Q} q^\varepsilon MN \left(1 + \frac{q}{MN}\right)^{1/2} q^{-1/8},$$

and likewise the estimate (1.5) improves significantly over

$$(1.7) \quad \sum_{n \geq 1} \lambda_f(n) \text{Kl}_2(an; q) W\left(\frac{n}{N}\right) \ll_{\varepsilon,Q} q^\varepsilon N \left(1 + \frac{q}{N}\right)^{1/2} q^{-1/8},$$

both of which were obtained by Fouvry, Kowalski and Michel as special cases of [2, Thm. 1.16] and [3, Thm. 1.2].

1.3. Applications. The bounds (1.6) and (1.7) have been applied recently in a number of problems, and the bounds (1.1) and (1.3) lead to further improvements. The main source for these improvements is the new input of Proposition 1.2, but a bit of extra work is necessary.

As a first application, we can improve our work on the error term for the fourth moment of Dirichlet series $L(s, \chi)$ of characters χ to a prime modulus q ([1, Theorem 1.1]).

Theorem 1.5. *There exists a polynomial $P_4 \in \mathbf{R}[X]$ of degree 4, such that*

$$\frac{1}{q-1} \sum_{\chi \pmod{q}} |L(\chi, 1/2)|^4 = P_4(\log q) + O(q^{-1/20+\varepsilon})$$

for all primes q , where the implied constant depends only on $\varepsilon > 0$. If the Ramanujan–Petersson conjecture holds for Fourier coefficients of Hecke–Maaß forms of level 1, then the exponent $1/20$ can be replaced by $1/16$.

Remark 1.6. In [1, Theorem 1.1], the exponents were respectively $1/32$ (unconditionally) and $1/24$ (assuming the Ramanujan–Petersson conjecture). The first breakthrough in this respect is due to M. Young [8] who obtained an asymptotic formula with exponents $5/512$ (resp. $1/80$).

Remark 1.7. The proof of Theorem 1.5 follows the same lines as [1, §6.3], except that instead of the bound [1, (5.5)] (i.e. (1.6) above) we use Proposition 1.2. It is of some interest to record here in outline how this improved exponent arises. The problem of the fourth moment leads to evaluating non-trivially the shifted convolution type sum

$$(1.8) \quad \sum_{\substack{m \asymp M, n \asymp N \\ m \equiv n \pmod{q}}} d(m)d(n) \approx \sum_{\substack{m_1, m_2, n_1, n_2 \\ m_1 m_2 \equiv n_1 n_2 \pmod{q}}} 1$$

with d the usual divisor function and with $MN = M_1 M_2 N_1 N_2 \approx q^2$. The spectral theory of automorphic forms provides a good error term when M and N are relatively close in the logarithmic scale. Otherwise, assuming that $N = N_1 N_2 \geq M = M_1 M_2$, we apply the Poisson summation formula to both variables n_1 and n_2 (equivalently, the Voronoi summation formula applied to the variable $n = n_1 n_2$), getting two variables of dual size $n_1^* \sim q/N_1$ and $n_2^* \sim q/N_2$ and a smooth quadrilinear sum of Kloosterman sums

$$\sum_{m_1, m_2, n_1^*, n_2^*} \text{Kl}_2(m_1 m_2 n_1^* n_2^*; q),$$

which is evaluated by various means, in particular using the smooth bilinear sum bound (1.3). In our specific case, the bound (1.3) amounts to applying the Poisson formula to two of the four variables m_1, m_2, n_1^*, n_2^* . This leads back to a sum of the type (1.8), which is then bounded trivially. This argument is not circular, and allows for an improvement, because we (implicitly) apply the process to variables different from the ones we started from (for instance to m_1 and n_1^* instead of n_1^* and n_2^*).

Our second application is an improvement of the first bound in [2, Cor. 1.13] for Kloosterman sums over primes in short intervals:

Theorem 1.8. *Let q be a prime number. Let $Q \geq 1$ be a parameter and let W be a function satisfying (1.2). Then for every X such that $2 \leq X \leq q$ and every $\varepsilon > 0$, we have*

$$(1.9) \quad \sum_{p \text{ prime}} W\left(\frac{p}{X}\right) \text{Kl}_2(p; q) \ll_{\varepsilon} q^{1/4+\varepsilon} Q^{1/2} X^{2/3}.$$

In addition, for every prime q , every X such that $2 \leq X \leq q$ and every $\varepsilon > 0$, we have

$$(1.10) \quad \sum_{\substack{p \leq X \\ p \text{ prime}}} \text{Kl}_2(p; q) \ll_{\varepsilon} q^{1/6+\varepsilon} X^{7/9}.$$

In both cases, the implicit constant depends only on ε .

