Abstract

We prove that the number of primes in an interval of length N is at most $2N/(\operatorname{Log} N + 3.53)$ when N is large enough. This is obtained 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”.

Improving on the Brun-Titchmarsh Theorem[*] [†]

Oliver Ramaré
Jan Christoph Schlage-Puchta

March 25, 2008

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*

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

Such theorems have been termed "Brun-Titchmarsh" Theorem by Linnik in [4]. Indeed, Titchmarsh proved such a theorem for $q = 1$, with a $\operatorname{Log}\operatorname{Log}(N/q)$ term instead of the 2 above, to establish the asymptotic for the number of divisors of the $p+1$, 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 this work, Selberg also shows that the constant $2 + o(1)$ is optimal in the above, if we are to 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 $2N/(\operatorname{Log} N + c)$ with an unspecified but very negative c is due to van Lint & Richert in [5] though [7] mentions without any proof such a result around equation (6) therein. Bombieri gave in [1] the value $c = -3$ and Montgomery & Vaughan the value $c = 5/6$ in [6]. The section 22 of "lectures on sieves" [9] gives a proof of $c = 2.81$, a proof from which we shall take several elements.

[*]AMS classification: 11N13, 11N35, 11N36
[†]Keywords: Brun-Titchmarsh Theorem

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 are as large as N), we can handle the remainder in a more detailled way. One can see this process as a weighted large sieve inequality, with weights adapted to the problem. We note that we shall in between be confronted with the problem of majorizing a step function by a polynomial, and that this problem appears to be numerically more difficult than meet 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 our attention to our own method!

As a matter of notations, we denote by $\sigma(d)$ the sum of the (positive) divisors of d while, 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 datas*

1. *a finite family $(\varphi_i^*)_{i\in I}$ of points*[1],
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 $\omega_{j,i} = \overline{\omega_{i,j}}$,*

all of them given so that

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

We are to comment on this definition. If the family $(\varphi_i^*)_{i\in I}$ were orthogonal, we could ask for equality with $M_i = \|\varphi_i^*\|^2$. As it turns out, 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

[1] The reader may wonder why we chose to call the members of this family with a star. It is to be consistent and to avoid confusion with notations that will appear later on.

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

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

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

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 are to lie in the so-called *Gershgorin's disc* centered on 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 $f \in \mathcal{H}$, let us set $\xi_i = [f|\varphi_i^*]/M_i$. We have*

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

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: For the proof, simply write

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

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

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

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. $\diamond\diamond\diamond$

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

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

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

$$\|f\|^2 - 2\Re \sum_i \overline{\xi_i}[f|\varphi_i^*] + \sum_i M_i|\xi_i|^2 + \Big(\sum_i |\xi_i| n_i\Big)^2 + \sum_{i,j} \xi_i\overline{\xi_j}\omega_{i,j} \geq 0. \quad (3)$$

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 these ξ_i's so as to minimize

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

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

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

provided

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

However, determining 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

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

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 be a non-negative real number and $(n_i)_i$ be non-negative real numbers. Assume that $[f|\varphi_i^*]$'s are real numbers. Set $\xi_i = ([f|\varphi_i^*] - Yn_i)/M_i$ for all i. Then we have*

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

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 $\mathfrak{f}$ be a positive integer and let us define by $\rho = \phi(\mathfrak{f})/\mathfrak{f}$. We study here the following two functions of the real non-negative variable u:

$$\begin{cases} \theta_{\mathfrak{f}}^-(u) &= \min\limits_{y \in \mathbb{R}} \min\limits_{\substack{0 \leq x \leq u \\ x \in \mathbb{R}}} \Bigg(\sum\limits_{\substack{y < n \leq y + x, \\ (n,\mathfrak{f})=1}} 1 - \rho x \Bigg), \\ \theta_{\mathfrak{f}}^+(u) &= \max\limits_{y \in \mathbb{R}} \max\limits_{\substack{0 \leq x \leq u \\ x \in \mathbb{R}}} \Bigg(\sum\limits_{\substack{y < n \leq y + x, \\ (n,\mathfrak{f})=1}} 1 - \rho x \Bigg). \end{cases}$$

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 next lemma.

