Morfismos, Vol. 9, No. 2, 2005, pp. 1–48

Riemann and his zeta function *

Elena A. Kudryavtseva [1] Filip Saidak Peter Zvengrowski

Abstract

An exposition is given, partly historical and partly mathematical, of the Riemann zeta function $\zeta(s)$ and the associated Riemann hypothesis. Using techniques similar to those of Riemann, it is shown how to locate and count non-trivial zeros of $\zeta(s)$. Relevance of these investigations to the theory of the distribution of prime numbers is discussed.

2000 Mathematics Subject Classification: 11M06, 11M26, 11A41, 11N05.
Keywords and phrases: meromorphic functions, Riemann zeta function, gamma function, Riemann hypothesis.

1 Introduction

The aim of this note is to give a straightforward introduction to some of the mysteries associated with the Riemann zeta function $\zeta(s)$ of a complex variable s and the Riemann hypothesis (usually written RH) about the location of its zeros, both from an historical and a mathematical perspective. The mathematical development will be largely self contained, and understandable to readers having a basic acquaintance with real and complex analysis. We hope to elucidate the answers to the following questions:

(a) What is the Riemann zeta function?

(b) What is the RH?

(c) Why is this conjecture considered so important?

**Invited Article.*

[1] Partially supported by a PIMS Visiting Research Fellowship.

(d) Using techniques available to Riemann, how can one actually locate zeros of ζ?

(e) How much did Riemann know about RH (did he even consider it relevant)?

(f) What is the history of this problem since Riemann?

(g) What is some of the current research being done on RH? (In particular, recent work of the authors will be briefly mentioned in this context.)

The zeta function is intimately connected with the distribution of the primes. Indeed one of Riemann's primary motivations for studying it was to prove the Prime Number Theorem, cf. (13). Discussion about the distribution of primes will therefore be included (cf. §4). Another extremely important aspect of the Riemann zeta function is its very significant generalizations, however we only give the briefest of introductions to this.

The outline of the paper is as follows. Section 2 clarifies the notations used. In §3 meromorphic functions f and their zeros are introduced, including results for the functions $f = \sin,\ \Gamma,\ \zeta$ that will be used in the sequel. The RH is stated here. In §4 the history of the zeta function and the distribution of primes, from Euclid [27] through Riemann [76], is sketched. The next two sections develop the mathematical theory of ζ and its zeros starting with basic results such as the intermediate value theorem from real analysis and the argument principle from complex analysis, and leading to the location of the first three zeros of ζ along the critical line. As mentioned in (d), this will be done using techniques that Riemann himself may well have used. In §7 we return to the historical perspective, discuss Weierstrass' contributions, and address questions (c), (e), (f), and (g), including Siegel's very important 1932 study [80] of Riemann's *Nachlass*. Appendices A, B give short proofs, respectively, of the Prime Number Theorem and von Mangoldt's Theorem. Details of the proof of the Riemann–von Mangoldt explicit formula 4.6 appear in Appendix C, as well as further discussion, partly speculative, of question (e).

2 Notation

All notations used in this paper are standard, however we list some of them here for completeness and convenience.

$\log z = \ln z$	the natural logarithm of a complex number z
$\lfloor x \rfloor$	the greatest integer $\leq x$, also floor of x
$\lceil x \rfloor$	the nearest integer to a real number x, $x \notin \frac{1}{2} + \mathbb{Z}$
$\{x\}$	$x - \lfloor x \rfloor$, the fractional part of x
$f(x) \sim g(x)$	f is asymptotic to g, i.e. $\lim\limits_{x\to\infty} \dfrac{f(x)}{g(x)} = 1$
$\approx$	approximately equal, for two complex numbers
γ	Euler's constant, $\gamma := \lim\limits_{N\to\infty} \left(\sum\limits_{n=1}^{N} \dfrac{1}{n} - \log N \right) \approx 0.5772$
$\sum\limits_{n}, \sum\limits_{p}, \prod\limits_{p}$	respectively $\sum\limits_{n=1}^{\infty}, \sum\limits_{p \text{ prime}}, \prod\limits_{p \text{ prime}}$
$\sum\limits_{n\leq x}, \sum\limits_{p\leq x}$	respectively $\sum\limits_{n=1}^{\lfloor x \rfloor}, \sum\limits_{p \text{ prime},\ p\leq x}$
$\sum\limits_{\rho}$	the sum taken over all zeros ρ of ζ (or ξ) in the entire critical strip, with their multiplicities, and in order of increasing $\lvert \mathrm{Im}(\rho) \rvert$

For functions f, g of a complex (or real) variable z, where g is positive real valued, we define

(a) $f(z) = o(g(z))$ if $\lim_{|z|\to\infty} |f(z)|/g(z) = 0$,

(b) $f(z) = O(g(z))$ if there exists a constant $C > 0$ such that $|f(z)| \leq Cg(z)$ as $|z| \to \infty$.

For any piecewise continuous function f of a real variable x, with only jump discontinuities, we define $\tilde{f}(x) := \frac{1}{2}\lim\limits_{\varepsilon\to 0}(f(x+\varepsilon) + f(x-\varepsilon))$.

3 Meromorphic functions

Recall that an *entire* function $f\colon \mathbb{C} \to \mathbb{C}$ is one that is complex differentiable (i.e. holomorphic, equivalently analytic) at each $s \in \mathbb{C}$. One calls a function *meromorphic* on a subset $A \subseteq \mathbb{C}$ if it is defined on some open neighbourhood U of A, except at a discrete (possibly empty) subset $S \subseteq A$, and is holomorphic everywhere on $U \setminus S$ with poles at the points of S. The remainder of this section gives examples of meromorphic functions (as well as their poles and zeros) that will be important in the subsequent development of the Riemann zeta function $\zeta(s)$. Throughout this note we write $s = \sigma + it$ for a complex variable,

$\sigma = \mathrm{Re}(s)$, $t = \mathrm{Im}(s)$ being respectively the real and imaginary parts of s, a tradition that goes back (at least) to Landau [57] in 1909.

Example 3.1 Let $f(s) = \frac{p(s)}{q(s)}$ be a *rational* function, where p, q are relatively prime polynomials of degree m, n respectively. Then f has m zeros and n poles, where zeros are always counted with their multiplicities, and poles with their orders. A zero of multiplicity 1, or a pole of order 1, is called *simple*. As a specific illustration of this type, consider $f(s) = \frac{s^3+1}{s^2-4is-4}$. Here f has the three (simple) zeros $s = -1$, $\frac{1}{2} \pm \frac{\sqrt{3}}{2}i$, and the pole of order two $s = 2i$. In this way, any two disjoint finite sets with assigned multiplicities (respectively orders) can be obtained for the zeros and poles of a meromorphic function, which can be taken to be a rational function.

As the following examples show, for more general meromorphic functions the number of zeros or poles can well be infinite.

Remark 3.2 For any non-constant meromorphic function the numbers of zeros and poles are at most countably infinite, since it is standard that the sets of zeros and poles must be topologically discrete subsets of $\mathbb{C}$ (cf. [58], III, §1 and V, §3). The latter property will be useful in the sequel.

Example 3.3 Let $f(s) = \sin(s)$. Then f is an entire function (hence meromorphic with no poles), and all zeros lie on the real axis, namely $s = n\pi$, $n \in \mathbb{Z}$. To verify the statement about the zeros, recall that

$$\sin(s) = \sin(\sigma + it) = \sin(\sigma)\cosh(t) + i\cos(\sigma)\sinh(t).$$

An easy calculation now shows that $|\sin(s)|^2 = \sin^2(\sigma) + \sinh^2(t)$, hence $\sin(s) = 0$ implies both $\sinh(t) = 0$ and $\sin(\sigma) = 0$, i.e. $t = 0$ and $\sigma = n\pi$.

We remark that the function sin satisfies the well known functional equations $\sin(s + \pi) = -\sin(s)$, $\sin(-s) = -\sin(s)$.

Example 3.4 $f(s) = \Gamma(s)$. The gamma function is usually introduced in real analysis courses via the integral due to Euler [28]:

$$\Gamma(s) = \int_0^\infty \frac{x^{s-1}}{e^x}\, dx = \int_0^\infty \frac{x^s}{e^x}\, \frac{dx}{x}\,, \qquad s \in \mathbb{R},\ s > 0. \tag{1}$$

The second form for this integral is called the *Mellin transform* [70] of $\frac{1}{e^x}$. (The integral (1) can also be viewed as a Laplace transform.) The requirement $s > 0$ guarantees convergence. An easy exercise in integration by parts shows that $\Gamma(s) = (s-1)\Gamma(s-1)$ for $s > 1$. Also $\Gamma(1) = 1$ is clear, hence $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$. Exactly the same arguments work if $s = \sigma + it \in \mathbb{C}$, $\sigma > 0$, and thus the above integral defines a holomorphic function $\Gamma(s)$ for $\sigma > 0$, satisfying the functional equation $\Gamma(s) = (s-1)\Gamma(s-1)$, $s \in \mathbb{C}$ and $\sigma > 1$. Using the functional equation and the principle of analytic continuation extends Γ to a meromorphic function on $\mathbb{C}$ with simple poles at $s = 0, -1, -2, \ldots$.

Other basic properties of the gamma function are the reflection formula of Euler

$$(2) \qquad \Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}, \qquad s \notin \mathbb{Z} \quad (\text{i.e. } \sin(\pi s) \neq 0),$$

and the duplication formula of Legendre

$$(3) \qquad \Gamma(s)\Gamma(s+\frac{1}{2}) = \sqrt{\pi}\; 2^{1-2s}\; \Gamma(2s), \qquad 2s \notin \mathbb{Z}_{\leq 0}.$$

From (2) it is easy to see that $\Gamma(s) \neq 0$ for all s in its domain. An interesting historical discussion of the gamma function is given in [21], and proofs of (2), (3) can be found in the classic work of Artin [3] as well as many other texts. For example, in [58], XV, §2, proofs are given based on the Weierstrass product formula (30).

Example 3.5 $f(s) = \zeta(s)$. The Riemann zeta function will be described in much more detail in §4 and thereafter. Here we introduce it by three equivalent formulae, the usual Dirichlet series, the Euler product, and a Mellin transform expression similar to (1):

$$(4) \quad \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \frac{1}{1-p^{-s}} = \frac{1}{\Gamma(s)} \int_0^\infty \frac{x^s}{e^x - 1}\, \frac{dx}{x}, \qquad \sigma > 1.$$

The integral representation of $\zeta(s)$ in (4), at least for $s \in \mathbb{R}$, $s > 1$, is due to Abel [1] in 1823. Riemann [76] obtained it by making the change of variable $x = tn$ in the definition (1), and then summing for all $n \geq 1$, as shown by the following sequence of formulae, for $\sigma > 1$:

$$\Gamma(s) = \int_0^\infty \frac{x^{s-1}}{e^x}\, dx, \quad \frac{\Gamma(s)}{n^s} = \int_0^\infty \frac{t^{s-1}}{e^{tn}}\, dt, \quad \text{and}$$

$$\Gamma(s)\zeta(s) = \int_0^\infty \frac{t^{s-1}}{e^t - 1}\, dt.$$

In all three cases in (4) the condition $\sigma > 1$ is necessary for convergence. Again, by analytic continuation (cf. (10), or (16), (17)), $\zeta(s)$ can be extended to a meromorphic function on $\mathbb{C}$ with a single (simple) pole at $s = 1$, and satisfying a functional equation relating $\zeta(1-s)$ and $\zeta(s)$, cf. (18). Using the functional equation we shall see in §5 that ζ has (simple) zeros at $s = -2, -4, -6, \ldots$. These are called the *trivial zeros*, and we shall also see that the functional equation implies that all other zeros, the *non-trivial zeros*, lie in the *critical strip* $0 \le \sigma \le 1$. The line $\sigma = \frac{1}{2}$ is called the *critical line.*

The **Riemann Hypothesis** asserts that, for any non-trivial zero $s = \sigma + it$ of ζ, $\sigma = 1/2$, i.e. all non-trivial zeros of $\zeta(s)$ lie on the critical line.

Remark 3.6 All functions f considered in 3.3, 3.4, 3.5 satisfy $f(s) \in \mathbb{R}$ for all real s in their domain, which is equivalent to $f(\bar{s}) = \overline{f(s)}$ for any meromorphic function on $\mathbb{C}$. This means that all their zeros are real or occur in conjugate pairs $\sigma \pm it$.

4 History from Euclid to Riemann

Let us now look at some of the fundamental ideas and theorems that played an important rôle in the historical development of the theory of the Riemann zeta function, up to and including Riemann's monumental 1859 paper [76], which is also quite remarkable since it is only eight pages long (see Appendix of [25] for an English translation). It should be mentioned that while much of this work (including Riemann's), coming before modern standards of mathematical rigour were introduced to analysis by Weierstrass and his successors (notably Hardy), falls short of what would be considered acceptable proofs today, this in no way detracts from the originality and significance of this pioneering work.

The Fundamental Theorem of Arithmetic, originating in Book IX of Euclid's *Elements* [27] (Proposition 14), states that every $n \in \mathbb{N}$ has a unique representation, up to order, as a product of prime numbers

$$n = p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} = \prod_{p_i | n} p_i^{m_i}, \qquad m_i \ge 1. \tag{5}$$

Here the *existence* of a prime factorization easily follows by induction. *Uniqueness*, likely first proved by Gauss in 1801 [37] (although tacitly assumed by many prior authors), also can be proved by induction and Euclid's lemma (Proposition 30 of [27]), i.e. $p|ab \Longrightarrow p|a$ or $p|b$. Euclid used the existence of a prime decomposition to show that there are infinitely many primes [27] (Proposition 20).

The Fundamental Theorem of Algebra, proved by Gauss in his Thesis [36] of 1799, states that if $P(z)$ is a polynomial of degree $n > 0$ with complex coefficients, then $P(z)$ has a unique (again up to order) factorization into n monic factors of degree 1 and a constant non-zero factor, over the complex numbers. In other words, $P(z)$ has n zeros (or roots) in $\mathbb{C}$, counted with multiplicities, and factors as

$$P(z) = a(z - z_1)(z - z_2) \cdots (z - z_n) = a \prod_{i=1}^{n} (z - z_i), \tag{6}$$

for $z_i \in \mathbb{C}$ and $a \in \mathbb{C} \setminus \{0\}$. Partial results (for situations in $\mathbb{R}$) had been obtained by Girard [39] in 1629 and Descartes (his Rule of Signs) [23] in 1637, and again this theorem was tacitly assumed by various authors prior to Gauss.