Remark 1.9. The range where these bounds are non-trivial is the same as that in [2, Cor. 1.13], namely the length of summation X should be greater than $q^{3/4+\varepsilon}$ if Q is fixed. The improvement therefore lies in the greater cancellation in this allowed range. For instance, when $X = q$, we gain a factor $q^{1/18-\varepsilon}$ over the trivial bound for the sum appearing in (1.10) instead of $q^{1/48-\varepsilon}$ in [2, Corollary 1.13].

Acknowledgement. We would like to thank the referee for very useful suggestions that improved the presentation of the paper.

2. CORRELATION SUMS OF KLOOSTERMAN SUMS AND DIVISOR-LIKE FUNCTIONS

In this section, we revisit Theorem 1.1 and establish Proposition 1.2. The idea behind the proof of Theorem 1.1 is that after applying the completion method twice over the m and n variables, the Kloosterman sum $\text{Kl}_2(amn; q)$ is transformed into the Dirac type function $q^{1/2}\delta_{mn \equiv a \pmod{q}}$, and taking the congruence condition into account one saves (in the most favourable situation) a factor $q^{1/2}/q = q^{-1/2}$ over the trivial bound.

In our smoothed setting, the completion method is replaced by two applications of the Poisson summation formula or more precisely by a single application of the *tempered Voronoi summation formula* of Deshouillers and Iwaniec, in the form established in [4, Prop. 2.2].

Let q be a prime number, and let $K: \mathbf{Z} \rightarrow \mathbf{C}$ be a q -periodic function. The *normalized Fourier transform* of K is the q -periodic function on \mathbf{Z} defined by

$$\widehat{K}(h) = \frac{1}{\sqrt{q}} \sum_{n \pmod{q}} K(n) e_q(hn)$$

and the *Voronoi transform* of K is the q -periodic function on \mathbf{Z} defined by

$$\check{K}(n) = \frac{1}{\sqrt{q}} \sum_{\substack{h \pmod{q} \\ (h,q)=1}} \widehat{K}(h) e_q(\bar{h}n).$$

Proposition 2.1 (Tempered Voronoi formula modulo primes). *Let q be a prime number, let $K: \mathbf{Z} \rightarrow \mathbf{C}$ be a q -periodic function, and let G be a smooth function on \mathbf{R}^2 with compact support and Fourier transform denoted by \widehat{G} . We have*

$$(2.1) \quad \sum_{m,n \in \mathbf{Z}} K(mn) G(m, n) = \frac{\widehat{K}(0)}{\sqrt{q}} \sum_{m,n \in \mathbf{Z}} G(m, n) + \frac{1}{q} \sum_{m,n \in \mathbf{Z}} \check{K}(mn) \widehat{G}\left(\frac{m}{q}, \frac{n}{q}\right).$$

The key point is that when K is a (multiplicatively shifted) Kloosterman sum, then \check{G} is a normalized delta-function:

Lemma 2.2. *For $(a, q) = 1$ and $K(n) = \text{Kl}_2(an; q)$ one has*

$$\widehat{K}(h) = \begin{cases} 0 & \text{if } q \mid h, \\ e_q(-a\bar{h}) & \text{if } q \nmid h, \end{cases}$$

and

$$\check{K}(n) = \begin{cases} \frac{q-1}{q^{1/2}} & \text{if } n \equiv a \pmod{q}, \\ -\frac{1}{q^{1/2}} & \text{otherwise.} \end{cases}$$

This lemma is proved by an immediate computation. We now begin with the proof of (1.3). Let q be a prime and let W be a function satisfying (1.2). By integration by parts, we then have

$$\widehat{W}(t) \ll_{j,\varepsilon} \min(1, q^{j\varepsilon} |t/Q|^{-j})$$

for $t \in \mathbf{R}$ and for any integer $j \geq 0$ and $\varepsilon > 0$, where the implied constant depends only on j and ε .

Defining $G(m, n) = W_1(m/M)W_2(n/N)$, we deduce that for any A and any $\varepsilon > 0$, we have

$$(2.2) \quad \widehat{G}\left(\frac{m}{q}, \frac{n}{q}\right) = M\widehat{W}_1\left(\frac{mM}{q}\right)N\widehat{W}_2\left(\frac{nN}{q}\right) \ll_{\varepsilon,A} q^\varepsilon MN \left(1 + \frac{|m|M}{qQ}\right)^{-A} \left(1 + \frac{|n|N}{qQ}\right)^{-A}.$$

