Selberg's Sieve - Bounding Twin Primes

Grace He

Department of Pure Mathematics
University of Waterloo

July 23rd, 2024

UNIVERSITY OF
WATERLOO

Recall Setup

Let us recall that

$$S(A, P, z) = \#\{a \in A : \ a \text{ is not divisible by any } p < z \text{ with } p \in \mathcal{P}\}$$

and Selberg's Sieve gives us

$$S(A, P, z) \leq \frac{X}{V(z)} + \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} |R_{[d_1, d_2]}|$$

where

$$V(z) = \sum_{\substack{d \leq z \\ d | P_z}} \frac{\mu^2(d)}{f_1(d)} \quad f(n) = \sum_{d|n} f_1(d) \ \text{ and } \ |A_d| = \frac{X}{f(d)} + R_d$$

Notice that in order to use Selberg's Sieve, we want to find an upper bound for $\frac{1}{V(z)}$, thus a lower bound for $V(z)$, which motivates the following lemma:

Lemma

Lemma

Let $\tilde{f}$ be a completely multiplicative function with $\tilde{f}(p) := f(p)$ for all primes p. Then we have

$$V(z) \geq \sum_{\substack{e \leq z \\ p|e \Rightarrow p|P_z}} \frac{1}{\tilde{f}(e)} \quad \textit{where} \quad P_z = \prod_{\substack{p \in \mathcal{P} \\ p \leq z}} p$$

Note: If $\mathcal{P}$ is the set of all primes, then the second condition $p \mid e \Rightarrow p \mid P_z$ is trivial. However, we will keep our $\mathcal{P}$ generic in this lemma.

Proof of Lemma

Proof. First, note that if the multiplicative function $f(n) = \sum_{d|n} f_1(d)$, then $f(1) = f_1(1) = 1$ and $f(p) = f_1(p) + 1$ for all primes p. Using this fact, we have, for $d \mid P_z$,

$$\frac{f(d)}{f_1(d)} = \prod_{p|d} \frac{f(p)}{f_1(p)} = \prod_{p|d} \frac{1}{\left(\frac{f(p)-1}{f(p)}\right)} = \prod_{p|d} \left(1 - \frac{1}{f(p)}\right)^{-1}$$

$$= \prod_{p|d} \left(1 + \frac{1}{f(p)} + \frac{1}{f(p)^2} + \cdots\right) = \sum_{p|k \Rightarrow p|d} \frac{1}{\tilde{f}(k)}$$

Now, we can write

$$V(z) = \sum_{\substack{d \le z \\ d|P_z}} \frac{\mu^2(d)}{f_1(d)} = \sum_{\substack{d \le z \\ d|P_z}} \frac{\mu^2(d)}{f(d)} \sum_{p|k \Rightarrow p|d} \frac{1}{\tilde{f}(k)} = \sum_{\substack{d \le z \\ d|P_z}} \sum_{p|k \Rightarrow p|d} \frac{1}{\tilde{f}(dk)}$$

Proof of Lemma Cont'd

To show that

$$V(z) = \sum_{\substack{d \le z \\ d | P_z}} \sum_{p|k \Rightarrow p|d} \frac{1}{\tilde{f}(dk)} \ge \sum_{\substack{e \le z \\ p|e \Rightarrow p|P_z}} \frac{1}{\tilde{f}(e)}$$

let

$$A = \frac{1}{\tilde{f}(e)} = \frac{1}{\tilde{f}(p_{n_1}^{r_1} \cdots p_{n_q}^{r_q})}$$

be a term in the right summation, where each $p_{n_i} \mid P_z$, $r_i \ge 1$. Then for $d = p_{n_1} \cdots p_{n_q}$ and $k = p_{n_1}^{r_1 - 1} \cdots p_{n_q}^{r_q - 1}$, we have

$$\frac{1}{\tilde{f}(dk)} = \frac{1}{\tilde{f}(p_{n_1} \cdots p_{n_q} p_{n_1}^{r_1-1} \cdots p_{n_q}^{r_q-1})} = \frac{1}{\tilde{f}(e)} = A$$

Clearly, all other terms in the left summation are positive, giving the desired inequality. $\square$

Twin Primes

Definition

A prime p is called a twin prime if $p+2$ is also a prime.

