

Improving on the Brun–Titchmarsh theorem

by

OLIVIER RAMARÉ (Lille) and
JAN-CHRISTOPH SCHLAGE-PUCHTA (Freiburg)

1. Introduction. The main result of this paper is the following theorem:

THEOREM 1.1. *There exists an N_0 such that for all $N \geq N_0$ and all $M \geq 1$ we have*

$$\pi(M + N) - \pi(M) \leq \frac{2N}{\log N + 3.53}.$$

Such theorems have been termed “Brun–Titchmarsh” theorems by Linnik in [4]. Indeed, Titchmarsh proved such a theorem for $q = 1$, with a $\log \log(N/q)$ term instead of the 2 above, to establish the asymptotics for the number of divisors of $p + 1$, p ranging through the primes; he used the method of Brun. The constant 2 (with a $o(1)$) appeared for the first time in [7]. In that work, Selberg also shows that the constant $2 + o(1)$ is optimal in the above, if we stick to sieve methods in a fairly general context. He expanded this theory, now known as the “parity principle”, in [8].

It is thus of interest to try to qualify the $o(1)$ in $2 + o(1)$. The first upper bound of the shape $2N/(\log N + c)$ with an unspecified but very negative c is due to van Lint & Richert [5] though [7] mentions without any proof such a result around equation (6) therein. Bombieri [1] gave the value $c = -3$ and Montgomery & Vaughan [6] the value $c = 5/6$. Section 22 of “lectures on sieves” [9] gives a proof of $c = 2.81$, from which we shall take several elements.

Our method goes through a sieving process which can be seen as a hybrid between the large sieve and the Selberg sieve, and draws on what we call “local models”. With no further input this would lead to Selberg’s results, with a tiny saving in the computations required. But having at our disposal a machinery that allows such a high level of sieving (the moduli we consider

2000 *Mathematics Subject Classification*: 11N13, 11N35, 11N36.

Key words and phrases: Brun–Titchmarsh theorem.

are as large as N), we can handle the remainder in a more detailed way. One can see this process as a weighted large sieve inequality, with weights adapted to the problem. We note that on the way we shall be confronted with the problem of majorizing a step function by a polynomial, and that this problem appears to be numerically more difficult than what meets the eye. And prior to that, we shall also be forced to use numerical means for want of a closed expression for our majorant. As a consequence of these two points, we are not in a position to assert that our result is best possible, even if we restrict attention to our own method!

As a matter of notation, we denote by $\sigma(d)$ the sum of the (positive) divisors of d , and, for any $r \geq 0$, we set $\eta_r(k) = \prod_{p|k} (p^{r+1} + 1)/(p - 1)$.

2. Hilbertian inequalities. Let us start with a complex vector space \mathcal{H} endowed with a hermitian product $[f | g]$, left linear and right sesquilinear.

The easiest exposition goes through a formal definition:

DEFINITION 2.1. By an *almost orthogonal system* in \mathcal{H} , we mean a collection of three sets of data:

1. a finite family $(\varphi_i^*)_{i \in I}$ of points ⁽¹⁾ in \mathcal{H} ,
2. a finite family $(M_i)_{i \in I}$ of positive real numbers,
3. a finite family $(\omega_{i,j})_{i,j \in I}$ of complex numbers with $\omega_{j,i} = \overline{\omega_{i,j}}$,

such that

$$(1) \quad \forall (\xi_i)_i \in \mathbb{C}^I, \quad \left\| \sum_i \xi_i \varphi_i^* \right\|^2 \leq \sum_i M_i |\xi_i|^2 + \sum_{i,j} \xi_i \overline{\xi_j} \omega_{i,j}.$$

Let us comment on this definition. If the family $(\varphi_i^*)_{i \in I}$ were orthogonal, we could ask for equality with $M_i = \|\varphi_i^*\|^2$. It turns out that in applications we have in mind, this family is not orthogonal, but almost so. It is this almost orthogonality that the above condition is meant to measure.

Our first lemma reads as follows:

LEMMA 2.1. *For any finite family $(\varphi_i^*)_{i \in I}$ of points of \mathcal{H} , the system build with $M_i = \sum_j |[\varphi_i^* | \varphi_j^*]|$ and $\omega_{i,j} = 0$ is almost orthogonal.*

So that, when $[\varphi_i^* | \varphi_j^*]$ is small for $i \neq j$, then M_i is indeed close to $\|\varphi_i^*\|^2$.

Proof. We write

$$\left\| \sum_i \xi_i \varphi_i^* \right\|^2 = \sum_{i,j} \xi_i \overline{\xi_j} [\varphi_i^* | \varphi_j^*]$$

and simply apply $2|\xi_i \overline{\xi_j}| \leq |\xi_i|^2 + |\xi_j|^2$. The lemma readily follows. ■

⁽¹⁾ The reader may wonder why we chose to denote the members of this family with a star. This is to be consistent and to avoid confusion with notations that will appear later on.

Here is an enlightening reading of this lemma: the hermitian form that appears has a matrix whose diagonal terms are the $\|\varphi_i^*\|^2$'s. A theorem of Gershgorin says that all eigenvalues of this matrix lie in the so-called *Gershgorin disc* centered at one $\|\varphi_i^*\|^2$ and with radius $\sum_{j \neq i} \|\varphi_i^* | \varphi_j^*\|$. This approach is due to [3]. It has a drawback: we do not know that each Gershgorin disc does indeed contain an eigenvalue, a flaw that is somehow repaired in the above lemma.

In general, and only under (1), we get the following kind of Parseval inequality:

LEMMA 2.2. *For any almost orthogonal system, and any $f \in \mathcal{H}$, set $\xi_i = [f | \varphi_i^*]/M_i$. Then*

$$\sum_i M_i^{-1} |[f | \varphi_i^*]|^2 \leq \|f\|^2 + \sum_{i,j} \xi_i \bar{\xi}_j \omega_{i,j}.$$

Once again, the orthogonal case is enlightening: when the (φ_i^*) are orthogonal, then we may take $M_i = \|\varphi_i^*\|^2$ and $\omega_{i,j} = 0$. The LHS becomes the square of the norm of the orthonormal projection of f on the subspace generated by the φ_i^* 's.

