MATH 238 TOPICS IN NUMBER THEORY

INTRODUCTION TO SIEVE THEORY

Contents

1. Warm up on primes 2
2. Arithmetic functions 3
3. It's prime time! 8
4. Eratosthenes 11
5. Selberg 13
6. Maynard/Tao/Zhang 18
7. Final things 25
8. Some answers 27

1. Warm up on primes

Exercise 1. Prove that every integer $n > 1$ is either prime or a product of primes.

Let's start with an old theorem of Euclid.

Exercise 2. Prove that there exists infinitely many prime numbers.

Now let's step it up a little, but this should still be something you've seen before.

Exercise 3 (Bézout's identity)**.** Given any two integers a and b with gcd $d = \gcd(a, b)$, there are integers x and y such that

$$d = ax + by.$$

Use the last result to prove the next.

Exercise 4 (Euclid's lemma)**.** Given integers a, b, and c, prove that if $a|bc$ and $\gcd(a, b) = 1$, then $a|b$.

And now for the fundamental theorem of arithmetic.

Exercise 5. Prove that every integer $n > 1$ can be represented as a product of prime factors in only one way, up to rearranging the factors.

Exercise 6. Let p_i denote the i-th prime number. Prove that the infinite series

$$\sum_{i=1}^{\infty} \frac{1}{p_i}$$

diverges as follows.

(1) Assume to the contrary that the series converges. Then there exists an integer k such that

$$\sum_{i=k+1}^{\infty} \frac{1}{p_i} < \frac{1}{2}.$$

Let $Q = p_1 \dots p_k$, and argue that for any $r \geq 1$, we have

$$\sum_{n=1}^{r} \frac{1}{1+nQ} \leq \sum_{j=1}^{\infty} \left(\sum_{i=k+1}^{\infty} \frac{1}{p_i} \right)^j.$$

(2) Show that the right-hand side of the inequality is bounded above by the convergent series

$$\sum_{j=1}^{\infty} \left(\frac{1}{2} \right)^j$$

so the left-hand side converges for all r.

(3) Obtain a contradiction by showing that the left-hand side in fact diverges.

2. Arithmetic functions

An arithmetic function is a function $f : \mathbf{N} \to \mathbf{R}$. Arithmetic functions are fundamental objects in number theory.

Definition 2.1. Define the Möbius function μ as $\mu(1) = 1$, and for every integer $n > 1$, define

$$\mu(n) = \begin{cases} 1 & \text{if } n \text{ is squarefree, and has an even number of prime factors} \\ -1 & \text{if } n \text{ is squarefree, and has an odd number of prime factors} \\ 0 & \text{if } n \text{ has a squared prime factor.} \end{cases}$$

Exercise 7. Prove that if $n > 1$, then

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

Hint: Show there is a bijection between the squarefree factors d of n and the subsets of all the set of all prime factors of n.

Definition 2.2. Define the Euler totient function $\varphi(n)$ to be the number of positive integers less than n that are relatively prime to n. In other words,

$$\varphi(n) = \sum_{\substack{i=1 \\ \gcd(i,n)=1}}^{n} 1$$

Exercise 8. For any $n \geq 1$, prove that

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

(1) Consider the sets $A(d) = \{k : (k, n) = d, 1 \leq k \leq n\}$. Show that

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

(2) Show that there is a bijection between the elements of $A(d)$ and the integers k/d such that $0 < k/d \leq n/d$, with $\gcd(k/d, n/d) = 1$. Use this to show that

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

and also

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

Exercise 9. For any $n \geq 1$, prove that

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

(1) First show that from the definition of $\varphi(n)$ we can write

$$\varphi(n) = \sum_{k=1}^{n} \left[\frac{1}{\gcd(n, k)} \right].$$

where $[x]$ indicates the floor function of x.

(2) Rewrite this as

$$\varphi(n) = \sum_{k=1}^{n} \sum_{d|(n,k)} \mu(d)$$

and rewrite this sum to conclude.

We will encounter a lot of products indexed by prime numbers. Here is a first taste.

Exercise 10. For any $n \geq 1$, prove that

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$$

(1) Explain why this is true if $n = 1$.
(2) If $n > 1$, let $n = p_1 p_2 \dots p_r$ be its prime factorization. Show that

$$\prod_{p|n} \left(1 - \frac{1}{p}\right) = 1 - \sum_{i=1}^{r} \frac{1}{p_i} + \sum_{i,j=1}^{r} \frac{1}{p_i p_j} + \dots + (-1)^r \frac{1}{p_1 \dots p_r}.$$

(3) Show that the sum is equal to

$$\sum_{d|n} \frac{\mu(d)}{d}$$

and conclude.

Definition 2.3. Let f and g be arithmetic functions. Define their Dirichlet product (or Dirichlet convolution) to be

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

The following should not take long.

Exercise 11. Show that for any arithmetic functions f, g, h, we have

(a) $f * g = g * f$
(b) $(f * g) * h = g * (f * h)$.
(c) $I * f = f * I = f$ where $I(n) = [1/n]$.

Now we are looking for the notion of an inverse.

Exercise 12. Let f be an arithmetic function such that $f(1) \neq 0$. Show that there is a unique function, denoted f^{-1} such that

$$f * f^{-1} = f^{-1} * f = I.$$

(1) Given f, we shall show that the equation $(f * f^{-1})(n) = I(n)$ has a unique solution for the values of $f^{-1}(n)$. We proceed by strong induction. First check the base case.
(2) Now assume that the values $f^{-1}(k)$ are uniquely determined for all $1 \leq k < n$. Show that the equation $(f * f^{-1})(n) = I(n)$ can be expressed as

$$\sum_{d|n} f\left(\frac{n}{d}\right) f^{-1}(d) = 0.$$

(3) Apply the induction assumption to argue that

$$f^{-1}(n) = \frac{-1}{f(n)} \sum_{\substack{d|n \\ d<n}} f\left(\frac{n}{d}\right) f^{-1}(d)$$

is the uniquely determined value of $f^{-1}(n)$, and use this to conclude.

Now for the Möbius inversion formula.

Exercise 13. Prove that

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

if and only if

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

(1) First show that if we define the arithmetic function $u(n) = 1$ for all n, then $\mu * u = I$, so that $\mu = u^{-1}$ and $u = \mu^{-1}$.
(2) To show the forward implication, consider $f * \mu$.
(3) To show the backward implication, consider $g * u$.

All that was to build up to the next arithmetic function.

Definition 2.4. Define the von Mangoldt function

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^m \text{ for some prime } p \text{ and } m \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Exercise 14. Show that for all $n \geq 1$, we have

$$\log n = \sum_{d|n} \Lambda(d).$$

Hint: Apply the logarithm to the prime factorization of n, and consider the resulting sum.

Exercise 15. Show that for all $n \geq 1$, we have by Möbius inversion

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d} = -\sum_{d|n} \mu(d) \log d.$$

Here is some notation that we will need.

Definition 2.5. If $g(x) > 0$ for all $x \geq a$, we write

$$f(x) = O(g(x))$$

if the quotient $f(x)/g(x)$ is bounded for all $x \geq a$, i.e., there exists a constant $M > 0$ such that

$$|f(x)| \leq Mg(x), \qquad x \geq a.$$

Also, $f(x) = h(x) + O(g(x))$ means that $f(x) - h(x) = O(g(x))$.

We need one more result before we go to town.

Exercise 16. We want to show that

$$\sum_{n\leq x}\Lambda(n)\left[\frac{x}{n}\right]=x\log x-x+O(\log x)$$

(a) Let h,f,g be arithmetic functions such that $h=f*g.$ and denote the partial sums

$$H(x)=\sum_{n\leq x}h(n),\quad F(x)=\sum_{n\leq x}f(n),\quad G(x)=\sum_{n\leq x}g(x).$$

Use the function

$$U(x)=\begin{cases}0 & 0<x<1\\ 1 & x\geq 1\end{cases}$$

and the associative law for $*$ to show that

$$H(x)=\sum_{n\leq x}f(n)G\left(\frac{x}{n}\right)=\sum_{n\leq x}g(n)F\left(\frac{x}{n}\right)$$

(b) Apply this to the cases $g(n)=1$ for all n, and $f(n)=\mu(n)$ or $\Lambda(n)$ to get

$$\sum_{n\leq x}\mu(n)\left[\frac{x}{n}\right]=1$$

and

$$\sum_{n\leq x}\Lambda(n)\left[\frac{x}{n}\right]=\log([x]!).$$

and use this to conclude.

(c) Use the property (which will be proved in the next exercise) that

$$\sum_{n\leq x}\log n=\int_1^x\log t\ dt+O(\log x).$$

The following formula is due to Euler.

Exercise 17. Let's use a bit of calculus. Given f such that the derivative f' is continuous over the interval $[a,b]$ where $0<a<b$, show that

$$\sum_{a<n\leq b}f(n)=\int_a^b f(t)dt+\int_a^b(t-[t])f'(t)dt+f(b)([b]-b)-f(a)([a]-a).$$

(1) Let $m=[a]$ and $k=[b]$. Then for integers n and $n-1$ in $[a,b]$ show that

$$\int_{n-1}^n[t]f'(t)dt=\int_{n-1}^n[n-1]f'(t)dt=\{nf(n)-(n-1)f(n-1)\}-f(n)$$

(2) Sum from $n=m+2$ to $n=k$ to show that

$$\int_{m+1}^k[t]f'(t)dt=kf(k)-(m+1)f(m+1)-\sum_{a<n\leq b}f(n).$$

(3) Combine this with the integration by parts

$$\int_a^b f(t)dt=bf(b)-af(a)-\int_a^b tf'(t)dt$$

to conclude.