Let

$$\pi_2(x) := \# \text{ of twin primes} \leq x$$

We would like to use Selberg's Sieve to obtain an upper bound for $\pi_2(x)$ as $x \to \infty$. In the setting of this problem, we define

$$A = \{n(n+2) : n \leq x\} \quad \text{and} \quad \mathcal{P} = \text{ set of all primes}$$

Each natural number n corresponds to a unique $a = n(n+2) \in A$, and vice versa; so we will sieve through A instead of the set $\{n \in \mathbb{N} : n \leq x\}$.

Understanding $S(A, \mathcal{P}, z)$

For $0 < z < x$, we have

$$P_z = \prod_{\substack{p \in \mathcal{P} \\ p < z}} p = \prod_{p \leq z} p$$

and so

$$S(A, \mathcal{P}, z) := \#\{n(n+2) : n \leq x,\ \ p \nmid n \text{ and } p \nmid (n+2) \text{ for all } p \leq z\}$$

- If $n \leq z, n(n+2)$ is not counted in $S(A, \mathcal{P}, z)$, ie. n is not counted.
- All twin primes $z < p \leq x$ are counted.

$$\pi_2(x) = \sum_{\substack{p \leq x \\ p+2 \in \mathcal{P}}} 1 = \pi_2(z) + \sum_{\substack{z < p \leq x \\ p+2 \in \mathcal{P}}} 1$$
$$\leq \pi_2(z) + S(A, \mathcal{P}, z) \leq z + \boxed{S(A, \mathcal{P}, z)}$$

Outline of Steps

Once again, Selberg's Sieve gives us

$$S(A, P, z) \leq \frac{X}{V(z)} + \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} |R_{[d_1, d_2]}|$$

To use Selberg's Sieve, we will need to

- Find X, estimation of the size of A (Clearly, $X = x$) ✓
- Estimate $|A_d|$ for $d \mid P_z$ to find our multiplicative function, f
- Find lower bound for $V(z)$
- Estimate error term

Estimating $|A_d|$

Let $d \mid P_z$, say $d = p_1 \cdots p_n$. Then we have

$$\begin{aligned} |A_d| &= \#\{n(n+2) : n \leq x \text{ and } d \mid n(n+2)\} \\ &= \#\{n(n+2) : n \leq x \text{ and } n(n+2) \equiv 0 \pmod{d}\} \end{aligned}$$

Let $N(q)$ be the number of solutions to $n(n+2) \equiv 0 \pmod{q}$. By the Chinese Remainder Theorem, $n(n+2) \equiv 0 \pmod{d}$ has the same number of solutions as

$$\begin{gathered} n(n+2) \equiv 0 \pmod{p_1} \\ \vdots \\ n(n+2) \equiv 0 \pmod{p_k} \end{gathered}$$

Let $\omega(d) = k$ be the number of prime factors of d. Since for each $1 \leq i \leq k,\ N(p_i) \leq 2$, we have that

$$N(d) = N(p_1) \cdots N(p_k) \leq 2^k = 2^{\omega(d)}$$

Estimating $|A_d|$ Cont'n

Further, since $N(d)$ is only the number of solutions modulo d and we want all solutions $\leq x$, we can estimate the total number of solutions, ie. the size of A_d by

$$|A_d| = \frac{x}{d} \cdot N(d) + R_d, \quad \text{where} \quad R_d \leq N(d) \leq 2^{\omega(d)}$$

Thus, we have our multiplicative function

$$f(d) = \frac{d}{N(d)}$$

which satisfies the conditions of Selberg's Sieve. And a simple fact for later:

$$f(p) = \frac{d}{N(d)} = \begin{cases} p & \text{if } p = 2 \\ p/2 & \text{if } p > 2 \end{cases}$$

Next Step

- Find X, estimation of the size of A (Clearly, $X = x$) ✓
- Estimate $|A_d|$ for $d \mid P_z$ to find our multiplicative function, f ✓
- Find lower bound for $V(z)$
- Estimate error term

