Chapter 5

An explicit formula for counting primes

So far, we have seen various ways of counting primes using combinatorial devices. We now introduce a different approach that transforms the problem of estimating $\pi(x)$ into a problem in complex analysis. The key idea is to package the primes all together and form an appropriate generating function.

Given an arithmetic function f, the most common generating function attached to f is arguably its power series

$$A(z) = \sum_{n \geqslant 1} f(n) z^n.$$

This series converges to a holomorphic function in a disk $|z| < R$. Moreover, $f(n)$ can be recovered from $A(z)$ via Cauchy's residue formula that implies

$$f(n) = \frac{1}{2\pi i} \oint_{|z|=r} \frac{A(z)}{z^{n+1}} \mathrm{d}z \qquad (n \in \mathbb{N},\ 0 < r < R). \tag{5.1}$$

We apply (5.1) when $f = 1_P$, the indicator function of the sequence of primes. The associated power series is

$$Q(z) := \sum_{p} z^p.$$

Summing (5.1) for $n = 0, 1, \ldots, N$ when $f = 1_P$ yields the inversion formula

$$\sum_{p \leqslant N} 1 = \sum_{0 \leqslant n \leqslant N} \frac{1}{2\pi i} \oint_{|z|=r} \frac{Q(z)}{z^{n+1}} \mathrm{d}z = \frac{1}{2\pi i} \oint_{|z|=r} \frac{Q(z)(1 - z^{N+1})}{z^{N+1}(1 - z)} \mathrm{d}z \tag{5.2}$$

for any $r \in (0, 1)$. Hence, a good understanding of the analytic behavior of $Q(z)$ can lead us to precise estimates for the counting function of the primes.

The above strategy arrives quickly at a dead end because it is not clear how to control the function $Q(z)$ without already knowing a lot about primes. As a matter of fact, the same objection can be raised for any generating function associated to the sequence of primes: how is it possible to determine its asymptotic behavior without already having a good grasp of the distribution of primes?

To break the vicious cycle, we analyze $Q(z)$ more closely. This function is naturally tied to the additive structure of the sequence of prime numbers. For example, note that

$$Q(z)^k = \sum_{p_1,\dots,p_k} z^{p_1+\cdots+p_k} = \sum_{n\geqslant 0} g_k(n) z^n,$$

where $g_k(n)$ is the number of ways to write n as the sum of k primes. However, primes are multiplicative objects, so it is more natural to study them from a multiplicative point of view. To this end, we observe that the logarithmic function is a group isomorphism from $(\mathbb{R}_{>0}, \times)$ to $(\mathbb{R}, +)$. We are thus naturally led to consider the generating function

$$\sum_p z^{\log p}.$$

This is no longer a power series because the exponents are not integers.

Note that $z^{\log p} = p^{\log z}$. Working with the complex logarithm causes technical difficulties. For this reason, we make the change of variables $s = -\log z$, so that our generating function becomes the Dirichlet series

$$\mathscr{P}(s) := \sum_p \frac{1}{p^s}.$$

In view of Mertens' second estimate (Theorem 3.4), this Dirichlet series has abscissa of convergence 1. In particular, Theorem 4.5 tells us that it defines a holomorphic function in the half-plane $\mathrm{Re}(s) > 1$.

Let us now consider the kth power of $\mathscr{P}$: we have

$$\mathscr{P}(s)^k = \sum_{p_1,\dots,p_k} \frac{1}{(p_1 \cdots p_k)^s} = \sum_{n\geqslant 1} \frac{r_k(n)}{n^s},$$

where $r_k(n)$ is the number of ways to write n as the product of k primes. In particular, r_k is supported on integers with $\leqslant k$ prime factors. In comparison, before we had no control over the support of g_k. We thus see right away that $\mathscr{P}(s)$ has better properties than $Q(z)$.

Taking the above argument one step further, Euler proved that $\mathscr{P}(s)$ can be written in terms of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$, which is for $\mathbb{N}$ what $\mathscr{P}(s)$ is for the sequence of primes. The key is Euler's

product formula

$$\zeta(s) = \prod_p \Big(1 - \frac{1}{p^s}\Big)^{-1} \qquad (\mathrm{Re}(s) > 1)$$

that we proved in the previous chapter. Taking logarithms, we infer that

$$\log \zeta(s) = \sum_p \sum_{m \geqslant 1} \frac{1}{m p^{ms}} = \sum_{m \geqslant 1} \frac{\mathscr{P}(ms)}{m} \tag{5.3}$$

which provides the link between $\mathscr{P}$ and ζ. The above formula is the starting point of analytic number theory, as it relates the function $\mathscr{P}$, for which we knew nothing about, to the function ζ. The latter is significantly simpler because it is defined as a summation over all integers, a very regular set. It thus seems plausible that we can obtain good estimates for $\mathscr{P}$ via this link.

As in the case of the function $Q(z)$ and the inversion formula (5.2), we want to find a passage from $\mathscr{P}(s)$ to $\pi(x) = \sum_{p \leqslant x} 1$. We start by writing

$$\mathscr{P}(s) = \int_1^\infty x^{-s} \mathrm{d}\pi(x) = s \int_0^\infty \pi(x) x^{-s-1} \mathrm{d}x. \tag{5.4}$$

Hence, we see that the function $\mathscr{P}(-s)/(-s)$ is the *Mellin transform* of the function $\pi(x)$. (A brief introduction to the necessary theory of the Mellin transform is given in the last section of Appendix B.) Mellin inversion allows us to go from (5.4) to the formula

$$\sum_{p<x} 1 + \frac{1_{x \text{ is prime}}}{2} = \frac{1}{2\pi i} \int_{(\alpha)} \mathscr{P}(s) \frac{x^s}{s} \mathrm{d}s, \tag{5.5}$$

where $\int_{(\alpha)} f(s)\mathrm{d}s$ denotes the *principal value* of $\int_{\mathrm{Re}(s)=\alpha} f(s)$, namely

$$\int_{(\alpha)} f(s)\mathrm{d}s = \lim_{T \to \infty} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\,\mathrm{Im}(s)| \leqslant T}} f(s)\mathrm{d}s. \tag{5.6}$$

Indeed, to see (5.5), we apply Theorem B.4 (whose hypotheses are met here with $\alpha_1 = -\infty$ and $\alpha_2 = -1$) and then make the change of variables $s \to -s$.

Jumping into the void

The inversion formula (5.5) expresses $\pi(x)$ in terms of the Riemann zeta function. However, it is not that useful as it stands for the estimation of $\pi(x)$. Indeed, we expect that there are about $x/\log x$ primes $\leqslant x$. On the other hand, we have $|x^s| = x^\alpha$ on the right side of (5.5). Since $\alpha > 1$, the size of x^s is bigger than the expected main term. This means that if we are to extract an asymptotic estimate for $\pi(x)$ from (5.5), we must understand the integrand in a way that is precise enough to establish significant cancellation among the different parts of the range of integration. Obtaining such sharp estimates on $\mathscr{P}$ without already controlling $\pi(x)$ seems impossible.

It thus seems that we have again reached an impasse. Riemann though had a brilliant idea to circumvent it. He realized that $\zeta(s)$ can be extended in a canonical way to values of s outside its domain of convergence using the theory of analytic and meromorphic continuation of complex analysis.[1] We do not need to delve too deeply into this theory; as we will see shortly, the special structure of ζ allows us to meromorphically continue it[2] to $\mathbb{C}$ relatively easily. The extension we obtain has only one singularity: a simple pole of residue 1 at $s = 1$. Such an extension must be unique by the identity principle. Thus ζ really is well-defined over $\mathbb{C}$. Using this fact and Cauchy's residue theorem, we can then replace the line of integration in (5.5) by a new contour that reaches to the left of the vertical line $\mathrm{Re}(s) = 1$, where x^s becomes of smaller magnitude than x. Hence, we can hope to obtain bounds for this new integral that are of genuinely smaller order than $x/\log x$. The main term to the approximation of $\pi(x)$ will arise from the singularities in the region encircled by the old and the new contour of integration. The end result of this calculation will be a formula for $\pi(x)$ in terms of the singularities of $\mathscr{P}$.

We devote the rest of this chapter to making the above discussion more precise and to laying Riemann's idea on rigorous mathematical grounds.

The meromorphic continuation of ζ

Perhaps the simplest way of meromorphically continuing ζ is to use the Euler-Maclaurin formula. Indeed, when $\mathrm{Re}(s) > 1$, $\zeta(s)$ is defined as the sum of the smooth function $1/n^s$ over $n \geqslant 1$, so Theorem 1.10 implies that

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{\{y\}}{y^{s+1}}\mathrm{d}y. \tag{5.7}$$

The integral on the right side converges absolutely for $\mathrm{Re}(s) > 0$ because $\{y\}$ is bounded. Thus, the right side of (5.7) supplies a meromorphic continuation of ζ to the half-plane $\mathrm{Re}(s) > 0$. The only singularity of ζ in this half-plane is a simple pole at $s = 1$ of residue 1 (a reflection of the divergence of the harmonic series $\sum_{n=1}^\infty 1/n$).

More generally, Exercise 1.10(b) implies that

$$\zeta(s) = \frac{s}{s-1} + \sum_{\ell=1}^{k} \frac{B_\ell}{\ell!}\prod_{j=0}^{\ell-2}(s+j) - \frac{\prod_{j=0}^{k-1}(s+j)}{k!}\int_1^\infty \frac{B_k(\{x\})}{x^{s+k}}\mathrm{d}x \tag{5.8}$$

[1] In fact, this theory was partly pioneered by Riemann himself.