Without the $\omega_{i,j}$'s and appealing to Lemma 2.1, this is due to Selberg, as mentioned in Section 2 of [2] and in [1].

Proof. Simply write

$$\left\| f - \sum_i \xi_i \varphi_i^* \right\|^2 \geq 0$$

and expand the square. We take care of $\|\sum_i \xi_i \varphi_i^*\|^2$ by using (1), getting

$$\|f\|^2 - 2\Re \sum_i \bar{\xi}_i [f | \varphi_i^*] + \sum_i M_i |\xi_i|^2 + \sum_{i,j} \xi_i \bar{\xi}_j \omega_{i,j} \geq 0.$$

We now choose ξ_i 's to the best of our interest, neglecting the bilinear form containing the $\omega_{i,j}$'s. We take $\xi_i = [f | \varphi_i^*]/M_i$. The lemma readily follows. ■

Combining Lemmas 2.2 and 2.1 yields what is usually known as “Selberg’s lemma” in this context. The introduction of the $\omega_{i,j}$'s, due to the authors, is to enable a refined treatment of the error term as well as a hybrid way between weighted large sieve and Selberg sieve results.

The value of ξ_i in the statement is usually of no importance, only its order of magnitude being relevant.

We now discuss a special problem to introduce our next theorem. In some cases, a partial treatment of the bilinear form is readily available in the shape of

$$(2) \quad \forall (\xi_i)_i \in \mathbb{C}^I, \quad \left\| \sum_i \xi_i \varphi_i^* \right\|^2 \leq \sum_i M_i |\xi_i|^2 + \left(\sum_i |\xi_i| n_i \right)^2 + \sum_{i,j} \xi_i \bar{\xi}_j \omega_{i,j}$$

for some positive M_i and n_i (here again, M_i is generally an approximation of $\|\varphi_i^*\|^2$). With such an inequality at hand, the above proof leads to

$$(3) \quad \|f\|^2 - 2\Re \sum_i \bar{\xi}_i [f | \varphi_i^*] + \sum_i M_i |\xi_i|^2 + \left(\sum_i |\xi_i| n_i \right)^2 + \sum_{i,j} \xi_i \bar{\xi}_j \omega_{i,j} \geq 0.$$

When using it, we take for φ_i^* a kind of local approximation of f , which implies that we can assume $[f | \varphi_i^*]$ to be a non-negative real number. It is then readily seen that the ξ_i 's that minimize the RHS are non-negative. Finally, we are led to choose those ξ_i 's so as to minimize

$$\|f\|^2 - 2 \sum_i \xi_i [f | \varphi_i^*] + \sum_i M_i \xi_i^2 + \left(\sum_i \xi_i n_i \right)^2.$$

We handle the optimization of (3) with calculus by setting $\xi_i = \zeta_i^2$. After some manipulations, we conclude that there exists a subset I' of I such that $\xi_i = 0$ if $i \in I \setminus I'$ and

$$(4) \quad \forall i \in I', \quad \xi_i = \frac{[f | \varphi_i^*] - X n_i}{M_i}, \quad X = \frac{\sum_{j \in I'} n_j [f | \varphi_j^*] / m_j}{1 + \sum_{j \in I'} n_j^2 / m_j}$$

provided

$$(5) \quad \forall i \in I', \quad [f | \varphi_i^*] / n_i \geq X.$$

However, determining an optimal I' is difficult: the index i appears on the left-hand side of (5), but also on its right-hand side since the definition of X depends on whether this index belongs to I' or not. It is easier to set

$$(6) \quad \xi_i = \frac{[f | \varphi_i^*] - Y n_i}{M_i},$$

for a Y to be chosen but which guarantees $\xi_i \geq 0$. The optimal Y is of course $Y = X$.

Once we have inferred the form of these weights, we can simply plug them in the proof of Lemma 2.2 without even mentioning (2). Here is the theorem we have reached:

THEOREM 2.1. *Let an almost orthogonal system be given with notations as above and let $f \in \mathcal{H}$. Let also Y and $(n_i)_i$ be non-negative real numbers. Assume that $[f | \varphi_i^*]$'s are real numbers. Set $\xi_i = ([f | \varphi_i^*] - Y n_i) / M_i$ for all i . Then*

$$\sum_i M_i \xi_i^2 + 2Y \sum_i n_i \xi_i - \sum_{i,j} \xi_i \bar{\xi}_j \omega_{i,j} \leq \|f\|^2.$$

Of course, the preliminary discussion tells us that it will be better to have $\xi_i \geq 0$, but the statement is valid as is, and may offer some more flexibility.

3. Integers coprime to a fixed modulus in an interval. Let f be a positive integer and define $\rho = \phi(f)/f$. We study the following two functions of the real non-negative variable u :

$$\theta_f^-(u) = \min_{y \in \mathbb{R}} \min_{\substack{0 \leq x \leq u \\ x \in \mathbb{R}}} \left(\sum_{\substack{y < n \leq y+x \\ (n,f)=1}} 1 - \rho x \right),$$

$$\theta_f^+(u) = \max_{y \in \mathbb{R}} \max_{\substack{0 \leq x \leq u \\ x \in \mathbb{R}}} \left(\sum_{\substack{y < n \leq y+x \\ (n,f)=1}} 1 - \rho x \right).$$

The introduction of these two functions is inspired from Section 22 of “lectures on sieves” in [9]. In order to compute them, we need to restrict both x and y to integer values. This is the role of the next lemma.

