ANALYTIC NUMBER THEORY

BEN GREEN

Contents

Overview. These are notes for a first course in analytic number theory, most recently taught in Michaelmas 2024 as C3.8 at the University of Oxford. They have evolved over around 20 years from various courses I have given at Oxford and Cambridge.

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 throughout. This has many advantages: most particularly, one can state a very clean 'explicit formula', and the handling of the error terms in the prime number theorem is arguably clearer. The main disadvantage is perhaps that proceeding in this way requires a little more mathematical maturity on the part of the reader.

Students should note that in recent years the course C3.8 has been taught by Prof. Maynard, who takes a slightly different approach. Therefore, questions on past exam papers from 2019–2023 may have parts which are not quite accessible using only these notes. The course and questions in 2017 and 2018 were set by me and are relevant to this version of the course. Finally, in 2016 and earlier the course was taught by Prof. Heath–Brown, and at that time the syllabus was rather different.

0.1. **Notation.** As with any course, a certain amount of notation will be introduced as we go along.

One very important point to be made at the outset is that $\log x$ *always* means the natural logarithm of x. Some people consider the notation $\ln x$, which can be found in some books, tasteless. $\log^C x$ is the same as $(\log x)^C$.

We will very often use the notation $\lfloor t \rfloor$, which means the greatest integer less than or equal to t.

Throughout the course we will be using *asymptotic notation.* This is vital in handling the many inequalities and rough estimates we will encounter. Here is a summary of the notation we will see. We suggest the reader not worry too much about this now; we will gain plenty of practice with this notation.

- $A \ll B$ means that there is an absolute constant $C > 0$ such that $|A| \leqslant CB$. In this notation, A and B will typically be variable quantities, depending on some other parameter. For example, $x + 1 \ll x$ for $x \geqslant 1$, because $|x+1| \leqslant 2x$ in this range. It is important to note that the constant C may be different in different instances of the notation.
- $A = O(B)$ means the same thing.
- $B \gg A$ means the same thing.

The three pieces of notation above are the crucial ones to get used to for now, but in due course we will also see the following:

- $A \ll_{k,l,m} B$ means that $|A| \leqslant CB$, but now C is allowed to depend on some other parameters k, l, m. For example, $kx \ll_k x$, $k+l+m = O_{k,l,m}(1)$.
- $A \ll B$ is the same as $B \gg A$.
- $O(A)$ means some quantity bounded in magnitude by CA for some absolute constant $C > 0$. In particular, $O(1)$ simply means a quantity bounded by an absolute positive constant. For example, $\frac{5x}{1+x} = O(1)$ for $x \geqslant 0$.
- $A = o(B)$ means that $|A| \leqslant \varepsilon B$ as some other parameter becomes large enough. The other parameter will usually be clear from context. For example, $\frac{1}{\log x} = o(1)$ (as $x \to \infty$).
- $A = o_{k,l,m}(B)$ means that $|A| \leqslant \varepsilon B$ as some other parameter becomes large enough, but how large it needs to be may depend on the other parameters k, l, m. For example, $\frac{k}{\log x} = o_k(1)$.
- $A \sim B$, which we read as "A is asymptotic to B" means that $A = (1 + o(1))B$.

0.2. **Quantities.** 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.

0.3. **Further reading.** The notes are self-contained, but anyone with a wider interest in the subject will want to read other sources. Here are a few suggestions. Davenport's book [1] is the classic reference, and it is still an excellent read. Kowalski's book [2] (in French) uses smooth weights rather more like the presentation here. The treatise of Iwaniec and Kowalski [3] is essential for anyone seriously committed to analytic number theory, but it contains an order of magnitude more material than this course, and presents at a high level. Finally, the relatively recent book of Koukoulopoulos [4] is a good read (again covering a lot more material than we do).

1. Basic facts about the primes

1.1. **Euclid's proof.** This course is largely about the *prime numbers* 2, 3, 5, 7, ... The most basic fact about them is the following, proven by Euclid over 2000 years ago.

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. □

Of course, it is possible to ask more refined questions. Does the sequence of prime numbers grow extremely rapidly (like the powers of two $1, 2, 2^2, 2^3, \dots$), or slowly like the odd numbers $1, 3, 5, 7, \dots$, 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. Rather 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, we may thus assert that $\frac{X}{\log X} \ll \pi(X) \ll \frac{X}{\log X}$, and the upper bound is saying that $\pi(X) = O(X/\log X)$. The lower bound immediately implies the infinitude of primes (and with a much better bound than Euclid's proof). 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

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

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. (We remark that 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

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

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

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

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 sufficiently large,

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

Taking logs gives

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

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 $p > n$ 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 $n = 2, 2^2, 2^3 \dots, 2^{k-1}$ and summing gives

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

We claim that the bracketed expression is $\ll 1/k$. 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

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

Therefore $\pi(2^k) \leqslant C2^k/k$ for some C.

If now X is arbitrary, 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* $\pi(X) \sim \frac{X}{\log X}$.

We now give a reformulation of the prime number theorem using a weight function, which turns out to me 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.

The reason this is a natural definition will become gradually apparent. For now, however, let us note that the prime number theorem has an equivalent formulation in terms of the von Mangoldt function.

We define

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

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

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

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

$$b(X) \leqslant a(X). \tag{1.3}$$

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

$$a(X) - b(X) \leqslant \frac{1}{X} \sum_{p \leqslant X} (\log X - \log p), \tag{1.4}$$

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

$$\frac{1}{X} \sum_{p \leqslant X(\log X)^{-2}} (\log X - \log p) < \frac{1}{\log X} \leqslant \frac{1}{c_1 \log X} a(X). \tag{1.5}$$

On the second range, we have

$$\log X - \log p \leqslant \log\big((\log X)^2\big) \leqslant 2 \log\log X. \tag{1.6}$$

Thus

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

Combining this with (1.4) and (1.5) gives

$$a(X) - b(X) \leqslant \Big(\frac{1}{c_1 \log X} + \frac{2 \log\log X}{\log X}\Big) a(X).$$

This last inequality implies that $b(X) \geqslant (1-\varepsilon)a(X)$ provided that X is large enough in terms of ε. Combined with (1.3), this immediately implies the result. □

To conclude this chapter, let us record an immediate consequence of (1.3) 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 section. However, this is far from the only interesting arithmetic function. Here are some other commonly occurring arithmetical functions:

- The von Mangoldt function Λ;
- The Möbius function μ is defined by $\mu(n) = (-1)^k$ if $n = p_1 \cdots p_k$ for distinct primes $p_1, \dots, p_k$, 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 arithmetical functions then we define

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

Multiplicativity. We say that an arithmetic function f is multiplicative if $f(mn) = f(m)f(n)$ whenever m, n are coprime. (Note carefully that 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 is certainly part of the general culture of analytic number theory, and helps us place Λ in a more general context.

The following result is known as *Möbius inversion.*

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

Proof. Observe that Dirichlet convolution is commutative and associative. Note that $\mu \star 1 = \delta$, where $\delta(n) = 1$ if $n = 1$ and 0 otherwise. This is a routine check: if $n = p_1^{\alpha_1} \dots p_k^{a_k}$ 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

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

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 $k \log p$ 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 arithmetical 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 arithmetical function. Then the *Dirichlet Series* of f is

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

At the moment this is just a "formal" Dirichlet series; if f grows extremely rapidly, the series may not make sense as an actual number for any value of s at all. In practice, f will have reasonable growth. Commonly, for example, $|f(n)| = n^{o(1)}$: 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 is not completely obvious). In this case the series for $D_f(s)$ converges, and defines a holomorphic function of s, in the domain $\operatorname{Re} s > 1$.

Indeed if $\operatorname{Re} s \geqslant 1 + \delta$ then for $N > N_0(\delta)$ sufficiently large we have

$$\Big| \sum_{n=N+1}^{\infty} f(n) n^{-s} \Big| \leqslant \sum_{n=N+1}^{\infty} n^{\delta/2} n^{-1-\delta} \ll N^{-\delta/2}.$$

Thus if we set

$$D_f^{(N)}(s) := \sum_{n=1}^{N} f(n) n^{-s}$$

then $D_f^{(N)}(s) \rightarrow D_f(s)$ uniformly in $\operatorname{Re} s \geqslant 1 + \delta$, which implies that D_f is holomorphic on the interior of any such domain.

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

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

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

Euler product. 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

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}. \tag{3.1}$$

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 (3.1).

Equation (3.1) is called the Euler product for ζ and it is valid for $\operatorname{Re} s > 1$.

An amusing consequence of the above is yet another proof (our third) that there are infinitely many primes. Suppose not. Then, letting $s \to 1^+$ in (3.1) we see that $\lim_{s\to 1^+} \zeta(s) < \infty$. However this is not the case, since

$$\lim_{s \to 1^+} \sum_{n=1}^{N} n^{-s} = \sum_{n=1}^{N} n^{-1},$$

and the harmonic series diverges.

It turns out that the Dirichlet series of many of the basic arithmetical 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.

Proposition 3.1. *We have the following facts about Dirichlet series.*

(i) *If the Dirichlet series of f is $F(s)$, and that of g is $G(s)$, then the Dirichlet series of the convolution $f \star g$ is FG.*
(ii) *If $\operatorname{Re} s > 1$ then $\sum_n \mu(n) n^{-s} = 1/\zeta(s)$;*
(iii) *If $\operatorname{Re} s > 1$ then $\sum_n \Lambda(n) n^{-s} = -\zeta'(s)/\zeta(s)$.*

Proof. For (i), note that the Dirichlet Series of $f \star g$ is

$$\sum_n \Big(\sum_{ab=n} f(a)g(b) \Big) n^{-s} = \sum_{a,b} f(a)g(b)(ab)^{-s}.$$

This proves the result.

For (ii), write $D(s)$ for the Dirichlet series of the Möbius function. Since $\mu * 1 = \delta$, it follows from (i) that $D(s)\zeta(s) = 1$, which is the stated result.

For (iii), we use without detailed proof the fact that the Dirichlet series for $\zeta(s)$ may be differentiated term by term in the domain $\operatorname{Re} s > 1$. This gives

$$\zeta'(s) = -\sum_n \log n \cdot n^{-s};$$

that is, the Dirichlet series of log is $-\zeta$. Since $\Lambda = \mu \star \log$, it follows from this and (i), (ii) that the Dirichlet series of Λ is indeed $-\zeta'/\zeta$, as claimed. $\square$

Remark. *To justify differentiating the series, it would probably be best to define $F_N(s) := \sum_{n=1}^N n^{-s}$, so $F_N'(s) = -\sum_{n=1}^N \log n \cdot n^{-s}$. Then, note that $F_N'(s)$ converges uniformly on $\mathrm{Re}\, s \geqslant 1 + \delta$ to some holomorphic function $g(s)$, which is given by the series $-\sum_{n=1}^\infty \log n \cdot n^{-s}$. Now define $F(s) := \int_{[2 \to s]} g(w)dw$, where $[2 \to s]$ means the straight line segment from 2 to s. By the fundamental theorem of calculus, $F'(s) = g(s)$. On the other hand, $\int_{[2\to s]} F_N'(w)dw = F_N(s) - F_N(2)$, so by the uniform convergence we have

$$\zeta(s) - \zeta(2) = \lim_{N\to\infty}(F_N(s) - F_N(2)) = \lim_{N\to\infty}\int_{[2\to s]} F_N'(w)dw$$
$$= \int_{[2\to s]} \lim_{N\to\infty} F_N'(w)dw = F(s).$$

Differentiating gives $\zeta'(s) = F'(s) = g(s)$. For more details on this kind of argument in the real-variable case, you can consult my undergraduate analysis notes, Proposition 5.1.*

Corollary 3.2. *We have $\zeta(s) \neq 0$ when $\mathrm{Re}\, s > 1$.*

Proof. This follows immediately from (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 $\mathrm{Re}\, s > 1$, 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* (holomorphic on the whole complex plane). 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 $0 \leqslant \mathrm{Re}\, s \leqslant 1$. To understand ζ in terms of its zeros, it is natural (in the light of ones experience 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 it is not quite possible to do this, a slight variant of it does hold.

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 3.3. *The Riemann ζ-function extends to a meromorphic function, analytic 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 *$0 \leqslant \mathrm{Re}\, s \leqslant 1$. Finally, we have the Weierstrass product expansion*

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

where the product is over all zeros of ζ and A, B are some constants (which can be evaluated explicitly).

Let us remark that at least one of the facts used in the proof of this theorem – the functional equation, Theorem 5.2 – is of great importance in its own right. As a further remark, 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 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** $\mathrm{Re}\, s > 0$**.** To conclude this introductory discussion of the ζ-function, we give a quick proof that it may be meromorphically continued a little to the left of its current domain of definition.

Proposition 3.4. *The Riemann zeta function* $\zeta(s)$ *has a meromorphic continuation to the right half-plane* $\mathrm{Re}(s) > 0$*, holomorphic except for a simple pole at* $s = 1$*.*

Proof. For $\mathrm{Re}\, s > 1$ we have

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

(Here, $\{x\} := x - \lfloor x \rfloor$.) The integral here defines a holomorphic function on $\mathrm{Re}\, s > 0$ by differentiating under the integral; see Proposition A.3 for more details. □

The formula for $\zeta(s)$ found here will sometimes be useful in its own right.

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. If $f : G \to \mathbf{C}$ is a "suitably nice" function and $\gamma \in \widehat{G}$ is a character then the Fourier transform $\widehat{f}(\gamma)$ is defined to be

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

where μ_G is the *Haar measure* on G, a certain uniquely-defined measure with natural properties. 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. If $f : \mathbf{R} \to \mathbf{C}$ is integrable then we define

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

(Stricly speaking, this is an abuse of notation: the Fourier transform is really defined "at the character $x \mapsto e^{i\xi x}$", and not at ξ.)