In the early 1730s, Euler found new, ingenious ways to combine (5) and (6) with theorems from analysis in order to prove new results in number theory. In 1737, in his *Variae observationes* [32] he used (5) to prove that the function $\zeta(s)$ has the product representation (4) for all real $s > 1$. Its significance comes from the fact that, for the very first time, one has an explicit link between prime numbers, natural numbers, and analysis, that can be used to study the distribution of primes (Euler mainly considered the special case where s is an integer, $s > 1$). As an immediate application of this product representation, the divergence of the harmonic series $\sum_n \frac{1}{n} = \zeta(1)$ gives a new proof of Euclid's theorem on the infinitude of primes. As a second application, taking logarithms of both sides of the product representation in (4), Euler himself was able to obtain, for $s > 1$,

$$\begin{aligned} \log \zeta(s) &= \sum_p \log \left(1 - \frac{1}{p^s}\right)^{-1} \\ &= \sum_p \frac{1}{p^s} + \sum_p \sum_{m=2}^{\infty} \frac{1}{mp^{ms}} = \sum_p \frac{1}{p^s} + R(s), \end{aligned} \tag{7}$$

where $0 < R(s) < \sum_p \sum_{m=2}^{\infty} \frac{1}{2p^m} = \frac{1}{2}\sum_p \frac{1}{p(p-1)} < \frac{1}{2}\sum_{n=2}^{\infty} \frac{1}{n(n-1)} = \frac{1}{2}$. Letting $s \to 1^+$, we now see that the divergence of the harmonic series $\sum_n \frac{1}{n} = \zeta(1)$ implies that $\sum_p \frac{1}{p} = \infty$, a non-trivial statement of Euler [32] about the density of the primes.

Remark 4.1 Using Euler-Maclaurin summation (cf. §6), discovered by Euler [29] in 1732, and Maclaurin [67] in 1742, Euler found:

$$(8) \qquad \sum_{n \le x} \frac{1}{n} = \int_1^x \frac{dt}{t} - \int_1^x \frac{\{t\}}{t^2} dt + 1 - \frac{\{x\}}{x} \in (\log x, 1 + \log x),$$

and then guessed the 1874 theorem of Mertens [71]:

$$\sum_{p \le x} \frac{1}{p} \sim \log\log x.$$

The inequality (similar to (7)) $\log \sum_{n \le x} \frac{1}{n} < \sum_{p \le x} \frac{1}{p} + \frac{1}{2}\sum_{p \le x} \frac{1}{p(p-1)}$ and (8) imply a weaker, but still very interesting lower bound

$$\sum_{p \le x} \frac{1}{p} > \log\left(\prod_{p \le x}\left(1 - \frac{1}{p}\right)^{-1}\right) - \frac{1}{2}\sum_{p \le x} \frac{1}{p(p-1)} > \log\log x - \frac{1}{2}$$

for all $x \ge 2$. This gives a second non-trivial statement about the density of the primes, strengthening the conclusion $\sum_p \frac{1}{p} = \infty$ from (7).

In 1734 (see [30] and [31]), Euler showed that $\zeta(2) = \frac{\pi^2}{6}$, a difficult question proposed by Cavalieri's student Mengoli [97] as early as 1650. Factoring the function $\sin x$ in terms of its zeros: $0, \pm\pi, \pm 2\pi, \ldots$, as if it were a polynomial in (6), Euler found its product representation (see also 7.3), and he equated it with the Taylor series of $\sin x$,

$$\begin{aligned} \sin x &= x \prod_{n=1}^{\infty} \left[\left(1 - \frac{x}{\pi n}\right)\left(1 + \frac{x}{\pi n}\right)\right] = x \prod_{n=1}^{\infty}\left(1 - \frac{x^2}{\pi^2 n^2}\right) \\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots . \end{aligned}$$

Comparing the coefficients at x^3 immediately gives us $-\sum_n \frac{1}{\pi^2 n^2} = -\frac{1}{3!}$. Another proof of the formula for $\zeta(2)$ can be obtained using the Fourier series expansion of $f(x) = x^2$, $-\pi \le x \le \pi$, evaluated at $x = \pi$. See also 5.2 for another method.

Remark 4.2 In 1837, Dirichlet [24] generalized parts of Euler's work on the zeta function in two significant ways. First, in (4), he now thought of $s \in \mathbb{R}$, $s > 1$, whereas Euler had mainly considered cases where $s \in \mathbb{Q}$, $s > 1$, see also (18). Second, Dirichlet introduced the generalization of (4) and of (17)

$$(9) \qquad L(s,\chi) = \sum_n \frac{\chi(n)}{n^s} = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}, \qquad s > 1,$$

where χ is a Dirichlet character modulo a prime q (we do not define this here, it is not necessary for the further discussion). Using this he generalized Euler's argument and proved his celebrated theorem [24] that, for any coprime a and b, we have

$$\sum_{p \equiv b \pmod a} \frac{1}{p} = \infty.$$

He thereby proved a famous conjecture of Legendre [60], that any arithmetical progression $\{an + b \,|\, n \in \mathbb{N}\}$, where a, b are relatively prime integers, contains an infinitude of prime numbers.

From the work of Euler and Dirichlet, it became clear that analytical methods were a powerful tool in number theory.

The main reason Riemann, who was a student of Dirichlet, was able to make tremendous advances in the theory of the zeta function, was the growth of the new field of complex analysis, created by Fourier, Cauchy, Gauss, and others in the period 1800-1830. In his thesis, submitted to the University of Göttingen in 1851, Riemann himself vastly enlarged this new branch of analysis. Such basic notions as the Cauchy-Riemann equations, the Riemann mapping theorem, and Riemann surfaces are among his many contributions to the subject, especially to that part now called geometric function theory. Probably no mathematician, for at least the 50 years following Riemann's death (at age 39, in 1866), came close to his mastery of geometric function theory, which he used to good advantage in his work on the zeta function.

In 1859, Riemann defined $\zeta(s)$ as a function of a complex variable s. The first step was to extend (or to *analytically continue*) the definition (4) of $\zeta(s)$ to all of $\mathbb{C} \setminus \{1\}$. This can be accomplished by noticing that, for $\sigma > 0$, $n^{-s} = s\int_n^\infty x^{-s-1}dx$, and so

$$(10) \qquad \zeta(s) = \sum_{n=1}^\infty \frac{1}{n^s} = \sum_{n=1}^\infty \left(s \int_n^\infty \frac{dx}{x^{s+1}}\right) = s \sum_{n=1}^\infty \int_n^\infty \frac{dx}{x^{s+1}}$$

$$= s\int_1^\infty \Big(\sum_{n\le x} 1\Big)\frac{dx}{x^{s+1}} \;=\; s\int_1^\infty \frac{\lfloor x\rfloor}{x^{s+1}}\,dx \;=\; s\int_1^\infty \frac{x-\{x\}}{x^{s+1}}\,dx$$

$$= \frac{s}{s-1} \;-\; s\int_1^\infty \frac{\{x\}}{x^{s+1}}\,dx \qquad \text{for} \quad \sigma > 1\,.$$

Since $\{x\} \in [0,1)$, it follows that the last integral converges for $\sigma > 0$, and defines a continuation of $\zeta(s)$ to the half-plane $\sigma = \mathrm{Re}(s) > 0$. If one continues this process, one can extend $\zeta(s)$ to a holomorphic function on all of $\mathbb{C} \setminus \{1\}$. One sees from (10) that $s = 1$ is a simple pole with residue 1. See also (16) for Riemann's original technique, or (17).

Remark 4.3 Note that for s real, $s > 0$, the last integral in (10) is always positive real. It follows at once from (10) that $\zeta(s) < 0$, $s \in (0,1)$, and clearly $\zeta(s) > 1$ for $s \in (1,\infty)$.

Remark 4.4 The next step of this continuation process gives

$$\zeta(s) \;=\; \frac{s}{s-1} \;-\; \frac{1}{2} \;-\; \frac{s(s+1)}{2}\int_1^\infty \frac{\{x\}^2-\{x\}}{x^{s+2}}\,dx, \quad \sigma > -1.$$

It follows at once that $\zeta(0) = -\frac{1}{2}$. Applying Stirling's formula (22) and the definition of γ, one also shows $\zeta'(0) = -\frac{1}{2}\log(2\pi)$.

In order to see a deeper connection between $\zeta(s)$ and prime numbers, let us now follow Riemann's ideas [76] and employ the logarithmic derivative of the Euler product (4), using (7):

$$(11) \qquad -\frac{\zeta'(s)}{\zeta(s)} = \sum_p \sum_{m=1}^\infty \frac{\log p}{p^{ms}} \;=\; \sum_{n=2}^\infty \frac{\Lambda(n)}{n^s} \qquad \text{for} \quad \sigma > 1,$$

where $\Lambda(n)$ is von Mangoldt's function [68], defined as $\Lambda(p^m) = \log p$ for a prime power p^m, and 0 otherwise. We also define the Chebyshev function (cf. [13])

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

Note that (11) can also be written, for $\sigma > 1$, as

$$(12) \quad -\frac{\zeta'(s)}{\zeta(s)} \;=\; \sum_p \sum_{m=1}^\infty \frac{\log p}{p^{ms}} \;=\; \int_1^\infty x^{-s}d\psi(x) \;=\; s\int_1^\infty \frac{\psi(x)}{x^{s+1}}\,dx.$$

We now turn to the early developments related to the Prime Number Theorem (PNT), stated in (13) below, a much stronger statement about the asymptotic distribution of the primes (e.g. compared to $\sum_p \frac{1}{p} = \infty$). In a letter written to the astronomer (his former student) Encke in 1849 (cf. [38] or Appendix B of [40]), Gauss stated that he observed as early as 1792 or 1793 (when he was only 16) that the density of primes around a number x appears to be on the average inversely proportional to $\log x$, and therefore the logarithmic integral $\mathrm{Li}(x) := \int_2^x \frac{dt}{\log t}$ should provide a good approximation to the prime counting function $\pi(x) := \sum_{p \le x} 1$. Gauss' work on this question, both as a youth and in his 1848 letter, was empirical in nature and unpublished.

Prime Number Theorem (PNT):

$$\pi(x) \sim \mathrm{Li}(x). \tag{13}$$

Remark 4.5 The PNT was proved in 1896, cf. Appendix A. The RH is closely related to refining the PNT further by estimating the error in the approximation (13), indeed, it is equivalent to this error being $O(\sqrt{x}\log x)$, cf. [54] and §5.5 of [25]. See also A.2 and A.3.

The first published account is probably due to Legendre [60] in 1798, again based on empirical observations. Legendre conjectured that, for x large, $\pi(x) \approx x/(\log x - 1.08366)$. Let us mention here that Legendre's formula clearly implies $\pi(x) \sim \frac{\log x}{x}$, and this in turn is equivalent to PNT, due to (15). Gauss, in his 1849 letter, compared Legendre's formula to $\mathrm{Li}(x)$ for values of $x = 5 \times 10^5$, 10^6, $1.5 \times 10^6, \dots, 3 \times 10^6$. He noted that while the Legendre formula seemed to have smaller deviations from $\pi(x)$, these deviations seemed to be growing more rapidly than for $\mathrm{Li}(x)$, and therefore it was "quite possible they may surpass them" (i.e. the deviations of the Legendre formula would eventually become larger than for $\mathrm{Li}(x)$).

The two memoirs published by Chebyshev [12], [13] in 1848 and 1850 comprised the first mathematical attack on the PNT. In [13] he defined the function $\psi(x)$ (above), also defined $\theta(x) := \sum_{p \le x} \log p$, realized that $\psi(x)$ is nearly equal to x (e.g. for $n = 10^2$, 10^3, 10^4, 10^5, $\psi(n)/n =$ 0.94045, 0.99668, 1.00134, 1.00051 respectively), and can be estimated more easily than $\pi(x)$. He was able to prove that, for x sufficiently large, it satisfies the inequality $Ax < \psi(x) < Bx$ with $A = 0.92129$, $B = 1.10555$. He used this to show that, for x large,

$$0.89\frac{x}{\log x} \;<\; \pi(x) \;<\; 1.11\frac{x}{\log x} \tag{14}$$

(see also [25], §1.1).[2] As an application, Chebyshev [13] gave the first proof of Bertrand's *postulate*[3], as well as obtained the following interesting result: for any positive non-increasing function $F = F(n)$, $n \in \mathbb{N}$, the series $\sum_n F(n)$ and $\sum_p F(p) \log p$ either both converge or both diverge. Chebyshev [12] also proved that $\lim_{x\to\infty} \frac{\pi(x)}{x/\log x}$, if it exists (possibly infinite), equals 1, and $\lim_{x\to\infty} (\frac{x}{\pi(x)} - \log x)$, if it exists (possibly infinite), equals -1. As a consequence, based on the asymptotic expansion of $\mathrm{Li}(x)$:

$$\mathrm{Li}(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} + \ldots + (n-1)!\frac{x}{\log^n x} + o\left(\frac{x}{\log^n x}\right), \tag{15}$$

for any fixed n, he showed (by taking $n = 1$ and $n = 2$) that $\frac{x}{\log x - 1}$ provides the best approximation to $\pi(x)$ among all formulae of the form $\frac{x}{A \log x - B}$ (in particular, both $\frac{x}{\log x}$ and Legendre's empirical formula), provided that the above two limits exist. Validity of the latter assumption was finally confirmed in 1899, see §7 and A.2.

Riemann's original motivation for his study of the zeta function was to obtain an explicit formula such as (49) for $\pi(x)$, similar to 4.6 below, and to prove the PNT. Being aware of Chebyshev's work (indeed Chebyshev had met Riemann's mathematical mentor Dirichlet in 1852[4]), Rie-

[2] Chebyshev's method was used later to give an elementary proof of PNT, see Appendix A.

[3] In his group-theoretical investigation in 1845, Bertrand used the following proposition which he only verified within the limits of tables of primes: for each integer $x \geq 7$ there exists a prime $p \in (\frac{x}{2}, x-2]$, cf. [57], §4. In fact Chebyshev [13] strengthened this by proving $\pi(2x) - \pi(x) > \frac{3}{5} \frac{x}{\log(2x)}$ for x sufficiently large.

[4] "In the summer of 1852, however, he [Chebyshev] was sent on an official mission, lasting six months, to visit the cities of Berlin, London and Paris. The main purpose of this was the inspection of factories and workshops, in order to learn about the use of steam engines and other types of machinery. ...

However while studying new technologies in the daytime he found opportunities in the evenings to meet the foremost mathematicians in the places he was visiting. For example in Berlin he spent a considerable time with Dirichlet, in London with Cayley and Sylvester, and in Paris he was warmly received by Liouville, who introduced him to other French mathematicians. ..." Excerpted from [51], Ch. 4.

"It was of great interest for me to become acquainted with the celebrated geometer Lejeune-Dirichlet. ... [I] found an occasion each day to talk with this geometer concerning [applications of calculus to number theory] as well as other questions on pure and applied analysis. ... [I attended] with particular pleasure one of his lectures on theoretical mechanics." Excerpted from Chebyshev's report on his trip to Western Europe [14], p. XVII.

mann came up with the revolutionary idea of applying Fourier analysis (of which he was a master) in order to get more precise information about $\pi(x)$ via a function $\Pi(x)$, cf. (46), which is analogous to Chebyshev's $\psi(x)$. For the purpose of Fourier analysis we also modify ψ slightly (by a standard procedure) to $\tilde{\psi}$, as given in §2.

