Notes on the large sieve

G.J.O. Jameson

Contents

Introduction

The "large sieve", in its arithmetic form, was originated by Linnik [Li] in 1941. It was developed and applied in a long series of papers by Rényi (1947–1959), e.g. [Ren].

Papers by Roth [Ro] and Bombieri [Bom] paved the way for the recognition that these results rested on an underlying analytic inequality on sums of squares of values of a trigonometric polynomial. This was first formulated explicitly by Davenport and Halberstam [DH] in 1966, and has come to be known as the "analytic large sieve inequality". Later writers have refined this result by establishing the best constant.

In section LS1, we give Gallagher's quick proof of the analytic result (without the best constant) as well as the proof by Montgomery and Vaughan of the optimal result. The modern arithmetic version, updating the results of Linnik and Rényi, is derived in section LS2: it is only here that the reason for the term "sieve" becomes apparent. It seems worth formulating some general lemmas for the cases where only one or two congruence classes are removed, since this is the case in most applications.

In section LS3, we present a number of applications that can be derived easily, aiming for simplicity rather than best constants (which in some cases require distinctly harder methods). The examples are well known, but I have not found a compilation of quite this sort in the sources that I have consulted. For readers familiar with other sieve methods, I hope to demonstrate that the large sieve delivers these applications at least as easily.

Prerequisites are undergraduate level real analysis and linear algebra, and some number theory (generally at what might be judged undergraduate level, but varying with the applications).

Only sketchy historical references and attributions are given. For more thorough historical information, and a very extensive list of references, see [Mont1].

LS1. The analytic large sieve inequality

Statement of the basic theorem

Following the custom of analytic number theory, we use the notation $e(t)$ for $e^{2\pi it}$. Note that $e(t)e(u) = e(t+u)$ and $e(n) = 1$ for integers n.

Consider a trigonometric polynomial

$$f(t) = \sum_{n=1}^{N} x_n e(nt).$$

We shall be concerned with inequalities of the form

$$\sum_{r=1}^{R} |f(t_r)|^2 \leq C \sum_{n=1}^{N} |x_n|^2. \tag{1}$$

Since f has period 1, we may as well assume that the points t_r are all in the interval $(0, 1]$, and that they are listed in increasing order: $0 < t_1 \leq t_2 \leq \ldots \leq t_R \leq 1$.

First, we note the obvious: $|f(t)| \leq \sum_{n=1}^{N} |x_n|$, hence, by the Cauchy-Schwarz inequality, $|f(t)|^2 \leq N \sum_{n=1}^{N} |x_n|^2$, so $\sum_{r=1}^{R} |f(t_r)|^2 \leq RN \sum_{n=1}^{N} |x_n|^2$. Our aim is to show that a much stronger inequality holds if we prevent reinforcement of the values $|f(t_r)|^2$ by not allowing the points t_r to coincide, or to differ by an integer. For real t, we denote by $\Delta(t)$ the distance to the nearest integer:

$$\Delta(t) = \inf\{|t - n| : n \in \mathbb{Z}\}.$$

Note that $\Delta(t) \leq \frac{1}{2}$. We require the points t_r to be "well separated" in the sense that $\Delta(t_r - t_s) \geq \delta$ for $r \neq s$. We may assume that $t_1 < t_2 < \ldots < t_R < t_1 + 1$. This divides $[t_1, t_1 + 1]$ into R subintervals, each of length at least δ, so $\delta \leq 1/R$ (with equality occurring when the points are equally spaced).

Two further remarks are useful at this point. First, the problem is equivalent to determining the norm of a certain matrix. The (operator) norm of a matrix A is $\|A\| = \sup\{\|Ax\| : \|x\| = 1\}$, where $\|x\|$ denotes the (Euclidean) norm of the vector $x = (x_n)$, given by $\|x\|^2 = \sum_{n=1}^{N} |x_n|^2$. (Unfortunately, the notation $\|t\|$ is often used for our $\Delta(t)$, a culture clash!). Since

$$f(t_r) = \sum_{n=1}^{N} e(nt_r) x_n,$$

the constant C in (1) is $\|V\|^2$, where V is the matrix $[e(nt_r)]$ $(1 \leq n \leq N,\ 1 \leq r \leq R)$.

Secondly, the modulus of $f(t)$ is unchanged if it is multiplied by $e(kt)$, putting it into the form $\sum_{n=k+1}^{k+N} y_n e(nt)$. So the problem is the same if the range of values of n is translated by k: the role of N is only to state the length of this range.

We are now ready to state the basic theorem. It is known, for reasons yet to be revealed, as the "analytic large sieve inequality".

Theorem LS1.1 *Let $f(t) = \sum_{n=1}^{N} x_n e(nt)$, and let t_r ($1 \le r \le R$, where $R \ge 2$) be points such that $\Delta(t_r - t_s) \ge \delta$ for $r \ne s$. Then*

$$\sum_{r=1}^{R} |f(t_r)|^2 \le \left(N - 1 + \frac{1}{\delta}\right) \sum_{n=1}^{N} |x_n|^2.$$

The same applies if the range of n is $[M+1, M+N]$ for any M.

A simple example shows that there is a fairly wide range of cases in which the stated constant $N - 1 + 1/\delta$ is exact.

Example. Let $R \ge 2$ and $t_r = r/R$ for $1 \le r \le R$, so that $\delta = 1/R$. Let $f(t) = e(0) + e(Rt)$, so that $N = R + 1$ and $\sum_{n=0}^{R} |x_n|^2 = 2$. Then $f(t_r) = 2$ for each r, so $\sum_{r=0}^{R} f(t_r)^2 = 4R$. The ratio is $2R$, equal to $N - 1 + 1/\delta$.

Theorem LS1.1 was first formulated in this style, but with a weaker constant, by Davenport and Halberstam [DH] in 1966. The constant was improved to $N+1/\delta$ by Montgomery and Vaughan [MV1] in 1973. The -1 was contributed by Selberg (unpublished), using a different method, but as we show below, it can be established by a minor addition to the proof of Montgomery and Vaughan.

We now give (a) a short proof of the theorem with a weaker constant, (b) a longer proof, giving the constant stated. Readers who are more interested in applications are at liberty to skip either proof (or even both) and proceed to section LS2.

Gallagher's proof

Here we present Gallagher's proof of Theorem LS1.1 [Gall], which is quick and elegant, at the cost of not delivering the best constant.

Note that for any c, $\int_c^{1+c} e(nt)\, dt = 0$ for integers $n \ne 0$. Since

$$|f(t)|^2 = \sum_{n=1}^{N} |x_n|^2 + \sum_{n=1}^{N} \sum_{m \ne n} x_m \overline{x_n} e[(m-n)t],$$

it follows that

$$\int_c^{1+c} |f(t)|^2 dt = \sum_{n=1}^{N} |x_n|^2.$$

We now prove a lemma showing how to convert estimates for integrals into estimates of functional values.

Lemma LS1.2. *Let g be a differentiable function (real or complex) on $[a-h, a+h]$. Then*

$$|g(a)| \le \frac{1}{2h} \int_{a-h}^{a+h} |g(t)|\, dt + \frac{1}{2} \int_{a-h}^{a+h} |g'(t)|\, dt.$$

Proof. Let

$$\rho(t) = \begin{cases} t-a+h & \text{for } a-h < t < a, \\ t-a-h & \text{for } a < t < a+h. \end{cases}$$

Integration by parts on the intervals $[a-h, a]$ and $[a, a+h]$ leads easily to

$$\int_{a-h}^{a+h} \rho(t) g'(t)\, dt = 2hg(a) - \int_{a-h}^{a+h} g(t)\, dt.$$

Since $|\rho(t)| \le h$, the statement follows. □

Proof of LS1.1 with constant $\pi N + 1/\delta$. As mentioned above, we can translate the range of n. In particular, we can move it to an interval J contained in $[-\frac{1}{2}N, \frac{1}{2}N]$, now taking $f(t)$ to be $\sum_{n\in J} x_n e(nt)$, with J as stated. Also, we may assume that $t_1 < t_2 < \cdots < t_R$, with $t_r - t_{r-1} \ge \delta$, and $t_1 + 1 \ge t_R + \delta$, so that if $t_1 - \frac{1}{2}\delta = c$, then $t_R + \frac{1}{2}\delta \le 1 + c$. So the intervals $[t_r - \frac{1}{2}\delta, t_r + \frac{1}{2}\delta]$ do not overlap, and are contained in $[c, 1+c]$. By the Lemma, applied to $f(t)^2$ on each of these intervals,

$$\sum_{r=1}^{R} |f(t_r)|^2 \le \frac{1}{\delta} \int_c^{1+c} |f(t)|^2\, dt + \int_c^{1+c} |f(t) f'(t)|\, dt.$$

Now $\int_c^{1+c} |f(t)|^2 dt = \sum_{n\in J} |x_n|^2$. Also, since $f'(t) = 2\pi i \sum_{n\in J} n x_n e(nt)$ and $|n| \le N/2$ for $n \in J$, we have

$$\int_c^{1+c} |f'(t)|^2\, dt = 4\pi^2 \sum_{n\in J} n^2 |x_n|^2 \le \pi^2 N^2 \sum_{n\in J} |x_n|^2.$$

By the Cauchy-Schwarz inequality for integrals,

$$\int_c^{1+c} |f(t) f'(t)|\, dt \le \pi N \sum_{n\in J} |x_n|^2.$$

The statement follows. □

The proof of Montgomery and Vaughan

For many purposes, Gallagher's result is quite good enough. However, it is not optimal. We now outline the method of Montgomery and Vaughan [MV1] which does lead to the optimal result. Their proof uses the following generalization of Hilbert's inequality:

Proposition LS1.3. *Let λ_j $(1 \le j \le n)$ be real numbers such that $\Delta(\lambda_j - \lambda_k) \ge \delta$ whenever $j \ne k$, and let H be the matrix defined by*

$$h_{j,k} = \begin{cases} \operatorname{cosec}[\pi(\lambda_j - \lambda_k)] & \text{if } j \ne k, \\ 0 & \text{if } j = k. \end{cases}$$

Then $\|H\| \le 1/\delta$.