Alternatively, prove the more general identity due to Abel: If $A(x) = \sum_{n \leq x} a(n)$ then

$$\sum_{a<n\leq b} a(n)f(n) = A(b)f(b) - A(a)f(a) - \int_a^b A(t)f'(t)dt.$$

Here's the idea.

(1) With notation like above, write the lefthand side

$$\sum_{n=m+1}^{k} a(n)f(n) = \sum_{n=m+1}^{k} (A(n) - A(n-1))f(n)$$

and then write this as

$$\sum_{n=m+1}^{k-1} A(n)(f(n) - f(n+1)) + A(k)f(k) - A(m)f(m+1).$$

(2) Then by the fundamental theorem of calculus this is

$$-\sum_{n=m+1}^{k-1} \int_n^{n+1} A(t)f'(t)dt + A(k)f(k) - A(m)f(m+1)$$

whence

$$-\int_{m+1}^{k} A(t)f'(t) + A(b)f(b) - \int_k^b A(t)f'(t)dt - A(a)f(a) - \int_a^{m+1} A(t)f'(t)dt$$

which is equal to the righthand side.

(3) Recover Exercise 17 by taking $a(n) = 1$ for all $n \geq 1$, so that $A(x) = [x]$, and use the integration by parts

$$\int_a^b tf'(t)dt = bf(b) - af(a) - \int_a^b f(t)dt.$$

3. It's prime time!

Definition 3.1. Let $x > 0$. Define $\pi(x)$ to be the number of primes less than or equal to x, i.e.,

$$\pi(x) = \sum_{p \leq x} 1.$$

Since we know there are infinitely many primes, it follows that $\pi(x) \to \infty$ as $x \to \infty$. C.F. Gauss conjectured in 1792 that

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$

This conjecture was proved in 1896 independently by Hadarmard and de la Vallée Poussin, and is referred to as the prime number theorem. It is the cornerstone of prime number theory. Our first goal is to show that the prime number theorem is equivalent to the statement that

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} \Lambda(n) = 1. \tag{1}$$

To start the journey, we'll need a function defined by Chebyshev.

Definition 3.2. Let $x > 0$. Define

$$\psi(x) = \sum_{n \leq x} \Lambda(n).$$

Then (1) is the same as saying

$$\lim_{x \to \infty} \frac{\psi(x)}{x} = 1.$$

To get a handle on $\pi(x)$, let's try to count some primes.

Exercise 18. (a) Let $x > 0$. Eratosthenes' sieve gives us

$$[x] - \sum_{p \leq x} \left[\frac{x}{p}\right] + \sum_{p_1 < p_2 \leq \sqrt{x}} \left[\frac{x}{p_1 p_2}\right] - \sum_{p_1 < p_2 < p_3 \leq \sqrt{x}} \left[\frac{x}{p_1 p_2 p_3}\right] + \cdots = \sum_{d} \mu(d) \left[\frac{x}{d}\right]$$

where the sum over d runs over integers d whose prime divisors $p|d$ satisfy $p \leq \sqrt{x}$. Show that this is equal to $\pi(x) - \pi(\sqrt{x}) + 1$.

(b) Rewrite the first expression as

$$x \sum_{d} \frac{\mu(d)}{d} - \sum_{d} \mu(d) \left\{\frac{x}{d}\right\},$$

where $\{x\}$ denotes the fractional part of x, ie.m $\{x\} = x - [x]$. Use the same logic in Exercise 10 to rewrite the first term to get

$$x \prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right) - \sum_{d} \mu(d) \left\{\frac{x}{d}\right\}.$$

The first term we'll think of as the main term, and the second term the error term. Let's get an estimate for the main term.

(c) Suppose you know that the main term satisfies

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) = \frac{e^{-\gamma}}{\log x} \left(1 + O\left(\frac{1}{\log x}\right)\right). \tag{2}$$

for some constant γ. Explain why this violates the expectation given the prime number theorem.

Now let's prove (2).

Exercise 19. First we need to show that

$$\sum_{p\le x} \log p \left[\frac{x}{p}\right] = x\log x + O(x).$$

(1) We'll need to use the result of Exercise 16, as you might guess. First rewrite

$$\sum_{n\le x} \Lambda(n)\left[\frac{x}{n}\right] = \sum_{p} \sum_{\substack{m=1 \\ p^m\le x}}^{\infty} \Lambda(p^m)\left[\frac{x}{p^m}\right]$$

and then write it as

$$\sum_{p}\sum_{m=1}^{\infty} \log p\left[\frac{x}{p^m}\right] = \sum_{p}\log p\left[\frac{x}{p}\right] + \sum_{p}\sum_{m=2}^{\infty}\log p\left[\frac{x}{p^m}\right]$$

(2) Show that the last sum is $O(x)$ by showing that it is bounded above by

$$\sum_{p}\sum_{m=2}^{\infty}\log p\frac{x}{p^m} = x\sum_{p\le x}\frac{\log p}{p(p-1)}$$

which is in turn bounded above by

$$x\sum_{n=1}^{\infty}\frac{\log n}{n(n-1)} = O(x).$$

Exercise 20. Second, we need to show that

$$\sum_{p\le x}\frac{\log p}{p} = \log x + O(1).$$

(1) Convince yourself first that

$$\sum_{p\le x}\log p\left[\frac{x}{p}\right] = \sum_{p\le x}\log p\left(\frac{x}{p}+O(1)\right) = x\sum_{p\le x}\frac{\log p}{p} + O\left(\sum_{p\le x}\log p\right)$$

Then assume that the error term is equal to $O(x)$ and apply Exercise 19.

(2) To prove that the error term is equal to $O(x)$, specialize the proof of Theorem 4.8(b) in Apostol. Let

$$S(x) = \sum_{p\le x}\log p, \qquad T(x) = \sum_{p\le x}\log p\left[\frac{x}{p}\right]$$

Show that

$$T(x) - 2T\left(\frac{x}{2}\right) \ge \sum_{x/2<p\le x}\left[\frac{x}{p}\right]\log p = S(x) - S\left(\frac{x}{2}\right)$$

Then show that

$$T(x) - 2T\left(\frac{x}{2}\right) = O(x)$$

so

$$S(x) - S\left(\frac{x}{2}\right) \le Kx$$

for some constant K. Then replace x with $x/2, x/4, x/8, \dots$ to conclude that

$$S(x) \leq Kx\left(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\dots\right) = 2Kx.$$

Exercise 21. One more thing. We want to show that for some constant A we have for all $x \geq 2$,

$$\sum_{p\leq x}\frac{1}{p} = \log\log x + A + O\left(\frac{1}{\log x}\right).$$

(1) We'll use Abel summation with

$$A(x) = \sum_{p\leq x}\frac{\log p}{p}$$

and define

$$b(n) = \begin{cases} 1 & n \text{ prime} \\ 0 & \text{otherwise} \end{cases}.$$

Then write

$$\sum_{p\leq x}\frac{1}{p} = \sum_{n\leq x}\frac{b(n)}{n}, \qquad A(x) = \sum_{n\leq x}\frac{b(n)}{n}\log n$$

Take $f(t) = 1/\log t$ and get

$$\sum_{p\leq x}\frac{1}{p} = \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t\log^2 t}dt.$$

(2) Apply Exercise 20 as $A(x) = \log x + R(x)$ to the last expression to get

$$\sum_{p\leq x}\frac{1}{p} = 1 + O\left(\frac{1}{\log x}\right) + \int_2^x \frac{1}{t\log t}dt + \int_2^x \frac{R(t)}{t\log^2 t}dt.$$

Show that the third term is

$$\log\log x - \log\log 2$$

and the fourth term is equal to

$$\int_2^\infty \frac{R(t)}{t\log^2 t}dt - \int_x^\infty \frac{R(t)}{t\log^2 t}dt = \int_2^\infty \frac{R(t)}{t\log^2 t}dt + O\left(\frac{1}{\log x}\right).$$

(3) Conclude with the constant

$$A = 1 - \log\log 2 + \int_2^\infty \frac{R(t)}{t\log^2 t}dt.$$

Exercise 22. Finally! Write

$$\prod_{p\leq x}\left(1-\frac{1}{p}\right) = \exp\left(\sum_{p\leq x}\log\left(1-\frac{1}{p}\right)\right)$$

and apply the last exercise to conclude that it is equal to

$$\frac{e^{-\gamma}}{\log x}\left(1+O\left(\frac{1}{\log x}\right)\right).$$

Whew, that was a work out! Good work! We're all primed and ready to go sieving.