A fundamental link between the functions $\zeta(s)$ and $\psi(x)$ can be obtained by *inverting* (12) to get an analytic expression for $\tilde{\psi}(x)$. In fact, starting from (12), the classical Fourier inversion formula[5] implies, for any fixed $a > 1$,

$$\tilde{\psi}(x) \;=\; \lim_{T\to\infty} \frac{1}{2\pi i} \int_{a-iT}^{a+iT} \left(-\frac{\zeta'(s)}{\zeta(s)}\right) x^s \frac{ds}{s}\,, \qquad x > 0.$$

Now consider the closed rectangular contour $\mathcal{C}$ with vertices $a \pm iT$, $-(2n+1) \pm iT$, with counterclockwise orientaion, where $T \to \infty$ is suitably chosen (cf. C.1 (a)) and $n \in \mathbb{N}$, $n \geq T \log T$. With careful estimations of the modulus of the integrand on the horizontal edges of $\mathcal{C}$, as well as the left hand edge, one shows that the contribution of these three edges approaches 0 with $T \to \infty$ as above, and hence (cf. C.1)

$$\tilde{\psi}(x) \;=\; \lim_{T\to\infty} \frac{1}{2\pi i} \int_{\mathcal{C}} \left(-\frac{\zeta'(s)}{\zeta(s)}\right) x^s \frac{ds}{s}\,, \qquad x > 1.$$

The latter integral is easily evaluated using the residue formula (cf. [58], VI, §1). The poles of the integrand inside $\mathcal{C}$ occur at $s = 1$, at the non-trivial zeros ρ of ζ, at the trivial zeros $-2n$ of ζ, and at $s = 0$. The residues are, respectively, x (since $s = 1$ is a simple pole of ζ, cf. (10)), $-m(\rho)\frac{x^\rho}{\rho}$, $m(\rho) \in \mathbb{N}$ being the multiplicity of ρ, $\frac{x^{-2n}}{2n}$ (since all trivial zeros of ζ are simple, cf. 5.1), and $-\frac{\zeta'(0)}{\zeta(0)} = -\log(2\pi)$, see 4.4. This leads directly to formula (38), which in turn leads to the important "explicit formula", stated by Riemann [76] in slightly different form (47), and proved in both Riemann's form and the following form by von Mangoldt [68]:

---

[5]When $\sigma = a > 1$, we can rewrite (12) as

$$f(t) = \int_0^\infty e^{-ity} g(y)\, dy, \quad \text{where} \quad f(t) := -\frac{\zeta'(a+it)}{\zeta(a+it)} \cdot \frac{1}{a+it}, \quad g(y) := \frac{\tilde{\psi}(e^y)}{e^{ay}},$$

$t \in \mathbb{R}$, $y \in \mathbb{R}$. By the Fourier inversion formula, $g(y) = \lim\limits_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} e^{ity} f(t)\, dt$.

4.6 Explicit formula (Riemann–von Mangoldt [68], 1895) For $x > 1$,

$$\tilde{\psi}(x) \;=\; x \;-\; \sum_{\rho} \frac{x^{\rho}}{\rho} \;+\; \frac{1}{2}\log\left(\frac{x^2}{x^2-1}\right) \;-\; \log(2\pi),$$

the sum being extended over all zeros ρ (with multiplicities) of ζ in the entire critical strip, in order of increasing $|\rho|$ (compare §2). □

Evidently, the explicit formula 4.6 gives a very precise description of the error committed in the approximation $\psi(x) \sim x$, and more importantly, it relates (e.g. in Appendix A) the estimation of this error to the location of the non-trivial zeros. Since the vertical distribution is reasonably well known, see (33), 7.5, the horizontal location of the zeros becomes of paramount importance, see also Remark A.3.

We close this historical discussion by appending Riemann's formula in [76] to obtain the analytic continuation of ζ to all of $\mathbb{C} \setminus \{1\}$:

$$\zeta(s) \;=\; \frac{\Gamma(1-s)}{2\pi i}\; H(s), \quad \text{where} \quad H(s) \;:=\; \int_{\mathcal{C}} \frac{(-z)^s}{e^z-1}\,\frac{dz}{z}\,, \tag{16}$$

$s \in \mathbb{C} \setminus \mathbb{N}$, and $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \mathcal{C}_3$ is the contour from $+\infty$ to $+\infty$ shown in Figure 1 with $\mathcal{C}_2$ a circle of radius ε, $0 < \varepsilon < 2\pi$.

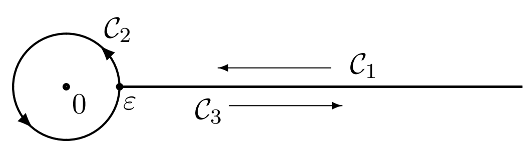

Figure 1: The Contour $\mathcal{C}$ for the Hankel Integral

Nowadays $H(s)$ is known as the *Hankel integral* ([58], XV, §4), and the exponential function $(-z)^s = e^{s\log(-z)}$ is defined by taking $\log(-z)$ to be the principal value of log on $\mathbb{C}$ with the negative real axis deleted. It follows that $\mathrm{Im}(\log(-z))$ varies from $-\pi$ to $+\pi$ on $\mathcal{C}_2$, and hence one defines $\log(-z) = \log|z| - \pi i$ on $\mathcal{C}_1$, $\log(-z) = \log|z| + \pi i$ on $\mathcal{C}_3$. The radius ε of $\mathcal{C}_2$ is taken less than 2π so that $z = 0$ is the only zero of the denominator $e^z - 1$ inside or on $\mathcal{C}$. The exponential decay of the integrand and Lemma 1.1 of [58], XV (which allows one to differentiate under the integral sign) show that H is an entire function.

Formula (16) for $\sigma > 1$ follows from (2) and the Mellin transform expression (4) of ζ, by showing that

$$H(s) \;=\; \int_{\mathcal{C}_1} + \int_{\mathcal{C}_2} + \int_{\mathcal{C}_3} \;=\; 2i\sin(\pi s)\int_0^\infty \frac{x^s}{e^x-1}\,\frac{dx}{x}\,, \qquad \sigma > 1,$$

where $\lim\limits_{\varepsilon\to 0^+}(\int_{\mathcal{C}_1} + \int_{\mathcal{C}_3})$ equals the right hand expression, while $\int_{\mathcal{C}_2}$ equals $2\pi i$ times the average value of $\frac{(-z)^s}{e^z-1} = (-z)^{s-1}\frac{-z}{e^z-1}$ on $\mathcal{C}_2$, and hence has limit 0 as $\varepsilon \to 0^+$. Thus (16) gives the analytic continuation of ζ to $\mathbb{C}\setminus\{1\}$.

5 Symmetry and the associated ξ function

In 1749, Euler returned to the subject in his paper [35], see also his 1748 book [33]. This time he considered the closely related function[6]

$$\phi(s) \;:=\; \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n^s} \;=(1-2^{1-s})\,\zeta(s)\,. \tag{17}$$

Using the associated power series $\phi(s,x) = \sum_n \frac{(-1)^{n+1}}{n^s}x^n$, which is absolutely convergent for $|x| < 1$, $s \in \mathbb{R}$, and taking limits as $x \to 1^-$ (cf. Hardy [46], §2.3, for details), he proved that, for any integer $m \geq 2$, we have

$$\lim_{x\to 1^-}\frac{\phi(1-m,x)}{\phi(m,x)} \;=\; \begin{cases} (-1)^{m/2+1}\frac{2^m-1}{\pi^m(2^{m-1}-1)}(m-1)! & \text{if } m \text{ is even,}\\ 0 & \text{if } m \text{ is odd.}\end{cases}$$

He then formally replaced $\lim_{x\to 1^-}\phi(s,x)$ by $\phi(s)$ and, with the help of the cosine function, rewrote this in the simple form

$$\frac{\phi(1-m)}{\phi(m)} \;=\; -\frac{2^m-1}{\pi^m(2^{m-1}-1)}(m-1)!\cos\frac{\pi m}{2} \qquad \text{for all } m \in \mathbb{N}\setminus\{1\}.$$

At this point Euler states his belief that the same should remain true for all real numbers, i.e.

$$\frac{\phi(1-s)}{\phi(s)} \;=\; -\frac{2^s-1}{\pi^s(2^{s-1}-1)}\Gamma(s)\cos\frac{\pi s}{2} \qquad \text{for } s \in \mathbb{R}\setminus\{1,0,-1,-2,\dots\}\,.$$

[6] The series in (17) uniformly converges in the half plane $\sigma \geq \varepsilon$ for any $\varepsilon > 0$ [57], §42, and thereby determines an analytic continuation of $\zeta(s)$ to $\sigma > 0$.

Due to (2), this is equivalent to saying that, for all $s \in \mathbb{R} \setminus \mathbb{Z}_{\geq 0}$, we have

$$\zeta(s) \;=\; \left[2^s \pi^{s-1} \Gamma(1-s) \sin \frac{\pi s}{2}\right] \zeta(1-s), \tag{18}$$

the so called *functional equation* of ζ. Euler did not know how to prove this intriguing assertion, but he verified it for several non-integer values of s, e.g. $s = \frac{1}{2}$, $\frac{3}{2}$, and $\frac{5}{2}$.

In 1859 Riemann [76] was the first to indicate that (18) is true, indeed for all $s \in \mathbb{C} \setminus \mathbb{Z}_{\geq 0}$. Today, many proofs of this important result exist (see [25], [55], [57], [58], or [85]). In [76] Riemann first expressed $\zeta(s)$ in terms of the Hankel integral $H(s)$ for all $s \in \mathbb{C} \setminus \mathbb{N}$, cf. (16). He then evaluated $H(s)$, for $\sigma < 0$, by reversing the orientation of the contour $\mathcal{C}$ shown in Figure 1, and applying the residue formula (cf. [58], VI, §1) to the domain $\mathcal{D}$ exterior to $\mathcal{C}$ (taking account of the poles of the integrand in this domain at $z = 2\pi k i$, $k \in \mathbb{Z} \setminus \{0\}$, where $e^z - 1 = 0$). This yields the functional equation (18) for $\sigma < 0$, and therefore for all $s \in \mathbb{C} \setminus \mathbb{Z}_{\geq 0}$, since the difference of the two sides of (18) is a meromorphic function on $\mathbb{C}$ having a non-discrete set of zeros, see 3.2. Since $\mathcal{D}$ is not bounded, to make Riemann's argument rigorous one can replace $\mathcal{D}$ by its intersection with a large square $|\mathrm{Re}(z)| < (2n+1)\pi$, $|\mathrm{Im}(z)| < (2n+1)\pi$, and take the limit of the integral as $n \to \infty$. Notice that $|e^z - 1| > 1/2$ on the boundary $\mathcal{Q}$ of the square, hence $\lim\limits_{n\to\infty} \int_{\mathcal{Q}} \frac{(-z)^s}{e^z - 1} \frac{dz}{z} = 0$, $\sigma < 0$ (see also [25], §1.6).

Remark 5.1 Strictly speaking (18) does not hold when $\Gamma(1-s)$ is undefined, which as we saw in 3.4 is true precisely for the poles at $s = 1, 2, 3, \dots$. However, for $s = 3, 5, 7, \dots$, $\zeta(s)$ is some positive real number and $|\sin(\frac{\pi s}{2})| = 1$, so an easy continuity argument shows $\zeta(1-s)$ then must equal 0, i.e. $0 = \zeta(-2) = \zeta(-4) = \dots$. As mentioned in 3.5, these are called the "trivial" zeros of the zeta function. We now see that they are the only possible zeros on the real axis, and are simple zeros. Since any convergent infinite product with non-zero factors cannot equal 0, the Euler product formula (4) already shows $\zeta(s) \neq 0$, $\sigma > 1$. Then (18) and the observations about the zeros of the functions $\sin, \Gamma$ in Examples 3.3, 3.4, show that $\zeta(s) \neq 0$, $\sigma < 0$, apart from the trivial zeros on the negative real axis. The assertion in Example 3.5, that all non-trivial zeros lie in the critical strip $0 \leq \sigma \leq 1$, is thus proved.

Remark 5.2 Alternatively, the result about the trivial zeros of ζ in Remark 5.1 can be thought of as a special case of the following explicit formula, which can be easily derived from the Hankel integral (16), cf. [25], §1.5:

$$\zeta(-n) = (-1)^n \frac{B_{n+1}}{n+1}, \qquad n = 0, 1, 2, \dots .$$

Here B_n is the nth Bernoulli number [7], defined by

$$\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} \frac{B_n z^n}{n!}.$$

For example, $B_0 = 1$, $B_2 = 1/6$, $B_4 = -1/30$, $B_6 = 1/42$, $B_8 = -1/30$, $B_{10} = 5/66$, $B_1 = -1/2$, $B_3 = B_5 = B_7 = \dots = 0$. In view of the functional equation (18) with $s = -2k + 1$, the above formula for $\zeta(-n)$, with $n = 2k - 1$, is equivalent to Euler's famous formula in [34]:

$$\zeta(2k) = (-1)^{k+1} \frac{(2\pi)^{2k} B_{2k}}{2 \cdot (2k)!}, \qquad k \in \mathbb{N}.$$

Having derived the functional equation (18), Riemann proceeded at once to obtain a more symmetric form by defining

$$\xi(s) := \frac{1}{2} s(s-1)\zeta(s)\Gamma(\frac{s}{2})\pi^{-s/2} = (s-1)\zeta(s)\Gamma(\frac{s}{2}+1)\pi^{-s/2}. \tag{19}$$

Proposition 5.3 (Riemann [76], 1859) *The function ξ satisfies*

- *(a)* $\xi(s) = \xi(1-s)$,
- *(b)* ξ *is an entire function, and* $\xi(\overline{s}) = \overline{\xi(s)}$,
- *(c)* $\xi(\frac{1}{2} + it) \in \mathbb{R}$,
- *(d) If* $\xi(s) = 0$, *then* $0 \le \sigma \le 1$,
- *(e)*[7] $\xi(0) = \xi(1) = 1/2$,
- *(f)* $\xi(s) > 0$ *for all* $s \in \mathbb{R}$.

Outline of proof: Using the properties (2), (3) of the gamma function, deriving the functional equation (a) for ξ from that of ζ, i.e. (18), is a straightforward exercise. The second expression in the definition (19) shows at once that ξ is holomorphic for $\sigma \ge 0$, since the simple pole of ζ at 1 is removed by the factor $s-1$, and there are no other poles for $\sigma > 0$. But then (a) implies ξ holomorphic on all of $\mathbb{C}$. The second part of (b) follows trivially from (19) and 3.6. Combining (a) with (b) and 3.6 gives (c), and similarly for (d) by first noting $\xi(s) \neq 0$, $\sigma > 1$. The known values $\Gamma(1) = 1$, $\zeta(0) = -1/2$ (see 3.4, 4.4) imply (e) for $\xi(0)$, and the functional equation (a) then gives the result for $\xi(1)$.

Finally, to prove (f), first note from (1) that $\Gamma(s) > 0$ for all $s \in \mathbb{R}$, $s > 0$. Combining this with Remark 4.3 and the definition (19) of ξ proves (f) for $s > 0$, $s \neq 0, 1$. Combining this with (e) then proves (f) for all $s \geq 0$, whence the functional equation (a) shows that (f) holds for all $s \in \mathbb{R}$. □

Corollary 5.4 *The zeros of the function ξ are identical to the non-trivial zeros of the function ζ.* □

It is now possible to understand what Riemann meant when he stated (his version of) the RH, which we quote in the original German, followed by an English translation ([25], Appendix): "Man findet nun in der That etwa so viele reele Nullstellen innerhalb dieser Grenzen, und es ist sehr wahrscheinlich, dass alle Wurzeln reel sind" (One finds in fact about this many real roots within these bounds, and it is very likely that all of the roots are real). At this stage (the third page) of his paper [76], Riemann is referring to the function $\xi(1/2 + iu)$ of the complex variable u. The fact that all zeros of this function are real (i.e. $u \in \mathbb{R}$) is equivalent to the fact that all zeros of $\xi(s)$ have real part $\mathrm{Re}(s) = \sigma = 1/2$, which by Corollary 5.4 is equivalent to RH.