Bounding $V(z)$ - Notations

First, let use define some notations

Definition

For $n \in \mathbb{N}$, define

$$\tau_1(n) := \#\ \textit{odd divisors of } n$$

And so for $n = 2^s p_1^{e_1} \cdots p_m^{e_m}$, we have $\tau_1(n) = (e_1 + 1) \cdots (e_m + 1)$

Definition

For $n \in \mathbb{N}$, define

$$\tau(n) := \#\ \textit{divisors of } n$$

Note that if d is square free, then $\tau(d) = 2^{\omega(d)}$.

Bounding $V(z)$

Let $\tilde{f}$ be a completely multiplicative function with $\tilde{f}(p) = f(p)$ for all primes p, as defined in our lemma. Then the lemma tells use that

$$
\begin{aligned}
V(z) \geq \sum_{\substack{n \leq z \\ p|n \Rightarrow p|P_z}} \frac{1}{\tilde{f}(n)} &= \sum_{n \leq z} \frac{1}{\tilde{f}(2)^s \tilde{f}(p_1)^{e_1} \cdots \tilde{f}(p_m)^{e_m}} \\
&= \sum_{n \leq z} \frac{1}{2^s (p_1/2)^{e_1} \cdots (p_m/2)^{e_m}} \\
&= \sum_{n \leq z} \frac{2^{e_1} \cdots 2^{e_m}}{n} \\
&\geq \sum_{n \leq z} \frac{(e_1 + 1) \cdots (e_m + 1)}{n} \\
&= \sum_{n \leq z} \frac{\tau_1(n)}{n}
\end{aligned}
$$

Bounding $V(z)$ Cont'd

Next, we have

$$\begin{aligned}\sum_{n\le z}\tau_1(n) &= \sum_{n\le z}\sum_{\substack{d|n\\(d,2)=1}} 1 = \sum_{\substack{d\le z\\(d,2)=1}}\sum_{\substack{n\le z\\d|n}} 1 = \sum_{\substack{d\le z\\(d,2)=1}}\left[\frac{z}{d}\right]\\ &= \sum_{\substack{d\le z\\(d,2)=1}}\frac{z}{d} - \sum_{\substack{d\le z\\(d,2)=1}}\left\{\frac{z}{d}\right\} \ge \sum_{\substack{d\le z\\(d,2)=1}}\frac{z}{d} - \sum_{\substack{d\le z\\(d,2)=1}} 1\\ &\ge z\sum_{\substack{d\le z\\(d,2)=1}}\frac{1}{d} - z\end{aligned}$$

Bouding $V(z)$ - Partial Summation

Theorem

Let $c_1, c_2, \ldots$ be a sequence of complex numbers and set

$$S(x) := \sum_{d \leq x} c_d.$$

Let d_0 be a fixed positive integer. If $c_j = 0$ for $j < d_0$ and $f : [d_0, \infty) \to \mathbb{C}$ has continuous derivative in $[d_0, \infty)$, then for x an integer $> d_0$ we have

$$\sum_{d \leq x} c_d f(d) = S(x)f(x) - \int_{d_0}^{x} S(t)f'(t)\, dt.$$

Bounding $V(z)$ Cont'd

For the summation

$$\sum_{\substack{d \le z \\ (d,2)=1}} \frac{1}{d}, \quad \text{we choose } c_d = \begin{cases} 1 & \text{if } (d,2) = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and } f(d) = \frac{1}{d}$$

Then $d_0 = 1$ will allow us to use the partial summation technique:

$$\begin{aligned}
\sum_{\substack{d \le z \\ (d,2)=1}} \frac{1}{d} &= \frac{1}{z} \sum_{\substack{d \le z \\ (d,2)=1}} 1 + \int_1^z \left(\frac{1}{t^2} \sum_{\substack{d \le t \\ (d,2)=1}} 1 \right) dt \\
&= \underbrace{\frac{1}{z} \left\lceil \frac{z}{2} \right\rceil}_{\ge 0} + \int_1^z \left(\frac{1}{t^2} \left\lceil \frac{t}{2} \right\rceil \right) dt \\
&\ge \int_1^z \left(\frac{1}{2t} - \frac{1}{t^2} \right) dt \ge \frac{1}{2} \log z - \int_1^\infty \frac{1}{t^2} dt = \frac{1}{2} \log z - c
\end{aligned}$$