The derivation of Theorem LS1.1 from this result is quite straightforward. We present it next, and then return (for those who wish) to the proof of Proposition LS1.3.

Recall first that for a matrix A, with adjoint A^* and transpose A^T, we have $\|A^*\| = \|A^T\| = \|A\|$, since A^* is the complex conjugate of A^T and (in the notation of Hilbert space theory)

$$\|A\| = \sup\{|\langle Ax, y\rangle| : \|x\| = \|y\| = 1\} = \sup\{|\langle x, A^*y\rangle| : \|x\| = \|y\| = 1\} = \|A^*\|.$$

Proof of Theorem LS1.1. We have seen that $C = \|V\|^2$, where V is the matrix defined by $v_{r,n} = e(nt_r)$ for $1 \le n \le N$, $1 \le r \le R$. We shall evaluate the norm of the transposed matrix. Given scalars y_r, let

$$T(y) = \sum_{n=1}^{N} \left| \sum_{r=1}^{R} e(nt_r) y_r \right|^2 .$$

Then $\|V\|^2$ is the least constant C for which we have $T(y) \le C \sum_{r=1}^{R} |y_r|^2$ for all choices of y_r. Note first that, by the geometric series,

$$\sum_{n=1}^{N} e(nt) = e(t)\frac{e(Nt) - 1}{e(t) - 1} = \frac{e[(N + \frac{1}{2})t] - e(\frac{1}{2}t)}{2i \sin \pi t}.$$

Hence we have

$$\begin{aligned} T(y) &= \sum_{n=1}^{N} \sum_{r=1}^{R} \sum_{s=1}^{R} y_r \overline{y}_s e[n(t_r - t_s)] \\ &= N \sum_{r=1}^{R} |y_r|^2 + \sum_{r \ne s} y_r \overline{y}_s \sum_{n=1}^{N} e[n(t_r - t_s)] \\ &= N \sum_{r=1}^{R} |y_r|^2 + \sum_{r \ne s} y_r \overline{y}_s \frac{e[(N + \frac{1}{2})(t_r - t_s)] - e[\frac{1}{2}(t_r - t_s)]}{2i \sin \pi(t_r - t_s)}, \end{aligned}$$

in which $\sum_{r\neq s}$ means summation over all pairs (r,s) with $r \neq s$. Now for any a,

$$\sum_{r\neq s} y_r \overline{y}_s \frac{e[a(t_r - t_s)]}{2\sin \pi(t_r - t_s)} = \sum_{r\neq s} \frac{z_r \overline{z}_s}{2\sin \pi(t_r - t_s)},$$

where $z_r = y_r e(at_r)$. Since $|z_r| = |y_r|$, Proposition LS1.3 shows that the modulus of this expression is not greater than $(1/2\delta)\sum_{r=1}^{R} |y_r|^2$. Apply this with $a = N + \frac{1}{2}$ and $a = \frac{1}{2}$ to obtain

$$|T(y)| \leq \left(N + \frac{1}{\delta}\right) \sum_{r=1}^{R} |y_r|^2.$$

This completes the proof, apart from showing that N can be replaced by $N - 1$. (This refinement is completely unimportant in all the applications that follow!) The following neat proof is due to Paul Cohen. Choose $K > 1$, and let

$$g(t) = f(Kt) = \sum_{n=1}^{N} x_n e(nKt),$$

which we can write as $\sum_{k=K}^{KN} y_k e(kt)$, where $\sum_{k=K}^{KN} |y_k|^2 = \sum_{n=1}^{N} |x_j|^2$. Since f has period 1,

$$K\sum_{r=1}^{R} |f(t_r)|^2 = \sum_{k=1}^{K}\sum_{r=1}^{R} |f(t_r + k)|^2 = \sum_{k=1}^{K}\sum_{r=1}^{R} \left|g\left(\frac{t_r + k}{K}\right)\right|^2.$$

The numbers $(t_r + k)/K$ are separated by δ/K, so, by the result already proved,

$$K\sum_{r=1}^{R} |f(t_r)|^2 \leq \left((KN - K + 1) + \frac{K}{\delta}\right) \sum_{n=1}^{N} |x_n|^2.$$

Now divide by K and let K tend to infinity to obtain the result. □

We return to the proof of Proposition LS1.3. First, some further general facts about norms of matrices. It is well-known that if $A = (a_{j,k})$ is symmetric or skew-symmetric, then *quadratic* forms are sufficient to determine the norm: $\|A\| = \sup\{|S(x)| : \|x\| = 1\}$, where $S(x) = \sum_{j=1}^{n}\sum_{k=1}^{n} a_{j,k} x_j \overline{x}_k$. (Complex x_j must be allowed; if A is skew-symmetric, then $S(x)$ is of the form iR, where R is real.)

We need the following Lemma, due to Schur, on matrices with non-negative entries.

Lemma LS1.4. *Let $A = (a_{j,k})$ be a matrix (finite or infinite) such that $a_{j,k} \geq 0$ for all j, k and*

$$\sum_k a_{j,k} \leq K_1 \quad \textit{for all } j \qquad \textit{(all row sums} \leq K_1\textit{)},$$

$$\sum_j a_{j,k} \leq K_2 \quad \textit{for all } k \qquad \textit{(all column sums} \leq K_2\textit{)}.$$

Then $\|A\| \le (K_1 K_2)^{1/2}$.

Proof. We use bilinear forms to establish the norm. Since the $a_{j,k}$ are non-negative, it is sufficient to consider non-negative, real vectors $x = (x_j)$ and $y = (y_k)$ (in ℓ_2 in the infinite case). In the following, the sums can be either finite or infinite. Let

$$u(x,y) = \sum_j \sum_k a_{j,k} x_j y_k = \sum_j \sum_k (a_{j,k}^{1/2} x_j)(a_{j,k}^{1/2} y_k).$$

By the Cauchy-Schwarz inequality (applied to the double sum), $u(x,y) \le (CD)^{1/2}$, where

$$C = \sum_j \sum_k a_{j,k} x_j^2 = \sum_j x_j^2 \sum_k a_{j,k} \le K_1 \sum_j x_j^2,$$

$$D = \sum_j \sum_k a_{j,k} y_k^2 = \sum_k y_k^2 \sum_j a_{j,k} \le K_2 \sum_k y_k^2.$$

So $u(x,y) \le (K_1 K_2)^{1/2} \|x\|.\|y\|$. □

Note. If A is symmetric, then the two hypotheses say the same (with $K_2 = K_1 = K$, say) and the conclusion is $\|A\| \le K$. An alternative proof for this case uses *quadratic* forms (sufficient since A is symmetric) and the inequality $\frac{1}{2} x_j x_k \le x_j^2 + x_k^2$.

We shall use the following special case:

Proposition LS1.5. *Let A_n be the matrix $(a_{j,k})$ $(1 \le j,k \le n)$, where*

$$a_{j,k} = \begin{cases} 1/(j-k)^2 & \text{for } j \ne k, \\ 0 & \text{for } j = k, \end{cases}$$

Then $\|A_n\| \le \pi^2/3$.

Proof. The matrix is symmetric, and $\sum_{k=1}^n a_{j,k} \le 2 \sum_{r=1}^\infty (1/r^2) = \pi^2/3$ for each j. □

We will deduce Proposition LS1.3 from another Hilbert-type inequality, also due to Montgomery and Vaughan:

Proposition LS1.6. *Suppose that λ_j $(1 \le j \le n)$ are real numbers such that $|\lambda_j - \lambda_k| \ge \delta$ for $j \ne k$. Let G_n be the matrix $(g_{j,k})$, where*

$$g_{j,k} = \begin{cases} 1/(\lambda_j - \lambda_k) & \text{if } j \ne k, \\ 0 & \text{if } j = k. \end{cases}$$

Then $\|G_n\| \le \pi/\delta$.

Here we give a relatively simple proof of this theorem which may not be well known. It is in the spirit of Hilbert's original proof of his inequality. Actually, it only delivers a slightly weaker result, with an intervening constant $2/\sqrt{3}$.

Proof of Proposition LS1.6 with an extra constant. We may assume that $\lambda_1 < \lambda_2 < \cdots < \lambda_n$ (this amounts to re-ordering the variables). Also, after multiplying by a constant, we may assume that $\delta = 1$. By these assumptions, $\lambda_k - \lambda_j \geq k - j$ for $k > j$.

Since G_n is skew-symmetric, it is sufficient to consider quadratic forms: given (complex) x_j $(1 \leq j \leq n)$, let $S(x) = \sum_{j=1}^n \sum_{k=1}^n g_{j,k} x_j \overline{x}_k$. Then $\|G_n\| = \sup\{|S(x)| : \|x\| = 1\}$.

