LECTURE NOTES: Sieves I[1]

Alex Kontorovich

1. Introduction

The first sieving procedure for producing primes is credited to Eratosthenes (∼200 BCE), who made a simple but important observation: if $n < x$ and n has no prime factors $< x^{1/2}$, then n is prime. So to make a table of the primes up to 100, one needs only to strike out (sieve) numbers divisible by 2, 3, 5, and 7 (the primes below $10 = \sqrt{100}$). A very slight generalization of the above is that: if $n < x$ has no prime factors $< x^{1/(k+1)}$ then n is a product of at most k primes. We call such a number a k-almost prime, or k-AP from now on.

The goal of this lecture is to prove

Theorem 1.1 (Brun 1919 [Bru19])**.**

$$\sum_{p \text{ a twin prime}} \frac{1}{p} < \infty.$$

Brun's beautiful theorem is actually a *negative* result. Recall that Euler (1737) proved that there are infinitely many primes by showing that $\sum 1/p$ diverges (and in the process giving birth to analytic number theory). The infinity of primes was of course not new, but a century later Dirichlet (1837) was able to prove that the sum of reciprocals of primes in (admissible) arithmetic progressions also diverges, thereby settling the longstanding problem that their cardinality is infinite. So one might wonder whether one can prove the infinity of twin primes this way; Brun's theorem says: "No."

Theorem 1.1 is an immediate consequence of integration by parts and

Theorem 1.2.

$$\pi_2(x) := \#\{p < x : p + 2 \text{ is prime}\} \ll \frac{x}{(\log x)^2}.$$

Brun also showed that $\#\{n < x : n, n+2 \text{ are 9-APs}\} \gg \frac{x}{(\log x)^2}$, an impressive approximation to the Twin Prime Conjecture. The "9" was gradually reduced to Chen's famous theorem [Che73] that there are infinitely many primes p so that $p + 2$ is 2-AP. Selberg explained the parity barrier: sieves alone cannot reduce the "2" to a "1", because they cannot distinguish numbers with an even or odd number of factors. But Vinogradov, in (essentially) resolving the ternary Goldbach problem, showed how extra ingredients (bilinear forms) can be used to produce actual primes. We will return to this point in a later lecture.

Brun's method is combinatorial, and quite delicate and technical. Selberg later gave a treatment, the so-called Upper Bound (or Λ^2) Sieve, which is extremely beautiful and elegant (but does not in general give lower bounds on almost primes); this is the direction we pursue here. While working on this specific problem, we'll introduce notation to hint at how this sieve works in a more general setting.

[1] September 19, 2011

2. The Sieve

Let $\mathcal{A} = \{a_n\}$ be a sequence of non-negative numbers, in our case

$$a_n := \begin{cases} 1 & \text{if } n = m(m+2) \\ 0 & \text{otherwise} \end{cases}$$

is an indicator of whether n is a value of the polynomial $f(m) = m(m+2)$. We will need to be able to tell when a number has no prime factors less than some parameter z, say, so we introduce the product

$$P_z := \prod_{p<z} p.$$

So n has no prime factors $< z$ if $(n, P_z) = 1$. We introduce the function

$$S(x, z) := \sum_{\substack{n<x(x+2) \\ (n,P_z)=1}} a_n.$$

If $p < x$ is a twin prime, then either $p < x^{1/3}$, or $n = p(p+2)$ has no prime factors $< x^{1/3}$; hence

$$\pi_2(x) \le x^{1/3} + S(x, x^{1/3}),$$

so it remains to bound $S(x, z)$ with $z = x^{1/3}$.