Lemma 3.1 *We have*

$$\begin{cases} \theta_{\mathfrak{f}}^-(u) &= \min\limits_{\ell \in \mathbb{N}} \left(\min\limits_{\substack{k \in \mathbb{N}, \\ 0 \leq k \leq u}} \Bigg(\sum\limits_{\substack{\ell+1 \leq n \leq \ell+k-1, \\ (n,\mathfrak{f})=1}} 1 - \rho k \Bigg), \sum\limits_{\substack{\ell+1 \leq n \leq \ell+[u], \\ (n,\mathfrak{f})=1}} 1 - \rho u \right), \\ \theta_{\mathfrak{f}}^+(u) &= \max\limits_{\substack{k,\ell \in \mathbb{N}, \\ k < u+1}} \Bigg(\sum\limits_{\substack{\ell \leq n \leq \ell+k-1, \\ (n,\mathfrak{f})=1}} 1 - \rho(k-1) \Bigg) \end{cases}$$

The function $\theta^+_{\mathfrak{f}}$ is a non-decreasing step function which is left continuous with jump points at integer points. The function $\theta^-_{\mathfrak{f}}$ is non-increasing continuous : it alternates linear pieces of directing coefficient $-\rho$ and constant pieces. Changes occur at integer points. Both are constant if $u \geq \mathfrak{f}$.

PROOF: We start with $\theta^+_{\mathfrak{f}}$. 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

$$\theta^+_{\mathfrak{f}}(u) = \max_{y\in\mathbb{R}} \max_{\substack{k\in\mathbb{N},\\ k\leq u+\{y\}}} \biggl(\sum_{y<n\leq [y]+k} w(n) + \rho(-k+\{y\})\biggr).$$

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

As for $\theta^-_{\mathfrak{f}}$, we start similarly by fixing y. 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 get that $\theta^-_{\mathfrak{f}}(u)$ equals

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

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

$$\min_{\ell\in\mathbb{N}}\biggl(\sum_{\ell+1\leq n\leq \ell+u} w(n) - \rho u\biggr). \tag{7}$$

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 $y = \ell+1-\{u\}$. This last contribution turns out to be exactly the same as the one in (7). ◇◇◇

Next we consider the function

$$\theta^*_{\mathfrak{f}}(v) = \max(\theta^+_{\mathfrak{f}}(1/v), -\theta^-_{\mathfrak{f}}(1/v)) \tag{8}$$

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

Case of $\mathfrak{f} = 210$

Here is our function:

$$\theta^*_{210}(1/u) = \begin{cases} 1 & \text{if } 0 < u \le 1 \\ 54/35 & \text{if } 1 < u \le 3 \\ 57/35 & \text{if } 3 < u \le 7 \\ 76/35 & \text{if } 7 < u \le 9 \\ 79/35 & \text{if } 9 < u \le 79/8 \end{cases} \quad \begin{cases} 8u/35 & \text{if } 79/8 \le u \le 10 \\ 16/7 & \text{if } 10 < u \le 13 \\ 82/35 & \text{if } 13 < u \le 17 \\ 94/35 & \text{if } 17 < u \le 41/2 \\ 8u/35 - 2 & \text{if } 41/2 \le u \le 22 \\ 106/35 & \text{if } 22 < u \le 210 \end{cases}$$