Let $f(t) = \sum_{j=1}^n x_j e(\lambda_j t)$, so that

$$|f(t)|^2 = \sum_{j=1}^n |x_j|^2 + \sum_{j=1}^n \sum_{k \neq j} x_j \overline{x}_k e[(\lambda_j - \lambda_k)t].$$

With $c > 0$ to be chosen later, we will use the fact that $I \geq 0$, where

$$I = \int_0^c (c-t)|f(t)|^2 \, dt.$$

(The use of $c - t$ instead of t makes the following work slightly simpler.) Now for $\lambda \neq 0$,

$$\int_0^c e(\lambda t) \, dt = \frac{e(\lambda c) - 1}{2\pi i \lambda},$$

$$\begin{aligned} \int_0^c (c-t) e(\lambda t) \, dt &= -\frac{c}{2\pi i \lambda} + \frac{1}{2\pi i \lambda} \int_0^c e(\lambda t) \, dt \\ &= -\frac{c}{2\pi i \lambda} - \frac{e(\lambda c) - 1}{4\pi^2 \lambda^2}. \end{aligned}$$

Also, $\int_0^c (c-t) \, dt = \frac{1}{2} c^2$. So

$$I = \tfrac{1}{2} c^2 \sum_{j=1}^n |x_j|^2 - \frac{c}{2\pi i} S(x) + R,$$

where

$$R = -\frac{1}{4\pi^2} \sum_{j=1}^n \sum_{k \neq j} \frac{x_j \overline{x}_k}{(\lambda_j - \lambda_k)^2} [e(\lambda_j c - \lambda_k c) - 1].$$

So

$$|R| \leq \frac{2}{4\pi^2} \sum_{j=1}^n \sum_{k \neq j} \frac{|x_j x_k|}{(\lambda_j - \lambda_k)^2} \leq \frac{1}{2\pi^2} \sum_{j=1}^n \sum_{k \neq j} \frac{|x_j x_k|}{(j-k)^2}.$$

By LS1.5,

$$|R| \leq \frac{1}{2\pi^2} \, \frac{\pi^2}{3} \sum_{j=1}^n |x_j|^2 = \frac{1}{6} \sum_{j=1}^n |x_j|^2.$$

So the inequality $I \geq 0$ translates into

$$\frac{c}{2\pi i} S(x) \leq \left(\tfrac{1}{2} c^2 + \tfrac{1}{6}\right) \sum_{j=1}^n |x_j|^2.$$

This applies equally to $S(\overline{x}) = -S(x)$, so in fact

$$|S(x)| \le \pi \left(c + \frac{1}{3c} \right) \sum_{j=1}^{n} |x_j|^2.$$

To minimize $c + 1/3c$, take $c = 1/\sqrt{3}$, giving $c + 1/3c = 2/\sqrt{3}$. (Note that if we simplified the proof by taking $c = 1$, the constant obtained would be $4/3$.) □

At the cost of rather more work, Montgomery and Vaughan obtained the result as stated, without the factor $2/\sqrt{3}$. Versions of their proof can be seen in [Mont1] and [MV2]. We will now take the liberty of assuming the result without the factor $2/\sqrt{3}$, and show how to derive Proposition LS1.3.

Proof of Proposition LS1.3. We use the well-known identity

$$\pi \operatorname{cosec} \pi\lambda = \lim_{n\to\infty} \sum_{m=-n}^{n} \left(1 - \frac{|m|}{n} \right) \frac{(-1)^m}{\lambda + m}, \tag{2}$$

which can be proved by applying Fejér's convergence theorem to the Fourier series for $\cos \lambda t$.

Choose $n > 1$ and scalars x_j. For $1 \le j \le n$ and $1 \le r \le n$, define

$$\lambda_{j,r} = \lambda_j + r,$$

$$x_{j,r} = (-1)^r x_j.$$

If $(j, r) \ne (k, s)$, then

$$|\lambda_{j,r} - \lambda_{k,s}| = |\lambda_j - \lambda_k - (s - r)| \ge \delta,$$

(if $j = k$, then $r \ne s$, so $|r - s| \ge 1$). Let

$$S_n(x) = \sum_{(j,r)\ne(k,s)} \frac{x_{j,r}\overline{x}_{k,s}}{\lambda_{j,r} - \lambda_{k,s}}$$

(the summation is over all j, k, r, s such that $(j, r) \ne (k, s)$). By Propositon LS1.6, applied to the scalars $\lambda_{j,r}$,

$$|S_n(x)| \le \frac{\pi}{\delta} \sum_{j=1}^{n} \sum_{r=1}^{n} |x_{j,r}|^2 = \frac{\pi n}{\delta} \sum_{j=1}^{n} |x_j|^2.$$

Now

$$S_n(x) = \sum_{(j,r)\ne(k,s)} \frac{(-1)^{r+s} x_j \overline{x}_k}{\lambda_j - \lambda_k + (r - s)}.$$

For fixed j, the $[(j, r), (j, s)]$ terms combine to

$$|x_j|^2 \sum_{r=1}^{n} \sum_{s\ne r} \frac{(-1)^{r+s}}{r - s} = 0,$$

so

$$S_n(x) = \sum_{j=1}^{n} \sum_{k \neq j} \sum_{r=1}^{n} \sum_{s=1}^{n} \frac{(-1)^{r+s} x_j \overline{x}_k}{\lambda_j - \lambda_k + (r-s)}.$$

For a chosen m with $-n \leq m \leq n$, there are $n - |m|$ (equal) terms with $r - s = m$ (e.g. if $m \geq 0$, these are given by $1 \leq s \leq n - m$ with $r = s + m$). Hence

$$S_n(x) = \sum_{j=1}^{n} \sum_{k \neq j} \sum_{m=-n}^{n} (n - |m|) \frac{(-1)^m x_j \overline{x}_k}{\lambda_j - \lambda_k + m}.$$

By (2),

$$\sum_{j=1}^{n} \sum_{k \neq j} \operatorname{cosec} \pi(\lambda_j - \lambda_k) x_j \overline{x}_k = \lim_{n \to \infty} \frac{1}{n} S_n(x),$$

and therefore the modulus of this expression is not greater than $(1/\delta) \sum_{j=1}^{n} |x_j|^2$. □

Further note on the constant in LS1.6. Our proof of LS1.6 actually gave $\|G_n\| \leq c\pi/\delta$, where $c = 2/\sqrt{3}$. This would lead to the value $N - 1 + c/\delta$ in Theorem LS1.1. In the asymptotic versions of the applications below, c/δ is fed into the "error" term estimated in terms of order of magnitude, so the value of c makes no difference to the statement.

The weighted version. The following refinement of Theorem LS1.1, making allowance for the points t_r not being equally spaced, was also proved in [MV1]: *Let* $\delta_r = \min_{s \neq r} \Delta(t_r - t_s)$. *Then*

$$\sum_{r=1}^{R} \left(N + \frac{3}{2\delta_r} \right)^{-1} |f(t_r)|^2 \leq \sum_{n=1}^{N} |x_n|^2.$$

It is not known whether the inelegant extra factor $\frac{3}{2}$ is really needed.

A more systematic account of Hilbert-type inequalities is given in the author's website notes [Jam2].

Another proof of Theorem LS1.1, due to Selberg, is based on harmonic analysis. The present writer finds it considerably harder. See [Mont1] or [Ten, section 1.4.5].

LS2. The arithmetic large sieve inequality

Special case: the Farey fractions

Nearly all number-theoretic applications of the large sieve use the following special case, in which the points t_j are the "Farey fractions".

For an integer $q \geq 1$, write G_q for the set of integers r such that $1 \leq r \leq q$ and $\gcd(r, q) = 1$. Note that the number of members of G_q is Euler's $\phi(q)$. We take the points t_j to be the numbers r/q, with $q \leq Q$ (where Q is to be chosen) and $r \in G_q$. Note that $0 < t_j \leq 1$. Consider two distinct such numbers $t_j = r_j/q_j$ and $t_k = r_k/q_k$. If $q_j = q_k = q$, then $|t_j - t_k| \geq 1/q$. If $q_j \neq q_k$, then $q_j q_k \leq Q(Q-1)$, so

$$|t_j - t_k| = \frac{|r_j q_k - r_k q_j|}{q_j q_k} \geq \frac{1}{q_j q_k} \geq \frac{1}{Q(Q-1)}.$$

The same applies if t_k (say) is replaced by $t_k + 1$, so $\Delta(t_j - t_k) \geq 1/Q(Q-1)$ whenever $j \neq k$, and Theorem LS1.1 gives:

Theorem LS2.1. *Let $I = \{M+1, M+2, \ldots, M+N\}$. Let numbers x_n $(n \in I)$ be given, and let $f(t) = \sum_{n \in I} x_n e(nt)$. Then*

$$\sum_{q \leq Q} \sum_{r \in G_q} \left| f\left(\frac{r}{q}\right) \right|^2 \leq [N + Q(Q-1)] \sum_{n \in I} |x_n|^2.$$

The result is usually stated with Q^2 instead of $Q(Q-1)$, but there is one application below where $Q(Q-1)$ will lead to a tidier result. There is no need for Q to be an integer. Indeed, a typical choice for Q will be $N^{1/2}$, so that $N + Q^2 = 2N$.

Of course, Gallagher's version of Theorem LS1.1 delivers πN instead of N. This would only make a minor difference in the applications described later.

