AN INTRODUCTION TO THE SELBERG SIEVE

STEVE LESTER

1. Introduction: Bounding the number of primes

In the next lecture we will give applications of Selberg's Λ^2-upper bound sieve. In particular we will use the sieve to give an upper bound for the number of twin primes less than x. We will see that

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

One may compare this to the conjecture of Hardy and Littlewood that

$$\pi_2(x) \sim C\frac{x}{(\log x)^2}$$

where $C > 0$ is the twin prime constant and is given by

$$C = 2\prod_{p \neq 2} \frac{\left(1-\frac{2}{p}\right)}{\left(1-\frac{1}{p}\right)^2}.$$

Before moving on to these more interesting applications our goal is to give a simpler and more straightforward application, illustrating the power of the Selberg Λ^2-sieve. Our aim is to show

$$\pi(x) \ll \frac{x}{\log x}.$$

Recall the following from Zeev's lecture:

- Let $P(z) = \prod_{p \leq z} p$ and

  $$\mathcal{S}(x,z) = \#\{n \leq x : \gcd(n, P(z)) = 1\}.$$

  Then

  $$\pi(x) \leq \mathcal{S}(x,z) + z.$$

- For real numbers λ_d with $\lambda_1 = 1$ and $\lambda_d = 0$ for $d > z$ we have

  $$\mathcal{S}(x,z) \leq xQ(\lambda) + R(z) \tag{1}$$

  where

  $$Q(\lambda) = \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P(z)}} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1, d_2]}.$$

Date: May 1, 2015.

and

$$R(z) \ll \left(\sum_{\substack{d \le z \\ d | P(z)}} |\lambda_d| \right)^2$$

In this lecture we will show how to minimize the quadratic form $Q(\lambda)$ with the constraint $\lambda_d = 1$ and bound $R(z)$.

Proposition 1.1. *The minimum value of the quadratic form $Q(\lambda)$ is*

$$S(z) = \sum_{d \le z} \frac{\mu^2(d)}{\phi(d)}$$

and the minimizing vector is given by

$$\lambda_e = \frac{e}{\varphi(e)} \sum_{\substack{d \le z \\ d | P(z) \\ e | d}} \mu\left(\frac{d}{e}\right) \mu(d) \varphi(d)$$

for $e \le z$.

The minimizing vector λ_e clearly is zero for $e > z$, since the condition on the sum is empty. It also satisfies $\lambda_1 = 1$ and is supported on squarefrees. These latter two properties are (perhaps) not immediately apparent from the definition but will be seen later.

Proposition 1.2. *For the minimizing vector λ_d as above*

$$|\lambda_d| \le 1.$$

Combining the two propositions with (1) immediately implies that

$$\pi(x) \le \frac{x}{S(z)} + O(z^2).$$

In Zeev's lecture we saw that by partial summation

$$S(z) \gg \log z.$$

So taking $z = (x/\log x)^{1/2}$ gives

$$\pi(x) \ll x/(\log x).$$

2. Minimizing quadratic forms

Let's now discuss the problem (in general) of minimizing a quadratic form

$$Q(x_1, x_2, \ldots, x_n) = \sum_{1 \le i,j \le n} b_{i,j} x_i x_j,$$

subject to a constraint $x_1 = 1$, which is the setting we are interesting in. Selberg brilliantly solved this problem for the specific quadratic form $Q(\lambda)$ in an amazingly simple way.

The key step in minimizing the quadratic form is a diagonalization procedure. Even though one can always diagonalize the form using linear algebra it is difficult to do explicitly if the number of variable is large (which it is in our setting). Once we have diagonalized the form the optimization problem is *easy* to solve using Lagrange multipliers (or other methods).

First observe that by taking

$$a_{i,j} = \frac{b_{i,j} + b_{j,i}}{2}$$

we may write

$$Q(\mathbf{x}) = \sum_{1 \le i,j \le n} a_{i,j} x_i x_j,$$

where $a_{i,j} = a_{j,i}$. Thus the matrix

$$A = (a_{i,j})_{i,j=1}^{n} = A^T$$

is symmetric and we can write

$$Q(x_1, \ldots, x_n) = \mathbf{x}^T A \mathbf{x}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. A quadratic form is called **diagonal** if A is a diagonal matrix.

