

ON SUMS OVER PRIMES

ANTAL BALOG

*Mathematical Institute, Hungarian Academy of Sciences
Budapest, Hungary*

1

Sums over primes are of great importance in the theory of numbers. Many problems can be reduced to estimating sums of the following kind:

$$(1) \quad \sum_{p \leq x} f(p) \quad \text{or} \quad \sum_{n \leq x} \Lambda(n) f(n).$$

For instance the ternary Goldbach problem ([15], [16]) requires a non-trivial bound for the sum

$$(2) \quad \sum_{p \leq x} e^{2\pi i \alpha p}.$$

The Bombieri–A. I. Vinogradov theorem ([14], [15]) depends on the behaviour of the sum

$$(3) \quad \sum_{\chi} \left| \sum_{p \leq x} \chi(p) \right|,$$

where the first sum runs over a certain collection of Dirichlet characters, and analytic proofs of many problems ([1], [2], [3], [4], [9], [10], [11]) are based on investigating some integrals

$$(4) \quad \int_{-T}^T \left| \sum_{p \leq x} p^{-it} \right| g(t) dt.$$

In a large class of problems $f(n)$ is an error term, i.e., it has the form

$$(5) \quad f(n) = F(n) - v(n)$$

where $F(n)$ is a non-negative function and $v(n)$ is its expected value ([1], [2], [3], [4], [7], [8], [12], [13]). After expanding $f(n)$ into Fourier-series, or

using the orthogonality of characters, or the Perron formula, we arrive at a sum over primes in the forms (2') or (3) or (4) respectively. Here

$$(2') \quad \sum_{p \leq x} e^{2\pi i g(p)}$$

with some smooth real function $g(n)$ ([7], [12], [13], [15]).

It seemed desirable to find general methods for treating these sums. The first method is due to I. M. Vinogradov ([16]), who developed it about 1937.

Nowadays we have some alternative methods to estimate a sum (1) but all of them are based on the same idea. It is to relate the initial sum (1) to bilinear forms

$$(6) \quad B = \sum_{\substack{mn \leq x \\ m \sim M}} a_m b_n f(mn)$$

where $m \sim M$ indicates $M \leq m < 2M$. In fact, there are several ways to estimate a sum like (6) and the resulting bounds usually hold for arbitrary collections of complex numbers. We distinguish two types of such sums. In the general case we speak about type II bilinear forms, and if the b_n 's are 1 identically then we call the form type I.

2

Before showing methods to relate sum (1) to sum (6), we consider some conditions that $f(n)$ and M must satisfy in order to obtain non-trivial bounds for B .

We have for type I bilinear forms

$$(7) \quad |B| \leq \sum_{m \sim M} |a_m| \sum_{n \leq x/m} |f(mn)|$$

and for type II bilinear forms

$$(8) \quad |B|^2 \leq \left(\sum_{m \sim M} |a_m|^2 \right) \left(\sum_{n \leq x/M} |b_n|^2 \right) \times \\ \times \left(\max_{m \sim M} \sum_{n \leq x/m} |f(mn)|^2 + \max_{\substack{m \sim M \\ m' \neq m}} \sum_{\substack{m' \sim M \\ n \leq x/m \\ n \leq x/m'}} | \sum_{n \leq x/m} f(mn) \bar{f}(m'n) | \right).$$

(The latter can be derived from the Cauchy-Schwartz inequality.) Of course the deeper the method we apply the deeper the result we obtain — but the general shape of the conditions on $f(n)$ and M can be seen even in the case of this fairly trivial treatment. For the sake of simplicity we assume that $|f(n)| \leq 1$, $|a_m| \leq 1$ and $|b_n| \leq 1$. The trivial bound for B is $|B| \leq x$. For our

purpose the innermost sums in (7) and (8) must be small, i.e., $f(n)$ must have an "oscillating" property:

$$(9) \quad \left| \sum_{n \leq y} f(mn) \right| \ll Y(y, m)$$

and a "quasi-orthogonal" property:

$$(10) \quad \left| \sum_{n \leq y} f(mn) \bar{f}(m'n) \right| \ll Y(y, m, m')$$

with some functions $Y(y, m)$ and $Y(y, m, m')$ having a smaller order of magnitude than y . For example, if (9) and (10) hold with $Y = y^{1/2}$ uniformly in m and m' , then we get

$$(11) \quad \begin{aligned} |B| &\ll (xM)^{1/2} && \text{in case I,} \\ |B| &\ll xM^{-1/2} + x^{3/4} M^{1/4} && \text{in case II.} \end{aligned}$$

This shows that we can gain a factor of x if M is not too large (in case I) and also not too small (in case II). This situation is typical.

