The distribution of prime numbers

Dimitris Koukoulopoulos
University of Montreal

Last update: November 28, 2016

Contents

Chapter 0

Preliminaries

0.1 Asymptotic notation

In Analytic Number Theory, we often have to deal with irregular functions. For example, if we count the number of integers in the interval $[1, x]$ with $x \geq 1$, then we find that this number equals $\lfloor x \rfloor$, the integer part of x, also called the floor function of x. This is a step function with jumps of length 1 at every integer. We thus want to approximate by a nice continuous function. The natural candidate is the function x: we have $\lfloor x \rfloor = x - \{x\}$, where $\{x\}$ is the fractional part of x. Instead of carrying the exact expression for the error term, we observe that $0 \leq \{x\} < 1$ and will often write $\lfloor x \rfloor = x + O(1)$, where $O(1)$ means that the error term in the approximation of $\lfloor x \rfloor$ by x belongs to the class of bounded functions. In general, we write

$$f(x) = O(g(x)) \quad (x \in I) \quad \text{or, equivalently, that} \quad f(x) \ll g(x) \quad (x \in I)$$

if there is a constant $c = c(f, g, I)$ such that

$$|f(x)| \leq c \cdot g(x) \quad \text{for each } x \in I.$$

In this case, we say that the order of magnitude of f in I is smaller than this of g. If $f(x) \ll g(x)$ and $g(x) \ll f(x)$ for $x \in I$, then we write

$$f(x) \asymp g(x) \quad (x \in I)$$

and we say that f and g have the same order of magnitude in I.

Notice that $O(1)+O(1) = O(1)$, since the sum of two bounded functions is also bounded. Similarly, we have $O(g(x)) \pm O(g(x)) = O(g(x))$. This one of the main reasons why this notation is so useful: it allows us to write complicated inequalities in a very compact way. Moreover, it allows to write inequalities as *asymptotic equalities*: if $|f(x) - g(x)| \leq c \cdot h(x)$ for $x \in I$, where c is some constant, then we write

$$f(x) = g(x) + O(h(x)) \quad (x \in I).$$

Finally, we discuss two more asymptotic notations that concern the limiting behaviour of functions: we write

$$f(x) = o(g(x)) \quad (x \to x_0) \quad \Leftrightarrow \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$$

and

$$f(x) \sim g(x) \quad (x \to x_0) \quad \Leftrightarrow \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 1,$$

where in both the above definitions $x_0 \in \widehat{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and g is non-zero in a neighbourhood of x_0.

Exercises

Exercise 0.1.1. Consider the following functions:

$$f_1(x) = x^{1/\log\log x}, \quad f_2(x) = e^{\sqrt{\log x}}, \quad f_3(x) = x, \quad f_4(x) = (\log x)^A, \quad f_5(x) = \sqrt{x},$$
$$f_6(x) = e^x, \quad f_7(x) = \frac{x}{(\log x)^A}, \quad f_8(x) = \frac{x}{e^{\sqrt{\log x}}}, \quad f_9(x) = \log\log x,$$

where $A > 0$ is fixed but arbitrarily large. Order the functions in terms of their order of magnitude as $x \to \infty$, namely find a permutation $\sigma \in S_9$ such that $f_{\sigma(1)}(x) \ll f_{\sigma(2)}(x) \ll \cdots \ll f_{\sigma(9)}(x)$ when $x \to \infty$.

Exercise 0.1.2. Show the following asymptotic estimates :

(a) $\log(1+\delta) = \delta + O(\delta^2)$ for $\delta \in [-1/2, 1/2]$;

(b) $\sqrt{x+1} = \sqrt{x} + O(1/\sqrt{x})$ for $x \geq 1$;

(c) $e^\delta = 1 + O(\delta)$ for $|\delta| \leq 1$;

(d) If $p > 1$, then $\sum_{n>x} 1/n^p \ll \frac{1}{(p-1)x^{p-1}}$ for $x \geq 1$;

(e) If $\{a_n\}_{n=1}^\infty$ is a sequence of non-negative numbers such that

$$a_{n+1} \leq \lambda a_n \quad (n \geq 1),$$

where $\lambda \in (0,1)$ (that is to say, $\{a_n\}_{n\geq 1}$ decreases faster than exponentially), then

$$\sum_{n \geq N} a_n \asymp a_N.$$

0.2 The Riemann-Stieltjes integral

We recall here the definition of the Riemann-Stieltjes integral and two important properties of it that will be very useful when dealing with various sums. Recall that a partition $\mathcal{P}$ of an interval $[a,b]$ is a finite set of points $\mathcal{P} = \{x_0, x_1, \ldots, x_n\}$ such that $a = x_0 < x_1 < \cdots < x_n = b$. Moreover, we say that a partition $\mathcal{P}$ is finer than a partition $\mathcal{Q}$ if $\mathcal{P} \supset \mathcal{Q}$.

Definition 0.2.1. *Let $f, \alpha : [a,b] \to \mathbb{C}$ be two functions. Assume that there is a number $I \in \mathbb{C}$ such that for every $\epsilon > 0$ there is a partition $\mathcal{P}_\epsilon$ of $[a,b]$ with the property that*

$$\left| \sum_{j=1}^n f(\xi_j)(\alpha(x_j) - \alpha(x_{j-1})) - I \right| < \epsilon$$

for any partition $\mathcal{P} = \{x_0, x_1, \dots, x_n\}$ *of* $[a, b]$ *that is finer than* $\mathcal{P}_\epsilon$ *and for any choice of* $\xi_j \in [x_{j-1}, x_j]$, $1 \le j \le n$, *then we say that* f is Riemann-integrable with respect to α (on the interval $[a, b]$) *and write* $f \in \mathcal{R}(\alpha; [a, b])$. *The number* I *is uniquely determined and called the* Riemann-Stieltjes integral of f with respect to α (over the interval $[a, b]$). *It is denoted by* $\int_a^b f(x)\mathrm{d}\alpha(x)$, *or by* $\int_a^b f\mathrm{d}\alpha$.

We will need the following four theorems about the Riemann-Stieltjes integral. For their proofs and for more details on the theory of the Riemann-Stieltjes integral, we refer the reader to [1, Chapter 7].

Theorem 0.2.2. *If* $f \in \mathcal{R}(\alpha; [a, b])$ *and* $\alpha \in C^1([a, b])$, *then*

$$\int_a^b f(x)d\alpha(x) = \int_a^b f(x)\alpha'(x)\mathrm{d}x,$$

where the second integral is a regular Riemann integral.

Theorem 0.2.3. *If* $f \in \mathcal{R}(\alpha; [a, b])$ *and* $\alpha \in \mathcal{R}(f; [a, b])$, *then*

$$\int_a^b f(x)d\alpha(x) = f(x)\alpha(x)\Big|_{x=a}^b - \int_a^b \alpha(x)\mathrm{d}f(x).$$

Theorem 0.2.4. *Let* $\alpha : [a, b] \to \mathbb{C}$ *be a step function, so that there is a partition* $\mathcal{P} = \{x_0, x_1, \dots, x_n\}$ *of* $[a, b]$ *such that* α *is constant in each interval of the form* (x_{j-1}, x_j), $1 \le j \le n$, *and let*

$$\alpha_j = \begin{cases} \alpha(a^+) - \alpha(a) & \text{if } j = 0, \\ \alpha(x_j^+) - \alpha(x_j^-) & \text{if } 1 \le j \le n-1, \\ \alpha(b) - \alpha(b^-) & \text{if } j = n. \end{cases}$$

If $f : [a, b] \to \mathbb{R}$ *is a function such that there is no* $x \in [a, b]$ *for which both* f *and* α *are simultaneously discontinuous from the right or from the left, then*

$$\int_a^b f(x)\mathrm{d}\alpha(x) = \sum_{j=0}^n f(x_j)\alpha_j.$$

Combining the above results, we prove the following identity, which will often refer to as "summation by parts" or "partial summation".

Theorem 0.2.5. *Let* $(a_n)_{n\ge1}$ *be a sequence of complex numbers and* $f \in C^1([1, +\infty))$. *If* $A(x) = \sum_{n\le x} a_n$, *then*

$$\sum_{y<n\le z} a_n f(n) = A(t)f(t)\Big|_{t=y}^z - \int_y^z A(t)f'(t)\mathrm{d}t.$$

Proof. Consider the integral $\int_y^z f(t)\mathrm{d}A(t)$. The function $A(t)$ is a step function that has jumps by a_n whenever $n \in \{\lfloor y\rfloor + 1, \dots, \lfloor z\rfloor\} = \mathbb{Z} \cap (y, z]$. Together with Theorem 0.2.4, this implies that

$$\int_y^z f(t)\mathrm{d}A(t) = \sum_{y<n\le z} f(n)a_n.$$

On the other hand, integration by parts (i.e. Theorem 0.2.3) yields the formula

$$\int_y^z f(t)\mathrm{d}A(t) = f(t)A(t)\Big|_{t=y}^z - \int_y^z A(t)\mathrm{d}f(t).$$

Finally, we apply Theorem 0.2.2 to rewrite the rightmost integral in the above relation. This proves the claimed identity. □

Remark 0.2.1. It is also possible to prove Theorem 0.2.5 without going through the theory of Riemann-Stieltjes integration: we set $N = \lfloor y \rfloor$ and $M = \lfloor z \rfloor$, and apply Abel's summation formula

$$\begin{aligned}\sum_{n=N+1}^{M} a_n f(n) &= \sum_{n=N+1}^{M} (A(n) - A(n-1))f(n) \\ &= \sum_{n=N+1}^{M} A(n)f(n) - \sum_{n=N}^{M-1} A(n)f(n+1) \\ &= A(n)f(n+1)\Big|_{n=N}^{M} - \sum_{n=N+1}^{M} A(n)(f(n+1) - f(n)).\end{aligned}$$

We then observe that the Fundamental Theorem of Integral Calculus implies the identities

$$A(n)(f(n+1) - f(n)) = \int_n^{n+1} A(t)f'(t)\mathrm{d}t,$$

$$A(N)f(N+1) - A(y)f(y) = A(N)(f(N+1) - f(y)) = \int_y^{N+1} A(t)f'(t)\mathrm{d}t$$

and

$$A(M)f(M+1) - A(z)f(z) = A(M)(f(M+1) - f(z)) = \int_z^{M+1} A(t)f'(t)\mathrm{d}t,$$

thus proving Theorem 0.2.5.

Exercises

Exercise 0.2.1. If $(a_n)_{n\geq 1}$ is a sequence of complex numbers and $A(x) = \sum_{n\leq x} a_n$, then show that

$$\sum_{n\leq x} a_n \log n = A(x)\log x - \int_1^x \frac{A(t)}{t}\mathrm{d}t.$$

Deduce that

$$\sum_{n\leq x} \log n = x\log x - x + O(\log x) \quad (x \geq 2).$$

Chapter 1

Arithmetic functions

1.1 The ring of arithmetic functions

An *arithmetic function* is a function $f : \mathbb{N} \to \mathbb{C}$. The set of all arithmetic functions is denoted by $\mathcal{A}$. An arithmetic function f is called *multiplicative* if $f(1) = 1$ and

$$f(mn) = f(m)f(n) \quad \text{whenever} \quad (m, n) = 1, \tag{1.1.1}$$

whereas f is called *completely multiplicative* if $f(1) = 1$ and relation (1.1.1) holds for all m and n, without the requirement that they are co-prime. We write $\mathcal{M}$ for the set of all multiplicative functions and $\mathcal{M}^*$ for the set of all completely multiplicative functions. Important examples of multiplicative functions include:

- the functions n^s, where s is a fixed complex number;
- the divisor functions

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

- the Möbius function

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

- the Euler totient function

$$\varphi(n) = \#\{1 \le a \le n : (a, n) = 1\};$$

- the sum-of-divisors function

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

A related notion is that of an *additive function*, that is to say of an arithmetic function $f : \mathbb{N} \to \mathbb{C}$ satisfying the functional relation

$$f(mn) = f(m) + f(n) \quad \text{whenever} \quad (m, n) = 1. \tag{1.1.2}$$

Clearly, if f is an additive function, then v^f is a multiplicative function, for any fixed $v > 0$, and we can even take $v \in \mathbb{C} \setminus \{0\}$ if f only takes integer values. Two of the most important examples of additive functions are

$$\omega(n) := \sum_{p|n} 1 \quad \text{and} \quad \Omega(n) := \sum_{p^\alpha \| n} \alpha.$$

An important operation on $\mathcal{A}$, which allows us to create new functions from existing ones, is the *Dirichlet convolution*: 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

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

Combining all of the above, we conclude that $(\mathcal{A}, +, *)$ forms a commutative and unitary ring called *the ring of arithmetic functions*. Moreover, an arithmetic function f is a unit in $\mathcal{A}$ (that is to say, there is $g : \mathbb{N} \to \mathbb{C}$ with $f * g = \mathbb{1}$, also called the Dirichlet inverse of f) exactly when $f(1) \neq 0$. In particular, any multiplicative function has a Dirichlet inverse. Combining all of the above, we conclude that $(\mathcal{M}, *)$ is an abelian group. Moreover, $(\mathcal{A}, +, *)$ forms a ring called *the ring of arithmetic functions*.

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*

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1, \end{cases} \tag{1.1.3}$$

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 (1.1.3) is that if f is completely multiplicative, then its convolution inverse is given by μf.

When we later study the distribution of primes, a prominent role will be played by *von Mangoldt's function*

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

We note that

$$\log n = \sum_{p^\alpha \| n} \log(p^\alpha) = \sum_{p^k | n} \log p = (\Lambda * 1)(n).$$

So we deduce by Möbius inversion that

$$\Lambda = \mu * \log . \tag{1.1.4}$$

Note that

$$\sum_{d|n} \mu(d) \log(n/d) = (\log n) \sum_{d|n} \mu(d) - \sum_{d|n} \mu(d) \log d,$$

whence we also deduce the formula

$$\Lambda = -\mu \log * 1. \tag{1.1.5}$$

Exercises

Exercise 1.1.1. Prove all of the assertions made above about the ring of arithmetic functions.

Exercise 1.1.2. Let $f(n) = \varphi(n)/n$, and let $\{n_k\}_{k=1}^{\infty}$ be the sequence of values n at which f attains a "record low"; i.e., $n_1 = 1$ and, for $k \geq 2$, n_k is defined as the smallest integer $> n_{k-1}$ with $f(n_k) < f(n)$ for all $n < n_k$. (For example, since the first few values of the sequence $f(n)$ are $1, 1/2, 2/3, 1/2, 4/5, 1/3, \ldots$, we have $n_1 = 1$, $n_2 = 2$, and $n_3 = 6$, and the corresponding values of f at these arguments are $1, 1/2$ and $1/3$.) Find (with proof) a general formula for n_k and $f(n_k)$.

Exercise 1.1.3. Assume f is multiplicative. Prove that:

(a) $f^{-1}(n) = \mu(n) f(n)$ for square-free n;

(b) $f^{-1}(p^2) = f(p)^2 - f(p^2)$ for every prime p.

Exercise 1.1.4. Let

$$f(n) = \#\{(n_1, n_2) \in \mathbb{N}^2 : [n_1, n_2] = n\},$$

where $[n_1, n_2]$ is the least common multiple of n_1 and n_2. Show that f is multiplicative and evaluate f at prime powers.

Exercise 1.1.5. (a) Show that

$$\tau_k(p^\alpha) = \binom{\alpha + k - 1}{k - 1}$$

and deduce that

$$k^{\omega(n)} \leq \tau_k(n) \leq k^{\Omega(n)}.$$

(b) Prove that, for every fixed $\epsilon > 0$ and $k \geq 2$, we have that

$$\tau_k(n) \ll_{\epsilon,k} n^\epsilon \quad (n \geq 1).$$

Exercise 1.1.6. For each $k \in \mathbb{N}$, we define the k-th generalized von Mangoldt function

$$\Lambda_k := \mu * \log^k,$$

so that Λ_1 is the usual von Mangoldt function Λ.

(a) Prove that
$$\Lambda_{k+1} = \Lambda_k \log + \Lambda_k * \Lambda.$$

(b) Prove that Λ_k is supported on integers n with $\omega(n) \leq k$.

(c) Show that $0 \leq \Lambda_k(n) \leq (\log n)^k$ for each $n \in \mathbb{N}$.

(d) Show that if $n = p_1 \cdots p_k$, where $p_1, \ldots, p_k$ are distinct primes, then
$$\Lambda_k(n) = (\log p_1) \cdots (\log p_k).$$

1.2 Summation by parts

Often, arithmetic functions behave in a irregular way as n varies. For example, if f is multiplicative, then its value at n depends intimately on the multiplicative structure of n. Even if two integers n and m are close together from an archimedian point of view (i.e. $|m-n|$ is small), their multiplicative structures might be completely different, thus leading to different f values. For example, $\tau(119) = 4$, but $\tau(120) = 32$. Such irregularities tend to disappear when we average our arithmetic function over long intervals. This brings us closer to a statistical point of view, where we study f *on average*. To this end, we consider the partial sums of f

$$\sum_{n \leq x} f(n).$$

These are step functions in x with jump discontinuities whenever $x \in \mathbb{N}$ with $f(x) \neq 0$. So, rather than aiming to obtain exact expressions for $\sum_{n \leq x} f(n)$, our goal will be to estimate this quantity *asymptotically* for different cases of interesting functions f.

A simple but fundamental tool in understanding partial sums of arithmetic functions is the summation by parts formula (cf. Theorem 0.2.5). We recast it here in a more versatile form:

Theorem 1.2.1. *Let $(a_n)_{n \geq 1}$ be a sequence of complex numbers and $f \in C^1([a,b])$. Assume further that*

$$\sum_{n \leq x} a_n = M(x) + R(x) \quad (a \leq x \leq b)$$

where $M \in C^1([a,b])$ and R is the remainder term in the approximation of $\sum_{n \leq x} a_n$ by $M(x)$. If $a \leq y \leq z \leq b$, then

$$\sum_{y < n \leq z} a_n f(n) = \int_y^z f(t) M'(t) \mathrm{d}t + R(t) f(t) \Big|_{t=y}^{z} - \int_y^z R(t) f'(t) \mathrm{d}t.$$

Proof. This follows by Theorem 0.2.5 and integration by parts. □

Applying Theorem 1.2.1 requires knowledge on the partial sums of a_n. This is easy to guarantee if $a_n = 1$. In this special case, Theorem 1.2.1 is the famous Euler-McLaurin formula:

Theorem 1.2.2 (Euler-McLaurin summation formula)**.** *If $f \in C^1([y,z])$, then*

$$\sum_{y<n\leq z} f(n) = \int_y^z f(t)\mathrm{d}t - \{t\}f(t)\Big|_{t=y}^{z} + \int_y^z \{t\}f'(t)\mathrm{d}t.$$

In particular, if $f \in C^1([1,+\infty))$, then for every $x \geq 1$ we have that

$$\sum_{n\leq x} f(n) = \int_1^x f(t)\mathrm{d}t + f(1) - \{x\}f(x) + \int_1^x \{t\}f'(t)\mathrm{d}t.$$

Remark 1.2.1. We may think of $\sum_{y<n\leq z} f(n)$ as a Riemann sum for the integral $\int_y^z f(t)\mathrm{d}t$. Of course, here we are summing over points that are 1 apart to approximate the integral, so it could be that the difference $\sum_{y\leq n\leq z} f(n) - \int_y^z f(t)\mathrm{d}t$ is rather large. However, if f does not oscillate too wildly (that is to say, if its derivative is usually small), then this difference should be quite small. The Euler-McLaurin formula is a quantification of this idea.

Proof. We note that $\sum_{n\leq x} 1 = \lfloor x \rfloor = x - \{x\}$. The result then follows by Theorem 1.2.1. □

As an application, we prove the following theorem:

Theorem 1.2.3. *There is a constant $\gamma \in \mathbb{R}$, called the Euler-Mascheroni constant, such that*

$$\sum_{n\leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right) \quad (x \geq 1).$$

Proof. By Theorem 1.2.2, we have that

$$\begin{aligned}\sum_{n\leq x} \frac{1}{n} &= \int_1^x \frac{\mathrm{d}t}{t} + 1 - \frac{\{x\}}{x} - \int_1^x \frac{\{t\}}{t^2}\mathrm{d}t \\ &= \log x + 1 - \int_1^x \frac{\{t\}}{t^2}\mathrm{d}t + O\left(\frac{1}{x}\right).\end{aligned}$$

The integral $\int_1^\infty (\{t\}/t^2)\mathrm{d}t$ converges absolutely. Moreover, we have the tail bound

$$\left|\int_x^\infty \frac{\{t\}}{t^2}\mathrm{d}t\right| \leq \int_x^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{x}.$$

Therefore, setting

$$\gamma := 1 - \int_1^\infty \frac{\{t\}}{t^2}\mathrm{d}t \tag{1.2.1}$$

completes the proof of the theorem. □

A more involved application of Theorem 1.2.2 is given in Stirling's approximation for $n!$:

Theorem 1.2.4 (Stirling's formula)**.** *For $n \geq 1$, we have that*

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Proof. Taking logarithms and applying the Euler-McLaurin formula, we find that

$$\begin{aligned}\log(n!) = \sum_{j=1}^{n} \log j &= \int_1^n \log x \,\mathrm{d}x + \log 1 - \{n\}\log n + \int_1^n \frac{\{t\}}{t}\mathrm{d}t \\ &= n\log n - n + 1 + \int_1^n \frac{\{t\}}{t}\mathrm{d}t,\end{aligned}$$

since $n \in \mathbb{N}$ here and thus $\{n\} = 0$. Next, set

$$F(x) = \int_0^x (\{t\} - 1/2)\mathrm{d}t$$

and note that $F(x) \ll 1$ by the periodicity of $\{t\} - 1/2$ and the fact that it has mean 0 over a complete period. Integration by parts implies that

$$\begin{aligned}\int_1^n \frac{\{t\}}{t}\mathrm{d}t = \frac{\log n}{2} + \int_1^n \frac{\{t\} - 1/2}{t}\mathrm{d}t &= \frac{\log n}{2} + \frac{F(t)}{t}\Big|_{t=1}^{n} + \int_1^n \frac{F(t)}{t^2}\mathrm{d}t \\ &= \frac{\log n}{2} + \int_1^n \frac{F(t)}{t^2}\mathrm{d}t.\end{aligned}$$

(Justify why we can integrate by parts even though F is not differentiable everywhere.) The integral $\int_1^\infty (F(t)/t^2)dt$ converges absolutely by the estimate $F(t) \ll 1$ and its tails satisfy the estimate

$$\int_n^\infty \frac{F(t)}{t^2}\mathrm{d}t \ll \int_n^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{n}.$$

This proves the theorem provided that we can show that $e^c = \sqrt{2\pi}$. The proof of this identity is outlined in Exercise 1.2.5 below. □

One of our main objectives in these notes is to estimate sums involving prime numbers. The simplest such sum is

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

the counting function of primes up to x. As we will see, the most basic form of the Prime Number Theorem states that

$$\pi(x) \sim \frac{x}{\log x} \quad (x \to \infty). \tag{1.2.2}$$

If $f \in C^1([1, +\infty))$, then partial summation (see Theorem 0.2.5 with $z = x$ and $y = 2^-$) implies that

$$\sum_{p \leq x} f(p) = f(x)\pi(x) - \int_2^x \pi(t)f'(t)\mathrm{d}t.$$

For example, using (1.2.2) and the above identity, can estimate the function

$$\theta(x) := \sum_{p \le x} \log p,$$

an important variant of $\pi(x)$ introduced by Chebyshev. A closely related function, also introduced by Chebyshev is

$$\psi(x) := \sum_{n \le x} \Lambda(n) = \sum_{p^k \le x} \log p.$$

The reader is invited to the equivalence of (1.2.2) to the asymptotic estimates $\theta(x) \sim x$ and $\psi(x) \sim x$ in Exercise 1.2.4.

Exercises

Exercise 1.2.1. (a) Prove that

$$\sum_{n \le x} \sqrt{n} = \frac{2}{3} x^{3/2} + O(\sqrt{x}) \quad (x \ge 1)$$

(b) Prove that there is some constant c such that

$$\sum_{n \le x} \frac{1}{\sqrt{n}} = 2\sqrt{x} + c + O\left(\frac{1}{\sqrt{x}}\right) \quad (x \ge 1).$$

Exercise 1.2.2. (a) Observe that if $f : [1, +\infty) \to [0, \infty)$ is a decreasing function, then

$$f(n) \le \int_{n-1}^{n} f(t) \mathrm{d}t,$$

and deduce that

$$\sum_{n > x} f(n) \le \int_{\lfloor x \rfloor}^{\infty} f(t) \mathrm{d}t.$$

(b) Prove that

$$\sum_{n > x} \frac{1}{n^{1+\delta}} \le \frac{1}{\delta (x-1)^{\delta}} \quad (\delta > 0,\ x \ge 2).$$

Exercise 1.2.3. Prove that

$$\sum_{n \le x} \omega(n) = x \sum_{p \le x} \frac{1}{p} + O(\pi(x)),$$

as well as that

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

Exercise 1.2.4. Prove that the following estimates are equivalent:

(a) $\pi(x) \sim x/\log x \quad (x \to \infty)$

(b) $\theta(x) \sim x \quad (x \to \infty)$

(c) $\psi(x) \sim x \quad (x \to \infty)$

Exercise 1.2.5. (a) Show that, for each integer $n \geq 0$,

$$I_n := \int_0^{\pi/2} (\cos x)^n \mathrm{d}x = \begin{cases} \dfrac{\pi}{2} \cdot \dfrac{1 \cdot 3 \cdots (2k-1)}{2 \cdot 4 \cdots (2k)} = \dfrac{\pi}{2} \cdot \dfrac{\binom{2k}{k}}{4^k} & \text{if } n = 2k, \\ \dfrac{2 \cdot 4 \cdots (2k)}{1 \cdot 3 \cdots (2k+1)} = \dfrac{4^k}{(2k+1)\binom{2k}{k}} & \text{if } n = 2k+1, \end{cases}$$

(b) Show that

$$\lim_{n \to \infty} \frac{I_{n+1}}{I_n} = 1.$$

(c) Deduce that the constant c in the proof of Theorem 1.2.4 satisfies $e^c = \sqrt{2\pi}$.

1.3 Convolution sums

One of the most powerful techniques in estimating partial sums of arithmetic functions f is to decompose them as the convolution $f = g * h$ of two other arithmetic functions whose partial sums we already understand. Then we note that

$$\sum_{n \leq x} f(n) = \sum_{n \leq x} \sum_{ab=n} g(a)h(b) = \sum_{ab \leq x} g(a)h(b). \tag{1.3.1}$$

The sum on the right hand side of (1.3.1) can be re-arranged in many possible ways. We discuss the two main techniques below.

The convolution method

If $h(b)$ decays fast in b, or is supported on small integers b, then we re-arrange the sum on the right hand side of (1.3.1) as

$$\sum_{n \leq x} f(n) = \sum_{b \leq x} h(b) \sum_{a \leq x/b} g(a).$$

Since the dominant contribution to this sum is supported on relatively small b's compared to x, the a variable is usually summed over a long interval $[1, x/b]$. So, inserting our asymptotic estimates on the partial sums of g allows to obtain some control on the partial sums of f.

The above idea works best when f, g and h are multiplicative. In order to implement it, we must choose the functions g and h efficiently. Given a multiplicative function f whose partial sums we want to estimate, we choose a simpler function g that satisfies two properties:

- f is 'close' to g;
- we already know how to estimate the partial sums of g.

By multiplicativity, f and g are close to each other if $f(p) \approx g(p)$ for all primes p (or, perhaps, in an average sense). Then we define h implicitly via the relation $f = g * h$. In particular, $h(p) = f(p) - g(p)$, so h has small prime values (pointwise or on average). For example, if $f = \varphi$, then $f(p) = p - 1 \approx p$, so we may take $g(n) = n$, since we already know that $\sum_{n \le x} n = x^2/2 + O(x)$.

We illustrate in detail the technique described above, often called *the convolution method*, by proving the following classical estimate:

Theorem 1.3.1. *For $x \ge 1$,*

$$\#\{n \le x : n \text{ square-free}\} = \frac{6}{\pi^2} \cdot x + O(\sqrt{x}).$$

Proof. The characteristic function of square-free integers is μ^2. We want to write $\mu^2 = g * h$, where g is a 'nice' function we already understand and h is a 'small' function. Since $\mu^2(p) = 1$, we take $g = 1$ and define h implicitly via the equation $\mu^2 = 1 * h$. Note that $\mu^2(p) = 1 + h(p)$, whence $h(p) = 0$. In general, $h = \mu * \mu^2$, and a simple calculation implies that

$$h(p^k) = \begin{cases} -1 & \text{if } k = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore

$$\begin{aligned} \sum_{n \le x} \mu^2(n) = \sum_{ab \le x} h(b) = \sum_{ad^2 \le x} \mu(d) = \sum_{d \le \sqrt{x}} \mu(d) \sum_{a \le x/d^2} 1 = \sum_{d \le \sqrt{x}} \mu(d) \left(\frac{x}{d^2} + O(1) \right) \\ = x \sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2} + O(\sqrt{x}). \end{aligned}$$

We want to extend the summation to all d. Estimating the tales as

$$\left| \sum_{d > \sqrt{x}} \frac{\mu(d)}{d^2} \right| \le \sum_{d > \sqrt{x}} \frac{1}{d^2} \ll \frac{1}{\sqrt{x}},$$

by Exercise 1.2.2(b), we deduce that

$$\sum_{n \le x} \mu^2(n) = x \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} + O(\sqrt{x}).$$

Finally, as we can see by a double application of Theorem 1.3.2 below, we have that

$$\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \prod_p \left(1 - \frac{1}{p^2} \right) = \left(\sum_{n=1}^{\infty} \frac{1}{n^2} \right)^{-1} = \frac{6}{\pi^2}.$$

□

To complete the proof of Theorem 1.3.1, we need to justify the last calculation. As the following theorem establishes, the complete series of a multiplicative function can be written as a product over primes called *Euler product*. The name is in honor of Leonhard Euler, who discovered the first instance of Theorem 1.3.2 when $f(n) = 1/n^\alpha$ with $\alpha > 1$, a case we will revisit when studying the Riemann zeta function in Chapter 4.

Theorem 1.3.2. *Let f be a multiplicative function. Then the series $\sum_{n=1}^{\infty} f(n)$ converges absolutely if, and only if, the series $\sum_{p \text{ prime}, k\geq 1} f(p^k)$ converges absolutely. Moreover, in the case when they both converge absolutely, we have that*

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

In particular, if f is completely multiplicative, then

$$\sum_{n=1}^{\infty} f(n) = \prod_{p} \frac{1}{1 - f(p)}.$$

Proof. If $\sum_{n\geq 1} f(n)$ converges absolutely, then it is clear that the same is true for the subsum over prime powers $\sum_{p \text{ prime}, k\geq 1} f(p^k)$ too. Conversely, assume that $\sum_{p \text{ prime}, k\geq 1} f(p^k)$ converges absolutely. Then the fundamental theorem of arithmetic and positivity imply that

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

So we conclude that $\sum_{n\geq 1} f(n)$ converges absolutely too, as desired.

It remains to prove the claimed Euler product representation in the case that $\sum_{n\geq 1} f(n)$ converges absolutely. We will re-arrange the terms of the series using a multiplicative ordering: the fundamental theorem of arithmetic implies that

$$\sum_{\substack{p|n \,\Rightarrow\, p\leq M \\ \nu_p(n)\leq N}} f(n) = \prod_{p\leq M} \left(1 + f(p) + f(p^2) + \cdots + f(p^N)\right)$$

Letting $N \to \infty$ implies that

$$\sum_{p|n \,\Rightarrow\, p\leq M} f(n) = \prod_{p\leq M} \left(1 + f(p) + f(p^2) + \cdots\right).$$

Next, we let $M \to \infty$ to derive the claimed formula for $\sum_{n=1}^{\infty} f(n)$. Finally, if f is completely multiplicative, then we note that

$$1 + f(p) + f(p^2) + \cdots = 1 + f(p) + f(p)^2 + \cdots = \frac{1}{1 - f(p)}.$$

This completes the proof of the theorem. □

The hyperbola method

The convolution method described above is of limited applicability because it requires the function h to be small. Dirichlet discovered another way of re-arranging the right hand side of (1.3.1) that allows us to estimate more general convolutions sums: if $f = g * h$ and $A, B \geq 1$ are two parameters with $AB = x$, we have the general formula

$$\begin{aligned}\sum_{n\leq x} f(n) &= \sum_{ab\leq x} g(a)h(b)\\ &= \sum_{\substack{ab\leq x\\ a\leq A}} g(a)h(b) + \sum_{\substack{ab\leq x\\ a>A}} g(a)h(b)\\ &= \sum_{a\leq A} g(a) \sum_{b\leq x/a} h(b) + \sum_{b\leq B} h(b) \sum_{A<a\leq x/b} g(a) \qquad (1.3.2)\\ &= \sum_{a\leq A} g(a) \sum_{b\leq x/a} h(b) + \sum_{b\leq B} h(b) \sum_{a\leq x/b} g(a) - \sum_{a\leq A} g(a) \sum_{b\leq B} h(b). \qquad (1.3.3)\end{aligned}$$

If A and B are both large but not too close to x, then all sums in the right hand side of (1.3.2) are long, so that it is easier to estimate them accurately if we understand the partial sums of g and h.

The technique described above is called the *hyperbola method*, borrowing its name from the fact that the range of summation $ab \leq x$ consists of all lattice points $(a, b) \in \mathbb{N}^2$ under the hyperbola $ab = x$. It was invented by Dirichlet while proving the following classical result:

Theorem 1.3.3. *For $x \geq 1$, we have*

$$\sum_{n\leq x} \tau(n) = x \log x + (2\gamma - 1)x + O(\sqrt{x}).$$

Proof. We have that $\tau = 1 * 1$. Applying (1.3.3), we have that

$$\begin{aligned}\sum_{n\leq x} \tau(n) &= \sum_{a\leq A} \sum_{b\leq x/a} 1 + \sum_{b\leq B} \sum_{a\leq x/b} 1 - \sum_{a\leq A} 1 \sum_{b\leq B} 1\\ &= \sum_{a\leq A} \Big(\frac{x}{a} + O(1)\Big) + \sum_{b\leq x/A} \Big(\frac{x}{b} + O(1)\Big) - (A + O(1))(B + O(1))\\ &= x\left(\sum_{a\leq A} \frac{1}{a} + \sum_{b\leq B} \frac{1}{b}\right) - AB + O(A + B).\end{aligned}$$

By a double application of Theorem 1.2.3 and recalling that $AB = x$, we deduce that

$$\begin{aligned}\sum_{n\leq x} \tau(n) &= x\left(\log(AB) + 2\gamma + O\left(\frac{1}{A} + \frac{1}{B}\right)\right) - AB + O(A + B)\\ &= x \log x + (2\gamma - 1)x + O(A + B).\end{aligned}$$

The optimal choice is $A = B = \sqrt{x}$, which yields the claimed result. □

Exercises

Exercise 1.3.1. Prove that

$$\sum_{n\le x}\frac{\varphi(n)}{n}=\frac{6}{\pi^2}x+O(\log x)\quad(x\ge 2).$$

and that

$$\sum_{n\le x}\varphi(n)=\frac{6}{\pi^2}x^2+O(x\log x)\quad(x\ge 2).$$

Exercise 1.3.2. Prove that there is some constant c such that

$$\sum_{n\le x}\frac{\mu^2(n)}{\varphi(n)}=\log x+c+O\left(\frac{\log x}{x}\right)\quad(x\ge 2).$$

Exercise 1.3.3. Prove that there is a constant c such that

$$\sum_{n\le x}2^{\omega(n)}=\frac{6}{\pi^2}x\log x+cx+O(x^{2/3})\quad(x\ge 1).$$

[*Hint*: Show that $2^\omega=1*\mu^2$.]

Exercise 1.3.4. Prove that, for every fixed integer $k\ge 3$, there is a polynomial P_k of degree $k-1$ such that

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

Moreover, calculate the leading coefficient of P_k.

Chapter 2

Elementary prime number theory

In 1792 or 1793, when Gauss was 15 or 16 years old, he conjectured that the density of prime numbers around x is about $1/\log x$. Translated in the language of calculus, this means that

$$\pi(x) \sim \mathrm{li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t}.$$

It is not so hard to show that

$$\mathrm{li}(x) \sim \frac{x}{\log x} \quad (x \to \infty).$$

More precisely, integrating by parts N times, we may prove the asymptotic expansion

$$\mathrm{li}(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} + \frac{2!x}{\log^3 x} + \cdots + \frac{(N-1)!x}{\log^N x} + O_N\left(\frac{x}{\log^{N+1} x}\right).$$

The reader is invited to prove these claims as an exercise. The asymptotic relation

$$\pi(x) \sim \frac{x}{\log x} \quad (x \to \infty),$$

proven independently in 1896 by de la Vallée-Poussin and Hadamard became to be known as the *Prime Number Theorem*. We will prove this result in Section 4.4 in a precise quantitative form. In this chapter, we develop various elementary techniques that enable to get some initial control on $\pi(x)$.

As we saw in Exercise 1.2.4, we may state the Prime Number Theorem in terms of the asymptotic behaviour of Chebyshev's functions $\theta(x)$ and $\psi(x)$. What is far less obvious is that the Prime Number Theorem is equivalent to proving that

$$\sum_{n \le x} \mu(n) = o(x) \quad (x \to \infty). \tag{2.0.1}$$

There is a good heuristic explanation for the above estimate: $\mu(n)$ counts n with the weight 1 if it square-free integer with an even number of prime factors, whereas it counts it with the weight -1 if it is square-free integer with an odd number of prime factors. There should not be any bias towards an even or an odd number of prime factors of a theorem, which naturally leads us to conjecturing (2.0.1). A quick way to connect primes and the Möbius function is the convolution identity (1.1.4). In Section 2.3, we will see how to use (1.1.4) to prove that the Prime Number Theorem is equivalent to the estimate (2.0.1).

2.1 Chebyshev's and Mertens's estimates

The first important rigorous result on $\pi(x)$ was proven by Chebyshev, who proved, among other things, that Gauss's guess was of the right order of magnitude:

Theorem 2.1.1 (Chebyshev)**.** *For $x \geq 2$,*

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

Proof. We start with the implicit upper bound. Let

$$B = \binom{2n}{n} = \frac{(2n)!}{n!^2} = \frac{(2n)(2n-1)\cdots(n+1)}{n!}.$$

and note that if $p \in (n, 2n]$, then $p|B$. Therefore, the product $\prod_{n<p\leq 2n} p$ divides B. Since we also have that

$$B \leq \sum_{j=0}^{2n} \binom{2n}{j} = 4^n,$$

we deduce that

$$\prod_{n<p\leq 2n} p \leq 4^n \quad \Longrightarrow \quad \sum_{n<p\leq 2n} \log p \leq n \log 4 \quad \Longrightarrow \quad \pi(2n) - \pi(n) = \sum_{n<p\leq 2n} 1 \leq \frac{n \log 4}{\log n}$$

If $x \geq 2$, then there is $k \geq 1$ such that $x \in (2^{k-1}, 2^k]$. Then

$$\pi(x) \leq \pi(2^k) = \sum_{j=1}^{k} (\pi(2^j) - \pi(2^{j-1})) \ll \sum_{j=1}^{k} \frac{2^j}{j} \asymp \frac{2^k}{k} \asymp \frac{x}{\log x}.$$

Finally, we show the implicit lower bound. We will study the factorization of the number B, which we write

$$B = \prod_{p\leq 2n} p^{e_p},$$

with $e_p \geq 0$. We claim that if $p \leq 2n$, then $p^{e_p} \leq 2n$. Indeed, if $\nu_p(m)$ denotes the p-adic valuation of the number m (that is to say, the highest power of p that divides m), then we note that $e_p = v_p(B) = v_p((2n)!) - 2v_p(n!)$. Moreover, we have the identity

$$v_p(m!) = \sum_{j\geq 1} \#\{k \leq m : p^j | k\} = \sum_{j\geq 1} \#\{p^j \ell \leq m : \ell \geq 1\} = \sum_{j\geq 1} \left\lfloor \frac{m}{p^j} \right\rfloor,$$

which implies that

$$e_p = \sum_{j\geq 1} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2 \left\lfloor \frac{n}{p^j} \right\rfloor \right) = \sum_{\substack{j\geq 1 \\ p^j \leq 2n}} \left(\left\lfloor \frac{2n}{p^j} \right\rfloor - 2 \left\lfloor \frac{n}{p^j} \right\rfloor \right),$$

where we used the fact that if $p^j > 2n$, then $\lfloor 2n/p^j \rfloor = \lfloor n/p^j \rfloor = 0$. For each $x \in [0,1)$, we have that

$$\lfloor 2x \rfloor - 2\lfloor x \rfloor = \begin{cases} 0 & \text{if } 0 \le x < 1/2, \\ 1 & \text{if } 1/2 \le x < 1. \end{cases}$$

In particular, $\lfloor 2x \rfloor - 2\lfloor x \rfloor \le 1$ for each $x \in [0,1)$ and thus for each $x \in \mathbb{R}$ by periodicity. Therefore

$$e_p \le \#\{j \ge 1 : p^j \le 2n\} = \#\left\{ j \in \mathbb{N} : 1 \le j \le \frac{\log(2n)}{\log p} \right\} = \left\lfloor \frac{\log(2n)}{\log p} \right\rfloor \le \frac{\log(2n)}{\log p},$$

which proves our assertion that $p^{e_p} \le 2n$. We may now complete the proof: we have that

$$B = \prod_{p \le 2n} p^{e_p} \le \prod_{p \le 2n} (2n) = (2n)^{\pi(2n)} \quad \Longrightarrow \quad \pi(2n) \ge \frac{\log B}{\log(2n)} \asymp \frac{n}{\log n},$$

since $B \asymp 4^n/\sqrt{n}$ by Stirling's formula (Theorem 1.2.4). Choosing $n = \lfloor x/2 \rfloor$ completes the proof of the lower bound as well. □

Using Chebyshev's result and Stirling's formula, Mertens proved the asymptotic behaviour of various sums involving primes. However, note that in Mertens's estimates, the primes are weighted with a function that decays fast to zero thus diminishing the effect of each individual prime. In comparison, each prime is counted with weight 1 in $\pi(x)$, thus playing a much more important role.

Theorem 2.1.2 (Mertens)**.**

(a) For $x \ge 1$, we have that

$$\sum_{n \le x} \frac{\Lambda(n)}{n} = \log x + O(1) \quad \textit{and} \quad \sum_{p \le x} \frac{\log p}{p} = \log x + O(1).$$

(b) There is a constant $c \in \mathbb{R}$ such that

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

(c) For $x \ge 2$, we have that

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

Proof. (a) Partial summation implies that

$$\sum_{n \le x} \log n = x \log x + O(x) \quad (x \ge 1).$$

Using the convolution identity (1.1.4), we have that the left hand side also equals

$$\sum_{n\le x}\log n=\sum_{ab\le x}\Lambda(a)=\sum_{a\le x}\Lambda(a)\cdot\left\lfloor\frac{x}{a}\right\rfloor=\sum_{a\le x}\Lambda(a)\cdot\left(\frac{x}{a}+O(1)\right)$$
$$=x\sum_{a\le x}\frac{\Lambda(a)}{a}+O\left(\sum_{a\le x}\Lambda(a)\right).$$

The error term is $O(x)$ by Chebyshev's estimate $\pi(x)\ll x/\log x$. Therefore

$$\sum_{a\le x}\frac{\Lambda(a)}{a}=\frac{1}{x}\sum_{n\le x}\log n+O(1)=\frac{x\log x+O(x)}{x}+O(1)=\log x+O(1),$$

as claimed. Finally, it is easy to see that

$$\sum_{n\le x}\frac{\Lambda(n)}{n}-\sum_{p\le x}\frac{\log p}{p}\le\sum_{p,k\ge 2}\frac{\log p}{p^k}\ll 1,$$

whence the second relation follows too.

(b) This follows by part (a) and partial summation. More precisely, we apply Theorem 1.2.1 with $a=2^-$, $b=\infty$, $a_n=\mathbf{1}_{n\text{ prime}}(\log n)/n$, $f(n)=1/\log n$, $M(x)=\log x$ and

$$R(x)=\sum_{p\le x}\frac{\log p}{p}-\log x\ll 1.$$

Then

$$\sum_{p\le x}\frac{1}{p}=\int_2^x\frac{\mathrm{d}t}{t\log t}+\frac{R(t)}{\log t}\Big|_{t=2^-}^{x}+\int_2^x\frac{R(t)}{t\log^2 t}\mathrm{d}t$$
$$=\log\log x-\log\log 2-1+O\left(\frac{1}{\log x}\right)+\int_2^x\frac{R(t)}{t\log^2 t}\mathrm{d}t.$$

We note that the integral $\int_2^\infty\frac{R(t)}{t\log^2 t}\mathrm{d}t$ converges absolutely by the estimate $R(t)\ll 1$. So, if we set $c=-\log\log 2-1+\int_2^\infty\frac{R(t)}{t\log^2 t}\mathrm{d}t$, then

$$\sum_{p\le x}\frac{1}{p}=\log\log x+c+O\left(\frac{1}{\log x}+\int_x^\infty\frac{\mathrm{d}t}{t\log^2 t}\right)=\log\log x+c+O\left(\frac{1}{\log x}\right).$$

(c) We have that

$$\epsilon_p:=\log\left(1-\frac{1}{p}\right)+\frac{1}{p}\ll\frac{1}{p^2}$$

by the Taylor expansion of the logarithm about 1. Therefore

$$\log\prod_{p\le x}\left(1-\frac{1}{p}\right)=\sum_{p\le x}\log\left(1-\frac{1}{p}\right)=-\sum_{p\le x}\frac{1}{p}+\sum_{p\le x}\epsilon_p$$
$$=-\log\log x+c+\sum_p\epsilon_p+O\left(\frac{1}{\log x}+\sum_{p>x}|\epsilon_p|\right)$$
$$=-\log\log x+c'+O\left(\frac{1}{\log x}\right),$$

where $c' := c + \sum_p \epsilon_p$. It remains to show that $c' = -\gamma$. This is proven by analyzing the behaviour of the Riemann zeta function about 1. The proof is illustrated in the following exercise. □

Exercises

Exercise 2.1.1. Prove that $|\psi(x) - \theta(x)| \ll \sqrt{x}$ for all $x \geq 1$.

Exercise 2.1.2. Complete the proof of Theorem 2.1.2(c) as folllows:

(a) For $\epsilon \in (0, 1]$ and $x \geq 2$, show that

$$\sum_{p \leq x} \log\left(1 - \frac{1}{p}\right) = \sum_{p \leq x} \log\left(1 - \frac{1}{p^{1+\epsilon/\log x}}\right) + O(\epsilon).$$

(b) For $s \in (1, 2]$, show that

$$\sum_{p} \log\left(1 - \frac{1}{p^s}\right) = \log(s-1) + O(|s-1|).$$

(c) For $\epsilon \in (0, 1]$ and $x \geq 2$, show that

$$\sum_{p > x} \log\left(1 - \frac{1}{p^{1+\epsilon/\log x}}\right) = -\int_{\epsilon}^{\infty} \frac{e^{-u}}{u} \mathrm{d}u + O\left(\frac{1}{\log x}\right).$$

[*Hint:* Theorem 2.1.2(a).] Conclude that

$$\sum_{p \leq x} \log\left(1 - \frac{1}{p}\right) = -\log\log x + \int_0^{\infty} \frac{e^{-u} - \mathbf{1}_{[0,1]}(u)}{u} \mathrm{d}u + O\left(\epsilon + \frac{1}{\log x}\right),$$

where $\mathbf{1}_A$ denotes the characteristic function of the set A.

(d) Note that

$$\gamma = \lim_{N \to \infty} \left(-\log N + \int_0^1 (1 + x + \cdots + x^{N-1}) \mathrm{d}x\right).$$

Make the change of variables $x = 1 - u/N$ and deduce that $\gamma = \int_0^{\infty} \frac{\mathbf{1}_{[0,1]}(u) - e^{-u}}{u} \mathrm{d}u$ to complete the proof of Theorem 2.1.2(c).

Exercise 2.1.3. For $n \geq 3$, prove the following inequalities:

(a) $\varphi(n) \gg n/\log\log n$ [*Hint:* Exercise 1.1.2];

(b) $\omega(n) \ll \log n/\log\log n$;

(c) $\tau_k(n) \leq n^{c_k/\log\log n}$, where c_k is a positive constant depending at most on k.

Moreover, prove that these inequalities are sharp, up to the implied constants.

2.2 Applications of elementary prime number estimates

In this section, we present two important applications of Chebyshev's and Mertens's estimates. Our first application is the following classical result:

Theorem 2.2.1 (Hardy-Ramanujan)**.** *There exists constants A and B such that*

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

uniformly for $x \ge 2$ and $r \in \mathbb{N}$.

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

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

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

$$\sum_{a \ge 1} \sum_{p \le x^{1/(a+1)}} \frac{1}{p^a \log(x/p^a)} \le \frac{\log\log x + O(1)}{\log x}.$$

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. □

The above theorem suggests that the distribution of $\omega(n)$ (or rather of $\omega(n)-1$) among integers $n \le x$ is approximately Poisson of parameter $\log\log x + O(1)$. A result towards this direction was proven by Landau, and the reader is invited to reprove it in Exercise 2.2.1. Sathe extended Landau's asymptotic in [21, 22] (but the right hand side in Exercise 2.2.1 needs to be multiplied by an appropriate function that is $\asymp 1$), and Selberg [23] produced a significantly simpler and more versatile proof.

The Hardy-Ramanujan inequality has an important consequence about the statistical properties of integers: it allows us to prove that $\omega(n) \sim \log\log x$ for almost all integers $n \le x$, which is the context of Corollary 2.2.2 below. This result can be considered as the birth of Probabilistic Number Theory, where the focus is often on properties of *typical integers*, that is to say properties that hold for almost all $n \le x$ as $x \to \infty$.

Corollary 2.2.2. *Fix some $\epsilon > 0$. Then we have that*

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

Proof. By Theorem 2.2.2, we have that

$$\#\{n \le x : \omega(n) > (1+\epsilon)\log\log x\} \le \sum_{r > (1+\epsilon)\log\log x} \frac{Ax(\log\log x + B)^{r-1}}{(r-1)!}.$$

The summands decay exponentially in this range, with ratio $< 1/(1+\epsilon)$. Thus

$$\#\{n \le x : \omega(n) > (1+\epsilon)\log\log x\} \ll_\epsilon \frac{x(\log\log x + B)^{\lfloor (1+\epsilon)\log\log x\rfloor}}{(\lfloor (1+\epsilon)\log\log x\rfloor)!} = o_{x\to\infty}(x)$$

by Stirling's formula. Similarly, we have that

$$\begin{aligned}\#\{n \le x : \omega(n) < (1-\epsilon)\log\log x\} &\le \sum_{r < (1-\epsilon)\log\log x} \frac{Ax(\log\log x + B)^{r-1}}{(r-1)!} \\ &\ll_\epsilon \frac{x(\log\log x + B)^{\lfloor (1-\epsilon)\log\log x\rfloor}}{(\lfloor (1-\epsilon)\log\log x\rfloor)!} \\ &= o_{x\to\infty}(x)\end{aligned}$$

since the summands increase exponentially for $r < (1-\epsilon)\log\log x$. □

The last application of Chebyshev's estimates concerns bounds for partial sums of non-negative multiplicative functions. The main input for proving such bounds is the following very useful theorem. If $f(p) \ll 1$ for all primes p, then Chebyshev's estimate implies that first inequality in (2.2.1) is indeed satisfied, whereas the second inequality is typically routine to check.

Theorem 2.2.3. *Let $f : \mathbb{N} \to [0, +\infty)$ be a multiplicative function such that*

$$\sum_{p \le x} f(p)\log p \le Ax \quad (x \ge 1), \quad \text{and} \quad \sum_{\substack{p \text{ prime} \\ k \ge 2}} \frac{f(p^k)\log(p^k)}{p^k} \le B. \tag{2.2.1}$$

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

$$\frac{1}{x}\sum_{n\leq x} f(n) \leq \frac{A+B+1}{\log x}\sum_{n\leq x}\frac{f(n)}{n} \ll_{A,B} \exp\left\{\sum_{p\leq x}\frac{f(p)-1}{p}\right\}.$$

Remark 2.2.1. Taking $f(n) = \tau_k(n)$, we see that the above theorem is best possible in this generality, up to multiplicative constants (cf. Theorem 1.3.3 and Exercise 1.3.4).