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

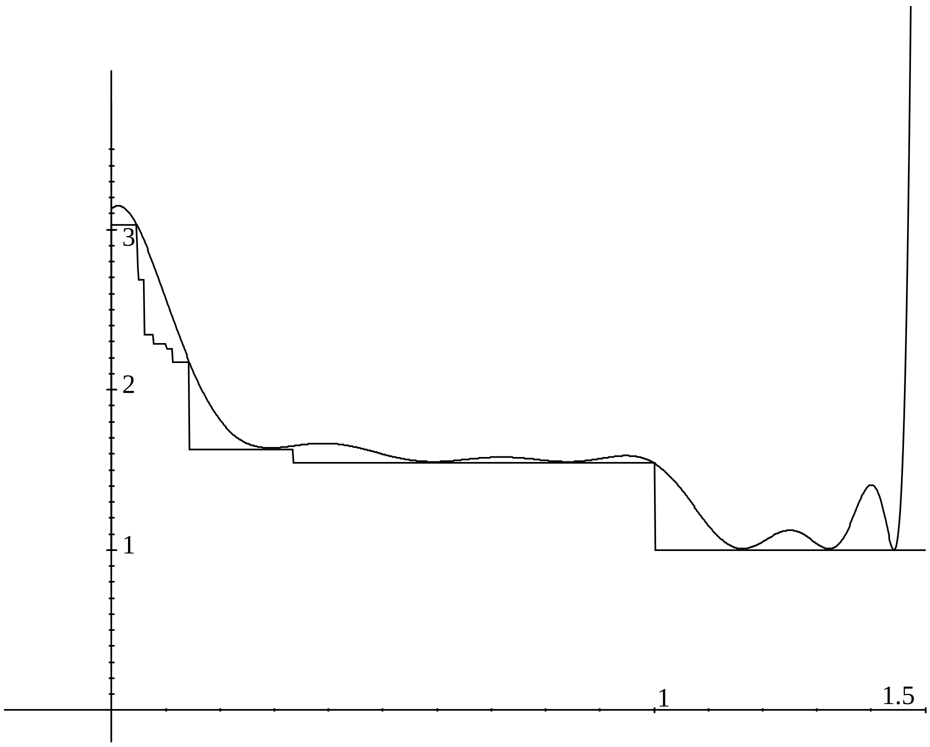

Figure 1: Comparison of θ^*_{210} and the optimizing polynomial

Polynomial approximation of $\theta^*_{\mathfrak{f}}(v)$

We shall require a good polynomial upper bound for $\theta *_{\mathfrak{f}}$:

$$\theta^*_{\mathfrak{f}}(v) \le \sum_{0 \le r \le R} b_r v^r. \tag{9}$$

Finding such an approximation turned out to be much more tricky than expected. Our first idea has been to start with a polynomial approximation of $\theta^*_{\mathfrak{f}}(v)$ of the form

$$\Big|\theta^*_{\mathfrak{f}}(v) - \sum_{0\le r\le R} \tilde{b}_r v^r\Big| \le \epsilon$$

for $0 \le v \le V$ 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 by much, despite the fact that we used polynomials of very high degree (up to 200). We finally decided for a different scheme exposed 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, it simply is 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 $d \le \sqrt{N}$. 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 is patterned for almost orthogonal φ_q's, which is not the case of the sequence $(\varphi_d)_d$: if $q|d$, knowing that a given integer is coprime with d implies it is coprime with q, so there is redundancy of information. It 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

$$\frac{d}{\phi(d)}\varphi_d = \sum_{q|d} \varphi^*_q \tag{10}$$

where

$$\varphi^*_q(n) = \mu(q)c_q(n)/\phi(q) \tag{11}$$

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

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

Verifying (10) is easy:

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

In our problem, we shall select a fixed integer $\mathfrak{f}$ 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 with $\mathfrak{f}$. This being chosen, our hermitian product on sequences over $[M+1, M+N]$ is defined by

$$[g|h] = \sum_{M+1\le n\le M+N} w(n)g(n)\overline{h(n)}. \tag{13}$$

Furthermore, we take the moduli q in the set

$$\{q \ / \ \sigma(q) \le S, \mu^2(q) = 1, (q, \mathfrak{f}) = 1\}, \tag{14}$$

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

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

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 with $\mathfrak{f}$. Write $(q, q', \Delta) = \delta$, so that $[\varphi_q^*|\varphi_{q'}^*]$ equals

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

where

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

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

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

At this level, we say that

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

by (9). We infer that

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

This leads to

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

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 $\mathfrak{f}\Delta$. 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

$$\sum_{\substack{(\ell,\mathfrak{f}\Delta)=1\\(\ell',\mathfrak{f}\Delta)=1}} |\xi_{\delta_1\delta_2\ell}|\eta_r(\ell)|\xi_{\delta_1\delta_3\ell'}|\eta_r(\ell')\biggl|\prod_{p|(\ell,\ell')}\frac{1+2p^{r+1}+p^{r+2}}{(1+p^{r+1})^2}-1\biggr|$$
$$\ll_r \sum_{p\geq P}\sum_{\substack{(m,p\mathfrak{f}\Delta)=1,\\(m',p\mathfrak{f}\Delta)=1}} |\xi_{\delta_1\delta_2pm}|\eta_r(pm)|\xi_{\delta_1\delta_3pm'}|\eta_r(pm')$$
$$\ll_r \sum_{p\geq P} p^{2r}\sum_{m,m'}|\xi_{\delta_1\delta_2pm}|\eta_r(m)|\xi_{\delta_1\delta_3pm'}|\eta_r(m')$$

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) and using Lemma 5.1, we get the above to be not more than

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

This will give rise to the error term

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

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

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

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

$$\Bigl\|\sum_q \xi_q\varphi_q^*\Bigr\|^2 \leq \rho N\sum_q |\xi_q|^2/\phi(q)$$
$$+\sum_{0\leq r\leq R}\frac{b_r}{N^r}\sum_{\delta_1\delta_2\delta_3|\Delta}\eta_r^\flat(\delta_1)\sum_{\substack{(\ell,\mathfrak{f}\Delta)=1,\\(\ell',\mathfrak{f}\Delta)=1}}|\xi_{\delta_1\delta_2\ell}|\eta_r(\delta_2\ell)|\xi_{\delta_1\delta_3\ell'}|\eta_r(\delta_3\ell')$$
$$+\mathcal{O}\biggl(\left(\frac{Z}{\rho N}\right)^2\prod_{p|\Delta}(1+p^{-1})^2\sum_{1\leq r\leq R}\frac{S^{2r+2}|b_r|}{N^rP}\biggr).$$

5 Some arithmetical auxiliaries

We need to evaluate some rather unusual averages.

