18.785: Analytic Number Theory, MIT, spring 2007 (K.S. Kedlaya)
The Selberg sieve

1 Review of notation

Let $f : \mathbb{N} \to \mathbb{C}$ be an arithmetic function, and suppose we want to estimate the sum of f over primes. More precisely, let P be a set of primes, and put

$$P(z) = \prod_{p \leq z, p \in P} p.$$

If we define

$$S(x,z) = \sum_{n \leq x, (n, P(z)) = 1} f(n),$$
$$A_d(x) = \sum_{n \leq x, n \equiv 0 \pmod{d}} f(n)$$

(with the dependence on P and f suppressed from the notation), we have

$$S(x,z) = \sum_{d | P(z)} \mu(d) A_d(x).$$

Let $g(d)$ be a multiplicative function with

$$g(p) \in [0,1) \qquad (p \in P);$$
$$g(p) = 0 \qquad (p \notin P),$$

and write

$$A_d(x) = g(d)x + r_d(x).$$

Then

$$S(x,z) = V(z)x + R(x,z)$$
$$V(z) = \prod_{p | P(z)} (1 - g(p))$$
$$R(x,z) = \sum_{d | P(z)} r_d(x).$$

2 The Selberg upper bound sieve

In the previous unit, we used the combinatorial sieve to construct an arithmetic function $\lambda^+ : \mathbb{N} \to \mathbb{R}$ such that

$$\lambda^+(1) = 1$$
$$\sum_{d|n} \lambda^+(d) \geq 0 \quad (n > 1).$$

By setting

$$V^+(z) = \sum_{d|P(z)} \lambda^+(d) g(d)$$
$$R^+(x,z) = \sum_{d|P(z)} \lambda^+(d) r_d(x),$$

we were able to obtain the bound

$$V^-(z)x + R^-(x,z) \le S(x,z) \le V^+(z)x + R^+(x,z), \tag{1}$$

but controlling V^+ and R^+ was rather painful.

Selberg proposed instead to construct an arithmetic function $\rho : \mathbb{N} \to \mathbb{R}$ with $\rho(1) = 1$ and

$$\sum_{d|n} \lambda^+(n) = \left(\sum_{d|n} \rho(d) \right)^2 .$$

In other words, let ρ be any arithmetic function with $\rho(1) = 1$, and put

$$\lambda^+(n) = \sum_{d_1,d_2 : \operatorname{lcm}(d_1,d_2)=n} \rho(d_1)\rho(d_2).$$

We will typically want $\lambda^+(d) = 0$ for $d \ge y$, for some prespecified number y; to enforce this, we may insist that $\rho(n) = 0$ for $n \ge \sqrt{y}$. We call the resulting λ^+ an *L^2-sieve of level* y, or more commonly a *Selberg (upper bound) sieve of level* y.

Let us drop x from consideration by agreeing to only consider functions f with finite support. (That is, we replace f by the function vanishing above x.) If we again set

$$S(z) = \sum_{(n,P(z))=1} f(n)$$
$$V^+(z) = \sum_{d|P(z)} \lambda^+(d) g(d)$$
$$= \sum_{d_1,d_2|P(z)} \rho(d_1)\rho(d_2) g(\operatorname{lcm}(d_1,d_2))$$

$$R^+(z) = \sum_{d|P(z)} \lambda^+(d) r_d(x)$$
$$= \sum_{d_1,d_2|P(z)} \rho(d_1)\rho(d_2) r_{\operatorname{lcm}(d_1,d_2)}(x),$$

we again have

$$S(z) \le V^+(z)x + R^+(z). \tag{2}$$

Ignoring the error term $R^+(z)$ for the moment, one can ask about optimizing the main term $V^+(z)x$ in the bound (2). This amounts to viewing $V^+(z)$ as a quadratic form and then minimizing it.

For simplicity, we will assume that $g(p) \in (0,1)$ for $p \in P$, and $g(p) = 0$ for $p \notin P$. (Before we only wanted $g(p) \in [0,1)$ for $p \in P$, but there is no harm in adding to P those primes p for which $g(p) = 0$ into P.) Let h be a multiplicative function with

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

We can then diagonalize the quadratic form as follows: first, put $c = \gcd(d_1, d_2)$, $a = d_1/c$, $b = d_2/c$ to obtain

$$\begin{aligned} V^+(z) &= \sum_{a,b,c:abc|P(z)} \rho(ac)\rho(bc)g(abc) \\ &= \sum_{c|P(z)} g(c)^{-1} \sum_{a,b:abc|P(z)} (g(ac)\rho(ac))(g(bc)\rho(bc)). \end{aligned}$$

Note that since $P(z)$ is squarefree, the condition $abc|P(z)$ forces $\gcd(a,b) = 1$. We now perform inclusion-exclusion on $\gcd(a,b)$ to obtain

$$\begin{aligned} V^+(z) &= \sum_{c|P(z)} g(c)^{-1} \sum_{d|P(z)/c} \mu(d) \left(\sum_{m|P(z)/(cd)} g(cdm)\rho(cdm) \right)^2 \\ &= \sum_{c|P(z)} g(c)^{-1} \sum_{d|P(z)/c} \mu(d) \left(\sum_{m|P(z):m\equiv 0\,(cd)} g(m)\rho(m) \right)^2. \end{aligned}$$

We next substitute $e, f/e$ in for c, d, and reorder the sum:

$$\begin{aligned} V^+(z) &= \sum_{f|P(z)} \sum_{e|f} \mu(f/e) g(e)^{-1} \left(\sum_{m|P(z):m\equiv 0\,(f)} g(m)\rho(m) \right)^2 \\ &= \sum_{f|P(z)} h(f)^{-1} \left(\sum_{m|P(z):m\equiv 0\,(f)} g(m)\rho(m) \right)^2. \end{aligned}$$

