Math 259: Introduction to Analytic Number Theory

pseudo-syllabus

0. Introduction: What is analytic number theory?

1. Distribution of primes before complex analysis: classical techniques (Euclid, Euler); primes in arithmetic progressions via Dirichlet characters and L-series; Čebyšev's estimates on $\pi(x)$.

2. Distribution of primes using complex analysis: $\zeta(s)$ and $L(s, \chi)$ as functions of a complex variable, and the proof of the Prime Number Theorem and its extension to Dirichlet; blurb for Čebotarev density; functional equations; the Riemann hypothesis, extensions, generalizations and consequences.

3. Selberg's quadratic sieve and applications.

4. Analytic estimates on exponential sums (van der Corput etc.); prototypical applications: Weyl equidistribution, upper bounds on $|\zeta(s)|$ and $|L(s, \chi)|$ on vertical lines, lattice point sums.

5. Lower bounds on discriminants, conductors, etc. from functional equations; geometric analogue: how many points can a curve of genus $g \to \infty$ have over a given finite field?

6. Analytic bounds on coefficients of modular forms and functions; applications to counting representations of integers as sums of squares, etc.

Prerequisites While Math 259 will proceed at a pace appropriate for a graduate-level course, its prerequisites are perhaps surprisingly few: complex analysis at the level of Math 113, and linear algebra and basic number theory (up to say arithmetic in the field $\mathbf{Z}/p\mathbf{Z}$ and Quadratic Reciprocity). Some considerably deeper results (such as estimates on Kloosterman sums) will be cited but may be regarded as black boxes for our purposes. If you know about algebraic number fields or modular forms or curves over finite fields, you'll get more from the course at specific points, but these points will be in the nature of scenic detours that are not required for the main journey.

Texts Lecture notes will be handed out periodically, and can also be found on the course webpage. There is no textbook: this class is an introduction to several different flavors of analytic methods in number theory, and I know of no one work that covers all this material. Thus I intend to expand and edit the lecture notes to put together a textbook, which may become available by the next time I teach the class... Supplementary readings such as Serre's *A Course in Arithmetic* and Titchmarsh's *The Theory of the Riemann Zeta-Function* will be suggested as we approach their respective territories.

Office Hours 335 Sci Ctr, Thursdays 3–4:30 PM (occasionally shortened by Colloquium or faculty meetings), or e-mail me at `elkies@math` (`elkies@math.harvard.edu` from outside Harvard) to ask questions or set up an alternative meeting time.

Grading There will be no required homework, though the lecture notes will contain recommended exercises. If you are taking Math 259 for a grade (i.e., are not a post-Qual math graduate student exercising your EXC option), tell me so we can work out an evaluation and grading procedure. This will most likely be either an expository final paper or an in-class presentation on some aspect of analytic number theory related to but just beyond what we cover in class. Which grading method is appropriate will be determined once the class size has stabilized after "Shopping Period". The supplementary references will be one good source for paper or presentation topics.

Math 259: Introduction to Analytic Number Theory

What is analytic number theory?

One may reasonably define analytic number theory as the branch of mathematics that uses analytical techniques to address number-theoretical problems. But this "definition", while correct, is scarcely more informative than the phrase it purports to define. (See [Wilf 1982].) What kind of problems are suited to "analytical techniques"? What kind of mathematical techniques will be used? What style of mathematics is this, and what will its study teach you beyond the statements of theorems and their proofs? The next few sections briefly answer these questions.

The problems of analytic number theory. The typical problem of analytic number theory is an enumerative problem involving primes, Diophantine equations, or similar number-theoretic objects, and usually concerns what happens for large values of some parameter. Such problems are of long-standing intrinsic interest, and the answers that analytic number theory provides often have uses in mathematics (see below) or related disciplines (notably in various algorithmic aspects of primes and prime factorization, including applications to cryptography). Examples of problems that we shall address are:

- How many 100-digit primes are there, and how many of these have the last digit 7? More generally, how do the prime-counting functions $\pi(x)$ and $\pi(x; a \bmod q)$ behave for large x? [For the 100-digit problems we would take $x = 10^{99}$ and $x = 10^{100}$, $q = 10$, $a = 7$.]
- Given a prime $p > 0$, a nonzero $c \bmod p$, and integers a_1, b_1, a_2, b_2 with $a_i < b_i$, how many pairs (x_1, x_2) of integers are there such that $a_i < x_i < b_i$ $(i = 1, 2)$ and $x_1 x_2 \equiv c \bmod p$? For how small an H can we guarantee that if $b_i - a_i > H$ then there is at least one such pair?
- Is there an integer n such that the first eleven digits of $n!$ are 31415926535? Are there infinitely many such n? How many such n are there of at most 1000 digits?
- Given integers n, k, how many ways are there to represent n as a sum of k squares? For instance, how many integer solutions has the equation $a^b + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + h^2 = 10^{100}$?

As often happens in mathematics, working on such down-to-earth questions quickly leads us to problems and objects that appear to belong to completely different mathematical disciplines:

- Analyze the Riemann zeta function $\zeta(s) := \sum_{n=1}^{\infty} 1/n^s$ and Dirichlet L-functions such as

$$L(s) := 1 - 3^{-s} - 7^{-s} + 9^{-s} + 11^{-s} - 13^{-s} - 17^{-s} + 19^{-s} + - - + \cdots$$

as functions of a *complex* variable s.

- Prove that the "Kloosterman sum"

$$K(p;a,b) := \sum_{x=1}^{p-1} \exp\left(\frac{2\pi i}{p}(ax + bx^{-1})\right)$$

(with x^{-1} being the inverse of x mod p) has absolute value at most $2\sqrt{p}$.

- Show that if a function $f : \mathbf{R} \to \mathbf{R}$ satisfies reasonable smoothness conditions then for large N the absolute value of the exponential sum

$$\sum_{n=1}^{N} \exp(if(n))$$

grows no faster than N^θ for some $\theta < 1$ (with θ depending on the conditions imposed on f).

