Selberg’s sieve of irregular density

by

J. B. FRIEDLANDER (Toronto) and H. IWANIEC (Piscataway, NJ)

In memory of our friend, Professor Andrzej Schinzel

1. Introduction. Although perhaps not as frequently seen as the corresponding upper bound, the Selberg lower bound sieve has proved to be a versatile member of the analytic number theory toolbox. One needs only to mention the spectacular progress on gaps between primes occasioned by the works [GPY] and [M] which develop and harness its power for problems of large sieve dimension.

In this paper we study certain aspects of the Selberg sieve, in particular when sifting by rather thin sets of primes. We derive new results for the lower bound sieve suited especially for this setup and we apply them in particular to give a new sieve-propelled proof of Linnik’s theorem on the least prime in an arithmetic progression in the case of the presence of exceptional zeros. In an accompanying paper, in preparation, we shall give a new proof in the opposite case, also in the spirit of Selberg’s works.

Let \( A = (a_n) \) be a finite sequence of numbers \( a_n \geq 0 \). Our goal is to estimate the sifting function

\[
S(A, z) = \sum_{(n,P(z))=1} a_n
\]

where \( P(z) \) is the product of all primes \( p < z \).

To this end we assume that the congruence sums

\[
A_d = \sum_{n \equiv 0 \pmod{d}} a_n
\]

2020 Mathematics Subject Classification: Primary 11M20; Secondary 11N13, 11N35.

Key words and phrases: Selberg sieve, exceptional characters.

Received 19 July 2022; revised 4 September 2022.

Published online 27 October 2022.
are well-approximated by
\begin{equation}
A_d = g(d)X + r_d \quad \text{if } d \mid P(z)
\end{equation}
where \( g(d) \) is a multiplicative function (called the \textit{density}) satisfying
\begin{equation}
0 \leq g(p) < 1
\end{equation}
and \( X > 0 \) is a fixed number such that the error terms \( r_d \) are small on average.

We assume that the “remainder”
\begin{equation}
R(y) = \sum_{d < y} \tau_3(d) |r_d|
\end{equation}
is considerably smaller than the contribution to \( S(A, z) \) from the main terms in (1.3), so it is negligible. The larger the value of \( y \) that can be used, the stronger the estimates that can be obtained. Here, the mysterious looking presence of the divisor function \( \tau_3 \) is due to our use of the Selberg sieve weights.

The \( A^2 \)-sieve of Selberg yields a lovely upper bound
\begin{equation}
S(A, z) \leq XJ^{-1} + R(\Delta^2)
\end{equation}
where
\begin{equation}
J = \sum_{d \leq \Delta, d \mid P(z)} h(d)
\end{equation}
and \( h(d) \) is the multiplicative function, supported on squarefree numbers, with
\begin{equation}
h(p) = \frac{g(p)}{1 - g(p)}.
\end{equation}
Obviously,
\begin{equation}
J \leq \prod_{p \mid P(z)} (1 + h(p)) = V(z)^{-1}
\end{equation}
where
\begin{equation}
V(z) = \prod_{p \mid P(z)} (1 - g(p)).
\end{equation}
Hence, the main term in the upper bound (1.6) exceeds \( XV(z) \). The remainder is insignificant if, say,
\begin{equation}
R(\Delta^2) \ll XV(z)(\log \Delta)^{-1}.
\end{equation}
But, we still need a lower bound for \( J \). A strong and useful bound is difficult to establish unless we make some further assumptions about the density function \( g(d) \) or its companion \( h(d) \). It is quite often the case that
$h(d)$ behaves nicely in the sense that the Dirichlet series
\begin{equation}
D(s) = \sum_{d \geq 1} h(d)d^{1-s}
\end{equation}
admits an analytic continuation to Re $s > 1/2$, with only a pole at $s = 1$.

Chapter 7 of [Opera] provides a plethora of major examples with proofs. In Section 7.9 an asymptotic formula for $J$ is given under the assumption
\begin{equation}
\sum_{p \leq x} g(p) \log p = \kappa \log x + O(1)
\end{equation}
for every $x \geq 2$, where $\kappa$ is a positive constant called the sieve dimension. This approximation tells us that $g(p)p$ is $\kappa$ on average provided that $p$ is sufficiently large in terms of some defining parameters of $g$, so that the error term in (1.13) can be ignored.

