AN ALTERNATE ARGUMENT FOR THE ARITHMETIC LARGE SIEVE INEQUALITY

EMMANUEL KOWALSKI

The classical arithmetic large sieve inequality states that, for any real numbers N, $Q \geqslant 1$, any choice of subsets $\Omega_p \subset \mathbf{Z}/p\mathbf{Z}$ for primes $p \leqslant Q$, we have

$$|\{n \leqslant N \mid n \,(\mathrm{mod}\, p) \notin \Omega_p \text{ for } p \leqslant Q\}| \leqslant \frac{\Delta}{H} \tag{1}$$

where

$$H = \sideset{}{^\flat}\sum_{q \leqslant Q} \prod_{p \mid q} \frac{|\Omega_p|}{p - |\Omega_p|},$$

and Δ is any constant for which the "harmonic" large sieve inequality holds: for any complex numbers (a_n), we have

$$\sum_{q \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \Bigl| \sum_{n \leqslant N} a_n e\Bigl(\frac{an}{q}\Bigr)\Bigr|^2 \leqslant \Delta \sum_{n \leqslant N} |a_n|^2, \tag{2}$$

the notation $\sideset{}{^\flat}\sum$ and $\sideset{}{^*}\sum$ denoting, respectively, a sum over squarefree integers, and one over integers coprime with the (implicit) modulus, which is q here.

By work of Montgomery-Vaughan and Selberg, it is known that one can take

$$\Delta = Q^2 - 1 + N$$

(see, e.g., [IK, Th. 7.7]).

There are a number of derivations of (1) from (2); for one of the earliest, see [M1, Ch. 3]. The most commonly used is probably the argument of Gallagher involving a "submultiplicative" property of some arithmetic function (see, e.g., [K, §II.2] for a very general version).

We will show in this note how to prove (1) quite straightforwardly from the *dual* version of the harmonic large sieve inequality: Δ is also any constant for which

$$\sum_{n \leqslant N} \Bigl| \sum_{q \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \beta(q,a) e\Bigl(\frac{an}{q}\Bigr)\Bigr|^2 \leqslant \Delta \sum_{q \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} |\beta(q,a)|^2, \tag{3}$$

holds for arbitrary complex numbers $(\beta(q,a))$. This is of some interest because, quite often,[1] the inequality (1) is proved by duality from (3), and because, in recent generalized versions of the large sieve (see [K]), it often seems that the analogue of (3) is the most natural inequality to prove – or least, the most easily accessible. So, in some sense, one could dispense entirely with (2) for many applications!

Note that the argument we give is not really a new proof; some ingredients of most (if not all!) of the previous ones occur – most particularly the Cauchy inequality. There are

[1] But not always – Gallagher's very short proof, found e.g. in [M2, Th. 1, p. 549], proceeds directly, as does the Montgomery-Vaughan proof.

other proofs of (1) working directly from the inequality (3) which can be found in the older literature on the large sieve, usually with explicit connections with the Selberg sieve (see the references to papers of Huxley, Kobayashi, Matthews and Motohashi in [M2, p. 561]), although none of those that the author has seen seems to give an argument which is exactly identical or as well motivated.

Indeed, maybe the most interesting aspect of our proof is that it very easy to motivate. It flows very nicely from an attempt to improve the earlier inequality

$$|\{n \leqslant N \mid n \,(\mathrm{mod}\, p) \notin \Omega_p \text{ for } p \leqslant Q\}| \leqslant \frac{\Delta}{K}, \qquad K = \sum_{p \leqslant Q} \frac{|\Omega_p|}{p}, \tag{4}$$

of Rényi, which is most easily proved using (3) instead of (1), as in [K, §II.4].