Proof. Note that

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

When $k = 1$, we have that

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

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

$$\begin{aligned}\sum_{k\geq 2}\sum_{mp^k\leq x} f(m)f(p^k)\log(p^k) &\leq \sum_{k\geq 2}\sum_{mp^k\leq x} f(m)f(p^k)\log(p^k)\cdot\frac{x}{mp^k}\\ &\leq x\sum_{k\geq 2}\sum_{m\leq x}\frac{f(m)f(p^k)\log(p^k)}{mp^k}\\ &= \left(\sum_{m\leq x}\frac{f(m)}{m}\right)\left(\sum_{k\geq 2,\, p\,\text{prime}}\frac{f(p^k)\log(p^k)}{p^k}\right) \leq Bx\sum_{m\leq x}\frac{f(m)}{m}.\end{aligned}$$

So we conclude that

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

To see the second inequality, note that

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

by the inequality $1+u \le e^u$ for $u \in \mathbb{R}$. Since

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

and

$$\log x \asymp \exp\left\{ \sum_{p \le x} \frac{1}{p} \right\}$$

by Theorem 2.1.2(c), the desired result follows. □

Theorem 2.2.3 is often sharp. Indeed, if f is a non-negative multiplicative function, then

$$\begin{aligned} \sum_{n \le x} f(n) &= \sum_{\substack{ab \le x \\ p|a \Rightarrow p \le x^\epsilon \\ p|b \Rightarrow p > x^\epsilon}} f(a)f(b) \\ &\ge \sum_{a \le x^\epsilon} f(a) \sum_{\substack{b \le x/a \\ p|a \Rightarrow p > x^\epsilon}} f(b). \end{aligned} \tag{2.2.2}$$

If we assume that $f(p) \ge c > 0$ for all prime p (or in a certain average sense), then

$$\sum_{\substack{b \le x/a \\ p|b \Rightarrow p > x^\epsilon}} f(b) \ge \sum_{x^\epsilon < p \le x/a} f(p) \gg_c \frac{x}{a \log x}$$

by Chebyshev's estimate, provided that $\epsilon < 1/2$ so that $x/a \ge x^{1-\epsilon} > x^\epsilon$. Thus

$$\sum_{n \le x} f(n) \gg_c \frac{x}{\log x} \sum_{n \le x^\epsilon} \frac{f(n)}{n}.$$

A matching lower bound on the partial sums of f is then derived by the following result:

Theorem 2.2.4. *Let $f : \mathbb{N} \to [0, +\infty)$ be a multiplicative function such that $0 \le f(p) \le A$ for all primes p. For $x \ge 2$, we have*

$$\sum_{n \le x} \frac{f(n)}{n} \gg_A \exp\left\{ \sum_{p \le x} \frac{f(p)}{p} \right\} ;$$

Proof. Without loss of generality, we may assume that $A \in \mathbb{N}$. We then define a multiplicative function g via the formula $g * \mu^2 f = A^\Omega$. We claim that $0 \le g(n) \le A^{\Omega(n)}$. By multiplicativity, it suffices to this when $n = p^m$ for some prime p and some $m \ge 1$. We argue by induction on m. We take $m = 0$ as our base case, which clearly holds. Now, assume that $0 \le g(p^m) \le A^m$ for some $m \ge 0$ and note that

$$A^{m+1} = (g * \mu^2 f)(p^{m+1}) = g(p^{m+1}) + g(p^m)f(p),$$

so that

$$g(p^{m+1}) = A^{m+1} - g(p^m)f(p).$$

Since $0 \le f(p) \le A$ and $0 \le g(p^m) \le A^m$, we deduce that $0 \le g(p^{m+1}) \le A^{m+1}$, which completes the induction.

Next, since $\mu^2 f * g = A^\Omega$ and $g, f \ge 0$, we find that

$$\left(\sum_{n\le x} \frac{f(n)}{n}\right)\left(\sum_{p|n \,\Rightarrow\, A<p\le x} \frac{g(n)}{n}\right) \ge \sum_{\substack{n\le x\\ p|n\Rightarrow p>A}} \frac{A^{\Omega(n)}}{n} \ge \sum_{\substack{n\le x\\ p|n\Rightarrow p>A}} \frac{\tau_A(n)}{n},$$

as well as that

$$\left(\sum_{\substack{n\le x\\ p|n\Rightarrow p\le A}} \frac{\tau_A(n)}{n}\right)\left(\sum_{\substack{n\le x\\ p|n\Rightarrow p>A}} \frac{\tau_A(n)}{n}\right) \ge \sum_{n\le x} \frac{\tau_A(n)}{n} \asymp_A (\log x)^A$$

by Exercise 1.3.4. These estimates together yield the inequality

$$\sum_{n\le x} \frac{f(n)}{n} \gg_A \left(\sum_{p|n\,\Rightarrow\, A<p\le x} \frac{g(n)}{n}\right)^{-1} (\log x)^A.$$

Finally, the fact that $0 \le g \le A^\Omega$ and the inequality $1 + u \le e^u$ imply that

$$\sum_{p|n\,\Rightarrow\, A<p\le x} \frac{g(n)}{n} \le \prod_{A<p\le x}\left(1 + \frac{g(p)}{p} + \frac{A^2}{p^2} + \frac{A^3}{p^3} + \cdots\right) \ll_A \exp\left\{\sum_{p\le x} \frac{g(p)}{p}\right\}.$$

Since $g(p) = A - f(p)$, by the definition of g, the claimed lower bound on $\sum_{n\le x} f(n)/n$ follows. □

Exercises

Exercise 2.2.1. Using the Prime Number Theorem in the form $\pi(x) \sim x/\log x$, prove that for each fixed $r \ge 1$ we have

$$\pi_r(x) \sim \frac{x(\log\log x)^{r-1}}{(r-1)!} \quad (x \to \infty).$$

Exercise 2.2.2. Given $\lambda > 0$, we set

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

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

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

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

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

(b) (Rankin's trick) Observe that

$$\#\{n \leq x : \omega(n) \geq \lambda \log\log x\} \leq \sum_{n\leq x} v^{\omega(n)-\lambda\log\log x}$$

for any $v \geq 1$, as well as that

$$\#\{n \leq x : \omega(n) \leq \lambda \log\log x\} \leq \sum_{n\leq x} w^{\omega(n)-\lambda\log\log x}$$

for any $0 < w \leq 1$. Estimate the above sums using Theorem 2.2.3 and optimize the choice of v and w. Compare the bounds you obtain with the results of part (a).

Exercise 2.2.3. (a) Show that the entries of the $N \times N$ multiplication table form a sparse set of the integers in the sense that

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

(b) Show that

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

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

Exercise 2.2.4. For fixed $r \in \mathbb{Z}$ and $k \in \mathbb{N}$, show that

$$\sum_{n\leq x} \tau_k(n)\left(\frac{\varphi(n)}{n}\right)^r \asymp_{r,k} x(\log x)^{k-1} \quad (x \geq 2).$$

Exercise 2.2.5. For each fixed $\epsilon > 0$, prove that

$$\#\{n \leq x : p|n \;\Rightarrow\; p \leq x^\epsilon\} \asymp_\epsilon x.$$

Exercise 2.2.6. (a) If $z \geq 1$ and $n \in \mathbb{N}$ with $z \geq \sqrt{\log n}$, then show that

$$\frac{n}{\varphi(n)} \asymp \prod_{\substack{p|n\\ p\leq z}}\left(1+\frac{1}{p}\right) = \sum_{\substack{d|n\\ p|d \,\Rightarrow\, p\leq z}} \frac{\mu^2(d)}{d}.$$

(b) Prove that

$$\sum_{x<n\leq x+y} \frac{n}{\varphi(n)} \asymp_\epsilon y \quad (x^\epsilon \leq y \leq x).$$

2.3 Primes and the Möbius function

In this section, we prove the equivalence between (2.0.1) and the Prime Number Theorem. For the convenience of notation, we set

$$M(x) := \sum_{n\le x} \mu(n).$$

We need a preliminary lemma:

Lemma 2.3.1. *For all $x \ge 1$, we have*

$$\left|\sum_{n\le x} \frac{\mu(n)}{n}\right| \le 1.$$

Proof. For $x < 4$, we verify this by direct computation. Assume that $x \ge 4$. We have

$$1 = \sum_{n\le x}(\mu * 1)(n) = \sum_{ab\le x} \mu(a) = \sum_{a\le x} \mu(a) \left\lfloor \frac{x}{a} \right\rfloor = x \sum_{a\le x} \frac{\mu(a)}{a} - \sum_{a\le x} \mu(a) \left\{ \frac{x}{a} \right\}.$$

So

$$\left|\sum_{a\le x} \frac{\mu(a)}{a}\right| \le \frac{1}{x} + \frac{1}{x}\left|\sum_{a\le x} \mu(a) \left\{ \frac{x}{a} \right\}\right| \le \frac{1}{x} + \frac{1}{x}\sum_{a\le x} \mu^2(a).$$

Since $x \ge 4$ and 4 is not square-free, the right hand side is $\le 1/x + (x-1)/x = 1$, which proves our claim. □

Theorem 2.3.2. *The following relations are equivalent:*

(a) $\psi(x) \sim x$ as $x \to \infty$;

(b) $M(x) = o(x)$ as $x \to \infty$.

Proof. $(a) \Rightarrow (b)$. It suffices to show that $\sum_{n\le x} \mu(n)\log n = o(x \log x)$. Indeed, partial summation implies that

$$M(x) = 1 + \int_{2^-}^{x} \frac{1}{\log t} \mathrm{d} \sum_{n\le t} \mu(n)\log n = 1 + \frac{\sum_{n\le x} \mu(n)\log n}{x\log x} + \int_2^x \sum_{n\le t} \mu(n)\log n \frac{\mathrm{d}t}{t\log^2 t}.$$

The integral on the right hand side is trivially

$$\ll \int_2^x \frac{t\log t}{t\log^2 t}\mathrm{d}t = \mathrm{li}(x) \sim \frac{x}{\log x} = o(x) \quad (x \to \infty).$$

So, if we know that $\sum_{n\le x} \mu(n)\log n = o_{x\to\infty}(x\log x)$, then we deduce that $M(x) = o_{x\to\infty}(x)$.

To prove that $\sum_{n\le x} \mu(n)\log n = o_{x\to\infty}(x\log x)$, we use the formula $\Lambda = -\mu \log *1$ to find that $-\mu\log = \Lambda * \mu$. In particular,

$$-\sum_{n\le x} \mu(n)\log n = \sum_{ab\le x} \Lambda(a)\mu(b). \tag{2.3.1}$$

The advantage of this formula is that it "does not lose any logarithms", that is to say if we trivially estimate the right hand side we recover the trivial bound $\sum_{n\le x}\mu(n)\log n \ll x\log x$. Moreover, though we will not need it, it is important to remark that (2.3.1) (together with partial summation, as in the first paragraph) expresses M as an average of its past, a typical feature of averages of multiplicative functions. Indeed, since $\Lambda(1)=0$, we find that

$$-\sum_{n\le x}\mu(n)\log n=\sum_{2<a\le x}\Lambda(a)M(x/a).$$

This is to be expected since if we know what value a multiplicative function f assumes at m and p with $(p,m)=1$, then we also know $f(pm)$. Relation (2.3.1) is an efficient way of expressing this phenomenon when $f=\mu$, and similar identities can be derived for general multiplicative functions and are abundant in their study.

Let us now see how to use (2.3.1) and our assumption that $\psi(x)\sim x$ to deduce that $S(x;\mu\log)=o_{x\to\infty}(x\log x)$. Fix $\epsilon>0$. We know that there is some $A\ge 1$ such that

$$|\psi(y)-y|\le \epsilon y \quad (y\ge A).$$

Therefore

$$\begin{aligned}\sum_{n\le x}\mu(n)\log n&=\left(\sum_{\substack{ab\le x\\ b\le x/A}}+\sum_{\substack{ab\le x\\ b>x/A}}\right)\Lambda(a)\mu(b)\\&=\sum_{b\le x/A}\mu(b)\psi(x/b)+\sum_{a\le A}\Lambda(a)\sum_{x/A<b\le x/a}\mu(b).\end{aligned}$$

Since $x/b\ge A$ for $b\le x/A$ and Lemma 2.3.1 implies that $|\sum_{b\le x/A}\mu(b)/b|\le 1$, we deduce that

$$\begin{aligned}\left|\sum_{n\le x}\mu(n)\log n\right|&\le\left|\sum_{b\le x/A}\frac{x\mu(b)}{b}\right|+\sum_{b\le x/A}\frac{\epsilon x}{b}+\sum_{a\le A}\Lambda(a)\frac{x}{a}\\&\le x+\epsilon x\left(1+\int_1^{x/A}\frac{\mathrm{d}t}{t}\right)+\sum_{a\le A}\Lambda(a)\frac{x}{a}\\&\le \epsilon x\log x+x\left(1-\log A+\sum_{a\le A}\frac{\Lambda(a)}{a}\right)\le 2\epsilon x\log x\end{aligned}$$

if x is large enough in terms of A and ϵ. This completes the proof that $S(x;\mu\log)=o_{x\to\infty}(x\log x)$ and hence that $M(x)=o_{x\to\infty}(x)$.

$(b)\Rightarrow(a)$. We need to show that $\Lambda-1$ is 0 on average. Note that

$$\Lambda-1=\mu*(\log-\tau)=\mu*(\log+c-\tau)-c\cdot\mathbb{1}$$

for any constant $c\in\mathbb{R}$. We set $f(n)=\log n+2\gamma-\tau(n)$, so that

$$\sum_{n\le x}f(n)=x\log x+(2\gamma-1)x+O(\log x)-\sum_{n\le x}\tau(n)\ll\sqrt{x}.$$

So

$$\begin{aligned}\psi(x) - x &= \sum_{n\le x}(\Lambda(n) - 1) + O(1)\\ &= \sum_{ab\le x}\mu(a)f(b) + O(1)\\ &= \sum_{a\le x/B}\mu(a)\sum_{b\le x/a}f(b) + \sum_{b\le B}f(b)\sum_{x/B<a\le x/b}\mu(a) + O(1)\\ &=: S_1 + S_2 + O(1),\end{aligned}$$

say. For S_1, we have

$$S_1 \ll \sum_{a\le x/B}\sqrt{\frac{x}{a}} \ll \sqrt{x}\cdot\sqrt{\frac{x}{B}} = \frac{x}{\sqrt{B}}.$$

Fix $\epsilon > 0$. By taking B is large enough in terms of ϵ, we have $|S_1| \le \epsilon x/2$. Now, let $\delta > 0$ be an auxiliary parameter. We know that if y is large enough in terms of δ, then

$$|M(y)| \le \delta y.$$

So, if x is large enough, then

$$\sum_{x/B<a\le x/b}\mu(a) \le |M(x/b)| + |M(x/B)| \le \frac{2\delta x}{b}.$$

Consequently,

$$|S_2| \le 2\delta x\sum_{b\le B}\frac{|f(b)|}{b}.$$

Taking δ small enough in terms of ϵ and B, we find that $|S_2| \le \epsilon x/2$. We conclude that for x large enough in terms of ϵ alone, we have that $|\psi(x) - x| \le \epsilon x + O(1)$, which completes the proof of (a). □

Finally, another equivalence can be given by the following result, which shows that Lemma 2.3.1 can be sharpened for large x. In particular, the Prime Number Theorem implies that the series $\sum_{n\ge1}\mu(n)/n$ converges and

$$\sum_{n=1}^{\infty}\frac{\mu(n)}{n} = 0.$$

Theorem 2.3.3. *The following relations are equivalent:*

(a) $M(x) = o(x)$ *as* $x \to \infty$*;*

(b) $\sum_{n\le x}\mu(n)/n = o_{x\to\infty}(1)$.

Proof. $(a) \Rightarrow (b)$. As in the proof of Lemma 2.3.1, we have that

$$1 = \sum_{ab \le x} \mu(a). \tag{2.3.2}$$

For any $A \in [1, x]$, Dirichlet's hyperbola method implies that

$$\begin{aligned} 1 &= \sum_{a \le x/A} \mu(a) \left\lfloor \frac{x}{a} \right\rfloor + \sum_{b \le A} \sum_{x/A < a \le x/b} \mu(a) \\ &= x \sum_{a \le x/A} \frac{\mu(a)}{a} - \sum_{a \le x/A} \mu(a) \left\{ \frac{x}{a} \right\} + \sum_{b \le A} (M(x/b) - M(x/A)). \end{aligned}$$

Fix $\epsilon > 0$. Our assumption that $M(x) = o(x)$ implies that there is some $x_0 \ge 1$ such that $|M(x)| < \epsilon x$ for $x \ge x_0$. Therefore, if $x \ge A x_0$, then

$$\begin{aligned} \left| \sum_{a \le x/A} \frac{\mu(a)}{a} \right| &\le \frac{1}{x} + \frac{1}{x} \left| \sum_{a \le x/A} \mu(a) \left\{ \frac{x}{a} \right\} \right| + \frac{1}{x} \sum_{b \le A} (|M(x/b)| + |M(x/A)|) \\ &\le \frac{1}{x} + \frac{1}{A} + \frac{1}{x} \sum_{b \le A} \left(\frac{\epsilon x}{b} + \frac{\epsilon x}{A} \right) \\ &\le \frac{2}{A} + \epsilon \left(1 + \sum_{b \le A} \frac{1}{b} \right) \\ &\le \frac{2}{A} + \epsilon (2 + \log A), \end{aligned}$$

since $\sum_{1 < b \le A} 1/b \le \int_1^A \mathrm{d}t/t = \log A$. We choose $A = 2/\epsilon$ to optimize the above inequality, so that

$$\left| \sum_{a \le x/A} \frac{\mu(a)}{a} \right| \le 4\epsilon + \epsilon \log \frac{2}{\epsilon} \quad (x \ge A x_0).$$

Replacing x by Ax completes the proof of the estimate $\sum_{n \le x} \mu(n)/n = o_{x \to \infty}(1)$.

$(b) \Rightarrow (a)$. This follows directly by partial summation: if $L(x) = \sum_{n \le x} \mu(n)/n$, then

$$M(x) = \sum_{n \le x} \mu(n) = \int_{1^-}^{x} t \, \mathrm{d}L(t) = x \cdot L(x) - \int_1^x L(t) \mathrm{d}t.$$

Now, fix $\epsilon > 0$. There is some x_0 such that $|L(x)| \le \epsilon$ for $x \ge x_0$. Therefore, if $x \ge x_0$, then the above relations and Lemma 2.3.1 imply that

$$|M(x)| \le \epsilon x + \int_1^{x_0} \mathrm{d}t + \int_{x_0}^{x} \epsilon \mathrm{d}t \le x_0 + 2\epsilon x,$$

which implies that $M(x) = o_{x \to \infty}(x)$. □

Exercises

Exercise 2.3.1. Show that the Prime Number Theorem implies the estimate

$$\#\{n \le x : n \text{ square-free}\} = \frac{6}{\pi^2}x + o(\sqrt{x}) \quad (x \to \infty).$$

Exercise 2.3.2. (a) Show that

$$\sum_{n \le x} \log(x/n) \asymp x \quad (x \ge 2).$$

(b) Prove that $\mu \log = -\mu * \Lambda$ and deduce that

$$M(x) \log x = -\sum_{d \le x} \Lambda(d) M(x/d) + O(x).$$

(c) Show that

$$\liminf_{x \to \infty} \frac{M(x)}{x} + \limsup_{x \to \infty} \frac{M(x)}{x} = 0.$$

Chapter 3

Sieve methods

3.1 Inclusion-exclusion

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 3.1.1. *If $n > 1$ is composite, then it has a prime divisor $p \leq \sqrt{n}$.*

Eratosthenes's algorithm for founding all primes up to x then goes as follows:

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

(b) Delete 1 from the list.

(c) 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.

(d) Delete all multiples of n.

(e) Go to step (3).

The termination of this algorithm is guaranteed by Theorem 3.1.1. 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 potential way of getting our hands on how big $\pi(x)$ is: if an integer $n \leq x$ is not divisible by any prime $\leq \sqrt{x}$, then either $n = 1$ or n is a prime number lying in $(\sqrt{x}, x]$. So

$$\begin{aligned} \pi(x) &= \#\{n \leq x : p|n \Rightarrow p > \sqrt{x}\} - 1 + \pi(\sqrt{x}) \\ &= \#\{n \leq x : p|n \Rightarrow p > \sqrt{x}\} + O(\sqrt{x}). \end{aligned} \tag{3.1.1}$$

Our attention thus turns to estimating the cardinality of the set appearing on the right hand side of (3.1.1). Let $\{p_1, p_2, \ldots, p_r\}$ be an indexing of the primes $\leq \sqrt{x}$. By the inclusion-

exclusion principle, we have

$$\begin{aligned}\#\{n\le x : p|n \Rightarrow p>\sqrt{x}\} &= \#\left(\bigcap_{i=1}^{r}\{n\le x : p_i\nmid n\}\right)\\ &= \#\{n\le x\} - \#\left(\bigcup_{i=1}^{r}\{n\le x : p_i|n\}\right)\\ &= \#\{n\le x\} - \sum_{i=1}^{r}\#\{n\le x : p_i|n\} + \sum_{1\le i<j\le r}\#\{n\le x : p_ip_j|n\}\\ &\quad - \sum_{1\le i<j<k\le r}\#\{n\le x : p_ip_jp_k|n\}\pm\cdots\\ &= \lfloor x\rfloor - \sum_{i=1}^{r}\left\lfloor\frac{x}{p_i}\right\rfloor + \sum_{1\le i<j\le r}\left\lfloor\frac{x}{p_ip_j}\right\rfloor - \sum_{1\le i<j\le r}\left\lfloor\frac{x}{p_ip_jp_k}\right\rfloor\pm\cdots \qquad (3.1.2)\end{aligned}$$

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

$$\begin{aligned}\pi(x) &\approx \#\{n\le x : p|n\Rightarrow p>\sqrt{x}\}\\ &\approx x - \sum_{i=1}^{r}\frac{x}{p_i} + \sum_{1\le i<j\le r}\frac{x}{p_ip_j} - \sum_{1\le i<j<k\le r}\frac{x}{p_ip_jp_k}\pm\cdots \qquad (3.1.3)\\ &= x\prod_{i=1}^{r}\left(1-\frac{1}{p_i}\right) = x\prod_{p\le\sqrt{x}}\left(1-\frac{1}{p}\right)\sim\frac{2e^{-\gamma}x}{\log x},\end{aligned}$$

by Mertens's estimate (cf. Theorem 2.1.2). However, as we discussed in Chapter 2 and we will show in Chapter 4.4, the correct asymptotic formula when $x\to\infty$ is $\pi(x)\sim x/\log x$. Since $2e^{-\gamma}=1.1229189671\ldots>1$, relation (3.1.3) overestimates $\pi(x)$, a typical feature of sieve methods, as we will see.

In order to understand why the above argument fails, it is convenient to recast formula (3.1.2) in a more compact way, using the Möbius function μ. If $m=p_1\cdots p_r=\prod_{p\le\sqrt{x}}p$, then

$$\mathbf{1}_{(n,m)=1} = \sum_{d|(n,m)}\mu(d), \qquad (3.1.4)$$

so that

$$\#\{n\le x : p|n\Rightarrow p>\sqrt{x}\} = \sum_{n\le x}\sum_{d|(n,m)}\mu(d) = \sum_{d|m}\mu(d)\sum_{\substack{n\le x\\ d|n}}1 = \sum_{d|m}\mu(d)\left\lfloor\frac{x}{d}\right\rfloor. \qquad (3.1.5)$$

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

$$\begin{aligned}\#\{n\le x : p|n\Rightarrow p>\sqrt{x}\} &= x\sum_{p|d\Rightarrow p\le\sqrt{x}}\frac{\mu(d)}{d} + O(2^{\pi(\sqrt{x})})\\ &= \frac{(2e^{-\gamma}+o(1))x}{\log x} + O(2^{\pi(\sqrt{x})}). \qquad (3.1.6)\end{aligned}$$

Note that the error term is of size $e^{c\sqrt{x}}$ by Chebyshev's estimate, an enormous function compared to our main term. Thus, our attempt to obtain an asymptotic formula for $\pi(x)$ fails dramatically. This happens for two interconnected reasons:

- The numbers d in the sum appearing on the right hand side (3.1.5) 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 (3.1.5) can get as big as

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

  where we used the Prime Number Theorem, which is enormous compared to x. Therefore the approximation $\lfloor x/d\rfloor = x/d + O(1)$ is very bad for most d in (3.1.5).

We will discuss in the next sections how these issues can be partially rectified.

3.2 Upper and lower bound sieves

Sieve methods can be thought as an attempt of truncating the identity (3.1.4). From now on we will start working in a more general context: given a set of primes

$$\mathcal{P} \subset \{p < z\},$$

our goal is to *sieve* all multiples of $\mathcal{P}$ from some set of integers $\mathcal{A}$. The parameter z will be referred to as the *level of the sifting range*.

The above discussion naturally leads us to the study of the characteristic function of integers that are co-prime to $m = \prod_{p\in\mathcal{P}} p$. To ease notation, we set

$$(n, \mathcal{P}) := \left(n, \prod_{p\in\mathcal{P}} p\right).$$

Brun, the pioneer of modern sieve methods, used the fact that

$$\sum_{\substack{d|(n,\mathcal{P})\\ \omega(d)=2j-1}} \mu(d) \le \sum_{d|(n,\mathcal{P})} \mu(d) \le \sum_{\substack{d|(n,\mathcal{P})\\ \omega(d)=2j}} \tag{3.2.1}$$

for any $j \in \mathbb{N}$, a simple consequence of inclusion-exclusion. This inequality restricts the divisors d on the left and right hand side of (3.2.1) to be $\le z^{2j}$, where z is the level of the sifting range.

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

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

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

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

is called a *lower bound sieve of level D*. The reason for this terminology is that, applying relations (S^+) and (S^-) with $(n, \mathcal{P})$ in place of n, we deduce that

$$\sum_{d|(n,\mathcal{P})} \mu^-(d) \le \mathbf{1}_{(n,\mathcal{P})=1} \le \sum_{d|(n,\mathcal{P})} \mu^+(d)$$

for any set of primes $\mathcal{P}$.

Sieve weights $\mu^\pm(d)$ that are produced in a combinatorial way as in (3.2.1) fall under the general category of *combinatorial sieves*. It is possible to produce many such sieve weights using *Buchstab's identity*

$$\mathbf{1}_{P^-(n)\ge z} = 1 - \sum_{p<z} \mathbf{1}_{P^-(n)=p} = 1 - \sum_{p<z} \mathbf{1}_{p|n,\, P^-(n/p)\ge p}, \tag{3.2.2}$$

where

$$P^-(n) := \min\{p|n\},$$

with the convention that $P^-(1) = \infty$. Here z is some parameter corresponding to the level of the sifting range of the particular problem we are dealing with.

In order to construct our weights, we need some notation. Given a family of sets

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

nested as

$$\Pi_j \subset \Pi_{j-2} \times \{p \text{ prime}\}^2 \quad (j \ge 3),$$

we set

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

and

$$\mu^\pm(d) := \mathbf{1}_{d\in\mathcal{D}^\pm} \cdot \mu(d). \tag{3.2.3}$$

We will show by successive application of Buchstab's inequality that these are indeed sieve weights. In fact, we will prove that

$$\sum_{\substack{d|n \\ p|d \,\Rightarrow\, p<z}} \mu^-(d) \le \mathbf{1}_{P^-(n)\ge z} \le \sum_{\substack{d|n \\ p|d \,\Rightarrow\, p<z}} \mu^+(d), \tag{3.2.4}$$

for all $z \ge 1$, which is sufficient. We first prove the right inequality: we have

$$\mathbf{1}_{P^-(n)\ge z} = 1 - \sum_{p_1<z} \mathbf{1}_{p_1|n,\, P^-(n/p_1)\ge p_1} \le 1 - \sum_{p_1\in\Pi_1} \mathbf{1}_{p_1|n,\, P^-(n/p_1)\ge p_1},$$

by dropping some terms by positivity. We re-apply Buchstab's identity to find that

$$\begin{aligned}\mathbf{1}_{P^-(n)\geq z} &\leq 1-\sum_{p_1\in\Pi_1}\left(\mathbf{1}_{p_1|n}-\sum_{p_2<p_1}\mathbf{1}_{p_1|n,\,P^-(n/p_1)=p_2}\right)\\ &= 1-\sum_{p_1\in\Pi_1}\mathbf{1}_{p_1|n}+\sum_{\substack{p_2<p_1<z\\ p_1\in\Pi_1}}\mathbf{1}_{p_1p_2|n,\,P^-(n/p_1p_2)\geq p_2}.\end{aligned}$$

We may not drop any terms now since the new summands have a positive weight, so we re-apply Buchstab's identity to the function $\mathbf{1}_{P^-(n/p_1)\geq p_1}$ to find that

$$\begin{aligned}\mathbf{1}_{P^-(n)\geq z} &\leq 1-\sum_{p_1\in\Pi_1}\mathbf{1}_{p_1|n}+\sum_{\substack{p_2<p_1<z\\ p_1\in\Pi_1}}\mathbf{1}_{p_1p_2|n}-\sum_{\substack{p_3<p_2<p_1<z\\ p_1\in\Pi_1}}\mathbf{1}_{p_1p_2p_3|n,\,P^-(n/p_1p_2p_3)\geq p_3}\\ &\leq 1-\sum_{p_1\in\Pi_1}\mathbf{1}_{p_1|n}+\sum_{\substack{p_2<p_1<z\\ p_1\in\Pi_1}}\mathbf{1}_{p_1p_2|n}-\sum_{(p_1,p_2,p_3)\in\Pi_3}\mathbf{1}_{p_1p_2p_3|n,\,P^-(n/p_1p_2p_3)\geq p_3}.\end{aligned}$$

Continuing as above, dropping summands every second step, we establish the right inequality in (3.2.4). Similarly, by dropping summands which come with a positive weight to obtain a lower bound, we establish the left inequality in (3.2.4) too, thus proving relations (S^+) and (S^-).

The above construction of combinatorial sieve weights offers a great deal of flexibility. For example, the choice $\Pi_j=\{p \text{ prime}\}^j$ for $j<r$ and $\Pi_j=\emptyset$ for $j\geq r$ corresponds to Brun's choice coming from inclusion-exclusion. Typically, we will be averaging $\mathbf{1}_{(a,\mathcal{P})=1}$ with a running over some special set of integers $\mathcal{A}$. The goal is to choose the sets Π_j tailored to our set $\mathcal{A}$ that satisfy two properties simultaneously: the sets Π_j are large enough, so that we are not throwing away too much information, while at the same time the sets $\mathcal{D}^{\pm}$ contain integers that are not too big (that is to say, the level D is not too big). It turns out that a good choice for the sets Π_j is given by

$$\begin{aligned}\Pi_j=\{(p_1,p_2,\dots,p_j) &: p_1>\cdots>p_j,\\ &p_1\cdots p_i<D/p_i^{\beta}, \text{ for all } i\in\{1,\dots,j\} \text{ with } i\equiv j \,(\mathrm{mod}\, 2)\}\end{aligned}\tag{3.2.5}$$

for some parameter β to be chosen conveniently, depending on the properties of the set $\mathcal{A}$. This leads to the so-called β sieve, pioneered by Rosser and brought to maturity by Iwaniec [13]. When we only to sieve primes $<z$, we take

$$\begin{aligned}\Pi_j=\{(p_1,p_2,\dots,p_j) &: z>p_1>\cdots>p_j,\\ &p_1\cdots p_i<D/p_i^{\beta}, \text{ for all } i\in\{1,\dots,j\} \text{ with } i\equiv j \,(\mathrm{mod}\, 2)\}\end{aligned}\tag{3.2.6}$$

The following lemma establishes that if the level of the weights D is relatively big compared to the level of the sifting range z, then the weights $\mu^{\pm}$ constructed above approximate μ very well for an appropriate choice of β, which are rather complicated. We will not go into the fine details of the optimal choice of β here. We refer the reader to [5] for more discussion.

Theorem 3.2.1 (Fundamental Lemma of Sieve Methods, I). *Let $\kappa \geq 0$ $c \geq 1$, $z \geq 1$ 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:*

(a) μ^+ and μ^- are both supported on $\{d \leq D : \mu^2(d) = 1,\ p|d \ \Rightarrow\ p < z\}$;

(b) $(\mu^- * 1)(n) \leq \mathbf{1}_{P^-(n)\geq z} \leq (\mu^+ * 1)(n)$ and $\mu^{\pm}(1) = 1$;

(c) If f is a multiplicative function satisfying $0 \leq f(p) \leq \min\{\kappa(1 + c/p), p - 1\}$ and $\lambda = \mu^{\pm}$, then

$$\sum_d \frac{\lambda(d)f(d)}{d} = \big\{1 + O_{\kappa,c}(e^{-cu})\big\} \prod_{p<z} \left(1 - \frac{f(p)}{p}\right).$$

Proof. For the most part, we follow an argument in [5]. We define $\mu^{\pm}$ as in (3.2.3) with the sets Π_j defined by (3.2.6) for some parameter $\beta = \beta(\kappa) > 1$ to be chosen later. Then property (a) is satisfied by the definition of $\mu^{\pm}$, and relation (3.2.4) implies property (b). It remains to show property (c).

Consider f as in the statement of the lemma. With a slight abuse of notation, given a sequence of primes $p_1, p_2, \ldots$, we write $z_j = (D/(p_1 \cdots p_{j-1}))^{1/(\beta+1)}$. Also, we define

$$V(y) = \sum_{p|d \ \Rightarrow p<y} \frac{\mu(d)f(d)}{d} = \prod_{p<y} \left(1 - \frac{f(p)}{p}\right) \quad (y \geq 1)$$

and

$$V_r = \sum_{\substack{z>p_1>\cdots>p_r\geq z_r \\ p_j<z_j \ (1\leq j<r,\, j\equiv r \,(\mathrm{mod}\, 2))}} \frac{f(p_1 \cdots p_r)}{p_1 \cdots p_r} \cdot V(p_r).$$

Then

$$\sum_{p|d \ \Rightarrow p<z} \frac{\mu^+(d)f(d)}{d} - V(z) = \sum_{\substack{r\geq 1 \\ r \,\mathrm{odd}}} V_r \tag{3.2.7}$$

and

$$V(z) - \sum_{p|d \ \Rightarrow p<z} \frac{\mu^-(d)f(d)}{d} = \sum_{\substack{r\geq 1 \\ r \,\mathrm{even}}} V_r. \tag{3.2.8}$$

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 $p_1, \ldots, p_r$ lying in the range of V_r. We claim that

$$p_1 \cdots p_j \leq D z^{-(u-1)\left(\frac{\beta-1}{\beta+1}\right)^{\left\lfloor \frac{j}{2} \right\rfloor}} \quad (0 \leq j \leq r,\ j \equiv r - 1 \,(\mathrm{mod}\, 2)). \tag{3.2.9}$$

We argue by induction: if $j = 0$ or $j = 1$, then relation (3.2.9) holds trivially, since $p_1 < z \leq D$. Assume now that (3.2.9) holds for some $j \in \{0, \ldots, r - 3\}$ that has opposite parity than

r. Then

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

which completes the inductive step and hence the proof of (3.2.9). Note that relation (3.2.9) and our assumption that $p_r \ge y_r$ imply that

$$p_r \ge \left(\frac{D}{p_1\cdots p_{r-1}}\right)^{\frac{1}{\beta+1}} \ge z^{\delta_r}, \tag{3.2.10}$$

where

$$\delta_r = \frac{u-1}{\beta+1}\left(\frac{\beta-1}{\beta+1}\right)^{\left\lfloor\frac{r-1}{2}\right\rfloor} \ge \frac{1}{\beta+1}\left(\frac{\beta-1}{\beta+1}\right)^{\frac{r-1}{2}} = \frac{1}{\sqrt{\beta^2-1}}\left(\frac{\beta-1}{\beta+1}\right)^{\frac{r}{2}}, \tag{3.2.11}$$

since $u \ge 2$ and Consequently, for every $\epsilon > 0$, we have that

$$\begin{aligned} V_r &\le \sum_{\substack{z>p_1>\cdots>p_r\ge z^{\delta_r}\\ p_1\cdots p_{r-1}p_r^{\beta+1}>D}} \frac{f(p_1)\cdots f(p_r)V(z^{\delta_r})}{p_1\cdots p_r} \\ &\le \frac{V(z^{\delta_r})}{e^{cu}} \sum_{z>p_1>\cdots>p_r\ge z^{\delta_r}} \frac{f(p_1)\cdots f(p_r)(p_1\cdots p_{r-1}p_r^{\beta+1})^{c/\log z}}{p_1\cdots p_r}, \end{aligned}$$

by Markov's inequality, since $D^{1/\log z} = e^u$. Since $f(p) \le \kappa(1+c/p)$, we have that

$$\begin{aligned} V(z^{\delta_r}) = V(z)\prod_{z^{\delta_r}<p\le z}\left(1-\frac{f(p)}{p}\right)^{-1} &\le V(z)\prod_{z^{\delta_r}<p\le z}\left(1-\frac{\min\{p-1,\kappa(1+c/p)\}}{p}\right)^{-1} \\ &= V(z)\prod_{z^{\delta_r}<p\le z}\left(1-\frac{1}{p}\right)^{\kappa}\left(1+O_{\kappa,c}\left(\frac{1}{p^2}\right)\right) \\ &\ll_{\kappa,c} V(z)\delta_r^{-\kappa}, \end{aligned}$$

by Mertens's theorem, since $1-\kappa/p+\kappa c/p^2 = (1-1/p)^{\kappa}(1+O(1/p^2))$ by Taylor's theorem. Thus

$$\begin{aligned} V_r &\ll_C \frac{V(z)\kappa^r\delta_r^{-\kappa}}{e^{cu}} \sum_{z^{\delta_r}<p_r\le z} \frac{1+c/p_r}{p_r^{1-c(\beta+1)/\log z}} \sum_{z>p_1>\cdots>p_{r-1}\ge z^{\delta_r}} \prod_{j=1}^{r-1}\frac{1+c/p_j}{p_j^{1-c/\log z}} \\ &\le \frac{V(z)\kappa^r\delta_r^{-\kappa}}{e^{cu}(r-1)!}\left(\sum_{z^{\delta_r}<p_r\le z}\frac{1+c/p_r}{p_r}\right)\left(\sum_{z>p\ge z^{\delta_r}}\frac{1+c/p}{p^{1-c/\log z}}\right)^{r-1}, \end{aligned}$$

where we used again the fact that $0 \le f(p) \le \kappa(1 + c/p)$ and the $(r-1)!$ factor comes from unordering the variables $p_1, \dots, p_{r-1}$. We note that

$$\sum_{z^{\delta_r} \le p \le z} \frac{1 + c/p}{p^{1-c/\log z}} = \sum_{z^{\delta_r} \le p \le z} \frac{1 + O_c(1/p + \log p/\log z)}{p} = \log \frac{1}{\delta_r} + O_c(1),$$

by Chebyshev's and Mertens's estimates. Thus

$$V_r \ll_{\kappa,\beta,c} \frac{V(z)\kappa^r \delta_r^{-\kappa}(\log(1/\delta_r) + O(1))^r}{e^{cu}(r-1)!}. \tag{3.2.12}$$

The inequality $x + y \le xe^{y/x}$, for x and y positive, and relation (3.2.11) imply that

$$\delta_r^{-\kappa}(\log(1/\delta_r) + O(1))^r \ll_{\beta,\kappa} \left\{ r \left(\frac{\beta+1}{\beta-1}\right)^{\kappa/2} \log\left(\frac{\beta+1}{\beta-1}\right) \right\}^r.$$

Together with Stirling's formula and (3.2.12), this implies that

$$V_r \ll_{\kappa,\beta,c} \frac{V(z)\sqrt{r}}{e^{cu}} \cdot \left\{ \left(\frac{\beta+1}{\beta-1}\right)^{\kappa/2} \frac{\kappa e}{2} \log\left(\frac{\beta+1}{\beta-1}\right) \right\}^r.$$

Choosing $\beta = \beta(\kappa)$ such that

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

we find that

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

We clearly then have that

$$\sum_{r \ge 1} V_r \ll_{C,c} \frac{V(z)}{e^{cu}},$$

thus completing the proof. □

3.3 Sieving general sets

The typical object we want to study in sieve methods, and for which we constructed our sieve weights, is the quantity

$$S(\mathcal{A}, \mathcal{P}) = \sum_{\substack{a \in \mathcal{A} \\ (a, \mathcal{P})=1}} w_a$$

where $\mathcal{A}$ is a finite subset of $\mathbb{Z}$, w_a are some non-negative weights (often, we take them all to be 1) and $\mathcal{P}$ is a subset of $\{p < z\}$. As in the previous section, the parameter z is called the level of the sifting range. We give some examples below to demonstrate the kind of problems we will be considering. In all examples below, we take $w_a = 1$ for all a.

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

$$\mathcal{A} = \{n(n+2) : n \le x\} \quad \text{and} \quad \mathcal{P} = \{p \le \sqrt{x+2}\}.$$

  Alternatively, we can take

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

- To count the number of representations of the even integer $2N$ as the sum of two primes (cf. Goldbach's conjecture), we take

$$\mathcal{A} = \{n(2N-n) : n \le 2N\} \quad \text{and} \quad \mathcal{P} = \{p \le \sqrt{2N}\}$$

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

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

- To count the integers all of whose prime factors are $\equiv 1 \pmod 4$, we take

$$\mathcal{A} = \{n \le x\} \quad \text{and} \quad \mathcal{P} = \{p \le x : p \equiv 3 \pmod 4\}.$$

  If we focus on integers that are in the progression $1 \pmod 4$, then we can take

$$\mathcal{A} = \{n \le x : n \equiv 1 \pmod 4\} \quad \text{and} \quad \mathcal{P} = \{p \le \sqrt{x} : p \equiv 3 \pmod 4\}.$$

The above examples indicate how flexible this terminology is. On the other hand, what we have already seen in the previous sections should make us sceptical of our abilities to handle levels z that are as big as in the above examples. Thus we will see that sieve methods can produce very good upper bounds on $S(\mathcal{A}, \mathcal{P})$, but the lower bounds will only be non-trivial when z is small enough.

Applying Möbius inversion, we find that

$$S(\mathcal{A}, \mathcal{P}) = \sum_{a \in \mathcal{A}} w_a \sum_{(a, \mathcal{P})=1} \mu(d). \tag{3.3.1}$$

Therefore, we can in principle obtain upper and lower bounds on $S(\mathcal{A}, \mathcal{P})$ using the sieve weights constructed in the previous section. However, we need to impose certain conditions on our set $\mathcal{A}$. Changing the order of summation in (3.3.1), we find that

$$S(\mathcal{A}, \mathcal{P}) = \sum_{d \mid m} \mu(d) W_d, \tag{3.3.2}$$

where

$$W_d := \sum_{\substack{a \in \mathcal{A} \\ d \mid a}} w_a.$$

In order to proceed, we assume that we have an approximation of the form

$$W_d = g(d) \cdot X + r_d, \tag{3.3.3}$$

for every integer d. We think of X as an approximation to the total weight W_1 and of $g(d) \cdot X$ as an approximation to W_d (in particular, $g(1) = 1$). The term r_d is the remainder term in this approximation of W_d by $g(d) \cdot X$.

We can think of W_d/W_1 as the probability that a given member of $\mathcal{A}$ is divisible by d. In particular, we should have that $0 \le g \le 1$. In fact, we will be assuming that $0 \le g(d) < 1$ for $d | \prod_{p \in \mathcal{P}} p$; otherwise, an element $a \in \mathcal{A}$ would be divisible by some prime $p \in \mathcal{P}$ with probability one, thus forcing $S(\mathcal{A}, \mathcal{P})$ to be very small (possibly zero). Sieve methods are particularly effective when the function g is multiplicative, which we will be assuming from now on. Thinking of $g(d)$ as the probability that a given member of $\mathcal{A}$ is divisible by d, the assumption that g is multiplicative means that the events $m|a$ and $n|a$ are approximately independent if m and n are coprime.[1] Note that the product $\prod_{p \in \mathcal{P}}(1 - g(p))$ can be then thought of as an approximation to the probability that $a \in \mathcal{A}$ is not divisible by any prime in $\mathcal{P}$, leading us to the prediction that

$$S(\mathcal{A}, \mathcal{P}) \approx X \cdot \prod_{p \in \mathcal{P}} (1 - g(p)).$$

As the following theorem shows, this is indeed true when $\mathcal{P}$ does not contain very large primes.

Theorem 3.3.1 (Fundamental Lemma of Sieve Methods, II)**.** *Let $\mathcal{A}$ be a finite set of integers, $(w_a)_{a \in \mathcal{A}}$ a set of non-negative weights satisfying* (3.3.3)*, where $g : \mathbb{N} \to [0, 1]$ is a multiplicative function with $0 \le g(p) \le \min\{1 - 1/p, C/p\}$. For $u \ge \epsilon > 0$ and any $c > 0$, we have that*

$$S(\mathcal{A}, \mathcal{P}) = \big\{1 + O_{c,\epsilon,C}(e^{-cu})\big\} \cdot X \cdot \prod_{p \in \mathcal{P}} (1 - g(p)) + O\Bigg(\sum_{\substack{p|d \ \Rightarrow \ p \in \mathcal{P} \\ d \le z^u}} \mu^2(d) |r_d| \Bigg).$$

Proof. First, consider the case when $u \ge 2$ and let $\mu^{\pm}$ be the two sequences from Theorem 3.2.1 applied with $D = z^u$. Then

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}) = \sum_{\substack{a \in \mathcal{A} \\ (a, \mathcal{P}) = 1}} w_a &\le \sum_{a \in \mathcal{A}} \sum_{d | (a, \mathcal{P})} \mu^+(d) = \sum_{d | p \ \Rightarrow \ p \in \mathcal{P}} \mu^+(d) W_d \\ &= X \sum_{d | p \ \Rightarrow \ p \in \mathcal{P}} \mu^+(d) g(d) + \sum_{d | p \ \Rightarrow \ p \in \mathcal{P}} \mu^+(d) r_d. \end{aligned}$$

By property (c) of Theorem 3.2.1 applied with $f(d)$ being equal to $g(d) \cdot d$ times the characteristic function of integers all of whose prime factors are in $\mathcal{P}$, we find that the main term equals

$$\{1 + O_{c,C}(e^{-cu})\} \cdot X \cdot \prod_{p \in \mathcal{P}} (1 - g(p)).$$

[1] In practice, this is only true if m and n are not too big. This failure of pseudo-independence is one of the reasons for the limitations of sieve methods.

For the error term, we note that $|\mu^+(d)| \le \mu^2(d) \cdot \mathbf{1}_{d\le D}$, whence we prove that

$$S(\mathcal{A},\mathcal{P}) \le \{1 + O_{c,\epsilon,C}(e^{-cu})\} \cdot X \cdot \prod_{p\in\mathcal{P}}(1-g(p)) + \sum_{\substack{p|d \Rightarrow p\in\mathcal{P}\\ d\le z^u}} \mu^2(d)|r_d|.$$

We prove similarly a lower bound, by using the lower bound sieve weights μ^-. This completes the proof of the theorem when $u \ge 2$.

Finally, if $\epsilon \le u \le 2$, then note that

$$0 \le S(\mathcal{A},\mathcal{P}) \le S(\mathcal{A},\mathcal{P}\cap\{p < z^{\epsilon/2}\}).$$

We then have that

$$\begin{aligned} S(\mathcal{A},\mathcal{P}\cap\{p<z^{\epsilon/2}\}) &\ll X\cdot \prod_{\substack{p\in\mathcal{P}\\ p<z^{\epsilon/2}}}(1-g(p)) + \sum_{\substack{p|d\Rightarrow p\in\mathcal{P}\\ d\le z^\epsilon}}\mu^2(d)|r_d| \\ &\ll_{\epsilon,C} X\cdot\prod_{p\in\mathcal{P}}(1-g(p)) + \sum_{\substack{p|d\Rightarrow p\in\mathcal{P}\\ d\le z^\epsilon}}\mu^2(d)|r_d|, \end{aligned}$$

where we used the case $u = 2$ of the theorem that we already proved. This completes the proof. □

We give a famous application of Theorem 3.3.1 to illustrate how it can be used in practice.

Theorem 3.3.2. *We have that*

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

Moreover, there is an integer r such that

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

Proof. Note that

$$\#\{p \le x : p+2 \text{ prime}\} \le \#\{n \le x : P^-(n(n+2)) \ge \sqrt{x}\} + O(\sqrt{x}) = S(\mathcal{A},\mathcal{P}) + O(\sqrt{x})$$

with $\mathcal{A} = \{n(n+2) : n \le x\}$, $w_a = 1$ for all a and $\mathcal{P} = \{p \le \sqrt{x}\}$. We want to prove that (3.3.3) holds for appropriate choices of X, g and r_d. Note that

$$W_d = \#\{n \le x : n(n+2) \equiv 0 \,(\mathrm{mod}\, d)\} = \nu(d)\cdot\frac{x}{d} + O(\nu(d)),$$

where $\nu(d) = \#\{n \in \mathbb{Z}/d\mathbb{Z} : n(n+2) \equiv 0 \,(\mathrm{mod}\, d)\}$. Thus (3.3.3) holds with $X = x$, $g(d) = \nu(d)/d$ and $r_d \ll \nu(d)$. Moreover, we note that $g(p) \le \min\{1-1/p, 2/p\}$. We may then apply Theorem 3.3.1: we take $u = 1$ to find that

$$S(\mathcal{A},\mathcal{P}) \ll x\prod_{p<\sqrt{x}}\left(1-\frac{\nu(p)}{p}\right) + \sum_{d\le\sqrt{x}}\mu^2(d)\nu(d).$$

Since $\nu(p) \leq 2$, we deduce that $\mu^2(d)\nu(d) \leq \tau(d)$. So the error term is $\ll \sqrt{x}\log^2 x$ by Theorem 1.3.3. We thus arrive to the inequality

$$\#\{p \leq x : p+2 \text{ prime}\} \ll x \prod_{p<\sqrt{x}} \left(1 - \frac{\nu(p)}{p}\right) + \sqrt{x}\log^2 x.$$

Finally, we estimate the product over the primes. We note that $\nu(2) = 1$ and $\nu(p) = 2$ for all odd primes p. Taking logarithms and using the Taylor expansion of the logarithm about 1 yields the estimate

$$\log \prod_{p<\sqrt{x}} \left(1 - \frac{\nu(p)}{p}\right) = \sum_{3\leq p<\sqrt{x}} \left(1 - \frac{2}{p}\right) + O(1) = \sum_{3\leq p<\sqrt{x}} \left(-\frac{2}{p} + O\left(\frac{1}{p^2}\right)\right) + O(1).$$

We then apply Mertens estimate to find that

$$\prod_{p<\sqrt{x}} \left(1 - \frac{\nu(p)}{p}\right) \asymp \frac{1}{\log^2 x}, \tag{3.3.4}$$

thus proving the desired upper bound on the number of twin primes up to x.

For the second part of the theorem, we note that if an integer m satisfies the inequality $P^-(m) > m^{1/(r+1)}$, then $\Omega(m) \leq r$. Consequently,

$$\#\{n \leq x : \Omega(n(n+2)) \leq r\} \geq S(\mathcal{A}, \{p < z\}) \tag{3.3.5}$$

with $z = (x(x+3))^{1/(r+1)}$ and $\mathcal{A}$ as above. We apply Theorem 3.2.1 with $c = 4$ and $u = r/4$, so that $cu = r$ and $z^u \leq \sqrt{x}$ to find that

$$S(\mathcal{A}, \{p < z\}) = (1 + O(e^{-r}))x \prod_{p<z} \left(1 - \frac{\nu(p)}{p}\right) + O(\sqrt{x}\log^2 x),$$

where the error term was estimated as above. If r is big enough (but fixed), the above inequality and (3.3.5) imply that

$$\#\{n \leq x : \Omega(n(n+2)) \leq r\} \geq \frac{x}{2} \prod_{p<z} \left(1 - \frac{\nu(p)}{p}\right) - O(\sqrt{x}\log^2 x).$$

Finally, we note that the product over primes is $\asymp 1/\log^2 z$ by using the argument leading to (3.3.4). This completes the proof of the theorem. □

Remark 3.3.1. As a direct corollary of the upper bound in Theorem 3.3.2, we deduce that the sum

$$\sum_{(p,p+2) \text{ twin primes}} \frac{1}{p}$$

converges. The value of this sum is called *Brun's constant* and its numerical calculation has an interesting history, as it led to the discovery of a bug in Intel's® PentiumTM microprocessor by Nicely.[2]

[2]Weisstein, Eric W. "Brun's Constant." From MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/BrunsConstant.html`

We conclude this section with a brief discussion of an important concept in sieve methods called the *sifting dimension*, usually denoted 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 $f(x) \in \mathbb{Z}[x]$ 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, it is not so hard to show that (3.3.3) holds with $g(d) = \nu_f(d)/d$ and $r_d \ll \nu_f(d)$, where $\nu_f(d)$ counts the roots of $f(x) \bmod d$.