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 $\widehat{\mathbf{Z}} = \mathbf{R}/\mathbf{Z}$. If $f : \mathbf{Z} \to \mathbf{C}$ is suitably nice we define

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

(This is again an abuse of notation, similar to the previous one.)

Example ($G = \mathbf{R}/\mathbf{Z}$: "Fourier series"). Here all the characters have the form $\theta \mapsto e^{2\pi i n\theta}$, where $n \in \mathbf{Z}$. Thus $\widehat{\mathbf{R}/\mathbf{Z}} = \mathbf{Z}$. If $f : \mathbf{R}/\mathbf{Z} \to \mathbf{C}$ is suitably nice we define

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

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 $x \mapsto e^{2\pi i r x/N}$, where $r \in \mathbf{Z}/N\mathbf{Z}$ (thus $\widehat{G} \cong G$). In this finite setting all functions are "suitably nice" and we define

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

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

Example ($G = \mathbf{R}^{\times}_{>0}$: Mellin transform). In fact, G is isomorphic to $\mathbf{R}$ via the logarithm map $x \mapsto \log x$. The characters are $x \mapsto x^{i\xi}$ for $\xi \in \mathbf{R}$. We have

$$\hat{f}(\xi) = \int_0^{\infty} f(x)x^{-i\xi} d^{\times}x,$$

where $d^{\times}x = \frac{dx}{x}$.

The Mellin transform will feature prominently later on. We will need to extend its domain of definition from $i\mathbf{R}$ to all of $\mathbf{C}$, defining

$$\tilde{f}(s) := \int_0^\infty f(x)x^s d^\times x$$

(we recover the previous definition on taking $s = -i\xi$). Often we will be look at a fixed vertical contour $s = \sigma + i\mathbf{R}$, in which case the Mellin transform really is a Fourier transform, in fact of the function $f(x)x^{-\sigma}$ on $\mathbf{R}^\times_{\geqslant 0}$.

Much of the theory of Fourier analysis is concerned with what is meant by "suitably nice", and 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 Schwarz 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 $\|f\|_1 := \int_{-\infty}^\infty |f(x)|dx$.

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

Lemma 4.1. *Suppose that $g : \mathbf{R} \to \mathbf{R}$ is a smooth function, all of whose derivatives lie in $L^1(\mathbf{R})$ (that is, are integrable) and decay at infinity. Then for any m 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 $|\hat{g}(\xi)| \ll_m |\xi|^{-m}$ for all m.

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

$$\hat{g}(\xi) := \int_{-\infty}^\infty g(x)e^{-i\xi x}dx.$$

Repeated integration by parts gives

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

The result then follows immediately from the triangle inequality. □

We now introduce a proper notion of "sufficiently nice" function which is appropriate to Fourier analysis on $\mathbf{R}$. This is the notion of a *Schwartz function*, which is a smooth function, all of whose derivatives decay rapidly at infinity.

Definition 4.2. Let $f \in C^\infty(\mathbf{R})$, that is to say suppose that f is infinitely differentiable. 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 $m, n \geqslant 0$.

We note in particular that, 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 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. This follows easily by differentiating $x^n f$ repeatedly using the product rule; each term is of the form $x^{n'}\partial^j f$ for some $n' \leqslant n$ and $j \geqslant 0$, and every such term decays quicker than any power of x.

Now differentiation under the integral in the definition of Fourier transform furnishes the following relation:

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

Since $x^n f$ is a Schwarz 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 $\hat{f} \in \mathcal{S}(\mathbf{R})$. □

Using Fubini's theorem, which is always valid when dealing with Schwartz functions, we can deduce the following important property of the Fourier transform.

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

$$\int_{\mathbf{R}} f\widehat{g} = \int_{\mathbf{R}} \widehat{f} g.$$

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 5.4.

Lemma 4.5. *Suppose that $t \in \mathbf{R}^+$, 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

$$\widehat{f}(\xi) = e^{-\xi^2/4\pi} \int_{-\infty}^{\infty} e^{-\pi(x + \frac{i\xi}{2\pi})^2}\, dx = e^{-\xi^2/4\pi} \int_{\Gamma_1} f(z)\, dz, \tag{4.1}$$

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 $\mathbb{C}$) 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 $\int_{\Gamma_{2,R}} f(z)\,dz$ and $\int_{\Gamma_{4,R}} f(z)\,dz$ both tend to 0 as $R \to \infty$, and so by Cauchy's theorem

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

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

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

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

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

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

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*

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

Remark. The factor 2π is a manifestation of the fact that we have two copies of $\mathbf{R}$ here, namely $\mathbf{R}$ and $\widehat{\mathbf{R}}$. In our definition of the Fourier transform we have implicitly given an isomorphism from $\mathbf{R}$ to $\widehat{\mathbf{R}}$. Under this isomorphism Lebesgue measure on $\mathbf{R}$ gets multiplied by a factor of 2π.

Proof. It suffices to establish this 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 $\hat{g}(\xi) = e^{ix\xi}\hat{f}(\xi)$, and so by applying the inversion formula at 0 to g yields

$$f(x) = g(0) = \int_{-\infty}^{\infty}\hat{g}(\xi)d\xi = \int_{-\infty}^{\infty}\hat{f}(\xi)e^{ix\xi}d\xi.$$

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

Using Lemma 4.4 we obtain

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

Taking limits as $\epsilon \to 0$, we see that

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

(To justify this rigorously, note the bound

$$\Big|\int_{-\infty}^{\infty}\widehat{f}(\xi)(1-e^{-\xi^2\epsilon^2/4\pi})\,d\xi\Big| \leqslant \|\hat{f}\|_1 \sup_{|\xi|\leqslant M}\big|1-e^{-\xi^2\epsilon^2/4\pi}\big|\,d\xi + 2\int_{|\xi|>M}|\hat{f}(\xi)|\,d\xi$$

for any M. The second integral tends to 0 as $M \to \infty$. For fixed M, $\sup_{|\xi| \leqslant M} \left|1 - e^{-\xi^2\epsilon^2/4\pi}\right| \to 0$ as $\varepsilon \to 0$.)

It suffices, then, to show that

$$f(0) = \lim_{\epsilon \to 0} \int_{-\infty}^{\infty} f(u) g_\epsilon(u)\, du.$$

If one thinks of g_ε as a spike around 0 with integral 1 (which it is!) then this seems eminently reasonable. Let us give a rigorous proof. Since $\int g_\epsilon = 1$, we see that it suffices to show that

$$\lim_{\epsilon \to 0} \int_{-\infty}^{\infty} (f(u) - f(0)) g_\epsilon(u)\, du = 0.$$

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

$$2\|f\|_\infty \int_{|u| \geqslant \sqrt{\varepsilon}} g_\epsilon(u) du.$$

Note that

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

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

$$\Big| \int_{-\infty}^{\infty} (f(u) - f(0)) g_\varepsilon(u)\, du \Big| \leqslant \sup_{|u| \leqslant \sqrt{\varepsilon}} |f(0) - f(u)| + \|f\|_\infty o_{\varepsilon \to 0}(1).$$

(where $o_{\varepsilon \to 0}(1)$ just means some quantity tending to 0 as $\varepsilon \to 0$, whose exact nature or speed of decay is irrelevant). Since f is continuous at 0, this indeed tends to zero as $\varepsilon \to 0$. □

Remark. 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*

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

Proof. Take $g := \overline{\widehat{f}}$ in Lemma 4.4, which stated that

$$\int_{-\infty}^{\infty} f\widehat{g} = \int_{-\infty}^{\infty} \widehat{f} g.$$

The right hand side is $\int |\widehat{f}|^2$. Meanwhile

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

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 same as $C^\infty(\mathbf{R}/\mathbf{Z})$, the space of smooth functions on $\mathbf{R}/\mathbf{Z}$. (There is no "decay at infinity" condition for this group, which 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. (There is no "smoothness" condition for this group, which is discrete.)

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

$$\widehat{f}(n) := \int_0^1 f(\theta) e^{-2\pi i n\theta}\, d\theta \tag{4.2}$$

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

$$\widehat{f}(\theta) = \sum_{n\in\mathbf{Z}} f(n) e^{-2\pi i\theta n} \tag{4.3}$$

for $\theta \in \mathbf{R}/\mathbf{Z}$. Although we are using the hat symbol in two rather different contexts at once, little confusion should hopefully result so long as the reader is careful to note which group each function under consideration is defined upon.

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. The first statement is fairly immediate by integration by parts. Specifically, we have

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

Thus (applying the preceding with k replaced by $k+1$) we have

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

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

For the second statement, let $r \in \{0, 1, 2, \dots\}$. We apply Corollary A.2 (or, more accurately, the comments immediately after it), taking

$$u_n(\theta) = (-2\pi i n)^r f(n) e^{-2\pi i\theta n}$$

and $u_n'(\theta) = (-2\pi i n)^{r+1} f(n) e^{-2\pi i \theta n}$. The conditions of that lemma are satisfied with $M_n = O_r(|n|^{-2})$ for $|n| \geqslant 1$, by the assumption that $f \in \mathcal{S}(\mathbf{Z})$.

The case $r = 0$ implies that $\hat{f}(\theta)$ is differentiable and its derivative is given by

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

Inductive applications then give that all derivatives exist and

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

This concludes the proof. □

Once again an application of Fubini's theorem (in this case an interchange of integration and summation) allows one to conclude the following, which is Plancherel's formula in this setting.

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

$$\sum_n \widehat{f}(n) g(n) = \int_0^1 f(\theta) \widehat{g}(\theta)\, d\theta.$$

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*

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

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

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

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 $N \in \mathbf{N}$,

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

By direct computation one may see that

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

A more conceptual way to establish this is to observe that $g_N = N h_N * h_N^\circ$, where $h_N(n) := N^{-1}1_{0 \leqslant n \leqslant N-1}$, h_N° is the 'opposite' function to h_N defined by $h_N^\circ(x) = h_N(x)$, and $*$ is the convolution on $\mathbf{Z}$ defined by

$$h_N * h_N^\circ(x) := \sum_y h_N(y)h_N^\circ(x - y).$$

The Fourier transform of g_N is then just the absolute value of the square of $\widehat{h}_N$. (With reference to the proof of Proposition 4.6, we note that the gaussian g_ϵ is also the convolution of a function with itself, that function being a gaussian of half the width.)

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

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

(Later on, it is notationally convenient to take the range of integration over $[-1/2, 1/2]$; 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

$$\lim_{N\to\infty}\int_{-1/2}^{1/2} f(\theta)K_N(\theta)\,d\theta = f(0). \tag{4.4}$$

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

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,

$$\int_{-1/2}^{1/2} K_N(\theta)d\theta = \int_{-1/2}^{1/2} \hat{g}_N(\theta)d\theta = \int_{-1/2}^{1/2} \sum_n g_N(n)e^{-2\pi i n\theta} d\theta$$
$$= \sum_n g_N(n) \int_{-1/2}^{1/2} e^{-2\pi i n\theta} d\theta = g_N(0) = 1.$$

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

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

Returning to the proof of (4.4), observe that from (ii) we have

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

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

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

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

$$\sup_{|\theta| \leqslant \delta} |f(\theta) - f(0)|. \tag{4.5}$$

The integral over the second range is bounded by

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

which, by (iii), is at most

$$\ll \|f\|_\infty \frac{1}{N\delta}. \tag{4.6}$$

Taking $N = \lceil 1/\delta^2 \rceil$ (say) and letting $\delta \to 0$, we see that both (4.5) and (4.6) tend to 0. This concludes the proof of Theorem 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 $F, G : \mathbf{R}/\mathbf{Z} \to \mathbb{C}$, defined by

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

and

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

Due to our conditions on f one may check that F and G lie in $\mathcal{S}(\mathbf{R}/\mathbf{Z})$. In both cases, the derivatives are given by term-by-term differentiation of the series, thus

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

and

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

We will compute the Fourier coefficients of F and G, the aim being to show that these are equal. On the one hand we have, for $m \in \mathbf{Z}$,

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

On the other hand, we have

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

which is also equal to $\widehat{f}(2\pi m)$. Thus indeed $\widehat{F}(n) = \widehat{G}(n)$, which by Corollary 4.13 implies that $F = G$ identically. In particular $F(0) = G(0)$, which is exactly the Poisson summation formula. □

5. The analytic continuation and functional equation

5.1. The Γ-function. Define the Γ-function by

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1}\, dt. \qquad (5.1)$$

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.

In Proposition 5.1 below, we establish the basic facts about Γ. Before doing this we record a result from complex analysis connected with "Weierstrass products". For a proof (not examinable) of this result, see Appendix C.

Proposition C.2 (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 (i.e. holomorphic on the whole complex plane), and has zeros at Ω with the correct multiplicities and nowhere else.

Of course, the proposition seems eminently reasonable and if Ω were finite it would be trivial. The factors of $e^{z/\rho}$ ensure that the product converges; note that $(1-w)e^w \approx 1 - w^2 \approx e^{-w^2}$ for small w, and so one may guess that convergence of the product is intimately tied to convergence of $\sum |\rho|^{-2}$.

Proposition 5.1. *We have the following basic facts about the* Γ*-function.*

(i) $\Gamma(z)$*, as defined by* (5.1)*, 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*

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

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

Remark. One can easily establish that γ exists by comparing the sum $\sum_{j=1}^n \frac{1}{j}$ with the integral $\int_1^n dx/x$. We have $\gamma = 0.577215665...$ Euler's constant γ is rather mysterious, and in fact it is not even known to be irrational.

Proof. (i) For this we 'differentiate under the integral'. A rigorous formulation of what we need is Proposition A.3. This should be applied with $T = (c, \infty)$ for $c > 0$. The required condition (A.1) is then that

$$\int_0^\infty e^{-t} t^{\operatorname{Re} z - 1} dt < \infty$$

for all z with $\operatorname{Re} z > c$. This is easily established by considering the integrals between $0 \leqslant t \leqslant 1$ and $1 \leqslant t < \infty$ separately. It follows from Proposition A.3 that $\Gamma(z)$ is differentiable on $\operatorname{Re} z > c$ (with derivative $\int_0^\infty e^{-t} t^{z-1} \log t dt$). Since c was arbitrary, the same holds in $\operatorname{Re} z > 0$.

