ANALYTIC NUMBER THEORY

TOM SANDERS[1]

Contents

Overview. These are notes for a first course in analytic number theory. The main aim is to give a proof of the prime number theorem, together with background on complex analysis, the Riemann ζ-function, and Fourier analysis.

A particular feature of the notes is the use of smooth cutoff functions. This has advantages we shall discuss through the course, but it also means it differs from some traditional presentations of the material.

Note to Oxford students Michaelmas 2026: Since AY2016/17 this course has been taught by Ben Green and James Maynard. James takes a slightly different approach to Ben, and we are following Ben's approach this year. This means that questions on past exam papers labelled Trinity 2019–2024 (when James taught the course) may have

[1] The original version of the notes and exercise sheets for this course were created by Ben Green. My thanks to him for providing them. Only minor modifications have been made to them.

parts that are not quite accessible using only these notes, but questions on exam papers labelled Trinity 2017, 2018, 2025, and 2026 are relevant to this version of the course. Prior to Michaelmas 2016 the course was quite different.

Notation. A certain amount of notation will be introduced as we go along.

We write $\log x$ for the natural logarithm of x. $\log^\alpha x$ is the same as $(\log x)^\alpha$.

We write $\lfloor t \rfloor$ for the floor of t, that is the greatest integer less than or equal to t.

Throughout the course we will be using the following *asymptotic notation*: $f(x) \ll g(x)$, $f(x) = O(g(x))$, and $g(x) \gg f(x)$ for $x \in D$ all mean that there is a constant $C > 0$ such that $|f(x)| \leqslant Cg(x)$ for all $x \in D$. Crucially, when we say C is constant we mean it does not vary on D. For example, $x + 1 \ll x$ for $x \geqslant 1$, because $|x + 1| \leqslant 2x$ for all $x \geqslant 1$. The constant C may be different in different instances of the notation, and the set D will often be implied.

Sometimes we will want to emphasise other variables k, l, m, $\ldots$ on which C may depend, in which case we write $f(x) \ll_{k,l,m,\ldots} g(x)$, $f(x) = O_{k,l,m,\ldots}(g(x))$, and $g(x) \gg_{k,l,m,\ldots} f(x)$. For example, $|k|x \ll_k x$, and $k + l = O_{k,l}(1)$.

We write $f(x) = o(g(x))$ as $x \to \infty$ to mean that for all $\varepsilon > 0$, there is M, which may depend on ε, such that if $x > M$ then $|f(x)| \leqslant \varepsilon g(x)$. For example, $\frac{1}{\log x} = o(1)$. As with D above, "$x \to \infty$" will often be implied, and if we want to emphasise some other variables k, l, m, $\ldots$ on which M may depend, then we write $f(x) = o_{k,l,m,\ldots}(g(x))$. For example, $\frac{|k|}{\log x} = o_k(1)$.

In understanding analytic number theory, it is important to develop a robust intuitive feeling for the rough size of certain quantities. For example, if X is large, $(\log X)^{10}$ is much smaller than $X^{1/10}$. The quantity $e^{\sqrt{\log X}}$ (which will appear again later) lies in between the two. In the notation above we can write this as $(\log X)^{10} = o(e^{\sqrt{\log X}})$ and $e^{\sqrt{\log X}} = o(X^{1/10})$, both as $X \to \infty$.

We write $f(x) \sim g(x)$, which we read as "$f(x)$ is asymptotic to $g(x)$", to mean that $f(x) = (1 + o(1))g(x)$.

There are a number of further examples on Exercise Sheet 0 and Exercise Sheet 1.

Further reading. These notes are self-contained, but there are some books covering the material: Davenport's book [Dav00] is the classic reference, and Koukoulopoulos' book [Kou19] is a more recent one. The book [IK04] of Iwaniec and Kowalski is a modern classic containing significantly more material than this course, presented at a high level.

1. Basic facts about the primes

1.1. **Euclid's proof.** This course is largely about the *prime numbers* 2, 3, 5, 7, ...

Theorem 1.1. *There are infinitely many primes.*

Proof. Suppose not, and that $p_1, \dots, p_N$ is a complete list of the primes. Consider the number $M = p_1 \cdots p_N + 1$. It must have a prime factor q. However, M is manifestly not divisible by any of $p_1, \dots, p_N$, and so $q \notin \{p_1, \dots, p_N\}$. This is a contradiction. $\square$

It is possible to ask more refined questions: Does the sequence of prime numbers grow extremely rapidly (like the powers of two 1, 2, 2^2, 2^3, ...), or slowly like the odd numbers 1, 3, 5, 7, ..., or somewhere in between? This is the question that will occupy us in this course.

1.2. **Elementary bounds.** If $X > 1$ is a real number then we write $\pi(X)$ for the number of primes less than or equal to X. "Elementary methods" suffice to get the correct order of magnitude for $\pi(X)$.

Theorem 1.2. *There are constants $0 < c_1 < 1 < c_2$ such that, for all sufficiently large X,*

$$c_1 \frac{X}{\log X} \leqslant \pi(X) \leqslant c_2 \frac{X}{\log X}.$$

Remarks. In asymptotic notation, $\frac{X}{\log X} \ll \pi(X) \ll \frac{X}{\log X}$, and $\pi(X) = O(\frac{X}{\log X})$.

The lower bound immediately implies the infinitude of primes. The upper bound implies, in particular, that the density of the primes up to X, that is to say $\pi(X)/X$, tends to zero as $X \to \infty$.

Proof. We begin with the lower bound. We look at the prime factorisation of $\binom{2n}{n}$, which we write as

$$\binom{2n}{n} = \prod_{p \leqslant 2n} p^{e_p(n)}.$$

Now recall (or look up) Legendre's formula, that is to say the fact that the power of p dividing $m!$ is $\sum_{i=1}^{\infty} \lfloor m/p^i \rfloor$, this sum being composed of $\lfloor m/p \rfloor$ multiples of p, $\lfloor m/p^2 \rfloor$ multiples of p^2, and so on. (The sum is actually finite, since $\lfloor m/p^i \rfloor = 0$ when $p^i > m$.)

It follows from Legendre's formula and the definition of binomial coefficients that

$$e_p(n) = \sum_{i=1}^{\infty} \Big\lfloor \frac{2n}{p^i} \Big\rfloor - 2\Big\lfloor \frac{n}{p^i} \Big\rfloor. \tag{1.1}$$

Now each term in (1.1) is at most 1, since $\lfloor 2x \rfloor - 2\lfloor x \rfloor \in \{0, 1\}$ for all real x. Moreover, the terms can only be non-zero for $i \leqslant \log(2n)/\log p$. It follows that

$$p^{e_p(n)} \leqslant p^{\log(2n)/\log p} = 2n.$$

Taking products over p (noting that all primes dividing $\binom{2n}{n}$ are $\leqslant 2n$) we obtain

$$\binom{2n}{n} \leqslant (2n)^{\pi(2n)}. \tag{1.2}$$

Now $\binom{2n}{n}$ is almost as big as 4^n; note that $\sum_{r=0}^{2n} \binom{2n}{r} = 4^n$ and, in this sum, the middle term $\binom{2n}{n}$ is the largest one. It follows from this observation and (1.2) that, for $n \geqslant 3$

$$2^n < \frac{4^n}{2n+1} \leqslant \binom{2n}{n} \leqslant (2n)^{\pi(2n)}.$$

Taking logs gives

$$\pi(2n) \geqslant \frac{\log 2}{2} \cdot \frac{2n}{\log(2n)}.$$

This is the lower bound in Theorem 1.2 in the case $X = 2n$ an even integer (with $c_1 = \frac{1}{2}\log 2$). The general case follows easily by considering the greatest even integer $\leqslant X$, and reducing c_1 slightly. We leave the details as an exercise.

We turn now to the upper bound in Theorem 1.2. We again consider the binomial coefficient $\binom{2n}{n}$, noting now that if n is an integer and if

$2n \geqslant p > n$ then p divides $\binom{2n}{n}$ precisely once. Therefore

$$n^{\pi(2n)-\pi(n)} \leqslant \prod_{n<p\leqslant 2n} p \leqslant \binom{2n}{n} \leqslant 4^n,$$

and so, taking logs,

$$\pi(2n) - \pi(n) \leqslant \frac{n \log 4}{\log n}.$$

Applying this with $n = 2, 2^2, 2^3 \ldots, 2^{k-1}$ and summing gives

$$\pi(2^k) \leqslant 2^k\Big(\frac{1}{k-1} + \frac{1}{2(k-2)} + \frac{1}{2^2(k-3)} + \cdots + \frac{1}{2^{k-2}}\Big).$$

We claim that the bracketed expression is $O(1/k)$. There are many ways to see this. For instance, the ratio between the terms $1/2^{i-1}(k-i)$ and $1/2^i(k-i-1)$ is $2(k-i-1)/(k-i)$, which is $\geqslant 4/3$ for $i \leqslant k-3$, so

$$\sum_{i=2}^{k-1} \frac{1}{2^{i-1}(k-i)} \leqslant \frac{1}{k-1}\Big(1 + \frac{3}{4} + (\frac{3}{4})^2 + \ldots\Big) + \frac{1}{2^{k-2}} \ll \frac{1}{k}.$$

Therefore $\pi(2^k) \ll 2^k/k$, which we write as $\pi(2^k) \leqslant C2^k/k$ for some C to make the next calculation clearer.

Finally, let 2^k be the smallest power of two which is greater than or equal to X. Then

$$\pi(X) \leqslant \pi(2^k) \leqslant \frac{C2^k}{k} \leqslant \frac{2CX}{\log_2 X} \leqslant c_2\frac{X}{\log X},$$

where $c_2 = 2C \log 2$. □

Remark. Euclid's proof of the infinitude of primes can be modified to give an explicit lower bound on $\pi(X)$, but it is very weak: we leave it as an exercise to show that $\pi(X) \gg \log\log X$ by this method.

1.3. **The prime number theorem.** The main aim of this course will be to prove the *prime number theorem*, which states that c_1 and c_2 in Theorem 1.2 can be taken arbitrarily close to 1.

Theorem 1.3 (Prime number theorem)**.** *Suppose that* $0 < c_1 < 1 < c_2$. *Then, if* X *is sufficiently large, we have*

$$c_1\frac{X}{\log X} \leqslant \pi(X) \leqslant c_2\frac{X}{\log X}.$$

Equivalently, $\pi(X) = (1 + o(1))\frac{X}{\log X}$*, or* $\pi(X) \sim \frac{X}{\log X}$.

The proof of this will involve tools from most of the course. The formal argument starts on p. 59.

We now give a reformulation of the prime number theorem using a weight function, which turns out to be much more natural than simply counting the primes for reasons we will see later.

The weight function in question is the *von Mangoldt function*. It is defined by

$$\Lambda(n) := \begin{cases} \log p & \text{if } n = p^k,\ p \text{ a prime;} \\ 0 & \text{otherwise.} \end{cases}$$

One should think of Λ as being roughly a function that assigns weight $\log p$ to the prime p; the contribution of the genuine prime powers p^k, $k \geqslant 2$, is negligible.

We define

$$\psi(X) := \sum_{n \leqslant X} \Lambda(n). \tag{1.3}$$

Proposition 1.4. *The prime number theorem is equivalent to the statement that $\psi(X) \sim X$, that is to say that the average value of $\Lambda(n)$ over $n \leqslant X$ is asymptotically 1.*

Proof. Write

$$a(X) := \frac{\pi(X)}{X/\log X}, \qquad b(X) := \psi(X)/X.$$

Thus the task is to show that $a(X) \to 1$ if and only if $b(X) \to 1$.

We first obtain an upper bound for $b(X)$ in terms of $a(X)$. For a given prime p, the maximum k for which $p^k \leqslant X$ is $k = \lfloor \log X/\log p \rfloor$. The contribution of this prime p to $\psi(X)$, $\sum_{k : p^k \leqslant X} \log p$, is therefore at most

$$\log p \left\lfloor \frac{\log X}{\log p} \right\rfloor \leqslant \log X.$$

Only those primes p with $p \leqslant X$ contribute at all, and so we have the inequality

$$\psi(X) \leqslant \pi(X) \log X,$$

and thus

$$b(X) \leqslant a(X). \tag{1.4}$$

Now we obtain a bound in the other direction. Assume throughout this argument that X is sufficiently large. Now

$$a(X) - b(X) \leqslant \frac{1}{X} \sum_{p \leqslant X} (\log X - \log p), \tag{1.5}$$

the contribution of the proper prime powers p^k, $k \geqslant 2$, to $-b(X)$ being negative. We split the sum over p into two ranges $p \leqslant X(\log X)^{-2}$ and $X(\log X)^{-2} < p \leqslant X$ (there is a certain amount of flexibility in these choices). We bound the contribution from the first range using a trivial bound and then the lower bound in Theorem 1.2, obtaining

$$\frac{1}{X} \sum_{p \leqslant X(\log X)^{-2}} (\log X - \log p) < \frac{1}{\log X} \leqslant \frac{1}{c_1 \log X} a(X). \tag{1.6}$$

On the second range, we have

$$\log X - \log p \leqslant \log\big((\log X)^2\big) \leqslant 2 \log\log X.$$

Thus

$$\begin{aligned} &\frac{1}{X} \sum_{X(\log X)^{-2} < p \leqslant X} (\log X - \log p) \\ &\qquad \leqslant \frac{2\log\log X}{X} \#\{p : X(\log X)^{-2} < p \leqslant X\} \\ &\qquad \leqslant \frac{2\log\log X}{\log X} a(X). \end{aligned}$$

Combining this with (1.5) and (1.6) gives

$$a(X) - b(X) \leqslant \Big(\frac{1}{c_1 \log X} + \frac{2\log\log X}{\log X}\Big) a(X) = o(1) \cdot a(X),$$

which with (1.4) gives the result. □

To conclude this chapter, we record an immediate consequence of (1.4) and Theorem 1.2:

Proposition 1.5. *We have the bound*

$$\sum_{n \leqslant X} \Lambda(n) = O(X).$$

2. Arithmetic functions

An *arithmetic function* is simply a function f from $\mathbf{N}$ to $\mathbf{C}$. The most important arithmetic function in this course is the von Mangoldt function Λ, introduced in the last chapter. However, this is far from the only interesting arithmetic function. Here are some other commonly occurring arithmetic functions:

- The δ-function, defined by $\delta(1) = 1$ and $\delta(n) = 0$ for all $n > 1$;
- The constant function 1;
- The Möbius function μ is defined by $\mu(n) = (-1)^k$ if $n = p_1 \cdots p_k$ for distinct primes $p_1, \ldots, p_k$, and $\mu(n) = 0$ otherwise;
- Euler's ϕ-function $\phi(n)$ is defined to be the number of integers $x \leqslant n$ which are coprime to n, which is the same thing as the order of the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^\times$, the group of invertible elements in $\mathbf{Z}/n\mathbf{Z}$;
- The divisor function $\tau(n)$ (sometimes written $d(n)$) is defined to be the number of positive integer divisors of n, including 1 and n itself;
- The sum-of-divisors function $\sigma(n)$ is defined to be $\sum_{d|n} d$.

Dirichlet convolution. If $f, g : \mathbf{N} \to \mathbf{C}$ are two arithmetic functions then we define

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

This operation is commutative and associative, and has the δ-function as an identity.

Multiplicativity. We say that an arithmetic function f is multiplicative if $f(1) = 1$ and $f(mn) = f(m)f(n)$ whenever m, n are coprime. Note we do not require $f(mn) = f(m)f(n)$ for *all* m, n. For example, the Möbius function μ is multiplicative, as may be checked immediately from the definition.

2.1. **Möbius inversion.**

Some of the material of this section is not really necessary for the main development of the course, and in particular for proving the prime number theorem. However, it helps us place Λ in a more general context.

Proposition 2.1 (Möbius inversion)**.** *Let $f, g : \mathbf{N} \to \mathbf{C}$ be two arithmetic functions. Then $g = f \star 1$ if and only if $f = g \star \mu$.*

Proof. Note that $\mu \star 1 = \delta$, where $\delta(n) = 1$ if $n = 1$ and 0 otherwise. This is a routine check: if $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ with $k \geqslant 1$ (and the p_i primes) then

$$\begin{aligned}\sum_{d|n} \mu(d) &= \sum_{\varepsilon_i \in \{0,1\}} \mu(p_1^{\varepsilon_1} \dots p_k^{\varepsilon_k}) \\ &= \sum_{\varepsilon_i \in \{0,1\}} \mu(p_1^{\varepsilon_1}) \cdots \mu(p_k^{\varepsilon_k}) = (1-1)\cdots(1-1) = 0.\end{aligned}$$

Here we used the multiplicativity of μ. Now if $g = f \star 1$ then,

$$g \star \mu = (f \star 1) \star \mu = f \star (1 \star \mu) = f \star \delta = f.$$

Conversely if $f = g \star \mu$ then similarly

$$f \star 1 = (g \star \mu) \star 1 = g \star (\mu \star 1) = g \star \delta = g.$$

This concludes the proof. □

This leads to an important link between the von Mangoldt function and the Möbius function.

Lemma 2.2. *We have $\Lambda = \mu \star \log$, that is to say*

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

Proof. Considering the prime factorization of n, we see that

$$\Lambda \star 1(n) = \sum_{d|n} \Lambda(d) = \log n.$$

Indeed, the only divisors of n which make a nontrivial contribution are the prime powers, and each contributes $\log p$ for a total contribution of $k \log p$ where p^k is the exact power of p dividing n.

The result then follows from Möbius inversion. □

3. Introducing the Riemann ζ function

3.1. **Dirichlet series.** An important tool for working with arithmetic functions – particularly when they arise from considerations that are somehow multiplicative – is that of *Dirichlet series.*

Let $f : \mathbf{N} \to \mathbf{C}$ be an arithmetic function. Then the *Dirichlet series* of f is

$$D_f(s) := \sum_n f(n)n^{-s}.$$

In other words,

$$D_f(s) := \lim_{N\to\infty} D_f^{(N)}(s) \text{ where } D_f^{(N)}(s) := \sum_{n=1}^{N} f(n)n^{-s}.$$

At the moment this limit might not exist for any s, for example if f grows extremely quickly. In practice, f will have reasonable growth. Commonly, for example, $|f(n)| = O_\varepsilon(n^\varepsilon)$ for all $\varepsilon > 0$. This is the case when f is the constant function 1, the Möbius function μ, the von Mangoldt function Λ, or the divisor function τ (by contrast to the first three, this last one takes a little extra work and is on Exercise Sheet 1). In this case we have the following:

Proposition 3.1. *Suppose that $f : \mathbf{N} \to \mathbf{C}$ has $|f(n)| = O_\varepsilon(n^\varepsilon)$ for all $\varepsilon > 0$. Then $D_f(s)$ is holomorphic whenever $\operatorname{Re} s > 1$, and*

$$D_f'(s) = -\sum_{n=1}^{\infty} \frac{f(n)\log n}{n^s} \text{ whenever } \operatorname{Re} s > 1. \tag{3.1}$$

Proof. Suppose that $\operatorname{Re} s > 1$ and take $\varepsilon = \frac{1}{2}(\operatorname{Re} s - 1)$. We have that

$$\left|\sum_{n=N+1}^{\infty} f(n)n^{-s}\right| \leqslant \sum_{n=N+1}^{\infty} \frac{|f(n)|}{n^{\operatorname{Re} s}} = O_s\left(\sum_{n=N+1}^{\infty} \frac{1}{n^{1+\varepsilon}}\right) = O_s\left(\frac{1}{N^\varepsilon}\right).$$

It follows that $D_f(s)$ converges absolutely for $\operatorname{Re} s > 1$.

Write $g(n) := -f(n)\log n$ so that $g(n) = O_\varepsilon(n^\varepsilon)$ for all $\varepsilon > 0$, and hence $D_g(s)$ exists for $\operatorname{Re} s > 1$. Recall from Lemma A.2 that

$$|n^{-h} - 1 + h\log n| \leqslant \frac{1}{2}|h|^2(\log n)^2 \max\{1, n^{-\operatorname{Re} h}\} \text{ for } h \in \mathbf{C}, n \in \mathbf{N}.$$

Suppose that $\operatorname{Re} s > 1$ and $|h| < \frac{1}{4}(\operatorname{Re} s - 1)$. Then

$$\begin{aligned}\left|D_f^{(N)}(s+h) - D_f^{(N)}(s) - hD_g^{(N)}(s)\right| \\ \leqslant \frac{1}{2}|h|^2 \sum_{n=1}^{N} \frac{|f(n)|(\log n)^2 \max\{1, n^{-\operatorname{Re} h}\}}{n^{\operatorname{Re} s}}.\end{aligned}$$

Let $\varepsilon = \frac{1}{4}(\operatorname{Re} s - 1)$, so that $f(n) = O_s(n^\varepsilon)$, $(\log n)^2 = O_s(n^\varepsilon)$, and $\max\{1, n^{-\operatorname{Re} h}\} = O_s(n^\varepsilon)$. Hence the right hand side above is

$$O_s\left(|h|^2 \sum_{n=1}^{N} n^{-1-\varepsilon}\right) = O_s(|h|^2).$$

Taking the limit as $N \to \infty$ on the left gives us

$$|D_f(s+h) - D_f(s) - hD_g(s)| = O_s(|h|^2),$$

and then dividing by h and letting $h \to 0$ we have that $D_f(s)$ is complex differentiable at s with derivative $D_g(s)$ for $\operatorname{Re} s > 1$. Since $\operatorname{Re} s > 1$ is an open set, D_f is holomorphic there. $\square$

3.2. **The ζ function.** Perhaps the most basic arithmetic function is the constant function $f = 1$. The Dirichlet series of this function is called the *Riemann ζ-function* and denoted

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}.$$

As just noted, the series converges and defines a holomorphic function in the domain $\operatorname{Re} s > 1$.

Informally unique factorization into primes convinces us of the identity

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots = \Big(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \dots\Big)\Big(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \dots\Big)\cdots,$$

where there is one product for each prime $p = 2, 3, 5, \dots$. Summing each of the geometric series suggests then that

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}. \tag{3.2}$$

This is called the *Euler product for ζ* and is valid for $\operatorname{Re} s > 1$. Let us pause to justify this formally. Let s, $\operatorname{Re} s > 1$, be fixed. Suppose $\operatorname{Re} s = 1 + \delta$. Fix an N, and suppose that $m > \log_2 N$. Then the product

$$P(m, N) := \prod_{p \leqslant N} \Big(\sum_{j=0}^{m} p^{-js}\Big)$$

may be expanded, and it equals $1 + 2^{-s} + \cdots + N^{-s}$ plus a number of further, distinct, terms n^{-s} with $n > N$. Thus

$$|P(m, N) - \zeta(s)| \leqslant \sum_{n=N+1}^{\infty} |n^{-s}| \ll N^{-\delta}$$

uniformly in m. Letting $m \to \infty$, we thus obtain

$$\Big|\prod_{p \leqslant N}(1-p^{-s})^{-1} - \zeta(s)\Big| \ll N^{-\delta}.$$

Finally, letting $N \to \infty$, we confirm the equality (3.2).