Bounding $V(z)$ Cont'd

Hence we have that

$$\sum_{n \leq z} \tau_1(n) \geq z \sum_{\substack{d \leq z \\ (d,2)=1}} \frac{1}{d} - z \geq \frac{1}{2} z \log z - \underbrace{(c+1)}_{D} z$$

Now, for

$$\sum_{n \leq z} \frac{\tau_1(n)}{n}, \quad \text{we choose } c_n = \tau_1(n) \text{ and } f(n) = \frac{1}{n}$$

Apply partial summation again, and we get

$$V(z) \geq \sum_{n \leq z} \frac{\tau_1(n)}{n} \geq \frac{1}{4} \log^2(z) + \left(\frac{1}{2} - D\right) \log z - D \gg \log^2(z)$$

Next Step

- Find X, estimation of the size of A (Clearly, $X = x$) ✓
- Estimate $|A_d|$ for $d \mid P_z$ to find our multiplicative function, f ✓
- Find lower bound for $V(z)$ ✓
- Estimate error term

Estimate Error Term

First, let us note that

$$\sum_{n\le x}\tau(n)=\sum_{n\le x}\sum_{\substack{d\le x\\ d|n}}1=\sum_{d\le x}\sum_{\substack{n\le x\\ d|n}}1=\sum_{d\le x}\left[\frac{x}{d}\right]\le x\sum_{d\le x}\frac{1}{d}$$

Taking $c_n=1$ and $f(t)=\frac{1}{t}$, we can use partial summation to get that

$$x\sum_{d\le x}\frac{1}{d}=x\left(\frac{1}{x}\cdot[x]+\int_1^x\frac{[t]}{t^2}dt\right)\le x(1+\log x)\ll x\log x$$

Hence,

$$\sum_{n\le x}\tau(n)\ll x\log x$$

Estimate Error Term Cont'd

Note that our error term when estimating $|A_d|$ satisfies

$$R(d) \leq N(d) \leq 2^{\omega(d)}$$

Thus, we have for the error term from Selberg's Sieve,

$$\begin{aligned}
\sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} R([d_1, d_2]) &\leq \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} 2^{\omega([d_1, d_2])} \leq \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} 2^{\omega(d_1)} 2^{\omega(d_2)} \\
&= \left(\sum_{\substack{d \leq z \\ d \text{ square free}}} 2^{\omega(d)} \right)^2 \leq \left(\sum_{d \leq z} 2^{\omega(d)} \right)^2 \\
&\leq \left(\sum_{d \leq z} \tau(d) \right)^2 \ll (z \log z)^2
\end{aligned}$$

Next Step

- Find X, estimation of the size of A (Clearly, $X = x$) ✓
- Estimate $|A_d|$ for $d \mid P_z$ to find our multiplicative function, f ✓
- Find lower bound for $V(z)$ ✓
- Estimate error term ✓

Finalé

We shall recall our bound on $\pi_2(x)$ from before:

$$\pi_2(x) \leq z + S(A, \mathcal{P}, z)$$

As well, from Selberg and all the work we've done, we have

$$S(A, \mathcal{P}, z) \leq \frac{X}{V(z)} + \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} |R_{[d_1, d_2]}| \ll \frac{x}{\log^2(z)} + (z \log z)^2$$

And so

$$\pi_2(x) \ll z + \frac{x}{\log^2(z)} + (z \log z)^2$$

Now, if we pick $z = x^{1/4}$, we have

$$\pi_2(x) \ll x^{1/4} + 16 \cdot \frac{x}{\log^2(x)} + \frac{1}{16} \sqrt{x} \log^2(x) \ll \frac{x}{\log^2(x)}$$

The End

Thank You!