LEMMA 3.1. *We have*

$$\theta_f^-(u) = \min_{l \in \mathbb{N}} \left(\min_{\substack{k \in \mathbb{N} \\ 0 \leq k \leq u}} \left(\sum_{\substack{l+1 \leq n \leq l+k-1 \\ (n,f)=1}} 1 - \rho k \right), \sum_{\substack{l+1 \leq n \leq l+[u] \\ (n,f)=1}} 1 - \rho u \right),$$

$$\theta_f^+(u) = \max_{\substack{k, l \in \mathbb{N} \\ k < u+1}} \left(\sum_{\substack{l \leq n \leq l+k-1 \\ (n,f)=1}} 1 - \rho(k-1) \right).$$

The function θ_f^+ is a non-decreasing step function which is left continuous with jump points at integer points. The function θ_f^- is non-increasing continuous; it alternates linear pieces of slope $-\rho$ and constant pieces. Changes occur at integer points. Both functions are constant for $u \geq f$.

Proof. We start with θ_f^+ . First fix y . The function $\sum_{y < n \leq y+x} w(n) - \rho x$ is linear non-increasing in x from 0 to $1 - \{y\}$, then from $1 - \{y\}$ to $2 - \{y\}$ and so on. Its maximum value is reached at $x = 0$ or $x = k - \{y\}$ for some integer k , thus

$$\theta_f^+(u) = \max_{y \in \mathbb{R}} \max_{\substack{k \in \mathbb{N} \\ k \leq u + \{y\}}} \left(\sum_{y < n \leq [y] + k} w(n) + \rho(-k + \{y\}) \right).$$

The condition $k \leq u + \{y\}$ is increasing in $\{y\}$ and so is the term to maximize. We may thus take y to be just below an integer l , reaching the expression we announced.

As for θ_f^- , we start similarly by fixing y . The minimum is reached at $x = k - \{y\} - 0$ or at $x = u$, where k is an integer and the -0 means we are to take x just below this value. We see that $\theta_f^-(u)$ equals

$$\min_{y \in \mathbb{R}} \left(\min_{\substack{k \in \mathbb{N} \\ k \leq u + \{y\}}} \left(\sum_{y < n \leq [y] + k - 1} w(n) + \rho(\{y\} - k) \right), \sum_{y < n \leq [y] + u} w(n) - \rho u \right).$$

As far as the last sum is concerned, the worst case is when y is an integer $l \geq 0$, so it reduces to

$$(7) \quad \min_{l \in \mathbb{N}} \left(\sum_{l+1 \leq n \leq l+u} w(n) - \varrho u \right).$$

For the first minimum, we distinguish between $k \leq [u]$ and $k = [u] + 1$ (which can only happen if u is not an integer). If $k \leq [u]$, we may take y to be integral. If $k = [u] + 1$, then $\{y\} \geq 1 - \{u\}$, which is indeed the worst case: we take $y = l + 1 - \{u\}$. This last contribution turns out to be exactly the same as the one in (7). ■

Next we consider the function

$$(8) \quad \theta_f^*(v) = \max(\theta_f^+(1/v), -\theta_f^-(1/v))$$

which is right continuous with jump points at the $1/m$'s, where m ranges over the integers from 1 to f . Of course, $\theta_f^*(1) = \theta_f^+(1) = 1$.

Case of $f = 210$. The function $\theta_{210}^*(1/u)$ equals

$$\left\{ \begin{array}{ll} 1 & \text{if } 0 < u \leq 1, \\ 54/35 & \text{if } 1 < u \leq 3, \\ 57/35 & \text{if } 3 < u \leq 7, \\ 76/35 & \text{if } 7 < u \leq 9, \\ 79/35 & \text{if } 9 < u \leq 79/8, \end{array} \right. \quad \left\{ \begin{array}{ll} 8u/35 & \text{if } 79/8 \leq u \leq 10, \\ 16/7 & \text{if } 10 < u \leq 13, \\ 82/35 & \text{if } 13 < u \leq 17, \\ 94/35 & \text{if } 17 < u \leq 41/2, \\ 8u/35 - 2 & \text{if } 41/2 \leq u \leq 22, \\ 106/35 & \text{if } 22 < u \leq 210. \end{array} \right.$$

The following plot displays the step function θ_{210}^* as well as the optimizing polynomial we shall compute in Section 7.

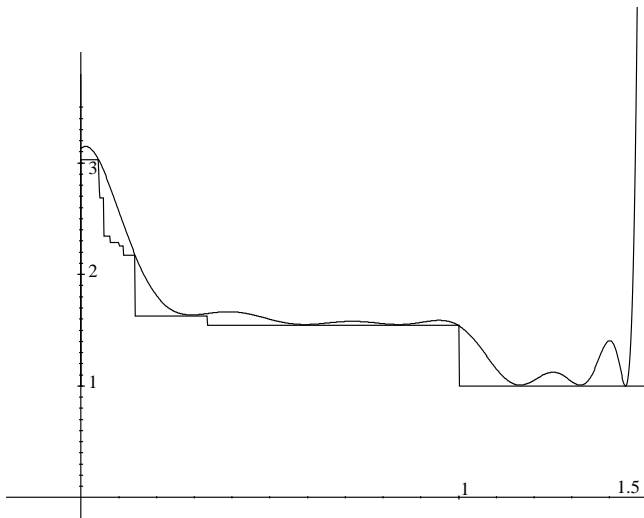


Fig. 1. Comparison of θ_{210}^* and the optimizing polynomial

Polynomial approximation of $\theta_f^(v)$.* We shall require a good polynomial upper bound for θ_f^* :

$$(9) \quad \theta_f^*(v) \leq \sum_{0 \leq r \leq R} b_r v^r.$$

Finding such an approximation turned out to be much more tricky than expected. Our first idea was to start with a polynomial approximation of $\theta_f^*(v)$ of the form

$$\left| \theta_f^*(v) - \sum_{0 \leq r \leq R} \tilde{b}_r v^r \right| \leq \varepsilon$$

for $0 \leq v \leq V$, from which an upper bound is easily derived by increasing the constant term. We carried out this scheme with Bernstein polynomials, with poor numerical results and fitting. We even tried to achieve such an approximation on a larger interval since endpoints are notoriously troublesome: this helped a bit but not much, despite the fact that we used polynomials of very high degree (up to 200). We finally decided for a different scheme described in the last section.