[2] The YouTube channel 3Blue1Brown has an excellent video about the meromorphic continuation of ζ that is called "Visualizing the Riemann hypothesis and analytic continuation". The video is located at the web address `https://www.youtube.com/watch?v=sD0NjbwqlYw`.

for $\mathrm{Re}(s) > 1$. Since the right side is meromorphic for $\mathrm{Re}(s) > -k+1$ with only a simple pole at $s = 1$ of residue 1, so is ζ. Letting $k \to \infty$ establishes the alleged meromorphic continuation of ζ to the entire complex plane.

Let us now examine what the above discussion tells us about the analytic character of $\mathscr{P}$. We start from relation (5.3). Since $\sum_{m\geqslant 2} \mathscr{P}(ms)/m = \sum_{m\geqslant 2,\, p} 1/(mp^{ms}) = O(1)$ for $\mathrm{Re}(s) \geqslant 1$, we find that $\mathscr{P}(s) = \log\zeta(s) + O(1)$ for $\mathrm{Re}(s) > 1$. In particular, $\mathscr{P}(s) \sim -\log(s-1)$ as $s \to 1$, that is to say, $\mathscr{P}(s)$ has a logarithmic singularity at $s = 1$. This type of singularity prohibits us from extending $\mathscr{P}$ to an analytic function around $s = 1$. In particular, we cannot apply Cauchy's residue theorem to an integral of the form $\int_C (\mathscr{P}(s)x^s/s)\mathrm{d}s$, where C is a closed contour going around 1. For this reason, extracting the main term for $\pi(x)$ from (5.5) is a bit hard (though certainly possible as Riemann himself explained in his 1859 manuscript).

The above obstacle is merely of a technical nature. To overcome it, recall that the asymptotic behavior of $\pi(x)$ can be extracted from that of Chebyshev's theta and psi functions

$$\theta(x) = \sum_{p\leqslant x} \log p \qquad \text{and} \qquad \psi(x) = \sum_{n\leqslant x} \Lambda(n).$$

Indeed, we saw in Examples 1.8 and 1.9 how to go back and forth between $\pi(x)$ and $\theta(x)$. In addition, Chebyshev's functions are very close to each other in virtue of Exercise 2.7 which implies that

$$|\theta(x) - \psi(x)| \ll \sqrt{x} \qquad (x \geqslant 2).$$

Therefore, instead of estimating $\pi(x)$, we may work with $\psi(x)$. We need an analogue of formula (5.5) for this function.

In general, a straightforward adaptation of the proof of (5.5) implies the following generalization: if f is an arithmetic function whose Dirichlet series F converges absolutely in the half-plane $\mathrm{Re}(s) > 1$, then

$$\sum_{n<x} f(n) + \frac{1_{x\in\mathbb{N}} f(x)}{2} = \frac{1}{2\pi i}\int_{(\alpha)} F(s)\frac{x^s}{s}\mathrm{d}s \qquad (x > 1,\ \alpha > 1). \tag{5.9}$$

This general identity is called the *Perron inversion formula*.

We apply (5.9) with $f = \Lambda$ whose summatory function is Chebyshev's psi function. The associated Dirichlet series is $-\zeta'/\zeta$. Since ζ is meromorphic over $\mathbb{C}$, so is $-\zeta'/\zeta$. They both have a simple pole of residue 1 at $s = 1$. Moreover, if z is a zero of ζ multiplicity m, then ζ'/ζ has a simple pole of residue m at $s = z$. Indeed, we may write $\zeta(s) = (s-z)^m g(s)$ with g analytic and non-zero in a neighborhood of z. Hence, $(\zeta'/\zeta)(s) = m/(s-z) + (g'/g)(s)$ and g'/g is analytic around z. This implies that

$$\mathrm{res}_{s=z}(\zeta'/\zeta)(s) = m. \tag{5.10}$$

As we will see in the next chapter, the zeroes of ζ fall under two categories: the *trivial zeroes*, which are located at $-2, -4, -6, \ldots$, and the *non-trivial zeroes*, which are located in the strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$. We denote a generic non-trivial zero by[3] $\rho = \beta + i\gamma$.

Remarkably, there is an *explicit formula* for $\psi(x)$ in terms of the non-trivial zeroes of the Riemann zeta function.[4]

Theorem 5.1. *For all $x, T \geqslant 2$, we have*

$$\psi(x) = x - \sum_{|\gamma| \leqslant T} \frac{x^\rho}{\rho} + O\bigg(\frac{x \log^2(xT)}{T} + \log x\bigg), \tag{5.11}$$

where the sum runs over the non-trivial zeroes of ζ with each zero repeated as many times as its multiplicity.

Before we explain why Theorem 5.1 is true, let us momentarily pause and make a few comments about it. This astonishing result reveals that primes, an elementary arithmetic object, have a "dual" complex-analytic object associated to them: the zeroes of ζ. These two objects of seemingly unrelated nature are interconnected in a fundamental way: the main term on the right-hand side of (5.11) approximates $\psi(x)$ better and better as $T \to \infty$, similarly to the Fourier expansion of a periodic function. Hence, the zeroes of ζ encode in principle everything we need to know about the distribution of primes (and vice versa). We may think of the zeroes as "frequencies" with which the counting function of prime numbers resonates. For this reason, they are of fundamental importance in mathematics.

Theorem 5.1 will play a key role in the proof of the Prime Number Theorem. Indeed, to establish the asymptotic formula $\psi(x) \sim x$, it suffices to bound $\sum_{|\gamma| \leqslant T} x^\rho/\rho$ and prove that it is of negligible size compared to x. Since $|x^\rho| = x^\beta$, this essentially reduces the Prime Number Theorem to showing that β is a bit less than 1 for all zeroes of ζ.

Cauchy's residue theorem and the explicit formula

Let us now give a rough sketch of the proof of Theorem 5.1. The complete details will be given in Chapter 8, after having developed the necessary tools.

We present the argument in a more general context. Recall the Perron inversion formula (5.9), valid for any arithmetic function f whose Dirichlet series F converges absolutely to the right of the line $\mathrm{Re}(s) = 1$. Similarly to

[3] The letter γ here is not to be confused with Euler-Mascheroni's constant defined by (1.13). This ambiguous notation is customary in the literature.

[4] The contribution of the trivial zeroes has been absorbed into the error term. There is an even more precise version of the explicit formula that takes into account trivial zeroes (see Exercise 8.2(a) and [31, Chapter 17]). The version stated in Theorem 5.1 is sufficient for most applications.

ζ and ζ'/ζ, the Dirichlet series F of many interesting arithmetic functions can be meromorphically continued to a half-plane $\mathrm{Re}(s) > \alpha_0$ with $\alpha_0 < 1$. In this case, the integral on the right-hand side of (5.9) can be studied using complex analysis as we explain below.

Fix $\alpha' \in (\alpha_0, 1) \setminus \{0\}$ such that $F(s)$ has no poles when $\mathrm{Re}(s) = \alpha'$. Such an α' always exists because F has at most countably many singularities in any given open region. Moreover, let $T = T(x)$ be large enough so that

$$\sum_{n \leqslant x} f(n) = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\mathrm{Im}(s)| \leqslant T}} F(s) \frac{x^s}{s} \mathrm{d}s + E, \tag{5.12}$$

with $E = o(\sum_{n \leqslant x} |f(n)|)$. The existence of such a T is guaranteed by (5.6). Furthermore, similarly to α', the parameter T can be chosen in such a way that F has no singularities on the lines $\mathrm{Im}(s) = \pm T$.

Let C_1 denote the contour of integration in (5.12), that is to say, the line segment from $\alpha - iT$ to $\alpha + iT$. We write symbolically $C_1 = [\alpha - iT, \alpha + iT]$. We deform C_1 to a new contour of integration consisting of the line segments $C_2 = [\alpha - iT, \alpha' - iT]$, $C_3 = [\alpha' - iT, \alpha' + iT]$ and $C_4 = [\alpha' + iT, \alpha + iT]$ (see Figure 5.1). We denote this new contour by $C_2 + C_3 + C_4$.[5] We claim that

$$\frac{1}{2\pi i} \int_{C_1} F(s) \frac{x^s}{s} \mathrm{d}s = \sum_{2 \leqslant j \leqslant 4} \frac{1}{2\pi i} \int_{C_j} F(s) \frac{x^s}{s} \mathrm{d}s + \sum_w \mathrm{res}_{s=w} \frac{F(s)x^s}{s}, \tag{5.13}$$

where the rightmost sum runs over all singularities of $F(s)/s$ in $\Omega := \{ s \in \mathbb{C} : \alpha' < \sigma < \alpha, |t| < T \}$. Indeed, the integrand $F(s)x^s/s$ is meromorphic in Ω and analytic in an open neighborhood of the boundary $\partial\Omega$. Since $\partial\Omega = C_1 - C_2 - C_3 - C_4$ when traversed counterclockwise, Cauchy's residue theorem implies that

$$\frac{1}{2\pi i} \oint_{\partial\Omega} F(s) \frac{x^s}{s} \mathrm{d}s = \sum_w \mathrm{res}_{s=w} \frac{F(s)x^s}{s}.$$

This proves our claim that (5.13) holds.

Combining (5.12) and (5.13), we infer that

$$\sum_{n \leqslant x} f(n) = \sum_w \mathrm{res}_{s=w} \frac{F(s)x^s}{s} + E + R,$$

where

$$R = \sum_{2 \leqslant j \leqslant 4} \frac{1}{2\pi i} \int_{C_j} F(s) \frac{x^s}{s} \mathrm{d}s.$$

We think of R as an error term because $|x^s/s| \leqslant x^\sigma/|t|$, so that the integrand $F(s)x^s/s$ is small on $C_2 \cup C_4$ because $|t| = T$ is large, and it is also small on

[5] In general, if C, C' are two contours with a given orientation, then $C + C'$ denotes the contour that first traces C and then C' in their respective orientation. Furthermore, $-C$ is the contour C traced in the opposite orientation.