The Farey fractions are by no means equally spaced! In fact, for distinct r/q and r'/q', one has $\Delta(r/q, r'/q') \geq 1/(qq')$, so the weighted version of Theorem LS1.1 becomes

$$\sum_{q \leq Q} \left(N + \tfrac{3}{2}qQ\right)^{-1} \sum_{r \in G_q} |f(r/q)|^2 \leq \sum_{n \in I} |x_n|^2.$$

In most applications, the problem is to give an upper estimate for the number of elements of a chosen subset E of I (we denote this number by $|E|$). In such cases, the numbers x_n will simply be 1 for $n \in E$ and 0 for other n, so that $\sum_{n \in I} |x_n|^2 = |E|$. Given that we are looking for an upper estimate, it will be seen that this quantity is on the wrong

side of the inequality in Theorem LS2.1! However, $|E|$ also equals $\sum_{n\in I} x_n = f(0)$, and we shall work towards introducing $f(0)^2$ on the left-hand side, so that cancellation will leave a spare $|E|$ on the left. However, we continue to work with general (complex) numbers x_n: this will be needed to keep the proof going at one point.

We mention that just one of the applications given in section LS3 (Theorem LS3.17) applies Theorem LS2.1 directly, without using the work that follows.

Sieving by congruence classes

Let us examine the sum $\sum_{r\in G_q} |f(r/q)|^2$ appearing in Theorem LS2.1. We start with the case where q is a prime, p. Then $G_p = \{1, 2, \ldots, p-1\}$. By the geometric series, we have

$$\sum_{r=0}^{p-1} e\left(\frac{rh}{p}\right) = \begin{cases} 0 & \text{if } p \text{ does not divide } h, \\ p & \text{if } p|h. \end{cases} \tag{3}$$

So the vectors e_h $(0 \le h \le p-1)$ defined by $e_h(r) = e(rh/p)$ form an orthogonal basis of $\mathbb{C}^p$ (not orthonormal: $\langle e_h, e_h\rangle$ is p, not 1!) The next remark just reproduces Parseval's identity for this basis. Given any complex numbers $y(r)$ for $0 \le r \le p-1$, we have, by (3),

$$\begin{aligned} \sum_{r=0}^{p-1} |y(r)|^2 &= \frac{1}{p}\sum_{r=0}^{p-1}\sum_{s=0}^{p-1} y(r)\overline{y(s)}\sum_{h=0}^{p-1} e\left(\frac{(s-r)h}{p}\right) \\ &= \frac{1}{p}\sum_{h=0}^{p-1}\sum_{r=0}^{p-1} y(r)e\left(-\frac{rh}{p}\right)\sum_{s=0}^{p-1}\overline{y(s)}e\left(\frac{sh}{p}\right) \\ &= \frac{1}{p}\sum_{h=0}^{p-1}\left|\sum_{r=0}^{p-1} y(r)e\left(-\frac{rh}{p}\right)\right|^2. \end{aligned}$$

With $y(r) = f(r/p)$ (for any function f), this becomes

$$\sum_{r=0}^{p-1}\left|f\left(\frac{r}{p}\right)\right|^2 = \frac{1}{p}\sum_{h=0}^{p-1}|\hat{f}(h)|^2, \tag{4}$$

where

$$\hat{f}(h) = \sum_{r=0}^{p-1} f\left(\frac{r}{p}\right) e\left(-\frac{rh}{p}\right). \tag{5}$$

Note that this sum includes $|f(0)|^2$, given by the term $r = 0$.

For our $f(t) = \sum_{n\in I} x_n e(nt)$, $\hat{f}(h)$ equates to a sum that picks out congruence classes mod p. In fact, substituting for $f(r/p)$ in (5) and applying (3), we have

$$\hat{f}(h) = \sum_{r=0}^{p-1}\sum_{n\in I} x_n e\left(\frac{rn}{p}\right) e\left(-\frac{rh}{p}\right)$$

$$
\begin{aligned}
&= \sum_{n\in I} x_n \sum_{r=0}^{p-1} e\left(\frac{(n-h)r}{p}\right) \\
&= pS(p,h), \qquad (6)
\end{aligned}
$$

where

$$S(p,h) = \sum\{x_n : n \in I \text{ and } n \equiv h \bmod p\}.$$

Clearly,

$$\sum_{h=0}^{p-1} S(p,h) = \sum_{n\in I} x_n = f(0).$$

We define further

$$Z(p) = \sum_{h=0}^{p-1} \left| S(p,h) - \frac{1}{p} f(0) \right|^2, \qquad (7)$$

the "variance" of this finite distribution. It is elementary (in the usual way for averages) that

$$Z(p) = \sum_{h=0}^{p-1} |S(p,h)|^2 - \frac{1}{p}|f(0)|^2. \qquad (8)$$

The contribution to the sum in Theorem LS2.1 is, of course, $\sum_{r=1}^{p-1} |f(r/p)|^2$. We have established the following expression for it:

Lemma LS2.2. *With this notation, we have*

$$\sum_{r=1}^{p-1} \left| f\left(\frac{r}{p}\right) \right|^2 = pZ(p). \qquad (9)$$

Proof. Denote the stated sum by S. By (4), (6) and (8),

$$S = \frac{1}{p} \sum_{h=0}^{p-1} |\hat{f}(h)|^2 - |f(0)|^2 = p \sum_{h=0}^{p-1} |S(p,h)|^2 - |f(0)|^2 = pZ(p). \qquad \square$$

Now, at last, we introduce the sieving process! For each prime p, suppose that there are $\rho(p)$ congruence classes (mod p) on which x_n is always zero. More exactly, let F_p be the set of h such that $0 \le h \le p-1$ and $x_n = 0$ for all $n \in I$ with $n \equiv h \bmod p$, and let $\rho(p)$ be the number of members of F_p. For example, if the sieving process simply consists of removing multiples of certain primes p, then $\rho(p) = 1$ for these p. (In the literature, the notation ω is often used where we have ρ, but we will reserve $\omega(n)$ for its usual meaning, the number of prime divisors of n.)

Clearly, $S(p,h) = 0$ for $h \in F_p$. Counting only these terms in (7), we see that

$$\sum_{r=1}^{p-1} \left| f\left(\frac{r}{p}\right) \right|^2 = pZ(p) \ge \frac{\rho(p)}{p} |f(0)|^2.$$

With a little care, we can improve this estimate, as follows:

Proposition LS2.3. *Let p be prime. Let $\rho(p)$ be defined as above, and let*

$$g(p) = \frac{\rho(p)}{p - \rho(p)}. \tag{10}$$

Then

$$\sum_{r=1}^{p-1} \left| f\left(\frac{r}{p}\right) \right|^2 \geq g(p)|f(0)|^2. \tag{11}$$

Proof. Write F_p^* for the set of $p - \rho(p)$ elements of $\{0, 1, \ldots, p-1\}$ not in F_p. Then $f(0) = \sum_{h \in F_p^*} S(p,h)$, so, by the Cauchy-Schwarz inequality,

$$|f(0)|^2 \leq [p - \rho(p)] \sum_{h=0}^{p-1} |S(p,h)|^2.$$

Hence, by (9),

$$\begin{aligned} \sum_{r=1}^{p-1} \left| f\left(\frac{r}{p}\right) \right|^2 &= p \sum_{r=0}^{p-1} |S(p,h)|^2 - |f(0)|^2 \\ &\geq \rho(p) \sum_{h=0}^{p-1} |S(p,h)|^2 \\ &= \frac{\rho(p)}{p - \rho(p)} |f(0)|^2. \qquad \square \end{aligned}$$

We remark that equality occurs in (11) if $S(p,h)$ has the same value (say K) for each $h \in F_p^*$. For then $f(0) = [p - \rho(p)]K$, so that $[p - \rho(p)] \sum_{r=0}^{p-1} S(p,h)^2 = [p - \rho(p)]^2 |K|^2 = |f(0)|^2$, which leads to equality in (11). To obtain an explicit example of this situation, just choose F_p and take x_n to be 1 for $n \in F_p^*$ and 0 for $n \in F_p$.

It would be possible to insert (11) into Theorem LS2.1, simply discarding all numbers in the sum except the primes. Historically, the resulting statement has been presented as a theorem. However, in most applications, it discards too much. We need a version of LS2.3 for composite numbers. This can be done, at least for square-free numbers, as we now show. First, an elementary lemma.

Lemma LS2.4. *Let* $\gcd(q, q') = 1$. *Then the set*

$$\{aq' + bq : a \in G_q,\ b \in G_{q'}\}$$

is equivalent (mod qq') to the set $G_{qq'}$.

Proof. Firstly, numbers of the form stated are coprime to q and q', hence to qq'. Conversely, choose $c \in G_{qq'}$. Take r, s such that $rq + sq' = 1$. Then $\gcd(s, q) = 1$, so $\gcd(cs, q) = 1$: let a be the element of G_q congruent to cs mod q. Similarly, let b be the element of $G_{q'}$ congruent to cr mod q'. Let $c' = aq' + bq$. Then $c' \equiv csq' \equiv c \bmod q$, and similarly $c' \equiv c \bmod q'$, so $c' \equiv c \bmod qq'$. □

We denote by $PD(q)$ the set of prime divisors of q (not a standard notation). As a notational device to pick out the square-free numbers, we use the Möbius function μ: recall that $\mu(q)^2$ is 1 if q is square-free and 0 otherwise (we only want μ^2, not μ itself).

