

# Selberg's Sieve I

Peiran Tao

Department of Pure Mathematics  
University of Waterloo

July 23rd, 2024



# Overview

1. Notations
2. Sieve of Eratosthenes
3. Selberg's Sieve

# Notations

1.  $\mathbb{N}$  = the set of natural numbers (positive integers).
2.  $\mathbb{P}$  = the set of all prime numbers.
3. For  $x > 0$ , let:

$$\pi(x) = \# \text{ of prime numbers } \leq x$$

to be the prime counting function.

4. For nonzero  $a, b \in \mathbb{N}$ , denote:

$$(a, b) := \gcd(a, b) \quad \text{and} \quad [a, b] := \text{lcm}(a, b)$$

# Sieve Method

**Sieve Methods** are techniques used to estimate the size of a set after elements with some undesirable property have been removed.

# Sieve of Eratosthenes

A classic application of sieve method is to estimate  $\pi(x)$ .

To estimate  $\pi(x)$  is the same as estimating the number of primes in  $A = [1, x] \cap \mathbb{N}$ .

Using the language of sieve method, to find all primes, we want to estimate the size of  $A$  after removing 1 and all composite numbers.

# Characterize composite numbers

## Theorem

*Let  $x \geq 2$  be a real number. Let  $n \in \mathbb{N}$  with  $2 \leq n \leq x$ . If  $n$  is composite, then  $n$  has a prime factor  $p$  with  $p \leq \sqrt{x}$ .*

**Proof:** Suppose the result is not true. Since  $n$  is composite, it must have at least two prime factors  $p, q$  (not necessarily distinct). Then  $p, q > \sqrt{x}$ , so:

$$n \geq pq > \sqrt{x}\sqrt{x} = x.$$

which is a contradiction. □

# Sieve of Eratosthenes

So, to remove all composite numbers, it suffices to remove all integers in  $A$  that do not satisfy the property in Lemma 1.1.

For  $x \geq 2$ , if we remove all the multiplies of primes  $\leq \sqrt{x}$  in  $A$ , the numbers that remain are primes numbers in  $(\sqrt{x}, x]$  and the number 1, thus:

$$\pi(x) - \pi(\sqrt{x}) + 1 = \pi(x, \sqrt{x}).$$

Here  $\pi(x, \sqrt{x})$  denote the number of  $n \leq x$  with no prime factors  $\leq \sqrt{x}$ .

# Generalization

Instead of removing multiples of primes  $\leq \sqrt{x}$ , we can replace  $\sqrt{x}$  with an arbitrary  $z > 0$ .

Moreover, we can impose some conditions on the primes.

## Definition

*Let  $A \subseteq \mathbb{N}$  be a finite subset of  $\mathbb{N}$ . Let  $\mathcal{P} \subseteq \mathbb{P}$  be a set of prime numbers and let  $z > 0$ . Define:*

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



# Generalization

If we define:

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

For  $a \in A$ :  $(a, P_z) = 1$  if and only if  $p \nmid a$  for all  $p \in \mathcal{P}$  with  $p \leq z$ .

Therefore, we can rewrite  $S(A, \mathcal{P}, z)$  as:

$$S(A, \mathcal{P}, z) = \sum_{\substack{a \in A \\ (a, P_z) = 1}} 1 = \sum_{a \in A} F(a).$$

where:

$$F(a) = \begin{cases} 1 & \text{if } (a, P_z) = 1, \\ 0 & \text{if } (a, P_z) > 1. \end{cases}$$

# Generalization

Let  $n \in \mathbb{N}$ . Define the **Möbius function**:

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \text{ is not squarefree,} \\ (-1)^r & \text{if } n = p_1 \cdots p_r \text{ is squarefree.} \end{cases}$$

## Lemma

Let  $\mu$  denote the Möbius function, then:

$$I(n) := \sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

# Generalization

By the lemma, we have:

$$I((a, P_z)) = \sum_{d|(a, P_z)} \mu(d) = \begin{cases} 1 & \text{if } (a, P_z) = 1, \\ 0 & \text{if } (a, P_z) > 1. \end{cases}$$

Hence, we have:

$$S(A, \mathcal{P}, z) = \sum_{a \in A} \sum_{d|(a, P_z)} \mu(d). \quad (1)$$

# Generalization

If we directly analyze the sum in (1), we can get the general Sieve of Eratosthenes, called the Legendre's Sieve.

But this talk is not called the Legendre's Sieve, so by contrapositive we are not going to analyze the sum directly.

# Selberg's trick

Look at the sum (1):

$$S(A, \mathcal{P}, z) = \sum_{a \in A} \sum_{d|(a, P_z)} \mu(d).$$

Note that  $\sum_{d|(a, P_z)} \mu(d)$  is either 1 or 0, so:

$$\sum_{d|(a, P_z)} \mu(d) \leq \left( \sum_{d|(a, P_z)} \lambda_d \right)^2.$$

for any sequence  $(\lambda_d) \subseteq \mathbb{R}$  with  $\lambda_1 = 1$ .