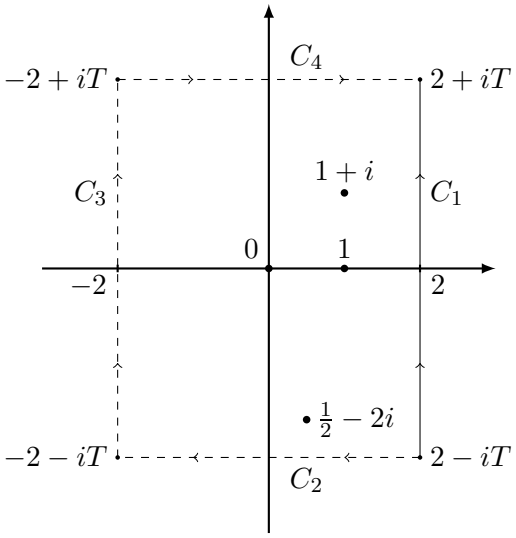

Figure 5.1. The poles of $\zeta(s)\zeta(s-i)\zeta(s+1/2+2i)/s$ inside the rectangle defined by the points $\pm 2 \pm iT$.

C_3 because $\sigma = \alpha' < 1$. In reality, we also need bounds on $F(s)$ to estimate R. Such estimates can be a bit tricky to obtain outside the region of absolute convergence. We will see methods of establishing them in Chapters 6, 8 and 11.

Assuming that R is indeed negligible, we are led to the guesstimate

$$\sum_{n \leqslant x} f(n) \approx \sum_{\substack{w \text{ is a pole of } F(s)/s \\ \alpha' < \mathrm{Re}(w) < \alpha, \, |\mathrm{Im}(w)| < T}} \mathrm{res}_{s=w} \frac{F(s)x^s}{s}. \tag{5.14}$$

Combining this heuristic with (5.10) explains why $\psi(x)$ should be closely approximated by the sum $x - \sum_{|\gamma| \leqslant T} x^\rho/\rho$ from Theorem 5.1. The rigorous proof of Theorem 5.1 will be given in Chapter 8, after having developed further the theory of the Riemann zeta function (in Chapter 6) and of the Perron inversion formula (in Chapter 7). We will then use the explicit formula for $\psi(x)$ together with a bound on the zeroes of ζ to establish the Prime Number Theorem in Chapter 8.

We conclude this chapter with some examples that showcase the utility and versatility of the ideas presented above.

Example 5.2. As a toy example, consider the function $f = 1$. We then have that $F = \zeta$, whose only singularity is a simple pole of residue 1 at $s = 1$. Thus, the only singularity of $\zeta(s)x^s/s$ in the half-plane $\mathrm{Re}(s) > 0$ is a simple pole of residue x at $s = 1$. This leads us to the prediction that

$\sum_{n\leqslant x} 1 \approx x$. This is of course true, since we know by elementary methods that $\sum_{n\leqslant x} 1 = x + O(1)$. □

Remark 5.3. In general, if F has a simple pole of residue r_w at a point w that is different than the origin, then

$$\operatorname{res}_{s=w} \frac{F(s)x^s}{s} = \frac{r_w x^w}{w}.$$

We can generalize this calculation further: if F has a pole of order m at $w \neq 0$, then there are coefficients $c_{w,0}, c_{w,1}, \dots, c_{w,m-1} \in \mathbb{C}$ such that

$$\operatorname{res}_{s=w} \frac{F(s)x^s}{s} = x^w\big(c_{w,m-1}(\log x)^{m-1} + c_{w,m-2}(\log x)^{m-2} + \cdots + c_{w,0}\big).$$

Indeed, let $F(s)/s = a_{w,m}/(s-w)^m + \cdots + a_{w,1}/(s-w) + \sum_{j\geqslant 0} b_{w,j}(s-w)^j$ be the Laurent expansion of $F(s)/s$ about $s = w$. In addition, we have the Taylor series expansion $x^s = x^w \sum_{j\geqslant 0}(s-w)^j(\log x)^j/j!$. Hence, the claimed formula for $\operatorname{res}_{s=w}(F(s)x^s/s)$ holds with $c_{w,j} = a_{w,j+1}/j!$. □

Example 5.4. Consider the divisor function τ, for which we have the convolution identity $\tau = 1 * 1$. Thus, its Dirichlet series is $\zeta(s)^2$, which has a meromorphic continuation to $\mathbb{C}$ with its only pole being a double pole of order 2 at $s = 1$. In view of relation (5.14) and Remark 5.3, we are led to predict that there are coefficients $c_0, c_1 \in \mathbb{C}$ such that

$$\sum_{n\leqslant x} \tau(n) \approx c_1 x \log x + c_0 x.$$

To calculate c_0 and c_1, note that $\zeta(s) = 1/(s-1) + \gamma + O(|s-1|)$ for $|s-1| \leqslant 1/2$ by Exercise 5.2, whereas $1/s = 1 - (s-1) + O(|s-1|^2)$. Hence

$$\frac{\zeta(s)^2}{s} = \frac{1}{(s-1)^2} + \frac{2\gamma - 1}{s-1} + O(1),$$

which implies that $c_1 = 1$ and $c_0 = 2\gamma - 1$. This agrees with Theorem 3.3. □

Example 5.5. Let f be the indicator function of square-full integers (see Exercise 1.6). In Exercise 3.11, we saw that the partial sums of f up to x have an asymptotic expansion with two main terms, of size $x^{1/2}$ and $x^{1/3}$, respectively. These terms can be guessed using (5.14): the multiplicativity of f implies that its Dirichlet series equals

$$F(s) = \prod_p \Big(1 + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots\Big) = \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)} \tag{5.15}$$

for $\operatorname{Re}(s) > 1$. Since ζ has a meromorphic continuation to $\mathbb{C}$, so does F. In addition, the only singularities of F in the half-plane $\operatorname{Re}(s) > 1/6$ are simple

poles at the points $s = 1/2$ and $s = 1/3$. They both arise from the simple pole of ζ at $s = 1$. Relation (5.14) then leads us to the prediction that

$$\#\{\, n \leqslant x : n \text{ square-full} \,\} \approx \frac{\zeta(3/2)}{\zeta(3)} x^{1/2} + \frac{\zeta(2/3)}{\zeta(2)} x^{1/3}. \qquad \square$$

Exercises

Exercise 5.1. Prove that

$$\zeta(s) = \frac{1}{1 - 2^{-s+1}} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} \qquad (\mathrm{Re}(s) > 0).$$

Exercise 5.2. When $0 < |s-1| \leqslant 1$, show the following estimates:

$$\begin{aligned} \zeta(s) &= \frac{1}{s-1} + \gamma + O(|s-1|)\,; \\ \log \zeta(s) &= -\log(s-1) + \gamma \cdot (s-1) + O(|s-1|^2) \qquad (s \notin [-1, 0])\,; \\ \frac{\zeta'}{\zeta}(s) &= -\frac{1}{s-1} + \gamma + O(|s-1|)\,. \end{aligned}$$

Exercise 5.3. Use (5.14) to predict the main term in the asymptotic formulas for $\sum_{n \leqslant x} \log n$, $\sum_{n \leqslant x} \varphi(n)$, $\sum_{n \leqslant x} \mu^2(n)$, $\sum_{n \leqslant x} \tau_3(n)$ and $\sum_{n \leqslant x} \tau(n)^2$. Compare your prediction with Theorems 1.12 and 3.2, and Exercises 3.8, 3.10 and 4.5, respectively.

Exercise 5.4.* Complete the proof of Theorem 3.4(c) as follows:

(a) Uniformly for $x \geqslant 2$ and $\varepsilon \in (0, 1]$, prove that

$$\sum_{p \leqslant x} \log\Big(1 - \frac{1}{p}\Big) = \sum_{p \leqslant x} \log\Big(1 - \frac{1}{p^{1+\varepsilon/\log x}}\Big) + O(\varepsilon).$$

(b) Uniformly for $x \geqslant 2$ and $\varepsilon \in (0, 1]$, prove that

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

[*Hint:* Taylor's theorem.]

(c) Deduce that the constant in (3.6) is $\kappa = \int_0^{\infty} u^{-1}(e^{-u} - 1_{[0,1]}(u))\mathrm{d}u$. [*Hint:* Use Exercise 5.2 to rewrite $\log \zeta(1 + \varepsilon/\log x)$.]

(d) Prove that $\gamma = \int_0^{\infty} u^{-1}(1_{[0,1]}(u) - e^{-u})\mathrm{d}u$.
[*Hint:* Note that $\gamma = \lim_{N\to\infty}(-\log N + \int_0^1 (1 + x + \cdots + x^{N-1})\mathrm{d}x)$ and let $x = 1 - u/N$.]

Chapter 23

Vinogradov's method

The parity barrier of sieve methods prevents us from getting tight bounds on $\sum_{p\leqslant x} a_p$ under the mere assumption of Axioms 1–3 for the sequence $\mathcal{A} = (a_n)_{n=1}^{\infty}$. In 1934, I. M. Vinogradov[1] developed a new method for estimating $\sum_{p\leqslant x} a_p$ when $\mathcal{A}$ satisfies certain additional hypotheses.

To simplify the exposition of Vinogradov's idea, let us assume that $|a_n| \leqslant 1$ for all n. We then have

$$\sum_{p\leqslant x} a_p = \sum_{\substack{n\leqslant x\\ P^-(n)>\sqrt{x}}} a_n + O(\sqrt{x}).$$

Applying a variant of Buchstab's identity (19.1) to the right-hand side yields that

$$\sum_{p\leqslant x} a_p = \sum_{\substack{n\leqslant x\\ P^-(n)>x^\varepsilon}} a_n - \sum_{x^\varepsilon<p\leqslant x} \sum_{\substack{n\leqslant x\\ P^-(n)=p}} a_n + O(\sqrt{x}), \tag{23.1}$$

where $\varepsilon > 0$ is at our disposal. If we assume that the sequence $\mathcal{A}$ satisfies a suitable version of Axioms 1–3, the first sum on the right-hand side of (23.1) can be estimated accurately using the Fundamental Lemma of Sieve Theory (Theorem 18.11) for small enough values of ε. Thus, it remains to handle the double sum over p and n.