We need the following standard

Lemma 2.1 (Möbius inversion)**.**

$$g = f * 1 \iff f = g * \mu.$$

Here $*$ denotes Dirichlet convolution, so

$$g(n) = \sum_{d|n} f(d)$$

iff

$$f(n) = \sum_{d|n} g(d)\mu(n/d).$$

Proof. Recall the Möbius function

$$\mu(n) = \begin{cases} 0 & \text{if } p^2 \mid n \text{ for some prime } p \\ (-1)^k & \text{if } n = p_1 p_2 \cdots p_k \text{ is squarefree} \end{cases}$$

which arises naturally as the Dirichlet coefficient of $1/\zeta(s)$. From the Euler product formula

$$\zeta(s) = \sum_n \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1},$$

we have

$$\frac{1}{\zeta(s)} = \prod_p \left(1 - \frac{1}{p^s}\right),$$

which, after multiplying out and collecting terms, gives the Dirichlet series

$$\frac{1}{\zeta(s)} = \sum_n \frac{\mu(n)}{n^s}.$$

So if

$$F(s) = \sum_n \frac{f(n)}{n^s},$$

then

$$\zeta(s)F(s) = \left(\sum_m \frac{1}{m^s}\right)\left(\sum_k \frac{f(k)}{k^s}\right) = \sum_n \frac{1}{n^s}\left(\sum_{mk=n} f(k)\cdot 1\right) = \sum_n \frac{g(n)}{n^s} =: G(s),$$

say. Then of course

$$F(s) = \frac{1}{\zeta(s)}G(s) = \left(\sum_m \frac{\mu(m)}{m^s}\right)\left(\sum_k \frac{g(k)}{k^s}\right) = \sum_n \frac{1}{n^s}\left(\sum_{mk=n} g(k)\cdot\mu(m)\right).$$

Comparing Dirichlet coefficients proves the claim. □

In particular, we have that $\zeta(s)\cdot\frac{1}{\zeta(s)} = 1$, so

$$\sum_{d|n}\mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{2.2}$$

giving us a nice indicator function.

Returning to $S(x,z)$, we use the above trick to indicate the condition $(n, P_z) = 1$, that is,

$$S(x,z) = \sum_{n<x(x+2)} a_n \left(\sum_{d|(n,P_z)} \mu(d)\right). \tag{2.3}$$

Reversing orders gives the formula of Legendre (1808):

$$S(x,z) = \sum_{d|P_z}\mu(d)\sum_{\substack{n<x(x+2)\\ n\equiv 0(\operatorname{mod} d)}} a_n = \sum_{d|P_z}\mu(d)|\mathcal{A}_d|.$$

Here it's obvious that Möbius is playing the role of inclusion-exclusion, and $|\mathcal{A}_d|$ measures the local distribution of our sequence.

2.1. **Interlude: PNT?.** We interrupt our train of thought momentarily to see if we can prove the Prime Number Theorem (PNT) this way. Here we would take $a_n \equiv 1$, and

$$S(x,z) = \sum_{\substack{n<x\\ (n,P_z)=1}} a_n = \sum_{d|P_z}\mu(d)|\mathcal{A}_d|.$$

with $z = x^{1/2}$. Now

$$|\mathcal{A}_d| = \#\{n < x : n \equiv 0(\operatorname{mod} d)\} = \left\lfloor\frac{x}{d}\right\rfloor = \frac{x}{d} + r_d \tag{2.4}$$

with $|r_d| < 1$. So we can write the above as

$$S(x, z) = x \sum_{d|P_z} \frac{\mu(d)}{d} + Small$$

Considering just the main term, we from multiplicativity see that

$$\sum_{d|P_z} \frac{\mu(d)}{d} = \prod_{p<z} \left(1 + \frac{\mu(p)}{p}\right) = \prod_{p<z} \left(1 - \frac{1}{p}\right).$$

Mertens (1874) studied the above product (the "rate" at which $\zeta(1)$ blows up), and showed that

$$\prod_{p<z} \left(1 - \frac{1}{p}\right)^{-1} \sim e^{\gamma} \log z. \tag{2.5}$$

Here γ is the Euler-Mascheroni constant,

$$\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} - \log n\right) \to \gamma \approx 0.577,$$

as $n \to \infty$. So we have

$$S(x, z) \sim x \left(\frac{e^{-\gamma}}{\log z}\right) = 2e^{-\gamma} \left(\frac{x}{\log x}\right),$$

for $z = x^{1/2}$. Note that $2e^{-\gamma} \approx 1.12$. But the PNT says this constant is 1! So we're off by a constant in the main term, meaning the error term is at least of the same order of magnitude as the main...