Let

$$S = \{n \leqslant N \mid n \,(\mathrm{mod}\, p) \notin \Omega_p \text{ for } p \leqslant Q\},$$

be the sifted set; we wish to estimate from above the cardinality of this finite set. From (3), the idea is to find an "amplifier" of those integers remaining in the sifted set, i.e., an expression of the form

$$A(n) = \sum_{q \leqslant Q} \sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \beta(q,a) e\Big(\frac{an}{q}\Big)$$

which is *large* (in some sense) when $n \in S$. Then an estimate for $|S|$ follows from the usual Chebychev-type manoeuvre.

To construct the amplifier $A(n)$, we look first at a single prime $p \leqslant Q$. If $n \in S$, we have $n \,(\mathrm{mod}\, p) \notin \Omega_p$. If we expand the characteristic function of Ω_p in terms of additive characters,[2] we have then

$$0 = \mathbf{1}_{\Omega_p}(n) = \sum_{a \,(\mathrm{mod}\, p)} \alpha(p,a) e\Big(\frac{an}{p}\Big), \qquad \alpha(p,a) = \frac{1}{p} \sum_{x \in \mathbf{Z}/p\mathbf{Z}} \mathbf{1}_{\Omega_p}(x) e\Big(\frac{ax}{p}\Big),$$

and the point is that the contribution of the constant function (0-th harmonic) is, indeed, relatively "large", because it is

$$\alpha(p,0) = \frac{|\Omega_p|}{p},$$

and exactly reflects the probability of a random element being in Ω_p. Thus for $n \,(\mathrm{mod}\, p) \notin \Omega_p$, we have

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} \beta(p,a) e\Big(\frac{an}{p}\Big) = c_p \tag{5}$$

with

$$c_p = \frac{|\Omega_p|}{p}, \qquad \beta(p,a) = -\alpha(p,a).$$

If we only use the contribution of the primes in (3), and the amplifier

$$A(n) = \sum_{p \leqslant Q} \sum_{a \,(\mathrm{mod}\, p)} \beta(p,a) e\Big(\frac{an}{p}\Big),$$

[2] We use this specific basis to use (3), but any orthonormal basis containing the constant function 1 would do the job, as in [K].

then by (3), we get

$$\sum_{n\in S}|A(n)|^2\leqslant\sum_{n\leqslant N}|A(n)|^2\leqslant\Delta\sum_{p\leqslant Q}\sum_{a\,(\mathrm{mod}\,p)}|\beta(p,a)|^2.$$

For $n\in S$, the size of the amplifier is

$$|A(n)|^2=\Bigl|\sum_{p\leqslant Q}\sum_{a\,(\mathrm{mod}\,p)}\beta(p,a)e\Bigl(\frac{an}{p}\Bigr)\Bigr|^2=\Bigl|\sum_{p\leqslant Q}c_p\Bigr|^2=K^2,$$

by (5), while on the other hand, by applying the Parseval identity in $\mathbf{Z}/p\mathbf{Z}$, we get

$$\begin{aligned}\sum_{p\leqslant Q}\sum_{a\,(\mathrm{mod}\,p)}|\beta(p,a)|^2&=\sum_{p\leqslant Q}\Bigl(\frac{1}{p}\sum_{x\in\mathbf{Z}/p\mathbf{Z}}|\mathbf{1}_{\Omega_p}(x)|^2-\alpha(p,0)^2\Bigr)\\&=\sum_{p\leqslant Q}c_p(1-c_p)\leqslant K.\end{aligned}$$

So we obtain

$$K^2|S|\leqslant\Delta K,$$

i.e., exactly Rényi's inequality (4), by this technique.