4. Eratosthenes

Now we are ready to formalize the preceding discussion. Fix a positive real number x. Let $\mathscr{A}$ be a sequence (a_n) for $n \leq x$, and let $\mathscr{P}$ be a set of primes. Let z be a positive real number, and set

$$P(z) = \prod_{\substack{p \in \mathscr{P} \\ p<z}} p.$$

Define $S(\mathscr{A}, \mathscr{P}, z)$ to be the cardinality of the set complement $\mathscr{A} \setminus \cup_{p|P(z)} \mathscr{A}_p$. The sieve problem is to estimate the size of this "sifting function"

$$\begin{aligned} S(\mathscr{A}, \mathscr{P}, z) &= \left|\mathscr{A} \setminus \cup_{p|P(z)} \mathscr{A}_p\right| \\ &= \sum_{\substack{n \leq x \\ (n, P(z))=1}} a_n. \end{aligned}$$

Exercise 23. Argue that the Möbius function satisfies

$$\sum_{\substack{d|n \\ d|P(z)}} \mu(d) = \begin{cases} 1 & \text{if } (n, P(z)) = 1 \\ 0 & \text{if } (n, P(z)) > 1 \end{cases}$$

and use this to show that

$$S(\mathscr{A}, \mathscr{P}, z) = \sum_{n \leq x} a_n \sum_{d|(n, P(z))} \mu(d) = \sum_{d|P(z)} \mu(d) \sum_{\substack{n \leq x \\ n \equiv 0 \bmod d}} a_n.$$

Moving on, let's denote

$$A_d(x) = \sum_{\substack{n \leq x \\ n \equiv 0 \bmod d}} a_n,$$

Assume we can write it in the form

$$A_d(x) = g(x)X + r_d(x)$$

where X is some approximation to

$$A(x) = A_1(x) = \sum_{n \leq x} a_n,$$

and $g(d)$ is the density function and $r_d(x)$ is a remainder term that should be small at least on average. In Eratosthenes' sieve, these were

$$A_d(x) = \left[\frac{x}{d}\right], \quad g(d) = \frac{1}{d}, \quad r_d(x) = -\left\{\frac{x}{d}\right\},$$

also $\mathscr{A} = \{n \in \mathbb{N} : n \leq x\}$ and $\mathscr{P}$ was the set of all primes. Putting these together, our sifting function becomes

$$S(\mathscr{A}, \mathscr{P}, z) = X \sum_{d|P(z)} \mu(d)g(d) + \sum_{d|P(z)} \mu(d) r_d(x).$$

The function $g(d)$ acts like a probability, approximating the fraction of the total mass of $\mathscr{A}$ which resides in the multiples of d.

Now for a quick stop along the way: The general version of Eratosthenes' sieve is as follows:

Theorem 4.1. *Let $\mathscr{A}$ be a sequence of nonnegative real numbers, and $\mathscr{P}$ a set of primes. Assume that the density function satisfies*

$$\sum_{p\le x, p\in\mathscr{P}} \frac{g(p)}{p}\log p = \kappa \log x + O(1),$$

for some $0\le\kappa$, and that the remainder terms satisfy $|r_d(x)|\le g(d)d$, for all $d\le x$. Then we have

$$S(\mathscr{A},\mathscr{P},z) = XV(z) + O\left(\left(X+\frac{x}{\log z}\right)(\log z)^{\kappa+1}\exp\left(-\frac{\log x}{\log z}\right)\right)$$

where

$$V(z) = \prod_{p\mid P(z)}(1-\frac{g(p)}{p}).$$

Let's take this for granted. Doing so, one can quickly recover the celebrated 1919 result of Viggo Brun, showing that

$$\sum_{\substack{p\\ p+2 \text{ prime}}} \frac{1}{p} < \infty. \tag{3}$$

Exercise 24. We first have to show that number of primes $p\le x$ such that $p+2$ is prime is equal to

$$O\left(\frac{x(\log\log x)^2}{\log^2 x}\right)$$

(1) Do so by taking $\mathscr{A}=\{n\in\mathbb{N}: n\le x\}$ and $\mathscr{P}$ the set of all primes. Let z be a positive real number that we will choose later. For each $p<z$, we consider the residue classes 0 and -2 mod p. Note that $\mathscr{A}_p$ is empty for $p>x+2$. Also take $X=x$, $\kappa=2$ and $g(p)=2$. (Why?) Apply this to Eratosthenes' sieve to get

$$S(\mathscr{A},\mathscr{P},z) = xV(z) + O\left(x(\log z)^3\exp\left(-\frac{\log x}{\log z}\right)\right)$$

(2) Argue that

$$V(z)\le \exp\left(-\sum_{p<z}\frac{2}{p}\right) = O\left((\log z)^{-2}\right)$$

Then choose z such that $\log z = \log x/A\log\log x$ for some large positive constant to deduce that

$$S(\mathscr{A},\mathscr{P},z) = O\left(\frac{x(\log\log x)^2}{\log^2 x}\right).$$

(3) Finally, conclude using the fact that the number of twin primes up to x is bounded by

$$\pi(z) + S(\mathscr{A},\mathscr{P},z) \le z + S(\mathscr{A},\mathscr{P},z).$$

Exercise 25. Prove (3) by using partial summation and the previous exercise, to show that

$$\sum_{\substack{p\\ p+2 \text{ prime}}} \frac{1}{p} = O\left(\int_2^\infty \frac{(\log\log t)^2}{t\log^2 t}dt\right)$$

Very nice work! We've developed some basic sieve technology and proved a classical result on twin primes. Our next result on twin primes will take the rest of our time. Here we go!

5. Selberg

Let's take stock of where we are at the moment. Given some set $\mathscr{A}$, we want to get rid of as much elements as possible that we don't want, in the form of subsets $\mathscr{A}_p$ where $p|P(z)$. The sets $\mathscr{A}$ and $\mathscr{P}$ should be chosen according to the quantity we are trying to estimate. In this class, we have been interested in the prime counting function $\pi(x)$, and the set of twin primes. (There are many other applications that we will not consider.)

Let's go back to Eratosthenes' sieve. Our count for primes p between $z$and x was

$$\Phi(x,z) := \sum_{n\le x}\sum_{d|(n,P(z))}\mu(d)$$

(the terms a_n are all equal to 1, since we are just counting). The sifting function $S(\mathscr{A},\mathscr{P},z)$ estimates the size of the remainder. One natural thing to do is to try to get an upper bound for $S(\mathscr{A},\mathscr{P},z)$.

Exercise 26. (a) (Easy) Given any sequence (λ_n) such that $\lambda_1 = 1$, we have for any $n \in \mathbf{N}$,

$$\sum_{d|n}\mu(d) \le \Big(\sum_{d|n}\lambda_d\Big)^2.$$

(b) Argue that

$$\Phi(x,z) \le \sum_{n\le x}\left(\sum_{d_1,d_2|(n,P(z))}\lambda_{d_1}\lambda_{d_2}\right) = \sum_{d_1,d_2|P(z)}\lambda_{d_1}\lambda_{d_2}\sum_{\substack{n\le x\\ \mathrm{lcm}(d_1,d_2)|n}} 1.$$

(c) Use the fact that the number of $n \le x$ such that $d|n$ is equal to

$$\Big[\frac{x}{d}\Big] = \frac{x}{d} + O(1)$$

to conclude that

$$\Phi(x,z) \le x\sum_{d_1,d_2|P(z)}\frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} + O\left(\sum_{d_1,d_2|P(z)}|\lambda_{d_1}||\lambda_{d_2}|\right),$$

where we have used the notation $\mathrm{lcm}(a,b) = [a,b]$. (Also $\gcd(a,b) = (a,b)$,)

Notice that if $\lambda_d = 0$ for all $d > z$ then we can write the latter as

$$\Phi(x,z) \le x\sum_{d_1,d_2<z}\frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} + O\left(\sum_{d_1,d_2<z}|\lambda_{d_1}||\lambda_{d_2}|\right). \tag{4}$$

and if $|\lambda_d| \le 1$ then the error term would be on the order of $O(z^2)$, and if $z < x$ then the error is smaller than in Eratosthenes's sieve. If we worked out the error term for $\pi(x)$ using Erathosthenes' sieve, we would conclude that

$$\pi(x) = O\left(\frac{x}{\log x}\log\log x\right)$$

which is way too big! Now we will develop Selberg's sieve, which will get us to

$$\pi(x) = O\left(\frac{x}{\log x}\right)$$

which is of the right order.

Let's first estimate the main term in (4). We will take for granted the following generalization of the Möbius inversion formula.

Lemma 5.1. *Let $D \subset \mathbf{N}$ be such that if $d \in D$ and $f|d$ then $f \in D$ also. Then given arithmetic functions f, g then we have*

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

if and only if

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

This is called the dual Möbius inversion formula.

Proof. Similar to Exercise 13. □

Now let's look at the main term.

Exercise 27. (a) Use the fact that

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

and Exercise 8 to write

$$\sum_{d_1,d_2\leq z} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1, d_2]} = \sum_{d_1,d_2\leq z} \frac{\lambda_{d_1}\lambda_{d_2}}{d_1 d_2}(d_1, d_2) = \sum_{d_1,d_2\leq z} \frac{\lambda_{d_1}\lambda_{d_2}}{d_1 d_2} \sum_{d|(d_1,d_2)} \varphi(d).$$

Then switch sums to get

$$\sum_{d\leq z} \varphi(d) \sum_{\substack{d_1,d_2\leq z \\ d|(d_1,d_2)}} \frac{\lambda_{d_1}\lambda_{d_2}}{d_1 d_2} = \sum_{d\leq z} \varphi(d) \left(\sum_{\substack{d'\leq z \\ d|d'}} \frac{\lambda_{d'}}{d'} \right)^2.$$

If we denote

$$u_d := \sum_{\substack{d'\leq z \\ d|d'}} \frac{\lambda_{d'}}{d'},$$

we can then write the last expresison as

$$\sum_{d\leq z} \varphi(d) u_d^2. \tag{5}$$

We would like to minimize this with respect to λ_d.