An amusing consequence of the above is yet another proof (our third) that there are infinitely many primes. Suppose not and the primes are $p_1, \dots, p_k$. Then by (3.2)

$$\zeta(1+\varepsilon) = \prod_{i=1}^{k}(1-p_i^{-(1+\varepsilon)})^{-1} \to \prod_{i=1}^{k}(1-p_i^{-1})^{-1}$$

as $\varepsilon \to 0$. However this is not the case, since

$$\zeta(1+\varepsilon) \geqslant \sum_{n=1}^{\lfloor \varepsilon^{-1}\rfloor} n^{-(1+\varepsilon)} \geqslant e^{-1}\sum_{n=1}^{\lfloor \varepsilon^{-1}\rfloor} n^{-1} \to \infty$$

as $\varepsilon \to 0$.

It turns out that the Dirichlet series of many of the basic arithmetic functions can be expressed in terms of ζ. In the following proposition we detail the most basic such relations; others may be found on the Exercise Sheet 2.

Proposition 3.2. *Suppose that f, g are arithmetic functions such that $|f(n)| = n^{o(1)}$ and $|g(n)| = n^{o(1)}$, and $s \in \mathbf{C}$ has $\operatorname{Re} s > 1$. Then*

(i) $|f \star g(n)| = n^{o(1)}$ *and* $D_f(s)D_g(s) = D_{f\star g}(s)$*;*
(ii) $D_\mu(s) = 1/\zeta(s)$*;*
(iii) $D_\Lambda(s) = -\zeta'(s)/\zeta(s)$*.*

Proof. For (i), we have

$$|f \star g(n)| \leqslant \sum_{d \mid n} |f(d)||g(n/d)| \leqslant |\tau(n)|n^{o(1)} = n^{o(1)},$$

where we use the divisor bound $\tau(n) = n^{o(1)}$ on Exercise Sheet 1. Hence $D_{f\star g}$ exists. Moreover,

$$|D^{(N^2)}_{f\star g}(s) - D^{(N)}_f(s)D^{(N)}_g(s)| = \left|\sum_{\substack{ab\leqslant N^2 \text{ and} \\ (a>N \text{ or } b>N)}} \frac{f(a)g(b)}{(ab)^s}\right|$$
$$\leqslant \sum_{a>N} \frac{|f(a)|}{a^{\operatorname{Re} s}} D_{|g|}(\operatorname{Re} s)$$
$$+ D_{|f|}(\operatorname{Re} s) \sum_{b>N} \frac{|g(b)|}{b^{\operatorname{Re} s}}.$$

Now $D_f(s)$ and $D_g(s)$ converge absolutely hence this last expression tends to 0, but $D^{(N^2)}_{f*g}(s)$ converges to $D_{f*g}(s)$, and $D^{(N)}_f(s)$ and $D^{(N)}_g(s)$ converge to $D_f(s)$ and $D_g(s)$ respectively. The identity is proved.

For (ii) we use $\mu \star 1 = \delta$. Note that $D_1(s) = \zeta(s)$ and $D_\delta(s) = 1$. By (i), it follows that $D_\mu(s)\zeta(s) = 1$, which is the stated result.

For (iii), by (3.1) we have $\zeta'(s) = -\sum_n \log n \cdot n^{-s}$, that is to say $D_{\log}(s) = -\zeta'(s)$. Since $\Lambda \star 1 = \log$, we have $D_\Lambda D_1 = D_{\log}$, and the result follows. □

Corollary 3.3. *We have* $\zeta(s) \neq 0$ *when* $\operatorname{Re} s > 1$.

Proof. This follows immediately from Proposition 3.2 (ii). □

3.3. **Looking forward.** One of the central results of the course is the fact that ζ, though it is currently defined only for $\operatorname{Re} s > 1$, extends to a meromorphic function on the complex plane, holomorphic except for a simple pole at $s = 1$. The function $(s-1)\zeta(s)$ is then entire. It has zeros ρ, and we shall show that they are of two types: the *trivial zeros* $-2, -4, -6, \dots$ and the *nontrivial zeros*, which all lie in the region $0 \leqslant \operatorname{Re} s \leqslant 1$. To understand ζ in terms of its zeros, it is natural (by analogy with polynomials) to ask to what extent $(s-1)\zeta(s)$ is related to $\prod_\rho (s-\rho)$, or even if we can make sense of the latter quantity. Though this is not quite possible, something like this can be made to work.

Here, then, is a summary of the main facts about ζ we will establish in this course that will be important for the later theory.

Theorem 3.4. *The Riemann ζ-function extends to a meromorphic function, holomorphic except for a simple pole at $s = 1$. It has trivial* *zeros at* $s = -2, -4, -6, \dots$, *all simple, and nontrivial zeros in the* critical strip $0 \leqslant \operatorname{Re} s \leqslant 1$. *Finally, there are constants* A *and* B *such that we have the Weierstrass product expansion*

$$(s-1)\zeta(s) = e^{As+B} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{s/\rho},$$

where the product is over all zeros of ζ.

The product expansion will be established in Theorem 6.2, and the remainder of the theorem shown in Corollary 5.5.

At least one of the facts used in the proof of this theorem – the functional equation (Theorem 5.4) – is of great importance in its own right.

We will not actually show in the main part of the course that there are any nontrivial zeros, but the existence of at least one is shown on Exercise Sheet 4. In fact, sharp asymptotics are known for the number of such zeros with imaginary part $\leqslant T$; see for instance Davenport's book.

3.4. **Meromorphic continuation to** $\operatorname{Re} s > 0$**.** To conclude this introductory discussion of the ζ-function, we give a useful alternative expression.

Proposition 3.5. *For* $\operatorname{Re} s > 1$ *we have*

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \{x\} x^{-s-1}\,\mathrm{d}x,$$

where $\{x\} := x - \lfloor x \rfloor$.

Proof.

$$\begin{aligned}
\zeta(s) = \sum_{n=1}^{\infty} n^{-s} &= \sum_{n=1}^{\infty} n(n^{-s} - (n+1)^{-s}) \\
&= s\sum_{n=1}^{\infty} n \int_n^{n+1} x^{-s-1}\,\mathrm{d}x \\
&= s\int_1^\infty \lfloor x \rfloor x^{-s-1}\,\mathrm{d}x \\
&= \frac{s}{s-1} - s\int_1^\infty \{x\} x^{-s-1}\,\mathrm{d}x.
\end{aligned}$$

□

The integral on the right is defined whenever $\operatorname{Re} s > 0$ and one can use it to show $\zeta(s)$ has a meromorphic continuation to $\operatorname{Re} s > 0$ which is holomorphic apart from a simple pole at $s = 1$.

4. Some Fourier analysis

A key role will be played in the rest of the course by Fourier analysis. We take the time to develop the parts of the subject that we need in this chapter.

4.1. **Introduction.** Many students will have met both the Fourier transform and "Fourier series" at undergraduate level. It turns out that these are just two instances of the same concept, that of a Fourier transform on a locally compact abelian group (LCAG). We will not go into any details of the theory in this generality – in fact there is not even any need to define a LCAG. The reader may, however, benefit from seeing the unified context at least in vague outline.

If G is a LCAG then we associate to G its *dual* $\widehat{G}$, which consists of all continuous homomorphisms (characters) from G to S^1, the unit circle $\{z \in \mathbf{C} : |z| = 1\}$. It is easy to see that $\widehat{G}$ is a group under pointwise multiplication.

It turns out that every LCAG G has an essentially unique Haar measure μ_G that is invariant under translation by G. Then for $f : G \to \mathbf{C}$ a μ_G-integrable function and $\gamma \in \widehat{G}$ the Fourier transform $\widehat{f}(\gamma)$ is defined to be

$$\widehat{f}(\gamma) := \int_G f(x)\overline{\gamma(x)}\,\mathrm{d}\mu_G(x) = \langle f, \gamma\rangle.$$

We will not need the general theory here, and instead illustrate with examples:

Example ($G = \mathbf{R}$). It turns out that all characters have the form $x \mapsto e^{i\xi x}$, for $\xi \in \mathbf{R}$. Thus $\widehat{\mathbf{R}} = \mathbf{R}$, that is $\mathbf{R}$ is self-dual. Haar measure can be taken to be the usual Lebesgue measure. If $f : \mathbf{R} \to \mathbf{C}$ is integrable then we define

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x)e^{-i\xi x}\,\mathrm{d}x. \tag{4.1}$$

Strictly speaking, $\widehat{\mathbf{R}}$ is not equal to $\mathbf{R}$, it is homeomorphically isomorphic to $\mathbf{R}$. A similar abuse appears in all the remaining examples too.

Example ($G=\mathbf{Z}$). It turns out that all characters have the form $n\mapsto e^{2\pi i\theta n}$, where $\theta\in\mathbf{R}/\mathbf{Z}$. Thus $\widehat{\mathbf{Z}}=\mathbf{R}/\mathbf{Z}$. Haar measure can be taken to be counting measure. If $f:\mathbf{Z}\to\mathbf{C}$ is integrable, which here just means $\sum_{n\in\mathbf{Z}}|f(n)|<\infty$, then we define

$$\widehat{f}(\theta):=\sum_{n\in\mathbf{Z}}f(n)e^{-2\pi i\theta n}.$$

Example ($G=\mathbf{R}/\mathbf{Z}$: "Fourier series"). Here all the characters have the form $\theta\mapsto e^{2\pi in\theta}$, where $n\in\mathbf{Z}$. Thus $\widehat{\mathbf{R}/\mathbf{Z}}=\mathbf{Z}$. Haar measure can be taken to be Lebesgue measure on the circle. If $f:\mathbf{R}/\mathbf{Z}\to\mathbf{C}$ is integrable we define

$$\widehat{f}(n):=\int_0^1 f(\theta)e^{-2\pi i\theta n}\,\mathrm{d}\theta.$$

The last two groups, $\mathbf{Z}$ and $\mathbf{R}/\mathbf{Z}$, are an example of a *dual pair*.

Example ($G=\mathbf{Z}/N\mathbf{Z}$: Discrete Fourier transform). Here all characters have the form $x\mapsto e^{2\pi irx/N}$, where $r\in\mathbf{Z}/N\mathbf{Z}$ (thus $\widehat{G}\cong G$). Haar measure can be taken to be the uniform probability measure on G, and in this finite setting every $f:\mathbf{Z}/N\mathbf{Z}\to\mathbf{C}$ is integrable, and we define

$$\widehat{f}(r):=\frac{1}{N}\sum_{x\in\mathbf{Z}/N\mathbf{Z}}f(x)e^{-2\pi irx/N}.$$

Conventionally, in analytic number theory one writes $e(t):=e^{2\pi it}$ for $t\in\mathbf{R}$; this makes the notation of these last three examples cleaner.

Example ($G=\mathbf{R}^\times_{>0}$: Mellin transform, special case). In fact, G is homeomorphically isomorphic to $\mathbf{R}$ via the map $x\mapsto\log x$. The characters are $x\mapsto x^{i\xi}$ for $\xi\in\mathbf{R}$. Here the Haar measure may be taken to be $\frac{\mathrm{d}x}{x}$ where $\mathrm{d}x$ is Lebesgue measure. For f integrable we define

$$\widehat{f}(\xi)=\int_0^\infty f(x)x^{-i\xi}\frac{\mathrm{d}x}{x}.$$

The Mellin transform will be defined later in Definition 4.15.

Much of the theory of Fourier analysis is concerned with such questions as when one can prove an inversion formula and what the decay of Fourier coefficients tells us about the smoothness of a function. This is a fascinating theory. However, for much of analytic number theory the deeper parts of this theory can be avoided.

In this course we will work with particularly nice classes of functions called Schwartz functions, where most of the analytic issues can be avoided and Fourier analysis takes on an almost "algebraic" flavour.

4.2. **Fourier analysis of Schwartz functions on R.** In this section, and for the rest of the course, we use the symbol ∂ for the differentiation operator. We also use the notation $\|f\|_1 := \int_{-\infty}^{\infty} |f(x)|\,\mathrm{d}x$.

The Fourier transform of a smooth, that is infinitely differentiable, function decays rapidly. This fundamental fact will be used repeatedly in this course.

Lemma 4.1. *Suppose that $g : \mathbf{R} \to \mathbf{C}$ is a smooth function, all of whose derivatives lie in $L_1(\mathbf{R})$ (that is, are integrable) and decay at infinity, meaning for all m, $\partial^m g(x) \to 0$ as $|x| \to \infty$. Then for any $m \geqslant 0$ (and $\xi \neq 0$) we have the decay estimate*

$$|\widehat{g}(\xi)| \leqslant |\xi|^{-m} \|\partial^m g\|_1.$$

Remark. In particular, if g is regarded as fixed then we have $|\widehat{g}(\xi)| \ll_m |\xi|^{-m}$ for all m.

Proof. Recall the definition of the Fourier transform, that is to say

$$\widehat{g}(\xi) := \int_{-\infty}^{\infty} g(x) e^{-i\xi x}\,\mathrm{d}x.$$

Repeated integration by parts gives

$$\widehat{g}(\xi) = \Big(-\frac{1}{i\xi}\Big)^m \int_{-\infty}^{\infty} \partial^m g(x) e^{-i\xi x}\,\mathrm{d}x.$$

The result then follows immediately from the triangle inequality. $\square$

We now introduce the notion of a *Schwartz function*, which is a smooth function, all of whose derivatives decay rapidly at infinity.

Definition 4.2. Suppose that f is smooth. We say that f belongs to Schwartz space $\mathcal{S}(\mathbf{R})$ if

$$\lim_{|x|\to\infty} |x|^n \partial^m f(x) = 0$$

for all integers $m, n \geqslant 0$.

In particular, since $\partial^m f(x) \ll |x|^{-2}$ for large x, every derivative of a Schwartz function f lies in $L_1(\mathbf{R})$ and hence has a well-defined Fourier transform.

One of the main reasons for introducing this definition is that the Fourier transform maps Schwartz functions to Schwartz functions.

Lemma 4.3. *Suppose that $f \in \mathcal{S}(\mathbf{R})$. Then $\widehat{f} \in \mathcal{S}(\mathbf{R})$.*

Proof. Note first of all that if $f \in \mathcal{S}(\mathbf{R})$ then, for any fixed $n \in \mathbf{Z}_{\geqslant 0}$, $x^n f \in \mathcal{S}(\mathbf{R})$ as well. Indeed, by the general general Leibniz rule $\partial^m(x^n f)$ is a sum of $O_m(1)$ terms of the form $x^k\partial^j f$, and every such term decays quicker than any power of x.

Now for $\theta \in \mathbf{R}$ we have $|e^{-i\theta} - 1 + i\theta| \leqslant \frac{1}{2}\theta^2$ (Lemma A.2), and so

$$\left|\widehat{f}(\xi + h) - \widehat{f}(\xi) + ih(xf)^\wedge(\xi)\right| \leqslant \frac{1}{2}\int_{\mathbf{R}} |f(x)||hx|^2 \,\mathrm{d}x.$$

Since $x^2 f \in \mathcal{S}(\mathbf{R}) \subset L_1(\mathbf{R})$, and the integral on the right is $O_{\xi,f}(h^2)$. Hence as $h \to 0$ we have $\partial\widehat{f}(\xi) = -i(xf)^\wedge(\xi)$. Inductively this gives

$$\partial^n\widehat{f}(\xi) = (-i)^n(x^n f)^\wedge(\xi).$$

Since $x^n f$ is a Schwartz function, all of its derivatives lie in $L_1(\mathbf{R})$. It follows from Lemma 4.1 that $\partial^n\widehat{f}(\xi)$ decays quicker than any polynomial, and hence $\widehat{f} \in \mathcal{S}(\mathbf{R})$. □

The Fourier transform also has a certain useful duality:

Lemma 4.4. *Suppose that $f, g \in L_1(\mathbf{R})$. Then*

$$\int_{\mathbf{R}} f(\xi)\widehat{g}(\xi)\,\mathrm{d}\xi = \int_{\mathbf{R}} \widehat{f}(x)g(x)\,\mathrm{d}x.$$

Proof. The function $(x, \xi) \mapsto f(\xi)g(x)e^{-i\xi x}$ is integrable on $\mathbf{R}\times\mathbf{R}$, and so by Fubini's theorem (Theorem A.4),

$$\begin{aligned}\int_{\mathbf{R}} f(\xi)\widehat{g}(\xi)\,\mathrm{d}\xi &= \int_{\mathbf{R}} f(\xi)\int_{\mathbf{R}} g(x)e^{-i\xi x}\,\mathrm{d}x\,\mathrm{d}\xi\\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} f(\xi)g(x)e^{-i\xi x}\,\mathrm{d}x\,\mathrm{d}\xi\\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} f(\xi)g(x)e^{-i\xi x}\,\mathrm{d}\xi\,\mathrm{d}x\\ &= \int_{\mathbf{R}}\left(\int_{\mathbf{R}} f(\xi)e^{-i\xi x}\,\mathrm{d}\xi\right)g(x)\,\mathrm{d}x = \int_{\mathbf{R}}\widehat{f}(x)g(x)\,\mathrm{d}x\end{aligned}$$

as required. □

We will need an explicit form for the Fourier transform of Gaussian functions. This is also used, in its own right, in the proof of Lemma 5.7.

Lemma 4.5. *Suppose that $t \in \mathbf{R}_{>0}$, and set $f(x) = e^{-\pi x^2 t}$. Then $\widehat{f}(\xi) = \frac{1}{\sqrt{t}} e^{-\xi^2/4\pi t}$.*

Proof. The general case follows from the case $t = 1$ by change of variables. Suppose, then, that $f(x) = e^{-\pi x^2}$. Observe that

$$\widehat{f}(\xi) = e^{-\xi^2/4\pi} \int_{-\infty}^{\infty} e^{-\pi(x+\frac{i\xi}{2\pi})^2}\, \mathrm{d}x = e^{-\xi^2/4\pi} \int_{\Gamma_1} f(z)\, \mathrm{d}z, \tag{4.2}$$

where Γ_1 is the line contour running from $-\infty + i\xi/2\pi$ to $\infty + i\xi/2\pi$. To evaluate this contour integral, integrate f (which is clearly defined on all of $\mathbf{C}$) around the box defined by contours $\Gamma_{1,R} = [R + i\xi/2\pi, -R + i\xi/2\pi]$, $\Gamma_{2,R} = [-R + i\xi/2\pi, -R]$, $\Gamma_{3,R} = [-R, R]$ and $\Gamma_{4,R} = [R, R + i\xi/2\pi]$. Clearly $\int_{\Gamma_{2,R}} f(z)\, \mathrm{d}z$ and $\int_{\Gamma_{4,R}} f(z)\, \mathrm{d}z$ both tend to 0 as $R \to \infty$, and so by Cauchy's integral theorem

$$\int_{\Gamma_1} f(z)\, \mathrm{d}z = -\lim_{R\to\infty} \int_{\Gamma_{1,R}} f(z)\, \mathrm{d}z = \lim_{R\to\infty} \int_{\Gamma_{3,R}} f(z)\, \mathrm{d}z = \int_{-\infty}^{\infty} e^{-\pi x^2}\, \mathrm{d}x.$$

This last integral is well known to be exactly 1. This concludes the proof of the lemma. □

Let $\varepsilon > 0$. Taking $t = \varepsilon^2/4\pi^2$ in the preceding lemma we see that if

$$g_\varepsilon(x) := \frac{1}{\varepsilon} e^{-\pi x^2/\varepsilon^2}$$

then $g_\varepsilon(u) = \widehat{\phi}_\varepsilon(u)$, where

$$\phi_\varepsilon(x) := \frac{1}{2\pi} e^{-x^2\varepsilon^2/4\pi}.$$

We are now in a position to prove one of the key results concerning Fourier analysis on $\mathbf{R}$, the *Fourier inversion formula*:

Proposition 4.6. *Suppose that $f \in \mathcal{S}(\mathbf{R})$. Then*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{ix\xi}\, \mathrm{d}\xi.$$

Proof. First note that it suffices to prove the result when $x = 0$, since if f is Schwartz then so is the "translate" function g defined by $g(t) = f(t + x)$. One may check that $\widehat{g}(\xi) = e^{ix\xi} \widehat{f}(\xi)$, and so applying the inversion formula at 0 to g yields

$$f(x) = g(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{g}(\xi)\, \mathrm{d}\xi = \frac{1}{2\pi} \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{ix\xi}\, \mathrm{d}\xi.$$

Whilst not a crucial step, this simplification affords some notational simplicity.

Using Lemma 4.4 we obtain

$$\int_{-\infty}^{\infty} f(u)g_\varepsilon(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} f(u)\widehat{\phi}_\varepsilon(u)\,\mathrm{d}u = \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(\xi)e^{-\xi^2\varepsilon^2/4\pi}\,\mathrm{d}\xi.$$

Since $|\widehat{f}(\xi)e^{-\xi^2\varepsilon^2/4\pi}| \leqslant |\widehat{f}(\xi)|$ and $\|\widehat{f}\|_1$ is bounded since $f \in \mathcal{S}(\mathbf{R})$, we may apply the dominated convergence theorem (Theorem A.3) and take limits as $\varepsilon \to 0$, giving

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(\xi)e^{-\xi^2\varepsilon^2/4\pi}\,\mathrm{d}\xi \to \frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{f}(\xi)\,\mathrm{d}\xi.$$

It suffices, then, to show that

$$f(0) = \lim_{\varepsilon\to 0}\int_{-\infty}^{\infty} f(u)g_\varepsilon(u)\,\mathrm{d}u.$$

This is a reflection of the fact that g_ε is a spike around 0 with integral 1. In detail, since $\int_{-\infty}^{\infty} g_\varepsilon(x)\,\mathrm{d}x = 1$, it suffices to show that

$$\lim_{\varepsilon\to 0}\int_{-\infty}^{\infty} (f(u) - f(0))g_\varepsilon(u)\,\mathrm{d}u = 0.$$

Divide the integral into the two ranges $|u| \leqslant \sqrt{\varepsilon}$ and $|u| \geqslant \sqrt{\varepsilon}$. The integral over the first range is bounded by

$$\sup_{|u|\leqslant\sqrt{\varepsilon}} |f(u) - f(0)|.$$

The integral over the second is at most

$$2\|f\|_\infty \int_{|u|\geqslant\sqrt{\varepsilon}} g_\varepsilon(u)\,\mathrm{d}u.$$

Note that

$$\int_{x\geqslant\sqrt{\varepsilon}} g_\varepsilon(x)\,\mathrm{d}x = \int_{x\geqslant 1/\sqrt{\varepsilon}} e^{-\pi x^2}\,\mathrm{d}x,$$

which clearly tends to 0 as $\varepsilon \to 0$. Putting these two facts together we obtain

$$\left|\int_{-\infty}^{\infty} (f(u) - f(0))g_\varepsilon(u)\,\mathrm{d}u\right| \leqslant \sup_{|u|\leqslant\sqrt{\varepsilon}} |f(0) - f(u)| + o_{\varepsilon\to 0}(\|f\|_\infty).$$

Since f is continuous at 0, this indeed tends to zero as $\varepsilon \to 0$. □

Remarks. The proof above works equally well for all f with $f \in L_1(\mathbf{R}) \cap C(\mathbf{R})$ and $\widehat{f} \in L_1(\mathbf{R})$.