4. Local models for the sequence of primes

4.1. Choice of the local system. Let us start with a general discussion on what “sieving” means. Sieving is about gaining information on a sequence from what we know of it modulo d for several d ’s. If one looks at the sequence of primes modulo d and if we neglect the prime divisors of d , we simply get the set of reduced residue classes modulo d . Thus, on one hand we have the characteristic function of the primes of the interval $[M + 1, M + N]$, say f , and on the other hand the characteristic function φ_d of the integers in this interval that are coprime to d for all $d \leq \sqrt{N}$. Notice here that it is enough to restrict our attention to squarefree d ’s.

On recalling what we did in Section 2, we could simply try to get an approximation of f in terms of the φ_d ’s. However, the study there concerns almost orthogonal φ_q ’s, which is not the case of the sequence $(\varphi_d)_d$: if $q \mid d$, knowing that a given integer is coprime to d implies it is coprime to q , so there is redundancy of information. This implies in turn that these functions are far from being linearly independent. We unscrew the situation in the following way. When d is squarefree, we set

$$(10) \quad \frac{d}{\phi(d)} \varphi_d = \sum_{q \mid d} \varphi_q^*$$

where

$$(11) \quad \varphi_q^*(n) = \mu(q)c_q(n)/\phi(q)$$

and $c_q(n)$ is the Ramanujan sum given by

$$(12) \quad c_q(n) = \sum_{a \bmod^* q} e(na/q) = \sum_{l|q} l\mu(q/l).$$

Verifying (10) is easy:

$$\sum_{q|d} \mu(q)c_q(n)/\phi(q) = \prod_{p|d} \left(1 - \begin{cases} 1 & \text{if } p \mid n \\ -1/(p-1) & \text{otherwise} \end{cases} \right).$$

In our problem, we shall select a fixed integer f that will be taken to be 210 at the end of the proof and consider the characteristic function w of the points in $[M + 1, M + N]$ that are coprime to f . This being chosen, our hermitian product on sequences over $[M + 1, M + N]$ is defined by

$$(13) \quad [g \mid h] = \sum_{M+1 \leq n \leq M+N} w(n)g(n)\overline{h(n)}.$$

Furthermore, we take the moduli q in the set

$$(14) \quad \{q : \sigma(q) \leq S, \mu(q)^2 = 1, (q, f) = 1\},$$

where $\sigma(q) = \sum_{d|q} d$. The reason for this choice will become apparent later on.

4.2. Study of the local models. Notice that

$$(15) \quad [\varphi_q^* \mid \varphi_{q'}^*] = \frac{\mu(q)}{\phi(q)} \frac{\mu(q')}{\phi(q')} \sum_n w(n)c_q(n)c_{q'}(n).$$

We note that when q and q' have a common factor, say δ , then $c_\delta(n)^2 = \phi((n, \delta))^2$ would factor out: this contribution is non-negative and we want to use this fact here. Let Δ be a squarefree integer coprime to f . Write $(q, q', \Delta) = \delta$, so that $[\varphi_q^* \mid \varphi_{q'}^*]$ equals

$$\frac{\mu(q)\mu(q')}{\phi(q)\phi(q')} \sum_{\substack{l|q/\delta \\ l'|q'/\delta \\ h|\delta}} l\mu(q/l)l'\mu(q'/l')(\mu \star \phi^2)(h) \left(\frac{\phi(f)}{f} \frac{N}{h[l, l']} + R_{h[l, l']}(M, N, f) \right)$$

where

$$(16) \quad R_d(M, N, f) = \sum_{\substack{M+1 \leq n \leq M+N \\ d|n}} w(n) - \frac{\phi(f)N}{fd}.$$

Recall that we have set $\varrho = \phi(f)/f$ in Section 3 to simplify typographical work. The reader will check that the main term (corresponding to $\varrho N/[l, l']$) vanishes when $q \neq q'$ and is $\varrho N/\phi(q)$ otherwise. We carry over this change to the bilinear form $\|\sum_q \xi_q \varphi_q^*\|^2$, which equals the diagonal term $\varrho N \sum_q |\xi_q|^2/\phi(q)$ augmented by

$$\mathfrak{R} = \sum_{\delta_1 \delta_2 \delta_3 | \Delta} \frac{\mu(\delta_2 \delta_3)}{\phi(\delta_1)^2 \phi(\delta_2) \phi(\delta_3)} \sum_{\substack{(l, f\Delta)=1 \\ (l', f\Delta)=1}} \frac{\mu(l) \xi_{\delta_1 \delta_2 l}}{\phi(l)} \frac{\mu(l') \xi_{\delta_1 \delta_3 l'}}{\phi(l')} \\ \times \sum_{\substack{d | l \delta_2 \\ d' | l' \delta_3 \\ h | \delta_1}} dd' \mu(l \delta_2 / d) \mu(l' \delta_3 / d') (\mu \star \phi^2)(h) R_{h[d, d']}(M, N, f).$$

At this level, we say that

$$|R_{h[d, d']}(M, N, f)| \leq \theta_f^*(h[d, d'] / N) \leq \sum_{0 \leq r \leq R} b_r(h[d, d'] / N)^r$$

by (9). We infer that

$$\mathfrak{R} \leq \sum_{0 \leq r \leq R} b_r N^{-r} \sum_{\delta_1 \delta_2 \delta_3 | \Delta} \frac{1}{\phi(\delta_1)^2 \phi(\delta_2) \phi(\delta_3)} \sum_{\substack{(l, f\Delta)=1 \\ (l', f\Delta)=1}} \frac{|\xi_{\delta_1 \delta_2 l}|}{\phi(l)} \frac{|\xi_{\delta_1 \delta_3 l'}|}{\phi(l')} \\ \times \sum_{\substack{d | l \delta_2 \\ d' | l' \delta_3 \\ h | \delta_1}} dd' (\mu \star \phi^2)(h) h^r [d, d']^r.$$