We next apply the Voronoi formula, Proposition 2.1, with $K(n) = \text{Kl}_2(an; q)$ to the left-hand side of (1.3). The first term on the right-hand side of (2.1) vanishes since $\widehat{K}(0) = 0$. By Lemma 2.2 and (2.2), the contribution of $mn \not\equiv a \pmod{q}$ in the second term is of order at most

$$\begin{aligned} \frac{MN}{q^{3/2-\varepsilon}} \sum_{m,n \in \mathbf{Z}} \left(1 + \frac{|m|M}{qQ}\right)^{-2} \left(1 + \frac{|n|N}{qQ}\right)^{-2} &\ll \frac{MN}{q^{3/2-\varepsilon}} \left(1 + \frac{qQ}{M}\right) \left(1 + \frac{qQ}{N}\right) \\ &\ll q^\varepsilon \left(\frac{MN}{q^{3/2}} + \frac{(M+N)Q}{q^{1/2}} + q^{1/2}Q^2\right). \end{aligned}$$

Similarly, the remaining terms $mn \equiv a \pmod{q}$ are, up to a constant, bounded by

$$q^\varepsilon \frac{MN}{q^{1/2}} \sum_{n \equiv a \pmod{q}} d(n) \left(1 + \frac{nMN}{q^2Q^2}\right)^{-2} \ll (q^2Q)^\varepsilon \left(\frac{MN}{q^{1/2}} + Q^2q^{1/2}\right).$$

This completes the proof of (1.3).

Next, we prove (1.4). We may suppose that

$$MN/8 < Y < 8MN,$$

since otherwise the sum of interest is empty. Then we see that for $M/2 < x < 2M$ and $N/2 < y < 2N$, we have the inequalities

$$\frac{\partial^{i+j} W_3(xy/Y)}{\partial x^i \partial y^j} \ll_{\varepsilon,i,j} (q^\varepsilon Q)^{i+j} M^{-i} N^{-j}$$

for all non-negative integers i, j . Hence the function $G(x, y) = W_1(x/M)W_2(y/N)W_3(xy/Y)$ satisfies the inequalities

$$\frac{\partial^{i+j} G(x, y)}{\partial x^i \partial y^j} \ll_{\varepsilon,i,j} (q^\varepsilon Q)^{i+j} x^{-i} y^{-j},$$

for $x, y > 0$, $\varepsilon > 0$ and integers $i, j \geq 0$. By repeated integration by parts of the definition of the Fourier transform

$$\widehat{G}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(x, y) e(-ux - vy) dx dy,$$

we obtain the bound

$$\widehat{G}\left(\frac{m}{q}, \frac{n}{q}\right) \ll_{\varepsilon,A} q^\varepsilon MN \left(1 + \frac{|m|M}{qQ}\right)^{-A} \left(1 + \frac{|n|N}{qQ}\right)^{-A}$$

for any A and any $\varepsilon > 0$, analogously to (2.2). The end of the proof of (1.4) is now similar to (1.3).

For future reference we record the following bound for type II sums of Kloosterman sums [2, Thm. 1.17].

Proposition 2.3. *Let q be a prime number. Let $1 \leq M, N \leq q$ and $(\alpha_m), (\beta_n)$ be sequences of complex numbers supported in $[M, 2M]$ and $[N, 2N]$ respectively. Let either $Q = 1$ and W be the constant function 1, or $Q \geq 1$ and W be a function satisfying (1.2). Then, for every $\varepsilon > 0$, we have*

$$\sum_{m,n} \alpha_m \beta_n \text{Kl}_2(mn; q) W\left(\frac{mn}{Y}\right) \ll_{\varepsilon} \|\alpha\|_2 \|\beta\|_2 (MN)^{1/2} \left(\frac{1}{M} + Q \frac{q^{1/2+\varepsilon}}{N}\right)^{1/2}.$$

This is a special case of [2, Thm. 1.17] when W is the constant 1. For smooth W , the same proof applies, except that we apply partial summation in [2, (3.2)] if $m_1 \neq m_2$ to remove the weight $W(m_1 n/Y)W(m_2 n/Y)$; this produces a factor Q that after taking square roots produces the above bound.

3. CORRELATION SUMS OF KLOOSTERMAN SUMS AND HECKE EIGENVALUES

In this section we prove Proposition 1.3. We replace the tempered Voronoi summation formula by the Voronoi summation formula for cusp forms, which we state in a form suited to our purpose.