To sum up, we can hope for a non-trivial bound on B if

$$(12) \quad \text{I: } \begin{cases} f(n) \text{ has an "oscillating" property, and} \\ M \text{ is not too large,} \end{cases}$$

$$(13) \quad \text{II: } \begin{cases} f(n) \text{ has a "quasi-orthogonal" property, and} \\ M \text{ is not too large and not too small.} \end{cases}$$

The acceptable ranges of M in (12) and (13) are determined by the factor we want to gain. As we can change the roles of the a_m 's and b_n 's in (8), the restrictions on M are symmetric with respect to $x^{1/2}$ (in case II). If $f(n)$ has the above properties, then the desired bound can be obtained by ascertaining

$$(14) \quad \begin{aligned} \text{I: } &M \ll z_3, \\ \text{II: } &z_1 \ll M \ll z_2 \text{ or } x/z_2 \ll M \ll x/z_1, \end{aligned}$$

where $z_1 < z_2 < x$ and $z_3 < x$. In addition, we can assume that $z_2 \geq x^{1/2}$.

Note that in the cases (3) and (4) $f(n)$ has only "average oscillating" and "average quasi-orthogonal" properties, but it is sufficient for our purposes, as the reader can verify himself.

3

Here we show some methods to relate sum (1) to sum (6). Our survey covers only the recently discovered methods, because they are not only stronger but also remarkably simpler than the old ones.

VAUGHAN'S IDENTITY ([12]–[15]). For the parameters $1 \leq u, v \leq x$

$$(15) \quad \sum_{n \leq x} \Lambda(n) f(n) = \sum_{n \leq v} \Lambda(n) f(n) + \sum_{\substack{mn \leq x \\ m \leq u}} \mu(m) \log n f(mn) - \\ - \sum_{\substack{mn \leq x \\ m \leq uv}} a_m f(mn) + \sum_{\substack{mn \leq x \\ m > u, n > v}} \mu(m) b_n f(mn)$$

where a_m and b_n are some real coefficients satisfying $|a_m| \leq \log m$, $|b_n| \leq \log n$.

(This can be proved by investigating the equation

$$(16) \quad -\frac{\zeta'}{\zeta} = F - M\zeta' - MF\zeta + \left(\frac{1}{\zeta} - M\right) (-\zeta' - F\zeta)$$

where $F(s) = \sum_{n \leq v} \Lambda(n) n^{-s}$ and $M(s) = \sum_{m \leq u} \mu(m) m^{-s}$.)

The first term in (15) is essentially trivial, the second and third are of type I (because $\log n$ is as good as 1) and the last is of type II. It is an easy task to check that the parameters u, v can be chosen in such a way that (14) holds for all bilinear forms occurring in (15) iff

$$(17) \quad z_1 \leq x^{1/2}, \quad (x/z_2)^2 \leq z_3,$$

or

$$(18) \quad \frac{x}{z_2} \frac{x}{z_3} \leq \frac{x}{z_1}.$$

If neither (17) nor (18) is satisfied, then there are some parts of the right-hand side of (15) for which we have no good upper bound. This limits the usefulness of Vaughan's method.

HEATH-BROWN'S IDENTITY ([10]). For a parameter $z \geq 1$ and an integer J satisfying

$$(19) \quad 2z^J \geq x$$

we have

$$(20) \quad \sum_{n \leq x} \Lambda(n) f(n) \\ = \sum_{j=1}^J (-1)^{j-1} \binom{J}{j} \sum_{\substack{m_1 \dots m_j n_1 \dots n_j \leq x \\ m_1 \leq z, \dots, m_j \leq z}} \mu(m_1) \dots \mu(m_j) \log n_1 f(m_1 \dots m_j n_1 \dots n_j).$$

(This can easily be proved by investigating the equation

$$(21) \quad \frac{\zeta'}{\zeta}(1-M\zeta)^J - \frac{\zeta'}{\zeta} = \sum_{j=1}^J (-1)^j \binom{J}{j} M^j \zeta' \zeta^{j-1}$$

where $M(s) = \sum_{m \leq z} \mu(m) m^{-s}$.

