Sieve methods

Dimitris Koukoulopoulos
University of Montreal

Last update: October 24, 2015

Contents

Conventions

We assume that the reader has already taken a first course in Analytic Number Theory and thus she is familiar with basic techniques, such as partial summation and Perron's inversion formula, as well as with basic results, such as Chebyshev's and Mertens's estimates, as well as the Prime Number Theorem for arithmetic progressions. However, we do give an almost self-contained proof of the latter in Appendix B.

Throughout these notes there are various exercises which are imbedded in the text (rather than at the end of each section or chapter). The most difficult ones have a "star".

We make use of some standard and of some less standard notation. We write $\mathbf{1}_A$ to denote the characteristic function of the set A. The symbol $\mathbb{P}$ denotes the set of prime numbers and the letter p, with or without subscripts, will always denote a member of $\mathbb{P}$. We write $f = O(g)$ or, equivalently, $f \ll g$ if there is a constant M such that $|f| \leq Mg$. The constant M will be absolute unless otherwise specified, e.g. by a subscript. Also, we write $f \asymp g$ if $f \ll g$ and $g \ll f$. For $n \in \mathbb{N}$ we use $P^+(n)$ and $P^-(n)$ to denote the largest and smallest prime factor of n, respectively, with the notational conventions that $P^+(1) = 1$ and $P^-(1) = +\infty$. We write $\omega(n)$ for the number of distinct prime factors of n and $\Omega(n)$ for the total number of prime factors of n, counted with multiplicity. As usually, μ denotes the Möbius function, defined to be $(-1)^{\omega(n)}$ if n is square-free and 0 otherwise, φ denotes Euler's totient function, which counts the size of the set $\{r \,(\mathrm{mod}\, n) : (r, n) = 1\}$, Λ denotes the von Mangoldt function, which is defined to be $\log p$ if $n = p^k$, for some prime p and some $k \geq 1$, and 0 otherwise, and τ_k denotes the k-divisor function, defined by $\tau_k(n) = \sum_{d_1 \cdots d_k = n} 1$. In particular, $\tau_2(n)$ is the number of divisors of n, which we simply denote by $\tau(n)$. We write $\pi(x)$ for the number of primes up to x and $\pi(x; q, a)$ for the number of primes up to x that lie in the arithmetic progression $a \,(\mathrm{mod}\, q)$. Finally, we give below references to the page where some additional basic notation is introduced.

Symbol	Page	Equation	Page
$\mathcal{A}_d$	23	(A1)	23
$S(\mathcal{A}, z)$	23	(A2)	24
$P(z)$	23	(A3)	25
$g(d)$	23	(A4a)	25
r_d	23	(A4b)	26
$V(z)$	24	(R)	33
β_κ	37	(R′)	58
		(r)	58

Chapter 0

Prelude: multiplicative functions

We start by covering some basic background material, which we will need in order to handle some technical parts of the theory of sieve methods. In particular, we will see various techniques for evaluating asymptotically the average value of a multiplicative function.

0.1 Arithmetic functions: basic definitions

An *arithmetic function* is a function $f : \mathbb{N} \to \mathbb{C}$. Two important classes of arithmetic functions are the so-called multiplicative functions, as well as the completely multiplicative functions. An arithmetic function $f : \mathbb{N} \to \mathbb{C}$ is called *multiplicative* if $f(1) = 1$ and

$$f(mn) = f(m)f(n) \quad \text{whenever} \quad (m, n) = 1,$$

whereas f is called *completely multiplicative* if $f(1) = 1$ and the above relation holds for all m and n, without the requirement that they are co-prime. Some important examples of multiplicative functions are the functions n^s, the divisor functions

$$\tau_k(n) := \#\{(d_1, \dots, d_k) \in \mathbb{N}^k : d_1 \cdots d_k = n\}$$

and the Möbius function

$$\mu(n) := \begin{cases} (-1)^r & \text{if } n \text{ is square free and has } r \text{ prime divisors,} \\ 0 & \text{otherwise.} \end{cases}$$

Given two arithmetic functions $f, g : \mathbb{N} \to \mathbb{C}$, we define a new arithmetic function $f * g : \mathbb{N} \to \mathbb{C}$ by the formula

$$(f * g)(n) = \sum_{ab=n} f(a)g(b) = \sum_{d|n} f(d)g(n/d).$$

The function $f * g$ is called the convolution of f and g. For example, we have that $\tau = 1 * 1$ and, in general, $\tau_k = \underbrace{1 * \cdots * 1}_{k \text{ times}}$.

Note that the operation $*$ is commutative and associative. Moreover, if f and g are both multiplicative functions, then their convolution $f * g$ is also multiplicative. The unit of the convolution operation is the completely multiplicative function

$$\mathbb{1}(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

Any arithmetic function f with $f(1) \neq 0$ has an inverse with respect to $*$, that is to say, there is $g : \mathbb{N} \to \mathbb{C}$ with $f * g = \mathbb{1}$. In particular, any multiplicative function has an inverse with respect to $*$. Combining all of the above, we conclude that the set of multiplicative functions together with the operation $*$ is an abelian group.

A particularly important example of a multiplicative function is the constant function 1. Its convolution inverse is the Möbius function. Equivalently, we have the *Möbius inversion formula*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1, \end{cases} \tag{0.1.1}$$

which follows by the inclusion-exclusion principle. Alternatively, one may observe that $1 * \mu$ is multiplicative and verify directly the above formula when n is a prime power. A direct consequence of (0.1.1) is that if f is completely multiplicative, then its convolution inverse is given by μf.

Exercise 0.1.1. Prove all of the above assertions.

0.2 Averages of multiplicative functions: basic techniques

In this section we cover some basic techniques for studying the average behaviour of multiplicative functions. The most basic such technique is the 'convolution method'. Basically, this method allows us to compute average values of a multiplicative function $f : \mathbb{N} \to \mathbb{C}$ by relating it to (approximating it by) a simpler multiplicative function g, whose average value we already understand. It turns out that a good way of doing this is by finding a nice function g such that $f(p) \approx g(p)$ most of the time. Then we write $f = g * h$ and we compute h by inverting g. For example, if $f(n) = n/\varphi(n)$, then $f(p) = p/(p-1) = 1 + O(1/p)$. So a good choice would be to set $g = 1$, in which case $h = \mu * f$, by the Möbis inversion formula. This method allows us to obtain the following result:

Theorem 0.2.1. *For $x \geq 2$, we have that*

$$\sum_{n \leq x} \frac{n}{\varphi(n)} = x \prod_{p} \left(1 + \frac{1}{p(p-1)}\right) + O(\sqrt{x}).$$

Proof. Let $f(n) = n/\varphi(n)$. We write $f = 1 * h$, so that $h = \mu * f$. Then

$$h(p^\nu) = \sum_{d|p^\nu} \mu(d) f(p^\nu/d) = f(p^\nu) - f(p^{\nu-1}) = \begin{cases} \frac{1}{p-1} & \text{if } \nu = 1, \\ 0 & \text{if } \nu \geq 2. \end{cases}$$

In particular, $|h(p^\nu)| \leq 1/\sqrt{p}$ for all $p > 3$, which also holds when $p = 2$ and $\nu \geq 2$. So $|h(n)| \leq \sqrt{2/n}$ for all $n \in \mathbb{N}$ and, consequently,

$$\begin{aligned} \sum_{n\leq x} f(n) &= \sum_{n\leq x} \sum_{d|n} h(d) = \sum_{d\leq x} h(d) \sum_{n\leq x,\, d|n} 1 = \sum_{d\leq x} h(d) \left(\frac{x}{d} + O(1)\right) \\ &= x \sum_{d=1}^{\infty} \frac{h(d)}{d} + O\left(\sum_{d>x} \frac{1}{d^{3/2}} + \sum_{d\leq x} \frac{1}{\sqrt{d}}\right) = x \sum_{d=1}^{\infty} \frac{h(d)}{d} + O(\sqrt{x}), \end{aligned}$$

which completes the proof of the theorem. □

Exercise 0.2.2. Let $\sigma(n) = \sum_{d|n} d$. Use the convolution method to show that

$$\sum_{n\leq x} \frac{1}{\sigma(n)} \sim c \log x \quad (x \to \infty),$$

for some appropriate constant c.

We now turn to another important example, the divisor function τ. As we said above, $\tau = 1 * 1$. So we have that

$$\begin{aligned} \sum_{n\leq x} \tau(n) = \sum_{n\leq x} \sum_{d|n} 1 = \sum_{d\leq x} \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d\leq x} \left(\frac{x}{d} + O(1)\right) &= x \left(\log x + \gamma + O\left(\frac{1}{x}\right)\right) + O(x) \\ &= x \log x + O(x). \end{aligned}$$

Dirichlet discovered that it possible to take advantage of the fact that both factors of $\tau = 1*1$ are 'nice' functions and improve significantly upon this result.

First, note that for every A and B with $AB = x$ we have the general formula

$$\sum_{n\leq x} (f*g)(n) = \sum_{a\leq A} f(a) \sum_{b\leq x/a} g(b) + \sum_{b\leq B} g(b) \sum_{a\leq x/b} f(b) - \left(\sum_{a\leq A} f(a)\right)\left(\sum_{b\leq B} g(b)\right). \tag{0.2.1}$$

If we know something about the average behaviour of both f and g, then may pick both A and B to be increasing functions of x. This allows us to reduce the length of sums we are considering and thus improve the error term in our formula for the summatory function of $f * g$. In particular, when $f = g = 1$, then choosing $A = B = \sqrt{x}$ yields the formula

$$\begin{aligned} \sum_{n\leq x} \tau(n) &= 2 \sum_{a\leq\sqrt{x}} \left\lfloor \frac{x}{a} \right\rfloor - \left\lfloor \sqrt{x} \right\rfloor^2 = 2 \sum_{a\leq\sqrt{x}} \left(\frac{x}{a} + O(1)\right) - (\sqrt{x} + O(1))^2 \\ &= 2x \left(\log \sqrt{x} + \gamma + O\left(\frac{1}{\sqrt{x}}\right)\right) - x + O(\sqrt{x}) \\ &= x \log x + (2\gamma - 1)x + O(\sqrt{x}). \end{aligned}$$

This argument can be generalized to deduce the following theorem.

Theorem 0.2.3. *Fix $k \geq 2$. There is a polynomial P_k of degree $k-1$ such that*

$$\sum_{n \leq x} \tau_k(n) = x \cdot P_k(\log x) + O_k(x^{1-1/k}) \quad (x \geq 1).$$

Proof. Exercise. □

Exercise 0.2.4. (a) Find an asymptotic formula for the summatory function of $2^{\omega(n)}$, i.e. for $\sum_{n \leq x} 2^{\omega(n)}$.

(b)* Try doing the same thing for the summatory function of $2^{\Omega(n)}$. What happens? Can you explain why?

Exercise 0.2.5.

(a) Show that, for every $x \geq 1$,

$$\#\{n \leq x : n \text{ is square-free}\} = x \cdot \prod_p \left(1 - \frac{1}{p^2}\right) + O(\sqrt{x}).$$

(b)* Show that the error term is in fact $o(\sqrt{x})$ as $x \to \infty$.

Next, we show the following very useful result.

Theorem 0.2.6. *Let $f : \mathbb{N} \to [0, +\infty)$ be a multiplicative function such that*

$$\sum_{p \leq x} f(p) \log p \leq Ax \quad (x \geq 1), \quad \text{and} \quad \sum_{\substack{p \text{ prime} \\ \nu \geq 2}} \frac{f(p^\nu)\log(p^\nu)}{p^\nu} \leq B.$$

Then, for $x \geq 2$, we have that

$$\sum_{n \leq x} f(n) \leq (A + B + 1)\frac{x}{\log x} \sum_{n \leq x} \frac{f(n)}{n} \quad \text{and} \quad \sum_{n \leq x} \frac{f(n)}{n} \leq e^B \prod_{p \leq x} \left(1 + \frac{f(p)}{p}\right).$$

Remark 0.2.7. Taking $f(n) = \tau_k(n)$, we see that the above theorem is best possible in this generality, up to multiplicative constants (cf. Theorem 0.2.3).

Proof. Note that

$$\begin{aligned}
(\log x) \sum_{n \leq x} f(n) &= \sum_{n \leq x} f(n) \log n + \sum_{n \leq x} f(n) \log \frac{x}{n} \\
&\leq \sum_{n \leq x} f(n) \sum_{p^\nu \| n} \log(p^\nu) + \sum_{n \leq x} f(n) \cdot \frac{x}{n} \\
&= \sum_{\nu \geq 1} \sum_{\substack{mp^\nu \leq x \\ p \nmid m}} f(mp^\nu) \log(p^\nu) + x \sum_{n \leq x} \frac{f(n)}{n} \\
&\leq \sum_{\nu \geq 1} \sum_{mp^\nu \leq x} f(m) f(p^\nu) \log(p^\nu) + x \sum_{n \leq x} \frac{f(n)}{n}.
\end{aligned}$$

When $\nu = 1$, we have that

$$\sum_{mp\le x} f(m)f(p)\log p = \sum_{m\le x} f(m) \sum_{p\le x/m} f(p)\log p \le \sum_{m\le x} f(m)\cdot A\frac{x}{m} = Ax\sum_{m\le x}\frac{f(m)}{m}.$$

Finally, we bound the rest of the summands by noting that

$$\begin{aligned}\sum_{\nu\ge 2}\sum_{mp^\nu\le x} f(m)f(p^\nu)\log(p^\nu) &\le x\sum_{\nu\ge 1}\sum_{mp^\nu\le x}\frac{f(m)f(p^\nu)\log(p^\nu)}{mp^\nu}\\ &\le x\sum_{\nu\ge 1}\sum_{m\le x}\frac{f(m)f(p^\nu)\log(p^\nu)}{mp^\nu}\\ &= \left(\sum_{m\le x}\frac{f(m)}{m}\right)\left(\sum_{\nu\ge 2,\, p\,\text{prime}}\frac{f(p^\nu)\log(p^\nu)}{p^\nu}\right) \le Bx\sum_{m\le x}\frac{f(m)}{m}.\end{aligned}$$

So we conclude that

$$\sum_{n\le x} f(n) \le (A+B+1)\frac{x}{\log x}\sum_{n\le x}\frac{f(n)}{n}.$$

To see the second part of the theorem, note that

$$\sum_{n\le x}\frac{f(n)}{n} \le \prod_{p\le x}\left(1+\frac{f(p)}{p}+\frac{f(p^2)}{p^2}+\cdots\right) \le \prod_{p\le x}\left(1+\frac{f(p)}{p}\right)\exp\left\{\frac{f(p^2)}{p^2}+\frac{f(p^3)}{p^3}+\cdots\right\},$$

by the inequality $1+u+v\le(1+u)e^v$, $u\ge 1, v\ge 0$. Since

$$\sum_{p}\left(\frac{f(p^2)}{p^2}+\frac{f(p^3)}{p^3}+\cdots\right) \le \frac{1}{\log 4}\sum_{\nu\ge 2,\, p\,\text{prime}}\frac{f(p^\nu)\log(p^\nu)}{p^\nu} \le \frac{B}{\log 4} \le B,$$

the desired result follows. □

Exercise 0.2.8. Fix $r\ge 0$ and $k\in\mathbb{N}$. Show that

$$\sum_{n\le x}\tau_k(n)\left(\frac{n}{\varphi(n)}\right)^r \ll_{r,k} x(\log x)^{k-1} \quad (x\ge 1)$$

and

$$\sum_{n\le x}\tau_k(n)\left(\frac{\varphi(n)}{n}\right)^r \gg_{r,k} x(\log x)^{k-1} \quad (x\ge 1)$$

Exercise 0.2.9. Show that

$$\frac{n}{\varphi(n)} \asymp \prod_{\substack{p\mid n\\ p\le y}}\left(1+\frac{1}{p}\right) \quad (y\ge\sqrt{\log n}).$$

Use this fact to show that

$$\sum_{x-y<n\le x}\frac{n}{\varphi(n)} \ll_\epsilon y \quad (x^\epsilon\le y\le x).$$

Exercise 0.2.10. Show that the error term in Theorem 0.2.1 can be strengthened to $O(\log x)$.

Exercise 0.2.11. This exercise generalizes and strengthens Theorem 0.2.1. Let $s \neq 0$ and $x \geq 1$. Show that

$$\sum_{\substack{a \leq x \\ (a,s)=1}} \frac{a}{\varphi(a)} = x \prod_{p \nmid s} \left(1 + \frac{1}{p(p-1)}\right) \prod_{p \mid s} \left(1 - \frac{1}{p}\right) + O\left(\frac{|s|}{\varphi(s)} \cdot \log(2x)\right).$$

Conclude that

$$\sum_{\substack{a \leq x \\ (a,s)=1}} \frac{1}{\varphi(a)} = \prod_{p \nmid s} \left(1 + \frac{1}{p(p-1)}\right) \prod_{p \mid s} \left(1 - \frac{1}{p}\right) \cdot \left\{ \log x - \gamma + \sum_{p \mid s} \frac{\log p}{p-1} - \sum_{p \nmid s} \frac{\log p}{p^2 - p + 1} \right\} + O\left(\frac{|s|}{\varphi(s)} \cdot \frac{\log(2x)}{x}\right).$$

0.3 Rankin's method

In this section we discuss briefly the so-called *Rankin's method*, which will play a prominent role at several places throughout these notes. The main idea of this method is that if f is a multiplicative function which takes non-negative values, then

$$\sum_{n \leq x} f(n) \leq \sum_{n \leq x} f(n) \left(\frac{x}{n}\right)^{\sigma} \leq x^{\sigma} \sum_{n=1}^{\infty} \frac{f(n)}{n^{\sigma}} = x^{\sigma} \prod_{p} \left(1 + \frac{f(p)}{p^{\sigma}} + \frac{f(p^2)}{p^{2\sigma}} + \cdots\right).$$

Using this simple trick, we obtain the following general result.

Theorem 0.3.1. *Let $f : \mathbb{N} \to \mathbb{C}$ be a multiplicative function, which we write as $f = 1 * g$. Consider $\sigma \in [0, 1)$ for which $\sum_{n \geq 1} g(n)/n^{\sigma}$ converges absolutely. Then we have that*

$$\sum_{n \leq x} f(n) = c_f x + O\left(x^{\sigma} \sum_{n=1}^{\infty} \frac{|g(n)|}{n^{\sigma}}\right),$$

where

$$c_f = \sum_{n=1}^{\infty} \frac{g(n)}{n} = \prod_{p} \left(1 - \frac{1}{p}\right) \left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \cdots\right).$$

Exercise 0.3.2. Use Theorem 0.3.1 to show that

$$\sum_{n \leq x} \frac{\mu^2(n) n}{\varphi(n)} = x + O(\sqrt{x} \log x) \quad (x \geq 2).$$

The following result is another application of Rankin's trick. In fact, it is precisely this form of the trick that we will see appearing again and again while discussing sieve methods.

Theorem 0.3.3. *Let $x \geq y \geq 3$ and define $u \geq 1$ by the relation $x = y^u$. If $y \geq (\log x)^3$, then we have that*

$$\sum_{\substack{P^+(n)\leq y\\ n>x}} \frac{1}{n} \ll (\log y) \cdot \frac{e^{O(u)}}{(u\log u)^u}.$$

Remark 0.3.4. Integers n all of whose prime divisors are $\leq y$ are called y-smooth. We will study their distribution more carefully in Chapter 6.

Proof. Clearly, we may assume that u is large enough. For every $\epsilon \in (0, 1/3]$, we have that

$$\begin{aligned}\sum_{\substack{P^+(n)\leq y\\ n>x}} \frac{1}{n} &\leq \frac{1}{x^\epsilon} \sum_{P^+(n)\leq y} \frac{1}{n^{1-\epsilon}} = \frac{1}{x^\epsilon} \prod_{p\leq y} \left(1 - \frac{1}{p^{1-\epsilon}}\right)^{-1}\\ &\ll \frac{\log y}{x^\epsilon} \prod_{p\leq y} \left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{p^{1-\epsilon}}\right)^{-1}\\ &\ll \frac{\log y}{x^\epsilon} \exp\left\{\sum_{p\leq y} \frac{p^\epsilon - 1}{p}\right\},\end{aligned}$$

since

$$\log \prod_{p\leq y} \left(1 - \frac{1}{p^\sigma}\right)^{-1} = \sum_{p\leq y} \log\left(1 - \frac{1}{p^\sigma}\right)^{-1} = \sum_{p\leq y}\sum_{m=1}^{\infty} \frac{1}{mp^{m\sigma}} = \sum_{p\leq y} \frac{1}{p^\sigma} + O(1)$$

for all $\sigma \geq 2/3$. Moreover, note that if $p \leq e^{1/\epsilon}$, then $p^\epsilon = 1 + O(\epsilon \log p)$ and consequently

$$\sum_{p\leq e^{1/\epsilon}} \frac{p^\epsilon - 1}{p} \ll \sum_{p\leq e^{1/\epsilon}} \frac{\epsilon \log p}{p} \ll 1.$$

So, if we impose the condition $\epsilon \geq 1/\log y$, then we arrive to the estimate

$$\sum_{\substack{P^+(n)\leq z\\ n>x}} \frac{1}{n} \ll \frac{\log y}{x^\epsilon} \exp\left\{\sum_{e^{1/\epsilon}<p\leq y} \frac{p^\epsilon - 1}{p}\right\} \leq \frac{\log y}{x^\epsilon} \exp\left\{\sum_{e^{1/\epsilon}<p\leq y} \frac{1}{p^{1-\epsilon}}\right\}.$$

By the Prime Number Theorem, we have that

$$\sum_{e^{1/\epsilon}<p\leq y} \frac{1}{p^{1-\epsilon}} = \int_{e^{1/\epsilon}}^{y} \frac{dt}{t^{1-\epsilon}\log t} + O\left(\frac{y^\epsilon}{(\log y)^2} + 1\right).$$

Moreover, the above integral equals

$$\int_{1/\epsilon}^{\log y} \frac{e^{\epsilon u}}{u} du = \int_{1}^{\epsilon \log y} \frac{e^u}{u} du = \frac{y^\epsilon}{\epsilon \log y} - e + \int_1^{\epsilon \log y} \frac{e^u}{u^2} du = \frac{y^\epsilon}{\epsilon \log y} + O\left(\frac{y^\epsilon}{(\epsilon \log y)^2}\right),$$

by two changes of variable and integration by parts. So we deduce that

$$\sum_{\substack{P^+(n)\le z\\ n>x}} \frac{1}{n} \ll \frac{\log y}{x^\epsilon} \exp\left\{\frac{y^\epsilon}{\epsilon \log y} + O\left(\frac{y^\epsilon}{(\epsilon \log y)^2}\right)\right\}.$$

Writing $x = y^u$ and setting $w = \epsilon \log y \in [1, (\log y)/3]$, we conclude that

$$\frac{1}{\log y} \sum_{\substack{P^+(n)\le y\\ n>x}} \frac{1}{n} \ll \exp\left\{\frac{e^w}{w} - uw + O\left(\frac{e^w}{w^2}\right)\right\}. \tag{0.3.1}$$

In order to optimize the above inequality, we choose $w \ge 1$ with $e^{w-1}/w = u$ (for every $u \ge 1$, there is such a w). We need that $w \le y^{1/3}$. This does hold if and only if

$$u = \frac{e^{w-1}}{w} \le \frac{3y^{1/3}}{\log y} \quad \Leftrightarrow \quad \log x \le 3y^{1/3}.$$

This last inequality is guaranteed by our assumption that $y \ge (\log x)^3$. Finally, note that $w - \log w - 1 = \log u$. In particular, $w \asymp \log u$ and thus $w = \log u + \log w = \log u + \log\log u + O(1)$. Therefore

$$\frac{e^w}{w} - uw + O\left(\frac{e^w}{w^2}\right) = u - u(\log u + \log\log u) + O(u),$$

which together with (0.3.1) completes the proof of the theorem. □

Exercise 0.3.5. Let $x \ge y \ge 2$ with $u \ge 1$. Show that, for every fixed $c \ge 1$, we have that

$$\sum_{\substack{P^+(n)\le y\\ n>x}} \frac{1}{n} \ll_c \frac{\log y}{e^{cu}}.$$

0.4 Averages of multiplicative functions: integral-delay equations

In this section we use a different method to show asymptotic formulas for the partial sums of multiplicative functions under some mild hypotheses.

Theorem 0.4.1. *Let $g : \mathbb{N} \to [0, +\infty)$ be a multiplicative function such that $g(p) \le C_5/p$ for all primes p, and*

$$-L + \kappa \log z \le \sum_{p\le z} g(p)\log p \le C_6 + \kappa \log z \quad (z \ge e^L), \tag{A3′}$$

for some $\kappa > 0$ and some constants $L, C_5, C_6 \ge 1$. Then

$$\sum_{n\le z} \mu^2(n) g(n) = \frac{\mathfrak{S}(z)}{\Gamma(\kappa+1)} \cdot (\log z)^\kappa \cdot \left\{1 + O_{\kappa,C_5,C_6}\left(\frac{L}{\log z}\right)\right\} \quad (z \ge 2),$$

where

$$\mathfrak{S}(z) = \prod_{p \le z} (1 + g(p)) \left(1 - \frac{1}{p}\right)^{\kappa}.$$

Moreover, if $z \ge e^L$ and we set $\mathfrak{S} = \mathfrak{S}(\infty)$, then

$$\sum_{n \le z} \mu^2(n) g(n) = \frac{\mathfrak{S}}{\Gamma(\kappa+1)} \cdot (\log z)^{\kappa} \cdot \left\{1 + O_{\kappa, C_5, C_6}\left(\frac{L}{\log z}\right)\right\}. \tag{0.4.1}$$

Proof. A natural way to try and prove this theorem would be to approximate $\mu^2(n)g(n)$ by the function $\kappa^{\omega(n)}/n$ or, even, $\tau_\kappa(n)/n$, where τ_κ is defined via the identity $\zeta(s)^\kappa = \sum_{n \ge 1} \tau_\kappa(n)/n^s$, the point being that both of these functions are equal to κ on prime numbers. This would theoretically reduce our task to estimating $\sum_{n \le x} \kappa^{\omega(n)}/n$ or $\sum_{n \le x} \tau_\kappa(n)/n$, which should be easier to analyze due to the regularity of the summands. However, this cannot be made to work because our assumptions on g are too weak to allow the convolution method to work. Instead we use a different approach, working directly with the partial sums of $\mu^2 g$.

Set $S(z) = \sum_{n \le z} \mu^2(n) g(n)$. The idea behind this theorem is the following: we have that

$$\sum_{n \le z} \mu^2(n) g(n) \log n = (\log z) S(z) - \int_1^z \frac{S(t)}{t} dt,$$

by partial summation. On the other hand, we have that

$$\begin{aligned} \sum_{n \le z} \mu^2(n) g(n) \log n \sum_{n \le z} \mu^2(n) g(n) \sum_{p \mid n} \log p &\approx \sum_{p \le z} g(p) (\log p) S(z/p) \\ &\approx \kappa \sum_{p \le z} \frac{\log p}{p} S(z/p) \approx \kappa \int_1^z S(z/t) \frac{dt}{t} = \kappa \int_1^z S(t) \frac{dt}{t}. \end{aligned}$$

So if $z = e^u$ and $\sigma(u) = S(e^u)/u^\kappa$, then the above formulas imply that

$$u^{\kappa+1} \sigma(u) \approx (\kappa+1) \int_0^u w^\kappa \sigma(w) dt. \tag{0.4.2}$$

Differentiating, we find that $\sigma'(u) \approx 0$, so $\sigma(u)$ is essentially constant. The calculation of this constant will be performed using different means in the end of the proof.

We shall now make the above argument rigorous. All implied constants might depend on κ, C_5 and C_6. Note that

$$\sum_{n \le z} \mu^2(n) g(n) \le \prod_{p \le z} (1 + g(p)) \asymp \mathfrak{S}(z) (\log z)^{\kappa} \quad (z \ge 2). \tag{0.4.3}$$

In particular, the theorem holds trivially when $z \le e^L$, so we only need to consider the case

$z > e^L$. Note that

$$\begin{aligned}\log\left\{\prod_{p>z}(1+g(p))\left(1-\frac{1}{p}\right)^{\kappa}\right\} &= \sum_{p>z}\left\{g(p)-\frac{\kappa}{p}+O\left(\frac{1}{p^2}\right)\right\}\\ &= O\left(\frac{1}{z}\right)+\int_z^{\infty}\left(\sum_{z<p\le t}\left(g(p)\log p-\frac{\kappa\log p}{p}\right)\right)\frac{dt}{t\log^2 t}\\ &= O\left(\frac{1}{z}\right)+\int_z^{\infty}O(L)\frac{dt}{t\log^2 t}=O\left(\frac{L}{\log z}\right),\end{aligned}$$

by partial summation and our assumptions on g, so that

$$\mathfrak{S}(z)=\mathfrak{S}\cdot\left(1+O\left(\frac{L}{\log z}\right)\right)\quad(z\ge e^L). \tag{0.4.4}$$

So, it suffices to show (0.4.1) for $z \ge e^L$.

For each $w \ge 1$, we have that

$$\begin{aligned}\sum_{n\le e^w}\mu^2(n)g(n)\log n=\sum_{n\le e^w}\mu^2(n)g(n)\sum_{p|n}\log p &= \sum_{p\le e^w}\log p\sum_{p|n\le e^w}\mu^2(n)g(n)\\ &= \sum_{p\le e^w}g(p)\log p\sum_{\substack{m\le e^w/p\\ p\nmid m}}\mu^2(m)g(m).\end{aligned}$$

Moreover,

$$\sum_{\substack{m\le e^w/p\\ p|m}}\mu^2(m)g(m)=g(p)\sum_{\substack{r\le e^w/p^2\\ p\nmid r}}\mu^2(r)g(r)\le g(p)\sum_{r\le e^w}\mu^2(r)g(r)\ll g(p)\mathfrak{S}(e^w)w^{\kappa},$$

by relation (0.4.3). Consequently,

$$\begin{aligned}\sum_{n\le e^w}\mu^2(n)g(n)\log n &= \sum_{p\le e^w}g(p)\log p\sum_{m\le e^w/p}\mu^2(m)g(m)+O\left(\mathfrak{S}(e^w)w^{\kappa}\sum_{p\le e^w}g(p)^2\log p\right)\\ &= \sum_{m\le e^w}\mu^2(m)g(m)\sum_{p\le e^w/m}g(p)\log p+O(\mathfrak{S}(e^w)w^{\kappa}),\end{aligned}$$

where we used our assumption that $g(p) \le C_5/p$. Inserting (A3′) into the above formula, we deduce that

$$\begin{aligned}\sum_{m\le e^w}\mu^2(m)g(m)\log m &= \sum_{m\le e^w}\mu^2(m)g(m)\left(\kappa\log(e^w/m)+O_{C_6}(L)\right)+O\left(\mathfrak{S}(e^w)Lw^{\kappa}\right),\\ &= \kappa\sum_{m\le e^w}\mu^2(m)g(m)\log(e^w/m)+O(\mathfrak{S}(e^w)Lw^{\kappa})\\ &= \kappa\int_1^{e^w}\frac{S(y)}{y}dt+O(\mathfrak{S}(e^w)Lw^{\kappa}),\end{aligned}$$

on using the formula $\log(e^w/m) = \int_m^{e^w} dy/y$ and inverting the order of summation and integration. By partial summation, the left hand side of the above formula is $wS(e^w) - \int_1^{e^w} S(y)dy/y$. So we find that

$$\begin{aligned} wS(e^w) &= (\kappa+1)\int_1^{e^w} \frac{S(y)}{y}dt + O(\mathfrak{S}(e^w)Lw^\kappa) \\ &= (\kappa+1)\int_0^{w} S(e^t)dt + O(\mathfrak{S}(e^w)Lw^\kappa) \\ &= (\kappa+1)\int_1^{w} S(e^t)dt + O(\mathfrak{S}(e^w)Lw^\kappa), \end{aligned}$$

where we used (0.4.3). Writing

$$S(e^t) = t^\kappa \sigma(t),$$

the above formula becomes

$$w^{\kappa+1}\sigma(w) - (\kappa+1)\int_1^w t^\kappa\sigma(t)dt \ll \mathfrak{S}(e^w)Lw^\kappa \quad (w \geq 1).$$

Set

$$E(w) := \sigma(w) - \frac{\kappa+1}{w^{\kappa+1}}\int_1^w t^\kappa\sigma(t)dt, \tag{0.4.5}$$

so that $E(w) \ll \mathfrak{S}(e^w)L/w$ for all $w \geq 1$. We multiply $E(w)$ by a weight function $k(w)$ and integrate over $w \in [1,u]$. In anticipation of the choice of k, and in order to simplify some calculations, we let $k(w) = f'(w)w^{\kappa+1}/(\kappa+1)$, where f' is the derivative of a twice differentiable increasing function $f : [1,+\infty) \to \mathbb{R}$ to be chosen later. (Assuming that k is of this form is clearly not a serious restriction.) We have that

$$\begin{aligned} \int_1^u \frac{E(w)f'(w)w^{\kappa+1}}{\kappa+1}dw &= \int_1^u \frac{\sigma(w)f'(w)w^{\kappa+1}}{\kappa+1} - \int_1^u f'(w)\int_1^w t^\kappa\sigma(t)dtdw \\ &= \int_1^u \frac{\sigma(w)f'(w)w^{\kappa+1}}{\kappa+1} - \int_1^u t^\kappa\sigma(t)\int_t^u f'(w)dwdt \\ &= \int_1^u \frac{\sigma(w)f'(w)w^{\kappa+1}}{\kappa+1} - \int_1^u t^\kappa\sigma(t)(f(u)-f(t))dt \\ &= \int_1^u \sigma(w)\left(\frac{f(w)w^{\kappa+1}}{\kappa+1}\right)'dw - f(u)\int_1^u t^\kappa\sigma(t)dt. \end{aligned}$$

Using (0.4.5) to rewrite the integral $\int_1^u t^\kappa\sigma(t)dt$ and rearranging the terms, we find that

$$\frac{f(u)u^{\kappa+1}}{\kappa+1}(\sigma(u)-E(u)) = \int_1^u \sigma(w)\left(\frac{f(w)w^{\kappa+1}}{\kappa+1}\right)'dw - \int_1^u \frac{E(w)f'(w)w^{\kappa+1}}{\kappa+1}dw. \tag{0.4.6}$$

We choose $f(w) = -(\kappa+1)/w^{\kappa+1}$ so that $k(w) = f'(w)w^{\kappa+1}/(\kappa+1) = 1/w$ and (0.4.6) becomes

$$\sigma(u) = E(u) + (\kappa+1)\int_1^u E(w)\frac{dw}{w}. \tag{0.4.7}$$

Since $E(w) \ll \mathfrak{S}(e^w)L/w$ for $w \geq 1$ and $\mathfrak{S}(e^w) \ll \mathfrak{S}$ for $w \geq L$, by relation (0.4.4), the integral in (0.4.7) converges absolutely as $u \to \infty$. Moreover,

$$\int_u^\infty E(w)\frac{dw}{w} \ll \mathfrak{S}L\int_u^\infty \frac{dw}{w^2} \ll \frac{\mathfrak{S}L}{u} \quad (u \geq L),$$

and, consequently,

$$\sigma(u) = (\kappa+1)\int_0^\infty E(w)\frac{dw}{w} + O\left(\frac{\mathfrak{S}L}{u}\right) =: I + O\left(\frac{\mathfrak{S}L}{u}\right) \quad (u \geq L). \tag{0.4.8}$$

Finally, we claim that

$$I = \frac{\mathfrak{S}}{\Gamma(\kappa+1)}, \tag{0.4.9}$$

an identity which completes the proof of (0.4.1) and, hence, of the theorem. In order to show (0.4.9) note that, as $s \to 0^+$, we have that

$$\sum_{n=1}^\infty \frac{\mu^2(n)g(n)}{n^s} = \prod_p \left(1 + \frac{g(p)}{p^s}\right) = \zeta(s+1)^\kappa \prod_p \left(1 + \frac{g(p)}{p^s}\right)\left(1 - \frac{1}{p^{s+1}}\right)^\kappa \sim \frac{\mathfrak{S}}{s^\kappa}, \tag{0.4.10}$$

by (A3′) and the fact that $\zeta(s) \sim 1/(s-1)$ as $s \to 1$. On the other hand, integration by parts implies that

$$\begin{aligned}\sum_{n=1}^\infty \frac{\mu^2(n)g(n)}{n^s} &= \int_{1^-}^\infty \frac{dt}{t^s}dS(t) = s\int_1^\infty \frac{S(t)}{t^{s+1}}dt = \int_0^\infty \frac{S(e^{u/s})}{e^u}du \\ &= \int_0^\infty \frac{I\cdot(u/s)^\kappa + O_g(1+(u/s)^{\kappa-1}))}{e^u}du = \frac{I\cdot\Gamma(\kappa+1)+O_g(s)}{s^\kappa},\end{aligned} \tag{0.4.11}$$

for every $s > 0$. Comparing (0.4.10) with (0.4.11) shows relation (0.4.9), thus completing the proof of the theorem. □

Chapter 1

Introduction to sieve methods

1.1 The sieve of Eratosthenes-Legendre

Sieve methods begin with Eratosthenes of Cyrene, who observed that it is possible to determine all primes up to a certain point. His starting point was the following simple theorem:

Theorem 1.1.1. *If $n > 1$ is composite, then there is a prime number $p \leq \sqrt{n}$ that divides n.*

Then, Eratosthenes's algorithm for founding all primes up to x has the following steps:

(1) List all integers in $[1, x]$.

(2) Delete 1 from the list.

(3) Find the smallest $n \in (1, \sqrt{x}]$ which has not been deleted yet and put a circle around it. If such an n does not exist, terminate the algorithm.

(4) Delete all multiples of n.

(5) Go to step (3).

The termination of this algorithm is guaranteed by Theorem 1.1.1. After its termination, the integers which have not yet been deleted together with the ones that are circled will be exactly the prime numbers in $[1, x]$. Eratosthenes' algorithm is called a sieve because the integers that are not deleted by it ('do not pass through it') are exactly the primes up to a given point.

Eratosthenes' sieve provides a way to count primes in various settings. We give below its simplest application to the basic question of how big is $\pi(x) = \#\{p \leq x\}$. We already know from Prime Number Theorem that

$$\pi(x) \sim \frac{x}{\log x} \quad (x \to \infty). \tag{1.1.1}$$

For comparison, let us see what Eratosthenes' sieve gives as an answer: consider all primes $p \leq \sqrt{x}$. If an integer $n \leq x$ is not divisible by any of these primes, then either $n = 1$ or n

is a prime number lying in $(\sqrt{x}, x]$. So

$$\begin{aligned}\pi(x) &= \#\{n \le x : p|n \Rightarrow p > \sqrt{x}\} - 1 + \pi(\sqrt{x}) \\ &= \#\{n \le x : p|n \Rightarrow p > \sqrt{x}\} + O(\sqrt{x}).\end{aligned} \tag{1.1.2}$$

Now we will try to understand what is the cardinality of the set appearing on the right hand side of (1.1.2). Let $\{p_1, p_2, \dots, p_r\}$ be an indexing of the set $\mathbb{P} \cap [1, \sqrt{x}]$. Then, by the inclusion-exclusion principle, we have that

$$\begin{aligned}\#\{n \le x : p|n \Rightarrow p > \sqrt{x}\} &= \#\left(\bigcap_{i=1}^{r} \{n \le x : p_i \nmid n\}\right) = \#\{n \le x\} - \#\left(\bigcup_{i=1}^{r} \{n \le x : p_i|n\}\right) \\ &= \#\{n \le x\} - \sum_{i=1}^{r} \#\{n \le x : p_i|n\} + \sum_{1 \le i < j \le r} \#\{n \le x : p_i p_j|n\} \\ &\quad - \sum_{1 \le i < j < k \le r} \#\{n \le x : p_i p_j p_k|n\} \pm \cdots \\ &= \lfloor x \rfloor - \sum_{i=1}^{r} \left\lfloor \frac{x}{p_i} \right\rfloor + \sum_{1 \le i < j \le r} \left\lfloor \frac{x}{p_i p_j} \right\rfloor - \sum_{1 \le i < j \le r} \left\lfloor \frac{x}{p_i p_j p_k} \right\rfloor \pm \cdots\end{aligned} \tag{1.1.3}$$

Since $\lfloor y \rfloor = y + O(1) \approx y$, it is reasonable to expect that

$$\begin{aligned}\#\{n \le x : p|n \Rightarrow p > \sqrt{x}\} &\approx x - \sum_{i=1}^{r} \frac{x}{p_i} + \sum_{1 \le i < j \le r} \frac{x}{p_i p_j} - \sum_{1 \le i < j < k \le r} \frac{x}{p_i p_j p_k} \pm \cdots \\ &= x \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right) = x \prod_{p \le \sqrt{x}} \left(1 - \frac{1}{p}\right).\end{aligned} \tag{1.1.4}$$

Now, Mertens proved that

$$\prod_{p \le z} \left(1 - \frac{1}{p}\right) = \frac{e^{-\gamma}}{\log z} \left(1 + O\left(\frac{1}{\log z}\right)\right) \quad (z \ge 2), \tag{1.1.5}$$

where γ denotes the Euler-Mascheroni constant. Inserting (1.1.5) into (1.1.4) and combining the resulting estimate with (1.1.2) leads to the prediction that

$$\pi(x) \sim \frac{2e^{-\gamma} x}{\log x} \quad (x \to \infty).$$

Comparing this estimate with (1.1.1), we see that it overestimates $\pi(x)$, since $2e^{-\gamma} = 1.1229189671\ldots > 1$, a typical feature of sieve methods, as we will see. In order to understand why this happens, it is convenient to recast formula (1.1.3) in a more compact way, using the Möbius function μ, so that (1.1.3) becomes

$$\#\{n \le x : p|n \Rightarrow p > \sqrt{x}\} = \sum_{p|d \Rightarrow p \le \sqrt{x}} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor. \tag{1.1.6}$$

The asymptotic formula $\lfloor x/d \rfloor = x/d + O(1)$ then implies that

$$\begin{aligned} \#\{n \le x : p|n \Rightarrow p > \sqrt{x}\} &= x \sum_{p|d \Rightarrow p \le \sqrt{x}} \frac{\mu(d)}{d} + O(2^{\pi(\sqrt{x})}) \\ &= \frac{(2e^{-\gamma} + o(1))x}{\log x} + O(4^{(1+o(1))\sqrt{x}/\log x}). \end{aligned} \tag{1.1.7}$$

So we see that our attempt to obtain an asymptotic formula for $\pi(x)$ fails dramatically, as the error term in (1.1.7) is much bigger than the main term. This happens for two interconnected reasons:

- The numbers d in the sum appearing on the right hand side (1.1.6) are in one-to-one correspondence with the divisors of $\prod_{p \le \sqrt{x}} p$, and there are too many of these.
- The numbers d in (1.1.6) can get as big as

$$\prod_{p \le \sqrt{x}} p = \exp\left\{ \sum_{p \le \sqrt{x}} \log p \right\} = e^{\sqrt{x}(1+o(1))},$$

  which is enormous compared to x. Therefore the approximation $\lfloor x/d \rfloor = x/d + O(1)$ is very bad for most d in (1.1.6).

Nevertheless, Legendre observed that it is possible to use the above idea to gain some information on the size of $\pi(x)$. His starting point was that, for any $z \in [1, x]$, the set of integers that have all their prime factors $> z$ contain the primes in the interval $(z, x]$. So

$$\pi(x) \le \pi(z) + \#\{n \le x : p|n \Rightarrow p > z\} \le z + \#\{n \le x : p|n \Rightarrow p > z\}.$$

Moreover, as before, the inclusion-exclusion principle and Mertens' estimate imply that

$$\begin{aligned} \#\{n \le x : p|n \Rightarrow p > z\} &= \sum_{p|d \Rightarrow p \le z} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor = \sum_{p|d \Rightarrow p \le z} \mu(d) \left(\frac{x}{d} + O(1) \right) \\ &= x \prod_{p \le z} \left(1 - \frac{1}{p} \right) + O\left(2^{\pi(z)}\right) \ll \frac{x}{\log z} + 2^z \end{aligned}$$

and, consequently,

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

This formula is valid for all $z \in [1, x]$. Taking $z = (\log x)/2$ then yields that

$$\pi(x) \ll \frac{x}{\log \log x}, \tag{1.1.8}$$

a non-trivial estimate for $\pi(x)$. Of course, this estimate is much worse that Chebyshev's estimate $\pi(x) \ll x/\log x$. However, as the following exercise shows, it is possible to use these ideas to obtain other interesting results.

Exercise 1.1.2 (The square-free sieve)**.** Use the ideas developed above to give a new proof of the estimate

$$\#\{n \leq x : n \text{ is square-free}\} = x \cdot \prod_p \left(1 - \frac{1}{p^2}\right) + O(\sqrt{x}) \quad (x \geq 2).$$

Compare the sieve-theoretic and multiplicative-theoretic proofs.

Exercise 1.1.3. Use Legendre's refinement of the sieve of Eratosthenes to show that

$$\pi(x) = o(x) \quad (x \to \infty),$$

without using (1.1.5).

Exercise 1.1.4. Find the average value of the greatest common divisor of a and b asymptotically, as a and b range over all integers up to x.

1.2 General set-up

The application of sieve-theoretic ideas to the study of $\pi(x)$ is perhaps not the most engaging example, since we already know quite a bit about $\pi(x)$ using the theory of the Riemann ζ function. However, there are various other prime-counting problems in which the theory of L-functions is not applicable. Some of the most famous ones are:

- The twin prime conjecture: Are there infinitely many pairs of integers $(n, n+2)$ which are both prime? (Such pairs are called twin primes.)
- Goldbach's conjecture: Can every even integer greater than 2 be written as the sum of two primes?
- Is there a prime number between two consecutive squares?
- Are there infinitely many primes of the form $n^2 + 1$?

To this day, all of the above problems remain wide open. However, it is possible to use sieve methods to make some progress towards them. Examples of results that can be proven using sieve methods include:

- There are infinitely many primes p such that $p+2$ has at most two prime factors (Chen, 1966).
- $\sum_{p,\, p+2 \text{ twin primes}} 1/p < \infty$ (Brun, 1908).
- For every large m, the interval $(m^2, (m+1)^2)$ contains an integer with at most 2 prime factors (Chen, 1975).
- There are infinitely many integers n such that $n^2 + 1$ has at most 2 prime factors (Iwaniec, 1980).
- There are infinitely many primes of the form $a^2 + b^4$ (Friedlander - Iwaniec, 2004).

In the context of sieve methods, all four problems listed in the beginning of this chapter as well as many others can be viewed in a unified way. To see this, we need to introduce some notation. Let $\mathcal{A}$ be a finite set of integers and $z \geq 1$ some real number. We set

$$P(z) = \prod_{p<z} p$$

and

$$S(\mathcal{A}, z) = \#\{a \in \mathcal{A} : (a, P(z)) = 1\}.$$

By Theorem 1.1.1, if we wish to extract primes (or product of primes) from the set $\mathcal{A}$, then we need to be able to obtain non-trivial lower bounds on $S(\mathcal{A}, z)$ with z as big as $\max\{p | \prod_{a\in\mathcal{A}} a\}^{1/2}$. For example,

- To count the number of twin prime pairs $(n, n+2)$ with $n \leq x$, we take

$$\mathcal{A} = \{n(n+2) : n \leq x\} \quad \text{and} \quad z = \sqrt{x+2}.$$

  Alternatively, we can take

$$\mathcal{A} = \{p+2 : p \leq x\} \quad \text{and} \quad z = \sqrt{x+2}.$$

- To count the number of representations of the even integer $2N$ as the sum of two primes, we take

$$\mathcal{A} = \{n(2N-n) : n \leq 2N\} \quad \text{and} \quad z = \sqrt{2N}.$$

- To count the number of primes in the interval $(m^2, (m+1)^2)$, we take

$$\mathcal{A} = \{n \in (m^2, (m+1)^2)\} \quad \text{and} \quad z = m+1.$$

- To count the number of primes of the form n^2+1 with $n \leq x$, we take

$$\mathcal{A} = \{n^2+1 : n \leq x\} \quad \text{and} \quad z = \sqrt{x^2+1} \sim x.$$

The above examples indicate how flexible this terminology is. Analyzing $S(\mathcal{A}, z)$ will be one of the main objective of this course. Recall the Möbius inversion formula:

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

So

$$S(\mathcal{A}, z) = \sum_{a\in\mathcal{A}} \sum_{d|(a,P(z))} \mu(d) = \sum_{d|P(z)} \mu(d)|\mathcal{A}_d|, \tag{1.2.1}$$

where

$$\mathcal{A}_d := \{a \in \mathcal{A} : a \equiv 0 \,(\mathrm{mod}\, d)\}.$$

In order to proceed we write

$$|\mathcal{A}_d| = g(d) \cdot X + r_d, \tag{A1}$$

for every integer d, where

- X is some positive number, which we think as an approximation to $|\mathcal{A}|$,
- $g : \mathbb{N} \to [0, 1]$ is a multiplicative function such that

  (A2) $$0 \le g(p) < 1 \quad (p \text{ prime}),$$

  which represents the 'probability' that a member of $\mathcal{A}$ is divisible by d,
- r_d is some real number, which we think as an error term.

The motivation behind the assumption that g is multiplicative comes from the assumption/belief that the events $\mathcal{A}_{d_1}$ and $\mathcal{A}_{d_2}$ are roughly independent if d_1 and d_2 are co-prime, something which very often happens in practice. Lastly, we set

$$V(z) = \prod_{p<z}(1 - g(p)),$$

the 'probability' that an element of $\mathcal{A}$ has no prime factors $< z$.

Inserting (A1) into (1.2.1), we obtain the *exact* formula

$$S(\mathcal{A}, z) = X \sum_{d|P(z)} \mu(d)g(d) + \sum_{d|P(z)} \mu(d)r_d = X \cdot V(z) + \sum_{d|P(z)} \mu(d)r_d \tag{1.2.2}$$

Ignoring for a moment the second term, we find that

$$S(\mathcal{A}, z) \approx X \cdot V(z). \tag{1.2.3}$$

Of course, in general, this is just wishful thinking: the 'error term' $\sum_{d|P(z)} r_d$ contains too many terms, so that even if we make the strong assumption that $r_d = O(1)$, the error term is $\ll 2^{\pi(z)}$. However, in most application we have that $g(p) \ll_{\mathcal{A}} 1/p$ and hence $X \cdot V(z) \gg X/(\log X)^{O_{\mathcal{A}}(1)}$ (see the next section). Consequently, arguing along these lines, we need to take $z \ll \log X$ for the approximation (1.2.3) to be accurate.

As a reality check, let us see how good the bound supplied by (1.2.3) is for the problem of counting twin prime pairs: Let $\mathcal{A} = \{n(n + 2) : n \le x\}$ and $z \le x$, so that

$$\#\{n \le x : n,\, n + 2 \text{ are both primes}\} \le z + S(\mathcal{A}, z), \tag{1.2.4}$$

for every $z \in [1, x]$. Note that

$$|\mathcal{A}_d| = \nu(d)\left(\frac{x}{d} + O(1)\right) = x \cdot \frac{\nu(d)}{d} + O(\nu(d)), \tag{1.2.5}$$

where

$$\nu(d) = \#\{n \in \mathbb{Z}/d\mathbb{Z} : n(n + 2) \equiv 0 \,(\mathrm{mod}\, d)\}.$$

The function ν is multiplicative by the Chinese Remainder Theorem. Setting $g(d) = \nu(d)/d$, we see that (A1) is satisfied with $r_d \ll \nu(d)$. Moreover, for every prime p, we have that

$$\nu(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p > 2. \end{cases} \tag{1.2.6}$$

Therefore we find that $V(z) \asymp 1/(\log z)^2$ and relation (1.2.2) yields the estimate

$$S(\mathcal{A}, z) \ll \frac{x}{(\log z)^2} + \sum_{d|P(z)} \mu^2(d)\nu(d) \ll \frac{x}{(\log z)^2} + 3^{\pi(z)} \leq \frac{X}{\log X} + 3^z.$$

Choosing $z = (\log x)/3$ in the above estimate and combining the resulting estimate with (1.2.4), we deduce that

$$\#\{n \leq x : n,\, n+2 \text{ are both prime}\} \ll \frac{x}{(\log\log x)^2}.$$

This bound is however worse than the semi-trivial bound

$$\#\{n \leq x : n,\, n+2 \text{ are both prime}\} \leq \pi(x) \ll \frac{x}{\log x}.$$

In conclusion, we see that the sieve of Eratosthenes-Legendre suffers a lot from a quantitative point of view. We will see how this deficiency was remedied, first by Viggo Brun and then by others. In particular, we shall show that (1.2.3) holds if $\log z = o(\log X)$ and $\mathcal{A}$ is 'structured'.

1.3 Sifting dimension

We conclude this introductory chapter to sieve methods with a brief discussion of a concept which plays an important role in them. This concept is called the *sifting dimension*, which we will denote by κ. Roughly speaking, κ corresponds to the average value of $g(p)p$, as p runs over all primes, provided of course that the latter exists. If $\mathcal{A} = \{f(n) : n \in I\}$, where f is a polynomial and I is some interval of the real line, then there is a more conceptual way to interpret the sifting dimension: it corresponds to the average number of congruence classes that we need to 'remove' modulo each prime in order to extract primes (or products of primes) from the indexing set I. Indeed, if $\mathcal{A} = \{n \leq x\}$, then in order to detect primes in $\mathcal{A}$, we need to 'remove' from $\mathbb{N} \cap [1, x]$ the congruence class $0 \,(\mathrm{mod}\, p)$ for each $p \leq \sqrt{x}$, that is to say, $\kappa = 1$. On the other hand, if $\mathcal{A} = \{n(n+2) : n \leq x\}$, then in order to detect products of two primes in $\mathcal{A}$ (and hence twin primes), we need to 'remove' from $\mathbb{N} \cap [1, x]$ the congruence classes $0 \,(\mathrm{mod}\, p)$ and $-2 \,(\mathrm{mod}\, p)$ for each $p \leq \sqrt{x+2}$. Hence[1], in this case, $\kappa = 2$. Generally speaking, the sieving problem becomes harder as κ increases.

One way to detect the condition that $g(p)p$ is κ on average is by imposing the condition

$$\text{(A3)} \qquad \sum_{p \leq x} g(p) \log p = \kappa \log x + O(1) \quad (x \geq 1).$$

Often, the above condition is unnecessarily strong and we can get away with the weaker condition

$$\text{(A4a)} \qquad \frac{V(w)}{V(w')} = \prod_{w \leq p < w'} (1 - g(p))^{-1} \leq K \left(\frac{\log w'}{\log w}\right)^{\kappa} \quad (3/2 \leq w \leq w').$$

[1]Note that when $p = 2$, the classes $0 \,(\mathrm{mod}\, 2)$ and $-2 \,(\mathrm{mod}\, 2)$ coincide. However, this does not affect things *on average*.

Note that (A4a) immediately implies that

$$g(p) \le 1 - \frac{1}{K} \quad (p \text{ prime}). \tag{1.3.1}$$

In practice (cf. relation (1.1.5)), we can often establish the stronger inequality

$$\frac{V(w)}{V(w')} \le \left\{1 + \frac{C_1}{\log w}\right\} \left(\frac{\log w'}{\log w}\right)^{\kappa} \quad (3/2 \le w \le w'), \tag{A4b}$$

for some $C_1 \ge 0$.

Exercise 1.3.1. Verify the 'sieve axioms' (A1), (A2), (A3) and (A4b) when $\mathcal{A}$ is the set $\{n(2N-n) : n \le 2N\}$, $\{x^2 < n \le (x+1)^2\}$, or $\{n^2+1 : n \le N\}$.

Exercise 1.3.2. Verify the 'sieve axioms' (A1), (A2), (A3) and (A4b) when $\mathcal{A} = \{p+2 : p \le x\}$. For this set, $S(\mathcal{A}, \sqrt{x+2})$ counts twin primes. This is also true if $\mathcal{A} = \{n(n+2) : n \le x\}$. What difference do you notice among these two choices of $\mathcal{A}$?

Chapter 2

Interlude: probabilistic number theory

Before we embark on the study of the combinatorial sieve, and in order to motivate Brun's pure sieve, we give a very brief introduction to probabilistic number theory.

2.1 The number of prime factors of an integer

In probabilistic number theory instead of studying the properties of a fixed integer, we focus on statistical properties of the integers. For example, we often talk of *a typical integer*, an expression which refers to an integer that enjoys properties shared among almost all other integers too[1]. More precisely, when we say that a typical integer n has some property $Q(n)$, we mean that the set of integers n which do not satisfy $Q(n)$ has density 0, in the sense that

$$\lim_{x\to\infty} \frac{1}{x}\#\{n \leq x : Q(n) \text{ does not hold}\} = 0.$$

The prototypical such result was proven by Hardy and Ramanujan:

Theorem 2.1.1. *Fix some $\epsilon > 0$. Then we have that*

$$\lim_{x\to\infty} \frac{1}{x}\#\{n \leq x : (1-\epsilon)\log\log x \leq \omega(n) \leq (1+\epsilon)\log\log x\} = 1.$$

The above theorem can be expressed loosely by saying that a typical integer n has about $\log\log n$ prime factors. Hardy and Ramanujan deduced theorem 2.1.1 from the following result.