Let's put

$$\xi(d) = \mu(d) \sum_{m|P(z):m\equiv 0\,(d)} g(m)\rho(m),$$

so that we have

$$V^+(z) = \sum_{d|P(z)} h(d)^{-1} \xi(d)^2.$$

Before we can minimize this quadratic form, we must first reexpress in terms of ξ the conditions we imposed on ρ. Namely, by Möbius inversion,

$$\rho(n) = \frac{\mu(n)}{g(n)} \sum_{d|P(z):d\equiv 0\,(n)} \xi(d),$$

so the condition $\rho(1) = 1$ is equivalent to

$$\sum_{d|P(z)} \xi(d) = 1,$$

and the condition $\rho(d) = 0$ for $d \geq \sqrt{y}$ is equivalent to

$$\xi(d) = 0 \qquad (d \geq \sqrt{y}).$$

That is, ξ is restricted to a hyperplane.

Here's where the L^2 part comes in. By the Cauchy-Schwartz inequality,

$$V^+(z) \geq H^{-1}, \qquad H = \sum_{d<\sqrt{y},d|P(z)} h(d)$$

and equality holds for

$$\xi(d) = h(d)H^{-1} \qquad (d < \sqrt{y}).$$

Backing up, we get

$$\rho(d) = \mu(d)\frac{h(d)}{g(d)}H^{-1} \sum_{n<\sqrt{y}/d:\gcd(d,n)=1} h(n).$$

Putting this together, we obtain the following.

Theorem 1 (Selberg)**.** *Let $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be an arithmetic function with finite support. Let P be a set of primes, and put $P(z) = \prod_{p\leq z,p\in P} p$. For $d|P(z)$, write*

$$A_d = \sum_{n\equiv 0\,(d)} f(n) = g(d)X + r_d(z)$$

for $X > 0$ and g a multiplicative function with $0 < g(p) < 1$ for all $p \in P$. Let $h(d)$ be a multiplicative function with $h(p) = g(p)(1 - g(p))^{-1}$ for all $p \in P$, and put

$$H = \sum_{d<\sqrt{y},d|P(z)} h(d)$$

for some $y > 1$. Then

$$S(z) = \sum_{(n,P(z))=1} f(n) \leq XH^{-1} + \sum_{d|P(z)} \lambda^+(d)r_d(z), \tag{3}$$

for

$$\lambda^+(n) = \sum_{d_1,d_2:\mathrm{lcm}(d_1,d_2)=n} \rho(d_1)\rho(d_2)$$

$$\rho(d) = \mu(d)\frac{h(d)}{g(d)}H^{-1} \sum_{n<\sqrt{y}/d:\gcd(d,n)=1} h(n).$$

As a somewhat miraculous corollary (due to van Lint and Richert), we obtain

$$0 \le \mu(d)\rho(d) \le 1 \tag{4}$$

(exercise); this makes it easy to estimate the error term in (3), e.g., by

$$|\lambda^+(d)| \le d^{(\log 3)/(\log 2)} \tag{5}$$

(exercise).

Exercises

1. Prove (4). (Hint: group terms in the definition of H according to the common divisor of d with some fixed number e.)
2. Deduce (5) from (4), by proving that $|\lambda^+(d)| \le 3^{\nu(d)}$, for $\nu(d)$ equal to the number of prime factors of d.
3. In the Selberg sieve, prove that if we extend g to a completely multiplicative function, then
$$H \ge \sum_{n<\sqrt{y}} g(n).$$
4. Prove that for some $c > 0$,
$$\sum_{n\le x} \frac{2^{\nu(n)}}{n} \ge c\log^2 x \qquad (x \ge 1).$$
(Hint: an elementary proof is possible, but one can also use analytic arguments on the Dirichlet series $\zeta^2(s)/\zeta(2s) = \sum_{n=1}^{\infty} 2^{\nu(n)} n^{-s}$.)
5. Let $d(n)$ denote the number of divisors of the positive integer n. Prove that
$$\sum_{n\le x} d(n) \sim x\log x.$$
(This is needed for the next problem.)

6. Use the Selberg sieve to prove that the number of twin primes $p \leq x$ is $O(x/\log^2 x)$. (Hint: put $f(n) = 1$ if $n = m(m+2)$ for some m and $f(n) = 0$ otherwise, then apply the Selberg sieve with $z = x^{1/4}$. You may need some of the earlier exercises as well.)

7. (Brun-Titchmarsh theorem) Prove that for any $\epsilon > 0$, there exists $x_0 = x_0(\epsilon)$ with the following property: for any positive integers m, N with $\gcd(m, N) = 1$, and any $x \geq \max\{N, x_0(\epsilon)\}$, the number of primes $p \leq x$ with $p \equiv m \pmod{N}$ is at most

$$\frac{(2+\epsilon)x}{\phi(N)\log(2x/N)}.$$

This is one of several problems in which the Selberg sieve applies to give you a result which is off by a factor of 2 from the expected best result.

8. Prove that

$$\sum_{n \leq x} \frac{n}{\phi(n)} = O(x),$$

then deduce by partial summation that

$$\sum_{n \leq x} \frac{1}{\phi(n)} = O(\log x),$$

(Hint: first prove that the sum $\sum_n 1/(n\gamma(n))$ converges, where $\gamma(n) = \prod_{p|n} p$.)

9. Use the previous two exercises to deduce that

$$\sum_{p \leq x} d(p-1) = O(x),$$

where $d(n)$ denotes the number of divisors of n.