However, in some practical sieve applications, the function $g(p)$ appears in segments in which it is not uniformly distributed. For example, in [Opera, Section 24.2] we have
\begin{equation}
\begin{cases}
g(p)p = 1 + \chi(p)(1 - 1/p) & \text{if } p \nmid q, \\
g(p)p = 0 & \text{if } p \mid q,
\end{cases}
\end{equation}
where $\chi \pmod{q}$ is a real, non-principal character. In this case (1.13) holds with $\kappa = 1$ but, as far as we know, with a terribly poor error term.

Fortunately, sieve methods can also produce useful estimates when one has access only to upper bounds for $g(p)$. In the above example, we can use the trivial bound $g(p)p < 2$ giving
\begin{equation}
\sum_{p \leq x} g(p) \log p \leq 2 \log x + O(1).
\end{equation}
To get an absolutely bounded error term we have here sacrificed the sieve dimension so the results are weaker. Still, if $z$ is small in the logarithmic scale (as in the Fundamental Lemma of sieve theory), compromising $\kappa$ does not significantly affect the output.

Even if $z$ is relatively large there can be significant consequences in the case of upper bounds. Unfortunately, the sieve of superficially enlarged dimension may yield a negative lower bound for $S(A, z)$ in a range of $z$ where it is expected to be positive.

In these notes we are concerned with $g(p)p$ fluctuating unpredictably within the segment
\begin{equation}0 \leq g(p)p < 2.
\end{equation}
We shall establish a positive lower bound
\begin{equation}S(A, z) \gg XV(z)
\end{equation}
for $z$ quite large, provided that $g(p)p$ is small on average. To this end we could go through the recurrence formula of Buchstab via the Fundamental Lemma. However, we can derive very explicit and neat results by means of Selberg’s lower-bound sieve method.

2. Selberg’s lower-bound sieve. Following Selberg [S], we have

$$S(A, z) \geq S^-(A, z)$$

where

$$S^-(A, z) = \sum_{n} a_n \left( 1 - \sum_{p \mid n, p < z} 1 \right) \left( \sum_{d \mid n} \rho_d \right)^2$$

with any real numbers $\rho_d, d \mid P(z), \rho_1 = 1$. We assume that $\rho_d$ are supported on squarefree $d \leq \Delta \leq z$, so the condition $d \mid P(z)$ is redundant. Opening the square and applying the approximations (1.3) we obtain

$$S^-(A, z) = X W + R$$

where

$$W = \sum_{d_1} \sum_{d_2} \rho_{d_1} \rho_{d_2} \left( g([d_1, d_2]) - \sum_{p < z} g([p, d_1, d_2]) \right)$$

and $R$ is the corresponding remainder

$$R = \sum_{d_1} \sum_{d_2} \rho_{d_1} \rho_{d_2} \left( r_{[d_1, d_2]} - \sum_{p < z} r_{[p, d_1, d_2]} \right).$$

In the quadratic form $W$ in the variables $\rho_d$ we make a linear change of variables, specifically setting

$$y_d = \frac{\mu(d)}{h(d)} \sum_{m \equiv 0 \pmod{d}} g(m) \rho_m.$$

Applying Möbius inversion, we find

$$\rho_{\ell} = \frac{\mu(\ell)}{g(\ell)} \sum_{d \equiv 0 \pmod{\ell}} h(d) y_d.$$

The quadratic form (2.4) in the new variables becomes

$$W = \sum_d h(d) y_d^2 - \sum_{p < z} g(p) \sum_{(d,p) = 1} h(d) (y_d - y_{pd})^2;$$

see [Opera, formula (7.109)].

Remark. We have tacitly assumed that $g(\ell), h(d)$ do not vanish to justify the transformations (2.6), (2.7). However, after obtaining (2.8) we no longer need this slight detour.
It is clear that the support conditions for $\rho_d$ and $y_d$ are the same, namely
(2.9) \[ d \leq \Delta, \quad d \text{ squarefree.} \]

The normalization $\rho_1 = 1$ becomes
(2.10) \[ \sum_d h(d)y_d = 1. \]

Now we choose
(2.11) \[ y_d = \frac{1}{H} \log \frac{\Delta}{d} \]
for $d \leq \Delta$, $d$ squarefree, and $y_d = 0$ otherwise, where
(2.12) \[ H = \sum_{d\leq\Delta} h(d) \log \frac{\Delta}{d}. \]

Recall that $h(d)$ is supported on squarefree numbers.