In fact it's much worse than that; if we analyze the leftover term we called "*Small*" above using just that $|r_d| < 1$, we get a bound for it roughly of size $O(2^x)$, so it totally swamps the main term. The moral is that this silly little floor function arising in (2.4) is actually very difficult to understand well, since x is very small relative to some large d's dividing P_z. (Actually this discrepancy can be used to great advantage, see e.g. [Mai85, FG89].) This concludes our interlude.

2.2. **Back to Twin Primes.** Returning to (2.3), Brun knew that stopping inclusion-exclusion at an even index is an under-count, whereas stopping at an odd index is an overestimate. Keeping very careful track of the resulting sums and making very judicious choices of cut-off parameters, he was able to produce his combinatorial sieve. As already stated, we will not take this approach (for which see, e.g. [IK04, §6.1-6.4]).

Instead, we back up to (2.3) and use Selberg's ingenious observation: if λ_d is any sequence of numbers with

$$\lambda(1) = 1, \tag{2.6}$$

then

$$\sum_{d|k} \mu(d) \leq \left(\sum_{d|k} \lambda(d)\right)^2.$$

At first glance, this looks just plain wrong. Once one has convinced oneself of its correctness, it looks completely useless and vacuous. Moreover, in a moment, it will appear that it's actually making things worse – the sieve level (to be discussed later) seems to be going up significantly. Nevertheless, it has the advantage that we have forgotten the difficult and random behavior of Möbius, and replaced it with an (almost) arbitrary sequence, the price being only a square. Putting this identity into (2.3) gives

$$S(x,z) \leq \sum_{n<x(x+2)} a_n \left(\sum_{d|(n,P_z)} \lambda(d) \right)^2 = \sum_{d_1|P_z} \sum_{d_2|P_z} \lambda(d_1)\lambda(d_2) \sum_{\substack{n<x(x+2) \\ n\equiv 0(\mathrm{mod}\,[d_1,d_2])}} a_n. \tag{2.7}$$

(Recall $[a,b]$ is the least common multiple of a and b.) We see easily that the function

$$\sum_{\substack{n<x(x+2) \\ n\equiv 0(\mathrm{mod}\,k)}} a_n = |\mathcal{A}_k|$$

is multiplicative in k. For $k=2$ takes the value

$$\begin{aligned} |\mathcal{A}_2| &= \#\{n < x(x+2) : n = m(m+2) \text{ for some } m \text{ and } n \equiv 0(\mathrm{mod}\,2)\} \\ &= \#\{m < x : m \equiv 0(\mathrm{mod}\,2)\} \\ &= \frac{x}{2} + r_2, \end{aligned}$$

with $|r_2| < 1$. For $k = p > 2$, the condition $m(m+2) \equiv 0(\mathrm{mod}\,p)$ means that either $m \equiv 0(\mathrm{mod}\,p)$ or $m \equiv -2(\mathrm{mod}\,p)$, so

$$|\mathcal{A}_p| = \frac{2}{p}x + r_p,$$

with $|r_p| < 2$. We can combine the above calculations into one by introducing a multiplicative function[2] w supported on squarefree numbers and defined on the primes by

$$w(p) := \begin{cases} 1 & \text{if } p = 2 \\ 2 & \text{if } 2 < p < z \\ 0 & \text{if } p > z, \end{cases}$$

so that

$$|\mathcal{A}_k| = \frac{w(k)}{k}x + r_k, \tag{2.8}$$

with

$$|r_k| \leq w(k). \tag{2.9}$$

Putting (2.8) into (2.7) gives

$$\begin{aligned} S(x,z) &\leq x \sum_{d_1|P_z} \sum_{d_2|P_z} \lambda(d_1)\lambda(d_2) \frac{w([d_1,d_2])}{[d_1,d_2]} + \sum_{d_1|P_z} \sum_{d_2|P_z} \lambda(d_1)\lambda(d_2) r_{[d_1,d_2]} \\ &= x \cdot Q + E, \end{aligned} \tag{2.10}$$

where

$$Q = \sum_{d_1|P_z} \sum_{d_2|P_z} \lambda(d_1)\lambda(d_2) \frac{w([d_1,d_2])}{[d_1,d_2]} \tag{2.11}$$

is a quadratic form in the λ's, and

$$E = \sum_{d_1|P_z} \sum_{d_2|P_z} \lambda(d_1)\lambda(d_2) r_{[d_1,d_2]} \tag{2.12}$$