Proposition 3.1 (Voronoi summation formula for cusp forms with arithmetic weights modulo primes). *Let q be a prime. Let W be a smooth function compactly supported in $]0, \infty[$ and let f be a holomorphic cuspidal Hecke eigenform of level 1 and weight k . Let $\varepsilon(f) = \pm 1$ denote the sign of the functional equation of the Hecke L -function $L(f, s)$ and let*

$$\widetilde{W}(y) = \int_0^{\infty} W(u) \mathcal{J}_k(4\pi\sqrt{uy}) du,$$

where

$$\mathcal{J}_k(u) = 2\pi i^k J_{k-1}(u).$$

Then, for any q -periodic arithmetic function $K: \mathbf{Z} \rightarrow \mathbf{C}$, we have

$$\sum_{n \geq 1} \lambda_f(n) K(n) W\left(\frac{n}{N}\right) = \frac{\widehat{K}(0)}{q^{1/2}} \sum_{n \geq 1} \lambda_f(n) W\left(\frac{n}{N}\right) + \varepsilon(f) \frac{N}{q} \sum_{n \geq 1} \lambda_f(n) \check{K}(n) \widetilde{W}\left(\frac{nN}{q^2}\right).$$

In particular, for a coprime to q , we have

$$\sum_{n \geq 1} \lambda_f(n) \text{Kl}_2(an; q) W\left(\frac{n}{N}\right) = \varepsilon(f) \frac{N}{q^{1/2}} \sum_{n \equiv a \pmod{q}} \lambda_f(n) \widetilde{W}\left(\frac{nN}{q^2}\right) - \varepsilon(f) \frac{N}{q^{3/2}} \sum_{n \geq 1} \lambda_f(n) \widetilde{W}\left(\frac{nN}{q^2}\right).$$

Proof. We expand $K(n)$ into additive characters

$$K(n) = \frac{1}{q^{1/2}} \sum_{a \pmod{q}} \widehat{K}(a) e_q(-an)$$

and apply the classical summation formula

$$\sum_{n \geq 1} \lambda_f(n) W\left(\frac{n}{N}\right) e\left(-\frac{an}{q}\right) = \varepsilon(f) \frac{N}{q} \sum_{n \geq 1} \lambda_f(n) e\left(\frac{\bar{a}n}{q}\right) \widetilde{W}\left(\frac{nN}{q^2}\right),$$

valid for all $N > 0$ and all a coprime to q ([6, Theorem A.4]). \square

We can now easily prove Proposition 1.3: integration by parts shows that for any $A \geq 0$ and $\varepsilon > 0$ we have

$$\widetilde{W}\left(\frac{nN}{q^2}\right) \ll_{k,A,\varepsilon} q^{\varepsilon} \left(1 + \frac{nN}{q^2}\right)^{-A}$$

(see [1, Lemma 2.4]), so that (using Deligne's bound $|\lambda_f(n)| \leq d(n) \ll_{\varepsilon} n^{\varepsilon}$), we get

$$\sum_n \lambda_f(n) \text{Kl}_2(an; q) W\left(\frac{n}{N}\right) \ll_{\varepsilon,k} (qN)^{\varepsilon} \left(q^{1/2} + \frac{N}{q^{1/2}}\right).$$

4. APPLICATION TO THE FOURTH MOMENT OF DIRICHLET L -FUNCTIONS

In this section we prove Theorem 1.5. The general strategy of the proof has been explained in detail in our paper [1]. We assume some familiarity with this paper, and refer in particular to [1, §1.2, §6.1, §6.3] for notations.

We begin with the unconditional bound. Let

$$B_{E,E}^{\pm}(M, N) = \frac{1}{(MN)^{1/2}} \sum_{\substack{m \equiv \pm n \pmod{q} \\ m \neq n}} d(m)d(n)W_1\left(\frac{m}{M}\right)W_2\left(\frac{n}{N}\right) - \frac{1}{q(MN)^{1/2}} \sum_{m,n} d(m)d(n)W_1\left(\frac{m}{M}\right)W_2\left(\frac{n}{N}\right).$$

Our objective is to prove that for $\eta = 1/20$ one has

$$(4.1) \quad B_{E,E}^{\pm}(M, N) - \text{MT}_{E,E}^{\text{od},\pm}(M, N) =: \text{ET}_{E,E}^{\pm}(M, N) \ll_{\varepsilon} q^{-\eta+o(1)},$$

where $\text{MT}_{E,E}^{\text{od},\pm}(M, N)$ is a suitable main term (described in [8]) and M, N range over a set of $O(\log^2 q)$ real numbers satisfying

$$1 \leq M \leq N, \quad MN \leq q^{2+o(1)}$$

(the first bound is by symmetry, the second is the length of the approximate functional equation). We set

$$N^* = q^2/N, \quad M = q^{\mu}, \quad N = q^{\nu}, \quad \nu^* = 2 - \nu,$$

so that

$$0 \leq \mu \leq \nu, \quad -\varepsilon \leq \nu^* - \mu.$$

In view of the bound [1, (3.18)], which reads

$$\text{ET}_{E,E}^{\pm}(M, N) \ll q^{\varepsilon} \left(\frac{N}{qM}\right)^{1/4} \left(1 + \left(\frac{N}{qM}\right)^{1/4}\right)$$