Theorem 2.1.2. *There exists constants A and B such that*

$$\pi_r(x) = \#\{n \leq x : \omega(n) = r\} \leq \frac{Ax}{\log x}\frac{(\log\log x + B)^{r-1}}{(r-1)!},$$

uniformly for $x \geq 1$ and $r \in \mathbb{N}$.

[1]This expression is often used in a non-rigorous fashion.

Proof. When $r = 1$, we have that

$$\pi_1(x) = \pi(x) + \sum_{2 \le k \le \frac{\log x}{\log 2}} \sum_{p^k \le x} 1 \le \pi(x) + \frac{\log x}{\log 2} \cdot \sqrt{x} \le \frac{cx}{\log x} \quad (x \ge 3),$$

for some constant c, by Chebyshev's estimate $\pi(x) \ll x/\log x$. We will show the theorem with $A = c$ and B large enough.

We argue by induction: assume that the result holds for some $r \in \mathbb{N}$. Let $n \le x$ be an integer with $r+1$ distinct prime factors, say $n = p_1^{a_1} \cdots p_{r+1}^{a_{r+1}}$ with $p_1 < p_2 < \cdots < p_{r+1}$. Then $p_j^{a_j+1} < p_j^{a_j} p_{r+1} \le x$ for $j \le r$, so that there are at least r ways to write $n = p^a m$ for some $p \le x^{1/(a+1)}$, $a \ge 1$ and m with $\omega(m) = r$. Consequently,

$$\begin{aligned} r\pi_{r+1}(x) &\le \sum_{a \ge 1} \sum_{p^a \le x^{1/(a+1)}} \pi_r(x/p^a) \le \sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{Ax/p^a}{\log(x/p^a)} \frac{(\log\log(x/p^a) + B)^{r-1}}{(r-1)!} \\ &\le \frac{Ax(\log\log x + B)^{r-1}}{(r-1)!} \sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{1}{p^a \log(x/p^a)}. \end{aligned}$$

Therefore, we need to show that $\sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{1}{p^a \log(x/p^a)} \le (\log\log x + O(1))/\log x$. Then choosing B large enough will complete the inductive step and hence the proof. Indeed, note that $1/(1-y) \le 1 + (a+1)y$ for $y \in [0, a/(a+1)]$. So

$$\begin{aligned} \sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{1}{p^a \log(x/p^a)} &= \frac{1}{\log x} \sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{1}{p^a \left(1 - \frac{a \log p}{\log x}\right)} \\ &\le \frac{1}{\log x} \sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{1}{p^a} \left(1 + \frac{(a+1)\log p}{\log x}\right) \\ &\le \frac{1}{\log x} \sum_{p \le \sqrt{x}} \frac{1}{p-1} \left(1 + \frac{\log p}{\log x} \sum_{a \ge 1} \frac{a+1}{2^{a-1}}\right) \\ &\le \frac{\log\log x + O(1)}{\log x}, \end{aligned}$$

which shows the desired result. □

Exercise 2.1.3. Given $\lambda > 0$, we set

$$Q(\lambda) := \lambda \log \lambda - \lambda + 1 = \int_1^\lambda \log t dt.$$

(a) Show that for $x \ge 2$ and $\lambda > 1$, we have that

$$\#\{n \le x : \omega(n) \ge \lambda \log\log x\} \ll_\lambda \frac{x}{(\log x)^{Q(\lambda)} \sqrt{\log\log x}}.$$

Similarly, show that for $0 < \lambda < 1$, we have that

$$\#\{n \le x : \omega(n) \le \lambda \log\log x\} \ll_\lambda \frac{x}{(\log x)^{Q(\lambda)} \sqrt{\log\log x}}.$$

(b) Deduce Theorem 2.1.1.

Exercise 2.1.4.

(a) Show that the entries of the $N \times N$ multiplication table form a sparse set of the integers in the sense that

$$\lim_{N\to\infty} \frac{1}{N^2} \#\{ab : a \le N,\, b \le N\} = 0$$

(b)* Show that

$$\#\{ab : a \le N,\, b \le N\} \ll \frac{N^2}{(\log N)^\delta \sqrt{\log\log N}},$$

where $\delta = Q(1/\log 2) = 1 - (1 + \log\log 2)/\log 2 = 0.08607\ldots$.

2.2 The Kubilius model

Theorem 2.1.2 leads to prediction that ω is distributed like a Poisson random variable on the set $\{n \le x\}$ and has mean value $\log\log x$. There is a nice heuristic argument in support of this statement based on the Kubilius model of the integers. Note that

$$\frac{1}{\lfloor x \rfloor} \#\{n \le x : d|n\} = \frac{\lfloor x/d \rfloor}{\lfloor x \rfloor} \approx \frac{1}{d}.$$

Motivated by this rough calculation, Kubilius assigned to the event $\{d|n\}$ the probability $1/d$. In addition, note that if d_1 and d_2 and co-prime, then

$$\mathbf{Prob}(\{d_1|n\} \cap \{d_2|n\}) = \mathbf{Prob}(\{d_1 d_2|n\}) = \frac{1}{d_1 d_2} = \mathbf{Prob}(\{d_1|n\})\,\mathbf{Prob}(\{d_2|n\}).$$

So it is reasonable to assume that the events $\{d_1|n\}$ and $\{d_2|n\}$ are independent of each other. This leads to the estimate

$$\begin{aligned}
&\mathbf{Prob}(\{\omega(n) = r \,|\, n \le x\}) \\
&\quad = \sum_{p_1<\cdots<p_r\le x} \mathbf{Prob}(\{p_1\cdots p_r|n\} \cap \{p \nmid n \ \forall p \in [1,x] \setminus \{p_1,\ldots,p_r\}\}) \\
&\quad = \sum_{p_1<\cdots<p_r\le x} \frac{1}{p_1\cdots p_r} \prod_{\substack{p\le x \\ p\notin\{p_1,\ldots,p_r\}}} \left(1-\frac{1}{p}\right) = \prod_{p\le x}\left(1-\frac{1}{p}\right) \sum_{p_1<\cdots<p_r\le x} \frac{1}{(p_1-1)\cdots(p_r-1)} \\
&\quad = \frac{1}{r!}\prod_{p\le x}\left(1-\frac{1}{p}\right) \sum_{\substack{p_1,\ldots,p_r\le x \\ \text{distinct}}} \frac{1}{(p_1-1)\cdots(p_r-1)} \approx \frac{1}{r!}\prod_{p\le x}\left(1-\frac{1}{p}\right)\left(\sum_{p\le x}\frac{1}{p}\right)^r \\
&\quad \asymp \frac{1}{\log x}\frac{(\log\log x + O(1))^r}{r!},
\end{aligned}$$

which confirms heuristically our prediction.

It is possible to go beyond Theorem 2.1.2 and show that $\omega(n;t) = \{p|n : p \leq t\}$ is also distributed like a Poisson random variable with mean value $\log\log t$, for $t \leq n$. So if $n = p_1^{a_1}p_2^{a_2}\cdots p_r^{a_r}$ is the prime factorization of a typical integer n, then we expect that $j = \omega(n;p_j) \approx \log\log p_j$, that is to say, the sequence $\{\log\log p\}_{p|n}$ is very close to being an arithmetic progression. This implies that

$$\log(p_1\cdots p_j) = \sum_{i=1}^{j} \log p_i \approx \log p_j, \tag{2.2.1}$$

since the sequence $\{\log p_i\}_{i=1}^{r}$ is close to being a geometric progression[2].

We shall be using statements like the ones above - or at least the spirit of them - to justify certain choices in the subsequent sections. As a matter of fact, we have already used something similar in the previous chapter. Indeed, if we assume that the events $\mathcal{A}_{d_1}$ and $\mathcal{A}_{d_2}$ are independent of each other whenever d_1 and d_2 are coprime, then $\mathbf{Prob}(\mathcal{A}_d)$ should be a is a multiplicative function. This justifies the assumption that $g(d) \approx |\mathcal{A}_d|/|\mathcal{A}| = \mathbf{Prob}(\mathcal{A}_d)$ is a multiplicative function.

For similar reasons, we expect that

$$\begin{aligned}\mathbf{Prob}(\{(a,P(z)) = 1 \mid a \in \mathcal{A}\}) &= \mathbf{Prob}\left(\bigcap_{p<z} \{a \not\equiv 0 \,(\mathrm{mod}\, p) \mid a \in \mathcal{A}\}\right)\\ &= \prod_{p<z} \mathbf{Prob}(\{a \not\equiv 0 \,(\mathrm{mod}\, p) \mid a \in \mathcal{A}\})\\ &\approx \prod_{p<z}(1-g(p)) = V(z).\end{aligned}$$

So we see that, even though the Kubilius model is very successful in predicting the distribution of ω, it fails in the case of $S(\mathcal{A},z)$. The reason for this failure lies in the assumption that the events $\mathcal{A}_{d_1}$ and $\mathcal{A}_{d_2}$ are independent of each other whenever d_1 and d_2 are coprime. Indeed, if for example $\mathcal{A} = \{n \leq x\}$ and $z = \sqrt{x}$, then an integer $a \leq x$ is divisible by at most three distinct primes $p \in (x^{1/3}, z]$, which certainly violates the independence assumption we made above[3]. In rough terms, this is why it is hard to obtain asymptotic formulas on $S(\mathcal{A},z)$ when z is large and, in fact, why the approximation $S(\mathcal{A},z) \approx X \cdot V(z)$ is generally false if z is large. However, we shall see in the next sections that the Kubilius model is accurate when $z = X^{o(1)}$.

[2]Indeed, if $\lambda > 1$, then $\lambda^N \leq 1 + \lambda + \lambda^2 + \cdots + \lambda^N \leq \frac{\lambda}{\lambda-1}\cdot\lambda^N$.

[3]This phenomenon is not significant in the study of ω, since n has at most 1000 prime factors in $[n^{1/1000}, n]$. So we may ignore these large primes without sacrificing a whole lot in precision.

Chapter 3

The combinatorial sieve

3.1 Brun's pure sieve

As we saw in Chapter 1, one of the main deficiencies of (1.2.2) is that the 'error term' $\sum_{d|P(z)} \mu(d) r_d$ contains too many terms. It was Viggo Brun who first realized that is possible to remedy this deficiency using some simple combinatorial ideas: Formula (1.2.1), and hence formula (1.2.2), is a consequence of the inclusion-exclusion principle and, when written out fully, it reads

$$\begin{aligned} S(\mathcal{A}, z) &= |\mathcal{A}| - \sum_{p_1<z} |\mathcal{A}_{p_1}| + \sum_{p_2<p_1<z} |\mathcal{A}_{p_1p_2}| - \sum_{p_3<p_2<p_1<z} |\mathcal{A}_{p_1p_2p_3}| \pm \cdots \\ &= |\mathcal{A}| + \sum_{j\geq 1} (-1)^j \sum_{p_j<\cdots<p_1<z} |\mathcal{A}_{p_1\cdots p_j}|. \end{aligned}$$

But it is a well-known fact that

$$|\mathcal{A}| + \sum_{j=1}^{2m-1} (-1)^j \sum_{p_j<\cdots<p_1<z} |\mathcal{A}_{p_1\cdots p_j}| \leq S(\mathcal{A}, z) \leq |\mathcal{A}| + \sum_{j=1}^{2m} (-1)^j \sum_{p_j<\cdots<p_1<z} |\mathcal{A}_{p_1\cdots p_j}| \tag{3.1.1}$$

for every $m \in \mathbb{N}$. This implies that, for every $r \in \mathbb{N}$, we have that

$$\begin{aligned} S(\mathcal{A}, z) &= |\mathcal{A}| + \sum_{1\leq j<r} (-1)^j \sum_{p_j<\cdots<p_1<z} |\mathcal{A}_{p_1\cdots p_j}| + O\left(\sum_{p_r<\cdots<p_1<z} |\mathcal{A}_{p_1\cdots p_r}| \right) \\ &= \sum_{\substack{d|P(z) \\ \omega(d)<r}} \mu(d) |\mathcal{A}_d| + O\left(\sum_{\substack{d|P(z) \\ \omega(d)=r}} |\mathcal{A}_d| \right) \\ &= X \sum_{\substack{d|P(z) \\ \omega(d)<r}} \mu(d) g(d) + O\left(\sum_{\substack{d|P(z) \\ \omega(d)\leq r}} |r_d| + X \sum_{\substack{d|P(z) \\ \omega(d)=r}} g(d) \right), \end{aligned} \tag{3.1.2}$$

by (A1). Even though the above relation holds for all r, Theorem 2.1.1 suggests that, in order to have any hope of succeeding, we must take r as a function of x and z (usually, $r \approx c \log\log x$ suffices).

Instead of estimating the main term in (3.1.2) directly, we observe that

$$V(z) = 1 + \sum_{j\geq 1}(-1)^j \sum_{p_j<\cdots<p_1<z} g(p_1\cdots p_j).$$

So

$$1 + \sum_{j=1}^{2m+1}(-1)^j \sum_{p_j<\cdots<p_1<z} g(p_1\cdots p_j) \leq V(z) \leq 1 + \sum_{j=1}^{2m}(-1)^j \sum_{p_j<\cdots<p_1<z} g(p_1\cdots p_j) \tag{3.1.3}$$

for every $m \in \mathbb{N}$ and, consequently,

$$\sum_{\substack{d|P(z)\\ \omega(d)<r}} \mu(d)g(d) = V(z) + O\left(\sum_{\substack{d|P(z)\\ \omega(d)=r}} g(d)\right). \tag{3.1.4}$$

Exercise 3.1.1. Show relation (3.1.3).

Inserting (3.1.4) into (3.1.2), we deduce that

$$\begin{aligned} S(\mathcal{A},z) &= X\cdot V(z) + O\left(X\sum_{\substack{d|P(z)\\ \omega(d)=r}} g(d) + \sum_{\substack{d|P(z)\\ \omega(d)\leq r}} |r_d|\right) \\ &= X\cdot V(z) + O\left(\frac{X}{r!}\left(\sum_{p<z} g(p)\right)^r + \sum_{\substack{d|P(z),\, d\leq z^r\\ \omega(d)\leq r}} |r_d|\right). \end{aligned} \tag{3.1.5}$$

Since

$$\sum_{p<z} g(p) \leq \sum_{p<z} \log\frac{1}{1-g(p)} = \frac{1}{\log V(z)},$$

we obtain the formula

$$S(\mathcal{A},z) = X\cdot V(z) + O\left(\frac{X\cdot|\log V(z)|^r}{r!} + \sum_{\substack{d|P(z),\, d\leq z^r\\ \omega(d)\leq r}} |r_d|\right).$$

By Stirling's formula, the main term in the above formula is bigger than the first error term if $r > 3.6|\log V(z)|$. Indeed, we have the following estimate:

Theorem 3.1.2. *Let $\mathcal{A}$ be a finite set of integers which satisfies* (A1) *and* (A2). *For $z \geq 1$ and $r \geq 3.6|\log V(z)|$, we have that*

$$S(\mathcal{A}, z) = X \cdot V(z) \left\{ 1 + O\left(\frac{1}{\sqrt{r}}\right) \right\} + O\left(\sum_{\substack{d|P(z) \\ d \leq z^r, \omega(d) \leq r}} |r_d| \right).$$

Controlling the second error in Theorem 3.1 is often very hard and depends on having at our disposal an estimate of the form

$$\sum_{d \leq D} \mu^2(d)|r_d| \leq \frac{C_2 X}{(\log X)^A} \tag{R}$$

with $D = z^r$ and A some large enough positive number. If such an estimate holds, then we say that $\mathcal{A}$ *has level of distribution* D. Assuming(A4a) and (R), we may simplify the statement of Theorem 3.1.2.

Theorem 3.1.3. *Let $\mathcal{A}$ be a finite set of integers which satisfies* (A1) *and* (A4a), *as well as* (R) *with $A = \kappa + 1$ and $D = X^\theta$ for some $\theta \in (0, 1]$. For $1 \leq z \leq X^{\theta/(4\kappa \log\log X)}$, we have that*

$$S(\mathcal{A}, z) = X \cdot V(z) \left\{ 1 + O\left(\frac{1}{\sqrt{\log\log X}}\right) \right\};$$

the implied constant depends at most of κ, K, θ and C_2.

Proof. Without loss of generality, we may assume that X is large enough. Note that relation (A2) holds by (1.3.1). The result then follows by Theorem 3.1.2 applied with $r = 4\kappa \log\log X \geq 3.6|\log V(z)|$, so that $z^r \leq X^\theta$. □

In the next section we will see that, by improving upon Brun's idea, it is possible to extend the above theorem to $z \leq X$. But before this, we give a nice application of Theorem 3.1.3, due to Brun.

Theorem 3.1.4. *We have that*

$$\sum_{p, p+2 \text{ twin primes}} \frac{1}{p} < \infty.$$

Remark 3.1.5. The value of the sum $\sum_{p, p+2 \text{ twin primes}} \frac{1}{p}$ is called Brun's constant. Since we know that $\sum_p 1/p = +\infty$, the above theorem tells us that primes p for which $(p, p+2)$ are twin primes are sparse among the sequence of primes.

Proof of Theorem 3.1.4. First, we apply Theorem 3.1.3 with $\mathcal{A} = \{n(n+2) : n \leq x\}$. Note that (A1) holds with $X = x$, $g(d) = \nu(d)/d$ and $r_d \ll \nu(d)$, where

$$\nu(d) = \#\{n \in \mathbb{Z}/d\mathbb{Z} : n(n+2) \equiv 0 \,(\mathrm{mod}\, d)\},$$

by (1.2.5). Moreover, (A4a) holds with $\kappa = 2$, by (1.2.6). Finally, (R) holds for any $A > 0$ and any $\theta < 1$, since $|\nu(d)| \le \tau_2(d)$ for d square-free. So Theorem 3.1.3, applied with $z = x^{1/(10 \log\log x)}$, yields the estimate

$$\#\{n \le x : n,\, n+2 \text{ twin primes}\} \le z + S(\mathcal{A}, z) \ll z + \frac{x}{(\log z)^2} \ll \frac{x(\log\log x)^2}{(\log x)^2}.$$

The desired result then follows by partial summation. □

3.2 Buchstab iterations and general upper & lower bound sieves

Buchstab observed that

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p<z} S(\mathcal{A}_p, p). \tag{3.2.1}$$

This identity can be used to obtain an entire family of combinatorial sieves, as follows: Consider sets

$$\Pi_j \subset \{(p_1, \dots, p_j) : z > p_1 > p_2 > \cdots > p_j \text{ primes}\} \quad (j \ge 1)$$

such that

$$\Pi_3 \subset \Pi_1 \times \{p < z\}^2,\ \Pi_5 \subset \Pi_3 \times \{p < z\}^2,\ \dots,\ \Pi_{2j+1} \subset \Pi_{2j-1} \times \{p < z\}^2,\ \dots$$

and

$$\Pi_4 \subset \Pi_2 \subset \{p < z\}^2,\ \Pi_6 \subset \Pi_4 \times \{p < z\}^2,\ \dots,\ \Pi_{2j+2} \subset \Pi_{2j} \times \{p < z\}^2,\ \dots$$

Moreover, set

$$\mathcal{D}^+ = \{1\} \cup \{d = p_1 \cdots p_r > 1 : (p_1, \dots, p_j) \in \Pi_j,\ 1 \le j \le r,\ j \text{ odd}\}$$

and

$$\mathcal{D}^- = \{1\} \cup \{d = p_1 \cdots p_r > 1 : (p_1, \dots, p_j) \in \Pi_j,\ 1 \le j \le r,\ j \text{ even}\}.$$

Then we have that

$$\begin{aligned}
S(\mathcal{A}, z) &= |\mathcal{A}| - \sum_{p_1<z} S(\mathcal{A}_{p_1}, p_1) \le |\mathcal{A}| - \sum_{p_1 \in \Pi_1} S(\mathcal{A}_{p_1}, p_1) \\
&= |\mathcal{A}| - \sum_{p_1 \in \Pi_1} |\mathcal{A}_{p_1}| + \sum_{\substack{p_2<p_1 \\ p_1 \in \Pi_1}} S(\mathcal{A}_{p_1p_2}, p_2) \\
&= |\mathcal{A}| - \sum_{p_1 \in \Pi_1} |\mathcal{A}_{p_1}| + \sum_{\substack{p_2<p_1 \\ p_1 \in \Pi_1}} |\mathcal{A}_{p_1p_2}| - \sum_{\substack{p_3<p_2<p_1 \\ p_1 \in \Pi_1}} S(\mathcal{A}_{p_1p_2p_3}, p_3) \\
&\le |\mathcal{A}| - \sum_{p_1 \in \Pi_1} |\mathcal{A}_{p_1}| + \sum_{\substack{p_2<p_1 \\ p_1 \in \Pi_1}} |\mathcal{A}_{p_1p_2}| - \sum_{(p_1,p_2,p_3) \in \Pi_3} S(\mathcal{A}_{p_1p_2p_3}, p_3) \\
&\le \cdots \le \sum_{d \mid P(z)} \mu^+(d) |\mathcal{A}_d|,
\end{aligned} \tag{3.2.2}$$

where

$$\mu^+(d) = \begin{cases} \mu(d) & \text{if } d \in \mathcal{D}^+, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.3}$$

Exercise 3.2.1. Show that

$$S(\mathcal{A}, z) \geq \sum_{d|P(z)} \mu^-(d)|\mathcal{A}_d|,$$

where

$$\mu^-(d) = \begin{cases} \mu(d) & \text{if } d \in \mathcal{D}^-, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2.4}$$

Exercise 3.2.2. Show that

$$(1 * \mu^-)(n) \leq \mathbf{1}_{P^-(n)>z} \leq (1 * \mu^+)(n).$$

This construction gives a great deal of flexibility. For example, the choice $\Pi_j = \{p < z\}^j$ for $j < r$ and $\Pi_j = \emptyset$ for $j \geq r$ corresponds to Brun's pure sieve. The goal is to choose the sets Π_j as large as possible, so that we are not throwing away too much information, while at the same time if $(p_1, \dots, p_j) \in \Pi_j$, then $p_1 \cdots p_j$ is not too big, so that we can control well the sum $\sum_{(p_1,\dots,p_j)\in\Pi_j} |\mathcal{A}_{p_1\cdots p_j}|$ using (A1) and (R).

It turns out that a good choice for the sets Π_j is given by

$$\Pi_j = \left\{ (p_1, p_2, \dots, p_j) : z > p_1 > \cdots > p_j,\ p_1 \cdots p_i < \frac{D}{p_i^\beta},\ 1 \leq i \leq j,\ i \equiv j \,(\mathrm{mod}\, 2) \right\}, \tag{3.2.5}$$

for some parameter β, which depends on the dimension of the sieve problem. This leads to the so-called β sieve, pioneered by Rosser and brought to maturity by Iwaniec [I]. In the next section, we shall see how our choice allows us to improve upon Theorem 3.1.2. For now, we give a heuristic argument [FI10, Section 6.4]) which motivates the above choice of Π_j.

Fix a dimension κ and assume that $\beta = \beta(\kappa)$ is the minimum number $u \geq 1$ such that

$$S(\mathcal{A}, z) \gg X \cdot V(z),$$

for all sieve problems of dimension κ such that $\mathcal{A}$ that satisfy (A1) and has level of distribution $D \geq z^u$. We consider such a triplet $(\mathcal{A}, \mathbb{P}, z)$ and make the further assumption that the level of distribution of $\mathcal{A}_d$ is D/d. Now, we have that

$$S(\mathcal{A}, z) = |\mathcal{A}| - \sum_{p_1<z} S(\mathcal{A}_{p_1}, p_1) = |\mathcal{A}| - \sum_{p_1<z} |\mathcal{A}_{p_1}| + \sum_{\substack{p_2<p_1 \\ p_1\in\Pi_1}} S(\mathcal{A}_{p_1p_2}, p_2).$$

From the sum over p_1 and p_2 we drop the terms with $D/(p_1p_2) < p_2^\beta$, since some of these are potentially much smaller than expected, and keep all of the terms with $D/(p_1p_2) \geq p_2^\beta$, since for these we have that

$$S(\mathcal{A}_{p_1p_2}, p_2) \gg |\mathcal{A}_{p_1p_2}| \cdot V(p_2) \approx g(p_1p_2)X \cdot V(p_2)$$

by our assumption and by (A1). This suggests setting

$$\Pi_2 = \{(p_1, p_2) : z > p_1 > p_2,\, p_1 p_2 < D/p_2^{\beta}\}.$$

Next, we have that

$$\begin{aligned} S(\mathcal{A}, z) &\geq |\mathcal{A}| - \sum_{p_1 < z} |\mathcal{A}_{p_1}| + \sum_{(p_1,p_2)\in\Pi_2} S(\mathcal{A}_{p_1p_2}, p_2) \\ &= |\mathcal{A}| - \sum_{p_1 < z} |\mathcal{A}_{p_1}| + \sum_{(p_1,p_2)\in\Pi_2} |\mathcal{A}_{p_1p_2}| - \sum_{\substack{p_3<p_2<p_1 \\ (p_1,p_2)\in\Pi_2}} S(\mathcal{A}_{p_1p_2p_3}, p_3) \\ &= |\mathcal{A}| - \sum_{p_1 < z} |\mathcal{A}_{p_1}| + \sum_{(p_1,p_2)\in\Pi_2} |\mathcal{A}_{p_1p_2}| - \sum_{\substack{p_3<p_2<p_1 \\ (p_1,p_2)\in\Pi_2}} |\mathcal{A}_{p_1p_2p_3}| + \sum_{\substack{p_4<\cdots<p_1 \\ (p_1,p_2)\in\Pi_2}} S(\mathcal{A}_{p_1\cdots p_4}, p_4) \end{aligned}$$

As before, we drop the terms with $D/(p_1p_2p_3p_4) < p_4^{\beta}$ and keep the rest. This leads us to the choice

$$\Pi_4 = \{(p_1, \ldots, p_4) : z > p_1 > \cdots > p_4,\, p_1 \cdots p_4 < D/p_4^{\beta}\}.$$

Continuing in the above fashion suggests selecting the sets Π_j as in (3.2.5).

In general, a sequence $\mathscr{S}^+ = \{\mu^+(d)\}_{d\leq D}$ such that

$$(S^+) \qquad \mu^+(1) = 1 \quad \text{and} \quad \sum_{d|n} \mu^+(d) \geq 0 \quad (n > 1)$$

is called an *upper bound sieve of level D*. Similarly, a sequence $\mathscr{S}^- = \{\mu^-(d)\}_{d\leq D}$ such that

$$(S^-) \qquad \mu^-(1) = 1 \quad \text{and} \quad \sum_{d|n} \mu^-(d) \leq 0 \quad (n > 1)$$

is called a *lower bound sieve of level D*. The reason for this terminology is that, under the above assumptions, we have that

$$S(\mathcal{A}, z) = \sum_{\substack{a\in\mathcal{A} \\ (a,P(z))=1}} 1 \leq \sum_{a\in\mathcal{A}} \sum_{d|(a,P(z))} \mu^+(d) = \sum_{d|P(z)} \mu^+(d)|\mathcal{A}_d|$$

and

$$S(\mathcal{A}, z) = \sum_{\substack{a\in\mathcal{A} \\ (a,P(z))=1}} 1 \geq \sum_{a\in\mathcal{A}} \sum_{d|(a,P(z))} \mu^-(d) = \sum_{d|P(z)} \mu^-(d)|\mathcal{A}_d|,$$

for any finite set of integers $\mathcal{A}$ and any real number z.

3.3 The fundamental lemma of sieve methods

In this section we show that choosing the sets Π_j as in (3.2.5) for an appropriate β, depending on the dimension of the sieving problem, allows us to evaluate $S(\mathcal{A}, z)$ asymptotically when $z = X^{o(1)}$, thus extending Theorem 3.1.3. Indeed, if

$$\beta_\kappa = 2\left(e^{\frac{1}{2\kappa}} - 1\right)^{-1} + 1 < 1 + 4\kappa, \tag{3.3.1}$$

then we have the following result.

Theorem 3.3.1 (Fundamental Lemma of Sieve Methods, I). *Let $\mathcal{A}$ be a finite set of integers which satisfies* (A1) *and* (A4a)*, for some $\kappa > 0$ and $K \geq 1$. For $z \geq 1$ and $u \geq \epsilon > 0$, we have that*

$$S(\mathcal{A}, z) = X \cdot V(z)\left\{1 + O\left(e^{-u\log u + O_{\kappa,K,\epsilon}(u)}\right)\right\} + O\left(\sum_{d|P(z),\, d\leq z^u} |r_d|\right).$$

If, in addition, (A4b) *holds for some $C_1 \geq 0$, then*

$$S(\mathcal{A}, z) \geq \frac{X \cdot V(z)}{8} + O\left(\sum_{d|P(z),\, d\leq z^{\beta_\kappa}} |r_d|\right),$$

provided that z is large enough in terms of κ, C_1 and ϵ.

As an immediate corollary, we have the following result.

Theorem 3.3.2 (Fundamental Lemma of Sieve Methods, II). *Let $\mathcal{A}$ be a finite set of integers which satisfies* (A1) *and* (A4a)*, for some $\kappa > 0$ and $K \geq 1$, and* (R) *for $A = \kappa + 1$ and $D = X^\theta$, $\theta \in (0, 1]$. If $z = X^{1/s}$ with $s \geq 1$, then we have that*

$$S(\mathcal{A}, z) = X \cdot V(z)\left\{1 + O_{\kappa,K}\left(e^{-\theta s\log s + O_{\kappa,K,\theta}(s)} + \frac{1}{\log X}\right)\right\}.$$

If, in addition, (A4b) *holds for some $C_1 \geq 0$, then*

$$S(\mathcal{A}, z) \geq \frac{X \cdot V(z)}{10} \quad (z_0 \leq z \leq X^{\theta/\beta_k}),$$

where z_0 is a sufficiently large constant that depends at most on κ and C_1.

Proof. The first part of the theorem follows by taking $u = \theta s$ and $\epsilon = \theta$ in the first part of Theorem 3.3.1, since relation (A4a) implies that $V(z) \gg_{\kappa,K} (\log X)^{-\kappa}$. For the second part, note that $z^{\beta_\kappa} \leq X^\theta$. So the desired lower bound on $S(\mathcal{A}, z)$ follows by the second part of Theorem 3.3.1. □

Theorem 3.3.1 is an easy consequence of the following technical result.

Theorem 3.3.3 (Fundamental Lemma of Sieve Methods, III)**.** *Let $\kappa > 0$, $z \geq 1$ and $D = z^u$ with $u \geq 2$. There exist two arithmetic functions $\mu^{\pm} : \mathbb{N} \to [-1, 1]$, depending at most on κ, z and D, such that:*

(1) μ^+ and μ^- are both supported in $\{d | P(z) : d \leq D\}$;

(2) $(\mu^- * 1)(n) = 1 = (\mu^+ * 1)(n)$ if $P^-(n) \geq z$;

(3) $(\mu^- * 1)(n) \leq 0 \leq (\mu^+ * 1)(n)$ if $P^-(n) < z$;

(4) If $g : \mathbb{N} \to [0, 1)$ is a multiplicative function satisfying (A4a) *with parameter κ as above and for some $K \geq 1$, then*

$$\sum_d \lambda(d)g(d) = V(z)\left\{1 + O\left(e^{-u\log u + O_{\kappa,K}(u)}\right)\right\} \quad (\lambda \in \{\mu^+, \mu^-\}).$$

(5) If $g : \mathbb{N} \to [0, 1)$ is a multiplicative function satisfying (A4b) *with parameter κ as above and for some $C_1 \geq 0$, then*

$$\sum_d \mu^-(d)g(d) \geq \frac{V(z)}{8},$$

provided that $u \geq \beta_\kappa$ and that z is large enough in terms of κ and C_1.

Deduction of Theorem 3.3.1 from Theorem 3.3.3. Note that we may assume throughout the proof that $u \geq 2$, since if $\epsilon \leq u \leq 2$, then we have that

$$\begin{aligned} 0 \leq S(\mathcal{A}, z) \leq S(\mathcal{A}, z^{\epsilon/2}) &\leq c_{\kappa,K} \cdot X \cdot V(z^{\epsilon/2}) + O\left(\sum_{d | P(z^{\epsilon/2}),\, d \leq z^\epsilon} |r_d|\right) \\ &\leq c_{\kappa,K} \cdot X \cdot K\left(\frac{2}{\epsilon}\right)^\kappa V(z) + O\left(\sum_{d | P(z),\, d \leq z^u} |r_d|\right) \end{aligned}$$

by the case $u = 2$ and relation (A4a), where $c_{\kappa,K}$ is some constant depending at most on κ and K. Therefore holds in this case too, possibly by enlarging the implied constants in the statement of Theorem 3.3.1. Now, assume that $u \geq 2$ and let $\mu^{\pm}$ be the two sequences from Theorem 3.3.3 applied with z and κ as in the statement of Theorem 3.3.1, and with $D = z^u$. Then

$$\begin{aligned} S(\mathcal{A}, z) = \sum_{\substack{a \in \mathcal{A} \\ (a, P(z)) = 1}} 1 &\leq \sum_{a \in \mathcal{A}} \sum_{d | (a, P(z))} \mu^+(d) = \sum_{d | P(z)} \mu^+(d) |\mathcal{A}_d| \\ &= X \sum_{d | P(z)} \mu^+(d)g(d) + \sum_{d | P(z)} \mu^+(d) r_d \\ &= X \cdot V(z)\left\{1 + O_{\kappa,K}\left(e^{-u\log u + O_{\kappa,K}(u)}\right)\right\} + O\left(\sum_{d | P(z),\, d \leq z^u} |r_d|\right), \end{aligned}$$

by part (4) of Theorem 3.3.2. Similarly, we find that

$$\begin{aligned} S(\mathcal{A},z) &\geq \sum_{d|P(z)} \mu^-(d)|\mathcal{A}_d| = X\sum_{d|P(z)} \mu^-(d)g(d) + \sum_{d|P(z)} \mu^-(d)r_d \\ &= X\cdot V(z)\left\{1+O_{\kappa,K}\left(e^{-u\log u+O_{\kappa,K}(u)}\right)\right\} + O\left(\sum_{d|P(z),\, d\leq z^u} |r_d|\right), \end{aligned}$$

which completes the deduction of Theorem 3.3.1. □

Proof of Theorem 3.3.3. For the most part, we follow an argument in [FI10]. We define $\mu^{\pm}$ as in (3.2.3) and (3.2.4) with $\beta=\beta_\kappa$. Then property (1) is satisfied by the definition of $\mu^{\pm}$ and Exercise 3.2.2 implies that properties (2) and (3) holds as well. It remains to show that properties (4) and (5) are also satisfied.

First, we show property (4). Consider $g:\mathbb{N}\to[0,1]$ that satisfies (A4a). With a slight abuse of notation, given a sequence of primes $p_1,p_2,\dots$, we write $y_j=(D/(p_1\cdots p_j))^{1/\beta}$. Also, we define

$$V_r = \sum_{\substack{z>p_1>\cdots>p_r\geq y_r\\ p_j<y_j,\, 1\leq j<r,\, j\equiv r \,(\mathrm{mod}\, 2)}} g(p_1\cdots p_r)V(p_r).$$

Then we have that

$$\sum_{d|P(z)} \mu^+(d)g(d) - V(z) = \sum_{\substack{r\geq 1\\ r\,\mathrm{odd}}} V_r \tag{3.3.2}$$

and

$$V(z) - \sum_{d|P(z)} \mu^-(d)g(d) = \sum_{\substack{r\geq 1\\ r\,\mathrm{even}}} V_r. \tag{3.3.3}$$

We fix an integer $r\geq 1$ and proceed to the estimation of V_r. This will be done by studying the complicated range of summation in V_r and replacing it by a much simpler one. In particular, we will show that the primes p_i are bounded from below by certain appropriate powers of z. Consider $p_1,\dots,p_r$ lying in the range of V_r. We claim that

$$p_1\cdots p_j \leq Dz^{-(u-1)\left(\frac{\beta-1}{\beta+1}\right)\left\lfloor\frac{j}{2}\right\rfloor} \quad (0\leq j\leq r,\, j\equiv r-1\,(\mathrm{mod}\,2)). \tag{3.3.4}$$

We argue by induction: if $j=0$ or $j=1$, then relation (3.3.4) holds trivially, since $p_1<z\leq D$. Assume now that (3.3.4) holds for some $j\in\{0,\dots,r-3\}$ that has opposite parity than r. Then

$$\begin{aligned} p_1\cdots p_{j+2} &< p_1\cdots p_j p_{j+1}^2 < p_1\cdots p_j\left(\frac{D}{p_1\cdots p_j}\right)^{\frac{2}{\beta+1}} \\ &= (p_1\cdots p_j)^{\frac{\beta-1}{\beta+1}} D^{\frac{2}{\beta+1}} \leq \left(Dz^{-(u-1)\left(\frac{\beta-1}{\beta+1}\right)\left\lfloor\frac{j}{2}\right\rfloor}\right)^{\frac{\beta-1}{\beta+1}} D^{\frac{2}{\beta+1}} \\ &= Dz^{-(u-1)\left(\frac{\beta-1}{\beta+1}\right)\left\lfloor\frac{j+2}{2}\right\rfloor}, \end{aligned}$$

which completes the inductive step and hence the proof of (3.3.4). Note that relation (3.3.4) and our assumption that $p_r \geq y_r$ imply that

$$p_r \geq \left(\frac{D}{p_1 \cdots p_{r-1}}\right)^{\frac{1}{\beta+1}} \geq z^{\delta_r}, \tag{3.3.5}$$

where

$$\delta_r = \frac{u-1}{\beta+1}\left(\frac{\beta-1}{\beta+1}\right)^{\lfloor \frac{r-1}{2} \rfloor} \geq \frac{1}{\beta+1}\left(\frac{\beta-1}{\beta+1}\right)^{\frac{r}{2}}. \tag{3.3.6}$$

Consequently, for every $\epsilon > 0$, we have that

$$\begin{aligned} V_r &\leq \sum_{\substack{z>p_1>\cdots>p_r\geq z^{\delta_r} \\ p_1\cdots p_{r-1}p_r^{\beta+1}>D}} g(p_1)\cdots g(p_r)V(z^{\delta_r}) \\ &\leq \frac{V(z^{\delta_r})}{D^\epsilon} \sum_{z>p_1>\cdots>p_r\geq z^{\delta_r}} g(p_1)\cdots g(p_r)(p_1\cdots p_{r-1}p_r^{\beta+1})^\epsilon \\ &\leq \frac{V(z^{\delta_r})}{D^\epsilon (r-1)!} \sum_{z>p_r\geq z^{\delta_r}} g(p_r)p_r^{\epsilon(\beta+1)} \sum_{z>p_1,\ldots,p_{r-1}>z^{\delta_r}} g(p_1)\cdots g(p_{r-1})(p_1\cdots p_{r-1})^\epsilon \\ &\leq \frac{KV(z)}{\delta_r^\kappa D^\epsilon (r-1)!} \sum_{z^{\delta_r}\leq p<z} g(p)p^{\epsilon(\beta+1)} \left(\sum_{z^{\delta_r}\leq p<z} g(p)p^\epsilon\right)^{r-1}. \end{aligned} \tag{3.3.7}$$

by relation (A4a) and Rankin's trick (see Section 0.3). Moreover, if λ is some number > 1 and $\epsilon \gg 1/\log z$, then we have that

$$\begin{aligned} \sum_{z^{\delta_r}\leq p<z} g(p)p^\epsilon &\leq \lambda \sum_{z^{\delta_r}\leq p<\lambda^{1/\epsilon}} g(p) + \sum_{1\leq t\leq \frac{\epsilon\log z}{\log\lambda}} \lambda^{t+1} \sum_{\lambda^{t/\epsilon}\leq p<\lambda^{(t+1)/\epsilon}} g(p) \\ &\leq \lambda \log\left(K\left(\frac{\log\lambda}{\epsilon\delta_r\log z}\right)^\kappa\right) + \sum_{1\leq t\leq \epsilon\log z} \lambda^{t+1}\log\left(K\left(\frac{t+1}{t}\right)^\kappa\right). \\ &\leq \lambda\log\left(K\left(\frac{1}{\delta_r}\right)^\kappa\right) + O_{K,\kappa,\lambda}(z^\epsilon) \leq \lambda\cdot\frac{\kappa r}{2}\log\left(\frac{\beta+1}{\beta-1}\right) + O_{K,\kappa,\lambda}(z^\epsilon), \end{aligned}$$

by (A4a). The inequality $x+y \leq xe^{y/x}$, for x and y positive, and relation (3.3.6) then imply that

$$\begin{aligned} \frac{1}{\delta_r^\kappa}\left(\sum_{z^{\delta_r}\leq p<z} g(p)p^\epsilon\right)^{r-1} &\leq (\beta+1)^\kappa \left(\frac{\beta+1}{\beta-1}\right)^{\frac{\kappa r}{2}} \left(\lambda\cdot\frac{\kappa r}{2}\log\left(\frac{\beta+1}{\beta-1}\right)\right)^{r-1} e^{O_{K,\kappa,\lambda}(z^\epsilon)} \\ &\leq \left(\frac{\lambda\rho r}{e}\right)^{r-1} e^{O_{K,\kappa,\lambda}(z^\epsilon)}, \end{aligned} \tag{3.3.8}$$

where

$$\rho = \left(\frac{\beta+1}{\beta-1}\right)^{\frac{\kappa}{2}} \frac{\kappa e}{2} \log\left(\frac{\beta+1}{\beta-1}\right) = \frac{e^{5/4}}{4} < 0.873, \tag{3.3.9}$$

by the choice of $\beta = \beta_\kappa$. Similarly, we find that

$$\sum_{z^{\delta_r} \le p < z} g(p) p^{\epsilon(\beta+1)} \le \lambda \cdot \frac{\kappa r}{2} \log\left(\frac{\beta+1}{\beta-1}\right) + O_{K,\kappa,\lambda}(z^{\epsilon(\beta+1)}) \ll_{\kappa,\lambda,K} r \cdot z^{\epsilon(\beta+1)}, \tag{3.3.10}$$

Inserting (3.3.8) and (3.3.10) into (3.3.7) and choosing $\lambda > 1$ so that $\lambda\rho \le 0.9$, we deduce that

$$V_r \le \frac{e^{O_{K,\kappa}(z^\epsilon)}}{D^\epsilon} \left(\frac{0.9}{e}\right)^{r-1} \frac{r^{r+1}}{r!}$$

Since $r!/r > r^{r-1}/e^r$, we conclude that

$$V_r \le \frac{e^{O_{K,\kappa}(z^\epsilon)}}{D^\epsilon} r^2 (0.9)^r.$$

Summing the above inequality over $r \ge 1$, we arrive to the estimate

$$\sum_{r \ge 1} V_r \le \frac{e^{O_{\kappa,K}(z^\epsilon)}}{D^\epsilon}.$$

Selecting $\epsilon = (\log u)/(\log z)$ completes the proof of property (4).

Finally, we show that property (5) holds. Assume that $u \ge \beta_k = \beta$ and consider $g : \mathbb{N} \to [0,1]$ that satisfies (A4b). The argument is similar with the previous one, but this time we avoid the use of Rankin's trick: We have that

$$\begin{aligned} V_r &\le \sum_{z > p_1 > \cdots > p_r \ge z^{\delta_r}} g(p_1)\cdots g(p_r) V(z^{\delta_r}) \le \frac{V(z^{\delta_r})}{r!} \sum_{z > p_1, \ldots, p_r \ge z^{\delta_r}} g(p_1)\cdots g(p_r) \\ &= \frac{V(z^{\delta_r})}{r!} \left(\sum_{z^{\delta_r} \le p < z} g(p)\right)^r. \end{aligned} \tag{3.3.11}$$

Now, let $\eta \in (0,1]$ be such that

$$\left(\frac{\beta+1}{\beta-1}\right)^{\frac{\kappa+\eta}{2}} \frac{(\kappa+\eta)e}{2} \log\left(\frac{\beta+1}{\beta-1}\right) \le 0.873$$

and

$$\left(\frac{\beta+1}{\beta-1}\right)^{\frac{\kappa+\eta}{2}} = \left(1 + \frac{2}{\beta-1}\right)^{\frac{\kappa+\eta}{2}} \le \left(1 + \frac{2}{4\kappa}\right)^{\frac{\kappa}{2}} < e^{\frac{1}{4}},$$

which is possible since $\beta = \beta_\kappa > 1 + 4\kappa$ and $\rho = e^{5/4}/4 < 0.873$, by relation (3.3.9). Note that for $3/2 \le w < z$ we have that

$$\begin{aligned}\frac{V(w)}{V(z)} \le \left(1 + \frac{C}{\log w}\right)\left(\frac{\log z}{\log w}\right)^\kappa &\le \left(1 + \frac{C}{(\log w)^{1-\eta}(\log z)^\eta}\right)\left(\frac{\log z}{\log w}\right)^{\kappa+\eta} \\ &\le 1.001\left(\frac{\log z}{\log w}\right)^{\kappa+\eta},\end{aligned}$$

provided that z is large enough. So

$$\sum_{z^{\delta_r} \le p < z} g(p) \le \log \frac{V(z^{\delta_r})}{V(z)} \le \log\left(\frac{1.001}{\delta_r^{\kappa+\eta}}\right).$$

Inserting the above inequalities into (3.3.11), we find that

$$V_r \le \frac{1.001 V(z)}{\delta_r^\kappa r!} \log^r\left(\frac{1.001}{\delta_r^{\kappa+\eta}}\right).$$

Furthermore, observe that, for $r \ge 2$ even, we have that

$$\delta_r = \frac{u-1}{\beta+1}\left(\frac{\beta-1}{\beta+1}\right)^{\frac{r-2}{2}} = \frac{u-1}{\beta-1}\left(\frac{\beta-1}{\beta+1}\right)^{\frac{r}{2}} \ge \left(\frac{\beta-1}{\beta+1}\right)^{\frac{r}{2}}$$

by our assumption that $u \ge \beta = \beta_\kappa$. Moreover, we have that $r! \ge r^r/(2e^{r-2})$, which can be proven inductively for $r \ge 2$. So we conclude that, for an even $r \ge 2$, we have that

$$\begin{aligned}\frac{V_r}{V(z)} &\le \frac{1.001}{r^r/(2e^{r-2})}\left(\frac{\beta+1}{\beta-1}\right)^{\frac{r(\kappa+\eta)}{2}}\left(\frac{r(\kappa+\eta)}{2}\log\left(\frac{\beta+1}{\beta-1}\right) + \log(1.001)\right)^r \\ &\le \frac{1.001}{2e^2}\left\{\left(\frac{\beta+1}{\beta-1}\right)^{\frac{\kappa+\eta}{2}}\frac{e(\kappa+\eta)}{2}\log\left(\frac{\beta+1}{\beta-1}\right) + \left(\frac{\beta+1}{\beta-1}\right)^{\frac{\kappa+\eta}{2}}\frac{e\log(1.001)}{r}\right\}^r \\ &\le \frac{1.001}{2e^2}\left(0.873 + e^{\frac{1}{4}}\cdot\frac{e\log(1.001)}{r}\right)^r = \frac{1.001(0.873)^r}{2e^2}\left(1 + \frac{e^{\frac{5}{4}}\log(1.001)}{0.873\, r}\right)^r \\ &\le \frac{1.001(0.873)^r}{2e^2}\exp\{e^{5/4}\log(1.001)/0.873\} \le 0.272\,(0.873)^r,\end{aligned}$$

by the choice of η. Consequently,

$$\frac{1}{V(z)}\sum_{\substack{r\ge 1\\ r\,\text{even}}} V_r \le 0.272 \sum_{\substack{r\ge 1\\ r\,\text{even}}} 0.873^r = 0.272 \cdot \frac{0.873^2}{1 - 0.873^2} < \frac{7}{8}.$$

Inserting the above inequality into (3.3.3) completes the proof of property (5) and hence of Theorem 3.3.3. □

Chapter 4

Some applications of sieve methods

In this chapter, we give some applications of the results we proved in the previous sections, particularly of the fundamental lemma of sieve methods. In order to show that (R) holds for certain sequences $\mathcal{A}$, we need to control the number of primes in arithmetic progressions *on average*. If we let

$$\mathrm{li}(x) := \int_2^x \frac{dt}{\log t},$$

then we expect that $\pi(x; q, a)$, the number of primes up to x in the arithmetic progression $a \pmod q$, should be very well approximated by $\mathrm{li}(x)/\varphi(q)$. The prime number theorem for arithmetic progressions implies that this is true if x is large enough in terms of q, specifically, when $x > e^{q^{\epsilon/3}}$. Bombier and Vinogradov showed (independently) that this remains true for much smaller x as well (for x as small as $q^{2+\epsilon}$), but on an average sense. Indeed, if we set

$$E(x; q) = \max_{y \le x} \max_{(a,q)=1} \left| \pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} \right|, \tag{4.0.1}$$

then their result is formulated as follows:

Theorem 4.0.4 (Bombieri-Vinogradov)**.** *Let $A > 0$. There is some constant $B = B(A)$ such that*

$$\sum_{q \le x^{1/2}/(\log x)^B} E(x; q) \ll_A \frac{x}{(\log x)^A} \quad (x \ge 2).$$

We will prove this theorem in Chapter 10. Note that the 'trivial' bound on the sum in question, which comes from the Brun-Titchmarsch inequality (see Theorem 4.1.4 below), is

$$\sum_{q \le x^{1/2}/(\log x)^B} \max_{y \le x} \max_{(a,q)=1} \left| \pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} \right| \ll \sum_{q \le x^{1/2}/(\log x)^B} \frac{x}{\varphi(q) \log x} \ll x \quad (x \ge 2).$$

So the Bombieri-Vinogradov theorem allows us to save an arbitrary power of log over this trivial estimate. Moreover, Theorem 4.0.4 is essentially of the same strength with the Generalized Riemann Hypothesis (GRH) on average. Indeed, GRH implies the bound

$$\sum_{q \le x^{1/2}/(\log x)^B} \max_{y \le x} \max_{(a,q)=1} \left| \pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} \right| \ll \sum_{q \le x^{1/2}/(\log x)^B} \sqrt{x} \log x \le \frac{x}{(\log x)^{B-1}} \quad (x \ge 2).$$

So taking $B = A + 1$, we deduce the Bombieri-Vinogradov theorem.

Elliott and Halberstam have conjectured that it is possible to improve significantly upon the Bombieri-Vinogradov theorem:

Conjecture 4.0.5 (Elliott-Halberstam)**.** *Fix $A > 0$ and $\epsilon > 0$. Then*

$$\sum_{q \le x^{1-\epsilon}} E(x;q) \ll_{\epsilon,A} \frac{x}{(\log x)^A} \quad (x \ge 2).$$

This conjecture is, in a certain sense, stronger than GRH.

Remark 4.0.6. As Friedlander and Granville [FG] have shown, if we replace $x^{1-\epsilon}$ with $x/(\log x)^B$ in the above conjecture, then the conclusion is not always correct.

Exercise 4.0.7. Show that the Bombieri-Vinogradov theorem is equivalent to the following statement: "For every $A > 0$, there is some $C > 0$, such that if $1 \le Q \le \sqrt{x}/(\log x)^C$, then

$$\# \left\{ q \le Q : E(x;q) \ge \frac{x}{\varphi(q) \log x} \cdot \frac{1}{(\log x)^A} \right\} \ll_A \frac{Q}{(\log x)^A}."$$

Exercise 4.0.8. Let $k \in \mathbb{N}$ and $A > 0$. Using the Bombieri-Vinogradov theorem, prove that there is $B = B(A,k)$ such that

$$\sum_{q \le x^{1/2}/(\log x)^B} \tau_k(q) E(x;q) \ll_{k,A} \frac{x}{(\log x)^A} \quad (x \ge 2).$$

4.1 Prime values of polynomials

We start with an application of the sieve to the twin prime problem.

Theorem 4.1.1 (Twin primes)**.** *For $x \ge 3$, we have that*

$$\#\{p \le x : p + 2 \text{ prime}\} \ll \frac{x}{(\log x)^2}$$

and

$$\#\{p \le x : \Omega(p+2) \le 8\} \gg \frac{x}{(\log x)^2} \quad (x \ge 3).$$

Proof. Let $\mathcal{A} = \{p + 2 : p \le x\}$ and note that

$$|\mathcal{A}_d| = \pi(x; d, -2) = 1$$

whenever $2|d$, and

$$|\mathcal{A}_d| = \pi(x; d, -2) = \frac{\mathrm{li}(x)}{\varphi(d)} + \left(\pi(x; d, -2) - \frac{\mathrm{li}(x)}{\varphi(d)} \right),$$

if $(d,2)=1$. So (A1) holds with $X=\mathrm{li}(x)$,

$$g(d)=\begin{cases}\dfrac{1}{\varphi(d)} & \text{if } (d,2)=1,\\ 0 & \text{otherwise,}\end{cases}$$

and

$$r_d=\begin{cases}\pi(x;d,-2)-\dfrac{\mathrm{li}(x)}{\varphi(d)} & \text{if } (d,2)=1,\\ 1 & \text{if } (d,2)>1.\end{cases}$$

In addition, note that

$$\begin{aligned}\prod_{w\le p<w'}\left(1-\frac{1}{\varphi(p)}\right)^{-1}&=\prod_{\max\{3,w\}\le p<w'}\left(1-\frac{1}{p-1}\right)^{-1}\\&=\prod_{\max\{3,w\}\le p<w'}\left(1-\frac{1}{p}\right)^{-1}\left(1+\frac{1}{p(p-2)}\right)\\&=\left\{1+O\left(\frac{1}{\log w}\right)\right\}\frac{\log w'}{\log w}\cdot\left\{1+O\left(\frac{1}{w}\right)\right\}\\&=\left\{1+O\left(\frac{1}{\log w}\right)\right\}\frac{\log w'}{\log w}\quad(3/2\le w\le w'),\end{aligned}\tag{4.1.1}$$

by Mertens' estimate (1.1.5), that is to say, relation (A4b) holds with $\kappa=1$. Lastly, relation (R) holds with $A=2$ and $D=x^\theta$, for any $\theta\in(0,1/2)$, by the Bombieri-Vinogradov Theorem (see Theorem 4.0.4). So Theorem 3.3.1 and relation (4.1.1) imply that

$$\#\{p\le x: p+2\,\text{prime}\}\le\sqrt{x}+S(\mathcal{A},\sqrt{x})\ll\sqrt{x}+\mathrm{li}(x)\prod_{p<\sqrt{x}}\left(1-\frac{1}{\varphi(p)}\right)\ll\frac{x}{(\log x)^2},$$

which completes the proof of the first part of the theorem.

For the second part, note that we may assume that x is large enough, since 3 and 5 are twin primes. Now, applying the second part of Corollary 3.3.2 with $\theta=10/21$ and $z=x^{1/8.6}<x^{10/(21\beta_1)}$ yields

$$\begin{aligned}\#\{p\le x:\Omega(p+2)\le 8\}\ge S(\mathcal{A},x^{1/8.6})&\ge\frac{\mathrm{li}(x)}{10}\prod_{p<x^{1/8.6}}\left(1-\frac{1}{\varphi(p)}\right)\\&\gg\frac{x}{\log x}\cdot\frac{1}{\log x}=\frac{x}{(\log x)^2}.\end{aligned}$$

This completes the proof of the theorem. □

Exercise 4.1.2.

(a) Show that

$$\#\{n\le x:(n,P(z))=1\}\ll\frac{x}{\log z}\quad(3/2\le z\le x)$$

and

$$\#\{n \le x : (n, P(z)) = 1\} \gg \frac{x}{\log z} \quad (3 \le 2z \le x).$$

(b)* Show the stronger estimates

$$\sum_{\substack{n \le x \\ P^-(n) \ge z}} \frac{n}{\varphi(n)} \ll \frac{x}{\log z} \quad (3/2 \le z \le x)$$

and

$$\sum_{\substack{n \le x \\ P^-(n) \ge z}} \frac{\mu^2(n)\varphi(n)}{n} \gg \frac{x}{\log z} \quad (3 \le 2z \le x).$$

Exercise 4.1.3 (Goldbach's problem)**.** Let $N \ge 2$ be an integer. Show that

$$\#\{(p, q) : p, q \text{ primes}, p + q = 2N\} \ll \frac{N}{\varphi(N)} \cdot \frac{N}{(\log N)^2}.$$

Moreover, if N is large, then prove that

$$\#\{p \le 2N : \Omega(2N - p) \le 8\} \gg \frac{N}{\varphi(N)} \cdot \frac{N}{(\log N)^2}.$$

The following result is an extremely useful inequality due to its uniformity and large range of applicability. In particular, it can give non-trivial results for the number of primes in short intervals that are, in certain cases, stronger than GRH (when $y/q < \sqrt{x}$).