[2] Recall f is multiplicative if whenever $(n,m)=1$, we have $f(nm) = f(n)f(m)$.

is an "error" term. (We'll be more careful now that we were burned on the PNT heuristic.)

Now we have an (in principle) simple task: choose the λ's so as to minimize Q, subject to the constraint (2.6). The reader is invited at this point to put down the notes and try to carry out the rest of the calculation herself. The punchline: Q can be diagonalized, and the λ's chosen optimally so that, via Mertens' Theorem,

$$Q \ll (\log z)^{-2} \ll (\log x)^{-2}.$$

Here we used $z = x^{1/3}$; in fact we could have taken $z = x^{1/2-\varepsilon}$ or $z = x^{1/1000}$, but note that taking $z = (\log x)^{1000}$ is insufficient to show Theorem 1.2.

2.3. **Evaluating Q.** To continue, here's an observation: what burned us before was that we introduced an extremely bad error in passing from $\lfloor x/d \rfloor$ to x/d, since $d \mid P_z$ can get very large. So here's a trick: since we have the freedom to choose λ, we'll insist that λ is supported on only small numbers:

$$\lambda(n) = 0, \text{ if } n > z. \tag{2.13}$$

(We could have given ourselves more flexibility above by introducing another parameter for the support of λ; in the end it will turn out that the already in-play parameter z is a good choice.)

Now we return to (2.11) and massage Q into diagonal form. For $d_1, d_2 \mid P_z$, write $d_1 = ac$ and $d_2 = bc$, where $c = (d_1, d_2)$ and $(a, b) = 1$. Then $[d_1, d_2] = abc$, and we have

$$\begin{aligned} Q &= \sum_{d_1 \mid P_z} \sum_{d_2 \mid P_z} \lambda(d_1)\lambda(d_2) \frac{w([d_1, d_2])}{[d_1, d_2]} \\ &= \sum_a \sum_b \sum_c \lambda(ac)\lambda(bc) \frac{w(abc)}{abc}. \end{aligned}$$

From the support of λ and w, the variables a, b, c range over divisors of P_z and are moreover forced to be coprime. Next we use multiplicativity of w to write $w(abc) = w(a)w(b)w(c)$. The λ's continue to guarantee that $(a, c) = (b, c) = 1$, but now we need to remember that $(a, b) = 1$, which we do with another application of (2.2):

$$\begin{aligned} Q &= \sum_c \frac{w(c)}{c} \sum_a \sum_{\substack{b \\ (a,b)=1}} \lambda(ac)\lambda(bc) \frac{w(a)}{a} \frac{w(b)}{b} \\ &= \sum_c \frac{w(c)}{c} \sum_a \sum_b \lambda(ac)\lambda(bc) \frac{w(a)}{a} \frac{w(b)}{b} \sum_{d \mid (a,b)} \mu(d) \\ &= \sum_{d \mid P_z} \mu(d) \sum_c \frac{w(c)}{c} \sum_{a \equiv 0 (\bmod d)} \sum_{b \equiv 0 (\bmod d)} \lambda(ac)\lambda(bc) \frac{w(a)}{a} \frac{w(b)}{b} \\ &= \sum_{d \mid P_z} \mu(d) \sum_c \frac{w(c)}{c} \left(\sum_{a \equiv 0 (\bmod d)} \lambda(ac) \frac{w(a)}{a} \right)^2. \end{aligned}$$