This leads to

$$\mathfrak{R} \leq \sum_{0 \leq r \leq R} b_r N^{-r} \sum_{\delta_1 \delta_2 \delta_3 | \Delta} \frac{\prod_{p | \delta_1} (1 + p^{r+1}(p-2)) \eta_r(\delta_2 \delta_3)}{\phi(\delta_1)^2} \\ \times \sum_{\substack{(l, f\Delta)=1 \\ (l', f\Delta)=1}} |\xi_{\delta_1 \delta_2 l}| \eta_r(l) |\xi_{\delta_1 \delta_3 l'}| \eta_r(l') \prod_{p | (l, l')} \frac{1 + 2p^{r+1} + p^{r+2}}{(1 + p^{r+1})^2}.$$

The factor that depends on (l, l') is somewhat troublesome. We handle it in the following way: for $r = 0$, it is equal to 1; otherwise let P be the smallest prime number that does not divide $f\Delta$. This prime factor will tend to infinity, and we approximate the factor depending on (l, l') essentially by $1 + \mathcal{O}(P^{-1})$. More precisely, we write

$$\sum_{\substack{(l, f\Delta)=1 \\ (l', f\Delta)=1}} |\xi_{\delta_1 \delta_2 l}| \eta_r(l) |\xi_{\delta_1 \delta_3 l'}| \eta_r(l') \left| \prod_{p | (l, l')} \frac{1 + 2p^{r+1} + p^{r+2}}{(1 + p^{r+1})^2} - 1 \right| \\ \ll_r \sum_{p \geq P} \sum_{\substack{(m, p f\Delta)=1 \\ (m', p f\Delta)=1}} |\xi_{\delta_1 \delta_2 pm}| \eta_r(pm) |\xi_{\delta_1 \delta_3 pm'}| \eta_r(pm') \\ \ll_r \sum_{p \geq P} p^{2r} \sum_{m, m'} |\xi_{\delta_1 \delta_2 pm}| \eta_r(m) |\xi_{\delta_1 \delta_3 pm'}| \eta_r(m').$$

The idea here is that the factor $\xi_{\delta_1\delta_2pm}$ forces m to be rather small. Indeed, anticipating on the values of ξ in (20) below and using Lemma 5.1, we find the above to be not more than

$$(17) \quad \left(\frac{Z}{\varrho N}\right)^2 \sum_{p \geq P} p^{2r} (S/p)^{2r+2} \ll_r \left(\frac{Z}{\varrho N}\right)^2 S^{2r+2} P^{-1}.$$

This will give rise to the error term

$$\left(\frac{Z}{\varrho N}\right)^2 \sum_{\delta_1\delta_2\delta_3|\Delta} \frac{\eta_r^b(\delta_1)\eta_r(\delta_2\delta_3)}{\sigma(\delta_1\delta_2)^{r+1}\sigma(\delta_1\delta_3)^{r+1}} \sum_{1 \leq r \leq R} \frac{S^{2r+2}|b_r|}{N^r P}$$

which, up to a multiplicative constant, is not more than

$$(18) \quad \left(\frac{Z}{\varrho N}\right)^2 \prod_{p|\Delta} (1+p^{-1})^2 \sum_{1 \leq r \leq R} \frac{S^{2r+2}|b_r|}{N^r P}.$$

The factor P^{-1} will indeed be enough to control this quantity. Hence, again anticipating on (20), we reach

$$\begin{aligned} \left\| \sum_q \xi_q \varphi_q^* \right\|^2 &\leq \varrho N \sum_q |\xi_q|^2 / \phi(q) \\ &\quad + \sum_{0 \leq r \leq R} \frac{b_r}{N^r} \sum_{\delta_1\delta_2\delta_3|\Delta} \eta_r^b(\delta_1) \sum_{\substack{(l,f\Delta)=1 \\ (l',f\Delta)=1}} |\xi_{\delta_1\delta_2l} \eta_r(\delta_2l)| \xi_{\delta_1\delta_3l'} |\eta_r(\delta_3l')| \\ &\quad + \mathcal{O}\left(\left(\frac{Z}{\varrho N}\right)^2 \prod_{p|\Delta} (1+p^{-1})^2 \sum_{1 \leq r \leq R} \frac{S^{2r+2}|b_r|}{N^r P}\right). \end{aligned}$$

5. Some arithmetical auxiliaries. We need to evaluate some rather unusual averages.

LEMMA 5.1. *Let f^* be a positive integer. Set $\varrho^* = \phi(f^*)/f^*$ and $t(q) = 1 - \sigma(q)/S^*$. For any real number S^* going to infinity, we have*

$$\sum_{\substack{q/\sigma(q) \leq S^* \\ (q,f^*)=1}} \frac{t(q)^2}{\phi(q)} = \varrho^* (\text{Log } S^* + \kappa(f^*)) + o(1)$$

with

$$\kappa(f^*) = \gamma + \sum_{p \geq 2} \frac{\text{Log } p}{p(p-1)} - \sum_p \frac{\text{Log}(1+p^{-1})}{p} + \sum_{p|f^*} \frac{\text{Log}(p+1)}{p} - \frac{3}{2}$$

($\kappa(210) = 1.11537\dots$) and

$$\sum_{\substack{q/\sigma(q) \leq S^* \\ (q,f^*)=1}} \eta_r(q)t(q) = \frac{\varrho^*}{2(r+1)} \prod_{p|f^*} \left(1 - \frac{p^r - 1}{p^{r+1}(p+1)}\right) S^{*(r+1)} + o(S^{*(r+1)}).$$

Proof. The first estimate comes from [9]. We closely follow Selberg’s proof to get

$$\sum_{\substack{q/\sigma(q) \leq S^* \\ (q, f^*)=1}} \frac{\eta_r(q)t(q)}{q^r} = \frac{\varrho}{2} \prod_{p|f^*} \left(1 - \frac{p^r - 1}{p^{r+1}(p+1)}\right) S^* + o(S^*)$$

from which we get

$$\begin{aligned} \sum_{\substack{q/\sigma(q) \leq S^* \\ (q, f^*)=1}} \eta_r(q)t(q) &= \frac{\varrho^*}{2(r+1)} \prod_{p|f^*} \left(1 - \frac{p^r - 1}{p^{r+1}(p+1)}\right) S^{*(r+1)} + o(S^{*(r+1)}). \blacksquare \end{aligned}$$

Note that the quantities we end up computing are the same as the ones that appear in [9] though we have one less to handle.