and which is proved using spectral theory, we may also assume that

$$(4.2) \quad \mu + \nu^* \leq 1 + 4\eta$$

for otherwise (4.1) is certainly true. Proceeding in the same way as in [1, §6.3], we apply Voronoi summation to reduced to the following bounds for $O(\log^4 q)$ sums of the shape

$$S^{\pm}(M_1, M_2, M_3, M_4) = \frac{1}{(qMN^*)^{1/2}} \sum_{m_1, m_2, m_3, m_4} W_1\left(\frac{m_1}{M_1}\right)W_2\left(\frac{m_2}{M_2}\right) \times W_3\left(\frac{m_3}{M_3}\right)W_4\left(\frac{m_4}{M_4}\right) \text{Kl}_2(\pm m_1 m_2 m_3 m_4; q) \ll q^{-\eta+o(1)},$$

where the W_i satisfy (1.2) with $Q = q^{\varepsilon}$, and the M_i written in the shape $M_i = q^{\mu_i}$, $i = 1, 2, 3, 4$, satisfy

$$\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4, \quad 0 \leq \mu_1 + \mu_2 + \mu_3 + \mu_4 = \mu + \nu', \quad \nu' \leq \nu^*.$$

By the trivial bound for Kloosterman sums (and recalling (4.2)), we may assume that

$$(4.3) \quad 1 - 2\eta \leq \mu + \nu' \leq \mu + \nu^* \leq 1 + 4\eta,$$

for otherwise (4.1) is true.

We use the same strategy as in [1, §6.3], except that we replace [1, (5.5)] by Proposition 1.2. Thus, if the largest variables m_3, m_4 are large enough, we apply (1.3) to them (fixing m_1, m_2);

otherwise, we find it more beneficial to group variables differently producing a bilinear sum of Kloosterman sums to which we apply Proposition 2.3.

Explicitly, using (1.3) we obtain that

$$\begin{aligned} S^\pm(M_1, M_2, M_3, M_4) &\ll q^{o(1)} \frac{M_1 M_2}{(q M N^*)^{1/2}} \left(q^{1/2} + \frac{M_3 M_4}{q^{1/2}} \right) \\ &\ll q^{o(1)} \left(\sqrt{\frac{M_1 M_2}{M_3 M_4}} + \frac{(M N')^{1/2}}{q} \right) \ll q^{o(1)} \left(\sqrt{\frac{M_1 M_2}{M_3 M_4}} + q^{-\eta} \right) \end{aligned}$$

since $q^{\frac{1}{2}(1+4\eta)-1} \leq q^{-\eta}$. We may therefore assume that

$$(4.4) \quad 0 \leq \mu_3 + \mu_4 - (\mu_1 + \mu_2) \leq 2\eta.$$

We now apply Proposition 2.3 with $M = M_4$ and $N = M_1 M_2 M_3$ so that $MN = q^{\mu+\nu'} \leq MN^*$ and derive

$$S^\pm(M_1, M_2, M_3, M_4) \ll q^{o(1)} \left(q^{\frac{\mu_1 + \mu_2 + \mu_3 - 1}{2}} + q^{-\frac{1}{4} + \frac{\mu_4}{2}} \right).$$

We claim that under the current assumptions both exponents on the right hand side are $\leq -\eta$, which completes the proof. Indeed, since $\mu_4 \geq \mu_i$ for $i = 1, 2, 3$, we obtain by (4.3) that

$$\left(1 + \frac{1}{3} \right) (\mu_1 + \mu_2 + \mu_3) \leq \mu_1 + \mu_2 + \mu_3 + \mu_4 \leq 1 + 4\eta \implies \mu_1 + \mu_2 + \mu_3 \leq \frac{3}{4} + 3\eta,$$

hence

$$\frac{\mu_1 + \mu_2 + \mu_3 - 1}{2} \leq -\frac{1}{8} + \frac{3}{2}\eta \leq -\eta.$$

Moreover, by (4.4) and (4.3) (since $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$) we have

$$\mu_4 \leq 2\eta + \mu_1 + \mu_2 - \mu_3 \leq 2\eta + \mu_1 \leq 2\eta + \frac{1}{3}(1 + 4\eta - \mu_4) = \frac{1}{3} + \frac{10}{3}\eta - \frac{1}{3}\mu_4,$$

which implies that $\mu_4 \leq \frac{1}{4} + \frac{5}{2}\eta$, and so

$$-\frac{1}{4} + \frac{\mu_4}{2} \leq -\frac{1}{8} + \frac{5}{4}\eta \leq -\eta.$$

If the Ramanujan–Pettersson conjecture is available, we can use [1, (1.7)] with $\theta = 0$ in place of [1, (3.2)] and replace (4.2) with $\mu + \nu^* \leq 1 + 2\eta$. Then the same strategy leads to the numerical value $\eta = 1/16$.