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

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

This relation may be used to extend Γ meromorphically to the entire 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 $\operatorname{Re} z > -n$, $n = 2, 3, \dots$. Note that in the process of extending to $\operatorname{Re} z > -1$ we introduce a pole at $z = 0$. 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

$$\frac{1}{\widetilde{\Gamma}(z)} := z e^{\gamma z} \prod_{n=1}^\infty (1 + z/n) e^{-z/n}. \tag{5.2}$$

By Proposition C.2, $\frac{1}{\widetilde{\Gamma}}$ is entire and has zeros only at $0, -1, -2, \dots$. Therefore $\widetilde{\Gamma}$ is meromorphic on $\mathbb{C}$ (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 it follows that if $\operatorname{Re} z > 0$ then

$$\lim_{m\to\infty} \int_0^\infty f_m(t) t^{z-1} dt = \int_0^\infty e^{-t} t^{z-1} dt = \Gamma(z). \tag{5.3}$$

We claim that the formula

$$\int_0^\infty f_m(t) t^{z-1}\, dt = \frac{m^z m!}{z(z+1)\dots(z+m)} \tag{5.4}$$

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

From (5.2) and some rearrangement we have

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

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

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

Comparing this with (5.3) tells us that indeed $\Gamma(z) = \tilde{\Gamma}(z)$ for $\operatorname{Re} z > 0$. $\square$

We note that, as an easy deduction from (ii), we have $\Gamma(k+1) = k!$ for all positive integers k.

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

$$\Xi(s) = \gamma(s)\zeta(s),$$

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, the functional equation for ζ.

Theorem 5.2. *The completed ζ-function Ξ is meromorphic in $\mathbf{C}$, and its only poles are simple ones at $s = 0$ and 1. It satisfies the* functional equation

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

As a corollary we obtain a the first part of Theorem 3.3, the main result of this chapter about the ζ-function.

Corollary 5.3. *The ζ-function has a meromorphic continuation to all of $\mathbb{C}$. It has a simple pole at $s = 1$ and no other poles. It has zeros at $s = -2, -4, -6, \ldots$ and no other zeros outside of the* critical strip $0 \leqslant \operatorname{Re} s \leqslant 1$.

Proof. The meromorphic continuation is obvious, since

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

and the three functions here are all meromorphic. The statements about zeros and poles are a consequence of the following information, which we have assembled at various earlier points of the course:

- $\Xi(s) = \Xi(1-s)$ and the only poles of Ξ are simple ones at $s = 0$ and $s = 1$;
- Γ has no zeros and simple poles at $0, -1, -2, \dots$;
- ζ has no zeros with $\operatorname{Re} s > 1$.

This completes the proof. □

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 $0 \leqslant \operatorname{Re} s \leqslant 1$.

Note that, 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 return to this later.

5.3. **Proof of the functional equation.** We turn now to the proof of Theorem 5.2. A key tool here will be the θ-function.
We will apply Lemma 4.14 to derive a functional equation for the θ-function. Let us now define that function. For z in the upper half-plane $\mathcal{H} := \{z : \operatorname{Im} z > 0\}$ set

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

It is quite easy to see that θ is analytic in $\mathcal{H}$.

Lemma 5.4. *Suppose that $t \in \mathbb{R}^+$. Then $\theta(it) = \frac{1}{\sqrt{t}}\theta(i/t)$.*

Remark. (may be of interest to some) Then this functional equation for θ may be analytically extended to all of $\mathcal{H}$ to give something like $\theta(-1/z) = \sqrt{z}\theta(z)$. We have been deliberately vague about the branch of square root to be taken here. In conjunction with the easily verified relation $\theta(z+2) = \theta(z)$, this means that the function $\theta(2z)$ is a *modular form* of weight $1/2$ for the group $\Gamma_0(4)$ of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $a, b, c, d \in \mathbf{Z}$, $ad - bc = 1$ and $4|c$.

Proof. Suppose that $t \in \mathbb{R}^+$, and set $f(x) = e^{-\pi x^2 t}$. We calculated (by direct calculation) the Fourier transform $\hat{f}(\lambda)$ in Lemma 4.5; indeed

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

Lemma 5.4 is simply a matter of applying the Poisson summation formula in this case. □

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 (we will see Mellin transforms again later).

Lemma 5.5. *Suppose that* $\mathrm{Re}\, s > 1$. *Then*

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

Proof. Observe that

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

Now simply sum over $n \in \mathbb{N}$. □

Proof of Theorem 5.2. Suppose that $\mathrm{Re}\, s > 1$. We split the integral in Lemma 5.5 into the ranges $[0, 1]$ and $[1, \infty)$, and then apply Lemma 5.4 to the first of these. We have

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

Thus

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

which implies that

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

A priori this formula is valid only for $\mathrm{Re}\, s > 1$. Note, however, that the right-hand side defines a function which is meromorphic in the whole complex plane, with simple poles at $s = 0$ and 1, and which 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

ζ'/ζ 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 C such that*

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

Here sum is over all zeros $Z = Z_{\text{triv}} \cup Z_{\text{nontriv}}$ of ζ: the trivial zeros $Z_{\text{triv}} := \{-2, -4, -6, \dots\}$ *and the* nontrivial *zeros Z_{nontriv} in the critical strip $0 \leqslant \operatorname{Re} s \leqslant 1$ (with multiplicity). 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 an exercise on Sheet 3.

The constant C can be computed explicitly, but this is not important for most applications; in fact $e^C = \pi\sqrt{2}/e$.

Proposition 6.1 is a consequence of a *product expansion* (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 an exercise on sheet 4.

Theorem 6.2 (Hadamard Product for ζ)**.** *There are constants A and B (which can be evaluated explicitly) 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 denotes both trivial and nontrivial zeros of ζ. We have $\sum_{\rho \in Z} |\rho|^{-2} < \infty$, so the product here is a Weierstrass product and converges as described in Proposition C.2.

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 $|f(z)| \ll_\varepsilon e^{C_\varepsilon |z|^{1+\varepsilon}}$ for all $\varepsilon > 0$. Then we say that f is an integral function of order 1.

Proposition 6.4. *Suppose that f is an integral function of order 1. 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 of course) such that
$$f(z)=z^r e^{Az+B}\prod_{\rho}\Big(1-\frac{z}{\rho}\Big)e^{z/\rho}.$$

Proposition 6.4 is, as we said, a result of complex analysis and has no arithmetic content. The proof is given in Sections 6.3 and 6.4.

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

Rather than prove that $(s-1)\zeta(s)$ is an integral function of order one, it is convenient to first include some Γ- and other factors.

Lemma 6.5. *The function $f(s)=s(1-s)\pi^{-s/2}\Gamma(s/2)\zeta(s)=s(1-s)\Xi(s)$ is an integral function of order one.*

Proof. We have already noted that Ξ is meromorphic except for simple poles at $s=0,1$, so f is entire. Therefore we need only establish a growth condition of the form
$$|f(s)|\ll e^{C_\varepsilon|s|^{1+\varepsilon}} \tag{6.1}$$
for all $\varepsilon>0$. In fact we will establish the stronger bound
$$|f(s)|\ll\exp(O(|s|\log|s|))$$
for $|s|\geqslant 3$, which is easily seen to imply (6.1). Moreover, from the functional equation we have $f(s)=f(1-s)$, and so it is enough to prove this when $\operatorname{Re}s\geqslant\frac{1}{2}$. (Here, we note that
$$|f(1-s)|=|f(s)|\ll\exp\big(O(|s|\log|s|\big)\ll\exp\big(O(|1-s|\log|1-s|)\big),$$
since $|s|\log|s|\ll|1-s|\log|1-s|$ for $|1-s|\geqslant 3$.) When $\operatorname{Re}s\geqslant\frac{1}{2}$ we have the integral representation
$$\zeta(s)=\frac{s}{s-1}-s\int_1^\infty\{x\}x^{-s-1}\,dx.$$
that we found in Proposition 3.4. This in fact implies (writing $\sigma:=\operatorname{Re}s$) that
$$\begin{aligned}|(s-1)\zeta(s)|&\leqslant|s|+|s(s-1)|\int_1^\infty x^{-\sigma-1}dx\\&=|s|+|s(s-1)|\frac{1}{\sigma}\ll|s|^2\end{aligned}$$
in this domain. For the Γ-factor we also have the integral representation
$$\Gamma(s/2)=\int_0^\infty e^{-t}t^{s/2-1}dt,$$
so
$$|\Gamma(s/2)|\leqslant\Gamma(\frac{1}{2}\operatorname{Re}s)\leqslant\lfloor\frac{1}{2}\operatorname{Re}s\rfloor!=\exp(O(|s|\log|s|))$$

This final bound follows from the fact that (for integer k) $k! \leqslant k^k = \exp(k \log k)$.

It follows from Proposition 6.4 that

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

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 being an exercise on sheet 3). The zeros ρ satisfy $\sum_\rho |\rho|^{-2} < \infty$. Note that 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 $\sum_{\rho \in Z_{\mathrm{triv}}} |\rho|^{-2} < \infty$, 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 formula (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 5.1 (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 for the domain $|z| < R$.

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

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

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

$$f(z) = C F(z) \prod_{\rho} g_\rho(z),$$

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 $B_{R+\epsilon}$ then F has no zeros in $B_{R+\epsilon}$. 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 $B_{R+\epsilon}$ by the formula

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

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

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

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

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

Taking real parts gives

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

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, note that 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, note that $|g_\rho(0)| = |\rho|/R$. 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 immediately gives

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

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 of ρ of the Riemann ζ-function with imaginary part at most T is $O(T \log T)$.*

Proof. Consider the function $f(s) = s(1-s)\pi^{-s/2}\Gamma(s/2)\zeta(s)$, which vanishes at all nontrivial zeros of ζ. We showed earlier in the chapter that f is entire and satisfies the growth condition $|f(s)| \ll e^{O(|s|\log|s|)}$. It follows from Corollary 6.7 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 $|\rho| \leqslant T+1$, and so the result follows. □

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

Lemma 6.9. *Suppose that g is an entire function with no zeros which satisfies the bound $|g(z)| = \exp(O(|z|^{3/2}))$ 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 absolute 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

$$\int_0^1 |\operatorname{Re} h(R_j e^{2\pi i\theta})| d\theta \leqslant 2CR_j^{3/2} - \int_0^1 \operatorname{Re} h(R_j e^{2\pi i\theta}) d\theta \ll R_j^{3/2}, \qquad (6.4)$$

using the fact that $\frac{h(z)}{z}$ is holomorphic, as in the proof of Jensen's formula.

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

$$2\operatorname{Re}(h(R_j e^{2\pi i\theta})) = \sum_{n\geqslant 0} c_n R_j^n e^{2\pi i n\theta} + \sum_{n \geqslant 0} \overline{c}_n R_j^n e^{-2\pi i n\theta}.$$

By orthogonality it follows that

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

and so by the triangle inequality

$$|c_m| \leqslant R_j^{-m} \frac{1}{2} \int_0^1 |\operatorname{Re} h(R_j e^{2\pi i\theta})| d\theta.$$

By (6.4) it follows 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. □

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

The reader should start by recall the statement of Proposition C.2, which describes the convergence properties of Weierstrass products $E_\Omega(z)$.

From the growth condition on f and Corollary 6.7 it follows very easily that

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

for every $\varepsilon > 0$. In particular (taking any $\varepsilon < 1$) we see that $\sum_{\rho \neq 0} |\rho|^{-2} < \infty$. Therefore we may apply Proposition C.2 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. If f has a zero of order r at $z = 0$, then we 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 $g(z) = e^{Az+B}$. Moreover we already have a tool for doing precisely this, namely Lemma 6.9. Applying this lemma, we see that all we need do is establish an upper bound

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

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} decays quickly, it is enough to establish the *lower* bound

$$|E_\Omega(z)| \geqslant e^{-C' R_j^{3/2}} \tag{6.6}$$

for $|z| = R_j$. 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 obviously 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 $2^j \leqslant |\rho| \leqslant 2^{j+1}$. Then, by (6.5) with $\varepsilon = 1$, $|S_j| \ll 2^{2j}$. It follows easily that R_j may be chosen in such a way that

$$|z - \rho| \gg 2^{-j} \tag{6.7}$$

whenever $|z| = R_j$, for all zeros ρ. Suppose from now on that $|z| = R_j$.

To get a lower bound on $E_\Omega(z)$, we divide the product into dyadic subproducts

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

$j' = 0, 1, 2, \ldots$. The contribution from $|\rho| \leqslant 1$ is clearly bounded below, independently of R, by some constant (which may depend on f).

Now if $j' < j - 10$ (say) then, for any zero ρ occurring in the product $E_\Omega^{(j')}(z)$, we have $|1 - z/\rho| \geqslant 1$. Furthermore $|e^{z/\rho}| \geqslant e^{-2^{j-j'}}$, and so

$$|E_\Omega^{(j')}(z)| \geqslant (e^{-2^{j-j'}})^{|S_{2^{j'}}|} \geqslant e^{-C2^j 2^{j'/10}}$$

(say), the last bound here being a consequence of (6.5) with $\varepsilon := 1/10$.