We have essentially diagonalized Q, but will continue to massage it into a slightly more palatable form. Pull the $w(c)/c$ term inside the a sum:

$$\begin{aligned} Q &= \sum_{d|P_z} \mu(d) \sum_{c} \frac{c}{w(c)} \left(\sum_{a\equiv 0 (\mathrm{mod}\, d)} \lambda(ac) \frac{w(ac)}{ac} \right)^2 \\ &= \sum_{d|P_z} \mu(d) \sum_{c} \frac{c}{w(c)} \left(\sum_{a\equiv 0 (\mathrm{mod}\, cd)} \lambda(a) \frac{w(a)}{a} \right)^2 . \end{aligned}$$

We are almost done; write $\ell = cd$, then

$$Q = \sum_{\ell} \left(\sum_{c|\ell} \frac{c}{w(c)} \mu(\ell/c) \right) \left(\sum_{a\equiv 0 (\mathrm{mod}\, \ell)} \lambda(a) \frac{w(a)}{a} \right)^2 .$$

Let

$$y(\ell) := \sum_{a\equiv 0 (\mathrm{mod}\, \ell)} \lambda(a) \frac{w(a)}{a} \tag{2.14}$$

be a linear change of variables from the λ's, and define

$$h(\ell) := \sum_{c|\ell} \frac{c}{w(c)} \mu(\ell/c), \tag{2.15}$$

so that

$$Q = \sum_{\ell} h(\ell) y(\ell)^2. \tag{2.16}$$

Now we just need to minimize Q with respect to the y's, but first we need to reformulate the constraint (2.6). To this end, we need the following funny-looking version of Möbius inversion:

Lemma 2.17.

$$g(n) = \sum_{a\equiv 0 (\mathrm{mod}\, n)} f(a)$$

iff

$$f(n) = \mu(n) \sum_{a\equiv 0 (\mathrm{mod}\, n)} \mu(a) g(a).$$

(As always we assume both functions are supported on square-free numbers.)

Proof. Plugging the first equation into the right-hand side of the second gives

$$\mu(n) \sum_{a\equiv 0 (\mathrm{mod}\, n)} \mu(a) g(a) = \mu(n) \sum_{a\equiv 0 (\mathrm{mod}\, n)} \mu(a) \sum_{b\equiv 0 (\mathrm{mod}\, a)} f(b).$$

If $a \equiv 0 (\operatorname{mod} n)$ then $a = nk$, similarly $b = a\ell = k\ell n$. Let $m = k\ell$. Then interchanging summations gives

$$\begin{aligned} &= \mu(n) \sum_m f(mn) \sum_{k|m} \mu(nk) \\ &= \mu(n)^2 \sum_m f(mn) \sum_{k|m} \mu(k) \\ &= f(n), \end{aligned}$$

where we used that the support of f is on square-free numbers. The other direction is proved similarly. □

Applied to (2.14), we invert y's to λ's:

$$\lambda(n) = \frac{n}{w(n)} \mu(n) \sum_{a \equiv 0 (\operatorname{mod} n)} \mu(a) y(a). \tag{2.18}$$

Note that y is supported on $n < z$ iff λ is. From (2.18), we can now reformulate (2.6) in terms of y's:

$$1 = \lambda(1) = \sum_a \mu(a) y(a).$$

Now it's completely elementary to minimize the diagonal quadratic form (2.16) subject to the above linear constraint. We get that the optimal choice for y is

$$y(n) = \frac{1}{H} \frac{\mu(n)}{h(n)},$$

where

$$H = \sum_n \frac{1}{h(n)}.$$

With this choice, we have

$$Q = \frac{1}{H},$$

and

$$\lambda(n) = \frac{1}{H} \frac{n}{w(n)} \mu(n) \sum_{a \equiv 0 (\operatorname{mod} n)} \frac{1}{h(a)}. \tag{2.19}$$

It remains to evaluate H, which we can do explicitly. Since h is multiplicative, we have

$$H = \sum_n \frac{1}{h(n)} = \prod_p \left(1 + \frac{1}{h(p)}\right) \tag{2.20}$$