# Selberg's trick

But obviously, we cannot choose  $(\lambda_d)$  to be an arbitrary sequence. We need to choose it so that the quadratic form with indeterminates  $\lambda_d$ :

$$\left( \sum_{d|(a, P_z)} \lambda_d \right)^2 = \sum_{d_1, d_2 | (a, P_z)} \lambda_{d_1} \lambda_{d_2}.$$

is minimal. Otherwise, our upper bound is too big, then this trick is useless.

# Selberg's Sieve

Now we can start the derivation for Selberg's Sieve.

$$\begin{aligned} S(A, \mathcal{P}, z) &= \sum_{\substack{a \in A \\ (a, P_z)=1}} 1 = \sum_{a \in A} \sum_{d|(a, P_z)} \mu(d) \\ &\leq \sum_{a \in A} \left( \sum_{d|(a, P_z)} \lambda_d \right)^2 \\ &= \sum_{a \in A} \sum_{d_1, d_2|(a, P_z)} \lambda_{d_1} \lambda_{d_2} \end{aligned}$$

# Selberg's Sieve

Note that:

$$\begin{aligned}d \mid (a, b) &\iff d \mid a \text{ and } d \mid b \\[a, b] \mid \ell &\iff a \mid \ell \text{ and } b \mid \ell\end{aligned}$$

Therefore:

$$\begin{aligned}S(A, \mathcal{P}, z) &\leq \sum_{a \in A} \sum_{\substack{d_1, d_2 \mid a \\ d_1, d_2 \mid P_z}} \lambda_{d_1} \lambda_{d_2} \\&= \sum_{d_1, d_2 \mid P_z} \lambda_{d_1} \lambda_{d_2} \sum_{\substack{a \in A \\ d_1, d_2 \mid a}} 1 \\&= \sum_{d_1, d_2 \mid P_z} \lambda_{d_1} \lambda_{d_2} \sum_{\substack{a \in A \\ [d_1, d_2] \mid a}} 1\end{aligned}$$



# Selberg's Sieve

The last sum:

$$\sum_{\substack{a \in A \\ [d_1, d_2] \mid a}} 1.$$

is exactly the number of  $a \in A$  such that  $[d_1, d_2] \mid a$ .

This suggests that it is helpful to study the size of the set:

$$A_d = \{a \in A : d \mid a\}.$$

for  $d \mid P_z$ .

# Selberg's Sieve

Suppose there is a multiplicative function  $f$  with  $f(p) > 1$  for all prime  $p \in \mathcal{P}$  such that:

$$|A_d| = \frac{X}{f(d)} + R_d. \quad (2)$$

1. Think of  $X$  as an estimation of  $|A|$ .
2. Think of (2) as an estimation of  $|A_d|$ , with  $1/f(d)$  the 'density' of  $A_d$  in  $A$ , and  $R_d$  as the error term to the estimation.

# Selberg's Sieve

$$S(A, \mathcal{P}, z) \leq \sum_{d_1, d_2 | P_z} \lambda_{d_1} \lambda_{d_2} |A_{[d_1, d_2]}|.$$

Recall that:

$$|A_d| = \frac{X}{f(d)} + R_d.$$

We get:

$$\begin{aligned} S(A, \mathcal{P}, z) &\leq \sum_{d_1, d_2 | P_z} \lambda_{d_1} \lambda_{d_2} \left( \frac{X}{f([d_1, d_2])} + R_{[d_1, d_2]} \right) \\ &= X \underbrace{\sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1} \lambda_{d_2}}{f([d_1, d_2])}}_T + \underbrace{\sum_{d_1, d_2 | P_z} \lambda_{d_1} \lambda_{d_2} R_{[d_1, d_2]}}_R \end{aligned}$$

# Selberg's Sieve

Hence we get:

$$S(A, \mathcal{P}, z) \leq XT + R.$$

Remember, our goal is to minimize this upper bound by choosing  $(\lambda_d)$  optimally.

Let us analyze  $T$  first.

# Möbius Inversion

## Lemma

Let  $f, F : \mathbb{N} \rightarrow \mathbb{C}$ . Then:

$$F(n) = \sum_{d|n} f(d) \iff f(n) = \sum_{d|n} F(d) \mu\left(\frac{n}{d}\right).$$

This is known as the **Möbius Inversion Formula**.