Define

$$(19) \quad C_r(\Delta) = \frac{\phi(\Delta)^2}{\Delta^2} \sum_{\delta_1 \delta_2 \delta_3 | \Delta} \frac{\prod_{p|\delta_1} (1 + p^{r+1}(p-2)) \eta_r(\delta_2 \delta_3)}{\phi(\delta_1)^2 \sigma(\delta_1)^{2r+2} \sigma(\delta_2)^{r+1} \sigma(\delta_3)^{r+1}}.$$

We have

LEMMA 5.2.

$$\begin{aligned} C_r(\Delta) &= \prod_{p|\Delta} \left(\frac{(p-1)^2}{p^2} + \frac{2(p-1)(p^{r+1}+1)}{p^2(p+1)^{r+1}} \right. \\ &\quad \left. + \frac{1+p^{r+1}(p-2)}{p^2(p+1)^{2r+2}} + \frac{(1+p^{r+1}(p-2))(p^{r+1}+1)}{(p-1)p^2(p+1)^{3r+3}} \right). \end{aligned}$$

Proof. We start with δ_3 :

$$\begin{aligned} \sum_{\delta_3 | \Delta / (\delta_1 \delta_2)} \frac{\eta_r(\delta_3)}{\sigma(\delta_3)^{r+1}} &= \prod_{p|\Delta / (\delta_1 \delta_2)} \left(1 + \frac{1+p^{r+1}}{(p+1)^{r+1}(p-1)} \right) \\ &= \prod_{p|\Delta / (\delta_1 \delta_2)} \frac{(p+1)^{r+1}(p-1) + p^{r+1} + 1}{(p+1)^{r+1}(p-1)}. \end{aligned}$$

Our sum reduces to

$$\begin{aligned} &\prod_{p|\Delta} \frac{((p+1)^{r+1}(p-1) + p^{r+1} + 1)(p-1)}{p^2(p+1)^{r+1}} \\ &\quad \times \sum_{\delta_1 \delta_2 | \Delta} \frac{\prod_{p|\delta_1} (1 + p^{r+1}(p-2)) \eta_r(\delta_2)}{\phi(\delta_1)^2 \sigma(\delta_1)^{2r+2} \sigma(\delta_2)^{r+1}} \prod_{p|\delta_1 \delta_2} \frac{(p+1)^{r+1}(p-1)}{(p+1)^{r+1}(p-1) + p^{r+1} + 1}. \end{aligned}$$

We continue with δ_2 :

$$\begin{aligned} \sum_{\delta_2|\Delta/\delta_1} \frac{1 + p^{r+1}}{(p-1)(p+1)^{r+1}} \frac{(p+1)^{r+1}(p-1)}{(p+1)^{r+1}(p-1) + p^{r+1} + 1} \\ = \prod_{p|\Delta/\delta_1} \left(1 + \frac{p^{r+1} + 1}{(p+1)^{r+1}(p-1) + p^{r+1} + 1} \right) \\ = \prod_{p|\Delta/\delta_1} \frac{(p+1)^{r+1}(p-1) + 2p^{r+1} + 2}{(p+1)^{r+1}(p-1) + p^{r+1} + 1}. \end{aligned}$$

Hence our quantity reduces to

$$\begin{aligned} \prod_{p|\Delta} \frac{((p+1)^{r+1}(p-1) + 2p^{r+1} + 2)(p-1)}{p^2(p+1)^{r+1}} \\ \times \sum_{\delta_1|\Delta} \prod_{p|\delta_1} \frac{1 + p^{r+1}(p-2)}{(p-1)^2(p+1)^{2r+2}} \frac{(p+1)^{r+1}(p-1) + p^{r+1} + 1}{(p+1)^{r+1}(p-1) + 2p^{r+1} + 2}, \end{aligned}$$

which reads

$$\begin{aligned} \prod_{p|\Delta} \frac{((p+1)^{r+1}(p-1) + 2p^{r+1} + 2)(p-1)}{p^2(p+1)^{r+1}} \\ \times \frac{(p-1)^2(p+1)^{2r+2}((p+1)^{r+1}(p-1) + 2p^{r+1} + 2) + (1 + p^{r+1}(p-2))((p+1)^{r+1}(p-1) + p^{r+1} + 1)}{(p-1)^2(p+1)^{2r+2}((p+1)^{r+1}(p-1) + 2p^{r+1} + 2)} \\ = \prod_{p|\Delta} \frac{(p-1)^2(p+1)^{2r+2}((p+1)^{r+1}(p-1) + 2p^{r+1} + 2) + (1 + p^{r+1}(p-2))((p+1)^{r+1}(p-1) + p^{r+1} + 1)}{(p-1)p^2(p+1)^{3r+3}}. \blacksquare \end{aligned}$$

6. Using the hermitian inequality. Optimizing in ξ is too difficult.

We stick to the simplest choice: $M_i = \varrho N / \phi(q)$, $[f | \varphi_i^*] / M_i = Z / (\varrho N)$, $n_i = \sigma(q) / \phi(q)$ and $Y = Z / S$ for a parameter S we shall choose later on. This leads to

$$(20) \quad \xi_q = \frac{Z}{\varrho N} t(q), \quad t(q) = 1 - \frac{\sigma(q)}{S}.$$