Any symmetric matrix, A, can be diagonalized in the following way

$$PAP^T = D$$

where P is an orthogonal matrix (so $P^T = P^{-1}$) and

$$D = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix}$$

is a diagonal matrix. Thus under the linear change of variables $\mathbf{y} = P\mathbf{x}$, we have

$$\begin{aligned} Q(\mathbf{x}) =& \mathbf{y}^T D \mathbf{y} \\ =& \sum_{j=1}^{n} d_j y_j^2 \end{aligned}$$

The condition $x_1 = 1$ and the relation $P^{-1}\mathbf{y} = \mathbf{x}$ gives a linear constraint

$$\sum_{i \le n} c_i y_i = 1.$$

Solving the optimization problem at this stage is easy by using the method of Lagrange multipliers. (We will do this explicitly later on.)

3. The proofs Propositions 1.1 and 1.2

We first will require a few auxiliary lemmas

Lemma 3.1.

$$\sum_{d|n} \varphi(d) = n \qquad \textit{and} \qquad \varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Proof. This follows from one of the homework exercises, but let's go over the proof again. Apply Möbius inversion to see that

$$\begin{aligned} \varphi(n) = & \sum_{\substack{d \le n \\ (d,n)=1}} 1 = \sum_{d \le n} \sum_{e|d, e|n} \mu(e) \\ = & \sum_{e|n} \mu(e) \sum_{\substack{d \le n \\ e|d}} 1 = \sum_{f|d} \mu(f) \left\lfloor \frac{n}{e} \right\rfloor. \end{aligned}$$

So that

$$\frac{\varphi(n)}{n} = \sum_{e|n} \frac{\mu(e)}{e}.$$

Consider the multiplicative function $\iota(n) = n$, so rewriting the above equation in terms of Dirichlet convolution gives

$$\phi(n) = (\mu * \iota)(n)$$

By Möbius inversion this implies that

$$(\mathbf{1} * \phi)(n) = ((\mathbf{1} * \mu) * \iota)(n) = (\delta * \iota)(n) = \iota(n) = n.$$

□

Lemma 3.2 (dual Möbius inversion)**.** *Let f be an arithmetic function. Also, let $\mathcal{D} \subset \mathbb{N}$ be a finite set such that for each $d \in \mathcal{D}$ if $e|d$ then $e \in \mathcal{D}$. Then*

$$g(n) = \sum_{\substack{n|d \\ d \in \mathcal{D}}} f(d)$$

iff

$$f(n) = \sum_{\substack{n|d \\ d \in \mathcal{D}}} g(d) \mu\left(\frac{d}{n}\right).$$

We will delay the proof of this lemma until the end of the section.

Proof of Proposition 1.1. Our goal is to minimize the quadratic form

$$Q(\lambda) = \sum_{\substack{d_1, d_2 \le z \\ d_1, d_2 | P(z)}} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1, d_2]}.$$

The first step is a diagonalization procedure, which reduces the problem to minimizing $Q(\lambda)$ subject to a linear constraint.

Using the equalities

$$\sum_{d|n} \varphi(d) = d$$

and $d_1 d_2/(d_1, d_2) = [d_1, d_2]$ we get that

$$\begin{aligned} x \sum_{\substack{d_1,d_2 \le z \\ d_1,d_2 | P(z)}} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1, d_2]} =& x \sum_{\substack{d_1,d_2 \le z \\ d_1,d_2 | P(z)}} \frac{\lambda_{d_1}\lambda_{d_2}}{d_1 d_2} \sum_{e|(d_1,d_2)} \varphi(e) \\ =& x \sum_{\substack{e \le z \\ e|P(z)}} \varphi(e) \sum_{\substack{e|d_1, e|d_2 \\ d_1,d_2 \le z \\ d_1,d_2|P(z)}} \frac{\lambda_{d_1}\lambda_{d_2}}{d_1 d_2} \\ =& x \sum_{\substack{e \le z \\ e|P(z)}} \varphi(e) \left(\sum_{\substack{e|d \\ d \le z \\ d|P(z)}} \frac{\lambda_d}{d} \right)^2 . \end{aligned}$$

Writing

$$\theta_e = \sum_{\substack{e|d \\ d \le z \\ d|P(z)}} \frac{\lambda_d}{d}$$

we have

$$Q(\lambda) = \sum_{\substack{e \le z \\ e|P(z)}} \varphi(e)\theta_e^2 .$$

So we have succeeded in diagonalizing the quadratic form!