- Investigate the coefficients of modular forms such as

$$\eta^8\eta_2^8 = q\prod_{n=1}^{\infty}(1-q^n)^8(1-q^{2n})^8 = q - 8q^2 + 12q^3 + 64q^4 - 210q^5 - 96q^6 \cdots.$$

Fortunately it will turn out that the route from (say) $\pi(x)$ to $\zeta(s)$ is not nearly as long and tortuous as that from $x^n + y^n = z^n$ to deformations of Galois representations...[1]

The techniques of analytic number theory. A hallmark of analytic number theory is the treatment of number-theoretical problems (usually enumerative, as noted above) by methods often relegated to the domain of "applied mathematics": elementary but clever manipulation of sums and integrals; asymptotic and error analysis; Fourier series and transforms; contour integrals and residues. While there is still good new work to be done along these lines, much contemporary analytic number theory also uses advanced tools from within and outside number theory (for instance, modular forms beyond the upper half-plane, Laplacian spectral theory). Nevertheless, in this introductory course we shall emphasize the classical methods characteristic of analytic number theory, on the grounds that they are rarely treated in this Department's courses, while our program already offers ample exposure to the algebraic/geometric tools. As already noted in the pseudo-syllabus, we shall on a few occasions invoke results that depend on deep (non-analytic) techniques, but we shall treat them as *deus ex mathematica*, developing only their analytic applications.

The style of analytic number theory. It has often been said that there are two kinds[2] of mathematicians: theory builders and problem solvers. In

[1] See for instance [Stevens 1994] and [Faltings 1995].

[2] Actually there are three kinds of mathematicians: those who can count, and those who cannot.

twentieth-century mathematics, these two styles are epitomized respectively by A. Grothendieck and P. Erdös. The Harvard math curriculum leans heavily towards the systematic, theory-building style; analytic number theory as usually practiced falls in the problem-solving camp. This is probably why, despite its illustrious history (Euclid, Euler, Riemann, Selberg, ...) and present-day vitality, analytic number theory has rarely been taught here — in the past fifteen years there have been only a handful of undergraduate seminars, research/Colloquium talks, and Catalog-listed courses. Now we shall see that there is more to analytic number theory than a bag of unrelated ad-hoc tricks, but it is true that partisans of contravariant functors, adèlic tangent sheaves, and étale cohomology will not find them in the present course. Still, even ardent structuralists can benefit from this course. First, specific results of analytic number theory often enter as necessary ingredients in the only known proofs of important structural results. Consider for example the arithmetic of elliptic curves: the many applications of Dirichlet's theorem on primes in arithmetic progression, and its generalization to Čebotarev's density theorem,[3] include the ground-breaking work of Kolyvagin and of Wiles and Taylor; in [Serre 1981] sieve methods are elegantly applied to the study of the distribution of traces of an elliptic curve;[4] in [Merel 1996] a result (Lemme 5) on the $x_1 x_2 \equiv c \bmod p$ problem is required to bound the torsion of elliptic curves over number fields. Second, the ideas and techniques apply widely. Sieve inequalities, for instance, are also used in probability theory to analyze nearly independent variables; the "stationary phase" methods for obtaining the asymptotic growth of the partition function are also used to estimate oscillatory integrals in enumerative combinatorics, quantum physics, special functions, and elsewhere; even the van der Corput estimates on exponential sums have found combinatorial application [CEP 1996]. Third, working on asymptotic results and error terms can be a healthy complement to the usual quest for exact answers that we might focus on too exclusively. Finally, An ambitious theory-builder should regard the absence thus far of a Grand Unified Theory of analytic number theory not as an insult but as a challenge. Both machinery- and problem-motivated mathematicians should note that some of the more exciting recent work in number theory depends critically on symbiosis between the two styles of mathematics. This course will introduce the main analytic techniques This text introduces the main analytic techniques needed to appreciate, and ultimately to extend, this work.

References

[CEP 1996] Cohn, H., Elkies, N.D., Propp, J.: Local statistics for random domino tilings of the Aztec diamond, Duke Math J. **85** #1 (Oct.96), 117–166.

[3] We shall describe Čebotarev's theorem briefly in the course but not develop it in detail. Given Dirichlet's theorem and the asymptotic formula for $\pi(x; a \bmod q)$, the extra work needed to get Čebotarev is not analytic but algebraic: the development of algebraic number theory and the arithmetic of characters of finite groups. Thus a full treatment of Čebotarev does not alas belong in this course.

[4] My doctoral work on the case of trace zero (see for instance [Elkies 1987]) also used Dirichlet's theorem.

[Elkies 1987] Elkies, N.D.: The existence of infinitely many supersingular primes for every elliptic curve over $\mathbf{Q}$. *Invent. Math.* **89** (1987), 561–567.

[Faltings 1995] Faltings, G.: The Proof of Fermat's Last Theorem by R. Taylor and A. Wiles. *Notices of the AMS*, July 1995, 743–746 (translated by U.F. Mayer from *Testausdruck DMV Mitteilungen* **27**, 3/1995).

[Merel 1996] Merel, L.: Bornes pour la torsion des courbes elliptiques sur les corps de nombres. *Invent. Math.* **124** (1996), 437–449.

[Serre 1981] Serre, J.-P.: Quelques applications du théorème de densité de Chebotarev. *IHES Publ. Math.* **54** (1981), 123–201.

[Stevens 1994] Stevens, G.: *Fermat's Last Theorem*, PROMYS T-shirt, Boston University 1994.

[Wilf 1982] Wilf, H.S.: What is an Answer? *Amer. Math. Monthly* **89** (1992), 289–292.