Let us consider two specific two examples to clarify the above point. If $\mathcal{A} = \{n \le x\}$ (corresponding to $f(x) = x$), 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 $p \le \sqrt{x}$, that is to say, $\kappa = 1$. On the other hand, if $\mathcal{A} = \{n(n+2) : n \le x\}$, where $f(x) = x(x+2)$, 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 $p \le \sqrt{x+2}$. Hence $\kappa = 2$ here. Generally speaking, the sieving problem becomes harder as κ increases. Moreover, the optimal choice of β when constructing the weights $\mu^{\pm}$ by (3.2.5) depends intimately on the size of κ.

Exercises

Exercise 3.3.1. Prove that there is a constant $C > 0$ such that

$$\#\{n \le x : P^-(n) > y\} \asymp \frac{x}{\log y} \quad (x \ge Cy,\ y \ge 2).$$

[*Hint :* For large y, use Chebyshev's estimate.]

Exercise 3.3.2. Prove that

$$\#\{n \le x : n^2 + 1 \text{ prime}\} \ll x \prod_{\substack{p \le x \\ p \equiv 1 \,(\mathrm{mod}\, 4)}} \left(1 - \frac{2}{p}\right) \quad (x \ge 2),$$

as well as that there is an integer r such that

$$\#\{n \le x : \Omega(n^2 + 1) \le r\} \gg x \prod_{\substack{p \le x \\ p \equiv 1 \,(\mathrm{mod}\, 4)}} \left(1 - \frac{2}{p}\right) \quad (x \ge 2).$$

Exercise 3.3.3. For any $s \in \mathbb{Z} \setminus \{0\}$, prove that

$$\#\{p \le x : p + 2s \text{ prime}\} \ll \frac{|s|}{\varphi(|s|)} \cdot \frac{x}{\log^2 x} \quad (x \ge 2).$$

Exercise 3.3.4. Prove that

$$\#\{n \le x : p | n \ \Rightarrow p \equiv 1 \,(\mathrm{mod}\, 4)\} \ll \frac{x}{\sqrt{\log x}}.$$

3.4 Selberg's sieve

In 1947, Selberg introduced a different approach to sieving. His idea relies on the following simple observation: if $\{\lambda_d\}_d$ is any sequence of real numbers supported on integers $d \le D$ that are composed of primes $p \in \mathcal{P}$, then then

$$\mathbf{1}_{(a,\mathcal{P})=1} \le \left(\sum_{d|a} \lambda_d \right)^2. \tag{3.4.1}$$

Opening the square, we find that

$$\mathbf{1}_{(a,\mathcal{P})=1} \le \sum_{d_1,d_2|a} \lambda_{d_1}\lambda_{d_2} = \sum_{d|a} \sum_{[d_1,d_2]=d} \lambda_{d_1}\lambda_{d_2}.$$

In particular, the sequence $\mu^+(d) = \sum_{[d_1,d_2]=d} \lambda_{d_2}\lambda_{d_2}$ forms an upper bound sieve whose level is D^2 if the weights λ_d are supported on integers $d \le D$. [3] Using the inequality (3.4.1), we obtain the upper bound

$$\begin{aligned} S(\mathcal{A},\mathcal{P}) \le \sum_{a\in\mathcal{A}} w_a \left(\sum_{d|(a,m)} \lambda_d \right)^2 &= \sum_{a\in\mathcal{A}} w_a \sum_{d_1,d_2|(a,m)} \lambda_{d_1}\lambda_{d_2} \\ &= \sum_{\substack{d_1,d_2\le D \\ d_1,d_2|m}} \lambda_{d_1}\lambda_{d_2} \cdot W_{[d_1,d_2]}. \end{aligned} \tag{3.4.2}$$

Our task is then to minimize the right hand side of (3.4.2), which is a quadratic form in the coefficients λ_d, under the constraint $\lambda_1 = 1$. This is to hard to be accomplished in this generality. Instead, Selberg inserted (3.3.3) and optimized the quadratic form

$$\sum_{\substack{d_1,d_2\le D \\ d_1,d_2|m}} \lambda_{d_1}\lambda_{d_2} \cdot g([d_1,d_2])$$

instead, which is an easier task.

We will demonstrate Selberg's optimization argument in a concrete application, called the *Brun-Titchmarsch inequality*, which serves as an excellent showcase of the power of sieve methods. Indeed, this inequality provides a sharp upper bound (up to the constant 2) for the number of primes $p \equiv a \pmod q$ with $p \in (x-y, x]$ as long as $y \ge q^{1+\epsilon}$ and $y \ge x^\epsilon$, a range that, as we will see in Chapter 10, is beyond the reach of the famous Generalized Riemann Hypothesis.

Theorem 3.4.1 (Brun-Titchmarsch inequality)**.** *For $1 \le q \le y \le x$ and $(a,q) = 1$, we have*

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

[3] The inequality (3.4.1) *a priori* produces upper bound sieves, but combining it with Buchstab's identity (3.2.2) it is possible to produce lower bound sieves. We will not pursue this here.

Proof. Note that

$$\pi(x;q,a)-\pi(x-y;q,a)\le S(\mathcal{A},\mathcal{P}),$$

where

$$\mathcal{A}=\{x<n\le x+y : n\equiv a\,(\mathrm{mod}\,q)\},\quad w_a=1\quad\text{and}\quad \mathcal{P}=\{p<z : p\nmid q\}.$$

Here, z is a parameter $\le x$ to be chosen later. We do not need to involve the primes $p|q$ in $\mathcal{P}$, because the condition $n\equiv a\,(\mathrm{mod}\,q)$ guarantees that $(n,q)=1$. Now, let λ_d be some parameters supported in the set

$$\mathcal{D}:=\{d\le D : \mu^2(d)=1,\ p|d\ \Rightarrow p\in\mathcal{P}\} \tag{3.4.3}$$

with $\lambda_1=1$, so that

$$\begin{aligned}\pi(x+y;q,a)-\pi(x;q,a)&\le \sum_{\substack{x<n\le x+y\\ n\equiv a\,(\mathrm{mod}\,q)}}\left(\sum_{d|a}\lambda_d\right)^2\\ &=\sum_{d_1,d_2\in\mathcal{D}}\lambda_{d_1}\lambda_{d_2}\cdot\#\left\{x<n\le x+y:\ \begin{array}{l} n\equiv a\,(\mathrm{mod}\,q)\\ n\equiv 0\,(\mathrm{mod}\,[d_1,d_2])\end{array}\right\}.\end{aligned}$$

Since $(d_1d_2,q)=1$, the Chinese Remainder Theorem implies that conditions $n\equiv a\,(\mathrm{mod}\,q)$ and $n\equiv 0\,(\mathrm{mod}\,[d_1,d_2])$ can be combined into a single congruence mod $[d_1,d_2]q$. Using the inequalities

$$\frac{y}{[d_1,d_2]q}-1\le \#\{x<n\le x+y : n\equiv b\,(\mathrm{mod}\,[d_1,d_2]q)\}\le \frac{y}{[d_1,d_2]q}+1,$$

we deduce that

$$\begin{aligned}\pi(x+y;q,a)-\pi(x;q,a)&\le \frac{y}{q}\sum_{d_1,d_2\in\mathcal{D}}\frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]}+\sum_{d_1,d_2\in\mathcal{D}}|\lambda_{d_1}\lambda_{d_2}|\\ &=\frac{y}{q}\sum_{d_1,d_2\in\mathcal{D}}\frac{\lambda_{d_1}\lambda_{d_2}}{d_1d_2}(d_1,d_2)+\left(\sum_{d\in\mathcal{D}}|\lambda_d|\right)^2.\end{aligned}$$

In order to diagonalize the quadratic form in the λ_d's, we note that

$$(d_1,d_2)=\sum_{m|(d_1,d_2)}\varphi(m)=\sum_{\substack{d_1|m\\ d_2|m}}\varphi(m),$$

a consequence of the convolution identity $n=(1*\varphi)(n)$. Writing $d_j=e_jm$, we find that

$$\begin{aligned}\pi(x+y;q,a)-\pi(x;q,a)&\le \frac{y}{q}\sum_{\substack{m\le D\\ p|m\implies p\in\mathcal{P}}}\frac{\varphi(m)}{m^2}\sum_{\substack{e_1,e_2\le D/m\\ p|e_1e_2\implies p\in\mathcal{P}}}\frac{\lambda_{me_1}\lambda_{me_2}}{e_1e_2}+\left(\sum_{d\le D}|\lambda_d|\right)^2\\ &=\frac{y}{q}\sum_{m\in\mathcal{D}}\frac{\varphi(m)}{m^2}\left(\sum_{e:\ me\in\mathcal{D}}\frac{\lambda_{me}}{e}\right)^2.\end{aligned}$$

This suggests making a change of variables. Indeed, we set

$$\xi_m = \sum_{e:\ me\in\mathcal{D}} \frac{\lambda_{me}}{e},$$

which is also a sequence supported in $\mathcal{D}$. Moreover, for any $d \in \mathcal{D}$, we have

$$\begin{aligned}\sum_{k:\ dk\in\mathcal{D}} \frac{\xi_{dk}\mu(k)}{k} &= \sum_{k:\ dk\in\mathcal{D}} \frac{\mu(k)}{k} \sum_{e:\ dke\in\mathcal{D}} \frac{\lambda_{dke}}{e} \\ &= \sum_{e,k:\ dek\in\mathcal{D}} \frac{\mu(k)\lambda_{dke}}{ke} \\ &= \sum_{f:\ df\in\mathcal{D}} \frac{\lambda_{df}}{f} \sum_{ek=f} \mu(k) = \lambda_d,\end{aligned} \tag{3.4.4}$$

by Möbius inversion. 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

$$\sum_{m\in\mathcal{D}} \frac{\xi_m\mu(m)}{m} = 1. \tag{3.4.5}$$

So our task now is to minimize

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

under condition (3.4.5). Using Lagrange multipliers, we find that this is achieved when

$$\xi_m = c \cdot \mathbf{1}_{\mathcal{D}}(m) \cdot \frac{\mu(m)m}{\varphi(m)}.$$

In particular, ξ_m is supported on square-free integers and, hence, so if λ_d. Inserting the above formula into (3.4.5), we find that

$$c \sum_{m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)} = 1.$$

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

$$\begin{aligned}\lambda_d = \sum_{k:\ dk\in\mathcal{D}} \frac{\mu(k)\xi_{dk}}{k} &= c \sum_{k:\ dk\in\mathcal{D}} \frac{\mu(k)\mu(dk)d}{\varphi(dk)} \\ &= \frac{\mu(d)d}{\varphi(d)} \sum_{k:\ dk\in\mathcal{D},\,(k,d)=1} \frac{\mu^2(k)}{\varphi(k)} \Bigg/ \sum_{m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)}.\end{aligned} \tag{3.4.6}$$

In particular, we note that

$$|\lambda_d| \le 1. \tag{3.4.7}$$

Indeed, we have that

$$\frac{d}{\varphi(d)} = \sum_{f|d} \frac{\mu^2(f)}{\varphi(f)},$$

so that

$$\frac{d}{\varphi(d)} \sum_{k:\ dk\in\mathcal{D},\,(k,d)=1} \frac{\mu^2(k)}{\varphi(k)} = \sum_{f|d} \sum_{k:\ dk\in\mathcal{D},\,(k,d)=1} \frac{\mu^2(kf)}{\varphi(kf)} \le \sum_{m:\ m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)}.$$

Combining the above, we deduce that

$$\begin{aligned}\pi(x+y;q,a)-\pi(x;q,a) &\le \frac{y}{q} \sum_{\substack{m\le D\\ p|m \implies p\in\mathcal{P}}} \frac{\varphi(m)\xi_m^2}{m^2} + |\mathcal{D}|^2\\ &= \frac{c^2 y}{q} \sum_{m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)} + |\mathcal{D}|^2\\ &= \frac{y}{q}\left(\sum_{m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)}\right)^{-1} + |\mathcal{D}|^2.\end{aligned}$$

In order to make the error term $|\mathcal{D}|^2$ small compared to the main term, we take $D = z \le \sqrt{y/q}$. As before, we notice that

$$\frac{q}{\varphi(q)} \sum_{m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)} = \sum_{d|q} \frac{\mu^2(d)}{\varphi(d)} \sum_{\substack{m\le D\\ (m,q)=1}} \frac{\mu^2(m)}{\varphi(m)} \ge \sum_{n\le D} \frac{\mu^2(n)}{\varphi(n)}.$$

Moreover, if $\mathrm{rad}(a) := \prod_{p|a} p$, then the identity $1/p + 1/p^2 + \cdots = 1/(p-1)$ implies that

$$\sum_{n\le D} \frac{\mu^2(n)}{\varphi(n)} = \sum_{n\le D} \mu^2(n) \sum_{a:\ \mathrm{rad}(a)=n} \frac{1}{a} = \sum_{a:\ \mathrm{rad}(a)\le D} \frac{1}{a} \ge \sum_{a\le D} \frac{1}{a} \ge \int_1^D \frac{\mathrm{d}t}{t} = \log D,$$

so that

$$\sum_{m\in\mathcal{D}} \frac{\mu^2(m)}{\varphi(m)} \ge \frac{\varphi(q)}{q} \log D.$$

(Using the convolution method, we may actually prove that $\sum_{m\in\mathcal{D}} \mu^2(m)/\varphi(m) \sim (\varphi(q)/q)\log D$, so the above inequality is very sharp.) Therefore

$$\pi(x+y;q,a)-\pi(x;q,a) \le \frac{y}{\varphi(q)\log D} + D^2.$$

In order to optimize this inequality, we take $D = \sqrt{y/q}/\log(y/q)$, which proves the theorem when $y \ge 2q$. Finally, when $q \le y \le 2q$, then we note that the theorem follows from the trivial bound $\pi(x+y;q,a) - \pi(x;q,a) \le y/q + 1 \le 2y/q$. □

Remark 3.4.1. We note that relation (3.4.6) implies that

$$\lambda_d \approx \mu(d)\frac{\log(D/d)}{\log D} \quad (d \le D).$$

In particular, we see that the λ_d's are close to a continuous function, since they approach 0 as $d \to D^-$. This is a very important feature of the weights in the Selberg sieve and it will motivate our choice of weights in the GPY sieve, in Chapter 9.

Exercises

Exercise 3.4.1. Use Selberg's sieve to prove that

$$\#\{p \le x : p+2 \text{ prime}\} \le (8 + o_{x\to\infty}(1)) \cdot \frac{cx}{\log^2 x}$$

where

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

is the so-called *twin prime constant.* [4]

[4] According to a probabilistic heuristic, we should have that the number of twin primes up to x is $\sim cx/\log^2 x$.

Chapter 4

Dirichlet series

To every arithmetic function f, we may associate its Dirichlet series

$$L(s,f) := \sum_{n=1}^{\infty} \frac{f(n)}{n^s},$$

defined for those $s \in \mathbb{C}$ that the sum converges. The most famous Dirichlet series is the *Riemann zeta function*

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s},$$

corresponding to the constant function $f = 1$. We will study ζ in detail in Sections 4.2 and 4.4, and again in Chapter 10.

We can think of $L(s,f)$ as a multiplicative Fourier transform for the summatory function of f. We illustrate this point when $|f| \leq 1$, so that $L(s,f)$ converges absolutely when $\mathrm{Re}(s) > 1$. For every such s, partial summation implies that

$$\begin{aligned} L(s,f) = \int_{1^-}^{\infty} \frac{1}{x^s} \mathrm{d} \sum_{n \leq x} f(n) &= \frac{\sum_{n \leq x} f(n)}{x^s}\Big|_{x=1^-}^{\infty} + s \int_{1^-}^{\infty} \sum_{n \leq x} f(n) \frac{\mathrm{d}x}{x^{s+1}} \\ &= s \int_{1}^{\infty} \sum_{n \leq x} f(n) \frac{\mathrm{d}x}{x^{s+1}}. \end{aligned}$$

Making the change of variable $x = e^u$ and writing $s = \sigma + it$, we find that

$$\frac{L(\sigma+it,f)}{\sigma+it} = \int_0^{\infty} \frac{\sum_{n \leq e^u} f(n)}{e^{\sigma u}} e^{-itu} \mathrm{d}u.$$

Therefore, we see that for each fixed c the function $t \to L(c+2\pi it, f)/(c+2\pi it)$ is the Fourier transform of the function $u \to \alpha(u) = e^{-cu} \sum_{n \leq e^u} f(n)$. So, if $c > 1$ is given, then Fourier inversion implies that

$$\frac{\alpha(u^+) + \alpha(u^-)}{2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{L(c+it,f)}{c+it} e^{itu} \mathrm{d}t,$$

where we use the principal value of the improprer integral, that is to say by $\int_{-\infty}^{\infty}$ we mean $\lim_{M\to\infty}\int_{-M}^{M}$. Multiplying by e^{cu} and setting $x = e^u$, we see that

$$\sum_{n<x} f(n) + \frac{f(x)}{2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{L(c+it,f)}{c+it} x^{\sigma+it}\mathrm{d}t = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=c} L(s,f)\frac{x^s}{s}\mathrm{d}s,$$

where $f(x) = f(n)$ if $x = n \in \mathbb{N}$ and $f(x) = 0$ otherwise, and ℓ is the vertical line of all $s \in \mathbb{C}$ with real part equal to s. This formula is due to Perron and it shows that studying the partial sums of f can be also accomplished by studying analytic properties of the associated Dirichlet series. We will return to Perron's formula and study it more rigorously in Section 4.3. First, we develop the basic theory of Dirichlet series in Section 4.1.

4.1 Convergence properties of Dirichlet series

The first result we will show is that Dirichlet series have domains of convergence that are half-planes:

Proposition 4.1.1. *Let $F(s) = \sum_{n=1}^{\infty} a_n/n^s$ be a Dirichlet series.*

(a) If $F(\sigma_0 + it_0)$ converges absolutely, then $F(s)$ converges absolutely for each $s \in \mathbb{C}$ with $\mathrm{Re}(s) \geq \sigma_0$.

(b) If $F(\sigma_0 + it_0)$ converges, then $F(s)$ converges for each $s \in \mathbb{C}$ with $\mathrm{Re}(s) > \sigma_0$.

Proof. (a) We simply observe that $|a_n/n^s| = |a_n|/n^{\mathrm{Re}(s)}$. So, if $\mathrm{Re}(s) \geq \sigma_0$, then $|a_n/n^s| \leq |a_n/n^{s_0}|$ and the absolute convergence of $F(s_0)$ is guaranteed by the comparison criterion.

(b) This follows by partial summation: we write $s_0 = \sigma_0 + it_0$ and $s = s_0 + \tau$, so that $\mathrm{Re}(\tau) > 0$. Then

$$\begin{aligned}\sum_{N<n\leq M} \frac{a_n}{n^s} &= \sum_{N<n\leq M} \frac{a_n}{n^{s_0+\tau}} \\ &= \int_N^M \frac{1}{x^\tau}\mathrm{d}\left(\sum_{N<n\leq x}\frac{a_n}{n^{s_0}}\right) \\ &= \frac{1}{M^\tau}\sum_{N<n\leq M}\frac{a_n}{n^{s_0}} + \tau\int_N^M \frac{1}{x^{\tau+1}}\left(\sum_{N<n\leq x}\frac{a_n}{n^{s_0}}\right)\mathrm{d}x.\end{aligned}$$

By Cauchy's criterion, for each ϵ, there is N_0 such that if $M \geq N \geq N_0$, then

$$\left|\sum_{N<n\leq M}\frac{a_n}{n^{s_0}}\right| < \epsilon.$$

So, if $M \geq N \geq N_0$, then

$$\left|\sum_{N<n\leq M}\frac{a_n}{n^s}\right| \leq \frac{\epsilon}{M^{\mathrm{Re}(\tau)}} + |\tau|\int_N^M \frac{\epsilon}{x^{\mathrm{Re}(\tau)+1}}\mathrm{d}x \leq \epsilon + \frac{\epsilon|\tau|}{\mathrm{Re}(\tau)}.$$

This shows that the series $\sum_{n\geq 1} a_n/n^s$ satisfies Cauchy's criterion, so it converges. □

Given a Dirichlet series $F(s) = \sum_{n=1}^{\infty} a_n/n^s$, we define its *abscissa of convergence*

$$\sigma_c = \sigma_c(F) := \inf\{\sigma \in \mathbb{R} : \exists t \in \mathbb{R} \text{ such that } F(\sigma + it) \text{ converges}\}$$

and its *abscissa of absolute convergence*

$$\sigma_a = \sigma_a(F) := \inf\{\sigma \in \mathbb{R} : F(\sigma) \text{ converges absolutely}\}.$$

Proposition 4.1.1(a) implies that F converges in the half-plane $\mathrm{Re}(s) > \sigma_c$ and Proposition 4.1.1(b) implies that F converges absolutely in the half-plane $\mathrm{Re}(s) > \sigma_a$. We have the following relation for the numbers σ_a and σ_c.

Proposition 4.1.2. *For a Dirichlet series $F(s) = \sum_{n=1}^{\infty} a_n/n^s$, we have*

$$\sigma_c(F) \le \sigma_a(F) \le \sigma_c(F) + 1.$$

Proof. See Exercise 4.1.1 □

Theorem 4.1.3. *Let $F(s) = \sum_{n=1}^{\infty} a_n/n^s$ be a Dirichlet series with abscissa of convergence σ_c. Then $\sum_{n=1}^{\infty} a_n/n^s$ converges uniformly on compact subsets of the half-plane $\mathrm{Re}(s) > \sigma_c$. In particular, F is a holomorphic function in the half-plane $\mathrm{Re}(s) > \sigma_c$.*

Proof. Let K be a compact subset of the half-plane $\mathrm{Re}(s) > \sigma_c$. Then there are numbers $\delta > 0$ and $B \ge 1$ such that if $s \in K$, then $\mathrm{Re}(s) \ge \sigma_c + \delta$ and $|s| \le B$. Now, Theorem 4.1.1(a) implies that F converges at $s_0 := \sigma_c + \delta/2$. In particular, for any $\epsilon > 0$, there is N_0 such that if $M \ge N \ge N_0$, then

$$\left| \sum_{N < n \le M} \frac{a_n}{n^{s_0}} \right| < \epsilon.$$

Now, following the proof of Proposition 4.1.1(a), we find that if $\mathrm{Re}(s) \ge \sigma_c + \delta \ge \mathrm{Re}(s_0) + \delta/2$, then

$$\left| \sum_{N < n \le M} \frac{a_n}{n^s} \right| \le \epsilon + \frac{\epsilon |s - s_0|}{\mathrm{Re}(s - s_0)} \le \epsilon + \frac{2\epsilon(B + |s_0|)}{\delta} \quad (M \ge N \ge N_0,\ s \in K)$$

by the choice of δ, s_0 and B. This implies that the series $\sum_{n \ge 1} a_n/n^s$ converges uniformly inside K. This proves that $\sum_{n \ge 1} a_n/n^s$ converges uniformly on compacts subsets of the half-place $\mathrm{Re}(s) > \sigma_c$. In particular, since this is a series of holomorphic functions, its sum $F(s)$ is holomorphic in the same half-plane. □

An important fact about Dirichlet series of multiplicative functions is that they have Euler product representations in their domain of absolute convergence:

Theorem 4.1.4. *Let f be a multiplicative function and $s \in \mathbb{C}$. Then the Dirichlet series $\sum_{n \ge 1} f(n)/n^s$ converges absolutely if, and only if, the series $\sum_{p \text{ prime}, k \ge 1} f(p^k)/p^{ks}$ converges absolutely. Moreover, in the case when they both converge absolutely, we have that*

$$\sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_p \left(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \cdots\right).$$

In particular, if f is completely multiplicative, then

$$\sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_p \left(1 - \frac{f(p)}{p^s}\right)^{-1}.$$

Proof. This is a rewrite of Theorem 1.3.2. □

Finally, we show that the Dirichlet series of a convolution $f * g$ factors as the product of the two individual series.

Theorem 4.1.5. *Let $f, g : \mathbb{N} \to \mathbb{C}$ be two arithmetic functions whose associated Dirichlet series $L(s, f)$ and $L(s, g)$ both converge absolutely at $s_0 \in \mathbb{C}$. Then the Dirichlet series $L(s, f * g)$ also converges absolutely at s_0 and we have that*

$$L(s_0, f * g) = L(s_0, f)L(s_0, g).$$

Proof. Let $s_0 = \sigma_0 + it_0$. We have that

$$\begin{aligned}\sum_{n=1} \left|\frac{(f*g)(n)}{n^{s_0}}\right| = \sum_{n=1} \frac{1}{n^{\sigma_0}} \left|\sum_{ab=n} f(a)g(b)\right| &\leq \sum_{n=1} \frac{1}{n^{\sigma_0}} \sum_{ab=n} |f(a)g(b)| \\ &= \sum_{ab=n} \frac{|f(a)g(b)|}{(ab)^{\sigma_0}} \\ &= \left(\sum_{a=1}^{\infty} \frac{|f(a)|}{a^{\sigma_0}}\right)\left(\sum_{b=1}^{\infty} \frac{|g(b)|}{b^{\sigma_0}}\right) < \infty;\end{aligned}$$

here, changing the order of summation and rearranging the series is justified by positivity. This proves that $L(s, f * g)$ converges absolutely at $s = s_0$. Finally, we have that

$$\sum_{n=1} \frac{(f*g)(n)}{n^{s_0}} = \sum_{n=1} \frac{1}{n^{s_0}} \sum_{ab=n} f(a)g(b) = \sum_{a,b\geq 1} \frac{f(a)g(b)}{(ab)^{s_0}} = \left(\sum_{a=1}^{\infty} \frac{f(a)}{a^{s_0}}\right)\left(\sum_{b=1}^{\infty} \frac{g(b)}{b^{s_0}}\right);$$

here, changing the order of summation and rearranging the series is justified by the absolute convergence of the double series. □

Exercises

Exercise 4.1.1. Prove Proposition 4.1.2. [*Hint:* what can you say about the sequence a_n/n^s if $\sum_{n\geq 1} a_n/n^s$ converges?]

Exercise 4.1.2. Let $F(s) = \sum_{n\geq 1}^{\infty} a_n/n^s$ be a Dirichlet series with abscissa of convergence σ_c. Prove that the abscissa of convergence for the series of derivatives $-\sum_{n\geq 1}^{\infty} a_n(\log n)/n^s$ is also σ_c. Deduce that

$$F'(s) = -\sum_{n=1}^{\infty} \frac{a_n \log n}{n^s}.$$

Exercise 4.1.3. Prove the following Dirichlet series identities:

(a) $L(s, \tau_k) = \zeta(s)^k$;

(b) $L(s, \sigma) = \zeta(s)\zeta(s-1)$, where σ is the sum-of-divisors function here;

(c) $L(s, \mu^2) = \zeta(s)/\zeta(2s)$;

(d) $L(s, \mu) = 1/\zeta(s)$.

Exercise 4.1.4. Prove that if f is a completely multiplicative function and $s \in \mathbb{C}$ is such that $L(s, f)$ converges absolutely, then

$$\sum_{n=1}^{\infty} \frac{f(n)\Lambda(n)}{n^s} = -\frac{L'}{L}(s, f).$$

Exercise 4.1.5. Consider the Dirichlet series $F(s) = \sum_{n=1}^{\infty}(-1)^{n-1}/n^s$. Prove that F has abscissa of absolute convergence $\sigma_a = 1$ and abscissa of convergence $\sigma_c = 0$. Furthermore, show the identity

$$F(s) = (1 - 2^{-s+1})\zeta(s) \quad (\mathrm{Re}(s) > 1).$$

Exercise 4.1.6. Let $F(s) = \sum_{n\geq 1} a_n/n^s$ and $G(s) = \sum_{n\geq 1} b_n/n^s$ be two Dirichlet series which converge and are equal in a half place $\mathrm{Re}(s) > c$, for some $c \in \mathbb{R}$. Show that $a_n = b_n$ for all $n \in \mathbb{N}$.

4.2 Analytic continuation of Dirichlet series

A crucial property that many Dirichlet series possess is that they can be analytically or meromorphically continued to the left of their half-plane of convergence. This can be done by various means. One of the easiest ways is to use available information on the partial sums of their coefficients. For example, we know that $\zeta(s) = \sum_{n\geq 1} 1/n^s$ has half-plane of convergence $\mathrm{Re}(s) > 1$. On the other hand, we know that $\sum_{n\leq x} 1 = \lfloor x \rfloor = x - \{x\}$. Using this fact and partial summation, we will show that we can meromorphically extend ζ to the half-plane $\mathrm{Re}(s) > 0$. [1] We prove this result in a more general context:

Theorem 4.2.1. *Let $(a_n)_{n=1}^{\infty}$ be a sequence of complex numbers for which there are parameters $\theta \in [0, 1)$, $c \in \mathbb{C}$ and $M \geq 1$ such that*

$$\sum_{n\leq x} a_n = cx + E(x) \quad \textit{with} \quad |E(x)| \leq Mx^{\theta} \quad (x \geq 1). \tag{4.2.1}$$

Then $F(s) = \sum_{n=1}^{\infty} a_n/n^s$ has a meromorphic continuation to the half-plane $\mathrm{Re}(s) > \theta$ via the formula

$$F(s) = \frac{cs}{s-1} + s\int_1^{\infty} \frac{E(x)}{x^{s+1}}\mathrm{d}x.$$

[1] Note that Exercise 4.1.5 provides an alternative way of establishing the meromorphic continuation of ζ to the half-plane $\mathrm{Re}(s) > 0$.

Proof. Notice that (4.2.1) implies that $a_n \ll x^\theta$, whence $F(s) = \sum_{n=1}^\infty a_n/n^s$ converges absolutely for $\mathrm{Re}(s) > 1+\theta$. Let $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$. For $N \in \mathbb{N}$, Theorem 1.2.1 implies that

$$\begin{aligned}\sum_{n=1}^N \frac{a_n}{n^s} &= a_1 + c\int_1^N \frac{\mathrm{d}x}{x^s} + \frac{E(x)}{x^s}\Big|_{x=1}^N + s\int_1^N \frac{E(x)}{x^{s+1}}\mathrm{d}x \\ &= a_1 - E(1) + \frac{c}{s-1} + \frac{E(N)}{N^s} + s\int_1^N \frac{E(x)}{x^{s+1}}\mathrm{d}x.\end{aligned}$$

The integral $\int_1^\infty E(x)x^{-s-1}\mathrm{d}x$ converges absolutely for $\mathrm{Re}(s) > \theta$ by (4.2.1). In particular, the sequence of holomorphic functions $f_N(s) = \int_1^N E(x)x^{-s-1}\mathrm{d}x$ converges uniformly in compact subsets of $\mathrm{Re}(s) > \theta$, so its limit $\int_1^\infty E(x)x^{-s-1}\mathrm{d}x$ is analytic for $\mathrm{Re}(s) > 0$. Therefore, letting $N \to \infty$, we deduce that

$$\begin{aligned}F(s) &= a_1 - E(1) + \frac{c}{s-1} + s\int_1^\infty \frac{E(x)}{x^{s+1}}\mathrm{d}x \\ &= \frac{cs}{s-1} + s\int_1^\infty \frac{E(x)}{x^{s+1}}\mathrm{d}x \quad (\mathrm{Re}(s) > 1+\theta),\end{aligned}$$

since $a_1 = c + E(1)$, by (4.2.1) with $x = 1$. The above formula is valid for $\mathrm{Re}(s) > 1+\theta$, so it can be taken as the definition of ζ in this region. On the other hand, the right hand side of the above formula is well-defined for $\mathrm{Re}(s) > \theta$ and it defines a meromorphic function there, with the only potential singularity being a simple pole of residue c at $s = 1$ (if $c = 0$, then there is obviously no singularity). This proves the theorem. □

Since $\sum_{n\le x} 1 = x - \{x\}$ and $0 \le \{x\} \le 1$, Theorem 4.2.1 implies that $\zeta(s)$ can be meromorphically continued to the half-plane $\mathrm{Re}(s) > 0$ with only a simple pole at $s = 1$ of residue 1 via the formula

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{\{u\}}{u^{s+1}}\mathrm{d}u. \tag{4.2.2}$$

Exercise 4.2.1 indicates how to push further the idea behind Theorem 4.2.1 and prove the meromorphic continuation to the entire complex-plane. See, also, Chapter 10.

As we will see in the next section, it is important to have control on Dirichlet series inside their domain of analytic continuation. This is the context of the next lemma.

Theorem 4.2.2. *Let $(a_n)_{n=1}^\infty$ be a sequence of complex numbers satisfying* (4.2.1) *and let $F(s) = \sum_{n=1}^\infty a_n/n^2$. Assume further that $|a_n| \le 1$ for all n. If $s = \sigma + it$ with $|s-1| \ge |c|$ and $\theta + \delta \le \sigma \le 2$ for some $\delta > 0$, then*

$$F(s) \ll_\delta (1 + (M(|t|+2))^{\frac{1-\sigma}{1-\theta}}) \min\left\{\frac{\sigma}{|\sigma-1|}, \log(M(|t|+2))\right\}.$$

Proof. First of all, note that the assumption that $|a_n| \le 1$ implies that $F(s)$ converges absolutely for $\mathrm{Re}(s) > 1$. Next, recall the formula for $F(s)$ given in Theorem 4.2.1, which is obtained by partial summation on $\sum_{n\ge1} a_n/n^s$. When we apply partial summation, we

implicitly assume that $1/n^s \approx \int_n^{n+1} \mathrm{d}x/x^s$ and $\sum_{n\le x} a_n \sim cx$. However, the first of these asymptotics is sharp only when $n > |s|$, and for the second one we need that $x > M^{1/(1-\theta)}$ for the main term in (4.2.1) to dominate. Thus partial summation will be much more efficient if applied to the tails of the Dirichlet series $\sum_{n>N} a_n/n^s$, with N large enough. Following the proof of Theorem 4.2.1, we may show that

$$\sum_{n>N} \frac{a_n}{n^s} = \frac{c\cdot N^{-s+1}}{s-1} + \frac{E(N)}{N^s} + s\int_N^\infty \frac{E(x)}{x^{s+1}}\mathrm{d}x$$

for any integer $N \ge 2$ and $\mathrm{Re}(s) > 1$. Therefore

$$F(s) = \sum_{n=1}^{N} \frac{a_n}{n^s} + \frac{c\cdot N^{-s+1}}{s-1} + \frac{E(N)}{N^s} + s\int_N^\infty \frac{E(x)}{x^{s+1}}\mathrm{d}x \tag{4.2.3}$$

for $\mathrm{Re}(s) > 1$. Since both sides of this equality are meromorphic for $\mathrm{Re}(s) > \theta$ with only a simple pole at $s = 1$ of residue c, they must be equal for $\mathrm{Re}(s) > \theta$. We take absolute values and use our assumptions that $|a_n| \le 1$ and that $|E(x)| \le Mx^\theta$ to find that

$$\begin{aligned} |F(s)| &\le \sum_{n=1}^{N} \frac{1}{n^\sigma} + N^{1-\sigma} + M\cdot N^{\theta-\sigma} + |s|\cdot\int_N^\infty \frac{M}{x^{\sigma-\theta+1}}\mathrm{d}x \\ &= \sum_{n=1}^{N} \frac{1}{n^\sigma} + N^{1-\sigma} + M\cdot N^{\theta-\sigma} + \frac{M|s|N^{\theta-\sigma}}{\sigma-\theta}, \end{aligned} \tag{4.2.4}$$

where we used assumption that $|s-1| \ge |c|$. For the first sum, we note that

$$\frac{1}{n^\sigma} \le \int_{n-1}^{n} \frac{\mathrm{d}x}{x^\sigma},$$

so that

$$\sum_{n=1}^{N} \frac{1}{n^\sigma} \le 1 + \int_1^N \frac{\mathrm{d}x}{x^\sigma} = 1 + \frac{N^{1-\sigma}-1}{1-\sigma} \ll (1+N^{1-\sigma})\min\left\{\frac{\sigma}{\sigma-1}, \log N\right\}. \tag{4.2.5}$$

Indeed, it is clear that we have the upper bound

$$1 + \frac{N^{1-\sigma}-1}{1-\sigma} \ll (1+N^{1-\sigma})\frac{\sigma}{|\sigma-1|}.$$

It is also clear we have the bound

$$1 + \frac{N^{1-\sigma}-1}{1-\sigma} \ll (1+N^{1-\sigma})\log N$$

if $|\sigma-1| \ge 1/\log N$. Finally, if $|1-\sigma| < 1/\log N$, then

$$N^{1-\sigma} = e^{(1-\sigma)\log N} = 1 + O((1-\sigma)\log N)$$

by the Taylor expansion of e^x about 0. This establishes (4.2.5) in all cases. Inserting this estimate into (4.2.4), we conclude that

$$|F(s)| \ll (1+N^{1-\sigma})\min\left\{\frac{\sigma}{|\sigma-1|}, \log N\right\} + \frac{M(|t|+2)N^{\theta-\sigma}}{\sigma-\theta}.$$

Taking $N = (M(|t|+2))^{1/(1-\theta)}$ completes the proof of the theorem. □

Exercises

Exercise 4.2.1. (a) Recall the function $F(x) = \int_0^x (\{t\} - 1/2)\mathrm{d}t$, defined in the proof of Stirling's formula (cf. Theorem 1.2.4). Using this function, prove that $\zeta(s) - 1/(s-1)$ can be extended to an analytic function in the half-planeRe$(s) > -1$ via the formula

$$\zeta(s) - \frac{1}{s-1} = \frac{1}{2} + s(s+1)\int_1^\infty \frac{F(x)}{x^{s+2}}\mathrm{d}x.$$

(b) For each $k \in \mathbb{N}$, show that there is a polynomial P_k and a bounded periodic function F_k such that

$$\zeta(s) - \frac{1}{s-1} = P_k(s) + s(s+1)\cdots(s+k-1)\int_1^\infty \frac{F_k(x)}{x^{s+k}}\mathrm{d}x.$$

Deduce that $\zeta(s) - 1/(s-1)$ can be analytically continued to an entire function.

Exercise 4.2.2. Prove that if

$$\sum_{n\le x}\mu(n) \ll x^\theta \quad (x \ge 1)$$

for some $\theta \in [0,1)$, then $\zeta(s) \neq 0$ for Re$(s) > \theta$.

Exercise 4.2.3. Consider the Dirichlet series $F(s) = \sum_{n=1}^\infty (-1)^{n-1}/n^s$. Prove that F has abscissa of convergence $\sigma_c = 0$ and

Exercise 4.2.4. Let $r(n) = \#\{(a,b) \in \mathbb{Z} : a^2 + b^2 = n\}$.

(a) Prove that r is a multiplicative function. [*Hint :* use the arithmetic of the Gaussian integers $\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{C}\}$.]

(b) Prove that the Dirichlet series $R(s) = \sum_{n=1}^\infty r(n)/n^s$ has abscissa of absolute convergence $\sigma_a = 1$.

(c) Prove that

$$\sum_{n\le x} r(n) = \pi x + O(\sqrt{x}),$$

and deduce that $R(s) - \pi/(s-1)$ has an analytic continuation to the half-plane Re$(s) > 1/2$. [*Hint :* Observe that the right hand side is the number of lattice points $(a,b) \in \mathbb{Z}^2$ in the circle centered at 0 and of radius $\sqrt{x}$.]

4.3 Perron's inversion formula

In Theorem 4.2.1, we showed that the asymptotics of f determine the analytic behaviour of $L(s,f)$. In this section, we will see that the converse also holds. This is based on the following quantitative version of *Perron's inversion formula*:

Theorem 4.3.1. *Let $f:\mathbb{N}\to\mathbb{C}$ is an arithmetic function with*

$$|f(n)|\ll n^{\alpha}(\log n)^{A} \tag{4.3.1}$$

for some $\alpha, A\geq 0$. For $x,T\geq 2$ and $c\geq \alpha+1+1/\log x$, we have that

$$\sum_{n\leq x} f(n)=\frac{1}{2\pi i}\int\limits_{\substack{\mathrm{Re}(s)=c\\ |\mathrm{Im}(s)|\leq T}} L(s,f)\frac{x^s}{s}\mathrm{d}s+O\left(\frac{x^c(\log x)^{A+1}}{T}+x^{\alpha}(\log x)^{A}\right);$$

the implied constant depends at most on A and the implied constant in (4.3.1).

The key input to the above result is supplied by the lemma below.

Lemma 4.3.2. *If we set*

$$\delta(y)=\begin{cases}0 & \text{if } 0<y<1,\\ 1/2 & \text{if } y=1,\\ 1 & \text{if } y>1,\end{cases}$$

then, for any $c>0$ and any $T\geq 1$, we have

$$\left|\delta(y)-\frac{1}{2\pi i}\int\limits_{\substack{\mathrm{Re}(s)=c\\ |\mathrm{Im}(s)|\leq T}}\frac{y^s}{s}\mathrm{d}s\right|\ll\begin{cases}\dfrac{y^c}{T|\log y|+1} & \text{if } y>0,\ y\neq 1,\\[2ex] \dfrac{c}{T} & \text{if } y=1.\end{cases}$$

Proof. First, we consider the case $y\in(0,1)$. We replace the contour of integration by a contour consisting of three line segments, going from $c-iT$ to $c'-iT$, then to $c'+iT$, and finally to $c+iT$, so that

$$\frac{1}{2\pi i}\int\limits_{\substack{\mathrm{Re}(s)=c\\ |\mathrm{Im}(s)|\leq T}}\frac{y^s}{s}\mathrm{d}s=\frac{1}{2\pi i}\int\limits_{\substack{c\leq \mathrm{Re}(s)\leq c'\\ \mathrm{Im}(s)=-T}}\frac{y^s}{s}\mathrm{d}s+\frac{1}{2\pi i}\int\limits_{\substack{\mathrm{Re}(s)=c'\\ |\mathrm{Im}(s)|\leq T}}\frac{y^s}{s}\mathrm{d}s-\frac{1}{2\pi i}\int\limits_{\substack{c\leq \mathrm{Re}(s)\leq c'\\ \mathrm{Im}(s)=T}}\frac{y^s}{s}\mathrm{d}s,$$

by Cauchy's theorem. The contribution of the two horizontal segments is

$$\ll\int_c^{c'}\frac{y^{\sigma}}{\sigma+T}\mathrm{d}\sigma\leq\frac{1}{T}\int_c^{c'}y^{\sigma}\mathrm{d}\sigma=\frac{y^c}{T|\log y|},$$

The contribution of the vertical segment, with $\mathrm{Re}(s)=c'$, is

$$\ll\int_{-T}^{T}\frac{y^{c'}}{c'+|t|}\mathrm{d}t\to 0\quad(c'\to+\infty).$$

Therefore, letting $c' \to \infty$, we deduce that

$$\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)|\le T}} \frac{y^s}{s} \mathrm{d}s \ll \frac{y^c}{T|\log y|}.$$

This proves the result when $T|\log y| \ge 1$. When $T|\log y| \le 1$, we replace the contour $\{s \in \mathbb{C} : \mathrm{Re}(s) = c,\ |\mathrm{Im}(s)| \le T\}$ by the part of the circle centred at 0 of radius $\sqrt{c^2+T^2}$ that lies to the right of the line $\mathrm{Re}(s) = c$. Then

$$\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)|\le T}} \frac{y^s}{s} \mathrm{d}s = \frac{1}{2\pi i} \int\limits_{\substack{|s|=\sqrt{c^2+T^2} \\ \mathrm{Re}(s)\ge c}} \frac{y^s}{s} \mathrm{d}s \ll y^c.$$

This proves the result when $T|\log y| \le 1$ as well.

The case $y > 1$ is similar. However, instead of shifting the contour to the left though, we shift it to the right so that y^s becomes smaller and smaller in magnitude. We then cross the pole of the function y^s/s at $s = 0$, of residue equal to 1. Therefore, for $c' < 0$, we have that

$$\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)|\le T}} \frac{y^s}{s} \mathrm{d}s = 1 - \frac{1}{2\pi i} \int\limits_{\substack{c'\le \mathrm{Re}(s)\le c \\ \mathrm{Im}(s)=-T}} \frac{y^s}{s} \mathrm{d}s + \frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c' \\ |\mathrm{Im}(s)|\le T}} \frac{y^s}{s} \mathrm{d}s + \frac{1}{2\pi i} \int\limits_{\substack{c'\le \mathrm{Re}(s)\le c \\ \mathrm{Im}(s)=T}} \frac{y^s}{s} \mathrm{d}s.$$

Estimating the three integrals of the right hand side as before proves that

$$\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)|\le T}} \frac{y^s}{s} \mathrm{d}s = 1 + O\left(\frac{y^c}{T|\log y|}\right),$$

which proves the result when $T|\log y| \ge 1$. Finally, when $T|\log y| \le 1$, then we replace the contour $\{s \in \mathbb{C} : \mathrm{Re}(s) = c,\ |\mathrm{Im}(s)| \le T\}$ by the part the circle centred at 0 of radius $\sqrt{c^2+T^2}$ that lies to the left of the line $\mathrm{Re}(s) = c$ and bound the error term as in the case $y < 1$.

It remains to handle the case $y = 1$. There we argue by direct computation: we have that

$$\begin{aligned}\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)|\le T}} \frac{\mathrm{d}s}{s} = \frac{1}{2\pi}\int_{-T}^{T} \frac{\mathrm{d}t}{c+it} = \frac{1}{2\pi}\int_0^T \left(\frac{1}{c+it} + \frac{1}{c-it}\right)\mathrm{d}t = \frac{1}{\pi}\int_0^T \frac{c}{c^2+t^2}\mathrm{d}t \\ = \frac{1}{\pi}\int_0^{T/c} \frac{\mathrm{d}t}{1+t^2} \\ = \frac{1}{2} + O\left(\frac{c}{T}\right),\end{aligned}$$

since $\int_0^\infty \mathrm{d}t/(1+t^2) = \pi/2$ and $\int_x^\infty \mathrm{d}t/(1+t^2) \le \int_x^\infty \mathrm{d}t/t^2 = 1/x$. This completes the proof of the lemma. □

We are finally ready to prove Perron's inversion formula:

Proof of Theorem 4.3.1. Recall the notation $\delta(y)$ from Lemma 4.3.2. Then

$$\begin{aligned}\sum_{n\le x} f(n) &= \sum_{n<x} f(n) + \frac{f(x)}{2} + O(x^\alpha (\log x)^A)\\ &= \sum_{n=1}^{\infty} f(n)\delta(x/n) + O(x^\alpha(\log x)^A)\\ &= \sum_{n=1}^{\infty} f(n)\left\{ \frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c+1+1/\log x\\ |\mathrm{Im}(s)|\le T}} \frac{(x/n)^s}{s}\mathrm{d}s + O\left(\frac{(x/n)^c}{1+T|\log \frac{x}{n}|}\right)\right\} + O(x^\alpha(\log x)^A)\\ &= \frac{1}{2\pi i}\int\limits_{\substack{\mathrm{Re}(s)=c\\ |\mathrm{Im}(s)|\le T}} L(s,f)\frac{x^s}{s}\mathrm{d}s + O(x^c\cdot R + x^\alpha(\log x)^A),\end{aligned}$$

where

$$R = \sum_{n=1}^{\infty} \frac{n^\alpha (\log n)^A}{n^c(1+T|\log \frac{x}{n}|)}.$$

It remains to handle the sum R.

The $O(1)$ terms with $|x-n|\le 1$ contribute $\ll (\log x)^A x^{\alpha-c}$.

When $1\le |x-n| < x/T$, then we note that $T<x$ and $T|\log(x/n)| \ll 1$. So these terms contribute

$$\sum_{1\le |x-n|<x/T} \frac{(\log n)^A}{n^{c-\alpha}(1+T|\log \frac{x}{n}|)} \ll \sum_{1\le |x-n|<x/T} \frac{(\log x)^A}{x^{c-\alpha}} \ll \frac{(\log x)^A}{x^{c-\alpha}}\cdot\frac{x}{T} \le \frac{(\log x)^A}{T}$$

by our assumption that $c\ge \alpha+1+1/\log x$.

When $n\in (x+x/T, 2x]$ and $n>x+1$, we observe that $T|\log x/n| \gg 1$ and, also, that $|\log(x/n)| \asymp (n-x)/x$, by Taylor's expansion of the logarithm about 1. Therefore these terms contribute

$$\begin{aligned}\sum_{x+\max\{1,x/T\}<n\le 2x} \frac{(\log n)^A}{n^{c-\alpha}(1+T|\log \frac{x}{n}|)} &\ll \sum_{x+1<n\le 2x} \frac{(\log x)^A}{x^{c-\alpha-1}T(n-x)}\\ &\ll \frac{(\log x)^{A+1}}{x^{c-\alpha-1}T}\end{aligned}$$

by partial summation used in a similar way as in the proof of Corollary 1.2.3.

Similarly, we find that

$$\sum_{x/2<n\le x-\max\{x/T,1\}} \frac{(\log n)^A}{n^{c-\alpha}(1+T|\log \frac{x}{n}|)} \ll \frac{(\log x)^{A+1}}{x^{c-\alpha-1}T}.$$

Finally, if $n \notin [x/2, 2x]$, then we observe that $|\log(x/n)| \geq \log 2$ and recall that $c \geq \alpha + 1/\log x$. Thus

$$\sum_{n\in\mathbb{N}\setminus[x/2,2x]} \frac{(\log n)^A}{n^{c-\alpha}(1+T|\log \frac{x}{n}|)} \ll \frac{1}{T}\sum_{n=1}^{\infty} \frac{(\log n)^A}{n^{1+1/\log x}} \ll_A \frac{(\log x)^{A+1}}{T}$$

by Theorem 1.2.2. Putting together the above estimates, we find that

$$R \ll (\log x)^A x^{\alpha-c} + \frac{(\log x)^{A+1}}{T}.$$

This completes the proof. □

Theorem 4.3.1 is the basis for a lot of estimates in Analytic Number Theory. Analogous results can be proven for a more general class of arithmetic functions, provided we have some good control on the size of $|f(n)|$ when ranges over short intervals.

The power of Theorem 4.3.1 is unleashed when paired with some ideas from Complex Analysis. We illustrate this in the case when $f = \tau$: as Exercise 4.3.2 states (see below), we have

$$\sum_{n\leq x} \tau(n) = \frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=1+1/\log x \\ |\mathrm{Im}(s)|\leq T}} \zeta(s)^2 \frac{x^s}{s} \mathrm{d}s + O_\epsilon\left(\frac{x^{1+\epsilon}}{T} + x^\epsilon\right), \tag{4.3.2}$$

for any fixed $\epsilon > 0$ and any $x, T \geq 2$. We easily see $|\zeta(\sigma+it)| \leq \zeta(\sigma) \asymp 1/(\sigma-1)$ for $1 < \sigma \leq 2$. So, taking absolute values in (4.3.2) and $T = x$, say, yields the estimate $\sum_{n\leq x} \tau(n) \ll x\log^2 x$, which is much worse than what we proved in Theorem 1.3.3.

Instead of bounding the right hand side of (4.3.2), we recall that $\zeta(s) - 1/(s-1)$ has an analytic continuation to the half-plane $\mathrm{Re}(s) > 0$. We could then use Cauchy's residue theorem to deform the contour $\{s \in \mathbb{C} : \mathrm{Re}(s) = 1 + 1/\log x, |\mathrm{Im}(s)| \leq T\}$ to the contour $C = [1+1/\log x - iT, c - iT] \cup [c - iT, c + iT] \cup [c + iT, 1 + 1/\log x + iT]$, where $0 < c < 1$. We pick up a residue contribution from the double pole of $g(s) = \zeta(s)^2 x^s/s$ at $s = 1$ and bound $g(s)$ on the new contour using Theorem 4.2.2. Notice that $g(s)$ is small when $|\mathrm{Im}(s)| = T$ because of the s in the denominator and it is also small when $\mathrm{Re}(s) = c$ because $|x^s| = x^c$ is small compared to x. More precisely, we have

$$\begin{aligned}\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)|\leq T}} \zeta(s)^2 \frac{x^s}{s} \mathrm{d}s &= \mathrm{Res}_{s=1}\left(\zeta(s)^2\frac{x^s}{s}\right) + \frac{1}{2\pi i}\int_C \zeta(s)^2 \frac{x^s}{s}\mathrm{d}s \\ &= \mathrm{Res}_{s=1}\left(\zeta(s)^2\frac{x^s}{s}\right) + O_c\left(x^c T^{2(1-c)}\log T + \frac{x\log^2 T}{T}\right)\end{aligned}$$

by Cauchy's residue theorem and by Theorem 4.2.2. We conclude that

$$\sum_{n\leq x} \tau(n) = \mathrm{Res}_{s=1}\left(\zeta(s)^2\frac{x^s}{s}\right) + O_{c,\epsilon}\left(\frac{x^{1+\epsilon}}{T} + x^c T^{2(1-c)}\right),$$

provided that $c \geq \epsilon$. We take $c = \epsilon$ and $T = x^{1/3}$ to conclude that

$$\sum_{n \leq x} \tau(n) = \mathrm{Res}_{s=1} \left(\zeta(s)^2 \frac{x^s}{s} \right) + O_\epsilon(x^{2/3+\epsilon}).$$

It remains to calculate the above residue.