To go further, we must exploit all the squarefree integers $q\leqslant Q$ (and not only the primes) to construct the amplifier. This is most easily described using the Chinese Remainder Theorem to write

$$\mathbf{Z}/q\mathbf{Z}\simeq\prod_{p\mid q}\mathbf{Z}/p\mathbf{Z},\qquad(\mathbf{Z}/q\mathbf{Z})^\times\simeq\prod_{p\mid q}(\mathbf{Z}/p\mathbf{Z})^\times,$$

and putting together the amplifiers modulo primes $p\mid q$: if $n\in S$ then $n\,(\mathrm{mod}\,p)\notin\Omega_p$ for all $p\mid q$, and hence multiplying out (5) over $p\mid q$, we find constants $\beta(q,a)\in\mathbf{C}$, defined for $(a,q)=1$ (because $\beta(p,a)$ is defined for a coprime with p), such that

$$\sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}\beta(q,a)e\Bigl(\frac{an}{q}\Bigr)=\prod_{p\mid q}c_p.$$

Moreover, because the product decomposition of the Chinese Remainder Theorem is compatible with the Hilbert space structure involved, we have

$$\sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}|\beta(q,a)|^2=\prod_{p\mod q}\sideset{}{^*}\sum_{a\,(\mathrm{mod}\,p)}|\beta(p,a)|^2=\prod_{p\mid q}c_p(1-c_p).$$

Arguing as before, we obtain from (3) – using all squarefree moduli $q\leqslant Q$ this time – that

$$|S|\leqslant\Delta\frac{A}{B^2},\tag{6}$$

with

$$A=\sideset{}{^\flat}\sum_{q\leqslant Q}\prod_{p\mid q}c_p(1-c_p),\qquad B=\sideset{}{^\flat}\sum_{q\leqslant Q}\prod_{p\mid q}c_p.$$

This is not quite (1), but we have some flexibility to choose another amplifier, namely, notice that this expression is not homogeneous if we multiply the coefficients $\beta(q,a)$ by scalars independent of a, and we can use this to find a better inequality. Precisely, let

$$\gamma(q,a)=\prod_{p\mid q}\gamma(p,a),\quad \text{with}\quad \gamma(p,a)=\lambda_p\beta(p,a).$$

Then we have the new amplification property

$$\sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}\gamma(q,a)e\Bigl(\frac{an}{q}\Bigr)=\prod_{p\mid q}\lambda_p c_p$$

with altered "cost" given by

$$\sideset{}{^*}\sum_{a\,(\mathrm{mod}\,q)}|\gamma(q,a)|^2=\prod_{p\mid q}\lambda_p^2 c_p(1-c_p),$$

so that, arguing as before, we get

$$|S|\leqslant\Delta\frac{A_1}{B_1^2}$$

with

$$A_1=\sideset{}{^\flat}\sum_{q\leqslant Q}\prod_{p\mid q}\lambda_p^2 c_p(1-c_p),\qquad B_1=\sideset{}{^\flat}\sum_{q\leqslant Q}\prod_{p\mid q}\lambda_p c_p.$$

If we wish to seem clever, we can select

$$\lambda_p=\frac{1}{1-c_p}=\frac{p}{p-|\Omega_p|},$$

as if by magic: since

$$c_p\lambda_p=\lambda_p^2 c_p(1-c_p)=\frac{|\Omega_p|}{p-|\Omega_p|},$$

this leads to $A_1=B_1=H$, hence to $|S|\leqslant\Delta H^{-1}$, which is (1). But this choice is not just a random one: the problem is to optimize a quadratic form (namely A_1) with a linear constraint given by B_1, and the choice above is the best one. This is checked, as usual, with Cauchy's inequality: writing

$$c_q=\prod_{p\mid q}c_p,\qquad \tilde{c}_q=\prod_{p\mid q}(1-c_p),\qquad \lambda_q=\prod_{p\mid q}\lambda_p$$

for ease of notation, we have

$$B_1^2=\Bigl(\sideset{}{^\flat}\sum_{q\leqslant Q}\lambda_q c_q\Bigr)^2\leqslant\Bigl(\sideset{}{^\flat}\sum_{q\leqslant Q}\lambda_q^2 c_q\tilde{c}_q\Bigr)\Bigl(\sideset{}{^\flat}\sum_{q\leqslant Q}\frac{c_q}{\tilde{c}_q}\Bigr)=A_1H,$$

with equality if λ_p is as given above, or in other words

$$\frac{A_1}{B_1^2}\geqslant\frac{1}{H},$$

with equality for this specific choice of λ_p.