If $j - 10 < j' < j + 10$ then we employ the bound $|1 - z/\rho| \geqslant c2^{-2j}$, a consequence of (6.7). For any such j' we have $|e^{z/\rho}| \gg 1$, and so

$$|E_\Omega^{(j')}(z)| \gg (c2^{-2j})^{|S_{2^{j'}}|} \gg e^{-C2^{3j/2}},$$

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

Finally, suppose that $j' \geqslant j + 10$. By a crude Taylor series expansion one has $|(1-w)e^w| > e^{-10|w|^2}$ for $|w| < 1/10$. Thus, for these values of j,

$$|E_\Omega^{(j')}(z)| > (e^{-10 \cdot 2^{2j-2j'}})^{|S_{2^{j'}}|} \geqslant e^{-C2^{2j}2^{-j'/2}},$$

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

It is a simple matter to check that the product of all these estimates for $E_\Omega^{(j')}(z)$, over all j', is bounded below by $e^{-C2^{3j/2}}$, for some absolute constant C. This concludes the proof.

7. Mellin transforms and the explicit formula

In this section 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 and it 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 more than a slight annoyance, since the prime powers are so sparse.

The prime number theorem asserts the $\psi(X) \sim X$, 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 $1_{[0,1]}$, 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*, that is to say infinitely differentiable, and also compactly supported (that is, zero outside some closed interval). This means that in order to apply it to the prime number theorem we must approximate $1_{[0,1]}$ by smooth, compactly supported functions. This is not especially difficult, but also not entirely straightforward – indeed, even the *existence* of a smooth compactly supported function other than 0 is not immediately obvious.

The explicit formula involves the Mellin transform. This was discussed briefly in Chapter 4. We reintroduce it now: it is convenient to denote it with a tilde rather than a hat, as we will require the Fourier transform on $\mathbf{R}$ in our arguments. Moreover, the Mellin transform as defined earlier had, as its domain, the imaginary axis $\{it : t \in \mathbf{C}\}$. Now we will extend the domain of definition to the whole complex plane, which takes us a little outside the scope of the discussion in Chapter 4.

Definition 7.1. Suppose that $W : \mathbf{R} \to \mathbf{R}$ is has compact support contained in $(0, \infty)$. For any complex number s we define the Mellin transform $\tilde{W}$ by

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

The assumption that W has compact support bounded away from 0 means that W is well-defined and entire, the derivative of W being obtained by differentiating under the integral.

Theorem 7.2 (Explicit formula)**.** *Let W be a smooth, compactly supported function, supported on $[1, \infty)$. Suppose that $X > 1$. Then*

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

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

Remarks. Since $\tilde{W}(1) = \int W$, one could if desired write the right-hand side as

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

where $\mathrm{ord}_\zeta(w)$ 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 assumption that W be compactly supported is very convenient for the proof, for instance because in this case it is clear that $\tilde{W}(s)$ is holomorphic. However, Theorem 7.2 does hold under weaker conditions. We will not discuss this here.

In some ways, W is most naturally thought of as a function on the multiplicative positive real line $\mathbf{R}_+$. To this end, we associate a function $w : \mathbf{R}_+ \to \mathbf{R}$ via $w := W \circ \exp$, that is to say

$$w(u) := W(e^u).$$

Saying $\mathrm{Supp}(W)$ is a compact subset of $(0, \infty)$ is then equivalent to saying that w is compactly supported, whilst the condition $\mathrm{Supp}(W) \subset [1, \infty)$ is equivalent to $\mathrm{Supp}(w) \subset [0, \infty)$. One may note that

$$\tilde{W}(s) = \int_0^\infty W(x) x^s \frac{dx}{x} = \hat{w}(is),$$

where $\hat{w}$ denotes the Fourier transform of w (but allowed to take arguments in $\mathbf{C}$, not just in $\mathbf{R}$). The explicit formula can be written in terms of w as

$$\sum_n \Lambda(n) w(\log n - \log X) = X \int_0^\infty w(\log x) dx - \sum_{\rho \in Z} X^\rho \hat{w}(i\rho).$$

Before turning to the proof of the explicit formula, let is make some remarks about its form. The term $X(\int W)$ is the "main term"; 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 $\sum_n W(\frac{n}{X}) \approx X(\int W)$. The remaining terms $\sum_{\rho \in Z} \tilde{W}(\rho)$ are, therefore, to be thought of as "error terms". It turns out that the contribution from $\rho \in Z_{\mathrm{triv}}$ is negligible, leaving the sum $\sum_{\rho \in Z_{\mathrm{nontriv}}} X^{\mathrm{Re}\,\rho} \tilde{W}(\rho)$. 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 $0 \leqslant \mathrm{Re}\,\rho \leqslant 1$. For all we know, it could be that some zero ρ lies on the line $\mathrm{Re}\,s = 1$, in which case one would expect the term $X^{\mathrm{Re}\,\rho} \tilde{W}(\rho)$ to have the same order of magnitude, $O(X)$, 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 section.

A key ingredient is the inversion formula for the Mellin transform, which is basically equivalent to the Fourier inversion formula.

Proposition 7.3. *Let $W : \mathbf{R} \to \mathbf{R}$ be a smooth function with compact support contained in $(0, \infty)$. Then for any $\sigma \in \mathbf{R}$ we have the Mellin inversion formula*

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

Proof. Suppose that $s = \sigma + it$. Then

$$\tilde{W}(s) = \widehat{w \exp^\sigma}(-t), \tag{7.1}$$

where w is defined as before, and now the Fourier transform takes a real argument $-t$. Here, exp is the exponential function, so $\exp^\sigma(x) := e^{\sigma x}$. The function $w \exp^\sigma$ is of course compactly supported and smooth, so we may apply the Fourier inversion formula to get

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

which is equivalent to the stated formula if one substitutes $u = \log x$. $\square$

Let us recall that we know how to obtain decay estimates for the Fourier transform of smooth functions: see for example Lemma 4.1. It is therefore not surprising that (7.1) will be useful in its own right for obtaining corresponding estimates for the Mellin transform. See Proposition 7.4 below for details.

Now we can outline the proof of the explicit formula as stated in Theorem 7.2. 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 earlier in the course.

By the Mellin inversion formula, Proposition 7.3 and this Dirichlet series expansion we have

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

We will rigorously justify this calculation in the course of our proofs.

We now reach 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 $\rho \in Z_{\mathrm{nontriv}}$ in the critical strip, and at the trivial zeros $Z_{\mathrm{triv}} = \{-2, -4, -6, \dots\}$. To this end we note that

$$\mathrm{Res}_{s=1}\frac{\zeta'}{\zeta} = -1,$$

$$\mathrm{Res}_{s=-2k}\frac{\zeta'}{\zeta} = 1,$$

and $\mathrm{Res}_{s=\rho}\frac{\zeta'}{\zeta}$ is equal to the multiplicity of the zero ρ.

To perform this contour integration rigorously we employ a rectangular contour $C_{k,T} = C_T^{(1)} \cup C_{k,T}^{(2)} \cup C_{k,T}^{(3)} \cup C_{k,T}^{(4)}$, where $C_T^{(1)} = [2 - iT, 2 + iT]$, $C_{k,T}^{(2)} = [2 + iT, -2k - 1 + iT]$, $C_{k,T}^{(3)} = [-2k - 1 + iT, -2k - 1 - iT]$ and $C_{k,T}^{(4)} = [-2k-1-iT, 2+iT]$, and then choose suitable values of k, T tending to infinity. Here k is an integer; with this choice, $C_{k,T}^{(3)}$ avoids the poles of

ζ'/ζ at the trivial zeros Z_{triv}. The choice of T is rather more delicate and in particular we need to make sure that neither $C^{(2)}_{k,T}$ nor $C^{(4)}_{k,T}$ pass close to any of the poles of ζ'/ζ at the nontrivial zeros Z_{nontriv}. Later on, we will choose a sequence $T_1, T_2, \dots$ of "good" values of T and corresponding integers k_j such that, in particular, neither $C^{(2)}_{k_j,T_j}$ nor $C^{(4)}_{k_j,T_j}$ contain any zeros of ζ, and such that $k_j, T_j \to \infty$.

For $\ell = 1, 2, 3, 4$ and for some choice of k, T define

$$I^{(\ell)}_{k,T} := -\frac{1}{2\pi i}\int_{C^{(\ell)}_{k,T}} \frac{\zeta'(s)}{\zeta(s)} X^s \tilde{W}(s)\, ds. \tag{7.3}$$

To calculate $\sum_{\ell=1}^4 I^{(\ell)}_{k,T}$ we apply the residue theorem and the preceding remarks, obtaining

$$\sum_{\ell=1}^{4} I^{(\ell)}_{k,T} = X\tilde{W}(1) - \sum_{\substack{\rho\in Z_{\text{nontriv}} \\ |\rho|<T}} X^{\rho}\tilde{W}(\rho) - \sum_{m=1}^{k} X^{-2m}\tilde{W}(-2m). \tag{7.4}$$

Note that

$$\tilde{W}(1) = \int_{\mathbf{R}} W.$$

Therefore as k, T tend to infinity along any subsequence, the right-hand side of (7.4) tends towards the right hand side of the explicit formula. To conclude the proof of the explicit formula it therefore suffices to show that

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

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

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

and

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

We will prove these statements (as well as define the sequences k_j, T_j) in the next section.

7.3. **Estimates for the Mellin transforms.** The following lemma gives a decay estimate for the Mellin transform of a smooth function. In this lemma, we have obeyed 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 $\beta + i\gamma$.

Lemma 7.4. *Suppose that $W : \mathbf{R}_+ \to \mathbf{R}$ is smooth and has compact support contained in $[1, \infty)$. Then*

(i) *Suppose that* $-\infty < \sigma \leqslant 2$. *Then we have*

$$|\tilde{W}(\sigma + it)| \ll_W 1 \tag{7.8}$$

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

$$|\tilde{W}(\sigma + it)| \ll_{m,W} (1 + |\sigma|)^m |t|^{-m} \tag{7.9}$$

uniformly for $|t| \geqslant 1$.

(ii) *Suppose that* $\operatorname{Supp} W \subseteq [1, 10]$. *We have the bounds*

$$|\tilde{W}(\rho)| \ll_m \sup_{0 \leqslant j \leqslant m} \|\partial^j W\|_1 |\rho|^{-m} \tag{7.10}$$

whenever $\rho \in Z_{\mathrm{nontriv}}$, $|\rho| \geqslant 2$, *and*

$$|\tilde{W}(\rho)| \ll \|W\|_1 \tag{7.11}$$

uniformly for all $\rho \in Z = Z_{\mathrm{triv}} \cup Z_{\mathrm{nontriv}}$.

Remark. We remark that we have explicitly formulated this result in 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.

Note that the bounds are of two types, namely 'trivial' bounds (7.8), (7.11) giving what are essentially very crude constant bounds, and the more sophisticated bounds (7.9), (7.10) which give decay of the Mellin transform in the vertical direction.

Proof. The proof of all parts uses (7.1), that is to say

$$\tilde{W}(\sigma + it) = \widehat{w \exp^\sigma}(-t), \tag{7.12}$$

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

$$|\hat{f}(t)| \leqslant |t|^{-m} \|\partial^m f\|_1 \tag{7.13}$$

for any $f \in \mathcal{S}(\mathbf{R})$. For (7.8) and (7.11), we use the case $m = 0$ of (7.13), or in other words the simple bound $|\hat{f}(t)| \leqslant \|f\|_1$. This gives

$$\big|\tilde{W}(\sigma + it)\big| \leqslant \int_0^\infty |w(u)| e^{\sigma u} du \leqslant \big(\max(\operatorname{Supp}(W))\big)^\sigma \|w\|_1.$$

The bound (7.8) follows immediately from this. The bound (7.11) follows from this and the additional observations that

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

as well as the fact that

$$\big(\max(\operatorname{Supp}(W))\big)^\sigma \leqslant 10$$

if $\rho = \sigma + it$ is a zero of ζ, since $\operatorname{Supp}(W) \subseteq [1, 10]$.

For the other statements, we use (7.13) with $m \geqslant 1$, which requires us to estimate the derivatives. By Leibniz's rule 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 (7.9), we use the bounds $E_1 \ll (1+|\sigma|)^m$, $E_2 \ll_W 1$ and $E_3 \ll_{W,m} 1$.

For (7.10), set $\rho = \beta + i\gamma$, where $0 \leqslant \beta \leqslant 1$, and note that $|\rho| \ll |\gamma|$; therefore it suffices to get bounds in terms of $|\gamma|^{-m}$ (which is what (7.13) gives) rather than $|\rho|^{-m}$. In the definition of E_1, E_2, E_3, σ must be replaced by β. We have the bounds $E_1 \ll_m 1$, which holds uniformly for $0 \leqslant \beta \leqslant 1$, and $E_2 \ll 1$, which again holds uniformly for $0 \leqslant \beta \leqslant 1$ since $\operatorname{Supp}(w) \subset [0, 3]$. To complete the proof, we must bound E_3 by showing that

$$\sup_{0\leqslant j\leqslant m}\|\partial^j w\|_1 \ll_m \sup_{0\leqslant j\leqslant m}\|\partial^j W\|_1. \tag{7.14}$$

To see this, observe that by Leibniz's rule $\partial^j w(u)$ is a sum of terms of the form $e^{j'u}\partial^{j'}W(e^u)$ for $j' \leqslant j$. But

$$\int_0^{\log 10} e^{j'u}\partial^{j'}W(e^u)du = \int_1^{10} x^{j'-1}\partial^{j'}W(x)dx \ll_{j'} \|\partial^{j'}W\|_1,$$

so the result follows. □

7.4. **Bounds for ζ'/ζ at judicuously 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 $C^{(2)}_{k_j,T_j}, C^{(4)}_{k_j,T_j}$. Very crude estimates suffice, ultimately because we are dealing with smooth functions W which enjoy very advantageous Mellin decay properties.

We isolate a sublemma from the proof (I called this 'an application of the pigeonhole principle' in lectures).

Lemma 7.5. *Let $S \subset \mathbf{R}$ be a sequence and let I be an interval. Suppose that $|S \cap I| = n$. Then there is some $x \in I$ with* $\operatorname{dist}(x, S) \geqslant |I|/2(n + 2)$.