From (2.15), we evaluate $h(p)$ explicitly:

$$h(p) = -1 + \frac{p}{w(p)} = \frac{p - w(p)}{w(p)}.$$

So we have

$$\frac{1}{H} = \prod_p \left(1 - \frac{w(p)}{p}\right) = \frac{1}{2} \prod_{3 \le p < z} \left(1 - \frac{2}{p}\right). \tag{2.21}$$

Note that

$$1 - \frac{2}{p} < 1 - \frac{2}{p} + \frac{1}{p^2} = \left(1 - \frac{1}{p}\right)^2.$$

So applying Mertens' Theorem (2.5) gives

$$Q = \frac{1}{H} < \frac{1}{2} \prod_{3 \le p < z} \left(1 - \frac{1}{p}\right)^2 \ll \frac{1}{(\log z)^2},$$

as claimed. It remains to control the error term.

2.4. **Estimating** E**.** Returning to (2.12), we first need to bound λ's, as follows. Putting absolute values into (2.19), using positivity of h, H, and w, and extending the range of a gives:

$$|\lambda(n)| \le \frac{1}{H} \frac{n}{w(n)} \sum_{a \equiv 0 \pmod n} \frac{1}{h(a)} \le \frac{1}{H} \frac{1}{h(n)} \frac{n}{w(n)} \sum_a \frac{1}{h(a)} = \frac{n}{w(n)h(n)}.$$

Recall also that λ and y are supported on $n < z$. Putting the above into (2.12), together with the bound (2.9) on r's, gives:

$$\begin{aligned}
|E| &\le \sum_{\substack{d_1 | P_z \\ d_1 < z}} \sum_{\substack{d_2 | P_z \\ d_2 < z}} |\lambda(d_1)||\lambda(d_2)| w([d_1, d_2]) \\
&\le \left(\sum_{\substack{d | P_z \\ d < z}} |\lambda(d)| w(d) \right)^2 \\
&\le \\
&\le \left(\sum_{\substack{d | P_z \\ d < z}} \frac{d}{w(d)h(d)} w(d) \right)^2 \\
&< z^2 \left(\sum_d \frac{1}{h(d)} \right)^2 = (zH)^2,
\end{aligned}$$

where we again used (2.20).

Returning to (2.10), note that we have proved

$$S(x, z) < x\frac{1}{H} + (zH)^2.$$

We already have lower bounds for H (via upper bounds for $1/H$), now we need upper bounds. This is again easy: note that

$$\left(1 - \frac{1}{p}\right)^2 \left(1 - \frac{2}{p^2}\right) = 1 - \frac{2}{p} - \frac{1}{p^2}\left(1 - \frac{4}{p} + \frac{2}{p^2}\right) < 1 - \frac{2}{p},$$

for $p > 100$ or so. (In general something like $(1 - k/p) \asymp (1 - 1/p)^k$ holds.) So returning to (2.21), we have

$$H = 2 \prod_{3 \le p < z} \left(1 - \frac{2}{p}\right)^{-1} \ll \prod_{p < z} \left(1 - \frac{1}{p}\right)^{-2} \prod_p \left(1 - \frac{2}{p^2}\right)^{-1} \ll (\log z)^2,$$

since the second product converges.

In conclusion, we've shown that

$$S(x, z) \ll \frac{x}{(\log z)^2} + z^2(\log z)^4.$$

This gives us what we want for any value of z in the range $x^\varepsilon < z < x^{1/2-\varepsilon}$.

References

[Bru19] V. Brun. Le crible d'Eratosthéne et le theoréme de Goldbach. *C. R. Acad. Sci. Paris*, 168:544–546, 1919. 1

[Che73] Jing Run Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica*, 16:157–176, 1973. 1

[FG89] John Friedlander and Andrew Granville. Limitations to the equi-distribution of primes. I. *Ann. of Math. (2)*, 129(2):363–382, 1989. 4

[IK04] Henryk Iwaniec and Emmanuel Kowalski. *Analytic Number Theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004. 4

[Mai85] Helmut Maier. Primes in short intervals. *Michigan Math. J.*, 32(2):221–225, 1985. 4