Writing $n = pm$, we find that

$$B := \sum_{x^\varepsilon<p\leqslant x} \sum_{\substack{n\leqslant x\\ P^-(n)=p}} a_n = \sum\sum_{\substack{x^\varepsilon<p\leqslant x\\ mp\leqslant x,\, P^-(m)\geqslant p}} a_{mp}.$$

[1]Not to be confused with A. I. Vinogradov from the Bombieri-Vinogradov theorem.

The right-hand side closely resembles a *bilinear sum*

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L} a_{k\ell}x_k y_\ell \tag{23.2}$$

for appropriate coefficients x_k and y_ℓ. There is a small technicality: the variables p and m are weakly tangled via the relations $pm \leqslant x$ and $P^-(m) \geqslant p$. We can easily decouple them though: we (roughly) have

$$B \approx \sum_{x^\varepsilon < 2^j \leqslant x^{1/2}} B_j \qquad \text{with} \qquad B_j = \sum_{\substack{2^{j-1} < p \leqslant 2^j \\ m \leqslant x/2^j,\, P^-(m) \geqslant 2^j}}\sum a_{mp}, \tag{23.3}$$

so that B is a sum of $O(\log x)$ bilinear sums (B_j is of the form (23.2) with $K = x/2^j$, $L = 2^j$, $x_k = 1_{P^-(k)\geqslant 2^j}$ and $y_\ell = 1_{\ell \text{ is prime}} 1_{\ell \in (2^{j-1}, 2^j]}$).

Vinogradov's groundbreaking idea is that, for certain special sequences $\mathcal{A}$, we can obtain strong estimates for the bilinear sum (23.2) no matter what the coefficients x_k and y_ℓ are, as long as they are of controlled size (e.g. if $|x_k|, |y_\ell| \leqslant 1$ for all k, ℓ) and as long as both K and L are large, so that we have genuine bilinearity.[2] We may thus forget the precise definition of x_k and y_ℓ. If this alleged bilinear estimate (which we can think of as "Axiom 4" of sieve theory) is available in a large enough region of K and L so that both terms on the right-hand side of (23.1) can be handled (the first one by Axioms 1–3 and the second one by Axiom 4), we can break the parity barrier and extract primes from the sequence $(a_n)_{n=1}^\infty$.

We will explain Vinogradov's method more rigorously in the subsequent sections. But first let us note that Axiom 3 of sieve methods can also be thought of as an estimate for a bilinear sum of the form (23.2), but with $y_\ell = 1$ for all ℓ. Indeed, if $(a_n)_{n=1}^\infty \subset [1, L]$ and we assume Axiom 1, then

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L} a_{k\ell}x_k = \sum_{k=1}^{K} x_k A_k = X\sum_{k=1}^{K}\frac{x_k\nu(k)}{k} + \sum_{k=1}^{K} x_k r_k,$$

where A_k is defined by (18.2). If we assume that $|x_k| \leqslant 1$ and that Axiom 3 holds with level of distribution $D \geqslant K$, then we can obtain a strong estimate for $\sum_{k\leqslant K} x_k r_k$. Conversely, if we can estimate this sum for any choice of x_k, we can also estimate it when x_k is the sign of r_k, which brings us right back to Axiom 3.

In conclusion, we may think of Axiom 3 as a bilinear estimate with the coefficients y_ℓ being smooth functions of ℓ. This point of view will be important in the next section.

[2] If, for instance, $K = 1$, then the expression in (23.2) becomes a sum over a single variable. We want to avoid such degenerate situations.

Two types of functions

Various technicalities in Vinogradov's method are simplified if instead of the sum $\sum_{p\leqslant x} a_p$ we work with $\sum_{n\leqslant x} a_n\Lambda(n)$. Indeed, the combinatorial identity $\Lambda = \mu * \log$ readily implies that

$$\sum_{n\leqslant x} a_n\Lambda(n) = \sum_{k\ell\leqslant x} a_{k\ell}\mu(k)\log\ell. \tag{23.4}$$

We thus see right away that $\sum_{n\leqslant x} a_n\Lambda(n)$ has some sort of bilinear structure. To bring the right-hand side of (23.4) into the form (23.2), we localize k into a dyadic interval $(2^{j-1}, 2^j]$, so that $\ell \leqslant x/2^{j-1}$. As we briefly mentioned before, the method of bilinear sums is efficient only when both k and ℓ are "long variables", that is to say, when 2^j and $x/2^j$ are both large (say when $D \leqslant 2^j \leqslant x/D$). On the other hand, when $2^j \leqslant D$, we can take advantage of the fact that the long variable ℓ is weighted with the smooth function log. Hence, this part of the sum can be handled too, provided that we have at our disposal an appropriate version of Axiom 3, as per the discussion in the end of the previous section. It remains to handle the summands with $x/D < 2^j \leqslant x$. If we can rewrite this part of the sum as a linear combination of sums that fit into one of the two above categories (i.e., a combination of some bilinear sums, and of some other ones with at least one smooth variable), we will have completed the estimation of $\sum_{n\leqslant x} a_n\Lambda(n)$.

This brings us to the heart of Vinogradov's method: given $x \geqslant 1$, we seek an identity of the form

$$\Lambda(n) = \sum_{1\leqslant j\leqslant J} (f_j * g_j)(n) + R(n) \qquad \text{for } n \leqslant x, \tag{23.5}$$

where the function R is a negligible "remainder term" in the sense that $\sum_{n\leqslant x} |a_n R(n)|$ is small compared to $\sum_{n\leqslant x} |a_n|$, and for each j the summands $f_j * g_j$ fall into one of the following two categories:

I) $\operatorname{supp}(f_j) \subseteq [1, y_j]$ for some y_j that is small compared to x and $g_j \in C^\infty(\mathbb{R}_{\geqslant 1})$. We then call $f_j * g_j$ a *quasi-smooth* or *type I function* and refer to the sum

$$\sum_{n\leqslant x} a_n(f_j * g_j)(n) = \sum_{k\leqslant y_j} f_j(k) \sum_{\ell\leqslant x/k} a_{k\ell} g_j(\ell)$$

as a *quasi-smooth*, *quasi-linear* or *type I sum*.

II) $\operatorname{supp}(f_j) \subseteq [1, y_j]$ and $\operatorname{supp}(g_j) \subseteq [1, z_j]$, where $D_j \leqslant y_j, z_j \leqslant x/D_j$ for some large D_j. We then call $f_j * g_j$ a *type II function* and its average

$$\sum_{n\leqslant x} a_n(f_j * g_j)(n) = \sum_{k\leqslant y_j,\, \ell\leqslant z_j,\, k\ell\leqslant x} \sum a_{k\ell} f_j(k) g_j(\ell)$$

a *bilinear* or *type II sum*.

Decomposing von Mangoldt's function

Vaughan's identity. One of the simplest and most useful ways to arrive at an identity of the form (23.5) was discovered by Vaughan. Given an arithmetic function f and a parameter V, we write

$$f_{\leqslant V}(n) := 1_{n\leqslant V} \cdot f(n) \qquad \text{and} \qquad f_{>V}(n) := 1_{n>V} \cdot f(n). \tag{23.6}$$

With the above notation, the identity $\Lambda = \mu * \log$ can be written as

$$\Lambda = \mu_{\leqslant V} * \log \, + \, \mu_{>V} * \log . \tag{23.7}$$

The first term on the right-hand side of (23.7) is of type I. But the second term is neither of type I nor of type II. To proceed, we replace $\mu_{>V}$ by $\mu_{\leqslant V}$ using Möbius inversion: we have

$$\mu_{>V} * 1 = \delta \, - \, \mu_{\leqslant V} * 1, \tag{23.8}$$

where we recall the notation $\delta(n) = 1_{n=1}$ from Chapter 3. As preparation for inserting (23.8) into (23.7), we write the latter formula as

$$\Lambda = \mu_{\leqslant V} * \log \, + \, \mu_{>V} * 1 * \Lambda.$$

Because Λ has unrestricted support, we first split it as $\Lambda = \Lambda_{\leqslant U} + \Lambda_{>U}$, where U is some parameter, and then apply (23.8) only to the part of Λ supported on $[1, U]$. We conclude that

$$\Lambda = \mu_{\leqslant V} * \log \, + \, \mu_{>V} * 1 * \Lambda_{>U} \, + \, (\delta \, - \, \mu_{\leqslant V} * 1) * \Lambda_{\leqslant U}.$$

We have thus proven *Vaughan's identity*:

Lemma 23.1. *For any $U, V \geqslant 1$, we have*

$$\Lambda = \mu_{\leqslant V} * \log \, - \, (\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1 \, + \, (\Lambda_{>U} * 1) * \mu_{>V} \, + \, \Lambda_{\leqslant U}. \tag{23.9}$$

The function $\Lambda_{\leqslant U}$ is supported on small integers and hence contributes a negligible amount to averages of Λ. The function $\mu_{\leqslant V} * \log$ is a quasi-smooth convolution: the first factor is a bounded function supported on integers $\leqslant V$. Similarly, the function $(\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1$ is also a quasi-smooth convolution, with the factor $\Lambda_{\leqslant U} * \mu_{\leqslant V}$ being supported on $[1, UV]$ and satisfying the pointwise bound $|\Lambda_{\leqslant U} * \mu_{\leqslant V}| \leqslant \Lambda * 1 = \log$. We denote the total contribution to Λ of these two type I functions by

$$\Lambda^\sharp := \mu_{\leqslant V} * \log \, - \, (\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1. \tag{23.10}$$

Finally, the function

$$\Lambda^\flat := (\Lambda_{>U} * 1) * \mu_{>V} \tag{23.11}$$

is of type II: its first factor is supported on integers $> U$ and its second one on integers $> V$.