Theorem 4.1.4 (Brun-Titchmarsch inequality)**.** *For $1 \le q \le y \le x$ and $(a, q) = 1$, we have that*

$$\pi(x; q, a) - \pi(x - y; q, a) \ll \frac{y}{\varphi(q) \log(2y/q)}.$$

Proof. If $y/2 < q \le y$, then the theorem follows by the trivial inequality

$$\pi(x; q, a) - \pi(x - y; q, a) \le \#\{x - y < n \le x\} \le 1 + \frac{y}{q}.$$

Assume now that $q \le y/2$ and let $\mathcal{A} = \{x - y < n \le x : n \equiv a \pmod q\}$. Fix for the moment an integer d. Note that if $(d, q) > 1$, then there are no solutions to the congruence $dm \equiv a \pmod q$, since $(a, q) = 1$ by assumption. Therefore $|\mathcal{A}_d| = 0$ in this case. On the other hand, if $(d, q) = 1$ and we let $\overline{d} \in [1, q]$ be the multiplicative inverse of d modulo q, then

$$\begin{aligned} |\mathcal{A}_d| = \#\left\{\frac{x-y}{d} < m \le \frac{x}{d} : dm \equiv a \pmod q\right\} &= \#\left\{\frac{x-y}{d} < m \le \frac{x}{d} : m \equiv \overline{d}a \pmod q\right\} \\ &= \#\left\{k \in \mathbb{Z} : \frac{x-y}{d} < kq + \overline{d}a \le \frac{x}{d}\right\} \\ &= \frac{y}{dq} + O(1). \end{aligned}$$

In any case, we find that (A1) is satisfied with $X = y/q$,

$$g(d) = \begin{cases} 0 & \text{if } (d,q) > 1, \\ 1/d & \text{if } (d,q) = 1. \end{cases}$$

which is easily seen to be multiplicative, and $r_d \ll 1$. Next, observe that

$$\prod_{w \le p < w'} (1 - g(p))^{-1} \le \prod_{w' \le p < w'} \left(1 - \frac{1}{p}\right)^{-1} \ll \frac{\log w'}{\log w},$$

that is to say, (A4a) holds with $\kappa = 1$. Finally, relation (R) holds trivially with $A = 2$ and $D = (y/q)^\theta$, for any $\theta < 1$. Therefore applying Corollary 3.3.3 with $z = \sqrt{y/q}$ and $u = 1$ implies that

$$\begin{aligned} \pi(x;q,a) - \pi(x-y;q,a) &\le \sqrt{\frac{y}{q}} + S(\mathcal{A}, \sqrt{y/q}) \ll \sqrt{\frac{y}{q}} + \frac{y}{q} \prod_{p < \sqrt{y/q}} (1 - g(p)) \\ &= \sqrt{\frac{y}{q}} + \frac{y}{q} \prod_{\substack{p < \sqrt{y/q} \\ p \nmid q}} \left(1 - \frac{1}{p}\right) \\ &\ll \sqrt{\frac{y}{q}} + \frac{y}{q \log(y/q)} \prod_{\substack{p < \sqrt{y/q} \\ p \mid q}} \left(1 + \frac{1}{p}\right) \ll \frac{y}{\varphi(q) \log(y/q)}, \end{aligned}$$

which completes the proof of the theorem. □

Theorem 4.1.5. *Let $F_1(x), \dots, F_r(x)$ be distinct irreducible polynomials over $\mathbb{Z}[x]$ with positive leading coefficient. Suppose that the polynomial $F = F_1 \cdots F_r$ has no fixed prime divisors, i.e. there is no prime p such that $p|F(n)$ for all integers n. Then we have that*

$$\#\{n \le x : F_1(n), \dots, F_r(n) \text{ are all primes}\} \ll_F \frac{x}{(\log x)^r}.$$

Moreover, if x is large enough and d denotes the degree of F, then

$$\#\{n \le x : \Omega(F(n)) \le 4rd\} \gg_F \frac{x}{(\log x)^r}.$$

Proof. Exercise.

Hint: If G is an irreducible polynomial over $\mathbb{Z}[x]$ and

$$\nu_G(d) = \#\{m \in \mathbb{Z}/d\mathbb{Z} : G(m) \equiv 0 \,(\mathrm{mod}\, d)\},$$

then there is a constant c_G such that

$$\sum_{p \le x} \frac{\nu_G(p)}{p} = \log\log x + c_G + O_G\left(\frac{1}{\log x}\right) \quad (x \ge 2).$$

□

4.2 The image of Euler's totient function

If p is a prime number, then we expect that $p-1$ behaves like a 'typical' integer, except for obvious restrictions such as that $2|p-1$ for all but one primes. For example, we have the following result:

Theorem 4.2.1. *There are absolute constants $A' > 0$ and $B' > 0$ such that, for every $x \geq 3$ and every integer $r \geq 1$, we have that*

$$\#\left\{3 < p \leq x : \omega\left(\frac{p-1}{2}\right) = r\right\} \leq \frac{A'x}{\log^2 x}\frac{(\log\log x + B')^{r-1}}{(r-1)!}$$

Proof. Exercise.

Hint: Show the stronger statement that

$$\begin{aligned} S_r(x,k) &:= \#\left\{p \leq x : p \equiv 1 \,(\mathrm{mod}\, 2k),\, \omega\left(\frac{p-1}{2k}\right) = r\right\} \\ &\leq \frac{A'x}{\varphi(k)\log^2(x/d)}\frac{(\log\log(x/k) + B')^{r-1}}{(r-1)!}, \end{aligned} \tag{4.2.1}$$

uniformly in $r \geq 1$ and $1 \leq k \leq x/2$.

□

Now, note that $\varphi(n) = \prod_{p^a \| n} p^{a-1}(p-1)$, so we expect $\varphi(n)$ to have many prime factors, many more than a typical integer of the same size. Our goal in this section is how fast the image of φ grows. So set $V(x) = \#\{\varphi(n) \leq x\}$. Clearly, $V(x) \geq \pi(x+1) \gg x/\log x$ for $x \geq 2$. Erdős showed that this lower bound is not so far from the truth:

Theorem 4.2.2 (Erdős, 1935)**.** *Let $\epsilon > 0$. For $x \geq 2$, we have that*

$$V(x) \ll_\epsilon \frac{x}{(\log x)^{1-\epsilon}}.$$

Proof. The idea of the proof is the following: we have that $\varphi(n) = \prod_{p^a \| n} p^{a-1}(p-1)$. Moreover, a 'typical' integer $n \leq x$ has about $\log\log x$ prime factors p by Theorem 2.1.2 and, for each such p, the number $p-1$ should have about $\log\log x$ prime factors by Theorem 4.2.1. So $\varphi(n)$ should have abnormally many prime factors.

Without loss of generality, we may assume that x is large enough. If $m = \varphi(n) \leq x$, then $n \leq cx\log\log x =: x_1$. Set $R = (\log\log x)/N$ for some integer $N \geq 2$, which is fixed for the moment, and $\mathbb{P} = \{p \text{ prime} : \omega(p-1) \leq 100N\}$. We expect that most m counted by $V(x)$ are images of integers n with $\omega(n) \leq R$. We bound the exceptional set by observing that

$$V(x) \leq V_1 + V_2 + V_3 + V_4, \tag{4.2.2}$$

where

$$\begin{aligned} V_1 &= \#\{n \leq x_1 : \omega(n) \leq R\} \\ V_2 &= \#\{n \leq x_1 : \omega(n) > R, \text{ there are } > R/2 \text{ primes in } \mathbb{P} \text{ that divide } n\} \\ V_3 &= \#\{m \leq x : \omega(m) > 10\log\log x\}, \\ V_4 &= \#\{m \leq x : \text{there is } d^2|m \text{ with } d > (\log x)^{10}\}. \end{aligned}$$

Indeed, if $m=\varphi(n)\le x$ is counted by $V(x)$ but not by V_1, V_2 or V_3, then $\omega(m)\le 10\log\log x$ and there are $> R/2$ primes in $\mathbb{P}^c$ that divide n. So, if we write $m=\prod_{i=1}^r p_i^{a_i}=\prod_{p^b\| n} p^{b-1}(p-1)$, where $p_1,\dots,p_r$ are the distinct prime factors of m, then

$$\Omega(m)-\omega(m)\ge \sum_{p|n}\omega(p-1)-10\log\log x\ge \frac{R}{2}\cdot 100N-10\log\log x=40\log\log x.$$

Consequently,

$$\prod_{i=1}^r p_i^{\lfloor a_i/2\rfloor}\ge \prod_{i=1}^r p_i^{\frac{a_i-1}{2}}\ge 2^{\sum_{i=1}^r\frac{a_i-1}{2}}=2^{\frac{\Omega(m)-\omega(m)}{2}}\ge 2^{20\log\log x}>(\log x)^{10},$$

that is to say, $\varphi(m)$ is divisible by a square $d>\log x$ and thus it is counted by V_4.

As we mentioned above, we expect that in the right hand side of (4.2.2) the main contribution comes from V_1, whereas V_2, V_3 and V_4 are error terms. We start by estimating V_4 which is the easiest. We have that

$$V_4\le \sum_{d>(\log x)^{10}}\#\{m\le x: d^2|m\}\le \sum_{d>(\log x)^{10}}\frac{x}{d^2}\ll \frac{x}{(\log x)^{10}},$$

which is admissible. Next, Theorem 2.1.2 and Stirling's formula imply that

$$\begin{aligned}V_3\le \sum_{r>10\log\log x}\pi_r(x)&\le \sum_{r>10\log\log x}\frac{Ax}{\log x}\frac{(\log\log x+B)^{r-1}}{(r-1)!}\\&\ll \frac{x}{\log x}\frac{(\log\log x+B)^{\lfloor 10\log\log x\rfloor}}{(\lfloor 10\log\log x\rfloor)!}\\&\asymp \frac{x}{\log x}\frac{(\log\log)^{10\log\log x}}{(10(\log\log x)/e)^{10\log\log x}\sqrt{\log\log x}}\\&\le \frac{x}{(\log x)^{10\log 10-9}}\le \frac{x}{(\log x)^{14}},\end{aligned}$$

which is also admissible. Similarly, we have that

$$\begin{aligned}V_1\le \sum_{r\le R}\pi_r(x_1)\le \sum_{r\le R}\frac{Ax}{\log x}\cdot\frac{(\log\log x+B)^{r-1}}{(r-1)!}&\ll \frac{x}{\log x}\frac{(\log\log x+B)^{\lfloor R\rfloor-1}}{(\lfloor R\rfloor-1)!}\\&\asymp \frac{x_1}{\log x_1}\frac{(\log\log x+B)^{\frac{\log\log x}{N}}}{((\log\log x)/(eN))!\sqrt{R}}\\&\le \frac{x_1}{(\log x)^{\frac{-1+\log N}{N}+1}},\end{aligned}$$

which is admissible, provided that N is large enough. Finally, we estimate V_2: note that if n is counted by V_2, then it is possible to write $n=ab$, where a has exactly $S=\lfloor R/2\rfloor$ prime factors all of which are in $\mathbb{P}$. Call $\mathcal{A}$ the set of such numbers a. Then

$$V_2\le \sum_{a\in\mathcal{A}}\#\{n\le x_1: a|n\}\le \sum_{a\in\mathcal{A}}\frac{x}{a}=x\sum_{\substack{p_1<\dots<p_S\\ p_i\in\mathbb{P}}}\frac{1}{p_1\cdots p_S}\le \frac{x}{S!}\left(\sum_{p\in\mathbb{P}}\frac{1}{p}\right)^S.$$

Now, Theorem 4.2.1 implies that

$$|\mathbb{P}\cap[1,u]| \le \sum_{r\le 100N} S_r(u,1) \le 1+\sum_{1\le r\le 100N-1}\frac{Cx}{(\log x)^2}\frac{(\log\log x+D)^{r-1}}{(r-1)!}$$
$$\le 1+\frac{100NCx}{(\log x)^2}(\log\log x+D)^{100N-2} \ll_N \frac{x}{(\log x)^{3/2}}$$

and, consequently, partial summation yields that

$$\sum_{p\in\mathbb{P}}\frac{1}{p} \ll_N 1.$$

So we deduce that

$$V_2 \le \frac{xe^{O_N(S)}}{S!} \asymp \frac{xe^{O_N(S)}}{(S/e)^S\sqrt{S}} \ll_N \frac{x}{(\log x)^{\frac{\log\log\log x}{2N}+O_N(1)}} \ll_N \frac{x}{(\log x)^2},$$

which is admissible. This completes the proof of the theorem. □

4.3 The Titchmarsch-Linnik divisor problem

Given $s\in\mathbb{Z}$ and $k\in\mathbb{N}$, what is the asymptotic behavior of the sum $\sum_{p\le x}\tau_k(p+s)$, as $x\to\infty$? As we saw before shifted primes of the form $p-1$ behave like typical integers. The same is also true for the shifted primes $p+s$. So we should expect that

$$\sum_{p\le x}\tau_k(p+s) \approx \mathbf{Prob}(n \text{ prime} \,|\, n\le x)\sum_{n\le x}\tau_k(n) \asymp \frac{1}{\log x}\cdot x(\log x)^{k-1} = x(\log x)^{k-2}.$$

Titchmarsch first studied this sum when $k=2$ and evaluated it asymptotically under the assumption of GRH. Subsequently, Linnik removed this assumption, so that the following result now holds unconditionally:

Theorem 4.3.1. *For $1\le|s|\le x$ and $x\ge 3$, we have that*

$$\sum_{p\le x}\tau(p+s) = x\prod_{p\nmid s}\left(1+\frac{1}{p(p-1)}\right)\prod_{p\mid s}\left(1-\frac{1}{p}\right)+O\left(\frac{x\log\log x}{(\log x)^2}\right).$$

Proof. We shall present a proof due to Rodriguez [Ro] based on the Bombieri-Vinogradov theorem (i.e. Theorem 4.0.4). We have that

$$T=\sum_{p\le x}\tau(p+s)=\sum_{p\le x}\sum_{ab=p+s}1=\sum_{p\le x}\left(2\sum_{\substack{a<\sqrt{p+s}\\ a\mid p+s}}1+\mathbf{1}_{p+s=\square}\right)$$
$$=2\sum_{p\le x}\sum_{a<\sqrt{p+s}}1+O(\sqrt{x})$$
$$=2\sum_{a<\sqrt{x+s}}(\pi(x,a,-s)-\pi(a^2-s,a,-s))+O(\sqrt{x}).$$

Note that if $(a,s)>1$ and $a|p+s$, then $p|s$. So

$$\sum_{\substack{a<\sqrt{x+s}\\(a,s)>1}}(\pi(x,a,-s)-\pi(a^2,a,-s))=\sum_{\substack{a<\sqrt{x+s}\\(a,s)>1}}\sum_{\substack{a^2-s<p\le x\\a|p+s}}1\le\sum_{p|s}\sum_{a|p+s}1$$
$$\ll\sum_{p|s}|s|^{1/4}\le\tau(s)|s|^{1/4}\ll H^{1/2}\le x^{1/2}.$$

So we deduce that

$$T=2\sum_{\substack{a<\sqrt{x+s}\\(a,s)=1}}(\pi(x,a,-s)-\pi(a^2-s,a,-s))+O(\sqrt{x}).$$

For $(a,s)=1$, the Brun-Titchmarsch inequality (Theorem 4.1.4) implies that

$$\pi(a^2-s,a,-s)=\pi(a^2-s,a,-s)-\pi(-s,a,-s)\ll\frac{a^2}{\varphi(a)\log(2a)}.$$

So

$$\sum_{\substack{a<\sqrt{x+s}\\(a,s)=1}}\pi(a^2-s,a,-s)\ll\sum_{a<2\sqrt{x}}\frac{a^2}{\varphi(a)\log(2a)}\ll\frac{x}{\log x},$$

by Theorem 0.2.1. Hence

$$T=2\sum_{\substack{a<\sqrt{x+s}\\(a,s)=1}}\pi(x,a,-s)+O\left(\frac{x}{\log x}\right).$$

Next, the Bombieri-Vinogradov Theorem with $A=1$ implies that there is some absolute constant $B>0$ such that

$$\sum_{p\le Q}\left|\pi(x,a,-s)-\frac{\mathrm{li}(x)}{\varphi(a)}\right|\ll\frac{x}{\log x},$$

where $Q=x^{1/2}/(\log x)^B$. So we find that

$$T=2\sum_{\substack{a\le Q\\(a,s)=1}}\frac{\mathrm{li}(x)}{\varphi(a)}+\sum_{\substack{Q<a\le\sqrt{x+s}\\(a,s)=1}}\pi(x,a,-s)+O\left(\frac{x}{\log x}\right).\tag{4.3.1}$$

Now, Exercise 0.2.11 yields that

$$\sum_{\substack{a\le Q\\(a,s)=1}}\frac{1}{\varphi(a)}=\prod_{p\nmid s}\left(1+\frac{1}{p(p-1)}\right)\prod_{p|s}\left(1-\frac{1}{p}\right)\left\{\log Q+O\left(1+\sum_{p|s}\frac{\log p}{p-1}\right)\right\}$$
$$=\prod_{p\nmid s}\left(1+\frac{1}{p(p-1)}\right)\prod_{p|s}\left(1-\frac{1}{p}\right)\left\{\frac{\log x}{2}+O\left(\log\log x+\sum_{p|s}\frac{\log p}{p-1}\right)\right\}.$$

Moreover, we have that

$$\sum_{p|s}\frac{\log p}{p-1}\le\sum_{p\le y}\frac{\log p}{p-1}+\frac{\log y}{y}\cdot\omega(s)\le\log y+O(1)+\frac{\log y}{y-1}\cdot\frac{\log|s|}{\log 2}.$$

Selecting $y=\log|s|+1$, we deduce that

$$\sum_{p|s}\frac{\log p}{p-1}\ll\log\log(3|s|)\le\log\log(3x),$$

and consequently

$$\sum_{\substack{a\le Q\\(a,s)=1}}\frac{1}{\varphi(a)}=\prod_{p\nmid s}\left(1+\frac{1}{p(p-1)}\right)\prod_{p|s}\left(1-\frac{1}{p}\right)\left\{\frac{\log x}{2}+O(\log\log x)\right\}.$$

Inserting this estimate into (4.3.1), and noting that $\mathrm{li}(x)=x/\log x+O(x/\log^2 x)$, we deduce that

$$T=x\prod_{p\nmid s}\left(1+\frac{1}{p(p-1)}\right)\prod_{p|s}\left(1-\frac{1}{p}\right)+\sum_{\substack{Q<a\le\sqrt{x+s}\\(a,s)=1}}\pi(x,a,-s)+O\left(\frac{x\log\log x}{\log x}\right).$$

This estimate reduces the theorem to showing that

$$\sum_{\substack{Q<a\le\sqrt{x+s}\\(a,s)=1}}\pi(x,a,-s)\ll\frac{x\log\log x}{\log x}. \tag{4.3.2}$$

The reason that this sum is so small is that $\log a$ lies in a short interval: $\log a$ is of size $\log x$ and it is restricted in an interval of length $\log\log x$. To see (4.3.2), we apply the Brun-Titchmarsch inequality and Exercise 0.2.9 to get that

$$\begin{aligned}\sum_{\substack{Q<a\le\sqrt{x+s}\\(a,s)=1}}\pi(x,a,-s)&\ll\sum_{\substack{Q<a\le\sqrt{x+s}\\(a,s)=1}}\frac{x}{\varphi(a)\log(x/a)}\ll\frac{x}{\log x}\sum_{\substack{Q<a\le 2\sqrt{x}\\(a,s)=1}}\frac{1}{\varphi(a)}\\&\le\frac{x}{\log x}\sum_{Q/2\le 2^m\le 2\sqrt{x}}\frac{1}{2^m}\sum_{2^m<a\le 2^{m+1}}\frac{a}{\varphi(a)}\ll\frac{x}{\log x}\sum_{Q/2\le 2^m\le 2\sqrt{x}}\frac{1}{2^m}\cdot 2^m\\&=\frac{x}{\log x}\cdot(\log\sqrt{x}-\log Q+1)\asymp\frac{x\log\log x}{\log x},\end{aligned}$$

which completes the proof of (4.3.2) and hence of the theorem. □

Exercise 4.3.2.

(a) Show that, for $1\le|s|\le x$ and $x\ge 3$, we have that

$$\sum_{p\le x}\tau_3(p+s)\asymp x\log x\prod_{p|s}\left(1-\frac{1}{p}\right)^2$$

(b)* Let $1 \leq |s| \leq x$ and $x \geq 3$. Show that under the Elliott- Halberstam conjecture we have that

$$\sum_{p \leq x} \tau_3(p+s) = C(s)x \log x + O(x),$$

for some appropriate constant $C(s)$.

Chapter 5

Selberg's sieve

5.1 An optimization problem

The combinatorial sieve was based on an iterative procedure, which we tried to optimize at every single step. In 1947 Selberg introduced a new approach to sieving which is based on global optimization. The starting point is the following simple idea: If $\{\lambda_d\}_{d\geq 1}$ is any sequence of real numbers with $\lambda_1 = 1$, and we denote with χ_z the characteristic function of integers n such that $(n, P(z)) = 1$, then

$$\chi_z(n) \leq \left(\sum_{d|(n,P(z))} \lambda_d \right)^2.$$

Note that this immediately provides an upper bound sieve: expanding the square, we find that

$$\chi_z(n) \leq \sum_{d|(n,P(z)} \mu^+(d), \quad \text{where} \quad \mu^+(d) = \sum_{\substack{d_1,d_2\in\mathbb{N} \\ [d_1,d_2]=d}} \lambda_{d_1}\lambda_{d_2}.$$

In particular, for any finite set of integers $\mathcal{A}$ satisfying (A1), we have that

$$\begin{aligned} S(\mathcal{A}, z) &\leq \sum_{d_1,d_2|P(z)} \lambda_{d_1}\lambda_{d_2}|\mathcal{A}_{[d_1,d_2]}| \\ &= X \sum_{d_1,d_2|P(z)} \lambda_{d_1}\lambda_{d_2} g([d_1, d_2]) + \sum_{d_1,d_2|P(z)} \lambda_{d_1}\lambda_{d_2} r_{[d_1,d_2]} \\ &\leq X \sum_{d_1,d_2|P(z)} \lambda_{d_1}\lambda_{d_2} g([d_1, d_2]) + \sum_{d|P(z)} |\mu^+(d) r_{[d_1,d_2]}| \\ &=: X \cdot G + R. \end{aligned}$$

This inequality turns upper bounds for $S(\mathcal{A}, z)$ to an optimization problem: we need to choose the parameters λ_d so that $XG + R$ is minimized. It turns out that this problem is too hard. Instead we optimize only the main term XG. In order to have some control on the error term R, we impose the additional constraint that $\lambda_d = 0$ for $d > \sqrt{D}$, so that μ^+ is supported on integers $d \leq D$.

Now we turn to the problem of optimizing the value of the bilinear form

$$G = \sum_{\substack{d_i | P(z),\, d_i \le \sqrt{D} \\ i \in \{1,2\}}} \lambda_{d_1} \lambda_{d_2} g([d_1, d_2]),$$

under the restriction that $\lambda_1 = 1$. For any prime p, if $p^{\nu_1} \| d_1$ and $p^{\nu_2} \| d_2$, then $p^{\max\{\nu_1,\nu_2\}} \| [d_1, d_2]$ and $p^{\min\{\nu_1,\nu_2\}} \| (d_1, d_2)$. This implies that

$$g((d_1, d_2)) g([d_1, d_2]) = g(d_1) g(d_2). \tag{5.1.1}$$

Let $\mathbb{P} = \{p \text{ prime} : g(p) \neq 0\}$ and $\mathbb{P}_z = \mathbb{P} \cap (1, z)$. Note that in optimizing G we may assume that the λ_d's are supported in

$$\mathcal{D}_z = \{d \le \sqrt{D} : d | \prod_{p \in \mathbb{P}_z} p\}.$$

Moreover, we assume that (A2) holds. From the above discussion, we have that

$$G = \sum_{d_1, d_2 \in \mathcal{D}_z} \lambda_{d_1} \lambda_{d_2} \frac{g(d_1) g(d_2)}{g((d_1, d_2))}.$$

Next, let $h(n) = \prod_{p|n} g(p)/(1 - g(p)) > 0$ so that $1/g(n) = (1 * (1/h))(n)$, whenever n is a square-free integer for which $g(n) \neq 0$. Then

$$\begin{aligned} G &= \sum_{d_1, d_2 \in \mathcal{D}_z} \lambda_{d_1} \lambda_{d_2} g(d_1) g(d_2) \sum_{m | (d_1, d_2)} \frac{1}{h(m)} \\ &= \sum_{m \in \mathcal{D}_z} \frac{1}{h(m)} \sum_{\substack{d_i \in \mathcal{D}_z,\, m | d_i \\ i \in \{1,2\}}} \lambda_{d_1} \lambda_{d_2} g(d_1) g(d_2) \\ &= \sum_{m \in \mathcal{D}_z} \frac{1}{h(m)} \left(\sum_{\substack{d \in \mathcal{D}_z \\ d \equiv 0 \,(\mathrm{mod}\, m)}} \lambda_d g(d) \right)^2. \end{aligned}$$

We make the change of variable

$$\xi_m = \sum_{\substack{d \in \mathcal{D}_z \\ d \equiv 0 \,(\mathrm{mod}\, m)}} \lambda_d g(d) \quad (m \in \mathcal{D}_z)$$

in order to diagonalize G. Then we have that

$$\begin{aligned} \sum_{\substack{m \in \mathcal{D}_z \\ m \equiv 0 \,(\mathrm{mod}\, d)}} \xi_m \mu(m/d) &= \sum_{\substack{m \in \mathcal{D}_z \\ m \equiv 0 \,(\mathrm{mod}\, d)}} \mu(m/d) \sum_{\substack{f \in \mathcal{D}_z \\ f \equiv 0 \,(\mathrm{mod}\, m)}} \lambda_f g(f) \\ &= \sum_{\substack{f \in \mathcal{D}_z \\ f \equiv 0 \,(\mathrm{mod}\, d)}} \lambda_f g(f) \sum_{m :\, d | m | f} \mu(m/d) \\ &= \sum_{\substack{f \in \mathcal{D}_z \\ f \equiv 0 \,(\mathrm{mod}\, d)}} \lambda_f g(f) \sum_{k | f/d} \mu(k) = \lambda_d g(d). \end{aligned} \tag{5.1.2}$$

This proves that there is a one-to-one correspondence between the variables λ_d and the variables ξ_m. In particular, the constraint $\lambda_1 = 1$ becomes

$$\sum_{m\in\mathcal{D}_z} \xi_m \mu(m) = 1. \tag{5.1.3}$$

So our task now is to minimize

$$G = \sum_{m\in\mathcal{D}_z} \frac{\xi_m^2}{h(m)}$$

under condition (5.1.3). Using Lagrange multipliers, we find that this is achieved when $\xi_m = c\cdot\mu(m)h(m)$, $m\in\mathcal{D}_z$, for some constant c. Then (5.1.3) implies that

$$c\sum_{m\in\mathcal{D}_z} h(m) = 1 \qquad \Longrightarrow \qquad c = \left(\sum_{m\in\mathcal{D}_z} h(m)\right)^{-1}.$$

So, we conclude that, for any $d\in\mathcal{D}_z$,

$$\begin{aligned}\lambda_d = \frac{1}{g(d)} \sum_{\substack{m\in\mathcal{D}_z\\ m\equiv 0 \,(\mathrm{mod}\, d)}} \xi_m\mu(m/d) &= \frac{1}{g(d)}\left(\sum_{m\in\mathcal{D}_z} h(m)\right)^{-1} \sum_{\substack{m\in\mathcal{D}_z\\ m\equiv 0 \,(\mathrm{mod}\, d)}} \mu(m/d)\mu(m)h(m)\\ &= \frac{\mu(d)}{g(d)}\cdot\left(\sum_{\substack{m\in\mathcal{D}_z\\ m\equiv 0 \,(\mathrm{mod}\, d)}} h(m)\right)\Bigg/\left(\sum_{m\in\mathcal{D}_z} h(m)\right).\end{aligned} \tag{5.1.4}$$

Additionally,

$$G = \sum_{m\in\mathcal{D}_z} \frac{1}{h(m)} c^2\mu^2(m)h^2(m) = c^2\sum_{m\in\mathcal{D}_z} h(m) = \left(\sum_{m\in\mathcal{D}_z} h(m)\right)^{-1}. \tag{5.1.5}$$

Finally, note that (5.1.4) implies that

$$|\lambda_d| \le 1. \tag{5.1.6}$$

Indeed, if $d\in\mathcal{D}_z$, then d is square-free and $1/g(d) = (1*(1/h))(d)$. Consequently,

$$\begin{aligned}&\frac{1}{g(d)} \sum_{\substack{m\in\mathcal{D}_z\\ m\equiv 0 \,(\mathrm{mod}\, d)}} h(m)\\ &\quad= \left(\sum_{f\mid d}\frac{1}{h(f)}\right)\left(\sum_{\substack{m\in\mathcal{D}_z\\ m\equiv 0 \,(\mathrm{mod}\, d)}} h(m)\right) = \left(\sum_{f\mid d}\frac{h(d)}{h(f)}\right)\left(\sum_{\substack{k\in\mathcal{D}_z\\ k\le\sqrt{D}/d,\ (k,d)=1}} h(k)\right)\\ &\quad= \left(\sum_{f\mid d} h(f)\right)\left(\sum_{\substack{k\in\mathcal{D}_z\\ k\le\sqrt{D}/d,\ (k,d)=1}} h(k)\right) = \sum_{\substack{m=kf\in\mathcal{D}_z\\ k\le\sqrt{D}/d,\ (k,d)=1,\ f\mid d}} h(kf) \le \sum_{m\in\mathcal{D}_z} h(m),\end{aligned}$$

which proves (5.1.6). Lastly, note that (5.1.6) implies that

$$|\mu^+(d)| \le \sum_{\substack{d_1,d_2|P(z)\\ [d_1,d_2]=d}} 1 \le 3^{\omega(d)} \le \tau_3(d).$$

Combining all of the above, we conclude that

Theorem 5.1.1. *Let $\mathcal{A}$ be a finite set of integers satisfying* (A1) *and* (A2) *and let D and z be positive real numbers. If $h(n) = \prod_{p|n} g(p)/(1-g(p))$, then*

$$S(\mathcal{A},z) \le X\left(\sum_{m\le\sqrt{D},\, m|P(z)} h(m)\right)^{-1} + \sum_{d\le D,\, d|P(z)} \tau_3(d)|r_d|.$$

Remark 5.1.2. Note that we no longer need to keep track of the fact that the first sum above runs over integers m with $g(m) \neq 0$, as this is captured by the vanishing of $h(m)$ whenever $g(m) = 0$.

As we will see in the next section, this theorem is essentially as strong as the upper bound implicit in Theorem 3.3.1. For now, let us point out that

$$\sum_{m\le\sqrt{D},\, m|P(z)} h(m) \le \prod_{p<z}(1+h(p)) = \prod_{p<z}\frac{1}{1-g(p)} = \frac{1}{V(z)}.$$

So the upper bound provided by Theorem 5.1.1 is always at least as big as $X \cdot V(z)$.

Remark 5.1.3. As in the combinatorial sieve, we need control of the error terms on average. We may then impose the condition that

$$\text{(R}') \qquad \sum_{d\le D,\, d|P(z)} \tau_3(d)|r_d| \le \frac{C_3 X}{(\log X)^B}.$$

This can be related to condition (R) under the assumption of a crude bound on r_d. Indeed, assume that

$$\text{(r)} \qquad |r_d| \le C_4 \cdot X g(d) \quad (d \le D,\, d|P(z))$$

If, in addition, (A4a) holds and $z \le X$, then

$$\begin{aligned}\sum_{d\le D,\, d|P(z)} \tau_3(d)^2|r_d| &\le C_4 X \sum_{d\le D,\, d|P(z)} \tau_3(d)^2 g(d) \le C_4 X \prod_{p<z}(1+9g(p))\\ &\le \frac{C_4 X}{V(z)^9} \le \frac{C_4 X}{V(X)^9} \le C_4 K^9 X\left(\frac{\log X}{\log 2}\right)^{9\kappa}.\end{aligned}$$

So if (R) holds, then the Cauchy-Schwarz inequality implies that

$$\sum_{d\le D,\, d|P(z)} \tau_3(d)|r_d| \le \left(\sum_{d\le D,\, d|P(z)} \tau_3(d)^2|r_d|\right)^{1/2} \left(\sum_{d\le D,\, d|P(z)} |r_d|\right)^{1/2}$$
$$\le \left(C_4K^9X\left(\frac{\log X}{\log 2}\right)^{9\kappa}\cdot\frac{C_2X}{(\log X)^A}\right)^{1/2}$$

and hence (R′) holds with $C_3 = (C_2C_4)^{1/2}(K/\log 2)^{9\kappa/2}$ and $B = (A-9\kappa)/2$.

5.2 The fundamental lemma: encore

In this section we give a proof of a version of the fundamental lemma using Selberg's sieve. The following theorem confirms that this new sieve is of similar same strength with the β-sieve.

Theorem 5.2.1 (Fundamental Lemma of Sieve Methods, IV)**.** *Let $\mathcal{A}$ be a finite set of integers which satisfies* (A1) *and* (A4b)*, for some $\kappa > 0$ and $C_1 \ge 0$. For $z \ge 1$ and $u \ge \epsilon > 0$, we have that*

$$S(\mathcal{A}, z) = X\cdot V(z)\left\{1 + O_{\kappa,K,\epsilon}\left(u^{-u/2}\right)\right\} + O\left(\sum_{d|P(z),\, d\le z^u} \tau_3(d)|r_d|\right).$$

Proof. Without loss of generality, we may assume that u is large enough. First, we show the upper bound. We claim that

$$V(z)\sum_{m\le z^{u/2},\, m|P(z)} h(m) = 1 + O\left(u^{-(u+1)/2}\right). \tag{5.2.1}$$

Together with Theorem 5.1.1, relation (5.2.1) certainly implies the desired upper bound. Since

$$\sum_{m|P(z)} h(m) = \prod_{p<z}(1+h(p)) = \prod_{p<z}\frac{1}{1-g(p)} = \frac{1}{V(z)},$$

it suffices to show that

$$V(z)\sum_{m|P(z),\, m>z^{u/2}} h(m) \ll_{\kappa,C_1} u^{-(u+1)/2} \tag{5.2.2}$$

Note that $P(z) \le e^{O(z)}$, so we may assume that $u \ll z/\log z$. We shall show (5.2.1) using

Rankin's trick, arguing as in Theorem 0.3.3. For every $\epsilon \in [1/\log z, 1]$, we have that

$$\sum_{m|P(z),\, m>z^{u/2}} h(m) \le z^{-\epsilon u/2} \sum_{m|P(z)} h(m) m^{\epsilon} = z^{-\epsilon u/2} \prod_{p<z} (1+h(p)p^{\epsilon}) \tag{5.2.3}$$

$$= \frac{z^{-\epsilon u/2}}{V(z)} \prod_{p<z} \left(\frac{1+h(p)p^{\epsilon}}{1+h(p)} \right) \tag{5.2.4}$$

$$= \frac{z^{-\epsilon u/2}}{V(z)} \prod_{p<z} \left(1+g(p)\left(p^{\epsilon}-1\right)\right). \tag{5.2.5}$$

Note that $p^{\epsilon} \le 1 + e\epsilon \log p$ for $p \le e^{1/\epsilon} \le z$ and thus

$$\begin{aligned} \log \prod_{p<e^{1/\epsilon}} (1+g(p)(p^{\epsilon}-1)) &\le e\epsilon \sum_{p<e^{1/\epsilon}} g(p)\log p \\ &\le \epsilon \sum_{-1\le m<\log(1/\epsilon)} e^{m+2} \sum_{e^m \le \log p < e^{m+1}} g(p) \\ &\le \epsilon \sum_{-1\le m<\log(1/\epsilon)} e^{m+2} \log \frac{V(\exp\{e^m\})}{V(\exp\{e^{m+1}\})} \\ &\le \epsilon \sum_{-1\le m<\log(1/\epsilon)} e^{m+2} \left(\frac{C_1}{e^m}+\kappa\right) \ll_{C_1,\kappa} 1, \end{aligned} \tag{5.2.6}$$

by (A4b). Moreover, note that for every integer r with $2 \le r \le \epsilon \log z + 1$, we have that

$$\begin{aligned} \log \prod_{e^{(r-1)/\epsilon} \le p < e^{r/\epsilon}} \left(1+g(p)\left(p^{\epsilon}-1\right)\right) &\le e^r \sum_{e^{(r-1)/\epsilon} \le p < e^{r/\epsilon}} g(p) \le e^r \log \frac{V(e^{(r-1)/\epsilon})}{V(e^{r/\epsilon})} \\ &\le e^r \left(\frac{\epsilon C_1}{r-1} + \kappa \log \frac{r}{r-1} \right) \\ &\ll_{C_1,\kappa} \frac{e^r}{r}. \end{aligned}$$

by (A4b), and consequently

$$\log \prod_{e^{1/\epsilon} \le p < z} (1+g(p)(p^{\epsilon}-1)) \le O_{C_1,\kappa}\left(\frac{z^{\epsilon}}{\epsilon \log z}\right).$$

Combining the above estimates, we deduce that

$$V(z) \sum_{m|P(z),\, m>z^{u/2}} h(m) \le \exp\left\{ -\frac{\epsilon u \log z}{2} + O_{C_1,\kappa}\left(\frac{z^{\epsilon}}{\epsilon \log z}\right)\right\},$$

for every $\epsilon \in [1/\log z, 1]$. We set $w = \epsilon \log z$ and choose $w \ge 1$ so that $e^{w-1}/w = u$. Then $w = \log u + \log\log u + O(1)$ and thus

$$\begin{aligned} -\frac{\epsilon u \log z}{2} + O_{C_1,\kappa}\left(\frac{z^{\epsilon}}{\epsilon \log z}\right) = -\frac{uw}{2} + O_{\kappa,C_1}\left(\frac{e^w}{w}\right) &= -\frac{u\log u + u\log\log u}{2} + O_{C_1,\kappa}(u) \\ &\le -\frac{(u+1)\log u}{2} + O_{C_1,\kappa}(1), \end{aligned}$$

which completes the proof of (5.2.2) and hence of the desired upper bound.

Finally, for the lower bound, we start with Buchstab's identity (3.2.1)

$$S(\mathcal{A},z)=|\mathcal{A}|-\sum_{p<z}S(\mathcal{A}_p,p), \tag{5.2.7}$$

which allows to turn the task of finding a lower bound for $S(\mathcal{A},z)$ to finding upper bounds for $S(\mathcal{A}_p,p)$, for $p<z$, which will be accomplished by an appeal to (5.2.2). We remark here that in the proof of (5.2.2) we only needed to (A1) for $d|P(z)$ and (A4b) with $3/2\le w\le w'\le z$. Note that for $d|P(p)$, we have that

$$|(\mathcal{A}_p)_d|=|\mathcal{A}_{[p,d]}|=|\mathcal{A}_{pd}|=Xg(pd)+r_{pd}=(g(p)X)\cdot g(d)+r_{dp}, \tag{5.2.8}$$

which shows that (A1) holds with $Xg(p)$ and r_{dp} in place of X and r_d, respectively. Moreover, relation (A4b) is clearly satisfied for $3/2\le w\le w'\le p$ with the same parameters κ and C_1. Therefore, relation (5.2.2) with $\mathcal{A}_p$, p and w_p in place of $\mathcal{A}$, z and u, respectively, implies that

$$S(\mathcal{A}_p,p)\le\left\{1+O_{C_1,\kappa}\left(w_p^{-(w_p+1)/2}\right)\right\}\cdot g(p)X\prod_{p'<p}(1-g(p'))+O\left(\sum_{d\le p^{w_p},\,d|P(p)}\tau_3(d)|r_{dp}|\right).$$

We pick w_p so that $p^{w_p}=z^u/p$, that is to say, $w_p=-1+u\log z/\log p$, so that if we set $h=u(\log z/\log p-1)$, then

$$\begin{aligned}(w_p+1)\log w_p&=O(1)+(w_p+1)\log(w_p+1)=O(1)+(u+h)\log(u+h)\\&\ge O(1)+u\log u+h\ge O(1)+u\log u+\frac{\log z}{\log p},\end{aligned}$$

by the Mean Value Theorem, so that $w_p^{(w_p+1)/2}\gg u^{u/2}e^{-\log z/\log p}$. So we conclude that

$$S(\mathcal{A}_p,p)\le\left\{1+O_{C_1,\kappa}\left(u^{-u/2}e^{-\log z/\log p}\right)\right\}\cdot g(p)X\prod_{p'<p}(1-g(p'))+O\left(\sum_{d\le z^u/p,\,d|P(p)}\tau_3(d)|r_{dp}|\right).$$

Inserting this inequality into (5.2.7) and using (A1) to write $|\mathcal{A}|$ in terms of g, X and r_1, we find that

$$\begin{aligned}S(\mathcal{A},z)\ge X\left(1-\sum_{p<z}g(p)V(p)\right)&+O_{C_1,\kappa}\left(\frac{X}{u^{u/2}}\sum_{p<z}\frac{g(p)V(p)}{e^{\frac{\log z}{\log p}}}\right)\\&+O\left(\sum_{m\le z^u,\,m|P(z)}\tau_3(m)|r_m|\right),\end{aligned} \tag{5.2.9}$$

since each $m|P(z)$ with $m\le z^u$ can be written uniquely as $m=dp$ with $d|P(p)$, $p<z$ and $d\le z^u/p$. Finally, observe that

$$1-\sum_{p<z}g(p)V(p)=V(z) \tag{5.2.10}$$

and that

$$\begin{aligned}\sum_{p<z}\frac{g(p)V(p)}{e^{\frac{\log z}{\log p}}} &\ll_{C_1} V(z)\sum_{p<z}\frac{g(p)V(p)}{e^{\frac{\log z}{\log p}}}\left(\frac{\log z}{\log p}\right)^{\kappa}\\ &\ll_{C_1,\kappa}\frac{V(z)}{\log z}\sum_{p<z}g(p)\log p \ll_{C_1,\kappa} V(z),\end{aligned}\tag{5.2.11}$$

by the argument leading to (5.2.6). Combining relations (5.2.9), (5.2.10) and (5.2.11) completes the proof of the desired lower bound as well. □

5.3 Applications

Our goal here is to obtain upper bounds on $S(\mathcal{A},z)$ that are as tight as possible for z as large as possible. Our starting point is Theorem 5.1.1 with $D=z^2$, which implies that

$$S(\mathcal{A},z)\le X\left(\sum_{m<z}\mu^2(m)h(m)\right)^{-1}+\sum_{d\le z^2}\tau_3(d)|r_d|.\tag{5.3.1}$$

This reduces upper bounds on $S(\mathcal{A},z)$ to lower bounds on the partial sums of $\mu^2(m)h(m)$. Theorem 0.4.1 then handles these averages.

Theorem 5.3.1. *Let $\mathcal{A}$ be a finite set of integers which satisfies* (A1) *and* (A3′) *for some $\kappa>0$, $C_6\ge 0$ and $L\ge 1$. Furthermore, assume that $g(p)\le C_5/p$ for all primes p, and that* (R′) *holds with $B=\kappa+1$ and $D=X^\theta$, for some $\theta\in(0,1)$. For $z\ge e^{L/\theta}$, we have that*

$$S(\mathcal{A},z)\le\mathfrak{S}(\mathcal{A})\cdot\frac{X}{(\log X)^\kappa}\cdot\left\{\frac{\Gamma(\kappa+1)}{(\theta/2)^\kappa}+O_{\kappa,C_5,C_6}\left(\frac{L}{\theta^{\kappa+1}\log z}\right)\right\},$$

where

$$\mathfrak{S}(\mathcal{A})=\prod_p(1-g(p))\left(1-\frac{1}{p}\right)^{-\kappa}.$$

Proof. This is a direct corollary of Theorems 5.1.1 and 0.4.1 (note that since g satisfies its hypotheses, so does h). □

Corollary 5.3.2. *Let $F_1(x),\dots,F_r(x)$ be distinct irreducible polynomials over $\mathbb{Z}[x]$ with positive leading coefficient. Suppose that the polynomial $F=F_1\cdots F_r$ has no fixed prime divisors, i.e. there is no prime p such that $p|F(n)$ for all integers n. Then we have that*

$$\#\{n\le x: F_1(n),\dots,F_r(n)\text{ are all primes}\}\le\frac{\mathfrak{S}(F)x}{(\log x)^r}\left\{2^r r!+O_F\left(\frac{\log\log x}{\log x}\right)\right\},$$

where

$$\mathfrak{S}(F)=\prod_p\left(1-\frac{\nu_F(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-r},$$

and $\nu_F(m)=\#\{n\in\mathbb{Z}/m\mathbb{Z}: F(n)\equiv 0\pmod m\}$.

Proof. Exercise. □

With regard to the frequency that tuples of polynomials are simultaneously prime numbers, Bateman-Horn conjectured that

Conjecture 5.3.3 (Bateman-Horn)**.** *Let $F_1(x), \dots, F_r(x)$ be distinct irreducible polynomials over $\mathbb{Z}[x]$ with positive leading coefficient. Suppose that the polynomial $F = F_1 \cdots F_r$ has no fixed prime divisors, i.e. there is no prime p such that $p|F(n)$ for all integers n. Then we have that*

$$\#\{n \leq x : F_1(n), \dots, F_r(n) \text{ are all primes}\} \sim_F \mathfrak{S}(F)x \prod_{i=1}^{r} \frac{1}{\log(x^{d_i})} \quad (x \to \infty),$$

where d_i denotes the degree of F_i.

Exercise 5.3.4. Using the ideas of Section 7.4, build a probabilistic model in favour of the above conjecture.

In certain cases, it is possible to improve upon Corollary 5.3.2.

Theorem 5.3.5. *Let $s \in \mathbb{N}$. For $x \geq 2$, we have that*

$$\#\{p \leq x : p + 2s \text{ is prime}\} \leq \left\{4 + O\left(\frac{\log\log x}{\log x}\right)\right\} \cdot \frac{cx}{(\log x)^2} \cdot \prod_{\substack{p|s \\ p>2}} \frac{p-1}{p-2},$$

where

$$c = 2\prod_{p>2}\left(1 - \frac{2}{p}\right)\left(1 - \frac{1}{p}\right)^{-2}$$

is the twin prime constant.

Proof. If $\mathcal{A} = \{p + 2s : p \leq x\}$, then we have that

$$\#\{p \leq x : p + 2s \text{ is prime}\} \leq S(\mathcal{A}, z) + z. \tag{5.3.2}$$

Moreover, relation (A1) holds with $X = \mathrm{li}(x)$,

$$g(d) = \begin{cases} 1/\varphi(d) & \text{if } (d, 2s) = 1, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$|r_d| \leq \begin{cases} E(x; d) & \text{if } (d, 2s) = 1, \\ \omega(2s) & \text{otherwise,} \end{cases}$$

where $E(x; d)$ is defined by (4.0.1). Therefore, if $D \leq \sqrt{x}/(\log x)^B$ for some large enough B, then the Bombieri-Vinogradov theorem (Theorem 4.0.4) and Remark 5.1.3 imply that

(R′) holds with $A = 3$. Consequently, Theorem 5.1.1 and relation (5.3.2) with $D = z^2 = \sqrt{x}/(\log x)^B$ yield that

$$\#\{p \le x : p + 2s \text{ is prime}\} \le \frac{\mathrm{li}(x)}{S} + O\left(\frac{x}{(\log x)^3}\right), \tag{5.3.3}$$

where

$$S = \sum_{\substack{n \le \sqrt{D} \\ (n,s)=1}} \mu^2(n)h(n) \quad \text{with} \quad h(n) = \prod_{p|n,\, p>2} \frac{1}{p-2}.$$

Note that

$$\prod_{p|s}(1+h(p)) \sum_{\substack{n \le \sqrt{D} \\ (n,s)=1}} \mu^2(n)h(n) = \left(\sum_{d|s} \mu^2(d)h(d)\right) \sum_{\substack{n \le \sqrt{D} \\ (n,s)=1}} \mu^2(n)h(n) \ge \sum_{m \le \sqrt{D}} \mu^2(m)h(m).$$

Moreover, Theorem 0.4.1 (or the convolution method) implies that

$$\sum_{m \le \sqrt{D}} \mu^2(m)h(m) = \frac{\log D}{2} \prod_p (1+h(p)) \left(1 - \frac{1}{p}\right) + O(1) = \frac{\log x}{4c} + O(\log\log x).$$

Hence

$$S \ge \left(\frac{\log x}{4c} + O(\log\log x)\right) \prod_{p|s} (1+h(p))^{-1} = \left(\frac{\log x}{4c} + O(\log\log x)\right) \prod_{p|s,\, p>2} \frac{p-2}{p-1}.$$

Inserting the above estimate into (5.3.3), we conclude that

$$\#\{p \le x : p + 2s \text{ is prime}\} \le \left\{4 + O\left(\frac{\log\log x}{\log x}\right)\right\} \cdot \frac{\mathrm{li}(x)}{\log x} \cdot \prod_{\substack{p|s \\ p>2}} \frac{p-1}{p-2} + O\left(\frac{x}{(\log x)^3}\right).$$

Since $\mathrm{li}(x) = x/\log x + O(x/\log^2 x)$ and $(p-1)/(p-2) > 1$, the theorem follows. □

The following result is an improved version of Theorem 4.1.4.

Theorem 5.3.6 (Brun-Titchmarsch inequality, II)**.** *For $1 \le q \le y \le x$ and $(a,q) = 1$, we have that*

$$\pi(x;q,a) - \pi(x-y;q,a) \le \frac{2y}{\varphi(q)\log(2y/q)} \left(1 + O\left(\frac{\log\log(3y/q)}{\log(2y/q)}\right)\right).$$

Proof. Exercise. □

Finally, following an argument due to Selberg [Se91, p. 226-233], we show how keeping track of the special structure of the parameters λ_d can yield an improved version of Theorem 5.3.6.

Theorem 5.3.7 (Brun-Titchmarsch inequality, III). *For $1 \le q \le y \le x$ and $(a,q)=1$, we have that*

$$\pi(x;q,a) - \pi(x-y;q,a) \le \frac{2y}{\varphi(q)(\log(y/q)+2.4)} + O(1).$$

Proof. Clearly, we may assume that $y \ge cq$, for some large enough constant c. First, we show the theorem when $q=1$. The general case will follow from this special case. We shall go through the proof of Theorem 5.1.1 again because we need to use the structure of the λ_d's in the error term too.

We start with the inequality

$$\pi(x) - \pi(x-y) \le z - 1 + \#\{x-y < n \le x : (n, P(z)) = 1\}. \tag{5.3.4}$$

Then, for any real numbers λ_d with $\lambda_1 = 1$, we have that

$$\begin{aligned}
\#\{x-y<n\le x : (n,P(z))=1\} &\le \sum_{x-y<n\le x} \left(\sum_{d|(n,P(z))} \lambda_d\right)^2 \\
&= \sum_{d_1,d_2|P(z)} \lambda_{d_1}\lambda_{d_2} \left(\left\lfloor \frac{x}{[d_1,d_2]} \right\rfloor - \left\lfloor \frac{x-y}{[d_1,d_2]} \right\rfloor \right) \\
&\le \sum_{d_1,d_2|P(z)} \lambda_{d_1}\lambda_{d_2} \frac{y}{[d_1,d_2]} + \sum_{d_1,d_2|P(z)} |\lambda_{d_1}\lambda_{d_2}| \\
&= y \sum_{d_1,d_2|P(z)} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} + \left(\sum_{d|P(z)} |\lambda_d|\right)^2.
\end{aligned} \tag{5.3.5}$$

Let $D \ge 1$. Arguing as in the proof of Theorem 5.1.1 (with $g(d) = 1/d$ and $h(d) = 1/\varphi(d)$ for d square-free), we find that if we let

$$\lambda_d = \mu(d) d \left(\sum_{\substack{m|P(z),\ m\le\sqrt{D} \\ m\equiv 0 \,(\mathrm{mod}\, d)}} \frac{1}{\varphi(m)} \right) \Bigg/ \left(\sum_{m|P(z),\, m\le\sqrt{D}} \frac{1}{\varphi(m)} \right),$$

then

$$\sum_{d_1,d_2|P(z)} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} = \left(\sum_{m|P(z),\, m\le\sqrt{D}} \frac{1}{\varphi(m)} \right)^{-1} = \frac{1}{S}. \tag{5.3.6}$$

Moreover, the choice of the parameters λ_d implies that

$$\begin{aligned}\sum_{d|P(z)} |\lambda_d| &= \sum_{d|P(z)} \frac{d}{S} \sum_{\substack{m|P(z),\ m\le\sqrt{D}\\ m\equiv 0 \,(\mathrm{mod}\, d)}} \frac{1}{\varphi(m)} = \frac{1}{S} \sum_{m|P(z),\ m\le\sqrt{D}} \frac{\sigma(m)}{\varphi(m)} \\ &\le \frac{1}{S} \sum_{m|P(z),\ m\le\sqrt{D}} \frac{\sigma(m)}{\varphi(m)} \le \frac{1}{S} \sum_{m\le\sqrt{D}} \frac{\mu^2(m)\sigma(m)}{\varphi(m)} \\ &\sim \frac{\sqrt{D}}{S} \prod_p \left(1+\frac{p+1}{p(p-1)}\right)\left(1-\frac{1}{p}\right),\end{aligned} \tag{5.3.7}$$

as $D \to \infty$, by the convolution method (see Section 0.2). Note that

$$\prod_p \left(1+\frac{p+1}{p(p-1)}\right)\left(1-\frac{1}{p}\right) = \prod_p \left(1+\frac{1}{p^2}\right) = \frac{\zeta(2)}{\zeta(4)} = \frac{15}{\pi^2} < \sqrt{2.31}.$$

So if D is large enough, then inserting (5.3.6) and (5.3.7) into (5.3.5), we deduce that

$$\#\{x-y<n\le x : (n,P(z))=1\} \le \frac{y}{S} + \frac{2.31D}{S^2}.$$

Next, note that if $D=z^2$, then

$$S = \sum_{d\le z} \frac{\mu^2(d)}{\varphi(d)} = \log z + \gamma + \sum_p \frac{\log p}{p(p-1)} + O\left(\frac{\log z}{z}\right) \ge \log z + 1.29$$

for z large enough. Consequently, writing $z=\sqrt{y/t}$ for some $t>0$, we find that

$$\begin{aligned}\#\{x-y<n\le x : (n,P(z))=1\} &\le \frac{y}{\log z + 1.29} + \frac{2.31z^2}{\log^2 z} \\ &= \frac{2y}{\log y - \log t + 2.58} + \frac{4.62\cdot 2y}{t(\log y - \log c)^2} \\ &= \frac{2y}{\log y}\cdot\left(1+\frac{\log t - 2.58 + 4.62/t}{\log y} + O_t\left(\frac{1}{\log y}\right)\right).\end{aligned}$$

We choose $t=4.62$, in which case $\log t - 2.58 - 4.62/t \le 2.409$, to conclude that

$$\#\{x-y<n\le x : (n,P(\sqrt{y/4.62}))=1\} \le \frac{2y}{\log y + 2.409} \quad (y\ge c_1). \tag{5.3.8}$$

Inserting this estimate into (5.3.4) when $z=\sqrt{y/4.62}$ completes the proof of the theorem in the case when $q=1$.

The case $q>1$ now follows from the case $q=1$: note that

$$\pi(x;q,a) - \pi(x-y;q,a) \le \frac{z}{q} + 1 + \sum_{\substack{x-y<n\le x\\ n\equiv a\,(\mathrm{mod}\, q)\\ (n,P(z))=1}} 1. \tag{5.3.9}$$

Write $P_q(z) = \prod_{p<z,\, p\nmid q} p$ and note that if $n \equiv a \,(\mathrm{mod}\, q)$, then $(n, P(z)) = 1$ if and only if $(n, P_q(z)) = 1$. Let $w \in [1, P_q(z)]$ be the multiplicative inverse of $q \,(\mathrm{mod}\, P_q(z))$, that is to say $wq \equiv 1 \,(\mathrm{mod}\, P_q(z))$. Then if we write $n = a + kq$, we find that

$$1 = (a + kq, P_q(z)) = (w(a + kq), P_q(z)) = (aw + k, P_q(z)).$$

Moreover, since $x - y < n \le x$, we have that $(x - a - y)/q < k \le (x - a)/q$. Thus if we set $m = k + aq$, then $x_1 - y/q < m \le x_1$, where $x_1 = (x - a)/q + aq$, and consequently

$$\sum_{\substack{x-y<n\le x\\ n\equiv a\,(\mathrm{mod}\, q)\\ (n,P(z))=1}} 1 = \sum_{\substack{x_1-y/q<m\le x_1\\ (n,P_q(z))=1}} 1.$$

Now note that if $z = \sqrt{(y/q)/4.62}$, then

$$\varphi(q) \sum_{\substack{x_1-y/q<m\le x_1\\ (m,P_q(z))=1}} 1 = \sum_{\substack{x_1-y/q<m\le x_1\\ (m,P_q(z))=1}} \sum_{\substack{1\le j\le q\\ (m+jP_q(z),q)=1}} 1 = \sum_{j=1}^{q} \sum_{\substack{x_1-y/q<m\le x_1\\ (m+jP_q(z),P(z))=1}} 1$$
$$\le q \cdot \frac{2y/q}{\log(y/q) + 2.409} \quad (y \ge c_1 q)$$

by (5.3.8). So letting $z = \sqrt{(y/q)/4.62}$ in (5.3.9), we deduce that

$$\pi(x; q, a) - \pi(x - y; q, a) \le \frac{\sqrt{y/q}}{q} + 1 + \frac{2y}{\varphi(q)(\log(y/q) + 2.409)} \quad (y \ge c_1 q),$$

and the theorem follows. □

Remark 5.3.8. Using different methods, Montgomery and Vaughan [MV] proved that

$$\pi(x; q, a) - \pi(x - y; q, a) \le \frac{2y}{\varphi(q) \log(y/q)} \quad (1 \le q < y \le x,\ (a, q) = 1).$$

Remark 5.3.9. Improving the constant 2 in the statement of Theorem 5.3.7 would have very important consequences. In particular, if there are positive constants ϵ and L such that

$$\pi(x; q, a) \le (2 - \epsilon) \cdot \frac{x}{\varphi(q) \log(x/q)} \quad (q \ge 2,\ (a, q) = 1,\ x \ge q^L),$$

then it is possible to show that there are no Landau-Siegel zeroes, that is to say, there is some constant c', depending at most on L and ϵ, such that the L-function $L(s, \chi)$ has no zeroes in $(1 - c'/\log q, 1)$, for every Dirichlet character $\chi \,(\mathrm{mod}\, q)$. However, Selberg's sieve alone cannot lead to such an improvement: as we will see in the next section, the constant 2 is, in general, best possible.