If $Z(s) = (s-1)\zeta(s)$, which is analytic for $\mathrm{Re}(s) > 0$, and $f(s) = Z(s)^2 x^s/s$, then

$$\mathrm{Res}_{s=1} \left(\zeta(s)^2 \frac{x^s}{s} \right) = g'(1) = x \cdot (Z(1)^2 \log x + 2Z'(1)Z(1) - 1).$$

By (4.2.2), we have that

$$Z(s) = s - s(s-1) \int_1^\infty \frac{\{t\}}{t^{s+1}} dt,$$

and

$$Z'(s) = 1 + s(s-1)(s+1) \int_1^\infty \frac{\{t\}}{t^{s+2}} dt - (2s-1) \int_1^\infty \frac{\{t\}}{t^{s+1}} dt,$$

so that $Z(1) = 1$ and

$$Z'(1) = 1 - \int_1^\infty \frac{\{t\}}{t^2} dt = \gamma,$$

the Euler-Mascheroni constant. Therefore

$$\mathrm{Res}_{s=1} \left(\zeta(s)^2 \frac{x^s}{s} \right) = x \log x + (2\gamma - 1)x,$$

the main term obtained by Dirichlet's hyperbola method. We then conclude that

$$\sum_{n \leq x} \tau(n) = x \log x + (2\gamma - 1)x + O_\epsilon(x^{1/2+\epsilon}) \quad (x \geq 1), \tag{4.3.3}$$

for every fixed $\epsilon > 0$.

Relation (4.3.3) is obviously worse than what we showed in Theorem 1.3.3, but the method leading to it has the advantage of great flexibility. In fact, if coupled with improved bounds on ζ, it can essentially recover Theorem 1.3.3. Moreover, when combined with the functional equation of ζ (c.f. Chapter 10), then it leads to the so-called Voronoi summation formula, which evaluates $\sum_{n \leq x} \tau(n) - (x \log x + (2\gamma - 1)x)$ as an exponential sum, thus making it amenable to techniques from the rich theory of exponential sums. In addition, we will use the above technique of shifting the contour of integration to prove the prime number theorem in the next section.

Exercises

Exercise 4.3.1. Prove that if $f : \mathbb{N} \to \mathbb{C}$ is a completely multiplicative function with $|f| \leq 1$, then

$$\sum_{n \leq x} f(n)\Lambda(n) = -\frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c \\ |\mathrm{Im}(s)| \leq T}} \frac{L'}{L}(s, f) \frac{x^s}{s} \mathrm{d}s + O\left(\frac{x(\log x)^2}{T} + \log x \right).$$

Exercise 4.3.2. For $k \in \mathbb{N}$, $\epsilon > 0$ and $x, T \geq 2$, prove that

$$\sum_{n \leq x} \tau_k(n) = \frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=1+1/\log x \\ |\mathrm{Im}(s)| \leq T}} \zeta(s)^k \frac{x^s}{s} \mathrm{d}s + O_{k,\epsilon}\left(\frac{x^{1+\epsilon}}{T} + x^{\epsilon}\right).$$

Exercise 4.3.3. Let f satisfying (4.3.1).

(a) If $\mathrm{Re}(s) > \alpha + 1$, then prove that

$$\frac{1}{x}\int_0^x \sum_{n \leq y} f(n) \mathrm{d}y = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=c} L(s,f)\frac{x^s}{s(s+1)}\mathrm{d}s.$$

[*Hint* : Prove a formula for the Perron integral $\int_{\mathrm{Re}(s)=c} w^{s+1}\mathrm{d}s/(s(s+1))$, where $w > 0$.]

(b) If $\mathrm{Re}(s) > \alpha + 1$, then prove that

$$\sum_{n \leq x} f(n)\log(x/n) = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=c} L(s,f)\frac{x^s}{s^2}\mathrm{d}s.$$

[*Hint* : Prove a formula for the Perron integral $\int_{\mathrm{Re}(s)=c} w^{s+1}\mathrm{d}s/s^2$, where $w > 0$.]

Exercise 4.3.4. Let $g : \mathbb{R} \to \mathbb{R}$ be a function in the Schwarz class and define its Mellin transform

$$G(s) = \int_0^{\infty} g(y) y^{-s+1} \mathrm{d}y.$$

(a) Prove that G defines an analytic for $\mathrm{Re}(s) > 0$.

(b) Fix $\delta, A > 0$. Prove that if $\mathrm{Re}(s) > \delta$, then $G(s) \ll_{\delta,A,g} 1/(1+|s|^A)$. [*Hint :* integration by parts.]

(c) Prove that G has a meromorphic continuation to $\mathbb{C}$ with the only possible poles at $s = 0, -1, -2, \ldots$. Moreover, prove that if $\mathrm{supp}(g) \subset (0, +\infty)$, then prove that G has an analytic continuation to $\mathbb{C}$.

(d) Prove that if $c > 0$ and $y > 0$, then

$$g(y) = \int_{\mathrm{Re}(s)=c} G(s) y^s \mathrm{d}y.$$

[*Hint :* Use Fourier inversion.]

(e) If $f : \mathbb{N} \to \mathbb{C}$ satisfies (4.3.1), $x \geq 1$ and $c > \alpha + 1$, then prove that

$$\sum_{n=1}^{\infty} f(n) g(n/x) = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=c} L(s,f) G(s) y^s \mathrm{d}s.$$

Exercise 4.3.5. (a) For any $k \geq 2$ and $x \geq 1$, prove that

$$\mathrm{Res}_{s=1}\left(\zeta(s)^k \frac{x^s}{s}\right) = P_k(\log x),$$

where P_k is a polynomial of degree $k-1$.

(b) Prove that the above polynomial is the same polynomial as in Exercise 1.3.4.

Exercise 4.3.6. An integer n is called square-full if $p^2|n$ whenever $p|n$. Let f denote the characteristic function of such integers.

(a) Show that

$$f(n) = \sum_{a^2b^3=n} \mu^2(b).$$

[*Hint:* Exercise 4.1.6.]

(b) Show that there are constants $c_1, c_2 > 0$ such that

$$\#\{n \leq x : n \text{ square-full}\} = c_1 x^{1/2} + c_2 x^{1/3} + O_\epsilon(x^{1/4+\epsilon}) \quad (x \geq 1),$$

for any fixed $\epsilon > 0$.

Exercise 4.3.7. Show that there is a quadratic polynomial P and some $\delta > 0$ such that

$$\sum_{n \leq x} 2^{\Omega(n)} = x \cdot P(\log x) + O(x^{1-\delta}) \quad (x \geq 1).$$

4.4 The prime number theorem

The goal of this section is to establish the Prime Number Theorem in the following form:

Theorem 4.4.1. *There is an absolute constant $c > 0$ such that*

$$\pi(x) = \mathrm{li}(x) + O(xe^{-c\sqrt{\log x}}).$$

By the discussion in the end of Section 1.1 and Exercise 2.1.1, we may instead show that

$$\sum_{n \leq x} \Lambda(n) = x + O(xe^{-c\sqrt{\log x}}).$$

Exercise 4.1.4 implies that the Dirichlet series of the von Mangoldt function equals $-\zeta'/\zeta$. Thus Theorem 4.3.1 yields the formula

$$\sum_{n \leq x} \Lambda(n) = \frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s)=c+1+1/\log x \\ |\mathrm{Im}(s)| \leq T}} \left(-\frac{\zeta'}{\zeta}(s)\right) \frac{x^s}{s} \mathrm{d}s + O\left(\frac{x(\log x)^2}{T} + \log x\right) \tag{4.4.1}$$

for all $x, T \geq 2$, where we used the simple fact that $|\Lambda(n)| \leq \log n$. The meromorphic continuation of $\zeta(s)$ to the half-plane $\mathrm{Re}(s) > 0$ implies that $-\zeta'/\zeta$ also has a meromorphic continuation to this plane. The pole of ζ at $s = 1$ induces a pole of $-\zeta'/\zeta$ at $s = 1$ of residue 1. All other poles of $-\zeta'/\zeta$ arise from potential zeroes of ζ. Notice that if we shift the contour in (4.4.1) to the left, the pole at $s = 1$ will contribute the expected main term x. Our task thus reduces to proving a zero-free region for ζ and bounds on its logarithmic derivative ζ'/ζ. Precisely, we will prove the following result:

Theorem 4.4.2. *Let $s = \sigma + it$ and $c > 0$ be the constant from Lemma 4.4.7. For $\sigma \geq 1 - c/\log(|t| + 100)$, we have that*

$$\frac{\zeta'}{\zeta}(s) = -\frac{1}{s-1} + O(\log(|t| + 2)).$$

Given the above result, proving the prime number theorem is quite easy:

Proof of Theorem 4.4.1. We start by applying (4.4.1) and then shift the contour to the line $\sigma = 1 - c_1/\log(T + 100)$, where c_1 is the constant c of Corollary (4.4.2). The residue contribution from the pole at $s = 1$ is x, and the complementary contour contributes $\ll x^{1-c_1/\log(T+100)} \log^2 T + x/T$. Thus

$$\sum_{n \leq x} \Lambda(n) = x + O\left(x^{1-c_1/\log(T+100)} \log^2 T + \frac{x \log^2 x}{T} \right).$$

Choosing $T = e^{\sqrt{\log x}}$ completes the proof. □

It remains to establish the key Theorem 4.4.2. By (4.2.2), we immediately see that

$$\zeta^{(j)}(s) = \frac{(-1)^j}{(s-1)^{j+1}} + O(1)$$

when $|s - 1| \leq 1/2$. Thus

$$\frac{\zeta'}{\zeta}(s) = -\frac{1}{s-1} + O(1)$$

provided that $|s - 1| \leq \epsilon$ for a small enough $\epsilon > 0$. So in order to prove (4.4.2), it suffices to consider the case when $|s - 1| \geq \epsilon$, in which case we need to prove that

$$\frac{\zeta'}{\zeta}(s) \ll \log(|t| + 2). \tag{4.4.2}$$

The proof of (4.4.2) is two-fold: we will prove an upper bound on ζ' and a lower bound on ζ. But first, we perform a technical manoeuvre and consider the logarithmic derivative of the sifted zeta function instead

$$\zeta_y(s) := \sum_{P^-(n) > y} \frac{1}{n^s} = \prod_{p > y} \left(1 - \frac{1}{p^s}\right)^{-1} = \zeta(s) \prod_{p \leq y} \left(1 - \frac{1}{p^s}\right).$$

We will eventually take $y = |t| + 100$. The reason why we work with ζ_y instead of ζ is the same with the reason why in Theorem 4.2.2 we applied partial summation only to the tails of $\sum_{n=1}^{\infty} a_n/n^s$: partial summation works well only for $n > |s| \sim |t|+100$. Considering the sifted version of ζ accomplishes this, while at the same preserving the multiplicative properties of ζ. This is the reason why the truncation to integers $n > y$ is performed via a sieve and not via a direct truncation of the sum; the reader can think of this as a multiplicative vs. an archimedian truncation.

Now, note that logarithmic differentiation implies that

$$\frac{\zeta'}{\zeta}(s) = \frac{\zeta_y'}{\zeta_y}(s) - \sum_{P^+(n)\le y} \frac{\Lambda(n)}{n^s}.$$

Moreover, if $s = \sigma + it$ with $\sigma \ge 1 - 1/\log y$, then

$$\left|\sum_{P^+(n)\le y} \frac{\Lambda(n)}{n^s}\right| \le \sum_{p\le y}\sum_{m\ge 1} \frac{\log p}{p^{m(1-1/\log y)}} = \sum_{p\le y} \frac{\log p}{p^{1-1/\log y}-1} \ll \sum_{p\le y} \frac{\log p}{p} \ll \log y. \tag{4.4.3}$$

Hence, (4.4.2) will follow if we can show that

$$\frac{\zeta_y'}{\zeta_y}(s) \ll \log y \quad (|t|\ge \epsilon,\ y \ge |t|+100,\ \sigma \ge 1 - c/\log y). \tag{4.4.4}$$

We will establish that

$$\zeta_y'(s) \ll \log y \quad (|t|\ge \epsilon,\ y \ge |t|+100,\ \sigma \ge 1 - 1/\log y) \tag{4.4.5}$$

and that

$$|\zeta_y(s)| \asymp 1 \quad (|t|\ge \epsilon,\ y \ge |t|+100,\ \sigma \ge 1 - c/\log y). \tag{4.4.6}$$

These bounds are clearly sufficient to deduce (4.4.2) and hence complete the proof of Theorems 4.4.2 and 4.4.1.

The following lemma establishes (4.4.5) and more.

Lemma 4.4.3. *Let $s = \sigma + it$ and $y \ge 100 + |t|$. For $j \in \{0,1\}$ and $\sigma \ge 1 - 1/\log y$, we have that*

$$\zeta_y^{(j)}(s) = \frac{(-1)^j}{(s-1)^{j+1}} \prod_{p\le y}\left(1-\frac{1}{p}\right) + O((\log y)^j).$$

Proof. It suffices to prove that

$$\sum_{\substack{n\le x\\ P^-(n)>y}} n^{-it} = \frac{x^{1-it}}{1-it}\prod_{p\le y}\left(1-\frac{1}{p}\right) + O\left(\frac{x^{1-2/\log y}}{\log y}\right). \tag{4.4.7}$$

for $x \ge y^4$; the lemma will then follow by partial summation. We use the Fundamental Lemma of Sieve Methods (cf. Lemma 3.2.1) to construct upper and lower sieve weights $(\mu^{\pm}(d))_{d\ge 1}$ supported on the set $\{d \le \sqrt{x} : d | P(y)\}$. Then

$$\begin{aligned}\sum_{\substack{n\le x\\ P^-(n)>y}} n^{-it} &= \sum_{n\le x} (1*\mu^+)(n) n^{-it} + O\left(\sum_{n\le x}(1*\mu^+ - 1*\mu^-)(n)\right)\\ &= \sum_{d\le\sqrt{x}} \mu^+(d) d^{-it} \sum_{m\le x/d} m^{-it} + O\left(\frac{x^{1-2/\log y}}{\log y}\right).\end{aligned}$$

We apply partial summation to the sum over m to find that

$$\begin{aligned}\sum_{m\le M} m^{-it} = 1 + \int_1^M u^{-it}\mathrm{d}(u-\{u\}) &= 1 + \int_1^M u^{-it}\mathrm{d}u - \{M\}M^{-it} + it\int_1^M u^{-it-1}\mathrm{d}u \\ &= \frac{M^{1-it}-1}{1-it} + O(1+|t|\log M) \\ &= \frac{M^{1-it}}{1-it} + O(1+|t|\log M).\end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{d\le\sqrt{x}} \mu^+(d)d^{-it}\sum_{m\le x/d} m^{-it} &= \frac{x^{1-it}}{1-it}\sum_{d\le\sqrt{x}}\frac{\mu^+(d)}{d} + O((1+|t|)\sqrt{x}\log x) \\ &= \frac{x^{1-it}}{1-it}(1-O(x^{-2/\log y}))\prod_{p\le y}\left(1-\frac{1}{p}\right) + O((1+|t|)\sqrt{x}\log x) \\ &= \frac{x^{1-it}}{1-it}\prod_{p\le y}\left(1-\frac{1}{p}\right) + O\left(\frac{x^{1-2/\log y}}{\log y} + (1+|t|)\sqrt{x}\log x\right).\end{aligned}$$

Finally, note that $x \ge y^4 \ge |t|^4$, so that the last error term is admissible. This completes the proof of (4.4.7) and hence of the lemma. □

The proof of (4.4.6) is more difficult. We first establish a series of results that will allow us to determine the effect of an abnormally small value of $\zeta_y(s)$. In particular, in Theorem 4.4.5 below, we will prove that if $|\zeta_y(1+it)|$ is abnormally small, then $p^{it} \approx -1$ for many small primes p. But then $p^{2it} \approx 1$ for many small primes p, thus forcing $|\zeta_y(1+2it)|$ to be abnormally large. This will lead to a contradiction, unless (4.4.6) is satisfied. We give the details below, starting with an auxiliary result. In its statement and for the rest of this section, we use the notation

$$\mathbb{U} := \{z \in \mathbb{C} : |z| \le 1\}.$$

Lemma 4.4.4. *Let $f : \mathbb{N} \to \mathbb{U}$ be a completely multiplicative function with sifted Dirichlet series $L_y(s,f) = \sum_{P^-(n)>y} f(n)/n^s$. Consider $y \ge 2$ and $s = \sigma + it$ with $\sigma > 1$. If $\sigma \ge 1 + 1/\log y$, then $|L_y(s,f)| \asymp 1$, whereas if $\sigma \le 1 + 1/\log y$ and we write $\sigma = 1 + 1/\log x$ with $x \ge y$, then*

$$\log L_y(s,f) = \sum_{y<p\le x}\frac{f(p)}{p^{1+it}} + O(1).$$

Proof. We have that

$$\log L_y(s,f) = \log\prod_{p>y}\left(1-\frac{f(p)}{p^s}\right)^{-1} = \sum_{p>y}\sum_{m=1}^{\infty}\frac{f(p)^m}{mp^{ms}} = \sum_{p>y}\frac{f(p)}{p^s} + O(1) \tag{4.4.8}$$

by our assumption that $|f| \le 1$. Note that

$$\sum_{p>e^{1/(1-\sigma)}}\frac{1}{p^\sigma} \ll 1 \tag{4.4.9}$$

by Chebyshev's estimate. This proves the lemma when $\sigma \geq 1+1/\log y$. On the other hand, if $\sigma = 1+1/\log x \leq 1+1/\log y$, then (4.4.8) and (4.4.9) imply that

$$\log L_y(s,f) = \sum_{y<p\leq x} \frac{f(p)}{p^{1+1/\log x+it}} + O(1).$$

Finally, observe that $p^{1/\log x} = 1+O(\log p/\log x)$ for $p \leq x$, and recall that $\sum_{p\leq x}(\log p)/p \ll \log x$ to complete the proof of the lemma. □

The conditions in the following lemma are motivated by the case when $f(n) = n^{-it}$ with $|t| \gg 1/\log y$, in which case $L_y(\sigma, f) = \zeta_y(\sigma+it) \ll (\log y)^j$ by Lemma 4.4.3.

Theorem 4.4.5. *Let $f : \mathbb{N} \to \mathbb{U}$ be a completely multiplicative function whose Dirichlet series $L(s,f)$ is continuously differentiable for $\sigma \in [1,+\infty)$. If $y \geq 2$ is such that $|L_y^{(j)}(\sigma,f)| \ll (\log y)^j$ uniformly for $j \in \{0,1\}$ and $\sigma \geq 1$, and we set $Y = y^{1/|L_y(1,f)|}$, then*

$$\sum_{y<p\leq Y} \frac{1+\mathrm{Re}(f(p))}{p} = O(1) \tag{4.4.10}$$

and

$$\sum_{u<p\leq v} \frac{\mathrm{Re}(f(p))}{p} = O(1) \quad (v \geq u \geq Y). \tag{4.4.11}$$

Proof. We may assume without loss of generality that $L_y(1,f) \neq 0$. Indeed, if this is not the case, we consider the functions $f_\epsilon(n) = f(n)n^{-\epsilon}$. Clearly, $L_y(\sigma, f_\epsilon) = L_y(\sigma+\epsilon, f)$, so that f_ϵ also satisfies the hypotheses of the theorem. In addition, $L_y(1,f_\epsilon) = L_y(1+\epsilon,f) \neq 0$, since the Dirichlet series $L_y(s,f)$ is given by an absolutely convergence Euler product that does not vanish in the domain of absolute convergence $\mathrm{Re}(s) > 1$. If we establish the result for f_ϵ, then the result also follows for f by letting $\epsilon \to 0^+$. This proves that we may restrict our attention to the case when $L_y(1,f) \neq 0$.

For $\sigma > 1$, the mean value theorem and our assumption on f imply that

$$L_y(\sigma,f) = L_y(1,f) + (\sigma-1)L_y'(\sigma_1,f) = L_y(1,f) + O((\sigma-1)\log y). \tag{4.4.12}$$

If C is large enough, then we conclude that $|L_y(\sigma,f)| \asymp |L_y(1,f)|$ whenever $\sigma \leq 1 + 1/(C\log Y)$. In particular, $|L_y(1+1/\log u, f)| \asymp |L_y(1+1/\log v, f)|$ for $y \geq u \geq Y_z^C$. Together with Lemma 4.4.4, this completes the proof of (4.4.11).

Next, we prove (4.4.10). For any $x \geq Y$, we have that

$$\sum_{y<p\leq Y} \frac{\mathrm{Re}(f(p))}{p} = \sum_{y<p\leq x} \frac{\mathrm{Re}(f(p))}{p} + O(1) = \log\left|L_y\left(1+\frac{1}{\log x}, g\right)\right| + O(1),$$

by (4.4.11) and Lemma 4.4.4. Letting $x \to \infty$, we deduce that

$$\sum_{y<p\leq Y} \frac{\mathrm{Re}(f(p))}{p} = \log|L_y(1,f)| + O(1) = -\sum_{y<p\leq Y} \frac{1}{p} + O(1),$$

by Mertens's theorem, thus proving (4.4.10). This completes the proof of the theorem. □

Corollary 4.4.6. *Let $f : \mathbb{N} \to \mathbb{U}$ be a completely multiplicative function and $y \geq 2$ such that $L_y(s, f)$ is continuously differentiable for $s \geq 1 - 1/\log y$ and $|L_y^{(j)}(\sigma, f)| \leq c(\log y)^j$ for $j \in \{0, 1\}$, where $c > 0$ is a constant. If $Y = y^{1/|L_y(1,f)|}$, then*

$$|L_y(\sigma, f)| \asymp \begin{cases} |L_y(1, f)| & \text{if } 1 - c/(2\log Y) \leq \sigma \leq 1 + 1/\log Y, \\ (\sigma - 1)\log y & \text{if } 1 + 1/\log Y < \sigma \leq 1 + 1/\log y, \\ 1 & \text{if } \sigma \geq 1 + 1/\log y. \end{cases}$$

Proof. When $\sigma > 1$, the claimed estimate follows immediately by Theorem 4.4.5 and Lemma 4.4.4. Finally, assume that $1 - c_1/(2\log Y) \leq \sigma \leq 1$. Then the mean value theorem and our assumption on f imply that

$$L_y(\sigma, f) - L_y(1, f) + (\sigma - 1)L_y'(\sigma_1, f) \tag{4.4.13}$$

for some $\sigma_1 \in [\sigma, 1]$, whence

$$|L_y(\sigma, f) - L_y(1, f)| \leq |1 - \sigma| c \log y \leq \frac{\log y}{2\log Y} = \frac{|L_y(1, f)|}{2}.$$

This completes the proof of the corollary. □

We are finally ready to establish (4.4.6) and complete the proof of the Prime Number Theorem:

Lemma 4.4.7. *Fix $\delta > 0$. There is a constant $c = c(\delta) \in (0, 1]$ such that if $s = \sigma + it$ and $y \geq 100 + |t|$ with $|t| \geq \delta/\log y$ and $\sigma \geq 1 - c/\log y$, then $|\zeta_y(s)| \asymp_\delta 1$.*

Proof. Let $Y = y^{1/|\zeta_y(1+it)|}$. Since $\zeta_y^{(j)}(\sigma' + it) \ll_\delta (\log y)^j$ for $\sigma' > 1 - 1/\log y$ by Lemma 4.4.3, we may apply Theorem 4.4.5 to the function $f(n) = n^{-it}$ to find that

$$|\zeta_y(\sigma + it)| \asymp \begin{cases} |\zeta_y(1 + it)| & \text{if } 1 - c_1/\log Y \leq \sigma \leq 1 + 1/\log Y, \\ ((\sigma - 1)\log y)^{-1} & \text{if } 1 + 1/\log Y < \sigma \leq 1 + 1/\log y, \\ 1 & \text{if } \sigma \geq 1 + 1/\log y. \end{cases}$$

for some $c_1 = c_1(\delta)$. Thus, the lemma will follow if we can prove that $|\zeta_y(1 + it)| \asymp_\delta 1$. It suffices to show that $Y \leq y^{O_\delta(1)}$.

Consider the distance function

$$\mathbb{D}(f, g; y, Y)^2 := \sum_{y < p \leq Y} \frac{|f(p) - g(p)|^2}{p} \tag{4.4.14}$$

and note that

$$\mathbb{D}(1, \mu(n)n^{it}; y, Y)^2 = \sum_{y < p \leq Y} \frac{2(1 + \mathrm{Re}(p^{-it}))}{p} \ll_\delta 1$$

by Lemma 4.4.5. Minkowski's inequality then implies that

$$\mathbb{D}(1, n^{2it}; y, Y) = \mathbb{D}(n^{-it}, n^{it}; y, Y) \leq \mathbb{D}(n^{-it}, \mu(n); y, Y) + \mathbb{D}(\mu(n), n^{it}; y, Y) \ll 1.$$

On the other hand,

$$\begin{aligned}\mathbb{D}(1, n^{2it}; y, Y)^2 &= \sum_{y<p\leq Y} \frac{2(1 - \mathrm{Re}(p^{-2it}))}{p} \\ &= 2\log\frac{\log Y}{\log y} - 2\log\left|\zeta_y\left(1 + 2it + \frac{1}{\log Y}\right)\right| + O(1) \\ &\geq 2\log\frac{\log Y}{\log y} - O_\delta(1)\end{aligned}$$

by Lemmas 4.4.4 and 4.4.3. We thus deduce that $Y \leq y^{O_\delta(1)}$, which completes the proof of the lemma. □

Exercises

Exercise 4.4.1. Prove that there is some constant $c > 0$ such that

$$\sum_{n\leq x} \mu(n) \ll xe^{-c\sqrt{\log x}} \quad (x \geq 2).$$

Chapter 5

Dirichlet characters

In this chapter we develop the theory of *Dirichlet characters* that play a key role in the study of primes in arithmetic progressions. Our goal is to decompose the characteristic function of integers n that are in the reduced congruence class $a \pmod q$ as a sum of certain multiplicative functions, called Dirichlet characters. This will allow us to study primes in arithmetic progressions using tools from the theory of multiplicative functions similar to the ones used to prove the Prime Number Theorem.

Precisely, a *Dirichlet character modulo some integer* q is a completely multiplicative function $\chi : \mathbb{N} \to \mathbb{C}$ 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 $\widetilde{\chi}(n \pmod q) = \chi(n)$. In group theoretic terms, $\widetilde{\chi}$ is a character of the abelian group $(\mathbb{Z}/q\mathbb{Z})^\times$. We will sometimes refer to $\widetilde{\chi}$ too as a Dirichlet character.

As we will see, there are $\varphi(q)$ Dirichlet characters mod q and we have the fundamental relation

$$\mathbf{1}_{n \equiv a \pmod q} = \frac{1}{\varphi(q)} \sum_{\chi \pmod q} \overline{\chi}(a) \cdot \chi(n), \tag{5.0.1}$$

which provides the alleged decomposition of the characteristic function of the congruence class $a \pmod q$.

5.1 Fourier analysis on finite abelian groups

We begin with a general discussion of the character theory of finite abelian groups. Let $(G, \cdot)$ be such a group and write $\mathcal{C}(G)$ for the set of all its characters, that is to say, the set of group homomorphisms $\chi : G \to \mathbb{C} \setminus \{0\}$. The set $\mathcal{C}(G)$ forms naturally a group with respect to the usual multiplication of complex valued functions. Its identity element is the

constant function 1 and it called the *principal character*. Every other character is called *non-principal*. Next, we note that

$$|\mathcal{C}(G)| = |G|.$$

This relation is obvious if G is cyclic: if $G = \mathbb{Z}/d\mathbb{Z}$, then every character is uniquely determined by its value at 1. Since $\chi(1)^d = \chi(d \cdot 1) = \chi(0) = 1$, it follows that $\chi(1)$ has to be a d-th root of unity, and consequently, there are precisely d characters. In the general case of a finite abelian group, the relation $|\mathcal{C}(G)| = |G|$ follows by writing G as the direct product of cyclic groups, say

$$G \cong \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z}, \tag{5.1.1}$$

and by the following lemma.

Lemma 5.1.1. *If $(G, \cdot), (G_1, \cdot)$ and $(G_2, \cdot)$ are abelian groups such that $G = G_1 \times G_2$, then the function*

$$\begin{aligned} \rho : \mathcal{C}(G_1) \times \mathcal{C}(G_2) &\longrightarrow \mathcal{C}(G) \\ (\chi_1, \chi_2) &\longrightarrow \rho_{\chi_1,\chi_2}, \\ \text{where} \quad \rho_{\chi_1,\chi_2}(g_1, g_2) &:= \chi_1(g_1)\chi_2(g_2), \end{aligned}$$

is a group isomorphism.

Proof. First of all, we note that ρ is well-defined: it is clear that ρ_{χ_1,χ_2} is a character of the group G. It is also easy to check that it is a group homomorphism. It remains to check that ρ is an isomorphism.

The function ρ is surjective: given $\chi \in \mathcal{C}(G)$, we define $\chi_1(g_1) := \chi((g_1, 1))$ and $\chi_2(g_2) := \chi((1, g_2))$. It is the easy to check that $\chi = \rho_{\chi_1,\chi_2}$.

The function ρ is injective: if $\rho_{\chi_1,\chi_2} = 1$, then $\chi_1(g_1)\chi_1(g_2) = 1$ for all $g_1 \in G_1$ and all $g_2 \in G_2$. In particular, $\chi_1(g_1)\chi_2(1) = 1$. Since $\chi_2(1) = 1$, we deduce that $\chi_1(g_1) = 1$ for all $g_1 \in G_1$. The proof that $\chi_2(g_2) = 1$ for all $g_2 \in G_2$ is similar. □

The most fundamental property of characters is that they satisfy the following orthogonality relations.

Theorem 5.1.2. *Let $(G, \cdot)$ be a finite abelian group. For every $\chi \in \mathbb{C}(G)$, we have that*

$$\frac{1}{|G|} \sum_{g \in G} \chi(g) = \begin{cases} 1 & \text{if } \chi = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1.2}$$

Also, for every $g \in G$, we have that

$$\frac{1}{|G|} \sum_{\chi \in \mathcal{C}(G)} \chi(g) = \begin{cases} 1 & \text{if } g = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1.3}$$

Proof. First, we prove relation (5.1.2). If $\chi = 1$, then (5.1.2) is trivially true. Now assume that $\chi \neq 1$. For every $h \in G$, we have that $hG = G$. So we have that

$$\chi(h) \sum_{g \in G} \chi(g) = \sum_{g \in G}^{q} \chi(hg) = \sum_{g \in G} \chi(g). \tag{5.1.4}$$

Since $\chi \neq 1$, there must be some $g \in G$ with $\chi(g) \neq 1$. Together with (5.1.4), this completes the proof of (5.1.2).

The proof of relation (5.1.3) is very similar. This relation is trivial when $g = 1$. Now, if $g \neq 1$, then there exists some character $\psi \in \mathcal{C}(G)$ such that $\psi(g) \neq 1$. It suffices to consider the case when $G = \mathbb{Z}/d_1\mathbb{Z} \times \cdots \mathbb{Z}/d_k\mathbb{Z}$. If $(a_1, \ldots, a_k) \neq \mathbf{0}$, then there is j such that $a_j \not\equiv 0 \pmod{d_j}$ and we may take $\psi(m_1, \ldots, m_k) := e^{2\pi i m_j/d_j}$.

Using the constructed character ψ, we find that

$$\psi(g) \sum_{\chi \in \mathcal{C}(G)} \chi(g) = \sum_{\chi \in \mathcal{C}(G)} \psi\chi(g) = \sum_{\chi \in \mathcal{C}(G)} \chi(g). \tag{5.1.5}$$

Since $\psi(g) \neq 1$ by assumption, relation (5.1.5) completes the proof of (5.1.3), and hence of the theorem. □

By (5.1.2), 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

$$\alpha = \sum_{\chi \in \mathcal{C}(G)} \langle \alpha, \chi \rangle_G \cdot \chi. \tag{5.1.6}$$

This relation allows us to do Fourier analysis on the group G.

5.2 Additive and multiplicative characters mod q

We will study in more detail the cases when $G = \mathbb{Z}/q\mathbb{Z}$ and $G = (\mathbb{Z}/q\mathbb{Z})^\times$. The characters of the first group are called *additive characters* mod q. Letting

$$e(x) := e^{2\pi i x}, \tag{5.2.1}$$

a notation we will be using throughout these notes, it is easy to see that the additive characters mod q are the functions $n \to e(an/q)$, indexed by $a \in \{0, 1, \ldots, q-1\}$.

It is also possible to give an explicit description of the set $\mathcal{C}((\mathbb{Z}/q\mathbb{Z})^\times)$ of Dirichlet characters mod q, also called *multiplicative characters*. For simplicity and with a slight abuse of notation, we write $\mathcal{C}(q)$ for the set of Dirichlet characters mod q viewed as functions $\chi : \mathbb{Z} \to \mathbb{C}$, keeping in mind the correspondence between $\mathcal{C}(q)$ and $\mathcal{C}((\mathbb{Z}/q\mathbb{Z})^\times)$. If $q = 2^e p_1^{e_1} \cdots p_k^{e_k}$ is the

prime factorization of q, where $p_1, \dots, p_k$ are distinct odd primes, then the Chinese Remainder Theorem and Lemma 5.1.1 imply that

$$\mathcal{C}(q) = \mathcal{C}(2^e) \cdot \mathcal{C}(p_1^{e_1}) \cdots \mathcal{C}(p_k^{e_k}), \tag{5.2.2}$$

where, given two sets of characters A and B, the notation $A \cdot B$ stands for the set $\{\chi\psi : \chi \in A, \psi \in B\}$. In view of (5.2.2), it remains to describe $\mathcal{C}(p^e)$ for a prime p and $e \geq 1$.

When p is odd, then we know that $(\mathbb{Z}/p^v\mathbb{Z})^\times$ is a cyclic group, that is to say there is some element g such that $(\mathbb{Z}/p^e\mathbb{Z})^\times = \{g^j \pmod{p^v} : j \in \mathbb{Z}\}$, called also a *primitive root* mod p^e. We then define the base g discrete logarithm $\pmod{p^e}$ by the relation

$$n \equiv g^{\log_g(n)} \pmod{p^e}.$$

Clearly, $\log_g(n)$ is well-defined only modulo $\varphi(p^e)$. Then $\mathcal{C}(p^e) = \{\chi_{p,j} : 0 \leq j < \varphi(p^e)\}$, where

$$\chi_{p,j}(n) := e\left(\frac{j \log_g(n)}{\varphi(p^e)}\right) \quad \text{if} \quad p \nmid n,$$

which is well-defined since $e(\cdot)$ is 1-periodic.

When $p = 2$, the situation is more complicated, since

$$(\mathbb{Z}/2^e\mathbb{Z})^\times = \begin{cases} \{1\} & \text{if } e \in \{0,1\}, \\ \{1 \pmod 4, -1 \pmod 4\} \cong \mathbb{Z}/2\mathbb{Z} & \text{if } e = 2, \\ \{\pm 5^j \pmod{2^e} : 0 \leq j < 2^{e-2}\} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{e-2}\mathbb{Z} & \text{if } e \geq 3. \end{cases}$$

From the above relation, we see that $\mathcal{C}(2^e) = \{\chi_{2,0}\}$, where $\chi_{2,0}(n) = \mathbf{1}_{(n,2)=1}$ for all $n \in \mathbb{N}$, the principal character mod 2^e. On the other hand, if $e = 2$, then $\mathcal{C}(2^2) = \{\chi_{2,0}, \chi_{2,1}\}$, where $\chi_{2,0}$ is as above and

$$\chi_{2,1}(n) = \begin{cases} 1 & \text{if } n \equiv 1 \pmod 4, \\ -1 & \text{if } n \equiv 3 \pmod 4, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, if $e \geq 3$, then $\mathcal{C}(2^e) = \{\chi_{2,r,s} : r \in \{0,1\},\ s \in \{0,1,\dots,2^{e-2}-1\}\}$ with

$$\chi_{2,r,s}(n) = e\left(\frac{ra}{2} + \frac{sb}{2^{e-2}}\right), \quad \text{where} \quad n \equiv (-1)^a 5^b \pmod{2^e}.$$

The Fourier transform mod q.

Of particular importance is the interaction between additive and multiplicative characters. To this end, given a q-periodic function $f : \mathbb{Z} \to \mathbb{C}$, we define its Fourier transform mod q

$$\widehat{f}(n) := \frac{1}{q}\sum_{a=1}^{q} f(a) e\left(-\frac{an}{q}\right),$$

which is also a q-periodic function. When $f = \chi$, this analyzes a multiplicative character as a linear combination of additive characters via the formula (5.1.6) with $G = \mathbb{Z}/q\mathbb{Z}$. In order to study the Fourier transform of a Dirichlet character χ, we define its *Gauss sum*

$$\mathcal{G}(\chi) := \sum_{a=1}^{q} \chi(a) e\left(\frac{a}{q}\right).$$

Note that we have the relation

$$\sum_{a=1}^{q} \overline{\chi}(a) e\left(\frac{an}{q}\right) = \mathcal{G}(\overline{\chi})\chi(n) \quad \text{whenever } (n,q) = 1. \tag{5.2.3}$$

Indeed, if $(n,q) = 1$, then n is invertible mod q. Write $\overline{n}$ for a representative of the inverse residue class of $n \pmod q$. Then $\{a\overline{n} : 1 \le a \le q\}$ is a complete set of residues mod q. The q-periodicity and the multiplicativity of χ imply that

$$\sum_{a=1}^{q} \overline{\chi}(a) e\left(\frac{an}{q}\right) = \sum_{a=1}^{q} \overline{\chi}(a\overline{n}) e\left(\frac{a\overline{n}n}{q}\right) = \sum_{a=1}^{q} \overline{\chi}(a)\overline{\chi}(\overline{n}) e\left(\frac{a}{q}\right) = \overline{\chi}(\overline{n})\mathcal{G}(\chi).$$

Finally, note that $\chi(n)\chi(\overline{n}) = \chi(n\overline{n}) = \chi(1) = 1$ by the periodicity of q. Since we also have that $|\chi(\overline{n})| = 1$, we deduce that $\overline{\chi}(\overline{n}) = 1/\chi(\overline{n}) = \chi(n)$, whence (5.2.3) follows.

Relation (5.2.3) can be also be written as

$$\overline{\widehat{\chi}(n)} = \frac{\mathcal{G}(\overline{\chi})}{q}\chi(n), \quad \text{whenever } (n,q) = 1, \tag{5.2.4}$$

which 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 called primitive characters. That is to say, if χ is a primitive Dirichlet character mod q, then we will show that

$$\overline{\widehat{\chi}} = \frac{\mathcal{G}(\overline{\chi})}{q}\chi, \tag{5.2.5}$$

which means that χ is a conjugate eigenvector of the Fourier transform $(\mathrm{mod}\, q)$ with conjugate eigenvalue equal to $\mathcal{G}(\overline{\chi})/q$.

5.3 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 of 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 to q, that is to say, χ is a genuine character 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.

Given a Dirichlet character $\chi \pmod q$, we define its *conductor* as the smallest positive integer $q'|q$ for which there exists a Dirichlet character $\chi' \pmod{q'}$ inducing χ. If the conductor of χ equals q, then χ is called *primitive*.

For example, the principal character $\chi_0 \pmod q$ is induced by the constant function 1 on $\mathbb{Z}$, which the only character mod 1. Therefore, χ_0 is imprimitive unless $q = 1$, in which case it is primitive.

We prove the following useful characterization of primitivity:

Lemma 5.3.1. *Let χ be a Dirichlet character mod q.*

(a) The character χ is imprimitive if, and only if, there is some $q_1|q$ with $1 \le q_1 < q$ such that $\chi(m) = \chi(n)$ whenever $m \equiv n \pmod{q_1}$ and $(mn, q) = 1$.

(b) The character χ is imprimitive if, and only if, there is some $q_1|q$ with $1 \le q_1 < q$ such that $\chi(n) = 1$ whenever $n \equiv 1 \pmod{q_1}$ and $(n, q) = 1$.

Proof. (a) If χ is imprimitive, it is clear that such a q_1 exists. Conversely, assume that there is $q_1|q$ such that $\chi(m) = \chi(n)$ whenever $m \equiv n \pmod{q_1}$ and $(mn, q) = 1$. We define a function $\widetilde{\chi}_1 : (\mathbb{Z}/q_1\mathbb{Z})^\times \to \mathbb{C}$ as follows: if $(a, q_1) = 1$, then there is $k \in \mathbb{Z}$ such that $(a + kq_1, q) = 1$. Indeed, we may locate such a k by using the Chinese Remainder Theorem to solve simultaneously the system of equations $kq_1 \equiv 1 - a \pmod p$, where p runs through all primes that divide q but not q_1. Once we have located such a k, we set $\widetilde{\chi}_1(a \pmod q) = \chi(a + kq_1)$. Our assumption on χ impies that this is a well defined function, which is clearly a character of $(\mathbb{Z}/q_1\mathbb{Z})^\times$ inducing χ. This proves that χ is imprimitive.

(b) If χ is imprimitive, it is clear that such a q_1 exists. Conversely, assume that there is $q_1|q$ such that $\chi(n) = 1$ whenever $n \equiv 1 \pmod{q_1}$ and $(n, q) = 1$. Consider then $m, n \in \mathbb{Z}$ that are coprime to q and congruent to each other mod q_1. Then $k = m\overline{n} \equiv 1 \pmod q$ and $\chi(k) = 1$, whence $\chi(m) = \chi(n)$. The fact that χ is imprimitive then follows by part (a). □

Using the above lemma, we prove the following fundamental property of primitive characters that establishes (5.2.5):

Theorem 5.3.2. *Let χ be a primitive Dirichlet character mod q. Then, for every $n \in \mathbb{N}$, we have that*

$$\chi(n) = \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{a=1}^{q} \overline{\chi}(a) e\left(\frac{an}{q}\right).$$

Proof. When $(n, q) = 1$, this is a consequence of the general Fourier inversion formula for characters, that is to say, relation (5.1.6). Assume now that $(n, q) = d > 1$, in which case

we need to show that

$$\sum_{a=1}^{q} \overline{\chi}(a) e\left(\frac{an}{q}\right) = 0.$$

Write $n = dn_1$ and $q = dq_1$, and note that

$$\sum_{a=1}^{q} \overline{\chi}(a) e\left(\frac{an}{q}\right) = \sum_{j=1}^{d} \sum_{b=1}^{q_1} \overline{\chi}(b + jq_1) e\left(\frac{(b + jq_1)n_1}{q_1}\right) = \sum_{b=1}^{q_1} e\left(\frac{bn_1}{q_1}\right) \sum_{j=1}^{d} \overline{\chi}(b + jq_1).$$

So it suffices to show that

$$\sum_{j=1}^{d} \chi(b + jq_1) = 0, \tag{5.3.1}$$

for all $b \in \{1, 2, \ldots, q_1\}$. Since χ is primitive, Lemma 5.3.1(b) 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 $\{bk + j(1 + kq_1) : 1 \leq j \leq d\}$ run over a complete set of representatives mod d, which implies that

$$\chi(1 + kq_1) \sum_{j=1}^{d} \chi(b + jq_1) = \sum_{j=1}^{d} \chi(b + jq_1).$$

Since $\chi(1 + kq_1) \neq 1$, relation (5.3.1) 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 5.3.3. *Let χ be a primitive Dirichlet character mod q. Then $|\mathcal{G}(\chi)| = \sqrt{q}$.*

Proof. By Theorem 5.3.2, we have that

$$|\mathcal{G}(\chi)|^2 |\chi(n)|^2 = \left| \sum_{a=1}^{q} \chi(a) e\left(\frac{an}{q}\right) \right|^2,$$

for all $n \in \{1, \ldots, q\}$. So

$$\begin{aligned} \varphi(q)|\mathcal{G}(\chi)|^2 &= \sum_{n=1}^{q} \left| \sum_{a=1}^{q} \chi(a) e\left(\frac{an}{q}\right) \right|^2 = \sum_{n=1}^{q} \left(\sum_{a=1}^{q} \chi(a) e\left(\frac{an}{q}\right) \right) \left(\sum_{b=1}^{q} \overline{\chi}(b) e\left(-\frac{bn}{q}\right) \right) \\ &= \sum_{a=1}^{q} \sum_{b=1}^{q} \chi(a)\overline{\chi}(b) \sum_{n=1}^{q} e\left(\frac{(a-b)n}{q}\right) = \sum_{a=1}^{q} |\chi(a)|^2 q = \varphi(q)q. \end{aligned}$$

This completes the proof of the theorem. □

Exercises

Exercise 5.3.1. Show that the function

$$f(q) = \sum_{n \in (\mathbb{Z}/q\mathbb{Z})^*} e(n/q)$$

is multiplicative and deduce that $f = \mu$.

Exercise 5.3.2. If $\chi \pmod q$ is induced by $\chi_1 \pmod{q_1}$, then prove that

$$\mathcal{G}(\chi) = \mu(q/q_1)\chi_1(q/q_1)\mathcal{G}(\chi_1).$$

Exercise 5.3.3. Given two primitive characters $\chi, \psi \pmod q$ and $(a, q) = 1$, show that

$$\sum_{n=1}^{q} \chi(n+a)\overline{\psi}(n) = \frac{\chi(-a)\overline{\psi}(-a)\mathcal{G}(\chi)\mathcal{G}(\overline{\chi}\psi)}{\mathcal{G}(\psi)}.$$

Using this formula and Exercise 5.3.1, deduce that

$$\sum_{n=1}^{q} \chi(n+a)\overline{\chi}(n) = \mu(q).$$

5.4 The Pólya-Vinogradov inequality

Theorem 5.4.1 (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. Since χ is non-principal, we must have that $q > 1$. First, we prove the theorem when χ is primitive. Theorem 5.3.2 and the periodicity of χ imply that

$$\begin{aligned}
\sum_{M < n \le M+N} \chi(n) &= \sum_{M < n \le M+N} \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{\substack{1 \le a \le q \\ (a,q)=1}} \overline{\chi}(a) e\left(\frac{an}{q}\right) \\
&= \sum_{M < n \le M+N} \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{\substack{-q/2 < a \le q/2 \\ (a,q)=1}} \overline{\chi}(a) e\left(\frac{an}{q}\right) \\
&= \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{\substack{-q/2 < a \le q/2 \\ (a,q)=1}} \overline{\chi}(a) \sum_{M < n \le M+N} e\left(\frac{an}{q}\right) \\
&\ll \frac{1}{|\mathcal{G}(\overline{\chi})|} \sum_{\substack{-q/2 < a \le q/2 \\ (a,q)=1}} \frac{1}{|1 - e(a/q)|}.
\end{aligned}$$

So Theorem 5.3.3, and the fact that $|1-e(x)|^2 = 2(1-\cos(2\pi x)) = 4\sin^2(\pi x) \asymp x^2$, 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. □

Exercises

Exercise 5.4.1 (The least quadratic nonresidue). Given an odd prime p, we let n_p be the smallest positive integer such $\left(\frac{n_p}{p}\right) = -1$, where $\left(\frac{n}{p}\right)$ denotes the Legendre symbol mod p. This number is called the *least quadratic nonresidue mod p*.

(a) Prove that n_p is well-defined and that it must be a prime number $\ll \sqrt{p}\log p$.

(b) Note that

$$\sum_{\substack{n\le x\\ p'\mid n \implies p'\le y}} \left(1+\left(\frac{n}{p}\right)\right) = 2\Psi(x,y)$$

for $1 \le y < n_p$ and $x \ge y$, where

$$\Psi(x,y) := \#\{n \le x : p'\mid n \implies p' \le y\},$$

where p' denotes here a generic prime number.

(c) Show that if $y \in (\sqrt{x}, x]$, then

$$\Psi(x,y) = \lfloor x\rfloor - \sum_{y<p'\le x} \left\lfloor \frac{x}{p'}\right\rfloor = x(1-\log u) + O\left(\frac{x}{\log x}\right), \quad \text{where} \quad u = \frac{\log x}{\log y}.$$

(d) Prove that $n_p \ll_\epsilon p^{1/(2\sqrt{e})+\epsilon}$, for any fixed $\epsilon > 0$, which improves the estimate coming from part (a).

Chapter 6

Primes in arithmetic progressions

We now turn our attention to the distribution of primes in arithmetic progressions. Our goal is to obtain an asymptotic formula for the quantity

$$\pi(x;q,a) := \#\{p \le x : p \equiv a \,(\mathrm{mod}\, q)\}$$

when $(a,q) = 1$. The case when $(a,q) > 1$ is not so interesting, since there are finitely many primes $p \equiv a \,(\mathrm{mod}\, q)$ in this case. We expect that primes are equidistributed among the various reduced residue classes $a \,(\mathrm{mod}\, q) \in (\mathbb{Z}/q\mathbb{Z})^\times$, for each q, that is to say

$$\pi(x;q,a) \sim \frac{\mathrm{li}(x)}{\varphi(q)} \tag{6.0.1}$$

for $(a,q) = 1$ and x large enough in terms of q. As a matter of fact, it would be reasonable to guess that as long as $x \ge q^{1+\epsilon}$, so that the arithmetic progressions $a \,(\mathrm{mod}\, q)$ has enough many elements $\le x$, relation (6.0.1) holds. Montgomery has even conjectured that error term is of square-root size, based on certain probabilistic heuristics, that is to say

$$\pi(x;q,a) = \frac{\mathrm{li}(x)}{\varphi(q)} + O_\epsilon\left(x^{1/2+\epsilon}q^{1/2}\right)$$

uniformly for $x, q \ge 1$ and $(a,q) = 1$, where ϵ is fixed and positive. Proving this conjecture is beyond the current technology. In fact, Montgomery's conjecture is even stronger than the Generalized Riemann Hypothesis that we will discuss in Chapter 10. The goal of this chapter is to prove that (6.0.1) holds when $q \le (\log x)^A$. Specifically, we will prove the following estimate, known as *the Siegel-Walfisz theorem*:

Theorem 6.0.1. *Fix $A > 0$. There is an (ineffective[1]) constant c_A such that if $x \ge 2$, $(a,q) = 1$ and $1 \le q \le (\log x)^A$, then*

$$\pi(x;q,a) = \frac{\mathrm{li}(x)}{\varphi(q)} + O(xe^{-c_A\sqrt{\log x}}).$$

[1]This means that the proof cannot produce a computable value for this constant.

We note that (5.0.1) impies that

$$\pi(x;q,a) = \frac{1}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a) \sum_{p \le x} \chi(p) \log p. \tag{6.0.2}$$

Moreover, for the principal character, we have that

$$\sum_{p \le x} \chi_0(p) = \sum_{p \le x} 1 + O(\omega(q)) = \mathrm{li}(x) + O(xe^{-c\sqrt{\log x}} + \log q), \tag{6.0.3}$$

by the Prime Number Theorem. Thus Theorem 6.0.1 will follow if we can prove that

$$\sum_{p \le x} \chi(p) \ll xe^{-c_A\sqrt{\log x}} \quad (2 \le q \le (\log x)^A,\ \chi \ne \chi_0).$$

As in the proof of the Prime Number Theorem, it suffices to prove that

$$\sum_{n \le x} \Lambda(n)\chi(n) \ll xe^{-c_A\sqrt{\log x}} \quad (2 \le q \le (\log x)^A,\ \chi \ne \chi_0), \tag{6.0.4}$$

possibly for a different constant c_A. Using Exercise 4.1.4 and Theorem 4.3.1, we find that

$$\sum_{n \le x} \Lambda(n)\chi(n) = \frac{1}{2\pi i} \int\limits_{\substack{\mathrm{Re}(s) = c+1+1/\log x \\ |\mathrm{Im}(s)| \le T}} \left(-\frac{L'}{L}(s,\chi)\right) \frac{x^s}{s} \mathrm{d}s + O\left(\frac{x(\log x)^2}{T} + \log x\right)$$

for $x, T \ge 2$. Theorem 4.2.1 and the Pólya-Vinogradov imply that $L(s,\chi)$ has an analytic continuation to the half-plane $\mathrm{Re}(s) > 0$. Hence its logarithmic derivative has a meromorphic continuation to this plane. There is no pole at $s = 1$ now, so the only possible singularities arise from zeroes of $L(s,\chi)$. Our task thus reduces to proving an upper bound $L'(s,\chi)$ and a lower bound on $L(s,\chi)$ (which, in particular, will imply it is non-zero in a certain range).