# The Main Term

By Möbius Inversion, there is  $f_1 : \mathbb{N} \rightarrow \mathbb{C}$  such that:

$$f(n) = \sum_{d|n} f_1(d).$$

Explicitly, we define:

$$f_1(n) = \sum_{d|n} f(d) \mu\left(\frac{n}{d}\right).$$

For  $n = p$  a prime, we get:

$$f_1(p) = \sum_{d|p} f(d) \mu\left(\frac{p}{d}\right) = f(1) \mu(p) + f(p) \mu(1) = f(p) - 1 > 0.$$

# The Main Term

## Lemma

*If  $f$  is multiplicative, then we have:*

$$f([d_1, d_2])f((d_1, d_2)) = f(d_1)f(d_2).$$

We have:

$$\begin{aligned} T &= \sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1} \lambda_{d_2}}{f([d_1, d_2])} \\ &= \sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1)f(d_2)} f((d_1, d_2)) \\ &= \sum_{d_1, d_2 | P_z} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1)f(d_2)} \sum_{\delta | (d_1, d_2)} f_1(\delta) \end{aligned}$$

# The Main Term

Now, we choose  $\lambda_d = 0$  for  $d > z$ . We have:

$$\begin{aligned} T &= \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1) f(d_2)} \sum_{\delta | (d_1, d_2)} f_1(\delta) \\ &= \sum_{\substack{\delta \leq z \\ \delta | P_z}} f_1(\delta) \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z \\ \delta | (d_1, d_2)}} \frac{\lambda_{d_1} \lambda_{d_2}}{f(d_1) f(d_2)} \\ &= \sum_{\substack{\delta \leq z \\ \delta | P_z}} f_1(\delta) \left( \sum_{\substack{d \leq z \\ d | P_z \\ \delta | d}} \frac{\lambda_d}{f(d)} \right)^2 \end{aligned}$$



# The Main Term

Define:

$$u_\delta = \sum_{\substack{d \leq z \\ d|P_z \\ \delta|d}} \frac{\lambda_d}{f(d)}.$$

Hence we get:

$$T = \sum_{\substack{\delta \leq z \\ \delta|P_z}} f_1(\delta) u_\delta^2.$$

Also, from the sum we see  $u_\delta = 0$  for  $\delta > z$ .

# The Main Term

It turns out, by another Inversion formula, we have:

$$\frac{\lambda_d}{f(d)} = \sum_{\substack{\delta|P_z \\ d|\delta}} \mu\left(\frac{\delta}{d}\right) u_\delta.$$

Plug in  $d = 1$  yields:

$$1 = \frac{\lambda_1}{f(1)} = \sum_{\delta|P_z} \mu(\delta) u_\delta = \sum_{\substack{\delta \leq z \\ \delta|P_z}} \mu(\delta) u_\delta.$$

To choose  $\lambda_d$ , it suffices to choose  $u_\delta$ .

# The Main Term

Define:

$$V(z) = \sum_{\substack{\delta \leq z \\ d|\overline{P}_z}} \frac{\mu^2(\delta)}{f_1(\delta)}.$$

Then we get:

$$\begin{aligned} & \sum_{\substack{\delta \leq z \\ d|\overline{P}_z}} f_1(\delta) \left( u_\delta - \frac{\mu(\delta)}{f_1(\delta)V(z)} \right)^2 + \frac{1}{V(z)} \\ &= \sum_{\substack{\delta \leq z \\ d|\overline{P}_z}} f_1(\delta) u_\delta^2 - \frac{2}{V(z)} \sum_{\substack{\delta \leq z \\ d|\overline{P}_z}} u_\delta \mu(\delta) + \frac{1}{V(z)^2} \sum_{\substack{\delta \leq z \\ d|\overline{P}_z}} \frac{\mu^2(\delta)}{f_1(\delta)} + \frac{1}{V(z)} \\ &= T - \frac{2}{V(z)} + \frac{1}{V(z)} + \frac{1}{V(z)} \end{aligned}$$

# The Main Term

Hence we have:

$$T = \sum_{\substack{\delta \leq z \\ \delta | P_z}} f_1(\delta) \left( u_\delta - \frac{\mu(\delta)}{f_1(\delta)V(z)} \right)^2 + \frac{1}{V(z)}.$$

# The Main Term

The first sum is non-negative as  $f_1(p) > 0$  for all  $p$ .

So,  $T$  is minimized when:

$$u_\delta = \frac{\mu(\delta)}{f_1(\delta)V(z)}.$$

So we can choose:

$$\lambda_d = f(d) \sum_{\substack{\delta|P_z \\ d|\delta}} \mu\left(\frac{\delta}{d}\right) u_\delta.$$

Therefore, we have:

$$T = \frac{1}{V(z)}.$$

# The Error Term

The error term depends on  $\lambda_d$ . It turns out that, given:

$$\lambda_d = f(d) \sum_{\substack{\delta | P_z \\ d | \delta}} \mu\left(\frac{\delta}{d}\right) u_\delta.$$

we must have  $|\lambda_d| \leq 1$  for all  $d$ . Hence:

$$R \leq \left| \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} \lambda_{d_1} \lambda_{d_2} R_{[d_1, d_2]} \right| \leq \sum_{\substack{d_1, d_2 \leq z \\ d_1, d_2 | P_z}} |R_{[d_1, d_2]}|.$$

# The final result

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

Given a problem, if we want to apply Selberg's Sieve, we need to:

1. Find suitable  $A, \mathcal{P}, z$ .
2. Estimate  $|A_d|$  for  $d \mid P_z$ .
3. Find a lower bound for  $V(z)$ .
4. Estimate the error term  $R$ .