Lemma 5.1 *Let $\mathfrak{f}^*$ be a positive integer. We set $\rho^* = \phi(\mathfrak{f}^*)/\mathfrak{f}^*$ and use $t(q) = 1 - \sigma(q)/S^*$. For any real number S^* going to infinity, we have*

$$\sum_{\substack{q/\sigma(q)\le S^*,\\ (q,\mathfrak{f}^*)=1}} \frac{t(q)^2}{\phi(q)} = \rho^*(\operatorname{Log} S^* + \kappa(\mathfrak{f}^*)) + o(1)$$

with

$$\kappa(\mathfrak{f}) = \gamma + \sum_{p\ge 2} \frac{\operatorname{Log} p}{p(p-1)} - \sum_{p} \frac{\operatorname{Log}(1+p^{-1})}{p} + \sum_{p|\mathfrak{f}^*} \frac{\operatorname{Log}(p+1)}{p} - \tfrac{3}{2}$$

($\kappa(210) = 1.115\,37\dots$) and

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

Proof: The first estimate comes from [9]. We follow closely Selberg's proof and get

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

from which we get

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

◇ ◇ ◇

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.

Let us define

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

We have

Lemma 5.2

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

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

$$\prod_{p|\Delta} \frac{\Big((p+1)^{r+1}(p-1)+p^{r+1}+1\Big)(p-1)}{p^2(p+1)^{r+1}}$$
$$\times \sum_{\delta_1\delta_2|\Delta} \frac{\prod_{p|\delta_1}\big(1+p^{r+1}(p-2)\big)\eta_r(\delta_2)}{\phi(\delta_1)^2\sigma(\delta_1)^{2r+2}\sigma(\delta_2)^{r+1}} \prod_{p|\delta_1\delta_2} \frac{(p+1)^{r+1}(p-1)}{(p+1)^{r+1}(p-1)+p^{r+1}+1}.$$

We continue with δ_2:

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

Hence our quantity reduces to

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

which reads

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

◇ ◇ ◇

6 Using the hermitian inequalilty

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

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

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 $\sum_{(\ell,\mathfrak{f}\Delta)=1}|\xi_{\delta_1\delta_2\ell}|\eta_r(\ell)$. 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

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

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

$$\begin{cases} A = \displaystyle\sum_{0\leq r\leq R}\frac{Z^2 b_r}{\rho^2 N^{r+2}}\sum_{\delta_1\delta_2\delta_3|\Delta} g(\delta_1)\sum_{\substack{(\ell,\mathfrak{f}\Delta)=1\\(\ell',\mathfrak{f}\Delta)=1}} t(\delta_1\delta_2\ell)\eta_r(\ell\delta_2)t(\delta_1\delta_3\ell')\eta_r(\ell'\delta_3),\\ B = \displaystyle\sum_{\delta_1\delta_2\delta_3|\Delta} g(\delta_1)\sum_{1\leq r\leq R}\sum_{p\geq P}\frac{|b_r|p^{2r}}{N^r}\sum_{\ell,\ell'}|\xi_{\delta_1\delta_2 p\ell}|\eta_r(\ell\delta_2)|\xi_{\delta_1\delta_3 p\ell'}|\eta_r(\ell'\delta_3).\end{cases}$$

We tidy this expression step by step:

$$N \geq Z\left(\operatorname{Log} S+\kappa(\mathfrak{f})\right)+Z$$
$$-Z\sum_{0\leq r\leq R}\frac{b_r S^{2r+2}}{\rho^2 N^{r+1}}\sum_{\delta_1\delta_2\delta_3|\Delta}\frac{\prod_{p|\delta_1}\big(1+p^{r+1}(p-2)\big)\eta_r(\delta_2\delta_3)}{\phi(\delta_1)^2\sigma(\delta_1)^{2r+2}\sigma(\delta_2)^{r+1}\sigma(\delta_3)^{r+1}}\frac{\rho^2\phi(\Delta)^2\mathfrak{S}_r(\mathfrak{f}\Delta)^2}{4\Delta^2(r+1)^2}$$
$$+\mathcal{O}\Big(\prod_{p|\Delta}(1+p^{-1})^2S^2ZN^{-1}P^{-1}\sum_{1\leq r\leq R}|b_r|(S^2/N)^r\Big)+o(N).$$