6.1 Bounds on $L(s,\chi)$: the general case

We need to control the logarithmic derivative of $L(s,\chi)$. As in the proof of the Prime Number Theorem, we will consider the sifted version of this Dirichlet series

$$L_y(s,\chi) := \sum_{P^-(n) > y} \frac{\chi(n)}{n^s} = L(s,\chi) \prod_{p \le y} \left(1 - \frac{\chi(p)}{p^s}\right).$$

By (4.4.3), we have that

$$\frac{L'}{L}(s,\chi) = \frac{L_y'}{L_y}(s,\chi) + O(\log y) \tag{6.1.1}$$

for $s = \sigma + it$ with $\sigma \ge 1 - 1/\log y$. As in the proof of Theorem 4.2.2, we will eventually take $y \asymp q(|t|+2)$ in order to make sure we are only summing over large enough integers (see the corresponding discussion for ζ_y). We need upper bounds on $L_y'(s,\chi)$ and lower bounds on $L_y(s,\chi)$. We start with the former, since they are much easier:

Lemma 6.1.1. *Let $\chi \pmod q$ be a non-principal Dirichlet character, $y \geq 2$ and $s = \sigma + it$. If $y \geq q(|t|+100)$ and $\sigma > 1 - 1/\log y$, then*

$$L_y^{(j)}(s,\chi) \ll (\log y)^j.$$

Proof. It suffices to prove that

$$\sum_{\substack{n\leq x\\ P^-(n)>y}} \chi(n)n^{-it} \ll \frac{x^{1-2/\log y}}{\log y} \tag{6.1.2}$$

for $x \geq y^4$; the lemma will then follow by partial summation. As in Lemma 4.4.3, we use the Fundamental Lemma of Sieve Methods (cf. Lemma 3.2.1) to construct upper and lower sieve weights $(\mu^{\pm}(d))_{d\geq 1}$ supported on the set $\{d \leq \sqrt{x} : d|P(y)\}$. Then

$$\begin{aligned}\sum_{\substack{n\leq x\\ P^-(n)>y}} \chi(n)n^{-it} &= \sum_{n\leq x}(1*\mu^+)(n)\chi(n)n^{-it} + O\left(\sum_{n\leq x}(1*\mu^+ - 1*\mu^-)(n)\right)\\ &= \sum_{d\leq\sqrt{x}} \mu^+(d)\chi(d)d^{-it}\sum_{m\leq x/d}\chi(m)m^{-it} + O\left(\frac{x^{1-2/\log y}}{\log y}\right).\end{aligned}$$

For the sum over m, note that

$$\begin{aligned}\sum_{m\leq M}\chi(m)m^{-it} = \int_{1^-}^{M} u^{-it}\mathrm{d}\sum_{m\leq u}\chi(m) &= M^{-it}\sum_{m\leq M}\chi(m) + it\int_1^M u^{-it-1}\sum_{m\leq u}\chi(m)\mathrm{d}u\\ &\ll (1+|t|\log M)\sqrt{q}\log q\end{aligned}$$

by the Pólya-Vinogradov inequality, whence (6.1.2) follows. □

We next prove our lower bound on $L_y(s,\chi)$. As for ζ_y, we use Theorem 4.4.5 to prove that if $|L_y(\sigma+it,\chi)|$ is abnormally small, then it must be the case that $\chi(p)p^{-it} \approx -1$ for many primes p. But then $\chi^2(p)p^{-2it} \approx 1$ for many primes p, thus leading to an abnormally large value of $|L_y(1+2it,\chi^2)|$. According to Lemmas 6.1.1 and 4.4.3, this can only happen when $\chi^2 = 1$ and t is small. Thus our lemma below only deals with the case when either χ is complex or t is not too close to 0. The complementary case is the subject of Section 6.2.

Lemma 6.1.2. *Let χ be a non-principal Dirichlet character modulo q, $s = \sigma + it$ and $y \geq q(|t|+100)$. There is a constant $c > 0$ such that if $\sigma \geq 1 - c/\log y$ and $|t| \geq \mathbf{1}_{\chi^2=\chi_0}/\log y$, then $|L_y(s,\chi)| \asymp 1$.*

Proof. Lemma 6.1.1 implies that the hypotheses of Corollary 4.4.6 hold for the function $f(n) = \chi(n)n^{-it}$. It thus suffices to prove that $|L_y(1+it,\chi)| \asymp 1$. If $Y := y^{1/|L_y(1+it,\chi)|}$, then it is actually sufficient to prove that $Y \leq y^{O(1)}$. Theorem 4.4.5 implies that

$$\mathbb{D}(\chi(n), \mu(n)n^{it}; y, Y) \ll 1,$$

where $\mathbb{D}$ is the distance function defined by (4.4.14). By Minkowski's inequality, we have that

$$\begin{aligned} D(\chi^2(n), n^{2it}; y, Y) &= \mathbb{D}(\chi(n)n^{-it}, \overline{\chi}(n)n^{it}; y, Y) \\ &\le \mathbb{D}(\chi(n)n^{-it}, \mu(n); y, Y) + \mathbb{D}(\mu(n), \overline{\chi(n)}n^{it}; y, Y) \\ &\ll 1. \end{aligned}$$

On the other hand, if $|t| \ge \mathbf{1}_{\chi^2=\chi_0}/\log y$, then

$$\begin{aligned} \mathbb{D}(\chi(n), n^{2it}; y, Y)^2 &= \sum_{y<p\le Y} \frac{1 - \mathrm{Re}(\chi(p)p^{-2it})}{p} \\ &= \log\frac{\log Y}{\log y} - \log\left|L_y\left(1 + 2it + \frac{1}{\log Y}, \chi^2\right)\right| + O(1) \\ &\ge \log\frac{\log Y}{\log y} - O(1), \end{aligned}$$

by Lemmas 4.4.4 and 6.1.1. Comparing the above estimates, we find that $Y \le y^{O(1)}$ as needed. □

Corollary 6.1.3. *Let χ be a non-principal Dirichlet character mod q and $s = \sigma + it$. There is an absolute constant $c > 0$ such that if $\sigma \ge 1 - c/\log(q(|t| + 100))$, then*

$$\frac{L'}{L}(s, \chi) \ll \log(q(|t| + 2)).$$

Proof. The result follows immediately by Lemmas 6.1.1 and 6.1.2. □

6.2 Bounds on $L(s, \chi)$ for real Dirichlet characters

The proof of Lemma 6.1.2 leaves an important gap in the study of real Dirichlet characters: it does not exclude the possibility that $L(s, \chi)$ gets very small when χ is a non-principal Dirichlet character and s is close to 1. According to Theorem 4.4.5, this would happen if $\chi(p) = -1$ for most small primes. The effect of a small value of $L(1, \chi)$ to the size of $L(s, \chi)$ for $s \approx 1$ is recored in the following lemma.

Lemma 6.2.1. *Let $\chi \pmod q$ be a real, non-principal character and $t \in [-1/\log q, 1/\log q]$. Then*

$$|L_q(1 + it, \chi)| \asymp \max\{L_q(1, \chi), |t|\log q\}.$$

Proof. Lemma 6.1.2 implies that $|L_y(1 + it, \chi)| \asymp 1$ when $|t| \ge 1/\log y$. Combining this estimate with Theorem 4.4.5 applied to $f(n) = \chi(n)n^{-it}$, we find that

$$\sum_{e^{1/|t|}<p\le x} \frac{\chi(p)}{p^{1+it}} \ll 1 \quad (x \ge e^{1/|t|}).$$

Together with Lemma 4.4.4, this implies that

$$\log L_q(1+it,\chi) = \lim_{x\to\infty} \log L_q\left(1+\frac{1}{\log x}+it,\chi\right) = \sum_{q<p\le e^{1/|t|}} \frac{\chi(p)}{p^{1+it}} + O(1). \tag{6.2.1}$$

For $p \le e^{1/|t|}$, we have that $p^{-it} = 1 + O(|t|\log p)$, so that

$$\log L_q(1+it,\chi) = \sum_{q<p\le e^{1/|t|}} \frac{\chi(p)}{p} + O(1).$$

by Chebyshev's estimate. If $Q = q^{1/L_q(1,\chi)}$, then Theorem 4.4.5 implies that

$$\log L_q(1+it,\chi) = \sum_{q<p\le \min\{Q,e^{1/|t|}\}} \frac{\chi(p)}{p} + O(1) = -\sum_{q<p\le \min\{Q,e^{1/|t|}\}} \frac{1}{p} + O(1),$$

and the lemma follows. □

We use the aforementioned results to deduce a preliminary version of relation (6.0.3).

Theorem 6.2.2. *Let χ be a non-principal character mod q and write $Q = q^{1/|L_q(1,\chi)|}$. There is an absolute constant $c > 0$ such that*

$$\sum_{n\le x} \chi(n)\Lambda(n) \ll xe^{-c\sqrt{\log x}} + x^{1-c/\log q} + \mathbf{1}_{\chi^2=\chi_0}\cdot x^{1-c/\log Q}.$$

Proof. Clearly, this estimate is trivial when $x < q^2$, so we may assume that $x \ge q^2$. Moreover, if $L(1,\chi) = 0$, in which case χ must be real by Lemma 6.1.2, then $Q = \infty$, so the theorem is again trivial. So we assume that $x \ge q$ and that $L(1,\chi) \ne 0$.

Instead of using Theorem 4.3.1, we will use the result of Exercise 4.3.3(a) in order to ensure faster decay of the complex integrals we are considering. (The readers are encouraged to go over the argument using Theorem 4.3.1 and find out for themselves why this yields a worse estimate.) We have that

$$\int_0^x \sum_{n\le y} \chi(n)\Lambda(n)\mathrm{d}y = \frac{1}{2\pi i}\int_{\mathrm{Re}(s)=2} \left(-\frac{L'}{L}(s,\chi)\right)\frac{x^{s+1}}{s(s+1)}\mathrm{d}s.$$

We then shift the line of integration to the contour C of $s = \sigma + it$ such that

$$\sigma = 1 - \frac{c_1|L_{q(|t|+100)}(1+it,\chi)|}{\log q},$$

where c_1 is a small enough constant. Corollary 4.4.6 implies that C lies inside a zero-free region for $L(s,\chi)$. Moreover, combining the same result with Lemmas 6.1.2 and 6.2.1, we find that for σ as above

$$\begin{aligned} |L_{q(|t|+100)}(\sigma+it,\chi)| &\asymp |L_{q(|t|+100)}(1+it,\chi)| \\ &\asymp \begin{cases} 1 & \text{if } |t| \ge \mathbf{1}_{\chi^2=\chi_0}/\log q, \\ |t|\log q & \text{if } \chi^2=\chi_0 \text{ and } 1/\log Q \le |t| \le 1/\log q, \\ L_q(1,\chi) & \text{if } \chi^2=\chi_0 \text{ and } |t| \le 1/\log Q. \end{cases} \end{aligned} \tag{6.2.2}$$

We then apply Cauchy's residue theorem (we imagine here that the contour shifting is taking place on the Riemann sphere, so the vertical portions are just the point at infinity) to find that

$$\int_2^x \sum_{n\le y} \chi(n)\Lambda(n)\mathrm{d}y = \mathbf{1}_{L(1,\chi)=0}\cdot x^2 + \frac{1}{2\pi i}\int_C \left(-\frac{L'}{L}(s,\chi)\right)\frac{x^{s+1}}{s(s+1)}\mathrm{d}s + O(x^{3/2}(\log x)^2).$$

For $s \in C$, relation (4.4.3) and Lemma 6.1.1 imply that

$$\frac{L'(s,\chi)}{L(s,\chi)} = \frac{L'_{q(|t|+100)}(s,\chi)}{L_{q(|t|+100)}(s,\chi)} + O(\log(q(|t|+100))) \ll \frac{\log(q(|t|+100))}{|L_{q(|t|+100)}(s,\chi)|} + \log(q(|t|+100)).$$

Together with (6.2.2), this implies that the portion of the integral over C with $|t| \ge e^{\sqrt{\log x}}$ contributes $\ll x^2(\log q)/e^{\sqrt{\log x}}$, the portion with $q \le |t| \le e^{\sqrt{\log x}}$ contributes $\ll x^2 e^{-c_2\sqrt{\log x}}$, the portion with $\mathbf{1}_{\chi^2=\chi_0}/\log q \le |t| \le q$ contributes $\ll x^{2-c_3/\log q}$. Moreover, if χ is real, the portion with $1/\log Q \le |t| \le 1/\log q$ contributes

$$\ll \int_{1/\log Q}^{1/\log q} \frac{x^{2-c_3 t}}{t}\mathrm{d}t \ll x^{2-c_3/\log Q}$$

and the portion with $|t| \le 1/\log Q$ contributes

$$\ll \int_0^{1/\log Q} x^{2-c_4/\log Q}(\log Q)\mathrm{d}t \ll x^{2-c_4/\log Q}.$$

Putting these estimates together, we find that

$$\int_0^x \sum_{n\le y} \chi(n)\Lambda(n)\mathrm{d}y \ll x^2 e^{-c_5\sqrt{\log x}} + x^{2-c_5/\log q} + \mathbf{1}_{\chi^2=\chi_0} x^{2-c_5/\log Q}, \tag{6.2.3}$$

Going from (6.2.3) to the stated inequality for $\sum_{n\le y}\chi(n)\Lambda(n)$ is achieved by a discrete differentiation argument: if

$$\delta = \frac{1}{2}\max\left\{e^{-c_5\sqrt{\log x}}, x^{-c_5/\log q}, \mathbf{1}_{\chi^2=\chi_0}x^{-c_5/\log Q}\right\} \in (1/\sqrt{x}, 1/2],$$

then applying (6.2.3) with x and $x - \sqrt{\delta}x$ and taking the difference yields

$$\int_{x-\sqrt{\delta}x}^x \sum_{n\le y} \chi(n)\Lambda(n)\mathrm{d}y \ll \delta x^2.$$

By the Brun-Titchmarsch inequality (cf. Theorem 3.4.1), we have that

$$\left|\sum_{y<n\le x} \chi(n)\Lambda(n)\right| \le \sum_{x-\sqrt{\delta}x<n\le x} \Lambda(n) \ll \sqrt{\delta}x,$$

for all $y \in [x-\sqrt{\delta}x, x]$, whence

$$\int_{x-\sqrt{\delta}x}^{x} \sum_{n\le y} \chi(n)\Lambda(n)\mathrm{d}y = \sqrt{\delta}x \sum_{n\le x} \chi(n)\Lambda(n) + O(\delta x^2).$$

Since the left hand side of the above relation is $\ll \delta x^2$, we deduce that

$$\sum_{n\le x} \chi(n)\Lambda(n) \ll \sqrt{\delta}x,$$

which completes the proof of the theorem with $c = c_5/2$. □

Remark 6.2.1. As Exercise 6.2.1 below shows, improving the constant 2 in Theorem 3.4.1 would imply that $L_q(1,\chi) \gg 1$. have very important consequences. However, there is a fundamental barrier in sieve methods that makes such an improvement very hard and, currently, out of reach. This is called the *parity barrier*, which we explain here.

Let $T(\mathcal{A},\mathcal{P})$ be the upper bound that Selberg's sieve yields for $S(\mathcal{A},\mathcal{P})$ that is to say

$$T(\mathcal{A},\mathcal{P}) = \min_{\lambda_d} w_a \sum_{a\in\mathcal{A}} \left(\sum_{d|a} \lambda_d\right)^2,$$

where the minimum runs over all sequences of real numbers λ_d that are supported on integers $d \le D$ all of whose prime factors lie in $\mathcal{P}$ and such that $\lambda_1 = 1$. For simplicity, we assume that $w_a = 1$ for all a. Selberg proved that there are sets of integers $\mathcal{A}_-$ and $\mathcal{A}_+$, and a set of primes $\mathcal{P}$ for which

$$\frac{S(\mathcal{A}^-,\mathcal{P})}{T(\mathcal{A}^-,\mathcal{P})} = o(1) \quad \text{and} \quad \frac{S(\mathcal{A}^+,\mathcal{P})}{T(\mathcal{A}^+,\mathcal{P})} \sim 2, \tag{6.2.4}$$

at least heuristically. These sets are constructed by setting

$$\mathcal{A}^+ = \{n \le x : \Omega(n) \equiv 1 \,(\mathrm{mod}\, 2)\}, \quad \mathcal{A}^- = \{n \le x : \Omega(n) \equiv 0 \,(\mathrm{mod}\, 2)\}, \quad \mathcal{P} = \{p \le \sqrt{x}\}.$$

Then we have that

$$S(\mathcal{A}^-,\mathcal{P}) = 1 \quad \text{and} \quad S(\mathcal{A}^+,\mathcal{P}) = 1 + \pi(x) - \pi(\sqrt{x}) \sim \frac{x}{\log x}.$$

However, if $d \le x$, then we should have that

$$\begin{aligned}\#\{n \in \mathcal{A}^{\pm} : d|n\} = \sum_{\substack{n\le x\\ d|n}} \frac{1 \mp (-1)^{\Omega(n)}}{2} &= \frac{x}{2d} + O(1) + \frac{(-1)^{\Omega(d)}}{2} \sum_{m\le x/d} (-1)^{\Omega(m)} \\ &= \frac{x}{d} + O\left(\frac{xe^{-c_\epsilon\sqrt{\log(x/d)}}}{d}\right),\end{aligned}$$

where the last equality can be proven using (4.4.7). Taking $D = \sqrt{x}/\log x$, this implies that

$$T(\mathcal{A}^+,\mathcal{P}) \sim T(\mathcal{A}^-,\mathcal{P}) \sim \frac{1}{2} T(\{1,\dots,\lfloor\{\sqrt{}\, x\}, \mathcal{P}) \sim \frac{x}{\log x},$$

where the last estimate follows by Exercise 4.4.1.

Exercises

Exercise 6.2.1. Taking $y = x$ in Theorem 3.4.1, we find that

$$\pi(x; q, a) \leq \frac{(2+\epsilon)x}{\varphi(q)\log x} \quad (x \geq q^{3/\epsilon}),$$

if q is large enough. Prove that if the Brun-Tichmarsch inequality could be improved to

$$\pi(x; q, a) \leq \frac{(2-\epsilon)x}{\varphi(q)\log x} \quad (x \geq q^{L})$$

when q is large enough, for some positive constants $\epsilon > 0$ and $L \geq 2$, then we could deduce that $L_q(1, \chi) \gg 1$ for real non-principal characters.

6.3 Siegel's theorem

In order to prove (6.0.3) and thus Theorem 6.0.1, it remains to prove a lower bound on $L_q(1, \chi)$ for real and non-principal characters χ. We will prove the following result due to Siegel:

Theorem 6.3.1. *Let $\chi \pmod q$ be a real non-principal Dirichlet character. For any fixed $\epsilon > 0$, we have that*

$$L(1, \chi) \gg_\epsilon q^{-\epsilon}.$$

The implied constant is ineffective.

Since

$$L_q(1, \chi) = L(1, \chi) \prod_{p \leq q} \left(1 - \frac{\chi(p)}{p}\right) \gg \frac{L(1, \chi)}{\log q}$$

by Mertens's estimate, it is clear that putting together Theorems 6.2.2 and 6.3.1 completes the proof of (6.0.3), and thus of Theorem 6.0.1.

Let us now proceed to the proof of Theorem 6.3.1. We start with a weak but effective lower bound on $L(1, \chi)$:

Theorem 6.3.2. *If χ is a non-principal real Dirichlet character mod q, then*

$$L(1, \chi) \gg \frac{1}{\sqrt{q}\log^2 q}.$$

Proof. First, note that the Euler-McLaurin summation formula implies that

$$\sum_{n \leq x} \left(1 - \frac{n}{x}\right) = \int_0^x \left(1 - \frac{t}{x}\right) \mathrm{d}t - \frac{1}{x}\int_0^x \{t\}\mathrm{d}t = \frac{x}{2} - \frac{1}{x}\int_0^x \{t\}\mathrm{d}t.$$

Now, consider the sum $\sum_{n \leq x}(1 * \chi)(n)(1 - n/x)$. On the one hand, note that

$$(1 * \chi)(p^{2k}) = 1 + \chi(p) + \chi(p)^2 + \cdots + \chi(p)^{2k-1} + \chi(p)^{2k} \geq 1$$

as well as that

$$(1*\chi)(p^{2k-1}) = 1+\chi(p)+\chi(p)^2+\cdots+\chi(p)^{2k-1} \geq 0.$$

Therefore

$$\sum_{n\leq x}(1*\chi)(n)\left(1-\frac{n}{x}\right) \geq \sum_{m^2\leq x}\left(1-\frac{m^2}{x}\right) \geq \frac{1}{2}\sum_{m\leq\sqrt{x/2}} 1 \geq \frac{\sqrt{x}+O(1)}{4}. \tag{6.3.1}$$

On the other hand, we will obtain an asymptotic evaluation of $\sum_{n\leq x}(1*\chi)(n)(1-n/x)$ in terms of $L(1,\chi)$: we have

$$\begin{aligned}\sum_{n\leq x}(1*\chi)(n)\left(1-\frac{n}{x}\right) &= \sum_{a\leq x}\chi(a)\sum_{b\leq x/a}\left(1-\frac{a}{x/b}\right)\\ &= \sum_{a\leq x}\chi(a)\left(\frac{x}{2a}-\frac{1}{x/a}\int_0^{x/a}\{t\}\mathrm{d}t\right)\\ &= \frac{x}{2}L(1,\chi)-\frac{x}{2}\sum_{a>x}\frac{\chi(a)}{a}-\frac{1}{x}\int_0^x\left(\sum_{a\leq\min\{x,x/t\}}a\chi(a)\right)\{t\}\mathrm{d}t.\end{aligned}$$

Using the Pólya-Vinogradov inequality and partial summation, we find that $\sum_{a>x}\chi(a)/a \ll \sqrt{q}(\log q)/x$ as well as that $\sum_{a\leq y}a\chi(a) \ll y\sqrt{q}\log q$. Thus

$$\sum_{n\leq x}(1*\chi)(n)\left(1-\frac{n}{x}\right) = \frac{xL(1,\chi)}{2}+O\left(\frac{\sqrt{q}\log q}{x}+\sqrt{q}(\log q)(\log x)\right).$$

Comparing this formula with (6.3.1) and taking $x = cq(\log q)^4$ for a large enough constant c completes the proof of the theorem. □

We need one more preliminary result due to Pintz [19]:

Lemma 6.3.3. *Let $f:\mathbb{N}\to\mathbb{R}$ be a multiplicative function such that $1*f\geq 0$. Assume further that there exist parameters $r\in\mathbb{N}$, $\theta\in[0,1)$ and $M\geq 2$ such that $|f|\leq\tau_r$ and*

$$\left|\sum_{n\leq x}f(n)\right| \leq Mx^\theta\log^r(xM) \quad (x\geq 1). \tag{6.3.2}$$

If $L(1-\eta,f)\geq 0$ for some $\eta\in[0,(1-\theta)/(4-2\theta)]$, then $L(1,f)\gg_{c,r}\eta M^{-\frac{\eta(4-2\theta)}{1-\theta}}$.

Proof. All implied constants might depend on θ and r. For any $A\in[1,x]$, Dirichlet's

hyperbola method implies that

$$\begin{aligned}\sum_{n\le x}(1*f)(n) &= \sum_{a\le A} f(a)\sum_{b\le x/a} 1 + \sum_{b\le x/A}\sum_{A<a\le x/b} f(a)\\ &= \sum_{a\le A} f(a)\left(\frac{x}{a}+O(1)\right) + O\left(M\sum_{b\le x/A}(x/b)^{\theta}(\log(Mx/b))^r\right)\\ &= x\sum_{a=1}^{\infty}\frac{f(a)}{a} + O\left((A+MxA^{\theta-1})(\log(Mx))^r\right),\end{aligned}$$

where we used Exercise 1.3.4 and the estimate $\sum_{a>A} f(a)/a \ll MA^{\theta-1}\log^r(MA)$ that follows by (6.3.2) and partial summation. We choose $A = x^{1/(2-\theta)}$ to arrive to the estimate

$$\sum_{n\le x}(1*f)(n) = x\cdot L(1,f) + O(Mx^{1/(2-\theta)}\log^r(xM)) \quad (x\ge 2). \tag{6.3.3}$$

We then follow the argument leading to (4.2.3) with $(1*f)(n)$ in place of a_n, $L(1,f)$ in place of c and $1/(2-\theta)$ in place of θ to deduce that

$$\sum_{n=1}^{N}\frac{(1*f)(n)}{n^{1-\eta}} = \frac{L(1,f)N^{\eta}}{\eta} + L(1-\eta,1*f) + O\left(\frac{M\cdot N^{1/(2-\theta)}}{N^{1-\eta}}\right),$$

for any $N\in\mathbb{N}$ and $0\le\eta\le(1-\theta)/(4-2\theta)$, where we used (6.3.3) to bound the error term. The left hand side is ≥ 1. Moreover,

$$L(1-\eta,1*f) = \zeta(1-\eta)L(1-\eta,f)\le 0$$

by the hypotheses of the lemma and the fact that $\zeta(1-\eta)<0$, something that can be seen by applying Lemma 4.1.5 with $f=1$. We thus conclude that

$$\frac{L(1,f)N^{\eta}}{\eta} \ge 1 - O\left(\frac{M}{N^{\frac{1-\theta}{2-\theta}-\eta}}\right).$$

Taking $N = \lfloor c_1 M^{\frac{4-2\theta}{1-\theta}}\rfloor$ for a large enough constant c_1 completes the proof of the lemma. □

We are now ready to prove Siegel's theorem:

Proof of Theorem 6.3.1. If χ' is the primitive character inducing χ, then

$$L(1,\chi) = \prod_{p\mid q}\left(1-\frac{\chi'(p)}{p}\right)L(1,\chi') \ge \prod_{p\mid q}\left(1-\frac{1}{p}\right)L(1,\chi') \gg \frac{L(1,\chi')}{\log\log q}.$$

This clearly allows us to consider only the case when χ is a primitive character.

Let $\epsilon\in[0,1/100]$. If $L(s,\chi)$ does not vanish in $[1-\epsilon/10,1]$, then $L(1-\epsilon/10,\chi)>0$, by continuity, and thus Lemma 6.3.3, which can be applied with $f=\chi$, $r=1$ and $M=\sqrt{q}$, implies that $L(1,\chi)\gg\epsilon q^{-4\epsilon/10}\gg_{\epsilon} q^{-\epsilon}$. So if there is no character χ for which $L(s,\chi)$ has a zero in $[1-\epsilon/10,1]$, then the theorem follows.

Assume, next, that we are in the opposite case, and let χ_1 be a primitive real character of minimal conductor q_1 such that $L(s,\chi_1)$ vanishes in $[1-\epsilon/10,1]$, say at $1-\eta$. For χ_1, we use the lower bound $L(1,\chi_1) \gg_{q_1} 1$, which follows by Theorem 6.3.2. Since q_1 is fixed here, this is an admissible lower bound (though we can say nothing about the implicit constant).

Consider now a primitive character $\chi \neq \chi_1$ of conductor q. As before, if $q < q_1$, then $L(1,\chi) \gg_\epsilon q^{-\epsilon}$, since $L(s,\chi)$ does not vanish in $[1-\epsilon/10,1]$. Finally, we consider the case $q \geq q_1$. Set $f = \chi_1 * \chi * \chi_1\chi$. Then Dirichlet's hyperbola method and the Pólya-Vinogradov inequality imply that

$$\sum_{n\leq x} f(n) \ll q^{2/3}x^{2/3}(\log x)^{1/3}(\log q)^{2/3}.$$

(We leave the details of this inequality to the reader as an exercise.) We also have that $1 * f \geq 0$, as well as that $L(s,f) = L(s,\chi)L(s,\chi_1)L(s,\chi_1\chi)$, so that $L(1-\eta,f) = 0$. Applying Lemma 6.3.3 with $M = q^{2/3}$, $\theta = 2/3$ and $r = 3$, we find that $L(1,f) \gg \eta q^{-8\eta}$. On the other hand,

$$L(1,f) = L(1,\chi)L(1,\chi_1)L(1,\chi_1\chi) \ll L(1,\chi)L(1,\chi_1)\log q,$$

since $L(1,\chi_1\chi) \ll \log q$ by (10.2.3). Finally, note that the Mean Value Theorem implies that there is some $\sigma \in (1-\eta,1)$ such that

$$L(1,\chi_1) = L(1-\eta,\chi_1) + \eta L'(\sigma,\chi_1) = \eta L'(\sigma,\chi_1) \ll \eta q^{\eta/2}\log^2 q,$$

where the bound $L'(1,\chi_1) \ll \log^2 q$ is proven by following the proof of Theorem 4.2.2 and using the Pólya-Vinogradov inequality. We deduce that

$$L(1,\chi) \gg \frac{1}{q^{8.5\eta}\log^3 q} \geq \frac{1}{q^{8.5\epsilon/10}\log^3 q} \gg_\epsilon \frac{1}{q^\epsilon},$$

thus completing the proof of the theorem. □

Chapter 7

Linnik's theorem

The Siegel-Walfisz theorem provides good quantitative control on $\pi(x;q,a)$ as long as $x \gg_\epsilon \exp\{q^\epsilon\}$. However, as we discussed in the previous chapter, the asymptotic $\pi(x;q,a) \sim \mathrm{li}(x)/\varphi(q)$ should hold in the much wider range $x \geq q^{1+\epsilon}$. In particular, for any fixed $\epsilon > 0$ and large q, each reduced arithmetic progression $a \pmod q$ should contain a prime $p \leq q^{1+\epsilon}$. A landmark result in Analytic Number Theory, proved by Linnik, establishes a weak version of this inequality:

Theorem 7.0.1. *There is an absolute and effectively computable constant $L \geq 1$ such that whenever $(a,q) = 1$, there is a prime $p \leq q^L$ congruent to a mod q.*

The proof of Theorem 7.0.1 is split into two parts: first, we will establish the following estimate that deals with the contribution of all but one Dirichlet characters.

Theorem 7.0.2. *Let $q \geq 4$. If $\chi_1 \pmod q$ is a real, non-principal Dirichlet character with*

$$L_q(1,\chi_1) = \min\{L_q(1,\chi) : \chi \text{ real and non-principal character mod } q\},$$

then

$$\sum_{\substack{y<p\leq z \\ p\equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{1}{\varphi(q)} \sum_{y<p\leq z} \frac{1+\chi_1(ap)}{p} + O\left(\frac{1}{\varphi(q)}\right),$$

for all $z \geq y \geq q^2$ and $(a,q) = 1$.

Clasically, proving results like Theorem 7.0.2 requires access to some very technical and hard results called *log-free zero-density estimates*. However, Granville and Soundararajan discovered a way to replace these estimates by techniques from the theory of multiplicative functions. We follow their argument, while our choice of the logarithmic weights $1/p$ allows us to simplify certain details in the proof.

Assuming Theorem 7.0.2 has been established, Theorem 7.0.1 follows in either of the following cases:

- *Case 1:* $\chi_1(a) = -1$. Then we use the fact that

$$\sum_{y<p\leq z} \frac{\chi_1(p)}{p} = \log L_y\left(1+\frac{1}{\log z}, \chi_1\right) \leq O(1),$$

a consequence of Lemmas 4.4.4 and 6.1.1.

- *Case 2:* $\chi_1(a) = 1$ and $L_q(1, \chi_1) \geq L^{-0.99}$. Then we take $y = q^{L^{0.99}}$ and $z = q^L$ and apply Theorem 4.4.5 to deduce Theorem 7.0.2, provided that L is large enough.

Thus, it remains to consider the case when $\chi_1(a) = 1$ and $L_q(1, \chi_1) \leq L^{-0.99}$. We will use a sieve-theoretic argument due to Iwaniec and Friedlander. First, we state a stronger version of Theorem 4.4.5 that we will prove in Section 7.1:

Lemma 7.0.3. *Let $\chi \pmod q$ be a real, non-principal character. If $Q = q^{1/L_q(1,\chi)}$, then*

$$\sum_{y<p\leq z} \frac{1+\chi(p)}{p} \ll \frac{\log z}{\log Q} + y^{-1/\log q} \quad (q \leq y \leq z \leq Q).$$

The above lemma and our assumption that $L_q(1, \chi) \geq L^{-0.99}$ imply that

$$\sum_{q^{\log L}<p\leq q^{L^{0.49}}} \frac{1+\chi_1(p)}{p} \ll \frac{1}{\sqrt{L}}. \tag{7.0.1}$$

This implies that for almost all $p \in (q^{\log L}, q^{L^{0.49}}]$, we have that $\chi_1(p) = -1$. We then take $x = q^{L^{0.49}}$ and consider the sum

$$P(x) := \sum_{\substack{n<x,\, n\equiv a \pmod q \\ p|n \Rightarrow p\geq\sqrt{x}}} (1 * \chi_1)(n).$$

We want to show that $P(x, \chi_1) > 0$, which clearly implies Theorem 7.0.2. The presence of the weight $(1 * \chi_1)(n)$ means that we have *pre-sifted* all primes p with $\chi_1(p) = -1$, that is to say the vast majority of primes. For this reason, we can obtain an asymptotic formula:

$$P(x) = \frac{2x \cdot L(1,\chi_1)(1+O(L^{-1/2}))}{q} \prod_{\substack{\ell<q^{\log L} \\ \ell|q}} \left(1 - \frac{1+\chi_1(\ell)}{\ell} + \frac{\chi_1(\ell)}{\ell^2}\right) + O(x^{11/12} q^{1/2} \log q). \tag{7.0.2}$$

Since $L(1, \chi_1) \gg 1/(\sqrt{q} \log^2 q)$ by Theorem 6.3.2, this completes the proof of Theorem 7.0.1 in this last case too.

7.1 Exceptional characters

In this section, we study the behaviour of exceptional characters, that is to say characters with an abnormally small value of $L_q(1, \chi)$, and establish Lemma 7.0.3 as well as another auxiliary result.

Proof of Lemma 7.0.3. Since $1 * \chi \geq 0$, we have that

$$\sum_{y<p\leq z} \frac{1+\chi(p)}{p} \leq \sum_{\substack{y<n\leq z \\ P^-(n)>q}} \frac{(1*\chi)(n)}{n}. \tag{7.1.1}$$

So, it suffices to estimate the sifted partials sums of $1 * \chi$. We claim that

$$\sum_{\substack{n\le x\\ P^-(n)>q}} (1*\chi)(n) = x\cdot L_q(1,\chi)\cdot \prod_{p\le q}\left(1-\frac{1}{p}\right) + O\left(\frac{x^{1-1/\log q}}{\log q}\right). \tag{7.1.2}$$

Clearly, inserting the above estimate via partial summation to (7.1.1) suffices to prove the lemma.

Let us now prove (7.1.2). By the hyperbola method, we have that

$$\sum_{\substack{n\le x\\ P^-(n)>q}} (1*\chi)(n) = \sum_{\substack{a\le\sqrt{x}\\ P^-(a)>q}} \chi(a) \sum_{\substack{b\le x/a\\ P^-(b)>q}} 1 + \sum_{\substack{b\le\sqrt{x}\\ P^-(b)>q}} \sum_{\substack{\sqrt{x}<a\le x/b\\ P^-(b)>q}} \chi(b).$$

We apply (6.1.2) to the rightmost sum and we estimate $\#\{b\le x/a : P^-(b)>y\}$ by the Fundamental Lemma of Sieve Methods (cf. Lemma 3.3.1) to deduce that

$$\begin{aligned}\sum_{\substack{n\le x\\ P^-(n)>q}} (1*\chi)(n) &= \sum_{\substack{a\le\sqrt{x}\\ P^-(a)>q}} \chi(a)\cdot\left\{\frac{x}{a}\prod_{p\le q}\left(1-\frac{1}{p}\right) + O\left(\frac{(x/a)^{1-2/\log q}}{\log q}\right)\right\}\\ &\qquad + O\left(\sum_{\substack{b\le\sqrt{x}\\ P^-(b)>q}} \frac{(x/b)^{1-2/\log q}}{\log q}\right)\\ &= x\prod_{p\le q}\left(1-\frac{1}{p}\right)\sum_{\substack{a\le\sqrt{x}\\ P^-(a)>q}}\frac{\chi(a)}{a} + O\left(\frac{x^{1-1/\log q}}{\log q}\right).\end{aligned}$$

Finally, we note that

$$\sum_{\substack{a>\sqrt{x}\\ P^-(a)>q}} \frac{\chi(a)}{a} \ll \frac{x^{1-1/\log q}}{\log q},$$

by (6.1.2) and partial summation. This completes the proof of (7.1.2) and, thus, of Lemma 7.0.3. □

We conclude with the following result that proves that exceptional characters 'repel' one another.

Theorem 7.1.1. *Let $\chi_1 \pmod{q_1}$ and $\chi_2 \pmod{q_2}$ be two real, non-principal characters that are not induced by the same primitive character. If $L_y(1,\chi_2) \ge L_y(1,\chi_1)$ for some $y \ge \max\{q_1,q_2\}$, then $L_y(1,\chi_2) \asymp 1$.*

Proof. Set $Y_j = y^{1/L_y(1,\chi_j)}$, so that $Y_2 \le Y_1$. Recall the definition of the distance function $\mathbb{D}(\cdot,\cdot;\cdot,\cdot)$. Theorem 4.4.5 then implies that

$$\mathbb{D}(\chi_j,\mu;y,Y_j) \ll 1,$$

so that

$$\mathbb{D}(\chi_1, \chi_2; y, Y_2) \ll 1$$

by Minkowski's inequality. On the other hand, Lemmas 4.4.4 and 6.1.1 yield that

$$\begin{aligned}\mathbb{D}(\chi_1, \chi_2; y, Y_2)^2 &= 2 \sum_{y<p\le Y_2} \frac{1-\chi_1\chi_2(p)}{p} \\ &= 2\log\frac{\log Y_2}{\log y} - 2\log L_y\left(1+\frac{1}{\log Y_2}, \chi\right) + O(1) \\ &\ge 2\log\frac{\log Y_2}{\log y} + O(1).\end{aligned}$$

This implies that $\log Y_2 \ll \log y$, from where we deduce our claim that $L_y(1, \chi_2) \gg 1$. □

7.2 All but one

In this section, we prove Theorem 7.0.2. Let χ_1 be as in its statement and set

$$\mathcal{C}_q = \{\chi \,(\mathrm{mod}\, q) : \chi \ne \chi_0, \chi_1\}$$

Theorem 7.1.1 clearly implies that $L_q(1, \chi) \gg 1$ for $\chi \in \mathcal{C}_q$. Applying Lemma 4.4.4 and Theorem 4.4.5, we then find that

$$|L_y(\sigma, \chi)| \asymp 1 \quad (\chi \in \mathcal{C}_q,\ \sigma > 1,\ y \ge q). \tag{7.2.1}$$

Next, we rewrite the sum in the left hand side of Theorem 7.0.2: the Brun-Titchmarsch inequality and the argument leading to Lemma 4.4.4 imply that

$$\varphi(q) \sum_{\substack{y<p\le z\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{1}{p} = O(1) + \varphi(q)\left(\sum_{\substack{P^-(n)>y\\ n\equiv a\,(\mathrm{mod}\, q)}} \frac{\Lambda(n)}{n^{1+1/\log z}\log n} - \sum_{\substack{P^-(n)>y\\ n\equiv a\,(\mathrm{mod}\, q)}} \frac{\Lambda(n)}{n^{1+1/\log y}\log n}\right).$$

Using characters to rewrite the condition $n \equiv a \,(\mathrm{mod}\, q)$, we find that

$$\varphi(q) \sum_{\substack{y<p\le z\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{1}{p} = O(1) + \sum_{\chi\,(\mathrm{mod}\, q)} \overline{\chi}(a)\left(\log L_y\left(1+\frac{1}{\log z}, \chi\right) - \log L_y\left(1+\frac{1}{\log y}, \chi\right)\right).$$

When $\chi \in \{\chi_0, \chi_1\}$, we use Lemma 4.4.4 to find that

$$\log L_y\left(1+\frac{1}{\log z}, \chi\right) - \log L_y\left(1+\frac{1}{\log y}, \chi\right) = \sum_{y<p\le z} \frac{\chi(p)}{p} + O(1);$$

otherwise, we note that

$$\log L_y\left(1+\frac{1}{\log z}, \chi\right) - \log L_y\left(1+\frac{1}{\log y}, \chi\right) = -\int_y^z \frac{L_y'}{L_y}\left(1+\frac{1}{\log u}, \chi\right)\frac{\mathrm{d}u}{u\log^2 u}.$$

Putting the above estimates together, we arrive to the formula

$$\varphi(q) \sum_{\substack{y<p\le z\\ p\equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \sum_{y<p\le z} \frac{1+\chi_1(ap)}{p} - R + O(1),$$

where

$$R := \int_y^z \sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a) \frac{L_y'}{L_y}\left(1+\frac{1}{\log u}, \chi\right) \frac{\mathrm{d}u}{u\log^2 u}.$$

We want to bound R by using the bilinear structure of the summands (we have the product $L_y'(s,\chi)\cdot(1/L_y)(s,\chi)$). We could, for example, apply Cauchy-Schwarz, extend the summation to all $\chi \,(\mathrm{mod}\, q)$ and use Parserval's identity

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{n=1}^N a_n\chi(n)\right|^2 = \varphi(q) \sum_{\substack{1\le n\le N\\ (n,q)=1}} |a_n|^2$$

to bound R. However, we first need to incorporate into our argument the fact that the summation runs over $\chi\in\mathcal{C}_q$ only. In order to do this, we want to re-write R to be a sum of the Dirichlet series of the form $L_y(s,f)L_y(s,g)/L_y(s,\chi)$. Then we will bound from below $L_y(s,\chi)$ using (7.2.1) and apply Cauchy-Scwarz to the other two factors. (This idea goes back to Halász and, in its current formulation, to Elliott.) To insert this extra factor, we note that

$$\begin{aligned}\frac{1}{L_y(1+1/\log u,\chi)} &= \frac{1}{L_y(1+1/\log y,\chi)} + \int_y^u \frac{L_y'}{L_y^2}\left(1+\frac{1}{\log w},\chi\right)\frac{\mathrm{d}w}{w\log^2 w}\\ &= 1 + \sum_{\substack{n>y\\ P^-(n)>y}} \frac{\mu(n)\chi(n)}{n^{1+1/\log y}} + \int_y^u \frac{L_y'}{L_y^2}\left(1+\frac{1}{\log w},\chi\right)\frac{\mathrm{d}w}{w\log^2 w}.\end{aligned}$$

We use the above formula to rewrite the $1/L_y$ factors in the definition of R: we find that

$$\begin{aligned}R = &\int_y^z \sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a) L_y'(1+1/\log u,\chi)\frac{\mathrm{d}u}{u\log^2 u}\\ &+ \int_y^z \sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a) L_y'\left(1+\frac{1}{\log u},\chi\right) \sum_{\substack{n>y\\ P^-(n)>y}} \frac{\mu(n)\chi(n)}{n^{1+1/\log y}}\frac{\mathrm{d}u}{u\log^2 u}\\ &+ \int_{y\le w\le u\le z} \sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a) \frac{L_y'(1+1/\log u,\chi)}{L_y(1+1/\log w,\chi)}\cdot\frac{L_y'(1+1/\log w,\chi)}{L_y(1+1/\log w,\chi)}\cdot\frac{\mathrm{d}u}{u\log^2 u}\cdot\frac{\mathrm{d}w}{w\log^2 w}.\end{aligned}$$

The first summand has no L-functions in the denominator, so it can be dealt with directly: Lemma 6.1.1 and orthogonality imply that

$$\begin{aligned}\sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a) L_y'(1+1/\log u,\chi) &= \sum_{\chi\ne\chi_0} \overline{\chi}(a) L_y'(1+1/\log u,\chi) + O(\log y)\\ &= \varphi(q) \sum_{\substack{n>y,\, P^-(n)>y\\ n\equiv a \,(\mathrm{mod}\, q)}} \frac{\log n}{n^\sigma} - \sum_{\substack{n>y\\ P^-(n)>y}} \frac{\log n}{n^\sigma} + O(\log y).\end{aligned}$$

Applying the Fundamental Lemma twice (cf. Lemma 3.3.1 with $y^u = x^{1/3}$), we find that

$$\begin{aligned}\#\{n \le x : P^-(n) > y,\ n \equiv a \,(\mathrm{mod}\, q)\} &= \frac{x}{\varphi(q)}\left(1 + O(x^{-1/\log y})\right)\prod_{p \le y}\left(1 - \frac{1}{p}\right) + O(x^{1/3}) \\ &= \frac{\#\{n \le x : P^-(n) > y\}}{\varphi(q)} + O\left(\frac{x^{1-1/\log y}}{\varphi(q)\log y} + x^{1/3}\right).\end{aligned}$$

for $x \ge y \ge q$. Together with partial summation, this yields the estimate

$$\sum_{\chi \in \mathcal{C}_q} \overline{\chi}(a) L_y'(1 + 1/\log u, \chi) \ll \log y$$

for $y \ge q^2$, so that

$$\begin{aligned}R = O(1) &+ \int_y^z \sum_{\chi \in \mathcal{C}_q} \overline{\chi}(a) L_y'\left(1 + \frac{1}{\log u}, \chi\right) \sum_{\substack{n > y \\ P^-(n) > y}} \frac{\mu(n)\chi(n)}{n^{1+1/\log y}} \cdot \frac{\mathrm{d}u}{u \log^2 u} \\ &+ \int_{y \le w \le u \le z} \sum_{\chi \in \mathcal{C}_q} \overline{\chi}(a) \frac{L_y'(1 + 1/\log u, \chi)}{L_y(1 + 1/\log w, \chi)} \cdot \frac{L_y'(1 + 1/\log w, \chi)}{L_y(1 + 1/\log w, \chi)} \cdot \frac{\mathrm{d}u}{u \log^2 u} \cdot \frac{\mathrm{d}w}{w \log^2 w}.\end{aligned}$$

Next, we take absolute values and apply (7.2.1) and Cauchy-Schwarz to find that

$$\begin{aligned}R &\ll 1 + \int_y^z \sum_{\chi \in \mathcal{C}_q} \left|L_y'\left(1 + \frac{1}{\log u}, \chi\right)\right| \cdot \left|\sum_{\substack{n > y \\ P^-(n) > y}} \frac{\mu(n)\chi(n)}{n^{1+1/\log y}}\right| \frac{\mathrm{d}u}{u \log^2 u} \\ &\quad + \int_{y \le w \le u \le z} \sum_{\chi \in \mathcal{C}_q} \left|L_y'\left(1 + \frac{1}{\log u}, \chi\right)\right| \cdot \left|L_y'\left(1 + \frac{1}{\log w}, \chi\right)\right| \cdot \frac{\mathrm{d}u}{u \log^2 u} \cdot \frac{\mathrm{d}w}{w \log^2 w} \\ &\le 1 + \int_y^z S\left(1 + \frac{1}{\log u}\right)^{1/2} T\left(1 + \frac{1}{\log y}\right)^{1/2} \frac{\mathrm{d}u}{u \log^2 u} \\ &\quad + \int_{y \le w \le u \le z} S\left(1 + \frac{1}{\log u}\right)^{1/2} S\left(1 + \frac{1}{\log w}\right)^{1/2} \frac{\mathrm{d}u}{u \log^2 u} \cdot \frac{\mathrm{d}w}{w \log^2 w},\end{aligned}$$

where

$$S(\sigma) = \sum_{\chi \in \mathcal{C}_q} |L_y'(\sigma, \chi)|^2 \quad \text{and} \quad T(\sigma) = \sum_{\chi \in \mathcal{C}_q} \left|\sum_{\substack{n > y \\ P^-(n) > y}} \frac{\mu(n)\chi(n)}{n^\sigma}\right|^2.$$

Therefore, Theorem 7.0.2 follows by the following estimate:

Lemma 7.2.1. *(a) For $y \ge q^2 > 1$ and $1 < \sigma \le 1 + 1/\log y$, we have that*

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{\substack{n > y \\ P^-(n) > y}} \frac{\mu(n)\chi(n)}{n^\sigma}\right|^2 \ll \frac{1}{(\sigma - 1)^2 (\log y)^2}$$

and

$$\sum_{\substack{\chi \,(\mathrm{mod}\, q)\\ \chi \neq \chi_0}} |L_y'(\sigma,\chi)|^2 \ll (\log y)^2.$$

Proof. By orthogonality, we have that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{\substack{n>y\\ P^-(n)>y}} \frac{\mu(n)\chi(n)}{n^\sigma} \right|^2 = \varphi(q) \sum_{\substack{P^-(n_1)>y\\ n_1>y}} \frac{\mu(n_1)}{n_1^\sigma} \sum_{\substack{P^-(n_2)>y\\ n_2>y\\ n_2\equiv n_1 \,(\mathrm{mod}\, q)}} \frac{\mu(n_2)}{n_2^\sigma}$$

For $(b,q)=1$, the Fundamental Lemma (cf. Lemma 3.3.1) implies that

$$\#\{n \le x : n \equiv b \,(\mathrm{mod}\, q),\ P^-(n) > y\} \ll \frac{x}{\varphi(q)\log y} \quad (x \ge y \ge q^2). \tag{7.2.2}$$

Together with partial summation, this yields the estimate

$$\sum_{\substack{P^-(n_2)>y,\, n_2>y\\ n_2\equiv n_1 \,(\mathrm{mod}\, q)}} \frac{\mu(n_2)}{n_2^\sigma} \ll \frac{1}{\varphi(q)(\sigma-1)\log y},$$

whence we deduce that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{\substack{n>y\\ P^-(n)>y}} \frac{\mu(n)\chi(n)}{n^\sigma} \right|^2 \ll \frac{1}{(\sigma-1)\log y} \sum_{\substack{n>y\\ P^-(n)>y}} \frac{1}{n^\sigma} \ll \frac{1}{(\sigma-1)^2(\log y)^2},$$

by (7.2.2) with $q=1$ and partial summation.

