15 The Riemann zeta function and prime number theorem

We now divert our attention from algebraic number theory for the moment to talk about zeta functions and L-functions. These are analytic objects (complex functions) that are intimately related to the global fields we have been studying. We begin with the progenitor of all zeta functions, the Riemann zeta function.

For the benefit of those who have not taken complex analysis (which is not a formal prerequisite for this course) the next section briefly recalls some of the basic definitions and facts we we will need; these are elementary results covered in any introductory course on complex analysis, and we state them only at the level of generality we need, which is minimal. Those familiar with this material should feel free to skip to Section 15.2, but may want to look at Section 15.1.2 on convergence, which will be important in what follows.

15.1 A quick recap of some basic complex analysis

The complex numbers $\mathbb{C}$ are a topological field whose topology is defined by the distance metric $d(x,y) = |x-y|$ induced by the standard absolute value $|z| := \sqrt{z\bar{z}}$; all implicit references to the topology on $\mathbb{C}$ (open, compact, convergence, limits, etc.) are made with this understanding. For the sake of simplicity we restrict our attention to functions $f\colon \Omega \to \mathbb{C}$ whose domain Ω is an open subset of $\mathbb{C}$ (so Ω denotes an open set throughout this section).

15.1.1 Holomorphic and analytic functions

Definition 15.1. Let $f\colon \Omega \to \mathbb{C}$ be a function. The *derivative* of f at a point $z_0 \in \Omega$ is

$$f'(z_0) := \lim_{z\to z_0} \frac{f(z)-f(z_0)}{z-z_0}$$

whenever this limit exists. If it does we say that f is *differentiable at* z_0. If f is differentiable at every point in an open neighborhood of z_0, we say that f is *holomorphic* at z_0 (this is a stronger condition than being differentiable at z_0). If f is holomorphic at every point in an open set $U \subseteq \Omega$ we say that f is *holomorphic on* U. If f is holomorphic on Ω we say that f is a *holomorphic function*. Holomorphic functions on $\mathbb{C}$ are also called *entire functions*.

The derivative satisfies all the properties you would expect. The set of holomorphic functions on Ω form a $\mathbb{C}$-algebra $\mathcal{O}(\Omega)$ that is closed under composition on which differentiation is a linear operator and the usual product rule $(fg)' = f'g + fg'$ and chain rule $(fg)' = f'(g)g'$ hold. All polynomials are entire, as is the exponential function.

Theorem 15.2 (Holomorphic Identity Theorem)**.** *Let f and g be two complex functions that are both holomorphic on a connected open set Ω, and let U be a nonempty open subset of Ω. If the restrictions $f|_U$ and $g|_U$ coincide, then so do $f|_\Omega$ and $g|_\Omega$.*

Theorem 15.2 allows us to unambiguously extend the domain of a holomorphic function $f\colon U \to \mathbb{C}$ to a connected open set Ω containing U, provided we can find a holomorphic function $g\colon \Omega \to \mathbb{C}$ for which $g|_U = f$. The function g is necessarily the unique holomorphic function on Ω that restricts to f, and we call it the *analytic continuation* of f to Ω (it would make more sense to call it the *holomorphic continuation* of f to Ω, but as explained in Remark 15.6 below, the terms *analytic* and *holomorphic* may be used interchangeably).

Example 15.3. Let $f\colon U \to \mathbb{C}$ be the function defined by the series $\sum_{n\geq 0} z^n$, which converges on the unit disk $U := \{z : |z| < 1\}$. The series defining f diverges for $|z| \geq 1$, but the function $g(z) = 1/(1-z)$ agrees with f on U and is holomorphic on the connected open set $\Omega = \mathbb{C} - \{1\}$; thus g is the analytic continuation of f to Ω. The values of $g(z)$ at points $z \notin U$ cannot be meaningfully interpreted in the terms of the series defining f, but it is nevertheless amusing to do so; consider $g(2) = -1 =: 1 + 2 + 4 + 8 + \cdots$, for example. One can argue that if one wishes to assign a value to this divergent series, the only reasonable choice is -1, since this is the value that *must* be assigned by any holomorphic function that extends the domain of f to a connected open set containing 2.

Definition 15.4. A function $f\colon \Omega \to \mathbb{C}$ is *analytic at* z_0 if there is an open neighborhood $U \subseteq \Omega$ of z on which

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

Equivalently, f is holomorphic at z_0 and agrees with its *Taylor series expansion*

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(z)}{n!}(z - z_0)^n$$

at all points $z \in U$; here $f^{(n)}(z_0)$ denotes the nth derivative of f at z_0. If f is analytic at every point in an open set U we say that f is *analytic on* U.

Theorem 15.5. *A complex function is holomorphic on an open set U if and only if it is analytic on U.*

Remark 15.6. Theorem 15.5 implies that the terms "holomorphic" and "analytic" can be used interchangeably; modern usage tends to favor the former, but historically the latter was more commonly used.