Math 259: Introduction to Analytic Number Theory

Elementary approaches I: Variations on a theme of Euclid

Like much of mathematics, the history of the distribution of primes begins with Euclid:

Theorem (Euclid [IX, 20])**.** *There are infinitely many primes.*

Euclid's justly famed argument, while often presented as a proof by contradiction, is readily framed as an effective (albeit rather inefficient) construction:

Proof: Given primes $p_1, p_2, \ldots, p_n$, let $P_n = \prod_{k=1}^{n} p_n$, define $N_n = P_n + 1$, and let p_{n+1} be the smallest factor of N_n. Then p_{n+1} is a prime no larger than N_n and different from $p_1, \ldots, p_n$. Thus $\{p_k\}_{k>1}$ is an infinite sequence of distinct primes, Q.E.D.

This answers Yes to the first asymptotic question to ask about

$$\pi(x) := \#\{p \leq x : p \text{ is a positive prime}\} = \sum_{\substack{0<p<x \\ p \text{ prime}}} 1,$$

namely whether $\pi(x) \to \infty$ as $x \to \infty$. Moreover, the proof also gives an explicit upper bound on p_n, and thus a lower bound on $\pi(x)$.

Theorem. *For each integer $n > 0$, there are more than n primes $p < 2^{2^n}$. Equivalently, we have*[1]

$$\pi(x) > \log_2 \log_2 x$$

for all $x > 1$.

Proof: In the proof of Euclid's theorem, we may take $p_1 = 2$, and observe that

$$p_{n+1} \leq N_n = 1 + \prod_{k=1}^{n} p_n \leq 2 \prod_{k=1}^{n} p_n.$$

if equality were satisfied at each step we would have $p_n = 2^{2^{n-1}}$. Thus by induction we see that

$$p_n \leq 2^{2^{n-1}},$$

and of course the inequality is strict once $n > 1$. Therefore if $x \geq 2^{2^{n-1}}$ then $p_k < x$ for $k = 1, 2, \ldots, n$, and so $\pi(x) \geq n$, Q.E.D.

The $P_n + 1$ trick has been adapted to prove some special cases of Dirichlet's theorem on primes in arithmetic progression, which asserts that for coprime integers $q > 0$ and a there are infinitely many primes $p \equiv a \bmod q$. (We shall give the proof later in the course.) Of course the case 1 mod 2 is trivial given Euclid. For $-1 \bmod q$ with $q = 3, 4, 6$, start with $p_1 = q - 1$ and define $N_n = qP_n - 1$.

[1] Q: What sound does a drowning analytic number theorist make?
A: log log log log ... [R. Murty, via B. Mazur]

More generally, for any quadratic character χ there are infinitely many primes p with $\chi(p) = -1$; as a special case, given an odd prime q_0, there are infinitely many primes p which are quadratic nonresidues of q_0. [I'm particularly fond of this argument because I was able to adapt it as the punchline of my doctoral thesis; see [Elkies 1987].] The case of $\chi(p) = +1$ is only a bit trickier.[2] For instance, to prove Dirichlet for $(q, a) = (4, 1)$, let $p_1 = 5$ and $N_n = 4P_n^2 + 1$, and invoke Fermat's theorem on the prime factors of $x^2 + y^2$. Again this argument even yields an explicit lower bound on

$$\pi(x, 1 \bmod 4) := \#\{p \le x : p \text{ is a positive prime congruent to } 1 \bmod 4\},$$

namely[3]

$$\pi(x, 1 \bmod 4) > C \log\log x$$

for some positive constant C.

But Euclid's approach and its variations, however elegant, are not sufficient for our purposes. For one thing, numerical evidence suggests — and we shall soon prove — that $\log_2 \log_2 x$ is a gross underestimate on $\pi(x)$. For another, one cannot prove all cases of Dirichlet's theorem using only variations on the Euclid argument.[4] Our next elementary approaches will address at least the first deficiency.

Exercises

1. Let G be a subgroup of $(\mathbf{Z}/q\mathbf{Z})^*$ other than $(\mathbf{Z}/q\mathbf{Z})^*$ itself. Prove that there are infinitely many primes whose residue modulo q is not in G.

2. Exhibit an explicit value of C such that $\pi(x, 1 \bmod 4) > C \log\log x$ for all $x > 1$.

3. Use cyclotomic polynomials to show more generally that for any q_0, prime or not, there exist infinitely many primes congruent to 1 mod q_0. [Attributed to Euler in [Dickson 1919, Ch.XVIII], a chapter which gives much more information on the history of work on the distribution of primes up to about 1900. Note that $4P_n^2 + 1$ is the fourth cyclotomic polynomial evaluated at $2P_n$.] Show that again the number of such primes $< x$ grows at least as fast as some multiple of $\log\log x$.

4. Show that there are infinitely many primes congruent to 4 mod 5, once more with a $\log\log$ lower bound.

5. [A much later proof of the infinitude of primes that curiously gives the same bound $\pi(x) > \log_2 \log_2(x)$.] Recall that the m-th Fermat number F_m is defined by $F_m = 2^{2^m} + 1$ $(m = 0, 1, 2, \ldots)$. Prove that F_m and $F_{m'}$ are relatively prime

[2] But enough so that a problem from a recent Qualifying Exam for our graduate students asked to prove that there are infinitely many primes congruent to 1 mod 4.

[3] Even a drowning analytic number theorist knows that $\log\log$ and $\log_2 \log_2$ are asymptotically within a constant factor of each other. What is that factor?

[4] This is <u>not</u> a theorem, of course. How could one even define "variation of the Euclid argument" rigorously? But a Euclid-style argument for the infinitude of primes congruent to 2 mod 5 or mod 7 would already be quite impressive.

unless $m = m'$. Conclude that there are at least n primes $p \leq F_{n-1}$, and thus that $\pi(x) > \log_2 \log_2 x$.