A very useful feature of $\Lambda^\flat$ is that one of its factors is the Möbius function that is completely aperiodic (see Corollary 13.4 and Exercise 23.4). As a

result, $\Lambda^\flat$ typically contributes to the error term in the estimation of the sum $\sum_{n\leqslant x} a_n\Lambda(n)$, so that the main term comes from $\Lambda^\sharp$. We thus think of $\Lambda^\sharp$ as the "structured" part of Λ. It resembles a sieve-type weight and we need a suitable version of Axiom 3 to estimate its averages. On the other hand, we think of $\Lambda^\flat$ as an "unstructured/random" error term, and we usually treat it using bilinear methods.

Remark 23.2. By definition, we have

$$\Lambda^\flat(n) = \sum_{\substack{k\ell=n\\ k>U,\,\ell>V}}\sum (\Lambda_{>U} * 1)(k)\mu(\ell).$$

When $n \leqslant x$, we have $U < k = n/\ell \leqslant x/V$. However, we often need better control of the support of the variables k and ℓ. To achieve this goal, we cover the interval $(U, x/V]$ by dyadic intervals $(2^{j-1}, 2^j]$, where $2^j \in (U, 2x/V]$. If $k \in (2^{j-1}, 2^j]$, we also have that $\ell = n/k \leqslant x/2^{j-1}$. This leads us to the more accurate decomposition

$$\Lambda^\flat(n) = \sum_{U<2^j\leqslant 2x/V} (f_j * g_j)(n) \quad \text{for} \quad n \leqslant x, \tag{23.12}$$

where $f_j(k) = (\Lambda_{>U} * 1)(k)1_{2^{j-1}<k\leqslant 2^j}$ and $g_j(\ell) = \mu(\ell)1_{V<\ell\leqslant x/2^{j-1}}$. □

Presieving Λ. In many occasions, it is advantageous to use a variant of Vaughan's identity whose summands enjoy slightly different properties. A simple way of obtaining such a variant is by *presieving* Λ. Indeed, since primes do not have small prime factors, we write

$$\Lambda(n) = \Lambda(n)\cdot 1_{P^-(n)>y} + \Lambda(n)\cdot 1_{P^-(n)\leqslant y}.$$

We expect $\Lambda(n)\cdot 1_{P^-(n)\leqslant y}$ to be small on average because it is supported on prime powers p^m with $p \leqslant y$. Next, we decompose the function $\Lambda(n)\cdot 1_{P^-(n)>y}$ by first replacing Λ by $\mu * \log$. This yields the identity

$$\Lambda(n)1_{P^-(n)>y} = \sum_{\substack{k\ell=n\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell. \tag{23.13}$$

The fact that $\log(1) = 0$ means that the above sum is supported on integers $\ell > 1$. Since we also know that $P^-(\ell) > y$, we must have $\ell > y$. We thus see that we automatically have a long ℓ variable weighted with the smooth function log times the indicator function of integers free of prime factors $\leqslant y$. Even though the latter is not a smooth function, it is quasi-smooth when y is small enough. The reason is that Theorem 19.1 allows us to approximate the function $n \to 1_{P^-(n)>y} = 1_{(n,P(y))=1}$ by convolutions $\lambda^\pm * 1$, where $\lambda^\pm$ take values in $[-1, 1]$ and have small support. Hence, for all practical purposes, we may think of the function $\ell \to 1_{P^-(\ell)>y}\log\ell$ as a quasi-smooth function.

Motivated by the above discussion, we split the right-hand side of (23.13) according to the size of k, which leads us to the following decomposition:

$$\Lambda = \Lambda^{\sharp}_{\text{sieve}} + \Lambda^{\flat}_{\text{sieve}} + R_{\text{sieve}}, \tag{23.14}$$

where

$$\Lambda^{\sharp}_{\text{sieve}}(n) = \sum_{\substack{k\ell=n,\, k\leqslant D\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell, \tag{23.15}$$

$$\Lambda^{\flat}_{\text{sieve}}(n) = \sum_{\substack{k\ell=n,\, k>D,\, \ell>y\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell \tag{23.16}$$

and $R_{\text{sieve}}(n) = 1_{P^-(n)\leqslant y}\Lambda(n)$. Note that $\Lambda^{\sharp}_{\text{sieve}}$ is essentially of type I, $\Lambda^{\flat}_{\text{sieve}}$ is of type II and R_{sieve} is of negligible size on average, since

$$\sum_{n\leqslant x} R_{\text{sieve}}(n) = \sum_{p\leqslant y,\, p^m\leqslant x}\sum \log p \leqslant \sum_{p\leqslant y}\log x \leqslant y\log x. \tag{23.17}$$

A choice of y and D that works for many applications is

$$y = \exp\{(\log x)^{\theta_1}\} \quad\text{and}\quad D = \exp\{(\log x)^{\theta_2}\}, \tag{23.18}$$

where $0 < \theta_1 < \theta_2 < 1$ can be chosen freely.

The main advantage of (23.14) compared to Vaughan's identity is that the functions $\Lambda^{\sharp}_{\text{sieve}}$ and $\Lambda^{\flat}_{\text{sieve}}$ are presieved with all primes $\leqslant y$. This rather technical feature of (23.14) plays a key role in the proof of Linnik's theorem in Chapter 27. We will also see in Exercise 26.4 how it leads to a better version of the Bombieri-Vinogradov theorem.

A secondary advantage of (23.14) versus Vaughan's identity is that its "main term" $\Lambda^{\sharp}_{\text{sieve}}$ consists of a single type I function. This fact makes various calculations easier and will come into play in Chapter 24.

On the other hand, Vaughan's identity offers much more freedom in the choice of the parameters U and V. Therefore, we have more control over the support of the functions appearing in the type I and type II sums, which is very important in certain applications. In contrast, the parameters y and D in (23.14) must be chosen carefully so that we have enough room to apply the Fundamental Lemma of Sieve Theory. In particular, y must be $x^{o(1)}$.

Remark 23.3. It is possible to create a new combinatorial decomposition of Λ that combines the best attributes of Vaughan's identity and of (23.14). This is done by presieving Vaughan's identity, that is to say, by multiplying all summands of (23.9) with the function $n \to 1_{P^-(n)>y}$. □

There are a lot more combinatorial decompositions of von Mangoldt's function than the ones we discussed above. A formula of particular importance is Heath-Brown's identity, given in Exercise 23.5 below. It is not

exactly of the form (23.5). Hence working with it is a bit more complicated, the task being understanding how to rearrange its terms and bring it to the form (23.5). However, Heath-Brown's identity has the important feature that all of the long functions appearing in it are smooth.

A further analysis of the subject of combinatorial decompositions of Λ can be found in [114, Chapter 13] or [59, Chapter 17]. Finally, a more sieve-theoretic approach to Vinogradov's method that is more in line with the discussion in the introduction of this chapter is presented in Harman's book on *prime-detecting sieves* [96].

The additive Fourier transform of the primes

To exemplify Vinogradov's method, we employ it to study a concrete and rather important example: the exponential sum

$$\sum_{p\leqslant x} e(\alpha p).$$

This sum is intimately related with the additive properties of primes and we will use it in the next chapter to study ternary arithmetic progressions in the primes. To get an idea of its size, we begin by studying it assuming the Generalized Riemann Hypothesis. This will serve as a guide for what kind of bounds to look for when we estimate it later via Vinogradov's method.

First, let us consider the special case when α is a rational number, say $\alpha = a/q$ with $(a,q)=1$. Then

$$\begin{aligned}\sum_{p\leqslant x} e(ap/q) &= \sum_{b\in(\mathbb{Z}/q\mathbb{Z})^*} e(ab/q)\pi(x;q,b) + \sum_{p\leqslant x,\, p\mid q} e(ap/q)\\ &= \frac{\operatorname{li}(x)}{\varphi(q)} \sum_{b\in(\mathbb{Z}/q\mathbb{Z})^*} e(ab/q) + O(\sqrt{x}q\log(qx))\end{aligned} \tag{23.19}$$

by Exercise 11.2 and partial summation. Making the change of variables $n \equiv ab \pmod q$, we see that the sum over b is the Gauss sum of the principal character mod q, which equals $\mu(q)$ (see Exercises 10.1 and 10.5). Therefore

$$\sum_{p\leqslant x} e(pa/q) = \frac{\mu(q)}{\varphi(q)} \cdot \operatorname{li}(x) + O(\sqrt{x}q\log(qx)).$$

To estimate $\sum_{p\leqslant x} e(p\alpha)$ for irrational α, we find a good rational approximation to it using the following classical result.