The innermost sum of the right-hand side of (20) is a multilinear form which can be rewritten into bilinear forms of type I or type II. We may subdivide the ranges of summation into intervals $m_1 \sim M_1, \dots, n_j \sim N_j$. When we divide the variables into two classes in order to arrive at a bilinear form, we can take the sizes of M_1, \dots, N_j into account. Thus identity (20) is more flexible than identity (15), and, indeed, it is easy to find a special choice of z_1, \dots, z_3 for which neither (17) nor (18) is satisfied but all bilinear forms made from (20) satisfy (14). This has some applications ([6], [10]). Note that an iterated application of Vaughan's identity leads to an identity similar to but more complicated than (20) ([5]).

4

Now we turn to our main goal, namely to the introduction of the so-called sieve identity method. This was developed by Harman ([8]), who extended the original idea of Heath-Brown ([10]).

In a large class of applications $f(n)$ has the form (5). If we can use Vaughan's identity or Heath-Brown's identity, then we get a result in the form

$$(22) \quad \sum_{p \leq x} F(p) = \sum_{p \leq x} v(p) + O(\dots)$$

with an acceptable error term. Usually the main term can be derived from the Prime Number Theorem because $v(n)$ is a smooth function. If in transforming (1) into (6) we arrive at bilinear forms not satisfying (14), then the above identities cannot be used. The main advantage of the sieve identity method is that we may admit a few "bad" bilinear forms if they have only non-negative coefficients. Omitting these terms, we can get a lower bound in the form

$$(23) \quad \sum_{p \leq x} F(p) \gg \sum_{p \leq x} v(p) + O(\dots)$$

which is generally satisfactory ([3], [8], [10]).

We will assume in the sequel that

$$(24) \quad x/z_2 < x/z_1 \leq x^{1/2} \leq z_1 < z_2 \leq z_3 < x,$$

and that $f(n)$ has the form (5), and we will use the following notation: $p(n)$ is the smallest prime factor of n ($n > 1$), $P(z) = \prod_{p < z} p$ and $\chi(n, z)$ is the characteristic function of the numbers coprime to $P(z)$, i.e.,

$$(25) \quad \chi(n, z) = \begin{cases} 1 & \text{if } (n, P(z)) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

We have the fundamental Buchstab identity

$$(26) \quad \sum_{n \leq x} \chi(n, z) g(n) = \sum_{n \leq x} \chi(n, z') g(n) - \sum_{z' \leq p < z} \sum_{pn \leq x} \chi(n, p) g(pn).$$

(This can be proved by ordering the numbers according to their smallest prime factor.) This is the basic identity in our method although we never use it explicitly.

The last term on the right-hand side of (26) is almost a type II bilinear form, but its coefficients are not entirely independent. However, this dependence can be controlled and in fact we have the same bounds for the type II* "bilinear" form

$$(27) \quad B^* = \sum_{\substack{mn \leq x \\ m \sim M}} a_m b_n \chi(n, z(m)) f(mn)$$

that we had for (6) when f was in the form (5). An iterated application of (26) yields sums

$$(28) \quad \sum_{z' \leq p_r < \dots < p_1 < z} \sum_{n \leq x/p_1 \dots p_r} \chi(n, p_r) g(np_1 \dots p_r)$$

and

$$(29) \quad \sum_{z' \leq p_r < \dots < p_1 < z} \sum_{n \leq x/p_1 \dots p_r} \chi(n, z') g(np_1 \dots p_r).$$

Although Heath-Brown's identity has similar flexibility, here all the coefficients have the same sign. Thus we can omit the terms which cannot be treated satisfactory. We will restate this argument in a more exact form.