It is possible to axiomatize the properties of the system (g_ε) that made this work, obtaining the notion of a sequence of "approximations to the identity". We shall not do so here.

A consequence of this and Lemma 4.4 is *Plancherel's formula*:

Proposition 4.7. *Suppose that $f \in \mathcal{S}(\mathbf{R})$. Then*

$$\int_{-\infty}^{\infty} |f(x)|^2 \,\mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2 \,\mathrm{d}\xi.$$

Proof. Take $g := \overline{\widehat{f}}$ in Lemma 4.4, we get

$$\int_{-\infty}^{\infty} f(x)\widehat{g}(x) \,\mathrm{d}x = \int_{-\infty}^{\infty} \widehat{f}(\xi) g(\xi) \,\mathrm{d}\xi = \int_{-\infty}^{\infty} |\widehat{f}(\xi)|^2 \,\mathrm{d}\xi.$$

Meanwhile

$$\widehat{g}(x) = \int \overline{\widehat{f}(\xi)} e^{-i\xi x} \,\mathrm{d}\xi = 2\pi \overline{f(x)},$$

by the inversion formula. The result follows. □

4.3. **Fourier analysis on Z and R/Z.** In the proof of the functional equation for ζ we will make use of some Fourier analysis on the group $\mathbf{Z}$ and its dual $\mathbf{R}/\mathbf{Z}$. We will only develop this theory for sufficiently smooth "Schwartz" functions, and it parallels the theory over $\mathbf{R}$ – just discussed very closely. This is an illustration of the fact that the natural place for harmonic analysis is on a general LCAG.

Definition 4.8. We define $\mathcal{S}(\mathbf{R}/\mathbf{Z})$ to be the space of smooth functions on $\mathbf{R}/\mathbf{Z}$. For the avoidance of doubt this exactly means that $f \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$ if and only if $\widetilde{f} : \mathbf{R} \to \mathbf{C}; \theta \mapsto f(\theta + \mathbf{Z})$ is infinitely differentiable. For $\mathcal{S}(\mathbf{R}/\mathbf{Z})$ there is no "decay at infinity" condition because the group is compact.

We define $\mathcal{S}(\mathbf{Z})$ to be the space of functions $f : \mathbf{Z} \to \mathbf{C}$ with the property that $\lim_{|n| \to \infty} |n|^k |f(n)| = 0$ for all k. For $\mathcal{S}(\mathbf{Z})$ there is no "smoothness" condition for the group since it is discrete.

Recall the definition of the Fourier transform on $\mathbf{R}/\mathbf{Z}$:

$$\widehat{f}(n) := \int_0^1 f(\theta) e^{-2\pi i n\theta} \,\mathrm{d}\theta \tag{4.3}$$

for $n \in \mathbf{Z}$. Recall also the definition of the Fourier transform on $\mathbf{Z}$:

$$\widehat{f}(\theta) = \sum_{n \in \mathbf{Z}} f(n) e^{-2\pi i \theta n} \tag{4.4}$$

for $\theta \in \mathbf{R}/\mathbf{Z}$. Although we are using the hat symbol in two rather different contexts at once, these can be distinguished by looking at the group upon which the function under consideration is defined.

Lemma 4.9. *If $f \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$ then $\widehat{f} \in \mathcal{S}(\mathbf{Z})$. If $f \in \mathcal{S}(\mathbf{Z})$ then $\widehat{f} \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$.*

Proof. For the first statement

$$\widehat{f}(n) = \int_0^1 f(\theta) e^{-2\pi i n \theta} \, \mathrm{d}\theta = (-2\pi i n)^{-(k+1)} \int_0^1 \partial^{k+1} f(\theta) e^{-2\pi i n \theta} \, \mathrm{d}\theta.$$

Thus we have

$$|n^k \widehat{f}(n)| \ll \frac{1}{n} \int_0^1 |\partial^{k+1} f(\theta)| \, \mathrm{d}\theta \ll \frac{1}{n},$$

so indeed $\lim_{|n| \to \infty} n^k \widehat{f}(n) = 0$.

For the second, recall that for $\theta \in \mathbf{R}$ we have $|e^{-i\theta} - 1 + i\theta| \leqslant \frac{1}{2}\theta^2$ (Lemma A.2). Hence, if $f \in \mathcal{S}(\mathbf{Z})$ we have $nf(n), n^4 f(n) \in \mathcal{S}(\mathbf{Z})$ and so

$$\left| \widehat{f}(\theta + h) - \widehat{f}(\theta) + 2\pi i h (nf)^\wedge(\theta) \right| \leqslant \frac{1}{2} \sum_{n \in \mathbf{Z}} 4\pi^2 h^2 n^2 |f(n)| = O_f(h^2).$$

Taking limits as $h \to 0$ we have that $\widehat{f}$ is differentiable and $\partial \widehat{f} = -2\pi i (nf)^\wedge$. Now we show by induction that every $f \in \mathcal{S}(\mathbf{Z})$ is r-times differentiable for every $r \in \mathbf{Z}_{\geqslant 0}$. The base case is vacuous and then the preceding showed the inductive step since $nf \in \mathcal{S}(\mathbf{Z})$. The result is proved. $\square$

This time the dominated convergence theorem (Theorem A.3) allows one to conclude the following analogue of Lemma 4.4 by interchanging the order of integration and summation.

Lemma 4.10. *Suppose that $f \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$ and $g \in \mathcal{S}(\mathbf{Z})$. Then*

$$\sum_{n \in \mathbf{Z}} \widehat{f}(n) g(n) = \int_0^1 f(\theta) \widehat{g}(\theta) \, \mathrm{d}\theta.$$

Now it is time to prove the inversion formula. In this setting, with two different groups $\mathbf{Z}$ and $\mathbf{R}/\mathbf{Z}$, there are two different inversion formulæ. One is almost trivial:

Proposition 4.11. *Suppose that $f \in \mathcal{S}(\mathbf{Z})$. Then*

$$f(n) = \int_0^1 \widehat{f}(\theta) e^{2\pi i n\theta}\, \mathrm{d}\theta.$$

Proof. Substitute in the definition of $\widehat{f}(\theta)$ and swap the order of summation and integration. Then use the fact that

$$\int_0^1 e^{2\pi i m\theta}\, \mathrm{d}\theta = 0$$

when $m \in \mathbf{Z} \setminus \{0\}$. □

The other inversion formula is less trivial (it is also the one that we need in establishing the Poisson summation formula). It is proved using a type of "approximation to the identity" rather analogous to the use of Gaussians in the proof of Proposition 4.6.

Proposition 4.12. *Suppose that $f \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$. Then*

$$f(\alpha) = \sum_{n \in \mathbf{Z}} \widehat{f}(n) e^{2\pi i n\alpha}.$$

Proof. We will use Lemma 4.10 and will deal with the case $\alpha = 0$ for notational simplicity (the general case may, as before, be deduced from this rather easily). Define, for $N \in \mathbf{N}$,

$$g_N(n) := \big(1 - \frac{|n|}{N}\big)\mathbf{1}_{|n| \leqslant N}.$$

By direct computation one may see that

$$\widehat{g}_N(\theta) = \frac{1}{N}\left|\frac{1 - e^{-2\pi i\theta N}}{1 - e^{-2\pi i\theta}}\right|^2 = \frac{\sin^2(\pi\theta N)}{N\sin^2(\pi\theta)}.$$

Write

$$K_N(\theta) := \frac{\sin^2(\pi\theta N)}{N\sin^2(\pi\theta)}.$$

This is called the *Fejér kernel.*

From Lemma 4.10 we know that

$$\sum_{n \in \mathbf{Z}} \widehat{f}(n) g_N(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\theta) K_N(\theta)\, \mathrm{d}\theta.$$

(Later on, it is notationally convenient to take the range of integration over $[-\frac{1}{2}, \frac{1}{2}]$; since the integrand has period 1, this is permissible.) It is quite easy to see that

$$\lim_{N\to\infty} \sum_{n\in\mathbf{Z}} \widehat{f}(n) g_N(n) = \sum_{n\in\mathbf{Z}} \widehat{f}(n).$$

It remains, then, to show that

$$\lim_{N\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\theta) K_N(\theta)\, \mathrm{d}\theta = f(0). \tag{4.5}$$

Since for large N the mass of $K_N(\theta)$ is concentrated near $\theta \approx 0$, this is intuitively reasonable in the same way the analogous statement was in the proof of Proposition 4.6. The details are remarkably similar as well, and depend on the following properties of the Fejér kernel, which again constitute a notion of "approximation to the identity":

(i) $K_N(\theta) \geqslant 0$
(ii) $\int_{-\frac{1}{2}}^{\frac{1}{2}} K_N(\theta)\ \mathrm{d}\theta = 1$
(iii) $\int_{\frac{1}{2}\geqslant|\theta|\geqslant\delta} K_N(\theta)\ \mathrm{d}\theta \ll \frac{1}{N\delta}$.

To prove (ii), one would do best to recall the definition of $K_N(\theta)$ as $\widehat{g}_N(\theta)$, rather than attempt to integrate the closed form of K_N directly. Indeed,

$$\begin{aligned}\int_{-\frac{1}{2}}^{\frac{1}{2}} K_N(\theta)\, \mathrm{d}\theta = \int_{-\frac{1}{2}}^{\frac{1}{2}} \widehat{g}_N(\theta)\, \mathrm{d}\theta &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \sum_n g_N(n) e^{-2\pi i n\theta}\, \mathrm{d}\theta \\ &= \sum_n g_N(n) \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i n\theta}\, \mathrm{d}\theta = g_N(0) = 1.\end{aligned}$$

To prove (iii) we use the fact that $|\sin t| \geqslant 2|t|/\pi$ for $|t| \leqslant \pi/2$. Indeed we have

$$\int_{\frac{1}{2}\geqslant|\theta|\geqslant\delta} K_N(\theta)\, \mathrm{d}\theta \ll \frac{1}{N}\int_{\delta}^{\frac{1}{2}} \frac{1}{|\theta|^2}\, \mathrm{d}\theta \ll \frac{1}{\delta N}.$$

Returning to the proof of (4.5), observe that from (ii) we have

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(\theta) K_N(\theta)\, \mathrm{d}\theta - f(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} (f(\theta) - f(0)) K_N(\theta)\, \mathrm{d}\theta.$$

Split the range of integration into two ranges, $|\theta| \leqslant \delta$ and $\delta < |\theta| \leqslant \frac{1}{2}$, where we will specify δ later. The integral over the first range is bounded by

$$\sup_{|\theta|\leqslant\delta}|f(\theta)-f(0)|\int_{-\frac{1}{2}}^{\frac{1}{2}}|K_N(\theta)|\,\mathrm{d}\theta$$

which, by (i) and (ii), is at most

$$\sup_{|\theta|\leqslant\delta}|f(\theta)-f(0)|. \tag{4.6}$$

The integral over the second range is bounded by

$$2\|f\|_\infty\int_{\delta<|\theta|\leqslant\frac{1}{2}}|K_N(\theta)|\,\mathrm{d}\theta$$

which, by (iii), is at most

$$\ll \|f\|_\infty\frac{1}{N\delta}. \tag{4.7}$$

Taking $N=\lceil 1/\delta^2\rceil$ (say) and letting $\delta\to 0$, we see that both (4.6) (since f is continuous at 0) and (4.7) tend to 0. This concludes the proof of Proposition 4.12. □

We conclude this section by stating, as a corollary, the only result concerning Fourier series that is actually needed in the course (in the proof of the Poisson summation formula). This is usually called the *uniqueness principle* for Fourier coefficients.

Corollary 4.13. *Suppose that $f_1, f_2\in\mathcal{S}(\mathbf{R}/\mathbf{Z})$ are two functions with the property that $\widehat{f_1}(n)=\widehat{f_2}(n)$ for all $n\in\mathbf{Z}$. Then $f_1\equiv f_2$ identically.*

Proof. This is immediate from Proposition 4.12. □

4.4. **The Poisson summation formula.** An important ingredient in the proof of the functional equation for the ζ-function is a version of the *Poisson summation formula*, which is the following statement.

Lemma 4.14. *Suppose that $f\in\mathcal{S}(\mathbf{R})$. Then*

$$\sum_{n\in\mathbf{Z}}f(n)=\sum_{n\in\mathbf{Z}}\widehat{f}(2\pi n).$$

Proof. Consider two functions $F,G:\mathbf{R}/\mathbf{Z}\to\mathbf{C}$, defined by

$$F(\theta):=\sum_{n\in\mathbf{Z}}\widehat{f}(2\pi n)e^{2\pi i n\theta}$$

and

$$G(\theta):=\sum_{k\in\mathbf{Z}}f(\theta+k).$$

We shall show F and G are well-defined and lie in $\mathcal{S}(\mathbf{R}/\mathbf{Z})$. In the first case, because $f \in \mathcal{S}(\mathbf{R})$ we have $\widehat{f} \in \mathcal{S}(\mathbf{R})$, and so *a fortiori* for every $k \geqslant 0$, $|n|^k\widehat{f}(2\pi n) \to 0$ as $|n| \to \infty$. It follows that $g : \mathbf{Z} \to \mathbf{C}; n \mapsto \widehat{f}(2\pi n)$ has $g \in \mathcal{S}(\mathbf{Z})$, but then $F(\theta) = \widehat{g}(-\theta)$ and so $F \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$.

The function G is a little different. By Taylor's theorem for every $\theta, h \in \mathbf{R}$ we have

$$|f(\theta+h) - f(\theta) - h\partial f(\theta)| \leqslant \frac{1}{2}h^2 \sup\{|\partial^2 f(\xi)| : |\xi - \theta| \leqslant |h|\}.$$

Note that we use the integral form of the remainder and estimate that. The Lagrange form of the remainder does not in general hold for complex-valued functions. The fact that $f \in \mathcal{S}(\mathbf{R})$ means that the sum in G converges absolutely, and also $\partial f \in \mathcal{S}(\mathbf{R})$ so

$$H(\theta) := \sum_{k\in\mathbf{Z}} \partial f(\theta + k)$$

also converges absolutely. Hence for $|h| \leqslant 1$ we have

$$\begin{aligned}
|G(\theta+h) &- G(\theta) - hH(\theta)| \\
&\leqslant \sum_{k\in\mathbf{Z}} |f(\theta+h+k) - f(\theta+k) - h\partial f(\theta+k)| \\
&\leqslant \frac{1}{2}h^2 \sum_{k\in\mathbf{Z}} \sup\{|\partial^2 f(x)| : |x| \geqslant |\theta+k| - 1\}.
\end{aligned}$$

Since $\partial^2 f \in \mathcal{S}(\mathbf{R})$ we also have $|\partial^2 f(x)| = O((1+|x|)^{-2})$, and hence this last term is $O_\theta(|h|^2)$. Dividing by h and taking limits we get that G is differentiable and $G' = H$. Since H has the same form as G we can repeat this process and by induction we conclude that G is smooth.

We now compute the Fourier coefficients of F and G, the aim being to show that these are equal. On the one hand we have, for $m \in \mathbf{Z}$, by the dominated convergence theorem (Theorem A.3) that

$$\begin{aligned}
\widehat{F}(m) &:= \int_0^1 \Big(\sum_{n\in\mathbf{Z}} \widehat{f}(2\pi n)e^{2\pi i n\theta}\Big)e^{-2\pi i m\theta}\,\mathrm{d}\theta \\
&= \sum_{n\in\mathbf{Z}} \widehat{f}(2\pi n)\int_0^1 e^{2\pi i(n-m)\theta}\,\mathrm{d}\theta = \widehat{f}(2\pi m).
\end{aligned}$$

On the other hand, again by dominated convergence theorem (Theorem A.3), we have

$$\begin{aligned}\widehat{G}(m) &:= \int_0^1 \Big(\sum_{k\in\mathbf{Z}} f(\theta+k)\Big)e^{-2\pi i m\theta}\,\mathrm{d}\theta \\ &= \int_0^1 \Big(\sum_{k\in\mathbf{Z}} f(\theta+k)\Big)e^{-2\pi i m(\theta+k)}\,\mathrm{d}\theta \\ &= \sum_{k\in\mathbf{Z}}\int_k^{k+1} f(x)e^{-2\pi i m x}\,\mathrm{d}x = \widehat{f}(2\pi m).\end{aligned}$$

Hence $\widehat{F}(m) = \widehat{G}(m)$ for all $m \in \mathbf{Z}$, which by Corollary 4.13 implies that $F = G$ identically. In particular $F(0) = G(0)$, which is exactly the Poisson summation formula. □

4.5. **The Mellin transform.** The Fourier transform deals with unitary characters – that is continuous homomorphisms into S^1. This leads to excellent unitary results such as Plancherel's formula above. It will also be useful for us to deal with slightly more general continuous homomorphism into $\mathbf{C}^*$, where the corresponding transform is called the *Mellin transform*:

Definition 4.15. Suppose that $W : \mathbf{R}_{>0} \to \mathbf{C}$ has

$$\int_0^\infty |W(x)|x^t\,\mathrm{d}x < \infty \text{ for all } t \in \mathbf{R}. \tag{4.8}$$

Then for any complex number s we define the Mellin transform $\tilde{W}$ by

$$\tilde{W}(s) := \int_0^\infty W(x)x^s\frac{\mathrm{d}x}{x}.$$

The condition (4.8) ensures that $\tilde{W}(s)$ is defined for all $s \in \mathbf{C}$. Moreover, we have the following:

Proposition 4.16. *Suppose that $W : \mathbf{R}_{>0} \to \mathbf{C}$ satisfies (4.8). Then $\tilde{W}(s)$ is entire.*

Proof. Note that

$$\begin{aligned}&\left|\tilde{W}(s+h) - \tilde{W}(s) - h\int_0^\infty W(x)x^s\log x\frac{\mathrm{d}x}{x}\right| \\ &\qquad\leqslant \frac{1}{2}|h|^2\int_0^\infty |W(x)|x^{\mathrm{Re}\,s}(\log x)^2\max\{1, x^{\mathrm{Re}\,h}\}\frac{\mathrm{d}x}{x}.\end{aligned} \tag{4.9}$$

Now $(\log x)^2 \leqslant x + x^{-1}$ and so, if $|h| \leqslant 1$ we have

$$x^{\operatorname{Re} s}(\log x)^2 \max\{1, x^{\operatorname{Re} h}\} \leqslant x^{|\operatorname{Re} s|+2} + x^{-|\operatorname{Re} s|-2}.$$

It follows from (4.8) that the right hand side of (4.9) is $O_s(|h|^2)$ and hence $\tilde{W}$ is complex differentiable at s. Since s was arbitrary, $\tilde{W}$ is holomorphic as claimed. □

Note that if $W : \mathbf{R}_{>0} \to \mathbf{C}$ is smooth and compactly supported then certainly it satisfies (4.8), and for these functions we have the *Mellin inversion formula*:

Proposition 4.17. *Let $W : \mathbf{R}_{>0} \to \mathbf{C}$ be a smooth function with compact support. Then for any $\sigma \in \mathbf{R}$ we have the Mellin inversion formula*

$$W(x) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} \tilde{W}(s)x^{-s}\,\mathrm{d}s.$$

Proof. Note that

$$\tilde{W}(\sigma + it) = \widehat{w\exp^\sigma}(-t), \tag{4.10}$$

where $w(u) := W(e^u)$ and $\exp^\sigma(u) := e^{\sigma u}$. The function $w\exp^\sigma$ is compactly supported, since W is compactly supported and $u \mapsto e^u$ is a homeomorphism. It is also smooth since W is smooth, $u \mapsto e^u$ is smooth, and $\exp^\sigma$ is smooth. In particular it is in $\mathcal{S}(\mathbf{R})$ and hence we may apply the Fourier inversion formula (Proposition 4.6) to get

$$w(u) = e^{-\sigma u}\frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{w\exp^\sigma}(-t)e^{-itu}\,\mathrm{d}t = e^{-\sigma u}\frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{W}(s)e^{-itu}\,\mathrm{d}t,$$

which is equivalent to the stated formula if one substitutes $u = \log x$ and writes $s = \sigma + it$. □

It is not obvious that there are any smooth functions with compact support that are not identically zero. It turns out that there are as we show in Appendix B. We will revisit this discussion when we develop the explicit formula (Section 7).

5. The meromorphic continuation and functional equation

Entire functions can be thought of as a generalisation of polynomials. For polynomials there is a useful fact that every polynomial p can be written in the form $p(z) = a_0(z - \rho_1)\cdots(z - \rho_n)$ for some zeros

$\rho_1, \ldots, \rho_n$. This is the fact that $\mathbf{C}$ is algebraically closed. We also have a converse: given zeros $\rho_1, \ldots, \rho_n \in \mathbf{C}$ the function $a_0(z-\rho_1)\cdots(z-\rho_n)$ is a polynomial with exactly those zeros.

For entire functions the situation is more complicated: First we must have some constraint on the zeros since we know by the identity theorem if they have a limit point then the function will be identically zero. Secondly, the obvious analogue taking an infinite product above does not in general converge. The next proposition remedies the situation for *some* sets of zeros using what are called *Weierstrass products*.