This leads to

$$N/Z \geq \operatorname{Log} S+\kappa(\mathfrak{f})+1-\sum_{0\leq r\leq R}\frac{b_r(S^2/N)^{r+1}}{4(r+1)^2}C_r(\Delta)\mathfrak{S}_r(\mathfrak{f}\Delta)^2$$
$$+\mathcal{O}\Big(\prod_{p|\Delta}(1+p^{-1})^2P^{-1}\sum_{1\leq r\leq R}|b_r|(S^2/N)^{r+1}\Big)+o(1)$$

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

$$N/Z-\tfrac{1}{2}\operatorname{Log} N \geq \tfrac{1}{2}\operatorname{Log}(S^2/N)+\kappa(\mathfrak{f})+1-\sum_{0\leq r\leq R}\frac{b_r(S^2/N)^{r+1}}{4(r+1)^2}C_r(\infty/\mathfrak{f})$$
$$+\mathcal{O}\Big(\prod_{p|\Delta}(1+p^{-1})^2P^{-1}\sum_{1\leq r\leq R}|b_r|(S^2/N)^{r+1}\Big)+o(1).$$

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

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

together with

$$c_r = C_r(\infty/\mathfrak{f}) = \lim_{x\to\infty} c_r(x). \tag{21}$$

Lemma 7.1 *We have for $x > 2r+2$ the estimate*

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

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

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

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 we have for $p \geq r$ 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 monotonically 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

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

◇◇◇

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 $M = -1\,000$, $V = 1.5$ and looking for polynomials of degree 25, we obtain a polynomial P with $\mathcal{F}(P) \leq 0.547\,38$. The resulting polynomial is

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

We determined this polynomial by using the lpsolve linear programming package and a C-program of our own. There has been numerous precisions issues and instabilities that we were no able to understand, less to tackle to our satisfaction. For instance, many of the coefficients are on the artificial boundary $a_i \geq -1\,000$. 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 revert to 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 $u = 1.442\,618\ldots$, where $P(u) - \theta^*_{210}(u) = -0.008\,338\ldots$. Hence, putting $P^*(u) = P(u) + 0.0084$, we obtain a polynomial which is strictly larger than θ^*_{210}, and we have

$$\mathcal{F}(P^*) = \mathcal{F}(P) + \frac{0.0084}{4} \leq 0.547\,4 + 0.002\,1 = 0.549\,5,$$

which implies that

$$N/Z - \frac{1}{2}\operatorname{Log} N \geq \tfrac{1}{2}\operatorname{Log} 1.5 + \kappa(210) + 1 - \mathcal{F}(P^*) \geq 1.768\,6$$

and therefore

$$\pi(M+N)-\pi(M) \leq \frac{2N}{\operatorname{Log} N + 3.537\,2}$$

as announced.

References

[1] E. Bombieri. A note on the large sieve method. *Acta Arith.*, 18:401–404, 1971.

[2] E. Bombieri. Le grand crible dans la théorie analytique des nombres. *Astérisque*, 18:103pp, 1987.

[3] P.D.T.A. Elliott. Some remarks concerning the large sieve. *Acta Arith.*, 18:405–422, 1971.

[4] Yu.V. Linnik. The dispersion method in binary additive problems. *Leningrad*, page 208pp, 1961.

[5] J.E. van Lint and H.E. Richert. On primes in arithmetic progressions. *Acta Arith.*, 11:209–216, 1965.

[6] H.L. Montgomery and R.C. Vaughan. The large sieve. *Mathematika*, 20(2):119–133, 1973.

[7] A. Selberg. On elementary problems in prime number-theory and their limitations. *C.R. Onzième Congrès Math. Scandinaves, Trondheim, Johan Grundt Tanums Forlag*, pages 13–22, 1949.

[8] A. Selberg. Remarks on sieves. In *Collected Papers*, pages 205–216, University of Colorado, Boulder, Colorado, August 14–18 1972. Number Theory Conference.

[9] A. Selberg. Collected Papers. *Springer-Verlag*, II:251pp, 1991.

Olivier Ramaré
Laboratoire Paul Painlevé
Université Lille 1
59 655 Villeneuve d'Ascq, cedex
France
`ramare@math.univ-lille1.fr`

Jan-Christoph Schlage-Puchta
Mathematisches Institut Freiburg
Freiburg
Germany