5.4 The parity problem in sieve methods

Selberg noticed that Theorem 5.1.1 is best possible. More precisely, by constructing sets $\mathcal{A}$ for which the true size of $S(\mathcal{A}, z)$ matches the upper bound provided by Theorem 5.1.1.

More precisely, let $\mathcal{A}$ be a finite set of integers that satisfies (A1), (A3) with $\kappa = 1$, and (R′) with $B = 2$ and $D = X^\theta$, for some $\theta = 1 - o(1)$ as $X \to \infty$. Assume in addition that for the multiplicative function g in relation (A1) we have that $g(p) \ll 1/p$, so that $h(p) = g(p)/(1 - g(p)) = g(p) + O(1/p^2)$. Then h satisfies the hypotheses of Theorem 0.4.1 and consequently

$$\sum_{m \le x} \mu^2(m)h(m) \sim (\log x) \prod_p \frac{1 - 1/p}{1 - g(p)} \quad (x \to \infty).$$

Then the upper bound supplied by Theorem 5.1.1 with $D = X^\theta$ and $z \in (\sqrt{X}, X]$ is asymptotically

$$S(\mathcal{A}, z) \le (1 + o(1)) \frac{2X}{\log X} \prod_p \frac{1 - g(p)}{1 - 1/p} \quad (X \to \infty). \tag{5.4.1}$$

Moreover, note that we trivially have that $S(\mathcal{A}, z) \ge 0$.

In the converse direction, Selberg constructed sets $\mathcal{A}$ for which (5.4.1) is satisfied, as well as sets $\mathcal{A}$ for which $S(\mathcal{A}, z) = o(X/\log X)$, as $X \to \infty$. These extremal examples are

$$\mathcal{A}^{(1)} = \{n \le x : \Omega(n) \text{ is odd}\} \quad \text{and} \quad \mathcal{A}^{(0)} = \{n \le x : \Omega(n) \text{ is even}\}.$$

Then we have that

$$\begin{aligned} \left|\mathcal{A}_d^{(j)}\right| = \sum_{\substack{n \le x \\ d|n}} \frac{1 + (-1)^{j+\Omega(n)}}{2} &= \frac{x}{2d} + O(1) + \frac{(-1)^{j+\Omega(d)}}{2} \sum_{m \le x/d} (-1)^{\Omega(m)} \\ &= \frac{x}{2d} + O\left(\frac{x}{d} e^{-c\sqrt{\log(x/d)}}\right), \end{aligned}$$

by the Prime Number Theorem. So (A1) is satisfied with $X = x/2$ and $g(d) = 1/d$, and (A3) holds with $\kappa = 1$. Morever, relation (R′) holds with $B = 2$ and $D = x/e^{M(\log\log x)^2}$, for some sufficiently large M. So the discussion of the previous paragraph implies that

$$0 \le S(\mathcal{A}^{(j)}, \sqrt{x}) \le (1 + o(1)) \frac{2X}{\log X} = (1 + o(1)) \frac{x}{\log x}.$$

However, note that

$$S(\mathcal{A}^{(1)}, \sqrt{x+1}) = 1 + \#\{\sqrt{x+1} \le p \le x : \Omega(p) \text{ is odd}\} = 1 + \pi(x) - \pi(\sqrt{x}) \sim \frac{x}{\log x},$$

whereas

$$S(\mathcal{A}^{(0)}, \sqrt{x+1}) = 1 + \#\{\sqrt{x+1} \le p \le x : \Omega(p) \text{ is even}\} = 1.$$

So we have found two examples of sets $\mathcal{A}$ which satisfy exactly the same set of sieve-theoretic axioms but for which the size of $S(\mathcal{A}, z)$ is vastly different for a certain choice of

z. This was accomplished by taking the integers $n \in [1, x]$ and splitting them two subsets of roughly the same size, according to the parity of $\Omega(n)$. This inability of sieve methods to distinguish between the elements of the same set that have an even or an odd number of prime factors is referred to as *the parity problem in sieve methods* and it is a stumbling block in many important number-theoretic questions, such as the twin prime conjecture. As Friedlander and Iwaniec showed [FI98], it is possible to overcome this obstacle by adding an extra axiom to (A1), (A3) and (R′), which eliminates examples such as the sets $\mathcal{A}^{(0)}$ and $\mathcal{A}^{(1)}$ defined above. This extra axiom guarantees that the characteristic function of $\mathcal{A}$ does not correlate with the Möbius function, that is to say, we impose conditions of the form

$$\sum_{a \in \mathcal{A}} \mu(a) = o(|\mathcal{A}|) \quad (|\mathcal{A}| \to \infty)$$

or, at least, of the same spirit.

Chapter 6

Smooth numbers

So far we have dealt with sieving problems where we try to estimate the size of a set $\mathcal{A}$ after having removed all multiplies of small primes from it, say of primes $< z$, in the hope of detecting prime numbers in $\mathcal{A}$. However, another natural problem, which also occurs frequently is to study the size of a set $\mathcal{A}$ after having removed all multiplies of large primes from it, say of primes $> y$. Such numbers are called y-smooth numbers[1], and their study has various important applications. We shall study the problem of counting y-smooth numbers in the simple case when $\mathcal{A} = \{n \leq x\}$. To this end, given real numbers x and y, we define

$$\Psi(x, y) = \#\{n \leq x : P^+(n) \leq y\}.$$

Following a heuristic argument based on probabilistic grounds (see Section 2.2), one might guess that $\Psi(x, y) \asymp x \cdot \frac{\log y}{\log x}$. However, as we will see, in reality the size of $\Psi(x, y)$ is much smaller.

6.1 Iterative arguments and integral-delay equations

As one might expect, estimating $\Psi(x, y)$ becomes harder as y becomes smaller in terms of x. But for large values of x, it is not terribly hard to estimate $\Psi(x, y)$. Indeed, observe that when $y \geq x$, we trivially have that

$$\Psi(x, y) = \lfloor x \rfloor = x + O(1), \tag{6.1.1}$$

[1]The term 'friable numbers', that is to say, numbers that can be easily broken into many little pieces, is used often too.

so there is not much to say in this case. Assume now that $y \in [\sqrt{x}, x]$. Note that an integer $n \leq x$ can have at most one prime divisor in $(\sqrt{x}, x]$. Therefore

$$\begin{aligned}\Psi(x,y) &= \lfloor x \rfloor - \#\{n \leq x : \exists p \in (y,x] \text{ such that } p|n\} \\ &= \lfloor x \rfloor - \sum_{y<p\leq x} \#\{n \leq x : p|n\} = \lfloor x \rfloor - \sum_{y<p\leq x} \left\lfloor \frac{x}{p} \right\rfloor \\ &= x + O(1) - \sum_{y<p\leq x} \left(\frac{x}{p} + O(1) \right) \\ &= x - x \log\left(\frac{\log x}{\log y} \right) + O\left(\frac{x}{\log x} \right),\end{aligned} \tag{6.1.2}$$

by Mertens' estimate. So, we have found an asymptotic formula for $\Psi(x,y)$ for all $y > \sqrt{x}$.

Let us now try to estimate $\Psi(x,y)$ when $y \in [x^{1/3}, x^{1/2})$. In order to do this, we use $\Psi(x,\sqrt{x})$ as an approximation for $\Psi(x,y)$, and try to understand how big is their difference: we have that

$$\begin{aligned}\Psi(x,y) &= \Psi(x,\sqrt{x}) - \#\{n \leq x : y < P^+(n) \leq \sqrt{x}\} \\ &= \Psi(x,\sqrt{x}) - \sum_{y<p\leq\sqrt{x}} \#\{n \leq x : P^+(n) = p\} \\ &= \Psi(x,\sqrt{x}) - \sum_{y<p\leq\sqrt{x}} \#\{m \leq x/p : P^+(m) \leq p\} \\ &= \Psi(x,\sqrt{x}) - \sum_{y<p\leq\sqrt{x}} \Psi(x/p,p).\end{aligned} \tag{6.1.3}$$

Now, note that $\sqrt{x/p} < p \leq x/p$ for all $p \in (y, \sqrt{x}] \subset (x^{1/3}, x^{1/2}]$. So, we may use (6.1.2) to estimate all terms appearing in (6.1.2). Hence we arrive to the estimate

$$\begin{aligned}\Psi(x,y) &= x(1-\log 2) + O\left(\frac{x}{\log x}\right) - \sum_{y<p\leq\sqrt{x}} \frac{x}{p} \left\{ 1 - \log\left(\frac{\log(x/p)}{\log p} \right) + O\left(\frac{1}{\log(x/p)} \right) \right\} \\ &= x(1-\log 2) - \sum_{y<p\leq\sqrt{x}} \frac{x}{p} \left\{ 1 - \log\left(\frac{\log(x/p)}{\log p} \right) \right\} + O\left(\frac{x}{\log x} \right) \\ &= x(1-\log 2) - \int_y^{\sqrt{x}} \left\{ 1 - \log\left(\frac{\log(x/t)}{\log t} \right) \right\} \frac{dt}{t \log t} + O\left(\frac{x}{\log x} \right),\end{aligned} \tag{6.1.4}$$

by Mertens' estimate and partial summation, which is an asymptotic formula for $\Psi(x,y)$ for $y \in [x^{1/3}, x^{1/2})$.

Relations (6.1.2) and (6.1.4) suggest introducing the parameter

$$u = \frac{\log x}{\log y}.$$

Indeed, if we set

$$\rho(u) = \begin{cases} 1 & \text{if } 0 \le u < 1, \\ 1 - \log u & \text{if } 1 \le u \le 2, \end{cases} \tag{6.1.5}$$

then (6.1.1) and (6.1.2) can be rewritten as

$$\Psi(x, x^{1/u}) = x\rho(u) + O\left(\frac{x}{\log x}\right) \quad (0 \le u \le 2).$$

Moreover, with this notation, relation (6.1.4) can be rewritten as

$$\begin{aligned} \Psi(x, x^{1/u}) &= x\rho(2) - \int_y^{\sqrt{x}} \rho\left(\frac{\log(x/t)}{\log t}\right) \frac{dt}{t \log t} + O\left(\frac{x}{\log x}\right) \\ &= x\rho(2) - x\int_2^u \rho(w-1)\frac{dw}{w} + O\left(\frac{x}{\log x}\right) \\ &= x\left\{1 - \int_1^u \rho(w-1)\frac{dw}{w}\right\} + O\left(\frac{x}{\log x}\right). \end{aligned} \tag{6.1.6}$$

So we see an iterative procedure building up. Indeed, we may define a function $\rho : [0, +\infty) \to \mathbb{R}$ by letting $\rho(u) = 1$ for $u \le 1$ and then defining ρ inductively for $u > 1$ via the relation

$$\rho(u) = 1 - \int_1^u \rho(t-1)\frac{dt}{t}. \tag{6.1.7}$$

Note that when $u \in [1, 2]$, we find that $\rho(u) = 1 - \log u$, which matches (6.1.5). The function ρ is called Dickman's function. Its defining equation (6.1.7) is called an integral-delay equation. It can be viewed as a continuous analogue of the Buchstab-like identity

$$\Psi(x, y) = \Psi(x, z) - \sum_{y < p \le z} \Psi(x/p, p) \quad (y \le z \le x), \tag{6.1.8}$$

which can be derived using the argument leading to (6.1.3). So it should come to no surprise that $\rho(u) = \lim_{x\to\infty} \Psi(x, x^{1/u})/x$. Indeed, this is a consequence of the next theorem:

Theorem 6.1.1. *Let $x \ge y \ge 1$ and set $x = y^u$. Then*

$$\Psi(x, y) = x\rho(u) + O\left(\frac{x}{\log y}\right). \tag{6.1.9}$$

Before we prove this theorem, we discuss briefly some properties of Dickman's function. First of all, note that differentiating (6.1.7) yields the estimate

$$u\rho'(u) = -\rho(u-1) \quad (u \ge 1). \tag{6.1.10}$$

Integrating the above relation, we find that

$$\int_1^u w\rho'(w)dw = -\int_1^u \rho(w-1)dw = -\int_0^{u-1} \rho(t)dt.$$

On the other hand, the left hand side of the above identity is equal to

$$\int_1^u w\rho'(w)dw = u\rho(u) - \rho(1) - \int_1^u \rho(w)dw = u\rho(u) - \int_0^u \rho(t)dt.$$

Putting together the above formulas, we conclude that

$$u\rho(u) = \int_{u-1}^u \rho(t)dt \quad (u \geq 1). \tag{6.1.11}$$

This formula is particularly useful in the study of the Dickman-de Brujin function. For example, using this formula, it is not very hard to show that

$$0 < \rho(u) \leq \frac{1}{\Gamma(u+1)} \quad (u \geq 0). \tag{6.1.12}$$

To see the lower bound, we argue by contradiction: let u_0 be the smallest $u \geq 0$ with $\rho(u) \leq 0$. Then we necessarily have that $u_0 > 1$, since $\rho(u) = 1$ for $u \in [0, 1]$. So (6.1.11) implies that

$$0 \geq \rho(u_0) = \frac{1}{u_0}\int_{u_0-1}^{u_0} \rho(t)dt > 0,$$

which is a contradiction. This proves the lower bound in (6.1.12) does hold. Now, the upper bound follows by (6.1.11) and induction: indeed, the lower bound and (6.1.10) imply that ρ is a decreasing function. Therefore (6.1.11) yields that $u\rho(u) \leq \rho(u-1)$, and the inequality $\rho(u) \leq 1/\Gamma(u+1)$ follows by inducting on $\lfloor u \rfloor$ and the fact that when $u \in [0, 1]$, we have that $\rho(u) = 1 \leq 1/\Gamma(u+1)$.

The following theorem gives a more accurate estimate on the rate of decay of u.

Theorem 6.1.2. *For $u \geq 10$, we have that*

$$\rho(u) = e^{-u\log(u\log u)+O(u)}.$$

Proof. First, we show the lower bound, which is simpler. Define $\xi(u)$ by $e^{\xi(u)} = u\xi(u) + 1$. Clearly, ξ is an increasing function and $\xi(1) = 0$. Moreover, since

$$\xi(u) = \log(u\xi(u) + 1) = \log u + \log\xi(u) + O(1) \quad (u \geq 2), \tag{6.1.13}$$

we find that $\xi(u) \asymp \log u$. Inserting this estimate into (6.1.13), we conclude that

$$\xi(u) = \log(u\log u) + O(1). \tag{6.1.14}$$

Next, we claim that

$$\rho(u) \geq \frac{c}{e^{u\xi(u)}} \quad (u \geq 1), \tag{6.1.15}$$

where c is some absolute constant. Note that if we can establish (6.1.15), then the lower bound follows by (6.1.14). In order to show (6.1.15), we argue by contradiction: assume that

(6.1.15) fails, and let u_0 be the smallest counterexample to it. By choosing c small enough, we may assume that $u_0 \geq 2$. Then

$$\begin{aligned}\frac{c}{e^{u_0\xi(u_0)}} > \rho(u_0) \geq \frac{1}{u_0}\int_{u_0-1}^{u_0}\frac{dt}{e^{t\xi(t)}} &\geq \frac{1}{u_0}\int_{u_0-1}^{u_0}\frac{dt}{e^{t\xi(u_0)}} \\ &= \frac{1}{u_0\xi(u_0)}\left(\frac{1}{e^{(u_0-1)\xi(u_0)}} - \frac{1}{e^{u_0\xi(u_0)}}\right) \\ &= \frac{1}{e^{u_0\xi(u_0)}},\end{aligned}$$

by the definition of $\xi(u)$, which is a contradiction. So relation (6.1.15) does hold, and the lower bound in our theorem follows.

For the upper bound we follow a similar argument: define $\psi(u)$ by $e^{\psi(u)+2} = u\psi(u)$. Then relation (6.1.14) also holds with ψ in place of ξ. Moreover, arguing as above, we can show that

$$\rho(u) \leq \frac{C}{e^{u\psi(u)}} \quad (u \geq 1),$$

where C is some large absolute constant. This completes the proof of the upper bound as well. □

Exercise 6.1.3. Show that

$$\rho(u) = \left(\frac{e+o(1)}{u\log u}\right)^u \quad (u \to \infty).$$

Hint: For each fixed $\epsilon > 0$, show that

$$\left(\frac{e-\epsilon}{u\log u}\right)^u \ll \rho(u) \ll \left(\frac{e+\epsilon}{u\log u}\right)^u \quad (u \geq 2).$$

Exercise 6.1.4. Consider the Laplace transform of Dickman's function

$$\widehat{\rho}(s) = \int_0^\infty \rho(t)e^{-st}dt \quad (s \in \mathbb{C}).$$

Show that $\widehat{\rho}$ satisfies the differential equation

$$\frac{d}{ds}(s\widehat{\rho}(s)) = e^{-s}\widehat{\rho}(s).$$

Derive a formula for $\widehat{\rho}$ (this formula might not be closed).

Next, we show Theorem 6.1.1:

Proof of Theorem 6.1.1. We argue by induction. However, instead of using relation (6.1.8), we establish another iterative formula for $\Psi(x,y)$: using the identity $\log = 1 * \Lambda$, we have

that

$$\sum_{\substack{n\le x\\ P^+(n)\le y}} \log n = \sum_{\substack{n\le x\\ P^+(n)\le y}} \sum_{d\mid n} \Lambda(n) = \sum_{\substack{d\le x\\ P^+(d)\le y}} \Lambda(d)\Psi(x/d,y)$$
$$= \sum_{p\le y} (\log p)\Psi(x/p,y) + O\left(\sum_{\nu\ge 2,\, p\le y} \frac{x\log p}{p^\nu}\right)$$
$$= \sum_{p\le y} (\log p)\Psi(x/p,y) + O(x).$$

On the other hand, we have that

$$\sum_{\substack{n\le x\\ P^+(n)\le y}} \log n = (\log x)\Psi(x,y) + O\left(\sum_{n\le x} \log(x/n)\right) = (\log x)\Psi(x,y) + O(x).$$

Putting together the above estimates, we conclude that

$$(\log x)\Psi(x,y) = \sum_{p\le y} (\log p)\Psi(x/p,y) + O(x) \quad (2\le y\le x). \tag{6.1.16}$$

It is not hard to show that a similar formula also holds for Dickman's function: indeed, by Mertens' theorem and partial summation, we find that

$$\sum_{p\le y} \frac{\log p}{p}\rho\left(\frac{\log(x/p)}{\log y}\right) = \int_1^y \rho\left(\frac{\log(x/t)}{\log y}\right)\frac{dt}{t} + O\left(\int_1^y \left|\rho'\left(\frac{\log(x/t)}{\log y}\right)\right| \frac{dt}{t\log y}\right)$$
$$= (\log y)\int_{u-1}^u \rho(w)dw + O\left(\int_{u-1}^u |\rho'(w)|dw\right).$$

The first integral is equal to $u\rho(u)$, by (6.1.11), and the second one is equal to $\rho(u-1)-\rho(u) \le 1$, since $\rho'(w)$ for all $w \ge 1$, by (6.1.10) and (6.1.12). So we conclude that

$$(\log x)\rho(u) = \sum_{p\le y} \frac{\log p}{p}\rho\left(\frac{\log(x/p)}{\log y}\right) + O(1). \tag{6.1.17}$$

Combining (6.1.16) and (6.1.17), we conclude that

$$(\log x)\left(\frac{\Psi(x,y)}{x} - \rho(u)\right) = \sum_{p\le y} \frac{\log p}{p}\left\{\frac{\Psi(x/p,y)}{x/p} - \rho\left(\frac{\log(x/p)}{\log y}\right)\right\} + O(1), \tag{6.1.18}$$

for all $x \ge y \ge 2$.

We are now ready to show the theorem. Fix $y \ge 2$. We need to prove that there is some absolute constant C such that

$$E(x,y) = \left|\frac{\Psi(x,y)}{x} - \rho\left(\frac{\log x}{\log y}\right)\right| \le \frac{C}{\log y}, \tag{6.1.19}$$

for all $x \geq y$. We may assume that $y \geq e^{C/2}$; else, relation (6.1.19) holds trivially for all $x \geq 2$. Moreover, (6.1.19) holds when $x \in [y, y^2]$, by the discussion in the beginning of this chapter. In order to establish for larger x, we argue inductively: assume that (6.1.19) holds for all $x \in [y, 2^m y^2]$, for some $m \geq 0$, and consider $x \in (2^m y^2, 2^{m+1} y^2]$. Write $x = y^u$ and note that $y \leq x/p \leq 2^m y$, for every prime $p \in [2, y]$, so the induction hypothesis and (6.1.18) imply that

$$(\log x)E(x,y) \leq \sum_{p \leq y} \frac{\log p}{p} \cdot \frac{C}{\log y} + O(1) = \frac{C(\log y + O(1))}{\log y} + O(1) = C + O(1),$$

by Mertens' theorem and our assumption that $y \geq e^{C/2}$. Since $x \geq y^2$, we deduce that

$$E(x,y) \leq \frac{C}{\log x} + O\left(\frac{1}{\log x}\right) \leq \frac{C}{2\log y} + O\left(\frac{1}{\log y}\right) \leq \frac{C}{\log y},$$

provided that C is large enough, which we may assume. This completes the inductive step and hence the proof of the theorem. □

6.2 Rankin's method: encore

Finally, in this section we show a result which is cruder than Theorem 6.1.1 when u is small but provides stronger results when u is large. We follow the argument given in [Fo, Part 4]. The main idea is to use Rankin's trick, as in Theorem 0.3.3. However, we first need to rewrite $\Psi(x,y)$ appropriately and to do this we use ideas coming from the integral-delay equations satisfied by multiplicative functions (see Section 0.4). A cruder argument is also possible, but with an extra factor of $\log y$ on our estimate for $\Psi(x,y)$.

Theorem 6.2.1. *Let $x \geq y \geq 3$ and set $x = y^u$. If $y \geq (\log x)^3$, then we have that*

$$\Psi(x,y) \leq x \cdot \frac{e^{O(u)}}{(u \log u)^u}.$$

Proof. Without loss of generality, we may assume that u and x are large enough. As in the proof of Theorem 6.1.1, we start with the formula

$$\sum_{\substack{n \leq x \\ P^+(n) \leq y}} \log n = \sum_{\substack{m \leq x \\ P^+(m) \leq y}} \sum_{\substack{d \leq x/m \\ P^+(d) \leq y}} \Lambda(d). \tag{6.2.1}$$

Now, fix some $\epsilon \in [1/\log y, 1/3]$, and note that for $1 \leq n \leq x$ we have that

$$\log x = \log n + \log(x/n) \leq \log n + \frac{1}{1-\epsilon} \cdot \frac{x^{1-\epsilon}}{n^{1-\epsilon}} \leq \log n + \frac{3x^{1-\epsilon}}{n^{1-\epsilon}},$$

by the inequality $\log t \leq t$, for $t > 0$. Together with (6.2.1), this implies that

$$(\log x)\Psi(x,y) \ll \sum_{\substack{n \leq x \\ P^+(n) \leq y}} \frac{x^{1-\epsilon}}{n^{1-\epsilon}} + \sum_{\substack{m \leq x \\ P^+(m) \leq y}} \sum_{\substack{d \leq x/m \\ P^+(d) \leq y}} \Lambda(d). \tag{6.2.2}$$

Next, note that

$$\sum_{\substack{d\le x/m\\ P^+(d)\le y}}\Lambda(d)=\sum_{p\le\min\{y,x/m\}}(\log p)\sum_{\substack{\nu\ge1\\ p^\nu\le x/m}}1\ll\sum_{p\le\min\{y,x/m\}}\log(x/m).$$

So, if $x/y<m\le x$, then we find that

$$\sum_{\substack{d\le x/m\\ P^+(d)\le y}}\Lambda(d)\ll\frac{x}{m}\le\frac{y^\epsilon x^{1-\epsilon}}{m^{1-\epsilon}},$$

whereas, if $1\le m\le x/y$, then

$$\sum_{\substack{d\le x/m\\ P^+(d)\le y}}\Lambda(d)\ll\frac{y\log\frac{x}{m}}{\log y}=y+\frac{y\log\frac{x}{my}}{\log y}\le\frac{y^\epsilon x^{1-\epsilon}}{m^{1-\epsilon}}+\frac{y}{\log y}\cdot\frac{1}{1-\epsilon}\cdot\frac{x^{1-\epsilon}}{y^{1-\epsilon}m^{1-\epsilon}}\ll\frac{y^\epsilon x^{1-\epsilon}}{m^{1-\epsilon}}.$$

In any case, we have that

$$\sum_{\substack{d\le x/m\\ P^+(d)\le y}}\Lambda(d)\ll\frac{y^\epsilon x^{1-\epsilon}}{m^{1-\epsilon}}.$$

Inserting this estimate into (6.2.2), we arrive to the estimate

$$\Psi(x,y)\ll\frac{y^\epsilon x^{1-\epsilon}}{\log x}\sum_{\substack{n\le x\\ P^+(n)\le y}}\frac{1}{n^{1-\epsilon}}\le\frac{y^\epsilon x^{1-\epsilon}}{\log x}\prod_{p\le y}\left(1-\frac{1}{p^{1-\epsilon}}\right)^{-1}\ll\frac{y^\epsilon x^{1-\epsilon}}{u}\exp\left\{\sum_{p\le y}\frac{p^\epsilon-1}{p}\right\}.$$

We bound this last sum as in the proof of Theorem 0.3.3 and choose $\epsilon=w/\log y$, where $e^{w-1}/w=u$, to complete the proof of the theorem. □

Chapter 7

Gaps between primes

Let $p_1, p_2, p_3, \dots$ be the sequence of prime numbers in increasing order. Our goal in this chapter is to study the spacing of this sequence and, in particular, how small and how large the gaps between two consecutive primes can get. The Prime Number Theorem implies that $p_n \sim n \log n$ as $n \to \infty$ or, equivalently, that $\sum_{k \le n}(p_{k+1} - p_k) = p_{n+1} - p_1 \sim n \log n$. Consequently,

$$\begin{aligned}\sum_{1<k\le n} \frac{p_{k+1}-p_k}{\log k} &= \int_1^n \frac{1}{\log t}\, d\left(\sum_{1<k\le t}(p_{k+1}-p_k)\right) \\ &= \frac{p_n - p_2}{\log n} - \int_1^n \left(\sum_{1<k\le t}(p_{k+1}-p_k)\right)\frac{dt}{t\log^2 t} \\ &= \frac{p_n - p_2}{\log n} - \int_1^n \frac{(p_{\lfloor t+1\rfloor} - p_2)dt}{t\log^2 t} \sim n \quad (n\to\infty),\end{aligned}$$

that is to say, the mean value of $(p_{k+1} - p_k)/\log k$ is 1. However, it could be possible that this ratio deviates significantly from its mean. Indeed, if the twin prime conjecture is true, then it immediately follows that $p_{k+1} - p_k = 2$ for infinitely many values of k. One of the main results of this chapter, which will be proven in Section 7.0.2, is a major step towards this conjecture.

Theorem 7.0.2. *For each $m \in \mathbb{N}$, we have that*

$$\liminf_{n\to\infty}(p_{n+m} - p_n) \ll e^{4m} m^5.$$

This result is due to Maynard [May15a] and Tao [Tab], who built upon previous work of Goldston, Pintz and Yildirim [GPY]. In the special case $m = 1$, the first person to show that $\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$ was Zhang [Z]. We will discuss more the history of Theorem 7.0.2 in Section 7.1, where its proof will be presented.

Exercise 7.0.3. Show that

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log n} \le \frac{7}{8}.$$

Hint: Fix $\delta > 0$. Starting from the formula

$$(1+o(1))N\log N = \sum_{k=1}^{\infty} 2k \sum_{\substack{N<n\le 2N\\ p_{n+1}-p_n=2k}} 1 \quad (N\to\infty),$$

show that

$$\sum_{k\le \frac{(1+\delta)\log N}{2}} ((1+\delta)\log N - 2k) \sum_{\substack{N<n\le 2N\\ p_{n+1}-p_n=2k}} 1 \ge (\delta+o(1))N\log N,$$

as $N\to\infty$. Next, use the sieve to bound $\#\{N < n \le 2N : p_{n+1}-p_n = 2k\}$ from above and deduce the desired result.

On the other hand, one could imagine that it would be feasible to construct long strings of consecutive numbers that are all compositive. For example, all numbers in the sequence $n!+2, n!+3, \dots, n!+n$ are composite. Consequently, if $p_r < n!+2 < p_{r+1}$ are consecutive primes, then $p_{r+1}-p_r \ge n$. Moreover, Bertrand's postulate implies that $p_r > n!/2$ and therefore $\log r \sim \log p_r \sim \log(n!) \sim n\log n$ as $n\to\infty$, by Stirling's formula, that is to say, $n \sim \log r/\log\log r$. So this construction produces gaps $d_r \gtrsim \log r/\log\log r$, for infinitely many integers r. This is not quite enough to yield gaps that are longer than the average ones. However, Westzynthius used a different construction to construct large gaps between primes, that is to say infinitely many integers n such that $p_{n+1}-p_n/\log n$ can get arbitrarily large. His ideas were strengthened by Erős and subsequently by Rankin. Following their arguments, we will show the following slightly stronger result (Rankin's result had the constant $e^\gamma/2$ in place of e^γ):

Theorem 7.0.4. *There are infinitely many integers n such that*

$$\frac{p_{n+1}-p_n}{\log n} \ge (e^\gamma + o_{n\to\infty}(1))\cdot \frac{(\log\log n)(\log\log\log\log n)}{(\log\log\log n)^2}$$

Erdős offered \$10,000 for improving the constant e^γ to an arbitrarily large constant, the largest Erdős prize ever. This problem was solved independently in two papers that appeared simultaneously, one by Maynard [May] and another one by Ford, Green, Konyagin and Tao [FGKT]:

Theorem 7.0.5. *We have that*

$$\limsup_{n\to\infty}\left(\frac{p_{n+1}-p_n}{\log n}\Big/\frac{(\log\log n)(\log\log\log\log n)}{(\log\log\log n)^2}\right) = \infty.$$

The proof of Theorem 7.0.5 we will present is the one due to Maynard, as it is simpler and its methods fit more naturally within the context of this chapter. The argument will be given in Section 7.3.

7.1 Bounded gaps between primes

In this section we prove Theorem 7.0.2. In order to detect $m+1$ primes close together, we use the following construction: Let $0 \le s_1 < s_2 < \cdots < s_k$ be integers such that the k-tuple

$\boldsymbol{s} = (s_1, \dots, s_k)$ is *admissible*, that is to say, its elements do not cover all congruences classes modulo any prime. Then, for an integer $N \geq s_k$ and a sequence of non-negative weights $\{w_n\}_{n=1}^{\infty}$, we consider the sum

$$S = \sum_{N < n \leq 2N} \left(\sum_{j=1}^{k} \mathbf{1}_{\mathbb{P}}(n + s_j) - m \right) w_n.$$

Clearly, if $S > 0$, then there are $m+1$ primes distinct primes in $(N + s_1, 2N + s_k]$ within an interval of length $s_k - s_1$. So the sum S is our "short gap detector". The key to making this approach work is to judiciously choose the weights w_n in a way that we achieve two things:

- Most of the contribution to the sum S comes from integers n for which there is a high probability that several of the numbers $n + s_1, \dots, n + s_k$ are simultaneously primes. Indeed, if, for example, $w_n = 1$ for all n, then $S \sim kN/\log N - mN < 0$ as $N \to \infty$, so this is not a good choice for w_n.
- We can actually estimate S unconditionally (e.g. without appealing to the twin prime conjecture, which would result to our argument entering a vicious cycle). Indeed, if w_n is the characteristic function of integers n such that $n + s_1, \dots, n + s_k$ are all primes, then $S \sim (k-m)(\log N) \cdot \#\{N < n \leq 2N : w_n = 1\}$, so in order to show that $S > 0$, we need to show that $w_n = 1$ often with $k = m + 1$, thus entering a circular argument.

These two restrictions suggest setting $w_n = (1 * \mu^+)(Q(n))$, where

$$Q(n) := \prod_{i=1}^{k} (n + s_j)$$

and μ^+ is an upper bound sieve. Indeed, the original idea presented here is due to Goldston, Pintz and Yildirim [GPY], who defined μ^+ using the same idea as in Selberg's sieve, by letting

$$w_n = \left(\sum_{d \mid (P(z), Q(n))} \lambda_d \right)^2, \tag{7.1.1}$$

with the parameters λ_d being at our disposal, with the condition that they are supported on integers $d \leq D$ (we don't have to assume that $\lambda_1 = 1$ here). If we set $D = N^{\theta/2}$, then opening the square and interchanging the order of summation, we find that we need a bound of the form

$$\sum_{q \leq x^{\theta}} \max_{(a,q)=1} \left| \pi(x; q, a) - \frac{\mathrm{li}(x)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^A}, \tag{7.1.2}$$

for all $x \geq 2$ and all $A \geq 1$. Note the Bombieri-Vinogradov theorem implies that (7.1.2) holds with $x^{1/2}/(\log x)^B$ and B sufficiently large in place of x^{θ}. In particular, any $\theta < 1/2$ is permissible. However, this was not enough to deduce that $\liminf_{n \to \infty} (p_{n+1} - p_n) < \infty$ in the original approach by Goldston, Pintz and Yildirim. Indeed, their method required (a

weaker version of) (7.1.2) with some fixed $\theta > 1/2$ that was not available at the time. This missing ingredient was supplied by Zhang in May 2013 in his breakthrough paper [Z], where he proved a variation of (7.1.2) with $\theta = 1/2 + 1/584$ that was sufficient for him to deduce that $\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$.

Very shortly after Zhang's paper was published, another method was proposed by James Maynard [May15a]. Maynard, instead of trying to prove stronger level-of-distribution results about the primes, took an alternative path and introduced a multidimensional variation of the GPY weights. His idea, also discovered independently by Tao [Tab], was to consider weights of the form

$$w_n = \left(\sum_{\substack{d_j | n+s_j \\ 1\le j\le k}} \lambda_{\boldsymbol{d}} \right)^2 .$$

It turned out that this simple idea has very far reaching consequences. As we will see, these modified weights need very weak level-of-distribution results as input. In fact, we only need to know that (7.1.2) holds for some $\theta > 0$, a much weaker result that the available Bombieri-Vinogradov result that allows us to take $\theta = 1/2 - \epsilon$. This new flexibility that the Maynard-Tao variation of the Goldston-Pintz-Yildirim weights permits makes them very applicable to a wide variety of set-ups. Indeed, after the publication of Maynard's paper, the subject has witnessed an explosion of activity.

We now proceed to the proof of Theorem 7.0.2. We will add a small technical twist to the construction of Maynard's weights, by performing a preliminary sieve up to z, where this is a parameter at our disposal (we will eventually take z to be a large power of $\log N$ - a similar idea is also used in [May15a], but the pre-sieving parameter is smaller). Indeed, we set

$$w_n = \mathbf{1}_{P^-(Q(n))>z} \cdot \left(\sum_{\substack{d_j | n+s_j \\ 1\le j\le k}} \lambda_{\boldsymbol{d}} \right)^2 ,$$

where the sequence $\lambda_{\boldsymbol{d}}$ is supported on those k-tuples $\boldsymbol{d} = (d_1, \dots, d_k)$ with $d_1 \cdots d_k \le D$. We will eventually take $D = N^{1/4}/(\log N)^{\log z}$. We need to estimate the sum

$$S(N, z) := \sum_{\substack{N<n\le 2N \\ P^-(Q(n))>z}} \left(\sum_{j=1}^{k} \mathbf{1}_{\mathbb{P}}(n + s_j) - m \right) \left(\sum_{\substack{d_j | n+s_j \\ 1\le j\le k}} \lambda_{\boldsymbol{d}} \right)^2 .$$

In the estimation of $S(N, z)$ in the following two lemmas, we use the notations

$$\nu(d) := \#\{m \in \mathbb{Z}/d\mathbb{Z} : Q(m) \equiv 0 \,(\mathrm{mod}\, d)\},$$

$$M := \max_{\boldsymbol{d}} |\lambda_d| \qquad \text{and} \qquad X := \int_N^{2N} \frac{dt}{\log t} = \frac{N}{\log N} + O\left(\frac{N}{(\log N)^2} \right).$$

All implicit constants in this section might depend on k and the choice of the k-tuple $\boldsymbol{s}$.

Lemma 7.1.1. *If $D \leq N^{1/2-\epsilon}$ and $(\log N)^{A+3k+1} \leq z \leq N^{1/\log\log N}$, then*

$$\sum_{\substack{N<n\leq 2N\\ P^-(Q(n))>z}} \left(\sum_{\substack{d_j \mid n+s_j\\ 1\leq j\leq k}} \lambda_{\boldsymbol{d}}\right)^2 = N\cdot\prod_{p\leq z}\left(1-\frac{\nu(p)}{p}\right)\sum_{\substack{P^-(a_j)>z\\ 1\leq j\leq k}} a_1\cdots a_k \left(\sum_{\substack{d_j\equiv 0 \,(\mathrm{mod}\, a_j)\\ P^-(d_j)>z\\ 1\leq j\leq k}} \frac{\lambda_{\boldsymbol{d}}}{d_1\cdots d_k}\right)^2 + O_{A,\epsilon}\left(\frac{NM^2}{(\log N)^A}\right).$$

Proof. Call T the sum in question. Applying the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.1) with u defined via the relation $z^u = N^\epsilon$, we find that

$$\begin{aligned} T &= \sum_{\substack{(d_ie_i,d_je_jP(z))=1\\ 1\leq i,j\leq k,\ i\neq j}} \lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}\cdot \#\left\{ \begin{array}{ll} N<n\leq 2N & : \ n\equiv -s_j \,(\mathrm{mod}\,[d_j,e_j]),\\ P^-(Q(n))>z & : \ 1\leq j\leq k \end{array}\right\}\\ &= \sum_{\substack{(d_ie_i,d_je_jP(z))=1\\ 1\leq i,j\leq k,\ i\neq j}} \lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}\cdot\left(\left(1+O(u^{-u/2})\right)\frac{N\cdot\prod_{p\leq z}\left(1-\frac{\nu(p)}{p}\right)}{[d_1,e_1]\cdots[d_k,e_k]} + O\left(N^\epsilon(\log N)^{k-1}\right)\right)\\ &= N\cdot\prod_{p\leq z}\left(1-\frac{\nu(p)}{p}\right)\sum_{\substack{(d_ie_i,d_je_jP(z))=1\\ 1\leq i,j\leq k,\ i\neq j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} + O_{\epsilon,A}\left(\frac{M^2N}{(\log N)^A}\right), \end{aligned}$$

We need to remove the conditions $(d_ie_i, d_je_j) = 1$ for $i \neq j$. We do this by noting that if a pair of k-tuples $\boldsymbol{d}, \boldsymbol{e}$ is such that $(d_1\cdots d_k, P(z)) = (e_1\cdots e_k, P(z)) = 1$ but $(d_ie_i, d_je_j) > 1$, then there must be a prime $p > z$ dividing both $[d_i, e_i]$ and $[d_j, e_j]$. We therefore conclude that

$$\sum_{\substack{(d_ie_i,d_je_jP(z))=1\\ 1\leq i,j\leq k,\ i\neq j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} = \sum_{\substack{(d_ie_i,P(z))=1\\ 1\leq i\leq k}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} + O\left(\frac{M^2(\log N)^{3k}}{z}\right),$$

which is admissible by our assumption that $z \geq (\log N)^{3k+A+1}$. Finally, we note that

$$\frac{1}{[d,e]} = \frac{(d,e)}{de} = \frac{1}{de}\sum_{a\mid(d,e)}\varphi(a).$$

Therefore

$$\sum_{\substack{P^-(d_ie_i)>z\\ 1\leq i\leq k}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} = \sum_{P^-(a_1\cdots a_k)>z} \varphi(a_1)\cdots\varphi(a_k)\left(\sum_{\substack{d_j\equiv 0\,(\mathrm{mod}\,a_j)\\ P^-(d_j)>z\\ 1\leq j\leq k}} \frac{\lambda_{\boldsymbol{d}}}{d_1\cdots d_k}\right)^2.$$

If $P^-(a) > z$ and $a \leq e^z$, then it is easy to see that

$$1 \leq \frac{a}{\varphi(a)} = \prod_{p|a}\left(1+\frac{1}{p-1}\right) \leq \left(1+\frac{1}{z-1}\right)^{\omega(a)} \leq \exp\left\{\frac{\omega(a)}{z-1}\right\} = 1+O\left(\frac{\log a}{z}\right),$$

since $\omega(a) \ll \log a$. So we may replace $\varphi(a_j)$ by a_j for all $j \in \{1,\dots,k\}$ by producing a total error term of size $(\log N)^{3k+1}/z$, which is admissible by our assumption on z. □

Lemma 7.1.2. *If* $(\log N)^{A+3k+1} \leq z \leq N^{1/(10\log\log N)}$, $D \leq N^{1/4}/(\log N)^{\log z}$ *and* $j_0 \in \{1,\dots,k\}$, *then*

$$\sum_{\substack{N<n\leq 2N\\ P^-(Q(n))>z}} \mathbf{1}_{\mathbb{P}}(n+s_{j_0})\left(\sum_{\substack{d_j|n+s_j\\ 1\leq j\leq k}}\lambda_{\boldsymbol{d}}\right)^2 = X\cdot\prod_{p\leq z}\left(1-\frac{\nu(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-1}\sum_{\substack{P^-(a_j)>z\\ 1\leq j\leq k\\ a_{j_0}=1}} a_1\cdots a_k$$
$$\times\left(\sum_{\substack{P^-(d_j)>z,\, d_{j_0}=1\\ d_j\equiv 0\,(\mathrm{mod}\, a_j)\\ 1\leq j\leq k}}\frac{\lambda_{\boldsymbol{d}}}{d_1\cdots d_k}\right)^2 + O_{A,\epsilon}\left(\frac{NM^2}{(\log N)^A}\right).$$

Proof. Call T_{j_0} the sum in question. For simplicity, we consider the case $j_0 = k$; the proof of the other cases is identical. Observe that the fact that $n+s_k$ is a prime number greater than N forces $n+s_k$ to be co-prime to integers $d \leq D \leq N$. Therefore, for such an n, the sum $\sum_{d_j|n+s_j\,(1\leq j\leq k)}\lambda_{\boldsymbol{d}}$ must have $d_k = 1$, and the condition $P^-(Q(n)) > z$ is reduced to $P^-(Q^*(n)) > z$, where

$$Q^*(n) = \prod_{j=1}^{k-1}(n+s_j).$$

(This decreases the dimension of the sieve we are considering by 1.) So, opening the square, changing the order of summation and setting $p = n+s_k$, we find that

$$T_k = \sum_{\substack{d_j,e_j\\ (d_ie_i,d_je_jP(z))=1\\ 1\leq i,j<k,\ i\neq j}}\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}\cdot\#\left\{\begin{array}{ll} N+s_k<p\leq 2N+s_k & :\ p\equiv s_k-s_j\,(\mathrm{mod}\,[d_j,e_j]),\\ P^-(Q^*(p-s_k))>z & :\ 1\leq j<k\end{array}\right\}.$$

For each fixed $\boldsymbol{d},\boldsymbol{e}$, we use the Fundamental Lemma of Sieve Methods with $u = \log\log N$ to control the cardinality of the above set of primes. If we set

$$\nu^*(a) := \#\left\{x \in (\mathbb{Z}/a\mathbb{Z})^* : Q^*(x-s_k) \equiv 0\,(\mathrm{mod}\, a)\right\},$$

so that $\nu^*(p) = \nu(p)-1$, then we find that the main term equals

$$\frac{XV}{\varphi([d_1,e_1])\cdots\varphi([d_k,e_k])},$$

where

$$V := \prod_{p\le z}\left(1-\frac{\nu^*(p)}{p-1}\right) = \prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-1},$$

and the error term is

$$\ll \frac{X}{(\log N)^{A+3k}\varphi([d_1,e_1])\cdots\varphi([d_k,e_k])} + \sum_{\substack{a\le z^{\log\log N}\\ a\mid P(z)}} (k-1)^{\omega(a)}E(a[d_1,e_1]\cdots[d_k,e_k]),$$

where

$$E(q) := \max_{(a,q)=1}\left|\sum_{\substack{N+s_k<p\le 2N+s_k\\ p\equiv a\,(\mathrm{mod}\,q)}} 1-\frac{X}{\varphi(q)}\right| = \max_{(a,q)=1}\left|\sum_{\substack{N<p\le 2N\\ p\equiv a\,(\mathrm{mod}\,q)}} 1-\frac{X}{\varphi(q)}\right| + O(1).$$

Therefore

$$T_k = XV\sum_{\substack{(d_ie_i,d_je_jP(z))=1\\ 1\le i,j<k,\ i\ne j}} \frac{\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{\varphi([d_1,e_1])\cdots\varphi([d_{k-1},e_{k-1}])} + O\left(\frac{M^2N}{(\log N)^A} + M^2\sum_{q\le N^{1/2}/(\log N)^{\log z}}\tau_{2k}(q)E(q)\right).$$

The Bombieri-Vinogradov theorem (see, also, Exercice 4.0.8) implies that the sum over q in the error term is $\ll N/(\log N)^A$. So we conclude that

$$T_k = XV\cdot\sum_{\substack{(d_ie_i,d_je_jP(z))=1\\ 1\le i,j<k,\ i\ne j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{\varphi([d_1,e_1])\cdots\varphi([d_k,e_k])} + O\left(\frac{M^2N}{(\log N)^A}\right).$$

As in the proof of Lemma 7.1.1, we may remove the conditions $(d_ie_i,d_je_j)=1$ and we may replace $\varphi([d_j,e_j])$ by $[d_j,e_j]$ at the cost of introducing an error of total size $O(M^2N(\log N)^{3k+1}/z)$, which is admissible since $z\ge(\log N)^{A+3k+1}$. Moreover, using the formula

$$\frac{1}{[d,e]} = \frac{1}{de}\sum_{a\mid(d,e)}\varphi(a),$$

we find that

$$\sum_{\substack{P^-(d_ie_i)>z\\ 1\le i<k}} \frac{\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{[d_1,e_1]\cdots[d_{k-1},e_{k-1}]} = \sum_{\substack{P^-(a_i)>z\\ 1\le i<k}} \varphi(a_1)\cdots\varphi(a_{k-1})\left(\sum_{\substack{P^-(d_i)>z\\ d_i\equiv 0\,(\mathrm{mod}\,a_i)\\ 1\le i<k}} \frac{\lambda_{\boldsymbol{d},1}}{d_1\cdots d_{k-1}}\right)^2.$$

Finally, we replace $\varphi(a_j)$ by a_j, thus introducing a total error of size $O(M^2N(\log N)^{3k+1}/z)$, which is admissible since $z\ge(\log N)^{A+3k+1}$. This completes the proof of the lemma. □

Motivated by the two lemmas above, we make a change of variable which diagonalizes the quadratic form appearing in Lemma 7.1.1: for a k-tuple $\boldsymbol{a} = (a_1, \dots, a_k)$, we set

$$\frac{\xi_{\boldsymbol{a}}}{a_1 \cdots a_k} = \sum_{\substack{P^-(d_j) > z \\ d_j \equiv 0 \,(\mathrm{mod}\, a_j) \\ 1 \le j \le k}} \frac{\lambda_{\boldsymbol{d}}}{d_1 \cdots d_k}. \tag{7.1.3}$$

Clearly, $\xi_{\boldsymbol{a}}$ is supported on k-tuples with $a_1 \cdots a_k \le D$ and $P^-(a_1 \cdots a_k) > z$.

We want to compute the expression appearing in Lemma 7.1.2 in terms of the new parameters $\xi_{\boldsymbol{a}}$. As in Section 5.1 (see relation (5.1.2)), if $P^-(d_1 \cdots d_k) > z$, we have the inversion formula

$$\frac{\lambda_{\boldsymbol{d}}}{d_1 \cdots d_k} = \sum_{\substack{a_j \equiv 0 \,(\mathrm{mod}\, d_j) \\ 1 \le j \le k}} \frac{\xi_{\boldsymbol{a}}}{a_1 \cdots a_k} \prod_{j=1}^{k} \mu(a_j/d_j). \tag{7.1.4}$$

Consequently, if $P^-(a_1 \cdots a_k) > z$ with $a_{j_0} = 1$, then we find that

$$\begin{aligned}
\sum_{\substack{P^-(d_1 \cdots d_k) > z \\ d_j \equiv 0 \,(\mathrm{mod}\, a_j) \\ 1 \le j \le k,\ d_{j_0} = 1}} \frac{\lambda_{\boldsymbol{d}}}{d_1 \cdots d_k} &= \sum_{\substack{P^-(d_1 \cdots d_k) > z \\ d_j \equiv 0 \,(\mathrm{mod}\, a_j) \\ 1 \le j \le k,\ d_{j_0} = 1}} \sum_{\substack{b_j \equiv 0 \,(\mathrm{mod}\, d_j) \\ 1 \le j \le k}} \frac{\xi_{\boldsymbol{b}}}{b_1 \cdots b_k} \prod_{j=1}^{k} \mu(b_j/d_j) \\
&= \sum_{\substack{P^-(b_1 \cdots b_k) > z \\ b_j \equiv 0 \,(\mathrm{mod}\, a_j) \\ 1 \le j \le k}} \frac{\xi_{\boldsymbol{b}}}{b_1 \cdots b_k} \sum_{\substack{d_j \equiv 0 \,(\mathrm{mod}\, a_j) \\ d_j/a_j \mid b_j/a_j \\ 1 \le j \le k,\ d_{j_0} = 1}} \prod_{j=1}^{k} \mu\left(\frac{b_j/a_j}{d_j/a_j}\right) \\
&= \sum_{\substack{b_j = a_j \\ 1 \le j \le k,\, j \ne j_0 \\ P^-(b_{j_0}) > z}} \frac{\mu(b_{j_0}) \xi_{\boldsymbol{b}}}{b_1 \cdots b_k} = \sum_{P^-(b) > z} \frac{\mu(b) \xi_{a_1, \dots, a_{j_0 - 1}, b, a_{j_0 + 1}, \dots, a_k}}{a_1 \cdots a_k b}.
\end{aligned} \tag{7.1.5}$$

This computation suggests setting

$$\xi_{\boldsymbol{a}} := \mathbf{1}_{P^-(a_1 \cdots a_k) > z} \cdot \lambda(a_1 \cdots a_k) \cdot f\left(\frac{\log a_1}{\log D}, \dots, \frac{\log a_k}{\log D}\right) \cdot \prod_{p \le z} \left(1 - \frac{1}{p}\right)^{-k},$$

where $\prod_{p \le z} (1 - 1/p)^{-k}$ is a normalisation factor, $\lambda(n) = (-1)^{\Omega(n)}$ is Liouville's function (which is placed here to annihilate the sign change caused by $\mu(b)$ in (7.1.5)) and f is a smooth function supported on the simplex

$$T_k := \{\boldsymbol{t} \in [0,1]^k : t_1 + \cdots + t_k \le 1\}.$$

With this choice, we have that the following result.

Lemma 7.1.3. *If $z = (\log N)^{6k+2}$ and $D = N^{1/4}/(\log N)^{\log z}$, then*

$$\frac{S(N,z)}{\mathfrak{S}(\boldsymbol{s})N(\log D)^k} = \frac{1}{4}\sum_{j_0=1}^{k}\int\left(\int f(\boldsymbol{t})dt_{j_0}\right)^2 \prod_{\substack{1\le j\le k\\ j\neq j_0}} dt_j - m\int f(\boldsymbol{t})^2 d\boldsymbol{t} + O\left(\frac{1}{\sqrt{\log N}}\right),$$

where

$$\mathfrak{S}(\boldsymbol{s}) := \prod_p \left(1-\frac{\nu(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k};$$

the implied constant depends at most on ϵ, f, k and $\boldsymbol{s}$.

Except for Lemmas 7.1.1 and 7.1.2, the key input to the proof of the above lemma comes from the following result:

Lemma 7.1.4. *Let D and z be two parameters with $\sqrt{\log D} \le z \le D$. If $g : \mathbb{R}^k \to \mathbb{R}$ is a smooth function supported on $\{\boldsymbol{t} \in [1,+\infty)^k : t_1\cdots t_k \le D\}$ and such that $\frac{\partial g}{\partial t_j} \ll 1/t_j$, then*

$$\sum_{P^-(n_1\cdots n_k)>z} \frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \prod_{p\le z}\left(1-\frac{1}{p}\right)^k \int \frac{g(t_1,\dots,t_k)}{t_1\cdots t_k}d\boldsymbol{t} + O\left(\left(\frac{\log D}{\log z}\right)^{k-1}\right);$$

the implied constant depends at most on k and g.