Proof. By translating and rescaling, we may assume without loss of generality that $I = [0, 1]$. Consider the $n + 1$ disjoint open intervals centred on $j/(n+2)$, $j = 1, \dots, n+1$ and with length $1/(n+2)$. These are all contained in I. By the pigeonhole principle and the assumption, one of these intervals contains no point of S. One may then take x to be the midpoint of this interval. □

Lemma 7.6. *There is a sequence of T tending to infinity such that*

$$\mathrm{dist}(s, Z) \geqslant 1/T$$

uniformly for all s with $\operatorname{Im} s = T$.

Proof. Let U be a sufficiently large real parameter. We will show that it is possible to find $T \in [U, 2U]$ with the stated property. Certainly, for any s with $\operatorname{Im} s \geqslant U$ we have (provided $U \geqslant 1$) that $\mathrm{dist}(s, Z_{\mathrm{triv}}) \geqslant 1$ and so it suffices to handle the nontrivial zeros. We showed earlier that $\sum_\rho |\rho|^{-3/2} < \infty$, which certainly implies that the number of nontrivial zeros $\rho = \beta + i\gamma$ in the critical strip (thus with $0 \leqslant \beta \leqslant 1$) with $U \leqslant \gamma \leqslant 2U$ is $O(U^{3/2})$. (In fact, in Proposition 6.8 we obtained the stronger bound $O(U \log U)$.) By Lemma 7.5, applied with S being the sequence of imaginary parts of the nontrivial zeros of ζ, $I = [U, 2U]$ and $n \ll U^{3/2}$, we see that there is some $T \in [U, 2U]$ such that $|T - \gamma| \gg U^{-1/2}$, and hence such that $|T - \gamma| \geqslant 1/T$ provided U is sufficiently large, for the ordinate γ of any nontrivial zero. Finally, note that if $\rho = \beta + i\gamma$ and if $\operatorname{Im} s = T$ then $|s - \rho| \geqslant |T - \gamma|$. □

We call a value of T for which the conclusion of Lemma 7.6 holds *good*. 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 $C^{(2)}_{k,T}, C^{(4)}_{k,T}$. We actually need rather more than this, namely a reasonable upper bound for ζ'/ζ on these contours. That said, a fairly crude bound will do, and we provide such a bound in the next proposition.

Proposition 7.7. *Suppose that $|s| \geqslant 3$. Then*

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

where Z is the set of all zeros of ζ.

Remark. This is a crude bound, and stronger ones could be obtained. For our purposes, however, any bound of the form $|s|^C \,\mathrm{dist}(s, Z)^{-C'}$ would be sufficient.

Proof. We use the partial fraction expansion

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

The first two terms are clearly $O(1)$ in the domain $|s| \geqslant 3$. Note that the right-hand side is $\gg 1$, since $\mathrm{dist}(s, Z) \geqslant \mathrm{dist}(s, -2) \geqslant 1$.

To bound the sum over the zeros in the partial fraction expansion, our main ingredient will be the estimate

$$\sum_{\rho \in Z} |\rho|^{-2} < \infty. \tag{7.15}$$

We will split into 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 hence $|s-\rho| \geqslant \frac{1}{2}|\rho|$. Hence

$$\sum_{|\rho| \geqslant 2|s|} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) \leqslant |s| \sum_{|\rho| \geqslant 2|s|} \frac{1}{|s-\rho||\rho|} \leqslant 2|s| \sum_{|\rho| \geqslant 2|s|} \frac{1}{|\rho|^2} \ll |s|, \quad (7.16)$$

the last step of course following from (7.15). To estimate the contribution from the remaining zeros, those with $|\rho| < 2|s|$, we proceed extremely crudely. By (7.15) it follows immediately 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 simply from the fact that 0 is not a zero of ζ. (Evaluating $\zeta(0)$ is a question on Sheet 3.) Therefore

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

and the proposition follows. □

Choose our sequences of k_j, T_j as follows. Let T_j be any sequence of good values of T with $\lim_{j \to \infty} T_j = \infty$. Set $k_j := \lceil T_j^{1/4} \rceil$. Now we complete the proof of the explicit formula by verifying (7.5), (7.6) and (7.7) with these choices. For brevity 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

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

(This is rigorous, being an application of the Mellin inversion formula; note also that the sum over n is effectively a finite sum since W is compactly supported). Now we truncate the integral to $2 \pm iT$, or in other words to the contour $C^{(1)}_{k,T}$. Recall from (7.9) that we have the decay estimate $|\tilde{W}(2+it)| \ll_W |t|^{-2}$, so the error in doing this is bounded above by

$$\ll_W \sum_n \Lambda(n)(n/X)^{-2} \int_T^\infty |t|^{-2} dt \ll \frac{X^2}{T} \sum_n \Lambda(n) n^{-2} \ll \frac{X^2}{T},$$

which clearly tends to 0 as $T \to \infty$.

For the integral from $2-iT$ to $2+iT$ we interchange summation, obtaining

$$\sum_n \Lambda(n)\frac{1}{2\pi i}\int_{2-iT}^{2+iT}\tilde{W}(s)(n/X)^{-s}ds = \frac{1}{2\pi i}\int_{2-iT}^{2+iT}\Big(\sum_n \Lambda(n)n^{-s}\Big)X^s\tilde{W}(s)$$
$$= \frac{1}{2\pi i}\int_{2-iT}^{2+iT} -\frac{\zeta'(s)}{\zeta(s)}X^s\tilde{W}(s)ds = I^{(1)}_{k,T}.$$

Note here that $\sum_n \Lambda(n)n^{-s} = -\zeta'(s)/\zeta(s)$ is uniformly bounded on the contour by $\zeta(2)$. The statement (7.5) now follows.

Proof of (7.6). If $s \in C^{(2)}_{k,T}\cup C^{(4)}_{k,T}$ then, since T is good, $\mathrm{dist}(s,Z) \gg \frac{1}{T}$. It follows from Proposition 7.7 that, uniformly for $s \in C^{(2)}_{k,T}\cup C^{(4)}_{k,T}$, we have

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

It follows that

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

The bound (7.9) with $m=5$ tells is that

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

uniformly for $-\infty < \sigma \leqslant 2$. Therefore we obtain

$$I^{(2)}_{k,T}, I^{(4)}_{k,T} \ll_W k^3T^{-1}\int_{-2k-1}^{2}(1+|\sigma|)^5X^\sigma d\sigma \ll_{W,X} k^3T^{-1},$$

where here we used the fact that $X>1$. Thus

$$I^{(2)}_{k,T}, I^{(4)}_{k,T} \ll_{W,X} k^3T^{-1}.$$

Since $k = \lceil T^{1/4}\rceil$, statement (7.6) follows.

Proof of (7.7). If $s \in C^{(3)}_{k,T}$ then $\mathrm{dist}(s,Z) \geqslant 1$ and so Proposition 7.7 furnishes the bound $\frac{\zeta'(s)}{\zeta(s)} \ll |s|^3$. If $s = -2k-1+it$ then (crudely) this is $\ll k^3\max(1,|t|)^3$.

From (7.9) with $m=5$ 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

$$I^{(3)}_{k,T} = \int_{-2k-1-iT}^{-2k-1+iT}\frac{\zeta'(s)}{\zeta(s)}X^s\tilde{W}(s)ds \ll_W k^8X^{-2k-1}.$$

From this and the fact that $X>1$ we immediately conclude (7.7), that is to say that indeed

$$\lim_{k,T\to\infty} I^{(3)}_{k,T} = 0$$

Note that for the estimation of $I^{(3)}$ there was no need to restrict to T good.

8. The prime number theorem

In the last section 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 $\sigma \to 1$. Taking logs*, we have

$$\log \zeta(s) = -\sum_p \log(1 - p^{-s}) = \sum_p \sum_{m=1}^{\infty} \frac{1}{mp^{ms}} = \sum_p \sum_{m=1}^{\infty} \frac{1}{mp^{m\sigma}} e^{-imt \log p}. \tag{8.1}$$

Now we invoke the inequality

$$3 + 4\cos\theta + \cos 2\theta = 2(1 + \cos\theta)^2 \geqslant 0. \tag{8.2}$$

Applying this with $\theta = mt \log p$ and comparing with (8.1), we get

$$3 \log \zeta(\sigma) + 4 \operatorname{Re} \log \zeta(\sigma + it) + \operatorname{Re} \log \zeta(\sigma + 2it) \geqslant 0,$$

and thus

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

Now suppose that $\zeta(1 + it) = 0$, and let $\sigma \to 1$ in the inequality just proved. We have $\zeta(\sigma) \sim (\sigma - 1)^{-1}$, but $|\zeta(\sigma + it)|^4 \ll (\sigma - 1)^4$ (by Taylor exapnsion 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. ∎

Remark. *There is a bit of subtlety about what is meant by taking logs here. Note that complex logs are only defined up to an ambiguity in $2\pi i\mathbf{Z}$. In the above, the $-\log(1 - p^{-s})$ terms are *defined* using the series expansion $-\log(1 - z) = \sum_{j \geqslant 1} z^j / j$, which converges for $|z| < 1$, and so is certainly valid with $z = p^{-s}$. It is a known fact from complex analysis that, with this definition, $e^{-\log(1-z)} = (1 - z)^{-1}$. The logarithm of $\prod_p (1 - p^{-s})^{-1}$ is *defined* to be $-\sum_p \log(1 - p^{-s})$. What we used in the proof is that the exponential of this is $\prod_p (1 - p^{-s})^{-1}$, which follows from $e^{\sum_p a_p} = \prod_p e^{a_p}$

(where $a_p := (1-p^{-s})^{-1}$). Note that we should not expect $\log \prod_p (1-p^{-s})^{-1}$ to be given by a series expansion, but we never used any such statement in the argument.*

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. *Suppose that* $W : \mathbf{R} \to \mathbf{R}$ *is smooth and that* $\mathrm{Supp}(W) \subset [1, 10]$. *Then*

$$\lim_{X \to \infty} \frac{1}{X} \sum_n \Lambda(n) W\big(\frac{n}{X}\big) = \int W.$$

Proof. We regard W as fixed throughout the proof and do not indicate dependencies on W. By the explicit formula it suffices to show that

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

We handle the contribution from the trivial zeros and the nontrivial ones separately. By (7.11) we have $\tilde{W}(\rho) \ll 1$ uniformly for $\rho \in Z_{\mathrm{triv}}$, and so

$$\sum_{\rho \in Z_{\mathrm{triv}}} X^{\rho - 1} \tilde{W}(\rho) \ll_W \sum_{j=1}^\infty X^{-2j-1}.$$

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

Turning now to the contribution from the nontrivial zeros, we use (7.10) with $m = 2$, which tells us that $\tilde{W}(\rho) \ll |\rho|^{-2}$ uniformly for $\rho \in Z_{\mathrm{nontriv}}$. Noting also that $|X^{\rho-1}| \leqslant 1$ (since $\mathrm{Re}\, \rho \leqslant 1$) and recalling that $\sum_\rho |\rho|^{-2} < \infty$, it follows that for some $K = K(\varepsilon)$ we have

$$\Big| \sum_{\rho \in Z_{\mathrm{nontriv}} : |\rho| \geqslant K} X^{\rho - 1} \tilde{W}(\rho) \Big| \leqslant \frac{\varepsilon}{2}.$$

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 we have

$$\sum_{\rho \in Z_{\mathrm{nontriv}} : |\rho| < K} X^{\rho - 1} \tilde{W}(\rho) \ll_K X^{\beta(K) - 1},$$

where $\beta(K) = \sup_{\rho \in Z_{\mathrm{nontriv}} : |\rho| \leqslant K} \mathrm{Re}\, \rho$. By Proposition 8.1, we have $\beta(K) < 1$. If X is big enough, this contribution is less than $\varepsilon/2$. It follows that if X is large enough in terms of ε then

$$\Big| \sum_{\rho \in Z_{\mathrm{nontriv}}} X^{\rho - 1} \tilde{W}(\rho) \Big| < \varepsilon,$$

which is what we wanted to prove. □

Proposition 8.2 is a kind of "smoothed" prime number theorem. Now we show how the prime number theorem itself follows from it. As shown earlier in the course, the prime number theorem is equivalent to the asymptotic

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

Here, we will show that if X is big enough in terms of ε then

$$\sum_{n \leqslant X} \Lambda(n) \leqslant (1 + \varepsilon)X. \tag{8.3}$$

A corresponding lower bound may be established in exactly the same way.

Take a smooth function $W : \mathbf{R}_+ \to \mathbf{R}$ such that $\mathrm{Supp}(W) \subset [2-\varepsilon, 4+\varepsilon]$, $0 \leqslant W(x) \leqslant 1$ for all x, and $W(x) = 1$ for $2 \leqslant x \leqslant 4$. A construction of such a function is given on Example Sheet 3 (Q4 and a slight modification of Q8). A construction that I regard as more 'conceptual' is given in Appendix B. Applying the explicit formula to W tells us that if Y is sufficiently large in terms of ε then

$$\sum_{2Y < n \leqslant 4Y} \Lambda(n) \leqslant \sum_n \Lambda(n) W\big(\frac{n}{Y}\big) \leqslant \Big(\frac{\varepsilon}{3} + \int W\Big) Y \leqslant \big(1 + \frac{2\varepsilon}{3}\big) 2Y.$$

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

$$\sum_{X/2^m < n \leqslant X} \Lambda(n) \leqslant \big(1 + \frac{2\varepsilon}{3}\big) X.$$

Combining this with the result of Proposition 1.5 gives

$$\sum_{n \leqslant X} \Lambda(n) \leqslant \big(1 + \frac{2\varepsilon}{3}\big) X + O\big(\frac{X}{2^m}\big),$$

provided X is sufficiently large in terms of ε and m. By choosing m sufficiently large in terms of ε, we can ensure that the second term is $\leqslant \varepsilon X/3$. Therefore, if X is big enough in terms of ε, we have

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

as required.