For the second estimate, we note that

$$-L_y'(\sigma,\chi) = \sum_{\substack{n>y\\ P^-(n)>y}} \frac{\chi(n)\log n}{n^\sigma} = \sum_{k=1}^{\infty} \sum_{\substack{y^k < n \le y^{k+1}\\ P^-(n)>y}} \frac{\chi(n)\log n}{n^\sigma}.$$

Fix k for the moment and let $\mu_k^\pm$ be the sieve weights supplied by the Fundemental Lemma (cf. Lemma 3.2.1) when $D = y^{k/3}$ and set

$$\delta_k(n) = (\mu_k^+ * 1)(n) - \mathbf{1}_{P^-(n)>y}, \quad \text{so that} \quad 0 \le \delta_k \le (\mu_k^+ - \mu_k^-) * 1.$$

We have that

$$\begin{aligned}\sum_{y^k < n \le y^{k+1}} \frac{\chi(n)(\mu_k^+ * 1)(n)\log n}{n^\sigma} &= \sum_{d \le y^{k/2}} \frac{\mu_k^+(d)\chi(d)}{d^\sigma} \sum_{y^k/d < m \le y^{k+1}/d} \frac{\chi(m)\log(dm)}{m^\sigma}\\ &\ll \sum_{d \le y^{k/3}} \frac{1}{d^\sigma} \cdot \frac{\sqrt{q}(\log q)\log(y^k)}{(y^k/d)^\sigma}\\ &\ll \frac{\sqrt{q}(\log q)\log(y^k)}{y^{2k/3}},\end{aligned}$$

by the Pólya-Vinogradov inequality and partial summation. This implies that

$$L'_y(\sigma,\chi) = O\left(\frac{\sqrt{q}(\log q)\log y}{y^{2/3}}\right) + \sum_{k=1}^{\infty}\sum_{y^k<n\le y^{k+1}}\frac{\chi(n)\delta_k(n)\log n}{n^\sigma}.$$

We then apply the Cauchy-Schwarz inequality to find that

$$\begin{aligned}|L'_y(\sigma,\chi)|^2 &\le O\left(\frac{q(\log q)^2(\log y)^2}{y^{4/3}}\right) + 2\left|\sum_{k=1}^{\infty}\frac{1}{k}\cdot k\sum_{y^k<n\le y^{k+1}}\frac{\chi(n)\delta_k(n)\log n}{n^\sigma}\right|^2 \\ &\ll \frac{q(\log q)^2(\log y)^2}{y^{4/3}} + \sum_{k=1}^{\infty}k^2\left|\sum_{y^k<n\le y^{k+1}}\frac{\chi(n)\delta_k(n)\log n}{n^\sigma}\right|^2.\end{aligned} \tag{7.2.3}$$

We apply Parseval's identity to the sum over χ on the right hand side to find that

$$\sum_{\chi\,(\mathrm{mod}\,q)}\left|\sum_{y^k<n\le y^{k+1}}\frac{\chi(n)\delta_k(n)\log n}{n^\sigma}\right|^2 = \varphi(q)\sum_{\substack{y^k<n_1,n_2\le y^{k+1}\\(n_1,q)=1\\ n_2\equiv n_1\,(\mathrm{mod}\,q)}}\frac{\delta_k(n_1)\delta_k(n_2)(\log n_1)(\log n_2)}{(n_1n_2)^\sigma}.$$

For any $(a,q)=1$ and $y^k < N \le y^{k+1}$, we have that

$$\begin{aligned}\sum_{\substack{n\le N\\ n\equiv a\,(\mathrm{mod}\,q)}}\delta_k(n) \le \sum_{\substack{n\le N\\ n\equiv a\,(\mathrm{mod}\,q)}}(\mu_k^+ * 1 - \mu_k^- * 1)(n) &= \sum_{\substack{d\le y^{k/3}\\(d,q)=1}}(\mu_k^+ - \mu_k^-)(d)\sum_{\substack{n\le N\\ n\equiv a\,(\mathrm{mod}\,q)\\ d\mid n}}1 \\ &= \sum_{\substack{d\le y^{k/3}\\(d,q)=1}}(\mu_k^+ - \mu_k^-)(d)\left(\frac{N}{dq}+O(1)\right) \\ &\ll \frac{Ne^{-k}}{\varphi(q)\log y} + y^{k/3},\end{aligned}$$

by the Fundamental Lemma (cf. Lemma 3.2.1). Partial summation then implies that

$$\sum_{\substack{y^k<n\le y^{k+1}\\ n\equiv a\,(\mathrm{mod}\,q)}}\frac{\delta_k(n)\log n}{n^\sigma} \ll \frac{k^2e^{-k}\log y}{\varphi(q)} + \frac{\log(y^k)}{y^{2k/3}},$$

so that

$$\sum_{\chi\,(\mathrm{mod}\,q)}\left|\sum_{y^k<n\le y^{k+1}}\frac{\chi(n)\delta_k(n)\log n}{n^\sigma}\right|^2 \ll \frac{k^4\log^2 y}{e^{2k}} + \frac{q^2\log^2(y^k)}{y^{4k/3}}.$$

Inserting the above estimate into (7.2.3), summing over $k \ge 1$ and recalling that $y \ge q^2$ completes the proof of the lemma. □

7.3 The exceptional case

In this section, we prove relation (7.0.2) and complete the proof of Theorem 7.0.1. Throughout the proof, we reserve the letters p and ℓ to denote prime numbers. We first apply Buchstab's identity (3.2.2) to find that

$$\begin{aligned} P(x,\chi_1) &= \sum_{\substack{n<x,\, n\equiv a \,(\mathrm{mod}\, q)\\ P^-(n)\geq q^{\log L}}} (1*\chi_1)(n) - \sum_{q^{\log L}\leq p<\sqrt{x}} \sum_{\substack{n<x,\, n\equiv a \,(\mathrm{mod}\, q)\\ P^-(n)=p}} (1*\chi_1)(n) \\ &= T(x, q^{\log L}; a) - \sum_{\substack{q^{\log L}\leq p<\sqrt{x}\\ j\geq 1}} (1*\chi_1)(p^j)\cdot T(M/p^j, p; a\overline{p}), \end{aligned} \tag{7.3.1}$$

where

$$T(M,y;b) := \sum_{\substack{m<M,\, m\equiv b \,(\mathrm{mod}\, q)\\ P^-(m)>y}} (1*\chi_1)(m)$$

and $\overline{p}$ denotes the multiplicative inverse of $p \,(\mathrm{mod}\, q)$. We shall apply the Fundamental Lemma to obtain an asymptotic formula for $S(M,y;b)$ when $(b,q)=1$.

We need to estimate

$$T_d(M;b) := \sum_{\substack{m<M\\ m\equiv b \,(\mathrm{mod}\, q)\\ n\equiv 0 \,(\mathrm{mod}\, d)}} (1*\chi_1)(m).$$

Clearly, $T_d(M;b)=0$ if $(d,q)>1$. Assume now that $(d,q)=1$. If we extend χ_1 to all real numbers by setting $\chi_1(\alpha)=0$ when $\alpha\notin\mathbb{Z}$, then

$$(1*\chi_1)(m) = \sum_{gh=m} \chi_1(g) = \sum_{\substack{gh=m\\ g<h}} (\chi_1(g)+\chi_1(h)) + \chi_1(\sqrt{m}).$$

If $m\equiv b \,(\mathrm{mod}\, q)$, then we note that

$$\chi_1(h) = \chi_1(m/g) = \chi_1(m)\chi_1(g) = \chi_1(b)\chi_1(g),$$

whence

$$(1*\chi_1)(m) = (1+\chi_1(b)) \sum_{\substack{gh=m\\ g<h}} \chi_1(g) + \chi_1(\sqrt{m}).$$

Therefore

$$\begin{aligned} T_d(M;b) &= (1+\chi_1(b)) \sum_{\substack{m<M\\ m\equiv b \,(\mathrm{mod}\, q)\\ m\equiv 0 \,(\mathrm{mod}\, d)}} \sum_{\substack{gh=m\\ k<\ell}} \chi_1(g) + \sum_{\substack{r^2<M\\ r^2\equiv b \,(\mathrm{mod}\, q)\\ r^2\equiv 0 \,(\mathrm{mod}\, d)}} \chi(r) \\ &= (1+\chi_1(b)) \sum_{g<\sqrt{M}} \chi_1(g) \sum_{\substack{g<h\leq M/g\\ \ell\equiv b\overline{g} \,(\mathrm{mod}\, q)\\ \ell\equiv 0 \,(\mathrm{mod}\, d/(d,g))}} 1 + O(\sqrt{M}) \\ &= (1+\chi_1(b)) \sum_{g<\sqrt{x}} \chi_1(g)\cdot\left(\frac{M/g-g}{qd/(d,g)} + O(1)\right) + O(\sqrt{M}). \end{aligned}$$

The Pólya-Vinogradov inequality and partial summation imply that

$$\sum_{g<\sqrt{M}} \chi_1(g)\cdot g \ll \sqrt{Mq}\log q.$$

Hence

$$T_d(M;b) = \frac{(1+\chi_1(b))M}{dq}\sum_{g<\sqrt{M}} \frac{\chi_1(k)(d,g)}{g} + O(\sqrt{Mq}\log q).$$

Since

$$(d,g) = \sum_{r|(d,g)} \varphi(r),$$

if we write $g = rk$, we arrive to the estimate

$$T_d(M;b) = \frac{(1+\chi_1(b))M}{dq}\sum_{r|d} \frac{\varphi(r)\chi_1(r)}{r}\sum_{k<\sqrt{M}/r} \frac{\chi_1(k)}{k}.$$

Applying the Pólya-Vinogradov inequality and partial summation again, we find that

$$\sum_{k\geq\sqrt{M}/r} \frac{\chi_1(k)}{k} \ll \frac{r\sqrt{q}\log q}{\sqrt{M}}.$$

We conclude that

$$T_d(M;b) = M\cdot\frac{(1+\chi_1(b))L(1,\chi_1)}{d}\sum_{r|d} \frac{\varphi(r)\chi_1(r)}{r} + O(\sqrt{Mq}\log q).$$

We are now ready to apply the Fundamental Lemma of Sieve Methods, in the form of Lemma 3.3.1. We take $X = (1+\chi_1(b))M\cdot L(1,\chi_1)$, $g(d) = d^{-1}\sum_{r|d}\varphi(r)\chi_1(r)/r$ and $D = x^{1/3}$ to find that

$$T(M,y;b) = M\cdot L(1,\chi_1)(1+\chi_1(b)+O(M^{-1/\log y}))\prod_{\substack{\ell<y\\ \ell\nmid q}}\left(1-\frac{1+\chi_1(\ell)}{\ell}+\frac{\chi_1(\ell)}{\ell^2}\right)$$
$$+ O(M^{5/6}q^{1/2}\log q)$$

for all $M\geq 1$ and $y\geq 2$, where we recall that ℓ denotes a generic prime number. In particular,

$$T(x,q^{\log L};a) = x\cdot L(1,\chi_1)(2+O(1/L))\prod_{\substack{\ell<q^{\log L}\\ \ell\nmid q}}\left(1-\frac{1+\chi_1(\ell)}{\ell}+\frac{\chi_1(\ell)}{\ell^2}\right) + O(x^{5/6}q^{1/2}\log q).$$

Similarly, if $q^{\log L} < p \leq \sqrt{x}$, then

$$T(x/p^j,q^{\log L};\overline{p}a) \ll \frac{x}{p^j}L(1,\chi_1)\prod_{\substack{\ell<p\\ \ell\nmid q}}\left(1-\frac{1+\chi_1(\ell)}{\ell}+\frac{\chi_1(\ell)}{\ell^2}\right) + \frac{x^{5/6}q^{1/2}\log q}{p^{5j/6}}$$
$$\asymp \frac{x}{p^j}L(1,\chi_1)\prod_{\substack{\ell<q^{\log L}\\ \ell\nmid q}}\left(1-\frac{1+\chi_1(\ell)}{\ell}+\frac{\chi_1(\ell)}{\ell^2}\right) + \frac{x^{5/6}q^{1/2}\log q}{p^{5j/6}}$$

where we used (7.0.1). Inserting the above estimates into (7.3.1), we deduce that

$$P(x,\chi_1) = x\cdot L(1,\chi_1)(2+O(\delta+1/L))\prod_{\substack{\ell<q^{\log L}\\ \ell\nmid q}}\left(1-\frac{1+\chi_1(\ell)}{\ell}+\frac{\chi_1(\ell)}{\ell^2}\right)+O(x^{11/12}q^{1/2}\log q),$$

where

$$\delta = \sum_{q^{\log L}<p\le\sqrt{x}}\frac{(1*\chi_1)(p^j)}{p^j} \ll \frac{1}{\sqrt{L}}+\frac{1}{q^{\log L}} \ll \frac{1}{\sqrt{L}}.$$

This completes the proof of (7.0.2) and hence of Theorem 7.0.1.

Chapter 8

The large sieve and the Bombieri-Vinogradov theorem

In this chapter, we develop the theory of the large sieve and we use it to prove the following fundamental result:

Theorem 8.0.1 (Bombieri-Vinogradov)**.** *Fix $A \geq 1$. There is a constant $B = B(A)$ such that if $1 \leq Q \leq x^{1/2}/(\log x)^B$, then*

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

Before we proceed with the theory of the large sieve, we note that the Bombieri-Vinogradov theorem can be deduced from the following character sum estimate:

Theorem 8.0.2 (Bombieri-Vinogradov theorem, II)**.** *Let $1 \leq Q \leq x^{2/3}$. Then*

$$\sum_{Q < q \leq 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y \leq x} \left| \sum_{n \leq y} \Lambda(n)\chi(n) \right| \ll (\log x)^6 (x + x^{1/2}Q^2 + x^{4/5}Q^{13/10}),$$

where the notation $\sideset{}{^*}\sum$ means that the sum runs over primitive characters only.

Deduction of Theorem 8.0.1 from Theorem 8.0.2. Fix $A \geq 1$ and let Q and x be as in Theorem 8.0.1. By (6.0.2) and (6.0.3), we have that

$$\pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} = \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} \overline{\chi}(a) \sum_{p \leq y} \chi(p) + O\left(\frac{y}{\varphi(q) e^{c\sqrt{\log y}}} + \frac{\log q}{\varphi(q)} \right).$$

Moreover, if $\chi' \,(\mathrm{mod}\, d)$, with $d|q$, $d > 1$, is the primitive character inducing $\chi \neq \chi_0$, then arguing as in (6.0.3) implies that

$$\sum_{p \leq y} \chi(p) = \sum_{p \leq y} \chi'(p) + O(\log q).$$

Furthermore, partial summation implies that

$$\begin{aligned}\sum_{p\le y}\chi'(p) &= \int_{2^-}^{y}\frac{1}{\log u}\mathrm{d}\sum_{p\le u}\chi'(p)\log p\\ &= \frac{\sum_{p\le y}\chi(p)\log p}{\log y}+\int_2^y\frac{\sum_{p\le u}\chi(p)\log p}{u\log^2 u}\mathrm{d}u\\ &\ll \max_{u\le y}\left|\sum_{p\le u}\chi(p)\log p\right|.\end{aligned}$$

Since we also have that

$$\sum_{p\le u}\chi(p)\log p=\sum_{n\le u}\Lambda(n)\chi(n)+O\left(\sum_{\substack{p^m\le u\\ 2\le m\le \log u/\log p}}\log p\right)=\sum_{n\le u}\Lambda(n)\chi(n)+O(u^{1/2})$$

by Chebysev's estimate, we deduce that

$$\begin{aligned}\max_{\substack{y\le x\\(a,q)=1}}\left|\pi(y;q,a)-\frac{\mathrm{li}(y)}{\varphi(q)}\right| \le \frac{1}{\varphi(q)}\sum_{d|q,\ d>1}\sideset{}{^*}\sum_{\chi'\,(\mathrm{mod}\,d)}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|\\ +O\left(\frac{y}{\varphi(q)e^{c\sqrt{\log y}}}+\sqrt{x}+\log q\right).\end{aligned}$$

Consequently,

$$\sum_{q\le Q}\max_{\substack{y\le x\\(a,q)=1}}\left|\pi(y;q,a)-\frac{\mathrm{li}(y)}{\varphi(q)}\right|\le\sum_{q\le Q}\frac{1}{\varphi(q)}\sum_{d|q,\ d>1}\sideset{}{^*}\sum_{\chi'\,(\mathrm{mod}\,d)}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|+O\left(\frac{x\log x}{e^{c\sqrt{\log x}}}\right)$$

For each fixed $d>1$ and q a multiple of d, we note that $\varphi(q)\ge\varphi(d)\varphi(q/d)$, so that

$$\sum_{q\le Q}\max_{\substack{y\le x\\(a,q)=1}}\left|\pi(y;q,a)-\frac{\mathrm{li}(y)}{\varphi(q)}\right|\ll\sum_{1<d\le Q}\frac{\log x}{\varphi(d)}\sideset{}{^*}\sum_{\chi'\,(\mathrm{mod}\,d)}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|+\frac{x\log x}{e^{c\sqrt{\log x}}}.$$

When $1<d\le(\log x)^C$ and $\chi'\,(\mathrm{mod}\,d)$ is primitive (in particular, χ' is non-principal), then (6.0.4) implies that

$$\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|=\max_{\sqrt{x}<y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|+O(\sqrt{x})\ll\frac{x}{e^{c'\sqrt{\log x}}},$$

where c' is some appropriate (ineffective) positive constant. Therefore

$$\sum_{q\le Q}\max_{\substack{y\le x\\(a,q)=1}}\left|\pi(y;q,a)-\frac{\mathrm{li}(y)}{\varphi(q)}\right|\ll_{A,C}\sum_{(\log x)^C<d\le Q}\frac{\log x}{\varphi(d)}\sideset{}{^*}\sum_{\chi'\,(\mathrm{mod}\,d)}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|+\frac{x}{(\log x)^A}.$$

We divide the range of summation of d into $O(\log Q)$ intervals of the form $(D, 2D]$, with D a power of 2 and note that $\varphi(d) \gg (\varphi(d)/d)D$ for each $d \in (D, 2D]$. Therefore

$$\begin{aligned}
&\sum_{q\le Q}\max_{\substack{y\le x\\(a,q)=1}}\left|\pi(y;q,a)-\frac{\mathrm{li}(y)}{\varphi(q)}\right|\\
&\ll_{A,C}\sum_{D=2^k\in((\log x)^C/2,Q]}\sum_{D<d\le 2D}\frac{d}{\varphi(d)}\sideset{}{^*}\sum_{\chi' \,(\mathrm{mod}\, d)}\max_{y\le x}\left|\sum_{n\le y}\Lambda(n)\chi'(n)\right|+\frac{x}{(\log x)^A}\\
&\ll\sum_{D=2^k\in((\log x)^C/2,Q]}\frac{(\log x)^8}{D}\cdot(x+x^{1/2}D^2+x^{4/5}D^{13/10})+\frac{x}{(\log x)^A}\\
&\ll\frac{x}{(\log x)^{C-8}}+x^{1/2}(\log x)^8Q+x^{4/5}Q^{3/10}+\frac{x}{(\log x)^A}.
\end{aligned}$$

Since $Q \le x^{1/2}/(\log x)^B$, taking $B = C = A + 8$ then yields Theorem 8.0.1. □

Exercises

Exercise 8.0.1. Use Selberg's sieve in conjunction with the Bombieri-Vinogradov theorem to prove that

$$\#\{p \le x : p + 2 \text{ prime}\} \le (4 + o_{x\to\infty}(1)) \cdot \frac{cx}{\log^2 x},$$

where

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

is the twin prime constant. Compare this upper bound with what was obtained in Exercise 3.4.1.

8.1 Quasi-orthogonality and the large sieve

Fix $N \ge 1$, 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 $e(x) = e^{2\pi ix}$, 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

$$\sum_{n=1}^N e(\alpha n)\overline{e(\beta n)} = \frac{e(\alpha-\beta)-e((N+1)(\alpha-\beta))}{1-e(\alpha-\beta)} \ll \frac{1}{\|\alpha-\beta\|}. \tag{8.1.1}$$

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, no sequence of complex numbers $(a_n)_{n\leq N}$ should be able to correlate strongly with the sequence $(e(n\alpha_r))_{n\leq N}$, for many $r \in \{1, \dots, R\}$. The following theorem confirms this intuition:

Theorem 8.1.1 (Large sieve - trigonometric version)**.** *Let $\{a_n\}_{n=1}^N$ be a sequence of complex numbers and $\{\alpha_1, \dots, \alpha_R\}$ be a set of δ-spaced real numbers. Then*

$$\sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 \leq \left(N + 1 + \frac{\pi\sqrt{6}}{3\delta} \right) \sum_{n=1}^{N} |a_n|^2.$$

Remark 8.1.1. Selberg [24] and, independently, Montgomery and Vaughan [18] showed that

$$\sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 \leq \left(N + \frac{1}{\delta} - 1 \right) \sum_{n=1}^{N} |a_n|^2, \tag{8.1.2}$$

for $\{\alpha_r\}_{r\leq R}$ as above, which is best possible in this generality, as the two examples below indicate:

- If $R = \delta = 1$ and $a_n = e(-n\alpha_1)$ for all n, then both sides of (8.1.2) equal N^2.
- If N is fixed, $\alpha_j = j/R$ for $1 \leq j \leq R$, and $\delta = 1/R$, then the left hand side of (8.1.2) multiplied by δ is a Riemann sum for the integral

$$\int_0^1 \left| \sum_{n=1}^{N} a_n e(n\alpha) \right|^2 \mathrm{d}\alpha = \sum_{1\leq n,m\leq N} a_n \overline{a_m} \int_0^1 e((n-m)\alpha)\mathrm{d}\alpha = \sum_{n=1}^{N} |a_n|^2.$$

  So we have that

$$\delta \sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 \sim \int_0^1 \left| \sum_{n=1}^{N} a_n e(n\alpha) \right|^2 \mathrm{d}\alpha = \sum_{n=1}^{N} |a_n|^2$$

  as $\delta \to 0^+$. In fact, if $N = 1$ and $a_1 = 1$, then

$$\sum_{r=1}^{R} \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|^2 = R = (N + 1/\delta - 1) \sum_{n=1}^{N} |a_n|^2.$$

Proof of Theorem 8.1.1. Since

$$\left| \sum_{n=1}^{N} a_n e((n+h)\alpha_r) \right| = \left| \sum_{n=1}^{N} a_n e(n\alpha_r) \right|$$

for any shift $h \in \mathbb{Z}$, it suffices to show that

$$\sum_{r=1}^{R} \left| \sum_{-N_1 \le n < N-N_1} a_n e(n\alpha_r) \right|^2 \le (N + 1/\delta) \sum_{-N_1 \le n < N-N_1} |a_n|^2,$$

where

$$N_1 := \lfloor N/2 \rfloor .$$

In addition, we may assume that $-1/2 < \alpha_1 \le \cdots < \alpha_r \le 1/2$. Our assumption that these are δ-spaced points then implies that

$$\|\alpha_r - \alpha_s\| \ge \delta \|r - s\|_R, \quad \text{where} \quad \|n\|_R := \min\{|n - kR| \,:\, k \in \mathbb{Z}\}. \tag{8.1.3}$$

Now, standard facts about Hilbert spaces imply that there is some number M such that

$$\sum_{r=1}^{R} \left| \sum_{-N_1 \le n < N-N_1} a_n e(n\alpha_r) \right|^2 \le M \sum_{-N_1 \le n < N-N_1} |a_n|^2,$$

for any sequence of complex numbers $\{a_n\}_{-N_1 \le n < N-N_1}$. Indeed, $\sqrt{M}$ is the norm of the operator $T : \mathbb{C}^N \to \mathbb{C}^R$ induced by the $R \times N$ matrix $A = (e(n\alpha_r))_{r \le R, -N_1 \le n < N-N_1}$. By duality, the norm of this operator equals the norm of the operator $T^* : \mathbb{C}^R \to \mathbb{C}^N$ induced by $A^* = (e(-n\alpha_r))_{-N_1 \le n < N-N_1, r \le R}$, that is to say M is the smallest number such that

$$\sum_{-N_1 \le n < N-N_1} \left| \sum_{r=1}^{R} b_r e(n\alpha_r) \right|^2 \le M \sum_{r=1}^{R} |b_r|^2.$$

(Prove this claim by first showing that $\langle A\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, A^*\boldsymbol{y} \rangle$ whenever $\boldsymbol{x} \in \mathbb{C}^N$ and $\boldsymbol{y} \in \mathbb{C}^R$.) We will increase the sum over n in order to smoothen the cut-off at the end of the range. Let $g : \mathbb{Z} \to [0, 1]$ be a sequence of finite support such that

$$g(n) \ge \begin{cases} 1 & \text{if } n \in [-N_1, N - N_1 - 1] \subset [-N_1, N_1], \\ 0 & \text{otherwise.} \end{cases}$$

Then it suffices to show that

$$\sum_{n \in \mathbb{Z}} g(n) \left| \sum_{r=1}^{R} b_r e(n\alpha_r) \right|^2 \le M \sum_{r=1}^{R} |b_r|^2.$$

Opening the square, we have that

$$\begin{aligned} \sum_{n \in \mathbb{Z}} g(n) \left| \sum_{r=1}^{R} b_r e(n\alpha_r) \right|^2 &= \sum_{n \in \mathbb{Z}} g(n) \sum_{1 \le r,s \le R} b_r \overline{b_s} e(n(\alpha_r - \alpha_s)) \\ &= \sum_{1 \le r,s \le R} b_r \overline{b_s} \sum_{n \in \mathbb{Z}} g(n) e(n(\alpha_r - \alpha_s)) \\ &= G(0) \sum_{r=1}^{R} |b_r|^2 + \sum_{\substack{1 \le r,s \le R \\ r \ne s}} b_r \overline{b_s} G(\alpha_r - \alpha_s), \end{aligned}$$

where

$$G(x) := \sum_{n\in\mathbb{Z}} g(n)e(nx)$$

is the Fourier series associated to the sequence $(g(n))_{n\in\mathbb{Z}}$. Given a parameter $H \in \mathbb{N}$ to be chosen later, we take

$$g(n) = \begin{cases} 1 & \text{if } |n| \le N_1, \\ 1 - \dfrac{|n| - N_1}{H} & \text{if } N_1 < |n| \le N_1 + H, \\ 0 & \text{if } |n| > N_1 + H. \end{cases}$$

It is then easy to see that

$$G(x) = \frac{(N_1 + H)F_{N_1+H}(x) - NF_N(x)}{H},$$

where

$$F_J(x) := \frac{1}{J}\sum_{j=0}^{J-1}\sum_{|k|\le j} e(kx) = \sum_{|j|<J}\left(1 - \frac{|j|}{J}\right)e(jx) = \frac{1}{J}\left(\frac{\sin(J\pi x)}{\sin(\pi x)}\right)^2$$

is the Fejer kernel. In particular,

$$|G(x)| \le \frac{2}{H\sin^2(\pi x)} \le \frac{1}{2H\|x\|^2},$$

by the inequality $\sin(\pi y) \ge 2y$ for $y \in [0, 1/2]$. Since we also have that

$$G(0) = \frac{(N_1 + H)^2 - N_1^2}{H} = 2N_1 + H \le N + H,$$

we deduce that

$$\sum_{n\in\mathbb{Z}} g(n)\left|\sum_{r=1}^{R} b_r e(n\alpha_r)\right|^2 \le (N + H)\sum_{r=1}^{R}|b_r|^2 + \sum_{\substack{1\le r,s\le R\\ r\ne s}} \frac{|b_r b_s|}{2\delta^2 H\|r - s\|_R^2},$$

where we used (8.1.3). Extending the definition of the sequence b_r for $r \in \mathbb{Z}$ via the formula $b_{r+kR} = b_r$, for $r \in \{1, \dots, R\}$ and $k \in \mathbb{Z}$, we find that

$$\sum_{\substack{1\le r,s\le R\\ r\ne s}} \frac{|b_r b_s|}{\|r - s\|_R^2} \le \sum_{1\le h\le R/2} \frac{1}{h^2}\sum_{r=1}^{R}(|b_r b_{r+h}| + |b_r b_{r-h}|).$$

Cauchy-Schwarz implies that

$$\sum_{r=1}^{R}|b_r b_{r\pm h}| \le \left(\sum_{r=1}^{R}|b_r|^2\right)^{1/2}\left(\sum_{r=1}^{R}|b_{r\pm h}|^2\right)^{1/2} = \sum_{r=1}^{R}|b_r|^2,$$

so we conclude that

$$\sum_{\substack{1\le r,s\le R\\ r\neq s}} \frac{|b_r b_s|}{\|r-s\|_R^2} \le 2\sum_{h=1}^{\infty}\frac{1}{h^2}\sum_{r=1}^{R}|b_r|^2 = \frac{\pi^2}{3}\sum_{r=1}^{R}|b_r|^2.$$

Therefore

$$\sum_{n\in\mathbb{Z}} g(n)\left|\sum_{r=1}^{R} b_r e(n\alpha_r)\right|^2 \le \left(N+H+\frac{\pi^2}{6H\delta^2}\right)\sum_{r=1}^{R}|b_r|^2,$$

for any $H\in\mathbb{N}$. Taking $H=\lceil \pi/(\delta\sqrt{6})\rceil$ completes the proof of the theorem. □

In order to put Theorem 8.1.1 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 8.1.2. *For $Q\ge 1$ and $\{a_n\}_{n=1}^N\subset\mathbb{C}$, we have that*

$$\sum_{q\le Q}\sum_{\substack{1\le a\le q\\ (a,q)=1}}\left|\sum_{n=1}^{N} a_n e(na/q)\right|^2 \le \left(N+1+\frac{\pi\sqrt{6}}{3}Q^2\right)\sum_{n=1}^{N}|a_n|^2.$$

Theorem 8.1.3 (Large sieve - character sum version)**.** *For $Q\ge 1$ and $\{a_n\}_{n=1}^N\subset\mathbb{C}$, we have that*

$$\sum_{q\le Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\left|\sum_{n=1}^{N} a_n\chi(n)\right|^2 \le \left(N+1+\frac{\pi\sqrt{6}}{3}Q^2\right)\sum_{n=1}^{N}|a_n|^2,$$

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 8.1.2, we use Theorem 5.3.2 (i.e. we use Fourier inversion with respect to the additive characters). This theorem implies that, for every primitive character $\chi\,(\mathrm{mod}\,q)$,

$$\sum_{n=1}^{N} a_n\chi(n) = \sum_{n=1}^{N} a_n\frac{1}{\tau(\overline{\chi})}\sum_{a=1}^{q}\overline{\chi}(a)e(an/q) = \frac{1}{\tau(\overline{\chi})}\sum_{a=1}^{q}\overline{\chi}(a)S(a/q).$$

Since for such a character we also have that $|\tau(\overline{\chi})| = \sqrt{q}$, we find that

$$\sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{n=1}^{N} a_n \chi(n)\right|^2 = \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{a=1}^{q} \overline{\chi}(a) S(a/q)\right|^2 \leq \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)} \left|\sum_{a=1}^{q} \overline{\chi}(a) S(a/q)\right|^2$$
$$= \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)} \sum_{a_1}^{q} \sum_{a_2=1}^{q} \overline{\chi}(a_1)\chi(a_2) S(a_1/q)\overline{S(a_2/q)}$$
$$= \frac{1}{q} \sum_{a_1}^{q} \sum_{a_2=1}^{q} S(a_1/q)\overline{S(a_2/q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a_1)\chi(a_2) = \frac{\varphi(q)}{q} \sum_{\substack{1 \leq a \leq q \\ (a,q)=1}} |S(a/q)|^2,$$

by Theorem 5.1.2. Multiplying the above relation by $q/\varphi(q)$, summing the resulting inequality over $q \leq Q$, and applying Corollary 8.1.2 completes the proof of the theorem. □

8.2 Proof of the Bombieri-Vinogradov theorem

8.2.1 Vaughan's identity

In view of Theorem 8.1.3, a plausible strategy for proving Theorem 8.0.2 would be to apply the Cauchy-Schwarz inequality. Ignoring the maximum over $y \leq x$ for the moment, this inequality and Theorem 8.1.3 yield that

$$\sum_{Q < q \leq 2Q} \frac{q}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{n \leq x} \chi(n)\Lambda(n)\right| \ll Q \left(\sum_{Q < q \leq 2Q} \frac{q}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{n \leq x} \chi(n)\Lambda(n)\right|^2\right)^{\frac{1}{2}}$$
$$\ll Q(Q + \sqrt{x})\sqrt{x} \log x.$$

This is barely not sufficient for deducing Theorem 8.0.2: indeed, the above inequality can be rewritten as

$$\frac{1}{Q} \sum_{Q < q \leq 2Q} \frac{q}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)}^{*} \left|\sum_{n \leq x} \chi(n)\Lambda(n)\right| \ll Q\sqrt{x \log x} + x\sqrt{\log x}.$$

However, the right hand side in the above estimate is never $\ll x/(\log x)^A$, a crucial ingredient in the deduction of Theorem 8.0.2.

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 8.0.2 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 $g = \log$, so that $f * g$ is 'quasi-linear', or both f and g are supported on large integers, so that $f * g$ is genuinely 'biliner'. 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 (cf. Theorem 5.4.1). 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 8.2.1 (Vaughan's identity)**.** *Let $U \geq 1$ and $V \geq 1$ be two parameters. For any $n > U$, we have that*

$$\Lambda(n) = \sum_{\substack{ab=n\\ a\leq V}} \mu(a)\log b - \sum_{\substack{ab=n\\ a\leq UV}} \left(\sum_{\substack{cd=a\\ c\leq V,\, d\leq U}} \mu(c)\Lambda(d) \right) - \sum_{\substack{ab=n\\ a>U,\, b>V}} \Lambda(a) \left(\sum_{\substack{d|b\\ d\leq V}} \mu(d) \right)$$

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,

$$\begin{aligned}
\sum_{\substack{k\ell=n\\ k>V}} \mu(k)\log \ell &= \sum_{\substack{k\ell=n\\ k>V}} \mu(k) \sum_{m|\ell} \Lambda(m) = \sum_{m|n} \Lambda(m) \sum_{\substack{k|n/m\\ k>V}} \mu(k) \\
&= \sum_{\substack{mr=n\\ r>1}} \Lambda(m) \sum_{\substack{k|r\\ k>V}} \mu(k) = - \sum_{\substack{mr=n\\ r>1}} \Lambda(m) \sum_{\substack{k|r\\ k\leq V}} \mu(k) \\
&= - \sum_{\substack{mr=n\\ m\leq U}} \Lambda(m) \sum_{\substack{k|r\\ k\leq V}} \mu(k) - \sum_{\substack{mr=n\\ m>U,\, r>1}} \Lambda(m) \sum_{\substack{k|r\\ k\leq V}} \mu(k) \\
&= - \sum_{\substack{mk\ell=n\\ k\leq V,\, m\leq U}} \mu(k)\Lambda(m) - \sum_{\substack{mr=n\\ m>U,\, r>V}} \Lambda(m) \sum_{\substack{k|r\\ k\leq V}} \mu(k),
\end{aligned}$$

which completes the proof of the Lemma. □

Exercises

Exercise 8.2.1. Prove the identity

$$\frac{-\zeta'}{\zeta} = F - \zeta FG - \zeta' G + \left(-\frac{\zeta'}{\zeta} - F \right)(1 - \zeta G),$$

where

$$F(s) = \sum_{n\leq U} \frac{\Lambda(n)}{n^s} \quad \text{and} \quad G(s) = \sum_{m\leq V} \frac{\mu(m)}{m^s}.$$

Deduce Lemma 8.2.1.

8.2.2 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 8.2.1. We start with the following general lemma.

Lemma 8.2.2 (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 8.2.1. The above lemma yields a non-trivial result as soon as $x > (D\sqrt{q})^{1+\epsilon}$.

Proof of Lemma 8.2.2. 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 5.4.1) and partial summation imply that

$$\sum_{b\le y} \chi(b)(\log b)^s = \int_1^y (\log t)^s \mathrm{d}\left(\sum_{b\le t}\chi(b)\right) = (\log y)^s \sum_{b\le y}\chi(b) - s\int_1^y \frac{(\log t)^{s-1}}{t}\left(\sum_{b\le t}\chi(b)\right)\mathrm{d}t$$
$$\ll \sqrt{q}(\log q)\left((\log y)^s + s\int_1^y \frac{(\log t)^{s-1}}{t}\mathrm{d}t\right) \ll \sqrt{q}(\log q)(\log y)^s.$$

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 8.2.2. Moreover, since

$$\left| \sum_{\substack{cd=a \\ c\le V,\, d\le U}} \mu(c)\Lambda(d) \right| \le \sum_{d|a} \Lambda(d) = \log a,$$

applying Lemma 8.2.2 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 8.2.1, we conclude that

$$\begin{aligned} \sum_{n\le y} \Lambda(n)\chi(n) &= \sum_{\substack{ab\le y \\ a>U,\, b>V}} \chi(ab)\Lambda(a) \left(\sum_{\substack{d|b \\ d\le V}} \mu(d) \right) + O(UV\sqrt{q}(\log x)^2) \\ &= \sum_{\substack{mn\le y \\ m>U,\, n>V}} \chi(mn)\Lambda(m) \left(\sum_{\substack{d|n \\ d\le V}} \mu(d) \right) + O(UV\sqrt{q}(\log x)^2), \end{aligned} \tag{8.2.1}$$

for all $y \le x$. Consequently, if we set

$$\alpha_m = \Lambda(m) \quad \text{and} \quad \beta_n = \sum_{\substack{d|n \\ d\le V}} \mu(d),$$

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

$$S(x;Q) \ll \sum_{Q<q\le 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y\le x} \left| \sum_{\substack{mn\le y \\ m>U,\, n>V}} \chi(mn)\alpha_m\beta_n \right| + UVQ^{5/2}(\log x)^2. \tag{8.2.2}$$

This reduces Theorem 8.0.2 to a bilinear sum estimate,w which will be demonstrated in the next section.

8.2.3 The bilinear part of von Mangoldt's function

We will treat the sum on the right hand side of (8.2.2) using Cauchy-Schwarz and Theorem 8.1.3. However, before we can apply Cauchy-Schwarz, we need to separate the variables m

and n. First, we perform a dyadic decomposition of the location of the variables y, m and n, which allows us to partially keep track of the fact that $mn \leq y$. Indeed, we break the range of y in relation (8.2.2) into $O(\log x)$ intervals of the form $(Y, 2Y]$, the range of summation of m into $O(\log x)$ dyadic intervals $(M, 2M]$, and the range of summation of n into $O(\log x)$ dyadic intervals $(N, 2N]$. Then taking the maximum over all these $O(\log x)^3$ possibilities, we find that

$$S(x;Q) \ll (\log x)^3 \max_{\substack{1\leq Y\leq x,\\ M\geq U,\, N\geq V\\ MN\leq 2Y}} \sum_{Q<q\leq 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{Y<y\leq 2Y} \left| \sum_{\substack{mn\leq y\\ M<m\leq 2M\\ N<n\leq 2N}} \chi(mn)\alpha_m\beta_n \right| + UVQ^{5/2}(\log x)^2, \tag{8.2.3}$$

We are now ready to separate m and n. This is accomplished by applying Theorem 4.3.1 with $c = \epsilon/2$, which is admissible by Exercise 1.1.5, yielding the estimate

$$\sum_{\substack{mn\leq y\\ M<m\leq 2M\\ N<n\leq 2N}} \chi(mn)\alpha_m\beta_n = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=1+\epsilon/2+1/\log y\\ |\mathrm{Im}(s)|\leq Y}} \sum_{\substack{M<m\leq 2M\\ N<n\leq 2N}} \frac{\chi(mn)\alpha_m\beta_n}{(mn)^s} \frac{y^s}{s} \mathrm{d}s + O_\epsilon(Y^\epsilon).$$

We shift the contour to the line $\mathrm{Re}(s) = 1/2$. There are no poles, and the contribution of the vertical lines of the rectangle is easily seen to be $\ll_\epsilon Y^\epsilon$, where we used Exercise 1.1.5 again. Therefore

$$\sum_{\substack{mn\leq y\\ M<m\leq 2M\\ N<n\leq 2N}} \chi(mn)\alpha_m\beta_n \ll_\epsilon \sqrt{y} \int_{-Y}^{Y} \left| \sum_{M<m\leq 2M} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right| \cdot \left| \sum_{N<n\leq 2N} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right| \frac{\mathrm{d}t}{1+|t|} + Y^\epsilon,$$

so that

$$\begin{aligned} &\sum_{Q<q\leq 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y\leq x} \left| \sum_{\substack{mn\leq y\\ M<m\leq 2M\\ N<n\leq 2N}} \chi(mn)\alpha_m\beta_n \right| \\ &\ll \sqrt{Y} \int_{-Y}^{Y} \sum_{Q<q\leq 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{M<m\leq 2M} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right| \cdot \left| \sum_{N<n\leq 2N} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right| \frac{\mathrm{d}t}{1+|t|} + Q^2Y^\epsilon. \end{aligned}$$

Inserting the above estimate into (8.2.2), and majoring the integrand by its maximum over all $t \in [-Y, Y]$, we deduce that

$$S(x;Q) \ll \sqrt{x}(\log x)^4 \max_{\substack{M\geq U,\, N\geq V\\ MN\leq 2x,\, |t|\leq x}} \sum_{Q<q\leq 2Q} \frac{q}{\varphi(q)} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \left| \sum_{M<m\leq 2M} \frac{\chi(m)\alpha_m}{m^{1/2+it}} \right| \cdot \left| \sum_{N<n\leq 2N} \frac{\chi(n)\beta_n}{n^{1/2+it}} \right| + UVQ^{5/2}(\log x)^2.$$

Finally, Theorem 8.1.3 implies that

$$\sum_{Q<q\leq 2Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\left|\sum_{M<m\leq 2M}\frac{\chi(m)\alpha_m}{m^{1/2+it}}\right|^2\ll (M+Q^2)\sum_{M<m\leq 2M}\frac{|\alpha_m|^2}{m}\ll (M+Q^2)\log x,$$

since $|\alpha_m|\leq\Lambda(m)$ for all m, and similarly

$$\sum_{Q<q\leq 2Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\left|\sum_{N<n\leq 2N}\frac{\chi(n)\beta_n}{n^{1/2+it}}\right|^2\ll (N+Q^2)\sum_{N<n\leq 2N}\frac{|\beta_n|^2}{n}\ll (N+Q^2)(\log x)^3,$$

since $|\beta_n|\leq\tau(n)$ for all n (here, we used the fact that $\sum_{n\leq x}\tau(n)^2\ll x(\log x)^3$; the reader is invited to prove this estimate by writing $\tau^2=\tau_4*g$ and applying the convolution method). So the Cauchy-Schwarz inequality yields the estimate

$$\begin{aligned}
S(x;Q)&\ll\sqrt{x}(\log x)^6\max_{\substack{1\leq Y\leq x\\ M\geq U/2,\,N\geq V/2\\ MN\leq 2Y}}\sqrt{(M+Q^2)(N+Q^2)}+UVQ^{5/2}(\log x)^2\\
&\ll\sqrt{x}(\log x)^6\max_{\substack{1\leq Y\leq x\\ M\geq U/2,\,N\geq V/2\\ MN\leq 2Y}}\left(\sqrt{MN}+Q^2+Q(\sqrt{M}+\sqrt{N})\right)+UVQ^{5/2}(\log x)^2\\
&\ll\sqrt{x}(\log x)^6\left(\sqrt{x}+Q^2+Q\sqrt{\frac{x}{U}}+Q\sqrt{\frac{x}{V}}\,\right)+UVQ^{5/2}(\log x)^2\\
&\leq(\log x)^6\left(x+x^{1/2}Q^2+\frac{xQ}{U^{1/2}}+\frac{xQ}{V^{1/2}}+UVQ^{5/2}\right).
\end{aligned}$$

since $MN\ll x$, $M\ll x/V$ and $N\ll x/U$, for all M and N as above. We choose $U=V=x^{2/5}/Q^{3/5}\in[1,x]$ for $1\leq Q\leq x^{2/3}$ to deduce that

$$S(x;Q)\ll(\log x)^6(x+x^{1/2}Q^2+x^{4/5}Q^{13/10}).$$

This completes the proof of Theorem 8.0.2, and hence of the Bombieri-Vinogradov theorem.

Exercises

Exercise 8.2.2. Show that

$$S(x;Q)\ll(\log x)^6(x+x^{1/2}Q^2+x^{5/6}Q)\quad(1\leq Q\leq x^{2/3}),$$

which is an improvement over Theorem 8.0.2 when $Q\geq x^{1/3}$.

Hint: Write

$$\sum_{\substack{ab=n\\ a\leq UV}}\left(\sum_{\substack{cd=a\\ c\leq V,\,d\leq U}}\mu(c)\Lambda(d)\right)=\sum_{\substack{ab=n\\ a\leq U}}\left(\sum_{\substack{cd=a\\ c\leq V,\,d\leq U}}\mu(c)\Lambda(d)\right)+\sum_{\substack{ab=n\\ U<a\leq UV}}\left(\sum_{\substack{cd=a\\ c\leq V,\,d\leq U}}\mu(c)\Lambda(d)\right).$$

8.3 The arithmetic form of the large sieve

We conclude this chapter by showing that the large sieve inequality can be recast as an inequality about a sifted set, thus justifying the terminology 'large sieve':

Theorem 8.3.1. *For each prime p, let R_p be some subset of $\mathbb{Z}/p\mathbb{Z}$. Given parameters M, N and z, let $\mathcal{N}$ be a subset of the integers in $(M, M+N]$ avoiding the sets R_p, that is to say if $n \in \mathcal{N}$, then $n \notin R_p \pmod p$ for all $p < z$. We then have that*

$$\#\mathcal{N} \le \left(N + \frac{\pi\sqrt{6}}{3}z^2\right) \Bigg/ \left(\sum_{m<z} \mu^2(m)h(m)\right),$$

where h is multiplicative with $h(p) = |R_p|/(p - |R_p|)$.

Remark 8.3.1. The easiest way to see the link between the quantity $S(\mathcal{A}, \mathcal{P})$ and the above theorem is in the special case when

$$\mathcal{N} = \{x < n \le 2x : F(n) \text{ is prime}\}$$

where $F(x)$ is some polynomial with integer coefficients. Then $\mathcal{N}$ avoids the sets

$$R_p = \{m \in \mathbb{Z}/p\mathbb{Z} : F(m) \equiv 0 \pmod p\}$$

with $p < x$. So we see that Theorem 8.3.1 can be translated to an upper bound sieve estimate in important cases such as the above one. Moreover, Theorem 8.3.1 has the significant advantage that it does not depend on the assumption of hypotheses such as (3.3.3). However, the real strength of Theorem 8.3.1 is revealed when the sifting dimension κ that we discussed in the end of Section 3.2 becomes unbounded.

Proof of Theorem 8.3.1. Set

$$\mathcal{M} = \{M < n \le M + N : n \notin R_p \pmod p, \text{ for all } p < z\}.$$

We claim that for all sequences of complex numbers, and for all square-free integers $q < z$, we have that

$$\sum_{\substack{1\le a\le q\\(a,q)=1}} \left|\sum_{n\in\mathcal{M}} a_n e(an/q)\right|^2 \ge h(q)\left|\sum_{n\in\mathcal{M}} a_n\right|^2. \tag{8.3.1}$$

When $q = 1$, this holds trivially. For $q > 1$, we argue by induction on $\omega(q)$. First, we establish (8.3.1) when $q = p$ is prime. Applying Parseval's identity for additive characters, we find that

$$\sum_{a=1}^{p-1}\left|\sum_{n\in\mathcal{M}} a_n e(an/p)\right|^2 = p \sum_{\substack{n_1,n_2\in\mathcal{M}\\ n_1\equiv n_2 \pmod p}} a_{n_1}\overline{a_{n_2}} - \left|\sum_{n\in\mathcal{M}} a_n\right|^2 = p\sum_{b=0}^{p-1}\left|\sum_{\substack{n\in\mathcal{M}\\ n\equiv b \pmod p}} a_n\right|^2 - \left|\sum_{n\in\mathcal{M}} a_n\right|^2.$$

This identity can be written more conceptually using probability theory: we equip the group $\mathbb{Z}/p\mathbb{Z}$ with the uniform counting measure (that is to say, $\mathbb{P}(A) = |A|/p$ for $A \subset \mathbb{Z}/p\mathbb{Z}$), and we consider the random variable $X : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$, defined by $X(b) = \sum_{n\in\mathcal{M},\, n\equiv b \,(\mathrm{mod}\, p)} a_n$, so that

$$\sum_{a=1}^{p-1} \left| \sum_{n\in\mathcal{M}} a_n e(an/p) \right|^2 = p^2 \cdot \mathrm{Var}(X),$$

with $\mathrm{Var}(X) = \mathbb{E}[|X - \mathbb{E}[X]|^2] = \mathbb{E}[|X|^2] - |\mathbb{E}[X]|^2$, as usual. If X were uniformly distributed among the residue classes of $\mathbb{Z}/p\mathbb{Z}$, we would have that $\mathrm{Var}(X) = 0$. However, we know that $\mathcal{M}$ is supported on $n \notin R_p \,(\mathrm{mod}\, p)$, so that $X(b) = 0$ if $b \in R_p$. We take advantage of this fact via the Cauchy-Schwarz inequality:

$$|\mathbb{E}[X]|^2 = \left| \frac{1}{p} \sum_{b\in\mathbb{Z}/p\mathbb{Z}\setminus R_p} X(b) \right|^2 \le \frac{p - |R_p|}{p} \cdot \mathbb{E}[|X|^2],$$

so that

$$\mathrm{Var}(X) \ge \left(\frac{p}{p - |R_p|} - 1 \right) |\mathbb{E}[X]|^2 = \frac{h(p)}{p^2} \left| \sum_{n\in\mathcal{M}} a_n \right|^2,$$

that is to say, (8.3.1) does hold when p is a prime $< z$.

Next, assume that (8.3.1) 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 $\omega(q) = j + 1$. We may write $q = q_1 q_2$ with $\omega(q_i) \le j$ for $i \in \{1, 2\}$. Furthermore, note that the set $\{a_1 q_2 + a_2 q_1 : 1 \le a_i \le q_i,\ (a_i, q_i)\ (i \in \{1, 2\})\}$ is a set of representatives for the set of residues $\{1 \le a \le q : (a, q) = 1\}$. Hence

$$\sum_{\substack{1\le a\le q\\(a,q)=1}} \left| \sum_{n\in\mathcal{M}} a_n e(an/q) \right|^2 = \sum_{\substack{1\le a_1\le q_1\\(a_1,q_1)=1}} \sum_{\substack{1\le a_2\le q_2\\(a_2,q_2)=1}} \left| \sum_{n\in\mathcal{M}} a_n e\left(\frac{a_1 n}{q_1} + \frac{a_2 n}{q} \right) \right|^2.$$

For each fixed a_1 as above, we apply (8.3.1) with q_2 in place of q and $a_n e(a_1 n/q_1)$ in place of a_n, which holds by the induction hypothesis. So

$$\sum_{\substack{1\le a_2\le q_2\\(a_2,q_2)=1}} \left| \sum_{n\in\mathcal{M}} a_n e\left(\frac{a_1 n}{q_1} + \frac{a_2 n}{q} \right) \right|^2 \ge h(q_2) \left| \sum_{n\in\mathcal{M}} a_n e\left(\frac{a_1 n}{q_1} \right) \right|^2.$$

Summing the above inequality over a_1 and applying (8.3.1) 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 (8.3.1) is true. This completes the inductive step, and hence the proof of (8.3.1).

Finally, applying this relation with a_n being the characteristic function of the set $\mathcal{M}$ 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 8.1.2. This completes the proof of Theorem 8.3.1. □

We conclude this section with a couple of applications of Theorem 8.3.1 by studying the least quadratic non-residue mod a prime p

$$n_p := \min\left\{a \in \mathbb{N} : \left(\frac{a}{p}\right) \neq 1\right\}.$$

Vinogradov conjectured that $n_p \ll_\epsilon p^\epsilon$ and in Exercise 5.4.1 we proved that $n_p \ll_\epsilon p^{1/(2\sqrt{e})+\epsilon}$. Using Theorem 8.3.1, we will show that Vinogradov's conjecture is true for most primes p.

Theorem 8.3.2. *Fix $\epsilon > 0$. Then we have that*

$$\#\{p \le x : n_p > x^\epsilon\} \ll_\epsilon 1.$$

Proof. Let $\mathcal{N} = \{m \le x^2 : p|m \ \Rightarrow\ p \le x^\epsilon\}$, and note that

$$|\mathcal{N}| \gg_\epsilon x^2 \tag{8.3.2}$$

by Exercise 2.2.5. Fix, for the moment, a prime p with $n_p > x^\epsilon$. If $n \in \mathcal{N}$, then every prime divisor q of n is $\le x^\epsilon < n_p$, whence $(q|p) = 1$. By the multiplicative of the Legendre symbol we then infer that $(n|p) = 1$ for all $n \in \mathcal{N}$.

The above discussion implies that if we reduce the set $\mathcal{N}$ modulo a prime p, then it avoids the set

$$R_p = \begin{cases} \{0\} \cup \{k \,(\mathrm{mod}\, p) : (k|p) = -1\} & \text{if } n_p > x^\epsilon,\\ \emptyset & \text{otherwise.}\end{cases}$$

We then apply Theorem 8.3.1 with $z = x$ and $N = x^2$ to find that

$$|\mathcal{N}| \ll x^2/S, \tag{8.3.3}$$

where

$$S = \sum_{m<x} \mu^2(m) \prod_{p|m} \frac{|R_p|}{p - |R_p|}.$$

Comparing (8.3.2) and (8.3.3), we find that $S \ll_\epsilon 1$. On the other hand, if $m = p > 2$ with $n_p > x^\epsilon$, then $|R_p|/(p - |R_p|) = (p+1)/(p-1) \ge 1$. Hence

$$S \ge \#\{2 < p < x : n_p > x^\epsilon\},$$

which completes the proof of the theorem. □

Exercises

Exercise 8.3.1. For each fixed $\epsilon > 0$, prove that

$$\#\{p \le x : n_p > p^\epsilon\} \ll_\epsilon \log\log x.$$

Exercise 8.3.2. (a) Let R_p and $\mathcal{N}$ be as in the statement of Theorem 8.3.1. If $|R_p| \ge cp + O(1)$ with $c \in (0,1)$, then prove that

$$\#\mathcal{N} \ll_c \sqrt{N}(\log N)^{\frac{1-2c}{1-c}}.$$

(b) Find a sequence $\mathcal{N} \subset [1, N]$ of cardinality $\gg \sqrt{N}$ that avoids $p/2+O(1)$ residue classes mod p, for each prime p.