15.1.2 Convergence

Recall that a series $\sum_{n=1}^{\infty} a_n$ of complex numbers *converges absolutely* if the series $\sum_n |a_n|$ of nonnegative real numbers converges. An equivalent definition is that the function $a(n) := a_n$ is integrable with respect to the counting measure μ on the set of positive integers $\mathbb{N}$; indeed if the series is absolutely convergent then

$$\sum_{n=1}^{\infty} a_n = \int_{\mathbb{N}} a(n)\mu,$$

and if the series is not absolutely convergent, then the integral not defined. Absolute convergence is effectively built-in to the definition of the Lebesgue integral, which requires that for $a(n) = x(n) + iy(n)$ to be integrable, the positive real functions $|x(n)|$ and $|y(n)|$ must both be integrable (equivalently, summable), and separately computes the sums of the positive and negative subsequences of $(x(n))$ and $(y(n))$ as suprema over finite subsets.

While equivalent, the measure-theoretic perspective has some advantages. It makes it immediately clear that we may replace the index set $\mathbb{N}$ with any set of the same cardinality, since the counting measure depends only on the cardinality of $\mathbb{N}$, not its ordering. We are thus free to sum over any countable index set, including $\mathbb{Z}$, $\mathbb{Q}$, any finite product of countable sets, and any countable coproduct of countable sets (such as countable direct sums of $\mathbb{Z}$); such sums are ubiquitous in number theory but generally have no canonical (or even meaningful) interpretation as limits of partial sums in the usual sense and must be understood as absolutely convergent sums in which the index set need not be ordered. The measure-theoretic view makes also makes it clear that we may convert any sum of the form $\sum_{X\times Y}$ into an iterated sum $\sum_X\sum_Y$ (or vice versa), via Fubini's theorem.

We say that an infinite product $\prod_n a_n$ of nonzero complex numbers is *absolutely convergent* when the sum $\sum_n \log a_n$ is, in which case $\prod_n a_n := \exp(\sum_n \log a_n)$.[1] This implies that an absolutely convergent product cannot converge to zero, and the sequence (a_n) must converge to 1 (no matter how we order the a_n). All of our remarks above about absolutely convergent series apply to absolutely convergent products as well.

A series or product of complex functions $\{f_n(z)\}$ is *absolutely convergent on S* if the series or product of complex numbers $\{f_n(z)\}$ is absolutely convergent for all $z \in S$.

A sequence of complex functions (f_n) *converges uniformly on S* if there is a function f such that for every $\epsilon > 0$ there is an integer N for which $\sup_{z\in S}|f_n(z) - f(z)| < \epsilon$ for all $n \geq N$. A series of complex functions $\sum f_n$ converges uniformly if the corresponding sequence of partial sums converges uniformly. A sequence or series of complex functions *converges locally uniformly* on S if it converges uniformly on every compact subset of S.

Proposition 15.7. *A series of holomorphic functions that converges absolutely and locally uniformly on an open set U converges to a holomorphic function on U. In particular, any power series that converges absolutely and locally uniformly on U uniquely determines a holomorphic function on U.*

15.1.3 Meromorphic functions

Definition 15.8. A function $f\colon \Omega \to \mathbb{C}$ is *meromorphic* on an open set U if there is a discrete set $S \subseteq U$ such that f is holomorphic on $U - S \subseteq \Omega$ and for all $z_0 \in S$ we have $\lim_{z\to z_0}|f(z)| = \infty$. The points $z_0 \in S$ are *poles* of f.

If z_0 is a pole of $f\colon \Omega \to \mathbb{C}$ then it is necessarily the case that $z_0 \not\in \Omega$, but there is an open neighborhood U of z_0 for which $U - \Omega = z_0$. Note that if f is meromorphic on its domain then it is technically a holomorphic function (if f has a pole at z_0 then z_0 is not in its domain).[2] The term *meromorphic function* generally refers to a function that is meromorphic on the interior of the closure of its domain, which is the largest set on which a function can be meromorphic. We may speak of the analytic continuation of a meromorphic function $f\colon U \to \mathbb{C}$ to a meromorphic function $g\colon \Omega \to \mathbb{C}$, where Ω is a connected open set containing U and $g|_U = F$.

Every meromorphic function is infinitely differentiable, and its derivatives are all meromorphic on the same set and have the same poles (but with higher orders). For any open set Ω the meromorphic functions on Ω form a field $F(\Omega)$ in which the usual quotient rule $(f/g)' = (f'g - g'f)/g^2$ holds (the domains of the functions in $F(\Omega)$ will be open sets U for which $S = \Omega - U$ is discrete). The field $F(\Omega)$ is the fraction field of the integral domain $\mathcal{O}(\Omega)$, and it contains all rational functions on Ω. It follows from Theorem 15.5 that every $f \in F(\Omega)$ has a Laurent series expansion

$$\sum_{n\geq n_0} a_n(z - z_0)^n$$

[1] In this definition we fix a branch of $\log z$, say $\log z := \log|z| + i \operatorname{Arg} z$ with $\operatorname{Arg} z \in (-\pi, \pi)$.

[2] One can instead consider functions that take values in $\mathbb{P}^1(\mathbb{C})$.

about any point $z_0 \in \Omega$, in which we may assume $a_{n_0} \neq 0$. Associated to each z_0 is a discrete valuation that assigns the integer n_0 to f; this discrete valuation is typically denoted $\operatorname{ord}_{z_0}(f)$ and called the *order of vanishing* of f at z_0; it is positive when f has a zero z_0 and negative when f has a pole at z_0. Equivalently, $\operatorname{ord}_{z_0}(f)$ is the greatest integer n_0 for which the function $f(z)/(z - z_0)^{n_0}$ is holomorphic at z_0. The coefficient of a_{-1} in the Laurent series expansion of f at z_0 is called the *residue* of f at z_0 and denoted $\operatorname{res}_{z_0}(f)$. When $\operatorname{ord}_{z_0}(f) = -1$ we say that f has a *simple pole* at z_0, and in this case res_{z_0} is the uniquely complex number a for which $f - a/(z - z_0)$ is holomorphic at z_0.