Proof. All implied constants might depend on g and on k. We may assume that $z \le D^{1/(1000k)}$; otherwise, the result is trivially true. We note that

$$\sum_{P^-(n_1\cdots n_k)>z} \frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \sum_{\substack{P^-(n_j)>z\\ n_j>z^{100}\\ 1\le j\le k}} \frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} + O\left(\left(\frac{\log D}{\log z}\right)^{k-1}\right).$$

We split the range of summation $\{\boldsymbol{n} : n_1\cdots n_k \le D,\ n_j > z^{100}\ (1\le j\le k)\}$ into small cubes of the form $B = \prod_{j=1}^k (x_j, x_j+\sqrt{x_j}]$. We note that

$$\frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \frac{g(x_1,\dots,x_k)}{x_1\cdots x_k} + O\left(\frac{1}{z^{100}\cdot x_1\cdots x_k}\right) = \frac{I}{\sqrt{x_1\cdots x_k}} + O\left(\frac{1}{z^{100}\cdot x_1\cdots x_k}\right)$$

by our assumption on g, where

$$I := \int_B \frac{g(t_1,\dots,t_k)}{t_1\cdots t_k}d\boldsymbol{t}.$$

Therefore

$$\begin{aligned}\sum_{\substack{P^-(n_1\cdots n_k)>z\\ \boldsymbol{n}\in B}} \frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} &= \frac{I}{\sqrt{x_1\cdots x_k}}\prod_{j=1}^{k}\sum_{\substack{x_j<n_j\le x_j+\sqrt{x_j}\\ P^-(n_j)>z}} 1 + O\left(\frac{1}{z^{100}(\log z)^k\sqrt{x_1\cdots x_k}}\right)\\ &= I\cdot\left(1+O\left(\max_{1\le j\le k}(x_j^{-1/\log z})\right)\right)\cdot\prod_{p\le z}\left(1-\frac{1}{p}\right)^k\\ &\quad + O\left(\frac{1}{z^{100}(\log z)^k\sqrt{x_1\cdots x_k}}\right)\\ &= I\cdot\prod_{p\le z}\left(1-\frac{1}{p}\right)^k + O\left(\frac{(\log z)^{-k}}{\sqrt{x_1\cdots x_k}}\left(\frac{1}{z^{100}}+\max_{1\le j\le k}(x_j^{-1/\log z})\right)\right)\end{aligned}$$

by the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.1). Hence, summing over all cubes $B\subset\{\boldsymbol{n}: n_1\cdots n_k\le D,\, n_j>z^{100}\,(1\le j\le k)\}$ completes the proof of the lemma. □

Proof of Lemma 7.1.3. Note that our choice of $\xi_{\boldsymbol{a}}$ and an upper bound sieve imply that

$$M=\max_{\boldsymbol{d}}|\lambda_{\boldsymbol{d}}|\ll(\log N)^k.$$

So, if we set

$$g(a_1,\dots,a_k)=f\left(\frac{\log a_1}{\log D},\dots,\frac{\log a_k}{\log D}\right)$$

and

$$V=\prod_{p\le z}\left(1-\frac{1}{p}\right),$$

then Lemmas 7.1.1 and 7.1.4, and our assumption that $(\log N)^{6k+2}\le z\le e^{\sqrt{\log N}}$ imply that

$$\begin{aligned}\sum_{\substack{N<n\le 2N\\ P^-(Q(n))>z}}\left(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2 &= \frac{N}{V^{2k}}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\sum_{P^-(a_1\cdots a_k)>z}\frac{g(\boldsymbol{a})^2}{a_1\cdots a_k}+O\left(\frac{N}{(\log N)^{k+1}}\right)\\ &= \frac{N}{V^k}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\left(\int\frac{g(\boldsymbol{u})^2}{u_1\cdots u_k}d\boldsymbol{u}+O((\log N)^{k-1/2})\right).\end{aligned}$$

Moreover,

$$V^{-k}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)=\mathfrak{S}(\boldsymbol{s})\left(1+O\left(\frac{1}{z}\right)\right) \tag{7.1.6}$$

and

$$\int\frac{g(\boldsymbol{u})^2}{u_1\cdots u_k}d\boldsymbol{u}=(\log D)^k\int_{T_k}f(\boldsymbol{t})^2d\boldsymbol{t},$$

so that

$$\sum_{\substack{N<n\le 2N\\ P^-(Q(n))>z}}\left(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2=\mathfrak{S}(\boldsymbol{s})N(\log D)^k\left(\int f(\boldsymbol{t})^2d\boldsymbol{t}+O\left(\frac{1}{\sqrt{\log N}}\right)\right). \tag{7.1.7}$$

Next, if we set

$$S_{j_0}(\boldsymbol{a})=\sum_{P^-(b)>z}\frac{\mu^2(b)g(a_1,\dots,a_{j_0-1},b,a_{j_0+1},\dots,a_k)}{b},$$

where $j_0\in\{1,\dots,k\}$, then Lemma 7.1.2 and relations (7.1.5) and (7.1.6) imply that

$$\begin{aligned}\sum_{\substack{N<n\le 2N\\ P^-(Q(n))>z}}\mathbf{1}_{\mathbb{P}}(n+s_{j_0})\left(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2&=\frac{X}{V^{2k+1}}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\sum_{\substack{P^-(a_1\cdots a_k)>z\\ a_{j_0}=1}}\frac{S_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}\\&\quad+O\left(\frac{N}{(\log N)^{k+1}}\right)\\&=\frac{\mathfrak{S}(\boldsymbol{s})N}{V^{k+1}\log N}\left(\sum_{\substack{P^-(a_1\cdots a_k)>z\\ a_{j_0}=1}}\frac{S_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}+O\left(\frac{1}{\log N}\right)\right).\end{aligned}$$

If $\mu^2(b)=0$ and $P^-(b)>z$, then b is divisible by the square of a prime $>z$. Therefore,

$$\begin{aligned}S_{j_0}(\boldsymbol{a})&=\sum_{P^-(b)>z}\frac{g(a_1,\dots,a_{j_0-1},b,a_{j_0+1},\dots,a_k)}{b}+O\left(\frac{\log N}{z}\right)\\&=V\cdot\left(\int\frac{g(a_1,\dots,a_{j_0-1},u,a_{j_0+1},\dots,a_k)}{u}du+O\left(\sqrt{\log N}\right)\right)\\&=V\cdot(\log D)\cdot\left(G_{j_0}(a_1,\dots,a_k)+O\left(\frac{1}{\sqrt{\log N}}\right)\right)\end{aligned}$$

by Lemma 7.1.3, where

$$G_{j_0}(a_1,\dots,a_k):=\int g(a_1,\dots,a_{j_0-1},D^{t_{j_0}},a_{j_0+1},\dots,a_k)dt_{j_0}\ll 1.$$

Therefore

$$\sum_{\substack{N<n\le 2N\\ P^-(Q(n))>z}}\mathbf{1}_{\mathbb{P}}(n+s_{j_0})\left(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2=\frac{\mathfrak{S}(\boldsymbol{s})N(\log D)^2}{V^{k-1}\log N}\left(\sum_{\substack{P^-(a_1\cdots a_k)>z\\ a_{j_0}=1}}\frac{G_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}+O\left(\frac{1}{\sqrt{\log N}}\right)\right).$$

Finally, applying again Lemma 7.1.4, we find that

$$\sum_{\substack{P^-(a_j)>z\\1\le j\le k\\a_{j_0}=1}}\frac{G_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}=V^{k-1}\left(\int\limits_{\substack{\boldsymbol{u}\in\mathbb{R}^k\\u_{j_0}=1}}\frac{G_{j_0}(\boldsymbol{u})^2}{u_1\cdots u_k}\prod_{\substack{1\le j\le k\\j\neq j_0}}du_j+O((\log N)^{k-3/2})\right)$$

$$=V^{k-1}(\log D)^{k-1}\left(\int\left(\int f(\boldsymbol{t})dt_{j_0}\right)^2\prod_{\substack{1\le j\le k\\j\neq j_0}}dt_j+O\left(\frac{1}{\sqrt{\log N}}\right)\right),$$

which completes the proof of Lemma 7.1.3. □

In view of Lemma 7.1.3, it is clear that our goal is to choose f supported on T_k and maximizing the ratio

$$\rho(f):=\frac{1}{k}\sum_{j=1}^{k}\frac{\int\left(\int f(\boldsymbol{t})dt_j\right)^2dt_1\cdots dt_{j-1}dt_{j+1}\cdots dt_k}{\int f(\boldsymbol{t})^2d\boldsymbol{t}}.$$

If we can show that, for k large enough, $\rho(f)>4m/k$, then we deduce that $\liminf_{n\to\infty}(p_{n+m}-p_n)<\infty$.

As a warm-up, using calculus of variations, we show that the maximiser of $\rho(f)$ is an eigenvector of the linear operator

$$(\mathcal{L}_kf)(\boldsymbol{t}):=\frac{1}{k}\sum_{j=1}^{k}\int f(t_1,\dots,t_{j-1},u,t_{j+1},\dots,t_k)du$$

and the corresponding eigenvalue is the maximum possible ratio $\rho=\rho(f)$. Indeed, for a function f supported on T_k, we have that

$$\rho(f)=\frac{\langle\mathcal{L}_kf,f\rangle}{\langle f,f\rangle},$$

where

$$\langle g,h\rangle:=\int_{T_k}g(\boldsymbol{t})h(\boldsymbol{t})d\boldsymbol{t}.$$

If, now, f is a maximiser of the function $\rho(\cdot)$, then the function $\epsilon\to\rho(f+\epsilon g)$ has a maximum at $\epsilon=0$ for any smooth $g:\mathbb{R}^k\to\mathbb{R}$ supported on T_k. So its derivative at $\epsilon=0$ must vanish, which implies that

$$\langle\mathcal{L}_kf,g\rangle+\langle\mathcal{L}_kg,f\rangle=2\rho(f)\langle f,g\rangle.$$

It is easy to see that $\mathcal{L}_k$ is a self-adjoint operator, so this implies that

$$\langle\mathcal{L}_kf,g\rangle=\rho(f)\langle f,g\rangle.$$

Taking $g(u_1,\dots,u_k)$ to be a smooth approximation to the function $\mathbf{1}_{B_n}(u_1,\dots,u_k)/\operatorname{Vol}(B_n)$ for a shrinking family of k-dimensional cubes $(B_n)_{n\geq 1}$ centered at a fixed point $\boldsymbol{t}$, we deduce that $\mathcal{L}_k f = \rho(f)\cdot f$, as claimed.

Now, note that the symmetric function

$$\widetilde{f}(t_1,\dots,t_k) := \sum_{\sigma\in S_k} f(t_{\sigma(1)},\dots,t_{\sigma(k)})$$

is also an eigenvalue of the operator $\mathcal{L}_k$ of eigenvalue $\rho(f)$. Therefore, $\rho(\widetilde{f}) = \rho(f)$, which means that $\widetilde{f}$ is also a maximiser for the function $\rho(\cdot)$.

In view of the above discussion, we may restrict our attention to symmetric functions f, in which case

$$\rho(f) = \frac{\int \left(\int f(\boldsymbol{t})dt_k\right)^2 dt_1\cdots dt_{k-1}}{\int f(\boldsymbol{t})^2 d\boldsymbol{t}}.$$

Moreover, we may also drop the assumption that f is smooth, since the integral of every measurable function can be approximated well-enough by integrals of smooth functions. So our goal becomes to estimate

$$M_k := \sup\{\rho(f) : f:\mathbb{R}^k\to\mathbb{R},\ \operatorname{supp}(f)\subset T_k,\ f \text{ symmetric and measurable}\}$$

An asymptotic estimation for M_k is given in Lemma 7.1.6 below. For explicit bounds on M_k, the reader is invited to consult the paper by Maynard [May15a] as well [Pol].

Remark 7.1.5. The original weights of Goldston, Pintz and Yildirim essentially correspond to taking the supremum over the restricted set of functions of the form $f(\boldsymbol{t}) = F(t_1+\cdots+t_k)$, where F is supported on $[0,1]$. Then we have that

$$\rho(f) = (k-1)\cdot\frac{\int_0^1 u^{k-2}(\int_u^1 F)^2 du}{\int_0^1 u^{k-1}F(u)^2 du}.$$

It is possible to show that $\rho(f)\leq 4/k$ in this special case [Sou], which means that the weights of Goldston, Pintz and Yildirim cannot yield bounded gaps between primes. On the other hand, choosing $F(t) = (1-t)^\ell$, we find that

$$\begin{aligned}\rho(f) = \frac{k-1}{(\ell+1)^2}\cdot\frac{\int_0^1 u^{k-2}(1-u)^{2\ell+2}du}{\int_0^1 u^{k-1}(1-u)^{2\ell}du} &= \frac{k-1}{(\ell+1)^2}\cdot\frac{\dfrac{(k-2)!(2\ell+2)!}{(k+2\ell+1)!}}{\dfrac{(k-1)!(2\ell!)}{(k+2\ell)!}}\\ &= \frac{2(2\ell+1)}{(\ell+1)(k+2\ell+1)} \sim \frac{4}{k}\end{aligned}$$

if $\ell = o(k)$ and $\ell, k\to\infty$. This means that if we could have inserted a slightly stronger input to the computations in Lemma 7.1.2, which would have allowed to take D slightly larger, we would have been able to prove Theorem 7.0.2 when $m=1$ with these weights. This is precisely what Zhang did. As we will see in the lemma below, using the higher dimensional structure of f allow us to show that $\rho(f)$ can get much bigger.

Lemma 7.1.6. *For large integers k, we have that*

$$\frac{\log k - 4\log\log k + O(1)}{k} \le M_k \le \frac{\log k + \log\log k + O(1)}{k}$$

Proof. For the lower bound, we consider functions of the form

$$f(t_1, \dots, t_k) = \mathbf{1}_{T_k}(t_1, \dots, t_k) \prod_{j=1}^{k} g(kt_j),$$

where $g : [0, +\infty) \to [0, +\infty)$ is a function supported on the interval $[0, T]$ and such that $\int_0^\infty g(t)^2 dt = 1$. Then

$$\int f(\boldsymbol{t})^2 d\boldsymbol{t} \le \prod_{j=1}^{k} \int_0^\infty g(kt_j)^2 dt_j = \frac{1}{k^k},$$

so that

$$\begin{aligned}
\rho(f) &\ge \frac{1}{k} \int g(t_1)^2 \cdots g(t_{k-1})^2 \left(\int_0^{k-t_1-\cdots-t_{k-1}} g(t_k) dt_k \right)^2 dt_1 \cdots dt_{k-1} \\
&\ge \frac{\left(\int_0^\infty g(t)dt\right)^2}{k} \int_{t_1+\cdots+t_{k-1} \le k-T} g(t_1)^2 \cdots g(t_{k-1})^2 dt_1 \cdots dt_{k-1} \\
&= \frac{\left(\int_0^\infty g(t)dt\right)^2}{k} \cdot \mathbf{Prob}\left(X_1 + \cdots + X_{k-1} \le k - T\right),
\end{aligned}$$

where $X_1, \dots, X_{k-1}$ are independent random variables with density function g^2. Let

$$\mu = \mathbb{E}[X_1] = \int t g(t)^2 dt$$

and $Y_i = X_i - \mu$, $1 \le i \le k$, so that $Y_1, \dots, Y_k$ are mean-zero independent random variables that are identically distributed. If we assume that $(k-1)\mu < k - T$, then Chebyshev's inequality implies that

$$\begin{aligned}
\mathbf{Prob}\left(X_1 + \cdots + X_{k-1} > k - T\right) &= \mathbf{Prob}\left(Y_1 + \cdots + Y_{k-1} > k - T - (k-1)\mu\right) \\
&\le \frac{1}{(k - T - (k-1)\mu)^2} \operatorname{Var}[Y_1 + \cdots + Y_{k-1}] \\
&= \frac{(k-1)\operatorname{Var}[Y_1]}{(k - T - (k-1)\mu)^2} \le \frac{k\mathbb{E}[X_1^2]}{(k - T - (k-1)\mu)^2}
\end{aligned}$$

Since

$$\mathbb{E}[X_1^2] = \int t^2 g(t)^2 dt \le T \int t g(t)^2 dt = T\mu$$

by our assumption that g is supported in $[0, T]$, we deduce that

$$\rho(f) \ge \frac{\left(\int_0^\infty g(t)dt\right)^2}{k} \cdot \left(1 - \frac{kT\mu}{(k - T - k\mu)^2}\right)$$

for any measurable function $g \geq 0$ supported on $[0,T]$ with $\int g(t)^2 dt = 1$ and $\mu = \int tg(t)^2 dt \leq 1 - T/k$. We choose

$$g(t) = c \cdot \frac{\mathbf{1}_{[0,T]}(t)}{1+At}.$$

(See Remark 7.1.7 for an explanation behind this choice of g.) In order to have that $\int g(t)^2 dt = 1$, we take

$$c^2 = \left(\int_0^T \frac{dt}{(1+At)^2} \right)^{-1} = \frac{A}{1 - 1/(1+AT)} = A + \frac{1}{T}.$$

We then have that

$$\begin{aligned} \mu = \int_0^T \frac{c^2 t}{(1+At)^2} dt = \frac{c^2}{A^2} \int_0^{AT} \frac{t}{(1+t)^2} dt &= \frac{c^2}{A^2} \left(\log(1+AT) - 1 + \frac{1}{1+AT} \right) \\ &= \frac{\log(AT)}{A} \left(1 + O\left(\frac{1}{\log(AT)} \right) \right). \end{aligned}$$

This suggests choosing $A \sim \log T$. We take $T = k/(\log k)^3$ and $A = \log k$, so that

$$\mu = \frac{\log(k/(\log k)^2)}{\log k} \left(1 + O\left(\frac{1}{\log k} \right) \right) = 1 - \frac{2 \log\log k}{\log k} + O\left(\frac{1}{\log k} \right) \leq 1 - \frac{T}{k} - \frac{\log\log k}{\log k}$$

for k large enough. In particular,

$$\frac{kT\mu}{(k - T - k\mu)^2} = \frac{T\mu}{k(1 - T/k - \mu)^2} \ll \frac{1}{\log k}$$

and therefore

$$\begin{aligned} k \cdot \rho(f) \geq \left(\int_0^\infty g(t) dt \right)^2 \cdot \left(1 - \frac{T\mu}{k(1-T/k-\mu)^2} \right) &= \frac{c^2 \log^2(1+AT)}{A^2} \left(1 + O\left(\frac{1}{\log k} \right) \right) \\ &= \frac{\log^2(k/(\log k)^2)}{\log k} \left(1 + O\left(\frac{1}{\log k} \right) \right) \\ &= \log k - 4 \log\log k + O(1), \end{aligned}$$

as claimed.

Finally, we prove the upper bound on $\rho(f)$. Let f be a symmetric measurable function supported on T_k. Motivated by the choice for f above, we use the Cauchy-Schwarz inequality in the following fashion:

$$\begin{aligned} \left(\int f(t_1, \dots, t_k) dt_k \right)^2 &= \left(\int_0^1 f(t_1, \dots, t_k) dt_k \right)^2 \\ &\leq \left(\int (1 + kAt_k) f(t_1, \dots, t_k)^2 dt_k \right) \cdot \left(\int_0^1 \frac{dt_k}{1+kAt_k} \right) \\ &= \frac{\log(1+kA)}{kA} \int (1 + kAt_k) f(t_1, \dots, t_k)^2 dt_k. \end{aligned}$$

Therefore

$$\int \left(\int f(t_1, \dots, t_k) dt_k \right)^2 dt_1 \cdots dt_{k-1} \leq \frac{\log(1+kA)}{kA} \int (1 + kAt_k) f(t_1, \dots, t_k)^2 d\boldsymbol{t}.$$

By symmetry,

$$\int \left(\int f(t_1, \dots, t_k) dt_k \right)^2 dt_1 \cdots dt_{k-1} \leq \frac{\log(1+kA)}{kA} \int (1 + kAt_j) f(t_1, \dots, t_k)^2 d\boldsymbol{t}$$

for all $j \in \{1, \dots, k\}$. So, summing over j and using the fact that $t_1 + \cdots + t_k \leq 1$ in the support of f, we find that

$$\begin{aligned} k \cdot \rho(f) &\leq \frac{\log(1+kA)}{kA} \cdot \frac{\int_{T_k} (k + kA(t_1 + \cdots + t_k)) f(t_1, \dots, t_k)^2 d\boldsymbol{t}}{\int f(\boldsymbol{t})^2 d\boldsymbol{t}} \\ &\leq \frac{(1+A)\log(1+kA)}{A}. \end{aligned}$$

for any $A > 0$. So, setting $A = \log k$ yields that

$$k \cdot \rho(f) \leq \log k + \log\log k + O(1)$$

for all symmetric functions f supported on T_k. This completes the proof of the lemma. □

Remark 7.1.7.

It is now easy to complete the proof of Theorem 7.0.2:

Proof of Theorem 7.0.2. Combining Lemmas 7.1.3 and 7.1.6, we find that there is a choice of the parameters $\lambda_{\boldsymbol{d}}$ such that

$$\frac{S(N,z)}{\mathfrak{S}(\boldsymbol{s}) N (\log D)^k} \geq \frac{\log k - 4\log\log k + O(1) - 4m}{4} + O\left(\frac{1}{\sqrt{\log N}}\right).$$

So, if $k = \lfloor Cm^4 e^{4m} \rfloor$ for a large enough constant C, then $S(N,z) > 0$ for large enough N, which implies that $\liminf_{n\to\infty}(p_{n+m} - p_n) \leq s_k - s_1$. We take s_j to be the j-th prime that is $> k$, which clearly form an admissible set. Then $s_k \lesssim k \log k \ll e^{4m} m^5$ by the Prime Number Theorem, which completes the proof of Theorem 7.0.2. □

7.2 Large gaps between primes

This section is devoted to the proof of Theorem 7.0.5. The construction of large gaps between primes is based on the following lemma.

Lemma 7.2.1. *Let $N \geq 1$, and assume that there is $z \geq 3$ and progressions $a_p \pmod p$, $p < z$, such that if $n \leq N$, then $n \equiv a_p \pmod p$, for some $p \leq y$. Then there exists $x \in (P(z), 2P(z)]$ for which there are no primes in $(x, x+N]$.*

Proof. Consider $x \in \{P(z)+1, P(z)+2, \dots, 2P(z)\}$ such that $x \equiv -a_p \pmod p$, for all $p < z$. If $n \in [1, N] \cap \mathbb{Z}$, then $n \equiv a_p \pmod p$, for some $p < z$, that is to say, there exists a prime $p < z$ that divides $x + n$. In particular, $x + n$ cannot be a prime number, for all $1 \le n \le N$. This completes the proof of the lemma. □

We are now in position to show Theorem 7.0.4.

Proof of Theorem 7.0.4. Let $z \ge 1$, and $N \in [z, z^2]$. To each prime $p < z$, we will assign a congruence $a_p \pmod p$ such that $\{n \le N\} \subset \bigcup_{p<z} a_p \pmod p$, as in Lemma 7.2.1. Therefore, in order to show Theorem 7.0.4, we need to be able to take

$$z \sim \frac{e^{-\gamma} N (\log\log N)^2}{(\log N)(\log\log\log N)},$$

so that

$$N \sim \frac{e^{\gamma} z (\log z)(\log\log\log z)}{(\log\log z)^2} \sim \frac{e^{\gamma} (\log X)(\log\log X)(\log\log\log\log X)}{(\log\log\log X)^2}$$

with $X = P(z)$.

We will choose $a_p \pmod p$ using different arguments, according to whether $p \le N^{1/u}$, $N^{1/u} < p \le N^{1-1/M}$ or $N^{1-1/M} < p \le z$, where u and M are defined by

$$u^u = \log N \quad \Longrightarrow \quad u \sim \frac{\log\log N}{\log\log\log N} \tag{7.2.1}$$

and

$$N^{1/M} = \frac{\log N}{\log\log N} \quad \Longrightarrow \quad M \sim \frac{\log N}{\log\log N}. \tag{7.2.2}$$

Intermediate primes: When $p \in (N^{1/u}, N^{1-1/M}]$, we choose $a_p = 0$. Call S the set of integers $n \le N$ which do not belong to any of the congruence classes $a_p \pmod p$, $p \in (N^{1/u}, N^{1-1/M}]$. Clearly,

$$S = \{n \le N : P^+(n) \le N^{1/u}\} \cup \{n \le N : \exists p | n \text{ with } p > N^{1-1/M}\},$$

and consequently

$$\begin{aligned} |S| \le \Psi(N, N^{1/u}) + \sum_{N^{1-1/M} < p \le N} \frac{N}{p} &\le \frac{N e^{O(u)}}{(u \log u)^u} + \frac{N}{M} + O\left(\frac{N}{(\log N)^2}\right) \\ &= \frac{N}{M} + O\left(\frac{N}{(\log N) e^{\sqrt{\log\log N}}}\right) \sim \frac{N \log\log N}{\log N} \end{aligned}$$

by Theorem 6.2.1, the Prime Number Theorem and our choice of u and M.

Small primes: For each $p \le N^{1/u}$, we select the progressions $a_p \pmod p$ "greedily": we let $a_2 \pmod 2$ be such that

$$\#\{n \in S : n \equiv a_2 \pmod 2\} = \max_{j \in \{1,2\}} \#\{n \in S : n \equiv j \pmod 2\}.$$

Having chosen a_2, we set $S_2 = \{n \in S : n \not\equiv a_2 \pmod p\}$, and we select $a_3 \pmod 3$ such that

$$\#\{n \in S_2 : n \equiv a_3 \pmod 3\} = \max_{j\in\{1,2,3\}} \#\{n \in S_2 : n \equiv j \pmod 3\},$$

and we set $S_3 = \{n \in S_2 : n \not\equiv a_3 \pmod 3\}$. Continuing this way, we find that there are progressions $a_p \pmod p$, $p \le N^{1/u}$, such that the set

$$S' := \{n \in S : \exists p \le N^{1/u} \text{ for which } n \not\equiv a_p \pmod p\}$$

has cardinality

$$|S'| \le |S| \cdot \prod_{p\le N^{1/u}} \left(1 - \frac{1}{p}\right) \lesssim \frac{e^{-\gamma} u N \log\log N}{(\log N)^2} \sim e^{-\gamma} \cdot \frac{N(\log\log N)^2}{(\log N)^2(\log\log\log N)}. \tag{7.2.3}$$

Large primes: Finally, to each $n \in S'$, we assign a prime $p \in (N^{1-1/M}, z]$ and the arithmetic progression $a_p \pmod p$ in which n lies modulo p. In order to be able to assign to each n a different prime $p \in (N^{1-1/M}, z)$, we need to have that $|S'| \le \pi(z) - \pi(N^{1-1/M}) \sim z/\log z$, since $N^{1-1/M} = o(z)$. In view of (7.2.3), this reduces to knowing that

$$\frac{z}{\log z} \gtrsim e^{-\gamma} \cdot \frac{N(\log\log N)^2}{(\log N)^2(\log\log\log N)} \quad \Leftrightarrow \quad z \gtrsim e^{-\gamma} \cdot \frac{N(\log\log N)^2}{(\log N)(\log\log\log N)}.$$

So, choosing

$$z = (e^{-\gamma} + \epsilon) \cdot \frac{N(\log\log N)^2}{(\log N)(\log\log\log N)}$$

for some positive $\epsilon = \epsilon(N)$ tending to 0 sufficient slowly as $N \to \infty$ completes the proof of Theorem 7.0.4. □

7.3 Even larger gaps between primes

The proof of Theorem 7.0.5 has the same general structure as the proof of Theorem 7.0.4, but now we need to show that we can take

$$z = \frac{1}{C} \cdot \frac{N(\log\log N)^2}{(\log N)(\log\log\log N)}$$

in Lemma 7.2.1, where C is a fixed but arbitrarily large constant. Again, we choose $a_p = 0$ for the "intermediate" residue classes $p \in (N^{1/u}, N^{1-1/M}]$, where u and M are defined by (7.2.1) and (7.2.2), respectively. Moreover, for $p \le N^{1/u}$, we choose $a_p = 1$. This has essentially the same effect as choosing the a_p's greedily because if n is an integer with no prime factors in $(N^{1/u}, N^{1-1/M}]$, then $n - 1$ looks like a 'random' integer. After these first steps have been performed, we are left with the set of integers $\mathcal{N}_1 \cup \mathcal{N}_2$, where

$$\mathcal{N}_1 := \{n \le N : P^+(n) \le N^{1/u},\ P^-(n-1) > y\}$$

and

$$\mathcal{N}_2 := \{mq \le N : m \le N^{1/M},\ q > N^{1-1/M} \text{ prime},\ P^-(mq-1) > N^{1/u}\}.$$

(Here and for the rest of this section, the letter q will always denote a prime number, and the same will be true later on for the letter ℓ.) Recall that $N^{1/M} = (\log N)/\log\log N$. We further write $\mathcal{N}_2$ as a disjoint union $\mathcal{N}_2' \cup \mathcal{N}_2''$, where

$$\mathcal{N}_2' := \{mq \le N : N^{1/M}/\log\log N < m \le N^{1/M},\ q > N^{1-1/M},\ P^-(mq-1) > N^{1/u}\}.$$

As in the proof of Theorem 7.0.5, our choice of u implies that

$$|\mathcal{N}_1| \ll \frac{N}{e^{\sqrt{\log\log N}}} = o\left(\frac{z}{\log z}\right).$$

Moreover, for an even integer $m \le N^{1/M}$, set

$$\mathcal{Q}_m := \{q \le N/m : P^-(aq-1) > N^{1/u}\},$$

so that S_2 is the disjoint union of the sets $\{mq : q \in \mathcal{Q}_m\}$, $m \le N^{1/M}$, with $\mathcal{N}_2'$ consisting of those with $N^{1/M}/\log\log N < m \le N^{1/M}$. An upper bound sieve implies that

$$|\mathcal{Q}_m| \ll \frac{uN}{\varphi(m)(\log N)^2} \sim \frac{N\log\log N}{\varphi(m)(\log N)^2\log\log\log N},$$

so that

$$\begin{aligned}|\mathcal{N}_2'| &\ll \frac{N\log\log N}{(\log N)^2\log\log\log N} \sum_{N^{1/M}/\log\log N < m \le N^{1/M}} \frac{1}{\varphi(m)} \\ &\ll \frac{N\log\log N}{(\log N)^2\log\log\log N} \cdot \log\log\log N = o\left(\frac{z}{\log z}\right).\end{aligned}$$

So we may pick residue classes $a_p \pmod p$ for the primes $p \in (N^{1-1/M}, z/2]$ to cover $\mathcal{N}_1 \cup \mathcal{N}_2'$. We are then left with the challenge of covering S_2' using residue classes $a_p \pmod p$ for the primes $p \in (z/2, z]$.

Note that if $m \le N^{1/M}/\log\log N$, then

$$\frac{N}{m} - N^{1-1/M} = \frac{N(1-m/N^{1/M})}{m} \sim \frac{N}{m}.$$

So the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.1) and the Bombieri-Vinogradov theorem imply that

$$\begin{aligned}|\mathcal{Q}_m| &\sim c\cdot e^{-\gamma}\cdot\left(\prod_{\substack{p\mid m\\ p>2}}\frac{p-1}{p-2}\right)\cdot\frac{N/m}{(\log N^{1/u})(\log(N/m))} \\ &\sim c\cdot e^{-\gamma}\cdot\left(\prod_{\substack{p\mid m\\ p>2}}\frac{p-1}{p-2}\right)\cdot\frac{N}{m(\log N)^2}\cdot\frac{\log\log N}{\log\log\log N}\end{aligned}$$

for even integers $m \leq N^{1/M}/\log\log N$, where c is the twin prime constant, that is to say

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

In particular, writing $m = 2n$, we find that

$$\begin{aligned}|\mathcal{N}_2'| &\sim \frac{ce^{-\gamma}}{2}\cdot\frac{N\log\log N}{(\log N)^2\log\log\log N}\sum_{n\leq N^{1/M}/(2\log\log N)}\frac{1}{n}\prod_{\substack{p\mid n\\ p>2}}\frac{p-1}{p-2}\\ &\sim \frac{e^{-\gamma}}{2}\cdot\frac{N\log\log N}{M(\log N)(\log\log\log N)}\\ &\sim \frac{e^{-\gamma}}{2}\cdot\frac{N(\log\log N)^2}{(\log N)^2(\log\log\log N)}\\ &\sim \frac{C}{2e^{\gamma}}\cdot\frac{z}{\log z},\end{aligned}$$

where the sum over n is estimate using the convolution method (see, for example, the proof of Theorem 0.2.1 and Exercise 0.2.2). Therefore, we need to choose the residue classes $a_p \pmod p$ for $p \in (z/2, z]$ in a way that, on average, each one will cover $> C/e^{\gamma}$ elements of $\mathcal{N}_2'$. In order to do so, we will use the Maynard-Tao sieve weights from Section 7.1 to construct a probability measure on $\prod_{z/2<p\leq z}\mathbb{Z}/p\mathbb{Z}$ that will be biased on choices of residue classes $a_p \pmod p$, $z/2 < p \leq z$, that each cover many elements of $\mathcal{N}_2'$ simultaneously. Here is the key result:

Proposition 7.3.1. *Fix $\eta > 0$ and let $m \leq N^{1/M}/\log\log N$ be even. If $I_m \subset [z/2, z]$ is an interval whose length is between $\eta|\mathcal{Q}_m|\log N$ and $2\eta|\mathcal{Q}_m|\log N$, and $C_k > k^5$, then there there are choices of residue classes $a_p \pmod p$, $p \in I_m$, whose union covers $\mathcal{Q}_m$.*

Theorem 7.0.5 is now a direct corollary of Proposition 7.3.1 and of the above discussion. As we mentioned before, our goal is to construct a probability measure on the space $\prod_{p\in I_m}\mathbb{Z}/p\mathbb{Z}$. Indeed, we set

$$s_j = p_{\pi(C_k)+j}\prod_{p\leq C_k} p,$$

where C_k is a large auxiliary integer to be chosen later. For the convenience of notation, we also set

$$s_0 = 0,$$

and we assume that $C_k > 2k$, so that the polynomial

$$Q_{p,m}(x) := \prod_{j=0}^{k}[(x+ps_j)(m(x+ps_j)-1)]$$

does not have a fixed prime divisor (recall that m is even here). We further decompose $Q_{p,m} = Q_{p,m}^{(1)}\cdot Q_{p,m}^{(2)}$, where

$$Q_{p,m}^{(1)}(x) := \prod_{j=0}^{k}(m(x+ps_j)-1) \quad \text{and} \quad Q_{p,m}^{(2)}(x) := \prod_{j=0}^{k}(x+ps_j).$$

We also set

$$w = e^{\sqrt{\log N}}$$

and fix two upper bound sieves $(\mu_j(d))_{d\geq 1}$ such that μ_1 has dimension $k+1$, μ_2 has dimension 1, and

$$\begin{aligned}\mathrm{supp}(\mu_1) &\subset \{d \leq N^{1/100} : p|d \implies w < p \leq N^{1/u}\},\\ \mathrm{supp}(\mu_2) &\subset \{d \leq N^{1/100} : p|d \implies N^{1/u} < p \leq N^{1/1000}\}.\end{aligned}$$

Finally, we let $\lambda_{\boldsymbol{d}}$ be some sieve parameters supported on k-tuples $\boldsymbol{d} = (d_1, \dots, d_k)$ with $d_1 \cdots d_k \leq N^{1/100}$. Eventually, we will take $\lambda_{\boldsymbol{d}}$ exactly as in Section 7.0.2. Then we define the probability density function

$$\delta_{p,m}(a) := \frac{1}{\Delta_{p,m}} \sum_{\substack{n\in N/m\\ P^-(Q_{p,m}(n))>w\\ n\equiv a \,(\mathrm{mod}\, p)}} (1*\mu_1)(Q^{(1)}_{p,m}(n)) \cdot (1*\mu_2)(n) \left(\sum_{\substack{d_j|n+ps_j\\ 1\leq j\leq k}} \lambda_{\boldsymbol{d}}\right)^2,$$

with $\Delta_{p,m}$ being the normalizing factor

$$\Delta_{p,m} := \sum_{\substack{n\leq N/m\\ P^-(Q_{p,m}(n))>w}} (1*\mu_1)(Q^{(1)}_{p,m}(n)) \cdot (1*\mu_2)(n) \left(\sum_{\substack{d_j|n+ps_j\\ 1\leq j\leq k}} \lambda_{\boldsymbol{d}}\right)^2.$$

Then the probability measure on $\prod_{p\in I_m} \mathbb{Z}/p\mathbb{Z}$ is simply defined by

$$\delta_m((a_p)_{p\in I_m}) := \prod_{p\in I_m} \delta_{p,m}(a).$$

The choice of the parameters $\lambda_{\boldsymbol{d}}$ will be similar to the one leading to the proof of Theorem 7.0.2: we choose them so that many of the numbers $n+ps_j$ are biased towards being primes, while the weight $\mathbf{1}_{\mathcal{N}_{p,m}}(n)(1*\mu_2)(Q^{(1)}_{p,m}(n))(1*\mu_3)(n)$ guarantees that our sum is essentially supported on integers n for which $mn-1, m(n+ps_1)-1, \dots, mn(n+ps_k)-1$ have no prime factors $\leq N^{1/u}$ and n has no primes $\leq N^{1/1000}$ (so, there is a positive probability it will be prime). We then have the following crucial estimate.

Lemma 7.3.2. *There are choices of $\lambda_{\boldsymbol{d}}$ such that the following holds: if m, η and I_m are as in the statement of Proposition 7.3.1, $q \in \mathcal{Q}_m \cap (s_k z, \infty)$ and N is large enough in terms of k and η, then there are absolute constants $c, c' > 0$ such that*

$$\sum_{p\in I_m} \delta_{p,m}(q) \geq c\eta \log k - c' \frac{\eta k^3 \log k}{C_k} \cdot e^{G_m(q)},$$

where

$$G_m(q) := \sum_{\substack{1\leq j,j',j''\leq k\\ j,j',j'' \text{ distinct}}} \sum_{\substack{C_k<\ell\leq (\log N)^2\\ \ell|(qm(s_j-s_{j'})-(s_j-s_{j''}))}} \frac{2k}{\ell}. \tag{7.3.1}$$

Before we turn to the proof of Lemma 7.3.2, let us see how it implies Proposition 7.3.1.

Deduction of Proposition 7.3.1 from Lemma 7.3.2. Let $\mathcal{Q}'_m$ be the set of $q \in \mathcal{Q}_m$ such that $G_m(q) \le 1$ and $q > s_k z$. Markov's inequality and the sieve imply that

$$\begin{aligned}
|\mathcal{Q}_m \setminus \mathcal{Q}'_m| &\le \sum_{\substack{q \le N/m \\ P^-(mq-1) > N^{1/u}}} G_m(q) + \#\{q \le s_k z : P^-(mq-1) > N^{1/u}\} \\
&\ll \sum_{\substack{1 \le j,j',j'' \le k \\ j,j',j'' \text{ distinct}}} \sum_{C_k < \ell \le (\log N)^2} \frac{k}{\ell} \cdot \#\left\{ q \le \frac{N}{m} : \begin{array}{l} P^-(mq-1) > N^{1/u} \\ qm(s_j - s_{j'}) \equiv (s_j - s_{j''}) \pmod{\ell} \end{array} \right\} \\
&\qquad + \frac{u s_k z m}{\varphi(m)(\log N)^2} \\
&\ll \sum_{\substack{1 \le j,j',j'' \le k \\ j,j',j'' \text{ distinct}}} \sum_{C_k < \ell \le (\log N)^2} \frac{k}{\ell} \cdot \left(\frac{1}{\ell} + \mathbf{1}_{\ell \mid (s_j - s_{j'})(s_j - s_{j''})} \right) \cdot |\mathcal{Q}_m| + \frac{m s_k z}{N} \cdot |\mathcal{Q}_m| \\
&\ll \left(\frac{k^4 \log C_k}{C_k} + \frac{s_k}{\log\log\log N} \right) \cdot |\mathcal{Q}_m|
\end{aligned}$$

for all $m \le N^{1/M}/\log\log N$, where we used the fact that if a prime $\ell > C_k$ divides $(s_j - s_{j'})(s_j - s_{j''})$, then it must be one of the $O(\log C_k)$ divisors of the number $(p_{\pi(C_k)+j} - p_{\pi(C_k)+j'})(p_{\pi(C_k)+j} - p_{\pi(C_k)+j''})$.

Now, let J_m be the left half of the interval I_m and fix $q \in \mathcal{Q}'_m$. The probability that q is not covered by a random selection of congruences $(a_p)_{p \in J_m} \in \prod_{p \in J_m} \mathbb{Z}/p\mathbb{Z}$ is

$$\prod_{p \in J_m} (1 - \delta_{p,m}(q)) \le \exp\left\{ - \sum_{p \in J_m} \delta_{p,m}(q) \right\} < \frac{\eta}{5}$$

if N and k are large enough in terms of η, by Lemma 7.3.2. Hence, the expected cardinality of the random set

$$\mathcal{R}_m((a_p)_{p \in J_m}) := \left\{ q \in \mathcal{Q}'_m : q \notin \bigcup_{p \in J_m} \{a_p \pmod p\} \right\},$$

which is the uncovered part of $\mathcal{Q}'_m$, is $< \eta |\mathcal{Q}_m|/4$. This implies that there is a choice of $(a_p)_{p \in J_m}$ such that

$$|\mathcal{R}_m((a_p)_{p \in J_m})| < \frac{\eta |\mathcal{Q}_m|}{4} \le \frac{\mathrm{meas}(I_m)}{4 \log N} \sim \frac{\#\{p \in I_m \setminus J_m\}}{2},$$

where we used the Prime Number Theorem. Finally, we may use one congruence class per prime in $I_m \setminus J_m$ to cover the remaining set $(\mathcal{Q}_m \setminus \mathcal{Q}'_m) \cup \mathcal{R}((a_p)_{p \in J_m})$, provided that k is large enough in terms of η and that $C_k > k^5$. This completes the proof of Proposition 7.3.1. □

So, it remains to prove that we may choose the parameters $\lambda_{\boldsymbol{d}}$ in a way that will make Lemma 7.3.2 true. For each $p \in I_m$, we only look at the terms of the sum defining $\delta_{p,m}(q)$ with

$n = q - ps_j$ for some $j \in \{1, \dots, k\}$. Note that $q - ps_j \le q \le N/m$ and $q - ps_j \ge q - zs_k > 0$, so that the condition $n \in [1, N/m]$ is always satisfied for these integers, and the same is true for the condition $n \equiv q \pmod p$. Therefore

(7.3.2)
$$\sum_{p\in I_m} \delta_{p,m}(q) \ge \sum_{j=1}^{k} \sum_{\substack{p\in I_m \\ P^-(Q_{p,m}(q-ps_j))>w}} \frac{(1*\mu_1)(Q^{(1)}_{p,m}(q-ps_j))\cdot(1*\mu_2)(q-ps_j)}{\Delta_{p,m}} \left(\sum_{\substack{d_i \mid q+p(s_i-s_j) \\ 1\le i\le k}} \lambda_{\boldsymbol{d}} \right)^2.$$

So we see that we need an upper bound for $\Delta_{p,m}$ and a lower bound for the resulting sums over $p \in I_m$. Anticipating the choice of the parameters $\lambda_{\boldsymbol{d}}$, and in order to simplify the statements of the results, we make the change of variables

(7.3.3)
$$\frac{\xi_{\boldsymbol{a}}}{a_1 \cdots a_k} := \sum_{\substack{P^-(d_j)>w \\ d_j \equiv 0 \pmod{a_j} \\ 1\le j\le k}} \frac{\lambda_{\boldsymbol{d}}}{d_1 \cdots d_k}.$$

We further set

$$\xi_{\boldsymbol{a}} := \frac{\mathbf{1}_{P^-(a_1\cdots a_k)>w}}{\prod_{\ell\le w}(1-1/\ell)^k} \cdot \lambda(a_1 \cdots a_k) \cdot f\left(\frac{\log a_1}{\log(N^{1/100})}, \dots, \frac{\log a_k}{\log(N^{1/100})} \right),$$

where λ is Liouville's functions and f is a smooth function supported on the simplex

$$T_k = \{\boldsymbol{t} \in [0, +\infty)^k : t_1 + \cdots + t_k \le 1\}.$$

As in Section 7.1, we have the inversion formula (7.1.4), which also implies that

$$\max_{\boldsymbol{d}} |\lambda_{\boldsymbol{d}}| \ll_k (\log N)^k \cdot \sup_{\boldsymbol{t}} |f(\boldsymbol{t})| \ll_{k,f} (\log N)^k.$$

With this notation, we have the following two lemmas.

Lemma 7.3.3. *Let $m \le N$ and $p \in (z/2, z]$. If $C_k > e^k$ and N is large enough in terms of k and f, then*

$$\Delta_{p,m} \ll e^{H_{p,m}} \cdot |\mathcal{Q}_m| \cdot \left(\frac{e^\gamma (\log C_k)^2 u}{100} \right)^k \cdot \int f(\boldsymbol{t})^2 d\boldsymbol{t},$$

where the implied constant is absolute and

$$H_{p,m} = \sum_{\substack{1\le i,j\le k \\ i\ne j}} \sum_{\substack{C_k<\ell\le(\log N)^2 \\ \ell \mid pm(s_i-s_j)-1}} \frac{2k+2}{\ell}.$$

Proof. All implicit constants might depend on k, f and A, unless otherwise specified. Write $\mathcal{P}(s,t)$ for the set of integers all of whose prime factors are in the interval $(s,t]$. Opening

the square and the two convolutions $1 * \mu_2$ and $1 * \mu_3$ in the definition of $\Delta_{p,m}$, we find that

$$\Delta_{p,m} = \sum_{\substack{d_j,e_j \mid n+ps_j \\ (d_ie_i,d_je_jP(w))=1 \\ 1\le i,j\le k,\ i\ne j}} \sum_{\substack{f_1\in\mathcal{P}(w,N^{1/u}) \\ f_2\in\mathcal{P}(N^{1/u},N^{1/1000})}} \mu_1(f_1)\mu_2(f_2)\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}$$
$$\times \#\left\{ \begin{array}{l|l} & Q^{(1)}_{p,m}(n) \equiv 0 \,(\mathrm{mod}\, f_1), \\ n \le N/m & n \equiv 0 \,(\mathrm{mod}\, f_2), \\ P^-(Q_{p,m}(n)) > w & n + ps_j \equiv 0 \,(\mathrm{mod}\, [d_j, e_j]), \\ & 1 \le j \le k \end{array} \right\}.$$

If $(f_1f_2, d_1e_1\cdots d_ke_k) = 1$, then the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.1) applied with $z = w$ and $u = \sqrt{\log N}/10$ implies that the cardinality of the above set of integers n is

$$\left(1 + O\left(e^{-\sqrt{\log N}}\right)\right) \cdot \frac{N}{m} \cdot \frac{\nu^{(1)}_{p,m}(f_1)}{f_1f_2[d_1, e_1]\cdots[d_k, e_k]} \prod_{\ell\le w}\left(1 - \frac{\nu_{p,m}(\ell)}{\ell}\right).$$

On the other hand, if $(f_1f_2, d_1e_1\cdots d_ke_k) > 1$, then we note that there must exist a prime $r > w$ dividing f_1f_2 and $d_1e_1\cdots d_ke_k$. In particular, $r|Q^{(1)}_{p,m}(n)$ and $r|n + ps_i$ for some $i \in \{1, \dots, k\}$, which implies that $r|Q^{(1)}_{p,m}(-ps_i)$. Therefore, in this case the cardinality of the above set of integers n is

$$\ll \sum_{i=1}^{k} \sum_{\substack{r\mid(f_1f_2,d_1e_1\cdots d_ke_k) \\ r\mid Q^{(1)}_{p,m}(-ps_i),\ r>w}} \frac{N}{m} \cdot \frac{(k+1)^{\omega(f_1)}}{[f_1f_2, [d_1, e_1]\cdots[d_k, e_k]]}$$

Combining these observations, we deduce that

$$\Delta_{p,m} = \frac{N}{m} \cdot \prod_{\ell\le w}\left(1 - \frac{\nu_{p,m}(\ell)}{\ell}\right) \cdot \sum_{\substack{d_j,e_j \mid n+ps_j \\ (d_ie_i,d_je_jP(w))=1 \\ 1\le i,j\le k,\ i\ne j}} \sum_{\substack{f_1\in\mathcal{P}(w,N^{1/u}) \\ f_2\in\mathcal{P}(N^{1/u},N^{1/1000}) \\ (f_1f_2,d_1e_1\cdots d_ke_k)=1}} \frac{\nu^{(1)}_{p,m}(f_1)\mu_1(f_1)\mu_2(f_2)\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{f_1f_2[d_1, e_1]\cdots[d_k, e_k]}$$
$$+ O\left(\frac{N}{m(\log N)^{10}} + (\log N)^{O(1)} \cdot E\right),$$

where

$$E := \sum_{i=1}^{k} \sum_{\substack{r>w \\ r\mid Q^{(1)}_{p,m}(-ps_i)}} \sum_{\substack{d_j,e_j \\ (d_ie_i,d_je_jP(w))=1 \\ 1\le i,j\le k,\ i\ne j \\ d_1\cdots d_k, e_1\cdots e_k \le N^{1/100}}} \sum_{\substack{f_1\in\mathcal{P}(w,N^{1/u}) \\ f_2\in\mathcal{P}(N^{1/u},N) \\ f_1,f_2\le N^{1/100} \\ r\mid(f_1f_2,d_1e_1\cdots d_ke_k)}} \frac{N}{m} \cdot \frac{(k+1)^{\omega(f_1)}}{[f_1f_2, [d_1, e_1]\cdots[d_k, e_k]]}$$
$$\ll \sum_{\substack{r>w \\ r\mid Q^{(1)}_{p,m}(-ps_i)}} \frac{N(\log N)^{O(1)}}{rm} \ll \frac{N(\log N)^{O(1)}}{wm},$$

since $Q_{p,m}^{(1)}(-ps_i)$ has $\ll \log N$ prime factors in total. Next, we remove again the condition $(f_1f_2, d_1e_1\cdots d_ke_k) = 1$ from our new formula for $\Delta_{p,m}$. Since any common prime factor of f_1f_2 and $d_1e_1\cdots d_ke_k$ must be $> w$ by the fact that $(d_1e_1\cdots d_ke_k, P(w)) = 1$, this produces an error term of size $\ll N(\log N)^{O(1)}/w$, which is admissible. In conclusion,

$$\Delta_{p,m} = \frac{N}{m}\cdot S_1\cdot S_2\cdot \prod_{\ell\le w}\left(1-\frac{\nu_{p,m}(\ell)}{\ell}\right)\cdot \sum_{\substack{d_j,e_j\\(d_ie_i,d_je_jP(w))=1\\1\le i,j\le k,\ i\neq j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} + O\left(\frac{N}{m(\log N)^{10}}\right),$$

where

$$S_1 := \sum_{f_1\in\mathcal{P}(w,N^{1/u})}\frac{\mu_1(f_1)\nu_{p,m}^{(1)}(f_1)}{f_1} \quad\text{and}\quad S_2 := \sum_{f_1\in\mathcal{P}(N^{1/u},N^{1/1000})}\frac{\mu_2(f_2)}{f_2}.$$

As in the proof of Lemma 7.1.3, we have that

$$\begin{aligned}\sum_{\substack{d_j,e_j\\(d_ie_i,d_je_jP(w))=1\\1\le i,j\le k,\ i\neq j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} &= \sum_{P^-(a_1\cdots a_k)>w}\frac{y_{\boldsymbol{a}}^2}{a_1\cdots a_k} + O\left(\frac{(\log N)^{O(1)}}{w}\right)\\ &\asymp (\log(N^{1/100}))^k\prod_{\ell\le w}\left(1-\frac{1}{\ell}\right)^{-k}\int f(\boldsymbol{t})^2d\boldsymbol{t}\\ &\asymp \left(\frac{e^\gamma(\log N)(\log w)}{100}\right)^k\int f(\boldsymbol{t})^2d\boldsymbol{t},\end{aligned}$$

by Mertens's formula. The implied constants in the second and third line are absolute, provided that N is large enough in terms of k. Moreover, the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.3) implies that[1]

$$S_1 \asymp \prod_{w<\ell\le N^{1/u}}\left(1-\frac{\nu_{p,m}^{(1)}(\ell)}{\ell}\right) \asymp \left(\frac{\log w}{\log(N^{1/u})}\right)^{k+2}$$

since $\nu_{p,m}^{(1)}(\ell) = k+2$, unless ℓ is one of the $O(\log N)$ prime divisors of the discriminant of the polynomial $Q_{p,m}$, and

$$S_2 \asymp \prod_{N^{1/u}<\ell\le N^{1/1000}}\left(1-\frac{1}{\ell}\right) \asymp \frac{1}{u}$$

Both constants are absolute here, as long as N is large enough in terms of k. Finally, note that the definition of s_j implies that $\nu_{p,m}(\ell) = 1 + \mathbf{1}_{\ell\nmid m}$ for primes $\ell \le C_k$. Moreover,

[1]Strictly speaking, Lemma 3.3.3 cannot be used directly to estimate S_1 and S_2. It is easy to deduce the claimed estimates from it though. For example, we may take $\mu_1 := \widetilde{\mu}\cdot\mathbf{1}_{P^-(n)>w}$, where $\widetilde{\mu}_1$ is an upper bound sieve supported on the set $\{d\le N^{1/100} : d|P(N^{1/u})\}$ and apply Lemma 3.3.3 with the multiplicative function $g(n) = \mathbf{1}_{P^-(n)>w}/n$. The sum S_2 is handle in an analogous way.

if $C_k < \ell \le w$, then $\nu_{p,m}(\ell) = 2k+2$, unless ℓ divides one of the numbers $(s_i - s_j)$ or $mp(s_i - s_j) - 1$, for some $0 \le i, j \le k$ with $i \ne j$. Since

$$\sum_{\substack{0\le i,j\le k\\ i\ne j}} \sum_{\substack{C_k<\ell\le w\\ \ell\mid s_i-s_j}} \frac{k}{\ell} \ll \frac{k^3 \log C_k}{C_k} \ll 1$$

and

$$\sum_{\substack{0\le i,j\le k\\ i\ne j}} \sum_{\substack{(\log N)^2<\ell\le w\\ \ell\mid mp(s_i-s_j)-1}} \frac{k}{\ell} \ll \frac{k^3 \log N}{(\log N)^2} \ll 1$$

for large enough N, we deduce that

$$\begin{aligned}\prod_{\ell\le w}\left(1-\frac{\nu_{p,m}(\ell)}{\ell}\right) &= \prod_{\ell\le C_k}\left(1-\frac{1+\mathbf{1}_{\ell\mid m}}{\ell}\right)\cdot\prod_{C_k<\ell\le w}\left(1-\frac{1}{\ell}\right)^{2k+2}\cdot\prod_{C_k<\ell\le w}\frac{1-\nu_{p,m}(\ell)/\ell}{(1-1/\ell)^{2k+2}}\\ &\ll \frac{m}{\varphi(m)}\cdot\frac{1}{(\log C_k)^2}\cdot\left(\frac{\log C_k}{\log w}\right)^{2k+2}\cdot e^{H_{p,m}},\end{aligned}$$

where the constant is absolute by our assumption that $C_k > e^k$. Since

$$|\mathcal{Q}_m| \asymp \frac{uN}{\varphi(m)(\log N)^2},$$

the lemma follows. □

Lemma 7.3.4. *If m, η and I_m are as in the statement of Proposition 7.3.1, $q \in \mathcal{Q}_m \cap (s_k z, \infty)$, $j_0 \in \{1,\dots,k\}$ and N is large enough in terms of k, A and η, then*

$$\begin{aligned}&\sum_{\substack{p\in I_m\\ P^-(Q_{p,m}(q-ps_{j_0}))>w}} (1*\mu_1)(Q^{(1)}_{p,m}(q-ps_{j_0}))\cdot(1*\mu_2)(q-ps_{j_0})\left(\sum_{\substack{d_i\mid q+p(s_i-s_{j_0})\\ 1\le i\le k}}\lambda_{\boldsymbol{d}}\right)^2\\ &\gg \eta\cdot|\mathcal{Q}_m|\cdot\left(\frac{e^\gamma(\log C_k)^2u}{100}\right)^k\cdot\int\left(\int f(\boldsymbol{t})dt_{j_0}\right)^2 dt_1\cdots dt_{j_0-1}d_{t_{j_0+1}}\cdots dt_k,\end{aligned}$$

where the implied constant is absolute. Moreover, if $\ell_0 \le (\log N)^2$ is a prime and $b_0 \in \{1,\dots,\ell_0-1\}$, then

$$\begin{aligned}&\sum_{\substack{p\in I_m\\ p\equiv b_0 \,(\mathrm{mod}\,\ell_0)\\ P^-(Q_{p,m}(q-ps_{j_0}))>w}} (1*\mu_1)(Q^{(1)}_{p,m}(q-ps_{j_0}))\cdot(1*\mu_2)(q-ps_{j_0})\left(\sum_{\substack{d_i\mid q+p(s_i-s_{j_0})\\ 1\le i\le k}}\lambda_{\boldsymbol{d}}\right)^2\\ &\ll e^{G_m(q)}\cdot\frac{\eta|\mathcal{Q}_m|}{\ell_0}\cdot\left(\frac{e^\gamma(\log C_k)^2u}{100}\right)^k\cdot\int\left(\int f(\boldsymbol{t})dt_{j_0}\right)^2 dt_1\cdots dt_{j_0-1}d_{t_{j_0+1}}\cdots dt_k,\end{aligned}$$

where the implied constant is again absolute and $G_m(q)$ is defined by (7.3.1).

Proof. The proof is a combination of the proofs of Lemmas 7.1.2 and 7.3.3. We outline the argument when $j_0 = k$, the other cases being similar. All implicit constants might depend on k, f and A, unless otherwise specified. Also, recall the notation $\mathcal{P}(s,t)$ from the proof of Lemma 7.3.3.