Remark 5.5 Riemann used the letter t for the complex variable that we have denoted by $1/2 + iu$ above (so as to avoid any confusion with the previous use of t throughout this paper). In fact Riemann's choice of the letter t was somewhat unfortunate and has led to some confusion in the literature, as well as a minor error in Riemann's paper, see the footnotes to 5.3 (e) and (47).

It is also now possible to anticipate Riemann's strategy in [76] for locating zeros of ζ (equivalently of ξ) in the critical strip, which we will carry out in detail in the next section. Estimating the real number $\xi(1/2 + it)$ for various real values of t in an interval $0 \leq t \leq T$, at least closely enough to determine its sign, will guarantee the existence of at least N zeros (along this portion of the critical line) when the sign changes N times, by the intermediate value theorem. Further, the argument principle (a standard result in complex analysis, stated immediately below as Theorem 5.6), and some further estimation of a suitable contour integral, will allow us to count the number

$$N(T) := \left|\{s \in \mathbb{C} \mid 0 \le \mathrm{Re}(s) \le 1,\, 0 \le \mathrm{Im}(s) \le T,\, \zeta(s) = 0\}\right| \tag{20}$$

where each zero is counted with its multiplicity. When $N \ge N(T)$, it follows that there are exactly N zeros in this portion of the critical strip, all lying on the critical line and simple.

Theorem 5.6 (Cauchy's principle of the argument) *Let f be meromorphic on a simple closed curve $\mathcal{C}$ and in its interior. Further assume f has no zero or pole on $\mathcal{C}$. Then*

$$\frac{1}{2\pi}\Delta_{\mathcal{C}} \arg(f(z)) = Z - P,$$

where Z equals the number of zeros (with multiplicities counted), and P the number of poles (with orders counted), of f in the interior of $\mathcal{C}$, and $\Delta_{\mathcal{C}} \arg(f(z))$ equals the net change in the argument $\arg(f(z))$ as z makes one counterclockwise circuit of $\mathcal{C}$. □

In our application of 5.6 we will have $f = \xi$, thus $P = 0$. It is important to also note that

$$\Delta_{\mathcal{C}} \arg(f(z)) = \mathrm{Im}\left(\int_{\mathcal{C}} \frac{f'(z)}{f(z)}\, dz\right) = \frac{1}{i}\int_{\mathcal{C}} \frac{f'(z)}{f(z)}\, dz. \tag{21}$$

Furthermore, the first equality in (21) holds more generally for any path $\mathcal{C}$, not necessarily closed.

6 Location of the first three zeros of ζ

Following the strategy outlined before 5.6, let us choose $T = 28$. We shall show that $N \ge 3$ and $N(28) = 3$.

6.1 Demonstration that $N \ge 3$

We already know (Proposition 5.3 (f)) that $\xi(1/2 + it)$ is positive for $t = 0$, and now outline a method that will show $\xi(1/2 + 18i) < 0$, $\xi(1/2 + 23i) > 0$, $\xi(1/2 + 27i) < 0$. Thus there must be at least three zeros on the portion of the critical line $s = 1/2 + it$, $0 < t < 28$. Our technique to approximate the ξ values, at least accurately enough to determine the signs, is based on the Euler-Maclaurin summation method and simple computations that can be done by hand. It is clear that with modern computers similar computations can easily be carried out for much larger values of T.

The Euler-Maclaurin summation formula arises from the approximation of a discrete sum by a definite integral, and can be found in many references (cf. [25], §6.2, or [77], §3). The theory is more or less elementary and involves Bernoulli numbers as well as their generalization to Bernoulli polynomials. A simple example, familiar from elementary calculus, is the approximation of the harmonic sum $\sum_{j=1}^{n} 1/j$ by the definite integral $\int_1^n (1/x)dx = \log n$, see (8), or the last equality in (10) for another example. We will content ourselves in this section with a couple of further examples which illustrate the method and apply to our proposed computations. A nice feature of the method is that it enables one to estimate partial sums of potentially divergent series, with a strict control of the error term.

Example 6.2 The sharp Stirling series for $\log \Gamma(s)$. The formula, derived by Stirling [84] in 1730 (for $s \in \mathbb{R}$, $s > 0$), states that, if $s = re^{i\theta}$, $r > 0$, $-\pi < \theta < \pi$, then

$$\log \Gamma(s) = (s - \frac{1}{2}) \log s - s + \frac{1}{2} \log(2\pi) + \sum_{k=1}^{n} \frac{B_{2k}}{(2k-1)2k\, s^{2k-1}} + R_{2n}(s),$$

where one has the strong upper bound (due to Stieltjes [83], see also [25], §6.3) for the error term

$$|R_{2n}(s)| \leq \left(\frac{1}{\cos(\theta/2)} \right)^{2n+2} \left| \frac{B_{2n+2}}{(2n+1)(2n+2)s^{2n+1}} \right| .$$

It may not be obvious that this infinite series is actually divergent. The divergence is due to the fact that the Bernoulli numbers actually grow very rapidly, for example $B_{26} = \frac{8\,553\,103}{6} \approx 1\,425\,517.17$, or more generally[8]

$$|B_{2n}| \sim 4\sqrt{\pi n} \left(\frac{n}{\pi e} \right)^{2n} .$$

[8]This asymptotic formula for the Bernoulli numbers is very accurate. For example, for $n = 13$, it gives $B_{26} \approx 1\,420\,956$, compare 6.2. It does not seem to appear in the literature, but can be deduced from [59] or [20].

Nevertheless, one can use the first few terms of the series to estimate $\log \Gamma(s)$ very accurately, i.e. with very small remainder. As a consequence, we also obtain the "classical" Stirling formula

$$(22) \qquad \Gamma(s) \sim \sqrt{\frac{2\pi}{s}}\,\frac{s^s}{e^s}\,, \qquad \sigma \geq 0, \quad |s| \to \infty.$$

As a specific example (that will be used later), take $s_0 = 5/4 + 9i$ and $n = 1$. Then

$$\log \Gamma(s_0) = (s_0 - 1/2)\log(s_0) - s_0 + \frac{1}{2}\log(2\pi) + \frac{1/6}{1 \cdot 2 \cdot s_0} + R_2(s_0),$$

where the inequality $|s_0| > 9$ and the Stieltjes remainder formula give $|R_2(s_0)| < 4 \cdot (1/30)/(3 \cdot 4 \cdot 9^3) \approx 1.52416 \times 10^{-5}$. Evaluating the above then gives $\log \Gamma(5/4+9i) \approx -11.5698+11.9265i$, where the magnitude of the remainder shows that the accuracy is to about six significant digits. Exponentiating this gives $\Gamma(5/4+9i) \approx 10^{-6}(7.57806-5.64057i)$, again with about six digit accuracy.

We remark that several calculations with complex numbers are involved in the above evaluation, and also the use of the well known formula $\log(r \cdot e^{i\theta}) = \log r + i\,\theta$. This must be applied carefully since θ is only unique $\mathrm{mod}(2\pi)$; we take the branch of the logarithm function (for $\log(\frac{5}{4} + 9i)$) where $0 \leq \theta < \pi/2$. Mathematical software can differ on the choice of branch, so an answer differing by $2m\pi i$, for some integer m, can easily occur. For example MAPLE gives $\log \Gamma(5/4 + 9i) \approx -11.56982768 - 0.6398651938i$, to ten digit accuracy. Of course, this difference of $4\pi i$ becomes irrelevant once the exponential is taken.

Before turning to our next example, we state an Euler-Maclaurin summation formula for $\Gamma'(s)/\Gamma(s)$, essentially the derivative of the first formula in 6.2, with $n = 0$, that will be of use in other parts of this paper:

$$(23) \qquad \frac{\Gamma'(s)}{\Gamma(s)} = \log s - \frac{1}{2s} + R_0'(s),$$

where $|R_0'(s)| \leq \sec^3(\theta/2) \cdot \left|\frac{B_2}{2s^2}\right|$, and $s = re^{i\theta}$, $r > 0$, $-\pi < \theta < \pi$, cf. [25], §6.3. The corresponding estimations for all $n \geq 0$ are given in [77], §8.

Example 6.3 Estimating $\zeta(s)$. In somewhat similar fashion to the previous example, Euler-Maclaurin summation can be applied to the tail of the Dirichlet series to obtain an accurate estimation of $\zeta(s)$, even for s in the critical strip (where the Dirichlet series diverges). It gives us

$$\zeta(s) = \sum_{j=1}^{N-1} \frac{1}{j^s} + \frac{N^{1-s}}{s-1} + \frac{1}{2N^s} + \frac{B_2 \cdot s}{2N^{s+1}} + \dots$$
$$+ \frac{B_{2n} \cdot s(s+1)\dots(s+2n-2)}{(2n)!N^{s+2n-1}} + R_{2n,N}(s),$$

where the error (due to Backlund [6], see also [25], §6.4) is controlled by

$$|R_{2n,N}(s)| \le \left| \frac{s+2n+1}{\sigma+2n+1} \cdot \frac{B_{2n+2} \cdot s(s+1)\dots(s+2n)}{(2n+2)!N^{s+2n+1}} \right|, \quad \sigma > -2n.$$

Example 6.4 Computation of $\zeta(1/2+18i)$. Specializing the previous example, with $N=6$, $n=4$, we have

$$\begin{aligned}\zeta(s) &= 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \frac{1}{(s-1)6^{s-1}} + \frac{1}{2\cdot 6^s} \\ &+ \frac{\frac{1}{6}\cdot s}{2!\cdot 6^{s+1}} - \frac{\frac{1}{30}\cdot s(s+1)(s+2)}{4!\cdot 6^{s+3}} + \frac{\frac{1}{42}\cdot s(s+1)\dots(s+4)}{6!\cdot 6^{s+5}} \\ &- \frac{\frac{1}{30}\cdot s(s+1)\dots(s+6)}{8!\cdot 6^{s+7}} + R_{8,6}(s),\end{aligned}$$

where

$$|R_{8,6}(s)| \le \left| \frac{s+9}{\frac{1}{2}+9} \cdot \frac{\frac{5}{66}\cdot s(s+1)\dots(s+8)}{10!\cdot 6^{s+9}} \right|, \quad \sigma > -8.$$

Evaluating this at $s = s_1 := 1/2+18i$ with modern computational tools is easily done, but it is worthwile at least thinking about how much work it would have been for Riemann, Backlund, or Gram to do this by hand. In particular, evaluating the exponentials such as $1/6^{s_1+1} = 6^{-3/2-18i}$ involves using the well known identity $m^{x+iy} = m^x \cdot (\cos(y\log m) + i\sin(y\log m))$. The estimation of the remainder $R_{8,6}(s_1)$ is somewhat simpler, e.g. one can use $|s_1(s_1+1)\dots(s_1+8)| < |s_1+8|^9 < 20^9$. The outcome of the calculations is $\zeta(1/2+18i) \approx 2.32922 - 0.18865i$, with error less than 10^{-3}. Thus the value is accurate to about three significant digits.

We now return to the original goal of calculating $\xi(1/2+18i) = (-1/2+18i)\zeta(1/2+18i)\Gamma(5/4+9i)\pi^{-1/4-8i}$. The difficult parts are already done in Examples 5.1 and 5.3, and the calculation $\pi^{-1/4-8i} \approx -0.4798582 + 0.5778631i$ is routine (with seven significant digits accuracy). One then finds $\xi(1/2+18i) \approx -10^{-4} \times 2.986$ with about three significant digits accuracy. This proves $\xi(1/2+18i) < 0$. With calculations quite similar to those above, and again about three digits accuracy, one finds $\xi(1/2+23i) \approx 10^{-6} \times 5.622 > 0$, $\xi(1/2+27i) \approx -10^{-7} \times 5.656 < 0$. The first goal of this section, showing that $N \geq 3$, is thus accomplished.

6.5 Demonstration that $N(28) = 3$

To commence the second objective of this section let us apply the Principle of the Argument 5.6 to $\xi(s)$ using the simple closed rectangular curve $\mathcal{D} = \mathcal{D}(T)$ with vertices -1, 2, $2+Ti$, $-1+Ti$, traversed in that order. Let us also write $\mathcal{C} = \mathcal{C}(T)$ for the contour consisting of the portion of $\mathcal{D}$ from 2, to $2+Ti$, to $1/2+Ti$. Finally, define

$$\vartheta(t) := \operatorname{Im}\left(\log\Gamma\left(\frac{1}{4}+\frac{it}{2}\right)\right) - \frac{t}{2}\log\pi. \tag{24}$$

This function has the following estimation, due to Stirling's formula (6.2) with $n = 0$:

$$\vartheta(T) = \frac{T}{2}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2} - \frac{\pi}{8} + O\left(\frac{1}{T}\right), \qquad T \to \infty. \tag{25}$$

This approximation can be further refined up to $O(1/T^{2n+1})$ by using (6.2), for any n (see also [25], §6.5, or [53], III, §4).

Riemann had the asymptotic estimate $N(T) \sim (T/2\pi)\log(T/2\pi) - T/2\pi$ (without proof, however see also Theorem 7.5 and B.2), but the following exact formula of Backlund [5] is a substantial improvement:

Proposition 6.6 (Backlund [5], 1914) *(a) For any $T > 0$ such that $\zeta(s) \neq 0$ on $\mathcal{C}$,*

$$N(T) = \frac{1}{\pi}\vartheta(T) + 1 + \frac{1}{\pi}\operatorname{Im}\left(\int_{\mathcal{C}} \frac{\zeta'(s)}{\zeta(s)}\, ds\right). \tag{26}$$

(b) If also $\operatorname{Re}(\zeta(s)) \neq 0$ on $\mathcal{C}$ then

$$N(T) = \lceil \frac{1}{\pi}\vartheta(T) + 1 \rfloor.$$

Remark 6.7 Actually, in both (a) and (b), the non-vanishing hypothesis need only be checked on the horizontal portion of $\mathcal{C}$, i.e. where $t = T$ and $\frac{1}{2} \le \sigma \le 2$. For (a) this is obvious since the vertical portion lies outside the critical strip. For (b), it is easily verified, using the Dirichlet series (4) and Euler's formula for $\zeta(2)$, cf. §4, that $\mathrm{Re}(\zeta(2+it)) > 2 - \zeta(2) > 0$ for any $t \in \mathbb{R}$. The latter implies that the absolute variation of $\arg(\zeta(2+it))$ is $< \pi$ on any segment $[t_1, t_2]$.

We will next sketch a proof of Proposition 6.6, but first let us note that, for $T = 28$, 6.6 (b) gives $N(T) = \lceil 3.078\ldots \rfloor = 3$, which will then complete the objective of this section, namely showing that ζ has three simple zeros on the critical line up to $1/2 + 28i$. The fact that $\mathrm{Re}(\zeta(s)) \ne 0$ on the horizontal portion $s = \sigma + 28i$, $1/2 \le \sigma \le 2$, of $\mathcal{C}(28)$ is somewhat delicate and can be proved similarly to 6.4, cf. [25], §6.6. We omit the details here and simply remark that it follows, in particular, that $\xi(s)$ is nowhere zero on the closed curve $\mathcal{D}$.