15.1.4 Contour integration

We shall restrict our attention to integrals along contours defined by piecewise-smooth parameterized curves; this covers all the cases we shall need.

Definition 15.9. A *parameterized curve* is a continuous function $\gamma\colon [a, b] \to \mathbb{C}$ whose domain is a compact interval $[a, b] \subseteq \mathbb{R}$. We say that γ is *smooth* if it has a continuous nonzero derivative on $[a, b]$, and *piecewise-smooth* if $[a, b]$ can be partitioned into finitely many subintervals on which the restriction of γ is smooth. We say that γ is *closed* if $\gamma(a) = \gamma(b)$, and *simple* if it is injective on $[a, b)$ and $(a, b]$. Henceforth we will use the term *curve* to refer to any piecewise-smooth parameterized curve γ, or to its oriented image of in the complex plane (directed from $\gamma(a)$ to $\gamma(b)$), which we may also denote γ.

Definition 15.10. Let $f\colon \Omega \to \mathbb{C}$ be a continuous function and let γ be a curve in Ω. We define the *contour integral*

$$\int_\gamma f(z)dz := \int_a^b f(\gamma(t))\gamma'(t)dt,$$

whenever the integral on the RHS (which is defined as a Riemann sum in the usual way) converges. Whether $\int_\gamma f(z)dz$ converges, and if so, to what value, does not depend on the parameterization of γ: if γ' is another parameterized curve with the same (oriented) image as γ, then $\int_{\gamma'} f(z)dz = \int_\gamma f(z)dz$.

We have the following analog of the fundamental theorem of calculus.

Theorem 15.11. *Let $\gamma\colon [a, b] \to \mathbb{C}$ be a curve in an open set Ω and let $f\colon \Omega \to \mathbb{C}$ be a holomorphic function Then*

$$\int_\gamma f'(z)dz = f(\gamma(b)) - f(\gamma(a)).$$

Recall that the Jordan curve theorem implies that every simple closed curve γ partitions $\mathbb{C}$ into two components, one of which we may unambiguously designate as the *interior* (the one on the left of our *positively oriented* curves). We say that γ is *contained* in an open set U if both γ and its interior lie in U.

Theorem 15.12 (Cauchy's Theorem). *Let U be an open set containing a simple closed curve γ. For any function f that is holomorphic on U we have*

$$\int_\gamma f(z)dz = 0.$$

Cauchy's theorem generalizes to meromorphic functions.

Theorem 15.13 (Cauchy Residue Formula)**.** *Let U be an open set containing a simple closed curve γ. Let f be a functions that is meromorphic on U, let $z_1, \ldots, z_n$ be the poles of f that lie in the interior of γ, and suppose that no pole of f lies on γ. Then*

$$\int_\gamma f(z)dz = 2\pi i \sum_{i=1}^n \operatorname{res}_{z_i}(f).$$

To see where the $2\pi i$ comes from, consider $\int_\gamma \frac{dz}{z}$ with $\gamma(t) = e^{it}$ for $t \in [0, 2\pi]$.

Cauchy's residue formula can be used to recover the coefficients $f^{(n)}(a)/n!$ appearing in the Laurent series expansion of a meromorphic function at a (apply it to $f(z)/(z-a)^{n+1}$). One of many useful consequences of this is Liouville's theorem, which can be proved by showing that the Laurent series expansion of a bounded holomorphic function on $\mathbb{C}$ about any point has only one nonzero coefficient (the constant coefficient) and an infinite radius of convergence.

Theorem 15.14 (Liouville's theorem)**.** *Bounded entire functions are constant.*

We also have the following converse of Cauchy's theorem.

Theorem 15.15 (Morera's Theorem)**.** *Let f be a continuous function and on an open set U, and suppose that for every simple closed curve γ contained in U we have*

$$\int_\gamma f(z)dz = 0.$$

Then f is holomorphic on U.

Corollary 15.16. *A sequence or series of functions holomorphic on an open set U that converges locally uniformly on U converges to a holomorphic function on U.*

Theorem 15.17 (Weierstrass M-test)**.** *Let (f_n) be a sequence of functions holomorphic on an open set U, and suppose there are positive real numbers M_n for which the series $\sum_{n\ge 1} M_n$ converges and such that $|f_n(z)| \le M_n$ for all $z \in C$ for every compact $C \subseteq U$. The series $\sum_n f_n(z)$ converges to a holomorphic function f on U for which $f'(z) = \sum_{n\ge1} f_n'(z)$.*

15.2 The Riemann zeta function

Definition 15.18. The *Riemann zeta function* is the complex function defined by

$$\zeta(s) := \sum_{n\ge 1} n^{-s},$$

for $\operatorname{Re}(s) > 1$. The series converges absolutely and locally uniformly for $\operatorname{Re}(s) > 1$ and thus defines a holomorphic function on $\operatorname{Re}(s) > 1$, since each $n^{-s} = e^{-s\log n}$ is entire.