The sieve constituents $\rho_d$ are bounded. More precisely, as in the pure $\Lambda^2$-sieve of Selberg, for every squarefree $\ell \leq \Delta$ we argue as follows:
\[
H = \sum_{k|\ell} \sum_{d < \Delta \atop (d, \ell) = k} h(d) \log \frac{\Delta}{d} = \sum_{k|\ell} h(k) \sum_{m < \Delta/k \atop (m, \ell) = 1} h(m) \log \frac{\Delta}{mk}
\]
\[
\geq \left( \sum_{k|\ell} h(k) \right) \sum_{m < \Delta/\ell \atop (m, \ell) = 1} h(m) \log \frac{\Delta}{m\ell}.
\]

On the other hand, pulling the factor $h(\ell)$ out from the sum (2.7), we get
\[
\mu(\ell)\rho_\ell H = \frac{h(\ell)}{g(\ell)} \sum_{m < \Delta/\ell \atop (m, \ell) = 1} h(m) \log \frac{\Delta}{m\ell}.
\]

Combining these results we find that
(2.13) \[ |\rho_\ell| \leq 1. \]

Now, we can estimate the remainder (2.5). We obtain
(2.14) \[ |R| \leq R(\Delta^2) + 2(\log \Delta) R(z\Delta^2). \]

Next, we proceed to an estimation of our main term $W$. For our choice of $y_d$ given by (2.11) we find that
(2.15) \[ y_p - y_{pd} = H^{-1} \min \left( \log p, \log \frac{\Delta}{d} \right) \]
and the formula (2.8) becomes
(2.16) \[ H^2 W = K(\Delta) - \sum_{p < z} g(p) \sum_{d < \Delta \atop (d, p) = 1} h(d) \left\{ \min \left( \log p, \log \frac{\Delta}{d} \right) \right\}^2 \]
where, for any \( u \leq \Delta \),

\[
K(u) = \sum_{d \leq u} h(d) \left( \log \frac{\Delta}{d} \right)^2.
\]

Note that, abbreviating \( K = K(\Delta) \), we have \( H^2 \leq JK \) by Cauchy’s inequality.

Having in mind that \( g(p) \) is unstable at small primes, we split the range of \( p \) in (2.16) into two segments \( p \leq w \) and \( w < p < z \) with some \( 2 \leq w < z \) at our disposal. Then we estimate \( \min(\ldots, \ldots) \) by \( \log p \) and \( \log(\Delta/d) \), respectively. We get

\[
(2.18) \quad H^2 W \geq K(\Delta) \left( 1 - \sum_{w < p < z} g(p) \right) - \sum_{p \leq w} g(p) (\log p)^2 \sum_{d \leq \Delta \atop (d,p)=1} h(d).
\]

We estimate the contribution of \( p \leq w, \ d \leq w, \ (d,p) = 1 \) as follows (note that \( g(p)h(d) \leq h(pd) \)):

\[
\sum_{p \leq w, d \leq w \atop (p,d)=1} h(pd)(\log p)^2 \leq (\log w) \sum_{d \leq w^2} h(d) \log d.
\]

If \( \Delta \geq w^6 \) we have \( (\log w)(\log d) \leq \frac{1}{2} (3\alpha \log(\Delta/d))^2 \), where

\[
(2.19) \quad \alpha = \log w/\log \Delta.
\]

Hence, this contribution is bounded by \( \frac{9}{2} \alpha^2 K(w^2) \) where, as in (2.17),

\[
(2.20) \quad K(w^2) = \sum_{d \leq w^2} h(d)(\log(\Delta/d))^2.
\]

We drop the condition \( (d,p) = 1 \) in the remaining sum

\[
(2.21) \quad J(w, \Delta) = \sum_{w < d < \Delta} h(d).
\]