Lemma 23.4 (Dirichlet's approximation theorem)**.** *Let $\alpha \in \mathbb{R}$ and $Q \geqslant 1$. There is a reduced fraction a/q with $q \leqslant Q$ and*

$$\left|\alpha - \frac{a}{q}\right| < \frac{1}{qQ}.$$

Proof. Consider the $\lfloor Q \rfloor + 1$ numbers αq with $0 \leqslant q \leqslant Q$. We reduce them mod 1 to place them in the interval $[0,1)$. By the pigeonhole principle, there must exist $0 \leqslant q_1 < q_2 \leqslant Q$ such that $\|\alpha q_2 - \alpha q_1\| \leqslant 1/(\lfloor Q \rfloor + 1) < 1/Q$. We then take $q' = q_2 - q_1$ and a' to be the unique integer in $[\alpha q' - 1/2, \alpha q' + 1/2)$, so that $1 \leqslant q' \leqslant Q$ and $|\alpha q' - a'| = \|\alpha q'\| < 1/Q$. Letting a/q be the fraction a'/q' in reduced form completes the proof. □

Fix Q and a/q as in Lemma 23.4. If we write $\alpha = \beta + a/q$, then

$$\begin{aligned}\sum_{p\leqslant x} e(\alpha p) &= \int_{2^-}^{x} e(\beta y)\mathrm{d}\sum_{p\leqslant y} e(ap/q)\\ &= \frac{\mu(q)}{\varphi(q)}\int_2^x \frac{e(\beta y)}{\log y}\mathrm{d}y + O\big((1+|\beta|x)\sqrt{x}q\log(qx)\big)\end{aligned} \tag{23.20}$$

by partial summation. Since $|\beta| \leqslant 1/(qQ)$, taking $Q = \sqrt{x}(\log x)^3$ yields

$$\sum_{p\leqslant x} e(p\alpha) = \frac{\mu(q)}{\varphi(q)}\int_2^x \frac{e(\beta y)}{\log y}\mathrm{d}y + O\Big(\frac{x}{(\log x)^2} + \sqrt{x}q\log x\Big). \tag{23.21}$$

In particular, we see that if α is close to a rational number of denominator $q \in [(\log x)^2, \sqrt{x}/(\log x)^3]$, then there is significant cancellation among the numbers $e(\alpha p)$ with $p \leqslant x$, which makes $\sum_{p\leqslant x} e(\alpha p)$ smaller than $\pi(x)$.

The above calculation is a manifestation of an important principle stemming from the *Hardy-Littlewood circle method* that we will study in detail in Chapter 24: the Fourier transform

$$\sum_{n\leqslant x} c_n e(n\alpha) \tag{23.22}$$

of various interesting arithmetic sequences $(c_n)_{n\leqslant x}$ is big when α lies close to a rational number of small denominator, and it is small otherwise. The rough heuristic to explain this dichotomy is that when α is far from any fraction of small denominator, the sequence $(e(n\alpha))_{n\leqslant x}$ lacks any meaningful arithmetic structure, so that it cannot correlate with any "reasonably regular" sequence $(c_n)_{n\leqslant x}$.

A central problem in analytic number theory is to establish strong estimates for the exponential sum $\sum_{n\leqslant x} c_n e(n\alpha)$: an asymptotic formula when α is close to a fraction of small denominator, and a non-trivial upper bound otherwise. In particular, we would like to do so when c_n is the indicator function of the primes without appealing to the unproven Generalized Riemann Hypothesis.

Type I exponential sums

In view of the decomposition of Λ into type I and type II functions, the estimation of $\sum_{p\leqslant x} e(\alpha p)$ boils down to the estimation of $\sum_{n\leqslant x}(f*g)(n)e(\alpha n)$,

when $f * g$ is a function of type I or II. We begin by studying the first category of functions.

Let us begin by handling the simplest non-trivial type I function: the constant function 1. Arguing as in (10.12), we have

$$\Big|\sum_{n\leqslant x} e(\alpha n)\Big| = \Big|e(\alpha)\cdot\frac{1-e(\alpha\lfloor x\rfloor)}{1-e(\alpha)}\Big| \leqslant \frac{1}{2\|\alpha\|}, \tag{23.23}$$

where we recall that $\|\alpha\|$ denotes the distance of α from the nearest integer. We thus immediately see that, as long as $\|\alpha\| = o(1/x)$, the sum $\sum_{n\leqslant x} e(\alpha n)$ is small compared to the trivial bound

$$\Big|\sum_{n\leqslant x} e(\alpha n)\Big| \leqslant \sum_{n\leqslant x} 1 \leqslant x. \tag{23.24}$$

Using partial summation, we may easily pass from (23.23) and (23.24) to an estimate for the Fourier transform of the function $\log^v$, where v is any fixed positive real number. Indeed, we have

$$\begin{aligned}\sum_{n\leqslant x}(\log n)^v e(n\alpha) &= \int_{1^-}^{x}(\log t)^v \mathrm{d}\sum_{n\leqslant t} e(n\alpha)\\ &\ll (\log x)^v\cdot\min\{x, \|\alpha\|^{-1}\}\end{aligned} \tag{23.25}$$

uniformly for $x \geqslant 1$ and $v \geqslant 0$. Similar estimates are true if we replace $\log^v$ by a more general smooth function but we will not need them.

The above observations and the simplest version of Dirichlet's hyperbola method allow us to establish non-trivial estimates for general exponential sums of type I when α is close to a fraction a/q of large denominator (say, with $q \geqslant (\log x)^A$ for some large A). The notation $\|f\|_\infty$ in the statement of Theorem 23.5 below stands for the supremum norm of f. Finally, its proof features an important concept in the study of exponential sums: we say that a set of real numbers $\{\alpha_1, \dots, \alpha_r\}$ is *δ-spaced mod* 1 if

$$\|\alpha_i - \alpha_j\| \geqslant \delta \quad \text{whenever} \quad i \neq j. \tag{23.26}$$

Theorem 23.5. *Let $f : \mathbb{N} \to \mathbb{C}$ be supported on $[1, y]$, $v \geqslant 0$, $x \geqslant 2$, $\alpha \in \mathbb{R}$ and a/q be a reduced fraction with $|\alpha - a/q| \leqslant 1/q^2$. Then*

$$\sum_{n\leqslant x}(f * \log^v)(n)e(n\alpha) \ll \Big(y + \frac{x}{q} + q\Big)(\log x)^{v+1}\|f\|_\infty. \tag{23.27}$$

Proof. If $q = 1$, $q > x$ or $y > x$, we simply note that $|(f * \log^v)(n)| \leqslant \|f\|_\infty \tau(n)(\log n)^v$ and use Theorem 3.3. Assume now that $2 \leqslant q \leqslant x$ and

$y \leqslant x$. Opening the convolution and applying (23.25) yields

$$\sum_{n\leqslant x}(f*\log^v)(n)e(n\alpha) = \sum_{k\leqslant y} f(k) \sum_{\ell\leqslant x/k} (\log \ell)^v e(\ell\cdot k\alpha)$$

$$\ll (\log x)^v \|f\|_\infty \sum_{k\leqslant y} \min\big\{x/k, 1/\|k\alpha\|\big\}. \tag{23.28}$$

We cover the last sum by subsums of length $\tilde{q} := \lfloor q/2\rfloor$ defined by

$$S_m := \sum_{m\tilde{q}<k\leqslant(m+1)\tilde{q}} \min\big\{x/k, 1/\|k\alpha\|\big\}.$$

Since $(a,q)=1$, the numbers ka/q with $m\tilde{q} < k \leqslant (m+1)\tilde{q}$ are all distinct mod 1. Hence, $\|k_1a/q - k_2a/q\| \geqslant 1/q$ whenever $m\tilde{q} < k_1 < k_2 \leqslant (m+1)\tilde{q}$. On the other hand, if we write $\alpha = a/q + \beta$, then $|k_1\beta - k_2\beta| \leqslant \tilde{q}|\beta| \leqslant (q/2)/q^2 = 1/(2q)$. As a consequence, we find that the numbers $k\alpha$ with $m\tilde{q} < k \leqslant (m+1)\tilde{q}$ are $(2q)^{-1}$-spaced mod 1. We index them as $\alpha_1, \dots, \alpha_{\tilde{q}}$ in a way that $\|\alpha_1\| \leqslant \cdots \leqslant \|\alpha_{\tilde{q}}\|$. For each integer $j \in [1,\tilde{q}]$, the interval $(-\frac{j-1}{4q}, \frac{j-1}{4q})$ can contain at most $j-1$ of the reductions mod 1 of the numbers $\alpha_1, \dots, \alpha_{\tilde{q}}$. Hence, we must have that $\|\alpha_j\| \geqslant (j-1)/(4q)$ for $j = 1, \dots, \tilde{q}$.

When $m \geqslant 1$, the above discussion and the fact that $x/k < x/(m\tilde{q})$ whenever $k > m\tilde{q}$ yield the inequality

$$S_m \leqslant \frac{x}{m\tilde{q}} + \sum_{2\leqslant j\leqslant q/2} \frac{4q}{j-1} \ll \frac{x}{mq} + q\log q. \tag{23.29}$$

However, when $m = 0$, we cannot use the above argument as it currently stands because we do not have a good bound for the summand of S_0 corresponding to the integer k with $k\alpha = \alpha_1$. Note though that if $1 \leqslant k \leqslant q/2$, then $|k\beta| \leqslant (q/2)/q^2 = 1/(2q)$ and $\|ka/q\| \geqslant 1/q$. Therefore, $\|k\alpha\| \geqslant 1/(2q)$ for all $k \in \mathbb{Z} \cap [1, q/2]$. In particular, $\|\alpha_1\| \geqslant 1/(2q)$ when $m = 0$, and thus

$$S_0 \leqslant 2q + \sum_{2\leqslant j\leqslant q/2} \frac{4q}{j-1} \ll q\log q. \tag{23.30}$$

Combining (23.29) with (23.30), and noticing that there are $\leqslant y/\tilde{q} \ll y/q$ integers $m \in [1, y/\tilde{q}]$ allows us to estimate the expression in (23.28) and complete the proof of the theorem. □

Type II exponential sums

Let us now consider the exponential sum $\sum_{n\leqslant x}(f*g)(n)e(\alpha n)$ for a type II function $f*g$. For concreteness, we assume momentarily that $\mathrm{supp}(f) \subseteq$

$[1,y]$ and $\mathrm{supp}(g) \subseteq [1,z]$ with $y = x^\theta$ and $z = x^{1-\theta}$ for some $\theta \in (0,1)$. We then find that

$$\sum_{n\leqslant x}(f*g)(n)e(\alpha n) = \sum_{\substack{k\leqslant y,\, \ell\leqslant z\\ k\ell\leqslant x}}\sum f(k)g(\ell)e(\alpha k\ell). \tag{23.31}$$

The advantage of this formula is that it transforms the Fourier transform of $f*g$ into a double sum that we can interpret as *an average of many sums*. For instance, we may arrange the summation as

$$\sum_{n\leqslant x}(f*g)(n)e(\alpha n) = \sum_{k\leqslant y} f(k) \sum_{\ell\leqslant \min\{z,x/k\}} g(\ell)e(\alpha k\ell). \tag{23.32}$$

In practice, we do not know much about the function g, so that for a given k we cannot hope to do much better than the trivial upper bound

$$\Big|\sum_{\ell\leqslant \min\{z,x/k\}} g(\ell)e(\alpha k\ell)\Big| \leqslant \sum_{\ell\leqslant \min\{z,x/k\}} |g(\ell)|. \tag{23.33}$$

(Consider for instance the case when $g(\ell) = e(-\alpha\ell)$ and $k = 1$.) However, it turns out that (23.33) can be improved for *most* k, something that we can take advantage of since we are averaging over many values of k.