Theorem 15.19 (Euler product)**.** *For $\operatorname{Re}(s) > 1$ we have the identity*

$$\zeta(s) = \sum_{n\ge1} n^{-s} = \prod_p (1 - p^{-s})^{-1},$$

where the the absolutely convergent product ranges over primes; thus $\zeta(s) \neq 0$ for $\operatorname{Re}(s) > 1$.

The product in the theorem above ranges over primes p. This is a standard practice in analytic number theory that we will follow: the symbol p always denotes a prime, and any sum or product over p is understood to be over primes.

Proof. The one-line proof is that unique factorization and absolute convergence imply

$$\sum_{n\geq 0} n^{-s} = \sum_{e_2,e_3,\ldots\geq 0} (2^{e_2}3^{e_3}\cdots)^{-s} = \prod_p \sum_{e\geq 0} p^{-es} = \prod_p (1-p^{-s})^{-1}.$$

However the middle equality deserves further justification. For each integer $m \geq 1$, let S_m be the set of m-smooth numbers: positive integers with prime factors $p \leq m$. Now define

$$\zeta_m(s) := \sum_{n\in S_m} n^{-s},$$

which converges absolutely and locally uniformly on $\mathrm{Re}(s) > 1$. If $p_1, \ldots, p_k$ are the primes up to m, then we may write the absolutely convergent sum as

$$\zeta_m(s) = \sum_{n\in S_m} n^{-s} = \sum_{e_1,\ldots,e_k\geq 0} (p_1^{e_1}\cdots p_k^{e_k})^{-s} = \sum_{e_1\geq 0} p_1^{-e_1 s} \sum_{e_2\geq 0} p_2^{-e_2 s} \cdots \sum_{e_k\geq 0} p_k^{e_k s}.$$

For $\mathrm{Re}(s) > 1$ we have $\sum_{e\geq 0} p^{-es} = 1 + p^{-s} + p^{-2s} + \cdots = (1-p^{-s})^{-1}$, for any prime p. Applying this k times yields the finite product

$$\zeta_m(s) = \prod_{p\leq m} (1-p^{-s})^{-1}.$$

We now note that for any $\delta > 0$ the sequence of functions $\zeta_m(s)$ converges uniformly on $\mathrm{Re}(s) > 1 + \delta$ to $\zeta(s)$; indeed, for any $\epsilon > 0$ and any such s we have

$$|\zeta_m(s) - \zeta(s)| \leq \left|\sum_{n\geq m} n^{-s}\right| \leq \sum_{n\geq m} |n^{-s}| = \sum_{n\geq m} n^{-\mathrm{Re}(s)} \leq \int_m^\infty x^{-1-\delta} dx \leq \frac{1}{\delta} m^{-\delta} < \epsilon$$

for all sufficiently large m. It follows that the sequence $\zeta_m(s)$ converges locally uniformly to $\zeta(s)$ on $\mathrm{Re}(s) > 1$, and therefore the sequence of functions $P_m(s) := \prod_{p\leq m}(1-p^{-s})^{-1}$ does as well. The sequence $\log P_m(s)$ clearly converges to $\log \prod_p (1-p^{-s})^{-1}$, and

$$\sum_p |\log(1-p^{-s})^{-1}| = \sum_p \left|\sum_{e\geq 1} \frac{1}{e} p^{-es}\right| \leq \sum_p \sum_{e\geq 1} |p^{-s}|^e = \sum_p (|p^s| - 1)^{-1} < \infty$$

is absolutely convergent (hence finite), thus $\prod_p (1-p^{-s})^{-1}$ is absolutely convergent (hence nonzero); here we have used the $\log(1-z) = -\sum_{n\geq 1} z^n$ for $|z| < 1$. □

Theorem 15.20 (Analytic continuation I)**.** *For* $\mathrm{Re}(s) > 1$ *we have*

$$\zeta(s) = \frac{1}{s-1} + \phi(s),$$

where $\phi(s)$ *is a holomorphic function on* $\mathrm{Re}(s) > 0$*. Thus* $\zeta(s)$ *extends to a meromorphic function on* $\mathrm{Re}(s) > 0$ *that has a simple pole at* $s = 1$ *with residue* 1 *and no other poles.*

Proof. For $\operatorname{Re}(s) > 1$ we have

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

We now define $\phi_n(s) := \int_n^{n+1}(n^{-s}-x^{-s})dx$. For any s and $x\in[n,n+1]$ we have

$$|n^{-s}-x^{-s}| = \left|\int_n^x st^{-s-1}dt\right| \leq \int_n^x \frac{|s|}{t^{1+\operatorname{Re}(s)}}dt \leq \frac{|s|}{n^{1+\operatorname{Re}(s)}}$$

Thus

$$|\phi_n(x)| \leq \int_n^{n+1}\left|n^{-s}-x^{-s}\right|ds \leq \frac{|s|}{n^{1+\operatorname{Re}(s)}}.$$

We now note that for $\operatorname{Re}(s)\geq\epsilon>0$ we have

$$\sum_{n\geq 1}\frac{|s|}{n^{1+\operatorname{Re}(s)}} < \infty.$$

Each ϕ_n is holomorphic, so by the Weierstrass M-test, the series $\sum_{n\geq 1}\phi_n(s)$ converges to a function $\phi(s)$ that is holomorphic on $\operatorname{Re}(s)>0$ (and clearly equal to $\zeta(s)-\frac{1}{s-1}$). □

We now wish to show that $\zeta(s)$ has no zeros on $\operatorname{Re}(s)=1$, this is the key to proving the prime number theorem. For this we rely on the following lemma.