Exercise 28. (a) Justify applying the dual Möbius inversion to u_d to get

$$\frac{\lambda_d}{d} = \sum_{d|d'} \mu\left(\frac{d'}{d}\right) u_{d'}.$$

Recall that $\lambda_1 = 1$ and $\lambda_d = 0$ for $d > z$. Use this to show that $u_d = 0$ for any $d > z$, and

$$\sum_{d\leq z} \mu(d) u_d = 1,$$

and use this to write

$$\sum_{d\le z}\varphi(d)u_d^2=\sum_{d\le z}\varphi(d)\left(u_d-\frac{\mu(d)}{\varphi(d)V(z)}\right)^2+\frac{1}{V(z)},$$

where

$$V(z):=\sum_{n\le z}\frac{\mu^2(n)}{\varphi(n)}.$$

(b) Observe then that choosing

$$u_d=\frac{\mu(d)}{\varphi(d)V(z)}$$

minimizes the value of (5), giving a minimal value $1/V(z)$. Doing so, conclude that

$$\lambda_d=d\sum_{d|e}\mu\left(\frac{e}{d}\right)\frac{\mu(e)}{\varphi(e)V(z)},$$

and hence

$$\Phi(x,z)\le\frac{x}{V(z)}+O\left(\sum_{d_1,d_2\le z}|\lambda_{d_1}||\lambda_{d_2}|\right).$$

It remains to estimate the error term. This one is quicker.

Exercise 29. Using our choice of λ_d, show that

$$V(z)\lambda_d=d\sum_{t\le\frac{z}{d}}\frac{\mu(t)\mu(dt)}{\varphi(dt)}=d\sum_{\substack{t\le\frac{z}{d}\\(t,d)=1}}\frac{\mu^2(t)\mu(d)}{\varphi(d)\varphi(t)}$$

here we have used the property that $\varphi(ab)=\varphi(a)\varphi(b)$ for any pair of coprime integers $(a,b)=1$. Then rewrite this as

$$\mu(d)\prod_{p|d}\left(1+\frac{1}{p-1}\right)\sum_{\substack{t\le\frac{z}{d}\\(t,d)=1}}\frac{\mu^2(t)}{\varphi(t)}$$

using Exercise 10. Infer that

$$|V(z)||\lambda(d)|\le|V(z)|,$$

and so $|\lambda_d|\le 1$ for all d.

(The 'infer' follows from the property that

$$\prod_{p|d}\left(1+\frac{1}{p-1}\right)=\prod_{p|d}\left(1+\frac{1}{\varphi(p)}\right)$$
$$=1+\sum_{p_i}\frac{1}{\varphi(p_i)}+\cdots+\frac{1}{\varphi(p_1)\dots\varphi(p_k)}=\sum_{m|d}\frac{\mu^2(m)}{\varphi(m)}$$

then

$$\prod_{p|d}\left(1+\frac{1}{p-1}\right)\sum_{\substack{t\le\frac{z}{d}\\(t,d)=1}}\frac{\mu^2(t)}{\varphi(t)}$$
$$=\sum_{m|d}\frac{\mu^2(m)}{\varphi(m)}\sum_{\substack{t\le\frac{z}{d}\\(t,d)=1}}\frac{\mu^2(t)}{\varphi(t)}$$
$$=\sum_{m|d}\sum_{\substack{t\le\frac{z}{d}\\(t,d)=1}}\frac{\mu^2(m)}{\varphi(m)}\frac{\mu^2(t)}{\varphi(t)}$$

since $(d,t)=1$, we write this as

$$\sum_{m|d}\sum_{\substack{dt\le z\\(t,d)=1}}\frac{\mu^2(mt)}{\varphi(mt)}\le\sum_{n\le z}\frac{\mu^2(n)}{\varphi(n)}=V(z)$$

where the inequality follows because the lefthand sum is contained in the right hand sum.)

Exercise 30. (a) Use the last two exercises to conclude that

$$\Phi(x,z)\le\frac{x}{V(z)}+O(z^2).$$

To get an upper bound for $\pi(x)\le\Phi(x,z)+z$, we need a lower bound for $V(z)$ and choose z appropriately.

(b) Argue that

$$V(z)=\sum_{n\le z}\frac{\mu^2(n)}{\varphi(n)}\ge\sum_{n\le z}\frac{\mu^2(n)}{n}=\sum_{n\le z}\frac{1}{n}-\sum_{n\le z}{}'\frac{1}{n}$$

where the summation $\sum'$ is taken over non-squarefree integers n.

Remark 5.2. the sieve of Eratosthenes can be used to show that the squarefree numbers up to x is asymptotically equal to $\frac{6}{\pi^2}x+O(\sqrt{x})$ which is greater than $\frac{1}{2}x$ But here is a different argument I like:

$$\sum_{n\le z}\frac{\mu^2(n)}{n}=\prod_{p\le z}\left(1+\frac{1}{\mu(p)}\right)=\prod_{p\le z}\left(1+\frac{1}{p}\right)=\prod_{p\le z}\frac{1-\frac{1}{p^2}}{1-\frac{1}{p}}$$

this is asymptotically equal to

$$\frac{\log x}{e^{-\gamma}}\prod_{p\le z}\left(1-\frac{1}{p^2}\right)$$

and note that

$$\prod_{p\le z}\left(1-\frac{1}{p^2}\right)\le\prod_{p}\left(1-\frac{1}{p^2}\right)=\sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{6}{\pi^2}$$

The last formula is due to Euler. It is the value of the zeta function $\zeta(2)$.

Then it follows that

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

Hence

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

and so if we pick

$$z = \left(\frac{x}{\log x}\right)^2$$

then we can conclude that

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

Remark 5.3. What we showed is that

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

meaning that there is some constant M such that $\pi(x) \leq Mx/\log x$. To get the big-O estimate, one actually also needs to show the lower bound,

$$\pi(x) \gg \frac{x}{\log x},$$

meaning that there is some other constant m such that $\pi(x) \geq mx/\log x$. Then choosing $K = \max(M, m)$ gives the big-O result. This is a theorem due to Chebyshev, which was obtained long before the prime number theorem was proved. It is the basis of the elementary proof of the PNT by Erdös and Selberg.

Now here is the general statement of the Selberg sieve. Let $\mathscr{A}$ be any finite sequence (a_n) of nonnegative numbers, $\mathscr{P}$ a set of distinct primes. Fix $z > 0$. For any $d|P(z)$, denote $\mathscr{A}_d = \cap_{p|d}\mathscr{A}_p$. We want to prove the following theorem.

Theorem 5.4 (Selberg)**.** *Assume that there exists real numbers $X > 0$ and R_d, and a multiplicative function f such that $f(p) > 1$ for all $p \in \mathscr{P}$.*

$$|\mathscr{A}_d| = \frac{X}{f(d)} + R_d.$$

Let h be the function determined by the inversion formula,

$$h(n) = \sum_{n|d} \mu\left(\frac{n}{d}\right) f(d), \qquad f(n) = \sum_{n|d} h(d),$$

and set

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

Then we have the upper bound

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

What a doozy! But as the general form of Eratosthenes' sieve, we will take this as a black box.

6. Maynard/Tao/Zhang

6.1. **Twin primes, double the fun.** Now we embark on the celebrated proof of the bounded gaps between primes, following Maynard's proof (2015).

Let p_n denote the sequence of consecutive primes. Goldston, Pintz, and Yıldırım [GPY, 2009] improved on Selberg's sieve to show that

$$\liminf_n \frac{p_{n+1} - p_n}{\log p_n} = 0$$

which tells us that the gaps between primes do not all grow arbitrary large. Zhang's breakthrough result (2014) is that

$$\liminf_n (p_{n+1} - p_n) \leq 70\ 000\ 000$$

which shows that the there are infinitely many bounded gaps between primes. This excitement led many others to refine the bound, going down to 4860 by optimizing and refining Zhang's arguments. Maynard (and Tao), using independent methods showed that

$$\liminf_n (p_{n+1} - p_n) \leq 600.$$

Combining these methods, the bound was brought down to 246, which is the current state of humanity's knowledge. Assuming a generalized form of the Elliot-Halberstam conjecture, the bound can be brought down to 6.

6.2. **The GPY sieve.** All of these methods begin with the GPY sieve, so let's start there. First, the primes in arithmetic progression satisfy an analogue of the prime number theorem. Let $\pi(x; q, a)$ be the number of primes in the arithmetic progression $a + qn$ with $(a, q) = 1$. Then

$$\pi(x; q, a) \sim \frac{\pi(x)}{\varphi(q)}.$$

Now given $\theta > 0$, we say that the primes have *level of distribution* θ if for all $A > 0$ we have

$$\sum_{q \leq x^\theta} \max_{(a,q)=1} \left| \pi(x; q, a) - \frac{\pi(x)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^A}.$$

Then the Bombieri Vinogradov theorem states that the primes have level of distribution θ for every $\theta < 1/2$. (The Elliot-Halberstam conjecture states that this true for all $\theta < 1$.)

GPY showed that if the above was true for some $\theta > 1/2$, then

$$\liminf_n (p_{n+1} - p_n) < \infty$$

Maynard's work removes the barrier of $\theta = 1/2$.

Let us turn to the Maynard's version of the GPY sieve.

Definition 6.1. Let $\mathscr{H} = \{h_1, \ldots, h_k\}$ be a set of distinct nonnegative integers. We say $\mathscr{H}$ is *admissible* if for every prime p, there exists an integer a_p such that for all $h \in \mathscr{H}$, $a_p \neq h \bmod p$.