Introducing the above estimates into (2.18), we obtain

**Lemma 2.1.** Let \( z > w \geq 2 \) and \( \Delta \geq w^6 \). We have

\[
(2.22) \quad H^2 W \geq \left( 1 - \sum_{w < p \leq z} g(p) \right) K - \frac{9}{2} \alpha^2 K(w^2) - J(w, \Delta) G(w)
\]

where

\[
(2.23) \quad G(w) = \sum_{p \leq w} g(p)(\log p)^2.
\]

3. Two assumptions. So far, the inequality (2.22) holds without severe restrictions on the density functions \( g(d), \ h(d) \). To proceed further we accept two assumptions.
Assumption 1. If \( w \) is larger than some absolute constant, then
\[
G(w) \leq \frac{3}{2} (\log w)^2.
\]

Note that if \( g(p)p \) fluctuates in the interval \([0, 2]\), then
\[
G(w) \leq \sum_{p \leq w} \frac{2}{p} (\log p)^2 = (\log w)^2 + O(\log w),
\]
so in this case we need \( w \) to be sufficiently large to make the error term \( O(\log w) \) strictly smaller than \( \frac{1}{2} (\log w)^2 \).

Let \( K(w^2, \Delta) = K(\Delta) - K(w^2) \) denote the part of (2.17) that is complementary to (2.20):
\[
K(w^2, \Delta) = \sum_{w^2 < d \leq \Delta} h(d)(\log(\Delta/d))^2.
\]

Assumption 2. For \( \Delta \geq w^6 \) we have
\[
J(w, \Delta)(\log \Delta)^2 \leq 9K(w^2, \Delta).
\]

We shall illustrate how to verify (3.3) in special circumstances. But first we enjoy using both assumptions. By (3.1) and (3.3) we have
\[
J(w, \Delta)G(w) \leq \frac{27}{2} \alpha^2 K(w^2, \Delta).
\]
Recall that \( \alpha = \log w/\log \Delta \leq 1/6 \). Hence Lemma 2.1 yields
\[
H^2 W \geq \nu K
\]
where
\[
\nu = 1 - \sum_{w < p \leq z} g(p) - \frac{27}{2} \alpha^2.
\]
This lower bound is only interesting if \( \nu \) is positive. Then \( W \) is positive and we can apply the inequality \( H^2 \leq JK \) where
\[
J = J(\Delta) = \sum_{d < \Delta} h(d) \leq V(\Delta)^{-1}
\]
by (1.9). Hence

Theorem 1. Let \( \Delta \geq w^6 \) and \( g(d) \) be such that (3.1) and (3.3) hold. If \( z > w \) and \( \nu > 0 \), then
\[
JW \geq \nu.
\]

Remark. One is unlikely to obtain \( \nu > 0 \) in normal situations. However, this can happen in exceptional circumstances, as for \( g(p) \) given by (1.14) with an exceptional real character \( \chi \) (mod \( q \)). These notes were designed mainly to handle such exceptional cases.
4. Verification of Assumption 2. Although \( h(p)p \) is unpredictable at small primes, \( h(d)d \) can be quite regular for large \( d \) in the sense that

\[
\sum_{d \leq x} h(d)d = cx + O(q^{1/4}x^{3/4})
\]

holds for every \( x \geq 2 \) with some constants \( c > 0, q \geq 2 \) and an absolute implied constant in the error term. Having this formula, we can establish (3.3) by asymptotic evaluation of both sides.

Recall that \( \alpha = \log w / \log \Delta \leq 1/6 \). On the left side we get

\[
J(w, \Delta) = \int_{w}^{\Delta} x^{-1} d(cx + O(q^{1/4}x^{3/4}))
\]

\[
= c \log \frac{\Delta}{w} + O(q^{1/4}w^{-1/4}) = (1 - \alpha)c(\log \Delta + O(1))
\]

provided that \( w \gg qc^{-4} \). On the right side we get

\[
K(w^2, \Delta) = \int_{w^2}^{\Delta} x^{-1} \left( \frac{\Delta}{x} \right)^2 d(cx + O(q^{1/4}x^{3/4}))
\]

\[
= \frac{c}{3} \left( \log \frac{\Delta}{w^2} \right)^3 + O(q^{1/4}w^{-1/4}(\log \Delta)^2)
\]

\[
= \frac{c}{3} (1 - 2\alpha)^3(\log \Delta)^2(\log \Delta + O(1)).
\]

Hence, the ratio of the right side to the left side of (3.3) is

\[
\frac{3}{1 - \alpha} (1 - 2\alpha)^3 + O\left( \frac{1}{\log \Delta} \right) \geq \frac{16}{15} + O\left( \frac{1}{\log \Delta} \right) > 1
\]

provided that \( w \) is sufficiently large to compensate for the error term \( O(1/\log \Delta) \). This proves (3.3) if \( w \gg qc^{-4} \).