Proposition 5.1. *Suppose that $\rho_1, \rho_2, \ldots$ is a sequence of non-zero complex numbers with $\sum_{n=1}^{\infty} |\rho_n|^{-2} < \infty$. Then the function*

$$\prod_{n=1}^{\infty} (1 - \frac{z}{\rho_n}) e^{z/\rho_n}$$

is well-defined, entire, and has zeros in the right places with the right multiplicities, meaning that $z \in \mathbf{C}$ is a zero with multiplicity $m(z) := \#\{j : \rho_j = z\}$ where $m(z) = 0$ means z is not *a zero.*

Proof. Let $K := \sum_{n=1}^{\infty} |\rho_n|^{-2} < \infty$, and write R_N for the tail $R_N := \sum_{n=N}^{\infty} |\rho_n|^{-2}$ so that by hypothesis $R_N \to 0$. Write

$$E(z) := \prod_{n=1}^{\infty} \left(1 - \frac{z}{\rho_n}\right) e^{z/\rho_n} \text{ and } E_N(z) := \prod_{n=1}^{N} \left(1 - \frac{z}{\rho_n}\right) e^{z/\rho_n}.$$

First note that $|(1-z)e^z| \leqslant e^{\frac{1}{2}|z|^2}$ for all $z \in \mathbf{C}$, hence

$$|E_N(z)| \leqslant e^{\frac{1}{2}K|z|^2}. \tag{5.1}$$

We also have $|(1-z)e^z - 1| \leqslant |z|^2$ whenever $|z| \leqslant 1$. Since $R_n \to 0$ we have $\rho_n \to \infty$, and so may take M large enough that $|z| \leqslant |\rho_n|$ for all $n \geqslant M$. Hence by Lemma A.1, for $N > M$ we have

$$\left| \prod_{n=M+1}^{N} \left(1 - \frac{z}{\rho_n}\right) e^{z/\rho_n} - 1 \right| \qquad \leqslant e^{R_{M+1}|z|^2} - 1.$$

It follows that

$$|E_N(z) - E_M(z)| \leqslant (e^{R_{M+1}|z|^2} - 1) e^{\frac{1}{2}K|z|^2},$$

and by continuity of the exponential function, the first factor on the right tends to 0 as $M \to \infty$. We conclude that the sequence $E_1(z), E_2(z), \ldots$ is Cauchy and so converges pointwise for all $z \in \mathbf{C}$.

The function E_N is entire and so by Taylor's theorem

$$E_N(z+h) = E_N(z) + hE_N'(z) + h^2\int_0^1 (1-t)E_N''(z+th)\,\mathrm{d}t. \quad (5.2)$$

By Cauchy's integral formula

$$E_N''(z+th) = 2\int_0^1 E_N(z+th+e^{2\pi i\theta})e^{-4\pi i\theta}\,\mathrm{d}\theta.$$

Inserting this into (5.2) and using (5.1) we get

$$\left|\frac{E_N(z+h)-E_N(z)}{h} - E_N'(z)\right| \leqslant |h|e^{\frac{1}{2}K(1+|h|+|z|)^2}.$$

Suppose that $h_j \to 0$ with $h_j \neq 0$. Then, for j, k large enough that $|h_j|, |h_k| \leqslant 1$ we have

$$\left|\frac{E_N(z+h_j)-E_N(z)}{h_j} - \frac{E_N(z+h_k)-E_N(z)}{h_k}\right|$$
$$\leqslant (|h_j|+|h_k|)e^{\frac{1}{2}K(2+|z|)^2}.$$

Letting $N \to \infty$ we conclude that $\frac{1}{h_j}(E(z+h_j)-E(z))$ is Cauchy. Hence it converges to a limit (necessarily the same for every such sequence by interlacing), and E is complex differentiable at z. Since z was arbitrary, E is holomorphic.

It remains to check that E has the right zeros. It is enough to show that

$$T_N(z) := \prod_{n=N}^{\infty}(1-z/\rho_n)e^{z/\rho_n}$$

is nonzero at z once N is sufficiently large. Concretely, since $|\rho_n| \to \infty$ we may let N be sufficiently large that $|\rho_n| > 2|z|$ for all $n \geqslant N$. Then $|(1-z/\rho_n)e^{z/\rho_n}| \geqslant e^{-|z|^2|\rho_n|^{-2}}$ (since $|z/\rho_n| \leqslant \frac{1}{2}$). Hence $|T_N(z)| \geqslant e^{-|z|^2K} > 0$. The result is proved. □

We shall find it helpful to also use some different notation: If Ω is a countable multiset, not containing 0, and such that $\sum_{\rho\in\Omega}|\rho|^{-2} < \infty$, then the preceding proposition tells us the function

$$E_\Omega(z) := \prod_{\rho\in\Omega}(1-\frac{z}{\rho})e^{z/\rho}$$

is well-defined, entire, and has zeros with the correct multiplicities and nowhere else. The sum and product over Ω are taken with multiplicity.

5.1. **The Γ-function.**

Definition 5.2. Define the Γ-function by

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1}\,\mathrm{d}t \tag{5.3}$$

for $\operatorname{Re} z > 0$. Observe that if $\operatorname{Re} z > 0$ then, since $|t^{z-1}| = t^{\operatorname{Re} z - 1}$, the integral converges.

Proposition 5.3 (Basic facts about Γ)**.**

(i) $\Gamma(z)$*, as defined by* (5.3)*, is holomorphic in* $\operatorname{Re} z > 0$.

(ii) $\Gamma(z)$ *extends to a meromorphic function on all of* $\mathbf{C}$*, satisfying the functional equation* $z\Gamma(z) = \Gamma(z+1)$*. It has simple poles at* $z = 0, -1, -2, \dots$ *and no other poles.*

(iii) Set $\Omega = \{-1, -2, -3, \dots\}$*. Then we have the* Weierstrass formula

$$\frac{1}{\Gamma(z)} = z e^{\gamma z} E_\Omega(z) = z e^{\gamma z} \prod_{n=1}^\infty (1 + z/n) e^{-z/n}$$

for all complex z*, where*

$$\gamma := \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right) \tag{5.4}$$

is Euler's constant*. In particular,* $\Gamma(z)$ *is never 0.*

Remark. One can establish that γ exists by comparing the sum $\sum_{j=1}^n \frac{1}{j}$ with the integral $\int_1^n \mathrm{d}x/x$. We have $\gamma = 0.577\dots$. Euler's constant γ is rather mysterious, and in fact it is not even known to be irrational.

Proof. (i) Recall (from Lemma A.2) the inequality

$$|t^h - 1 - h\log t| \leqslant \frac{1}{2}|h|^2 (\log t)^2 \max\{1, t^{\operatorname{Re} h}\}$$

for $h \in \mathbf{C}$ and $t > 0$. Then

$$\left|\Gamma(z+h) - \Gamma(z) - h\int_0^\infty e^{-t} t^{z-1} \log t\,\mathrm{d}t\right| \leqslant \frac{1}{2}|h|^2 \int_0^\infty e^{-t} t^{\operatorname{Re} z - 1} (\log t)^2 \max\{1, t^{\operatorname{Re} h}\}\,\mathrm{d}t.$$

If $|h| < \frac{1}{2}\operatorname{Re} z$ then the integral on the right is $O_z(1)$ by splitting into $0 < t \leqslant 1$ and $1 < t < \infty$. Dividing by h and letting $h \to 0$ we conclude that Γ is differentiable at all z with $\operatorname{Re} z > 0$ and hence Γ is holomorphic in this region.

(ii) Suppose first that $\operatorname{Re} z > 0$. Integrating by parts we have, provided $\operatorname{Re} z > 0$,

$$\begin{aligned}\Gamma(z+1) &= [-t^z e^{-t}]_0^\infty + z\int_0^\infty e^{-t}t^{z-1}\,\mathrm{d}t \\ &= z\Gamma(z).\end{aligned}$$

This relation may be used to extend Γ meromorphically to the whole complex plane. One begins by extending it to the domain $\operatorname{Re} z > -1$ via the formula $\Gamma(z) := \Gamma(z+1)/z$. One can then iterate, extending in turn to the domains $\operatorname{Re} z > -n$, $n = 2, 3, \dots$. In the process of extending to $\operatorname{Re} z > -1$ we introduce a pole at $z = 0$ (a pole because $\Gamma(1) = 1$). This propagates along so that we get simple poles at all negative integers $n = -1, -2, \dots$ as well. $\Gamma(z)$ is meromorphic, and the poles just described are the only ones.

(iii) Define the function $\widetilde{\Gamma}$ by

$$\frac{1}{\widetilde{\Gamma}(z)} := ze^{\gamma z}\prod_{n=1}^\infty (1+z/n)e^{-z/n}. \tag{5.5}$$

By Proposition 5.1, $\frac{1}{\widetilde{\Gamma}}$ is entire and has zeros only at $0, -1, -2, \dots$. Therefore $\widetilde{\Gamma}$ is meromorphic on $\mathbf{C}$ (with poles at $0, -1, -2, \dots$).

Our aim, of course, is to show that $\Gamma(z) = \tilde{\Gamma}(z)$ for all $z \in \mathbf{C}$. By the identity principle and the fact that both Γ and $\tilde{\Gamma}$ are meromorphic, it suffices to establish this when $\operatorname{Re} z > 0$.

Consider, for positive integer m and for $t \in \mathbf{R}_{\geqslant 0}$, the function

$$f_m(t) := (1 - t/m)^m \mathbf{1}_{[0,m]}(t).$$

Note that $f_m(t) \leqslant e^{-t}$ for all m (since $1 - x \leqslant e^{-x}$ when $x \geqslant 0$) and that $\lim_{m\to\infty} f_m(t) = e^{-t}$ for every t. By the dominated convergence theorem (Theorem A.3) it follows that if $\operatorname{Re} z > 0$ then

$$\lim_{m\to\infty}\int_0^\infty f_m(t)t^{z-1}\,\mathrm{d}t = \int_0^\infty e^{-t}t^{z-1}\,\mathrm{d}t = \Gamma(z). \tag{5.6}$$

We claim that the formula

$$\int_0^\infty f_m(t)t^{z-1}\,\mathrm{d}t = \frac{m^z m!}{z(z+1)\dots(z+m)} \tag{5.7}$$

holds. To see this, first make the substitution $t = mu$ to write the integral as $m^z I(m,z)$, where

$$I(m,z) := \int_0^1 (1-u)^m u^{z-1}\,du.$$

Integration by parts gives

$$I(m,z) = \frac{m}{z} I(m-1, z+1).$$

Applying this repeatedly yields

$$I(m,z) = \frac{m}{z} \cdot \frac{m-1}{z+1} \cdots \frac{1}{z+m-1} I(0, z+m),$$

and the claim (5.7) follows upon noting that $I(0, z+m) = \frac{1}{z+m}$.

From (5.5) and some rearrangement we have

$$\int_0^\infty f_m(t) t^{z-1} \, \mathrm{d}t = \tilde{\Gamma}(z) e^{(\gamma - 1 - \frac{1}{2} - \cdots - \frac{1}{m} + \log m) z} \prod_{n=m+1}^{\infty} (1 + z/n) e^{-z/n}.$$

Since the infinite product converges for every z, it follows from this and the definition of γ that

$$\lim_{m \to \infty} \int_0^\infty f_m(t) t^{z-1} \, \mathrm{d}t = \tilde{\Gamma}(z).$$

Comparing this with (5.6) tells us that indeed $\Gamma(z) = \tilde{\Gamma}(z)$ for $\mathrm{Re}\, z > 0$. □

Remark. From (ii), we have $\Gamma(k+1) = k!$ for all positive integers k.

5.2. **The functional equation for ζ.** Define the *completed ζ-function* by

$$\Xi(s) = \gamma(s) \zeta(s) \text{ for } \mathrm{Re}\, s > 1, \tag{5.8}$$

where the *gamma factor* $\gamma(s)$ is defined to equal $\pi^{-s/2}\Gamma(s/2)$. Our objective in this section is to prove the following theorem:

Theorem 5.4. *The completed ζ-function Ξ has a unique meromorphic continuation to $\mathbf{C}$, and its only poles are simple ones at $s = 0$ and 1. It satisfies the* functional equation

$$\Xi(s) = \Xi(1-s).$$

The formal proof starts on p. 36.

As a corollary we obtain the first part of Theorem 3.4, the main result of this chapter about the ζ-function.

Corollary 5.5. *The ζ-function has a unique meromorphic continuation to all of $\mathbf{C}$ extending the identity (5.8). It has a simple pole at $s = 1$, no other poles, and $\zeta(0) \neq 0$. It has zeros at $s = -2, -4, -6, \ldots$*

and no other zeros outside of the critical strip $0 \leqslant \operatorname{Re} s \leqslant 1$. *Furthermore, if ρ is a zero in the critical strip then so are $\overline{\rho}$, $1-\rho$, and $1-\overline{\rho}$.*

Proof. We may rewrite (5.8) as

$$\zeta(s) = \pi^{s/2}(\Gamma(s/2))^{-1}\Xi(s) \text{ for } \operatorname{Re} s > 1.$$

Ξ has a meromorphic continuation to $\mathbf{C}$ by Theorem 5.4, and Γ by Proposition 5.3 (ii). $\pi^{s/2}$ is entire, and hence $\zeta(s)$ has a, necessarily unique, meromorphic continuation to $\mathbf{C}$.

By Proposition 5.3 (iii) $\Gamma(s/2)$ has no zeros. On the other hand $\pi^{s/2}$ has no poles, so the only possible poles of ζ are poles of Ξ. The latter only has simple poles at 0 and 1 by Theorem 5.4, and since $\Gamma(s/2)$ has a simple pole at 0 but not at 1, and $\pi^{s/2}$ has no zeros at all, we see that ζ has a simple pole at 1 and $\zeta(0) \neq 0$.

From Corollary 3.3 we have $\zeta(s) \neq 0$ for $\operatorname{Re} s > 1$. It follows that $\Xi(s) \neq 0$ for $\operatorname{Re} s > 1$, since neither $\Gamma(s/2)$ nor $\pi^{-s/2}$ have zeros in $\operatorname{Re} s > 1$. By the functional equation for Ξ we therefore conclude that $\Xi(s) \neq 0$ for $\operatorname{Re} s < 0$. Since $\pi^{s/2}$ also has no zeros in $\operatorname{Re} s < 0$, it follows that $\zeta(s) = 0$ if and only if $s/2$ is a pole of Γ. This exactly tells us that ζ has simple (because the poles of Γ are simple) zeros at $\{-2, -4, \dots\}$ and nowhere else in $\operatorname{Re} s < 0$. In particular, other than $\{-2, -4, \dots\}$, all the zeros of ζ lie in the critical strip.

By the functional equation for Ξ if ρ is a zero in the critical strip then so is $1-\rho$. Finally, the function $\overline{\zeta(\overline{s})}$ is meromorphic as can be seen, for example, by expanding ζ in a Taylor series about any point in $\mathbf{C} \setminus \{1\}$. But for $\operatorname{Re} s > 1$, we see that $\overline{\zeta(\overline{s})} = \zeta(s)$ from the Dirichlet series. Hence by the identity theorem $\zeta(s) = \overline{\zeta(\overline{s})}$ for all $s \in \mathbf{C} \setminus \{1\}$. In particular ρ is a zero of ζ if and only if $\overline{\rho}$ is a zero of ζ. □

Definition 5.6. The set of all zeros of ζ will be denoted by Z. We further divide this set into the set of *trivial zeros* Z_{TRIV}, by which we mean $-2, -4, -6, \dots$, and *nontrivial zeros* Z_{NONTRIV}, by which we mean all the other zeros. By the above corollary every zero in Z_{NONTRIV} lies in the critical strip $0 \leqslant \operatorname{Re} s \leqslant 1$.

We will often be summing over elements in Z, and we will want to do so with multiplicity. For this purpose we think of Z as a multiset, though it is conjectured that all zeros of ζ are simple and hence that Z is, in fact, a set.

Remarks. With our knowledge at this point, Z_{NONTRIV} could be empty. In fact it is infinite and one can obtain quite good control on the number of nontrivial zeros with imaginary part at most T. We will see this later in Proposition 6.8.

On the other hand, while we know that Z_{NONTRIV} is discrete (*i.e.* it has no accumulation points in $\mathbf{C}$), the only points in the critical strip we currently know not to be zeros are 0 and 1. One of the main results of the course – Proposition 8.1 – will be showing that in fact the line $\operatorname{Re} s = 1$ (and so also $\operatorname{Re} s = 0$) does not intersect Z_{NONTRIV}.

5.3. **Proof of the functional equation.** We turn now to the proof of Theorem 5.4. A key tool here will be the θ-function: For z in the upper half-plane $\mathcal{H} := \{z : \operatorname{Im} z > 0\}$ set

$$\theta(z) := \sum_{n\in\mathbf{Z}} e^{i\pi n^2 z}.$$

This is a uniform limit of continuous functions on $\operatorname{Im} z \geqslant \delta$ for any $\delta > 0$, and hence it is continuous. In fact it is holomorphic on the upper half-plane but we will not need that.

Lemma 5.7. *Suppose that $t \in \mathbf{R}_{>0}$. Then $\theta(it) = \frac{1}{\sqrt{t}}\theta(i/t)$.*

Proof. Set $f(x) = e^{-\pi x^2 t}$, and note that $f \in \mathcal{S}(\mathbf{R})$. In Lemma 4.5 we computed

$$\widehat{f}(\xi) = \frac{1}{\sqrt{t}} e^{-\xi^2/4\pi t},$$

and the lemma now follows by applying the Poisson summation formula Lemma 4.14. □

This symmetry lets us establish some bounds. First we have

$$\left|\frac{\theta(it) - 1}{2}\right| \leqslant \frac{1}{e^{\pi t} - 1} \text{ for } t > 0. \tag{5.9}$$

For $t \in (0, 1]$ we have $\theta(i/t) = 1 + O(e^{-\pi/t})$ from this, which combined with Lemma 5.7 gives

$$\left|\frac{\theta(it) - 1}{2}\right| = O(t^{-1/2}) \text{ whenever } t \in (0, 1]. \tag{5.10}$$

The next lemma relates the ζ-function to the θ-function. It turns out that the ζ-function is, roughly speaking, the *Mellin transform* of θ (see Definition 4.15 later).

Lemma 5.8. *Suppose that* $\operatorname{Re} s > 1$. *Then*

$$\Xi(s) = \int_0^\infty \Big(\frac{\theta(ix)-1}{2}\Big)x^{s/2}\,\frac{\mathrm{d}x}{x}.$$

Proof. First note that since $\operatorname{Re} s > 1$, the integrand is $O(x^{\frac{1}{2}(\operatorname{Re} s-1)})$ for $0 < x \leqslant 1$ by (5.10) and $O(x^{\operatorname{Re} s/2}e^{-\pi x})$ for $x \geqslant 1$ so that the integrand is integrable. Now observe that

$$\int_0^\infty e^{-\pi n^2 x}x^{s/2}\,\frac{\mathrm{d}x}{x} = \pi^{-s/2}n^{-s}\int_0^\infty e^{-u}u^{s/2}\,\frac{du}{u} = \pi^{-s/2}\Gamma(s/2)n^{-s}.$$

Summing over $n \in \mathbf{N}$ and applying the dominated convergence theorem (Theorem A.3) to interchange the sum and the integral gives the result. $\square$

Proof of Theorem 5.4. Suppose that $\operatorname{Re} s > 1$. We split the integral in Lemma 5.8 into the ranges $[0,1]$ and $[1,\infty)$, and then apply Lemma 5.7 to the first of these. We have

$$\begin{aligned}
\int_0^1 \Big(\frac{\theta(ix)-1}{2}\Big)x^{s/2}\,\frac{\mathrm{d}x}{x} &= \frac{1}{2}\int_0^1 \theta(ix)x^{s/2}\,\frac{\mathrm{d}x}{x} - \frac{1}{s}\\
&= \frac{1}{2}\int_0^1 \theta(\frac{i}{x})x^{(s-1)/2}\,\frac{\mathrm{d}x}{x} - \frac{1}{s}\\
&= \frac{1}{2}\int_1^\infty \theta(iu)u^{(1-s)/2}\,\frac{du}{u} - \frac{1}{s}\\
&= \int_1^\infty \Big(\frac{\theta(iu)-1}{2}\Big)u^{(1-s)/2}\,\frac{du}{u} - \frac{1}{s} - \frac{1}{1-s}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\int_0^\infty \Big(\frac{\theta(ix)-1}{2}\Big)x^{s/2}\,\frac{\mathrm{d}x}{x}\\
&\qquad = \int_1^\infty \Big(\frac{\theta(ix)-1}{2}\Big)(x^{s/2}+x^{(1-s)/2})\,\frac{\mathrm{d}x}{x} - \frac{1}{s} - \frac{1}{1-s},
\end{aligned}$$

which implies that

$$\Xi(s) = \int_1^\infty \Big(\frac{\theta(ix)-1}{2}\Big)(x^{s/2}+x^{(1-s)/2})\,\frac{\mathrm{d}x}{x} - \frac{1}{s} - \frac{1}{1-s}. \tag{5.11}$$

A priori this formula is valid only for $\operatorname{Re} s > 1$. Note, however, that the integral on the right is just

$$\tilde{W}\left(\frac{1}{2}s\right) + \tilde{W}\left(\frac{1}{2}(1-s)\right)$$

where

$$W(x) := \mathbf{1}_{[1,\infty)}(x)\Big(\frac{\theta(ix)-1}{2}\Big).$$

By (5.9) we have $\frac{1}{2}|\theta(ix)-1| = O(e^{-\pi x})$ for all $x \geqslant 1$, and hence W satisfies the hypotheses of Proposition 4.16, and so $\tilde{W}$ is entire. It follows that the right hand side in (5.11) defines a function that is meromorphic in the whole complex plane, with simple poles at $s = 0$ and 1. This is a meromorphic continuation of Ξ which is then necessarily unique. It manifestly satisfies the claimed relation $\Xi(s) = \Xi(1-s)$. □

6. The partial fraction expansion

6.1. **Statement.** In prime number theory, the ζ-function itself is less important than the logarithmic derivative ζ'/ζ. This is, of course, because $-\zeta'/\zeta$ is the Dirichlet series of the von Mangoldt function Λ, that is to say $-\zeta'(s)/\zeta(s) = \sum_n \Lambda(n) n^{-s}$.

The heart of the link between primes and zeros of ζ is the *partial fraction expansion* of ζ'/ζ.

Proposition 6.1. *There is a constant A such that*

$$\frac{\zeta'(s)}{\zeta(s)} = A - \frac{1}{s-1} + \sum_{\rho\in Z}\Big(\frac{1}{s-\rho} + \frac{1}{\rho}\Big),$$

where Z is the (multi-)set of zeros of ζ, as in Definition 5.6. Moreover, $\sum_{\rho \in Z} |\rho|^{-2} < \infty$.

The fact that $\sum_{\rho\in Z} |\rho|^{-2} < \infty$ means that the sum over ρ converges for $s \notin Z$ and defines a meromorphic function on $\mathbf{C}$; this is on Exercise Sheet 3.