Proposition LS2.5. *Let $g(p)$ be defined for prime p by (10). For composite q, define*

$$g(q) = \mu(q)^2 \prod_{p \in PD(q)} g(p) \tag{12}$$

(also, let $g(1) = 1$). Then, for all q,

$$\sum_{r \in G_q} \left| f\left(\frac{r}{q}\right) \right|^2 \geq g(q)|f(0)|^2. \tag{13}$$

Note. The function g is multiplicative. Indeed, we could have defined it by simply stipulating that it is multiplicative and that $g(p^k) = 0$ for $k \geq 2$.

Proof. The statement is trivial for $q = 1$ (since $G_1 = \{1\}$), and is given by LS2.3 when q is prime. So it is enough to show that if it holds for coprime square-free numbers q and q', then it holds for qq'.

If x_n is replaced by $x_n e(nu)$, then $f(t)$ becomes $f(t + u)$ (this is where we needed to allow complex x_n). Hence the statement for q is equivalent to

$$\sum_{a \in G_q} \left| f\left(\frac{a}{q} + u\right) \right|^2 \geq g(q)|f(u)|^2$$

for any u. By this fact and Lemma LS2.4,

$$\begin{aligned}
\sum_{c \in G_{qq'}} \left| f\left(\frac{c}{qq'}\right) \right|^2 &= \sum_{a \in G_q} \sum_{b \in G_{q'}} \left| f\left(\frac{a}{q} + \frac{b}{q'}\right) \right|^2 \\
&\geq g(q) \sum_{b \in G_{q'}} \left| f\left(\frac{b}{q'}\right) \right|^2 \\
&\geq g(q)g(q')|f(0)|^2 \\
&= g(qq')|f(0)|^2,
\end{aligned}$$

as required. □

Inserting this result into Theorem LS2.1, and recalling that $f(0) = \sum_{n\in I} x_n$, we finally derive the fundamental result of this section, the "arithmetic large sieve inequality".

Theorem LS2.6 *Let $I = \{M+1, M+2, \ldots, M+N\}$. Let (complex) numbers x_n $(n \in I)$ be given. Let $g(q)$ be defined by (10) and (12), and let $G(Q) = \sum_{q\le Q} g(q)$. Then, for any $Q > 1$,*

$$G(Q)\left|\sum_{n\in I} x_n\right|^2 \le [N + Q(Q-1)]\sum_{n\in I}|x_n|^2.$$

If E is a subset of I, and x_n is 1 for $n \in E$ and 0 otherwise, then both $\sum_{n\in E} x_n$ and $\sum_{n\in I}|x_n|^2$ equal $|E|$. After cancellation of $|E|$, we have the following corollary, which is the form in which the theorem will be applied in the examples to follow:

Corollary LS2.7. *If E is a subset of I, then*

$$|E| \le \frac{N + Q(Q-1)}{G(Q)}. \qquad \square$$

Clearly, the method is limited to providing an upper bound for $|E|$; it has nothing to say about a lower bound.

To apply LS2.7 in any particular case, we clearly have to find a lower estimate for $G(Q)$. This can be an interesting problem in its own right, often involving estimations of the partial sums of number-theoretic functions.

The cases where $\rho(p)$ is 1 or 2

In many applications, including all the ones discussed below, we have either $\rho(p) = 1$ or $\rho(p) = 2$ for a designated set P_1 of primes (typically, the primes $p \le N^{1/2}$). We now give estimations of $G(Q)$ specific to these two cases, using variants of the Euler product. We use the following standard notation for arithmetic functions:

$\tau(n)$ = the number of divisors of n;

$\omega(n)$ = the number of prime divisors of n;

$\Omega(n)$ = the number of prime factors of n, counted with multiplicity;

$k(n)$ = the square-free kernel of n (i.e. the product of its prime divisors):

$\pi(x)$ = the number of primes not greater than x.

From the usual expression for $\tau(n)$, it is easily seen that $2^{\omega(n)} \le \tau(n) \le 2^{\Omega(n)}$, with equality when n is square-free.

Also, given a set P_1 of primes, $\mathbb{N}(P_1)$ denotes the set of all integers (including 1) that are products of members of P_1.

With $\rho(p)$ given for primes p, the function g continues to be defined by (10) and (12), and $G(Q)$ means $\sum_{q\le Q} g(q)$.

Proposition LS2.8. *Let P_1 be a set of primes. Suppose that $\rho(p) = 1$ for $p \in P_1$ and $\rho(p) = 0$ for other primes p. Then*

$$G(Q) = \sum \left\{ \frac{1}{n} : n \in \mathbb{N}(P_1) \text{ and } k(n) \le Q \right\}.$$

Proof. Clearly, $g(q)$ is only non-zero for square-free q in $\mathbb{N}(P_1)$. First, recall that $g(1) = 1$. Now take $q = p_1 p_2 \dots p_k$ in $\mathbb{N}(P_1)$ with $1 < q \le Q$. Then

$$g(q) = \prod_{j=1}^{k} \frac{1}{p_j - 1} = \prod_{j=1}^{k} \left(\frac{1}{p_j} + \frac{1}{p_j^2} + \cdots \right).$$

This equals

$$\sum_{k(n)=q} \frac{1}{n},$$

since if $k(n) = q$, then $n = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$, where each $r_j \ge 1$, and $1/n$ is a term in the expansion of the product. (We remark in passing that $g(q)$ also equals $1/\phi(q)$.) □

We shall usually use this result in the following slightly weaker form:

$$G(Q) \ge \sum_{\substack{n \in \mathbb{N}(P_1) \\ n \le Q}} \frac{1}{n}.$$

We turn to the case where $\rho(p) = 2$, which is a bit more complicated.

Proposition LS2.9. *Let P_1 be a set of odd primes, and let $P_1^* = P_1 \cup \{2\}$. Suppose that $\rho(2) = 1$, $\rho(p) = 2$ for $p \in P_1$ and $\rho(p) = 0$ for other primes p. Then*

$$G(Q) \ge \frac{1}{3} \sum \left\{ \frac{\tau(n)}{n} : n \in \mathbb{N}(P_1^*) \text{ and } k(n) \le Q \right\}.$$

Proof. Note that $g(2) = 1$ and $g(p) = 2/(p-2)$ for $p \in P_1$. First consider square-free q in $\mathbb{N}(P_1)$, say $q = p_1 p_2 \dots p_k$. Then

$$g(q) = \prod_{j=1}^{k} \frac{2}{p_j - 2} = \prod_{j=1}^{k} \left(\frac{2}{p_j} + \frac{2^2}{p_j^2} + \cdots \right). \tag{14}$$

This clearly equals

$$\sum_{k(n)=q} \frac{2^{\Omega(n)}}{n},$$

and it is elementary that $2^{\Omega(n)} \geq \tau(n)$.

Now consider an even integer $q > 2$ with $g(q) \neq 0$. Then $q = 2p_1p_2 \dots p_k$, where the p_j are in P_1, and $g(q)$ is still given by (14). The integers n with $k(n) = q$ are of the form $n_r = 2^r m$, where $r \geq 1$ and $m = p_1^{r_1} p_2^{r_2} \dots p_k^{r_k}$, with each $r_j \geq 1$. One term in the expansion of $g(q)$ is again $2^{\Omega(m)}/m \geq \tau(m)/m$. Now $\tau(n_r) = (r+1)\tau(m)$, hence

$$\sum_{r=1}^{\infty} \frac{\tau(n_r)}{n_r} = \frac{\tau(m)}{m} \sum_{r=1}^{\infty} \frac{r+1}{2^r} = 3\frac{\tau(m)}{m}.$$

This reasoning works equally for the case $q = 2$, with $m = 1$. The statement follows. □

We have actually shown the following stronger (if more complicated) statement:

$$\begin{aligned} G(Q) \;\geq\; & \sum \left\{ \frac{2^{\Omega(n)}}{n} : n \in \mathbb{N}(P_1),\ k(n) \leq Q \right\} \\ & + \frac{1}{3} \sum \left\{ \frac{\tau(n)}{n} : n \in \mathbb{N}(P_1^*),\ n \text{ even},\ k(n) \leq Q \right\}. \end{aligned}$$

To apply LS2.9, we will need lower estimates of sums involving $\tau(n)/n$. For a sum of the form $\sum_{n \leq x} \tau(n)/n$, an instant, but inaccurate, estimate can be given as follows, using the well-known fact that $\sum_{n \leq x} 1/n > \log x$:

$$\sum_{n \leq x} \frac{\tau(n)}{n} \geq \left(\sum_{j \leq x^{1/2}} \frac{1}{j} \right)^2 \geq \tfrac{1}{4} (\log x)^2.$$

With only a little more trouble, we can give the following estimate, in which the constant $\frac{1}{2}$ is in fact optimal. We use Abel's summation formula in the form

$$\sum_{n \leq x} a(n)[f(x) - f(n)] = \int_1^x A(t) f'(t)\, dt,$$

where $A(x) = \sum_{n \leq x} a(n)$ (e.g. [Jam1, Proposition 1.3.6]).