We invoke Lemma 5.1 to compute the relevant mean values, and for instance, we use $S^* = S / \sigma(\delta_1 \delta_2)$ and $\mathfrak{f}^* = \mathfrak{f} \Delta$ to evaluate $\sum_{(l, \mathfrak{f} \Delta)=1} |\xi_{\delta_1 \delta_2 l}| \eta_r(l)$. There appear constants in the form of an Euler product, say $\mathfrak{S}_r(\mathfrak{f}^*)$, which we again approximate by $1 + \mathcal{O}(P^{-1})$. Let us give some details. In a first

step we reach

$$Z \geq \left(\frac{Z}{\varrho N}\right)^2 \varrho^2 N(\text{Log } S + \kappa(\mathfrak{f})) + \frac{2Z^2}{\varrho NS} \sum_{(q,\mathfrak{f})=1} \frac{\sigma(q)t(q)}{\phi(q)} - A + \mathcal{O}(B)$$

with $g(\delta) = \prod_{p|\delta}(1 + p^{r+1}(p - 2))/(p - 1)^2$ and

$$A = \sum_{0 \leq r \leq R} \frac{Z^2 b_r}{\varrho^2 N^{r+2}} \sum_{\delta_1 \delta_2 \delta_3 | \Delta} g(\delta_1) \sum_{\substack{(l,\mathfrak{f}\Delta)=1 \\ (l',\mathfrak{f}\Delta)=1}} t(\delta_1 \delta_2 l) \eta_r(l \delta_2) t(\delta_1 \delta_3 l') \eta_r(l' \delta_3),$$

$$B = \sum_{\delta_1 \delta_2 \delta_3 | \Delta} g(\delta_1) \sum_{1 \leq r \leq R} \sum_{p \geq P} \frac{|b_r| p^{2r}}{N^r} \sum_{l, l'} |\xi_{\delta_1 \delta_2 p l} \eta_r(l \delta_2)| |\xi_{\delta_1 \delta_3 p l'} \eta_r(l' \delta_3)|.$$

We tidy this expression step by step:

$$N \geq Z(\text{Log } S + \kappa(\mathfrak{f})) + Z$$

$$- Z \sum_{0 \leq r \leq R} \frac{b_r S^{2r+2}}{\varrho^2 N^{r+1}} \sum_{\delta_1 \delta_2 \delta_3 | \Delta} \frac{\prod_{p|\delta_1}(1 + p^{r+1}(p - 2)) \eta_r(\delta_2 \delta_3)}{\phi(\delta_1)^2 \sigma(\delta_1)^{2r+2} \sigma(\delta_2)^{r+1} \sigma(\delta_3)^{r+1}} \frac{\varrho^2 \phi(\Delta)^2 \mathfrak{S}_r(\mathfrak{f}\Delta)^2}{4\Delta^2(r + 1)^2}$$

$$+ \mathcal{O}\left(\prod_{p|\Delta}(1 + p^{-1})^2 S^2 Z N^{-1} P^{-1} \sum_{1 \leq r \leq R} |b_r|(S^2/N)^r\right) + o(N).$$

This leads to

$$N/Z \geq \text{Log } S + \kappa(\mathfrak{f}) + 1 - \sum_{0 \leq r \leq R} \frac{b_r (S^2/N)^{r+1}}{4(r + 1)^2} C_r(\Delta) \mathfrak{S}_r(\mathfrak{f}\Delta)^2$$

$$+ \mathcal{O}\left(\prod_{p|\Delta}(1 + p^{-1})^2 P^{-1} \sum_{1 \leq r \leq R} |b_r|(S^2/N)^{r+1}\right) + o(1).$$

And since $\mathfrak{S}_r(\mathfrak{f}\Delta) = 1 + \mathcal{O}(P^{-1})$, we finally obtain

$$N/Z - \frac{1}{2} \text{Log } N \geq \frac{1}{2} \text{Log}(S^2/N) + \kappa(\mathfrak{f}) + 1 - \sum_{0 \leq r \leq R} \frac{b_r (S^2/N)^{r+1}}{4(r + 1)^2} C_r(\infty/\mathfrak{f})$$

$$+ \mathcal{O}\left(\prod_{p|\Delta}(1 + p^{-1})^2 P^{-1} \sum_{1 \leq r \leq R} |b_r|(S^2/N)^{r+1}\right) + o(1).$$

At this level, we send Δ (and P) to infinity and we are left with finding an optimal value for S^2/N . It would be satisfactory to have an expression for the final constant, but we are not able to attain such a precision. In particular, the b_r 's should not appear in the final expression. We are, however, able to get numerical results.

7. Optimizing the polynomial via linear programming. It is better at this level to change notation slightly and set

$$c_r(x) = \prod_{7 < p \leq x} \left(\frac{(p-1)^2}{p^2} + \frac{2(p-1)(p^{r+1}+1)}{p^2(p+1)^{r+1}} + \frac{1+p^{r+1}(p-2)}{p^2(p+1)^{2r+2}} + \frac{(1+p^{r+1}(p-2))(p^{r+1}+1)}{(p-1)p^2(p+1)^{3r+3}} \right),$$

together with

$$(21) \quad c_r = C_r(\infty/f) = \lim_{x \rightarrow \infty} c_r(x).$$

LEMMA 7.1. *For $x > 2r + 2$ we have the estimate*

$$c_r(x) \geq c_r \geq c_r(x) \left(1 - \frac{2r+2}{x} \right).$$

Proof. Denote by $F(p)$ the factor in $c_r(x)$ corresponding to the prime number p . We then have

$$\begin{aligned} \frac{(p-1)^2}{p^2} + \frac{2(p-1)(p^{r+1}+1)}{p^2(p+1)^{r+1}} &\leq F(p) \\ &\leq \frac{(p-1)^2}{p^2} + \frac{2(p-1)(p^{r+1}+1)}{p^2(p+1)^{r+1}} + \frac{2}{p^2(p+1)^r}, \end{aligned}$$

and the left-hand expression can be written as

$$1 - \frac{2 - 2p + 2p^{1+r} - 2p^{2+r} - (1+p)^r + p(1+p)^r + 2p^2(1+p)^r}{p^2(1+p)^{r+1}},$$

from which it is obvious that $F(p) < 1$ for $r \geq 1$. This can be checked directly when $r = 0$. On the other hand, for $p \geq r$ we have the estimate

$$p^r = ((p+1) - 1)^r = \sum_{\nu=0}^r (-1)^\nu \binom{r}{\nu} (p+1)^{r-\nu} \geq (p+1)^r - r(p+1)^{r-1},$$

since the binomial sum is alternating and decreasing, which implies, for $r \geq 1$, the estimate

$$\begin{aligned} F(p) &\geq 1 - \frac{2 - 2p + 2p^{1+r} - (1+p)^r + p(1+p)^r + 2rp^2(1+p)^{r-1}}{p^2(1+p)^{r+1}} \\ &\geq 1 - \frac{(2r+2)p^2(p+1)^{r-1}}{p^2(p+1)^{r+1}} \geq \left(1 - \frac{1}{p^2} \right)^{2r+2}. \end{aligned}$$