The constant A can be computed explicitly by setting $s = 0$, so that $A = \frac{\zeta'(0)}{\zeta(0)} - 1 = \log 2\pi - 1 = 0.837\dots$. The computation of $\zeta(0)$ is on Exercise Sheet 3.

Proposition 6.1 is a consequence of a *product expansion* given in Theorem 6.2, and repeated use of the fact that

$$\frac{(fg)'}{fg} = \frac{f'}{f} + \frac{g'}{g}.$$

In fact, we need some version of this for *infinite products*; the justification of this is on Exercise Sheet 4.

Theorem 6.2. *There are constants A and B such that*
$$(s-1)\zeta(s) = e^{As+B}\prod_{\rho\in Z}\Big(1-\frac{s}{\rho}\Big)e^{s/\rho},$$
where again Z is the (multi-)set of zeros of ζ, as in Definition 5.6. Moreover, we have $\sum_{\rho\in Z}|\rho|^{-2} < \infty$, so the product here is a Weierstrass product and converges as described in Proposition 5.1.

As with Proposition 6.1, the constants A and B are known: $A = 0.837\ldots$ as in Proposition 6.1, and $B = -\log 2 = -0.693\ldots$.

Theorem 6.2 is quite reminiscent of the well-known fact that if $f : \mathbf{C} \to \mathbf{C}$ is a polynomial then $f(z) = C\prod_\rho(z-\rho)$, where the product is over the (finite) collection of zeros of f. Of course, $(s-1)\zeta(s)$ is not a polynomial. However, Theorem 6.2 may still be obtained as a consequence of a quite general theorem about entire functions, rather than any particular specific properties of ζ. We state such a theorem now.

Definition 6.3. Suppose that $f : \mathbf{C} \to \mathbf{C}$ is an entire function satisfying a growth condition $|f(z)| = e^{O_\varepsilon(|z|^{1+\varepsilon})}$ for all $\varepsilon > 0$ and $|z| \geqslant 1$. Then we say that f is an integral function of order (at most) one.

Proposition 6.4. *Suppose that f is an integral function of order (at most) one. Suppose that f has a zero of order r at 0, and write Ω for the set of other zeros of f. Then $\sum_{\rho\in\Omega}|\rho|^{-1-\varepsilon} < \infty$ for every $\varepsilon > 0$ and there are constants A and B (depending on f) such that*
$$f(z) = z^r e^{Az+B}\prod_{\rho\in\Omega}\Big(1-\frac{z}{\rho}\Big)e^{z/\rho}.$$

The proof is given in Sections 6.3 and 6.4.

6.2. **The product expansion of ζ.** In this section we will deduce Theorem 6.2 from Proposition 6.4, the product expansion for integral functions of order (at most) one.

Rather than prove that $(s-1)\zeta(s)$ is an integral function of order (at most) one, it is convenient to first include some Γ- and other factors. Define
$$f(s) = s(1-s)\pi^{-s/2}\Gamma(s/2)\zeta(s) = s(1-s)\Xi(s). \tag{6.1}$$

Lemma 6.5. *The function f is entire and*

$$|f(s)| \leqslant e^{O(|s|\log|s|)} \text{ for } |s| \geqslant 3,$$

so in particular f is an integral function of order (at most) one.

Proof. The function f is entire since Ξ is meromorphic except for two simple poles at $s = 0$ and $s = 1$. We have from (5.11) that

$$\Xi(s) = \int_1^\infty \Big(\frac{\theta(ix) - 1}{2}\Big)(x^{s/2} + x^{(1-s)/2})\,\frac{\mathrm{d}x}{x} - \frac{1}{s} - \frac{1}{1-s}.$$

On the other hand (5.9) tells us that $|\theta(ix) - 1| = O(e^{-\pi x})$ for $x \geqslant 1$, and hence for $|s| \geqslant 3$ we have

$$\begin{aligned}|f(s)| &\ll |s|^2 \int_1^\infty e^{-\pi x}(x^{\frac{1}{2}\operatorname{Re} s - 1} + x^{-\frac{1}{2}(1+\operatorname{Re} s)})\,\mathrm{d}x \\ &\leqslant |s|^2\Gamma(|s|+1) \leqslant |s|^2(|s|+2)^{|s|+2} \ll e^{O(|s|\log|s|)}.\end{aligned}$$

The result is proved. □

Proof of Theorem 6.2. Lemma 6.5 means we can apply Proposition 6.4 to get

$$f(s) = e^{As+B}\prod_\rho \Big(1 - \frac{s}{\rho}\Big)e^{s/\rho}, \tag{6.2}$$

where the product is over the zeros ρ of f. In making this claim we have observed that $f(0) \neq 0$; the simple zero of s cancels the simple pole of $\Gamma(s/2)$, and $\zeta(0) \neq 0$ (the calculation of its exact value is on Exercise Sheet 3). The zeros ρ satisfy $\sum_\rho |\rho|^{-2} < \infty$. These zeros ρ are precisely the nontrivial zeros Z_{NONTRIV} of $\zeta(s)$, namely those in the critical strip $0 \leqslant \operatorname{Re} s \leqslant 1$, because the factor of $\Gamma(\frac{1}{2}s)$ cancels out the trivial zeros of ζ at $-2, -4, -6, \dots$. It is clear that $\sum_{\rho \in Z_{\text{TRIV}}} |\rho|^{-2} < \infty$, so $\sum_{\rho\in Z} |\rho|^{-2} < \infty$.

Of course,

$$(s-1)\zeta(s) = -\frac{\pi^{s/2}}{s}\frac{1}{\Gamma(s/2)}f(s).$$

We may now recover the claimed product expansion (with different constants A, B) for $(s-1)\zeta(s)$ from (6.2) and the Weierstrass product for $1/\Gamma$, that is to say Proposition 5.3 (iii). □

6.3. **The size of a holomorphic function and its zeros.** To prove Proposition 6.4 we need various facts about the relation between the size of a holomorphic function and the number of its zeros. We collect the facts we need in this section.

Write $B_R := \{z \in \mathbf{C} : |z| < R\}$.

Theorem 6.6 (Jensen's formula)**.** *Let $R, \varepsilon > 0$. Suppose that f is holomorphic on $B_{R+\varepsilon}$, and that $f(z) \neq 0$ for $R \leqslant |z| < R+\varepsilon$ and for $z = 0$. Then*

$$\int_0^1 \log|f(Re^{2\pi i\theta})|\,\mathrm{d}\theta = \log|f(0)| + \sum_\rho \log\frac{R}{|\rho|}, \tag{6.3}$$

where the sum is over zeros ρ of f in B_R, counted with multiplicity.

Proof. Observe that if the identity is true for functions f_1 and f_2 then it is also true for $f_1 f_2$. Write

$$g_\rho(z) = \frac{R(z-\rho)}{R^2 - \overline{\rho}z}$$

and define a meromorphic function F by

$$f(z) = CF(z)\prod_\rho g_\rho(z),$$

where C is chosen so that $F(0) = 1$. If ε is chosen so small that the poles $z = R^2/\overline{\rho}$ lie outside of $B_{R+\varepsilon}$ then F has no zeros in $B_{R+\varepsilon}$. Jensen's formula being manifestly true for constant functions, it suffices to check it for F and for the functions g_ρ. Now we may define a single-valued, holomorphic logarithm of F in $B_{R+\varepsilon}$ by the formula

$$\log F(z) := \int_{[0\to z]} \frac{F'(w)}{F(w)}\,\mathrm{d}w.$$

Since $F(0) = 1$ the function $z^{-1}\log F(z)$ is also holomorphic in $B_{R+\varepsilon}$. Hence by Cauchy's integral theorem we have

$$\int_{\partial B_R} \frac{\log F(z)}{z}\,\mathrm{d}z = 0.$$

Parametrising the circle ∂B_R by $z = Re^{2\pi i\theta}$, $0 \leqslant \theta < 1$ we get

$$\int_0^1 \log F(Re^{2\pi i\theta})\,\mathrm{d}\theta = 0.$$

Taking real parts gives

$$\int_0^1 \log |F(Re^{2\pi i\theta})|\,\mathrm{d}\theta = 0.$$

This is one side of (6.3); the other side is clearly zero. Thus we have verified the formula for F.

Turning our attention to the functions g_ρ, if $|z| = R$ then

$$|g_\rho(z)| = \left|\frac{\overline{z}(z-\rho)}{R^2 - \overline{\rho}z}\right| = 1.$$

Thus the left-hand side of (6.3) equals 0. As for the right-hand side, $|g_\rho(0)| = |\rho|/R$. Thus the right-hand side is zero as well. □

The next result applies Jensen's formula to get a fairly tight relation between the size of a holomorphic function and the number of its zeros.

Corollary 6.7. *Let f be an entire function with $f(0) = 1$. Then for any R the number of zeros ρ of f with $|\rho| < R$ is at most*

$$2 \sup_{|z|\leqslant 3R} \log |f(z)|.$$

Proof. Pick a radius $R_0 \in [2R, 3R]$ such that $f(z) \neq 0$ whenever $|z| = R_0$. This is possible since otherwise f would have infinitely many zeros in the compact set $|z| \leqslant 3R$, contrary to the identity principle. Jensen's formula (Theorem 6.6) immediately gives

$$\begin{aligned}\sum_\rho \log \frac{R_0}{|\rho|} &= \int_0^1 \log |f(R_0 e^{2\pi i\theta})|\,\mathrm{d}\theta \\ &\leqslant \sup_{|z|=R_0} \log |f(z)| \leqslant \sup_{|z|\leqslant 3R} \log |f(z)|,\end{aligned}$$

where the sum is over all zeros ρ inside B_{R_0}. Each term $\log \frac{R_0}{|\rho|}$ is positive, and the terms corresponding to zeros ρ with $|\rho| < R$ contribute at least $\log 2 > \frac{1}{2}$. □

A corollary of this and some of our earlier estimates is a bound for the number of zeros of ζ.

Proposition 6.8. *Let $T \geqslant 2$. The number of nontrivial zeros ρ of the Riemann ζ-function with $|\operatorname{Im}\rho| \leqslant T$ is $O(T \log T)$.*

Proof. Consider the function $f(s) = s(1-s)\pi^{-s/2}\Gamma(s/2)\zeta(s)$ as defined in (6.1), which vanishes at all nontrivial zeros of ζ. We showed in

Lemma 6.5 that $|f(s)| \leqslant e^{O(|s|\log|s|)}$, and so it follows from Corollary 6.7 (technically applied to $f(s)/f(0)$, and this can be done since $f(0) \neq 0$ since $\zeta(0) \neq 0$, and Γ has a simple pole at 0 which is cancelled by the s-factor in the definition of $f(s)$) that the number of nontrivial zeros of ζ with $|\rho| \leqslant T+1$ is $O(T\log T)$. If ρ is nontrivial and $|\operatorname{Im}\rho| \leqslant T$ then $|\rho| \leqslant T+1$, which gives the result. □

The next result does not use Jensen's formula. It tells us the structure of entire functions of moderate growth with *no* zeros.

Proposition 6.9. *Suppose that g is an entire function with no zeros which satisfies the bound $|g(z)| = e^{O(|z|^{3/2})}$ whenever $|z| = R_j$, $j = 1, 2, \dots$, where $R_j \to \infty$ as $j \to \infty$. Then $g(z) = e^{Az+B}$ for some constants A, B.*

Proof. Multiplying through by a constant, we may assume that $g(0) = 1$. Since g is nonvanishing, it has a holomorphic branch of logarithm $h(z) := \log g(z)$ with $h(0) = 0$, exactly as in the proof of Jensen's formula. We *cannot* immediately conclude that $|h(z)| \ll |z|^{3/2} = R_j^{3/2}$ for $|z| = R_j$, but it does at least follow that $\operatorname{Re} h(z) = \log|g(z)| \leqslant CR_j^{3/2}$ for some constant C, and hence

$$|\operatorname{Re} h(z)| \leqslant 2CR_j^{3/2} - \operatorname{Re} h(z)$$

(consider the cases $\operatorname{Re} h(z) \geqslant 0$ and $\operatorname{Re} h(z) \leqslant 0$ separately). It follows that

$$\begin{aligned}\int_0^1 |\operatorname{Re} h(R_je^{2\pi i\theta})|\,\mathrm{d}\theta &\leqslant 2CR_j^{3/2} - \int_0^1 \operatorname{Re} h(R_je^{2\pi i\theta})\,\mathrm{d}\theta \\ &= 2CR_j^{3/2} - \operatorname{Re}\int_0^1 h(R_je^{2\pi i\theta})\,\mathrm{d}\theta = 2CR_j^{3/2},\end{aligned} \tag{6.4}$$

by Cauchy's integral theorem.

Now if the Taylor expansion of $h(z)$ is $\sum_{n=0}^\infty c_nz^n$ then we have

$$2\operatorname{Re}(h(R_je^{2\pi i\theta})) = \sum_{n\geqslant 0} c_nR_j^ne^{2\pi in\theta} + \sum_{n\geqslant 0}\overline{c}_nR_j^ne^{-2\pi in\theta}.$$

Writing $c_{-n} := \overline{c_n}$ for $n > 0$, and noting that $c_0 = 0$ since $h(0) = 0$, these are just the Fourier coefficients of $f(\theta) := 2\operatorname{Re}(h(R_je^{2\pi i\theta}))$ by Corollary 4.13. ($f \in \mathcal{S}(\mathbf{R}/\mathbf{Z})$ since h is holomorphic, so smooth, and

$e^{2\pi i\theta}$ is smooth, $z \mapsto \operatorname{Re} z$ is smooth (but *not* holomorphic), and the composition of smooth functions is smooth.) It follows that

$$c_m = 2R_j^{-m} \int_0^1 \operatorname{Re}(h(R_j e^{2\pi i\theta}))e^{-2\pi i m\theta}\,\mathrm{d}\theta,$$

and so by the triangle inequality and (6.4), that

$$|c_m| \ll R_j^{3/2-m}.$$

Letting $j \to \infty$, this implies that $c_2 = c_3 = \cdots = 0$, and so $h(z) = c_0 + c_1 z$ is linear. This concludes the proof. $\square$

6.4. **Weierstrass products: proof.** The aim of this section is to give a detailed proof of:

Proposition 6.4. *Suppose that f is an integral function of order (at most) one. Suppose that f has a zero of order r at 0, and write Ω for the set of other zeros of f. Then $\sum_{\rho\in\Omega} |\rho|^{-1-\varepsilon} < \infty$ for every $\varepsilon > 0$ and there are constants A and B (depending on f) such that*

$$f(z) = z^r e^{Az+B} \prod_{\rho\in\Omega} \Big(1 - \frac{z}{\rho}\Big)e^{z/\rho}.$$

We begin by recalling the properties of $E_\Omega(z)$:

Proposition 5.1 (Weierstrass product)**.** *Suppose that Ω is such that $0 \notin \Omega$ and $\sum_{\rho\in\Omega} |\rho|^{-2} < \infty$ (where the sum over Ω is taken with multiplicity). Then the function*

$$E_\Omega(z) := \prod_{\rho\in\Omega} \Big(1 - \frac{z}{\rho}\Big)e^{z/\rho}$$

is well-defined, entire, and has zeros at Ω with the correct multiplicities and nowhere else.

Proof of Proposition 6.4. We may write $f(z) = Cz^r h(z)$ for an entire function h with $h(0) = 1$. Since f is an integral function of order (at most) one and z^{-r} is bounded for $|z| \geqslant 1$, we have $|h(z)| = e^{O_{\varepsilon'}(|z|^{1+\varepsilon'})}$ for all $\varepsilon' > 0$ and $|z| \geqslant 1$. Apply Corollary 6.7 to h to get that

$$\#\{\rho : R \leqslant |\rho| \leqslant 2R\} \ll_{\varepsilon'} R^{1+\varepsilon'}, \tag{6.5}$$

for every $\varepsilon' > 0$. In particular, given $\varepsilon > 0$ and taking $\varepsilon' = \varepsilon/2$ in (6.5), we see that

$$\sum_{\rho \neq 0} |\rho|^{-1-\varepsilon} \leqslant \sum_{\substack{\rho \neq 0 \\ |\rho| \leqslant 1}} |\rho|^{-1-\varepsilon} + \sum_{j=0}^{\infty} \#\{\rho : 2^j < |\rho| \leqslant 2^{j+1}\}(2^j)^{-1-\varepsilon}$$
$$= O(1) + \sum_{j=0}^{\infty} O_\varepsilon(2^{-j\varepsilon/2}) < \infty.$$

This gives the first part of the conclusion. Taking $\varepsilon = 1$ we may apply Proposition 5.1 with Ω being the set of zeros of f other than 0, obtaining an entire function $E_\Omega(z)$ which vanishes (with the correct multiplicity) precisely at Ω. Define

$$g(z) := \frac{f(z)}{z^r E_\Omega(z)}.$$

By construction, g is an entire function of z with no zeros. To complete the proof of Proposition 6.4, we need only show that $g(z) = e^{Az+B}$. We already have a tool for doing precisely this, namely Proposition 6.9. Applying this proposition, we see that all we need do is establish an upper bound

$$|g(z)| \leqslant e^{O(R_j^{3/2})}$$

for $|z| = R_j$, for some sequence of radii $R_j \to \infty$.

Since we already have a bound on f (by assumption) and z^{-r} is bounded, it is enough to establish the *lower* bound

$$|E_\Omega(z)| \geqslant e^{-O(R_j^{3/2})} \text{ for all } z \text{ with } |z| = R_j. \qquad (6.6)$$

This is slightly delicate, and must involve a careful choice of the R_j, since $E_\Omega(z)$ vanishes at the points of Ω. In the light of this, it makes sense to choose the radii R_j to lie away from any of the zeros ρ.

We will show that for every j we can choose an $R_j \in [2^j, 2^{j+1}]$ such that (6.6) is satisfied. Write S_j for the set of zeros with $2^j \leqslant |\rho| < 2^{j+1}$. Then, (6.5) with $\varepsilon = 1$, certainly implies that $|S_j| < C2^{2j}$. It follows that R_j may be chosen in such a way that for every zero ρ,

$$|z - \rho| \geqslant \frac{2}{21C} 2^{-j} \gg 2^{-j} \text{ whenever } |z| = R_j. \qquad (6.7)$$

To see this, note that for (6.7) to hold each zero $\rho \in S_{j-1} \cup S_j \cup S_{j+1}$ excludes radii in the interval $[|\rho| - \frac{2}{21C}2^{-j}, |\rho| + \frac{2}{21C}2^{-j}]$. These intervals have length $\frac{4}{21C}2^{-j}$, and so their union has measure at most $(|S_{j-1}| +$

$|S_j| + |S_{j+1}|) \cdot \frac{4}{21C} 2^{-j} < 2^j$. However the interval $[2^j, 2^{j+1}]$ has measure 2^j so there must be a radius not excluded.

Having chosen our radii, suppose that z has $|z| = R_j$. To get a lower bound on $E_\Omega(z)$, we write

$$E_\Omega(z) = E_\Omega^*(z) \cdot \prod_{j'=0}^{\infty} E_\Omega^{(j')}(z) \tag{6.8}$$

where

$$E_\Omega^{(j')}(z) := \prod_{\rho \in S_{j'}} (1 - z/\rho) e^{z/\rho} \text{ for } j' = 0, 1, 2, \ldots,$$

and

$$E_\Omega^*(z) := \prod_{|\rho|<1} (1 - z/\rho) e^{z/\rho}.$$

Since f is entire it has finitely many zeros ρ with $|\rho| < 1$ and so it follows that

$$|E_\Omega^*(z)| \geqslant e^{-O(R_j)}. \tag{6.9}$$

For the infinite product in (6.8) we have three ranges for j':

(i) $j' < j - 1$: for any zero ρ occurring in the product $E_\Omega^{(j')}(z)$, we have $|1 - z/\rho| \geqslant 1$ and $|e^{z/\rho}| \geqslant e^{-2^{j-j'}}$. Hence

$$|E_\Omega^{(j')}(z)| \geqslant (e^{-2^{j-j'}})^{|S_{j'}|} \geqslant e^{-O(2^{5j/4})},$$

the last bound here being a consequence of (6.5) with $\varepsilon' := 1/4$.

(ii) $j - 1 \leqslant j' \leqslant j + 1$: Here we have $|e^{z/\rho}| \gg 1$ and (6.7) gives $|1 - z/\rho| \gg 2^{-2j}$, hence

$$|E_\Omega^{(j')}(z)| \geqslant (c2^{-2j})^{|S_{j'}|} \gg e^{-O(2^{3j/2})},$$

this last bound following from (6.5) with any value of $\varepsilon' < \frac{1}{2}$.

(iii) $j' > j + 1$. Note that $|(1 - w)e^w| > e^{-|w|^2}$ for $|w| < 1/2$. In view of the size of j' we may take $|z/\rho| < \frac{1}{2}$, and so

$$|E_\Omega^{(j')}(z)| \geqslant \prod_{\rho \in S_{j'}} e^{-|z/\rho|^2} \geqslant (e^{-2^{2(j+1)-2j'}})^{|S_{j'}|} \geqslant e^{-O(2^{2j}2^{-j'/2})},$$

applying (6.5) with $\varepsilon' = \frac{1}{2}$.

It remains to note that

$$\left| \prod_{j'<j-1} E_\Omega^{(j')}(z) \right| \geqslant \left(e^{-O(2^{5j/4})} \right)^{j-1} = e^{-O(2^{3j/2})}$$

and

$$\Big|\prod_{j'>j+1} E_\Omega^{(j')}(z)\Big| \geqslant \prod_{j'>j+1} e^{-O(2^{2j}2^{-j'/2})} = e^{-O(2^{3j/2})}.$$

Taking the product of these two with case (ii) above and (6.9) gives the required bound (6.6). This concludes the proof. ☐

7. The explicit formula

In this chapter we will take steps towards clarifying the relationship between primes and the zeros of the Riemann ζ-function, by proving the so-called *explicit formula*.

7.1. **Definitions and statement of the formula.** Recall that the von Mangoldt function Λ is defined by

$$\Lambda(n) := \begin{cases} \log p & \text{if } n = p^m \text{ is a prime power} \\ 0 & \text{otherwise.} \end{cases}$$

The fact that $\Lambda(n) \neq 0$ for prime powers p^m as well as for the primes themselves is almost never an issue, since the prime powers are so sparse.

The prime number theorem asserts that $\psi(X) \sim X$, where

$$\psi(X) := \sum_{n \leqslant X} \Lambda(n).$$

Another way of writing this is

$$\psi(X) = \sum_n \Lambda(n) W\big(\frac{n}{X}\big),$$

where $W : \mathbf{R} \to \mathbf{R}$ is $\mathbf{1}_{[0,1]}$, the function defined to equal 1 on $[0,1]$, and zero elsewhere. The explicit formula is a formula for sums like this in terms of the zeros of ζ. However, we will only prove the formula when W is smooth. This means that in order to apply it to the prime number theorem we must approximate $\mathbf{1}_{[0,1]}$ by smooth, compactly supported functions which will take some additional work.