Lemma LS2.10. *For all $x > 1$,*

$$\sum_{n \leq x} \frac{\tau(n)}{n} \geq \tfrac{1}{2} (\log x)^2.$$

Proof. We have $\tau(n)/n = (a * a)(n)$, where $a(n) = 1/n$. Then $A(x) \geq \log x$ for all $x \geq 1$. By the expression for the partial sum of a convolution, ([Jam1, Proposition 1.8.4]).

$$\sum_{n \leq x} \frac{\tau(n)}{n} = \sum_{n \leq x} a(n) A\left(\frac{x}{n}\right) \geq \sum_{n \leq x} a(n)(\log x - \log n).$$

By Abel's summation formula,

$$\sum_{n\le x} a(n)(\log x - \log n) = \int_1^x \frac{A(t)}{t}\,dt \ge \int_1^x \frac{\log t}{t}\,dt = \tfrac{1}{2}(\log x)^2. \qquad \square$$

Of course, more accurate estimations are known, but they would not lead to any improvement in the applications that follow.

Lemma LS2.11. *For all $x > 1$,*

$$\sum_{\substack{n\le x\\ n \text{ odd}}} \frac{\tau(n)}{n} \ge \tfrac{1}{8}(\log x)^2.$$

Proof. Similar to LS2.10, using the arithmetic function

$$b(n) = \begin{cases} 1/n & \text{for } n \text{ odd},\\ 0 & \text{for } n \text{ even}, \end{cases}$$

for which $B(x) \ge \frac{1}{2}\log x$. $\square$

LS3. Some applications

Primes in intervals

Our first application is an upper bound for the number of primes in an interval $[M+1, M+N]$ of length N. If M is either 1 or N, then a bound of the form $CN/\log N$ is given quite easily by Chebyshev's estimate ([Ap], [Jam1]). The point is to give an estimate valid for all M (actually, we assume that $M \ge N^{1/2}$, but it is then easy to dispose of the remaining case).

Theorem LS3.1. *For any M, N with $M \ge N^{1/2}$, we have*

$$\pi(M+N) - \pi(M) \le \frac{4N}{\log N}.$$

For any given $\varepsilon > 0$, the factor 4 can be replaced by $(2+\varepsilon)$ for all sufficiently large N.

Proof. Let E be the set of primes in $[M+1, M+N]$, and let P_1 be the set of primes $p \le N^{1/2}$. For $p \in P_1$, E contains no multiples of p, since $M+1 > N^{1/2}$. So (in our notation) $\rho(p) = 1$ for $p \in P_1$. Now $\mathbb{N}(P_1)$ clearly contains all integers in $[1, N^{1/2}]$. By LS2.8, for any $Q \le N^{1/2}$, we have

$$G(Q) > \sum_{n\le Q} \frac{1}{n} > \log Q.$$

So, by LS2.7,

$$|E| \le \frac{N+Q^2}{\log Q}.$$

The choice $Q = N^{1/2}$ gives $|E| \le 4N/\log N$. If, instead, we take $Q = N^{1/2}/\log N$, we obtain a bound of the form

$$\frac{N}{\log N}\left(2 + O\left(\frac{\log\log N}{\log N}\right)\right),$$

hence the second version of the statement. □

Note 1: The case $M < N^{1/2}$. For this case, the theorem gives the bound $4N/\log N$ for the number of primes between $N^{1/2}$ and $M+N$. To this, we only need to add $N^{1/2}$ or (slightly better) Chebyshev's estimate in the form $\pi(N^{1/2}) \le 4N^{1/2}/\log N$.

Note 2. The estimation really applies to a larger set than E, namely the set of integers in $[M+1, M+N]$ whose prime factors are all greater than $N^{1/2}$.

Twin primes and related problems

"Twin primes" are prime pairs p, $p+2$. We denote by $J(N)$ be the number of primes $p \le N$ such that $p+2$ is also prime (thereby counting pairs). Given that the prime number theorem says, informally, that the density of primes around N is about $1/(\log N)$, one might expect $J(N)$ to be something like $N/(\log N)^2$. As is typical for all sieve methods, we will actually give an upper bound for $J(N) - J(N^{1/2})$.

A similar problem is to estimate the number of primes $p \le N$ such that $2p-1$ is also prime: denote this by $J_1(N)$. We shall see that the same method applies.

Of course, the large sieve provides an upper bound, but not a lower one. In fact, for both cases, no non-trivial lower bound is known: it is a long-standing unsolved problem whether there are infinitely many pairs of primes as described!

Theorem LS3.2. *For all $N > 1$,*

$$J(N) - J(N^{1/2}) \le \frac{48N}{(\log N)^2}.$$

For large enough N, the factor 48 can be replaced by $24+\varepsilon$. The same estimate applies to $J_1(N)$.

Proof. With N fixed, let E be the set of n such that $N^{1/2} < n \le N$ and both n and $n+2$ are prime, so that $|E| = J(N) - J(N^{1/2})$. Then E contains no even numbers, so $\rho(2) = 1$. Let P_1 be the set of primes p in $[3, N^{1/2}]$. For each such p, the set E contains no

numbers congruent to 0 or -2 mod p, so $\rho(p) = 2$. By LS2.9 and LS2.10, for any $Q \le N^{1/2}$,

$$G(Q) \ge \frac{1}{3} \sum_{n \le Q} \frac{\tau(n)}{n} \ge \frac{1}{6} (\log Q)^2,$$

hence

$$|E| \le \frac{6(N + Q^2)}{(\log Q)^2}.$$

The choice $Q = N^{1/2}$ gives the estimate stated, and $Q = N^{\frac{1}{2}}/(\log N)$ gives the second estimate for large enough N.

The same method applies to $J_1(N)$. The second congruence class excluded is q mod p, where $p = 2q - 1$: if $n \equiv q$ mod p, then $2n - 1$ is a multiple of p. □

Of course, the same applies (for example) to primes p such that $2p+1$ is prime ("Sophie Germain primes").

The more detailed version of LS2.9, together with LS2.11, gives

$$G(Q) \ge \sum_{\substack{n \le Q \\ n \text{ odd}}} \frac{\tau(n)}{n} + \frac{1}{3} \sum_{\substack{n \le Q \\ n \text{ even}}} \frac{\tau(n)}{n} \ge \tfrac{1}{4} (\log Q)^2.$$

Hence the 48 can be replaced by 32 and the 24 by 16.

By a more elaborate method, expressing $ng(n)$ as a convolution of the form $2^\omega * h$, one can determine the C such that $G(Q) \sim C(\log Q)^2$ as $Q \to \infty$, and hence the best asymptotic constant afforded by this method. See [Ten, section 1.4.6]. The constant is not much smaller than our 16, and in the absence of a lower bound, this refinement is arguably of somewhat limited interest.

Numbers with all prime factors congruent to 1 mod 4

Let P_1 be the set of primes congruent to 1 mod 4 and P_2 the set of primes congruent to -1 mod 4. Also, let $P_j^* = P_j \cup \{2\}$ for $j = 1, 2$, and write $P_j(N)$ for $P_j \cap [1, N]$.

Further, write $\mathbb{N}_j$ for $\mathbb{N}(P_j)$, the set of integers (including 1) that are products of primes in P_j, also $\mathbb{N}_j(N) = \mathbb{N}_j \cap [1, N]$ and $\mathbb{N}_j(M, N) = \mathbb{N}_j \cap [M + 1, N]$. (similarly for $\mathbb{N}_j^*$). We will estimate the number of members of $\mathbb{N}_j(M, N)$. This is the purest of "sieving" questions, since, for example, $\mathbb{N}_1$ is exactly the set remaining after removing multiples of all the primes in P_2^*. So for the estimation of $|\mathbb{N}_1(M, N)|$, we have $\rho(p) = 1$ for $p \in P_2^*$, and LS2.8 says that $G(Q) \ge H_2^*(Q)$, where

$$H_j(Q) = \sum_{n \in \mathbb{N}_j(Q)} \frac{1}{n}, \qquad H_j^*(Q) = \sum_{n \in \mathbb{N}_j^*(Q)} \frac{1}{n}.$$

Of course, we shall have to depend on some result affirming that there are plenty of members of both P_1 and P_2. More exactly, we assume the following, a combination of the theorems of Dirichlet and Mertens [Ap, chapter 7].

Proposition LS3.3. *There are constants c_j, C_j $(j = 1, 2)$ such that as $Q \to \infty$*

$$\sum_{p \in P_j(Q)} \frac{1}{p} = \tfrac{1}{2} \log\log Q + c_j + O\left(\frac{1}{\log Q}\right),$$

$$\prod_{p \in P_j(Q)} \left(1 - \frac{1}{p}\right)^{-1} \sim C_j (\log Q)^{1/2}. \qquad \square$$

The constants C_j can be identified, but we will not attempt this here. Clearly, there are (possibly larger) constants A_j such that $\prod_{p \in P_j(Q)} (1 - 1/p)^{-1} \leq A_j (\log Q)^{1/2}$ for all Q. By the Euler product,

$$\prod_{p \in P_j(Q)} \left(1 - \frac{1}{p}\right)^{-1} = \sum \left\{ \frac{1}{n} : n \in \mathbb{N}[P_j(Q)] \right\}.$$

Now $\mathbb{N}_j(Q)$ is a subset of $\mathbb{N}[P_j(Q)]$ (this is obvious when you remember what the notation means!), hence:

Lemma LS3.4. *With A_j as above, we have (for all $Q \geq 2$),*

$$H_j(Q) \leq A_j (\log Q)^{1/2} \qquad (j = 1, 2). \qquad \square$$

At the same time, we have the following obvious lower bound:

Lemma LS3.5. *We have $H_1(Q) H_2^*(Q) \geq \log Q$ (and similarly for $H_1^*(Q) H_2(Q)$).*

Proof. Every $n \leq Q$ is uniquely expressible as st, where $s \in \mathbb{N}_1(Q)$ and $t \in \mathbb{N}_2^*(Q)$, so $H_1(Q) H_2^*(Q) \geq \sum_{n \leq Q} \frac{1}{n}$. $\square$

By LS3.4 and LS3.5, we have at once:

Lemma LS3.6. *For all $Q \geq 2$, we have*

$$H_1^*(Q) \geq \frac{1}{A_2} (\log Q)^{1/2}, \qquad H_2^*(Q) \geq \frac{1}{A_1} (\log Q)^{1/2}. \quad \square$$

Some minor variants of these estimates are worth mentioning. Firstly, there is an opposite inequality to the one in LS3.5 (showing that the given one is not too wasteful):

$$H_1(Q) H_2^*(Q) \leq \sum_{n \leq Q^2} \frac{1}{n} \leq 2 \log Q + 1.$$

Secondly, in the same way,

$$2H_j(Q) = \left(1 + \frac{1}{2} + \frac{1}{2^2} + \cdots\right) H_j(Q) \geq H_j^*(Q),$$

so $H_1(Q) \geq (\log Q)^{1/2}/(2A_2)$.