9. The zero-free region. Error terms.

9.1. **The classical zero-free region.** In the last section we showed that ζ has no zeros on the line $\mathrm{Re}\, s = 1$, 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 $\mathrm{Re}\, s = 1$ 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*

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

Remark. Since there are no zeros on the line $\operatorname{Re} s = 1$, one can if desired state a zero-free region of the form

$$\beta < 1 - \frac{c'}{\log(|\gamma| + 2)}$$

for *all* γ.

Proof. We may assume that $|\gamma| \geqslant 2$, since for small γ the result is a consequence of the fact that there are no zeros on the line $\operatorname{Re} s = 1$ (established in the last chapter) and a compactness argument.

Of course, 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 section) it follows that

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

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 holomorphic, and hence continuous, it is bounded on the compact interval $[1, 2]$. That is, if $1 < \sigma \leqslant 2$ then

$$-3 \operatorname{Re} \frac{\zeta'(\sigma)}{\zeta(\sigma)} \leqslant \frac{3}{\sigma - 1} + C_1, \tag{9.2}$$

for some absolute constant C_1. To get bounds for the other two terms in (9.1), we use the partial fraction expansion

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

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 $\ll \frac{1}{k}$, so we get a contribution of $O(\log |t|)$. On the second range the summand is $\ll \frac{|t|}{k^2}$, and so we get a contribution of $O(1)$. It follows that

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

By an essentially identical analysis we also have

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

Combining all this information, we obtain

$$4\,\mathrm{Re} \sum_{\rho \in Z_{\mathrm{nontriv}}} \Big(\frac{1}{\sigma + it - \rho} + \frac{1}{\rho}\Big) + \mathrm{Re} \sum_{\rho \in Z_{\mathrm{nontriv}}} \Big(\frac{1}{\sigma + 2it - \rho} + \frac{1}{\rho}\Big) \leqslant C_2 \log |t| + \frac{3}{\sigma - 1} \tag{9.3}$$

for some C_2.

Since $\mathrm{Re}\,\frac{1}{a+ib} = \frac{a}{a^2+b^2}$ and $\sigma > 1 > \mathrm{Re}\,\rho > 0$ for every $\rho \in Z_{\mathrm{nontriv}}$, 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

$$\frac{4}{\sigma - \beta} \leqslant C_2 \log |\gamma| + \frac{3}{\sigma - 1}.$$

Setting $\sigma = 1 + 1/2C_2 \log |\gamma|$ and rearranging then gives

$$\beta \leqslant 1 - \frac{1}{14 C_2 \log |\gamma|}.$$

This has been proved whenever $|\gamma| \geqslant 2$, and this completes the proof. □

We remark that the classical zero-free region is not the largest one known. In fact, the stronger inequality

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

is known to hold for any $\varepsilon > 0$, a result of Vinogradov and Korobov. The methods necessary to prove this lie considerably deeper.

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. 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 $1_{[2,4]}$ by smooth functions.

Let $\varepsilon > 0$. We claim that there is a smooth function $W_\varepsilon : \mathbf{R}_+ \to \mathbf{R}$ with $\operatorname{Supp}(W_\varepsilon) \subset [2-\varepsilon, 4+\varepsilon]$, $0 \leqslant W_\varepsilon(x) \leqslant 1$ for all x and $W_\varepsilon(x) = 1$ for $2 \leqslant x \leqslant 4$, satisfying the derivative bounds

$$\|\partial W_\varepsilon\|_1 \ll 1, \qquad \|\partial^2 W_\varepsilon\|_1 \ll \frac{1}{\varepsilon}, \tag{9.4}$$

This is reasonably "clear by picture". One way to proceed more rigorously is to first define $W_{1/2}$ and then to set

$$\begin{array}{ll} W_\varepsilon(2+t) = W_{1/2}(2 + \frac{t}{2\varepsilon}) & |t| < \varepsilon, \\ W_\varepsilon(4+t) = W_{1/2}(4 + \frac{t}{2\varepsilon}) & |t| < \varepsilon, \\ W_\varepsilon(x) = 1 & 2+\varepsilon \leqslant x \leqslant 4-\varepsilon. \end{array}$$

That is, W_ε is a kind of contracted version of $W_{1/2}$. We leave the proof of the bounds (9.4) to the reader.

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

$$\begin{aligned} \psi(4Y) - \psi(2Y) &\leqslant \sum_n \Lambda(n) W_\varepsilon(\frac{n}{Y}) = Y \int W - \sum_\rho Y^\rho \tilde{W}_\varepsilon(\rho) \\ &\leqslant 2Y(1+\varepsilon) - \sum_\rho Y^\rho \tilde{W}_\varepsilon(\rho) \\ &\leqslant 2Y(1+\varepsilon) + \sum_\rho |Y^\rho||\tilde{W}_\varepsilon(\rho)|. \end{aligned} \tag{9.5}$$

As in the last chapter, the contribution from the trivial zeros is $\ll Y^{-2}$ and can be ignored. To estimate the sum over $\rho \in Z_{\mathrm{nontriv}}$, we split into exponential ranges. For $j \geqslant 1$, define Z_j to be the set of zeros $\rho \in Z_{\mathrm{nontriv}}$ with $e^j \leqslant |\rho| < e^{j+1}$. Define Z_0 to be the set of zeros with $|\rho| < e$; there are only finitely many such zeros (in fact, though we have not proven this, there are none).

Consider first the ranges Z_j with $j \geqslant 1$. The number of zeros in such a range is $\ll je^j$ by Proposition 6.8. If $\rho = \beta + i\gamma$ is such a zero then, by the classical zero-free region, $\beta \leqslant 1 - c/j$, and hence $|Y^\rho| \leqslant Y^{1-c/j}$.

Applying (7.10) with $m = 2$ gives

$$\tilde{W}_\varepsilon(\rho) \ll \frac{1}{\varepsilon}|\rho|^{-2} \ll \frac{1}{\varepsilon}e^{-2j}. \tag{9.6}$$

It follows that

$$\begin{aligned}\sum_{j\geqslant 1}\sum_{\rho\in Z_j}|Y^\rho||\tilde{W}_\varepsilon(\rho)| &\ll \frac{1}{\varepsilon}\sum_{j\geqslant 1} je^{-j}Y^{1-c/j} \\ &\ll \frac{1}{\varepsilon}\sup_{j\geqslant 1}\big(e^{-j/2}Y^{1-c/j}\big)\sum_{j\geqslant 1} je^{-j/2} \\ &\ll \frac{1}{\varepsilon}\sup_{j\geqslant 1}\big(e^{-j/2}Y^{1-c/j}\big). \end{aligned} \tag{9.7}$$

Now since (by the AM-GM inequality)

$$\frac{c}{j} + \frac{j}{2\log Y} \geqslant 2\Big(\frac{c}{2\log Y}\Big)^{1/2},$$

we have

$$\sup_{j\geqslant 1}\big(e^{-j/2}Y^{1-c/j}\big) \leqslant Ye^{-c'\sqrt{\log Y}} \tag{9.8}$$

where $c' := \sqrt{2c}$. Finally

$$\sum_{\rho\in Z_0}|Y^\rho||\tilde{W}_\varepsilon(\rho)| \ll Y^{1-c''}. \tag{9.9}$$

where $c'' = 1 - \max_{\rho\in Z_0}\operatorname{Re}\rho > 0$, using here the trivial estimate $|\tilde{W}_\varepsilon(\rho)| \ll 1$. Combining (9.7) to (9.9) together in (9.5) yields

$$\psi(4Y) - \psi(2Y) \leqslant 2Y + O(\varepsilon Y) + O\big(\frac{1}{\varepsilon}Ye^{-c'\sqrt{\log Y}}\big).$$

Now we must choose a good value of ε. With the choice $\varepsilon = e^{-\frac{1}{2}c'\sqrt{\log Y}}$ we obtain

$$\psi(4Y) - \psi(2Y) \leqslant 2Y + O(Ye^{-\frac{1}{2}c'\sqrt{\log Y}}).$$

By a telescoping sum argument, using the above with $X/4$, $X/8, \dots$, we see that

$$\psi(X) \leqslant X + O(Xe^{-\frac{1}{2}c'\sqrt{\log X}}).$$

(We leave the precise details of this telescoping sum argument to the reader.)

By entirely analogous arguments, only using a minorant to the interval $1_{[2,4]}$ rather than the majorants W_ε, we may obtain the corresponding lower bound

$$\psi(X) \geqslant X - O(Xe^{-\frac{1}{2}c'\sqrt{\log X}}).$$

This completes the proof of the prime number theorem with classical error term.

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

If the Riemann hypothesis holds then we can, of course, improve the estimate $|Y^\rho| \leqslant Y^{1-c/j}$ in the above argument to $|Y^\rho| \leqslant Y^{1/2}$, and (9.7) is replaced by the bound $\ll \frac{1}{\varepsilon} Y^{1/2}$. This leads to somewhat suboptimal bounds, and to recover the situation we should supplement (9.6) with the bound

$$\hat{W}_\varepsilon(\rho) \ll e^{-j}, \tag{9.10}$$

which comes from the case $m = 1$ of (7.10) and the derivative bounds (9.4). We then have

$$\sum_{j \geqslant 1} \sum_{\rho \in Z_j} |Y^\rho||\tilde{W}_\varepsilon(\rho)| \ll Y^{1/2} \sum_{j \geqslant 1} j \min(1, \frac{1}{\varepsilon} e^{-j}).$$

Set $J := 10 \log(1/\varepsilon)$. Using the first bound in the $\min(,)$ for $j \leqslant J$ and the second for $j > J$ gives

$$\sum_{j \geqslant 1} \sum_{\rho \in Z_j} |Y^\rho||\tilde{W}_\varepsilon(\rho)| \ll Y^{1/2} J^2.$$

(Note here that the sum of je^{-j} over $j > J$ is $\ll Je^{-J} < \varepsilon^9$, say, with each successive term being at most half the previous one.) The contribution from $\rho \in Z_0$ is simply $\ll Y^{1/2}$.

Analogously to before, we obtain

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

Choosing $\varepsilon = Y^{-2/3}$ (say) yields

$$\psi(4Y) - \psi(2Y) \leqslant 2Y + O(Y^{1/2} \log^2 Y).$$

Telescoping the sum in the usual manner gives

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

Once again, a corresponding lower bound may be obtained using an analogous argument.

10. *An introduction to sieve theory

10.1. **Introduction.** In this chapter we introduce a different kind of method in the study of prime numbers, the sieve. Sieve theory is an enormous topic, but we will only consider one very particular problem: that of bounding from above the quantity $\pi(X + Y) - \pi(X)$, the number of primes between X and $X + Y$.

We note to begin with that, at least if X is much larger than Y, the prime number theorem is totally useless. Moreover, we certainly do not expect any kind of asymptotic formula for this quantity, since there could be no primes at all in such an interval.

It is instructive to consider the efficacy of the most naive sieve method, the sieve of Erathosthenes (a.k.a. inclusion-exclusion). Writing $f(d) := \#\{n : X < n \leqslant X+Y, d \mid n\}$, an upper bound for $\pi(X+Y) - \pi(X)$ is

$$f(1) - f(2) - f(3) - \cdots + f(6) + f(10) + \cdots = \sum_{I \subset [k]} (-1)^{|I|} f(\prod_{i \in I} p_i),$$

where $p_1, \dots, p_k$ are the first k primes.

We have

$$f(d) = \frac{Y}{d} + O(1),$$

and in general it is not possible to say anything useful about the $O(1)$ term. Therefore these error terms of $O(1)$ add up to something that cannot be bounded by better than $O(2^k)$, and we have

$$\begin{aligned} \pi(X+Y) - \pi(X) &\leqslant Y \sum_{I \subset [k]} (-1)^{|I|} \prod_{i \in I} \frac{1}{p_i} + O(2^k) \\ &= Y \prod_{i=1}^{k} \big(1 - \frac{1}{p_i}\big) + O(2^k). \end{aligned}$$

Clearly we cannot take k larger than a multiple of $\log Y$ and hope to get a useful result. However, $\prod_{i=1}^{k}(1 - \frac{1}{p_i})$ is of size roughly $1/\log k$, and so the best bound this method can give is

$$\pi(X+Y) - \pi(Y) \ll \frac{Y}{\log\log Y}.$$

The main aim of this section is to prove the following result, giving the best possible order of magnitude.

Theorem 10.1. $\pi(X+Y) - \pi(X)$*, the number of primes between* X *and* $X+Y$ *is at most* $(2 + o(1))\frac{Y}{\log Y}$ *(with the* $o(1)$ *being as* $Y \to \infty$*).*

Remark. It is in fact known that the number of primes in this range is at most $2\pi(Y)$. The argument we give here can be used, with a little more care, to prove a result almost as good as this, namely that $\pi(X+Y) - \pi(X) \ll (2 + o(1))Y/\log Y$. It is a major unsolved problem to improve the constant 2, though it is conjecture that it can be replaced by 1. However, Hensley and Richards have shown that one should not expect that $\pi(X+Y) - \pi(X) \leqslant \pi(Y)$, in the sense that they have proved that this inequality fails if one assumes some widely-believed conjectures about configurations of primes.