Lemma 15.21. *For all $x,y\in\mathbb{R}$ with $x>1$ we have $|\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)|\geq 1$.*

Proof. From the Euler product $\zeta(s)=\prod_p(1-p^{-s})^{-1}$, we see that for $\operatorname{Re}(s)>1$ we have

$$\log|\zeta(s)| = -\sum_p \log|1-p^{-s}| = -\sum_p \operatorname{Re}\log(1-p^{-s}) = \sum_p\sum_{n\geq 1}\frac{\operatorname{Re}(p^{-ns})}{n},$$

where we have used the general facts $\log|z| = \operatorname{Re}\log z$ and $\log(1-z) = -\sum_{n\geq 1}\frac{z^n}{n}$ for $|z|<1$ (note that $\operatorname{Re}(s)>1$ implies $|p^{-s}| = p^{-\operatorname{Re}(s)}<1$). Applying this to $s=x+iy$ yields

$$\log|\zeta(x+iy)| = \sum_p\sum_{n\geq 1}\frac{\cos(ny\log p)}{np^{nx}}$$

Thus

$$\log|\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)| = \sum_p\sum_{n\geq 1}\frac{3+4\cos(ny\log p)+\cos(2ny\log p)}{np^{nx}}.$$

We now note that the identity $\cos(2\theta) = 2\cos^2\theta - 1$ implies

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

Taking $\theta = ny\log p$ yields $\log|\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)|\geq 0$, which proves the lemma. □

Corollary 15.22. *$\zeta(s)$ has no zeros on* $\mathrm{Re}(s) \geq 1$*.*

Proof. We know from Theorem 15.19 that $\zeta(s)$ has no zeros on $\mathrm{Re}(s) > 1$, so suppose $\zeta(1+iy) = 0$ for some $y \in \mathbb{R}$. Then $y \neq 0$, since $\zeta(s)$ has a pole at $s = 1$, and we know that $\zeta(s)$ does not have a pole at $1 + 2iy \neq 1$, by Theorem 15.20. We therefore must have

$$\lim_{x\to 1} |\zeta(x)^3\zeta(x+iy)^4\zeta(x+2iy)| = 0, \tag{1}$$

since $\zeta(s)$ has a simple pole at $s = 1$, a zero at $1 + iy$, and no pole at $1 + 2iy$, but this contradicts Lemma 15.21. □

15.3 The Prime Number Theorem

The prime counting function $\pi\colon \mathbb{R} \to \mathbb{Z}_{\geq 0}$ is defined by

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

it counts the number of primes up to x. The Prime Number Theorem (PNT) states that

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

The notation $f(x) \sim g(x)$ means $\lim_{x\to\infty} f(x)/g(x) = 1$; one says that f is *asymptotic to*, in other words, the functions f and g grow at the same asymptotic rate.

This conjectured growth rate for $\pi(x)$ dates back to Gauss and Legendre in the late 18th century. In fact Gauss believed the asymptotically equivalent but more accurate statement[3]

$$\pi(x) \sim \mathrm{Li}(x) := \int_2^\infty \frac{dx}{\log x}.$$

However it was not until a century later that the prime number theorem was independently proved by Hadamard [2] and de la Vallée Poussin [7] in 1896, building on the work of Riemann [5], who in 1860 showed that there is a precise connection between the zeros of $\zeta(s)$ and the distribution of primes (we shall say more about this later), but was unable to prove the prime number theorem.

The proof we will give is more recent and due to Newman [4], but it relies on the same properties of the Riemann zeta function that were exploited by both Hadamard and de la Vallée, the most essential of which is the fact that $\zeta(s)$ has no zeros on $\mathrm{Re}(s) \geq 1$ (Corollary 15.22). A wonderfully concise version of Newman's proof by Zagier can be found in [9]; we shall be slightly more expansive here. We should note that there are also "elementary" proofs of the prime number theorem obtained by Erdös [1] and Selberg [6] in the 1940s that do not use the Riemann zeta function, but they are elementary only in the sense that they do not use complex analysis; the elementary proofs are actually more complicated than those that use complex analysis.

Rather than work directly with $\pi(x)$, it is more convenient to work with the log-weighted prime-counting function defined by Chebyshev[4]

$$\vartheta(x) := \sum_{p\leq x} \log p,$$

whose growth rate differs from that of $\pi(x)$ by a logarithmic factor.

[3] More accurate in the sense that $|\pi(x) - \mathrm{Li}(x)|$ grows more slowly than $|\pi(x) - \frac{x}{\log x}|$ as $x \to \infty$.

[4] As with most Russian names, there is no canonical way to write Chebyshev in the latin alphabet and one finds many variations in the literature; in English, the spelling Chebyshev is now the most widely used.