Theorem LS3.7. *With A_j as in LS3.4, we have for all M and all $N \geq 2$,*

$$|\mathbb{N}_j(M,N)| \leq 2\sqrt{2}\, A_j \frac{N}{(\log N)^{1/2}}.$$

For sufficiently large N, we can replace 2 by $(1+\varepsilon)$. For $|\mathbb{N}_j^*(M,N)|$, these bounds are doubled.

Proof. By LS2.7 and LS2.8,

$$|\mathbb{N}_1(M,N)| \leq \frac{N+Q^2}{H_2^*(Q)} \leq A_1 \frac{N+Q^2}{(\log Q)^{1/2}}.$$

Take $Q = N^{1/2}$ or $Q = N^{1/2}/(\log N)$ to obtain the two statements (for $j=1$). □

These results generalise to the case where P_1 is the set of primes congruent to a mod k. The $\frac{1}{2}$ is replaced by $1/\phi(k)$ in LS3.3, and by $1 - 1/\phi(k)$ in LS3.7.

We digress here to sketch how one can give both upper and lower bounds for $|\mathbb{N}_j(N)|$ (though not $|\mathbb{N}_j(M,N)|$) without the large sieve. Recall that $\Lambda * u = \ell$, where Λ is the von Mangoldt function, $u(n) = 1$ and $\ell(n) = \log n$ for all n. Let u_j be the indicator function of $\mathbb{N}_j$:

$$u_j(n) = \begin{cases} 1 & \text{if } n \in \mathbb{N}_j, \\ 0 & \text{otherwise} \end{cases}$$

Then u_j is completely multiplicative, and hence $(\Lambda u_j) * u_j = \ell u_j$. Now write

$$L_j(N) = \sum_{n \in \mathbb{N}_j(N)} \log n = \sum_{n \leq N} u_j(n)\ell(n), \qquad \psi_j(N) = \sum_{n \leq N} u_j(n)\Lambda(n).$$

There is a constant K such that $\psi_j(N) \leq KN$ for all $N \geq 1$ and $j = 1, 2$ (by Chebyshev's estimate, $K \leq 2$; the prime number theorem for residue classes says that K is asymptotically $\frac{1}{2}$). By summation of the convolution $\ell u_j = (\Lambda u_j) * u_j$,

$$L_j(N) = \sum_{n \leq N} u_j(n)\psi_j\left(\frac{N}{n}\right) \leq KN \sum_{n \leq N} \frac{u_j(n)}{n} = KNH_j(N) \leq KA_jN(\log N)^{1/2}.$$

Now by Abel summation,

$$L_j(N) = \sum_{n \leq N} u_j(n) \log n = |\mathbb{N}_j(N)| \log N - \int_1^N \frac{|\mathbb{N}_j(t)|}{t}\, dt.$$

Since $|\mathbb{N}_j(t)| \le t$, the integral term is less than N, and we deduce that

$$\mathbb{N}_j(N) \le (KA_j + 1)\frac{N}{(\log N)^{1/2}}.$$

Given the lower estimate $\psi_j(N) \ge K'N$ for large enough N, the same method leads to lower bounds for $|\mathbb{N}_j(N)|$ of the form $M_jN/(\log N)^{1/2}$, showing that the bound given by the large sieve is of the right order of magnitude. (An "Omega" result, stating that such lower bounds apply for arbitrarily large N, can be derived quickly by Abel summation from our lower bounds for $H_j(N)$.)

There is an asymptotic version of these results: for certain constants K_j (which can be identified), $|\mathbb{N}_j(N)| \sim K_jN/(\log N)^{1/2}$ as $N \to \infty$. The proof uses Karamata's Tauberian theorem to give asymptotic estimates for $H_j(N)$. Again, the results generalise to primes belonging a congruence classe mod k. See [Wir] and [Ten, exercise 9, p. 247].

Primes of the form $n^2 + 1$

It is another famous unsolved problem whether there are infinitely many integers n such that $n^2 + 1$ is prime. We shall give an upper bound for the number of such integers not greater than N, which we denote by $F(N)$. We shall actually consider $F(N) - F(N^{1/4})$, the number of members of

$$E_N = \{n : N^{1/4} < n \le N \text{ and } n^2 + 1 \text{ prime}\}.$$

It is elementary that any odd prime factors of $n^2 + 1$ must be congruent to 1 mod 4. We continue with the notation P_j, P_j^*, $\mathbb{N}_j$ (etc.) of the previous subsection.

Lemma LS3.8. *For the set E_N, we have $\rho(2) = 1$ and $\rho(p) = 2$ for $p \in P_1(N^{1/2})$.*

Proof. If $n > 1$ is odd, then $n^2 + 1$ is even, so not prime, hence n is not in E_N. So $\rho(2) = 1$.

Let $p \in P_1(N^{1/2})$. Then there exists u such that $u^2 \equiv -1 \bmod p$. If $n \equiv \pm u \bmod p$, then $p|(n^2 + 1)$. If also $n > N^{1/4}$, then $n^2 + 1 > N^{1/2}$, so $p \ne n^2 + 1$. Hence $n^2 + 1$ is not prime, so $n \notin E_N$. So $\rho(p) = 2$. □

The set we are estimating is distinctly larger than E_N: it includes all numbers $n \le N$ such that the smallest prime factor of $n^2 + 1$ is larger than $N^{1/2}$. For example, when $N = 100$, after removing odd numbers and those congruent to 2 or 3 mod 5, we are left with the numbers congruent to 0, 4 or 6 mod 10. There are 30 such numbers not greater than 100, of which only 17 have $n^2 + 1$ prime.

We follow the steps of the previous subsection, modified to take account of the fact that $\rho(p)$ is 2 instead of 1. The "Dirichlet-Mertens" theorem becomes: *there exist constants B_j $(j = 1, 2)$ such that for all $N \geq 2$,*

$$\prod_{p \in P_j(N)} \left(1 - \frac{2}{p}\right)^{-1} \leq B_j \log N.$$

This product clearly equals

$$\sum \left\{ \frac{2^{\Omega(n)}}{n} : n \in \mathbb{N}[P_j(N)] \right\}.$$

Let

$$K_j(N) = \sum_{n \in \mathbb{N}_j(N)} \frac{\tau(n)}{n}, \qquad K_j^*(N) = \sum_{n \in \mathbb{N}_j^*(N)} \frac{\tau(n)}{n}.$$

Since $2^{\Omega(n)} \geq \tau(n)$ and $\mathbb{N}_j(N)$ is a subset of $\mathbb{N}[P_j(N)]$, we deduce:

Lemma LS3.9. *For all $N \geq 2$, we have $K_j(N) \leq B_j \log N$ $(j = 1, 2)$.* □

Lemma LS3.10. *We have $K_1^*(N) K_2(N) \geq \frac{1}{2}(\log N)^2$.*

Proof. Every $n \leq N$ is uniquely expressible as st, where $s \in \mathbb{N}_1^*(N)$ and $t \in \mathbb{N}_2(N)$, so $K_1^*(N) K_2(N)$ contains the term

$$\frac{\tau(s)}{s} \frac{\tau(t)}{t} = \frac{\tau(n)}{n}.$$

By LS2.10, $K_1^*(N) K_2(N) \geq \sum_{n \leq N} \frac{\tau(n)}{n} \geq \frac{1}{2}(\log N)^2$. □

Corollary LS3.11. *We have $K_1^*(N) \geq (1/2B_2) \log N$.* □

Theorem LS3.12. *Let $F(N)$ be the number of integers $n \leq N$ with $n^2 + 1$ prime, and let B_2 be the constant in LS3.9. Then for all $N \geq 2$,*

$$F(N) - F(N^{1/4}) \leq 24 B_2 \frac{N}{\log N}.$$

The same holds for sufficiently large N with 24 replaced by $(12 + \varepsilon)$.

Proof. By LS2.9, $G(Q) \geq \frac{1}{3} K_1^*(Q) \geq (1/6B_2) \log Q$. Hence

$$|E_N| \leq 6 B_2 \frac{N + Q^2}{\log Q}.$$

Take Q to be $N^{1/2}$ or $N^{\frac{1}{2}}/(\log N)$ to obtain the statements. □

With a good deal more work, one can identify the constant C such that $G(Q) \sim C \log Q$ as $Q \to \infty$, and the method can be generalized to irreducible polynomials other than $n^2 + 1$: see [BD, 13.3].