Chapter 9

Bounded gaps between primes

Let $p_1, p_2, p_3, \dots$ be the sequence of prime numbers in increasing order. We wish to study the spacing distribution 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 $\sum_{k \leq n}(p_{k+1} - p_k) = p_{n+1} - p_1 \sim n \log n$. By partial summation, it is then easy to deduce that

$$\sum_{1<k\leq n} \frac{p_{k+1} - p_k}{\log k} \sim n \quad (n \to \infty),$$

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. The main theorem of this chapter is a result towards the twin prime conjecture:

Theorem 9.0.1. *For each $m \in \mathbb{N}$, we have that*

$$\liminf_{n\to\infty}(p_{n+m} - p_n) \ll e^{4m} m^5.$$

Theorem 9.0.1 is due to Maynard [17] and Tao [27], who built upon previous work of Goldston, Pintz and Yildirim [6]. In the special case $m = 1$, the first person to show that $\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$ was Zhang [29].

In order to detect $m + 1$ primes close together, we use the following construction: Let $0 \leq s_1 < s_2 < \cdots < s_k$ 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 "gap detector". The key to making this approach work is to judiciously choose the weights w_n in a way that achieves two things simultaneously:

- 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 estimate S unconditionally (e.g. without appealing to the twin prime conjecture, which make us enter 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 $S\sim (k-m)(\log N)\cdot\#\{N<n\le 2N : w_n=1\}$, 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.

Motivated by the above discussion and by the Selberg sieve weights discussed in Section 3.4, we take

$$w_n=\left(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2,$$

where $\boldsymbol{d}=(d_1,\dots,d_k)$. This is the choice made by Maynard and Tao, and it can be viewed as a multidimensional version of the original GPY sieve, named after Goldston, Pintz and Yildirim, who made the more restricted choice

$$\left(\sum_{d\mid\prod_{j=1}^k(n+s_j)}\lambda_d\right)^2.$$

Opening the square in the definition of w_n and inverting the order of summation, we see that

$$\sum_{N<n\le 2N}\mathbf{1}_{\mathbb{P}}(n+s_j)w_n=\sum_{\boldsymbol{d},\boldsymbol{e}}\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}\sum_{\substack{N-s_j<p\le 2N-s_j\\ p\equiv -s_i\,(\mathrm{mod}\,[d_i,e_i])\ (1\le i\le k)}}1$$

for each $j\in\{1,\dots,k\}$. We naturally appeal to the Bombieri-Vinogradov theorem to estimate this sum. In order for the Bombieri-Vinogradov theorem to be applicable, we assume that the parameters $\lambda_{\boldsymbol{d}}$ are supported on k-tuples $\boldsymbol{d}$ such that

$$d_1\cdots d_k\le D,\quad \text{where}\quad D\le N^{1/4}/(\log N)^B$$

for a sufficiently large B. We add an additional small technical twist to our short gap detector S. We wish to perform what is sometimes called a preliminary sieve. We want to only consider integers $n\in(N,2N]$ such that the integers $n+s_j$ have no small prime factors, up to some parameter z. To this end, we set

$$Q(x)=\prod_{j=1}^k(x+s_j)\quad\text{and}\quad W=\prod_{p<z}p,$$

so that the sum we wish to estimate is

$$S=\sum_{\substack{N<n\le 2N\\ (Q(n),W)=1}}\left(\sum_{j=1}^k\mathbf{1}_{\mathbb{P}}(n+s_j)-m\right)\left(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2.$$

The condition $(Q(n),W)=1$ will be dealt with by an application of the Fundamental Lemma of Sieve Methods.

9.1 Estimating the prime gap detector

In the estimation of the sum S, we use the notations

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

and

$$M := \max_{\boldsymbol{d}} |\lambda_d|.$$

All implicit constants in this section might depend on k and the choice of the k-tuple $\boldsymbol{s}$.

Lemma 9.1.1. *If $D \le N^{1/2-\epsilon}$ and $z = (\log N)^C$, then*

$$\sum_{\substack{N<n\le 2N\\(Q(n),W)=1}} \left(\sum_{d_j \mid n+s_j,\, 1\le j\le k} \lambda_{\boldsymbol{d}}\right)^2 = N \cdot \prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right) \sum_{(a_1\cdots a_k,W)=1} \frac{\xi_{\boldsymbol{a}}^2}{a_1\cdots a_k} + O_{C,\epsilon}\left(\frac{NM^2}{(\log N)^{C-3k-1}}\right),$$

where

$$\xi_{\boldsymbol{a}} := \sum_{(m_1\cdots m_k,W)=1} \frac{\lambda_{d_1 m_1,\ldots,m_k d_k}}{m_1\cdots m_k}.$$

Proof. Call T the sum in question and assume that N is large enough. Since $(Q(n),W)=1$, we must also have that $(d_j,W)=1$ for any $d_j \mid n+s_j$. In addition, note that if $d_i \mid n+s_i$ and $d_j \mid n+s_j$ for $i\neq j$, then $(d_i,d_j)\mid s_j-s_i$. However, all prime factors of d_i and d_j are $\ge z$, so the same must be true for the prime factors of (d_i,d_j). If N is large, the only possibility for (d_i,d_j) then is to equal 1. The above discussion implies that

$$T = \sum_{\substack{(d_ie_i,d_je_jW)=1\\1\le i,j\le k,\ i\neq j}} \lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}} \cdot \#\left\{ \begin{array}{ll} N<n\le 2N & : \quad n\equiv -s_j \,(\mathrm{mod}\,[d_j,e_j]),\\ (Q(n),W)=1 & \quad 1\le j\le k \end{array} \right\}.$$

To estimate the cardinality of the set in the right hand side, we use Theorem 3.3.1, where we take $\mathcal{A} = \{Q(n) : N<n\le 2N,\ n\equiv s\,(\mathrm{mod}\, q)\}$ with $q=\prod_{j=1}^k [d_j,e_j] \le N^{1-\epsilon}$ and an appropriate $s\in\mathbb{Z}$, $w_a=1$ for all a, $X=N/q$, $g(d)=\nu(d)/d$ and $r_d = O(\nu(d))$ for $d\mid W$. Then

$$\sum_{\substack{N<n\le 2N\\(Q(n),W)=1\\n\equiv a\,(\mathrm{mod}\, q)}} 1 = \frac{N}{q}\prod_{p\mid W}\left(1-\frac{\nu(p)}{p}\right) + O_{C,\epsilon}\left(\frac{N}{e^{\sqrt{\log N}}} + \sum_{d\le N^{1/2}} \mu^2(d)\nu(d)\right).$$

Since $\nu(d)\le \tau_k(d) \ll d^{1/3}$ for square-free d, the remainder term in the above formula is $\ll N/(\log N)^A$, for any fixed $A\ge 1$. Taking A large enough, we conclude that

$$T = N\prod_{p\mid W}\left(1-\frac{\nu(p)}{p}\right) \sum_{\substack{(d_ie_i,d_je_jW)=1\\1\le i,j\le k,\ i\neq j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1,e_1]\cdots[d_k,e_k]} + O\left(\frac{M^2N}{(\log N)^C}\right).$$

Finally, 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 $(d_1 \cdots d_k, W) = (e_1 \cdots e_k, W) = 1$ but $(d_ie_i, d_je_j) > 1$, then there must be a prime $p > z = (\log N)^C$ dividing both $[d_i, e_i]$ and $[d_j, e_j]$. We therefore conclude that

$$\sum_{\substack{(d_ie_i, d_je_jP(z))=1 \\ 1\leq i,j\leq k,\ i\neq j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1, e_1]\cdots[d_k, e_k]} = \sum_{\substack{(d_ie_i, P(z))=1 \\ 1\leq i\leq k}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1, e_1]\cdots[d_k, e_k]} + O\left(\frac{M^2}{(\log N)^{C-3k}}\right),$$

which an error of admissible size. Since we also have that

$$\frac{1}{[d, e]} = \frac{(d, e)}{de} = \frac{1}{de}\sum_{a|(d,e)} \varphi(a),$$

we deduce that

$$\sum_{\substack{P^-(d_ie_i)>z \\ 1\leq i\leq k}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{[d_1, e_1]\cdots[d_k, e_k]} = \sum_{P^-(a_1\cdots a_k)>z} \varphi(a_1)\cdots\varphi(a_k)\left(\sum_{\substack{d_j\equiv 0 \,(\mathrm{mod}\, a_j) \\ (d_j, W)=1 \\ 1\leq j\leq k}} \frac{\lambda_{\boldsymbol{d}}}{d_1\cdots d_k}\right)^2.$$

If $(a, W) = 1$ 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 $j \in \{1, \ldots, k\}$ by producing a total error term of size $(\log N)^{3k+1}/z$, which is admissible by our choice of z. This completes the proof of the lemma. □

Lemma 9.1.2. *If $z = (\log N)^C$, $D \leq N^{1/4}e^{-\sqrt{\log N}}$ and $j_0 \in \{1, \ldots, k\}$, then*

$$\sum_{\substack{N<n\leq 2N \\ (Q(n), W)=1}} \mathbf{1}_{\mathbb{P}}(n + s_{j_0})\left(\sum_{\substack{d_j|n+s_j \\ 1\leq j\leq k}} \lambda_{\boldsymbol{d}}\right)^2 = \int_N^{2N} \frac{\mathrm{d}t}{\log t} \cdot \prod_{p\leq z} \frac{1 - \nu(p)/p}{1 - 1/p} \sum_{\substack{(a_1\cdots a_k, W)=1 \\ a_{j_0}=1}} \frac{\zeta_{j_0,\boldsymbol{a}}^2}{a_1\cdots a_k} + O_{A,\epsilon}\left(\frac{NM^2}{(\log N)^{C-3k+2}}\right),$$

where

$$\zeta_{j_0,\boldsymbol{a}} = \sum_{\substack{(m_j, W)=1 \\ m_{j_0}=1}} \frac{\lambda_{a_1m_1,\ldots,a_km_k}}{m_1\cdots m_k}.$$

Proof. Call T_{j_0} the sum in question. For simplicity, we consider the case $j_0 = k$; the proof

of the other cases follows *mutatis mutandis*. Clearly,

$$T_k = \sum_{\substack{N-s_k<n\le 2N-s_k\\(Q(n),W)=1}} \mathbf{1}_{\mathbb{P}}(n+s_k)\Bigg(\sum_{\substack{d_j\mid n+s_j\\1\le j\le k}}\lambda_{\boldsymbol{d}}\Bigg)^2 + O(M^2N^{1/2})$$
$$= \sum_{\substack{N<p\le 2N\\(Q(p-s_k),W)=1}} \Bigg(\sum_{\substack{d_j\mid p+s_j-s_k\\1\le j\le k}}\lambda_{\boldsymbol{d}}\Bigg)^2 + O(M^2N^{1/2}),$$

since a prime $p > N$ does not have divisors in the interval $(1, D]$. In particular, we must have that $d_k = 1$, and the condition $(Q(p - s_k), W) = 1$ is reduced to $(Q^*(p), W) = 1$, where

$$Q^*(x) := \prod_{j=1}^{k-1}(x + s_j - s_k).$$

Opening the square and changing the order of summation, we find that

$$T_k = \sum_{\substack{d_j,e_j\\(d_ie_i,d_je_jP(z))=1\\1\le i,j<k,\ i\ne j}} \lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}\cdot \#\left\{\begin{array}{ll} N<p\le 2N & : \ p\equiv s_k-s_j \,(\mathrm{mod}\,[d_j,e_j]),\\ (Q^*(p),W)=1 & : \ 1\le j<k\end{array}\right\}$$
$$= \sum_{\substack{d_j,e_j\\(d_ie_i,d_je_jP(z))=1\\1\le i,j<k,\ i\ne j}} \lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}\cdot \#\left\{\begin{array}{ll} N<p\le 2N & : \ p\equiv s_k-s_j \,(\mathrm{mod}\,[d_j,e_j]),\\ (Q^*(p),W)=1 & : \ 1\le j<k\end{array}\right\}$$
$$+ O(M^2N^{1/2}),$$

as in the proof of Lemma 9.1.1. Again, as in the proof of this other lemma, we use Theorem 3.3.1, which we apply to the set $\mathcal{A} = \{Q^*(p) : N < p \le 2N,\ p \equiv s \,(\mathrm{mod}\, q)\}$ with

$$q = \prod_{j=1}^{k-1}[d_j, e_j] \le D^2 \le N^{1/2}e^{-2\sqrt{\log N}}$$

and an appropriate $s \in (\mathbb{Z}/q\mathbb{Z})^*$, and with set of weights $w_a = 1$ for all a. The parameter u in Theorem 3.3.1 will be defined via the relation $z^u = e^{\sqrt{\log N}}$. Consider $d|W$ with $d \le z^u = e^{\sqrt{\log N}}$, so that $(d, q) = 1$. If N is large enough, so that $(p, d) = 1$ for all primes $p > N$, we have that

$$\#\{a\in\mathcal{A} : d|a\} = \sum_{\substack{b\in(\mathbb{Z}/d\mathbb{Z})^*\\Q^*(b)\equiv 0\,(\mathrm{mod}\,d)}} \#\left\{N<p-s_k\le 2N : \begin{array}{l} p\equiv a\,(\mathrm{mod}\,q)\\ p\equiv b\,(\mathrm{mod}\,d)\end{array}\right\}$$
$$= \frac{\nu^*(d)}{\varphi(d)}\cdot\frac{\int_N^{2N}\mathrm{d}t/\log t}{\varphi(q)} + O(\nu^*(d)E(dq)),$$

where

$$\nu^*(d) = \#\{b\in(\mathbb{Z}/d\mathbb{Z})^* : Q^*(b)\equiv 0\,(\mathrm{mod}\,d)\},\quad \text{so that}\quad \nu^*(p) = \nu(p) - 1 \le k-1,$$

and

$$E(m) = \max_{(c,m)=1} \left| \sum_{\substack{N<p\le 2N \\ p\equiv c \,(\mathrm{mod}\, m)}} 1 - \frac{\int_N^{2N} \mathrm{d}t/\log t}{\varphi(m)} \right|.$$

Therefore, Theorem 3.2.1 implies that

$$\sum_{\substack{N<p\le 2N \\ (Q^*(p),W)=1 \\ p\equiv a \,(\mathrm{mod}\, q)}} 1 = \frac{V\cdot \int_N^{2N} \mathrm{d}t/\log t}{\varphi(q)} + O_C\left(\frac{N}{\varphi(q) e^{(\log\sqrt{N})^{1/4}}} + \sum_{\substack{d\le e^{\sqrt{\log N}} \\ d|W}} \tau_{k-1}(d) E(dq) \right),$$

where

$$V := \prod_{p|W} \left(1 - \frac{\nu^*(p)}{p-1}\right) = \prod_{p|W} \left(1 - \frac{\nu(p)}{p}\right) \left(1 - \frac{1}{p}\right)^{-1}.$$

Consequently,

$$\begin{aligned} T_k = XV \sum_{\substack{(d_i e_i, d_j e_j W)=1 \\ 1\le i,j<k,\ i\ne j}} \frac{\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{\varphi([d_1,e_1])\cdots\varphi([d_{k-1},e_{k-1}])} \\ + O_C\left(\frac{M^2 N}{e^{(\frac{1}{2}\log N)^{1/4}}} + M^2 \sum_{m\le D^2 e^{\sqrt{\log N}}} \tau_{3k}(m) E(m) \right), \end{aligned}$$

where we used the inequality

$$\sum_{[d_1,e_1]\cdots[d_k,e_k]=m} 1 \le \tau_{3k}(m).$$

We claim that the sum over m in the error term is $\ll N/(\log N)^A$; indeed, the Brun-Titchmarsch inequality implies that

$$E(m) \ll \sqrt{E(m)\cdot \frac{N}{\varphi(m)\log N}} \quad (m\le \sqrt{N}),$$

so the claimed estimate follows by the Cauchy-Schwarz inequality and the Bombieri-Vinogradov theorem. We thus conclude that

$$T_k = V\int_N^{2N} \frac{\mathrm{d}t}{\log t} \cdot \sum_{\substack{(d_i e_i, d_j e_j W)=1 \\ 1\le i,j<k,\ i\ne j}} \frac{\lambda_{\boldsymbol{d}}\lambda_{\boldsymbol{e}}}{\varphi([d_1,e_1])\cdots\varphi([d_k,e_k])} + O_C\left(\frac{M^2N}{(\log N)^C}\right).$$

As in the proof of Lemma 9.1.1, we may remove the conditions $(d_i e_i, d_j e_j) = 1$ and we may replace $\varphi([d_j,e_j])$ by $[d_j,e_j]$ at the cost of introducing an error of total size $O(M^2 N(\log N)^{3k-2-C})$, which is of admissible size. Moreover, using the formula

$$\frac{1}{[d,e]} = \frac{1}{de}\sum_{a|(d,e)} \varphi(a),$$

we find that

$$\sum_{\substack{(d_ie_i,W)=1\\1\le i<k}}\frac{\lambda_{\boldsymbol{d},1}\lambda_{\boldsymbol{e},1}}{[d_1,e_1]\cdots[d_{k-1},e_{k-1}]}=\sum_{(a_1\cdots a_{k-1},W)=1}\varphi(a_1)\cdots\varphi(a_{k-1})\Bigg(\sum_{\substack{(d_i,W)=1\\d_i\equiv 0\,(\mathrm{mod}\,a_i)\\1\le i<k}}\frac{\lambda_{\boldsymbol{d},1}}{d_1\cdots d_{k-1}}\Bigg)^2.$$

Finally, we replace $\varphi(a_j)$ by a_j, thus introducing a total error of size $O(M^2N(\log N)^{3k-2}/z)$, which is admissible. This completes the proof of the lemma. □

9.2 A reduction to an optimization problem

In this section, we will simplify significantly the estimate for our prime gap detector. Recall the definition of $\boldsymbol{\xi}_{\boldsymbol{a}}$ in the statement of Lemma 9.1.1. We want to rewrite the parameters $\lambda_{\boldsymbol{d}}$ with respect to the parameters $\xi_{\boldsymbol{a}}$. First of all, note that $\xi_{\boldsymbol{a}}$ is supported on k-tuples with $a_1\cdots a_k\le D$ and $(a_1\cdots a_k,W)=1$. As in Section 3.4 (see relation (3.4.4)), if $(d_1\cdots d_k,W)=1$, we have the inversion formula

$$\lambda_{\boldsymbol{d}}=\sum_{(b_1\cdots b_k,W)=1}\frac{\mu(b_1)\cdots\mu(b_k)\xi_{d_1b_1,\dots,d_kb_k}}{b_1\cdots b_k}. \tag{9.2.1}$$

Consequently, if $(a_1\cdots a_k,W)=1$ with $a_{j_0}=1$, then

$$\begin{aligned}\sum_{\substack{(d_1\cdots d_k,W)=1)\\d_{j_0}=1}}\frac{\lambda_{a_1d_1,\dots,a_kd_k}}{d_1\cdots d_k}&=\sum_{\substack{(d_1\cdots d_k,W)=1\\d_{j_0}=1}}\frac{1}{d_1\cdots d_k}\sum_{(b_1\cdots b_k,W)=1}\frac{\mu(b_1)\cdots\mu(b_k)\xi_{a_1d_1b_1,\dots,a_kd_kb_k}}{b_1\cdots b_k}\\&=\sum_{(m_1\cdots m_k,W)=1}\frac{\xi_{a_1m_1,\dots,a_km_k}}{m_1\cdots m_k}\sum_{\substack{b_jd_j=m_j\\1\le j\le k\\d_{j_0}=1}}\mu(b_1)\cdots\mu(b_k)\\&=\sum_{(m_{j_0},W)=1}\frac{\mu(m_{j_0})\xi_{a_1,\dots,a_{j_0-1},m_{j_0},a_{j_0+1},\dots,a_k}}{m_{j_0}}.\end{aligned} \tag{9.2.2}$$

This computation suggests setting

$$\xi_{\boldsymbol{a}}:=\mathbf{1}_{(a_1\cdots a_k,W)=1}\cdot\lambda(a_1\cdots a_k)\cdot f\left(\frac{\log a_1}{\log D},\dots,\frac{\log a_k}{\log D}\right)\cdot\prod_{p\le z}\left(1-\frac{1}{p}\right)^{-k},$$

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 (9.2.2)) and f is a smooth function supported on the simplex

$$\Delta_k:=\{\boldsymbol{t}\in[0,1]^k:t_1+\cdots+t_k\le 1\}.$$

With this choice, we have that the following result.

Lemma 9.2.1. *If $z = (\log N)^{5k+1}$, $k \geq 2$, $D = N^{1/4}e^{-\sqrt{\log N}}$ and*

$$S = \sum_{\substack{N<n\leq 2N\\(Q(n),W)=1}} \left(\sum_{j=1}^{k} \mathbf{1}_{\mathbb{P}}(n+s_j) - m\right)\left(\sum_{\substack{d_j \mid n+s_j\\1\leq j\leq k}} \lambda_{\boldsymbol{d}}\right)^2,$$

and

$$\mathfrak{S}(\boldsymbol{s}) = \prod_{p}\left(1-\frac{\nu(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k},$$

then

$$\frac{S}{\mathfrak{S}(\boldsymbol{s})N(\log D)^k} = \frac{1}{4}\sum_{j_0=1}^{k}\int\left(\int f(\boldsymbol{t})\mathrm{d}t_{j_0}\right)^2 \prod_{\substack{1\leq j\leq k\\ j\neq j_0}} dt_j - m\int f(\boldsymbol{t})^2\mathrm{d}\boldsymbol{t} + O\left(\frac{1}{\sqrt{\log N}}\right);$$

the implied constant depends at most on ϵ, f, k and $\boldsymbol{s}$.

Except for Lemmas 9.1.1 and 9.1.2, the key input to the proof of the above lemma comes from the following result:

Lemma 9.2.2. *Let D and z be two parameters with $z \leq (\log D)^C$. If $W = \prod_{p\leq z} p$ and $g : \mathbb{R}^k \to \mathbb{R}$ is a smooth function supported on $\{\boldsymbol{t} \in [1,+\infty)^k : t_1\cdots t_k \leq D\}$ and such that $\frac{\partial g}{\partial t_j} \ll 1/t_j$, then*

$$\sum_{(n_1\cdots n_k,W)=1}\frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \left(\frac{\varphi(W)}{W}\right)^k\left(\int\frac{g(t_1,\dots,t_k)}{t_1\cdots t_k}\mathrm{d}\boldsymbol{t} + O((\log D)^{k-1/2})\right);$$

the implied constant depends at most on k, C and g.

Proof. All implied constants might depend on g and on k. We note that

$$\begin{aligned}
&\sum_{(n_1\cdots n_k,W)=1}\frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} - \sum_{\substack{(n_j,W)=1\\ n_j>e^{\sqrt{\log D}}\\1\leq j\leq k}}\frac{g(n_1,\dots,n_k)}{n_1\cdots n_k}\\
&\ll \sum_{j_0=1}^{k}\sum_{\substack{(n_1\cdots n_k,W)=1\\ n_j\leq D\ (j\neq j_0)\\ n_{j_0}\leq e^{\sqrt{\log D}}}}\frac{1}{n_1\cdots n_k}\\
&\leq k\left(\prod_{\substack{p\leq e^{\sqrt{\log D}}\\ p\nmid W}}\left(1-\frac{1}{p}\right)^{-1}\right)\left(\prod_{\substack{p\leq D\\ p\nmid W}}\left(1-\frac{1}{p}\right)^{-1}\right)^{k-1}\\
&\ll (\log D)^{k-1/2}\left(\frac{\varphi(W)}{W}\right)^k,
\end{aligned}$$

which is of admissible size. We split the range of summation $\{\boldsymbol{n} : n_1 \cdots n_k \le D,\, n_j > e^{\sqrt{\log D}}\,(1 \le j \le k)\}$ into small cubes of the form $B = \prod_{j=1}^{k}(x_j, x_j + \sqrt{x_j}]$. We note that

$$\frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \frac{g(x_1,\dots,x_k)}{x_1\cdots x_k} + O\left(\frac{\max_{1\le j\le k} x_j^{-1/2}}{x_1\cdots x_k}\right) = \frac{I}{\sqrt{x_1\cdots x_k}} + O\left(\frac{e^{-\sqrt{\log D}/2}}{x_1\cdots x_k}\right)$$

by our assumption on g, where

$$I := \int_B \frac{g(t_1,\dots,t_k)}{t_1\cdots t_k}\mathrm{d}\boldsymbol{t}.$$

Therefore

$$\sum_{\substack{(n_1\cdots n_k, W)=1\\ \boldsymbol{n}\in B}} \frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \frac{I}{\sqrt{x_1\cdots x_k}} \prod_{j=1}^{k} \sum_{\substack{x_j < n_j \le x_j + \sqrt{x_j}\\ (n_j, W)=1}} 1 + O\left(\frac{e^{-\sqrt{\log D}/2}}{(\log z)^k \sqrt{x_1\cdots x_k}}\right).$$

Applying Theorem 3.2.1 with $\mathcal{A} = \{x_j < n_j \le x_j + \sqrt{x_j}\}$, $\mathcal{P} = \{p \le z\}$, $X = \sqrt{x_j}$, $g(d) = 1/d$ and $r_d = O(1)$, we find that

$$\sum_{\substack{x_j < n_j \le x_j + \sqrt{x_j}\\ (n_j, W)=1}} 1 = \frac{\varphi(W)}{W}\sqrt{x_j} + O\left(\frac{\sqrt{x_j}}{e^{(\log D)^{1/4}}}\right),$$

by our assumption that $\log x_j \ge \sqrt{\log D}$. Therefore

$$\sum_{\substack{(n_1\cdots n_k, W)=1\\ \boldsymbol{n}\in B}} \frac{g(n_1,\dots,n_k)}{n_1\cdots n_k} = \left(\frac{\varphi(W)}{W}\right)^k \cdot I + O\left(\frac{1}{(\log D)^2}\right).$$

Summing the above estimate over all cubes $B \subset \{\boldsymbol{n} : n_1\cdots n_k \le D,\, n_j > e^{\sqrt{\log D}}\,(1 \le j \le k)\}$ completes the proof of the lemma. □

Proof of Lemma 9.2.1. 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 9.1.1 and 9.2.2 imply that

$$\sum_{\substack{N<n\le 2N\\(Q(n),W)=1}}\left(\sum_{\substack{d_j\mid n+s_j\\1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2=\frac{N}{V^{2k}}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\sum_{(a_1\cdots a_k,W)=1}\frac{g(\boldsymbol{a})^2}{a_1\cdots a_k}+O(N)$$
$$=\frac{N}{V^k}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\left(\int\frac{g(\boldsymbol{u})^2}{u_1\cdots u_k}\mathrm{d}\boldsymbol{u}+O((\log N)^{k-1/2})\right),$$

since we have assumed that $z=(\log N)^{5k+2}$. Moreover,

$$V^{-k}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)=\mathfrak{S}(\boldsymbol{s})\left(1+O\left(\frac{1}{z}\right)\right) \tag{9.2.3}$$

and

$$\int\frac{g(\boldsymbol{u})^2}{u_1\cdots u_k}\mathrm{d}\boldsymbol{u}=(\log D)^k\int_{\Delta_k}f(\boldsymbol{t})^2\mathrm{d}\boldsymbol{t},$$

so that

$$\sum_{\substack{N<n\le 2N\\(Q(n),W)=1}}\left(\sum_{\substack{d_j\mid n+s_j\\1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2=\mathfrak{S}(\boldsymbol{s})N(\log D)^k\left(\int f(\boldsymbol{t})^2\mathrm{d}\boldsymbol{t}+O\left(\frac{1}{\sqrt{\log N}}\right)\right). \tag{9.2.4}$$

Next, if we set

$$S_{j_0}(\boldsymbol{a})=\sum_{(b,W)=1}\frac{\mu^2(b)g(a_1,\dots,a_{j_0-1},b,a_{j_0+1},\dots,a_k)}{b},$$

where $j_0\in\{1,\dots,k\}$, then Lemma 9.1.2 and relations (9.2.2) and (9.2.3) imply that

$$\sum_{\substack{N<n\le 2N\\(Q(n),W)=1}}\mathbf{1}_{\mathbb{P}}(n+s_{j_0})\left(\sum_{\substack{d_j\mid n+s_j\\1\le j\le k}}\lambda_{\boldsymbol{d}}\right)^2=\frac{X}{V^{2k+1}}\prod_{p\le z}\left(1-\frac{\nu(p)}{p}\right)\sum_{\substack{(a_1\cdots a_k,W)=1\\a_{j_0}=1}}\frac{S_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}+O(N)$$
$$=\frac{\mathfrak{S}(\boldsymbol{s})N}{V^{k+1}\log N}\left(\sum_{\substack{(a_1\cdots a_k,W)=1\\a_{j_0}=1}}\frac{S_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}+O((\log N)^{3/2})\right).$$

If $\mu^2(b)=0$ and $P^-(b)>z$, then b is divisible by the square of a prime $>z$. Therefore,

$$S_{j_0}(\boldsymbol{a})=\sum_{(b,W)=1}\frac{g(a_1,\dots,a_{j_0-1},b,a_{j_0+1},\dots,a_k)}{b}+O\left(\frac{\log N}{z}\right)$$
$$=V\cdot\left(\int\frac{g(a_1,\dots,a_{j_0-1},u,a_{j_0+1},\dots,a_k)}{u}\mathrm{d}u+O\left(\sqrt{\log N}\right)\right)$$
$$=V\cdot(\log D)\cdot\left(G_{j_0}(a_1,\dots,a_k)+O\left(\frac{1}{\sqrt{\log N}}\right)\right)$$

by Lemma 9.2.1, where

$$G_{j_0}(a_1,\dots,a_k) := \int g(a_1,\dots,a_{j_0-1},D^{t_{j_0}},a_{j_0+1},\dots,a_k)\mathrm{d}t_{j_0} \ll 1.$$

Therefore

$$\sum_{\substack{N<n\le 2N\\ P^-(Q(n))>z}} \mathbf{1}_{\mathbb{P}}(n+s_{j_0})\Bigg(\sum_{\substack{d_j\mid n+s_j\\ 1\le j\le k}}\lambda_{\boldsymbol{d}}\Bigg)^2 = \frac{\mathfrak{S}(\boldsymbol{s})N(\log D)^2}{V^{k-1}\log N}\Bigg(\sum_{\substack{(a_1\cdots a_k,W)=1\\ a_{j_0}=1}}\frac{G_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k}+O\left(\frac{1}{\sqrt{\log N}}\right)\Bigg).$$

Finally, applying again Lemma 9.2.2, we find that

$$\begin{aligned}\sum_{\substack{P^-(a_j)>z\\ 1\le j\le k\\ a_{j_0}=1}}\frac{G_{j_0}(\boldsymbol{a})^2}{a_1\cdots a_k} &= V^{k-1}\Bigg(\int_{\substack{\boldsymbol{u}\in\mathbb{R}^k\\ u_{j_0}=1}}\frac{G_{j_0}(\boldsymbol{u})^2}{u_1\cdots u_k}\prod_{\substack{1\le j\le k\\ j\ne j_0}}\mathrm{d}u_j+O((\log N)^{k-3/2})\Bigg)\\ &= V^{k-1}(\log D)^{k-1}\Bigg(\int\Bigg(\int f(\boldsymbol{t})\mathrm{d}t_{j_0}\Bigg)^2\prod_{\substack{1\le j\le k\\ j\ne j_0}}dt_j+O\left(\frac{1}{\sqrt{\log N}}\right)\Bigg).\end{aligned}$$

Since

$$\log D = \frac{\log N}{4}+O(\sqrt{\log N}),$$

Lemma 9.2.1 follows. □

9.3 Optimizing the choice of weights

In view of Lemma 9.2.1, it is clear that our goal is to choose f supported on Δ_k and maximizing the ratio

$$\rho(f) := \frac{1}{k}\sum_{j=1}^{k}\frac{\int\left(\int f(\boldsymbol{t})\mathrm{d}t_j\right)^2\mathrm{d}t_1\cdots\mathrm{d}t_{j-1}\mathrm{d}t_{j+1}\cdots\mathrm{d}t_k}{\int f(\boldsymbol{t})^2\mathrm{d}\boldsymbol{t}}.$$

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

$$(\mathcal{L}_k f)(\boldsymbol{t}) := \frac{1}{k}\sum_{j=1}^{k}\int f(t_1,\dots,t_{j-1},u,t_{j+1},\dots,t_k)\mathrm{d}u$$

and the corresponding eigenvalue is the maximum possible ratio $\rho=\rho(f)$. Indeed, for a function f supported on Δ_k, we have that

$$\rho(f) = \frac{\langle\mathcal{L}_k f,f\rangle}{\langle f,f\rangle},$$

where

$$\langle g, h\rangle := \int_{\Delta_k} g(\boldsymbol{t})h(\boldsymbol{t})\mathrm{d}\boldsymbol{t}.$$

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 Δ_k. So its derivative at $\epsilon = 0$ must vanish, which implies that

$$\langle \mathcal{L}_k f, g\rangle + \langle \mathcal{L}_k g, f\rangle = 2\rho(f)\langle f, g\rangle.$$

It is easy to see that $\mathcal{L}_k$ is a self-adjoint operator, so this implies that

$$\langle \mathcal{L}_k f, g\rangle = \rho(f)\langle f, g\rangle.$$

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

$$\rho(f) = \frac{\int \left(\int f(\boldsymbol{t})\mathrm{d}t_k\right)^2 \mathrm{d}t_1 \cdots \mathrm{d}t_{k-1}}{\int f(\boldsymbol{t})^2 \mathrm{d}\boldsymbol{t}}.$$

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

$$M_k := \sup\{\rho(f) : f : \mathbb{R}^k \to \mathbb{R},\ \operatorname{supp}(f) \subset \Delta_k,\ f \text{ symmetric and measurable}\}$$

An asymptotic estimation for M_k is given in Lemma 9.3.1 below. For explicit bounds on M_k, the reader is invited to consult the paper by Maynard [17] as well [20].

Remark 9.3.1. 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

$$\rho(f) = (k-1)\cdot \frac{\int_0^1 u^{k-2}(\int_u^1 F)^2 \mathrm{d}u}{\int_0^1 u^{k-1}F(u)^2\mathrm{d}u}.$$

It is possible to show that $\rho(f) \leq 4/k$ in this special case [25], 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

$$\rho(f) = \frac{k-1}{(\ell+1)^2} \cdot \frac{\int_0^1 u^{k-2}(1-u)^{2\ell+2}\mathrm{d}u}{\int_0^1 u^{k-1}(1-u)^{2\ell}\mathrm{d}u} = \frac{k-1}{(\ell+1)^2} \cdot \frac{\frac{(k-2)!(2\ell+2)!}{(k+2\ell+1)!}}{\frac{(k-1)!(2\ell!)}{(k+2\ell)!}}$$
$$= \frac{2(2\ell+1)}{(\ell+1)(k+2\ell+1)} \sim \frac{4}{k}$$

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 9.1.2, which would have allowed to take D slightly larger, we would have been able to prove Theorem 9.0.1 when $m = 1$ with these weights. This is precisely what Zhang did in [29]. 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 9.3.1. *For large integers k, we have that*

$$\frac{\log k - 4\log\log k + O(1)}{k} \leq M_k \leq \frac{\log k + \log\log k + O(1)}{k}$$

Proof. For the lower bound, we consider functions of the form

$$f(t_1, \dots, t_k) = \mathbf{1}_{\Delta_k}(t_1, \dots, t_k) \prod_{j=1}^{k} g(kt_j),$$

where $g : [0, +\infty) \to [0, +\infty)$ is a function supported on the interval $[0, T]$ and such that $\int_0^\infty g(t)^2\mathrm{d}t = 1$. Then

$$\int f(\boldsymbol{t})^2\mathrm{d}\boldsymbol{t} \leq \left(\int g(kt)^2\mathrm{d}t\right)^k = \frac{1}{k^k},$$

so that

$$\rho(f) \geq \frac{1}{k}\int g(t_1)^2 \cdots g(t_{k-1})^2 \left(\int_0^{k-t_1-\cdots-t_{k-1}} g(t_k)\mathrm{d}t_k\right)^2 \mathrm{d}t_1 \cdots \mathrm{d}t_{k-1}$$
$$\geq \frac{\left(\int_0^\infty g(t)\mathrm{d}t\right)^2}{k} \int_{t_1+\cdots+t_{k-1}\leq k-T} g(t_1)^2 \cdots g(t_{k-1})^2 \mathrm{d}t_1 \cdots \mathrm{d}t_{k-1}$$
$$= \frac{\left(\int_0^\infty g(t)\mathrm{d}t\right)^2}{k} \cdot \mathbf{Prob}\left(X_1 + \cdots + X_{k-1} \leq k - T\right),$$

where $X_1, \dots, X_{k-1}$ are independent random variables with density function g^2. Let

$$\mu = \mathbb{E}[X_1] = \int t g(t)^2\mathrm{d}t$$

and $Y_i = X_i - \mu$, $1 \leq i \leq k$, 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

$$\mathbb{E}[X_1^2] = \int t^2 g(t)^2 \mathrm{d}t \le T\int t g(t)^2 \mathrm{d}t = T\mu$$

by our assumption that g is supported in $[0,T]$, we deduce that

$$\rho(f) \ge \frac{\left(\int_0^\infty g(t)\mathrm{d}t\right)^2}{k}\cdot\left(1-\frac{kT\mu}{(k-T-k\mu)^2}\right)$$

for any measurable function $g\ge 0$ supported on $[0,T]$ with $\int g(t)^2\mathrm{d}t = 1$ and $\mu = \int t g(t)^2\mathrm{d}t \le 1-T/k$. We choose

$$g(t) = c\cdot\frac{\mathbf{1}_{[0,T]}(t)}{1+At}.$$

In order to have that $\int g(t)^2\mathrm{d}t = 1$, we take

$$c^2 = \left(\int_0^T \frac{\mathrm{d}t}{(1+At)^2}\right)^{-1} = \frac{A}{1-1/(1+AT)} = A+\frac{1}{T}.$$

We then have that

$$\begin{aligned}\mu = \int_0^T \frac{c^2 t}{(1+At)^2}\mathrm{d}t = \frac{c^2}{A^2}\int_0^{AT}\frac{t}{(1+t)^2}\mathrm{d}t &= \frac{c^2}{A^2}\left(\log(1+AT)-1+\frac{1}{1+AT}\right)\\ &= \frac{\log(AT)}{A}\left(1+O\left(\frac{1}{\log(AT)}\right)\right).\end{aligned}$$

This suggests choosing $A\sim\log T$. We take $T = k/(\log k)^3$ and $A = \log k$, so that

$$\mu = \frac{\log(k/(\log k)^2)}{\log k}\left(1+O\left(\frac{1}{\log k}\right)\right) = 1-\frac{2\log\log k}{\log k}+O\left(\frac{1}{\log k}\right) \le 1-\frac{T}{k}-\frac{\log\log k}{\log k}$$

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

$$\begin{aligned}k\cdot\rho(f) \ge \left(\int_0^\infty g(t)\mathrm{d}t\right)^2\cdot\left(1-\frac{T\mu}{k(1-T/k-\mu)^2}\right) &= \frac{c^2\log^2(1+AT)}{A^2}\left(1+O\left(\frac{1}{\log k}\right)\right)\\ &= \frac{\log^2(k/(\log k)^2)}{\log k}\left(1+O\left(\frac{1}{\log k}\right)\right)\\ &= \log k - 4\log\log k + O(1),\end{aligned}$$

as claimed.

Finally, we prove the upper bound on $\rho(f)$. Let f be a symmetric measurable function supported on Δ_k. Motivated by the choice for f above, we use the Cauchy-Schwarz inequality in the following fashion:

$$\begin{aligned}\left(\int f(t_1,\dots,t_k)\mathrm{d}t_k\right)^2 &= \left(\int_0^1 f(t_1,\dots,t_k)\mathrm{d}t_k\right)^2\\ &\le \left(\int (1+kAt_k)f(t_1,\dots,t_k)^2\mathrm{d}t_k\right)\cdot\left(\int_0^1 \frac{\mathrm{d}t_k}{1+kAt_k}\right)\\ &= \frac{\log(1+kA)}{kA}\int (1+kAt_k)f(t_1,\dots,t_k)^2\mathrm{d}t_k.\end{aligned}$$

Therefore

$$\int\left(\int f(t_1,\dots,t_k)\mathrm{d}t_k\right)^2\mathrm{d}t_1\cdots\mathrm{d}t_{k-1} \le \frac{\log(1+kA)}{kA}\int (1+kAt_k)f(t_1,\dots,t_k)^2\mathrm{d}\boldsymbol{t}.$$

By symmetry,

$$\int\left(\int f(t_1,\dots,t_k)\mathrm{d}t_k\right)^2\mathrm{d}t_1\cdots\mathrm{d}t_{k-1} \le \frac{\log(1+kA)}{kA}\int (1+kAt_j)f(t_1,\dots,t_k)^2\mathrm{d}\boldsymbol{t}$$

for all $j\in\{1,\dots,k\}$. So, summing over j and using the fact that $t_1+\cdots+t_k\le 1$ in the support of f, we find that

$$\begin{aligned}k\cdot\rho(f) &\le \frac{\log(1+kA)}{kA}\cdot\frac{\int_{\Delta_k}(k+kA(t_1+\cdots+t_k))f(t_1,\dots,t_k)^2\mathrm{d}\boldsymbol{t}}{\int f(\boldsymbol{t})^2\mathrm{d}\boldsymbol{t}}\\ &\le \frac{(1+A)\log(1+kA)}{A}.\end{aligned}$$

for any $A>0$. So, setting $A=\log k$ yields that

$$k\cdot\rho(f)\le \log k+\log\log k+O(1)$$

for all symmetric functions f supported on Δ_k. This completes the proof of the lemma. □

It is now easy to complete the proof of Theorem 9.0.1:

Proof of Theorem 9.0.1. Combining Lemmas 9.2.1 and 9.3.1, we find that there is a choice of the parameters $\lambda_{\boldsymbol{d}}$ such that

$$\frac{S(N,z)}{\mathfrak{S}(\boldsymbol{s})N(\log D)^k}\ge\frac{\log k-4\log\log k+O(1)-4m}{4}+O\left(\frac{1}{\sqrt{\log N}}\right).$$

So, if $k=\lfloor Cm^4e^{4m}\rfloor$ for a large enough constant C, then $S(N,z)>0$ for large enough N, which implies that $\liminf_{n\to\infty}(p_{n+m}-p_n)\le s_k-s_1$. 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 9.0.1. □

Chapter 10

The analytic theory of Dirichlet L-functions

This chapter is devoted to the study of the analytic properties of the Riemann zeta function and the Dirichlet L-functions. As we saw in the previous chapters, understanding better these functions has direct consequences on our understanding about the distribution of primes in arithmetic progressions. Note that $\zeta(s) = L(s, 1)$. Thus, we will work directly with the more general functions $L(s, \chi)$. Throughout this chapter, we will be assuming that χ is primitive of conductor q, so that $L(s, \chi) = \zeta(s)$ if, and only if, $q = 1$ (and thus $\chi = 1$).

Note that if $\chi' \pmod{q'}$ is imprimitive and induced by the primitive character $\chi \pmod{q}$, then

$$L(s, \chi') = L(s, \chi) \prod_{p|q} \left(1 - \frac{\chi(p)}{p^s}\right).$$

Thus, the theory for L-functions of primitive characters has a direct translation to L-functions of general characters.

10.1 The functional equation

Let $\chi \pmod{q}$ be a primitive character. We will show that $L(s, \chi)$ has a meromorphic continuation to $\mathbb{C}$ whose only pole is a simple pole at $s = 1$ of residue 1, arising only when $\chi = 1$. Moreover, $L(s, \chi)$ possesses a fundamental symmetry with respect to the so-called critical line $\mathrm{Re}(s) = 1/2$. Indeed, we set

$$\xi(s, \chi) = \left(\frac{s(s-1)}{2}\right)^{\mathbf{1}_{\chi=1}} \cdot \left(\frac{q}{\pi}\right)^{\frac{s+a}{2}} \Gamma\left(\frac{s+a}{2}\right) L(s, \chi),$$

where[1]

$$a := \begin{cases} 0 & \text{if } \chi(-1) = 1, \\ 1 & \text{if } \chi(-1) = -1. \end{cases} \tag{10.1.1}$$

[1]Note that $\chi(-1)^2 = \chi((-1)^2) = \chi(1) = 1$, so that $\chi(-1) \in \{-1, 1\}$. If $\chi(-1) = 1$, then we call χ an even character, whereas if $\chi(-1) = -1$, then we call it an odd character.

Then

$$\xi(1-s,\overline{\chi}) = \frac{i^a\sqrt{q}}{\mathcal{G}(\chi)}\xi(s,\chi). \tag{10.1.2}$$

We will prove the functional equation (10.1.2) as an application of the Poisson summation formula. Let $f:\mathbb{R}\to\mathbb{R}$ be a function in the Schwarz class, that is to say f is differentiable an infinite number of times and $f^{(j)}(x) \ll_{j,n} 1/(1+|x|^n)$ for $j\in\mathbb{N}\cup\{0\}$, $n\in\mathbb{N}$ and $x\in\mathbb{R}$. We then consider its *Mellin transform*

$$F(s) = \int_0^\infty f(y)y^{s-1}\mathrm{d}y,$$

which is well defined for $\mathrm{Re}(s)>0$. Making the change of variable $y\to ny/\sqrt{q}$, we find that

$$F(s) = \frac{n^s}{q^{s/2}}\int_0^\infty f\left(\frac{ny}{\sqrt{q}}\right)y^{s-1}\mathrm{d}s.$$

(The factor $1/\sqrt{q}$ is inserted for the simplification of the argument later on; we could have worked with the change of variables $y\to ny$ instead.) Multiplying the above formula by $q^s\chi(n)/n^s$ and the resulting identity over all $n\in\mathbb{N}$, we deduce that

$$q^{s/2}L(s,\chi)F(s) = \int_0^\infty S_f(y,\chi)y^{s-1}\mathrm{d}y \quad (\mathrm{Re}(s)>1),$$

where

$$S_f(y,\chi) := \sum_{n=1}^\infty \chi(n)f\left(\frac{ny}{\sqrt{q}}\right).$$

We break the range of integration as $[0,1]\cup[1,+\infty)$ and make the change of variable $y\to 1/y$ in the first portion. Thus

$$q^{s/2}L(s,\chi)F(s) = \int_1^\infty S_f(y,\chi)y^{s-1}\mathrm{d}y + \int_1^\infty S_f(1/y,\chi)y^{-s-1}\mathrm{d}y \quad (\mathrm{Re}(s)>1). \tag{10.1.3}$$

We want to express $S_f(1/y,\chi)$ in terms of $S_g(y,\psi)$ for another function g and another character f. We will do so using the Poisson summation formula. In order to pass from a sum running over $n\in\mathbb{N}$ to a sum running over $n\in\mathbb{Z}$, we impose the additional assumption that f and χ have the same parity, that is to say $f(-x)=\chi(-1)f(x)$. Then

$$2\cdot S_f(1/y,\chi) = \sum_{n\in\mathbb{Z}}\chi(n)f\left(\frac{n}{y\sqrt{q}}\right) - \mathbf{1}_{\chi=1}\cdot f(0). \tag{10.1.4}$$

In order to apply the Poisson summation formula, we split the summands according to their residue class mod q:

$$\sum_{n\in\mathbb{Z}}\chi(n)f\left(\frac{n}{y\sqrt{q}}\right) = \sum_{a=1}^q\chi(a)\sum_{\substack{n\in\mathbb{Z}\\ n\equiv a\,(\mathrm{mod}\,q)}} f\left(\frac{n}{y\sqrt{q}}\right) = \sum_{a=1}^q\chi(a)\sum_{m\in\mathbb{Z}} f\left(\frac{a}{y\sqrt{q}}+\frac{m\sqrt{q}}{y}\right).$$

The Fourier transform of the function $u \to f(a/(y\sqrt{q}) + u\sqrt{q}/y)$ is the function

$$\xi \to \frac{y \cdot e(a\xi/q)}{\sqrt{q}} \widehat{f}\left(\frac{y\xi}{\sqrt{q}}\right).$$

Therefore, the Poisson summation formula and Theorem 5.3.2 imply that

$$\begin{aligned}\sum_{n\in\mathbb{Z}} \chi(n) f\left(\frac{n}{y\sqrt{q}}\right) &= \frac{y}{\sqrt{q}} \sum_{a=1}^{q} \chi(a) \sum_{m\in\mathbb{Z}} e\left(\frac{am}{q}\right) \widehat{f}\left(\frac{ym}{\sqrt{q}}\right) \\ &= \frac{y}{\sqrt{q}} \sum_{m\in\mathbb{Z}} \widehat{f}\left(\frac{ym}{\sqrt{q}}\right) \sum_{a=1}^{q} \chi(a) e\left(\frac{am}{q}\right) \\ &= \frac{y\mathcal{G}(\chi)}{\sqrt{q}} \sum_{m\in\mathbb{Z}} \overline{\chi}(m) \widehat{f}\left(\frac{ym}{\sqrt{q}}\right).\end{aligned}$$

Since f and $\widehat{f}$ have the same parity, applying (10.1.4) twice (once for f and once for $\widehat{f}$), we find that

$$\begin{aligned}S_f(1/y, \chi) &= \frac{y\mathcal{G}(\chi)}{2\sqrt{q}} \sum_{m\in\mathbb{Z}} \overline{\chi}(m) \widehat{f}\left(\frac{ym}{\sqrt{q}}\right) - \frac{\mathbf{1}_{\chi=1}}{2} \cdot f(0) \\ &= \frac{y\mathcal{G}(\chi)}{\sqrt{q}} \cdot S_{\widehat{f}}(y, \chi) + \frac{\mathbf{1}_{\chi=1}}{2} \cdot \left(\widehat{f}(0) \cdot y - f(0)\right).\end{aligned}$$

Inserting the above identity to (10.1.3), we conclude that

$$\begin{aligned}q^{s/2} L(s, \chi) F(s) = &\int_1^\infty S_f(y, \chi) y^{s-1} \mathrm{d}y + \frac{\mathcal{G}(\chi)}{\sqrt{q}} \int_1^\infty S_{\widehat{f}}(y, \overline{\chi}) y^{(1-s)-1} \mathrm{d}y \\ &+ \mathbf{1}_{\chi=1} \cdot \left(\frac{\widehat{f}(0)}{s-1} - \frac{f(0)}{s}\right).\end{aligned} \tag{10.1.5}$$

Finally, we use (10.1.5) with a particular choice of f to deduce (10.1.2). We take $f(x) = x^a e^{-\pi x^2}$, where a is defined by (10.1.1). For this choice of f, we have that

$$F(s) = \int_0^\infty e^{-\pi x^2} x^{s+a-1} \mathrm{d}x = \frac{1}{\pi^{(s+a)/2}} \int_0^\infty e^{-y} y^{(s+a)/2-1} \mathrm{d}y = \frac{\Gamma((s+a)/2)}{2\pi^{(s+a)/2}}.$$

Moreover, we claim that

$$\widehat{f} = \chi(-1) i^a f. \tag{10.1.6}$$

Indeed, if $a = 0$, then

$$\widehat{f}(\xi) = \int_{\mathbb{R}} e^{-\pi x^2 - 2\pi i x\xi} \mathrm{d}x = e^{-\pi\xi^2} \int_{\mathbb{R}} e^{-\pi(x+i\xi)^2} \mathrm{d}x = e^{-\pi\xi^2} = f(\xi), \tag{10.1.7}$$

whereas if $a = 1$, so that $\chi(-1) = -1$, then we note that

$$\widehat{f}(\xi) = -\frac{1}{2\pi} \int_{\mathbb{R}} (e^{-\pi x^2})' e^{-2\pi i\xi} \mathrm{d}x = -i\xi \int_{\mathbb{R}} e^{-\pi x^2} e^{-2\pi i\xi} \mathrm{d}x = -i\xi e^{-\pi\xi^2} = -i \cdot f(\xi),$$

as needed.

By the above discussion, (10.1.5) becomes

$$\begin{aligned}\frac{(q/\pi)^{(s+a)/2}\Gamma((s+a)/2)L(s,\chi)}{2q^{a/2}} &= \int_1^\infty S_f(y,\chi)y^{s-1}\mathrm{d}y \\ &\quad + \frac{\chi(-1)i^a\mathcal{G}(\chi)}{\sqrt{q}}\int_1^\infty S_f(y,\overline{\chi})y^{(1-s)-1}\mathrm{d}y \\ &\quad + \frac{\mathbf{1}_{\chi=1}}{2s(s-1)}.\end{aligned} \tag{10.1.8}$$

The right hand side is a meromorphic function over the entire complex plane, whose only singularities are at $s=0$ and $s=1$, occurring only when $\chi=1$. Since Γ is a meromorphic function that does not vanish and whose only singularities are simple poles at $0,-1,-2,\dots$ (cf. Appendix A), we deduce the following:

(a) If $\chi\neq 1$, then $L(s,\chi)$ has an analytic continuation over the entire complex plane.