Fix an admissible set $\mathscr{H}$. Let $\chi_{\mathbb{P}}(n) = 1$ if n is prime and 0 otherwise. Given $\rho > 0$ and a nonnegative sequence w_n, consider the sum

$$S(N, \rho) = \sum_{N \leq n < 2N} \left(\sum_{i=1}^{k} \chi_{\mathbb{P}}(n + h_i) - \rho \right) w_n.$$

If we can show that $S(N,\rho) > 0$, then at least one term in the sum over n must have a positive contribution. Since w_n is nonnegative, this means that there is some integer $n \in [N, 2N)$ such that at least $[\rho+1]$ of the $n+h_i$ are prime. Thus if $S(N,\rho) > 0$ for all N large, then there are infinitely many integers n for which at least $[\rho+1]$ of the $n+h_i$ are prime, so there are infinitely many intervals of bounded length containing $[\rho+1]$ primes.

The weights w_n are chosen to mimic the λ_n in Selberg's sieve. Estimating $S(N,\rho)$ is a k-dimensional sieve problem, and GPY considered for example

$$w_n = \Big(\sum \lambda_d\Big)^2, \quad \lambda_d = \mu(d)F(\log R/d)$$

where the sum runs over divisors $d < R$ of the product $(n+h_1)\dots(n+h_k)$, for some $R > 0$ and smooth function $F(x)$. Maynard chooses instead

$$w_n = \Big(\sum_{\substack{d_i|(n+h_i)\\ 1\le i\le k}} \lambda_{d_1,\dots,d_k}\Big)^2$$

where the weights are allowed to depend on each d_i individually.

Remark 6.2. Moreover, we will set

$$w_n = 0$$

unless n belongs to a fixed residue class $v_0 \bmod W$ where

$$W = \prod_{p\le D_0} p, \quad D_0 = \log\log\log N$$

so that $W \ll (\log\log N)^2$ by the PNT. This is a technical modification to remove the some of the effect of small prime divisors. The Chinese Remainder Theorem states that if $n_1,\dots,n_k$ are pairwise relatively prime, and $a_1,\dots,a_k$ are integers, then there exists an integer x such that

$$x \equiv a_i \pmod{n}_i$$

for $i = 1,\dots,k$. (Show that the case $k = 2$ follows from Bezout's identity.) Applying this, we can choose v_0 such that $v_0 + h_i$ is relatively prime to W for each $i = 1,\dots,k$, since $\mathscr{H}$ is admisible.

6.3. **Outline of proof.** Based on these choices, estimating $S(N,\rho)$ amounts to estimating the sums

$$S_1 = \sum_{\substack{N\le n<2N\\ n\equiv v_0 \bmod W}} \Big(\sum_{\substack{d_i|(n+h_i)\\ 1\le i\le k}} \lambda_{d_1,\dots,d_k}\Big)^2$$

$$S_2 = \sum_{\substack{N\le n<2N\\ n\equiv v_0 \bmod W}} \left(\sum_{i=1}^k \chi_{\mathbb{P}}(n+h_i)\right)\Big(\sum_{\substack{d_i|(n+h_i)\\ 1\le i\le k}} \lambda_{d_1,\dots,d_k}\Big)^2,$$

whence

$$S := S(N,\rho) = S_2 - \rho S_1.$$

We choose the weights $\lambda_{d_1,\dots,d_k}$ as follows. Suppose the primes have level of distribution $\theta > 0$. Let $R = N^{\theta/2-\delta}$ for some fixed $\delta > 0$. Then define

$$\lambda_{d_1,\dots,d_k} = \left(\prod_{i=1}^k \mu(d_i)d_i\right)\sum_{i=1}^k \sum_{\substack{r_i\\ d_i|r_i,\\ (r_i,W)=1}} \frac{\mu(r_1\dots r_k)^2}{\varphi(r_1)\dots\varphi(r_k)} F\left(\frac{\log r_1}{\log R},\dots,\frac{\log r_k}{\log R}\right)$$

if $(d_1 \dots d_k, W) = 1$ and $\lambda_{d_1,\dots,d_k} = 0$ otherwise. Also take F to be a smooth function on $\mathbf{R}^k$ that is zero outside of the set $\mathcal{R}_k = \{(x_1, \dots, x_k) \in [0,1]^k : \sum_{i=1}^k x_i \le 1\}$.

Proposition 6.3 (Proposition 4.1)**.** *Let w_n and $\lambda_{d_1,\dots,d_k}$ be chosen as above. Then*

$$S_1 = \frac{(1+o(1))\varphi(W)^k N(\log R)^k}{W^{k+1}} I_k(F)$$

$$S_2 = \frac{(1+o(1))\varphi(W)^k N(\log R)^{k+1}}{W^{k+1}\log N} \sum_{m=1}^k J_k^{(m)}(F)$$

where

$$I_k(F) = \int_0^1 \cdots \int_0^1 F(t_1, \dots, t_k)^2$$

$$J_k^{(m)}(F) = \int_0^1 \cdots \int_0^1 \left(F(t_1, \dots, t_k)dt_m\right)^2 dt_1 \dots dt_{m-1} dt_{m+1} \dots dt_k$$

provided that $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$ for each m.

If S_2 is large compared to S_1, the GPY method will show that there are infinitely many integers n such that several of the integers $n + h_i$ are prime. The following proposition makes this precise. First, let $\mathcal{S}_k$ be the set of integrable functions $F : [0,1]^k \to \mathbf{R}$ such that $I_k(F) \neq 0$, $J_k^{(m)}(F) \neq 0$, for $1 \le m \le k$, and $F(x) = 0$ for all $x \notin \mathcal{R}_k$. Let

$$M_k = \sup_{F \in \mathcal{S}_k} \frac{1}{I_k(F)} \sum_{m=1}^k J_k^{(m)}(F),$$

and let $r_k = \lceil \theta M_k/2 \rceil$, the smallest integer greater than $\theta M_k/2$.[1]

Proposition 6.4 (Proposition 4.2)**.** *Suppose the primes have level of distribution $\theta > 0$. Let $\mathscr{H}$ be an admissible set, and $\delta > 0$. Then there exists infinitely many integers n such that at least r_k of the $n + h_i, 1 \le i \le k$ are prime. In particular,*

$$\liminf_n (p_{n+r_k-1} - p_n) \le \max_{1 \le i,j \le k} (h_i - h_j).$$

We first show how Maynard's Proposition 4.1 proves Proposition 4.2, and how Proposition 4.2 proves the bounded gaps.

Proof of Prop 6.4. By the definition of M_k, we can choose $F_0 \in \mathcal{S}_k$ such that

$$\frac{1}{I_k(F_0)} \sum_{m=1}^k J_k^{(m)}(F_0) > (M_k - \delta) > 0$$

for some small δ. Since F_0 is integrable, there is a smooth function F_1 such that

$$\frac{1}{I_k(F_1)} \sum_{m=1}^k J_k^{(m)}(F_1) > (M_k - 2\delta) > 0.$$

Roughly, we approximate F_0 with F_1, and hence the integral of F_0^2 with the integral of F_1^2.

[1] This is the ceiling function.

Now, using the Proposition 6.3 above, we write

$$S = \frac{\varphi(W)^k N(\log R)^k}{W^{k+1}} \left(\frac{\log R}{\log N} \sum_{m=1}^{k} J_k^{(m)}(F_1) - \rho I_k(F_1) + o(1) \right)$$

and from our choice of F_1 we have

$$S \geq \frac{\varphi(W)^k N(\log R)^k I_k(F_1)}{W^{k+1}} \left(\left(\frac{\theta}{2} - \delta \right)(M_k - 2\delta) - \rho + o(1) \right)$$

where we recall that $R = N^{\theta/2-\delta}$.

If $\rho = \theta M_k/2 - \epsilon$ then, by choosing δ suitably small (depending on ϵ), we can make

$$\left(\frac{\theta}{2} - \delta \right)(M_k - 2\delta) - \rho > 0$$

and hence $S > 0$ for all large N. Thus there exists infinitely many integers n for which at least $[\rho + 1]$ of the $n + h_i$ are prime. And since $[\rho + 1] = \lceil \theta M_k/2 \rceil$ for ϵ small enough, the proposition follows. □

Thus if the primes have a fixed level of distribution θ, then to show that the existence of many of the $n + h_i$ being prime for infinitely many n we only require a suitable lower bound for M_k. We recall the result needed for twin primes.

Proposition 6.5 (Proposition 4.3(1))**.** *Let M_k be given as above. Then $M_{105} > 4$.*

We now state the theorem on bounded gaps between primes.

Theorem 6.6 (Maynard)**.**

$$\liminf_n (p_{n+1} - p_n) \leq 600.$$

Proof. We show how the theorem follows from Propositions 6.4 and 6.5. By the Bombieri-Vinogradov theorem, the primes have level of distribution $\theta = \frac{1}{2} - \epsilon$ for all $0 < \epsilon < \frac{1}{2}$. Thus we have for ϵ sufficiently small,

$$\theta M_{105}/2 > 1,$$

so by Proposition 6.4, we have

$$\liminf_n (p_{n+1} - p_n) \leq \max_{1 \leq i,j \leq 105} (h_i - h_j)$$

since $r_{105} = 2$, for any admissible set $\mathscr{H}$. By numerical computation, one can choose $\mathscr{H}$ such that $0 \leq h_1 \leq \dot{\leq} h_k$ and $h_k - h_1 \leq 600$. □

It remains to prove Propositions 6.4 and 6.5. The plan is as follows: we first rewrite the sums S_1 and S_2 by a change of variable, which will lead us to a variable y. Then finding a proper choice of y will yield the proof of Proposition 6.4. For Proposition 6.5, we will use calculus techniques to find the bound that we need.