The Brun-Titchmarsh theorem

Given a number $k \geq 3$ and a number a coprime to k, denote by $\pi(N, k, a)$ the number of primes not greater than N that are congruent to a mod k. We shall give an upper bound for $\pi(M+N, k, a) - \pi(M, k, a)$, thereby generalizing Theorem LS3.1. As one might expect, the previous bound will be essentially divided by $\phi(k)$. We assume that $M^2 > N/k$; as before, this is the case of real interest.

The primes we are counting are contained in the set of numbers of the form $nk + a$ satisfying:

(1) $M < nk + a \leq M + N$,

(2) $nk + a$ is not a multiple of any prime $p \leq (N/k)^{1/2}$.

Let E be the set of numbers n satisfying (1) and (2). We will give an upper bound for $|E|$. Condition (1) says that

$$\frac{M-a}{k} < n \leq \frac{M-a}{k} + \frac{N}{k}.$$

The number of such n is not greater than $N/k + 1$.

Let P_2 be the set of prime divisors of k, and let P_1 be the set of all primes not greater than $(N/k)^{1/2}$ that are not in P_2.

Lemma LS3.13. *For the set E, we have $\rho(p) = 1$ for $p \in P_1$.*

Proof. There is a unique n_0 in $[0, p-1]$ such that $n_0 k \equiv -a \bmod p$. So if $n \equiv n_0 \bmod p$, then p divides $nk + a$, hence $n \notin E$. □

Note that if $p \in P_2$, then $nk + a$ is not a multiple of p for any n, so no congruence class mod p is excluded from E.

Let

$$H_j(Q) = \sum_{\substack{n \in \mathbb{N}(P_j) \\ n \leq Q}} \frac{1}{n}$$

for $j = 1, 2$. By LS2.8, for any $Q \leq (N/k)^{1/2}$, we have $G(Q) \geq H_1(Q)$.

Lemma LS3.14. *We have* $H_1(Q) \geq \dfrac{\phi(k)}{k} \log Q$.

Proof. By the finite Euler product,

$$\frac{k}{\phi(k)} = \prod_{p \in P_2} \left(1 - \frac{1}{p}\right)^{-1} = \sum_{n \in \mathbb{N}(P_2)} \frac{1}{n} \geq H_2(Q).$$

Since every $n \le Q$ is expressible as rs, where $r \in \mathbb{N}(P_1)$ and $s \in \mathbb{N}(P_2)$, we have $H_1(Q)H_2(Q) \ge \sum_{n \le Q} \frac{1}{n} > \log Q$. □

Theorem LS3.15. *If* $\gcd(a,k) = 1$ *and* $M^2 > N/k > 1$, *then*

$$\pi(M+N,k,a) - \pi(M,k,a) \le \frac{4N}{\phi(k)(\log N - \log k)}.$$

For sufficiently large N,

$$\pi(M+N,k,a) - \pi(M,k,a) \le \frac{(2+\varepsilon)N}{\phi(k)\log N}.$$

Proof. This time we take advantage of the fact that $Q(Q-1)$, rather than Q^2, appears in the original statement of LS2.7. Our results combine to give

$$|E| \le \frac{k}{\phi(k)} \frac{(N/k) + 1 + Q(Q-1)}{\log Q} \le \frac{k}{\phi(k)} \frac{(N/k) + Q^2}{\log Q}.$$

Take $Q = (N/k)^{1/2}$ or $Q = (N/k)^{\frac{1}{2}}/(\log N)$ to obtain the statements. □

By a more careful development of the method, it is shown in [MV1] that the 4 can actually be replaced by 2 for all $N > k$.

An estimate for Dirichlet characters

This is a direct application of Theorem LS2.1, independent of the rest of section LS2. It requires familiarity with Dirichlet characters and Gauss sums (as presented, for example, in [Ap]). For a fixed modulus q, we denote by $\mathrm{ch}(q)$ the set of all Dirichlet characters mod q, and by $\mathrm{pch}(q)$ the set of all primitive ones.

For χ in $\mathrm{ch}(q)$, the *Gauss sum* $G(n,\chi)$ is defined by

$$G(n,\chi) = \sum_{r=1}^{q} \chi(r) e(rn/q).$$

We use the following well-known facts:

(DCH1) (orthogonality): for $r, s \in G_q$,

$$\sum_{\chi \in \mathrm{ch}(q)} \overline{\chi(r)}\chi(s) = \begin{cases} \phi(q) & \text{if } r = s, \\ 0 & \text{if } r \ne s. \end{cases}$$

(DCH2) if χ is primitive, then $G(n,\chi) = \overline{\chi(n)} G(1,\chi)$ for all n.

(DCH3) if χ is primitive, then $|G(1,\chi)| = q^{1/2}$.

Lemma LS3.16. *Let $I = \{M+1, M+2, \dots, M+N\}$. Let numbers x_n $(n \in I)$ be given. Let $f(t) = \sum_{n\in I} x_n e(nt)$, and for any Dirichlet character χ, let $T(\chi) = \sum_{n\in I} x_n \chi(n)$. Then, for any fixed q,*

$$\sum_{\chi \in \mathrm{pch}(q)} |T(\chi)|^2 \le \frac{\phi(q)}{q} \sum_{r \in G_q} \left| f\left(\frac{r}{q}\right)\right|^2 .$$

Proof. If χ is primitive, so is $\overline{\chi}$, so, by (DCH2) and (DCH3),

$$|T(\chi)| = \frac{1}{q^{1/2}} |V(\chi)|,$$

where

$$\begin{aligned} V(\chi) &= \sum_{n\in I} x_n G(n, \overline{\chi}) \\ &= \sum_{n\in I} x_n \sum_{r=1}^{q} \overline{\chi}(r) e\left(\frac{rn}{q}\right) \\ &= \sum_{r=1}^{q} \overline{\chi}(r) f\left(\frac{r}{q}\right). \end{aligned}$$

Now considering *all* characters mod q (which increases the sum in question), we have, by (DCH1),

$$\begin{aligned} \sum_{\chi \in \mathrm{ch}(q)} |V(\chi)|^2 &= \sum_{r=1}^{q} \sum_{s=1}^{q} f\left(\frac{r}{q}\right) \overline{f\left(\frac{s}{q}\right)} \sum_{\chi \in \mathrm{ch}(q)} \overline{\chi(r)} \chi(s) \\ &= \phi(q) \sum_{r \in G_q} \left| f\left(\frac{r}{q}\right)\right|^2 . \end{aligned}$$

The statement follows. □

By Theorem LS2.1, we have at once:

Theorem LS3.17. *Let x_n and $T(\chi)$ be as in LS3.16. Then for any Q,*

$$\sum_{q \le Q} \frac{q}{\phi(q)} \sum_{\chi \in \mathrm{pch}(q)} |T(\chi)|^2 \le (N + Q^2) \sum_{n\in I} |x_n|^2. \qquad \square$$

The reader may wonder whether this result is of any interest. It has been used by Rényi, Gallagher and Bombieri to obtain estimations of average and maximum deviations in the prime number theorem for arithmetic progressions (see [Dav]). These results, in turn, have been used as a step in a partial result towards Goldbach's conjecture: every sufficiently large even number can be expressed in the form $p + q$, where p is prime and q is a product of at most two primes.

REFERENCES

[Ap] T.M. Apostol, *Introduction to Analytic Number Theory*, Springer (1976).

[BD] P.T. Bateman and H.G. Diamond, *Analytic Number Theory*, World Scientific (2004).

[Bom] E. Bombieri, On the large sieve, *Mathematika* **12** (1965), 201–225.

[Dav] H. Davenport, *Multiplicative Number Theory*, 2nd ed., Springer (1980).

[DH] H. Davenport and H. Halberstam, The values of a trigonometric polynomial at well spaced points, *Mathematika* **13** (1966), 91–96.

[Gall] P.X. Gallagher, The large sieve, *Mathematika* **14** (1967), 14–20.

[Gr] G. Greaves, *Sieves in Number Theory*, Springer (2001).

[Jam1] G.J.O. Jameson, *The Prime Number Theorem*, Cambridge Univ. Press (2003).

[Jam2] G.J.O. Jameson, Hilbert's inequality and related results, `www.maths.lancs.ac.uk/~jameson`

[Lin] Yu. V. Linnik, The large sieve, *Dokl. Akad. Nauk SSSR* **30** (1941), 292–294 (Russian).

[Mont1] H.L. Montgomery, The analytic principle of the large sieve, *Bull. Amer. Math. Soc.* **84** (1978), 547–567.

[Mont2] H.L. Montgomery, *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, CBMS no. 84, Amer. Math. Soc. (1990).

[MV1] H.L. Montgomery and R.C. Vaughan, The large sieve, *Mathematika* **20** (1973), 119–134.

[MV2] H.L. Montgomery and R.C. Vaughan, Hilbert's inequality, *J. London Math. Soc.* (2) **8** (1974), 73-82.

[Ren] A. Rényi, On the large sieve of Ju. V. Linnik, *Compositio Math.* **8** (1950), 68–75.

[Ro] K.F. Roth, On the large sieves of Linnik and Rényi, *Mathematika* **12** (1969), 1–9.

[Ten] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Univ. Press (1995).

[Wir] E. Wirsing, Über die Zahlen, deren Primteiler einer gegebenen Menge angehören, *Arch. Math.* **7** (1956), 263–272.

updated 24 October 2017