Note that since q is a prime $> z \geq N^{1/100}$, it cannot have any divisors $\leq N^{1/100}$. In particular, we must have that $d_k = 1$. Similarly, since $q \in \mathcal{Q}_m$ and $q > N^{1/u}$, we know that $P^-(q(qm-1)) > N^{1/u}$. Therefore the condition $P^-(Q_{p,m}(q-ps_k)) > w$ reduces to $P^-(\widetilde{Q}_{p,m}(q-ps_k)) > w$, where

$$\widetilde{Q}_{p,m}(x) := \prod_{i=0}^{k-1}[(x+ps_i)(m(x+ps_i)-1)].$$

In the same fashion, we have that $(1*\mu_1)(Q^{(1)}_{p,m}(q-ps_j)) = (1*\mu_1)(\widetilde{Q}^{(1)}_{p,m}(q-ps_j))$ by our assumption that μ_1 is supported on the set $\mathcal{P}(w, N^{1/u})$, where

$$\widetilde{Q}^{(1)}_{p,m}(x) := \prod_{i=0}^{k-1}(m(x+ps_i)-1).$$

So, if we set

$$S = \sum_{\substack{p\in I_m \\ p\equiv b_0 \,(\mathrm{mod}\, \ell_0) \\ P^-(Q_{p,m}(q-ps_k))>w}} (1*\mu_1)(Q^{(1)}_{p,m}(q-ps_k))\cdot(1*\mu_2)(q-ps_k)\left(\sum_{\substack{d_i \mid q+p(s_i-s_k) \\ 1\leq i<k}} \lambda_{\boldsymbol{d},1}\right)^2,$$

where we now allow ℓ_0 to be either 1 or a prime number $\leq (\log N)^2$ and b_0 is coprime to ℓ_0, we find that

$$S = \sum_{\substack{d_i,e_i \mid n+ps_j \\ (d_ie_i, d_je_jP(w))=1 \\ 1\leq i,j<k,\ i\neq j}} \sum_{\substack{f_1\in\mathcal{P}(w,N^{1/u}) \\ f_2\in\mathcal{P}(N^{1/u},N^{1/1000})}} \mu_1(f_1)\mu_2(f_2)\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}$$
$$\times \#\left\{ \begin{array}{ll} p\in I_m & \widetilde{Q}^{(1)}_{p,m}(q-ps_k)\equiv 0\,(\mathrm{mod}\, f_1), \\ p\equiv b_0\,(\mathrm{mod}\,\ell_0) & : \ q-ps_k\equiv 0\,(\mathrm{mod}\, f_2), \\ P^-(\widetilde{Q}_{p,m}(q-ps_k))>w & q+p(s_i-s_k)\equiv 0\,(\mathrm{mod}\,[d_j,e_j]), \\ & 1\leq j<k \end{array} \right\}.$$

Note that if $\widetilde{Q}_{b_0,m}(q-b_0s_k)\equiv 0\,(\mathrm{mod}\,\ell_0)$, then $S=0$. So we may assume that $\widetilde{Q}_{b_0,m}(q-b_0s_k)\not\equiv 0\,(\mathrm{mod}\,\ell_0)$. As in the proof of Lemma 7.3.3, we note that if r is a common prime factor of f_1f_2 and of $d_1e_1\cdots d_ke_k$, then it must also divide

$$D = \prod_{j=1}^{k}\prod_{i=0}^{k}((s_k-s_j)(mq-1)-q(s_k-s_i)).$$

So the contribution to S of integers f_1, f_2 that are not coprime to $d_1e_1\cdots d_ke_k$ is

$$\ll \sum_{\substack{r\mid D\\ r>w}} \sum_{\substack{d_i,e_i\mid n+ps_j\\ (d_ie_i,d_je_jP(w))=1\\ 1\le i,j<k,\ i\neq j}} \sum_{\substack{f_1\in\mathcal{P}(w,N^{1/u})\\ f_2\in\mathcal{P}(N^{1/u},N^{1/1000})\\ r\mid(f_1f_2,d_1e_1\cdots d_ke_k)}} \mathrm{meas}(I_m)\cdot\frac{(k-1)^{\omega(f_1)}}{[f_1f_2,d_1e_1\cdots d_ke_k]} \ll \frac{N(\log N)^{O(1)}}{mw}.$$

Next, if $(f_1f_2,d_1e_1\cdots d_ke_k)=1$, then we estimate the cardinality of primes $p\in I_m$ that appears in the above expression for S using the Fundamental Lemma of Sieve Methods (cf. 3.3.3) as in the proof of Lemma 7.1.2, controlling the total error using the Bombieri-Vinogradov theorem. The resulting estimate is

$$S=\frac{\#\{p\in I_m\}}{\varphi(\ell_0)}\cdot\prod_{\substack{\ell\le w\\ \ell\neq\ell_0}}\left(1-\frac{\widetilde{\nu}(\ell)}{\ell}\right)\sum_{\substack{d_i,e_i,f_1,f_2\\ (d_ie_i,d_je_jf_1f_2P(w))=1\\ 1\le i,j<k,\ i\neq j}}\frac{\widetilde{\nu}^{(1)}(f_1)\mu_1(f_1)\mu_2(f_2)\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{\varphi(f_1)\varphi(f_2)\varphi([d_1,e_1])\cdots\varphi([d_{k-1},e_{k-1}])}$$
$$+O\left(\frac{N}{m(\log N)^A}\right),$$

where

$$\widetilde{\nu}(d):=\#\{n\,(\mathrm{mod}\,d):\widetilde{Q}_{n,m}(q-ns_k)\equiv 0\,(\mathrm{mod}\,d)\}$$

and

$$\widetilde{\nu}^{(1)}(d):=\#\{n\,(\mathrm{mod}\,d):\widetilde{Q}^{(1)}_{n,m}(q-ns_k)\equiv 0\,(\mathrm{mod}\,d)\}.$$

Next, we remove again the condition that $(f_1f_2,d_1e_1\cdots d_ke_k)=1$ as in Lemma 7.3.3 at the cost of a total error that is $\ll N(\log N)^{O(1)}/(mw)$, which is of admissible size. This separates the variables f_1,f_2,f_3 from each other and from $d_1,e_1,\dots,d_ke_k$, and we deduce that

$$S=\frac{\#\{p\in I_m\}}{\varphi(\ell_0)}\cdot S_1\cdot S_2\cdot\prod_{\substack{\ell\le w\\ \ell\neq\ell_0}}\left(1-\frac{\widetilde{\nu}(\ell)}{\ell-1}\right)\cdot\sum_{\substack{d_i,e_i\\ (d_ie_i,d_je_jP(w))=1\\ 1\le i,j<k,\ i\neq j}}\frac{\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{\varphi([d_1,e_1])\cdots\varphi([d_{k-1},e_{k-1}])}$$
$$+O\left(\frac{N}{m(\log N)^A}\right),$$

where

$$S_1:=\sum_{f_1}\frac{\widetilde{\nu}^{(1)}(f_1)\mu_1(f_1)}{\varphi(f_1)},\quad\text{and}\quad S_2:=\sum_{f_2}\frac{\mu_2(f_2)}{\varphi(f_2)}.$$

The sum over $\boldsymbol{d}$ and over $\boldsymbol{e}$ is estimated as in Section 7.1: we have that

$$\sum_{\substack{d_i,e_i\\(d_ie_i,d_je_jP(w))=1\\1\le i,j<k,\ i\ne j}} \frac{\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{\varphi([d_1,e_1])\cdots\varphi([d_{k-1},e_{k-1}])}$$

$$= \sum_{P^-(a_1\cdots a_{k-1})>w} \frac{1}{a_1\cdots a_{k-1}} \left(\sum_{P^-(b)>w} \frac{\mu(b)\xi_{\boldsymbol{a},b}}{b}\right)^2 + O\left(\frac{(\log N)^{O(1)}}{w}\right)$$

$$\asymp (\log(N^{1/100}))^{k+1} \prod_{\ell\le w}\left(1-\frac{1}{\ell}\right)^{-k-1} \int\left(\int f(\boldsymbol{t})dt_k\right)^2 dt_1\cdots dt_{k-1}$$

$$\asymp \frac{e^{k\gamma}(\log N)^{k+1}(\log w)^{k+1}}{100^k} \int\left(\int f(\boldsymbol{t})dt_k\right)^2 dt_1\cdots dt_{k-1}$$

by Mertens's formula, provided that N is large enough in terms of k. The implied constants in the third and fourth line are absolute. Moreover, the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.3) implies that

$$S_1 \asymp \prod_{w<\ell\le N^{1/u}}\left(1-\frac{\widetilde{\nu}^{(1)}(\ell)}{\ell-1}\right) \asymp \left(\frac{\log w}{\log(N^{1/u})}\right)^{k+1},$$

where the constant is uniform in k if N is large enough, since $\widetilde{\nu}^{(1)}(\ell)=k+1$ unless ℓ is one of the $O(\log N)$ divisors of the discriminant of the polynomial $n\to\widetilde{Q}_{n,m}(n-ps_k)$, and that

$$S_2 \asymp \prod_{N^{1/u}<\ell\le N^{1/1000}}\left(1-\frac{1}{\ell-1}\right) \asymp \frac{1}{u}.$$

Moreover, note that $\widetilde{\nu}(\ell)=0$ if $\ell\le C_k$ since $q(qm-1)$ have no divisors $\le N^{1/u}$. This implies that

$$\prod_{\substack{C_k<\ell\le w\\ \ell\nmid\ell_0}}\left(1-\frac{\widetilde{\nu}(\ell)}{\ell-1}\right) = \prod_{\substack{C_k<\ell\le w\\ \ell\nmid\ell_0}}\left(1-\frac{\widetilde{\nu}(\ell)}{\ell-1}\right) \asymp \prod_{C_k<\ell\le w}\left(1-\frac{1}{\ell}\right)^{2k}\cdot\prod_{C_k<\ell\le w}\frac{1-\widetilde{\nu}(\ell)/(\ell-1)}{(1-1/\ell)^{2k}}$$

$$\asymp \left(\frac{\log C_k}{\log w}\right)^{2k}\cdot\prod_{\substack{C_k<\ell\le w\\ \widetilde{\nu}(\ell)<2k}}\frac{1-\widetilde{\nu}(\ell)/\ell}{1-2k/\ell}$$

since we have assumed that $C_k>e^k\gg k^2$. Therefore

$$1\ll \prod_{\substack{C_k<\ell\le w\\ \widetilde{\nu}(\ell)<2k}}\frac{1-\widetilde{\nu}(\ell)/(\ell-1)}{(1-1/\ell)^{2k}} \ll \exp\left\{\sum_{\substack{C_k<\ell\le w\\ \widetilde{\nu}(\ell)<2k}}\frac{2k}{\ell}\right\} \ll \exp\left\{\sum_{\substack{C_k<\ell\le(\log N)^2\\ \widetilde{\nu}(\ell)<2k}}\frac{2k}{\ell}\right\}.$$

Note that

$$\widetilde{Q}_{n,m}(q-ns_k) = \prod_{j=0}^{k-1}[(q+n(s_j-s_k))(mq-1+mn(s_j-s_k))].$$

Therefore, if $\widetilde{\nu}(\ell) < 2k$ and $C_k < \ell \le w$, then ℓ must be a divisor of

$$\prod_{\substack{1\le j,j'\le k-1\\ j\neq j'}} [(s_j - s_{j'})(qm(s_j - s_{j'}) - (s_j - s_k))].$$

Finally, note that if $\ell | s_j - s_{j'}$ with $\ell > C_k$, then we must have that ℓ is one of the $O(\log C_k)$ divisors of the number $p_{\pi(C_k)+j} - p_{\pi(C_k)+j'}$. Since $k^3(\log C_k)/C_k \ll 1$ by our assumption that $C_k > e^k$, the lemma follows. □

Proof of Lemma 7.3.2. Since we are looking for a lower bound, we may restrict the summation in $\sum_{p\in I_m} \delta_{p,m}(q)$ to those primes $p \in I_m$ with $H_{p,m} \le 1$, where $H_{p,m}$ is defined in the statement of Lemma 7.3.3. Then relation (7.3.2) and Lemma 7.3.3 imply that

$$\sum_{p\in I_m} \delta_{p,m}(q) \gg \sum_{j=1}^{k} \sum_{\substack{p\in I_m\\ H_{p,m}\le 1\\ P^-(Q_{p,m}(q-ps_j))>w}} \frac{(1*\mu_1)(Q_{p,m}^{(1)}(q-ps_j))\cdot(1*\mu_2)(q-ps_j)}{|\mathcal{Q}_m|\left(\frac{e^\gamma(\log C_k)^2u}{100}\right)^k \int f(\boldsymbol{t})^2 d\boldsymbol{t}} \left(\sum_{\substack{d_i | q+p(s_i-s_j)\\ 1\le i\le k}} \lambda_{\boldsymbol{d}}\right)^2$$
$$= \frac{\Delta_1 - \Delta_2}{|\mathcal{Q}_m|\left(\frac{e^\gamma(\log C_k)^2u}{100}\right)^k \int f(\boldsymbol{t})^2 d\boldsymbol{t}},$$

where the implicit constant is absolute,

$$\Delta_1 := \sum_{j=1}^{k} \sum_{\substack{p\in I_m\\ P^-(Q_{p,m}(q-ps_j))>w}} (1*\mu_1)(Q_{p,m}^{(1)}(q-ps_j))\cdot(1*\mu_2)(q-ps_j) \left(\sum_{\substack{d_i | q+p(s_i-s_j)\\ 1\le i\le k}} \lambda_{\boldsymbol{d}}\right)^2$$

and

$$\Delta_2 := \sum_{j=1}^{k} \sum_{\substack{p\in I_m\\ H_{p,m}>1\\ P^-(Q_{p,m}(q-ps_j))>w}} (1*\mu_1)(Q_{p,m}^{(1)}(q-ps_j))\cdot(1*\mu_2)(q-ps_j) \left(\sum_{\substack{d_i | q+p(s_i-s_j)\\ 1\le i\le k}} \lambda_{\boldsymbol{d}}\right)^2.$$

The lower bound in Lemma 7.3.4 implies that

$$\Delta_1 \gg \eta|\mathcal{Q}_m|\left(\frac{e^\gamma(\log C_k)^2u}{100}\right)^k \sum_{j=1}^{k} \int \left(\int f(\boldsymbol{t})dt_j\right)^2 dt_1\cdots d_{j-1}dt_{j+1}\cdots dt_k,$$

whereas Markov's inequality and the upper bound in Lemma 7.3.4 imply that

$$\begin{aligned}\Delta_2 &\le \sum_{j=1}^{k} \sum_{\substack{p\in I_m\\ H_{p,m}>1\\ P^-(Q_{p,m}(q-ps_j))>w}} H_{p,m}\cdot(1*\mu_1)(Q_{p,m}^{(1)}(q-ps_j))\cdot(1*\mu_2)(q-ps_j)\left(\sum_{\substack{d_i\mid q+p(s_i-s_j)\\ 1\le i\le k}}\lambda_{\boldsymbol{d}}\right)^2\\ &= \sum_{j=1}^{k}\sum_{\substack{0\le i,i'\le k\\ i\ne i'}}\sum_{\substack{C_k<\ell\le(\log N)^2\\ \ell\nmid m(s_i-s_{i'})}}\frac{2k+2}{\ell}\\ &\quad\times\sum_{\substack{p\in I_m\\ pm(s_i-s_{i'})\equiv 1\,(\mathrm{mod}\,\ell)\\ P^-(Q_{p,m}(q-ps_j))>w}}(1*\mu_1)(Q_{p,m}^{(1)}(q-ps_j))\cdot(1*\mu_2)(q-ps_j)\left(\sum_{\substack{d_i\mid q+p(s_i-s_j)\\ 1\le i\le k}}\lambda_{\boldsymbol{d}}\right)^2\\ &\ll \frac{k^3e^{G_m(q)}}{C_k}\eta|\mathcal{Q}_m|\left(\frac{e^\gamma(\log C_k)^2u}{100}\right)^k\sum_{j=1}^{k}\int\left(\int f(\boldsymbol{t})dt_j\right)^2 dt_1\cdots d_{j-1}dt_{j+1}\cdots dt_k.\end{aligned}$$

Putting together the above estimates and choosing f such that the quantity

$$\sum_{j=1}^{k}\frac{\int\left(f(\boldsymbol{t})dt_j\right)dt_1\cdots d_{j-1}dt_{j+1}\cdots dt_k}{\int f(\boldsymbol{t})^2d\boldsymbol{t}}$$

is $\gg\log k$, which is possible by Lemma 7.1.6, completes the proof of Lemma 7.3.2. □

7.4 Cramér's model

We conclude this chapter with a heuristic discussion of the local distribution of primes and, in particular, of the gaps between them. In order to do so, we introduce the so-called Cramér model for the prime numbers and its refinement due to Granville. This model turns out to be very effective in making accurate predictions about the distribution of primes (local and global), unlike the Kubilius model which, as we saw in Section 2.2, is not very successful in this task.

First of all, let us recall a quantitative form of the Prime Number Theorem:

$$\pi(x)=\mathrm{li}(x)+O_A\left(\frac{x}{(\log x)^A}\right)=\int_2^x\frac{dt}{\log t}+O_A\left(\frac{x}{(\log x)^A}\right)\quad(x\ge 2).$$

This estimate can be also interpreted by saying that the density of primes around x is about $1/\log x$ (this is was what Gauss had conjectured in fact). So an integer n should be prime with probability about $1/\log n$. Of course, the primes are deterministic objects, so this statement is obviously false. However, modelling them this way leads to surprisingly accurate estimates. More precisely, we define a sequence of Bernoulli random variables $Y_2, Y_3, \ldots$ such

that

$$\begin{cases} \mathbf{Prob}(Y_n = 1) = \dfrac{1}{\log n}, \\ \mathbf{Prob}(Y_n = 0) = 1 - \dfrac{1}{\log n}. \end{cases} \tag{7.4.1}$$

Furthermore, we assume that the variables Y_n are independent from each other, since á priori there should not be any correlation between two integers being prime. This is the so-called Cramér model of the prime numbers, which we can use to make predictions about various questions regarding the primes.

For example, the random variable

$$\Pi(x) = \sum_{2 \le n \le x} Y_n,$$

is a model for $\pi(x)$. Now, note that

$$\mathbb{E}[\Pi(x)] = \sum_{2 \le n \le x} \mathbb{E}[Y_n] = \sum_{2 \le n \le x} \frac{1}{\log n} = \mathrm{li}(x) + O(1)$$

and

$$\mathrm{Var}[\Pi(x)] = \sum_{2 \le n \le x} \mathrm{Var}[Y_n] = \sum_{2 \le n \le x} \left(\frac{1}{\log n} - \frac{1}{\log n^2} \right) = \mathrm{li}(x) + O(1) \sim \frac{x}{\log x}.$$

So, the Law of the Iterated Logarithm predicts that

$$|\Pi(x) - \mathrm{li}(x)| \le (1+\epsilon)\sqrt{2\,\mathrm{Var}[\Pi(x)] \log\log(\mathrm{Var}[\Pi(x)])} \sim (1+\epsilon)\sqrt{\frac{2x \log\log x}{\log x}}$$

almost surely, for every fixed $\epsilon > 0$. In particular, with probability 1, the Riemann Hypothesis is true for our model $\Pi(x)$.

Next, we use Cramér's model to study the distribution of the difference of two consecutive primes. Let $p_1, p_2, \ldots$ denote the sequence of prime numbers in increasing order. Then, according to Cramér's model, a model for

$$\#\{p_m \le x : p_{m+1} = p_m + k\}$$

is given by

$$\begin{aligned} \mathbb{E}\left[\sum_{n \le x} Y_n Y_{n+k} \prod_{j=1}^{k-1} (1 - Y_{n+j}) \right] &= \sum_{n \le x} \frac{1}{\log n} \log(n+k) \prod_{j=1}^{k-1} \left(1 - \frac{1}{\log(n+j)} \right) \\ &\sim \frac{x}{(\log x)^2} \left(1 - \frac{1}{\log x} \right)^{k-1} \sim \frac{x}{(\log x)^2} e^{-\frac{k}{\log x}}, \end{aligned}$$

uniformly in $1 \le k \le \log x$, as $x \to \infty$. So, for fixed $\alpha < \beta$, we expect that

$$\begin{aligned}\#\{p_m \le x : \alpha < \frac{p_{m+1}-p_m}{\log m} \le \beta\} &\approx \#\{p_m \le x : \alpha \log x < p_{m+1}-p_m \le \beta \log x\} \\ &\sim \frac{x}{(\log x)^2} \sum_{\alpha \log x < k \le \beta \log x} e^{-\frac{k}{\log x}} \\ &\sim \frac{x}{\log x}(e^{-\alpha}-e^{-\beta}) \sim \pi(x)\int_\alpha^\beta e^{-t}dt,\end{aligned}$$

as $x \to \infty$. This leads to the prediction that the sequence $(p_{m+1}-p_m)/\log m$ should follow an exponential distribution.

Now, let us use this model to study another problem: counting twin primes. A model for the number of twin primes $(p, p+2)$ with $p \le x$ then is $\Pi_2(x) = \sum_{2\le n\le x} Y_n Y_{n+2}$. Since

$$\mathbb{E}[\Pi_2(x)] = \sum_{2\le n\le x} \mathbb{E}[Y_n Y_{n+2}] = \sum_{2\le n\le x} \mathbb{E}[Y_n]\cdot\mathbb{E}[Y_{n+2}] = \sum_{2\le n\le x} \frac{1}{(\log n)\log(n+2)} \sim \frac{x}{\log^2 x},$$

as $x \to \infty$. So this leads to the prediction that the numbers of twin primes $(p, p+2)$ with $p \le x$ is asymptotic to $x/\log^2 x$. However, it is widely believed that

$$\#\{n \le x : (n, n+2) \text{ are twin primes}\} \sim \frac{x}{\log^2 x}\cdot 2\prod_{p>2}\left(1-\frac{1}{p}\right)^{-2}\left(1-\frac{2}{p}\right). \tag{7.4.2}$$

The reason why Cramér's model failed to give the right asymptotic formula for the number of twin primes is the assumption that the variables Y_n are independent from each other. Indeed, two consecutive integers larger than 2 can never be prime simultaneously. So there is a strong correlation between Y_n and Y_{n+1}. Similarly, if $n > 3$, then n, $n+2$ and $n+4$ cannot be simultaneously prime, since at least one of them is divisible by 3. So there the variables Y_n, Y_{n+2} and Y_{n+4} are correlated. Similar constraints arise for all small primes and we need to adjust our model appropriately if we want to make accurate predictions.

The way we modify Cramér's is by ensuring that n and $n+2$ are in the right classes modulo all small primes. This is done by applying what is called a preliminary sieve and restricting n and $n+2$ to *a priori* have no fixed prime factors. This increases the probability that the randomly chosen n and $n+2$ are both primes. More precise, we set $\mathcal{N} = \{n \in \mathbb{N} : P^-(n) > z\}$ and we consider a new sequence of independent Bernoulli random variables $\{Z_n\}_{n\in\mathcal{N}}$ such that

$$\begin{cases}\mathbf{Prob}(Z_n = 1) = \dfrac{1}{\log n}\displaystyle\prod_{p\le z}\left(1-\frac{1}{p}\right)^{-1}, \\ \mathbf{Prob}(Z_n = 0) = 1 - \dfrac{1}{\log n}\displaystyle\prod_{p\le z}\left(1-\frac{1}{p}\right)^{-1}.\end{cases} \tag{7.4.3}$$

(Strictly speaking, we have to assume that n is large enough in terms of z, otherwise the above probabilities might not be numbers in $[0,1]$. This is a minor technical assumption.)

Then our model for the number of twin primes up to x is the function

$$\widetilde{\Pi}_2(x) = \sum_{\substack{n\leq x\\ n,n+2\in\mathcal{N}}} Z_n Z_{n+2}.$$

We have that

$$\begin{aligned}\mathbb{E}[\widetilde{\Pi}_2(x)] &= \sum_{\substack{n\leq x\\ n,n+2\in\mathcal{N}}} \mathbb{E}[Z_n Z_{n+2}] = \sum_{\substack{n\leq x\\ n,n+2\in\mathcal{N}}} \mathbb{E}[Z_n]\cdot\mathbb{E}[Z_{n+2}]\\ &= \sum_{\substack{n\leq x\\ n,n+2\in\mathcal{N}}} \frac{1}{(\log n)\log(n+2)} \prod_{p\leq z}\left(1-\frac{1}{p}\right)^{-2}\\ &\sim \frac{x}{\log^2 x}\prod_{p\leq z}\left(1-\frac{1}{p}\right)^{-2}\left(1-\frac{\nu(p)}{p}\right),\end{aligned}$$

where

$$\nu(p) := \#\{m \,(\mathrm{mod}\, d) : m(m+2)\equiv 0\,(\mathrm{mod}\, p)\} = \begin{cases}1 & \text{if } p=2,\\ 2 & \text{otherwise,}\end{cases}$$

by the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.1). So we arrive to the refined prediction that

$$\#\{n\leq x : (n,n+2) \text{ are twin primes}\} \sim \frac{x}{\log^2 x}\cdot 2\prod_{2<p\leq z}\left(1-\frac{1}{p}\right)^{-2}\left(1-\frac{2}{p}\right).$$

Letting $z\to\infty$, we recover the conjectured estimate (7.4.2).

Exercise 7.4.1. Let $N\geq 1$. Use the Cramér-Granville model to predict an asymptotic formula for the number of pairs of primes (p,q) such that $p+q=2N$.

Finally, it is possible to use Cramér's model (or its refinement) to make predictions about how big the difference of two consecutive primes can be. To do this, we appeal to the Borel-Cantelli lemma. Before we state this lemma, recall that, for a sequence of sets $\{A_n\}_{n=1}^{\infty}$, all of which are subsets of some ambient space Ω, we have that

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m = \{\omega\in\Omega : \omega\in A_m \text{ for infinitely many } m\}$$

and

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m = \{\omega\in\Omega : \omega\in A_m \text{ for all but finitely many } m\}.$$

Then we have the following result:

Lemma 7.4.2 (Borel-Cantelli)**.** *Let $(\Omega,\mathcal{F},\mathbf{Prob})$ denote a probability space and consider $\{A_n\}_{n=1}^{\infty}$, a sequence of events in the σ-algebra $\mathcal{F}$.*

(1) If $\sum_{n=1}^{\infty} P(A_n) < \infty$*, then* $\mathbf{Prob}(\limsup_{n\to\infty} A_n) = 0$.

(2) If the events A_n *are independent and* $\sum_{n=1}^{\infty} P(A_n) = \infty$*, then* $\mathbf{Prob}(\limsup_{n\to\infty} A_n) = 1$.

Now, let Ω be the sample space where the random variables Y_n, defined by (7.4.1), live. Given an increasing function $h : \mathbb{N} \to [1, \infty)$, we set

$$A_n(h) = \{\omega \in \Omega : Y_n(\omega) = 1,\ Y_{n+j}(\omega) = 0\ (1 \le j \le h(n))\} \quad (n \ge 2),$$

If we can show that

$$\mathbf{Prob}\left(\limsup_{n\to\infty} A_n(h)\right) = 0, \tag{7.4.4}$$

then it readily follows that

$$\mathbf{Prob}\left(\liminf_{n\to\infty} A_n(h)^c\right) = 1,$$

that is to say, almost surely all but finitely many of the events $A_n(h)^c$ occur simultaneously. This leads to the prediction that, for all but finitely many n, we have that $p_{n+1} - p_n \le h(n)$.

Assume that $h(n) \le n/\log n$ for all n, since we already know by the Prime Number Theorem that $p_{n+1} \le p_n + O(p_n/\log^2 p_n)$. Then

$$\mathbf{Prob}(A_n(h)) = \frac{1}{\log n} \prod_{j=1}^{\lfloor h(n)\rfloor} \left(1 - \frac{1}{\log(n+j)}\right) \asymp \frac{1}{\log n}\left(1 - \frac{1}{\log n}\right)^{h(n)} \le \frac{1}{\log n} e^{-h(n)/\log n},$$

for all $n \ge 2$. If $h(n) = (1+\epsilon)\log^2 n$, then

$$\mathbf{Prob}(A_n(h)) \ll \frac{1}{n^{1+\epsilon}\log n}.$$

Summing this inequality over all $n \ge 2$, we find that

$$\sum_{n=2}^{\infty} \mathbf{Prob}(A_n(h)) < \infty.$$

So applying Lemma 7.4.2, we deduce that (7.4.4) holds for this choice of h. Consequently, the discussion of the previous paragraph leads to the prediction that, for every fixed $\epsilon > 0$, we have that $p_{n+1} - p_n \le (1+\epsilon)\log^2 n$ for $n \ge n_0(\epsilon)$.

Using a similar but slightly more involved argument, we arrive to the prediction that $p_{n+1} - p_n > (1-\epsilon)\log^2 n$ for infinitely many n. Let

$$m_n = \left\lfloor (1-\epsilon)\sum_{k=2}^{n} \log^2 k \right\rfloor = (1-\epsilon)(n\log^2 n - n\log n + n) + O(1),$$

by Stirling's formula and partial summation, and consider the events

$$B_n(\epsilon) = \{\omega \in \Omega : Y_{m_n}(\omega) = 1,\ Y_{m_n+j}(\omega) = 0\ (1 \le j \le (1-\epsilon)\log n - 1)\} \quad (n \ge 10).$$

As before, we have that

$$\mathbf{Prob}(B_n(h)) = \prod (A_{m_n}(h)) \asymp \frac{1}{\log m_n}\left(1 - \frac{1}{\log m_n}\right)^{(1-\epsilon)\log^2 n} \asymp \frac{e^{(1-\epsilon)\log^2 n/\log m_n}}{\log n} \gg_\epsilon \frac{1}{n^{1-\epsilon/2}},$$

and consequently,

$$\sum_{n\geq 10} \mathbf{Prob}(B_n(h)) = \infty.$$

Since the events $B_n(h)$ are independent, by our assumption that the random variables Y_n are independent, Lemma 7.4.2 implies that

$$\mathbf{Prob}\left(\limsup_{n\to\infty} B_n(h)\right) = 1,$$

that is to say, infinitely many of the events $B_n(h)$ occur simultaneously almost surely. This leads to the prediction that infinitely many of the intervals $[m_n, m_{n+1})$ contain precisely one prime number, or equivalently, that $p_{n+1} - p_n > (1-\epsilon)\log^2 n$ for infinitely many n, as we mentioned above.

Putting together the above predictions, Cramér was led to conjecture that

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{(\log n)^2} = 1.$$

Later, building on the work of Maier [Mai85], Granville [Gr95] gave evidence that this conjecture might not be true. More precisely, using the refinement of Cramér's model that we gave above, he conjectured that

$$\limsup_{n\to\infty} \frac{p_{n+1} - p_n}{(\log n)^2} \geq 2e^{-\gamma} = 1.12291896713\cdots > 1. \tag{7.4.5}$$

In any case, we expect that $p_{n+1} - p_n \ll (\log n)^2$, for all $n \geq 2$.

Exercise 7.4.3. Use the refinement of Cramér's model given by (7.4.3) to give evidence in support of (7.4.5).

Chapter 8

Irregularities in the distribution of primes

So far we were concentrating our efforts into proving that the primes behave in the 'expected way'. As it is discussed in Section 7.4, we expect the maximal gap between two consecutive primes p_n and p_{n+1} to be of the order of $(\log p_n)^2$. So, if $y \geq (\log x)^{2+\epsilon}$, then it seems natural to conjecture that the interval $(x, x+y]$ contains the 'right' amount of prime numbers, that is to say,

$$\pi(x+y) - \pi(x) \sim \frac{y}{\log x}, \tag{8.0.1}$$

as $x \to \infty$. Indeed, Selberg showed a partial (conditional) result towards this direction:

Theorem 8.0.4 (Selberg)**.** *Assume that the Riemann Hypothesis is true. For every fixed $\epsilon > 0$ and $\delta > 0$, we have that*

$$\lim_{x\to\infty} \frac{1}{X} \operatorname{meas}\left(\left\{X \leq x \leq 2X : \left|\pi(x+y) - \pi(x) - \frac{y}{\log x}\right| \leq \frac{\epsilon y}{\log x},\ y = (\log x)^{2+\delta}\right\}\right) = 1.$$

However, in 1985 Maier arrived to the groundbreaking conclusion that (8.0.1) fails infinitely often when y is a power of $\log x$:

Theorem 8.0.5 (Maier)**.** *For every fixed $\alpha > 2$, we have that*

$$\liminf_{x\to\infty} \frac{\pi(x + (\log x)^\alpha) - \pi(x)}{(\log x)^{\alpha-1}} < 1 < \limsup_{x\to\infty} \frac{\pi(x + (\log x)^\alpha) - \pi(x)}{(\log x)^{\alpha-1}}. \tag{8.0.2}$$

Remark 8.0.6. Note that if relation (8.0.2) holds for some fixed α_0, then it also holds for all $\alpha \in (0, \alpha_0]$ by the pigeonhole principle.

We will show Theorem 8.0.5 in Section 8.2. Before this, in Section 8.1, we will introduce and study the so-called Buchstab function, which is central in the proof of Theorem 8.0.5.

8.1 Buchstab's function

Buchstab's function arises naturally in the following basic sieve problem: given x and z, let

$$\Phi(x,z) = \#\{n \le x : P^-(n) > z\}.$$

We know that

$$\begin{aligned}\Phi(x,z) = 1 + \pi(x) - \pi(z) &= \int_z^x \frac{dt}{\log t} + O\left(\frac{x}{e^{c\sqrt{\log x}}}\right) \\ &= \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2} + \frac{z}{\log z}\right) \quad (\sqrt{x} \le z \le x).\end{aligned} \tag{8.1.1}$$

On the other hand, Theorem 3.3.2 and Mertens' estimate imply that

$$\Phi(x,z) = \frac{e^{-\gamma}x}{\log z}\left\{1 + O\left(\frac{1}{\log x} + e^{-s\log s + O(s)}\right)\right\} \quad (1 \le z \le x,\ x = z^s). \tag{8.1.2}$$

We want to understand the transition from (8.1.1) to (8.1.2) as z becomes smaller compared to x. As in Chapter 6, in order to achieve this, we use a variation of Buchstab's identity:

$$\Phi(x,z) = \Phi(x,z') - \sum_{z < p \le z'} \Phi(x/p, p) \quad (1 \le z \le z'). \tag{8.1.3}$$

When $x^{1/3} \le z < x^{1/2}$, applying formula (8.1.3) with $z' = \sqrt{x}$ yields

$$\begin{aligned}\Phi(x,z) &= \Phi(x,\sqrt{x}) + \sum_{z < p \le \sqrt{x}} \Phi(x/p, p) \\ &= \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right) + \sum_{z < p \le \sqrt{x}} \left(\frac{x/p}{\log(x/p)} + O\left(\frac{x/p}{\log^2(x/p)} + \frac{z}{\log z}\right)\right) \\ &= \frac{x}{\log x} + x\int_z^{\sqrt{x}} \frac{1}{u\log(x/u)}\frac{du}{\log u} + O\left(\frac{x}{(\log x)^2}\right) \\ &= \frac{x}{\log x} + x\int_z^{\sqrt{x}} \frac{1}{\frac{\log x}{\log u} - 1}\frac{du}{u(\log u)^2} + O\left(\frac{x}{(\log x)^2}\right) \\ &= \frac{x}{\log x} + \frac{x}{\log x}\int_2^{\frac{\log x}{\log z}} \frac{dt}{t-1} + O\left(\frac{x}{(\log x)^2}\right).\end{aligned} \tag{8.1.4}$$

Motivated by relations (8.1.1), (8.1.2), (8.1.3) and (8.1.4), we define *Buchstab's function* $w : [0, +\infty) \to \mathbb{R}$ by letting $w(u) = 0$ for $u \le 1$, $w(u) = 1/u$ for $1 < u \le 2$, and then defining w inductively for $u > 2$ via the relation

$$w(u) = \frac{1}{u} + \frac{1}{u}\int_1^{u-1} w(t)dt. \tag{8.1.5}$$

This define a continuous and differentiable function for $u > 2$. The following theorem gives the expected relation between $\Phi(x,z)$ and $\omega(u)$.

Theorem 8.1.1. *For $1 \le z \le \sqrt{x}$ with $x = z^u$, we have that*

$$\Phi(x,z) = \frac{x\omega(u)}{\log z} + O\left(\frac{x}{(\log z)^2}\right).$$

Before we prove Theorem 8.1.1, we discuss briefly the properties of Buchstab's function. By induction and relation (8.1.5), we immediately find that

$$\frac{1}{u} \le w(u) \le 1 \quad (u > 1). \tag{8.1.6}$$

Moreover, mutiplying (8.1.5) by u and differentiating the resulting identity, we find that

$$w'(u)u = -w(u) + w(u-1) = -\int_{u-1}^{u} w(t)dt \quad (u > 2). \tag{8.1.7}$$

Taking absolute values, we deduce that

$$|w'(u)| \le \frac{1}{u}\int_{u-1}^{u} |w'(t)|dt \quad (u > 2).$$

So, arguing as in the proof of Theorem 6.1.2, we deduce that

$$w'(u) \ll e^{-u\log(u\log u)+O(u)} \quad (u > 2). \tag{8.1.8}$$

Proof of Theorem 8.1.1. First, we show that Buchstab's function satisfies an identity similar to (8.1.3). Indeed, the prime number theorem implies that

$$\sum_{p\le x} \frac{1}{p\log p} = c_1 + \int_2^x \frac{dt}{t(\log t)^2} + R(x) \quad \text{with} \quad R(x) \ll e^{-c_2\sqrt{\log x}},$$

for some appropriate constants $c_1 \in \mathbb{R}$ and $c_2 > 0$. Together with relations (8.1.5), (8.1.6) and (8.1.8), this implies that for $z' = x^{1/u'} \in [z, \sqrt{x}]$,

$$\begin{aligned}\sum_{z<p\le z'} w\left(\frac{\log x}{\log p} - 1\right)\frac{1}{p\log p} &= \int_z^{z'} w\left(\frac{\log x}{\log t} - 1\right)\frac{dt}{t(\log t)^2} + R(t)w\left(\frac{\log x}{\log t} - 1\right)\Big|_{t=z}^{z'} \\ &\quad + \int_z^{z'} R(t)w'\left(\frac{\log x}{\log t} - 1\right)\frac{\log x}{t(\log t)^2}dt \\ &= \frac{1}{\log x}\int_{u'}^{u} w(s-1)ds + O(e^{-c_2\sqrt{\log z}}) \\ &= \frac{w(u)u - w(u')u'}{\log x} + O(e^{-c_2\sqrt{\log z}}) = \frac{w(u)}{\log z} - \frac{w(u')}{\log z'} + O(e^{-c_2\sqrt{\log z}}).\end{aligned}$$

Together with (8.1.3), the above relation implies that

$$R(x,z) = R(x,z') + \sum_{z<p\le z'} R(x/p,p) + O\left(\frac{x}{e^{c_2\sqrt{\log z}}}\right), \tag{8.1.9}$$

where

$$R(y,t) := \Phi(y,t) - \frac{y}{\log t} w\left(\frac{\log y}{\log t}\right).$$

We will prove Theorem 8.1.1 using the above formula and an inductive argument, much like the proof of Theorem 6.1.1. It suffices to prove that there is some absolute constant C such that

$$|R(x,z)| \le \frac{Cx}{(\log z)^2} \quad (x = z^u \ge z^2 \ge 4). \tag{8.1.10}$$

Clearly, we may assume that z is large enough. Also, if $u \in [2,3)$, then note that (8.1.10) follows by (8.1.4). Next, assume that relation (8.1.10) holds for all $u \in [2,U)$, for some $U \ge 3$, and consider $u \in [U, U+1)$. Note that $(\log(x/p))/(\log p) \le u - 1 < U$, for all $p > z$, so the induction hypothesis and (8.1.9) with $z' = \sqrt{x}$ imply that

$$\begin{aligned} |R(x,z)| &\le |R(x,\sqrt{x})| + \sum_{z<p\le\sqrt{x}} \frac{Cx}{p(\log p)^2} + O\left(\frac{x}{e^{c_2\sqrt{\log z}}}\right) \\ &\le Cx \sum_{p>z} \frac{1}{p(\log p)^2} + O\left(\frac{x}{(\log x)^2} + \frac{x}{e^{c_2\sqrt{\log z}}}\right), \end{aligned}$$

where we used (8.1.1). So the prime number theorem yields that

$$\begin{aligned} |R(x,z)| &\le Cx\left(\int_z^\infty \frac{dt}{t(\log t)^3} + O(e^{-c_3\sqrt{\log z}})\right) + O\left(\frac{x}{(\log x)^2} + \frac{x}{e^{c_2\sqrt{\log z}}}\right) \\ &= \frac{Cx}{2(\log z)^2} + O\left(\frac{x}{(\log x)^2} + \frac{(C+1)x}{e^{\min\{c_2,c_3\}\sqrt{\log z}}}\right). \end{aligned}$$

Choosing C large enough implies that (8.1.10) when $u \in [U, U+1)$ too, thus completing the inductive step. This concludes the proof of (8.1.10), and hence of Theorem 8.1.1. □

Theorem 8.1.2. *We have that*

$$w(u) = e^{-\gamma} + O\left(e^{-u\log(u\log u)+O(u)}\right) \quad (u \ge 2).$$

Moreover, the difference $w(u) - e^{\gamma}$ changes signs infinitely often as $u \to \infty$.

Proof. By relation (8.1.8), we find that the integral $\int_2^\infty w'(t)dt$ converges absolutely. So

$$w(u) = w(2) + \int_2^u w'(t)dt = w(2) + \int_2^\infty w'(t)dt + O\left(e^{-u\log(u\log u)+O(u)}\right) \quad (u > 2). \tag{8.1.11}$$

In particular, the limit $\lim_{u\to\infty} w(u)$ exists. On the other hand, Theorem 3.3.2 and Mertens' estimate imply that

$$\Phi(x,z) = \frac{e^{-\gamma}x}{\log z}\left\{1 + O\left(\frac{1}{\log z} + e^{-u\log u+O(u)}\right)\right\}.$$

Comparing the above formula with Theorem 8.1.1, we deduce that $\lim_{u\to\infty} w(u) = e^{-\gamma}$. Together with relation (8.1.11), this implies that

$$w(u) = e^{-\gamma} + O\left(e^{-u\log(u\log u)+O(u)}\right) \quad (u \geq 2), \tag{8.1.12}$$

and the first part of the theorem follows. To see the second part, write $E(u) = w(u) - e^{-\gamma}$ and note that relation (8.1.7) can be rewritten as

$$(tE(t))' = E(t-1) \quad (t > 2).$$

Integrating over $t \in [u, \infty)$, we find that

$$uE(u) = -\int_u^\infty E(t-1)dt = -\int_{u-1}^\infty E(t)dt \quad (u \geq 2).$$

Now, if $u_+ = \sup\{u \geq 2 : E(t) > 0 \text{ for all } t \geq u\}$, then we must have that $u_+ = \infty$; otherwise we have that

$$0 < (u_+ + 1)E(u_+ + 1) = -\int_{u_+}^\infty E(t)dt < 0,$$

a contradiction. Similarly, if $u_- = \sup\{u \geq 2 : E(t) < 0 \text{ for all } t \geq u\}$, then we have that $u_- = \infty$. These two facts together imply that E changes sign infinitely often, thus completing the proof of the theorem. □

8.2 Maier's matrix method

In this section, we prove Theorem 8.0.5. The key idea is that primes cannot be simultaneously very well distributed in arithmetic progressions and in short intervals. In order to capture this, we consider the matrix

$$M(m,h,q) = \begin{pmatrix} 1+(m+1)q & 2+(m+1)q & \cdots & h+(m+1)q \\ 1+(m+2)q & 2+(m+2)q & \cdots & h+(m+2)q \\ \vdots & \vdots & & \vdots \\ 1+2mq & 2+2mq & \cdots & h+2mq \end{pmatrix}, \tag{8.2.1}$$

where m, h and q are fixed positive integers. Note that the i-th row of this matrix contains all integers in the short interval $(1+(m+i)q, h+(m+i)q]$, whereas its j-th column contains all integers in the arithmetic progression $\{n \equiv j \pmod q : j+mq < n \leq j+2mq\}$. Now, if we know that primes are well distributed in the arithmetic progressions $\{n \equiv j \pmod q : j+mq < n \leq j+2mq\}$, $1 \leq j \leq h$, $(j,q) = 1$, then we expect that

$$\begin{aligned} \#\{p \text{ prime} : p \text{ appears in } M(m,q,h)\} &\sim \sum_{\substack{1\leq j\leq h \\ (j,q)=1}} \frac{1}{\varphi(q)} \int_{j+mq}^{j+2mq} \frac{dt}{\log t} \\ &\sim \frac{mq}{\varphi(q)\log(mq)} \#\{1 \leq j \leq h : (j,q) = 1\}. \end{aligned}$$

On the other hand, if we know that each short interval $(1+(m+i)q, h+(m+i)q]$, $1 \le i \le m$, contains the expected proportion of primes, then

$$\#\{p \text{ prime} : p \text{ appears in } M(m,q,h)\} \sim \sum_{i=1}^{m} \frac{h}{\log(mq)} = \frac{mh}{\log(mq)}.$$

Therefore, if we can take q and h such that

$$\frac{\#\{1 \le j \le h : (j,q) = 1\}}{hq/\varphi(q)} \ge c > 1 \tag{8.2.2}$$

as $q, h \to \infty$, then we obtain a contradiction. Similarly, if we can take q and h such that

$$\frac{\#\{1 \le j \le h : (j,q) = 1\}}{hq/\varphi(q)} \le c' < 1 \tag{8.2.3}$$

as $q, h \to \infty$, then we obtain a contradiction. Such a modulus q will be provided by the second part of Theorem 8.1.2. Indeed, if we take $q = \prod_{p \le z} p$ for some appropriate $z \ge 1$, then Theorems 8.1.1 and 8.1.2, together with Mertens' estimate $\prod_{p \le z}(1-1/p) \sim e^{-\gamma}/\log z$, imply that (8.2.2) and (8.2.3) both hold for infinitely many values of h. In order to be able to deduce Theorem 8.0.5, we need to be able to show that the primes are well-distributed in the arithmetic progressions $\{n \equiv j \pmod q : j + mq < n \le j + 2mq\}$, for $1 \le j \le h$ with $(j,q) = 1$. The following result ensures that this is indeed the case for infinitely many This is ensured by the

Lemma 8.2.1. *There exists a constant $c > 0$ such that there are infinitely many values of $z \ge 1$ for which the modulus $q = \prod_{p \le z}$ satisfies the estimate*

$$\pi(x;q,a) = \frac{\mathrm{li}(x)}{\varphi(q)} \left\{1 + O\left(x^{-c/\log q} + e^{-c\sqrt{\log x}}\right)\right\} \quad (x \ge q,\ (a,q) = 1).$$

Proof. Without loss of generality, we may assume that $x \ge q^L$ for a sufficiently large L; otherwise, the result follows from the Brun-Titchmarsch inequality. For any $T \ge 1$, we have that

$$\psi(x;q,a) := \sum_{\substack{n \le x \\ n \equiv a \pmod q}} \Lambda(n) = \frac{x}{\varphi(q)} + \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \sum_{\substack{\rho:\, L(\rho,\chi)=0 \\ |\mathrm{Im}(\rho)| \le T}} \frac{x^\rho}{\rho} + O\left(\frac{x \log^2(qx)}{T}\right). \tag{8.2.4}$$

(This is a standard consequence of Perron's formula and the residue theorem in complex analysis; see, for example, [Da, Chapters 17, 19].) Since the zeroes of $L(s,\chi)$ are symmetric about the line $\mathrm{Re}(s) = 1/2$, a consequence of the functional equation for Dirichlet L-functions, relation (8.2.4) can be rewritten as

$$\psi(x;q,a) = \frac{x}{\varphi(q)} + \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \sum_{\substack{\rho:\, L(\rho,\chi)=0 \\ \mathrm{Re}(\rho) \ge 1/2,\, |\mathrm{Im}(\rho)| \le T}} \left(\frac{x^\rho}{\rho} + \frac{x^{1-\rho}}{1-\rho}\right) + O\left(\frac{qx \log^2(qx)}{T}\right). \tag{8.2.5}$$

Now, Linnik (see, for example, [IK, Chapter 18]) showed that there exists an absolute positive constant c_1 such that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \sum_{\substack{\rho:\, L(\rho,\chi)=0 \\ \mathrm{Re}(\rho)\geq\sigma,\, |\mathrm{Im}(\rho)|\leq T}} 1 \ll (qT)^{c_1(1-\sigma)} \quad (\sigma \geq 1/2,\ T \geq 1). \tag{8.2.6}$$

We will show that this estimate, together with an assumption about a certain zero-free region for all $L(s,\chi)$ implies the conclusion of the lemma. More precisely, we assume that there exists some constant $c_2 > 0$ such that

$$L(\sigma + it, \chi) \neq 0 \quad \text{for all } \sigma \geq 1 - \frac{c_2}{\log(q + |t|)} \quad \text{and all } \chi \,(\mathrm{mod}\, q), \tag{8.2.7}$$

that is to say, that there is no Landau-Siegel zero for any Dirichlet character $(\mathrm{mod}\, q)$. Then choosing $T = q^2 e^{\sqrt{\log x}}$ in relation (8.2.4), and then applying (8.2.6) and partial summation implies that

$$\psi(x; q, a) = \frac{x}{\varphi(q)} + O\left(\frac{x e^{-c_3\sqrt{\log x}} + x^{1-c_3/\log q}}{\varphi(q)}\right) \quad (x \geq q^L), \tag{8.2.8}$$

for some positive constant c_3 that depends at most on c_2, provided that L is large in terms of the absolute constant c_1. Partial summation the implies that the conclusion of the lemma holds, provided that q satisfies (8.2.7). So it remains to show that (8.2.7) holds for infinitely many moduli q of the form $q = \prod_{p\leq z} p$.

We know (see, for example, [Da, p. 93]) that for each modulus $q \geq 2$ there is at most one possible counterexample to (8.2.7), that is to say, there exists a constant $c_4 > 0$ which has the following property: there is at most one real non-principal Dirichlet character $(\mathrm{mod}\, q)$ which has at most one zero $\rho = \beta + i\gamma$ with $\beta \geq 1 - c_4/\log(q + |\gamma|)$. Moreover, if such a zero ρ exists, then it is necessarily real and simple, that is to say $\rho = \beta \geq 1 - c_4/\log q$.

Now, consider $q = \prod_{p\leq z} p$ for which an exceptional zero as above exists, say at β. By Betrand's postulate, there exists some $q' = \prod_{p\leq z'} p \geq q$ such that

$$1 - \frac{c_4}{\log q'} \leq \beta \leq 1 - \frac{c_4}{2\log q'}.$$

By the discussion in the previous paragraph, the modulus q' satisfies relation (8.2.7) with $c_2 = c_4/2$. (The character χ induces a character $\chi' \,(\mathrm{mod}\, q')$, which has a zero $\beta \geq 1 - c_4/\log(q')$. Therefore, this is the unique character failing (8.2.7) when $c_2 = c_4$.) In any case, we see that we can construct arbitrarily large moduli of the form $q = \prod_{p\leq z} p$ for which relation (8.2.7) is true. As we saw above, such moduli satisfy the conclusion of the lemma, which completes the proof. □

Lemma 8.2.1 together with the argument we gave in the beginning of the section yield Theorem 8.0.5. We give the complete argument below.

Proof of Theorem 8.0.5. By Remark 8.0.6, it suffices to show the theorem for an unbounded sequence of arbitrarily large values of λ. Fix for the moment some $\lambda \geq 3$. Let $q = \prod_{p\leq z} p$ be

a sufficiently large modulus which satisfies the conclusion of Lemma 8.2.1. Set $m = q^L \geq e^{L^2}$, where L is a large integer, and let $h = \left\lfloor (\log q)^\lambda \right\rfloor \in [z^2, q]$. Consider the matrix $M(m, h, q)$, defined by (8.2.1), and let N be the number of its elements that are prime numbers. Then

$$\begin{aligned} N &= \sum_{\substack{1\leq j\leq h\\(j,q)=1}} \sum_{\substack{j+mq<p\leq j+2mq\\ p\equiv j \,(\mathrm{mod}\, q)}} 1 = \sum_{\substack{1\leq j\leq h\\(j,q)=1}} \left\{ \frac{1}{\varphi(q)} \int_{j+mq}^{j+2mq} \frac{dt}{\log t} + O\left(\frac{mq}{e^{cL}\varphi(q)\log(mq)} \right) \right\} \\ &= \sum_{\substack{1\leq j\leq h\\(j,q)=1}} \frac{mq}{\varphi(q)} \left\{ 1 + O\left(\frac{1}{L\log q} + \frac{1}{e^{cL}} \right) \right\} = \Phi(h,z)\frac{mq}{\varphi(q)} \left\{ 1 + O\left(\frac{1}{\log(mq)} + \frac{1}{e^{cL}} \right) \right\}. \end{aligned}$$

Theorem 8.1.1 and the fact that $w(t) \asymp 1$, for all $t \geq 2$, imply that

$$\Phi(h,z) = \frac{xw(\lambda')}{\log z} \left\{ 1 + O\left(\frac{1}{\log z} \right) \right\},$$

where $\lambda' = \frac{\log h}{\log z} \sim \lambda$, as $z \to \infty$. Since we also have that

$$\frac{\varphi(q)}{q} = \prod_{p\leq z} \left(1 - \frac{1}{p} \right) = \frac{e^{-\gamma}}{\log z} \left\{ 1 + O\left(\frac{1}{\log z} \right) \right\},$$

by Mertens' estimate, we deduce that

$$N = w(\lambda')e^\gamma \cdot \frac{mh}{\log(mq)} \left\{ 1 + O\left(\frac{1}{\log z} + \frac{1}{e^{cL}} \right) \right\}.$$

On the other hand, we have that

$$N = \sum_{k=m+1}^{2m} (\pi(kq+h) - \pi(kq))$$

So, there exists at least one $k \in \{m+1, \ldots, 2m\}$ such that

$$\begin{aligned} \pi(kq+h) - \pi(kq) &\geq w(\lambda')e^\gamma \cdot \frac{h}{\log(mq)} \left\{ 1 - O\left(\frac{1}{\log z} + \frac{1}{e^{cL}} \right) \right\} \\ &= w(\lambda')e^\gamma \cdot \frac{h}{\log(kq)} \left\{ 1 - O\left(\frac{1}{\log z} + \frac{1}{e^{cL}} \right) \right\}, \end{aligned} \tag{8.2.9}$$

and at least one $k' \in \{m+1, \ldots, 2m\}$ such that

$$\pi(k'q+h) - \pi(k'q) \leq w(\lambda')e^\gamma \cdot \frac{h}{\log(kq)} \left\{ 1 + O\left(\frac{1}{\log z} + \frac{1}{e^{cL}} \right) \right\}. \tag{8.2.10}$$

So if we choose $\lambda \geq 3$ with $w(\lambda) > e^{-\gamma}$, by Theorem 8.1.2, and $L = L(\lambda)$ and $z = z(\lambda)$ big enough, then we find that there exists at least one $k \in \{m+1, \ldots, 2m\}$ such that

$$\pi(kq+h) - \pi(kq) \geq \frac{1 + w(\lambda)e^\gamma}{2} \cdot \frac{h}{\log(kq)} > \frac{h}{\log(kq)}.$$

Therefore

$$\limsup_{x\to\infty} \frac{\pi(x+(\log x)^\lambda)-\pi(x)}{(\log x)^{\lambda-1}} \geq \frac{1+w(\lambda)e^\gamma}{2} > 1,$$

which proves the first part of the theorem. The second part follows similarly, by using relation (8.2.10) in place of (8.2.9). □

Chapter 9

The large sieve

In this chapter we will see a quite different approach to sieving, the so-called *large sieve*. There are three versions of it, each suited for different applications. Originally, the large sieve arose as an inequality involving trigonometric polynomials, which explored the idea of quasi-orthogonality. However, for number-theoretic purposes, the power of the large sieve is revealed when stated in its two other forms: the arithmetic version and the character sum version. As we will see, the former allows us to obtain quite good estimates in sieve problems of very large dimension, whereas using the later we can control the average distribution of interesting sets of integers in arithmetic progressions.

We begin with the arithmetic formulation of the large sieve, Theorem 9.1.1, which will allow for a direct comparison with the results of the previous chapters. Then we give the more classical trigonometric version of the large sieve, Theorem 9.2.1, and show how to deduce Theorem 9.1.1 from it. Finally, we conclude with the character sum version in Section 9.3. Arguably the most important application of this last version is the Bombieri-Vinogradov theorem, whose proof we give in the subsequent chapter.

9.1 Arithmetic version and applications

Theorem 9.1.1 (Large sieve - arithmetic version)**.** *Let $\mathcal{N} \subset \{M+1, \dots, M+N\}$. Let $\{R_p : p < z\}$ be a collection of sets such that $R_p \subset \mathbb{Z}/p\mathbb{Z}$ for each prime $p < z$. Then*

$$\#\{n \in \mathcal{N} : n \notin R_p \,(\mathrm{mod}\, p),\ \textit{for all } p < z\} \leq (\pi N + z^2) \Bigg/ \left(\sum_{m<z} \mu^2(m) h(m) \right),$$

where

$$h(m) = \prod_{p \mid m} \frac{|R_p|/p}{1 - |R_p|/p} = \prod_{p \mid m} \frac{|R_p|}{p - |R_p|}.$$

Remark 9.1.2. Let $F(x) \in \mathbb{Z}[x]$ and $\mathcal{N} \subset \{M+1, \dots, M+N\}$. If we set

$$\mathcal{A} = \{F(n) : n \in \mathcal{N}\}$$

and

$$R_p = \{m \in \mathbb{Z}/p\mathbb{Z} : F(m) \equiv 0 \,(\mathrm{mod}\, p)\},$$

then

$$\#\{n \in \mathcal{N} : n \notin R_p \,(\mathrm{mod}\, p),\ \forall p < z\} = \#\{n \in \mathcal{N} : p \nmid F(n),\ \forall p < z\} = S(\mathcal{A}, z).$$

So we see that Theorem 9.1.1 can be translated to a sieve estimate in important cases such as the above one. Moreover, it provides an upper bound on $S(\mathcal{A}, z)$ that is - up to the value of the constant C - as strong as Theorem 5.1.1. Finally, Theorem 9.1.1 has the significant advantage that it does not depend on the assumption of hypotheses such as (A1), (A3) and (R′), and it is particularly strong when the sifting dimension κ becomes unbounded.

The proof of Theorem 9.1.1 will be given in Section 9.2. We give below a couple of applications of it.