**Example.** Let \( g(p)p \) be given by (1.14). Then

\[
h(p)p = \left( 1 - \frac{1}{p} \right)^{-1} \left( 1 - \frac{\chi(p)}{p} \right)^{-1} \left( 1 + \chi(p) \left( 1 - \frac{1}{p} \right) \right)
\]

if \( p \nmid q \) and \( h(p) = 0 \) if \( p \mid q \). Hence, the series

\[
D(s) = \sum_{d} h(d)d^{1-s} = \prod_{p} (1 + h(p)p^{1-s}) = \zeta(s)L(s, \chi)E(s)
\]

has analytic continuation to \( \text{Re } s > 1/2 \). The local factors

\[
E_p(s) = (1 - p^{-s})(1 - \chi(p)p^{-s})(1 + h(p)p^{1-s})
\]

\[
= 1 + a_1p^{-s-1} + a_2p^{-2s} + a_3p^{-3s}
\]
have $a_1, a_2, a_3$ bounded. Hence (4.1) follows by standard contour integration with the constant
\[ c = \text{res}_{s=1} D(s) = L(1, \chi)E(1). \]
We have
\[ 1 + h(p) = (1 - g(p))^{-1} = \left( 1 - \frac{1}{p} \right)^{-1} \left( 1 - \frac{\chi(p)}{p} \right)^{-1} \]
if $p \nmid q$ and 1 if $p | q$. Hence $E(1) = \varphi(q)/q$ so the residue is
\[ (4.3) \quad c = L(1, \chi)\varphi(q)/q \gg \varphi(q)q^{-3/2}(\log q)^{-2}, \]
by the Dirichlet class number formula. Therefore, we have proved that (3.3) holds as long as
\[ (4.4) \quad \Delta \geq w^6, \quad w \geq q^3(\log q)^9 \]
and $q$ is sufficiently large.

5. Counting primes. We are going to apply the theorem to estimate the sum of $a_p$ over primes $p \leq x$. To this end we need
\[ (5.1) \quad \delta(w, z) = \sum_{w < p \leq z} g(p) \]
to be smaller than 1. Specifically, we take $\Delta = w^6$ so that $\alpha = 1/6$ and we make the following

Assumption 3.
\[ (5.2) \quad \delta(w, z) \leq 3/8. \]
We then have
\[ (5.3) \quad \nu = 1 - \delta(w, z) - 3/8 \geq 1/4. \]
so that (3.7) and (3.8) imply $W \geq \frac{1}{4}V$, where
\[ (5.4) \quad V = V(\Delta) = \prod_{p < \Delta} (1 - g(p)). \]
If $A = (a_n)$ is supported on $\sqrt{x} < n \leq x$, then, for $z = \sqrt{x} \geq w^7$, we have
\[ (5.5) \quad \sum_p a_p = S(A, z) \geq S^{-}(A, z) = XW + R \]
where the remainder $R$ is bounded by $R \log x$ with
\[ (5.6) \quad R = \sum_{d < \sqrt{xw^{1/2}}} \tau_3(d) |r_d|; \]
see (1.5) and (2.14). Hence we conclude the following:
Proposition 5.1. If $g$ is exceptional in the sense that (5.2) holds with $z = \sqrt{x} \geq w^7$, then
\[
\sum_{\sqrt{x} < p \leq x} a_p \geq \frac{1}{4} XV - R \log x.
\]