6.4. **The S_1 term.** Assume the primes have fixed level of distribution θ, and let $R = N^{\theta/2-\delta}$. We assume that $\lambda_{d_1,\dots,d_k} = 0$ unless $d = d_1 \cdots d_k < R$, $(d, W) = 1$, and $\mu(d)^2 = 1$. We note that this last condition means that $(d_i, d_j) = 1$ for all $i \neq j$.

Lemma 6.7. *Let*

$$y_{r_1,\dots,r_k} = \left(\prod_{i=1}^{k} \mu(r_i)\varphi(r_i)\right) \sum_{\substack{d_1,\dots,d_k \\ r_i | d_i \forall i}} \frac{\lambda_{d_1,\dots,d_k}}{d_1 \cdots d_k},$$

and $y_{\max} = \sup_{r_i} |y_{r_1,\dots,r_k}|$. *Then*

$$S_1 = \frac{N}{W} \sum_{r_1,\dots,r_k} \frac{y^2_{r_1,\dots,r_k}}{\varphi(r_1)\dots\varphi(r_k)} + O\left(\frac{y^2_{\max}\varphi(W)^k N(\log R)^k}{W^{k+1}D_0}\right).$$

Proof. Looking at S_1, we expand the square and switch sums, as we've done with the Selberg sieve,

$$S_1 = \sum_{\substack{N \leq n < 2N \\ n \equiv v_0 \bmod W}} \Big(\sum_{\substack{d_i | (n+h_i) \\ 1 \leq i \leq k}} \lambda_{d_1,\dots,d_k} \Big)^2 = \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}} \lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k} \sum_{\substack{N \leq n < 2N \\ n \equiv v_0 \bmod W \\ [d_i,e_i] | (n+h_i) \forall i}} 1.$$

Recall that $[a, b]$ denotes the lcm of a, b. Using the Chinese Remainder Theorem, we write the inner sum as over a single residue class modulo $q = W\prod_{i=1}^{k}[d_i, e_i]$ provided that the integers $W, [d_1, e_1], \dots, [d_k, e_k]$ are pairwise coprime. In this case the inner sum is $N/q + O(1)$. If they are not pairwise coprime, the inner sum is empty because $\mathscr{H}$ is admissible. We thus write

$$S_1 = \frac{N}{W} \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}}{}' \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{[d_1, e_1]\dots[d_k, e_k]} + O\Big(\sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k}}{}' |\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}| \Big)$$

where the $'$ indicates the sum is restricted such that $W, [d_1, e_1], \dots, [d_k, e_k]$ are pairwise coprime. Let us denote

$$\lambda_{\max} = \sup_{d_1,\dots,d_k} |\lambda_{d_1,\dots,d_k}|.$$

Since $\lambda_{d_1,\dots,d_k}$ is nonzero only when $d_1 \cdots d_k < R$, the error term contributes

$$\ll \lambda^2_{\max} \Big(\sum_{d<R} \tau_k(d) \Big)^2 \ll \lambda^2 R^2 (\log R)^{2k}$$

where $\tau_k(n)$ is the number of ways of writing n as a product of k natural numbers. We'll take the latter estimate for granted. Its contribution will be negligible.

For the main term, we wish to remove the dependence between d_i and e_i. Using the identity

$$\frac{1}{[d_i, e_i]} = \frac{1}{d_i e_i} \sum_{u_i | (d_i, e_i)} \varphi(u_i) = \frac{1}{d_i e_i} \sum_{u_i | d_i, e_i} \varphi(u_i)$$

we write the main term as

$$\frac{N}{W} \sum_{u_1,\dots u_k} \prod_{i=1}^{k} \varphi(u_i) \sum_{\substack{d_1,\dots,d_k \\ e_1,\dots,e_k \\ u_i | d_i, e_i}}{}' \frac{\lambda_{d_1,\dots,d_k}\lambda_{e_1,\dots,e_k}}{d_1 \cdots d_k e_1 \cdots e_k}.$$

Recall that $\lambda_{d_1,\dots,d_k}$ is supported on $d_1, \dots, d_k$ with $(d_i, W) = 1$ for each i and $(d_i, d_j) = 1$ for all $i \neq j$. Thus we can drop the requirement that W is coprime to each of the $[d_i, e_i]$ in the

summation, since they contribute nothing. Similarly, we may also drop the requirement that the d_i and e_i are respectively pairwise coprime. The remaining restriction from the requirement that $W, [d_1, e_1], \dots, [d_k, e_k]$ is coprime is that $(d_i, e_j) = 1$ for $i \neq j$.

We can remove this latter restriction by multiplying by

$$\sum_{s_{i,j} | d_i, e_j} \mu(s_{i,j})$$

for each $i \neq j$. We can then write the main term as

$$\frac{N}{W} \sum_{u_1, \dots u_k} \prod_{i=1}^{k} \varphi(u_i) \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k \\ u_i | d_i, e_i}} \frac{\lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k}}{d_1 \cdots d_k e_1 \cdots e_k} \sum_{\substack{s_{i,j} | d_i, e_j, \\ i \neq j}} \mu(s_{i,j})$$

and switching sums we get

$$\frac{N}{W} \sum_{u_1, \dots u_k} \prod_{i=1}^{k} \varphi(u_i) \sum_{s_{1,2}, \dots, s_{k,k-1}} \prod_{\substack{1 \leq i,j \leq k \\ i \neq j}} \mu(s_{i,j}) \sum_{\substack{d_1, \dots, d_k \\ e_1, \dots, e_k \\ u_i | d_i, e_i \\ s_{i,j} | d_i, e_j, i \neq j}} \frac{\lambda_{d_1, \dots, d_k} \lambda_{e_1, \dots, e_k}}{d_1 \cdots d_k e_1 \cdots e_k}.$$

We can restrict the sum over $s_{i,j}$ to be coprime to u_i and u_j, since the other terms make no contribution as $\lambda_{d_1, \dots, d_k} = 0$ unless $(d_i, d_j) = 1$. Similarly we can restrict the sum so that $s_{i,j}$ is coprime to $s_{i,a}$ and $s_{b,j}$ for $a \neq j$ and $b \neq i$. We denote the sum over $s_{1,2}, \dots s_{k,k-1}$ with these restrictions $\sum^*$.

We now introduce a change of variables. Let

$$y_{r_1, \dots, r_k} = \Big(\prod_{i=1}^{k} \mu(r_i) \varphi(r_i) \Big) \sum_{\substack{d_1, \dots, d_k \\ r_i | d_i}} \frac{\lambda_{d_1, \dots, d_k}}{d_1 \cdots d_k}.$$

(This change is invertible.) For $d_1, \dots, d_k$ with $d_1 \cdots d_k$ squarefree, we have that

$$\begin{aligned} \sum_{\substack{r_1, \dots, r_k \\ d_i | r_i}} \frac{y_{r_1, \dots, r_k}}{\prod_{i=1}^{k} \varphi(r_i)} &= \sum_{\substack{r_1, \dots, r_k \\ d_i | r_i}} \Big(\prod_{i=1}^{k} \mu(r_i) \Big) \sum_{\substack{e_1, \dots, e_k \\ r_i | e_i}} \frac{\lambda_{e_1, \dots, e_k}}{e_1 \cdots e_k} \\ &= \sum_{e_1, \dots, e_k} \frac{\lambda_{e_1, \dots, e_k}}{e_1 \cdots e_k} \sum_{\substack{r_1, \dots, r_k \\ d_i | r_i | e_i}} \prod_{i=1}^{k} \mu(r_i) = \frac{\lambda_{d_1, \dots, d_k}}{\prod_{i=1}^{k} \mu(d_i) d_i}. \end{aligned}$$

Where the last equality follows from the observation that the inner sum vanishes unless $d_i = e_i$ for all i. Thus any choice of $y_{r_1, \dots, r_k}$ supported on $r_1, \dots, r_k$ with the product $r = r_1 \cdots r_k$ squarefree and satisfying $r < R$ and $(r, W) = 1$ will give a suitable choice of $\lambda_{d_1, \dots, d_k}$. Let $y_{\max} = \sup_{r_1, \dots, r_k} |y_{r_1, \dots, r_k}|$. Using the property that

$$\frac{d}{\varphi(d)} = \sum_{e | d} \frac{1}{\varphi(e)},$$

we find that by taking $r' = \prod_{i=1}^{k} r_i / d_i$ we have

$$\lambda_{\max} \leq \sup_{\substack{d_1,\dots,d_k \\ d_1\dots d_k \text{ squarefree}}}$$

□

[Note: This is as far as we got in the course]

7. Final things

Instructions: Do all exercises by Dec 18. You may work together and ask me questions, but your answers must be your own. Explain all work clearly.

(1) Assume the prime number theorem:

$$\lim_{x\to\infty} \frac{\pi(x)\log x}{x} = 1.$$

Taking logarithms, show that this implies that

$$\lim_{x\to\infty} \frac{\pi(x)\log \pi(x)}{x} = 1.$$

Then if p_n is the n-th prime, so that $\pi(p_n) = n$, show that

$$\lim_{n\to\infty} \frac{p_n}{n\log n} = 1.$$

This says that the p_n is roughly of the size $n\log n$.