Theorem 15.23 (Chebyshev)**.** *We have $\pi(x) \sim \frac{x}{\log x}$ if and only $\vartheta(x) \sim x$.*

Proof. We clearly have $0 \le \vartheta(x) \le \pi(x)\log x$, thus

$$\frac{\vartheta(x)}{x} \le \frac{\pi(x)\log x}{x}.$$

For every $\epsilon > 0$ we have

$$\begin{aligned}\vartheta(x) \ge \sum_{x^{1-\epsilon}<p\le x} \log p &\ge (1-\epsilon)(\log x)\big(\pi(x) - \pi(x^{1-\epsilon})\big)\\ &\ge (1-\epsilon)(\log x)(\pi(x) - x^{1-\epsilon}),\end{aligned}$$

and therefore

$$\pi(x) \le \frac{1}{1-\epsilon}\frac{\vartheta(x)}{\log x} + x^{1-\epsilon}.$$

Thus for all $\epsilon > 0$ we have

$$\frac{\vartheta(x)}{x} \le \frac{\pi(x)\log x}{x} \le \frac{1}{1-\epsilon}\frac{\vartheta(x)}{x} + \frac{\log x}{x^{\epsilon}}.$$

The last term tends to 0 as $x \to \infty$, and the lemma follows: by choosing ϵ appropriately we can make the ratios of $\vartheta(x)$ to x and $\pi(x)$ to $x/\log x$ arbitrarily close together as $x \to \infty$, and if one of them tends to 1, then so must the other. □

In view of Chebyshev's result, the prime number theorem is equivalent to the statement $\vartheta(x) \sim x$, which is what we will prove. The first step is to show that the asymptotic growth rate of $\vartheta(x)$ is at most linear in x.

Lemma 15.24 (Chebyshev)**.** *For all $x \ge 1$ we have $\vartheta(x) \le (4\log 2)x$; thus $\vartheta(x) = O(x)$.*

Proof. For any integer $n \ge 1$, the binomial theorem implies

$$2^{2n} = (1+1)^{2n} = \sum_{m=0}^{2n}\binom{2n}{m} \ge \binom{2n}{n} = \frac{(2n)!}{n!n!} \ \ge \prod_{n<p\le 2n} p = \exp(\vartheta(2n) - \vartheta(n)),$$

Since $(2n)!$ is divisible by every prime $p \in (n, 2n]$ but $n!$ is not divisible by any such p. Taking logarithms on both sides yields the bound

$$\vartheta(2n) - \vartheta(n) \le 2n\log 2,$$

for all integers $n \ge 1$. For any integer $m \ge 1$ we have

$$\vartheta(2^m) = \sum_{n=1}^{m}\big(\vartheta(2^n) - \vartheta(2^{n-1})\big) \le \sum_{n=1}^{m} 2^n \log 2 \ \le\ 2^{m+1}\log 2.$$

For any real $x \ge 1$ we can choose an integer $m \ge 1$ so that $2^{m-1} \le x < 2^m$, and then

$$\vartheta(x) \le \vartheta(2^m) \le 2^{m+1}\log 2 = (4\log 2)2^{m-1} \le (4\log 2)x,$$

as claimed. □

In order to prove $\vartheta(x) \sim x$, we will use a general analytic criterion that is applicable to any non-decreasing real function.

Lemma 15.25. *Let $f\colon \mathbb{R}_{\geq 1} \to \mathbb{R}$ be a non-decreasing function for which the integral $\int_1^\infty \frac{f(t)-t}{t^2}dt$ converges. Then $f(x) \sim x$.*

Proof. Let $F(x) := \int_1^x \frac{f(t)-t}{t^2}dt$. The hypothesis is that $\lim_{x\to\infty} F(x)$ exists. This implies that for all $\lambda > 1$ and all $\epsilon > 0$ we must have $|F(\lambda x) - F(x)| < \epsilon$ for all sufficiently large x.

Consider any $\lambda > 1$. Suppose there is an unbounded sequence (x_n) such that $f(x_n) \geq \lambda x_n$ for all $n \geq 1$. Then for each x_n we have

$$F(\lambda x_n) - F(x_n) = \int_{x_n}^{\lambda x_n} \frac{f(t)-t}{t^2}dt \geq \int_{x_n}^{\lambda x_n} \frac{\lambda x_n - t}{t^2}dt = \int_1^\lambda \frac{\lambda - t}{t^2}dt = c,$$

for some $c > 0$, where we used the fact that f is non-decreasing to get the middle inequality. Taking $\epsilon < c$, we have $|F(\lambda x_n) - F(x_n)| = c > \epsilon$ for arbitrarily large x_n, a contradiction. Thus $f(x) < \lambda x$ for all sufficiently large x. A similar argument shows that $f(x) > \frac{1}{\lambda}x$ for all sufficiently large x. These inequalities hold for all $\lambda > 1$, so we must have $\lim_{x\to\infty} f(x)/x = 1$, equivalently, $f(x) \sim x$. □

We now recall the Laplace transform.

Definition 15.26. Let $h\colon \mathbb{R}_{>0} \to \mathbb{R}$ be a piecewise continuous function. The *Laplace transform* of h is the complex function defined by

$$(\mathcal{L}h)(s) := \int_0^\infty e^{-st}h(t)dt;$$

it is a holomorphic function on $\operatorname{Re}(s) > c$ for any $c \in \mathbb{R}$ for which $h(t) = O(e^{ct})$.