LEMMA 1 (Richert's Fundamental Identity). *Let $h(d)$ and $g(d)$ be arbitrary functions with*

$$(30) \quad h(1) = 1.$$

We define $\bar{h}(d)$ as

$$(31) \quad \bar{h}(d) = h\left(\frac{d}{p(d)}\right) - h(d) \quad (d > 1),$$

For any $z \geq 2$ we have

$$(32) \quad \sum_{d|P(z)} \mu(d)g(d) = \sum_{d|P(z)} \mu(d)h(d)g(d) + \sum_{d|P(z)} \mu(d)\bar{h}(d) \sum_{\delta|P(p(d))} \mu(\delta)g(d\delta).$$

Proof. $d|P(z)$ and $\delta|P(p(d))$ imply that $(d, \delta) = 1$ and $d\delta|P(z)$. Collecting the terms $t = d\delta$ for the second sum on the right-hand side we get

$$(33) \quad \sum_{d|P(z)} \mu(d)\bar{h}(d) \sum_{\delta|P(p(d))} \mu(\delta)g(d\delta) = \sum_{t|P(z)} \mu(t)g(t) \sum_{t=d\delta, \delta|P(p(d))} \bar{h}(d).$$

Let $t = p_1 \dots p_r$ where $p_1 > \dots > p_r$. $t = d\delta$ and $\delta|P(p(d))$ are simultaneously satisfied iff $d = p_1 \dots p_j$, $1 \leq j \leq r$. Using (30) and (31), we arrive at

$$(34) \quad \sum_{\substack{t=d\delta \\ \delta|P(p(d))}} \bar{h}(d) = \sum_{j=1}^r \bar{h}(p_1 \dots p_j) = 1 - h(t).$$

It is also true for $t = 1$, and (33) coupled with (34) gives (32).

COROLLARY 1. If $a_m = 0$ whenever $(m, P(z)) > 1$, then

$$(35) \quad \sum_{\substack{mn \leq x \\ m \sim M}} a_m \chi(n, z) g(mn) = \sum_{\substack{mdn \leq x \\ m \sim M \\ d|P(z), d < y}} a_m \mu(d) g(mdn) + \\ + \sum_{\substack{mdn \leq x \\ m \sim M, d|P(z) \\ d \geq y, d/p(d) < y}} a_m \mu(d) \chi(n, p(d)) g(mdn).$$

Proof. We apply Lemma 1 with

$$(36) \quad h(d) = \begin{cases} 1 & \text{if } d < y, \\ 0 & \text{otherwise,} \end{cases}$$

and with

$$(37) \quad G(d) = \sum_{\substack{mn \leq x/d \\ m \sim M}} a_m g(mnd)$$

in place of $g(d)$. In this case we have

$$(38) \quad \bar{h}(d) = \begin{cases} 1 & \text{if } d \geq y \text{ and } d/p(d) < y, \\ 0 & \text{otherwise.} \end{cases}$$

(35) follows from (32), (36), (38) and the following form of the sieve of Erathostenes:

$$(39) \quad \sum_{n \leq x} \chi(n, z) g(n) = \sum_{\substack{dn \leq x \\ d|P(z)}} \mu(d) g(dn).$$

Collecting the factors md on the right-hand side of (35), we can see that the first term is of type I and the second is of type II*. In our case ((5) and

(24)) we conclude that

$$(40) \quad \sum_{\substack{mn \leq x \\ m \sim M}} a_m \chi(n, z) F(mn) = \sum_{\substack{mn \leq x \\ m \sim M}} a_m \chi(n, z) v(mn) + O(\dots)$$

with an acceptable error term, whenever $M \ll z_2$ and $z \leq z_2/z_1$. Indeed, this follows directly if $z_1 \ll M \ll z_2$. If $M < z_1$ we can utilize (35) because we can choose $y = z_1/M$, and thus in the first term $md \leq 2z_1$ (this is acceptable by (14)) and in the second term $z_1 \leq md < 2Myp(d) \leq 2Myz \leq 2z_2$ (this is also acceptable by (14)).