Theorem 7.1 (Explicit formula)**.** *Let W be a smooth, compactly supported function, supported on $[1,\infty)$. Suppose that $X > 1$. Then*

$$\sum_n \Lambda(n) W\big(\frac{n}{X}\big) = X \int_{\mathbf{R}} W(x)\,\mathrm{d}x - \sum_{\rho \in Z} X^\rho \tilde{W}(\rho). \qquad (7.1)$$

The sum here is over the set $Z = Z_{\mathrm{TRIV}} \cup Z_{\mathrm{NONTRIV}}$ of all zeros of ζ.

Remark. Since $\tilde{W}(1) = \int_{\mathbf{R}} W(x)\,\mathrm{d}x$, one could if desired write the right-hand side as

$$-\sum_{w\in\mathbf{C}} \operatorname{ord}_\zeta(w) X^w \tilde{W}(w),$$

where $\operatorname{ord}_\zeta(w)$ is the order of ζ at w, that is to say r if w is a zero of order r, and $-r$ if w is a pole of order r.

The term $X\int_{\mathbf{R}} W(x)\,\mathrm{d}x$ is the "main term" in (7.1); one would expect the sum on the left to equal roughly this, at least if one assumes the prime number theorem, since Λ has average value 1 and $\sum_n W(\frac{n}{X}) \approx X(\int_{\mathbf{R}} W(x)\,\mathrm{d}x)$. The remaining terms $\sum_{\rho\in Z} X^\rho \tilde{W}(\rho)$ are, therefore, to be thought of as "error terms". It turns out that the contribution from $\rho \in Z_{\text{TRIV}}$ is negligible, leaving the sum $\sum_{\rho\in Z_{\text{NONTRIV}}} X^{\operatorname{Re}\rho}\tilde{W}(\rho)$. Let us imagine, for the moment (and we will prove various rigorous assertions about this in the next section) that $\tilde{W}(\rho)$ is on the order of 1. The contribution of this term rather depends on what we know about the location of the nontrivial zeros Z_{NONTRIV}; at the moment, all we know is that they lie in $0 \leqslant \operatorname{Re}\rho \leqslant 1$. For all we know, it could be that some zero ρ lies on the line $\operatorname{Re} s = 1$, in which case one would expect the term $X^{\operatorname{Re}\rho}\tilde{W}(\rho)$ to have the same order of magnitude, as the main term. Thus, once one has the explicit formula, further progress on the prime number theorem depends on pinning down further information about the nontrivial zeros.

7.2. **Proof of the explicit formula – overview.** In this section we outline the proof of the explicit formula, leaving some technical details to the next sections. Recall that the Mellin transform of W is

$$\tilde{W}(s) := \int_0^\infty W(x) x^s \frac{\mathrm{d}x}{x},$$

and that (as in (4.10) in our proof of the Mellin inversion formula (Proposition 4.17)) it can be expressed in terms of the Fourier transform:

$$\tilde{W}(\sigma + it) = \widehat{w\exp^\sigma}(-t),$$

where $w(u) := W(e^u)$ and $\exp^\sigma(u) := e^{\sigma u}$. However, we know from Chapter 4 how to obtain decay estimates for the Fourier transform of smooth functions – see for example Lemma 4.1 – and we can use these to get corresponding estimates for the Mellin transform. We will work out some relevant details of this in Lemma 7.2.

Now we can outline the proof of the explicit formula (Theorem 7.1). The proof begins with the Dirichlet series

$$\sum_n \Lambda(n)n^{-s} = -\frac{\zeta'(s)}{\zeta(s)},$$

valid for $\mathrm{Re}\, s > 1$. We established this in Proposition 3.2 (iii).

By the Mellin inversion formula, Proposition 4.17, and this Dirichlet series expansion we have

$$\begin{aligned}\sum_n \Lambda(n)W\big(\frac{n}{X}\big) &= \sum_n \Lambda(n)\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \tilde{W}(s)(n/X)^{-s}\,\mathrm{d}s\\ &= \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \Big(\sum_n \Lambda(n)n^{-s}\Big)X^s\tilde{W}(s)\,\mathrm{d}s\\ &= \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} -\frac{\zeta'(s)}{\zeta(s)}X^s\tilde{W}(s)\,\mathrm{d}s. \qquad (7.2)\end{aligned}$$

The interchange of summation and integration here is unjustified, but in the course of the proof of the explicit formula (Theorem 7.1) we will effectively do so. For now it provides motivation for the heart of the argument.

The idea is to start with (7.2) and to move the contour of integration far to the left, picking up residues due to the poles of ζ'/ζ at $s = 1$, at the non-trivial zeros $\rho \in Z_{\text{NONTRIV}}$ in the critical strip, and at the trivial zeros $Z_{\text{TRIV}} = \{-2, -4, -6, \dots\}$. To this end

$$\mathrm{Res}_{s=1}\frac{\zeta'}{\zeta} = -1, \mathrm{Res}_{s=-2k}\frac{\zeta'}{\zeta} = 1, \text{ and } \mathrm{Res}_{s=\rho}\frac{\zeta'}{\zeta} = \mathrm{ord}_\zeta(\rho) \qquad (7.3)$$

for $k \in \mathbf{N}$ and $\rho \in Z_{\text{NONTRIV}}$. (As before $\mathrm{ord}_\zeta(\rho)$ is the order of the zero ρ in ζ.) These residues produce the terms on the right hand side of the explicit formula (7.1).

To make all this precise we use a rectangular contour $\mathcal{C}_{k,T} = \mathcal{C}^{(1)}_{k,T} \cup \mathcal{C}^{(2)}_{k,T} \cup \mathcal{C}^{(3)}_{k,T} \cup \mathcal{C}^{(4)}_{k,T}$, where $\mathcal{C}^{(1)}_{k,T} = [2-iT, 2+iT]$, $\mathcal{C}^{(2)}_{k,T} = [2+iT, -2k-1+iT]$, $\mathcal{C}^{(3)}_{k,T} = [-2k-1+iT, -2k-1-iT]$ and $\mathcal{C}^{(4)}_{k,T} = [-2k-1-iT, 2-iT]$, and for $\ell = 1, 2, 3, 4$ write

$$I^{(\ell)}_{k,T} := -\frac{1}{2\pi i}\int_{\mathcal{C}^{(\ell)}_{k,T}} \frac{\zeta'(s)}{\zeta(s)}X^s\tilde{W}(s)\,\mathrm{d}s. \qquad (7.4)$$

We will choose values of k and T tending to infinity in such a way that we can estimate the $I_{k,T}^{(\ell)}$s. We will choose k an integer so that $\mathcal{C}_{k,T}^{(3)}$ avoids the poles of ζ'/ζ at the trivial zeros Z_{TRIV}. The choice of T is harder as we need to make sure that neither $\mathcal{C}_{k,T}^{(2)}$ nor $\mathcal{C}_{k,T}^{(4)}$ pass close to any of the poles of ζ'/ζ at the nontrivial zeros Z_{NONTRIV}. In fact we will produce a sequence of T_js that are "good" in the sense of avoiding these zeros, with $T_j \to \infty$, and then define a corresponding sequence of k_js also tending to infinity. With these choices we will be able to show the following:

$$\lim_{j\to\infty} I_{k_j,T_j}^{(1)} = \sum_n \Lambda(n) W\left(\frac{n}{X}\right) \tag{7.5}$$

(which also gives a rigorous justification of (7.2)) as well as

$$\lim_{j\to\infty} I_{k_j,T_j}^{(2)}, \lim_{j\to\infty} I_{k_j,T_j}^{(4)} = 0 \tag{7.6}$$

and

$$\lim_{j\to\infty} I_{k_j,T_j}^{(3)} = 0. \tag{7.7}$$

On the other hand we can also calculate $\sum_{\ell=1}^{4} I_{k_j,T_j}^{(\ell)}$ using the residue theorem and (7.3), obtaining

$$\sum_{\ell=1}^{4} I_{k_j,T_j}^{(\ell)} = X\tilde{W}(1) - \sum_{\substack{\rho\in Z_{\text{NONTRIV}} \\ |\operatorname{Im}\rho|<T_j}} X^{\rho}\tilde{W}(\rho) - \sum_{m=1}^{k_j} X^{-2m}\tilde{W}(-2m). \tag{7.8}$$

Note that

$$\tilde{W}(1) = \int_{\mathbf{R}} W(x)\,\mathrm{d}x.$$

Therefore as k_j and T_j tend to infinity the right-hand side of (7.8) tends towards the right hand side of the explicit formula (7.1). To conclude the proof of the explicit formula (Theorem 7.1) it suffices to show that we can choose sequences T_j and k_j tending to infinity such that (7.5) to (7.7) hold. Figure 1 illustrates the situation.

7.3. **Estimates for some Mellin transforms.** The following lemma gives a decay estimate for the Mellin transform of a smooth function. In this lemma, we have followed the usual cultural norm, which is that $s = \sigma + it$ denotes a general element of the complex plane, while ρ denotes a zero of ζ which, if ρ is nontrivial, we write as $\beta + i\gamma$.

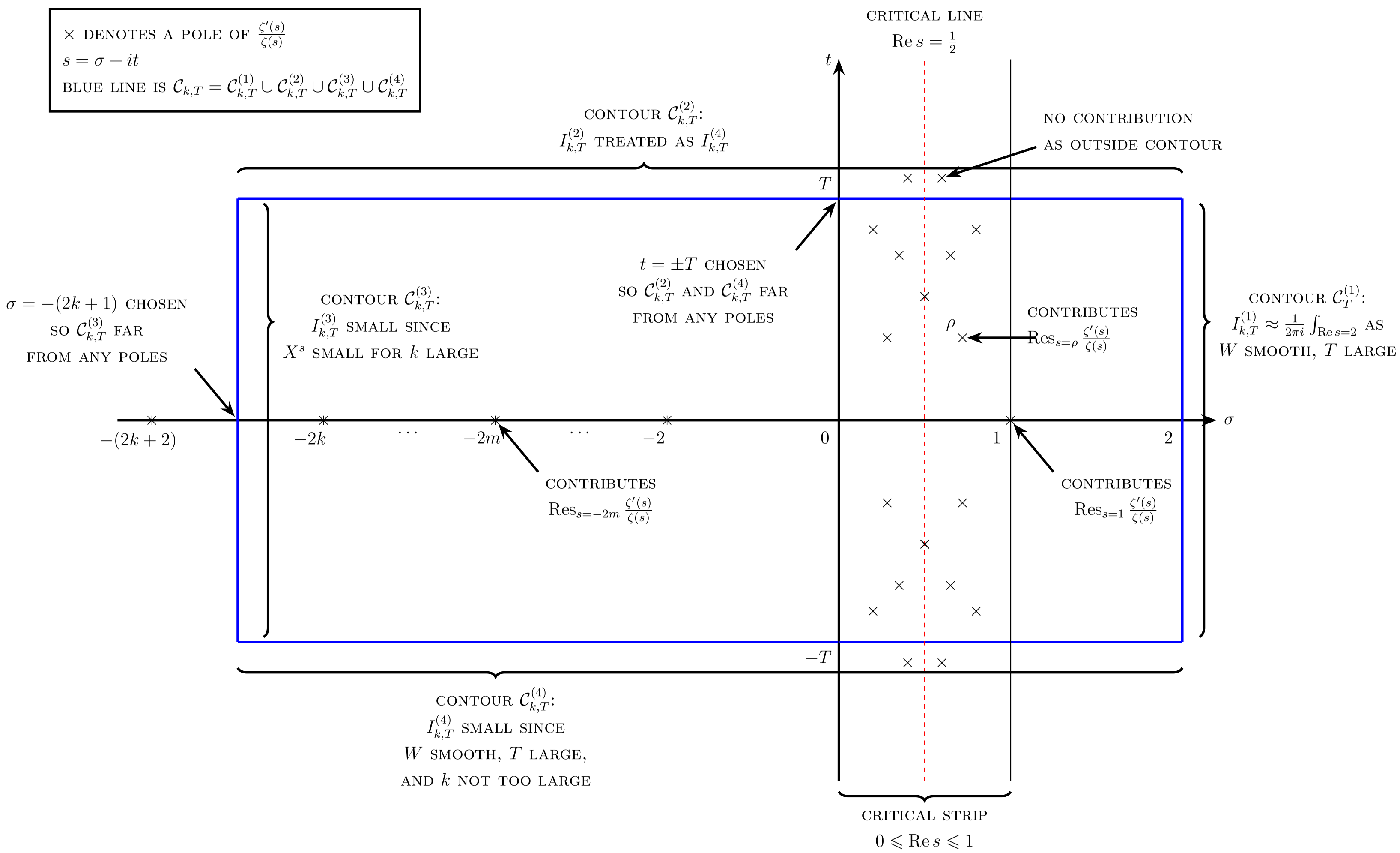

FIGURE 1. The complex plane for computation of $\frac{1}{2\pi i} \int_{\mathrm{Re}\, s=2} \frac{\zeta'(s)}{\zeta(s)} X^s \tilde{W}(s)\, \mathrm{d}s$ in the proof of the explicit formula (Theorem 7.1).

Lemma 7.2. *Suppose that $W : \mathbf{R}_{>0} \to \mathbf{R}$ is smooth and has compact support contained in $[1, \infty)$.*

(i) If $-\infty < \sigma \leqslant 2$, then

$$|\tilde{W}(\sigma + it)| \ll_W 1; \tag{7.9}$$

and, for any integer $m \geqslant 0$,

$$|\tilde{W}(\sigma + it)| \ll_{m,W} (1 + |\sigma|)^m |t|^{-m} \tag{7.10}$$

uniformly for $|t| \geqslant 1$.

(ii) If $\operatorname{supp} W \subseteq [1, 10]$, then

$$|\tilde{W}(\rho)| \ll_m \sup_{0 \leqslant j \leqslant m} \|\partial^j W\|_1 |\rho|^{-m} \tag{7.11}$$

whenever $\rho \in Z_{\text{NONTRIV}}$ has $|\rho| \geqslant 2$; and

$$|\tilde{W}(\rho)| \ll \|W\|_1 \tag{7.12}$$

uniformly for all $\rho \in Z = Z_{\text{TRIV}} \cup Z_{\text{NONTRIV}}$.

Remarks. We have formulated this result in two parts with future applications in mind. Part (i) is the bound we will use in *proving* the explicit formula, whereas (ii) are the bounds we will use in *applying* the explicit formula, where it is important to understand the penalty one must pay when approximating a rough cutoff by a smooth function W.

The bounds are of two types, namely 'trivial' bounds (7.9), (7.12) giving what are essentially very crude constant bounds, and the more sophisticated bounds (7.10), (7.11) which give decay of the Mellin transform in the vertical direction.

Proof. The proof of all parts uses (4.10), that is to say

$$\tilde{W}(\sigma + it) = \widehat{w \exp^{\sigma}}(-t), \tag{7.13}$$

where $w := W \circ \exp$. We also use Lemma 4.1, which states that

$$|\widehat{f}(t)| \leqslant |t|^{-m} \|\partial^m f\|_1 \tag{7.14}$$

for any $f \in \mathcal{S}(\mathbf{R})$. For (7.9) and (7.12), we use the case $m = 0$ of (7.14), or in other words the simple bound $|\widehat{f}(t)| \leqslant \|f\|_1$. This gives

$$\big|\tilde{W}(\sigma+it)\big| \leqslant \int_0^\infty |w(u)|e^{\sigma u}\,\mathrm{d}u \leqslant \begin{cases} \big(\sup(\operatorname{supp}(W))\big)^\sigma \|w\|_1 & \text{if } \sigma \geqslant 0 \\ \|w\|_1 & \text{if } \sigma < 0 \end{cases}. \tag{7.15}$$

The bound (7.9) follows from this since $\sigma \leqslant 2$. Note that the range of integration above may start at 0 because $\operatorname{supp} w \subset [0,\infty)$ since $\operatorname{supp} W \subset [1,\infty)$.

The bound (7.12) follows from (7.15) and

$$\|w\|_1 = \int_0^\infty |w(u)|\,\mathrm{d}u = \int_0^\infty |W(e^u)|\,\mathrm{d}u = \int_1^\infty \frac{|W(x)|}{x}\,\mathrm{d}x \leqslant \|W\|_1.$$

In the case of trivial ρ, we have $\sigma < 0$ and so we are in the second case of (7.15), and hence are done. For nontrivial ρ we have $0 \leqslant \sigma \leqslant 1$ and so are in the first case of (7.15), and moreover

$$\big(\sup(\operatorname{supp}(W))\big)^\sigma \leqslant 10$$

from the hypothesis $\operatorname{supp}(W) \subseteq [1,10]$. Again, we are done.

For the other statements, we use (7.14) with $m \geqslant 1$, which requires us to estimate the derivatives. By the general Leibniz rule (*i.e.* the higher order product rule for differentiation) and the triangle inequality,

$$\begin{aligned} \big\|\partial^m(w\exp^\sigma)\big\|_1 &\ll_m \sup_{0\leqslant j\leqslant m} \big\|(\partial^j w)(\partial^{m-j}\exp^\sigma)\big\|_1 \\ &= \sup_{0\leqslant j\leqslant m} \big\|(\partial^j w)\sigma^{m-j}\exp^\sigma\big\|_1 \\ &\leqslant E_1E_2E_3, \end{aligned}$$

where

$$E_1 := \sup_{0\leqslant j\leqslant m} |\sigma|^{m-j}, \quad E_2 := \sup_{u\in \operatorname{supp} w} e^{\sigma u}, \qquad E_3 := \sup_{0\leqslant j\leqslant m} \|\partial^j w\|_1.$$

For (7.10), we use $E_1 \leqslant (1+|\sigma|)^m$, $E_2 \leqslant (\sup\operatorname{supp} W)^2 \ll_W 1$, and $E_3 \ll_{W,m} 1$. Note that the bound on E_2 again uses the fact that $\operatorname{supp} w \subset [0,\infty)$, so the supremum is maximised when σ is maximised.

For (7.11), set $\rho = \beta + i\gamma$, where $0 \leqslant \beta \leqslant 1$, and so $|\rho| \ll |\gamma|$; therefore it suffices to get bounds in terms of $|\gamma|^{-m}$ (which is what (7.14) gives) rather than $|\rho|^{-m}$. In the definition of E_1, E_2, E_3, set $\sigma = \beta$. We have the bounds $E_1 \leqslant 1$, and $E_2 \ll 1$ since $\operatorname{supp}(w) \subset [0, \log 10]$ and $\beta \leqslant 1$. To complete the proof, we must bound E_3 by showing that

$$\sup_{0 \leqslant j \leqslant m} \|\partial^j w\|_1 \ll_m \sup_{0 \leqslant j \leqslant m} \|\partial^j W\|_1. \tag{7.16}$$

To see this, observe that by Faà di Bruno's formula (*i.e.* the higher order chain rule) $\partial^j w(u)$ is a sum of $O_j(1)$ terms of the form $e^{ku}\partial^l W(e^u)$ where $k, l \geqslant 0$ are integers with $k, l \leqslant j$. But

$$\int_0^{\log 10} e^{ku}\partial^l W(e^u)\, \mathrm{d}u = \int_1^{10} x^{k-1}\partial^l W(x)\, \mathrm{d}x \ll_j \|\partial^l W\|_1,$$

and the result follows. □

7.4. **Bounds for ζ'/ζ at judiciously chosen points.** Now it is time to start thinking about how to choose the ordinates T_j and the corresponding k_j in such a way that the poles of $\zeta'(s)/\zeta(s)$ do not interfere with the horizontal contours $\mathcal{C}^{(2)}_{k_j,T_j}, \mathcal{C}^{(4)}_{k_j,T_j}$. Very crude estimates suffice, ultimately because we are dealing with smooth functions W which enjoy very advantageous Mellin decay properties.

Lemma 7.3. *There is a sequence $T_1, T_2, \dots$ with $T_j \to \infty$ such that*

$$\operatorname{dist}(s, Z) \geqslant 1/T_j \text{ whenever } |\operatorname{Im} s| = T_j.$$

Proof. Suppose that we have defined $T_0 \leqslant T_1 \leqslant \dots \leqslant T_j$ for some $j \in \mathbf{Z}_{\geqslant 0}$ with $T_0 \geqslant 2$ sufficiently large. We will show that there is $T_{j+1} \in [2T_j, 3T_j]$ such that $\operatorname{dist}(s, Z) \geqslant 1/T_{j+1}$ whenever $|\operatorname{Im} s| = T_{j+1}$. This will prove the lemma.

Let $S := \{\operatorname{Im}\rho : \rho \in Z_{\text{NONTRIV}}\}$ and $S_j := S \cap [-4T_j, 4T_j]$. By Proposition 6.8 we have

$$|S_j| \leqslant \#\{\rho \in Z_{\text{NONTRIV}} : |\operatorname{Im}\rho| \leqslant 4T_j\} < C_1 T_j \log T_j,$$

where $C_1 > 1$ is some absolute constant.

For $\gamma \in S_j$ we want to avoid picking T_{j+1} in $[\gamma - \frac{1}{2C_1 \log T_j}, \gamma + \frac{1}{2C_1 \log T_j}]$. (In fact we want to avoid this for all $\gamma \in S$, but we will deal with this in a moment.) The length of each such interval is $\frac{1}{C_1 \log T_j}$ and so the measure of the set to be avoided is $< (C_1 T_j \log T_j) \cdot \frac{1}{C_1 \log T_j} \leqslant T_j$. On the other hand the interval $[2T_j, 3T_j]$ has measure T_j and so there is certainly a T_{j+1} avoiding all these intervals *i.e.* such that $|\gamma - T_{j+1}| \geqslant \frac{1}{2C_1 \log T_j}$ for all $\gamma \in S_j$; pick such a T_{j+1}.