(b) If $\chi=1$, then $L(s,\chi)=\zeta(s)$ has a meromorphic continuation over the entire complex plane, with its only pole being a simple pole at $s=1$ of residue 1. There is no pole at 0; the pole of $\xi(s,1)$ at $s=0$ is due to the pole of Γ there. In fact, it is possible to show that $\zeta(0)=-1$.

Moreover, (10.1.8) clearly implies that there is a certain symmetry under the change of variables $s\to 1-s$ and $\chi\to\overline{\chi}$. Indeed, applying (10.1.8) with $1-s$ in place of s and $\overline{\chi}$ in place of χ, we find that

$$\begin{aligned}\frac{(q/\pi)^{(1-s+a)/2}\Gamma((1-s+a)/2)L(s,\overline{\chi})}{2q^{a/2}} &= \int_1^\infty S_f(y,\overline{\chi})y^{(1-s)-1}\mathrm{d}y \\ &\quad + \frac{\chi(-1)i^a\mathcal{G}(\overline{\chi})}{\sqrt{q}}\int_1^\infty S_f(y,\chi)y^{s-1}\mathrm{d}y \\ &\quad + \frac{\mathbf{1}_{\chi=1}}{2s(s-1)}.\end{aligned} \tag{10.1.9}$$

Moreover, we have that

$$\frac{\chi(-1)i^a\mathcal{G}(\overline{\chi})}{\sqrt{q}} = \frac{i^a\overline{\mathcal{G}(\chi)}}{\sqrt{q}} = \frac{i^a\sqrt{q}}{\mathcal{G}(\chi)}$$

using the fact that $|\mathcal{G}(\chi)|^2=q$ (see Theorem 5.3.3), and that

$$\left(\frac{\chi(-1)i^a\mathcal{G}(\overline{\chi})}{\sqrt{q}}\right)^{-1} = \frac{\mathcal{G}(\chi)}{i^a\sqrt{q}} = \frac{\chi(-1)i^a\mathcal{G}(\chi)}{\sqrt{q}}$$

by the definition of a. Thus (10.1.9) becomes

$$\begin{aligned}\frac{(q/\pi)^{(1-s+a)/2}\Gamma((1-s+a)/2)L(s,\overline{\chi})}{2q^{a/2}} &= \frac{i^a\sqrt{q}}{\mathcal{G}(\chi)}\left(\int_1^\infty S_f(y,\chi)y^{s-1}\mathrm{d}y\right. \\ &\quad \left. + \frac{\chi(-1)i^a\mathcal{G}(\chi)}{\sqrt{q}}\int_1^\infty S_f(y,\overline{\chi})y^{(1-s)-1}\mathrm{d}y\right) \\ &\quad + \frac{\mathbf{1}_{\chi=1}}{2s(s-1)},\end{aligned}$$

whence (10.1.2) follows.

10.2 Counting zeroes in the critical strip

As we showed in the previous section, the function

$$\xi(s,\chi) = \left(\frac{s(s-1)}{2}\right)^{\mathbf{1}_{\chi=1}} \left(\frac{q}{\pi}\right)^{\frac{s+a}{2}} \Gamma\left(\frac{s+a}{2}\right) L(s,\chi),$$

is entire and satisfies the functional equation (10.1.2). The Euler product representation of $L(s,\chi)$ implies that it does not vanish for $\mathrm{Re}(s) > 1$. Consequently, $\xi(s,\chi)$ does not vanish for $\mathrm{Re}(s) > 1$, so it cannot vanish for $\mathrm{Re}(s) < 0$ either. Therefore, the only zeroes of $L(s,\chi)$ when $\mathrm{Re}(s) < 0$ are at the points $-2, -4, -6, \dots$ when $a = 0$ (i.e. when χ is even) and at the points $-1, -3, -5, \dots$ when $a = 1$ (i.e. when χ is odd). These zeroes are all simple. Moreover, if $a = 0$ and $\chi \neq 1$, then we also have that $L(0,\chi) = 0$. Since $L(1,\chi) \neq 0$, as we saw in Theorem ...., this zero is also simple.

The zeroes of $L(s,\chi)$ described above are called the *trivial zeroes* of $L(s,\chi)$. Any other zero of $L(s,\chi)$, which necessarily lies within the critical strip $\{s \in \mathbb{C} : 0 \leq \mathrm{Re}(s) \leq 1\}$, is called a *non-trivial zero* of $L(s,\chi)$; it is usually denoted by $\rho = \beta + i\gamma$, where $0 \leq \beta \leq 1$, or by $\rho_\chi = \beta_\chi + i\gamma_\chi$, when we want to underline the dependence on the particular Dirichlet character χ. The non-trivial zeroes of $L(s,\chi)$ are in one-to-one correspondence with the zeroes of $\xi(s,\chi)$.

Note that the functional equation and the obvious symmetry $\overline{L(s,\chi)} = L(\overline{s},\overline{\chi})$ imply that if $\rho = \beta + i\gamma$ is a non-trivial zero of $L(s,\chi)$, then so is $1 - \overline{\rho} = 1 - \beta + i\gamma$, whereas $\overline{\rho} = \beta - i\gamma$ and $1 - \rho = 1 - \beta - i\gamma$ are zeroes of $L(s,\overline{\chi})$. The famous *Riemann Hypothesis* states that the non-trivial zeroes of $\zeta(s)$ are on the symmetry line $\mathrm{Re}(s) = 1/2$. The *Generalized Riemann Hypothesis* states that this is true for the non-trivial zeroes of all Dirichlet L-functions $L(s,\chi)$ of primitive characters χ.

The density of zeroes of an entire function is controlled by its rate of growth (cf. Appendix B). For $\xi(s,\chi)$, we have the bound

$$|\xi(s,\chi)| \ll q \cdot e^{O(|s|\log(|s|+2))} \quad (s \in \mathbb{C}). \tag{10.2.1}$$

By the functional equation (10.1.2), it suffices to show this when $\mathrm{Re}(s) \geq 1/2$. Stirling's formula (cf. Theorem A.0.3) implies that

$$|\Gamma(s)| \ll e^{|s|\log|s| + O(|s|)} q^{1/2} \log q \quad (\mathrm{Re}(s) \geq 1/4). \tag{10.2.2}$$

Moreover, we have that

$$L(s,\chi) \ll \frac{\mathbf{1}_{\chi=1}}{|s-1|} + (1 + (\sqrt{q}(\log q)(|t|+2))^{1-\sigma}) \log(q(|t|+2)), \tag{10.2.3}$$

a consequence of Theorem 4.2.2 and of the Pólya-Vinogradov inequality. Putting together the above estimates, we see that (10.2.1) is indeed true. In particular, we see that $\xi(s,\chi)$ is

an entire function of order 1 (cf. Appendix B). Therefore, there are constants $A_\chi, B_\chi \in \mathbb{C}$ and an integer $m_\chi \geq 0$ such that

$$\xi(s,\chi) = e^{A_\chi + B_\chi s} \prod_\rho \left(1 - \frac{s}{\rho}\right) e^{s/\rho}, \tag{10.2.4}$$

where the product runs over the non-trivial zeroes of $L(s,\chi)$ **listed with multiplicity**. Moreover, the product converges uniformly in compacts. This is called the *Hadamard product* of $\xi(s,\chi)$.

From (10.2.4), we can already see that $L(s,\chi)$ has an infinity of non-trivial zeroes. In fact, we have the stronger relation $\sum_\rho 1/|\rho| = \infty$. For if $\sum_\rho 1/|\rho| < \infty$, then we would deduce the estimate $\xi(s,\chi) \ll e^{O(|s|)}$. This is impossible, since as $s \to +\infty$ over real numbers, we have that $L(s,\chi) \to 1$ and that $\log\Gamma(s) \sim s\log s$ by Stirling's formula, which implies the asymptotic formula $\xi(s,\chi) = e^{(1+o(1))s\log s}$, which is a contradiction. This proves our assertion that $\sum_\rho 1/|\rho| = \infty$. On the other hand, the theory of entire functions of finite order implies that $\sum_\rho 1/|\rho|^{1+\epsilon} < \infty$, for any fixed $\epsilon > 0$ (cf. Appendix B). As a matter of fact, a very precise estimate can be proven about the number of zeroes of up to a given height. To this end, we define the quantity

$$N(T,\chi) := \#\{\rho \in \mathbb{C} : 0 \leq \mathrm{Re}(\rho) \leq 1,\ |\mathrm{Im}(\rho)| \leq T,\ L(\rho,\chi) = 0\},$$

where each zero ρ is understood to be counted with multiplicity and to be non-trivial (i.e. a zero of $\xi(s,\chi)$). When $\chi = 1$, it is easy to see that

$$N(T,1) = 2 \cdot \#\{\rho \in \mathbb{C} : 0 \leq \mathrm{Re}(\rho) \leq 1,\ 0 < \mathrm{Im}(\rho) \leq T,\ \zeta(\rho) = 0\}.$$

Indeed, if ρ is a zero of ζ, then so is $\overline{\rho}$. Moreover, Theorem 4.1.5 implies that $\zeta(s) < 0$ for $s \in (0,1)$, and we also know that $\zeta(0), \zeta(1) \neq 0$, that is to say there are no real zeroes in the critical strip, whence the above relation follows. We will show the following rather precise estimate for $N(T,\chi)$:

Theorem 10.2.1. *For $T \geq 0$, we have that*

$$\frac{N(T,\chi)}{2} = \frac{T}{2\pi}\log\frac{qT}{2\pi e} + O(\log(q(T+2))).$$

We need two preliminary estimates, the first one of which is a weak form of Theorem 10.2.1:

Lemma 10.2.2. *For $T \geq 0$, we have that*

$$N(T+1,\chi) - N(T,\chi) \ll \log(q(T+2)).$$

Proof. Recall that if $\rho = \beta + i\gamma$ is a zero of $L(s,\chi)$, then so is $1-\beta+i\gamma$. Moreover, $\beta - i\gamma$ and $1-\beta-i\gamma$ are zeroes of $L(s,\overline{\chi})$. Thus, it suffices to show that the number of zeroes ρ of $L(s,\chi)$ with $T \leq \mathrm{Im}(\rho) \leq T+1$ and $\mathrm{Re}(\rho) \geq 1/2$ is $\ll \log(q(T+2))$. We apply Jensen's formula to $L(s,\chi)$ and the disk $D(z_0,R) := \{z \in \mathbb{C} : |z - z_0| \leq R\}$, where $z_0 = 2 + i(T+1/2)$

and $R \sim 1.6 > \sqrt{5/2}$ is chosen so that there are no zeroes on the boundary of $D(z_0, R)$: we have that

$$\begin{aligned}\#\{\rho : 1/2 \le \mathrm{Re}(\rho) \le 1,\ T \le \mathrm{Im}(\rho) \le T+1\} &\le \sum_{\rho \in D(z_0, \sqrt{5/2})} 1 \\ &\ll \sum_{\rho \in D(z_0, R)} \log \frac{R}{|\rho - z_0|} \\ &= \frac{1}{2\pi} \int_0^{2\pi} \log \left| \frac{L(z_0 + Re^{i\theta}, \chi)}{L(z_0, \chi)} \right| \mathrm{d}\theta.\end{aligned}$$

Inserting the bound $|L(2+it, \chi)| \asymp 1$ in the denominator (which follows by the Euler product representation) and the bound (10.2.3) to the numerator, we find that the right hand side of the above relation is $\ll \log(q(T+2))$, thus completing the proof of the lemma. □

Next, we show that $(L'/L)(s, \chi)$ can be well-approximated by a short sum of terms of the form $1/(s-\rho)$.

Lemma 10.2.3. *For $s = \sigma + it$ with $-1 \le \sigma \le 2$, we have*

$$\frac{L'}{L}(s, \chi) = -\frac{\mathbf{1}_{\chi=1}}{s-1} + \frac{\mathbf{1}_{a=0,\, \chi \neq 1}}{s} + \sum_{\rho:\ |\gamma - t| \le 1} \frac{1}{s-\rho} + O(\log(q(|t|+2))).$$

In particular,

$$\frac{L'}{L}(s, \chi) \ll \frac{\log(q(|t|+2))}{\min\{|s-\rho| : \rho = \beta + i\gamma \text{ with } |\gamma - t| \le 1\}} + \log(|t|+2) + \frac{\mathbf{1}_{\chi=1}}{|s-1|} + \frac{\mathbf{1}_{a=0,\, \chi \neq 1}}{|s|}.$$

Proof. We note that $\mathrm{Re}(s+4) \ge 3$. We apply the Hadamard product at $s+3$ and at s and take logarithmic derivatives to find that

$$\frac{\xi'}{\xi}(s, \chi) - \frac{\xi'}{\xi}(s+4, \chi) = \sum_\rho \left(\frac{1}{s-\rho} - \frac{1}{s+4-\rho} \right) = \sum_\rho \frac{4}{(s+\rho)(s+4-\rho)}.$$

On the other hand, since $\Gamma((s+4+a)/2) = \frac{1}{4}(s+2+a)(s+a)\Gamma((s+a)/2)$, we have that

$$\frac{\xi(s, \chi)}{\xi(s+4, \chi)} = \frac{4\pi^2 q^{-2}}{(s+2+a)(s+a)} \left(\frac{s(s-1)}{(s+4)(s+3)} \right)^{\mathbf{1}_{\chi=1}} \frac{L(s, \chi)}{L(s+4, \chi)},$$

so that

$$\begin{aligned}\frac{\xi'}{\xi}(s, \chi) - \frac{\xi'}{\xi}(s+4, \chi) = \frac{L'}{L}(s, \chi) - \frac{L'}{L}(s+4, \chi) + \mathbf{1}_{\chi=1} \cdot \left(\frac{1}{s-1} + \frac{1}{s} - \frac{1}{s+3} - \frac{1}{s+4} \right) \\ - \frac{1}{s+2+a} - \frac{1}{s+a}.\end{aligned}$$

Since we also have that $(L'/L)(s+4) \ll 1$ and $1/(s+j) \ll 1$ for $\mathrm{Re}(s) \geq -1$ and $j \geq 2$, we deduce that

$$\frac{L'}{L}(s,\chi) = -\frac{\mathbf{1}_{\chi=1}}{s-1} + \frac{\mathbf{1}_{a=0,\,\chi\neq 1}}{s} + \sum_{\rho} \frac{4}{(s-\rho)(s+4-\rho)} + O(1).$$

We note that

$$\sum_{\rho:\ |\gamma-t|>1} \frac{4}{(s-\rho)(s+4-\rho)} \ll \sum_{\rho:\ |\gamma-t|>1} \frac{1}{(\gamma-t)^2} \ll \log(q(|t|+2)),$$

by Lemma (10.2.2). Moreover,

$$\begin{aligned}\sum_{\rho:\ |\gamma-t|\leq 1} \frac{4}{(s-\rho)(s+4-\rho)} &= \sum_{\rho:\ |\gamma-t|\leq 1} \frac{1}{s-\rho} - \sum_{\rho:\ |\gamma-t|\leq 1} \frac{1}{s+4-\rho} \\ &= \sum_{\rho:\ |\gamma-t|\leq 1} \frac{1}{s-\rho} + O(\log(q(|t|+2)))\end{aligned}$$

applying Lemma (10.2.2) again. Putting together the above estimates complete the proof of the first part of the lemma. The second part follows immediately by the first part and by the fact that there are $\ll \log(q(|t|+2))$ zeroes $\rho = \beta + i\gamma$ with $\gamma \in [t-1, t+1]$, a consequence of Lemma 10.2.2. □

Proof of Theorem 10.2.1. It suffices to show the claimed formula for $N(T,\chi)$ when T does not coincide with the ordinate of a zero of $\xi(s,\chi)$, since varying T by $O(1)$ alters both sides in the statement of Theorem 10.2.1 by $O(\log(q(T+2)))$, a consequence of Lemma 10.2.2. Let R be the rectangle with vertices at $5/2 - iT, 5/2 + iT$, $-3/2 + iT$ and $-3/2 - iT$, traversed in the counterclockwise direction. Then $\xi(s,\chi)$ does not vanish on the boundary of R: it does not vanish on its left right parts, and it does not vanish on its top part by the choice of T.

Since $\xi(s,\chi)$ does not vanish on R and $N(T,\chi)$ counts non-trivial zeroes of $L(s,\chi)$, that is to say zeroes of $\xi(s,\chi)$, the argument principle implies

$$N(T,\chi) = \frac{1}{2\pi i}\int_R \frac{\xi'}{\xi}(s,\chi)\mathrm{d}s,$$

where we added the term $\mathbf{1}_{\chi\neq 1,\,a=0}$ to account for the trivial zero of $L(s,\chi)$ at 0 when $a=0$ that is not a zero of $\xi(s,\chi)$. Then

$$N(T,\chi) = \frac{N(T,\chi) + \overline{N(T,\chi)}}{2} = \frac{1}{2\pi}\int_R \mathrm{Im}\left(\frac{\xi'}{\xi}(s,\chi)\right)\mathrm{d}s.$$

We want to write the above formula in terms of the variation of the argument of $\xi(s,\chi)$. Since ξ does not vanish on R, a compactness argument implies that there is some $\epsilon \in (0, 1/10)$ such that ξ does not vanish in the open set $U := \{s \in \mathbb{C} : \mathrm{dist}(s,R) < \epsilon\}$. However, U is not simply connected and we cannot define $\log\xi$ and $\arg\xi$ there. Instead, we consider $V = U \setminus L$, where L is a line segment from the inner to the outer boundary of U that contains the point

$5/2$, so that V is an open and simply connected set containing $R \setminus \{5/2\}$. Then $\log \xi(s,\chi)$ and $\arg \xi(s,\chi) = \mathrm{Im}(\log \xi(s,\chi))$ are well-defined for $s \in V$ (we consider the principal branch of the logarithm, with $\log z \in \mathbb{R}$ for $z > 0$), and we find that

$$N(T,\chi) = \frac{\Delta_R \arg \xi(s,\chi)}{2\pi},$$

where $\Delta_R \arg \xi(s,\chi)$ denotes the total variation of the argument of $\xi(s,\chi)$ as s traverses R. The change in argument in the left half of R is the same as the change of argument in the right half, a consequence of the formulas

$$\overline{\xi(1-\sigma+it,\chi)} = \xi(1-\sigma-it,\overline{\chi}) = \frac{i^a\sqrt{q}}{\mathcal{G}(\chi)}\xi(\sigma+it,\chi).$$

If C denotes the right half of R, then

$$\begin{aligned}\frac{N(T,\chi)}{2} &= \frac{\Delta_C \arg \xi(s,\chi)}{2\pi} \\ &= \frac{\mathbf{1}_{\chi=1}\cdot \Delta_C \arg(s(s-1)) + \Delta_C((q/\pi)^{\frac{s+a}{2}}) + \Delta_C \arg \Gamma(\frac{s+a}{2}) + \Delta_C \arg L(s,\chi)}{2\pi},\end{aligned}$$

by the definition of ξ. We have that $\Delta_C \arg(s(s-1)) = O(1)$ and $\Delta_C \arg((q/\pi)^{\frac{s+a}{2}}) = T\log(q/\pi)$. Moreover, Stirling's formula implies that

$$\begin{aligned}\Delta_C \arg \Gamma\left(\frac{s+a}{2}\right) &= \mathrm{Im}\left(\log\Gamma\left(\frac{1+2a}{4}+\frac{iT}{2}\right)\right) - \mathrm{Im}\left(\log\Gamma\left(\frac{1+2a}{4}-\frac{iT}{2}\right)\right) \\ &= 2\cdot \mathrm{Im}\left(\left(\frac{1+2a}{4}+\frac{iT}{2}\right)\log\left(\frac{1+2a}{4}+\frac{iT}{2}\right) - \left(\frac{1+2a}{4}+\frac{iT}{2}\right)\right. \\ &\qquad \left. + \frac{1}{2}\log\frac{2\pi}{1/4+a/2+iT/2}\right) + O\left(\frac{1}{T}\right) \\ &= T\log\left|\frac{1+2a}{4}+\frac{iT}{2}\right| - \frac{T}{2} + O(1) \\ &= T\log\frac{T}{2} - T + O(1).\end{aligned}$$

Putting together the above estimates, we deduce that

$$\frac{N(T,\chi)}{2} = \frac{T}{2\pi}\log\frac{qT}{2\pi e} + O(1) + \frac{\Delta_C \arg L(s,\chi)}{\pi}.$$

Next, note that

$$\Delta_{[5/2-iT,5/2+iT]} \arg \zeta(s) = \mathrm{Im}(\log(L(5/2+iT,\chi)) - \mathrm{Im}(\log(L(5/2-iT,\chi)) \ll 1,$$

since

$$\log L(s,\chi) = -\sum_p \log\left(1-\frac{\chi(p)}{p^s}\right) = \sum_p\sum_{m=1}^{\infty}\frac{\chi(p)^m}{mp^{ms}} \quad (\mathrm{Re}(s)>1).$$

Finally, we have that

$$\begin{aligned}\Delta_{[5/2+iT,1/2+iT]} \arg L(s,\chi) &= \mathrm{Im}(\log(L(1/2+iT,\chi)) - \log(\zeta(5/2+iT,\chi)) \\ &= -\mathrm{Im}\int_{1/2}^{5/2} \frac{L'}{L}(\sigma+iT,\chi)\mathrm{d}\sigma.\end{aligned}$$

Using Lemmas 10.2.3 and 10.2.2, we find that

$$\begin{aligned}\Delta_{[5/2+iT,1/2+iT]} \arg L(s,\chi) &= -\sum_{\rho:\ |\gamma-T|\le 1} \mathrm{Im}\int_{1/2}^{5/2} \frac{1}{s-\rho}\mathrm{d}\sigma + O(\log(q(T+2))) \\ &= \sum_{\rho:\ |\gamma-T|\le 1} \Delta_{[5/2+iT,1/2+iT]}(s-\rho) + O(\log T) \\ &\ll N(T+1,\chi) - N(T-1,\chi)) + \log(q(T+2)) \\ &\ll \log(q(T+2)).\end{aligned}$$

Similarly, we have that $\Delta_{[1/2-iT5/2-iT]} \arg L(s,\chi) \ll \log(q(T+2))$. This completes the proof of Theorem 10.2.1. □

10.3 The explicit formula

In this section, we connect the distribution of primes in progressions to the distribution of zeroes of Dirichlet L-functions:

Theorem 10.3.1. *For $x, T \ge 2$ and a primitive character $\chi \pmod q$, we have that*

$$\begin{aligned}\sum_{n\le x}\Lambda(n)\chi(n) &= \mathbf{1}_{\chi=1}\cdot x - \sum_{\rho:\ |\mathrm{Im}(\rho)|\le T} \frac{x^\rho-1}{\rho} + O\left(\frac{x\log^2(xqT)}{T} + \log x + \frac{\log^2(qT)}{\sqrt{x}}\right) \\ &= \mathbf{1}_{\chi=1}\cdot x - \sum_{\substack{\rho:\ |\mathrm{Im}(\rho)|\le T \\ \mathrm{Re}(\rho)\ge 1/\log x}} \frac{x^\rho}{\rho} + O\left(\frac{x\log^2(xqT)}{T} + \log^2(qT) + (\log q)(\log x)\right).\end{aligned}$$

Proof. By varying x infinitesimally, we may assume that $x^\rho \ne 1$ for all zeroes ρ with $|\mathrm{Im}(\rho)| \le T+1$. Since $\Lambda(n) \ll \log n$, Lemma 4.3.1 implies that

$$\sum_{n\le x}\Lambda(n)\chi(n) = \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=1+1/\log x \\ |\mathrm{Im}(s)|\le T'}} \left(-\frac{L'}{L}(s,\chi)\right)\frac{x^s}{s}\mathrm{d}s + O\left(\frac{x\log^2 x}{T'} + \log x\right),$$

for all $2 \le T' \le x$, where $T' \in [T-1, T+1]$ does not coincide with the ordinate of a zero of ζ. In order to remove the singularity of x^s/s at $s=0$, we will replace x^s by $x^s - 1$. We note that

$$\frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=1+1/\log x \\ |\mathrm{Im}(s)|\le T'}} \left(-\frac{L'}{L}(s,\chi)\right)\frac{\mathrm{d}s}{s} = \sum_{n=1}^{\infty}\Lambda(n)\chi(n)\cdot\frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=1+1/\log x \\ |\mathrm{Im}(s)|\le T'}} \frac{n^{-s}}{s}\mathrm{d}s.$$

The term with $n = 1$ does not contribute since $\Lambda(1) = 1$, and if $n > 1$, then Lemma 4.3.1 implies that the integral is $\ll n^{-1-1/\log x}/(T' \log n)$. Thus

$$\sum_{n\le x}\Lambda(n)\chi(n) = \frac{1}{2\pi i}\int\limits_{\substack{\mathrm{Re}(s)=1+1/\log x\\ |\mathrm{Im}(s)|\le T'}}\left(-\frac{L'}{L}(s,\chi)\right)\frac{x^s-1}{s}\mathrm{d}s + O\left(\frac{x\log^2 x}{T'}+\log x\right).$$

The integrant is a meromorphic function. We are going to replace the contour $[1+1/\log x - iT', 1+1/\log x + iT']$ by the contour consisting of the line segments $L_1 = [1+1/\log x - iT', 1/\log x - iT']$, $L_2 = [-1/2 - iT', -1/2 + iT']$ and $L_3 = [-1/2 + iT', 1+1/\log x + iT']$ with this orientation. The difference of the two integrals is given by the residues of the integrant $f(s) = -(L'/L)(s,\chi)(x^s-1)/s$ inside the rectangle with vertices $1 + 1/\log x \pm iT'$, and $-1/2 \pm iT'$. The factor $(x^s-1)/s$ is analytic. If $\chi = 1$, then the pole of ζ at $s = 1$ induces a pole of f at $s = 1$ of residue $x - 1$. Moreover, for each zero ρ of $L(s,\chi)$ of multiplicity m_ρ, we obtain a pole of residue $-m_\rho(x^\rho - 1)/\rho$, unless $(x^s-1)/s$ vanishes at $s = \rho$. But we have assume that this does not happen for our choice of x. Therefore

$$\begin{aligned}\frac{1}{2\pi i}&\int\limits_{\substack{\mathrm{Re}(s)=1+1/\log x\\ |\mathrm{Im}(s)|\le T'}}\left(-\frac{L'}{L}(s,\chi)\right)\frac{x^s-1}{s}\mathrm{d}s\\ &= \mathbf{1}_{\chi=1}\cdot(x-1) - \sum_{\rho:\ |\mathrm{Im}(\rho)|\le T'}\frac{x^\rho-1}{\rho} - \mathbf{1}_{a=0,\,\chi\neq 1}\cdot\log x\\ &\quad + \frac{1}{2\pi i}\left(\int_{L_1}+\int_{L_2}+\int_{L_3}\right)\left(-\frac{L'}{L}(s,\chi)\right)\frac{x^s-1}{s}\mathrm{d}s,\end{aligned}$$

with the term $\mathbf{1}_{a=1,\,\chi\neq 1}\cdot\log x$ coming from the trivial zero at $s = 0$ when $a = 0$ and $\chi \neq 1$. We choose T' as follows: given $T \in [2, x]$, we take $T' \in [T-1, T+1] \cap [2, x]$ such that $|T' - \gamma| \gg 1/\log(qT)$ for all zeroes $\rho = \beta + i\gamma$. Such a T' exists by Lemma 10.2.2. By the choice of T' and Lemma 10.2.3, we find that $(L'/L)(\sigma + iT') \ll \log^2(qT)$ for all $\sigma \in [-1, 2]$. So the integrals over L_1 and L_3 contribute $\ll x(\log(qT))^2/T$. Finally, applying Lemma 10.2.3 again, we find that $(L'/L)(-1/2 + it) \ll \log(q(|t|+2))$, since $|\rho + 1 - it| \ge 1/2$ for all non-trivial zeroes ρ. Therefore, the contribution of the integral over L_2 is

$$\ll \int_{-T}^{T}\frac{\log(q(|t|+2))}{\sqrt{x}}\cdot\frac{\mathrm{d}t}{|t|+1} \ll \frac{\log q + \log^2 T}{\sqrt{x}}.$$

Putting together the above estimates, yields that

$$\sum_{n\le x}\Lambda(n)\chi(n) = \mathbf{1}_{\chi=1}\cdot(x-1) - \sum_{\rho:\ |\mathrm{Im}(\rho)|\le T'}\frac{x^\rho-1}{\rho} + O\left(\frac{x\log^2(xqT)}{T}+\log x\right).$$

We note that we may replace T' by T introducing an error term of size at most

$$\sum_{\rho:\ T-1\le|\mathrm{Im}(\rho)|\le T+1}\left|\frac{x^\rho-1}{\rho}\right| \ll \frac{x}{T}\cdot(N(T+1,\chi) - N(T-1,\chi)) \ll \frac{x\log(qT)}{T}$$

by Lemma 10.2.2, so that

$$\sum_{n\le x}\Lambda(n)\chi(n) = \mathbf{1}_{\chi=1}\cdot(x-1) - \sum_{\rho:\ |\mathrm{Im}(\rho)|\le T}\frac{x^\rho-1}{\rho} + O\left(\frac{x\log^2(xqT)}{T}+\log x\right).$$

This proves the first claimed formula. Finally, note that $(x^\rho-1)/\rho \ll x^{\mathrm{Re}(\rho)}\log x$, so that

$$\sum_{\substack{\rho:\ |\mathrm{Im}(\rho)|\le 1\\ \mathrm{Re}(\rho)\le 1/\log x}}\frac{x^\rho-1}{\rho} \ll (\log q)(\log x).$$

Moreover,

$$\left|\sum_{\substack{\rho:\ 1<|\mathrm{Im}(\rho)|\le T\\ \mathrm{Re}(\rho)\le 1/\log x}}\frac{x^\rho-1}{\rho}\right| + \left|\sum_{\substack{\rho:\ 1<|\mathrm{Im}(\rho)|\le T\\ \mathrm{Re}(\rho)> 1/\log x}}\frac{1}{\rho}\right| \ll \sum_{\rho:\ 1<|\mathrm{Im}(\rho)|\le T}\frac{1}{|\rho|} \ll \log^2(qT).$$

Putting together the above estimates completes the proof of the theorem. □

Exercise 10.3.1. Assuming that the Generalized Riemann Hypothesis holds, prove that

$$\psi(x;q,a) = \frac{x}{\varphi(q)} + O(\sqrt{x}\log^2 x) \quad (1\le q\le x,\ (a,q)=1).$$

In particular, deduce that $\psi(x;q,a)\sim x/\varphi(q)$ as long as $q\le x^{1/2}/(\log x)^{2+\epsilon}$ and $x\to\infty$.

Exercise 10.3.2. Prove that the Riemann Hypothesis for the zeta function is equivalent to the following estimates:

(a) For fixed $\epsilon>0$ and $x\ge 1$,
$$\psi(x) = x + O_\epsilon(x^{1/2+\epsilon})$$

(b) For all $x\ge 2$,
$$\psi(x) = x + O(\sqrt{x}\log^2 x).$$

(c) For fixed $\epsilon>0$ and $x\ge 1$,
$$M(x) = \sum_{n\le x}\mu(n) \ll_\epsilon x^{1/2+\epsilon}.$$

Exercise 10.3.3. Prove that there is some (small) constant $c>0$ such that whenever $\chi_1 \pmod{q_1}$ and $\chi_2 \pmod{q_2}$ are two real and non-principal Dirichlet characters whose associated L-functions have real zeroes at β_1 and at β_2, respectively, then

$$\min\{\beta_1,\beta_2\} \le 1 - \frac{c}{\log(q_1q_2)}.$$

Deduce that if $1-\beta_j \le c^2/\log q_j$ for both $j\in\{1,2\}$, then $\max\{q_1,q_2\}\ge \min\{q_1,q_2\}^{1/(2c)}$.

Appendix A

The Gamma function

The Gamma function is an extension of the factorial function. It is defined for $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 0$ via the formula

$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1} \mathrm{d}x.$$

Integrating by parts, we see that

$$\begin{aligned} \Gamma(s+1) &= \int_0^\infty (-e^{-x})' x^s \mathrm{d}x \\ &= -e^{-x} x^s \Big|_{x=0}^\infty + \int_0^\infty e^{-x} (x^s)' \mathrm{d}x \\ &= s\Gamma(s) \quad (\mathrm{Re}(s) > 0). \end{aligned} \tag{A.0.1}$$

In particular, iterating this formula, we see that $\Gamma(n+1) = n!$ for all $n \in \mathbb{Z}_{\geq 0}$. Moreover, the above formula can be use to meromorphically continue Γ to the entire complex plane: Indeed, applying it $n+1$ times, we deduce that

$$\Gamma(s) = \frac{\Gamma(s+n+1)}{s(s+1)\cdots(s+n)}, \tag{A.0.2}$$

which can be taken as the definition of Γ for $\mathrm{Re}(s) > -n-1$. It is clear from this formula that the only singularities of Γ are at $0, -1, -2, \ldots$ with $\mathrm{res}_{s=-n}\Gamma(s) = (-1)^n/n!$.

Lemma A.0.1. *For every $s \in \mathbb{C}$, we have that*

$$\lim_{n\to\infty} \frac{\Gamma(s+n)}{n^s \Gamma(n)} = 1.$$

Proof. Fix $\epsilon \in (0, 1/100]$ and write $s = \sigma + it$. If $f(x) = (\sigma + n - 1)\log x - x$, then $f'(x) = (\sigma+n-1)/x - 1 < -\epsilon/2$ for n large enough. Therefore

$$\int_{n+\epsilon n}^\infty e^{-x} x^{s+n-1} \mathrm{d}x \ll \int_{n+\epsilon n}^\infty e^{f(x)} \mathrm{d}x \ll e^{f(n+\epsilon n)}.$$

Since $f(n+\epsilon n) = n\log n - n - (\epsilon - \log(1+\epsilon))n + O_{\epsilon,\sigma}(\log n)$, an application of Stirling's formula (cf. Theorem 1.2.4) implies that

$$\lim_{n\to\infty} \frac{\Gamma(s+n)}{n^s\Gamma(n)} = \lim_{n\to\infty} \frac{\int_0^{n+\epsilon n} e^{-x}x^{s+n-1}\mathrm{d}x}{n^s\Gamma(n)}.$$

Similarly, we may show that the contribution of $x \le n - \epsilon n$ to the integral is $o_{n\to\infty}(n^s\Gamma(n))$. Therefore

$$\lim_{n\to\infty} \frac{\Gamma(s+n)}{n^s\Gamma(n)} = \lim_{n\to\infty} \frac{\int_{n-\epsilon n}^{n+\epsilon n} e^{-x}x^{s+n-1}\mathrm{d}x}{n^s\Gamma(n)}. \tag{A.0.3}$$

Note that $x^s = (1+O_s(\epsilon))n^s$ for $x \in [n-\epsilon n, n+\epsilon n]$. Therefore

$$\lim_{n\to\infty} \frac{\Gamma(s+n)}{n^s\Gamma(n)} = (1+O_s(\epsilon)) \lim_{n\to\infty} \frac{\int_{n-\epsilon n}^{n+\epsilon n} e^{-x}x^{n-1}\mathrm{d}x}{\Gamma(n)} = 1 + O_s(\epsilon),$$

where we applied again relation (A.0.3) with $s = 0$. Letting $\epsilon \to 0^+$ completes the proof of the lemma.[1] □

Theorem A.0.2. *For all $s \in \mathbb{C}$, we have that*

$$\Gamma(s) = \lim_{n\to\infty} \frac{n!n^s}{s(s+1)\cdots(s+n)} = \frac{1}{s}\prod_{n=1}^{\infty} \frac{(1+1/n)^s}{1+s/n} = \frac{e^{-\gamma s}}{s}\prod_{n=1}^{\infty} \frac{e^{s/n}}{1+s/n}.$$

Proof. The first equality follows immediately by (A.0.2) and (A.0.1), and the second equality follows immediately from the first one. Finally, for the third equality, note that

$$\begin{aligned}\log \prod_{n=1}^{N} \frac{1+1/n}{e^{1/n}} &= \sum_{n=1}^{N} \log\left(1+\frac{1}{n}\right) - \sum_{n=1}^{N}\frac{1}{n} \\ &= \log(N+1) - \log N - \gamma + O\left(\frac{1}{N}\right) \to -\gamma\end{aligned}$$

as $N \to \infty$, by Theorem 1.2.3. This completes the proof. □

Since Γ is a generalization of the factorial function, one might suspect that Stirling's formula can also be generalized. This is indeed the case:

Theorem A.0.3 (Stirling's formula)**.** *Fix $\delta > 0$. Uniformly for $s \in \mathbb{C}$ with $|s| \ge 1$ and $|\arg(s)| \le \pi - \delta$, we have that*

$$\Gamma(s) = \left(\frac{s}{e}\right)^s \sqrt{\frac{2\pi}{s}}\left(1 + O\left(\frac{1}{|s|}\right)\right).$$

[1] Technically, we should have worked with $\limsup_{n\to\infty}$ and $\liminf_{n\to\infty}$, since we don't know *a priori* the existence of the limit in question. This is a simple technicality that is easy to get around. We leave the details to the reader.

Proof. By Theorems A.0.2 and 1.2.4, we have that

$$\log\Gamma(s) = \lim_{n\to\infty}\left((n+s)\log n - n + \frac{\log(2\pi n)}{2} - \log s - \sum_{j=1}^{n}\log(s+j)\right).$$

The Euler-McLaurin formula then yields that

$$\begin{aligned}\sum_{j=1}^{n}\log(s+j) &= \int_0^n \log(s+x)\mathrm{d}x + \int_0^x \frac{\{x\}}{s+x}\mathrm{d}x \\ &= (s+n)\log(s+n) - n + \frac{\log(s+n)}{2} + \int_0^n \frac{\{x\}-1/2}{s+x}\mathrm{d}x.\end{aligned}$$

If we set $F(x) = \int_0^x(\{x\}-1/2)\mathrm{d}x \ll 1$, then integration by parts implies that

$$\begin{aligned}\sum_{j=1}^{n}\log(s+j) &= \int_0^n \log(s+x)\mathrm{d}x + \int_0^x \frac{\{x\}}{s+x}\mathrm{d}x \\ &= (s+n)\log(s+n) - n + \frac{\log(s+n)}{2} + \int_0^n \frac{F(x)}{(s+x)^2}\mathrm{d}x.\end{aligned}$$

Putting together the above formulas, we find that

$$\log\Gamma(s) = s\log s - s + \frac{\log(2\pi/s)}{2} + \int_0^\infty \frac{F(x)}{(s+x)^2}\mathrm{d}x.$$

It remains to show that the integral on the right hand side of the above equality is $\ll 1/|s|$. This can be easily seen to be the case. If $s = \sigma + it$ with $\sigma \geq 0$, then

$$\int_0^\infty \frac{F(x)}{(s+x)^2}\mathrm{d}x \ll \int_0^\infty \frac{1}{(x+\sigma)^2+t^2}\mathrm{d}x \ll \frac{1}{\sigma+|t|},$$

by a case analysis. Finally, if $s = -\sigma + it$ with $\sigma \geq 0$, then we note that $|t| \gg \sigma/\delta$, by our assumption that $|\arg(s)| \leq \pi - \delta$. So

$$\begin{aligned}\int_0^\infty \frac{F(x)}{(s+x)^2}\mathrm{d}x &\ll \int_0^\infty \frac{\mathrm{d}x}{(x-\sigma)^2+t^2} \\ &\ll \int_{x\geq 0,\ |x-\sigma|\geq\sigma/2} \frac{\mathrm{d}x}{x^2+t^2} + \int_{|x-\sigma|\leq\sigma/2}\frac{\mathrm{d}x}{t^2} \\ &\ll \frac{t+\sigma}{t^2} \ll_\delta \frac{1}{t} \ll_\delta \frac{1}{|s|},\end{aligned}$$

which completes the proof. □

Corollary A.0.4. *For $s = \sigma + it$ with $|s| \geq 2$ and $-1 \leq \sigma \leq 2$, we have that*

$$|\Gamma(s)| \asymp e^{-\pi|t|/2}(|t|+1)^{\sigma-1/2}.$$

Theorem A.0.5. *For all $s \in \mathbb{C}$, we have that*

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

Proof. We note that the function $f(s) = \Gamma(s)\Gamma(1-s)\sin(\pi s)$ is 1-periodic and entire. Moreover, if $0 \leq \mathrm{Re}(s) \leq 1$, then Corollary A.0.4 implies that $|f(s)| \asymp 1$, that is to say f is a bounded entire function. Liouville's theorem then yields that f is constant. Since $f(0) = \pi$, the theorem follows. □

Theorem A.0.6. *For all $s \in \mathbb{C}$, we have that*

$$\Gamma(2s) = \frac{2^{2s-1}}{\sqrt{\pi}}\Gamma(s)\Gamma\left(s + \frac{1}{2}\right).$$

Proof. We note that the function $f(s) = \Gamma(2s)2^{-2s}/(\Gamma(s)\Gamma(s+1/2))$ is 1-periodic and entire. (The function Γ does not vanish anywhere, by its product representation in Theorem A.0.2.) Moreover, if $0 \leq \mathrm{Re}(s) \leq 1$, then Corollary A.0.4 implies that $|f(s)| \asymp 1$, that is to say f is a bounded entire function. Liouville's theorem then yields that f is constant. Finally, we note that $f(0) = 1/(2\Gamma(1/2))$. By Theorem A.0.5, we know that $\Gamma(1/2)^2 = \pi$. Since $\Gamma(1/2) > 0$ by its integral definition, we find that $\Gamma(1/2) = \sqrt{\pi}$, and the theorem follows. □

Appendix B

Analytic functions of finite order

A non-constant analytic function $f : \mathbb{C} \to \mathbb{C}$ is called of *finite order* if there is some number $\alpha > 0$ such that

$$f(z) = O\left(e^{|z|^\alpha}\right) \quad (z \in \mathbb{C}). \tag{B.0.1}$$

For any such f, the infimum of all $\alpha > 0$ such that the above relation holds is called the *order* of f, which we denote by $\mathrm{ord}(f)$. We will show here that if $d = \lfloor \mathrm{ord}(f) \rfloor$, m is the order of vanishing of f at 0 and $z_1, z_2, \ldots$ are the zeroes of f that are different from 0, listed with multiplicity if necessary, then there are numbers $c_0, c_1, \ldots, c_d$ such that

$$f(z) = e^{c_0 + c_1 z + \cdots + c_d z^d} z^m \prod_{j=1}^{\infty} \left(1 - \frac{z}{z_j}\right) e^{L_d(z/z_j)}, \tag{B.0.2}$$

where we have set

$$L_d(z) = \sum_{j=1}^{d} \frac{z^j}{j}$$

and the product in the above relation converges uniformly in compacts.

Indeed, assume that f satisfies (B.0.1) for some $\alpha > 0$, for which we may that $\lfloor \alpha \rfloor = d$. Without loss of generality, we may assume that $f(0) = 1$; otherwise, we simply replace $f(z)$ by $f(z)/(cz^m)$ for an appropriate $c \in \mathbb{C}$, which has the same finite order as f. Let $Z_f(R)$ denote the number of zeroes of f inside the disk $D(0, R) := \{z \in \mathbb{C} : |z| < R\}$. If f does not vanish on the boundary of $D(0, R)$, then Jensen's formula implies that

$$\sum_{\substack{j \geq 1 \\ |z_j| < R}} \log \frac{R}{z_j} = \int_0^1 \log |f(Re^{2\pi i \theta})| \mathrm{d}\theta.$$

The left hand side is at least $(\log 2) \cdot Z_f(R/2)$, whereas the right hand side is at most $R^\alpha + O(1)$ by (B.0.1). Replacing R by $R/2$, we see that $Z_f(R) \ll R^\alpha$. In particular, the series $\sum_{z_j} 1/|z_j|^{\alpha+\epsilon}$ converges absolutely for any fixed $\epsilon > 0$. Since $(1 - z)e^{L_d(z)} = 1 + O_d(|z|^{d+1})$ for $|z| \leq 1/2$, the infinite product

$$P(z) := \prod_{j=1}^{\infty} \left(1 - \frac{z}{z_j}\right) e^{L_d(z/z_j)},$$

converges uniformly in compacts and thus defines an entire function. Moreover, since P has the exact same zeroes with f, their quotient f/P is an entire function. We claim, in fact, that $f/P = e^Q$, where Q is a polynomial of degree $\leq d$, which would complete the proof of (B.0.2).

The first step in showing that $\log(f/P)$ is a polynomial is showing that

$$\frac{f(z)}{P(z)} \ll (|z|+2)^{O(|z|^\alpha)} \quad (z \in \mathbb{C}). \tag{B.0.3}$$

Let $w \in \mathbb{C}$ with $|w| \geq 1$. We want to majorize $f(w)/P(w)$. Since $Z_f(r) \ll r^\alpha$ for $r \geq 1$, we know that there is $R \in [|w|, 2|w|]$ with $|R - |z_j|| \gg 1/R^\alpha$ for all j. By the maximum modulus principle, it suffices to show (B.0.3) for all $z \in \mathbb{C}$ with $|z| = R$. Since we already know that $f(z) \ll e^{|z|^\alpha}$, it is enough to show that $1/P(z) \ll R^{O(R^\alpha)}$ for $|z| = R$. For such a z, we write $P(z) = P_1(z)P_2(z)$, where

$$P_1(z) = z^m \prod_{\substack{j\geq 1 \\ |z_j|\leq 2R}} \left(1 - \frac{z}{z_j}\right) e^{L_a(z/z_j)},$$

and

$$P_1(z) = \prod_{\substack{j\geq 1 \\ |z_j|> 2R}} \left(1 - \frac{z}{z_j}\right) e^{L_a(z/z_j)},$$

We note that $L_a(z/z_j) \ll |z/z_j|^n$ for $|z_j| \leq 2R$ and thus

$$\sum_{\substack{j\geq 1 \\ |z_j|\leq 2R}} L_a(z/z_j) \ll R^n \sum_{\substack{j\geq 1 \\ |z_j|\leq 2R}} \frac{1}{|z_j|^n} \ll R^n \cdot R^{\alpha-n} \log R = R^\alpha \log R,$$

by the estimate $Z_f(r) \ll r^\alpha$ and partial summation. Moreover, since

$$\left|\frac{z}{z_j} - 1\right| = \frac{|z - z_j|}{|z_j|} \geq \frac{|R - |z_j||}{2R} \gg R^{-\alpha-1}$$

whenever $|z_j| \leq 2R$, we find that

$$|P_1(z)| \gg R^{-O(Z_f(2R))} e^{-O(R^\alpha \log R)} = e^{-O(R^\alpha \log R)}.$$

Finally, for $P_2(z)$, we note that $(1 - z/z_j)e^{L_a(z/z_j)} = \exp\{O(|z/z_j|^{n+1})\}$, so that

$$|P_2(z)| \gg \exp\left\{-c \sum_{\substack{j\geq 1 \\ |z_j|\leq 2R}} \frac{R^{n+1}}{|z_j|^{n+1}}\right\} \geq \exp\left\{-c' R^\alpha\right\},$$

for some positive constants c, c' that do not depend on R or z. Together with (B.0.1), the above estimates imply that $f(z)/P(z) \ll R^{O(R^\alpha)}$ for those R whenever $|z| = R$, which completes the proof of (B.0.3).

Now that we have established (B.0.3), it is not too hard to prove that $\log(f/P)$ is a polynomial of degree $\leq d$. Indeed, f/P is an entire function which does not vanish, so $g = \log(f/P)$ is entire. Moreover, $\mathrm{Re}(g(z)) \leq C|z|^\alpha \log(|z| + 2) + O(1)$ some absolute constant C. We want to show that g is a polynomial. Write

$$g(z) = \sum_{j=0}^{\infty} (a_j + ib_j) z^j,$$

so that

$$\mathrm{Re}(g(Re^{i\theta})) = \sum_{j=0}^{\infty} R^j a_j \cos(j\theta) - \sum_{j=1}^{\infty} R^j b_j \sin(j\theta).$$

Fourier inversion (or Cauchy's formula) then implies that

$$R^j a_j = \frac{1}{\pi} \int_0^{2\pi} \mathrm{Re}(g(Re^{i\theta})) \cos(j\theta) \mathrm{d}\theta$$

for $j \in \mathbb{N}$, so that

$$\begin{aligned} R^j |a_j| \leq \frac{1}{\pi} \int_0^{2\pi} |\mathrm{Re}(g(Re^{i\theta}))| \mathrm{d}\theta &= \frac{1}{\pi} \int_0^{2\pi} (|\mathrm{Re}(g(Re^{i\theta}))| + \mathrm{Re}(g(Re^{i\theta}))) \mathrm{d}\theta - 2a_0 \\ &\leq 4CR^\alpha \log(R+2) + O(1). \end{aligned}$$

This implies that $a_j = 0$ if $j > \alpha$. Similarly, we find that $b_j = 0$ for $j > \alpha$, which proves our claim that g is a polynomial of degree $\leq d = \lfloor \alpha \rfloor$, thus completing the proof of (B.0.2).

Bibliography

[1] T. M. Apostol, *Mathematical analysis.* Second edition. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1974.

[2] —, *Introduction to analytic number theory.* Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1976.

[3] 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.

[4] P. D. T. A. Elliott, *Multiplicative functions on arithmetic progressions. VII. Large moduli.* J. London Math. Soc. (2) 66 (2002), no. 1, 14–28.

[5] J. Friedlander and H. Iwaniec, *Opera de cribro.* American Mathematical Society Colloquium Publications, 57. American Mathematical Society, Providence, RI, 2010.

[6] D. A. Goldston, J. Pintz and C. Y. Yildirim, *Primes in tuples. I.* Ann. of Math. (2) 170 (2009), no. 2, 819–862.

[7] 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.

[8] —, *Primes in intervals of bounded length.* Bull. Amer. Math. Soc. (N.S.) 52 (2015), 171–222.

[9] A. Granville, A. Harper and K. Soundararajan, *Mean values of multiplicative functions over function fields.* Res. Number Theory 1 (2015), Art. 25, 18 pp.

[10] A. Granville and K. Soundararajan, *Pretentious multiplicative functions and an inequality for the zeta-function.* Anatomy of integers, 191–197, CRM Proc. Lecture Notes, 46, Amer. Math. Soc., Providence, RI, 2008.

[11] —, *Multiplicative number theory.* Snowbird MRC notes (unpublished), 2011.

[12] G. Halász, *On the distribution of additive and the mean values of multiplicative arithmetic functions.* Studia Sci. Math. Hungar. 6 (1971), 211–233.

[13] H. Iwaniec, *Rosser's sieve.* Acta Arith. 36 (1980), no. 2, 171–202.

[14] H. Iwaniec and E. Kowalski, *Analytic number theory.* American Mathematical Society Colloquium Publications, 53, American Mathematical Society, Providence, RI, 2004.

[15] D. Koukoulopoulos, *Pretentious multiplicative functions and the prime number theorem for arithmetic progressions.* Compos. Math. 149 (2013), no. 7, 1129–1149.

[16] 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`

[17] J. Maynard, Small gaps between primes. Ann. of Math. (2) 181 (2015), no. 1, 383–413.

[18] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory. I. Classical theory.* Cambridge Studies in Advanced Mathematics, 97. Cambridge University Press, Cambridge, 2007.

[19] J. Pintz, *Elementary methods in the theory of L-functions. I. Hecke's theorem.* Acta Arith. **31** (1976), no. 1, 53–60.

[20] 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

[21] L. G. Sathe, *On a problem of Hardy on the distribution of integers having a given number of prime factors. III.* J. Indian Math. Soc. (N.S.) 18, (1954). 27–42.

[22] —, *On a problem of Hardy on the distribution of integers having a given number of prime factors. IV.* J. Indian Math. Soc. (N.S.) 18, (1954). 43–81.

[23] A. Selberg, *Note on a paper by L. G. Sathe.* J. Indian Math. Soc. (N.S.) 18, (1954). 83–87.

[24] —, *Lectures on sieves.* Collected papers. Vol. II. With a foreword by K. Chandrasekharan. Springer-Verlag, Berlin, 1991.

[25] 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.

[26] T. Tao, *The parity problem is sieve methods.* Blog post: `terrytao.wordpress.com/2007/06/05/open-question-the-parity-problem-in-sieve-theory/`

[27] —, *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`

[28] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres.* Third edition, coll. Échelles, Belin, 2008.

[29] Y. Zhang, *Bounded gaps between primes.* Ann. of Math. (2) 179 (2014), no. 3, 1121–1174.