Proof of Proposition 6.6: The Principle of the Argument 5.6, together with (21) and the fact that ξ has no poles, give

$$N(T) = \frac{1}{2\pi i}\int_{\mathcal{D}} \frac{\xi'(s)}{\xi(s)}\,ds = \frac{1}{2\pi}\,\mathrm{Im}\left(\int_{\mathcal{D}} \frac{\xi'(s)}{\xi(s)}\,ds\right).$$

Now since, from Proposition 5.3 (f), ξ is positive real on the portion of $\mathcal{D}$ on the real axis, the argument of $\xi(s)$ does not change here, so by (21) this contributes nothing to the above integral. By the symmetry of both ξ and $\mathcal{D}$ in the critical line $\sigma = 1/2$, it follows that

$$N(T) = \frac{1}{\pi}\,\mathrm{Im}\left(\int_{\mathcal{C}} \frac{\xi'(s)}{\xi(s)}\,ds\right).$$

Considering the definition (19) of $\xi(s)$ and then taking its logarithmic derivative, we are able to write

$$\frac{\xi'(s)}{\xi(s)} = \frac{1}{s} + \frac{1}{s-1} - \frac{1}{2}\log\pi + \frac{\zeta'(s)}{\zeta(s)} + \frac{1}{2}\frac{\Gamma'(\frac{s}{2})}{\Gamma(\frac{s}{2})}\,. \tag{27}$$

Using the above definition (24) of $\vartheta(t)$, one can readily obtain (26). This completes (a).

As for (b), the assumption that $\mathrm{Re}(\zeta(s)) \ne 0$ on $\mathcal{C}$ clearly implies this quantity is in fact positive on $\mathcal{C}$. Since $\zeta(2) \in \mathbb{R}^+$, its argument starts at 0. Hence the absolute variation in its argument, over $\mathcal{C}$, is strictly less than $\pi/2$. This and (21) show that the last integral in (26) has absolute value strictly less than $1/2$. Since $N(T)$ must be an integer, we obtain (b). $\square$

Remark 6.8 In this section we have shown that $\zeta(1/2+it)$ has a zero for three values $t=\alpha_1,\alpha_2,\alpha_3$ with $0<\alpha_1<18,\ 18<\alpha_2<23,\ 23<\alpha_3<27$. With more calculations of the type we have made, it would be possible to narrow down the precise locations of the zeros. Riemann had estimated at least the first three zeros, although this does not appear in his paper. In 1903 Gram [42] located the first 15 zeros, for the first three one has $\alpha_1\approx 14.134725$, $\alpha_2\approx 21.022040$, $\alpha_3\approx 25.010856$, using methods similar to those we have used in this section. Riemann used the more efficient Riemann-Siegel formula, which was not available until Siegel's publication [80] in 1932 of Riemann's *Nachlass* (see also §7).

7 History of the zeta function since Riemann

The two decades following the publication of Riemann's paper [76], in 1859, were largly uneventful. Weierstrass, who was eleven years older than Riemann, but whose rise to fame —from an obscure schoolteacher to a professor at Berlin— happened in a way very different from Riemann's, began working and lecturing on complex numbers and the general theory of entire functions already during the 1860's. But it wasn't until 1876, when Weierstrass finally published his famous memoir [96], that mathematicians became aware of some of his revolutionary ideas and results. The first half of this section will discuss these ideas and how, together with the zeta function, they led to the estimation of the vertical location of the non-trivial zeros and to the proof in 1896 of the Prime Number Theorem (13), arguably the greatest achievement of 19th century mathematics (a short version of the original proof is given in Appendix A). In the second half we return to the discussion of Riemann's paper, the RH, Riemann's *Nachlass* (the 1932 study [80] by Siegel), and some of the subsequent history of the RH.

We say that an entire function f is an *entire function of finite order* if

$$\log|f(s)| = O(|s|^A), \quad \text{for some} \quad A>0. \tag{28}$$

The *order* of $f(s)$ is the lower bound of all A, for which the inequality (28) holds.

Among Weierstrass' many contributions were the following two important theorems:

Theorem 7.1 (Weierstrass [96], 1876) *Let $\{c_n\}$ be an infinite sequence of complex numbers, such that $0 < |c_1| \le |c_2| \le |c_3| \le \dots$, and assume that its only limit point is ∞. Then there exists an entire function $f(s)$ with zeros (with prescribed multiplicities) at precisely these complex numbers.* □

Remark 7.2 Note that in both this theorem and Theorem 7.3 zeros of arbitrary multiplicities are accounted for by taking e.g. $c_j = c_{j+1} = \dots = c_{j+r}$.

Theorem 7.3 (Weierstrass [96], 1876) *Every entire function $g(s)$ of order ≤ 1, which has no zeros in $\mathbb{C}$, can be written as $g(s) = e^{a+bs}$, where a and b are constants, while every entire function $f(s)$ of order ≤ 1, which has $N \le \infty$ zeros at $c_1, c_2, c_3, \dots \ne 0$, can be written in the form*

$$f(s) \ = \ e^{a+bs} \prod_{n=1}^{N} \left[\left(1 - \frac{s}{c_n} \right) e^{s/c_n} \right] \tag{29}$$

where a and b are constants, and the product converges absolutely (if $N = \infty$) for all $s \in \mathbb{C}$. □

Remark 7.4 Let γ be Euler's constant. Weierstrass proved the product formula

$$\frac{1}{\Gamma(s)} \ = \ s e^{\gamma s} \prod_{n=1}^{\infty} \left[\left(1 + \frac{s}{n} \right) e^{-s/n} \right]. \tag{30}$$

This, along with Riemann's paper, set the stage for the great work of Hadamard and de la Valée-Poussin in the 1890's. Recall the definition (19) of $\xi(s)$ and note that, applying (6.3) with $n = 0$, $N = 1$, and Stirling's formula (6.2) with $n = 0$, one can find constants C_1, C_2, C_3 such that, for all $s \in \mathbb{C} \setminus \{1\}$ with $\sigma \ge \frac{1}{2}$,

$$|s(s-1)\zeta(s)| < C_1 |s|^4, \quad \left| \Gamma\left(\frac{s}{2}\right) \right| < e^{C_3 |s| \log |s|}, \quad \left| \pi^{-s/2} \right| < e^{C_2 |s|}.$$

From this, using the properties 5.3 (a,b) of ξ, we have

$$|\xi(s)| \ < \ e^{C|s| \log |s|}, \quad s \in \mathbb{C}, \tag{31}$$

for a constant $C > 0$. Stirling's formula also tells us that the upper bound $|\xi(s)| < e^{C|s|}$ fails as $s = \sigma \to \infty$. Therefore $\xi(s)$ is of order 1, has infinitely many zeros, and can be written in Weierstrass' form as follows:

$$\xi(s) \;=\; e^{A+Bs} \prod_{n=1}^{\infty} \left[\left(1 - \frac{s}{\rho_n}\right) e^{s/\rho_n} \right], \tag{32}$$

where A and B are constants, and $\rho_n = \beta_n + i\gamma_n$ are all the zeros of ξ, arranged so that $|\gamma_1| \le |\gamma_2| \le |\gamma_3| \le \ldots$, and the ρ_j may repeat, as in Remark 7.2.

Entire functions of *arbitrary order* have product representations analogous to (29), as Weierstrass proved in [96]. His general theorem was made more explicit and applicable by Hadamard [43], in 1893. He used it, together with (31), to prove in [43] that $\zeta(s)$ and $\xi(s)$ have infinitely many zeros in the critical strip, and that there exist constants $a, A > 0$ such that

$$\gamma_n \;\ge\; a\frac{n}{\log n}, \qquad \text{equivalently} \qquad N(T) \le AT \log T, \tag{33}$$

for $n \ge 2$, $T \ge 2$. An important consequence is

$$\sum_{n=1}^{\infty} \frac{1}{|\rho_n|^c} \;<\; \infty \qquad \text{for all } c > 1. \tag{34}$$

Using (34) Hadamard [43] proved the following product formula similar to (32), see also [25]:

$$\xi(s) \;=\; \xi(0) \prod_{n=1}^{\infty} \left(1 - \frac{s}{\rho_n}\right). \tag{35}$$

In 1895, von Mangoldt [68] used Hadamard's results (34), (35), to obtain

$$\frac{\xi'(s)}{\xi(s)} = \sum_{\rho} \frac{1}{s-\rho}, \tag{36}$$

where validity of the termwise differentiation of the product in (35) follows from the uniform convergence of its logarithmic derivative in any disk $|s| \le R$, due to (34). He also estimated the vertical density of the roots ρ_n of ζ, for large $T > 0$:

$$N(T+1) \;-\; N(T) \;\le \sum_{T \le \gamma_n \le T+1} 1 \;<\; 2 \log T, \tag{37}$$

by noticing that (34), (36) imply

$$\operatorname{Im}\left(\int_{2+iT}^{2+i(T+1)} \frac{\xi'(s)}{\xi(s)} ds\right) = \sum_{\rho_n} \operatorname{Im}\left(\int_{2+iT}^{2+i(T+1)} \frac{ds}{s-\rho_n}\right)$$
$$> \sum_{T\le\gamma_n\le T+1} \operatorname{Im}\left(\int_{2+iT}^{2+i(T+1)} \frac{ds}{s-\rho_n}\right) \ge \sum_{T\le\gamma_n\le T+1} \arctan\left(\frac{1}{2}\right),$$

and by showing $\int_{2+iT}^{2+i(T+1)} \frac{\xi'(s)}{\xi(s)} ds = \frac{i}{2}\log T + O(1)$ via (27) and Stirling's formula (6.2) with $n=0$, together with (21) and the boundedness of the total variation of $\arg(\zeta(2+it))$ on $[T, T+1]$ (see 6.7 or [25], §3.4). With the help of (34), (36), and (37), von Mangoldt [68] proved the explicit formula 4.6, cf. [25], §3.2-3.5. See also Remark B.2.

In 1896, based on Hadamard's results (33)-(35), Hadamard and de la Vallée-Poussin proved independently $\lim\limits_{x\to\infty} \frac{\theta(x)}{x} = 1$ and the PNT, see Appendix A. An important step in both proofs was to show that no zero of $\zeta(s)$ has real part 1. In 1899 de la Vallée-Poussin made a further improvement (see A.2) which finally justified Chebyshev's prediction of the correct constant in the Legendre prime number formula (cf. §4).

Six years later von Mangoldt proved Riemann's estimate for the vertical distribution of the zeros of ζ, see 6.5, strengthening (33), (37):

Theorem 7.5 (von Mangoldt [69], 1905) *For $T \ge 2$,*

$$N(T) = \frac{T}{2\pi}\log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + O(\log T).$$

A proof is given in Appendix B.

Returning to our historical sketch, let us first make some concluding comments about Riemann's 1859 paper. Needless to say, this paper is written in an extremely terse and difficult style, with huge intuitive leaps and many proofs omitted. This led to (in retrospect quite unfair) criticism by Landau and Hardy in the early 1900's, who commented that Riemann had only made conjectures and had proved almost nothing. The situation was greatly clarified in 1932 when Siegel [80] published his paper, representing about two years of scholarly work studying Riemann's left over mathematical notes at the University of Göttingen, the so-called Riemann's *Nachlass*. From this study it became clear that Riemann had done an immense amount of work related to [76] that never appeared in his paper. One conclusion is that many formulae that lacked sufficient proof in [76] were in fact proved in these notes. A second is that the notes contained further discoveries of Riemann that were never even written up in [76]. One such is what is now called the Riemann-Siegel formula, which Riemann had written down and Siegel (with great difficulty) was able to prove, cf. [25] or [53]. This formula (which we omit) arises from a Hankel integral type expression for $\xi(s)$, and gives a refined method to calculate $\xi(1/2 + it)$, in comparision to the crude methods of §6.

In his 1859 paper Riemann only mentions RH briefly. To quote him once more, "Hiervon wäre allerdings ein strenger Beweis zu wünschen; ich habe indes die Aufsuchung desselben nach einigen flüchtigen vergeblichen Versuchen vorläufig bei Seite gelassen, da er für den nächsten Zweck meiner Untersuchung entbehrlich schien" (One would of course like to have a rigorous proof of this, but I have put aside the search for such a proof after some fleeting vain attempts because it is not necessary for the immediate objective of my investigation), cf. Appendix of [76]. However, towards the end of the paper there are some speculations that a more detailed mathematical analysis (see C.2, or [25], §1.17, 5.5) shows are indirectly related to RH and the improvement of the remainder term in PNT. Furthermore, it is not clear from Riemann's paper that he had any solid evidence for RH, but it is now known (Riemann's *Nachlass*) that he had calculated at least the first three non-trivial zeros and found them to lie on the critical line, much as was done in §5 above. As Ivič says [50], Ch. I, Notes, "it is apparent that he knew much more about $\zeta(s)$ than he cared to publish." It is also important to note that Riemann only lived until 1866, and that his health was very bad during his final years.

Starting from about 1890, the evidence for RH has rapidly increased. For example, we will see in Appendix A that the celebrated PNT proved in 1896 is equivalent to reducing the critical strip from $0 \leq \sigma \leq 1$ to $0 < \sigma < 1$, i.e. it can be thought of as a very small first step towards RH.

Hilbert included RH in his list of 23 problems, at his address to the International Congress in 1900. It is interesting that at the time of his address, Hilbert did not consider RH to be one of the most important problems of his list. However, some years later when asked, if he could sleep 500 years what his first question would be upon awakening, Hilbert replied "has the RH been solved?" Generalizations of RH have taken on equal significance. Starting with the Dirichlet L-functions the concept has been further generalized to Artin L-functions and to global L-functions, which have many similarities to ζ such as an Euler product formula and a functional equation, and are of basic importance in diverse areas of modern mathematics.

Since 1900 the progress towards solving the RH has been enormous, nevertheless it is still unsolved and appears on the Clay Institute list (in 2000) of seven questions for the new millenium. Some highlights of these developments are now outlined, with no attempt at completeness. We have already seen in this section that there are infinitely many zeros of ζ in the critical strip. Hardy [45] improved this in 1914, showing that in fact there are infinitely many zeros on the critical line. His collaboration with Littlewood and Ramanujan produced other important advances [47]. Bohr and Landau [8] proved in 1914 that the proportion of the zeros lying within ε from the critical line equals 1, for any $\varepsilon > 0$. Later in the 20th century Selberg [78], Bombieri [9], and Deligne [22] made very significant contributions. Selberg [78], for example, showed in 1942 that some positive proportion of the zeros lie on the critical line, and this was later improved by Levinson [62] to at least 1/3, and still later by Conrey [16] to at least 2/5. Deligne [22] in 1974 proved the related Weil Conjecture (an analogue of RH for zeta functions of general algebraic varieties over finite fields).