10.2. **Selberg's weights.** To get an upper bound on $\pi(X+Y) - \pi(X)$ we use an idea of Selberg which, in retrospect, is very simple. Let $(\lambda_d)_{d \geqslant 1}$ be an sequence of real numbers with $\lambda_1 = 1$, let $R < Y$ be a threshold to be specified later (we will choose it to be Y^c for c slightly less than $\frac{1}{2}$), and consider the weight function

$$\nu(n) := \Big(\sum_{d \mid n : d \leqslant R} \lambda_d\Big)^2.$$

Evdiently $\nu(n) \geqslant 0$ for all n. Moreover, if n is a prime $> R$, then the sum $\sum_{d|n:d\leqslant R}\lambda_d$ collapses to simply $\lambda_1 = 1$, and therefore $\nu(n) = 1$. In other words, $\nu(n)$ is a majorant for the characteristic function of the primes, and hence

$$\pi(X+Y) - \pi(X) \leqslant \sum_{X<n\leqslant X+Y} \nu(n) + R. \qquad (10.1)$$

(The $+R$ term comes from the possibility that some primes $\leqslant R$ may lie in the interval $[X, X+Y]$; we have used the very crude bound R for the number of these.) What has been gained? It turns out that the sum on the right is relatively tractable. Expanding out the square, we have

$$\sum_{X\leqslant n\leqslant X+Y} \nu(n) = \sum_{X\leqslant n<X+Y} \Big(\sum_{d|n;d\leqslant R} \lambda_d\Big)^2 = \sum_{d_1,d_2\leqslant R} \lambda_{d_1}\lambda_{d_2} \sum_{\substack{X\leqslant n\leqslant X+Y\\ d_1,d_2|n}} 1.$$

The inner sum, that is to say the number of $n \in (X, X+Y]$ divisible by d_1 and d_2, is $\frac{Y}{[d_1,d_2]} + O(1)$ (where here $[a,b]$ means the l.c.m. of a and b) and so

$$\begin{aligned}\pi(X+Y) - \pi(Y) &\leqslant \sum_{X\leqslant n\leqslant X+Y} \nu(n) + R \\ &= Y\sum_{d_1,d_2\leqslant R} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} + O(1)\sum_{d_1,d_2\leqslant R} |\lambda_{d_1}||\lambda_{d_2}| + R. \qquad (10.2)\end{aligned}$$

At this point the λ_d are still completely arbitrary, subject only to $\lambda_1 = 1$. The strategy henceforth is to choose them to minimise the quadratic form

$$Q(\vec{\lambda}) := \sum_{d_1,d_2\leqslant R} \frac{\lambda_{d_1}\lambda_{d_2}}{[d_1,d_2]} = \sum_{d_1,d_2\leqslant R} \frac{\lambda_{d_1}\lambda_{d_2}}{d_1 d_2}(d_1,d_2),$$

which forms the main term above, and hope that the error term looks after itself. Under what conditions can the error term be "expected to look after itself?" If we assume that the λ_d are "reasonably bounded" in the sense that

$$\lambda_d \ll d^{o(1)}, \qquad (10.3)$$

then the error term here is $\ll R^{2+o(1)}$. If we choose $R = Y^c$ for some $c < 1/2$ then this will, if we take ε sufficiently small, be much smaller than $Y/\log Y$, the main term in our result.

To carry out this strategy, we need to diagonalise the form Q. To do this, we note that

$$(d_1,d_2) = \sum_{\delta|(d_1,d_2)} \phi(\delta),$$

where ϕ is Euler's ϕ-function. We showed this earlier in the course. Substituting in to the definition of Q yields

$$Q(\vec{\lambda}) = \sum_{\delta \leqslant R} \phi(\delta) \Big(\sum_{\substack{\delta | d \\ d \leqslant R}} \frac{\lambda_d}{d} \Big)^2 = \sum_{\delta \leqslant R} \phi(\delta) u_\delta^2, \tag{10.4}$$

where

$$u_\delta := \sum_{\substack{\delta | d \\ d \leqslant R}} \frac{\lambda_d}{d}.$$

We claim that the change of variables is invertible, or in other words that we can express λ_d in terms of the u_δ. This is a slight twist of the Möbius inversion formula; indeed we claim that

$$\frac{\lambda_d}{d} = \sum_{\substack{d | \delta \\ \delta \leqslant R}} \mu\big(\frac{\delta}{d}\big) u_\delta. \tag{10.5}$$

This is easily verified: the right-hand side is

$$\sum_{d' \leqslant R} \frac{\lambda_{d'}}{d'} \Big(\sum_{d | \delta | d'} \mu\big(\frac{\delta}{d}\big) \Big),$$

and the inner sum vanishes except when $d' = d$.

Note in particular that the constraint $\lambda_1 = 1$ becomes

$$1 = \sum_{\delta \leqslant R} \mu(\delta) u_\delta. \tag{10.6}$$

Minimising $Q(\vec{\lambda})$, as given in (10.4), subject to (10.6), is a standard task. Indeed by Cauchy-Schwarz we have

$$1 = \sum_{\delta \leqslant R} \mu(\delta) u_\delta \leqslant \Big(\sum_{\delta \leqslant R} \phi(\delta) u_\delta^2 \Big)^{1/2} \Big(\sum_{\delta \leqslant R} \frac{\mu^2(\delta)}{\phi(\delta)} \Big)^{1/2},$$

and moreover equality can occur by taking $u_\delta \propto \frac{\mu(\delta)}{\phi(\delta)}$. Thus the minimum value of $Q(\vec{\lambda})$ subject to (10.6) is $1/D$, where

$$D := \sum_{\delta \leqslant R} \frac{\mu^2(\delta)}{\phi(\delta)} = \sum_{\delta \leqslant R : \delta \text{ squarefree}} \frac{1}{\phi(\delta)},$$

this being attained when

$$u_\delta = \frac{\mu(\delta)}{D\phi(\delta)}.$$

To complete the proof, it is enough to show that

$$D \geqslant \log R, \tag{10.7}$$

as well as the fact that, with this choice of the u_δ, the λ_d as specified in (10.5) satisfy (10.3), that is to say $|\lambda_d| \ll_\varepsilon d^\varepsilon$. We handle these tasks in turn.

Proof of (10.7)*.* By definition we have

$$D = \sum_{d \leqslant R : d \, \text{squarefree}} \frac{1}{\phi(d)}.$$

If d is squarefree then $\phi(d) = \prod_{p|d}(1 - \frac{1}{p})$ and therefore this can be written as

$$\sum_{d \leqslant R : d \, \text{squarefree}} \frac{1}{d} \prod_{p|d} (1 + \frac{1}{p} + \frac{1}{p^2} \cdots).$$

Now every $m \leqslant R$ can be written as a product $dp_1^{\alpha_1} \cdots p_k^{\alpha_k}$, where d is squarefree and $p_i | d$ (simply take $d = p_1 \cdots p_k$, where the p_i are the primes dividing m). Therefore

$$D \geqslant \sum_{m \leqslant R} \frac{1}{m} \geqslant \log R,$$

which is what we wanted to prove.

Proof of (10.3)*.* By the choice of u_δ and (10.5) we have

$$\lambda_d = \frac{d}{D} \sum_{d | \delta, \delta \leqslant R} \frac{\mu(\delta/d)\mu(\delta)}{\phi(\delta)}.$$

Note that the sum is supported where δ is squarefree. Thus, writing $\delta' := \delta/d$, we have $\phi(\delta) = \phi(\delta')\phi(d)$. The sum is therefore bounded above by

$$\frac{d}{D\phi(d)} \sum_{\delta' \leqslant R, \delta' \, \text{squarefree}} \frac{1}{\phi(\delta')} = \frac{d}{\phi(d)}.$$

(Note that if $\mu(\delta) \neq 0$ then δ', d are coprime and hence $\phi(\delta) = \phi(d)\phi(\delta')$, and certainly δ' must be squarefree.) Thus we only need prove that

$$\phi(d) \gg_\varepsilon d^{1-\varepsilon} \tag{10.8}$$

when d is squarefree. This is clear if we factor d as a product of primes: if $p > p_0(\varepsilon)$ is a sufficiently large prime then $\phi(p) = p - 1 \geqslant p^{1-\varepsilon}$, whilst the contribution from the smaller primes is just a constant.

Appendix A. Some analysis results

In this section we collect some results from analysis which are necessary for the rigorous justification of some of the results of the course.

We begin with some results on differentiation and limits. The next result is Proposition 5.1 in my notes from the Oxford first-year course Analysis III: Integration.

Proposition A.1. *Suppose that $f_n : [a,b] \to \mathbf{C}$, $n = 1, 2, \dots$ is a sequence of functions with the property that f_n is continuously differentiable on (a,b), that f_n converges pointwise to some function f on $[a,b]$, and that f_n' converges uniformly to some bounded function g on (a,b). Then f is differentiable and $f' = g$.*

Remark. In my notes, the result was stated for functions with values in $\mathbf{R}$, but it applies to complex-valued functions by taking real and imaginary parts.

Corollary A.2. *Suppose that $u_n : [a,b] \to \mathbf{C}$ is a sequence of functions with the property that u_n is continuously differentiable on (a,b), and such that $|u_n(x)|, |u_n'(x)| \leqslant M_n$ with $\sum_{n=1}^{\infty} M_n < \infty$. Then $\sum_{n=1}^{\infty} u_n$ is differentiable, with derivative $\sum_{n=1}^{\infty} u_n'$.*

Proof. Follows by applying the above with $f_n = u_1 + \cdots + u_n$ and the Weierstrass M-test. □

We note that the result also holds if n ranges over $\mathbf{Z}$ (with $\sum_{n\in\mathbf{Z}} u_n$ defined to be $\sum_{n=-\infty}^{\infty} u_n = \lim_{N\to\infty} \sum_{n=-N}^{N} u_n$). It also holds for functions from $\mathbf{R}/\mathbf{Z}$ to $\mathbf{C}$, as can be seen by lifting such a function to $[0,2]$ (say).

Results of the following kind were used several times in the course to argue that a function defined by an integral is holomorphic by 'differentiating under the integral'.

Proposition A.3. *Let $I \subseteq [0,\infty)$ be a (possibly infinite) interval. Let $f : I \to \mathbf{R}$ be a function such that*

$$\int_I |f(x)| x^t dx < \infty \tag{A.1}$$

for all t in some (possibly infinite) open interval $T \subset \mathbf{R}$. For z with $\operatorname{Re} z \in T$, define

$$F(z) := \int_I f(x) x^z dx.$$

Then $F(z)$ is holomorphic in the domain $\mathcal{D} := \{z : \operatorname{Re} z \in T\}$ and

$$F'(z) = \int_I f(x) x^z \log x \, dx.$$

Proof. We verify the definition of derivative directly. Let $z \in \mathcal{D}$, and let δ be small enough that the ball $B_\delta(z)$ is contained within $\mathcal{D}$. Thus, the task is to show that

$$\lim_{h\to 0} \int_I f(x) x^z \Big(\frac{x^h - 1}{h} - \log x \Big) dx = 0.$$

Since the derivative of $x^h = e^{h \log x}$ at $h = 0$ is $\log x$, we do have

$$\lim_{h\to 0} \frac{x^h - 1}{h} = \log x$$

pointwise. Therefore, by the dominated convergence theorem, it is enough to show that

$$\int_I |f(x)x^z| \sup_{|h|\leqslant\delta}\Big|\frac{x^h-1}{h} - \log x\Big| dx < \infty. \tag{A.2}$$

Now $(x^z-1)/h$ is a holomorphic function of h (it has a removable singularity at $h=0$) and so by the maximum modulus principle we have

$$\sup_{|h|\leqslant\delta}\Big|\frac{x^h-1}{h}\Big| = \sup_{|h|=\delta}\Big|\frac{x^h-1}{h}\Big| \ll_\delta 1 + x^\delta,$$

here assuming $x \geqslant 0$. It follows that the LHS of (A.2) is bounded by

$$\ll_\delta \int_I |f(x)x^z|(1 + x^\delta + |\log x|)dx \ll \int_I |f(x)x^z|(1 + x^\delta + x^{-\delta})dx.$$

(For this last step, note that $|\log x|$ is either $\ll x^\delta$ (for $x \geqslant 1$) or $\ll x^{-\delta}$ (for $0 < x \leqslant 1$). This is finite by the assumption (A.1) and the fact that $z \pm \delta \in \mathcal{D}$. $\square$

Appendix B. *Some smooth bump functions

In this appendix we discuss some smooth functions – for example, smooth approximations to the interval $[-1,1]$. Traditionally a "trick" is used to construct these involving the function

$$f(t) := \begin{cases} e^{1/(1-t^2)} & |t| < 1; \\ 0 & |t| > 1, \end{cases}$$

which can be shown (a standard undergraduate exercise) to lie in $C^\infty(\mathbf{R})$.

I dislike this because it is a "trick". Furthermore, though we know from Lemma 4.1 that the Fourier transform $\hat{f}$ satisfies the estimate $|\hat{f}(\xi)| \leqslant C_m|\xi|^{-m}$ for $|\xi| \geqslant 1$, it is surprisingly difficult to get any effective bound for C_m.

In this appendix we present a more natural construction. This is no doubt "classical", but I learned of it in connection with (the easy direction of) something called the Denjoy-Carleman theorem.

The idea is to construct a function f as an infinite convolution of normalised characteristic functions of intervals. If $\delta > 0$, write

$$\nu_\delta(x) := \frac{1}{2\delta}1_{|x|\leqslant\delta}.$$

Thus $\|\nu_\delta\|_1 = 1$.

It is extremely easy to compute the Fourier transform of ν_δ: we have

$$\hat{\nu}_\delta(\xi) = \frac{1}{2\delta}\int_{-\delta}^{\delta} e^{-ix\xi}\, dx = \frac{\sin \delta\xi}{\delta\xi}.$$

In particular we have the trivial bound

$$|\hat{\nu}_\delta(\xi)| \leqslant \frac{2}{\delta|\xi|}. \tag{B.1}$$

Recall that the convolution of two functions f and g, $f * g$, is defined by

$$(f * g)(x) := \int f(x-y)g(y)dy.$$

Suppose that $f : \mathbf{R} \to [0,1]$ is a function with $f(x) = 1$ for $|x| \leqslant 1-\eta$ and $f(x) = 0$ for $|x| \geqslant 1+\eta$. Then it is easy to see that $f * \nu_\delta$ takes values in $[0,1]$, has $f(x) = 1$ for $|x| \leqslant 1-\eta-\delta$, and $f(x) = 0$ for $|x| \geqslant 1+\eta+\delta$. This observation leads quickly to the following lemma.