Digression

Even a piece of mathematics as venerable as Euclid's proof of the infinitude of primes can continue to suggest very difficult problems. For instance, let p_n be the n-th prime, and let[5] $P_n = \prod_{i=1}^{n} p_i$. We know that $P_n + 1$ must contain a new prime factor, which cannot be p_{n+1} once $n > 1$ (if only because $P_n - 1$ must also contain a new prime factor). Does it happen infinitely often that p_{n+1} is a factor of $P_n + 1$? [This is the case for $n = 1$, 7, 232, 430, and no other $n < 5000$.] What of the primality of $P_n + 1$ itself? It is well-known that $P_n + 1$ is prime for $n = 1, 2, 3, 4, 5$, but $P_6 + 1 = 30031 = 59 \cdot 509$. Only fifteen $n > 5$ have been found for which $P_n + 1$ is prime, of which the smallest is 11 and the largest is 13494.[6] Again it is not known whether this happens infinitely often. Likewise for the primality of $P_n - 1$ and its divisibility by p_{n+1}. For another variation, define $q_1 = 2$ and, for $n > 0$, let q_{n+1} be the smallest prime factor of $(\prod_{i=1}^{n} q_i) + 1$. The sequence $\{q_n\}_{n=1}^{\infty}$ starts

$$2, 3, 7, 43, 13, 53, 5, 6221671, 38709183810571, 139, 2801, 11, \ldots$$

For instance, $q_5 = 13$ because $2 \cdot 3 \cdot 7 \cdot 43 + 1 = 1807 = 13 \cdot 139$. Is this "Euclid-Mullin sequence" [Sloane, A000945] a permutation of the sequence of primes? Probably yes, but proving this will likely be intractable for the foreseeable future. The same is true for the infinitude of primes of the form $P_n \pm 1$, and of n such that $p_{n+1} | P_n \pm 1$.

It should not even be obvious that one should expect that these four sets are all infinite. The heuristics supporting this expectation rely on results on the distribution of primes that we shall develop in the next few weeks.

References

[Dickson 1919] Dickson, L.E.: *History of the Theory of Numbers, Vol. I: Divisibility and Primality.* Washington: Carnegie Inst., 1919.

[Euclid] Euclid, *Elements.*

[Elkies 1987] Elkies, N.D.: The existence of infinitely many supersingular primes for every elliptic curve over **Q**, *Invent. Math.* **89** (1987), 561–568; See also: Supersingular primes for elliptic curves over real number fields, *Compositio Math.* **72** (1989), 165–172.

[Sloane] Sloane, N.J.A.: *On-Line Encyclopedia of Integer Sequences.*
`http://www.research.att.com/~njas/sequences`

[5] By analogy with the "factorial" $n! = \prod_{i=1}^{n} i$, this P_n is sometimes called the n-th "primorial".

[6] Sequence A014545 in [Sloane], where the primality of $P_{13494} + 1$ is attributed to Arlin Anderson, Oct.20, 2000. For the analogous question concerning $P_n - 1$, see Sequence A055704 and A006794.

Math 259: Introduction to Analytic Number Theory

Elementary approaches II: the Euler product

Euler [Eul] achieved the first major advance beyond Euclid's proof by combining his method of generating functions with another highlight of ancient Greek number theory, unique factorization into primes.

Theorem [Euler product]**.** *The identity*

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}. \tag{1}$$

holds for all s such that the left-hand side converges absolutely.

Proof: Here and henceforth we adopt the convention:

The notation $\prod_p$ or $\sum_p$ means a product or sum over prime *p.*

Every positive integer n may be written uniquely as $\prod_p p^{c_p}$, with each c_p a nonnegative integer that vanishes for all but finitely many p. Thus the formal expansion of the infinite product

$$\prod_{p \text{ prime}} \Big(\sum_{c_p=0}^{\infty} p^{-c_p s} \Big) \tag{2}$$

contains each term

$$n^{-s} = \Big(\prod_p p^{c_p} \Big)^{-s} = \prod_p p^{-c_p s}$$

exactly once. If the sum of the n^{-s} converges absolutely, we may rearrange the sum arbitrarily and conclude that it equals the product (2). On the other hand, each factor in this product is a geometric series whose sum equals $1/(1-p^{-s})$. This establishes the identity (2). □

The sum on the left-hand side of (2) is nowadays called the *zeta function*

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots = \sum_{n=1}^{\infty} n^{-s} \; ;$$

the formula (2) is called the *Euler product* for $\zeta(s)$. Euler did not actually impose the convergence condition: the rigorous treatment of limits and convergence was not yet available, and Euler either handled such issues intuitively or ignored them. If s is a real number — the only case that concerned Euler — then it is well known that $\sum_{n=1}^{\infty} n^{-s}$ converges if and only if $s > 1$, by comparison with $\int_1^{\infty} x^{-s}\, dx$ (that is, by the "Integral Test" of elementary calculus). We shall use

complex s as well, but the criterion for absolute convergence is still easy: if s has real part σ then

$$|n^{-s}| = |\exp(-s\log n)| = \exp(\mathrm{Re}(-s\log n)) = \exp(-\sigma\log n) = n^{-\sigma},$$

so the Euler product holds in the half-plane $\sigma > 1$.

Euler's next step was to set $s = 1$ in (2). This equates $\prod_p 1/(1-p^{-1})$ with the sum $\sum_{n=1}^{\infty} 1/n$ of the harmonic series. Since the sum diverges to $+\infty$, whereas each factor $\prod_p 1/(1-p^{-1})$ is finite, there are infinitely many factors. Therefore, there are infinitely many primes. This proof does not meet modern standards of rigor, but it is easy enough to fix: instead of setting s equal 1, let s approach 1 from above. The next result is an easy estimate on the behavior of $\zeta(s)$ for s near 1.

Lemma. *The inequalities*

$$\frac{1}{s-1} < \zeta(s) < \frac{1}{s-1} + 1 \tag{3}$$

hold for all $s > 1$.