We will use the following definitions:

DEFINITION 1. We say that a number d is *bad* whenever

$$(41) \quad z_2/z_1 < d \leq x/z_2, \quad x/z_1 < d \leq z_1 \quad \text{or} \quad z_2 < d.$$

DEFINITION 2. D_0 is the set of integers $d = p_1 \dots p_r$, $x^{1/2} > p_1 > \dots > p_r > z_2/z_1$, and all divisors of d (except 1) are bad and

$$(42) \quad p_1 \dots p_{2j-1} p_{2j}^2 \leq z_2 \quad \text{for} \quad j \leq [r/2].$$

Note that $1 \in D_0$.

DEFINITION 3. D_1 is the set of integers $d = p_1 \dots p_r$, $r \geq 1$, $x^{1/2} > p_1 > \dots > p_r > z_2/z_1$, d has a good divisor but d/p_r has none, and (42) holds for $j \leq [(r-1)/2]$.

DEFINITION 4. D_2 is the set of integers $d = p_1 \dots p_{2r}$, $r \geq 1$, $x^{1/2} > p_1 > \dots > p_{2r} > z_2/z_1$, all divisors of d (except 1) are bad, (42) holds for $j \leq r-1$, and

$$(43) \quad p_1 \dots p_{2r-1} p_{2r}^2 > z_2.$$

Note that D_0 , D_1 and D_2 are disjoint.

COROLLARY 2.

$$(44) \quad \sum_{n \leq x} \chi(n, x^{1/2}) g(n) = \sum_{\substack{dn \leq x \\ d \in D_0}} \mu(d) \chi(n, z_2/z_1) g(dn) + \\ + \sum_{\substack{dn \leq x \\ d \in D_1}} \mu(d) \chi(n, p(d)) g(dn) + \sum_{\substack{dn \leq x \\ d \in D_2}} \chi(n, p(d)) g(dn).$$

Proof. We apply Lemma 1 with

$$(45) \quad h(d) = \begin{cases} 1 & \text{if } d = d_0 d_1, d_0 \in D_0 \text{ and } d_1 | P(z_2/z_1), \\ 0 & \text{otherwise,} \end{cases}$$

and with

$$(46) \quad G(d) = \sum_{n \leq x/d} g(dn)$$

in place of $g(d)$. We have

$$(47) \quad \bar{h}(d) = \begin{cases} 1 & \text{if } d \in D_1 \cup D_2, \\ 0 & \text{otherwise.} \end{cases}$$

The corollary follows directly from (32).

(44) is our main sieve identity. By writing $g(n) = F(n)$ in (44) we get

$$(48) \quad \sum_{n \leq x} \chi(n, x^{1/2}) F(n) \geq \sum_{\substack{dn \leq x \\ d \in D_0}} \mu(d) \chi(n, z_2/z_1) F(dn) + \\ + \sum_{\substack{dn \leq x \\ d \in D_1}} \mu(d) \chi(n, p(d)) F(dn).$$

In this step we omitted the "bad" bilinear forms (note that $F(n)$ is non-negative), i.e., it is the crucial point of this method. As $d \in D_0$ implies $d \leq z_2$ (see (42)) the first term can be handled by (40). If $d \in D_1$ then d has a good divisor. After some technical difficulties ([3]) this good divisor can be separated and the second sum can be rewritten into the form

$$(49) \quad \sum_{\substack{\delta n \leq x \\ \delta \text{ good}}} a_\delta b_n \chi(n, p(\delta)) F(\delta n).$$

(Note that a good divisor always contains the smallest prime factor of d .) Good δ 's satisfy (14), and so (49) is a good bilinear form in the sense we have a non-trivial bound for

$$(50) \quad \sum_{\substack{\delta n \leq x \\ \delta \text{ good}}} a_\delta b_n \chi(n, p(d)) f(\delta n).$$