Similarly, starting from about 1890, the realization of the significance of RH has rapidly increased. One equivalent formulation of RH in number theory, the estimation of the error in the approximation of $\pi(x)$ by $\mathrm{Li}(x)$, has already been mentioned in 4.5, see also A.2, A.3. There are many further significant number theoretical implications of RH. For example, Bertrand's *postulate* that there exists a prime in $[n+1, 2n-2]$, $n > 3$ (first proved by Chebyshev [13], cf. (14)) was successively improved over ten times (cf. [50], Ch. 12, Notes), e.g. by Montgomery [72] in 1969 to the existence of a prime in $[n, n+n^{3/5+\varepsilon}]$, and by Lou and Yau [66] in 1992 to the existence of a prime in $[n, n+n^{6/11+\varepsilon}]$, for all $\varepsilon > 0$ and $n \geq n_0(\varepsilon)$. This in turn can be further strengthened using RH to the existence of a prime in $[n, n+cn^{1/2}\log n]$ (Cramér [18], 1920, see also [19], or [50], §12.6), and using Cramér's conjecture (i.e. $\overline{\lim\limits_{n\to\infty}} \frac{p_{n+1}-p_n}{\log^2 p_n} = 1$) even to $[n, n+c\log^2 n]$, cf. [19]. The latter cannot be strengthened much further, since, due to Westzynthius [98] in 1931, $\overline{\lim\limits_{n\to\infty}} \frac{p_{n+1}-p_n}{\log p_n} = \infty$. A further example involves, for a given prime p, estimating the least quadratic non-residue (mod p), written $n(p)$. Here Vinogradov's classical 1918 result [90], [91] that $n(p) < p^{\frac{1}{2\sqrt{e}}}\log^2 p$ for all sufficiently large p (see also [75]), improved in 1957 by Burgess [11] to $n(p) = O(p^\alpha)$ for any fixed $\alpha > \frac{1}{4\sqrt{e}}$, can be strengthened using the extended RH (i.e. the RH for the Dirichlet L-functions, cf. (9)). In this way, Ankeny [2] showed in 1952 that $n(p) = O(\log^2 p)$, and Bach [4] improved this in 1990 to $n(p) \leq 2\log^2 p$. This cannot be strengthened much further, since Graham and Ringrose [41] showed in 1990 that $\overline{\lim\limits_{p\to\infty}} \frac{n(p)}{\log p \log\log\log p} > 0$ unconditionally, while Montgomery [73] showed in 1971 using the extended RH that $\overline{\lim\limits_{p\to\infty}} \frac{n(p)}{\log p \log\log p} > 0$.

Intriguing (and important) equivalent conjectures abound, suggesting alternative approaches to RH. For an excellent survey of these as well as of recent progress on the problem cf. Conrey [17] and Bombieri [10] (his descriptive paper for the Millenium Problems). In a recent paper [77] by two of the authors, as well as in some earlier work of Spira [82], a slightly different "horizontal" approach to the question is taken. The functional equation shows that for any non-trivial zero $Q := 1/2 + \Delta + it$ in the critical strip ($0 \leq \Delta \leq 1/2$), one also has a zero at $P := 1/2 - \Delta + it$ (as well as at $\overline{P}, \overline{Q}$). In [77] very accurate upper and lower bounds for the ratio $|\zeta(P)/\zeta(Q)|$ are obtained. In particular, it is shown that $|\zeta(P)| \geq |\zeta(Q)|$. Clearly the inequality $|\zeta(P)| > |\zeta(Q)|,\ 0 < \Delta \leq 1/2$, would imply RH since both could not then be simultaneously 0.

From the point of view of gathering numerical evidence, the early work of Gram (cf. 6.8) and Backlund [5] was carried further by Hutchinson [48] in 1925 to show that the first 138 zeros (in the upper half plane) lie on the critical line. Once the Riemann-Siegel formula became available, this was soon improved to the first 1041 zeros by Titchmarsh and Comrie [86], [87]. Thanks to modern computational power, it is now known that at least the first 10^{10} zeros lie on the critical line (a number that is steadily increasing).

A Appendix: Prime Number Theorem

In this appendix we give a proof, incorporating ideas from the original proofs, of the celebrated Prime Number Theorem (13), conjectured by Gauss in 1793, and proved in 1896 by both Hadamard [44] and (independently) de la Vallée-Poussin [88] (and also [89]). Alternative proofs using elementary methods appeared some 50 years later, cf. [26], [79] (see also [65], [74], and [50], Ch. 12), where "elementary" means, in particular, without use of complex analysis, but not necessarily simpler.

A.1 Prime Number Theorem: $\pi(x) \sim \mathrm{Li}(x)$.

While it would be beyond the scope of this paper to furnish complete details of this proof, we shall fully describe the key step in the proof (reducing the critical strip from $0 \leq \sigma \leq 1$ to $0 < \sigma < 1$), as well as clearly indicate and discuss the other ingredients of this proof (for full details excellent sources are [25], [49], [52], [53], [57], [58], and others).

First of all we will use the Riemann-von Mangoldt explicit formula 4.6. Secondly, we will prove the PNT in the equivalent form stated in §4, namely $\psi(x) \sim x$. The proof that these are equivalent is straightforward and goes back to Chebyshev's ideas, cf. [25], §4.4. A third fact we shall use is that, for the non-trivial zeros ρ of ζ, $\sum_\rho \frac{1}{\rho^2}$ is absolutely convergent; this is an immediate consequence of Hadamard's formula (34).

Let us start the sketch by rewriting the explicit formula 4.6 in the form

$$\tilde{\psi}(x) = x - \sum_\rho \frac{x^\rho}{\rho} + \sum_n \frac{x^{-2n}}{2n} - \log(2\pi) \qquad \text{for } x > 1. \tag{38}$$

In 1896, de la Vallée-Poussin [88] showed that the term-by-term integration of both sides of (38) is a valid operation, and in fact, for $x > 1$, it leads to the formula

$$\psi_1(x) \ := \ \int_0^x \psi(t)dt \tag{39}$$

$$= \ \frac{x^2}{2} \ - \ \sum_\rho \frac{x^{\rho+1}}{\rho(\rho+1)} \ - \ \sum_n \frac{x^{-2n+1}}{2n(2n-1)} \ - \ x\log(2\pi) \ + \ \text{const.}$$

It is clear that, as $x \to \infty$, the last three terms on the right hand side of (39) are all $o(x^2)$.

Our next step is to show $\zeta(1+it) \neq 0$, i.e. there are no zeros of $\zeta(s)$ on the line $\sigma = 1$ (Hadamard showed this in [43], however we will follow the method of de la Vallée-Poussin in [89]). To see this, let $\sigma > 1$ and integrate (11) termwise (the constant of integration is clearly seen to equal 0 by taking $s = \sigma$ real and letting $\sigma \to \infty$), giving

$$\log\,\zeta(s) = \sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^s \log n}, \quad \sigma > 1.$$

Taking the real parts,

$$\mathrm{Re}(\log\,\zeta(s)) = \sum_{n=2}^{\infty} \frac{\Lambda(n)}{n^{\sigma}\cdot \log n}\cdot \cos(t\log n).$$

Using the trigonometric identity $3+4\cos t+\cos 2t = 2(1+\cos t)^2 \geq 0$, it follows that

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

and exponentiating this gives

$$|\zeta(\sigma)|^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \geq 1, \quad \text{for} \quad \sigma>1. \tag{40}$$

As we saw in (10), ζ has the single pole at $s=1$, and it is simple with residue 1. This is equivalent to $\lim_{s\to 1}(s-1)\zeta(s) = 1$. Now suppose that ζ has a zero of order $m \geq 1$ at $s_0 = 1+it_0$, then similarly this is equivalent to $\lim_{s\to s_0}(s-s_0)^{-m}\zeta(s) = c$ for some $c \in \mathbb{C}\setminus\{0\}$. Taking $s = \sigma + it_0, \sigma > 1$, we can rewrite (40) as

$$\begin{aligned} |\zeta(\sigma)|^3\cdot|\sigma-1|^3\cdot\frac{|\zeta(\sigma+it_0)|^4}{|s-s_0|^{4m}}\cdot|\zeta(\sigma+2it_0)| \ &\geq\ \frac{|\sigma-1|^3}{|s-s_0|^{4m}} \\ &=\ \frac{|\sigma-1|^3}{|\sigma-1|^{4m}} \ =\ \frac{1}{|\sigma-1|^{4m-1}}. \end{aligned}$$

Letting $\sigma\to 1^+$ in this inequality, and taking account of the two limits above, shows that there is a pole of order $\geq 4m-3 \geq 1$ at $s = 1+2it_0$. Since this is impossible, the claim $\zeta(1+it) \neq 0,\ t\in\mathbb{R}\setminus\{0\}$ is established.

Therefore, if ρ is a non-trivial zero of $\zeta(s)$, then $\mathrm{Re}(\rho) < 1$, and we have $|x^{\rho-1}| < 1$, while the infinite sum $\sum_{\rho}\frac{1}{\rho(\rho+1)}$ converges absolutely, cf. (34). This implies that $\sum_{\rho} x^{\rho-1}/\rho(\rho+1)$ converges uniformly in x, whence

$$\lim_{x\to\infty}\sum_{\rho}\frac{x^{\rho-1}}{\rho(\rho+1)} \ =\ \sum_{\rho}\lim_{x\to\infty}\frac{x^{\rho-1}}{\rho(\rho+1)} \ =\ \sum_{\rho} 0 \ =\ 0,$$

and hence the second term of the right hand side of (39) is also bounded by $o(x^2)$. Therefore, we can conclude $\psi_1(x) \sim \frac{x^2}{2}$. In general, if two functions are asymptotic, one cannot conclude their derivatives are asymptotic. However, in the situation at hand, one also knows the derivative $\psi = \psi_1'$ is a monotone non-decreasing function, and it is then straightforward (cf. [25], §4.3) to conclude that $\psi(x)\sim x$. □

Remark A.2 In 1899, de la Vallée-Poussin [89] made the above argument more explicit, and he was the first to obtain a zero-free region of $\zeta(s)$ having positive measure:

$$\zeta(s) \neq 0 \qquad \text{for} \qquad \mathrm{Re}(s) \ \geq \ 1 \ - \ \frac{c}{\log(|t|+2)},$$

where $c > 0$ is a constant. From this, using (34), as well as (39) and an inequality similar to (40), de la Vallée-Poussin obtained the following estimation of the error in the PNT:

$$\psi(x) = x + O\left(xe^{-A\sqrt{\log x}}\right) \quad \text{and} \quad \pi(x) = \mathrm{Li}(x) + O\left(xe^{-B\sqrt{\log x}}\right),$$

where $A, B > 0$ are constants, see also [25], §5.3. Due to the asymptotic expansion (15) of $\mathrm{Li}(x)$, this proves that $\mathrm{Li}(x)$ indeed provides a much better approximation to $\pi(x)$ than do $\frac{x}{\log x}$, $\frac{x}{\log x - 1}$, and the Legendre formula, see §4. Littlewood [64] in 1924 improved de la Vallée-Poussin's estimation for the zero-free region with the help of Weyl's method of evaluating "exponential sums" $\sum_{n=a}^{b} n^{-1-it}$ in the Euler-Maclaurin formula for $\zeta(1+it)$. This was substantially improved by Chudakov [15] in 1936, based on Vinogradov's powerful methods [92] in the 1930s for the estimation of exponential sums, and later by Vinogradov (announced [93] in 1942, published [94] in 1958) and Korobov (1958), cf. [50], Ch. 6, [25], §9.8, or [53], IV. Using the methods of Vinogradov and Korobov, Richert slightly improved their estimates (unpublished) to obtain the zero-free region $\sigma \geq 1 - c\log^{-2/3} t(\log\log t)^{-1/3}$, $t \geq t_0$, cf. [95]. These improvements led to corresponding improvements of de la Vallée-Poussin's estimate of the error term in PNT (cf. [95]). The best known estimates (obtained by Walfisz [95] from Richert's result) are $\psi(x) = x + O\left(xe^{-C\log^{3/5} x(\log\log x)^{-1/5}}\right)$, and $\pi(x) = \mathrm{Li}(x) + O\left(xe^{-C_1\log^{3/5} x(\log\log x)^{-1/5}}\right)$, for constants $C, C_1 > 0$.

Remark A.3 As Landau [56] (see also [57], §93-94) showed in 1909, additional information about the horizontal location of the non-trivial zeros can provide an improvement of de la Vallée-Poussin's estimate of the error in PNT (see A.2), as follows: if for all non-trivial zeros ρ_n of $\zeta(s)$ we have $\mathrm{Re}(\rho_n) \leq \Delta$, for some fixed $\frac{1}{2} \leq \Delta < 1$, then

$$\pi(x) \ = \ \mathrm{Li}(x) \ + \ O(x^{\Delta}\log x).$$

In particular, RH (corresponding to $\Delta = 1/2$) implies that the error in PNT is $O(x^{1/2}\log x)$ [54]. Actually, the converse is also true, cf. [57],

§93, 201 (see also 4.5 and [25], §5.5). In other words, RH could be obtained by somehow proving the PNT with a very sharp error term.

B Appendix: Von Mangoldt's theorem

In this appendix, a proof of von Mangoldt's Theorem 7.5 is given. This theorem describes the vertical distribution of the non-trivial zeros of $\zeta(s)$. Again, the proof uses the results of Weierstrass (cf. §7). It is based on Backlund's ideas (cf. Proposition 6.6), as well as [50], see also [25], [49], [53], [57].

Theorem B.1 (von Mangoldt [69], 1905) *For $T \geq 2$,*

$$N(T) = \frac{T}{2\pi} \log\left(\frac{T}{2\pi}\right) - \frac{T}{2\pi} + O(\log T). \tag{41}$$

Remark B.2 Earlier, in 1895, von Mangoldt [68] proved an analogue of (41) with a slightly weaker error term $O(\log^2 T)$. In fact, this was the first time that the correct main term for $N(T)$ was obtained, and it turned out to be exactly what Riemann claimed in [76]. Riemann also predicted the error correctly in [76], however his description was unclear (he was referring to relative error) and this led to subsequent misinterpretations in the literature (see also [25], §1.9).

Proof: Since the set of zeros of ζ is discrete, see 3.2, we may assume no zero of ζ has imaginary part T. Due to (25) and (26), it suffices to show $\operatorname{Im}\left(\int_{\mathcal{C}} \frac{\zeta'(s)}{\zeta(s)}\, ds\right) = O(\log T)$. The latter is equivalent to

$$I := \operatorname{Im}\left(\int_{1/2}^{2} \frac{\zeta'(\sigma+iT)}{\zeta(\sigma+iT)}\, d\sigma\right) = O(\log T),$$

due to (21) and the boundedness of the total variation of $\arg(\zeta(2+it))$ on $[0,T]$, see Remark 6.7.