We begin by noticing that the sum in the left-hand side of (23.33) can be interpreted as the Hermitian inner product over $\mathbb{C}$ of the vectors

$$\vec{g} = (g(\ell))_{\ell=1}^d \quad \text{and} \quad \vec{v}_k = (1_{k\ell\leqslant x}\cdot e(-k\ell\alpha))_{\ell=1}^d,$$

where $d = \lfloor z \rfloor$. The key observation is that if $\alpha \approx a/q$ with large q, then the vectors $\vec{v}_k$ are approximately orthogonal to each other, so that the fixed vector $\vec{g}$ cannot correlate strongly with many of them. Consequently, we expect that the trivial bound (23.33) can be improved significantly for most values of k.

To see the claim that the vectors $\vec{v}_k$ are mutually quasi-orthogonal, note that relation (23.23) implies the estimate

$$\langle \vec{v}_{k_1}, \vec{v}_{k_2}\rangle = \sum_{\ell\leqslant \min\{z,x/k_1,x/k_2\}} e(-k_1\ell\alpha)\overline{e(-k_2\ell\alpha)} \ll \frac{1}{\|(k_2-k_1)\alpha\|}. \tag{23.34}$$

Generalizing the argument used to prove Theorem 23.5, we will show that if α is far from fractions of small denominator, the quantity $\|(k_2-k_1)\alpha\|$ is away from 0 for most pairs (k_1,k_2) with $k_1 \neq k_2$, so that $\langle \vec{v}_{k_1}, \vec{v}_{k_2}\rangle$ is small.

The above ideas will be vastly generalized in Chapter 25, where we study bounds for general bilinear sums $\sum_{m=1}^M \sum_{n=1}^N a_{m,n}x_m y_n$. We will prove there that there is some Δ that depends at most on the coefficients $a_{m,n}$ such that

$$\Big|\sum_{m=1}^M\sum_{n=1}^N a_{m,n}x_m y_n\Big| \leqslant \Delta\cdot\Big(\sum_{m=1}^M |x_m|^2\Big)^{1/2}\Big(\sum_{n=1}^N |y_n|^2\Big)^{1/2}.$$

For now, we use this circle of ideas to derive a strong bound for the Fourier transform of type II functions. The notation $\|f\|_2$ in the statement of Theorem 23.6 stands for the ℓ^2-norm of f, that is to say, $\|f\|_2^2 = \sum_{n\geqslant 1} |f(n)|^2$.

Theorem 23.6. *Let $f, g : \mathbb{N} \to \mathbb{C}$ be two arithmetic functions such that $\mathrm{supp}(f) \subseteq [1, y]$ and $\mathrm{supp}(g) \subseteq [1, z]$. In addition, consider $\alpha \in \mathbb{R}$ and a reduced fraction a/q such that $|\alpha - a/q| \leqslant 1/q^2$. For all $x \geqslant 1$, we have*

$$\sum_{n\leqslant x}(f*g)(n)e(\alpha n) \ll \Big(q + y + z + \frac{yz}{q}\Big)^{1/2}\sqrt{\log(2q)}\cdot \|f\|_2\|g\|_2.$$

In particular, if $yz \leqslant 2x$ and $|f|, |g| \leqslant 1$, so that $\|f\|_2 \leqslant \sqrt{y}$ and $\|g\|_2 \leqslant \sqrt{z}$, then

$$\sum_{n\leqslant x}(f*g)(n)e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + \frac{x}{\sqrt{y}} + \frac{x}{\sqrt{z}} + \sqrt{xq}\Big)\sqrt{\log(2q)}.$$

Proof. Let S be the sum we want to bound, which we arrange in the "dual"[3] form to (23.32)

$$S = \sum_{\ell\leqslant z} g(\ell) \sum_{k\leqslant y,\, k\ell\leqslant z} f(k)e(k\ell\alpha).$$

We use the Cauchy-Schwarz inequality to remove the unknown function g:

$$|S|^2 \leqslant \|g\|_2^2 \sum_{\ell\leqslant z}\Big|\sum_{k\leqslant y,\, k\ell\leqslant x} f(k)e(k\ell\alpha)\Big|^2.$$

As a result, the variable ℓ is now weighted with the smooth function 1. Opening the square via the identity $|z|^2 = z\bar{z}$ yields that

$$\begin{aligned}|S|^2 &\leqslant \|g\|_2^2 \sum_{\substack{\ell\leqslant z,\ k_1,k_2\leqslant y\\ k_1\ell,\, k_2\ell\leqslant x}}\sum\sum f(k_1)\bar{f}(k_2)e((k_1-k_2)\ell\alpha)\\ &= \|g\|_2^2 \sum_{k_1,\,k_2\leqslant y}\sum f(k_1)\bar{f}(k_2) \sum_{\ell\leqslant \min\{z, x/k_1, x/k_2\}} e(\ell\cdot(k_1-k_2)\alpha). \qquad (23.35)\end{aligned}$$

We bound the innermost sum of (23.35) using (23.34) to find that

$$|S|^2 \ll \|g\|_2^2 \sum_{k_1,\,k_2\leqslant y}\sum |f(k_1)f(k_2)|\cdot \min\Big\{z, \frac{1}{\|(k_2-k_1)\alpha\|}\Big\}.$$

To remove one of the unknown factors $f(k_j)$, we use the inequality $|zw| \leqslant (|z|^2 + |w|^2)/2$. This implies that

$$|S|^2 \ll \|g\|_2^2 \sum_{j\in\{1,2\}}\sum_{k_1,\,k_2\leqslant y}\sum |f(k_j)|^2 \min\Big\{z, \frac{1}{\|(k_2-k_1)\alpha\|}\Big\}.$$

[3]This terminology will be explained in Chapter 25.

The theorem will then follow if we can prove the following estimate:

$$\sum_{k\leqslant y}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}\ll\Big(q+y+z+\frac{yz}{q}\Big)\log(2q), \tag{23.36}$$

uniformly for all $\eta\in\mathbb{R}$. This will be demonstrated by adapting the argument of the proof of Theorem 23.5.

If $q=1$, (23.36) follows by majorizing all summands by z. Let us consider now the more interesting case when $q\geqslant 2$. We let $\tilde{q}=\lfloor q/2\rfloor$ and break the interval $[1,y]$ into subintervals of length $\tilde{q}$ to find that

$$\sum_{k\leqslant y}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}\leqslant\sum_{m=0}^{\lfloor y/\tilde{q}\rfloor}\sum_{k=m\tilde{q}+1}^{(m+1)\tilde{q}}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}. \tag{23.37}$$

Fix $m\in\mathbb{Z}_{\geqslant 0}$. Arguing as in the proof of Theorem 23.5, we find that the numbers $k\alpha+\eta$ are $(2q)^{-1}$-spaced mod 1 when $m\tilde{q}<k\leqslant(m+1)\tilde{q}$. Hence, a straightforward adaptation of the proof of (23.29) implies that

$$\sum_{m\tilde{q}<k\leqslant(m+1)\tilde{q}}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}\leqslant z+\sum_{2\leqslant j\leqslant q/2}\frac{4q}{j-1}\ll z+q\log q.$$

Inserting this bound into (23.37) completes the proof of (23.36), and hence of the theorem. □

Remark 23.7. Remarkably, the estimate for $\sum_{n\leqslant x}(f*g)e(\alpha n)$ supplied by Theorem 23.6 is essentially sharp. For simplicity, we consider only the case when $y\geqslant z$, since the other one is symmetric.

Indeed, let $x\geqslant yz\geqslant x/2$ and choose $f(k)$ to be the complex conjugate of $\sum_{\ell\leqslant z}g(\ell)e(\alpha k\ell)$. We then have

$$\sum_{n\leqslant x}(f*g)e(\alpha n)=\sum_{k\leqslant y}\Big|\sum_{\ell\leqslant z}g(\ell)e(\alpha k\ell)\Big|^2=\|f\|_2^2.$$

If we now let $\{g(\ell)\}_{\ell\leqslant z}$ be a sequence of independent random variables with $\mathbb{P}(g(\ell)=1)=\mathbb{P}(g(\ell)=-1)=1/2$, we find that

$$\mathbb{E}\Big[\sum_{k\leqslant y}\Big|\sum_{\ell\leqslant z}g(\ell)e(k\ell\alpha)\Big|^2\Big]=\sum_{k\leqslant y}\sum_{\ell\leqslant z}1\asymp x.$$

In particular, there must exist a choice of $g(\ell)$ such that $\sum_{n\leqslant x}(f*g)e(\alpha n)=\|f\|_2^2\gg x$. Since $\|g\|_2^2=\lfloor z\rfloor$, we infer that

$$\sum_{n\leqslant x}(f*g)e(\alpha n)=\|f\|_2^2\gg\sqrt{x}\|f\|_2\asymp\sqrt{y}\cdot\|f\|_2\|g\|_2.$$

By swapping the roles of f and g, we can also find choices of them such that $|\sum_{n\leqslant x}(f*g)e(\alpha n)|\gg\sqrt{z}\cdot\|f\|_2\|g\|_2$.