Proof: For all $n > 0$ we have

$$\int_n^{n+1} x^{-s}\,dx = \frac{1}{s-1}\big(n^{1-s} - (n+1)^{1-s}\big),$$

whence

$$(n+1)^{-s} < \frac{n^{1-s} - (n+1)^{1-s}}{s-1} < n^{-s}.$$

Now sum over $n = 1, 2, 3, \ldots$. The sum of $(n^{1-s} - (n+1)^{1-s})/(s-1)$ telescopes to $1/(s-1)$. This sum is bounded above by $\sum_{n=1}^{\infty} n^{-s} = \zeta(s)$, and below by $\sum_{n=1}^{\infty}(n+1)^{-s} = \zeta(s) - 1$. This proves the inequalities (3). □

In fact one can obtain more accurate estimates are available using the "Euler-Maclaurin formula", but we do not yet need them. The lower bound in (3) shows that $\zeta(s) \to \infty$ as $s \to 1$ from above. Since each factor $(1-p^{-s})^{-1}$ in the Euler product remains bounded, we have vindicated Euler's argument for the infinitude of primes.

The divergence of $\prod_p p/(p-1)$, and the behavior of $\prod_p 1/(1-p^{-s})$ as $s \to 1+$, gives us much more specific information on the distribution of primes than we could hope to extract from Euclid's argument. For instance, we cannot have constants C, θ with $\theta < 1$ such that $\pi(x) < Cx^{\theta}$ for all x, because then the Euler product would converge for $s > \theta$. To go further along these lines it is convenient to use the logarithm of the Euler product:

$$\log\zeta(s) = \sum_p -\log(1-p^{-s}). \tag{4}$$

Euler again took $s = 1$ and concluded that $\sum_p 1/p$ diverges. Again we justify his conclusion by letting s approach 1 from above:

Theorem. *For any $s_0 > 1$ there exists M such that*

$$\Big|\sum_p p^{-s} - \log\frac{1}{s-1}\Big| < M \tag{5}$$

for all $s \in (1, s_0]$. In particular, $\sum_p 1/p$ diverges.

Proof: By our Lemma, $\log\zeta(s)$ is between $\log 1/(s-1)$ and $\log s/(s-1)$. Since $0 < \log s < s-1$, we conclude that $\zeta(s)$ differs from $\log 1/(1-s)$ by less than $s-1 < s_0 - 1$. In the right-hand side of (4), we approximate each summand $-\log(1-p^{-s})$ by p^{-s}. The error is at most p^{-2s}, so

$$\Big|\sum_p p^{-s} - \sum_p(-\log(1-p^{-s}))\Big| < \sum_p p^{-2s} < \zeta(2).$$

Hence (5) holds with $M = s_0 - 1 + \zeta(2)$. Letting $s \to 1$ we obtain the divergence of $\sum_p 1/p$. $\square$

Interlude on the "Big Oh" notation $O(\cdot)$. The point of (5) is that $\sum_p p^{-s}$ equals $\log\frac{1}{s-1}$ within a bounded error, not the specific upper bound M on this error — which is why we were content with a bound $s_0 - 1 + \zeta(2)$ weaker than what the method can give. In such approximate formulas we will usually be interested only in the existence of constants such as M, not in their exact values. To avoid distractions such as "$s_0 - 1 + \zeta(2)$", we henceforth use "big Oh" notation. In this notation, (5) appears as

$$\sum_p p^{-s} = \log\frac{1}{s-1} + O(1). \tag{6}$$

In general, $f = O(g)$ means that f, g are functions on some space S with g nonnegative, and there exists a constant M such that $|f(z)| \le Mg(z)$ for all $z \in S$. Thus $O(1)$ is a bounded function, so (6) is indeed equivalent to (5). so (E') means that there exists a constant C such that An equivalent notation, more convenient in some circumstances, is $f \ll g$ (or $g \gg f$). For instance, a linear map T between Banach spaces is continuous iff $Tv = O(|v|)$ iff $|v| \gg |Tv|$. Each instance of $O(\cdot)$ or $\ll$ or $\gg$ is presumed to carry its own implicit constant M. If the constant depends on some parameter(s), we may use the parameter(s) as a subscript to the "O" or "$\ll$". For instance, we may write $O_{s_0}(1)$ instead of $O(1)$ in (6); for any $\epsilon > 0$, we have $\log x \ll_\epsilon x^\epsilon$ on $x \in [1, \infty)$. For basic properties of $O(\cdot)$ and $\ll$ see the Exercises at the end of this section.

Back to $\pi(x)$. The estimate (6) for $\sum_p p^{-s}$ does not explicitly involve $\pi(x)$. We thus rearrange this sum as follows. Write p^{-s} as an integral $s\int_p^\infty y^{-1-s}\,dy$, and sum over p. Then y occurs in the interval of integration $[p, \infty)$ iff $p < y$, that is, with multiplicity $\pi(y)$. Therefore

$$\sum_p p^{-s} = s\int_1^\infty \pi(y) y^{-1-s}\,dy, \tag{7}$$

and (6) becomes an estimate for an integral involving $\pi(\cdot)$.

This transformation from the sum in (6) to the integral (7) is an example of a method we shall use often, known either as partial summation or integration by parts. To explain the latter name, consider that the sum may be regarded as the Stieltjes integral $\int_1^\infty y^{-s}\,d\pi(y)$, which integrated by parts yields (7); that is how we shall write this transformation from now on.

Our estimate (6) on the integral (7) does not suffice to prove the Prime Number Theorem, but does prove support for it: the estimate holds if we replace $\pi(x)$ with $x/\log x$. That is,[1]

$$\int_2^\infty \frac{y^{-s}}{\log y}\,dy = \log\frac{1}{s-1} + O(1) \qquad (1 < s \le 2).$$

To prove this, let $I(s) = \int_2^\infty \frac{y^{-s}}{\log y}\,dy$, and differentiate under the integral sign to obtain $I'(s) = -\int_2^\infty y^{-s}\,dy = 2^{1-s}/(1-s) = 1/(1-s) + O(1)$. Thus for $1 < s \le 2$ we have

$$I(s) = I(2) - \int_s^2 I'(\sigma)\,d\sigma = +\int_s^2 \frac{d\sigma}{\sigma - 1} + O(1) = \log\frac{1}{s-1} + O(1)$$

as claimed. While this does not prove the Prime Number Theorem, it does show that, for instance, if $c < 1 < C$ then there are arbitrarily large x, x' such that $\pi(x) > cx/\log x$ and $\pi(x') < Cx'/\log x'$.