Now, suppose that $|\operatorname{Im} s| = T_{j+1}$. If $\rho \in Z$ has $|\operatorname{Im} \rho| > 4T_j$ then $|s-\rho| > 4T_j - T_{j+1} \geqslant T_j \geqslant 1/T_{j+1}$. If $|\operatorname{Im}\rho| \leqslant 4T_j$, then either ρ is trivial in which case $\operatorname{Im}\rho = 0$ and $|s-\rho| \geqslant |\operatorname{Im} s| \geqslant T_j \geqslant 1/T_{j+1}$; or $\operatorname{Im}\rho \in S_j$, and hence also $-\operatorname{Im}\rho \in S_j$, and so

$$\begin{aligned}|s-\rho| &\geqslant \min\{|T_{j+1} - \operatorname{Im}\rho|, |-T_{j+1} - \operatorname{Im}\rho|\}\\ &= \min\{|T_{j+1} - \operatorname{Im}\rho|, |T_{j+1} + \operatorname{Im}\rho|\} \geqslant \frac{1}{2C_1 \log T_j} \geqslant 1/T_{j+1}.\end{aligned}$$

The last inequality tells us how large to take T_0: take it large enough that $T \geqslant C_1 \log T$ for all $T \geqslant T_0$. □

We say that T is *good* if

$$\operatorname{dist}(s, Z) \geqslant 1/T \text{ whenever } |\operatorname{Im} s| = T,$$

so $T_1, T_2, \dots$ from Lemma 7.3 is a sequence of good positive reals tending to infinity.

If T is good then certainly there are no elements of Z_{NONTRIV}, and hence no poles of ζ'/ζ, on the horizontal lines $\operatorname{Im} s = \pm T$, and hence in particular no such poles on the contours $\mathcal{C}^{(2)}_{k,T}, \mathcal{C}^{(4)}_{k,T}$. We actually need rather more than this, namely a reasonable upper bound for ζ'/ζ on these contours. That said, the following crude bound will do:

Proposition 7.4. *Suppose that $|s| \geqslant 3$. Then*

$$\frac{\zeta'(s)}{\zeta(s)} \ll |s|^3 \operatorname{dist}(s, Z)^{-1},$$

where Z is the set of all zeros of ζ.

Remark. Stronger bounds can be obtained, but for our purposes any bound of the form $|s|^{O(1)} \operatorname{dist}(s, Z)^{-O(1)}$ would be sufficient.

Proof. We use the partial fraction expansion

$$\frac{\zeta'(s)}{\zeta(s)} = C - \frac{1}{s-1} + \sum_{\rho \in Z} \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right)$$

from Proposition 6.1. The first two terms are $O(1)$ in the domain $|s| \geqslant 3$, and hence certainly $O(|s|^3 \operatorname{dist}(s, Z)^{-1})$ in this domain, since $\operatorname{dist}(s, Z) \leqslant \operatorname{dist}(s, -2) = O(|s|)$.

To bound the sum over the zeros, our main ingredient will be

$$\sum_{\rho \in Z} |\rho|^{-2} < \infty, \tag{7.17}$$

also from Proposition 6.1. We will split the sum into two parts. Suppose first that $|\rho| \geqslant 2|s|$. Then $|s/\rho| \leqslant \frac{1}{2}$, and hence $|\frac{s}{\rho} - 1| \geqslant \frac{1}{2}$ and so $|s-\rho| \geqslant \frac{1}{2}|\rho|$. It follows that

$$\Big|\sum_{|\rho|\geqslant 2|s|}\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big)\Big| \leqslant |s|\sum_{|\rho|\geqslant 2|s|}\frac{1}{|s-\rho||\rho|} \leqslant 2|s|\sum_{|\rho|\geqslant 2|s|}\frac{1}{|\rho|^2} \ll |s|;$$

the last step using (7.17). To estimate the contribution from the remaining zeros, those with $|\rho| < 2|s|$, we proceed extremely crudely. By (7.17) it follows that the number of such zeros is $O(|s|^2)$. The contribution from a given one is

$$\frac{|s|}{|s-\rho||\rho|} \ll |s|\operatorname{dist}(s,Z)^{-1},$$

where here we used the fact that $\frac{1}{|\rho|} \ll 1$, which follows from (7.17). Therefore

$$\sum_{|\rho|<2|s|}\Big(\frac{1}{s-\rho}+\frac{1}{\rho}\Big) \ll |s|^3\operatorname{dist}(s,Z)^{-1},$$

and the proposition follows. □

7.5. **Proof of the explicit formula – the details.** We are now ready to fill in the details from our overview in Section 7.2. We choose our sequences of k_j, T_j as follows. Let T_j be any sequence of good values of T with $\lim_{j\to\infty} T_j = \infty$ – the existence of such is ensured by Lemma 7.3. Set $k_j := \lceil T_j^{1/10}\rceil$. Now we complete the proof of the explicit formula by verifying (7.5) to (7.7) with these choices. To avoid notational overload we omit the subscript j, writing $(k,T) = (k_j, T_j)$, and when we talk about the limit we mean as $j \to \infty$.

Proof of (7.5). We begin by recalling the first step of the calculation leading to (7.2), namely

$$\sum_n \Lambda(n)W\big(\frac{n}{X}\big) = \sum_n \Lambda(n)\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty}\tilde{W}(s)(n/X)^{-s}\,\mathrm{d}s. \qquad (7.18)$$

This is just the Mellin inversion formula, Proposition 4.17. Now we truncate the integral to $\mathcal{C}^{(1)}_{k,T}$. Recall from (7.10) that we have the decay estimate $|\tilde{W}(2+it)| \ll_W |t|^{-2}$, so

$$\sum_n \Lambda(n)\frac{1}{2\pi i}\int_{2+iT}^{2+i\infty} \tilde{W}(s)(n/X)^{-s}\,\mathrm{d}s$$
$$\ll_W \sum_n \Lambda(n)(n/X)^{-2}\int_T^{\infty} |t|^{-2}\,\mathrm{d}t \ll_W \frac{X^2}{T}.$$

The last inequality uses that $\sum_n \Lambda(n)n^{-2} = O(1)$ and $\int_T^\infty t^{-2}\,\mathrm{d}t = O(1/T)$. We get a similar estimate for the integrals from $2-iT$ to $2-i\infty$, and so in (7.18) we have

$$\sum_n \Lambda(n)W\big(\frac{n}{X}\big)$$
$$= \sum_n \Lambda(n)\frac{1}{2\pi i}\int_{\mathcal{C}_{k,T}^{(1)}} \tilde{W}(s)(n/X)^{-s}\,\mathrm{d}s + O_W(X^2/T)$$
$$= \frac{1}{2\pi i}\int_{\mathcal{C}_{k,T}^{(1)}} \left(\sum_n \Lambda(n)n^{-s}\right) X^s\tilde{W}(s)\,\mathrm{d}s + O_W(X^2/T)$$
$$= I_{k,T}^{(1)} + O_W(X^2/T).$$

Interchanging the sum and integral is allowed by the dominated convergence theorem (Theorem A.3) because the Dirichlet series $\sum_n \Lambda(n)n^{-s}$ is uniformly bounded by $\sum_n n^{-2}\log n$ on $C_{k,T}^{(1)}$, $X^s\tilde{W}(s)$ is continuous and $\mathcal{C}_{k,T}^{(1)}$ is compact. The last equality is then by Proposition 3.2 (iii).

The statement (7.5) now follows. □

Proof of (7.6). If $s \in \mathcal{C}_{k,T}^{(2)} \cup \mathcal{C}_{k,T}^{(4)}$ then, since T is good, $\mathrm{dist}(s, Z) \gg \frac{1}{T}$. It follows from Proposition 7.4 that, for $s \in \mathcal{C}_{k,T}^{(2)} \cup \mathcal{C}_{k,T}^{(4)}$, we have

$$\frac{\zeta'(s)}{\zeta(s)} \ll |s|^3 T \ll k^3T^4.$$

Hence

$$I_{k,T}^{(2)}, I_{k,T}^{(4)} \ll k^3T^4\int_{-2k-1}^{2} X^\sigma|\tilde{W}(\sigma+iT)|\,\mathrm{d}\sigma.$$

The bound (7.10) with $m=5$ tells us that

$$|\tilde{W}(\sigma+iT)| \ll_W (1+|\sigma|)^5T^{-5},$$

and therefore

$$I_{k,T}^{(2)}, I_{k,T}^{(4)} \ll_W k^3T^{-1}\int_{-2k-1}^{2}(1+|\sigma|)^5X^\sigma d\sigma \ll_W X^2k^9T^{-1},$$

since $X \geqslant 1$. Since $k = \lceil T^{1/10}\rceil$, statement (7.6) follows. □

Proof of (7.7). If $s \in \mathcal{C}^{(3)}_{k,T}$, say $s = -(2k+1) + it$, then $\operatorname{dist}(s, Z) \geqslant 1$ and $|s| \geqslant 3$, so Proposition 7.4 gives $\frac{\zeta'(s)}{\zeta(s)} \ll |s|^3 \ll k^3 \max(1, |t|)^3$.

From (7.10) with $m = 5$ (and (7.9) for $|t| \leqslant 1$) we have

$$|\tilde{W}(-2k-1+it)| \ll_W k^5 \max(1, |t|)^{-5}.$$

Thus

$$\frac{\zeta'(s)}{\zeta(s)} X^s \tilde{W}(s) \ll_W k^8 \max(1, |t|)^{-2} X^{-2k-1},$$

which implies that

$$I^{(3)}_{k,T} = -\frac{1}{2\pi i} \int_{\mathcal{C}^{(3)}_{k,T}} \frac{\zeta'(s)}{\zeta(s)} X^s \tilde{W}(s)\, \mathrm{d}s \ll_W k^8 X^{-2k-1}.$$

From this and the fact that $X > 1$ we immediately conclude (7.7). (For the estimation of $I^{(3)}_{k,T}$ there was no need to restrict to T good.) $\square$

8. The prime number theorem

In the last chapter we established the explicit formula, which provides a relationship between the distribution of prime numbers and the nontrivial zeros ρ of the ζ-function. To make use of this result, and in particular to prove the prime number theorem, it is important to have some information about the location of those zeros.

The prime number theorem turns out to be more-or-less equivalent to the following statement.

Proposition 8.1. *There are no zeros of ζ on the line $\operatorname{Re} s = 1$.*

Proof. Consider the Euler product identity

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1},$$

valid for $\operatorname{Re} s > 1$. Set $s = \sigma + it$; we will let $\sigma \to 1$. For $0 \leqslant r < 1$ we have

$$\log|1 - re^{i\theta}| = -\sum_{n=1}^{\infty} \frac{r^n \cos n\theta}{n},$$

and hence for $\sigma > 1$ we have

$$\begin{aligned}
&\log|\zeta(\sigma)|^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)|\\
&\qquad = -\sum_p \big(3\log|1-p^{-\sigma}| + 4\log|1-p^{-\sigma-it}| + \log|1-p^{-\sigma-2it}|\big)\\
&\qquad = \sum_p \sum_{n=1}^{\infty} \frac{1}{np^{\sigma n}}(3+4\cos n\theta + \cos 2n\theta)\\
&\qquad = \sum_p \sum_{n=1}^{\infty} \frac{1}{np^{\sigma n}}(1+\cos n\theta)^2 \geqslant 0.
\end{aligned}$$

Exponentiating we conclude that

$$\zeta(\sigma)^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \geqslant 1. \tag{8.1}$$

Now suppose that $\zeta(1+it)=0$. Certainly $t \neq 0$. Now consider what happens as $\sigma \to 1$. We have $\zeta(\sigma) \sim (\sigma-1)^{-1}$, but $|\zeta(\sigma+it)|^4 \ll (\sigma-1)^4$ (by Taylor expansion about the putative zero at $1+it$). Since $\zeta(\sigma+2it)$ remains bounded as $\sigma \to 1$, we have

$$\zeta(\sigma)^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \to 0,$$

a contradiction to (8.1). □

We are now in a position to prove the following 'smoothed' version of the prime number theorem, from which the theorem itself will be a relatively easy deduction.

Proposition 8.2 (Smooth prime number theorem)**.** *Suppose that* $W : \mathbf{R} \to \mathbf{R}$ *is smooth and that* $\operatorname{supp}(W) \subset [1,10]$. *Then*

$$\lim_{X\to\infty} \frac{1}{X}\sum_n \Lambda(n) W\Big(\frac{n}{X}\Big) = \int_{\mathbf{R}} W(x)\,dx.$$

Proof. By the explicit formula (Theorem 7.1) it suffices to show that

$$\lim_{X\to\infty} \sum_{\rho\in Z} X^{\rho-1}\tilde{W}(\rho) = 0.$$

We handle the contribution from the trivial zeros and the nontrivial zeros separately. By (7.12) we have $\tilde{W}(\rho) \ll_W 1$ for $\rho \in Z_{\text{TRIV}}$, and so

$$\sum_{\rho\in Z_{\text{TRIV}}} X^{\rho-1}\tilde{W}(\rho) \ll_W \sum_{j=1}^{\infty} X^{-2j-1}. \tag{8.2}$$

This certainly tends to 0 as $X \to \infty$.

Turning now to the contribution from the nontrivial zeros. We will show that for all $\varepsilon > 0$ if X is large enough (in terms of ε and W) then

$$\Big| \sum_{\rho \in Z_{\text{NONTRIV}}} X^{\rho-1} \tilde{W}(\rho) \Big| < \varepsilon. \tag{8.3}$$

With this the result will be proved. Let $\varepsilon > 0$. From (7.11) with $m = 2$ we have $\tilde{W}(\rho) \ll_W |\rho|^{-2}$ for $\rho \in Z_{\text{NONTRIV}}$ and $|\rho| \geqslant 2$. Moreover, $|X^{\rho-1}| \leqslant 1$ since $\operatorname{Re} \rho \leqslant 1$. Hence for $K \geqslant 2$,

$$\left| \sum_{\rho \in Z_{\text{NONTRIV}} : |\rho| \geqslant K} X^{\rho-1} \tilde{W}(\rho) \right| \ll_W \sum_{\rho \in Z_{\text{NONTRIV}} : |\rho| \geqslant K} |\rho|^{-2}.$$

From the "moreover" in the partial fraction expansion (Proposition 6.1) we know $\sum_{\rho \in Z} |\rho|^{-2} < \infty$, and hence there is K (depending on W and ε) such that

$$\left| \sum_{\rho \in Z_{\text{NONTRIV}} : |\rho| \geqslant K} X^{\rho-1} \tilde{W}(\rho) \right| < \frac{1}{2}\varepsilon.$$

For the nontrivial zeros with $|\rho| \leqslant K$ it follows from (7.9) and the fact that there are only finitely many such zeros that

$$\sum_{\rho \in Z_{\text{NONTRIV}} : |\rho| < K} X^{\rho-1} \tilde{W}(\rho) \ll_{W,K} X^{\beta(K)-1}, \tag{8.4}$$

where $\beta(K) = \sup_{\rho \in Z_{\text{NONTRIV}} : |\rho| \leqslant K} \operatorname{Re} \rho$. Furthermore, since there are only finitely many nontrivial zeros ρ with $|\rho| \leqslant K$, the supremum in $\beta(K)$ is a maximum and since there are no zeros on $\operatorname{Re} s = 1$ (Proposition 8.1), we have $\beta(K) < 1$. It follows that if X is big enough, in terms of W, K, and ε then (8.4) is at most $\frac{1}{2}\varepsilon$. It follows that if X is large enough in terms of W and ε (since K just depends on these) then (8.3) holds and the result is proved. $\square$

Proof of the prime number theorem (Theorem 1.3). Let $\varepsilon \in (0, \frac{1}{4}]$ and W^+ and W^- be smooth such that $0 \leqslant W^-(x) \leqslant \mathbf{1}_{[2,4]}(x) \leqslant W^+(x) \leqslant \mathbf{1}_{[2-2\varepsilon, 4+4\varepsilon]}$ for all x, and

$$\int_{\mathbf{R}} W^+(x) - W^-(x) \, \mathrm{d}x \leqslant 4\varepsilon.$$

Figure 2 depicts this.

A construction of such functions is given in Lemma B.2 in the appendix. We need W^+ for the upper bound, and W^- can be used for the

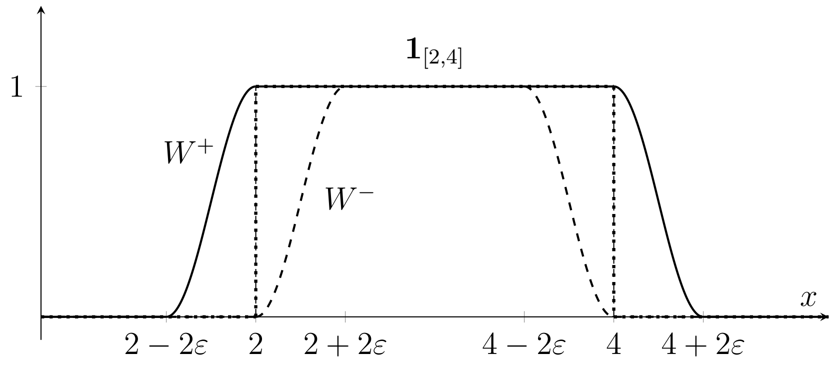

FIGURE 2. The smooth functions W^+ and W^-.

lower bound. Applying the smooth prime number theorem (Proposition 8.2) to W^+ and W^- (which is admissible since $\operatorname{supp} W^- \subset \operatorname{supp} W^+ \subset [1,5]$) tells us that if Y is sufficiently large in terms of ε then

$$\begin{aligned}\sum_{2Y<n\leqslant 4Y}\Lambda(n) &\leqslant \sum_n \Lambda(n)W^+\Big(\frac{n}{Y}\Big)\\ &\leqslant \Big(\varepsilon+\int_{\mathbf{R}} W^+(x)\,\mathrm{d}x\Big)Y \leqslant \Big(1+\frac{5\varepsilon}{2}\Big)2Y,\end{aligned}$$

and

$$\begin{aligned}\sum_{2Y<n\leqslant 4Y}\Lambda(n) &\geqslant \sum_n \Lambda(n)W^-\Big(\frac{n}{Y}\Big)\\ &\geqslant \Big(-\varepsilon+\int_{\mathbf{R}} W^-(x)\,\mathrm{d}x\Big)Y \geqslant \Big(1-\frac{5\varepsilon}{2}\Big)2Y.\end{aligned}$$

In particular, for Y large enough in terms of ε we have

$$\Big|\sum_{2Y<n\leqslant 4Y}\Lambda(n)-2Y\Big| \leqslant \frac{5\varepsilon}{2}\cdot 2Y. \tag{8.5}$$

Applying this with $Y = X/4, X/8, \dots, X/2^{m+1}$ and summing, we see that if X is sufficiently large in terms of ε and m then

$$\Big|\sum_{X/2^m<n\leqslant X}\Lambda(n)-X\Big| \leqslant \frac{5\varepsilon}{2}X+\frac{X}{2^m}. \tag{8.6}$$

Finally, by Proposition 1.5 we have

$$\sum_{n\leqslant X/2^m}\Lambda(n) = O(X/2^m),$$

and so

$$\left|\sum_{n\leqslant X}\Lambda(n)-X\right|\leqslant\frac{5\varepsilon}{2}X+O(X/2^m).$$

By choosing m sufficiently large in terms of ε, we can ensure that the second term is at most $\frac{1}{2}\varepsilon X$. Therefore, if X is big enough in terms of ε, we have

$$\left|\sum_{n\leqslant X}\Lambda(n)-X\right|\leqslant 3\varepsilon X.$$

In other words

$$\sum_{n\leqslant X}\Lambda(n)=(1+o(1))X,$$

which is equivalent to the prime number theorem by Proposition 1.4. □

9. The zero-free region and error terms

9.1. **The classical zero-free region.** In the last chapter we showed that ζ has no zeros on the line $\operatorname{Re}s=1$, and we saw that this implies the prime number theorem.

In this section we obtain, using a related argument, a region to the left of $\operatorname{Re}s=1$ in which there are no zeros of ζ. This may be used to prove a stronger version of the prime number theorem, with a good error term: we provide details of this in the next section.

The following is known as the *classical zero-free region.*

Theorem 9.1. *There is an absolute constant $c>0$ such that any zero $\rho=\beta+i\gamma$ with $|\gamma|\geqslant 2$ satisfies*

$$\beta<1-\frac{c}{\log|\gamma|}.$$

Proof. We showed in Proposition 3.2 (iii) that for $\operatorname{Re}s>1$ we have

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

Taking real parts gives, provided $\sigma>1$,

$$-\operatorname{Re}\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)}=\sum_{n=1}^{\infty}\Lambda(n)n^{-\sigma}\cos(t\log n).$$

Using the inequality

$$3+4\cos\theta+\cos 2\theta=2(1+\cos\theta)^2\geqslant 0$$

(exactly as in the last chapter) it follows that

$$-3\,\mathrm{Re}\,\frac{\zeta'(\sigma)}{\zeta(\sigma)} - 4\,\mathrm{Re}\,\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} - \mathrm{Re}\,\frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)} \geqslant 0 \tag{9.1}$$

for any $\sigma > 1$ and for any $t \in \mathbf{R}$.

Now we assemble some further inequalities. Since $\frac{\zeta'(s)}{\zeta(s)} + \frac{1}{s-1}$ is continuous at 1, and holomorphic for $\sigma > 1$, it is continuous on the compact interval $[1, 2]$ and so bounded. That is, if $1 < \sigma \leqslant 2$ then

$$-3\,\mathrm{Re}\,\frac{\zeta'(\sigma)}{\zeta(\sigma)} \leqslant \frac{3}{\sigma-1} + O(1). \tag{9.2}$$

To get bounds for the other two terms in (9.1), we use the partial fraction expansion (Proposition 6.1)

$$\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} = B - \frac{1}{\sigma+it-1} + \sum_{\rho\in Z}\Big(\frac{1}{\sigma+it-\rho} + \frac{1}{\rho}\Big).$$

From now on, suppose $1 \leqslant \sigma \leqslant 2$ and that $|t| \geqslant 2$. Then the first two terms are bounded by $O(1)$. To bound the sum over ρ, we treat the trivial zeros separately. If $1 \leqslant \sigma \leqslant 2$ and $|t| \geqslant 2$, the contribution from the trivial zeros is

$$\sum_k \sum_{\rho=-2k} \frac{\sigma+it}{\rho(\sigma+it-\rho)} \ll \sum_{k\geqslant 1} \frac{|t|}{k\sqrt{k^2+t^2}}.$$

To estimate this sum, we split it into $k \leqslant |t|$ and $k \geqslant |t|$. On the first range the summand is $O(1/k)$, so we get a contribution of $O(\log|t|)$. On the second range the summand is $O(|t|/k^2)$, and so we get a contribution of $O(1)$. It follows that

$$-4\,\mathrm{Re}\,\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} = O(\log|t|) - 4\,\mathrm{Re}\sum_{\rho\in Z_{\text{NONTRIV}}}\Big(\frac{1}{\sigma+it-\rho} + \frac{1}{\rho}\Big).$$

By an essentially identical analysis we also have

$$-\,\mathrm{Re}\,\frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)} = O(\log|t|) - \mathrm{Re}\sum_{\rho\in Z_{\text{NONTRIV}}}\Big(\frac{1}{\sigma+2it-\rho} + \frac{1}{\rho}\Big).$$

Combining all this information, we obtain

$$4\,\mathrm{Re}\sum_{\rho\in Z_{\text{NONTRIV}}}\Big(\frac{1}{\sigma+it-\rho} + \frac{1}{\rho}\Big) + \mathrm{Re}\sum_{\rho\in Z_{\text{NONTRIV}}}\Big(\frac{1}{\sigma+2it-\rho} + \frac{1}{\rho}\Big) \leqslant C\log|t| + \frac{3}{\sigma-1} \tag{9.3}$$

for some C.