5. SUMS OF KLOOSTERMAN SUMS ALONG THE PRIMES: PROOF OF THEOREM 1.8

5.1. Proof of inequality (1.9). We now recall the main ideas of the proof of [2, Thm. 1.5], since our proof will follow the same path until the moment we use Proposition 1.2. We will incorporate some shortcuts and combinatorial improvements to [2], mainly due to the assumption $X \leq q$. By [2, p. 1711–1716], we are reduced to proving the same bound as (1.9) for the sum

$$\mathfrak{S}_{W,X}(\Lambda, \text{Kl}_2) := \sum_n \Lambda(n) \text{Kl}_2(n; q) W\left(\frac{n}{X}\right),$$

where Λ is the von Mangoldt function. We now apply Heath-Brown’s identity [5] with integer parameter $J \geq 2$. This decomposes $\mathfrak{S}_{W,X}(\Lambda, \text{Kl}_2)$ into a linear combination, with coefficients bounded by $O_J(\log X)$, of $O(\log^{2J} X)$ sums of the shape

$$(5.1) \quad \begin{aligned} \Sigma(\mathbf{M}, \mathbf{N}) &= \sum_{m_1, \dots, m_J} \alpha_1(m_1) \alpha_2(m_2) \cdots \alpha_J(m_J) \\ &\times \sum_{n_1, \dots, n_J} V_1\left(\frac{n_1}{N_1}\right) \cdots V_J\left(\frac{n_J}{N_J}\right) W\left(\frac{m_1 \cdots m_J n_1 \cdots n_J}{X}\right) \text{Kl}_2(m_1 \cdots m_J n_1 \cdots n_J; q) \end{aligned}$$

where

- $\mathbf{M} = (M_1, \dots, M_J)$, $\mathbf{N} = (N_1, \dots, N_J)$ are J -tuples of parameters in $[1/2, 2X]^{2J}$ which satisfy

$$(5.2) \quad N_1 \geq N_2 \geq \dots \geq N_J, \quad M_i \leq X^{1/J}, \quad M_1 \cdots M_J N_1 \cdots N_J \asymp_J X;$$

- the arithmetic functions $m \mapsto \alpha_i(m)$ are bounded and supported in $[M_i/2, 2M_i]$;
- the smooth functions $x \mapsto V_i(x)$ satisfy (1.2) with parameter Q .

It now remains to study the sum $\Sigma(\mathbf{M}, \mathbf{N})$ defined in (5.1) for every (\mathbf{M}, \mathbf{N}) as above. We estimate $\Sigma(\mathbf{M}, \mathbf{N})$ in two ways.

Our first method is to bound $\Sigma(\mathbf{M}, \mathbf{N})$ by applying (1.4) to the largest smooth variables n_1 and n_2 in $\Sigma(\mathbf{M}, \mathbf{N})$ and a trivial summation over the other variables. We obtain

$$\Sigma(\mathbf{M}, \mathbf{N}) \ll q^\varepsilon Q^2 X \left(\frac{q^{1/2}}{N_1 N_2} + \frac{1}{q^{1/2}} \right),$$

which, by (5.2) and the assumption $X \leq q$, simplifies into

$$(5.3) \quad \Sigma(\mathbf{M}, \mathbf{N}) \ll q^\varepsilon Q^2 X (q^{1/2}/(N_1 N_2)).$$

Our second method is to apply Proposition 2.3 to $\Sigma(\mathbf{M}, \mathbf{N})$; in this way we obtain

$$(5.4) \quad \Sigma(\mathbf{M}, \mathbf{N}) \ll q^\varepsilon Q^{1/2} X \left(\frac{1}{M^{1/2}} + \frac{q^{1/4}}{(X/M)^{1/2}} \right)$$

for any factorization

$$M_1 \cdots M_J N_1 \cdots N_J = M \times N.$$

We have now to play with (5.3) and (5.4) in an optimal way to bound $\Sigma(\mathbf{M}, \mathbf{N})$. We follow the same presentation as in [2, §4.2]. We introduce the real numbers κ , x , μ_i , ν_j , $1 \leq i, j \leq J$, defined by

$$Q = q^\kappa, \quad X = q^x, \quad M_i = q^{\mu_i}, \quad N_j = q^{\nu_j}$$

and we set

$$(\mathbf{m}, \mathbf{n}) = (\mu_1, \dots, \mu_J, \nu_1, \dots, \nu_J) \in [0, x]^{2J}.$$