Remarks

Euler's result $\sum_p 1/p = \infty$ underlies for our expectation that p_{n+1} divides $1 + \prod_{i=1}^n p_n$ infinitely often. The residue of $\prod_{i=1}^n p_n \bmod p_{n+1}$ should behave like a random element of $(\mathbf{Z}/p_{n+1}\mathbf{Z})^*$, and thus should equal -1 with probability $1/(p-1)$. The expected value of the number of $n < N$ such that p_{n+1} divides $1 + \prod_{i=1}^n p_n$ is thus $\sum_{n=2}^N 1/(p-1) > \sum_{n=2}^N 1/p \to \infty$ as $N \to \infty$. We expect the same behavior for many other problems of the form "how many primes p are factors of $f(p)$?", notably $f(p) = ((p-1)! + 1)/p$ (the Wilson quotient), $f(p) = (a^p - a)/p$ (the Fermat quotient with fixed base $a > 1$), and $f(p) = p^{-2}\sum_{i=1}^{p-1} 1/i$ (the Wolstenholme quotient). We shall soon see that $\sum_p 1/p$ diverges very slowly: $\sum_{p<x} 1/p = \log\log x + O(1)$. Therefore, while we expect infinitely many solutions of $p|f(p)$ in each case, we expect that these solutions will be very scarce.

Euler's work on the zeta function includes also its evaluation at positive integers: $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, "etc." The silliest proof I know of the infinitude

[1] We shift the lower limit of integration to $y = 2$ to avoid the spurious singularity of $1/\log y$ at $y = 1$, and suppress the factor s because only the behavior as $s \to 1$ matters and multiplying by s does not affect it to within $O(1)$. We also made the traditional and convenient choice $s_0 = 2$; the value of s_0 does not matter, as long as $s_0 > 1$, because we are concerned with the behavior near $s = 1$, and by specifying s_0 we can dispense with a distracting subscript in O_{s_0}.

of primes is to pick one such integer s, and observe that if there were finitely many primes then $\zeta(s) = \prod_p (1-p^{-s})^{-1}$ would be rational, and thus so would π^s, contradicting Lindemann's theorem (1882) that π is transcendental. It is only a bit less silly to take $s = 2$ and use the irrationality of π^2, which though unknown to Euler was proved a few generations later by Legendre [Leg]. This can actually be used to obtain lower bounds on $\pi(x)$, but even with modern "irrationality measures" we can obtain no lower bounds on $\pi(x)$ better than the $\log\log x$ bound already available from Euclid's proof.

Less frivolously, we note that the integral $\int_1^\infty \pi(y) y^{-s}\, dy/y$ appearing in (7) is the *Mellin transform* of $\pi(y)$, evaluated at s. The Mellin transform may not be as familiar as the integral transforms of Fourier and Laplace, but the change of variable $y = e^u$ yields

$$\int_1^\infty \pi(y) y^{-s} \frac{dy}{y} = \int_0^\infty \pi(e^u) e^{-su}\, du,$$

which identifies the Mellin transform of $\pi(y)$ with the Laplace transform of $\pi(e^u)$. In general, if $f(u)$ is a nonnegative function whose Laplace transform $\mathcal{L}f(s) := \int_0^\infty f(u) e^{-su}\, du$ converges for $s > s_0$, then the behavior of $\mathcal{L}f(s)$ as $s \to s_0+$ detects the behavior of $f(u)$ as $u \to \infty$. In our case, $s_0 = 1$, so we expect that our estimate on $\int_1^\infty \pi(y) y^{-s}\, dy/y$ for s near 1 will give us information on the behavior of $\pi(x)$ for large x. Moreover, inverting the Laplace transform requires a contour integral over complex s; this suggests that we will need to consider $\log\zeta(s)$, and thus the solutions of $\zeta(s) = 0$, in the complex plane. We'll return to these ideas and the Mellin transform before long.

Exercises

Concerning the Big Oh (a.k.a. $\ll$) notation:

1. If $f \ll g$ and $g \ll h$ then $f \ll h$. If $f_1 = O(g_1)$ and $f_2 = O(g_2)$ then $f_1 f_2 = O(g_1 g_2)$ and $f_1 + f_2 = O(g_1 + g_2) = O(\max(g_1, g_2))$. Given a positive function g, the functions f such that $f = O(g)$ constitute a vector space.

2. If $f \ll g$ on $[a, b]$ then $\int_a^x f(y)\, dy \ll \int_a^x g(y)\, dy$ for $x \in [a, b]$. (We already used this to obtain $I(s) = \log(1/(s-1)) + O(1)$ from $I'(s) = 1/(1-s) + O(1)$.) In general differentiation does not commute with "$\ll$" (why?). Nevertheless, prove that $\zeta'(s) [= -\sum_{n=1}^\infty n^{-s} \log n]$ is $-1/(s-1)^2 + O(1)$ on $s \in (1, \infty)$.

3. So far all the implicit constants in the $O(\cdot)$ or $\ll$ we have seen are *effective*: we didn't bother to specify them, but we could if we really had to. Moreover the transformations in exercises 1,2 preserve effectivity: if the input constants are effective then so are the output ones. However, it can happen that we know that $f = O(g)$ without being able to name a constant C such that $|f| \le Cg$. Here is a prototypical example. Suppose $x_1, x_2, x_3, \ldots$ is a sequence of positive reals which we suspect are all ≤ 1, but all we can show is that if $i \ne j$ then $x_i x_j < x_i + x_j$. Prove that x_i are bounded, i.e., $x_i = O(1)$, but that as long as we do not find some x_i greater than 1, we cannot use this to exhibit a specific C such that $x_i < C$ for all i — and indeed if our suspicion that every $x_i \le 1$ is

correct then we'll never be able to find C.