If $z \in \mathbf{C}$ has $\operatorname{Re} z > 0$ then $\operatorname{Re} z^{-1} > 0$. Since there are no zeros on the line $\operatorname{Re} s = 1$ (Proposition 8.1), $\sigma > 1 > \operatorname{Re} \rho > 0$ for every $\rho \in Z_{\text{NONTRIV}}$, and hence every term on the left here is non-negative. Ignoring all terms except the contribution of the zero $\rho = \beta + i\gamma$ to the leftmost sum, and setting $t = \gamma$, we obtain

$$\frac{4}{\sigma - \beta} \leqslant C \log |\gamma| + \frac{3}{\sigma - 1}.$$

Setting $\sigma = 1 + 1/2C \log |\gamma|$ and rearranging then gives

$$\beta \leqslant 1 - \frac{1}{14C \log |\gamma|},$$

completing the proof. □

Remark. The classical zero-free region is not the largest one known. In fact, the stronger inequality

$$\beta < 1 - \frac{c_\varepsilon}{\log^{2/3+\varepsilon} |\gamma|}$$

is known to hold for any $\varepsilon > 0$, where $c_\varepsilon > 0$ is a constant that depends only on ε. This is a result of Vinogradov and Korobov.

9.2. **The prime number theorem with classical error term.** In this section we examine the implications of the classical zero-free region for the prime number theorem.

The main result is the following.

Theorem 9.2. *We have* $\psi(X) = X + O(Xe^{-c\sqrt{\log X}})$.

Here, $\psi(X) = \sum_{n \leqslant X} \Lambda(n)$, as usual (1.3). The letter c denotes a positive absolute constant, but it may vary from line to line.

The proof is very closely related to that in Chapter 8, except now we must be a little more precise in how we approximate $\mathbf{1}_{[2,4]}$ by smooth functions.

Proof. Let $\varepsilon \in (0, 1/2]$, and take W_ε to be W^+ from Lemma B.2 so that $\mathbf{1}_{[2,4]} \leqslant W_\varepsilon \leqslant 1$, and

$$\int_{\mathbf{R}} W_\varepsilon(x)\, \mathrm{d}x \leqslant 2 + 4\varepsilon, \|\partial W_\varepsilon\|_1 \ll 1, \text{ and } \|\partial^2 W_\varepsilon\|_1 \ll \frac{1}{\varepsilon}. \tag{9.4}$$

In what follows let Y be a sufficiently large parameter. By the explicit formula (Theorem 7.1) and the properties of W_ε we have

$$\begin{aligned}\psi(4Y)-\psi(2Y) &\leqslant \sum_n \Lambda(n)W_\varepsilon\Big(\frac{n}{Y}\Big) = Y\int_{\mathbf{R}} W_\varepsilon(x)\,\mathrm{d}x - \sum_{\rho\in Z} Y^\rho \tilde{W}_\varepsilon(\rho)\\ &\leqslant 2Y(1+2\varepsilon) - \sum_{\rho\in Z} Y^\rho\tilde{W}_\varepsilon(\rho)\\ &\leqslant 2Y(1+2\varepsilon) + \sum_{\rho\in Z}|Y^\rho||\tilde{W}_\varepsilon(\rho)|. \end{aligned} \tag{9.5}$$

We split the sum here into four using three different estimates for $\tilde{W}_\varepsilon(\rho)$. First, from (7.12) we have $\tilde{W}_\varepsilon(\rho)=O(1)$ for $\rho\in Z_{\text{TRIV}}\cup Z_{\text{NONTRIV}}$; secondly, from (7.11) with $m=1$ and $m=2$ and (9.4) we have $\tilde{W}_\varepsilon(\rho)\ll|\rho|^{-1}$ and $\tilde{W}_\varepsilon(\rho)\ll\varepsilon^{-1}|\rho|^{-2}$ respectively for $\rho\in Z_{\text{NONTRIV}}$. These then give the following:

$$\begin{aligned}&\sum_{\rho\in Z}|Y^\rho||\tilde{W}_\varepsilon(\rho)|\\ &\ll \underbrace{\sum_{\rho\in Z_{\text{TRIV}}}|Y^\rho|}_{S_1} + \underbrace{\sum_{\substack{\rho\in Z_{\text{NONTRIV}}\\ |\operatorname{Im}\rho|\leqslant 2}}|Y^\rho|}_{S_2} + \underbrace{\sum_{\substack{\rho\in Z_{\text{NONTRIV}}\\ 2\leqslant|\operatorname{Im}\rho|\leqslant T}}\frac{|Y^\rho|}{|\rho|}}_{S_3} + \underbrace{\varepsilon^{-1}\sum_{\substack{\rho\in Z_{\text{NONTRIV}}\\ |\operatorname{Im}\rho|\geqslant T}}\frac{|Y^\rho|}{|\rho|^2}}_{S_4},\end{aligned}$$

for any T; we will explain how to optimise T now. First, $S_1=O(Y^{-2})$, and contributes essentially nothing. For S_2 we know that ζ has finitely many nontrivial zeros with $|\operatorname{Im}\rho|\leqslant 2$, and since there are no zeros on $\operatorname{Re}s=1$ (Proposition 8.1) there is some absolute $c>0$ such that $S_2=O(Y^{1-c})$. For S_3 and S_4 we need to estimate sums of powers of nontrivial zeros: write

$$N(t):=\#\{\rho\in Z_{\text{NONTRIV}}:|\operatorname{Im}\rho|\leqslant t\}=\sum_{\rho\in Z_{\text{NONTRIV}}}\mathbf{1}_{\{|\operatorname{Im}\rho|\leqslant t\}}.$$

Then by Proposition 6.8 we have $N(t) = O(t \log t)$, and

$$\begin{aligned}\sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ |\operatorname{Im}\rho| \geqslant T}} \frac{1}{|\rho|^2} &\leqslant \sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ |\operatorname{Im}\rho| \geqslant T}} \frac{1}{|\operatorname{Im}\rho|^2} \\ &= 2 \sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ |\operatorname{Im}\rho| \geqslant T}} \int_{|\operatorname{Im}\rho|}^{\infty} t^{-3}\,\mathrm{d}t \\ &= 2 \int_T^{\infty} \sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ t \geqslant |\operatorname{Im}\rho| \geqslant T}} t^{-3}\,\mathrm{d}t \leqslant 2 \int_T^{\infty} \frac{N(t)}{t^3}\,\mathrm{d}t = O(T^{-1}\log T).\end{aligned}$$

We also have

$$\begin{aligned}\sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ 2 \leqslant |\operatorname{Im}\rho| \leqslant T}} \frac{1}{|\rho|} &\leqslant 2 \sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ 2 \leqslant |\operatorname{Im}\rho| \leqslant T}} \int_{|\operatorname{Im}\rho|}^{2T} t^{-2}\,\mathrm{d}t \\ &= 2 \int_2^{2T} \sum_{\substack{\rho \in Z_{\text{NONTRIV}} \\ \min\{t,T\} \geqslant |\operatorname{Im}\rho| \geqslant 2}} t^{-2}\,\mathrm{d}t \\ &\leqslant 2 \int_2^{2T} \frac{N(t)}{t^2}\,\mathrm{d}t = O(\log^2 T).\end{aligned}$$

When $\rho = \beta + i\gamma \in Z_{\text{NONTRIV}}$ has $2 \leqslant |\gamma| \leqslant T$ we have $\beta \leqslant 1 - c'/\log T$ in view of the classical zero-free region (Theorem 9.1), and hence $|Y^\rho| \leqslant Y^{1-c'/\log T}$. Here $c' > 0$ is an absolute constant.

In summary we conclude that

$$\begin{aligned}\psi(4Y) - \psi(2Y) &\leqslant 2Y(1+2\varepsilon) + O(Y^{-2}) + O(Y^{1-c}) \\ &\qquad + O\left(Y^{1-c'/\log T}\log^2 T\right) + O\left(\frac{\varepsilon^{-1} Y \log T}{T}\right).\end{aligned}$$

Setting $\varepsilon = \exp(-\sqrt{\log Y})$ and $T = \varepsilon^{-2}$ the error term on the right hand side here is $O(Y\exp(-c''\sqrt{\log Y}))$, for $c'' > 0$ yet another absolute constant.

Finally a telescoping sum argument (like we used to establish (8.6) from (8.5)), using the above with $X/4$, $X/8, \dots,$ gives

$$\psi(X) \leqslant X + O(Xe^{-c''\sqrt{\log X}}).$$

By similar arguments using W^- from Lemma B.2 instead of W^+, we may obtain the corresponding lower bound

$$\psi(X) \geqslant X - O(Xe^{-c''\sqrt{\log X}}),$$

which completes the proof. □

9.3. **The Riemann hypothesis and its implications for primes.** The Riemann hypothesis is the assertion that all the nontrivial zeros lie on the *critical line* $\operatorname{Re} s = \frac{1}{2}$.

If it holds then in the above argument,

$$S_2 = O(Y^{1/2}), S_3 = O(Y^{1/2}\log^2 T), S_4 = O\left(\frac{\varepsilon^{-1}Y^{1/2}\log T}{T}\right).$$

It follows that

$$\psi(4Y) - \psi(2Y) \leqslant 2Y + 4\varepsilon Y + O(Y^{-2}) + O(Y^{1/2}) \\ + O(Y^{1/2}\log^2 T) + O\left(\frac{\varepsilon^{-1}Y^{1/2}\log T}{T}\right).$$

There is quite a bit of flexibility here; taking $T = Y$ and $\varepsilon = Y^{-1}$ gives that the right hand side is $2Y + O(Y^{1/2}\log^2 Y)$. As before a telescoping sum and similar lower bound argument can then be used to show

$$\psi(X) = X + O(X^{1/2}\log^2 X).$$

Appendix A. Some analysis results

In this appendix we collect some results from analysis which we have needed. We begin with a basic inequality that is useful for controlling products:

Lemma A.1. *Let $(z_n)_{n=1}^N$ be a sequence of complex numbers. Then*

$$\Big|\prod_{n=1}^N (1+z_n) - 1\Big| \leqslant e^{\sum_{n=1}^N |z_n|} - 1.$$

Proof. We should like to use the fact that $1 + x \leqslant e^x$ for $x \in \mathbf{R}$. To do this we first expand out the product and then take the modulus:

$$\begin{aligned}\Big|\prod_{n=1}^N (1+z_n) - 1\Big| &= \Big|\sum_{i_1} z_{i_1} + \sum_{i_1<i_2} z_{i_1}z_{i_2} + \sum_{i_1<i_2<i_3} z_{i_1}z_{i_2}z_{i_3}\dots\Big| \\ &\leqslant \sum_{i_1}|z_{i_1}| + \sum_{i_1<i_2}|z_{i_1}||z_{i_2}| + \sum_{i_1<i_2<i_3}|z_{i_1}||z_{i_2}||z_{i_3}|\dots \\ &= \prod_{n=1}^N (1+|z_n|) - 1 \\ &\leqslant e^{\sum_i |z_i|} - 1.\end{aligned}$$

This concludes the proof. □

We will also find the following inequality about exponentials useful:

Lemma A.2. *Suppose that $h \in \mathbf{C}$ and $t > 0$. Then*
$$|t^h - 1 - h\log t| \leqslant \frac{1}{2}|h|^2(\log t)^2 \max\{1, t^{\operatorname{Re} h}\}.$$

Proof. Note the identity
$$e^z - 1 - z = z^2 \int_0^1 (1-u)e^{uz}\,\mathrm{d}u.$$
Setting $z = h\log t$ we have
$$\begin{aligned}|t^h - 1 - h\log t| &\leqslant |h|^2(\log t)^2 \int_0^1 (1-u)e^{u\operatorname{Re} h \log t}\,\mathrm{d}u\\ &\leqslant |h|^2(\log t)^2\max\{1, t^{\operatorname{Re} h}\}\int_0^1 (1-u)\,\mathrm{d}u.\end{aligned}$$
The lemma is proved. □

We make use of two results which refer to measurable functions. An important feature of measurable functions is that the pointwise limit of a sequence of measurable functions is measurable. This is to repair the situation with Riemann integrable functions where such a limit can fail to be Riemann integrable (and not only because the integrals become large).

All of the functions we consider in these notes are measurable.

Theorem A.3 (Dominated convergence theorem)**.** *Suppose that we have a sequence $f_1, f_2, \dots : \mathbf{R} \to \mathbf{C}$ of measurable functions converging pointwise to a function f, and such that $|f_n(x)| \leqslant g(x)$ for every x, and some $g \in L_1(\mathbf{R})$. Then $f_1, f_2, \dots, f \in L_1(\mathbf{R})$ and*
$$\int_{\mathbf{R}} f_n(x)\,\mathrm{d}x \to \int_{\mathbf{R}} f(x)\,\mathrm{d}x.$$

This is closely related to Fubini's theorem:

Theorem A.4 (Fubini's theorem)**.** *Suppose that $f : \mathbf{R}^2 \to \mathbf{C}$ is measurable and*
$$\int_{\mathbf{R}}\int_{\mathbf{R}} |f(x,y)|\,\mathrm{d}x\,\mathrm{d}y < \infty.$$
Then
$$\int_{\mathbf{R}}\int_{\mathbf{R}} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbf{R}}\int_{\mathbf{R}} f(x,y)\,\mathrm{d}y\,\mathrm{d}x.$$

The only situation where we use this is for continuous functions which also have all but ε of their mass supported on a compact set (for all $\varepsilon > 0$). In this setting this can be proved by a fairly direct approximation argument.

Appendix B. Smooth bump functions

Many of the examples of smooth functions that we initially think of – functions like the exponential function or polynomials – do not have compact support. This is because they are real analytic, meaning that in some neighbourhood of every point they have a convergent power series. Just like holomorphic functions on $\mathbf{C}$, real analytic functions satisfy the identity theorem and so can't have compact support (without being identically 0).

We are left with the question of whether there are any smooth functions with compact support. There are, and one quick way to see this is on Exercise Sheet 3. The example there has a rabbit-out-of-hat quality because it asks you to consider a particular function which can seem to come out of nowhere. It can be motivated, but in this section we take a different approach.

Our idea here is to construct a function f as an infinite convolution of normalised characteristic functions of intervals. The convolution of two functions f and g, $f * g$, is defined by

$$(f * g)(x) := \int f(x-y)g(y)\,\mathrm{d}y.$$

Convolution is an associative and commutative binary operation, and the map $g \mapsto f * g$ is linear in g. Moreover, the triangle inequality tells us that

$$\|f * g\|_\infty \leqslant \|f\|_\infty \|g\|_1. \tag{B.1}$$

This inequality is a special case of Young's inequality, and for us, coupled with linearity it tells us that if $f_n \to f$ uniformly then $f_n * g \to f * g$ uniformly for any probability density function g.

The normalised characteristic function of $[-\delta, \delta]$ is $\nu_\delta(x) := \frac{1}{2\delta}\mathbf{1}_{|x|\leqslant\delta}$. The fact they are normalised exactly means that these functions ν_δ are probability density functions. We will be looking at convolutions of functions with these. Importantly the fundamental theorem of calculus gives that

$$\partial(\nu_\delta * g)(x) = \frac{1}{2\delta}(g(x+\delta) - g(x-\delta)). \tag{B.2}$$

Writing $C^k(\mathbf{R})$ for the space of k-times continuously differentiable functions, we therefore (also using linearity of convolution and the fact that $C^k(\mathbf{R})$ is a vector space) have

$$\nu_\delta * f \in C^{k+1}(\mathbf{R}) \text{ whenever } f \in C^k(\mathbf{R}). \tag{B.3}$$

In view of this we would like to take an "infinite" convolution, but need to ensure that it has compact support. To do this we have one final observation: since $\operatorname{supp} f * g \subset \operatorname{supp} f + \operatorname{supp} g$ we have

$$\operatorname{supp} \nu_{\delta_1} * \cdots * \nu_{\delta_n} \subset [-(\delta_1 + \cdots + \delta_n), \delta_1 + \cdots + \delta_n]. \tag{B.4}$$

This leads us to the idea of taking smaller and smaller intervals such that $\delta_1 + \cdots + \delta_n$ converges.

Lemma B.1 (A bump function)**.** *There is a smooth function $W : \mathbf{R} \to \mathbf{R}$ that is non-negative, is supported in $[-1,1]$, and has $\int_{\mathbf{R}} W(x)\,\mathrm{d}x = 1$.*

Proof. Fix f continuous, non-negative, supported in $[-\frac{1}{2}, \frac{1}{2}]$, and with $\int_{\mathbf{R}} f(x)\,\mathrm{d}x = 1$. For example the tent function $f = \nu_{\frac{1}{4}} * \nu_{\frac{1}{4}}$. Let

$$g_n := f * \nu_{2^{-2}} * \cdots * \nu_{2^{-n}}.$$

Since f is continuous and compactly supported it is uniformly continuous and so for all $\varepsilon > 0$ there is $\delta > 0$ such that if $|x-y| < \delta$ then $|f(x) - f(y)| < \varepsilon$. Suppose that $2^{-n} < \delta$. Then for $n < m$ we have (B.1)

$$\begin{aligned}\|g_n - g_m\|_\infty &\leqslant \|(f - f * \nu_{2^{-(n+1)}} * \cdots * \nu_{2^{-m}}) * \nu_{2^{-2}} * \cdots * \nu_{2^{-n}}\|_\infty \\ &\leqslant \|f - f * \nu_{2^{-(n+1)}} * \cdots * \nu_{2^{-m}}\|_\infty.\end{aligned}$$

Now $\operatorname{supp} \nu_{2^{-(n+1)}} * \cdots * \nu_{2^{-m}} \subset (-\delta, \delta)$ by (B.4) and the fact that $2^{-(n+1)} + \cdots + 2^{-m} < 2^{-n} < \delta$. Hence $\nu_{2^{-(n+1)}} * \cdots * \nu_{2^{-m}}$ is a probability density function supported in $(-\delta, \delta)$ and so $f * \nu_{2^{-(n+1)}} * \cdots * \nu_{2^{-m}}(x)$ is a weighted average of f over the interval $(x-\delta, x+\delta)$, and hence within ε of $f(x)$ by the uniform continuity above. We conclude that

$$\|g_n - g_m\|_\infty \leqslant \varepsilon.$$

It follows that the sequence g_n is uniformly Cauchy and so converges uniformly to some function W. Similarly there are functions h_k for $k \geqslant 2$ such that $f * \nu_{2^{-(k+1)}} * \cdots * \nu_{2^{-n}}$ converges uniformly to h_k. By linearity of convolution and (B.1) we have

$$h_k * \nu_{2^{-2}} * \cdots * \nu_{2^{-k}} = W.$$

Given this, (B.3) tells us that $W \in C^{k-1}(\mathbf{R})$ for all $k \geqslant 2$. We also have $W \geqslant 0$, $\operatorname{supp} W \subset [-1,1]$ and

$$\begin{aligned}\left|\int_{\mathbf{R}} W(x)\,\mathrm{d}x - \int_{\mathbf{R}} g_n(x)\,\mathrm{d}x\right| &= \left|\int_{-1}^{1} (W(x) - g_n(x))\,\mathrm{d}x\right|\\ &\leqslant 2\|W - g_n\|_\infty \to 0,\end{aligned}$$

so $\int_{\mathbf{R}} W(x)\,\mathrm{d}x = \int_{\mathbf{R}} f(x)\,\mathrm{d}x = 1$. The convergence of the integral follows since $\operatorname{supp} g_n \subset [-1,1]$ is compact and uniform convergence on a compact set implies convergence of integrals. The lemma is proved. $\square$

Once we have one bump function, as above, we can dilate it and use it to create smooth approximations to lots of functions.

A function f is said to be a majorant for g if $g(x) \leqslant f(x)$ for all $x \in \mathbf{R}$; and a minorant for g if $f(x) \leqslant g(x)$ for all $x \in \mathbf{R}$.

Lemma B.2 (Smooth majorants and minorants of intervals)**.** *For all $\varepsilon \in (0,1]$ there are smooth functions W^+ and W^- from $\mathbf{R}$ to $\mathbf{R}$, such that*

(i) $0 \leqslant W^- \leqslant 1_{[2,4]} \leqslant W^+ \leqslant 1_{[2-2\varepsilon,4+4\varepsilon]}$;

(ii)
$$\int_{\mathbf{R}} W^+(x) - W^-(x)\,\mathrm{d}x \leqslant 4\varepsilon;$$

(iii) and
$$\|\partial^k W^\pm\|_1 = O_k(\varepsilon^{-(k-1)}) \textit{ for } k \in \mathbf{N}.$$

Proof. The picture here is as in Figure 2. Let W be as in Lemma B.1 and $W_\varepsilon(x) := \varepsilon^{-1}W(\varepsilon^{-1}x)$. By change of variables W_ε is a probability density function. We set $W^- := 1_{[2+\varepsilon,4-\varepsilon]} * W_\varepsilon$ and $W^+ := 1_{[2-\varepsilon,4+\varepsilon]} * W_\varepsilon$. These functions certainly satisfy (i) and (ii) (in fact stronger for (ii) where there is equality). By (B.2) (or an unnormalised variant for the interval $[2-\varepsilon, 4+\varepsilon]$) we have $\partial W^+(x) = W_\varepsilon(x - (2-\varepsilon)) -$

$W_\varepsilon(x-(4+\varepsilon))$. Now $\partial^{k-1}W_\varepsilon = \varepsilon^{-k}W^{(k-1)}(\varepsilon^{-1}x)$. Hence by the triangle inequality and (B.1)

$$\|\partial^k W^+\|_1 \leqslant 2\|\partial^{(k-1)}W_\varepsilon\|_1 = O_k(\varepsilon^{-(k-1)}).$$

Similarly for W^-, and the lemma is proved. □

References

[Dav00] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. doi:10.1007/978-1-4757-5927-3. Revised and with a preface by Hugh L. Montgomery.

[IK04] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004. doi:10.1090/coll/053.

[Kou19] D. Koukoulopoulos. *The Distribution of Prime Numbers*, volume 203 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2019. doi:10.1090/gsm/203.