(2) Let us write $f(x) = o(g(x))$ if $|f(x)/g(x)| \to 0$ as $x \to \infty$ and $g(x) \geq 0$. We write $p_n = n\log n + o(1)$. With this notation, use the previous exercise to show that

$$\sum_{p_n \leq x} \frac{p_n - p_{n-1}}{\log n} \leq (1 + o(1))\frac{x}{\log x}$$

which says that the average gap size of $p_n - p_{n-1}$ is $\log n$.

(3) Recall that the infimum of a subset of $\mathbf{R}$ is the greatest lower bound of that subset. And the limit inferior is

$$\liminf_n x_n = \lim_{n\to\infty} \left(\inf_{n\geq m} x_m \right).$$

Use the previous result to show that

$$\liminf_n \frac{p_n - p_{n-1}}{\log p_n} \leq 1,$$

which says that the gaps between consecutive primes p_n, p_{n-1} is less than $\log p_n$ infinitely often.

(4) Similarly, define the supremum of a subset of $\mathbf{R}$ is the least upper bound of that subset. And the limit superior is

$$\limsup_n x_n = \lim_{n\to\infty} \left(\sup_{n\geq m} x_m \right).$$

Let's show that

$$\limsup_n \frac{p_n - p_{n-1}}{\log p_n}$$

is unbounded, which means that the gaps between primes grow arbitrarily large. Prove this by considering the sequence

$$n! + 2, n! + 3, \ldots, n! + n$$

for any fixed $n > 1$. Use this to show that if p_k is the largest prime less than $n! + 2$, then the next prime p_{k+1} must be a greater than $n! + n$. So there is a subsequence q_n of primes

with prime gap n. [2]

(5) The Chinese Remainder Theorem states that if $n_1, \ldots, n_k$ are pairwise relatively prime, and $a_1, \ldots, a_k$ are integers, then there exists an integer x such that

$$x \equiv a_i \pmod{n}_i$$

for $i = 1, \ldots, k$. Use Bezout's identity to prove the case $k = 2$. Use this to show that for any admissible set $\mathscr{H}$ and $k = 2$, we can choose v_0 such that $v_0 + h_1$ and $v_0 + h_2$ are relatively prime to W, where W is defined above.

(6) Explain in your own words
   (a) how Propositions 6.4 and 6.5 prove the main theorem,
   (b) what is the role of the GPY sieve in the proof, and
   (c) how the main theorem

$$\liminf_n (p_n - p_{n-1}) \leq 600,$$

   is related to the Twin Prime conjecture.

[2] In 2016 Maynard, Tao et al. showed that there exists infinite many primes for whose gaps are greater than $c \log n \log\log n \log\log\log\log n / \log\log\log n$, for some $c > 0$. This is the problem of *large* prime gaps, which Erdös offered $10,000 for for improving on his result. Erdös liked to give cash bounties. In this spirit, Tao also is offering $10,000 for improving on this constant c. The Twin Prime Conjecture is about *small* prime gaps.

8. Some answers

3. Recall that using the Euclidean algorithm we may compute $\gcd(a, b)$ by a series of divisions, say

$$\begin{aligned} a &= bq_1 + r_1, \\ b &= r_1q_2 + r_2, \\ r_1 &= r_2q_3 + r_3, \\ &\vdots \\ r_{n-2} &= r_{n-1}q_n + r_n, \\ r_{n-1} &= r_nq_{n+1}. \end{aligned}$$

Note that $\gcd(a, b) = \gcd(bq_1 + r_1, b) = \gcd(b, r_1)$, and more generally we have

$$\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{n-1}, r_n) = r_n.$$

Now we proceed by induction. We will show that for any pair of integers a, b with Euclidean algorithm taking n steps to compute $\gcd(a, b)$, we can express

$$\gcd(a, b) = ax + by$$

for some integers x and y.

If $n = 1$, then $\gcd(a, b) = r_1$ and we can write

$$r_1 = a + (-q_1)b$$

and we are done. This proves the base case.

Now our induction assumption will be that for any pair of integers a', b' with Euclidean algorithm taking $n - 1$ steps to compute $\gcd(a'b')$, we can express the gcd as a linear combination in those two integers, we can write $\gcd(a', b') = a'x' + b'y'$ for some $x', y' \in \mathbb{Z}$.

So let now a, b be such that $\gcd(a, b)$ takes n steps to compute. Then $\gcd(b, r_1)$ takes $n-1$ steps to compute. Applying the induction assumption to $a' = r_1$ and $b' = b$, it follows that

$$\gcd(b, r_1) = r_1x' + by'$$

for some $x', y' \in \mathbb{Z}$. The lefthand side is equal to $\gcd(a, b)$, and rewriting the righthand side we have

$$\begin{aligned} \gcd(a, b) &= (a - bq_1)x' + by' \\ &= ax' + b(y' - qx') \\ &= ax + by \end{aligned}$$

as desired, with $x = x'$ and $y = y' - qx'$.

6. Assume to the contrary that the series converges, say to $S > 0$. If we denote the partial sums as

$$S_n = \sum_{i=1}^{n} \frac{1}{p_i},$$

then

$$\lim_{n\to\infty} S_n = S.$$

This is the same as saying that the difference

$$\lim_{n\to\infty} (S - S_n) = 0.$$

But this means that

$$\lim_{n\to\infty}\left(\sum_{i=1}^{\infty}\frac{1}{p_i}-\sum_{i=1}^{n}\frac{1}{p_i}\right)=\lim_{n\to\infty}\sum_{i=n+1}^{\infty}\frac{1}{p_i}=0.$$

If replace n with k (for notational purposes), it follows then that there exists an integer k large enough such that

$$\sum_{i=k+1}^{\infty}\frac{1}{p_i}<\frac{1}{2}. \tag{6}$$

Now let $Q=p_1\dots p_k$. We claim that that for any $r\geq 1$, we have

$$\sum_{n=1}^{r}\frac{1}{1+nQ}\leq\sum_{j=1}^{\infty}\left(\sum_{i=k+1}^{\infty}\frac{1}{p_i}\right)^j.$$

Notice that for any $n\geq 1$, the denominator $1+nQ$ is not divisible by any $p_1,\dots,p_k$, hence $1+nQ$ can only be divisible only by primes p_i where $i\geq k+1$. In particular, if $1+nQ$ has j prime factors in its unique prime factorization, it will appear in the expansion of the series

$$\left(\sum_{i=k+1}^{\infty}\frac{1}{p_i}\right)^j.$$

On the other hand, since the righthand-side runs over all j, it follows that for any fixed r, the sum

$$\sum_{n=1}^{r}\frac{1}{1+nQ}$$

on the left is contained in the sum on the right. The rest of the terms on the righthand side being positive, we see that the inequality holds for any r.

Applying the inequality (6) to each summand, we see immediately that

$$\sum_{j=1}^{\infty}\left(\sum_{i=k+1}^{\infty}\frac{1}{p_i}\right)^j<\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^j=\frac{1}{1-\frac{1}{2}}=2.$$

and second last equality follows from the geometric series. Then since

$$\sum_{n=1}^{r}\frac{1}{1+nQ}<2$$

uniformly for all r, it follows that the limit also converges as r tends to infinity, thus

$$\sum_{n=1}^{\infty}\frac{1}{1+nQ}<2.$$

But this is impossible because we know that the series diverges, for example

$$\sum_{n=1}^{\infty}\frac{1}{1+nQ}>\sum_{n=1}^{\infty}\frac{1}{n+nQ}=\frac{1}{1+Q}\sum_{n=1}^{\infty}\frac{1}{n},$$

which shows that the series is bounded below by the harmonic series, which diverges by the integral test $\int_1^{\infty}x^{-1}dx=\infty$. Thus we have arrived at a contradiction, so the original series does not converge.

22. First, applying exponent and log rules gives immediately

$$\prod_{p\le x}\left(1-\frac{1}{p}\right) = \exp\left(\sum_{p\le x}\log\left(1-\frac{1}{p}\right)\right)$$
$$= \exp\left(\sum_{p\le x}\left(\log(p-1)-\log p\right)\right)$$

Now in class we defined the quantity

$$C(n) = \sum_{k=1}^{n}\frac{1}{k} - \int_1^n \frac{1}{t}\,dt.$$

We showed that

$$\lim_{n\to\infty} C(n) < \infty$$

since $C(n)$ is bounded by the area between the two curves $1/x$ and $1/(x-1)$,

$$C(n) < \int_2^n \frac{1}{t-1}\,dt + 1 - \int_1^n \frac{1}{t}\,dt = \log\frac{n-1}{n} + 1$$

(the $+1$ arises because of $1/(t-1)$ is undefined at $t=1$, so has to be defined separately) so

$$\lim_{n\to\infty} C(n) < 1.$$

We will apply this in the form

$$\log p = \int_1^p \frac{1}{t}\,dt = \sum_{k=1}^{p}\frac{1}{k} + C(p).$$

Hence we may write

$$\log(p-1)-\log p = \left(\sum_{k=1}^{p-1}\frac{1}{k} + C(p-1)\right) - \left(\sum_{k=1}^{p}\frac{1}{k} + C(p)\right)$$
$$= -\frac{1}{p} - (C(p)-C(p-1))$$
$$= -\frac{1}{p} - C'(p).$$