Given a prime p, we define

$$n(p) = \min\left\{a \in \mathbb{N} : \left(\frac{a}{p}\right) \neq 1\right\}.$$

It is easy to see that $n(p)$ is a prime number $< p$. The least quadratic non-residue problem asks for estimates on $n(p)$. It is believed that $n(p) \ll_\epsilon (\log p)^{1+\epsilon}$, whereas the Generalized Riemann Hypothesis would imply that $n(p) \ll_\epsilon (\log p)^{2+\epsilon}$. The pointwise bound known is $n(p) \ll_\epsilon p^{1/(4\sqrt{e})+\epsilon}$, for every fixed $\epsilon > 0$. Using Theorem 9.1.1, we will show that for most primes p it is possible to do much better than this bound:

Theorem 9.1.3. *Fix $\epsilon > 0$. Then we have that*

$$\#\{p \leq N : n(p) > N^\epsilon\} \ll_\epsilon 1.$$

Proof. Let $\mathcal{N} = \{m \leq N^2 : P^+(m) \leq N^\epsilon\}$, and note that

$$|\mathcal{N}| \gg_\epsilon N^2 \tag{9.1.1}$$

by Theorem 6.1.1. For each prime $p < N$, let

$$R_p = \begin{cases} \left\{k \in \mathbb{Z}/p\mathbb{Z} : \left(\frac{k}{p}\right) = -1\right\} & \text{if } n(p) > N^\epsilon, \\ \emptyset & \text{otherwise,} \end{cases}$$

so that

$$|R_p| = \begin{cases} \frac{p-1}{2} & \text{if } n(p) > N^\epsilon \text{ and } p \geq 3, \\ 0 & \text{otherwise,} \end{cases}$$

Note that if p is such that $n(p) > N^\epsilon$, then $(\frac{q}{p}) = 1$ for all primes $q \leq N^\epsilon$, and consequently $(\frac{m}{p}) = 1$ for all $m \in \mathbb{N}$. This implies that

$$\mathcal{N} \subset \{m \leq N^2 : m \notin R_p \,(\mathrm{mod}\, p),\ \text{for all } p < N\} \tag{9.1.2}$$

The idea of the proof is that if $R_p \neq 0$ for too many primes $p < N$, then $\mathcal{N}$ will be forced to have abnormally small size, thus contradicting (9.1.1). Indeed, relations (9.1.1) and (9.1.2) imply that

$$N^2 \ll_\epsilon \#\{n \leq N^2 : n \notin R_p \,(\mathrm{mod}\, p),\ \text{for all } p < N\}.$$

So, if we let

$$S = \sum_{m<N} \mu^2(m) \prod_{p|m} \frac{|R_p|}{p - |R_p|},$$

then Theorem 9.1.1 yields the inequality

$$S \ll_\epsilon 1. \tag{9.1.3}$$

On the other hand, we have that

$$S \geq \sum_{p<N} \frac{|R_p|}{p - |R_p|} \geq \sum_{\substack{3\leq p<N \\ n(p)>N^\epsilon}} \frac{p-1}{p+1} \gg \#\{3 \leq p < N : n(p) > N^\epsilon\}.$$

Combining the above relation with (9.1.3) completes the proof of the theorem. □

Corollary 9.1.4. *Fix $\epsilon > 0$. Then we have that*

$$\#\{p \leq N : n(p) > p^\epsilon\} \ll_\epsilon \log\log N.$$

Proof. Exercise. □

Exercise 9.1.5. Fix $\epsilon > 0$. Show that, for every prime p, we have that

$$n(p) \ll_\epsilon p^{1/(2\sqrt{e})+\epsilon}.$$

Hint: Use Theorem A.3.1 and the fact that $(\frac{m}{p}) = 1$ for all m with $P^+(m) \leq n(p)$.

Finally, we give a last application to demonstrate the power of the large sieve when the sifting dimension grows. As a motivation, note that the set of squares occupies exactly $(p+1)/2$ congruence classes modulo each odd prime p, or equivalently, it avoids $(p-1)/2$ congruence classes modulo each odd prime p. Moreover, there are about $\sqrt{N}$ squares of size $\leq N$. Theorem 9.1.1 implies that this is in fact best possible:

Proposition 9.1.6. *Let $R_p \subset \mathbb{Z}/p\mathbb{Z}$ with $|R_p| = p/2 + O(1)$. Then*

$$\#\{n \leq N : n \notin R_p \,(\mathrm{mod}\, p), \text{ for all } p < \sqrt{N}\} \ll \sqrt{N}.$$

Proof. We may assume that $R_p \leq p-1$ for all $p < \sqrt{N}$; else, $R_{p_0} = \mathbb{Z}/p_0\mathbb{Z}$ for some prime p_0, and there are no integers $n \notin R_{p_0} \,(\mathrm{mod}\, p_0)$. Now, in view of Theorem 9.1.1, it suffices to show that

$$\sum_{m<\sqrt{N}} \mu^2(m) \prod_{p|m} \frac{|R_p|}{p - |R_p|} \gg \sqrt{N}. \tag{9.1.4}$$

Since $|R_p| = p/2 + O(1)$, we find that $|R_p|/(p - |R_p|) = 1 + O(1/p)$, and (9.1.4) follows by an application of the convolution method (see Section 0.2). □

9.2 Quasi-orthogonality and the trigonometric version of the large sieve

Fix $N \geq 1$, and consider the space of sequences of complex numbers $\boldsymbol{a} = \{a_n\}_{n=1}^N$ equipped with the inner product

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \sum_{n=1}^{N} a_n \overline{b_n}.$$

Given $x \in \mathbb{R}$, set $e(x) = e^{2\pi i x}$, and let $\|x\|$ denote the distance of x from its nearest integer, that is to say, $\|x\| = \min\{|x - n| : n \in \mathbb{Z}\}$. Note that

$$\sum_{n=1}^{N} e(\alpha n)\overline{e(\beta n)} = \frac{e(\alpha-\beta) - e((N+1)(\alpha-\beta))}{1 - e(\alpha-\beta)} \ll \frac{1}{\|\alpha - \beta\|}.$$

So we see that if $\|\alpha - \beta\|$ is large, then the sequences $\{e(n\alpha)\}_{n=1}^N$ and $\{e(\beta n)\}_{n=1}^N$ are nearly orthogonal.

Motivated by the above observation, we call a set of real numbers $\{\alpha_1, \dots, \alpha_R\}$ *δ-spaced* if $\|\alpha_r - \alpha_s\| \geq \delta$ for all $1 \leq r < s \leq R$. Given such a set, the sequences $\{e(n\alpha_r)\}_{n=1}^N$, $1 \leq r \leq R$, appropriately scaled, form a quasi-orthonormal set. Indeed, we have that

$$\sum_{n=1}^{N} e(n\alpha_r)\overline{e(n\alpha_s)} = \begin{cases} N & \text{if } r = s, \\ O(1/\delta) & \text{if } r \neq s. \end{cases}$$

Then standard facts about Hilbert spaces lead us to the prediction that, given any sequence of complex numbers $\{a_n\}_{n=1}^N$, we should have that

$$\sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 \leq M \sum_{n=1}^{N} |a_n|^2,$$

for some relatively small $M = M(\delta, N)$ that is close to N. The following theorem confirms this guess.

Theorem 9.2.1 (Large sieve - trigonometric version)**.** *Let $\{a_n\}_{n=1}^N$ be a sequence of complex numbers. Consider a set of δ-spaced real numbers $\{\alpha_1, \dots, \alpha_R\}$. Then*

$$\sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 \leq (\pi N + 1/\delta) \sum_{n=1}^{N} |a_n|^2. \tag{9.2.1}$$

Remark 9.2.2. Selberg [Se91] and, independently, Montgomery and Vaughan [MV] showed that the above theorem holds with $N + 1/\delta - 1$ in place of $\pi N + 1/\delta$, which is best possible in this generality, as the two examples below indicate:

- If R is fixed and $a_n = e(-n\alpha_1)$ for all n, then the left hand side of (9.2.1) is $\geq N^2$, whereas the right hand side of (9.2.1), which is asymptotically equal to $(N + 1/\delta - 1)\sum_{n=1} |a_n|^2$. In fact, if $R = 1$ and $\delta = 1$, then we have that

$$\sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 = N^2 = (N + 1/\delta - 1) \sum_{n=1}^{N} |a_n|^2.$$

- The constant $1/\delta$ is necessary: if N is fixed, $\alpha_j = j/R$ for $1 \le j \le R$, and $\delta = 1/R$, then the left hand side of (9.2) multiplied by δ is a Riemann sum for the integral

$$\int_0^1 \left|\sum_{n=1}^N a_n e(n\alpha)\right|^2 d\alpha = \sum_{1\le n,m\le N} a_n\overline{a_m}\int_0^1 e((n-m)\alpha)d\alpha = \sum_{n=1}^N |a_n|^2.$$

  So we have that

$$\delta\sum_{r=1}^R \left|\sum_{n=1}^N a_n e(n\alpha_r)\right|^2 \sim \int_0^1 \left|\sum_{n=1}^N a_n e(n\alpha)\right|^2 d\alpha = \sum_{n=1}^N |a_n|^2$$

  as $\delta \to 0^+$. In fact, if $N = 1$ and $a_1 = 1$, then

$$\sum_{r=1}^R \left|\sum_{n=1}^N a_n e(n\alpha_r)\right|^2 = R = (N + 1/\delta - 1)\sum_{n=1}^N |a_n|^2.$$

There are several approaches to proving Theorem 9.2.1, or other closely related results. We follow an argument due to Gallagher, which is based on the following key lemma.

Lemma 9.2.3. *Let $f : [c - \delta/2, c + \delta/2] \to \mathbb{C}$ be continuously differentiable function. Then*

$$|f(c)| \le \frac{1}{\delta}\int_{c-\delta/2}^{c+\delta/2} |f(t)|dt + \frac{1}{2}\int_{c-\delta/2}^{c+\delta/2} |f'(t)|dt.$$

Proof. We may assume that $c = 0$ and $\delta = 2$; if not, we replace $f(t)$ by $g(t) = f(c + t\delta/2)$. The idea of the proof is that $f(0)$ should be well approximated by the mean value $\frac{1}{2}\int_{-1}^1 f(t)dt$, and the quality of this approximation should be controlled by how large f' changes. Indeed, note that

$$\begin{aligned}\int_{-1}^1 f(t)dt - 2f(0) = \int_{-1}^1 (f(t) - f(0))dt &= f(1) - f(0) + f(-1) - f(0) - \int_{-1}^1 tf'(t)dt \\ &= \int_0^1 f'(t)dt - \int_{-1}^0 f'(t) - \int_{-1}^1 tf'(t)dt \\ &= \int_{-1}^1 \mathrm{sgn}(t)(1 - |t|)f'(t)dt,\end{aligned}$$

where $\mathrm{sgn}(t)$ denotes the sign of t. Consequently,

$$2f(0) = \int_{-1}^1 f(t)dt - \int_{-1}^1 \mathrm{sgn}(t)(1 - |t|)f'(t)dt.$$

Taking absolute values and using the triangle inequality then completes the proof of the theorem. □

Proof of Theorem 9.2.1. For each $r \in \{1, \dots, R\}$, we apply Lemma 9.2.3 with $c = \alpha_r$ and

$$f(t) = \left(\sum_{n=1}^{N} a_n e((n - N/2)t) \right)^2$$

to get that

$$|f(\alpha_r)| \le \frac{1}{\delta} \int_{\alpha_r - \delta/2}^{\alpha_r + \delta/2} |f(t)| dt + \frac{1}{2} \int_{\alpha_r - \delta/2}^{\alpha_r + \delta/2} |f'(t)| dt. \tag{9.2.2}$$

Since the points $\{\alpha_1, \dots, \alpha_r\}$ are δ-spaced, the intervals $(\alpha_r - \delta/2, \alpha_r + \delta/2)$ are disjoint (mod 1). But f is 1-periodic, so summing relation (9.2.2) yields that

$$\sum_{r=1}^{R} |f(\alpha_r)| \le \frac{1}{\delta} \int_0^1 |f(t)| dt + \frac{1}{2} \int_0^1 |f'(t)| dt.$$

In order to complete the proof of the theorem, note that

$$\begin{aligned} \int_0^1 |f(t)| dt = \int_0^1 \left| \sum_{-N/2 < n \le N/2} a_n e((n - N/2)t) \right|^2 &= \sum_{1 \le n, m \le N} a_n \overline{a_m} \int_0^1 e((n-m)t) dt \\ &= \sum_{-N/2 < n \le N/2} |a_n|^2 \end{aligned}$$

and, similarly,

$$\begin{aligned} \int_0^1 |f'(t)| dt &= 2 \int_0^1 \left| \left(\sum_{n=1}^{N} a_n e((n - N/2)t) \right) \left(\sum_{n=1}^{N} 2\pi \left(n - \frac{N}{2} \right) a_n e((n - N/2)t) \right) \right| dt \\ &\le 4\pi \left(\int_0^1 \left| \sum_{n=1}^{N} a_n e(nt) \right|^2 \right)^{1/2} \left(\int_0^1 \left| \sum_{n=1}^{N} \left(n - \frac{N}{2} \right) a_n e(nt) \right|^2 \right)^{1/2} \\ &= 4\pi \left(\sum_{n=1}^{N} |a_n|^2 \right) \left(\sum_{n=1}^{N} \left(n - \frac{N}{2} \right)^2 |a_n|^2 \right)^{1/2} \le 2\pi N \sum_{n=1}^{N} |a_n|^2. \end{aligned}$$

Combining the above estimates completes the proof of the theorem. □

In order to put Theorem 9.2.1 into use for arithmetic applications, we shall pick as our δ-spaced points the Farey fractions

$$\mathscr{F}_Q = \left\{ \frac{a}{q} : 1 \le a \le q \le Q, \ (a, q) = 1 \right\},$$

for some parameter $Q \ge 1$. Note that if a/q and a'/q' are distinct elements of $\mathscr{F}_Q$ written in lowest terms, then

$$\left| \frac{a}{q} - \frac{a'}{q'} + n \right| = \frac{|aq' - aq' + nqq'|}{qq'} \ge \frac{1}{qq'} \ge \frac{1}{Q^2},$$

for every integer n, that is to say, the set $\mathscr{F}_Q$ is $(1/Q^2)$-spaced. So we obtain the following corollary:

Corollary 9.2.4. *Let $Q \geq 1$, and $\{a_n\}_{n=1}^N$ be a sequence of complex numbers. Then we have that*

$$\sum_{q \leq Q} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} \left| \sum_{n=1}^N a_n e(na/q) \right|^2 \leq (\pi N + Q^2) \sum_{n=1}^N |a_n|^2.$$

We now show how Corollary 9.2.4 can be used to obtain control on the average of the error

$$\sum_{\substack{1 \leq n \leq N \\ n \equiv a \, (\mathrm{mod}\, p)}} a_n - \frac{1}{p} \sum_{n=1}^N a_n,$$

over prime moduli p.

Corollary 9.2.5. *Let $Q \geq 1$, and $\{a_n\}_{n=1}^N$ be a sequence of complex numbers. Then we have that*

$$\sum_{p \leq Q} p \sum_{b=1}^p \left| \sum_{\substack{1 \leq n \leq N \\ n \equiv b \, (\mathrm{mod}\, p)}} a_n - \frac{1}{p} \sum_{n=1}^N a_n \right|^2 \leq (\pi N + Q^2) \sum_{n=1}^N |a_n|^2.$$

Proof. For brevity, we write $S(\alpha) = \sum_{n=1}^N a_n e(n\alpha)$. Note that

$$\begin{aligned} \sum_{\substack{1 \leq n \leq N \\ n \equiv b \, (\mathrm{mod}\, p)}} a_n - \frac{1}{p} \sum_{n=1}^N a_n &= \frac{1}{p} \sum_{n=1}^N a_n \sum_{j=1}^p e(j(n-b)/p) - \frac{1}{p} \sum_{n=1}^N a_n \\ &= \frac{1}{p} \sum_{n=1}^N a_n \sum_{j=1}^{p-1} e(j(n-b)/p) = \frac{1}{p} \sum_{j=1}^{p-1} e(-jb/p) S(j/p). \end{aligned}$$

Consequently,

$$\begin{aligned} \sum_{b=1}^p \left| \sum_{\substack{1 \leq n \leq N \\ n \equiv b \, (\mathrm{mod}\, p)}} a_n - \frac{1}{p} \sum_{n=1}^N a_n \right|^2 &= \frac{1}{p^2} \sum_{b=1}^p \sum_{j_1=1}^{p-1} \sum_{j_2=1}^{p-1} e((j_2 - j_1)b/p) S(j_1/p) \overline{S(j_2/p)} \\ &= \frac{1}{p^2} \sum_{j_1=1}^{p-1} \sum_{j_2=1}^{p-1} S(j_1/p) \overline{S(j_2/p)} \sum_{b=1}^p e((j_2 - j_1)b/p) \\ &= \frac{1}{p} \sum_{j=1}^{p-1} |S(j/p)|^2. \end{aligned} \tag{9.2.3}$$

Multiplying the above identity by p, summing the resulting formula over $p \leq Q$, and applying Corollary 9.2.4 completes the proof of the corollary. □

Remark 9.2.6. Note that if we divide the inequality in the statement of Corollary 9.2.5 by N^2, we find that

$$\sum_{p\le Q}\frac{1}{p}\sum_{b=1}^{p}\left|\frac{1}{N/p}\sum_{\substack{1\le n\le N\\ n\equiv b\,(\mathrm{mod}\,p)}}a_n-\frac{1}{N}\sum_{n=1}^{N}a_n\right|^2\le\left(\frac{1}{N}+\frac{Q^2}{N^2}\right)\sum_{n=1}^{N}|a_n|^2.$$

If $|a_n|\le 1$ for all n, then the right hand side of the above inequality is $\le 1+Q^2/N$. Therefore, if $Q=o(N)$, then we find that, for most $p\le Q$,

$$\frac{1}{p}\sum_{b=1}^{p}\left|\frac{1}{N/p}\sum_{\substack{1\le n\le N\\ n\equiv b\,(\mathrm{mod}\,p)}}a_n-\frac{1}{N}\sum_{n=1}^{N}a_n\right|^2=o(1),$$

that is to say, a_n is well-distributed in most progressions $b\,(\mathrm{mod}\,p)$, for most primes $p\le Q$.

We conclude this section with the proof of Theorem 9.1.1.

Proof of Theorem 9.1.1. Set

$$\mathcal{M}=\{M<n\le M+N : n\notin R_p\,(\mathrm{mod}\,p),\ \text{for all } p<z\}.$$

We claim that for all sequences of complex numbers, and for all square-free integers $q<z$, we have that

$$\sum_{\substack{1\le a\le q\\ (a,q)=1}}\left|\sum_{n\in\mathcal{M}}a_n e(an/q)\right|^2\ge h(q)\left|\sum_{n\in\mathcal{M}}a_n\right|^2. \tag{9.2.4}$$

When $q=1$, this holds trivially. For $q>1$, we argue by induction on $\omega(q)$. First, we establish (9.2.4) when $q=p$ is prime. Our starting point is relation (9.2.3), which implies that

$$\sum_{a=1}^{p-1}\left|\sum_{n\in\mathcal{M}}a_n e(an/p)\right|^2=p\sum_{b=1}^{p}\left|\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}}a_n-\frac{1}{p}\sum_{n\in\mathcal{M}}a_n\right|^2.$$

Now, using the identity $|z-w|^2 = |z|^2 + |w|^2 - 2\mathrm{Re}(z\overline{w})$, we find that

$$\begin{aligned}
&\sum_{a=1}^{p-1}\left|\sum_{n\in\mathcal{M}} a_n e(an/p)\right|^2 \\
&\quad = p\sum_{b=1}^{p}\left\{\left|\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}} a_n\right|^2 + \frac{1}{p^2}\left|\sum_{n\in\mathcal{M}} a_n\right|^2 - \frac{2}{p}\mathrm{Re}\left(\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}} a_n \sum_{m\in\mathcal{M}}\overline{a_m}\right)\right\} \\
&\quad = p\sum_{b=1}^{p}\left|\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}} a_n\right|^2 + \left|\sum_{n\in\mathcal{M}} a_n\right|^2 - 2\mathrm{Re}\left(\sum_{n\in\mathcal{M}} a_n \sum_{m\in\mathcal{M}}\overline{a_m}\right) \\
&\quad = p\sum_{b=1}^{p}\left|\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}} a_n\right|^2 - \left|\sum_{n\in\mathcal{M}} a_n\right|^2.
\end{aligned}$$

On the other hand, the Cauchy-Schwarz inequality and the fact that if $b \in R_p\,(\mathrm{mod}\,p)$, then there are no elements of $\mathcal{M}$ that lie in the arithmetic progression $b\,(\mathrm{mod}\,p)$, imply that

$$\left|\sum_{n\in\mathcal{M}} a_n\right|^2 = \left|\sum_{b=1}^{p}\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}} a_n\right|^2 \le (p-|R_p|)\sum_{b=1}^{p}\left|\sum_{\substack{n\in\mathcal{M}\\ n\equiv b\,(\mathrm{mod}\,p)}} a_n\right|^2. \tag{9.2.5}$$

Combining the two last relations, we conclude that

$$\sum_{a=1}^{p-1}\left|\sum_{n\in\mathcal{M}} a_n e(an/p)\right|^2 \ge \left(\frac{p}{p-|R_p|}-1\right)\left|\sum_{n\in\mathcal{M}} a_n\right|^2 = h(p)\left|\sum_{n\in\mathcal{M}} a_n\right|^2,$$

that is to say, (9.2.4) does hold when p is a prime $< z$. Now assume that (9.2.4) holds for all square-free integers $q < z$ with $\omega(q) \le j$, where j is some positive integer. Let q be a square-free integer $< z$ with $\omega(q) = j+1$. Then we may write $q = q_1q_2$ with $\omega(q_i) \le j$ for $i \in \{1,2\}$. Furthermore, note that the set $\{a_1q_2 + a_2q_1 : 1 \le a_i \le q_i,\ (a_i,q_i)\ (i\in\{1,2\})\}$ is a set of representatives for the set of residues $\{1 \le a \le q : (a,q)=1\}$. Hence

$$\sum_{\substack{1\le a\le q\\ (a,q)=1}}\left|\sum_{n\in\mathcal{M}} a_n e(an/q)\right|^2 = \sum_{\substack{1\le a_1\le q_1\\ (a_1,q_1)=1}}\sum_{\substack{1\le a_2\le q_2\\ (a_2,q_2)=1}}\left|\sum_{n\in\mathcal{M}} a_n e\left(\frac{a_1n}{q_1}+\frac{a_2n}{q}\right)\right|^2.$$

For each fixed a_1 as above, we apply (9.2.4) with q_2 in place of q and $a_n e(a_1n/q_1)$ in place of a_n, which holds by the induction hypothesis. So

$$\sum_{\substack{1\le a_2\le q_2\\ (a_2,q_2)=1}}\left|\sum_{n\in\mathcal{M}} a_n e\left(\frac{a_1n}{q_1}+\frac{a_2n}{q}\right)\right|^2 \ge h(q_2)\left|\sum_{n\in\mathcal{M}} a_n e\left(\frac{a_1n}{q_1}\right)\right|^2.$$

Summing the above inequality over a_1 and applying (9.2.4) with q_1 in place of q, which also holds by the induction hypothesis, yields that

$$\sum_{\substack{1\le a\le q\\(a,q)=1}}\left|\sum_{n\in\mathcal{M}}a_n e(an/q)\right|^2 \ge h(q_1)h(q_2)\left|\sum_{n\in\mathcal{M}}a_n\right|^2.$$

Since h is a multiplicative function, we deduce that relation (9.2.4) is true. This completes the inductive step, and hence the proof of (9.2.4). Finally, applying this relation with a_n being the characteristic function of the set $\mathcal{N}$ implies that

$$|\mathcal{N}|^2\sum_{q<z}\mu^2(q)h(q) \le \sum_{q<z}\sum_{\substack{1\le a\le q\\(a,q)=1}}\left|\sum_{n\in\mathcal{M}}e(na/q)\right|^2 \le (\pi N+z^2)|\mathcal{N}|,$$

by Corollary 9.2.4. This completes the proof of Theorem 9.1.1. □

9.3 Character sum version

In this last section, we show another consequence of Corollary 9.2.4, which will play a central role in the proof of the Bombieri-Vinogradov theorem.

Theorem 9.3.1 (Large sieve - character sum version)**.** *Let $\{a_n\}_{n=1}^N$ be a sequence of complex numbers. For every $Q\ge 1$, we have that*

$$\sum_{q\le Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\left|\sum_{n=1}^N a_n\chi(n)\right|^2 \le (N+Q^2)\sum_{n=1}^N|a_n|^2,$$

where the notation $\sum^*$ means that the sum runs over primitive characters only.

Proof. For brevity, we write $S(\alpha)=\sum_{n=1}^N a_n e(n\alpha)$. In order to translate the statement of the theorem to an inequality involving the additive characters $n\to e(an/q)$ and apply Corollary 9.2.4, we use Theorem A.2.2 (i.e. we use Fourier inversion with respect to the additive characters; see Section A.1). This theorem implies that, for every primitive character $\chi\,(\mathrm{mod}\,q)$, we have

$$\sum_{n=1}^N a_n\chi(n)=\sum_{n=1}^N a_n\frac{1}{\tau(\overline{\chi})}\sum_{a=1}^q\overline{\chi}(a)e(an/q)=\frac{1}{\tau(\overline{\chi})}\sum_{a=1}^q\overline{\chi}(a)S(a/q).$$

Since for such a character we also have that $|\tau(\overline{\chi})| = \sqrt{q}$, we find that

$$\begin{aligned}\sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{n=1}^{N} a_n \chi(n)\right|^2 &= \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{a=1}^{q} \overline{\chi}(a) S(a/q)\right|^2 \le \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{a=1}^{q} \overline{\chi}(a) S(a/q)\right|^2 \\ &= \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)} \sum_{a_1}^{q} \sum_{a_2=1}^{q} \overline{\chi}(a_1)\chi(a_2) S(a_1/q)\overline{S(a_2/q)} \\ &= \frac{1}{q} \sum_{a_1}^{q} \sum_{a_2=1}^{q} S(a_1/q)\overline{S(a_2/q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a_1)\chi(a_2) = \frac{\varphi(q)}{q} \sum_{\substack{1 \le a \le q \\ (a,q)=1}} |S(a/q)|^2,\end{aligned}$$

by Theorem A.1.1. Multiplying the above relation by $q/\varphi(q)$, summing the resulting inequality over $q \le Q$, and applying Corollary 9.2.4 completes the proof of the theorem. □

Chapter 10

The Bombieri-Vinogradov theorem

In this chapter we prove the Bombieri-Vinogradov theorem. We shall prove this theorem in a slightly different but equivalent formulation: set

$$E'(x;a) = \max_{(a,q)=1} \max_{y\le x} \left| \sum_{\substack{n\le y \\ n\equiv a \,(\mathrm{mod}\, q)}} \Lambda(n) - \frac{y}{\varphi(q)} \right|,$$

where Λ is the von Mangoldt function. Then we have the following result.

Theorem 10.0.2 (Bombieri-Vinogradov theorem, II)**.** *Fix $A > 0$. There is $B = B(A) > 0$ such that*

$$\sum_{q\le x^{1/2}/(\log x)^B} E'(x;q) \ll_A \frac{x}{(\log x)^A},$$

Exercise 10.0.3. Deduce Theorem 4.0.4 from Theorem 10.0.2.

10.1 Reduction to Dirichlet characters

The first step in the proof of Theorem 10.0.2 is to reduce it to an estimate about Dirichlet characters. Indeed, the main result of this chapter will be the following result:

Theorem 10.1.1 (Bombieri-Vinogradov theorem, III)**.** *Let $1 \le Q \le x^{2/3}$. Then*

$$\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y\le x} \left| \sum_{n\le y} \chi(n)\Lambda(n) \right| \ll (\log x)^6 (x + x^{1/2}Q^2 + x^{4/5}Q^{13/10}),$$

where the notation $\sideset{}{^*}\sum$ means that the sum runs over primitive characters only.

Remark 10.1.2. If $Q \ge x^{3/7}$, then $x + x^{1/2}Q^2 + x^{4/5}Q^{13/10} \asymp x^{1/2}Q^2$. SoTheorem 10.1.1 implies that

$$\sum_{n\le x} \Lambda(n)\chi(n) \ll x^{1/2+\epsilon}$$

for most $q \in (Q, 2Q]$ and for most primitive characters $\chi \pmod q$, an estimate which is as good as the Generalized Riemann Hypothesis.

As we prove below, Theorem 10.0.2 follows by Theorem 10.1.1 and the following fundamental result:

Theorem 10.1.3 (Siegel-Walfisz)**.** *Fix $A > 0$. Let χ be a primitive Dirichlet character $\pmod q$. For $1 \le q \le (\log x)^A$, we have that*

$$\sum_{n \le x} \Lambda(n)\chi(n) = \delta(\chi)x + O_A\left(\frac{x}{e^{c\sqrt{\log x}}}\right),$$

where c is some absolute constant and

$$\delta(\chi) = \begin{cases} 1 & \text{if } \chi = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 10.1.4. As it is well-known, the implied constant in the above theorem cannot be computed effectively due to the potential presence of Landau-Siegel zeroes. This deficiency will be inherited to the Bombieri-Vinogradov theorem, as it will become clear below.

Deduction of Theorem 10.0.2 from Theorem 10.1.1. First, we rewrite $E'(x; q)$ in terms of primitive characters. Note that

$$\sum_{\substack{n \le y \\ (n,q)>1}} \Lambda(n) \le \sum_{p \mid q} \log p \sum_{\substack{m \ge 1 \\ p^m \le y}} 1 \ll \omega(q) \log y \ll (\log q)(\log y). \tag{10.1.1}$$

In particular,

$$\begin{aligned} \sum_{\substack{n \le y \\ (n,q)=1}} \Lambda(n) = \sum_{n \le y} \Lambda(n) + O((\log y)(\log q)) &= y + O\left(\frac{y}{e^{c\sqrt{\log y}}} + O((\log y)(\log q))\right) \\ &=: y + O(R_q(y)). \end{aligned}$$

So we have that

$$\begin{aligned} \sum_{\substack{n \le y \\ n \equiv a \pmod q}} \Lambda(n)\chi(n) - \frac{y}{\varphi(q)} &= \sum_{\substack{n \le y \\ n \equiv a \pmod q}} \Lambda(n)\chi(n) - \frac{1}{\varphi(q)} \sum_{\substack{n \le y \\ (n,q)=1}} \Lambda(n) + O(R_q(y)) \\ &= \sum_{n \le y} \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \overline{\chi}(a)\chi(n) - \frac{1}{\varphi(q)} \sum_{n \le y} \Lambda(n)\chi_0(n) + O(R_q(y)) \\ &\frac{1}{\varphi(q)} \sum_{\substack{\chi \pmod q \\ \chi \ne \chi_0}} \overline{\chi}(a) \sum_{n \le y} \Lambda(n)\chi(n) + O(R_q(y)). \end{aligned}$$

by Theorem A.1.1, where χ_0 denotes the principal character. Given a non-principal character $\chi \pmod q$, let χ' be the primitive character which induces it, say of conductor $d > 1$. Then

$$\left|\sum_{n\le y}\Lambda(n)\chi(n) - \sum_{n\le y}\Lambda(n)\chi'(n)\right| \le \sum_{\substack{n\le y\\(n,q)>1}}\Lambda(n)\chi(n) \ll (\log q)(\log y),$$

by (10.1.1). Therefore, if we set

$$R'_q(y) = \frac{|R_q(y)|}{\varphi(q)} + (\log q)(\log y),$$

then

$$\begin{aligned}\sum_{\substack{n\le y\\ n\equiv a \pmod q}}\Lambda(n)\chi(n) - \frac{y}{\varphi(q)} &= \frac{1}{\varphi(q)}\sum_{\substack{\chi \pmod q\\ \chi\neq\chi_0}}\overline{\chi}(a)\sum_{n\le y}\Lambda(n)\chi'(n) + O(R'_q(y))\\ &= \frac{1}{\varphi(q)}\sum_{\substack{d\mid q\\ d>1}}\sideset{}{^*}\sum_{\chi' \pmod d}\sum_{n\le y}\Lambda(n)\chi'(n)\sum_{\substack{\chi \pmod q\\ \chi \text{ is induced by } \chi'}}\overline{\chi}(a) + O(R'_q(y)).\end{aligned}$$

Since, each primitive character $\chi' \pmod d$ induces at most one character $\chi \pmod q$, taking absolute values we deduce that

$$\begin{aligned}\sum_{q\le x^{1/2}/(\log x)^B}E'(x;q) &\le \sum_{q\le x^{1/2}/(\log x)^B}\frac{1}{\varphi(q)}\sum_{\substack{d\mid q\\ d>1}}\sideset{}{^*}\sum_{\chi' \pmod d}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|\\ &\quad + O\left(\sum_{q\le x^{1/2}/(\log x)^B}R'_q(x)\right)\\ &= \sum_{1<d\le x^{1/2}/(\log x)^B}\sideset{}{^*}\sum_{\chi' \pmod d}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|\sum_{\substack{q\le x^{1/2}/(\log x)^B\\ d\mid q}}\frac{1}{\varphi(q)}\\ &\quad + O\left(\frac{x(\log x)}{e^{c\sqrt{\log x}}}\right).\end{aligned}$$

For the sum over q, note that $\varphi(ab) \ge \varphi(a)\varphi(b)$ for all a and b, and consequently

$$\sum_{\substack{q\le x^{1/2}/(\log x)^B\\ d\mid q}}\frac{1}{\varphi(q)} = \sum_{m\le x^{1/2}/(d(\log x)^B)}\frac{1}{\varphi(dm)} \le \frac{1}{\varphi(d)}\sum_{m\le x^{1/2}/(d(\log x)^B)}\frac{1}{\varphi(m)} \ll \frac{\log x}{\varphi(d)}.$$

Consequently,

$$\begin{aligned}
&\sum_{q\le x^{1/2}/(\log x)^B} E'(x;q)\\
&\ll \sum_{1<d\le x^{1/2}/(\log x)^B} \frac{\log x}{\varphi(d)} \sideset{}{^*}\sum_{\chi' \,(\mathrm{mod}\, d)} \max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right| + O\left(\frac{x(\log x)}{e^{c\sqrt{\log x}}}\right)\\
&\le \sum_{1\le 2^k\le x^{1/2}/(\log x)^B} \frac{\log x}{2^k} \sum_{2^k<d\le 2^{k+1}} \frac{d}{\varphi(d)} \sideset{}{^*}\sum_{\chi' \,(\mathrm{mod}\, d)} \max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right| + O\left(\frac{x(\log x)}{e^{c\sqrt{\log x}}}\right)\\
&\ll \max_{1\le Q\le x^{1/2}/(\log x)^B} \frac{(\log x)^2}{Q} \sum_{Q<d\le 2Q} \frac{d}{\varphi(d)} \sideset{}{^*}\sum_{\chi' \,(\mathrm{mod}\, d)} \max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right| + O\left(\frac{x(\log x)}{e^{c\sqrt{\log x}}}\right).
\end{aligned}$$

If $Q \le (\log x)^{A+8}$, then we apply Theorem 10.1.3 to find that

$$\sum_{Q<d\le 2Q} \frac{d}{\varphi(d)} \sideset{}{^*}\sum_{\chi' \,(\mathrm{mod}\, d)} \max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right| \ll_A \frac{xQ^2}{e^{c\sqrt{\log x}}} \ll_A \frac{xQ}{(\log x)^{A+2}}.$$

Otherwise, if $(\log x)^{A+8} \le Q \le x^{1/2}/(\log x)^B$, then Theorem 10.1.1 implies that

$$\sum_{Q<d\le 2Q} \frac{d}{\varphi(d)} \sideset{}{^*}\sum_{\chi' \,(\mathrm{mod}\, d)} \max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right| \ll (\log x)^6(x + x^{1/2}Q^2 + x^{4/5}Q^{13/10}) \ll \frac{xQ}{(\log x)^{A+2}}.$$

provided that $B \ge A + 8$. So selecting $B = A + 8$ completes the deduction of Theorem 10.0.2. □

10.2 Vaughan's identity

Before we delve into the details of the proof of Theorem 10.1.1, we discuss briefly a possible strategy for proving it. In view of Theorem 9.3.1, a plausible thing to do would be to apply the Cauchy-Schwarz inequality. Ignoring the maximum over $y \le x$ for the moment, this inequality and Theorem 9.3.1 yield that

$$\begin{aligned}
\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{n\le x}\chi(n)\Lambda(n)\right| &\ll Q\left(\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{n\le x}\chi(n)\Lambda(n)\right|^2\right)^{\frac{1}{2}}\\
&\ll Q(Q+\sqrt{x})\sqrt{x}\log x.
\end{aligned}$$

This is barely not sufficient for deducing Theorem 10.0.2: indeed, the above inequality can be rewritten as

$$\frac{1}{Q}\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{n\le x}\chi(n)\Lambda(n)\right| \ll Q\sqrt{x\log x} + x\sqrt{\log x}.$$

However, the right hand side in the above estimate is never $\ll x/(\log x)^A$, a crucial ingredient in the deduction of Theorem 10.0.2.

We will see that the above approach can only work if instead of $\Lambda(n)$ we have weights that have a certain bilinear structure, that is to say weights of the form $\sum_{k\ell=n} a_k b_\ell$, where a_k is supported on integers $k \asymp K$ and b_ℓ is supported on integers $\ell \asymp L$, where we also have that $KL = x$. Indeed, the key idea in proving Theorem 10.1.1 is to decompose Λ as the sum of convolutions $f * g$ of some arithmetic functions f and g which have one of the two following properties: either f is supported on small integers and g is a nice smooth function, such as $g = 1$ or $g = \log$, or they are of the form $f * g$ where both f and g are supported on large integers. In the first case, we take advantage of the cancellation coming from the sums $\sum_{n\leq t} \chi(n)g(n)$, which is a consequence of the smoothness of g and of the Pólya-Vinogradov inequality (see Theorem A.3.1). In the second case, we argue as in the previous paragraph, applying the Cauchy-Schwarz inequality together with the character sum version of the Large Sieve.

The aforementioned decomposition is given by the following lemma due to Vaughan. Note that the first two sums appearing are of the first kind (which are often referred to in the literature as *Type I sums*), whereas the third sum is of the second kind (which are often referred to in the literature as *Type II sums*).

Lemma 10.2.1 (Vaughan's identity)**.** *Let $U \geq 1$ and $V \geq 1$ be two parameters. For any $n > U$, we have that*

$$\Lambda(n) = \sum_{\substack{ab=n\\ a\leq V}} \mu(a)\log b - \sum_{\substack{ab=n\\ a\leq UV}} \left(\sum_{\substack{cd=a\\ c\leq V,\, d\leq U}} \mu(c)\Lambda(d) \right) - \sum_{\substack{ab=n\\ a>U,\, b>V}} \Lambda(a) \left(\sum_{\substack{d\mid b\\ d\leq V}} \mu(d) \right)$$

Proof. Note that

$$\Lambda(n) = \sum_{k\ell=n} \mu(k)\log \ell = \sum_{\substack{k\ell=n\\ k\leq V}} \mu(k)\log \ell + \sum_{\substack{k\ell=n\\ k>V}} \mu(k)\log \ell.$$

Moreover,

$$\begin{aligned}
\sum_{\substack{k\ell=n\\ k>V}} \mu(k)\log \ell &= \sum_{\substack{k\ell=n\\ k>V}} \mu(k) \sum_{m\mid \ell} \Lambda(m) = \sum_{m\mid n} \Lambda(m) \sum_{\substack{k\mid n/m\\ k>V}} \mu(k) \\
&= \sum_{\substack{mr=n\\ r>1}} \Lambda(m) \sum_{\substack{k\mid r\\ k>V}} \mu(k) = -\sum_{\substack{mr=n\\ r>1}} \Lambda(m) \sum_{\substack{k\mid r\\ k\leq V}} \mu(k) \\
&= -\sum_{\substack{mr=n\\ m\leq U}} \Lambda(m) \sum_{\substack{k\mid r\\ k\leq V}} \mu(k) - \sum_{\substack{mr=n\\ m>U,\, r>1}} \Lambda(m) \sum_{\substack{k\mid r\\ k\leq V}} \mu(k) \\
&= -\sum_{\substack{mk\ell=n\\ k\leq V,\, m\leq U}} \mu(k)\Lambda(m) - \sum_{\substack{mr=n\\ m>U,\, r>V}} \Lambda(m) \sum_{\substack{k\mid r\\ k\leq V}} \mu(k),
\end{aligned}$$

which completes the proof of the Lemma. □

Remark 10.2.2. Lemma 10.2.1 can be also proven starting from the straightforward identity

$$\frac{-\zeta'}{\zeta} = F - \zeta FG - \zeta' G + \left(-\frac{\zeta'}{\zeta} - F\right)(1 - \zeta G),$$

where ζ is the Riemann ζ function,

$$F(s) = \sum_{n \le U} \frac{\Lambda(n)}{n^s} \quad \text{and} \quad G(s) = \sum_{m \le V} \frac{\mu(m)}{m^s}.$$

10.3 The smooth part of von Mangoldt's function

In this section we show how to handle the Type I sums that appear in the decomposition of Λ given in Lemma 10.2.1. We start with the following general lemma.

Lemma 10.3.1 (Estimates for Type I sums)**.** *Let χ be a non-principal character $(\mathrm{mod}\, q)$. Also, let $f : \mathbb{N} \to \mathbb{C}$ be an arithmetic function that is supported on integers $d \le D$ and which satisfies the inequality $|f| \le \log^r$, for some $r \ge 0$. Then, for any $s \ge 0$, we have that*

$$\sum_{n \le x} (f * \log^s)(n)\chi(n) \ll D\sqrt{q}(\log(Dqx))^{r+s+1} \quad (x \ge 2).$$

Remark 10.3.2. The above lemma yields a non-trivial result as soon as $x > (D\sqrt{q})^{1+\epsilon}$.

Proof of Lemma 10.3.1. Note that

$$\sum_{n \le x} (f * \log^s)(n)\chi(n) = \sum_{ab \le x} f(a)\chi(a)(\log b)^s \chi(b) = \sum_{a \le x} f(a)\chi(a) \sum_{b \le x/a} \chi(b)(\log b)^s$$
$$\ll \sum_{a \le D} (\log a)^r \left| \sum_{b \le x/a} \chi(b)(\log b)^s \right|.$$

Now, for every $y \ge 1$, the Pólya-Vinogradov inequality (i.e. Theorem A.3.1) and partial summation imply that

$$\sum_{b \le y} \chi(b)(\log b)^s = \int_1^y (\log t)^s d\left(\sum_{b \le t} \chi(b)\right) = (\log y)^s \sum_{b \le y} \chi(b) - s \int_1^y \frac{(\log t)^{s-1}}{t} \left(\sum_{b \le t} \chi(b)\right) dt$$
$$\ll \sqrt{q}(\log q)\left((\log y)^s + s \int_1^y \frac{(\log t)^{s-1}}{t} dt\right) \ll \sqrt{q}(\log q)(\log y)^s.$$

So

$$\sum_{n \le x} (f * \log^s)(n)\chi(n) \ll \sum_{a \le D} (\log a)^r \sqrt{q}(\log q)\left(\log \frac{x}{a}\right)^s \ll D\sqrt{q}(\log q)(\log Dx)^{r+s},$$

which completes the proof of the lemma. □

Now, fix for the moment $1 \le U, V \le x$, to be chosen later. Let $y \le x$ and χ be a non-principal Dirichlet character modulo some integer $q \le x$. Then, we have that

$$\sum_{n \le \min\{y,U\}} \Lambda(n)\chi(n) \ll U,$$

and

$$\sum_{U < n \le y} \chi(n) \sum_{\substack{ab=n \\ a \le V}} \mu(a) \log b \ll V\sqrt{q}(\log x)^2$$

by Lemma 10.3.1. Moreover, since

$$\left| \sum_{\substack{cd=a \\ c \le V,\, d \le U}} \mu(c)\Lambda(d) \right| \le \sum_{d \mid a} \Lambda(d) = \log a,$$

applying Lemma 10.3.1 again, we deduce that

$$\sum_{U < n \le y} \chi(n) \sum_{\substack{ab=n \\ a \le UV}} \left(\sum_{\substack{cd=a \\ c \le V,\, d \le U}} \mu(c)\Lambda(d) \right) \ll UV\sqrt{q}(\log x)^2.$$

Combining the above estimates with Lemma 10.2.1, we conclude that

$$\begin{aligned} \sum_{n \le y} \Lambda(n)\chi(n) &= \sum_{\substack{ab \le y \\ a > U,\, b > V}} \chi(ab)\Lambda(a) \left(\sum_{\substack{d \mid b \\ d \le V}} \mu(d) \right) + O(UV\sqrt{q}(\log x)^2) \\ &= \sum_{\substack{mn \le y \\ m > U,\, n > V}} \chi(mn)\Lambda(m) \left(\sum_{\substack{d \mid n \\ d \le V}} \mu(d) \right) + O(UV\sqrt{q}(\log x)^2), \end{aligned} \tag{10.3.1}$$

for all $y \le x$. Consequently, if we set

$$\alpha_m = \Lambda(m) \quad \text{and} \quad \beta_n = \sum_{\substack{d \mid n \\ d \le V}} \mu(d),$$

then, if we set

$$S(x;Q) = \sum_{Q < q \le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y \le x} \left| \sum_{n \le y} \chi(n)\Lambda(n) \right|,$$

we have that

$$S(x;Q) \ll \sum_{Q < q \le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y \le x} \left| \sum_{\substack{mn \le y \\ m > U,\, n > V}} \chi(mn)\alpha_m\beta_n \right| + UVQ^{5/2}(\log x)^2. \tag{10.3.2}$$

This reduces Theorem 10.1.1 to a bilinear sum estimate,w which will be accomplished in the next section.

10.4 The bilinear part of von Mangoldt's function

We will treat the sum on the right hand side of (10.3.2) using Cauchy-Schwarz and Theorem 9.3.1. However, before we can apply Cauchy-Schwarz, we need to separate the variables m and n. Before we do this, we perform an initial step, which allows us to partially keep track of the fact that $mn \le x$: we break the range of summation of m in relation 10.3.2 into $O(\log x)$ dyadic intervals $(M, 2M]$, and the range of summation of n into $O(\log x)$ dyadic intervals $(N, 2N]$, where $U \le M \le x/N \le x/V$. Then taking the maximum over all these $O(\log x)^2$ possibilities, we find that

$$S(x;Q) \ll (\log x)^2 \max_{\substack{M\ge U,\, N\ge V\\ MN\le x}} \sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y\le x} \left| \sum_{\substack{mn\le y\\ M<m\le 2M\\ N<n\le 2N}} \chi(mn)\alpha_m\beta_n \right| + UVQ^{5/2}(\log x)^2, \tag{10.4.1}$$

We are now ready to separate m and n. This is accomplished by applying Perron's inversion formula in the following form (see [Da, p. 105-6] for a proof of it).

Lemma 10.4.1. *Let*

$$I(z) = \begin{cases} 0 & \text{if } 0 < z < 1,\\ 1/2 & \text{if } z = 1,\\ 1 & \text{if } z > 1. \end{cases}$$

Then, for $z > 0$, $c > 0$ and $T \ge 1$, we have that

$$\left| \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{z^s}{s} ds - I(z) \right| \le \begin{cases} z^c \min\{1, 1/(T|\log z|)\} & \text{if } z \ne 1,\\ c/T & \text{if } z = 1. \end{cases}$$

Now, let $y \le x$, and note that

$$\sum_{\substack{mn\le y\\ M<m\le 2M\\ N<n\le 2N}} \chi(mn)\alpha_m\beta_n = \sum_{\substack{mn\le \lfloor y\rfloor+1/2\\ M<m\le 2M\\ N<n\le 2N}} \chi(mn)\alpha_m\beta_n.$$

So Lemma 10.4.1 implies that

$$\begin{aligned} \sum_{\substack{mn\le y\\ M<m\le 2M\\ N<n\le 2N}} \chi(mn)\alpha_m\beta_n &= \frac{1}{2\pi i} \int_{1/2-ix^2}^{1/2+ix^2} \sum_{\substack{M<m\le 2M\\ N<n\le 2N}} \frac{\chi(mn)\alpha_m\beta_n}{(mn)^s} \frac{(\lfloor y\rfloor + 1/2)^s}{s} ds \\ &\quad + O\left(\frac{y^{1/2}}{x^2} \sum_{m\le 2M,\, n\le 2N} \frac{|\alpha_m\beta_n|}{\sqrt{mn}|\log \frac{\lfloor y\rfloor+1/2}{mn}|} \right) \\ &\ll \sqrt{y} \int_{-x^2}^{x^2} \left| \sum_{M<m\le 2M} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right| \cdot \left| \sum_{N<n\le 2N} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right| \frac{dt}{1+|t|} + 1, \end{aligned}$$

since $MN \le x$, $|\log \frac{\lfloor y\rfloor+1/2}{mn}| \gg 1/y$, $|\alpha_m| \le \log m$, and $|\beta_n| \le \tau(n) \ll n^{1/3}$, for all m and n. Therefore

$$\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y\le x} \left| \sum_{\substack{mn\le y\\ M<m\le 2M\\ N<n\le 2N}} \chi(mn)\alpha_m\beta_n \right|$$
$$\ll \sqrt{x} \int_{-x^2}^{x^2} \sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{U<m\le x/V} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right| \cdot \left| \sum_{V<n\le x/U} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right| \frac{dt}{1+|t|} + Q^2.$$

Inserting the above estimate into (10.3.2), and majoring the integrand by its maximum over all $t \in [-x^2, x^2]$, we deduce that

$$S(x;Q) \ll \sqrt{x}(\log x)^3 \max_{\substack{M\ge U,\, N\ge V\\ MN\le x,\, |t|\le x^3}} \sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{U<m\le x/V} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right| \cdot \left| \sum_{V<n\le x/U} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right|$$
$$+ UVQ^{5/2}(\log x)^2.$$

Finally, Theorem 9.3.1 implies that

$$\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{M<m\le 2M} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right|^2 \ll (M+Q^2) \sum_{M<m\le 2M} \frac{|\alpha_m|^2}{m} \ll (M+Q^2)(\log x)^2,$$

since $|\alpha_m| \le \log m$ for all m, and similarly

$$\sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{N<n\le 2N} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right|^2 \ll (N+Q^2) \sum_{N<n\le 2N} \frac{|\beta_n|^2}{n} \ll (N+Q^2)(\log x)^4,$$

since $|\beta_n| \le \tau(n)$ for all n. So the Cauchy-Schwarz inequality yields that

(10.4.2)
$$S(x;Q) \ll \sqrt{x}(\log x)^6 \max_{\substack{M\ge U,\, N\ge V\\ MN\le x}} \sqrt{(M+Q^2)(N+Q^2)} + UVQ^{5/2}(\log x)^2$$
$$\ll \sqrt{x}(\log x)^6 \max_{\substack{M\ge U,\, N\ge V\\ MN\le x}} \Big(\sqrt{MN} + Q^2 + Q(\sqrt{M}+\sqrt{N})\Big) + UVQ^{5/2}(\log x)^2$$
$$\ll \sqrt{x}(\log x)^6 \left(\sqrt{x} + Q^2 + Q\sqrt{\frac{x}{U}} + Q\sqrt{\frac{x}{V}}\,\right) + UVQ^{5/2}(\log x)^2,$$

since $M \le x/V$ and $N \le x/U$, for all M and N as above. For $Q \le x^{1/2}$, we choose $U = V = x^{2/5}/Q^{3/5}$, so that (10.4.2) becomes

$$S(x;Q) \ll (\log x)^6 (x + x^{1/2}Q^2 + x^{4/5}Q^{13/10}).$$

This completes the proof of Theorem 10.1.1, and hence of the Bombieri-Vinogradov theorem.

Exercise 10.4.2. Show that

$$S(x;Q) \ll (\log x)^6(x + x^{1/2}Q^2 + x^{5/6}Q) \quad (1 \le Q \le x^{2/3}),$$

which is an improvement over Theorem 10.1.1 when $Q \ge x^{1/9}$.

Hint: Write

$$\sum_{\substack{ab=n\\ a\le UV}} \left(\sum_{\substack{cd=a\\ c\le V,\, d\le U}} \mu(c)\Lambda(d) \right) = \sum_{\substack{ab=n\\ a\le U}} \left(\sum_{\substack{cd=a\\ c\le V,\, d\le U}} \mu(c)\Lambda(d) \right) + \sum_{\substack{ab=n\\ U<a\le UV}} \left(\sum_{\substack{cd=a\\ c\le V,\, d\le U}} \mu(c)\Lambda(d) \right).$$

Appendix A

Dirichlet characters

In this chapter we gather some basic facts about an important class of multiplicative functions, the Dirichlet characters. In general, a Dirichlet character modulo some integer q is a completely multiplicative function[1] $\chi : \mathbb{N} \to \mathbb{C}$ such that:

- χ is q-periodic, that is to say, $\chi(n+q) = \chi(n)$, for all $n \in \mathbb{N}$;
- χ is supported exactly on these integers that are co-prime to q, that is to say, $\chi(n) \neq 0$ if and only if $(n, q) = 1$.

The two above facts imply that χ induces canonically a group homomorphism $\widetilde{\chi} : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C} \setminus \{0\}$, simply by letting $\widetilde{\chi}(n \,(\mathrm{mod}\, q)) = \chi(n)$. In group theoretic terms, $\widetilde{\chi}$ is a character of the abelian group $(\mathbb{Z}/q\mathbb{Z})^\times$.

The above discussion gives a natural correspondence between Dirichlet characters mod q and characters[2] of the group $(\mathbb{Z}/q\mathbb{Z})^\times$. In the next section, we develop the basics of character theory of finite abelian groups, and then discuss it further in the case of the groups $(\mathbb{Z}/q\mathbb{Z})^\times$ and $\mathbb{Z}/q\mathbb{Z}$.

A.1 Fourier analysis on finite abelian groups

Let $(G, \cdot)$ be a finite abelian group. We denote with $C(G)$ its set of characters, that is to say, the set of group homomorphisms $\chi : G \to \mathbb{C} \setminus \{0\}$. The set $C(G)$ forms naturally a group with respect to the usual multiplication of complex valued functions. Its identity element is called the principal character of G. It is denoted by χ_0, and it is constantly equal to 1. The cardinality of $C(G)$ is precisely equal to the cardinality of G. This relation is obvious if G is cyclic: if $G = \mathbb{Z}/q\mathbb{Z}$, then every character is uniquely determined by its value at 1. Since $\chi(1)^q = \chi(q \cdot 1) = \chi(0) = 1$, it follows that $\chi(1)$ has to be a q-th root of unity, and consequently, there are precisely q characters. In the general case of a finite abelian group,

[1] That is to say, $\chi(mn) = \chi(m)\chi(n)$ for all integers m and n.

[2] In fact, this is historically the first occurrence of what in modern algebraic language is called group character. Note that, strictly speaking, $\widetilde{\chi}$ is a 1-dimensional representation of $(\mathbb{Z}/q\mathbb{Z})^\times$. However, since this group is abelian, there are no higher dimensional irreducible representations, so the set of irreducible representations of $(\mathbb{Z}/q\mathbb{Z})^\times$ is isomorphic with the set of characters of $(\mathbb{Z}/q\mathbb{Z})^\times$.

the relation $|C(G)| = |G|$ follows by writing G as the direct product of cyclic groups, say

$$G \simeq \mathbb{Z}/q_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/q_k\mathbb{Z}, \tag{A.1.1}$$

and noting that, as a consequence,

$$C(G) \simeq C(\mathbb{Z}/q_1\mathbb{Z}) \oplus \cdots \oplus C(\mathbb{Z}/q_k\mathbb{Z}).$$

The characters of an abelian group G satisfy the following important orthogonality relations.

Theorem A.1.1. *Let $(G, \cdot)$ be a finite abelian group. For every $\chi \in C(G)$, we have that*

$$\frac{1}{|G|}\sum_{g\in G}\chi(g) = \begin{cases} 1 & \text{if } \chi = \chi_0, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1.2}$$

Also, for every $g \in G$, we have that

$$\frac{1}{|G|}\sum_{\chi\in C(G)}\chi(g) = \begin{cases} 1 & \text{if } g = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{A.1.3}$$

Proof. First, we prove relation (A.1.2). If $\chi = \chi_0$, then (A.1.2) is trivially true. Now assume that $\chi \neq \chi_0$. For every $h \in G$, we have that $hG = G$. So we have that

$$\chi(h)\sum_{g\in G}\chi(g) = \sum_{g\in G}^{q}\chi(hg) = \sum_{g\in G}\chi(g). \tag{A.1.4}$$

Since $\chi \neq \chi_0$, there must be some $g \in G$ with $\chi(g) \neq 1$. Together with (A.1.4), this completes the proof of (A.1.2).