Recall that we work under Assumptions 1–3. The first two assumptions are verified in the case of $g(p)$ given by (1.14). The required conditions are those in (4.4). We choose
\[
w = q^3 (\log q)^9.
\]
Note that $V = V(\Delta) \geq M(\chi)q/(82\varphi(q))$ where
\[
M(\chi) = \prod_{p < q^2} \left(1 - \frac{1}{p}\right) \left(1 - \frac{\chi(p)}{p}\right)
\]
because for $q$ sufficiently large we have
\[
\prod_{q^2 \leq p < \Delta} \left(1 - \frac{1}{p}\right) \left(1 - \frac{\chi(p)}{p}\right) \geq \prod_{q^2 \leq p < \Delta} \left(1 - \frac{1}{p}\right)^2 \geq \frac{1}{82}.
\]

6. The exceptional character. Let $\lambda = 1 * \chi$, so $\lambda(p) = 1 + \chi(p)$. We have proved in [Opera (24.20)] that
\[
\sum_{q^3 < p \leq \sqrt{x}} \lambda(p)p^{-1} < (1 - \beta) \log x + O(q^{-3/4})
\]
if $x \geq q^6$, where $\beta$ is any real zero of $L(s, \chi)$ and the implied constant is absolute. Note that
\[
1 - \beta \gg q^{-1/2} (\log q)^{-2},
\]
so the error term in the inequality (6.1) is negligible unless $x$ is exceedingly large. Thus, our exceptional condition (5.2) holds in the segment
\[
q^6 \leq x \leq e^{1/(4(1-\beta))},
\]
which is non-empty if
\[
(1 - \beta) \log q \leq 1/24.
\]
Therefore, we shall say that the real primitive character $\chi$ is exceptional if $L(s, \chi)$ has a real zero $\beta$ satisfying (6.4). If also $\chi(a) = 1$ its effect pulls in an unfavourable direction.

Under this exceptional situation we are able to estimate the least prime in an arithmetic progression
\[
p \equiv a \pmod{q} \quad \text{with} \quad \chi(a) = 1.
\]
To this end, consider the sequence
\[
a_n = \lambda(n)
\]
Selberg’s sieve of irregular density

for $n \equiv a \pmod{q}$, $\sqrt{x} \leq n \leq x$. We set $a_n = 0$ otherwise. We have shown in [Opera, Section 24.2] that $A$ has our sieve properties with the density function (1.14) and

\[(6.7) \quad X = 2L(1, \chi)xq^{-1}.\]

The individual error terms satisfy

\[(6.8) \quad r_d \ll \tau_3(d)\sqrt{x/d}\]

if $d \leq x$, where the implied constant is absolute. Hence, our remainder (5.6) satisfies

\[(6.9) \quad R \log x \ll w^3x^{3/4}(\log x)^{10}.\]

We want this bound to be insignificant by comparison with the main term

\[(6.10) \quad \frac{1}{4}XV \gg xq^{-3/2}(\log x)^{-4}.\]

This is the case if $x \geq q^{43}$. We obtain

**Corollary 6.1.** Suppose $L(s, \chi)$ has a real zero $\beta$ with

\[(6.11) \quad (1 - \beta) \log q \leq 1/172.\]

Let $\chi(a) = 1$. Then, for $q^{43} \leq x \leq e^{1/(4(1-\beta))}$, we have

\[(6.12) \quad \pi(x; q, a) \geq L(1, \chi)M(\chi)\frac{x}{164\varphi(q)},\]

where $M(\chi)$ is the product (5.9).

**Remark.** There has been in recent years an increased interest in giving proofs of Linnik’s theorem which circumvent an appeal to some of the deep innovations employed in the original arguments. See, for example [K] Chapter 27 and the references therein.

**Acknowledgements.** We thank the referee for pointing out mistakes in some of our original computations and for suggestions about them.

Research of J. B. Friedlander was supported in part by NSERC grant A5123.

**References**

[Opera] J. B. Friedlander and H. Iwaniec, *Opera de Cribro*, Colloq. Publ. 57, Amer. Math. Soc., Providence, RI, 2010.

[GPY] D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Primes in tuples. I*, Ann. of Math. (2) 170 (2009), 819–862.

[K] D. Koukoulopoulos, *The Distribution of Prime Numbers*, Grad. Stud. Math. 203, Amer. Math. Soc., Providence, RI, 2019.

[M] J. Maynard, *Small gaps between primes*, Ann. of Math. (2) 181 (2015), 383–413.

[S] A. Selberg, *Lectures on sieves*, in: Collected Works, Vol. II, Springer, Berlin, 1991, 65–247.