From (27) and (36), we have

$$\frac{\zeta'(s)}{\zeta(s)} = -\frac{1}{s} - \frac{1}{s-1} + \frac{1}{2}\log\pi + \sum_{\rho} \frac{1}{s-\rho} - \frac{1}{2}\frac{\Gamma'(\frac{s}{2})}{\Gamma(\frac{s}{2})} \tag{42}$$

$$= \sum_{\rho} \frac{1}{s-\rho} + O(\log|t|), \qquad \sigma \geq -1,\ |t| \geq 2.$$

where the latter estimate follows from (23). Let us denote

$$S_1(s) := \sum_{|\gamma_n - t| \le 1} \frac{1}{s-\rho_n}, \quad S_2(s) := \sum_{|\gamma_n - t| > 1} \frac{1}{s-\rho_n},$$

and show that

$$S_2(s) = O(\log|t|), \qquad -1 \le \sigma \le 2, \quad |t| \ge 2. \tag{43}$$

Using the logarithmic derivative of the Euler product given in (4), as in (11), it is easily verified that $\frac{\zeta'(2+it)}{\zeta(2+it)} = O(1)$. Furthermore, due to (37),

$$|S_1(2+it)| \le \sum_{|\gamma_n - t| \le 1} \frac{1}{|2+it-\rho_n|} \le \sum_{|\gamma_n - t| \le 1} 1 = O(\log|t|).$$

Together with (42), these estimations imply $S_2(2+it) = O(\log|t|)$ for $|t| \ge 2$. This in turn gives

$$O(\log|t|) = \mathrm{Re}(S_2(2+it)) = \sum_{|\gamma_n - t| > 1} \mathrm{Re}\,\frac{1}{2+it-\rho_n}$$

$$= \sum_{|\gamma_n - t| > 1} \frac{2-\beta_n}{(2-\beta_n)^2 + (t-\gamma_n)^2} > \sum_{|\gamma_n - t| > 1} \frac{1}{4+(t-\gamma_n)^2}$$

for $|t| \ge 2$. Therefore, with the notation $\sum' := \sum_{|\gamma_n - t| > 1}$,

$$|S_2(s) - S_2(2+it)| \le \sum{}' \left| \frac{1}{s-\rho_n} - \frac{1}{2+it-\rho_n} \right|$$

$$= \sum{}' \frac{2-\sigma}{|s-\rho_n| \cdot |2+it-\rho_n|} < \sum{}' \frac{3}{(t-\gamma_n)^2} < \sum{}' \frac{15}{4+(t-\gamma_n)^2}$$

has order $O(\log|t|)$ as $-1 \le \sigma \le 2$, $|t| \ge 2$. Together with $S_2(2+it) = O(\log|t|)$ from above, this proves (43). From (42) and (43) we have

$$\frac{\zeta'(s)}{\zeta(s)} = \sum_{|\gamma_n - t| \le 1} \frac{1}{s-\rho_n} + O(\log|t|), \qquad -1 \le \sigma \le 2, \ |t| \ge 2. \tag{44}$$

Thus, in order to prove $I = O(\log T)$, it suffices to show

$$\sum_{|\gamma_n - T| \le 1} \mathrm{Im} \left(\int_{1/2}^{2} \frac{d\sigma}{\sigma + iT - \rho_n} \right) = O(\log T). \tag{45}$$

By (21), each summand of the left-hand expression equals the net change of $\arg(s-\rho_n)$ on $[1/2+iT, 2+iT]$, thus its absolute value is $< \pi$. By (37), the number of summands is $< 4\log T$, for large T. Thus, the modulus of the left hand side of (45) is $< 4\pi \log T$, for large T. This, in turn, implies $I = O(\log T)$. □

C Appendix: Riemann-von Mangoldt formula

In the first part C.1 of this appendix we outline some of the details that were omitted in the sketch of the reasoning leading to the Riemann–von Mangoldt explicit formula 4.6. The method is based on [57], §87, see also [50], §12.2. For an alternative method, which does not use contour integration, see [25], §3.2-3.5. In the second part C.2 we return to Riemann's paper [76] and suggest a few further ideas related to the explicit formula and RH.

C.1 The main step in the derivation of 4.6 that needs justification is showing that the integrals on the top, bottom, and left edges of the closed rectangular contour $\mathcal{C}$ with vertices $a \pm iT$, $-(2n+1) \pm iT$, $a > 1$ fixed (for convenience, we also assume $a \leq 2$), all approach 0 as $T \to \infty$ suitably (see (a) below) and $n \in \mathbb{N}$, $n \geq T \log T$. To do this, we start with (37) and (44). Using these one can show

(a) T can be chosen arbitrarily large with $|\gamma - T| > \frac{1}{4\log T}$, for any zero $\rho = \beta + i\gamma$ of ζ,

(b) $\zeta'(s)/\zeta(s) = O(\log^2 T)$, $s = \sigma + iT$, $-1 \leq \sigma \leq 2$ and T chosen as in (a).

We also need the estimation

(c) $\zeta'(s)/\zeta(s) = O(\log |s|)$, $\sigma \leq -1$, $|s + 2j| \geq \frac{1}{2}$ for $j \in \mathbb{N}$.

To prove (c) one starts with the logarithmic derivative of the functional equation (18), namely

$$\frac{\zeta'(s)}{\zeta(s)} \;=\; \log(2\pi) + \frac{\pi}{2}\cot\frac{\pi s}{2} - \frac{\Gamma'(1-s)}{\Gamma(1-s)} - \frac{\zeta'(1-s)}{\zeta(1-s)},$$

taking $\sigma \leq -1$ so that $1 - \sigma \geq 2$. It is easy to show $\cot \frac{\pi s}{2}$ is bounded for $|s + 2j| \geq \frac{1}{2}$, $j \in \mathbb{Z}$, i.e. on the region given in (c). Similarly, the boundedness of $\zeta'(1-s)/\zeta(1-s)$ for $1 - \sigma \geq 2$ is easily shown, e.g.

using (11). Finally, $\Gamma'(1-s)/\Gamma(1-s) = O(\log|1-s|) = O(\log|s|)$ for $1-\sigma \geq 2$, see (23), completing the derivation of (c).

Using (b) it is not hard to show that integral over the portion of the top and bottom edges of $\mathcal{C}$, where $-1 \leq \sigma \leq a$, tends to 0 as $T \to \infty$. And, using (c), one establishes the same for the left edge of $\mathcal{C}$ and the remaining portions of the horizontal edges. This completes the proof of the explicit formula 4.6.

C.2 Let us now return to Riemann's paper [76] and make a few concluding remarks. As mentioned in §4, and as the title of [76] suggests, Riemann's main objective was to obtain an explicit formula for $\pi(x)$, which was only proved later [68], see also [57], §88. He used the closely related function

$$\Pi(x) := \sum_{p^m \leq x} \frac{1}{m} = \sum_n \frac{1}{n}\pi(x^{1/n}). \tag{46}$$

Here the number of non-vanishing terms equals $N := \left\lfloor \frac{\log x}{\log 2} \right\rfloor$ (because $\pi(u) = 0$ for $u < 2$), and this together with (14) easily imply $\Pi(x) - \pi(x) = \frac{1}{2}\pi(x^{1/2}) + O(x^{1/3}) = O(\frac{x^{1/2}}{\log x})$, cf. [57], §5. By the method of Fourier inversion he obtained the explicit formula in the following form[9]:

$$\tilde{\Pi}(x) = \mathrm{li}(x) - \sum_\rho \mathrm{li}(x^\rho) + \int_x^\infty \frac{dx}{(x^2-1)x\log x} - \log 2, \tag{47}$$

for $x > 1$, where $\tilde{\Pi}$ is defined as in §2, and $\mathrm{li}(x) := \int_0^x \frac{dt}{\log t}$ is defined as the Cauchy principal value $\lim\limits_{\varepsilon \downarrow 0} \left(\int_0^{1-\varepsilon} + \int_{1+\varepsilon}^x\right)$, differing from $\mathrm{Li}(x)$ by the constant $\mathrm{li}(2) \approx 1.04516$, cf. [25], §1.14-1.16. In his paper $\tilde{\Pi}$ is denoted f, and $\tilde{\pi}$ (again defined as in §2) denoted F.

[9] To quote Riemann: "By setting these values in the expression for $f(x)$, one finds

$$f(x) = \mathrm{Li}(x) - \sum_\alpha [\mathrm{Li}(x^{(1/2)+\alpha i}) + \mathrm{Li}(x^{(1/2)-\alpha i})] + \int_x^\infty \frac{1}{x^2-1}\frac{dx}{x\log x} + \log\xi(0),$$

where the sum $\sum_\alpha$ is over all positive roots (or all roots with positive real parts) of the equation $\xi(\alpha) = 0$, ordered according to their size", cf. Appendix of [25].

Here $\alpha = (1/2-\rho)i$, and, in the notation of the present paper, Riemann's $\mathrm{Li}(x)$ means $\mathrm{li}(x)$, while Riemann's $\xi(u)$ means $\xi(1/2+ui)$. The term $\log\xi(0)$ is in fact erroneous, see 5.5 and the footnote to 5.3 (e).

Riemann inverted (46) by means of the Möbius inversion formula to obtain

$$\tilde{\pi}(x) = \tilde{\Pi}(x) + \sum_{n=2}^{\infty} \frac{\mu(n)}{n} \tilde{\Pi}(x^{1/n}) = \tilde{\Pi}(x) + \sum_{n=2}^{N} \frac{\mu(n)}{n} \tilde{\Pi}(x^{1/n}), \tag{48}$$

where $\mu(n)$ is 0 when n is divisible by a prime square, and otherwise $(-1)^r$ where r is the number of distinct prime divisors of n. Substituting (47) into (48) gives an explicit formula for $\tilde{\pi}(x)$, cf. [25], §1.17 and §5.4:

$$\tilde{\pi}(x) = \sum_{n=1}^{N} \frac{\mu(n)}{n} \mathrm{li}(x^{1/n}) - \sum_{n=1}^{N} \sum_{\rho} \frac{\mu(n)}{n} \mathrm{li}(x^{\rho/n}) + \text{lesser terms}, \tag{49}$$

thereby achieving the main goal of [76]. Notice that $\sum_\rho$ in (49) can be restricted to a (sufficiently large, depending on x) finite number of terms, if the right hand side of (49) is replaced by the nearest half integer to it (since the left hand side is always a half integer).

Having done this, Riemann speculates (on the final page of his paper) that the main term in (49) is given by the first (finite) sum:

$$\pi(x) \approx \mathrm{li}(x) + \sum_{n=2}^{N} \frac{\mu(n)}{n} \mathrm{li}(x^{1/n}) \tag{50}$$

$$= \mathrm{li}(x) - \frac{1}{2}\mathrm{li}(x^{1/2}) - \frac{1}{3}\mathrm{li}(x^{1/3}) - \frac{1}{5}\mathrm{li}(x^{1/5}) + \frac{1}{6}\mathrm{li}(x^{1/6}) + \dots,$$

and that the estimate $\pi(x) \approx \mathrm{li}(x)$ has negative error of order $O(x^{1/2})$. To quote him once more: "Thus the known approximation $F(x) = \mathrm{Li}(x)$ is correct only to an order of magnitude of $x^{1/2}$ and gives a value which is somewhat too large...". His prediction that $\mathrm{li}(x)$ should overestimate $\pi(x)$ is indeed the case for all x within present computational power. However, Littlewood [63] showed that the difference $\pi(x) - \mathrm{li}(x)$ changes sign infinitely often, indeed $\overline{\lim} \frac{\pi(x)-\mathrm{li}(x)}{\mathrm{li}(x^{1/2}) \log\log\log x} \geq \frac{1}{6}$ and $\underline{\lim} \frac{\pi(x)-\mathrm{li}(x)}{\mathrm{li}(x^{1/2}) \log\log\log x} \leq -\frac{1}{6}$ as $x \to \infty$, in particular $\mathrm{li}(x)$ will underestimate $\pi(x)$ for some sequence $x_n \to \infty$. Skewes [81] showed that $x_1 < 10_4(3)$, where $10_1(x) = 10^x$, $10_2(x) = 10^{10_1(x)}$, and so on. This bound has been improved to $x_1 < 1.65 \times 10^{1165}$ by Lehman [61], and afterwards further improved by others. It is interesting that, for $x \leq 10^7$, the estimate (50) is substantially more accurate than $\pi(x) \approx \mathrm{li}(x)$, as

Table III in [25], §1.17, shows. However, due to the Littlewood result, for x large, the "periodic terms" of the (essentially finite, as explained above) sum $\sum_\rho \mathrm{li}(x^{\rho/n})$ in (49) should be also taken into account in estimating $\pi(x)$, cf. [25], §5.4. Here the number of "significant" periodic terms is large with x, thus eventually (as $x \to \infty$) every periodic term $\mathrm{li}(x^{\rho/n})$ in (49) becomes as significant as the nonperiodic term $-\frac{1}{2}\mathrm{li}(x^{1/2})$.

It is perhaps even more interesting to consider Riemann's error estimate of $O(x^{1/2})$ in the above statement, for the approximation $\pi(x) \approx \mathrm{li}(x)$. This is done very carefully in [25], §1.17 and §5.5. It is shown that each individual periodic term $\mathrm{li}(x^{\rho_n})$ in (47) equals $\frac{x^{\rho_n}}{\rho_n \log x} + O(\frac{x^{\beta_n}}{\log^2 x})$ and would not be less in magnitude than $O(\frac{x^{\beta_n}}{\log x})$, as $x \to \infty$, where as usual $\rho_n = \beta_n + i\gamma_n$. If RH were false then $\beta_n > \frac{1}{2}$ for some n, thus the contribution of this term to the error in PNT would grow more rapidly than $O(x^{(1/2)+\varepsilon})$, for some $\varepsilon > 0$. Moreover, this would also apply to the total error in PNT, see 4.5, A.3, or [25], §5.5. Thus, it is very probable that at this stage (the final page) of his paper, Riemann assumed the validity of RH. To quote Bombieri [10], "it is quite likely that he saw how his hypothesis was central to the question of how good an approximation to $\pi(x)$ one may get from his formula."

Acknowledgement

The authors are grateful to R. K. Guy, J. P. Jones, and N.-P. Skoruppa for useful discussions.

E. A. Kudryavtseva
Department of Mathematics and Mechanics,
Moscow State University,
Moscow 119992,
Russia,
eakudr@mech.math.msu.su

F. Saidak
Department of Mathematics,
Wake Forest University,
Winston-Salem, NC 27109,
USA,
saidakf@wfu.edu

P. Zvengrowski
Department of Mathematics and Statistics,
University of Calgary,
Calgary, AB, T2N 1N4,
Canada,
zvengrow@ucalgary.ca

References

[1] Abel N. H., *Solution de quelques problèmes à l'aide d'intégrales définies*, Mag. Naturvidenskaberne **2** (1823). (Also in: "Oeuvres" 1, Christiania (1881), 11–27.)

[2] Ankeny N. C., *The least quadratic non residue*, Ann. of Math. (2) **55** (1952), 65–72.

[3] Artin E., The Gamma Function, Holt, Reinhart, and Winston, New York, 1964.

[4] Bach E., *Explicit bounds for primality testing and related problems*, Math. Comp. **55** No. 191 (1990), 355–380.

[5] Backlund R., *Sur les zéros de la fonction $\zeta(s)$ de Riemann*, C. R. Acad. Sci. Paris **158** (1914), 1979–1982.

[6] Backlund R., Über die Nullstellen der Riemannschen Zetafunktion, Dissertation, Helsingfors, 1916.

[7] Bernoulli J., *Ars Conjectandi*, Basel (1713), 95–98.

[8] Bohr H.; Landau E., *Ein Satz über Dirichletsche Reihen mit Anwendung auf die ζ-Funktion und die L-Funktionen*, Rend. Circ. Mat. Palermo **37** (1914), 269–272. (Also Bohr's "Collected Works", vol. I, B13.)

[9] Bombieri E., *On the large sieve*, Mathematika **12** (1965), 201–225.

[10] Bombieri E., *Problems of the Millennium: the Riemann Hypothesis*, http://www.claymath.org/millennium/Riemann_Hypothesis, 2000.

[11] Burgess D. A., *The distribution of quadratic residues and non-residues*, Mathematika **4** (1957), 106–112.