The proof of relation (A.1.3) is very similar. This relation is trivial when $g = 1$. Now, if $g \neq 1$, then there exists some character $\chi_1 \in C(G)$ such that $\chi_1(g) \neq 1$. Indeed, if G is cyclic, say $G = \mathbb{Z}/q\mathbb{Z}$, then this equivalent to saying that, for every $n \not\equiv \pmod q$, there is some $a \in \mathbb{Z}$ such that $an \not\equiv 0 \pmod q$, which is true. In general, the existence of χ_1 follows by the cyclic case and relation (A.1.1). Now, we have that

$$\chi_1(g)\sum_{\chi\in C(G)}\chi(g) = \sum_{\chi\in C(G)}\chi_1\chi(g) = \sum_{\chi\in C(G)}\chi(g). \tag{A.1.5}$$

Since $\chi_1(g) \neq 1$ by assumption, relation (A.1.5) completes the proof of (A.1.3), and hence of the theorem. □

By (A.1.2), the characters of a group G form an orthonormal set in the space of functions $\alpha : G \to \mathbb{C}$ with respect to the inner product

$$\langle \alpha, \beta \rangle_G = \frac{1}{|G|}\sum_{g\in G}\alpha(g)\overline{\beta}(g).$$

In particular, for every $\alpha : G \to \mathbb{C}$, we have the inversion formula

$$\alpha = \sum_{\chi \in C(G)} \langle \alpha, \chi \rangle_G \cdot \chi. \tag{A.1.6}$$

This relation allows us to do Fourier analysis on the group G.

Now, we consider the case of the group $\mathbb{Z}/q\mathbb{Z}$. As we mentioned above, the characters of this group are the functions $n \to e(an/q)$, where

$$e(x) := e^{2\pi i x} \quad \text{for } x \in \mathbb{R}.$$

These functions are also called *additive characters* $(\mathrm{mod}\, q)$, as opposed to the Dirichlet characters $(\mathrm{mod}\, q)$, which are also called *multiplicative characters* $(\mathrm{mod}\, q)$. Of particular importance is the interaction between these two different kind of characters. To this end, given a function $f : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$, we define its Fourier transform by

$$\widehat{f}(n) = \sum_{a=1}^{q} f(a) e(-an/q).$$

In order to study the Fourier transform of a Dirichlet character χ, we define its *Gauss sum* $\tau(\chi)$ by the formula

$$\tau(\chi) = \sum_{a=1}^{q} \chi(a) e(a/q).$$

Note that we have the relation

$$\sum_{a=1}^{q} \overline{\chi}(a) e(an/q) = \tau(\overline{\chi})\chi(n), \quad \text{whenever } (n,q) = 1, \tag{A.1.7}$$

or equivalently,

$$\overline{\widehat{\chi}(n)} = \tau(\overline{\chi})\chi(n), \quad \text{whenever } (n,q) = 1.$$

This gives the Fourier transform of χ in terms of the additive characters mod q for the frequencies n that are co-prime to q. In the next section we shall see that this formula can be expanded to all n for an important class of Dirichlet characters, the primitive characters, thus showing that a primitive Dirichlet character $\chi\,(\mathrm{mod}\, q)$ is a conjugate eigenvector of the Fourier transform $(\mathrm{mod}\, q)$ [3], with conjugate eigenvalue equal to $\tau(\overline{\chi})$.

A.2 Primitive characters

Two important notions concerning Dirichlet characters is the notion of a primitive character and the notion of the conductor of a character. They express the fact that a Dirichlet character $(\mathrm{mod}\, q)$ might be a Dirichlet character a smaller modulus $q'|q$ in disguise. The smallest such q' is called the conductor of q. If it happens that the conductor of χ is equal

[3] That is to say, it is an eigenvector of the conjugate of the Fourier transform $(\mathrm{mod}\, q)$

to q, that is to say, χ is a genuine character $(\mathrm{mod}\, q)$, then we say that χ is primitive. We give the formal definitions below.

Let $q_1|q_2$ and consider two Dirichlet characters χ_1 and χ_2, modulo q_1 and q_2, respectively. We say that χ_1 *induces* χ_2 if

$$\chi_2(n) = \begin{cases} \chi_1(n) & \text{if } (n, q_2) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Note that a Dirichlet character always induces itself.

Let χ be a Dirichlet character $(\mathrm{mod}\, q)$, and define $c(\chi)$ to be the smallest positive integer $q'|q$ for which there exists a Dirichlet character $\chi'\,(\mathrm{mod}\, q')$ which induces χ. Then $c(\chi)$ is called the conductor of χ. Finally, χ is called primitive if $c(\chi) = q$.

Exercise A.2.1. Let χ be a Dirichlet character $(\mathrm{mod}\, q)$.

(a) Show that χ is primitive if and only if for every natural number $q_1 < q$ there are integers m and n such that $m \equiv n\,(\mathrm{mod}\, q_1)$, $(n, q) = (m, q) = 1$ and $\chi(n) \neq \chi(m)$.

(b) Show that χ is primitive if and only if for every natural number $q_1 < q$ there is an integer $n \equiv 1\,(\mathrm{mod}\, q_1)$ such that $(n, q) = 1$ and $\chi(n) \neq 1$.

Theorem A.2.2. *Let χ be a primitive Dirichlet character mod q. Then, for every $n \in \mathbb{N}$, we have that*

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) e(an/q).$$

Proof. When $(n, q) = 1$, this is a consequence of the general Fourier inversion formula for characters, that is to say, relation (A.1.6). Assume now that $(n, q) = d > 1$, in which case we need to show that

$$\sum_{a=1}^{q} \overline{\chi}(a) e(an/q) = 0.$$

Write $n = dn_1$ and $q = dq_1$, and note that

$$\sum_{a=1}^{q} \overline{\chi}(a) e(an/q) = \sum_{j=1}^{d} \sum_{b=1}^{q_1} \overline{\chi}(b + jq_1) e\left(\frac{(b + jq_1)n_1}{q_1}\right) = \sum_{b=1}^{q_1} e(bn_1/q_1) \sum_{j=1}^{d} \overline{\chi}(b + jq_1).$$

So it suffices to show that

$$\sum_{j=1}^{d} \chi(b + jq_1) = 0, \tag{A.2.1}$$

for all $b \in \{1, 2, \ldots, q_1\}$. Since χ is primitive, Exercise A.2.1 implies that there is some $k \in \mathbb{Z}$ for which$(1 + kq_1, q) = 1$ and $\chi(1 + kq_1) \neq 1$. Consequently,

$$\chi(1 + kq_1) \sum_{j=1}^{d} \chi(b + jq_1) = \sum_{j=1}^{d} \chi(b + [bk + (1 + kq_1)j] \cdot q_1).$$

Since $(1+kq_1, q) = 1$, the numbers $\{bk + j(1+kq_1) : 1 \le j \le d\}$ run over a complete set of representatives mod d, and therefore

$$\chi(1+kq_1)\sum_{j=1}^{d}\chi(b+jq_1) = \sum_{j=1}^{d}\chi(b+jq_1).$$

Since $\chi(1+kq_1) \neq 1$, relation (A.2.1) follows. This completes the proof of the theorem. □

Using the above theorem, we can determine the size of the Gauss sum for primitive Dirichlet characters.

Theorem A.2.3. *Let χ be a primitive Dirichlet character mod q. Then $|\tau(\chi)| = \sqrt{q}$.*

Proof. By Theorem A.2.2, we have that

$$|\tau(\chi)|^2|\chi(n)|^2 = \left|\sum_{a=1}^{q}\chi(a)e(an/q)\right|^2,$$

for all $n \in \{1, \dots, q\}$. So

$$\begin{aligned}\varphi(q)|\tau(\chi)|^2 &= \sum_{n=1}^{q}\left|\sum_{a=1}^{q}\chi(a)e(an/q)\right|^2 = \sum_{n=1}^{q}\left(\sum_{a=1}^{q}\chi(a)e(an/q)\right)\left(\sum_{b=1}^{q}\overline{\chi}(b)e(-bn/q)\right)\\ &= \sum_{a=1}^{q}\sum_{b=1}^{q}\chi(a)\overline{\chi}(b)\sum_{n=1}^{q}e((a-b)n/q) = \sum_{a=1}^{q}|\chi(a)|^2 q = \varphi(q)q.\end{aligned}$$

This completes the proof of the theorem. □

Exercise A.2.4. Calculate the absolute value of the Gauss sum $\tau(\chi)$ for all Dirichlet characters $(\mathrm{mod}\, q)$.

A.3 The Pólya-Vinogradov inequality

Theorem A.3.1 (Pólya-Vinogradov inequality)**.** *Let χ be a non-principal character mod q. Then*

$$\sum_{M<n\le M+N}\chi(n) \ll \sqrt{q}\log q.$$

Proof. First, we prove the theorem when χ is primitive. Theorem A.2.2 and the periodicity of χ imply that

$$\begin{aligned}\sum_{M<n\le M+N}\chi(n) &= \sum_{M<n\le M+N}\frac{1}{\tau(\overline{\chi})}\sum_{\substack{1\le a\le q\\(a,q)=1}}\overline{\chi}(a)e(an/q) = \sum_{M<n\le M+N}\frac{1}{\tau(\overline{\chi})}\sum_{\substack{-q/2<a\le q/2\\(a,q)=1}}\overline{\chi}(a)e(an/q)\\ &= \frac{1}{\tau(\overline{\chi})}\sum_{\substack{-q/2<a\le q/2\\(a,q)=1}}\overline{\chi}(a)\sum_{M<n\le M+N}e(an/q) \ll \frac{1}{|\tau(\overline{\chi})|}\sum_{\substack{-q/2<a\le q/2\\(a,q)=1}}\frac{1}{|1-e(a/q)|}.\end{aligned}$$

So Theorem A.2.3, and the fact that $|1 - e(x)| \asymp |x|$, for all $x \in [-1/2, 1/2]$, imply that

$$\sum_{M<n\le M+N} \chi(n) \ll \frac{1}{\sqrt{q}} \sum_{\substack{-q/2<a\le q/2 \\ (a,q)=1}} \frac{1}{|a/q|} \le 2\sqrt{q} \sum_{1\le a\le q/2} \frac{1}{a} \ll \sqrt{q}\log q.$$

This completes the proof in the case that χ is primitive. Finally, if χ is induced by the primitive character $\chi_1 \pmod{q_1}$, then

$$\begin{aligned}\sum_{M<n\le M+N} \chi(n) &= \sum_{\substack{M<n\le M+N \\ (n,q)=1}} \chi(n) = \sum_{\substack{M<n\le M+N \\ (n,q)=1}} \chi_1(n) = \sum_{\substack{M<n\le M+N \\ (n,q/q_1)=1}} \chi_1(n) \\ &= \sum_{M<n\le M+N} \chi_1(n) \sum_{d\mid(n,q/q_1)} \mu(d) = \sum_{d\mid q/q_1} \mu(d)\chi_1(d) \sum_{M/d<m\le(M+N)/d} \chi_1(m) \\ &\ll \sum_{d\mid q/q_1} \sqrt{q_1}\log q_1 \le \tau(q/q_1)\sqrt{q_1}\log q \ll \sqrt{\frac{q}{q_1}}\cdot\sqrt{q_1}\log q = \sqrt{q}\log q,\end{aligned}$$

which completes the proof in this case as well. □

Appendix B

Primes in arithmetic progressions

The purpose of this chapter is to establish the Siegel-Walfisz theorem:

Theorem B.0.2. *Let $A > 0$. There is a (non-effective) constant c_A such that if $1 \le q \le (\log x)^A$ and $(a,q) = 1$, then*

$$\pi(x; q, a) = \frac{\mathrm{li}(x)}{\varphi(q)} + O\left(\frac{x}{e^{c_A\sqrt{\log x}}}\right).$$

The proof we will present is not the classical complex-analytic proof. Rather, we use the approach of [Kou], which is based on the theory of *pretentious multiplicative functions*.

As it is common, instead of working directly with the indicator function of the primes, we use von Mangoldt's function $\Lambda(n)$, which reduces Theorem B.0.2 to proving that

$$\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\, q)}} \Lambda(n) = \frac{x}{\varphi(q)} + O\left(\frac{x}{e^{c_A\sqrt{\log x}}}\right) \quad (q \le (\log x)^A,\ (a,q) = 1).$$

For every $x \ge 1$ and $(a,q) = 1$, the orthogonality of characters implies that

$$\sum_{\substack{n\le x\\ n\equiv a \,(\mathrm{mod}\ q)}} \Lambda(n) - \frac{x}{\varphi(q)} = \frac{1}{\varphi(q)} \sum_{\chi \ (\mathrm{mod}\ q)} \overline{\chi}(a) \sum_{n\le x} (\chi(n)\Lambda(n) - \delta(\chi)) + O\left(\frac{1}{\varphi(q)}\right),$$

where

$$\delta(\chi) := \begin{cases} 1 & \text{if } \chi \text{ is principal,} \\ 0 & \text{otherwise.} \end{cases}$$

So, it suffices to prove that

$$\sum_{n\le x} \Lambda_\chi(n) \ll \frac{x}{e^{\sqrt{\log x}}} \quad (x \ge \exp\{q^\epsilon\}) \tag{B.0.1}$$

for all characters $\chi \,(\mathrm{mod}\, q)$, where we have set

$$\Lambda_\chi(n) = \chi(n)\Lambda(n) - \delta(\chi).$$

We write $F_\chi(s)$ for the Dirichlet series corresponding to Λ_χ and we note that

$$F_\chi(s) = \sum_{n=1}^{\infty} \frac{\Lambda_\chi(n)}{n^s} = -\frac{L'}{L}(s,\chi) - \delta(\chi)\zeta(s),$$

where, given an arithmetic function $f : \mathbb{N} \to \mathbb{C}$, we employ the standard notation

$$L(s,f) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}.$$

Our strategy towards proving (B.0.1) is to multiply $\Lambda_\chi(n)$ by an appropriately high power of $\log n$ and control that sum instead. This manoeuvre will allow us to work with $F_\chi(s)$ when $\mathrm{Re}(s) > 1$, where no arguments about the analyticity and the location of the zeroes of $L(s,\chi)$ are needed. For technical reasons, we also introduce the smoothing $\log(x/n)$ to the sum we consider. Fourier inversion implies that

$$\begin{aligned} \sum_{n\le x} \Lambda_\chi(n)(\log n)^{k-1} \log\frac{x}{n} &= \frac{(-1)^{k-1}}{2\pi i} \int_{\mathrm{Re}(s)=1+1/\log x} F_\chi^{(k-1)}(s)\frac{x^s}{s^2}ds \\ &\ll x \cdot \int_{-\infty}^{\infty} \left| F_\chi^{(k-1)}\left(1+\frac{1}{\log x}+it\right)\right| \frac{dt}{1+t^2} \end{aligned} \tag{B.0.2}$$

for all $k \in \mathbb{N}$. In view of the above formula, we need to understand the size of the derivatives of F_χ. The following key lemma, which is based on an idea in [IK, p. 40], allows us to reduce this problem to upper bounds for the derivatives of $F_\chi(s)$ and a lower bound on $|F_\chi(s)|$.

Lemma B.0.3. *Let $M \ge 1$, D be an open subset of $\mathbb{C}$ and $s \in D$. Consider a function $F : D \to \mathbb{C}$ that is differentiable k times at s and its derivatives satisfy the bound $|F^{(j)}(s)| \le j!M^j$ for $1 \le j \le k$. If $F(s) \ne 0$, then*

$$\left| \left(\frac{F'}{F}\right)^{(k-1)}(s) \right| \le \frac{k!}{2}\left(\frac{2M}{\min\{|F(s)|,1\}}\right)^k.$$

Proof. We have the identity

$$\left(\frac{-F'}{F}\right)^{(k-1)}(s) = k! \sum_{a_1+2a_2+\cdots=k} \frac{(-1+a_1+a_2+\cdots)!}{a_1!a_2!\cdots}\left(\frac{-F'}{1!F}(s)\right)^{a_1}\left(\frac{-F''}{2!F}(s)\right)^{a_2}\cdots, \tag{B.0.3}$$

which can be easily verified by induction. In order to complete the proof of the lemma, we will show that

$$\sum_{a_1+2a_2+\cdots+ka_k=k} \frac{(a_1+a_2+\cdots+a_k)!}{a_1!a_2!\cdots a_k!} = \sum_{a_1+2a_2+\cdots+ka_k=k} \binom{a_1+a_2+\cdots+a_k}{a_1,a_2,\ldots,a_k} = 2^{k-1}. \tag{B.0.4}$$

Indeed, for each fixed k-tuple $(a_1,\ldots,a_k) \in (\mathbb{N}\cup\{0\})^k$ with $a_1+2a_2+\cdots+ka_k = k$, the multinomial coefficient $\binom{a_1+a_2+\cdots+a_k}{a_1,a_2,\ldots,a_k}$ represents the way of writing k as the sum of a_1

ones, a_2 twos, and so on, with the order of the different summands being important, e.g. if $k = 5$, $a_1 = 1$, $a_2 = 2$ and $a_3 = a_4 = a_5 = 0$, then there are three such ways to write 5: $5 = 2 + 2 + 1 = 2 + 1 + 2 = 1 + 2 + 2$. So we conclude that

$$\sum_{a_1+2a_2+\cdots+ka_k=k} \binom{a_1 + a_2 + \cdots + a_k}{a_1, a_2, \ldots, a_k} = \#\{\text{ordered partitions of } k\},$$

where we define an ordered partition of k to be a way to write k as the sum of positive integers, with the order of the different summands being important. To every ordered partition of $k = b_1 + \cdots + b_m$, we can associate a unique subset of $\{1, \ldots, k\}$ in the following way: consider the set $B \subset \{1, \ldots, k\}$ which contains $\{1, \ldots, b_1\}$, does not contain $\{b_1 + 1, \ldots, b_1 + b_2\}$, contains $\{b_1+b_2+1, \ldots, b_1+b_2+b_3\}$, does not contain $\{b_1+b_2+b_3+1, \ldots, b_1+b_2+b_3+b_4\}$, and so on. Then B necessarily contains 1 and, conversely, every subset of $\{1, \ldots, k\}$ containing 1 can arise this way. So we conclude that there are 2^{k-1} ordered partitions, and (B.0.4) follows, thus completing the proof of the lemma. □

Our next goal is to obtain upper bounds on the derivatives of $L(s, \chi)$ and a lower bound on $|L(s, \chi)|$. In fact, we shall perform a technical manoeuvre and switch to a sifted version of $L(s, \chi)$. We may do so since

$$\left(\frac{L'}{L}\right)^{(k-1)}(s, \chi) = \left(\frac{L'_y}{L_y}\right)^{(k-1)}(s, \chi) + O(c^k k!(\log y)^k),$$

where

$$L_y(s, \chi) := \sum_{P^-(n)>y} \frac{\chi(n)}{n^s}.$$

The reason for doing so is that, in general, we can only control $\sum_{n\le N} \chi(n)n^{it}$ for N large enough in terms of q and t, and it is possible that this sum is rather large for small values of N. This would force $L^{(j)}(s, \chi)$ to be large. But then, the same thing would be true for $L(s, \chi)$. However, it is rather hard to capture this correlation in the sizes of $L^{(j)}(s, \chi)$ and $L(s, \chi)$ in practice. By considering the sifted L-function $L_y(s, \chi)$, we ensure that the partial sums we consider $\sum_{n\le N} \chi(n)n^{it}$ all have $N > y$, so the small values of N can no longer cause any problems.

Lemma B.0.4. *Let χ be a Dirichlet character modulo q, $k \in \mathbb{N} \cup \{0\}$, and $s = \sigma + it$ with $\sigma > 1$ and $t \in \mathbb{R}$. For $y \ge 3/2$ we have that*

$$\left| L_y^{(k)}(s, \chi) + \frac{(-1)^{k+1}k!}{(s-1)^{k+1}} \frac{\delta(\chi)\varphi(q)}{q} \prod_{\substack{p\le y \\ p\nmid q}} \left(1 - \frac{1}{p}\right) \right| \ll \frac{k!(c\log(|t| + q + y))^{k+1}}{\log y}. \tag{B.0.5}$$

In particular, if $y \ge \max\left\{q + |t|, e^{\delta(\chi)/|t|}\right\}^{\epsilon}$ for some fixed $\epsilon > 0$, then

$$\left|L_y^{(k)}(s, \chi)\right| \ll k!(c_\epsilon \log y)^k. \tag{B.0.6}$$

Proof. Set $z = \max\{y, q^4, (|t|+1)^{100}\}$ and note that

$$(-1)^k L_y^{(k)}(s,\chi) = \sum_{\substack{n>z \\ P^-(n)>y}} \frac{\chi(n)(\log n)^k}{n^s} + O\left(\frac{(\log z)^{k+1}}{\log y}\right). \tag{B.0.7}$$

Next, we need to control the sum $\sum_{n\le x,\, P^-(n)>y} \chi(n)n^{-it}$. We use the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.3) to construct upper and lower bound weights $(\mu^{\pm}(d))_{d\ge 1}$ supported on the set $\{d \le \sqrt{x} : d|P(y)\}$. Then

$$\begin{aligned}\sum_{\substack{n\le x \\ P^-(n)>y}} \chi(n)n^{-it} &= \sum_{n\le x}(1*\mu^+)(n)\chi(n)n^{-it} + O\left(\sum_{n\le x}(1*\mu^+ - 1*\mu^-)(n)\right)\\ &= \sum_{d\le\sqrt{x}} \mu^+(d)\chi(d)d^{-it}\sum_{m\le x/d}\chi(m)m^{-it} + O\left(\frac{x^{1-1/\log y}}{\log y}\right).\end{aligned}$$

We break the inner sum into congruence classes mod q. For each $b \in (\mathbb{Z}/q\mathbb{Z})^*$, partial summation implies that

$$\sum_{\substack{m\le x/d \\ m\equiv b \,(\mathrm{mod}\, q)}} m^{-it} = \int_1^{x/d} u^{-it} d\left(\frac{u}{q} + O(1)\right) = \frac{(x/d)^{1-it}}{q(1-it)} + O(1+|t|\log x).$$

So, we conclude that

$$\begin{aligned}\sum_{\substack{n\le x \\ P^-(n)>y}} \chi(n)n^{-it} &= \frac{x^{1-it}}{1-it}\sum_{d\le\sqrt{x}}\frac{\mu^+(d)\chi(d)}{d}\cdot\frac{1}{q}\sum_{b\in(\mathbb{Z}/q\mathbb{Z})^*}\chi(b) + O\left(q\sqrt{x}(1+|t|\log x) + \frac{x^{1-1/\log y}}{\log y}\right)\\ &= \delta(\chi)\cdot\frac{\varphi(q)}{q}\cdot\frac{x^{1-it}}{1-it}\prod_{\substack{p\le y \\ p\nmid q}}\left(1-\frac{1}{p}\right) + O\left(\frac{x^{1-1/\log y}}{\log y}\right)\end{aligned}$$

for all $x \ge y$. Let $R_t(x)$ be the error term in the above approximation. Then

$$\begin{aligned}\sum_{\substack{n>z \\ P^-(n)>y}} \frac{\chi(n)(\log n)^k}{n^s} &= \int_z^\infty \frac{(\log u)^k}{u^\sigma} d\left(\frac{\delta(\chi)\varphi(q)}{q}\frac{u^{1-it}}{1-it}\prod_{\substack{p\le y \\ p\nmid q}}\left(1-\frac{1}{p}\right) + R_t(u)\right)\\ &= \frac{\delta(\chi)\varphi(q)}{q}\prod_{\substack{p\le y \\ p\nmid q}}\left(1-\frac{1}{p}\right)\int_z^\infty\frac{(\log u)^k}{u^s}du + \int_z^\infty\frac{(\log u)^k}{u^\sigma}dR_t(u).\end{aligned}$$

Since

$$\begin{aligned}\int_z^\infty \frac{(\log u)^k}{u^\sigma}dR_t(u) &= -\frac{(\log z)^k R_t(z)}{z^\sigma} + \int_z^\infty \frac{(\log u)^{k-1}(\sigma\log u - k)}{u^{\sigma+1}}R_t(u)du\\ &\ll \frac{(\log z)^k}{\log y} + \frac{\sigma+k}{\log y}\int_z^\infty\frac{(\log u)^k}{u^{\sigma+1/\log y}}du\\ &\le \frac{(\log z)^k}{\log y} + \frac{\sigma+k}{z^{\sigma-1}\log y}\int_z^\infty\frac{(\log u)^k}{u^{1+1/\log y}}du \ll \frac{(k+1)!(\log z)^{k+1}}{\log y},\end{aligned}$$

by partial summation, and

$$\int_z^\infty \frac{(\log u)^k}{u^s}du = \int_1^\infty \frac{(\log u)^k}{u^s}du + O((\log z)^{k+1}) = \frac{k!}{(s-1)^{k+1}} + O((\log z)^{k+1}),$$

by observing that $(\log u)^m u^{-s} = (\log u)^m \frac{d(u^{1-s}/(1-s))}{du}$ for all $m \geq 0$ and integrating by parts k times, relation (B.0.5) follows. Finally, relation (B.0.6) is a direct consequence of relation (B.0.5), since $|s-1| \geq |t| \geq \epsilon \cdot \delta(\chi)/\log y$ under the assumption that $y \geq e^{\epsilon\cdot\delta(\chi)/|t|}$. This completes the proof of the lemma. □

Next, we need a lower bound on $L_y(s,\chi)$ close to the line $\mathrm{Re}(s) = 1$. We need to introduce some notation a state a preliminary result. Given two multiplicative functions $f, g : \mathbb{N} \to \{z \in \mathbb{C} : |z| \leq 1\}$ and real numbers $x \geq y \geq 1$, we set

$$\mathbb{D}(f,g;y,x) = \left(\sum_{y<p\leq x} \frac{1-\mathrm{Re}(f(p)\overline{g(p)})}{p}\right)^{1/2}.$$

This quantity defines a certain measure of 'distance' between f and g, that is to say $\mathbb{D}(f,g;y,x)$ is small if f *pretends to be* g for primes in $(y,x]$. It was introduced by Granville and Soundararajan, partially in order to conceptualize and put under a unified framework several results in the literature. It is central in the theory of *pretentious multiplicative functions* and it satisfies the triangle inequality (see [GS, p. 207] for a slightly weaker form of this inequality).

Lemma B.0.5. *Let $f, g, h : \mathbb{N} \to \{z \in \mathbb{C} : |z| \leq 1\}$ be multiplicative functions and $x \geq y \geq 1$. Then*

$$\mathbb{D}(f,g;y,x) + \mathbb{D}(g,h;y,x) \geq \mathbb{D}(f,h;y,x).$$

Proof. It suffices to show that

$$\sqrt{\mathrm{Re}\left(1-z\overline{z'}\right)} + \sqrt{\mathrm{Re}\left(1-z'\overline{z''}\right)} \geq \sqrt{\mathrm{Re}\left(1-z\overline{z''}\right)}.$$

for all complex numbers z, z', z'' of modulus ≤ 1. We set $w = z\overline{z'}$, $w' = z'\overline{z''}$ and $r = |z'|$, so that the inequality we need to show becomes

$$\sqrt{\mathrm{Re}(1-w)} + \sqrt{\mathrm{Re}(1-w')} \geq \sqrt{\mathrm{Re}(1-ww'/r^2)}, \tag{B.0.8}$$

for all $r \in (0,1]$ and all complex numbers w, w' such that $|w|, |w'| \leq r$. We write $w = x+iy$ and $w' = a+ib$ and we consider r, x and a fixed for the moment. So our goal is to show that

$$\sqrt{1-(ax-by)/r^2} \leq \sqrt{1-x} + \sqrt{1-a},$$

for all b, y with $|b| \leq \sqrt{r^2-a^2}$ and $|y| \leq \sqrt{r^2-b^2}$. It is easy to see that $\sqrt{1-(ax-by)}$ is maximized when $b^2 = r^2-a^2$ and $y^2 = r^2-x^2$, so that (B.0.8) follows by the case $|w| = |w'| = r$. We set $w = re^{i\theta}$ and $w' = re^{i\varphi}$, and we define $s \in [0,1]$ via the relation

$$\frac{2s}{s^2+1} = r.$$

Then (B.0.8) reduces to showing that

$$\sqrt{\mathrm{Re}(1-re^{i\theta})}+\sqrt{\mathrm{Re}(1-re^{i\varphi})}\geq\sqrt{\mathrm{Re}(1-e^{i(\theta+\varphi)})}$$

or, equivalently, that

$$\frac{|s-e^{i\theta}|+|s-e^{i\varphi}|}{\sqrt{s^2+1}}\geq\frac{|e^{i\theta}-e^{i\varphi}|}{\sqrt{2}},$$

which is a consequence of the triangle inequality in $\mathbb{C}$ and the fact that $s\leq 1$. □

We can now give our lower bound for $|L_y(s,\chi)|$.

Lemma B.0.6. *Fix $\epsilon\in(0,1]$. Let χ be a Dirichlet character modulo q, $s=\sigma+it$ with $\sigma>1$ and $t\in\mathbb{R}$, and $y\geq q+|t|$. If $|t|\geq\epsilon/\log y$ or if χ is complex, then we have that $|L_y(s,\chi)|\asymp_\epsilon 1$. Finally, if χ is a non-principal, real character, and $|t|\leq 1/\log y$, then $|L_y(s,\chi)|\gg L_y(1,\chi)$.*

Proof. First, assume that either $|t|\geq\epsilon/\log y$ or χ is complex. Equivalently, $|t|\geq\epsilon\delta(\chi^2)/\log y$. Note that

$$\begin{aligned}\log\left|L_y\left(1+it+\frac{1}{\log x},\chi\right)\right|&=\sum_{p>y}\frac{\mathrm{Re}(\chi(p)p^{-it})}{p^{1+1/\log x}}+O(1)\\&=\sum_{y<p\leq x}\frac{\mathrm{Re}(\chi(p)p^{-it})}{p}+O(1)\\&=\mathbb{D}(\chi(n),\mu(n)n^{it};y,x)-\log\left(\frac{\log x}{\log y}\right)+O(1),\end{aligned}$$

upon observing that $p^{-1/\log x}=1+O(\log p/\log x)$ for $p\leq x$ and that $\sum_{p>x}1/p^{1+1/\log x}\ll 1$. Therefore we see that if $|L_y(1+1/\log x,\chi)$ is small, then $\chi(n)$ must pretend to be $\mu(n)n^{it}$. But then the triangle inequality implies that $\chi^2(n)$ must present to be n^{2it}, and this is impossible by our upper bounds on $|L_y(1+1/\log x+2it,\chi^2)|$ provided by Lemma B.0.4. Indeed, for every $x\geq y$ we have that

$$\begin{aligned}2\cdot\mathbb{D}(\chi(n),\mu(n)n^{it};y,x)&=\mathbb{D}(\chi(n)n^{-it},\mu(n);y,x)+\mathbb{D}(\mu(n),\overline{\chi}(n)n^{it};y,x)\\&\geq\mathbb{D}(\chi(n)n^{-it},\overline{\chi}(n)n^{it};y,x)=D(\chi^2(n),n^{2it};y,x)\end{aligned}$$

by Lemma B.0.5. Consequently,

$$\begin{aligned}\mathbb{D}^2(\chi(n),\mu(n)n^{it};y,x)&\geq\frac{1}{4}\log\left(\frac{\log x}{\log y}\right)+O(1)-\sum_{y<p\leq x}\frac{\mathrm{Re}(\chi^2(p)p^{-2it})}{p}\\&=\frac{1}{4}\log\left(\frac{\log x}{\log y}\right)+O(1)-\log\left|L_y\left(1+\frac{1}{\log x}+2it,\chi^2\right)\right|\\&\geq\frac{1}{4}\log\left(\frac{\log x}{\log y}\right)+O_\epsilon(1),\end{aligned}$$

by relation (B.0.6) with $k=0$, provided that $y \geq \max\left\{q+|t|, e^{\epsilon\cdot\delta(\chi^2)/|t|}\right\}$. Hence,

$$\left|L_y\left(1+\frac{1}{\log x}+it,\chi\right)\right| \gg \left(\frac{\log y}{\log x}\right)^{3/4} \quad \left(x\geq y\geq \max\left\{q+|t|, e^{\epsilon\cdot\delta(\chi^2)/|t|}\right\}\right), \tag{B.0.9}$$

We claim that this inequality is self-improving. Indeed, if x and y are as above, then

$$\begin{aligned}\sum_{y<p\leq x}\frac{\chi(p)}{p^{1+it}} = \sum_{p>y}\frac{\chi(p)}{p^{1+it+1/\log x}} + O(1) &= \sum_{p>y}\frac{\chi(p)}{p^{1+it+1/\log x}} - \sum_{p>y}\frac{\chi(p)}{p^{1+it+1/\log y}} + O(1)\\ &= \int_y^x \sum_{p>y}\frac{\chi(p)\log p}{p^{1+it+1/\log u}}\cdot\frac{du}{u(\log u)^2} + O(1)\\ &= -\int_y^x \frac{L_y'}{L_y}\left(1+it+\frac{1}{\log u},\chi\right)\frac{du}{u(\log u)^2} + O(1)\\ &\ll_\epsilon \int_y^x (\log y)^{3/4}(\log u)^{1/4}\frac{du}{u(\log u)^2} + 1 \ll 1,\end{aligned}$$

by (B.0.6) and (B.0.9). So we conclude that

$$\left|\sum_{y<p\leq x}\frac{\chi(p)}{p^{1+it}}\right| \ll 1 \quad \left(x\geq y\geq \max\left\{q+|t|, e^{\epsilon\cdot\delta(\chi^2)/|t|}\right\}\right). \tag{B.0.10}$$

The above relation for $x=\max\{e^{1/(\sigma-1)}, y\}$ implies that $|L_y(s,\chi)| \asymp 1$, which completes the proof of the first part of the lemma.

Finally, assume that χ is a real, non-principal character, and that $|t|\leq 1/\log y$. Let $x=\max\{e^{1/(\sigma-1)}, y\}$ and $z=\min\{x, e^{1/|t|}\} \geq y \geq q+|t|$. Then we have that

$$\left|\sum_{z<p\leq x}\frac{\chi(p)}{p^{1+it}}\right| \ll 1,$$

trivially if $z=x$, and by relation (B.0.10) with z in place of y otherwise, which holds, since in this case $z=e^{1/|t|}\geq e^{\delta(\chi^2)/|t|}$. So we deduce that

$$\sum_{y<p\leq x}\frac{\chi(p)}{p^{1+it}} = \sum_{y<p\leq z}\frac{\chi(p)}{p^{1+it}} + O(1) = \sum_{y<p\leq z}\frac{\chi(p)+O(|t|\log p)}{p} + O(1) = \sum_{y<p\leq z}\frac{\chi(p)}{p} + O(1).$$

Finally, for every $w\geq z\geq q+|t|$, we have that

$$\sum_{z<p\leq w}\frac{\chi(p)}{p} = \log\left|L_z\left(1+\frac{1}{\log w},\chi\right)\right| + O(1) \leq O(1),$$

by relation (B.0.6) with $k=0$. So

$$\sum_{y<p\leq x}\frac{\mathrm{Re}(\chi(p)p^{-it})}{p} = \sum_{y<p\leq z}\frac{\chi(p)}{p} + O(1) \geq \sum_{y<p\leq w}\frac{\chi(p)}{p} + O(1) \quad (w\geq z).$$

Therefore, we conclude that $|L_y(s,\chi)| \gg L_y(1+1/\log w,\chi)$ for all $w\geq z$. Letting $w\to\infty$ completes the proof of the last part of the lemma too. □

Finally, we prove an estimate for the derivatives of $(L'/L)(s,\chi)$, which will be key in the proof of Theorem B.0.2.

Lemma B.0.7. *Let χ be a Dirichlet character modulo q and $s=\sigma+it$ with $\sigma>1$ and $t\in\mathbb{R}$. For every $k\in\mathbb{N}$ we have that*

$$\left|\left(\frac{L'}{L}\right)^{(k-1)}(s,\chi)+\frac{\delta(\chi)(-1)^{k-1}(k-1)!}{(s-1)^k}\right|\ll\left(\frac{ck\log(2q+|t|)}{\delta(\chi)+(1-\delta(\chi))|L_{q+|t|}(s,\chi)|}\right)^k.$$

Proof. Set $y=q+|t|$ and fix some constant ϵ to be chosen later. We separate three cases.

Case 1: $\sigma\geq 1+\epsilon/\log y$. Note that $\delta(\chi)+(1-\delta(\chi))|L_{q+|t|}(s,\chi)|\ll 1$, trivially if $\delta(\chi)=1$, and by relation (B.0.6) with $k=0$ if $\delta(\chi)=0$. Since we also have that

$$\left|\left(\frac{L'}{L}\right)^{(k-1)}(s,\chi)+\frac{\delta(\chi)(-1)^{k-1}(k-1)!}{(s-1)^k}\right|\leq\sum_{n=1}^{\infty}\frac{\Lambda(n)(\log n)^{k-1}}{n^{1+\epsilon/\log y}}+\frac{(k-1)!}{(\epsilon/\log y)^k}\ll(c_1k\log y)^k$$

for some $c_1=c_1(\epsilon)$, the lemma follows.

Case 2: $|t|>\epsilon/\log y$. Note that

$$\left(\frac{L'}{L}\right)^{(k-1)}(s,\chi)=O\left((c_2k\log y)^k\right)+\left(\frac{L_y'}{L_y}\right)^{(k-1)}(s,\chi). \tag{B.0.11}$$

Furthermore, relation (B.0.6) implies that $|L_y^{(j)}(s,\chi)|\leq j!(c_3\log y)^j$ for all $j\in\mathbb{N}$, for some $c_3=c_3(\epsilon)$. Additionally, we have that $|L_y(s,\chi)|\asymp_\epsilon 1$ by Lemma B.0.6. So Lemma B.0.3 applied to $F(s)=L_y(s,\chi)$ yields that the right hand side of (B.0.11) is $\ll(c_4k\log y)^k$ for some $c_4=c_4(\epsilon)$, and the lemma follows (note that in this case $|s-1|\geq|t|\geq\epsilon/\log y$).

Case 3: $|s-1|\leq 2\epsilon/\log y$. Let

$$F(s)=(s-1)^{\delta(\chi)}L_y(s,\chi)\prod_{p\leq y}\left(1-\frac{1}{p}\right)^{-\delta(\chi)},$$

and observe that

$$F^{(j)}(s)=\left((s-1)^{\delta(\chi)}L_y^{(j)}(s,\chi)+\delta(\chi)jL_y^{(j-1)}(s,\chi)\right)\prod_{p\leq y}\left(1-\frac{1}{p}\right)^{-\delta(\chi)}\ll j!(c_5\log y)^j,$$

for all $j\in\mathbb{N}$, by relation (B.0.5). So Lemma B.0.3 implies that

$$\left(\frac{L_y'}{L_y}\right)^{(k-1)}(s,\chi)+\frac{\delta(\chi)(-1)^{k-1}(k-1)!}{(s-1)^k}=\left(\frac{F'}{F}\right)^{(k-1)}(s)\ll\left(\frac{c_6k\log y}{\min\{|F(s)|,1\}}\right)^k.$$

Together with (B.0.11), the above estimate reduces the desired result to showing that

$$\min\{|F(s)|,1\}\asymp\delta(\chi)+(1-\delta(\chi))|L_y(s,\chi)|. \tag{B.0.12}$$

If $\delta(\chi) = 0$, then (B.0.12) holds, since $|F(s)| = |L_y(s,\chi)| \ll 1$, by relation (B.0.6) with $k = 0$. Lastly, if $\delta(\chi) = 1$, then $|F(s)| = 1 + O(|s-1|\log y) = 1 + O(\epsilon)$ by relation (B.0.5) with $k = 0$ (note that $\prod_{p\le y,\, p|q}(1-1/p) = \varphi(q)/q$, since $y > q$ by assumption). So choosing ϵ small enough (independently of k, q and s), we find that $|F(s)| \asymp 1$, that is, (B.0.12) is satisfied in this case too. This completes the proof of (B.0.12) and hence of the lemma. □

The last ingredient missing in order to complete the proof of Theorem B.0.2 is Siegel's theorem:

Theorem B.0.8. *Let $\epsilon \in (0,1]$. With at most one exception, for all real, non-principal, primitive Dirichlet characters we have that $L(1,\chi) \gg \epsilon^3 q^{-\epsilon}$, where q denotes the conductor of χ; the implied constant is effectively computable.*

Before proving Theorem B.0.8, we need a preliminary lemma. The idea behind its proof can be traced back to [Pi].

Lemma B.0.9. *Let $c \ge 1$, $r \in \mathbb{N}$, and $Q \ge 2$. Consider a multiplicative function $f : \mathbb{N} \to \mathbb{R}$ such that $0 \le (1*f)(n) \le \tau_r(n)$ for all integers n, and*

$$\left|\sum_{n\le x} f(n)\right| \le c x^{4/5}\log x \quad (x \ge Q).$$

If $L(1-\eta, f) \ge 0$ for some $\eta \in [0, 1/20]$, then $L(1,f) \gg_{c,r} \eta/Q^{2\eta}$. If, in addition, $L(1-\eta,f) = 0$, then $L(1,f) \ll_{c,r} \eta(\log Q)^r$.

Proof. By partial summation, we have that

$$\sum_{A_1 < a \le A_2} \frac{f(a)}{a^{1-\eta}} \ll_c \frac{\log A_1}{A_1^{1/5-\eta}} \quad (A_2 \ge A_1 \ge Q),$$

and

$$\sum_{b\le B} \frac{1}{b^{1-\eta}} = \frac{B^\eta - 1}{\eta} + \gamma_\eta + O(B^{\eta-1}) \quad (B \ge 1), \quad \text{where } \gamma_\eta = 1 - (1-\eta)\int_1^\infty \frac{\{u\}}{u^{2-\eta}}du.$$

So for $x \ge Q^2$ we have that

$$\begin{aligned} S := \sum_{n\le x}\frac{(1*f)(n)}{n^{1-\eta}} &= \sum_{a\le\sqrt{x}}\frac{f(a)}{a^{1-\eta}}\sum_{b\le x/a}\frac{1}{b^{1-\eta}} + \sum_{b\le\sqrt{x}}\frac{1}{b^{1-\eta}}\sum_{\sqrt{x}<a\le x/b}\frac{f(a)}{a^{1-\eta}} \\ &= \sum_{a\le\sqrt{x}}\frac{f(a)}{a^{1-\eta}}\left(\frac{(x/a)^\eta - 1}{\eta} + \gamma_\eta + O_c\left((x/a)^{\eta-1}\right)\right) + O\left(\frac{\log x}{x^{1/10-\eta/2}}\sum_{b\le\sqrt{x}}\frac{1}{b^{1-\eta}}\right) \\ &= \sum_{a\le\sqrt{x}}\frac{f(a)}{a^{1-\eta}}\frac{(x/a)^\eta - 1}{\eta} + \gamma_\eta L(1-\eta,f) + O_{c,r}\left(\frac{\log^2 x}{x^{1/10-\eta}}\right), \end{aligned}$$

since $|f| = |\mu * (1 * f)| \le \tau_{r+1}$. Finally, for $A > \sqrt{x}$ we have that

$$\sum_{\sqrt{x}<a\le A} \frac{f(a)}{a^{1-\eta}} \frac{(x/a)^\eta - 1}{\eta} = -\int_{\sqrt{x}}^{A} \frac{(x/u)^\eta - 1}{\eta} d\left(\sum_{u<a\le A} \frac{f(a)}{a^{1-\eta}} \right) \ll_c \frac{\log^2 x}{x^{1/10-\eta}},$$

by integration by parts. Consequently,

$$S = \frac{x^\eta}{\eta} L(1, f) + \left(\gamma_\eta - \frac{1}{\eta} \right) L(1-\eta, f) + O_{c,r}\left(\frac{\log^2 x}{x^{1/10-\eta}} \right).$$

Note that $1 \le S \ll_r x^\eta \log^r x$, by our assumption that $0 \le 1 * f \le \tau_r$. Since $\gamma_\eta < 1 < 1/\eta$ for $\eta \in (0, 1)$, the claimed result then follows by taking $x = c_1 Q^2$ for some sufficiently large constant c_1 that depends at most on c and r. □

Proof of Theorem B.0.8. Let $\epsilon \in [0, 1]$. If $L(s, \chi)$ does not vanish in $[1 - \epsilon/30, 1]$, then $L(1-\epsilon/30, \chi) > 0$, by continuity, and thus Lemma B.0.9 with $Q = q^{5/4}$ implies that $L(1, \chi) \gg \epsilon q^{-\epsilon/12}$. So if there is no character χ for which $L(s, \chi)$ has a zero in $[1 - \epsilon/30, 1]$, then the theorem follows. Therefore we may assume that this is not the case. Let χ_1 be a primitive real character of minimum conductor q_1 such that $L(s, \chi_1)$ vanishes in $[1 - \epsilon/30, 1]$, say at $1 - \eta$. Consider $\chi \ne \chi_1$ of conductor q. As before, if $q < q_1$, then $L(1, \chi) \gg \epsilon q^{-\epsilon/12}$, since $L(s, \chi)$ does not vanish in $[1 - \epsilon/30, 1]$. Finally, we consider the case $q \ge q_1$. Set $f = \chi_1 * \chi * \chi_1\chi$. Then

$$\sum_{n\le x} f(n) \ll q^{4/3} x^{2/3} \log x \le x^{4/5} \log x \quad (x \ge q^{10}),$$

by applying Dirichlet's hyperbola method first to $\chi_1 * \chi$ and then to $f = (\chi_1 * \chi) * \chi_1\chi$. In a similar fashion, we find that $L(\sigma, f) = L(\sigma, \chi)L(\sigma, \chi_1)L(\sigma, \chi_1\chi)$ for $\sigma > 2/3$. In particular, $L(1-\eta, f) = 0$ and therefore $L(1, f) \gg \eta q^{-20\eta}$, by Lemma B.0.9. Since we also have that

$$L(1, f) = L(1, \chi)L(1, \chi_1)L(1, \chi_1\chi) \ll L(1, \chi) \cdot (\eta \log^2 q) \cdot (\log q),$$

by Lemma B.0.9 applied to χ_1 and the standard bound $L(1, \chi_1\chi) \ll \log q$, which follows by partial summation, we deduce the theorem. □

Also, we state the following classical result, which is a consequence of Dirichlet's class number formula [Da, p. 49-50]. See also [IK, p. 37] for an elementary proof, in the spirit of the proof of Lemma B.0.9.

Lemma B.0.10. *If χ is a non-principal real character mod q, then $L(1, \chi) > 0$.*

We are now ready to complete the proof of the Siegel-Walfisz theorem:

Proof of Theorem B.0.2. Recall the notations $\Lambda_\chi(n)$ and $F_\chi(s)$, and that our goal is to prove relation B.0.1. We claim that, for $k \in \mathbb{N}$, $\epsilon > 0$ and $s = \sigma + it$ with $\sigma > 1$ and $t \in \mathbb{R}$, we have that

$$F_\chi^{(k-1)}(s) \ll \begin{cases} (c_1 k \log(2q + |t|))^k & \text{if } |t| \ge 1/\log(3q), \\ (c_2 k q^{\epsilon/3})^k & \text{otherwise,} \end{cases} \tag{B.0.13}$$

where $c_2 = c_2(\epsilon)$ is an ineffective constant. Indeed, relation (B.0.5) with $y = 3/2$, $q = 1$, $\chi(n) = 1$ for all n, and $k-1$ in place of k, implies that

$$\left|\zeta^{(k-1)}(s) + \frac{(-1)^k(k-1)!}{(s-1)^k}\right| \ll k!(c_3 \log(|t|+2))^k.$$

Together with Lemma B.0.7, this yields the estimate

$$F_\chi^{(k-1)}(s) \ll \left(\frac{c_4 k \log(2q+|t|)}{\delta(\chi) + (1-\delta(\chi))|L_{q+|t|}(s,\chi)|}\right)^k.$$

This reduces (B.0.13) to showing that

$$\delta(\chi) + (1-\delta(\chi))|L_{q+|t|}(s,\chi)| \gg \begin{cases} 1 & \text{if } |t| \geq 1/\log(3q), \\ q^{-\epsilon/3} & \text{otherwise.} \end{cases} \tag{B.0.14}$$

If, now, $|t| \geq 1/\log(3q) \geq 1/(3\log(q+|t|)$ or χ is complex, then $|L_{q+|t|}(s,\chi)| \asymp 1$ by Lemma B.0.6, so (B.0.14) follows. Also, if $|t| \leq 1/\log(3q)$ and χ is principal, that is to say, $\delta(\chi) = 1$, then we have trivially that $\delta(\chi) + (1-\delta(\chi))|L_{q+|t|}(s,\chi)| = 1$, so (B.0.14) holds in this case too. Finally, if $|t| \leq 1/\log(3q) \leq 1$ and χ is real and non-principal, then $|t| \leq 1/\log(q+|t|)$, and thus Lemma B.0.6 implies that

$$\delta(\chi) + (1-\delta(\chi))|L_{q+|t|}(s,\chi)| = |L_{q+|t|}(s,\chi)| \gg L_{q+|t|}(1,\chi)$$

So, a continuity argument implies that

$$\begin{aligned} L_{q+|t|}(1,\chi) = \lim_{\sigma\to 1^+}\left\{L(\sigma,\chi)\prod_{p\leq q+|t|}\left(1-\frac{\chi(p)}{p^\sigma}\right)\right\} &= L(1,\chi)\prod_{p\leq q+|t|}\left(1-\frac{\chi(p)}{p}\right) \\ &\gg \frac{L(1,\chi)}{\log q} \gg_\epsilon \frac{1}{q^{\epsilon/3}} \end{aligned} \tag{B.0.15}$$

by Theorems B.0.8 and B.0.10, which shows (B.0.14) in this last case too. This completes the proof of (B.0.13).

Next, for every integer $k \geq 3$ and for every real number $y \geq 2$, we apply relations (B.0.2) and (B.0.13) to get that

$$\begin{aligned} \sum_{n\leq y}\Lambda_\chi(n)(\log n)^{k-1}\log\frac{y}{n} &\ll y\int_{|t|\geq\frac{1}{\log(3q)}}\frac{(c_5 k\log(2q))^2 + (c_5 k\log(|t|+1))^k}{t^2+1}dt + y(c_2 k q^{\epsilon/3})^k \\ &\ll y(c_6k^2)^k + y(c_6 k q^{\epsilon/3})^k. \end{aligned}$$

Now, set

$$\Delta(x) = x\sqrt{\log x}\left\{\left(\frac{c_6k^2}{\log x}\right)^{k/2} + \left(\frac{c_6 k q^{\epsilon/3}}{\log x}\right)^{k/2}\right\}$$

and note that $\Delta(x) \geq \sqrt{x}$, since $c_6k^2 > k \geq x^{-1/k}\log x$. We claim that

$$\sum_{n\leq x}\Lambda_\chi(n)(\log n)^{k-1} \ll \Delta(x)(\log x)^{k-1} \quad (x \geq 4). \tag{B.0.16}$$

If $\Delta(x) > x/2$, then (B.0.16) holds trivially. So assume that $\Delta(x) < x/2$. Applying (B.0.2) for $y = x$ and $y = x - \Delta(x)$ and subtracting one inequality from the other completes the proof of (B.0.16). Relation (B.0.16) and partial summation imply that

$$\sum_{n\le x}\Lambda_\chi(n) = O(\sqrt{x}) + \int_{\sqrt{x}}^{x}\frac{1}{(\log t)^{k-1}}d\left(\sum_{n\le t}\Lambda_\chi(n)(\log n)^{k-1}\right) \ll 2^k\Delta(x) \quad (x \ge 16).$$

Choosing

$$k = \min\left\{\frac{\sqrt{\log x}}{4c_6}, \frac{\log x}{4c_6 q^{\epsilon/3}}\right\} + O(1) = \frac{\sqrt{\log x}}{4c_6} + O(1),$$

where we used our assumption that $x \ge \exp\{q^\epsilon\}$, yields the estimate

$$\sum_{n\le x}\Lambda_\chi(n) \ll xe^{-c_7\sqrt{\log x}}, \tag{B.0.17}$$

which proves (B.0.1) and, consequently, completes the proof of the theorem. □

Bibliography

[Da] H. Davenport, *Multiplicative number theory.* Third edition. Revised and with a preface by Hugh L. Montgomery. Graduate Texts in Mathematics, 74. Springer-Verlag, New York, 2000.

[Fo] K. Ford, *Sieve methods.* Lecture notes (2011). `http://www.math.uiuc.edu/~ford/sieve_methods_Sp2011.html`

[FGKT] K. Ford, B. Green, S. Konyagin and T. Tao, *Large gaps between consecutive prime numbers.* Preprint (2014). arXiv:1408.4505

[FGKMT] K. Ford, B. Green, S. Konyagin, J. Maynard and T. Tao, *Long gaps between primes.* Preprint (2014). arXiv:1412.5029

[FG] J. Friedlander and A. Granville, *Limitations to the equi-distribution of primes. I.* Ann. of Math. (2) 129 (1989), no. 2, 363–382.

[FI98] J. Friedlander and H. Iwaniec, *Asymptotic sieve for primes.* Ann. of Math. (2) 148 (1998), no. 3, 1041–1065.

[FI10] —, *Opera de cribro.* American Mathematical Society Colloquium Publications, 57. American Mathematical Society, Providence, RI, 2010.

[GPY] D. A. Goldston, J. Pintz and C. Y. Yildirim, *Primes in tuples. I.* Ann. of Math. (2) 170 (2009), no. 2, 819–862.

[GGPY] D. A. Goldston, S. W. Graham, J. Pintz and C. Y. Yildirim, *Small gaps between primes or almost primes.* Trans. Amer. Math. Soc. 361 (2009), no. 10, 5285–5330.

[Gr95] A. Granville, *Harald Cramér and the distribution of primes.* Harald Cramér Symposium (Stockholm, 1993). Scand. Actuar. J. 1995, no. 1, 12–28.

[Gr15] —, *Primes in intervals of bounded length.* Bull. Amer. Math. Soc. (N.S.) 52 (2015), 171–222.

[Gr12] —, *Analytic number theory.* Lecture notes (2012). `http://www.dms.umontreal.ca/~andrew/Courses/MAT6627.A12.html`

[GS] A. Granville and K. Soundararajan, *Pretentious multiplicative functions and an inequality for the zeta-function.* Anatomy of integers, 191197, CRM Proc. Lecture Notes, 46, Amer. Math. Soc., Providence, RI, 2008.

[HR] H. Halberstam and H.–E. Richert, *Sieve methods.* London Mathematical Society Monographs, No. 4. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974.

[I] H. Iwaniec, *Rosser's sieve.* Acta Arith. 36 (1980), no. 2, 171–202.

[IK] H. Iwaniec and E. Kowalski, *Analytic number theory.* American Mathematical Society Colloquium Publications, 53, American Mathematical Society, Providence, RI, 2004.

[Kou] D. Koukoulopoulos, *Pretentious multiplicative functions and the prime number theorem for arithmetic progressions.* Compos. Math. 149 (2013), no. 7, 1129–1149.

[Kow] E. Kowalski, *Gaps between prime numbers and prime numbers in arithmetic progressions, after Y. Zhang and J. Maynard.* Survey (Bourbaki seminar, March 2014). The english version is available at `people.math.ethz.ch/~kowalski/zhang-bourbaki.pdf`

[Mai81] H. Maier, *Chains of large gaps between consecutive primes.* Adv. in Math. 39 (1981), no. 3, 257–269.

[Mai85] —, *Primes in short intervals.* Michigan Math. J. 32 (1985), no. 2, 221–225.

[May15a] J. Maynard, Small gaps between primes. Ann. of Math. (2) 181 (2015), no. 1, 383–413.

[May] —, Large gaps between primes. Preprint (2014). arXiv:1408.5110

[MV] H. L. Montgomery and R. C. Vaughan, *The large sieve.* Mathematika 20 (1973), 119–134.

[Pi] J. Pintz, *Elementary methods in the theory of L-functions. I. Hecke's theorem.* Acta Arith. **31** (1976), no. 1, 53–60.

[Pol] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes.* Research in the Mathematical Sciences 1:12 (2014). arXiv:1407.4897

[Ra] R. A. Rankin, *The difference between consecutive primes.* J. London Math. Soc. (1938) s1-13 (4): 242–247.

[Ro] G. Rodriquez, *Sul problema dei divisori di Titchmarsh.* (Italian. English summary). Boll. Un. Mat. Ital. (3) 20 1965, 358–366.

[Se47] A. Selberg, *On an elementary method in the theory of primes.* Norske Vid. Selsk. Forh., Trondhjem 19, (1947). no. 18, 64–67.

[Se91] —, *Lectures on sieves.* Collected papers. Vol. II. With a foreword by K. Chandrasekharan. Springer-Verlag, Berlin, 1991.

[Sou] K. Soundararajan, *Small gaps between prime numbers: the work of Goldston-Pintz-Yildirim.* Bull. Amer. Math. Soc. (N.S.) 44 (2007), no. 1, 1–18.

[Taa] T. Tao, *The parity problem is sieve methods.* Blog post: `terrytao.wordpress.com/2007/06/05/open-question-the-parity-problem-in-sieve-theory/`

[Tab] —, *Polymath8b: Bounded intervals with many primes, after Maynard.* Blog post: `terrytao.wordpress.com/2013/11/19/polymath8b-bounded-intervals-with-many-primes-after-maynard/#more-7155`

[Te] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres.* Third edition, coll. Échelles, Belin, 2008.

[Z] Y. Zhang, *Bounded gaps between primes.* Ann. of Math. (2) 179 (2014), no. 3, 1121–1174.