We'll encounter this sort of unpleasant ineffectivity (where it takes at least two outliers to get a contradiction) in Siegel's lower bound on $L(1,\chi)$; it arises elsewhere too, notably in Faltings' proof of the Mordell conjecture, where the number of rational points on a given curve of genus > 1 can be effectively bounded but their size cannot.

Applications of the Euler product for $\zeta(s)$:

4. Complete the proof that for each $c < 1$ there are arbitrarily large x such that $\pi(x) > cx/\log x$ and for each $C > 1$ there are arbitrarily large x' such that $\pi(x') < Cx'/\log x'$.

5. It is known that there exists a constant M such that $|\pi^2 - a/b| \gg 1/b^M$ for all positive integers a, b. Use this together with the Euler product for $\zeta(2)$ to prove that $\pi(x) \gg \log\log x$.

6. Prove that there are $N/\zeta(2) + O(N^{1/2})$ squarefree integers in $[1, N]$. Obtain similar estimates for the number of natural numbers $< N$ not divisible by n^s for any $n > 1$ ($s = 3, 4, 5, \ldots$).

It follows that an integer chosen uniformly at random from $[1, N]$ is squarefree with probability approaching $1/\zeta(2) = 6/\pi^2$ as $N\to\infty$. Informally, "a random integer is squarefree with probability $6/\pi^2$". We shall see that the error estimate $O(N^{1/2})$ can be improved, and that the asymptotic growth of the error hinges on the Riemann Hypothesis.

7. Prove that there are $N^2/\zeta(2) + O(N\log N)$ ordered pairs of relatively prime integers in $[1, N]$. What of relatively prime pairs (x_1, x_2) with $x_1 < N_1$ and $x_2 < N_2$? Generalize.

Again we may informally deduce that two random integers are coprime with probability $6/\pi^2$. Alternatively, we may regard a coprime pair (x_1, x_2) with $x_i \le N$ as a positive rational number x_1/x_2 of height at most N. Dropping the positivity requirement, we find that there are asymptotically $2N^2/\zeta(2)$ rational numbers of height at most N. This has been generalized to number fields other than $\mathbf{Q}$ by Schanuel [1979]; a function-field analogue, concerning rational functions of bounded degree on a given algebraic curve over a finite field, was announced by Serre [1989, p.19] and proved by DiPippo [1990] and Wan [1992] (independently but in the same way). The function-field result was the starting point of our estimate on the size of the nonlinear linear codes obtained from rational functions on modular curves [Elkies 2001]. Schanuel also obtained asymptotics for rational points of height at most N in projective space of dimension $s - 1$ over a number field K; when $K = \mathbf{Q}$ this recovers the asymptotic enumeration of coprime s-tuples of integers.

8. Prove that as $N\to\infty$ the number of ordered quadruples (a, b, c, d) of integers in $[1, N]$ such that $\gcd(a, b) = \gcd(c, d)$ is asymptotic to $2N^4/5$.

Can this be proved without invoking the values of $\zeta(2)$ or $\zeta(4)$? This can be regarded as a form of a question attributed to Wagstaff in [Guy 1981, B48]: "Wagstaff asked

for an elementary proof (e.g., without using properties of the Riemann zeta-function) that $\prod_p (p^2+1)/(p^2-1) = 5/2$."

References

[DiPippo 1990] DiPippo, S.A.: *Spaces of Rational Functions on Curves Over Finite Fields.* Ph.D. Thesis, Harvard, 1990.

[Elkies 2001] Elkies, N.D.: Excellent nonlinear codes from modular curves, pages 200–208 in *STOC'01: Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, Hersonissos, Crete, Greece.* Isomorphic with `math.NT/0104115` at `arXiv.org`.

[Euler] Euler, L.: ??

[Guy 1981] Guy, R.K.: *Unsolved Problems in Number Theory.* Springer, 1981.

[Legendre] Legendre, A.-M.: *Eléments.*

[Schanuel 1979] Schanuel, S.H.: Heights in number fields. *Bull. Soc. Math. France* **107**, 433–449 (1979).

[Serre 1989] Serre, J.-P.: *Lectures on the Mordell-Weil Theorem* (trans. M. Brown). F. Vieweg & Sohn, Braunschweig 1989.

[Wan 1992] Wan, D.: Heights and Zeta Functions in Function Fields. In *The Arithmetic of Function Fields*, pages 455–463. Berlin: W. de Gruyter, 1992.

Math 259: Introduction to Analytic Number Theory

Primes in arithmetic progressions: Dirichlet characters and L-functions

Dirichlet extended Euler's analysis from $\pi(x)$ to

$$\pi(x, a \bmod q) := \#\{p \le x : p \text{ is a positive prime congruent to } a \bmod q\}.$$

We introduce his approach with the example of the distribution of primes mod 4, that is, of $\pi(x, 1 \bmod 4)$ and $\pi(x, 3 \bmod 4)$. The sum of these is of course $\pi(x)-1$ once $x > 2$, and we have already obtained

$$s \int_1^\infty \pi(y) y^{-1-s}\, dy = \log \frac{1}{s-1} + O_{s_0}(1) \qquad (1 < s \le s_0) \tag{1}$$

from the Euler product for $\zeta(s)$. If we omit the factor $(1-2^{-s})^{-1}$, we obtain a product formula for

$$(1-2^{-s})\zeta(s) = 1 + 3^{-s} + 5^{-s} + 7^{-s} + \cdots .$$

If we try to estimate $\pi(\cdot, 1 \bmod 4)$ (or $\pi(\cdot, 3 \bmod 4)$) in the same way, we are led to the sum of n^{-s} over the integers all of whose prime factors are are congruent to 1 (or 3) mod 4, which is hard to work with. But we can analyze the *difference* $\pi(x, 1 \bmod 4) - \pi(x, 3 \bmod 4)$ using an Euler product for the *L-series*

$$L(s, \chi_4) := 1 - \frac{1}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + - \cdots = \sum_{n=1}^\infty \chi_4(n) n^{-s}.$$