Now use the dual Möbius inversion formula to see that

$$\frac{\lambda_e}{e} = \sum_{\substack{e|d \\ d \le z \\ d|P(z)}} \mu\left(\frac{d}{e}\right) \theta_d .$$

In particular,

$$1 = \sum_{\substack{e \le z \\ e|P(z)}} \mu(e)\theta_e . \tag{2}$$

This is our linear constraint. It remains to minimize

$$Q(\lambda) = \sum_{\substack{e \le z \\ e|P(z)}} \varphi(e)\theta_e^2$$

subject to (2). I claimed this last step was easy so let's solve it in two ways.

Method 1 Cauchy-Schwarz. Applying Cauchy-Schwarz to (2) we get

$$1 \leq \left(\sum_{\substack{e \leq z \\ e | P(z)}} \frac{\mu^2(e)}{\varphi(e)} \right)^{1/2} \left(\sum_{e \leq z} \phi(e) u_e^2 \right)^{1/2} = \left(\sum_{e \leq z} \frac{\mu^2(e)}{\varphi(e)} \right)^{1/2} \left(\sum_{e \leq z} \phi(e) \theta_e^2 \right)^{1/2}$$

where the last identity follows since $\mu^2(e) = 0$ for $e < z$ with $e \nmid P(z)$. Therefore,

$$\sum_{e \leq z} \varphi(e)\theta_e^2 \geq \frac{1}{\displaystyle\sum_{e \leq z} \frac{\mu^2(e)}{\varphi(e)}}$$

We may take

$$\theta_e = \frac{\mu(e)}{\varphi(e) \displaystyle\sum_{e \leq z} \frac{\mu^2(e)}{\varphi(e)}}$$

since this choice satisfies (2). Additionally,

$$\sum_{\substack{e \leq z \\ e | P(z)}} \frac{\mu^2(e)}{\varphi(e) \left(\displaystyle\sum_{e \leq z} \frac{\mu^2(e)}{\varphi(e)} \right)^2} = \frac{1}{\displaystyle\sum_{e \leq z} \frac{\mu^2(e)}{\varphi(e)}}$$

so this minimum is achieved.

Method 2 Lagrange Multipliers. Write

$$Q(\lambda) = \widetilde{Q}(\theta) = \sum_{\substack{e \leq z \\ e | P(z)}} \varphi(e)\theta_e^2 \qquad \text{and} \qquad L(\theta) = \sum_{\substack{e \leq z \\ e | P(z)}} \mu(e)\theta_e.$$

The equality

$$\nabla \widetilde{Q}(\theta) = C \nabla L(\theta)$$

implies that for each squarefree $e \leq z$

$$2\varphi(e)\theta_e = C\mu(e)$$

so

$$\theta_e = \frac{C}{2} \frac{\mu(e)}{\varphi(e)}.$$

Using (2) we see that this gives

$$\frac{C}{2} = \frac{1}{\displaystyle\sum_{d \leq z} \frac{\mu^2(d)}{\varphi(d)}}.$$

□

Proof of Proposition 1.2. Recall that

$$\frac{\lambda_e}{e} = \sum_{\substack{e|d\\ d\leq z\\ d|P(z)}} \mu\left(\frac{d}{e}\right) u_d = \frac{1}{\displaystyle\sum_{f\leq z}\frac{\mu^2(f)}{\varphi(f)}} \sum_{\substack{e|d\\ d\leq z\\ d|P(z)}} \mu\left(\frac{d}{e}\right)\frac{\mu(d)}{\varphi(d)}.$$

In the inner sum write $d = ef$ and note that $\mu(ef) = 0$ unless $(e, f) = 1$ so that by multiplicativity

$$\sum_{\substack{e|d\\ d\leq z\\ d|P(z)}} \mu\left(\frac{d}{e}\right)\frac{\mu(d)}{\varphi(d)} = \frac{\mu(e)}{\varphi(e)} \sum_{\substack{f\leq z/e\\ (f,e)=1\\ f|P(z)}} \frac{\mu^2(f)}{\varphi(f)} = \frac{\mu(e)}{\varphi(e)} \sum_{\substack{f\leq z/e\\ (f,e)=1}} \frac{\mu^2(f)}{\varphi(f)}.$$

Hence,

$$\lambda_e = \frac{\mu(e)}{\varphi(e)\displaystyle\sum_{f\leq z}\frac{\mu^2(f)}{\varphi(f)}} \sum_{\substack{f\leq z/e\\ (f,e)=1}} \frac{\mu^2(f)}{\varphi(f)}$$

As a consistency check, note that it is now clear that $\lambda_1 = 1$ and λ_e is supported on squarefrees.