Lemma B.1. *Let $\varepsilon > 0$, and let $n \in N$. Then there is a function $f = f_{n,\varepsilon} : \mathbf{R} \to [0,1]$ such that $f(x) = 1$ for $|x| \leqslant 1-\varepsilon$, $f(x) = 0$ for $|x| \geqslant 1+\varepsilon$, and $|\hat{f}(\xi)| \leqslant C(n,\varepsilon)|\xi|^{-n-1}$. We may take $C(n,\varepsilon) = 2^{2n+1}(n!)^2\varepsilon^{-n}$.*

Proof. Define

$$f := \nu_1 * \nu_{\delta_1} * \cdots * \nu_{\delta_n},$$

where $\delta_j := \varepsilon/2j^2$. (There is considerable flexibility here; the important feature of the sequence $1/j^2$ is that it is summable but *slowly*.) Since $\sum j^{-2} < 2$, this has the support properties claimed.

We have $\hat{f} = \hat{\nu}_1\hat{\nu}_{\delta_1}\dots\hat{\nu}_{\delta_n}$, and so from the trivial bound (B.1) we obtain

$$|\hat{f}(\xi)| \leqslant \frac{2^{n+1}}{\delta_1\dots\delta_n}|\xi|^{-n-1}.$$

The result follows. □

A slightly more refined analysis allows one to show that, for fixed $\varepsilon > 0$, the sequence $f_{n,\varepsilon}$ converges as $n \to \infty$. This gives a version of Lemma B.1 in which f depends only on ε, and not on n.

Lemma B.2. *Let $\varepsilon > 0$ be fixed. Let $(\delta_j)_{j=1}^{\infty}$ be a sequence of positive real numbers with $\sum_j \delta_j \leqslant \varepsilon$. Set $f_n = f_{n,\varepsilon} := \nu_1 * \nu_{\delta_1} * \cdots * \nu_{\delta_n}$. Then f_n converges uniformly to a function $f : \mathbf{R} \to [0,1]$ with $f(x) = 1$ for $|x| \leqslant 1-\varepsilon$ and $f(x) = 0$ for $|x| \geqslant 1+\varepsilon$. The Fourier transform $\hat{f}(\xi)$ satisfies*

$$|\hat{f}(\xi)| \leqslant \inf_n \frac{2^{n+1}}{\delta_1\dots\delta_n}|\xi|^{-n-1}.$$

Proof. Since ε is fixed throughout the argument, we write f_n instead of $f_{n,\varepsilon}$. First note that $f_1 = \nu_1 * \nu_{\delta_1}$ satisfies the Lipschitz property

$$|f_1(x) - f_1(x')| \leqslant \frac{2}{\delta_1}|x-x'|.$$

Indeed

$$\begin{aligned}|f_1(x) - f_1(x')| &= \int \nu_1(y)(\nu_{\delta_1}(x-y) - \nu_{\delta_1}(x'-y))\,dy \\ &\leqslant \frac{1}{2}\int |\nu_{\delta_1}(x-y) - \nu_{\delta_1}(x'-y)|dy,\end{aligned}$$

which can easily be computed to be at most the claimed bound. Next note that any such Lipschitz bound is preserved under convolution with an arbitrary non-negative function ν with integral 1; indeed if $|f(x)-f(x')| \leqslant C|x-x'|$ for all x, x' then

$$\begin{aligned} |f*\nu(x) - f*\nu(x')| &= \Big|\int \nu(y)(f(x-y) - f(x'-y))dy\Big| \\ &\leqslant \int \nu(y)\big|f(x-y)-f(x'-y)\big|dy \\ &\leqslant C|x-x'|. \end{aligned}$$

It follows that each of the functions f_n satisfies the same Lipschitz bound as f_1.

Now suppose that f satisfies the Lipschitz bound $|f(x)-f(x')| \leqslant C|x-x'|$, and that ϕ is a non-negative function with integral 1, supported on $[-\delta, \delta]$. Then

$$\begin{aligned} \|f - f*\phi\|_\infty &= \sup_x \Big|\int (f(x)-f(x-y))\phi(y)dy\Big| \\ &\leqslant \sup_{|y|\leqslant\delta} |f(x)-f(x-y)| \int \phi(y)dy \leqslant C\delta. \end{aligned}$$

It follows from these observations that

$$\|f_m - f_n\|_\infty \leqslant \frac{2}{\delta_1}\sum_{j=m+1}^{n} \delta_j.$$

In particular, (f_n) is Cauchy in the uniform norm and hence f_n tends to some continuous function f.

To obtain the stated bound on the Fourier transform, first note that $\hat{f}(\xi) = \lim_{n\to\infty} \hat{f}_n(\xi)$. This follows immediately from the fact that $f_n \to f$ uniformly, and the support of f_n is contained in $[-2,2]$ for all n. Note also that if $\|\nu\|_1 = 1$ then

$$|(f_n * \nu)^\wedge(\xi)| = |\hat{f}_n(\xi)||\hat{\nu}(\xi)| \leqslant |\hat{f}_n(\xi)|.$$

This means that $|\hat{f}_n(\xi)|$ is a non-increasing function of n. The claimed bound now follows using (B.1). □

Corollary B.3. *Let $\varepsilon > 0$. Then for any $\kappa > 0$ there is a continuous function $f = f_{\varepsilon,\kappa} : \mathbf{R} \to [0,1]$ such that $f(x) = 1$ for $|x| \leqslant 1-\varepsilon$, $f(x) = 0$ for $|x| \geqslant 1+\varepsilon$ and*

$$|\hat{f}(\xi)| \leqslant e^{-C_{\varepsilon,\kappa}|\xi|^{1-\kappa}}.$$

Remark. Taking $\kappa = 1/2$ is certainly sufficient for any application I know of, and usually the bound $|\hat{f}(\xi)| \ll_A |\xi|^{-A}$, with no explicit dependence of the implied constant on A, is enough.

Proof. Apply the preceding lemma with $\delta_j := c/j^{1+\kappa/2}$ for an appropriately small $c = c(\varepsilon,\kappa)$. For every n we have the bound

$$|\hat{f}(\xi)| \leqslant C^n (n!)^{1+\kappa/2}|\xi|^{-n}.$$

Choosing $n \sim |\xi|^{1-\kappa/2}$ gives the claimed bound. □

We have said nothing about the smoothness (or otherwise) of the functions f_n and f. One could examine this using the definition in terms of convolution. However, now that we have bounds on Fourier transforms, we may as well use the following lemma.

Lemma B.4. *Suppose that f is continuous, that $f \in L^1(\mathbf{R})$ and that $|\hat{f}(\xi)| \ll |\xi|^{-n}$. Then f is $(n-2)$-times differentiable.*

Proof. (Sketch) One first needs to know that the inversion formula Proposition 4.6 holds for this f; this is not automatic, since we only proved it for Schwartz functions. However, the same proof works here with minor modifications. In particular, applying our bound on the Fourier transform with $n = 2$ means that all of the functions in the application of Lemma 4.4 are absolutely integrable, and so the use of Fubini is permissible here too.

Once one has the inversion formula, one may obtain the derivatives of f by repeated differentiation under the integral. □

Appendix C. Infinite products

In this section we supply the proofs of some very basic results on infinite products from the course. The following key inequality underlies everything.

Lemma C.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. Expanding out the product gives

$$\begin{aligned} \Big| \prod_{n=1}^N (1 + z_n) - 1 \Big| &= \Big| \sum_{i_1} z_i + \sum_{i_1 < i_2} z_{i_1} z_{i_2} + \sum_{i_1 < i_2 < i_3} z_{i_1} z_{i_2} z_{i_3} \dots \Big| \\ &\leqslant \sum_{i_1} |z_i| + \sum_{i_1 < i_2} |z_{i_1}||z_{i_2}| + \sum_{i_1 < i_2 < i_3} |z_{i_1}||z_{i_2}||z_{i_3}| \dots \\ &\leqslant \sum_i |z_i| + \frac{1}{2}\Big(\sum_i |z_i| \Big)^2 + \frac{1}{6}\Big(\sum_i |z_i| \Big)^3 + \dots \\ &= e^{\sum_i |z_i|} - 1. \end{aligned}$$

This concludes the proof. □

The following result concerns what are known as *Weierstrass Products*

Proposition C.2. *Suppose that $\Omega \subset \mathbf{C}$ is a countable multiset, not containing 0, and such that $\sum_{\rho \in \Omega} |\rho|^{-2} < \infty$. 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 with the correct multiplicities and nowhere else.

Proof. For each pair of positive integers M, N with $M < N$, define the truncated products

$$E_\Omega^{(M,N)}(z) := \prod_{\substack{\rho \in \Omega \\ M < |\rho| \leqslant N}} (1 - \frac{z}{\rho}) e^{z/\rho}.$$

Write

$$E_\Omega^{(N)} = E_\Omega^{(0,N)} = \prod_{\substack{\rho \in \Omega \\ |\rho| \leqslant N}} (1 - \frac{z}{\rho}) e^{z/\rho}.$$

for short. To make sense of E_Ω, it suffices to show that the sequence $(E_\Omega^{(N)}(z))_{N=1}^\infty$ is uniformly Cauchy "on compacta", that is to say on compact subsets of $\mathbf{C}$. If this can be shown, it then follows from standard results in complex analysis that $E_\Omega(z) := \lim_{N \to \infty} E_\Omega^{(N)}(z)$ exists and is holomorphic.

A first step is to show that the $E_\Omega^{(N)}(z)$ are uniformly bounded on compacta. To establish this, note first that the z/ρ are uniformly bounded by some quantity $K = K(R)$ as z ranges over $|z| \leqslant R$, because some ball about 0 contains none of the ρ. By Taylor expansion, the function $F(w) := \frac{(1-w)e^w - 1}{w^2}$ has a removable singularity at 0, and hence is bounded on $|w| \leqslant K$; let us say that $|F(w)| \leqslant C_K$ for $|w| \leqslant K$, thus

$$\Big|(1 - \frac{z}{\rho}) e^{z/\rho} - 1\Big| \leqslant C_K \frac{|z|^2}{|\rho|^2} \leqslant C_K R^2 |\rho|^{-2} \tag{C.1}$$

for all $|z| \leqslant R$. It now follows from Lemma C.1 that the $E_\Omega^{(N)}(z)$ are indeed uniformly bounded on compacta.

Now we show the Cauchy property. We start with the observation that

$$\begin{aligned} \big|E_\Omega^{(N)}(z) - E_\Omega^{(M)}(z)\big| &= |E_\Omega^{(M)}(z)| \cdot \Big| \prod_{M < |\rho| \leqslant N} (1 - \frac{z}{\rho}) e^{z/\rho} - 1 \Big| \\ &\ll_R \Big| \prod_{M < |\rho| \leqslant N} (1 - \frac{z}{\rho}) e^{z/\rho} - 1 \Big| \end{aligned}$$

uniformly for $|z| \leqslant R$. By Lemma C.1 and (C.1), it follows that

$$\big|E_\Omega^{(N)}(z) - E_\Omega^{(M)}(z)\big| \ll_R e^{O_R(\sum_{M < |\rho| \leqslant N} |\rho|^{-2})} - 1.$$

The Cauchy property is now immediate from the convergence of $\sum_\rho |\rho|^{-2}$.

We have now established that $E_\Omega(z)$ is well-defined as a holomorphic function. It remains to show that it has zeros with the correct multiplicity at the points ρ and at no other points. To this end it is enough to show that

$$\prod_{\rho \in \Omega} (1 - \frac{z}{\rho}) e^{z/\rho} \neq 0 \quad \text{when } z \notin \Omega. \tag{C.2}$$

Indeed, this obviously implies that $E_\Omega(z) \neq 0$ when $z \notin \Omega$, but it also implies that $E_\Omega(z)$ has a zero of exactly the right multiplicity $m(\rho)$ at ρ by writing

$$E_\Omega(z) = \big((1 - z/\rho)e^{z/\rho}\big)^{m(\rho)} E_{\Omega\setminus\{\rho\}}(z).$$

To establish (C.2), fix z. We need only consider the product over $|\rho| > R$, for large enough R, the remaining part of the product being finite. Here we use Lemma C.1 once again, together with the estimate

$$|(1 - z/\rho)e^{z/\rho} - 1| \ll \frac{|z|^2}{|\rho|^2},$$

which follows from the same argument used to prove (C.1), noting that we can assume that $|w| = \frac{|z|}{|\rho|} \leqslant 1$ by taking R large enough. This gives

$$\Big|\prod_{R<|\rho|\leqslant N}(1 + \frac{z}{\rho})e^{z/\rho} - 1\Big| \leqslant e^{O(z^2 \sum_{|\rho|>R} |\rho|^{-2})} \leqslant \frac{1}{2}$$

provided that R is big enough. Thus, in this range,

$$\Big|\prod_{R<|\rho|\leqslant N}(1 + \frac{z}{\rho})e^{z/\rho}\Big| \geqslant \frac{1}{2}.$$

Note that this bound does not depend on the choice of N in any way, so we may let $N \to \infty$ to conclude the proof. □

References

[1] H. Davenport, *Multiplicative number theory,* Graduate Texts in Mathematics, **74**. Springer-Verlag, New York, 2000. xiv+177 pp.

[2] E. Kowalski, *Un cours de théorie analytique des nombres,* Cours Spécialisés **13**, Société Mathématique de France, Paris, 2004.

[3] E. Kowalski and H. Iwaniec, *Analytic Number Theory,* American Mathematical Society Colloquium Publications, **53**. American Mathematical Society, Providence, RI, 2004.

[4] D. Koukoulopoulos, *The distribution of prime numbers,* Grad. Stud. Math., **203**, American Mathematical Society, Providence, RI, 2019. xii + 356 pp.