The following properties of the Laplace transform are easy to verify:

- $\mathcal{L}(g + h) = \mathcal{L}g + \mathcal{L}h$, and for any $a \in \mathbb{R}$ we have $\mathcal{L}(ah) = \alpha\mathcal{L}h$.
- If $h(t) = a \in \mathbb{R}$ is constant then $\mathcal{L}h = \frac{a}{s}$.
- $\mathcal{L}(e^{at}h)(s) = (\mathcal{L}h)(s - a)$ for all $a \in \mathbb{R}$.

We now define the auxiliary function

$$\Phi(s) := \sum_p p^{-s}\log p,$$

which is related to $\vartheta(x)$ by the following lemma.

Lemma 15.27. $\mathcal{L}(\vartheta(e^t))(s) = \frac{\Phi(s)}{s}$ *is a holomorphic function on* $\operatorname{Re}(s) > 1$.

Proof. By Lemma 15.24, $\vartheta(e^t) = O(e^t)$, so $\mathcal{L}\vartheta(e^t)$ is holomorphic on $\operatorname{Re}(s) > 1$. Let p_n be the nth prime, and put $p_0 := 0$. The function $\vartheta(e^t)$ is constant on $(\log p_n, \log p_{n+1})$, so

$$\int_{\log p_n}^{\log p_{n+1}} e^{-st}\vartheta(e^t)dt = \vartheta(p_n)\int_{\log p_n}^{\log p_{n+1}} e^{-st}dt = \frac{1}{s}\vartheta(p_n)\big(p_n^{-s} - p_{n+1}^{-s}\big).$$

We then have

$$
\begin{aligned}
(\mathcal{L}\vartheta(e^t))(s) = \int_0^\infty e^{-st}\vartheta(e^t)dt &= \frac{1}{s}\sum_{n=1}^\infty \vartheta(p_n)\big(p_n^{-s} - p_{n+1}^{-s}\big)\\
&= \frac{1}{s}\sum_{n=1}^\infty \vartheta(p_n)p_n^{-s} - \frac{1}{s}\sum_{n=1}^\infty \vartheta(p_{n-1})p_n^{-s}\\
&= \frac{1}{s}\sum_{n=1}^\infty \big(\vartheta(p_n) - \vartheta(p_{n-1})\big)p_n^{-s}\\
&= \frac{1}{s}\sum_{n=1}^\infty p_n^{-s}\log p_n = \frac{\Phi(s)}{s}. \qquad \square
\end{aligned}
$$

Let us now consider the function $H(t) := \vartheta(e^t)e^{-t} - 1$. It follows from the lemma and standard properties of the Laplace transform that on $\mathrm{Re}(s) > 0$ we have

$$
(\mathcal{L}H)(s) = \mathcal{L}(\vartheta(e^t)e^{-t})(s) - (\mathcal{L}1)(s) = \mathcal{L}(\vartheta(e^t))(s+1) - \frac{1}{s} = \frac{\Phi(s+1)}{s+1} - \frac{1}{s}.
$$

Lemma 15.28. *The function* $\Phi(s) - \frac{1}{s-1}$ *extends to a meromorphic function on* $\mathrm{Re}(s) > \frac{1}{2}$ *that is holomorphic on* $\mathrm{Re}(s) \geq 1$.

Proof. The logarithmic derivative $\frac{\zeta'(s)}{\zeta(s)}$ of $\zeta(s)$ is meromorphic on $\mathrm{Re}(s) > 0$, since (the extension of) $\zeta(s)$ is. In term of the Euler product we have

$$
\begin{aligned}
-\frac{\zeta'(s)}{\zeta(s)} &= -\log\left(\prod_p (1-p^{-s})^{-1}\right)' = \left(\sum_p \log(1-p^{-s})\right)'\\
&= \sum_p \frac{p^{-s}\log p}{1-p^{-s}} = \sum_p \frac{\log p}{p^s - 1} = \sum_p \left(\frac{1}{p^s} + \frac{1}{p^s(p^s-1)}\right)\log p\\
&= \Phi(s) + \sum_p \frac{\log p}{p^s(p^s-1)}.
\end{aligned}
$$

The sum on the RHS converges absolutely and locally uniformly to a holomorphic function on $\mathrm{Re}(s) > 1/2$. The LHS is holomorphic on $\mathrm{Re}(s) > 1$, since $\zeta(s)$ has no zeros or poles in this region; moreover the LHS has only a simple pole of residue 1 at $s = 1$ on $\mathrm{Re}(s) = 1$, since $\zeta(s)$ has no zeros on $\mathrm{Re}(s) = 1$ and a simple pole of residue 1 at $s = 1$ (by Theorem 15.20). It follows that $\Phi(s) - \frac{1}{s-1}$ extends to a meromorphic function on $\mathrm{Re}(s) > \frac{1}{2}$ that is holomorphic on $\mathrm{Re}(s) \geq 1$. □

Corollary 15.29. *The functions* $\Phi(s+1) - \frac{1}{s}$ *and* $(\mathcal{L}H)(s) = \frac{\Phi(s+1)}{s+1} - \frac{1}{s}$ *both extend to meromorphic functions on* $\mathrm{Re}(s) > -\frac{1}{2}$ *that are holomorphic on* $\mathrm{Re}(s) \geq 0$.