Next, observe that for any $e \geq 1$

$$\sum_{f\leq z}\frac{\mu^2(f)}{\varphi(f)} = \sum_{\ell|e} \sum_{\substack{m\leq z\\ (m,e)=\ell}} \frac{\mu^2(m)}{\varphi(m)}.$$

In the inner sum write $m = \ell h$ (m is square free so $(\ell, h) = 1$) so that

$$\begin{aligned}\sum_{\ell|e} \sum_{\substack{m\leq z\\ (m,e)=\ell}} \frac{\mu^2(m)}{\varphi(m)} =& \sum_{\ell|e} \sum_{\substack{h\leq z/\ell\\ (h,\ell)=1,(h,e/\ell)=1}} \frac{\mu^2(\ell h)}{\varphi(\ell h)} = \sum_{\ell|e}\frac{\mu^2(\ell)}{\phi(\ell)} \sum_{\substack{h\leq z/\ell\\ (h,e)=1}} \frac{\mu^2(h)}{\varphi(h)}\\ &\geq \sum_{\ell|e}\frac{\mu^2(\ell)}{\phi(\ell)} \sum_{\substack{h\leq z/e\\ (h,e)=1}} \frac{\mu^2(h)}{\varphi(h)}\end{aligned}$$

So that

$$\sum_{\substack{h\leq z/e\\ (h,e)=1}} \frac{\mu^2(h)}{\varphi(h)} \Bigg/ \sum_{f\leq z}\frac{\mu^2(f)}{\varphi(f)} \leq \frac{1}{\displaystyle\sum_{\ell|e}\frac{\mu^2(\ell)}{\varphi(\ell)}}.$$

Collecting estimates, this implies

$$\begin{aligned}|\lambda_e| \leq \frac{e}{\varphi(e)}\cdot\frac{1}{\displaystyle\sum_{\ell|e}\frac{\mu^2(\ell)}{\varphi(\ell)}} =& \prod_{p|e}\left(\left(1-\frac{1}{p}\right)\cdot\left(1+\frac{1}{p-1}\right)\right)^{-1}\\ =& \prod_{p|e}\left(\left(\frac{p-1}{p}\right)\cdot\left(\frac{p}{p-1}\right)\right)^{-1} = 1.\end{aligned}$$

□

It remains to prove the reverse Möbius inversion lemma.

Proof of reverse Möbius inversion. We will prove that if

$$g(n) = \sum_{\substack{n|d \\ d \in \mathcal{D}}} f(d)$$

then

$$f(n) = \sum_{\substack{n|d \\ d \in \mathcal{D}}} g(d)\mu\left(\frac{d}{n}\right)$$

the other claim follows from a similar argument which we will omit. Write

$$\chi_{d=n} \begin{cases} 1 \text{ if } d = n, \\ 0 \text{ otherwise.} \end{cases}$$

Using the definition of $g(n)$ we get that

$$\begin{aligned} \sum_{\substack{n|d \\ d \in \mathcal{D}}} \mu\left(\frac{d}{n}\right) \sum_{\substack{d|e \\ e \in \mathcal{D}}} f(e) &= \sum_{\substack{n|d \\ d \in \mathcal{D}}} \mu\left(\frac{d}{n}\right) \sum_{\substack{c \\ cd \in \mathcal{D}}} \sum_{r \in \mathcal{D}} f(r)\chi_{r=cd} \\ &= \sum_{r \in \mathcal{D}} f(r) \sum_{\substack{m \\ mn \in \mathcal{D}}} \mu(m) \sum_{\substack{c \\ cmn \in \mathcal{D}}} \chi_{cm=\frac{r}{n}} \\ &= \sum_{r \in \mathcal{D}} f(r) \sum_{m|\frac{r}{n}} \mu(m) = f(n) \end{aligned}$$

□