(48) and (50) give

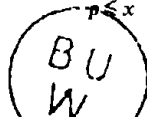
$$(51) \quad \sum_{n \leq x} \chi(n, x^{1/2}) F(n) \geq \sum_{\substack{dn \leq x \\ d \in D_0}} \mu(d) \chi(n, z_2/z_1) v(dn) + \\ + \sum_{\substack{dn \leq x \\ d \in D_1}} \mu(d) \chi(n, p(d)) v(dn) + O(\dots)$$

with an acceptable error term. Again using (44) we arrive at our main result

$$(52) \quad \sum_{n \leq x} \chi(n, x^{1/2}) F(n) \geq \sum_{n \leq x} \chi(n, x^{1/2}) v(n) - \sum_{\substack{dn \leq x \\ d \in D_2}} \chi(n, p(d)) v(dn) + O(\dots)$$

with an acceptable error term. By (25)

$$(53) \quad \sum_{n \leq x} \chi(n, x^{1/2}) g(n) = \sum_{p \leq x} g(p) - \sum_{p \leq x^{1/2}} g(p) + g(x).$$



(The last term appears when $x^{1/2}$ is prime.) Thus (52) implies (23) provided

$$(54) \quad \sum_{\substack{dn \leq x \\ d \in D_2}} \sum \chi(n, p(d)) v(dn) \leq (1-c) \sum_{p \leq x} v(p)$$

with some constant $0 < c < 1$. This is independent of the "unknown" function $F(n)$. In general, it can be verified by the following way. For a fixed $d \in D_2$ the sum over n 's can be estimated by elementary sieve methods and in some cases it can even be determined asymptotically. Provided that the bad intervals in (41) are not too wide, D_2 is a small set. These principles give (54).

Note that the use of Vaughan's identity generally requires the disappearance of the first two intervals in (41).

5

Finally we mention a few cases where the sieve identity method is more efficient than the Vaughan's identity.

THEOREM 1 (Harman [8]). *For any irrational α there exist infinitely many primes p with*

$$(55) \quad \|\alpha p\| < p^{-3/10}.$$

THEOREM 2 (Balog [3]). *For any $2/5 \leq \theta < 1$ and $\varepsilon > 0$ there exist infinitely many primes p with*

$$(56) \quad \|p^\theta\| < p^{-(1-\theta)/2+\varepsilon}.$$

(The use of Vaughan's identity gives (55) with $1/4$ in place of $3/10$ ([13]) and (56) only for $1/2 \leq \theta < 1$ ([2]).)

References

- [1] A. Balog, *Numbers with a large prime factor II*, to appear in Coll. Math. Soc. János Bolyai 34. Topics in Classical Number Theory.
- [2] —, *On the fractional part of p^θ* , to appear in Archiv der Math.
- [3] —, *On the distribution of $p^\theta \bmod 1$* , in preparation.
- [4] A. Balog, G. Harman, J. Pintz, *Numbers with a large prime factor III*, to appear in Quart J. Math. Oxford.
- [5] E. Fouvry, *Repartition des suites alons des progressions arithmetiques*, Thèse de Doctorat d'Etat a l'Université de Bordeaux I, 1981.
- [6] E. Fouvry and H. Iwaniec, *Primes in arithmetic progressions*, Acta Arith. 42 (1982), 197–218.
- [7] S. W. Graham, *The greatest prime factor of the integers in an interval*, J. London Math. Soc. (2) 24 (1981), 427–440.
- [8] G. Harman, *On the distribution of $\alpha p \bmod 1$* , to appear.
- [9] —, *Primes in short intervals*, to appear.

- [10] D. R. Heath-Brown, *Sieve identities and gaps between primes*, a talk at the Journées Arithmétiques de Metz, Sept. 1981.
- [11] D. R. Heath-Brown and H. Iwaniec, *On the difference between consecutive primes*, Invent. Math. 55 (1979), 49–69.
- [12] R. C. Vaughan, *Sommes trigonometriques sur les nombres premiers*, C. R. Acad. Sci. Paris, Ser. A. 19 (1977) (16), 981–984.
- [13] —, *On the distribution of $a \pmod{1}$* , Mathematika 24 (1977), 135–141.
- [14] —, *An elementary method in prime number theory*, Acta. Arith. 37 (1980), 111–115.
- [15] —, *On the estimation of trigonometrical sums over primes and related questions*, Mittag Leffler Institut Report, n. 9, 1977.
- [16] I. M. Vinogradov, *Some theorems concerning the theory of primes*, Mat. Sb. 2 (44), 2 (1937), 179–195.

*Presented to the Semester
Elementary and Analytic Theory of Numbers
September 1–November 13, 1982*