Remark. (1) The last "optimization" step is reminiscent of the Selberg sieve. Indeed, it is well known that the Selberg sieve is related to the large sieve, and particularly with the dual inequality (3), as explained in [HR, p. 125]. Note however that the coefficients we optimize for, being of an "amplificatory" nature, and different from the coefficents λ_d typically sought for in Selberg's sieve, which are akin to the Möbius function and of a "mollificatory" nature.

(2) The argument does not use any particular feature of the classical sieve, and thus extends immediately to provide a proof of the general large sieve inequality of [K, Prop. 2.3] which is directly based on the dual inequality [K, Lemma 2.8]; readers interested in the formalism of [K] are encouraged to check this.

Example. What are the amplifiers above in some simple situations? In the case – maybe the most important – where we try to count primes, we then take $\Omega_p = \{0\}$ to detect integers free of small primes by sieving, and (5) becomes

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, p)} \Bigl(-\frac{1}{p}\Bigr) \cdot e\Bigl(\frac{an}{p}\Bigr) = \frac{1}{p},$$

if $p \nmid n$. Then, for q squarefree, the associated detector is the identity

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} \frac{\mu(q)}{q} e\Bigl(\frac{an}{q}\Bigr) = \frac{1}{q},$$

if $(n, q) = 1$, or in other words, it amounts to the well-known formula

$$\sideset{}{^*}\sum_{a \,(\mathrm{mod}\, q)} e\Bigl(\frac{an}{q}\Bigr) = \mu(q)$$

for the values of a Ramanujan sum with coprime arguments. Note that in this case, the optimization process above replaced $c_p = \frac{1}{p}$ with

$$\lambda_p c_p = \frac{1}{p-1},$$

which is not a very big change – and indeed, for small sieves, the bound (6) is not far from (1), and remains of the right order of magnitude.

On the other hand, for an example in a large sieve situation, we can take Ω_p to be the set of squares in $\mathbf{Z}/p\mathbf{Z}$. The characteristic function (for odd p) is

$$\mathbf{1}_{\Omega_p}(x) = \sum_{a \,(\mathrm{mod}\, p)} \tau(p, a) e\Bigl(\frac{ax}{p}\Bigr)$$

with coefficients given – essentially – by Gauss sums

$$\tau(p, a) = \frac{1}{p}\Bigl(1 + \frac{1}{2} \sideset{}{^*}\sum_{x \,(\mathrm{mod}\, p)} e\Bigl(\frac{ax^2}{p}\Bigr)\Bigr).$$

Then c_p tends to $1/2$ as $p \to +\infty$, while $\lambda_p c_p$ tends to 1. This difference leads to a discrepancy in the order of magnitude of the final estimate: using standard results on bounds for sums of multiplicative functions, (6) and taking $Q = \sqrt{N}$, we get

$$|S| \ll \sqrt{N} (\log N)^{1/4},$$

instead of $|S| \ll \sqrt{N}$ that follows from (1).

References

[HR] H. Halberstam and H.E. Richert: *Sieve methods*, London Math. Soc. Monograph, Academic Press (London), 1974

[IK] H. Iwaniec and E. Kowalski: *Analytic number theory*, AMS Colloquium Publ. 53, 2004.

[K] E. Kowalski: *The large sieve and its applications: arithmetic geometry, random walks and discrete groups*, Cambridge Tracts in Math. 175, 2008.

[M1] H-L. Montgomery: *Topics in multiplicative number theory*, Lecture Notes Math. 227, Springer-Verlag 1971.

[M2] H-L. Montgomery: *The analytic principle of the large sieve*, Bull. A.M.S. 84 (1978), 547–567.