Finally, let us consider the case when $\alpha = a/q$, $f(k) = e(-ak/q)$ for $k \leqslant y$, and $g(\ell) = 1_{\ell \equiv 1 \,(\mathrm{mod}\, q)}$ for $\ell \leqslant z$. We then have

$$\sum_{n \leqslant x} (f * g) e(\alpha n) = \sum_{\substack{k \leqslant y,\, \ell \leqslant z \\ \ell \equiv 1 \,(\mathrm{mod}\, q)}} \sum 1 \asymp y \cdot (z/q + 1) \asymp \sqrt{yz/q + y} \cdot \|f\|_2 \|g\|_2.$$

To conclude, a general estimate for $\sum_{n \leqslant x} (f*g) e(\alpha n)$ can never be better than $\max\{y, z, yz/q\}^{1/2} \|f\|_2 \|g\|_2$, and Theorem 23.6 comes remarkably close to this bound. □

The additive Fourier transform of the primes: Encore

We shall now apply the methods we have developed to establish Vinogradov's famous estimate.

Theorem 23.8. *Let $\alpha \in \mathbb{R}$ and consider a reduced fraction a/q such that $|\alpha - a/q| \leqslant 1/q^2$. For all $x \geqslant 2$, we have*

$$\sum_{n \leqslant x} \Lambda(n) e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + x^{4/5} + \sqrt{xq} \Big) (\log x)^{5/2}.$$

Proof. We may assume that $q \leqslant x$; otherwise, the theorem follows by bounding all summands by $\log x$.

Let us decompose Λ using Vaughan's identity. First, we deal with $\Lambda^\sharp$. We apply Theorem 23.5 twice, once to the convolution $\mu_{\leqslant V} * \log$ (so $v = 1$ and $f = \mu_{\leqslant V}$ here, with $y = V$ and $\|f\|_\infty = 1$) and once to $(\mu_{\leqslant V} * \Lambda_{\leqslant U}) * 1$ (so $v = 0$ and $f = \mu_{\leqslant V} * \Lambda_{\leqslant U}$ here, with $y = UV$ and $|f| \leqslant 1 * \Lambda = \log$, whence $\|f\|_\infty \leqslant \log(UV)$). We thus conclude that

$$\sum_{n \leqslant x} \Lambda^\sharp(n) e(n\alpha) \ll \Big(UV + \frac{x}{q} + q \Big) \log^2(xUV). \tag{23.38}$$

Next, we deal with $\Lambda^\flat$. We rewrite this function using (23.12) and apply Theorem 23.6 to each summand $f_j * g_j$ of that identity. Since $q \leqslant x$, $\|f_j\|_2^2 \leqslant 2^j \log^2 x$ and $\|g_j\|_2^2 \leqslant x/2^{j-1}$, we find that

$$\sum_{n \leqslant x} \Lambda^\flat(n) e(n\alpha) \ll \sum_{U < 2^j \leqslant 2x/V} \Big(\frac{x}{\sqrt{q}} + \sqrt{2^j x} + \frac{x}{2^{j/2}} + \sqrt{xq} \Big) (\log x)^{3/2}.$$

We note that $\sqrt{2^j x} \ll x/\sqrt{V}$ and $x/2^{j/2} \ll x/\sqrt{U}$. Applying these bounds to each of the $O(\log x)$ choices of j yields the estimate

$$\sum_{n \leqslant x} \Lambda^\flat(n) e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + \frac{x}{\sqrt{U}} + \frac{x}{\sqrt{V}} + \sqrt{xq} \Big) (\log x)^{5/2}. \tag{23.39}$$

Since we also have that $|\sum_{n\leqslant x}\Lambda_{\leqslant U}(n)e(n\alpha)|\leqslant\sum_{n\leqslant U}\Lambda(n)\ll U$ and $q\leqslant\sqrt{xq}$ by our assumption that $q\leqslant x$, Vaughan's identity in combination with (23.38) and (23.39) implies that

$$\sum_{n\leqslant x}\Lambda(n)e(n\alpha)\ll\big(UV+x/\sqrt{q}+x/\sqrt{U}+x/\sqrt{V}+\sqrt{xq}\big)(\log x)^{5/2}.$$

Taking $U=V=x^{2/5}$ to optimize the above bound completes the proof. □

Theorem 23.8 confirms the prediction we made using the Generalized Riemann Hypothesis that the exponential sum $\sum_{n\leqslant x}\Lambda(n)e(\alpha n)$ can only be large when α is close to a rational number with small denominator. Indeed, if $|\alpha-a/q|\leqslant 1/q^2$ with $(\log x)^A\leqslant q\leqslant x/(\log x)^A$, we find that

$$\sum_{n\leqslant x}\Lambda(n)e(\alpha n)\ll_A x/(\log x)^{(A-5)/2}. \tag{23.40}$$

We will demonstrate the utility of this key estimate in the next chapter.

Conclusion

Vinogradov's method allows us to deal with very general sums of the form

$$\sum_{n\leqslant x}a_n\Lambda(n), \tag{23.41}$$

where $(a_n)_{n=1}^{\infty}$ is some interesting sequence. To estimate (23.41), we first use various combinatorial ideas such as convolution identities and Buchstab iterations to obtain an appropriate decomposition of Λ of the form (23.5). We then handle quasi-smooth sums, namely sums of the form $\sum_{n\leqslant x}a_n(f*g)(n)$ with f "small" and g smooth, using a mix of tools such as the summation formulas of Poisson and of Euler-Maclaurin, L-functions, sieves and estimates for exponential sums (e.g. the Pólya-Vinogradov inequality and other more advanced results beyond the scope of this book). Finally, we estimate bilinear sums, namely sums of the form $\sum_{n\leqslant x}a_n(f*g)(n)$ with f and g both supported on large integers, by employing methods arising from the theory of bilinear forms that we will fully develop in Chapter 25 (with the Cauchy-Schwarz inequality playing a central role), coupled with various exponential sum estimates. We thus see that this approach to the distribution of primes utilizes the full toolset we have at our disposal.

Exercises

Exercise 23.1. Consider $\alpha\in\mathbb{R}$ and a reduced fraction a/q such that $|\alpha-a/q|\leqslant 1/q^2$. Prove that

$$\sum_{n\leqslant x}\tau(n)e(\alpha n)\ll(\sqrt{x}+q+x/q)\log x\qquad(x\geqslant 2).$$

Exercise 23.2. Let $v \geqslant 0$, let f be an arithmetic function supported in $[1, y]$, and χ be a non-principal Dirichlet character mod q. For $x \geqslant 2$, prove that

$$\sum_{n \leqslant x} (f * \log^v)(n)\chi(n) \ll \sqrt{q}(\log q)(\log x)^v \sum_{k \leqslant y} |f(k)|.$$

Exercise 23.3. Let $r, s > 1$ be such that $1/r + 1/s = 1$. Assuming the set-up of Theorem 23.5, prove that

$$\sum_{n \leqslant x} (f * \log^v)(n)e(n\alpha) \ll_r \big(y^{1/r}q^{1/s} + q + x/q\big)(\log x)^v \|f\|_s,$$

where $\|f\|_s = (\sum_{k=1}^{\infty} |f(k)|^s)^{1/s}$.

Exercise 23.4.*

(a) For any $U, V \geqslant 1$, prove that

$$\mu = -\mu_{\leqslant U} * \mu_{\leqslant V} * 1 \ + \ \mu_{>U} * \mu_{>V} * 1 \ + \ \mu_{\leqslant U} \ + \ \mu_{\leqslant V}.$$

(b) Let $\alpha \in \mathbb{R}$, and let a/q be a reduced fraction with $|\alpha - a/q| \leqslant 1/q^2$. For every fixed $\varepsilon > 0$, show that

$$\sum_{n \leqslant x} \mu(n)e(n\alpha) \ll_\varepsilon \big(x/\sqrt{q} + \sqrt{xq}\big)(\log x)^3 + x^{4/5+\varepsilon} \qquad (x \geqslant 3).$$

[*Hint:* Select $U = V = \min\{x^{2/5}, q, x/q\}$ in part (a).]

(c) (Davenport) Fix $A \geqslant 1$. Prove that

$$\sum_{n \leqslant x} \mu(n)e(n\alpha) \ll_A x/(\log x)^A \quad (x \geqslant 2, \ \alpha \in \mathbb{R}).$$

Exercise 23.5 (Heath-Brown's identity)**.** Let $k \in \mathbb{N}$, $x \geqslant 1$ and $V \geqslant x^{1/k}$. For $n \leqslant x$, show that

$$\Lambda(n) = \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} (\log * \underbrace{1 * \cdots * 1}_{j-1 \text{ times}} * \underbrace{\mu_{\leqslant V} * \cdots * \mu_{\leqslant V}}_{j \text{ times}})(n).$$

[*Hint:* Let $f = \mu_{\leqslant V} * 1$ and $g = \mu_{>V} * 1$. On the one hand, we have $\Lambda * \underbrace{g * \cdots * g}_{k \text{ times}} = 0$ on $\mathbb{N}_{\leqslant x}$. On the other hand, $g = \delta - f$ with $\delta(n) = 1_{n=1}$.]