The conditions (5.2) are reinterpreted as

$$(5.5) \quad \sum_i \mu_i + \sum_j \nu_j = x \leq 1, \quad \mu_i \leq x/J, \quad \nu_1 \geq \nu_2 \geq \dots \geq \nu_J.$$

According to (5.3) and (5.4), we introduce the function (compare with [2, definition (4.5)]) $\eta(\mathbf{m}, \mathbf{n})$ defined by

$$(5.6) \quad \eta(\mathbf{m}, \mathbf{n}) := \max \left\{ (\nu_1 + \nu_2) - \frac{1}{2} - 2\kappa; \max_\sigma \min \left(\frac{\sigma}{2}, \frac{x - \sigma}{2} - \frac{1}{4} \right) - \frac{\kappa}{2} \right\},$$

where σ ranges over all possible sub-sums of the μ_i and ν_j for $1 \leq i, j \leq J$, that is, over the sums

$$\sigma = \sum_{i \in \mathcal{I}} \mu_i + \sum_{j \in \mathcal{J}} \nu_j,$$

for \mathcal{I} and \mathcal{J} ranging over all possible subsets of $\{1, \dots, J\}$.

With these conventions, as a consequence of (5.3) and (5.4) we have the inequality

$$\Sigma(\mathbf{M}, \mathbf{N}) \ll (qQ)^\varepsilon q^{-\eta(\mathbf{m}, \mathbf{n})} X,$$

and finally, summing over all possible (\mathbf{M}, \mathbf{N}) , we have the inequality

$$(5.7) \quad \mathcal{S}_{W, X}(\Lambda, \text{Kl}_2) \ll (qQ)^\varepsilon q^{-\eta} X,$$

where

$$\eta = \min_{(\mathbf{m}, \mathbf{n})} \eta(\mathbf{m}, \mathbf{n}),$$

where (\mathbf{m}, \mathbf{n}) satisfy (5.5).

The estimate (1.9) is trivial for $x < 3/4$, so we may assume that $3/4 \leq x \leq 1$. For $\varepsilon > 0$ sufficiently small, let \mathcal{J}_x be the interval

$$\mathcal{J}_x = [x/6 - \varepsilon, x/3 + \varepsilon],$$

and choose $J = 10$ to apply Heath-Brown's identity.

We now consider two different cases in the combinatorics of (\mathbf{m}, \mathbf{n}) .

- If (\mathbf{m}, \mathbf{n}) contains a subsum $\sigma \in \mathcal{J}_x$, then, by (5.6), we have the inequality

$$\eta(\mathbf{m}, \mathbf{n}) \geq \min\left(\frac{x/6}{2}, \frac{x - x/3}{2} - \frac{1}{4}\right) - \frac{\kappa}{2} - \frac{\varepsilon}{2},$$

which simplifies into

$$(5.8) \quad \eta(\mathbf{m}, \mathbf{n}) \geq \frac{x}{3} - \frac{1}{4} - \frac{\kappa}{2} - \frac{\varepsilon}{2}.$$

- If (\mathbf{m}, \mathbf{n}) contains no subsum $\sigma \in \mathcal{J}_x$, then the sum of all the μ_i and ν_j which are less than $x/6 - \varepsilon$ is also less than $x/6 - \varepsilon$ (this is a consequence of the inequality $2(x/6 - \varepsilon) < x/3 + \varepsilon$). In light of (5.5), this includes all μ_i , and so some ν_j must be greater than $x/3 + \varepsilon$. On the other hand, since $3(x/3 + \varepsilon) > x$, we deduce that at most two ν_i (more precisely, ν_1 or ν_1 and ν_2) are greater than $x/3 + \varepsilon$. Combining these remarks, we deduce the inequality

$$\nu_1 + \nu_2 \geq x - (x/6 - \varepsilon) = 5x/6 + \varepsilon,$$

which implies, by (5.6), the inequality

$$(5.9) \quad \eta(\mathbf{m}, \mathbf{n}) \geq \frac{5x}{6} - \frac{1}{2} - 2\kappa - \varepsilon.$$

By (5.7), (5.8) and (5.9), we deduce the inequality

$$(5.10) \quad \mathcal{S}_{W,X}(\Lambda, \text{Kl}_2) \ll (qQ)^\varepsilon (q^{1/4}Q^{1/2}X^{2/3} + q^{1/2}Q^2X^{1/6}).$$

In the above upper bound, the first term is larger than the second one if and only if $Q < q^{-1/6}X^{1/3}$, and in this case, we have $Q^\varepsilon < q^\varepsilon$. However, when $Q \geq q^{-1/6}X^{1/3}$, it is easy to see that the bound (1.9) is trivial since we have

$$q^{1/4}Q^{1/2}X^{2/3} \geq q^{1/4}(q^{-1/6}X^{1/3})^{1/2}X^{2/3} = q^{1/6}X^{5/6} \geq X,$$

since we suppose $X \leq q$. In conclusion, we may drop the second term on the right-hand side of (5.10). This remark completes the proof of (1.9).