[12] Chebyshev P. L., *Sur la fonction qui détermine la totalité des nombres premiers inférieurs à une limite donnée*, J. Math. Pures Appl. (1) **17** (1852), 341–365. (Available in both French and Russian Editions of Collected Works.)

[13] Chebyshev P. L., *Mémoire sur les nombres premiers*, J. Math. Pures Appl. (1) **17** (1852), 366–390. (Available in both French and Russian Editions of Collected Works.)

[14] Chebyshev P. L., Oeuvres de P. L. Tchebychef, tome II, A. Markoff et N. Sonin, Chelsea, New York, 1962.

[15] Chudakov N. G., *On zeros of Dirichlet's L-functions* (Russian), Matem. Sbornik **1** No. 43 (1936), 591–602.

[16] Conrey J. B., *More than two fifths of the zeros of the Riemann zeta function are on the critical line*, J. Reine Angew. Math. **399** (1989), 1–26.

[17] Conrey J. B., *The Riemann hypothesis*, Notices Amer. Math. Soc. (March 2003), 341–353.

[18] Cramér H., *Some theorems concerning prime numbers*, Arkiv för Mat. Astr. o. Fys. **15** No. 5 (1920), 1–32.

[19] Cramér H., *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arithmet. **2** No. 1 (1936), 23–46.

[20] d'Aniello C., *On some inequalities for the Bernoulli numbers*, Rend. Circ. Mat. Palermo (2) **43** No. 3 (1994), 329–332.

[21] Davis P. J., *Leonhard Euler's integral: A historical profile of the gamma function*, Amer. Math. Monthly **66** (1959), 849–869.

[22] Deligne P., *La conjecture de Weil. I*, Inst. Hautes Études Sci. Publ. Math. **43** (1974), 273–307.

[23] Descartes R., La Géométrie, Leyden, 1637.

[24] Dirichlet P. G. Lejeune, *Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendlich viele Primzahlen enthält*, Abh. Kgl. Preuss. Akad. Wiss. Berlin (gelesen in der Akad. der Wiss. 1837), 45–81, 1839 (Also in: "Werke", Bd. 1, (1889), 313–342.)

[25] Edwards H. M., Riemann's Zeta Function, Academic Press, San Diego, 1974.

[26] Erdős P., *On a new method in elementary number theory which leads to an elementary proof of the PNT*, Proc. Nat. Acad. Sci. U.S.A. **35** (1949), 374–384.

[27] Euclid, The thirteen books of Euclid's, Book IX (Engl. transl. by Heath T. L.), Cambridge Univ. Press, Cambridge, 1926. (Also in: "Opera Omnia", vol. 2.)

[28] Euler L., *De progressionibus transcendentibus seu quarum termini generales algebraice dari nequeunt* (presented to the St. Petersburg Acad. 1729), Comm. Acad. Sci. Petropol. **5** (1738), 36–57. (Also in: "Opera Omnia", Ser. 1, vol. 14, 1–24.)

[29] Euler L., *Methodus generalis summandi progressiones* (presented to the St. Petersburg Acad. 1732), Comm. Acad. Sci. Petropol. **6** (1738), 68–97. (Also in: "Opera Omnia", Ser. 1, vol. 14, 42–72.)

[30] Euler L., *De summis serierum reciprocarum* (read in the St. Petersburg Acad. 1735), Comm. Acad. Sci. Petropol. **7** (1740), 123–134. (Also in: "Opera Omnia", Ser. 1, vol. 14, 73–86.)

[31] Euler L., *De progressionibus harmonicis observationes* (presented to the St. Petersburg Acad. 1734), Comm. Acad. Sci. Petropol. **7** (1740), 150–161. (Also in: "Opera Omnia", Ser. 1, vol. 14, 87–100.)

[32] Euler L., *Variae observationes circa series infinitas* (presented to the St. Petersburg Acad. 1737), Comm. Acad. Sci. Imp. Petropol. **9** (1744), 160–188. (Also in: "Opera Omnia", Ser. 1, vol. 14, 217–244.)

[33] Euler L., Introductio in analysin infinitorum, Bousquet and Socios., Lausanne vol. 1, chap. 15, 1748. (Also in: "Opera Omnia", Ser. 1, vol. 8. Engl. transl. by J. Blanton, Introduction to Analysis of the Infinite, Book I, Springer-Verlag, 1988.)

[34] Euler L., Institutiones calculi differentialis cum eius usu in analysi finitorum ac doctrina serierum, Pt. 2, Ch. 5, 6. Acad. Imp. Sci. Petropol., St. Petersburg, 1755. (Also in: "Opera Omnia", Ser. 1, vol. 10.)

[35] Euler L., *Remarques sur un beau rapport entre les séries des puissances tant directes que réciproques* (lu en 1749) (presented 1761), Hist. Acad. Roy. Sci. Belles-Lettres Berlin **17** (1768), 83–106. (Also in: "Opera Omnia", Ser. 1, vol. 15, 70–90.)

[36] Gauss C. F., *Doctoral Dissertation*, Göttingen Univ., 1799.

[37] Gauss C. F., Disquisitiones arithmeticae, G. Fleisher, Lipsiae, 1801. (Also in: "Werke" **1**, 1899, 1–474.)

[38] Gauss C. F., Letter to Encke J. F. dated December 24, 1849. In: "Werke", 1. Aufl., Kgl. Ges. Wiss. Göttingen **2** (1863),444–447 (Engl. transl. by G. Ehrlich: Appendix B to [40]). The scan of the original: www.math.princeton.edu/~ytschink/.gauss/Briefe-B.pdf

[39] Girard A., L'Invention Nouvelle en l'Algébre, Amsterdam, 1629.

[40] Goldstein L. J., *A history of the prime number theorem*, Amer. Math. Monthly **80** No. 6 (1973), 599–615.

[41] Graham S. W.; Ringrose C. J., *Lower bounds for least quadratic nonresidues*, In: Analytic number theory (Allerton Park, IL, 1989), 269–309, Progr. Math. 85, Birkhauser Boston, Boston, MA, 1990.

[42] Gram J.-P., *Note sur les zéros de la fonction $\zeta(s)$ de Riemann*, Oversigt Kong. Danske Vid. Selsk. Forh. (1902), 3–15. (Also in: Acta Math. **27** (1903), 289–304.)

[43] Hadamard J., *Étude sur les propriétés des fonctions entières et en particulier d'une fonction considérée par Riemann*, J. Math. Pures Appl. (4) **9** (1893), 171–215.

[44] Hadamard J., *Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques*, Bull. Soc. Math. France **24** (1896), 199–220.

[45] Hardy G. H., *Sur les zéros de la fonction $\zeta(s)$ de Riemann*, C. R. Acad. Sci. Paris **158** (1914), 1012–1014.

[46] Hardy G. H., Divergent Series, Oxford Univ. Press, Oxford, 1949.

[47] Hardy G. H.; Littlewood J. E., *The zeros of Riemann's zeta-function on the critical line*, Math. Z. **10** (1921), 283–317.

[48] Hutchinson J. I., *On the roots of the Riemann zeta-function*, Trans. Amer. Math. Soc. **27** (1925), 49–60.

[49] Ingham A. E., The distribution of prime numbers, Cambridge Univ. Press, Cambridge, 1990.

[50] Ivič A., The Riemann Zeta-Function, J. Wiley & Sons, New York, 1985.

[51] James I., Remarkable mathematicians, Cambridge Univ. Press, Cambridge, 2002.

[52] Jameson G. J. O., The prime number theorem, London Math. Soc. student texts 53, Cambridge Univ. Press, Cambridge, 2003.

[53] Karatsuba A. A.; Voronin S. M., The Riemann Zeta-Function, de Gruyter Expositions in Mathematics, Berlin, 1992.

[54] von Koch H., *Sur la distribution des nombres premiers*, Acta Math. **24** (1901), 159–182.

[55] Kloosterman H. D., *Een integraal voor de ζ-functie van Riemann*, Christiaan Huygens Math. Tijdschrift **2** (1922), 172–177.

[56] Landau E., *Neue Beiträge zur analytischen Zahlentheorie*, Rend. Circ. Mat. Palermo **27** (1909), 46–58.

[57] Landau E., Handbuch der Lehre von der Verteilung der Primzahlen, Teubner, Leipzig, 1909. (Reprinted 1974 by Chelsea, New York.)

[58] Lang S., Complex Analysis, 4th ed., GTM 103, Springer-Verlag, New York, 1999.

[59] Leeming D. J., *An asymptotic estimate for the Bernoulli numbers and the Euler numbers*, Can. Math. Bull. **20** No. 1 (1977), 109–111.

[60] Legendre A. M., Essai sur la Théorie des Nombres, 1st ed., Duprat, Paris, 1798; 2nd ed., Courcier, Paris, 1808. Théorie des Nombres, 3rd ed., vol. 2, Didot, Paris, 1830; 4th ed., vol. 2, Hermann, Paris, 1900. (Reprinted, Librairie Sci. Tech., A. Blanchard, Paris, 1955.)

[61] Lehman R. S., *On the difference* $\pi(x)-\mathrm{li}(x)$, Acta Arith. **11** (1966), 397–410.

[62] Levinson N., *More than one-third of zeros of Riemann's zeta-function are on* $\sigma = 1/2$, Adv. Math. **13** (1974), 383–436.

[63] Littlewood J. E., *Sur la distribution des nombres premiers*, C. R. Acad. Sci. Paris **158** (1914), 1869–1872.

[64] Littlewood J. E., *On the zeros of Riemann zeta-function*, Proc. Cambr. Phyl. Soc. **22** (1924), 295–318.

[65] Littlewood J. E., *The quickest proof of the prime number theorem*, Acta Arith. **18** (1971), 83–86.

[66] Lou S.; Yau Q., *A Chebyshev's type of prime number theorem in a short interval* (II), Hardy–Ramanujan J. **15** (1992), 1–33.

[67] Maclaurin C., Treatise of Fluxions, Edinburgh, 1742, p. 672.

[68] von Mangoldt H., *Zu Riemanns Abhandlung 'Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse'*, J. Reine Angew. Math. **114** (1895), 255–305.

[69] von Mangoldt H., *Zur Verteilung der Nullstellen der Riemannschen Funktion* $\xi(t)$, Math. Ann. **60** (1905), 1–19.

[70] Mellin Hj., *Eine Formel für den Logarithmus transcendenter Funktionen von endlichem Geschlecht*, Acta Soc. Sci. Fenn. **29** No. 4 (1900).

[71] Mertens F., *Ein Beitrag zur analytischen Zahlentheorie*, J. Reine Angew. Math. **78** (1874), 46–62.

[72] Montgomery H. L., *Zeros of L-functions*, Invent. Math. **8** (1969), 346–354.

[73] Montgomery H. L., Topics in multiplicative number theory, Lecture Notes in Math. 227, Springer, Berlin, 1971.

[74] Newman J., *Simple analytic proof of the prime number theorem*, Amer. Math. Monthly **87** (1980), 693–696.

[75] Pólya G., *Über die Verteilung der quadratischen Reste und Nichtreste*, Nachr. Ges. Wiss. Göttingen Math.-Phys. (1918), 21–29.

[76] Riemann B., *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsber. Kgl. Preuss. Akad. Wiss. Berlin (1859), 671–680; 1860. (Also in: "Gesammelte mathematische Werke und wissenschaftlicher Nachlass", 2. Aufl., Teubner, Leipzig (1892), 145–155.)

[77] Saidak F.; Zvengrowski P., *On the modulus of the Riemann zeta function in the critical strip*, Math. Slovaca **53** No. 2 (2003), 145–172.

[78] Selberg A., *On the zeros of Riemann's zeta-function*, Skr. Norske Vid. Akad. Oslo I, No. 10 (1942).

[79] Selberg A., *An elementary proof of the prime-number theorem*, Ann. of Math. (2) **50** (1949), 305–313.

[80] Siegel C. L., *Über Riemanns Nachlass zur analytischen Zahlentheorie*, Quellen und Studien zur Geschichte der Math. Astron. und Phys., Abt. B: Studien 2 (1932), 45–80. (Also in: "Gesammelte Abhandlungen", **1** (1966), 275–310, Springer-Verlag, Berlin and New York.)

[81] Skewes S., *On the difference* $\pi(x) - \operatorname{li} x$. *II*, Proc. London Math. Soc. (3) **5** (1955), 48–70.

[82] Spira R., *An inequality for the Riemann zeta-function*, Duke Math. J. **32** (1965), 247–250.

[83] Stieltjes T. J., *Sur le developpement de* $\log \Gamma(a)$, J. Math. Pures Appl. **5** (1889), 425–444.

[84] Stirling J., Methodus differentialis: sive tractatus de summatione et interpolatione serierum infinitarum, Gul. Bowyer, London, 1730.

[85] Titchmarsh E. C., The Theory of the Riemann Zeta-Function, 2nd ed. (revised by D. R. Heath-Brown), Oxford Univ. Press, Oxford, 1986.

[86] Titchmarsh E. C., *The zeros of the Riemann zeta-function*, Proc. Roy. Soc., Ser. A **151** (1935), 234–255.

[87] Titchmarsh E. C., *The zeros of the Riemann zeta-function*, Proc. Roy. Soc., Ser. A **157** (1936), 261–263.

[88] de la Vallée-Poussin C.-J., *Recherches analytiques sur la théorie des nombres premiers* (premiére partie), Ann. Soc. Sci. Bruxelles (1) **20** (1896), 183–256, 281–397.

[89] de la Vallée-Poussin C.-J., *Sur la fonction* $\zeta(s)$ *de Riemann et le nombre des nombres premiers inférieurs à une limite donnée*, Mém. Courronnés et Autres Acad. Roy. Sci., des Lettres Beaux-Arts Belg. **59** No. 1 (1899–1900).

[90] Vinogradov J. M., *On the distribution of residues and non-residues of powers*, J. Phys.-Math. Soc. Perm **1** (1918), 94–98.

[91] Vinogradov I. M., *On a general theorem concerning the distribution of the residues and non-residues of powers*, Trans. Amer. Math. Soc. **29** No. 1 (1927), 209–217.

[92] Vinogradov J. M., *On Weyl's sums*, Matem. Sbornik **42** (1935), 521–530.

[93] Vinogradov I. M., *On the estimation of a trigonometrical sum* (Russian), Dokl. Akad. Nauk SSSR **34** No. 7 (1942), 199–200.

[94] Vinogradov I. M., *A new estimate for the function* $\zeta(1+it)$ (Russian), Izv. Akad. Nauk SSSR, Ser. Mat. **22** No. 1 (1958), 161–164.

[95] Walfisz, A., Weylsche Exponentialsummen in der Neueren Zahlentheorie, VEB Deutscher Verlag der Wiss., Berlin, 1963.

[96] Weierstrass K., *Zur Theorie der eindeutigen analytischen Funktionen*, Abh. Kgl. Preuss. Akad. Wiss. Berlin (1876), 11–60. (Also in: "Math. Werke", vol. 2, Mayer & Müller, Berlin (1895), 77–124.)

[97] Weil A., *Prehistory of the zeta-function.* In: Number theory, trace formulas and discrete groups (Oslo, 1987), Acad. Press, Boston, 1989, 1–9.

[98] Westzynthius E., *Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind*, Comment. Phys.-Math. Soc. Sci. Fenn. **5** (1931), No. 25, 1–37.