Proof. The first statement is immediate. For the second we note that

$$
(\mathcal{L}H)(s) = \frac{\Phi(s+1)}{s+1} - \frac{1}{s} = \frac{1}{s+1}\left(\Phi(s+1) - \frac{1}{s}\right) - \frac{1}{s+1}
$$

is holomorphic on $\mathrm{Re}(s) \geq 0$, since it is a sum of products of such functions. □

The final step of the proof relies on the following analytic result.

Theorem 15.30. *Let $f\colon \mathbb{R}_{\ge 0}\to\mathbb{R}$ be a bounded piecewise continuous function, and suppose its Laplace transform extends to a holomorphic function $g(s)$ on $\operatorname{Re}(s)\ge 0$. Then the integral $\int_0^\infty f(t)dt$ converges and is equal to $g(0)$.*

This theorem is an example of what is known as a *Tauberian theorem.* The Laplace transform

$$(\mathcal{L}f)(s) := \int_0^\infty e^{-st}f(t)dt,$$

is in general not defined on $\operatorname{Re}(s)\le c$, where c is the least c for which $f(t)=O(e^{ct})$. It may happen that the function $\mathcal{L}f$ has an analytic continuation to a larger domain; for example, if $f(t)=e^t$ then $(\mathcal{L}f)(s)=\frac{1}{s-1}$ extends to a holomorphic function on $\mathbb{C}-\{1\}$. But plugging values of s with $\operatorname{Re}(s)\le c$ into the integral usually does not work; in our $f(t)=e^t$ example, the integral diverges on $\operatorname{Re}(s)\le 1$. The theorem says that when $\mathcal{L}f$ extends to a holomorphic function on the entire half-plane $\operatorname{Re}(s)\ge 0$, its value at $s=0$ is exactly what would get by plugging 0 into the integral defining $\mathcal{L}f$, even though you are in general not allowed to do this.

This theorem is not difficult to prove, but as it is has no particular number-theoretic content, we will not take the time to do so; see [9] for a short proof. We now ready to prove the prime number theorem.

Theorem 15.31 (Prime Number Theorem)**.** $\pi(x)\sim\frac{x}{\log x}$.

Proof. The function $H(t)=\vartheta(e^t)e^{-t}-1$ is bounded (by Lemma 15.24) and piecewise continuous, and its Laplace transform extends to a holomorphic function on $\operatorname{Re}(s)\ge 0$, by Corollary 15.29. Theorem 15.30 then implies that the integral

$$\int_0^\infty H(t)dt = \int_0^\infty \left(\vartheta(e^t)e^{-t}-1\right)dt$$

converges. Replacing t with $\log x$, we see that

$$\int_1^\infty \left(\vartheta(x)\frac{1}{x}-1\right)\frac{dx}{x} = \int_1^\infty \frac{\vartheta(x)-x}{x^2}dx$$

converges, and Lemma 15.25 then implies $\vartheta(x)\sim x$, and this is equivalent to $\pi(x)\sim\frac{x}{\log x}$ by Theorem 15.23. □

One disadvantage of our proof is that it does not give an error term. Using more sophisticated methods, Korobov [3] and Vinogradov [8] independently obtained the bound

$$\pi(x)=\operatorname{Li}(x)+O\left(\frac{x}{\exp\left((\log x)^{3/5+o(1)}\right)}\right),$$

in which we note that the error term is bounded by $O(x/(\log x)^n)$ for all n but is not bounded by $O(x^{1-\epsilon})$ for any $\epsilon>0$. Assuming the Riemann Hypothesis, which states the all zeros of $\zeta(s)$ in the critical strip $0<\operatorname{Re}(s)<1$ lie on the line $\operatorname{Re}(s)=\frac{1}{2}$, one can prove

$$\pi(x)=\operatorname{Li}(x)+x^{1/2+o(1)}.$$

There thus remains a large gap between what we can prove about the distribution of prime numbers and what we believe to be true. Remarkably, other than refinements to the $o(1)$ term appearing in the Korobov-Vinogradov bound, essentially no progress has been made in this direction in the past 50 years.

References

[1] Paul Erdös, *On a new method in elementary number theory which leads to an elementary proof of the prime number theorem*, Proc. Nat. Acad. Scis. U.S.A. **35** (1949), 373–384.

[2] Jacques Hadamard, *Sur la distribution des zéros de la function $\zeta(s)$ et ses conséquences arithmétique*, Bull. Soc. Math. France **24** (1896), 199–220.

[3] Nikolai M. Korobov, *Estimates for trigonometric sums and their applications*, Uspechi Mat. Nauk **13** (1958), 185–192.

[4] David J. Newman, *Simple analytic proof of the Prime Number Theorem*, Amer. Math. Monthly **87** (1980), 693–696.

[5] Bernhard Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie, 1859.

[6] Alte Selberg, *An elementary proof of the Prime-Number Theorem*, Ann. Math. **50** (1949), 305–313.

[7] Charles Jean de la Vallée Poussin, *Reserches analytiques sur la théorie des nombres premiers*, Ann. Soc. Sci. Bruxelles **20** (1896), 183–256.

[8] Ivan M. Vinogradov, *A new estimate of the function $\zeta(1+it)$*, Izv. Akad. Nauk SSSR. Ser. Mat. **22** (1958), 161–164.

[9] Don Zagier, *Newman's short proof of the Prime Number Theorem*, Amer. Math. Monthly **104** (1997), 705–708.