5.2. Proof of inequality (1.10). The proof mimics the proof appearing in [2, §4.3]. By a simple subdivision, it is sufficient to prove the inequality

$$(5.11) \quad \sum_{\substack{X < p \leq \frac{3}{2}X \\ p \text{ prime}}} \text{Kl}_2(p; q) \ll q^{1/6+\varepsilon} X^{7/9}.$$

Let $\Delta < 1/2$ be some parameter, let W be a smooth function defined on $[0, +\infty[$ such that

$$\text{supp}(W) \subset [1 - \Delta, \frac{3}{2} + \Delta], \quad 0 \leq W \leq 1, \quad W(x) = 1 \text{ for } 1 \leq x \leq \frac{3}{2},$$

and such that the derivatives satisfy

$$x^j W^{(j)}(x) \ll_j Q^j,$$

with $Q = \Delta^{-1}$. By applying (1.9), we have

$$\sum_{\substack{X < p \leq \frac{3}{2}X \\ p \text{ prime}}} \text{Kl}_2(p; q) \ll \Delta X + 1 + \left| \sum_p W\left(\frac{p}{X}\right) \text{Kl}_2(p; q) \right| \\ \ll \Delta X + q^{1/4+\varepsilon} Q^{1/2} X^{2/3} \ll q^{1/6+\varepsilon} X^{7/9},$$

by the choice $\Delta = q^{1/6} X^{-2/9} < 1/2$ (the claim is trivial if $q^{1/6} \geq \frac{1}{2} X^{2/9}$). This completes the proof of (5.11).

REFERENCES

- [1] V. Blomer, É. Fouvry, E. Kowalski, Ph. Michel, and D. Milićević, *On moments of twisted L -functions*, Amer. J. Math. (to appear, [arXiv:1411.4467](https://arxiv.org/abs/1411.4467)).
- [2] É. Fouvry, E. Kowalski, and Ph. Michel, *Algebraic trace functions over the primes*, Duke Math. J. **163** (2014), no. 9, 1683–1736.
- [3] É. Fouvry, E. Kowalski, and Ph. Michel, *Algebraic twists of modular forms and Hecke orbits*, Geom. Funct. Anal. **25** (2015), no. 2, 580–657.
- [4] É. Fouvry, E. Kowalski, and Ph. Michel, *On the exponent of distribution of the ternary divisor function*, Mathematika **61** (2015), no. 1, 121–144.
- [5] D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), 1365–1377.
- [6] E. Kowalski, Ph. Michel, and J. VanderKam, *Rankin-Selberg L -functions in the level aspect*, Duke Math. J. **114** (2002), no. 1, 123–191.
- [7] I. Shparlinski and T. P. Zhang, *Cancellations amongst Kloosterman sums*, Acta Arith. (to appear, [arXiv:1601.05123](https://arxiv.org/abs/1601.05123)).
- [8] M. P. Young, *The fourth moment of Dirichlet L -functions*, Ann. of Math. (2) **173** (2011), no. 1, 1–50.

MATHEMATISCHES INSTITUT, UNIVERSITÄT GÖTTINGEN, BUNSENSTR. 3-5, 37073 GÖTTINGEN, GERMANY
E-mail address: vblomer@math.uni-goettingen.de

LABORATOIRE DE MATHÉMATIQUES D'ORSAY, UNIVERSITÉ PARIS-SACLAY, 91405 ORSAY, FRANCE
E-mail address: etienne.fouvry@math.u-psud.fr

ETH ZÜRICH – D-MATH, RÄMISTRASSE 101, CH-8092 ZÜRICH, SWITZERLAND
E-mail address: kowalski@math.ethz.ch

EPF LAUSANNE, CHAIRE TAN, STATION 8, CH-1015 LAUSANNE, SWITZERLAND
E-mail address: philippe.michel@epfl.ch

DEPARTMENT OF MATHEMATICS, BRYN MAWR COLLEGE, 101 NORTH MERION AVENUE, BRYN MAWR, PA 19010-2899, U.S.A.

Current address: Max-Planck-Institut für Mathematik, Vivatsgasse 7, D-53111 Bonn, Germany
E-mail address: dmilicevic@brynmawr.edu