Again, for $r = 0$ this bound can be checked directly. We now have

$$\begin{aligned} c_r &= c_r(x) \prod_{p > x} F(p) \geq c_r(x) \prod_{p > x} \left(1 - \frac{1}{p^2} \right)^{2r+2} \geq c_r(x) \left(1 - \sum_{n > x} n^{-2} \right)^{2r+2} \\ &\geq c_r(x) \left(1 - \frac{1}{x} \right)^{2r+2} \geq c_r(x) \left(1 - \frac{2r+2}{x} \right). \blacksquare \end{aligned}$$

Next we compute a polynomial P such that $P(Vn/300) > \theta_{210}^*(Vn/300)$ for all $n \leq 300$, where $V = S^2/N$ is some parameter, and such that the linear functional $\mathcal{F}^*(P)$ defined on monomials as

$$\mathcal{F}^*(x^r) = \frac{1}{4(r+1)^2} c_r(10^5)$$

is minimized. Since the domain described by the inequalities relating P and θ_{210}^* is non-compact, we further restrict all coefficients to be at least $-M$. Then we compute an upper bound for $\mathcal{F}(P)$, the linear functional defined on monomials through

$$\mathcal{F}(x^r) = \frac{1}{4(r+1)^2} c_r,$$

by replacing c_r by $c_r(1 - (2r + 2)/10^5)$ whenever the coefficient of x^r in P is negative. Note that smaller values of M yield worse approximations to the optimal value of the functional $\mathcal{F}^*(P)$, but restricting the size of negative coefficients diminishes the upper bound for $\mathcal{F}(P)$.

Using $M = -1000$, $V = 1.5$ and looking for polynomials of degree 25, we obtain a polynomial P with $\mathcal{F}(P) \leq 0.54738$. The resulting polynomial is

$$\begin{aligned} P(u) = & 3.117973 + 3.555433 u - 154.037413 u^2 + 732.936467 u^3 - 1000 u^4 \\ & - 1000 u^5 + 3227.305717 u^6 - 1000 u^7 - 1000 u^8 - 1000 u^9 \\ & - 1000 u^{10} + 3012.745710 u^{11} + 1227.721539 u^{12} - 1000 u^{13} - 1000 u^{14} \\ & - 1000 u^{15} - 1000 u^{16} + 1191.986883 u^{17} + 2708.564854 u^{18} - 1000 u^{19} \\ & - 1000 u^{20} - 1000 u^{21} + 675.282733 u^{22} + 1158.017142 u^{23} - 1000 u^{24} \\ & + 214.336183 u^{25}. \end{aligned}$$

We determined this polynomial by using the `lpsolve` linear programming package and a C-program of our own. There has been numerous precision issues and instabilities that we were not able to understand, less to tackle to our satisfaction. For instance, many of the coefficients are on the artificial boundary $a_i \geq -1000$. We delay a further study to a latter paper.

Once the polynomial is selected, we can simply consider it and study anew how it fits $\theta_{\mathfrak{f}}^*$. To do so, we apply Pari/GP.

By construction of P we know that $P(u) \geq \theta_{\mathfrak{f}}^*(u)$ for 300 well spaced points; however, this does not imply that this inequality holds true for all values of u . In fact, there are six regions in which P dips slightly below θ_{210}^* , these regions being close to the points $u = 0.29, 0.59, 0.84, 1.16, 1.32$ and 1.44 . The difference is greatest at $u = 1.442618\dots$, where $P(u) - \theta_{210}^*(u) = -0.008338\dots$. Hence, putting $P^*(u) = P(u) + 0.0084$, we obtain

a polynomial which is strictly larger than θ_{210}^* , and we have

$$\mathcal{F}(P^*) = \mathcal{F}(P) + \frac{0.0084}{4} \leq 0.5474 + 0.0021 = 0.5495,$$

which implies that

$$N/Z - \frac{1}{2} \text{Log } N \geq \frac{1}{2} \text{Log } 1.5 + \kappa(210) + 1 - \mathcal{F}(P^*) \geq 1.7686$$

and therefore

$$\pi(M + N) - \pi(M) \leq \frac{2N}{\text{Log } N + 3.5372}$$

as announced.

References

- [1] E. Bombieri, *A note on the large sieve*, Acta Arith. 18 (1971), 401–404.
- [2] —, *Le grand crible dans la théorie analytique des nombres*, Astérisque 18 (1987).
- [3] P. D. T. A. Elliott, *On inequalities of large sieve type*, Acta Arith. 18 (1971), 405–422.
- [4] Yu. V. Linnik, *The Dispersion Method in Binary Additive Problems*, Leningrad. Univ., 1961 (in Russian); English transl.: Amer. Math. Soc., 1963.
- [5] J. E. van Lint and H. E. Richert, *On primes in arithmetic progressions*, Acta Arith. 11 (1965), 209–216.
- [6] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika 20 (1973), 119–133.
- [7] A. Selberg, *On elementary problems in prime number-theory and their limitations*, in: C.R. Onzième Congrès Math. Scandinaves, Trondheim, Johan Grundt Tanums Forlag, 1949, 13–22.
- [8] —, *Remarks on sieves*, in: Collected Papers, 205–216.
- [9] —, *Collected Papers*, Springer, 1991.

Laboratoire Paul Painlevé
 Université Lille 1
 59655 Villeneuve d'Ascq Cedex, France
 E-mail: ramare@math.univ-lille1.fr

Mathematisches Institut Freiburg
 Freiburg, Germany
 E-mail: jcp@mathematik.uni-freiburg.de

*Received on 23.5.2007
 and in revised form on 9.1.2008*

(5447)