For convenience, we have written $C'(p) = (C(p)-C(p-1)$. And so we have

$$\exp\left(\sum_{p\le x}\left(\log(p-1)-\log p\right)\right) = \exp\left(\sum_{p\le x}\left(-\frac{1}{p}-C'(p)\right)\right)$$
$$= \exp\left(-\sum_{p\le x}\frac{1}{p} - \sum_{p\le x}C'(p)\right)$$

Observe that

$$B(x) := \sum_{p\le x} C'(p) < C([x])$$

and since the terms $C'(p)$ are positive, it follows that

$$\lim_{x\to\infty} B(x) < \lim_{x\to\infty} C([x]) < 1,$$

so the left hand side converges to a constant, call it B. Note that

$$B - B(x) = O\left(\frac{1}{x}\right)$$

So we may write

$$\exp\left(-\sum_{p\le x}\frac{1}{p} - \sum_{p\le x} C'(p)\right) = \exp\left(-\sum_{p\le x}\frac{1}{p} - B + O\left(\frac{1}{x}\right)\right)$$

where we note that $B(x) = O(1)$. Now to the first sum we apply Exercise 21,

$$\sum_{p\le x}\frac{1}{p} = \log\log x + A + O\left(\frac{1}{\log x}\right).$$

to get

$$\exp\left(-\log\log x - A + O\left(\frac{1}{\log x}\right) - B + O\left(\frac{1}{x}\right)\right).$$

Using Taylor series expansion we see that

$$e^{O\left(\frac{1}{\log x}\right)} = 1 + O\left(\frac{1}{\log x}\right).$$

so we have

$$\exp\left(-\log\log x - A - B + O\left(\frac{1}{\log x}\right)\right) = \frac{e^{-\gamma}}{\log x}\left(1 + O\left(\frac{1}{\log x}\right)\right)$$

since $1/x$ decays faster than $1/\log x$. And we may take the constant $\gamma = A - B$.

23. To switch the sum, you should mention that $d|(n, P(z)$ if and only if $d|n$ and $d|P(z)$. (The forward implication is true because of the definition of gcd. What about the backward implication?)

24.(1) We are looking for two residue classes $[0]$ and $[-2]$ mod p, so the "density" should be $2/p$, so $g(p) = 2$. It follows immediately from Ex 20 that

$$\sum_{p\le x}\frac{2}{p}\log p = 2\log x + O(1).$$

We take $X = x$ because that matches the main term in the discussion after Ex 23. Theorem 4.1 gives

$$S(\mathscr{A}, \mathscr{P}, z) = XV(z) + O\left(\left(X + \frac{x}{\log z}\right)(\log z)^{\kappa+1})\exp\left(-\frac{\log x}{\log z}\right)\right)$$

The main term is clear. For the error term, since $X = x$,

$$X + \frac{x}{\log z} = x + \frac{x}{\log z}$$

has leading term x so we simply write the error as

$$O\left(x(\log z)^{\kappa+1})\exp\left(-\frac{\log x}{\log z}\right)\right).$$

(2) Like Ex 22,

$$V(z) = \prod_{p \le z} \left(1 - \frac{2}{p}\right) = \exp\left(\sum_{p \le z} \log\left(1 - \frac{2}{p}\right)\right)$$

We write the righthand sum as

$$\sum_{p \le z} \big(\log(p-2) - \log p\big).$$

We want to argue like before, but let's make things a bit easier: let's write

$$\log n = \sum_{k=1}^{n} \frac{1}{k} + \gamma + O\left(\frac{1}{n}\right)$$

where we define γ to be the Euler-Mascheroni constant

$$\gamma = \sum_{k=1}^{\infty} \frac{1}{k} - \int_1^{\infty} \frac{1}{x} dx = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \log n\right).$$

(The error at n is bounded above by $1/n$.) So using this we have

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

Then summing over $p \le x$,

$$-\sum_{p \le z} \frac{2p-1}{p(p-1)} + O\left(\sum_{p \le z} \frac{1}{p-2}\right) = -\sum_{p \le z} \frac{2p-1}{p(p-1)} + O\left(\frac{1}{z}\right)$$

Note that

$$\frac{2p-1}{p(p-1)} \ge \frac{2}{p} = \frac{2p-2}{p(p-1)}$$

so

$$-\sum_{p \le x} \frac{2p-1}{p(p-1)} \le -\sum_{p \le x} \frac{2}{p}.$$

So this gives us

$$V(z) \le \exp\left(-\sum_{p \le z} \frac{2}{p} + O\left(\frac{1}{z}\right)\right)$$

which isn't quite what was stated, but we can still apply Ex 21 to write the exponential as

$$\exp\left(-2\log\log z - 2A + O\left(\frac{1}{\log z}\right) + O\left(\frac{1}{z}\right)\right)$$

Splitting off the first term and absorbing the $1/z$ term we get

$$(\log z)^{-2} \exp\left(-2A + O\left(\frac{1}{\log z}\right)\right)$$

and so

$$e^{-2A}(\log z)^{-2} \exp\left(O\left(\frac{1}{\log z}\right)\right) = e^{-2A}(\log z)^{-2}(1 + O\left(\frac{1}{\log z}\right)) = O(\log z)^{-2})$$

where we have used the Taylor series for exp.

Remark 8.1. Okay, so that was a little shady with the $O(1/z)$. To get it right one uses the Taylor series expansion

$$\log(1-x) = -\sum_{n=1}^{\infty} \frac{x^n}{n}$$

for $|x| < 1$. In our case,

$$\log(1-\frac{2}{p}) = -\sum_{n=1}^{\infty} \frac{2^n}{np^n} = -\frac{2}{p} - \sum_{n>1} \frac{2^n}{np^n}$$

and the last sum is a geometric series equal to $2/(p(1-p)^2)$, which is less than $1/(p(p-1))$. These formula follow from the usual calc rules.

Great. Now this gives us

$$\begin{aligned} S(\mathscr{A}, \mathscr{P}, z) &= xV(z) + O\left(x(\log z)^3 \exp\left(-\frac{\log x}{\log z}\right)\right) \\ &= O\left(x(\log z)^{-2}) + x(\log z)^3 \exp\left(-\frac{\log x}{\log z}\right)\right) \end{aligned}$$

let's choose $\log z = \log x/A \log\log x$ and see that we get

$$\begin{aligned} &\frac{x(A\log\log x)^2)}{\log^2 x} + x(\log x/A\log\log x)^3 \exp\left(-A\log\log x\right) \\ &= A^2 \frac{x(\log\log x)^2}{\log^2 x} + \frac{x\log^{3-A} x}{A(\log\log x)^3}. \end{aligned}$$

The first term is what we want, so we need to pick A large enough so that the first term dominates. For example, we pick $A = 5$, then we have

$$\frac{x}{\log^2 x} 5^2 (\log\log x)^2 + \frac{x}{\log^2 x} \frac{1}{A(\log\log x)^3}.$$

(3) To get the result, let $\pi_2(x)$ be twin prime up to x. Then the sieve gives

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

and the righthand side is

$$\exp\left(\frac{\log x}{A\log\log x}\right) + O\left(\frac{x(\log\log x)^2)}{\log^2 x}\right)$$

we have to show the first term is smaller than the error. So write

$$\exp\left(\frac{\log x}{A\log\log x}\right) = (\log x)^{\frac{1}{A\log\log x}} \ll \log x \ll \frac{x(\log\log x)^2)}{\log^2 x}.$$

where $\ll$ means that when x grows large the right hand side is larger, up to a constant. Alternatively, take logs on both sides to compare

$$\frac{1}{A\log\log x} \log\log x \leq \log x + \log(\log\log x)^2 - \log\log^2 x.$$

for x large enough.

(25) We use Abel summation with $a(n) = 1$ if n is twin prime and 0 otherwise, and $f(n) = 1/n$. Then $A(x) = \pi_2(x)$, and

$$\sum_{a<n\leq b} a(n)f(n) = A(b)f(b) - A(a)f(a) - \int_a^b A(t)f'(t)dt.$$

with $a = 2, b = x$ gives

$$\begin{aligned}\sum_{\substack{p\leq x\\ p+2 \text{ prime}}} \frac{1}{p} &= \frac{\pi_2(x)}{x} - \int_2^x \frac{\pi_2(t)}{t^2}dt \\ &= O\left(\left(\frac{\log\log x}{\log x}\right)^2\right) + O\left(\int_2^x \frac{(\log\log t)^2}{t\log^2 t}dt\right)\end{aligned}$$

since $\pi_2(2) = 0$. Taking the limit as x tends to ∞, the first term vanishes, and we have finally

$$\sum_{\substack{p\\ p+2 \text{ prime}}} \frac{1}{p} = O\left(\int_2^\infty \frac{(\log\log t)^2}{t\log^2 t}dt\right).$$

Here we have used that if $f(x) = O(g(x))$, then $\int f(x)dx = O(\int g(x)dx)$.

One last thing: to see that the integral converges, apply $u = \log t$ to get

$$\int_2^\infty \frac{(\log\log t)^2}{t\log^2 t}dt = \int_{\log 2}^\infty \frac{(\log u)^2}{u^2}du \leq \int_{\log 2}^\infty \frac{u^\epsilon}{u^2}du = \int_{\log 2}^\infty \frac{1}{u^{2-\epsilon}}du$$

for some appropriate constant $0 < \epsilon < 1$, and the last integral can be seen to converge. (Alternatively, integrate by parts a couple of times..)