Here χ_4 is the function

$$\chi_4(n) = \begin{cases} +1, & \text{if } n \equiv +1 \bmod 4; \\ -1, & \text{if } n \equiv -1 \bmod 4; \\ 0, & \text{if } 2|n. \end{cases}$$

This function is (strongly[1]) *multiplicative*:

$$\chi_4(mn) = \chi_4(m)\chi_4(n) \qquad (m, n \in \mathbf{Z}). \tag{2}$$

Therefore $L(s, \chi_4)$ factors as did $\zeta(s)$:

$$L(s, \chi_4) = \prod_{p \text{ prime}} \Big(\sum_{c_p=1}^\infty \chi(p^{c_p}) p^{-c_p s} \Big) = \prod_{p \text{ prime}} \frac{1}{1 - \chi(p) p^{-s}}. \tag{3}$$

By comparison with the Euler product for $\zeta(s)$ we see that the manipulations in (3) are valid for $s > 1$ (and in fact for s of real part > 1). Unlike $\zeta(s)$, the function $L(s, \chi_4)$ remains bounded as $s \to 1+$, because the sum $\sum_{n=1}^\infty \chi_4(n) n^{-s}$ may be grouped as

[1] Often a function f is called multiplicative when $f(mn) = f(m)f(n)$ only for coprime m, n; see the Exercises.

$$\Big(1-\frac{1}{3^s}\Big)+\Big(\frac{1}{5^s}-\frac{1}{7^s}\Big)+\Big(\frac{1}{9^s}-\frac{1}{11^s}\Big)+\cdots$$

in which the n-th term is $O(n^{-(s+1)})$ (why?). Indeed this regrouping lets us extend $L(\cdot,\chi_4)$ to a continuous function on $(0,\infty)$. Moreover, each term $(1-3^{-s})$, $(5^{-s}-7^{-s})$, $(9^{-s}-11^{-s})$,... is positive, so $L(s,\chi_4)>0$ for all $s>0$, in particular for $s=1$ (you probably already know that $L(1,\chi_4)=\pi/4$). The same analysis we used to get an estimate on the Mellin transform of $\pi(\cdot)$ from the Euler product for $\zeta(s)$ can now be used starting from (3) to obtain:[2]

$$s\int_1^\infty \pi(y,\chi_4)y^{-1-s}\,dy = O(1) \qquad (1<s\le 2), \tag{4}$$

where

$$\pi(y,\chi_4) := \pi(y,1 \bmod 4)-\pi(y,3 \bmod 4) = \sum_{p\le y}\chi_4(p).$$

Averaging (4) with (1), we find that

$$\begin{aligned} s\int_1^\infty \pi(y,1 \bmod 4)y^{-1-s}\,dy &= \frac{1}{2}\log\frac{1}{s-1}+O(1) \qquad (1<s\le 2),\\ s\int_1^\infty \pi(y,3 \bmod 4)y^{-1-s}\,dy &= \frac{1}{2}\log\frac{1}{s-1}+O(1) \qquad (1<s\le 2). \end{aligned}$$

This is consistent with $\pi(x,\pm 1 \bmod 4)\sim \frac{1}{2}x/\log x$, and corroborates our expectation that there should be on the average as many primes congruent to $+1 \bmod 4$ as $-1 \bmod 4$. Specifically, it shows that for $a=\pm 1$ the sets of primes congruent to $a \bmod 4$ has logarithmic density $1/2$ in the primes. This concept is defined as follows:

Definition. Suppose P is a set of positive integers such that $\sum_{n\in P}1/n$ diverges. A subset S of P is said to have *logarithmic density* δ if

$$\Big(\sum_{n\in S}n^{-s}\Big)\Big/\Big(\sum_{n\in P}n^{-s}\Big)\to\delta$$

as $s\to 1+$. Taking for P the set of primes, we see that a set S of primes has logarithmic density δ if and only if

$$\sum_{p\in S}p^{-s}\sim\delta\log\frac{1}{s-1}$$

as $s\to 1+$.

This notion of "logarithmic density" has the properties we would expect from a density: $\delta\in[0,1]$; a set of positive density is nonempty; if disjoint sets P_1,P_2

[2] Again, the choice of $s_0>1$ does not matter, because we are concerned with the behavior near $s=1$; thus we have made the traditional and convenient choice $s_0=2$, rather than continue with an unspecified s_0 and a distracting subscript in O_{s_0}.

have logarithmic densities δ_1, δ_2, then $P_1 \cup P_2$ has logarithmic density $\delta_1 + \delta_2$; and if P_1, P_2 are sets of logarithmic densities δ_1, δ_2 and $P_1 \subseteq P_2$, then $\delta_1 \leq \delta_2$. See the first Exercise for further information.

We can use the notion of logarithmic density to state Dirichlet's theorem as follows:

Theorem [Dirichlet]**.** *For any positive integer q, and any integer a coprime to q, the primes congruent to $a \bmod q$ constitute a set of logarithmic density $1/\varphi(q)$ in the primes.*

Here φ is the Euler phi ("totient") function, $\varphi(q) = |(\mathbf{Z}/q)^*|$. We have just proved the cases $(q, a) = (4, \pm 1)$ of Dirichlet's theorem. The same method disposes of $(q, a) = (3, \pm 1)$, using

$$(1 - 3^{-s})\zeta(s) = 1 + 2^{-s} + 4^{-s} + 5^{-s} + 7^{-s} + 8^{-s} + \cdots$$

and

$$L(s, \chi_3) := 1 - \frac{1}{2^s} + \frac{1}{4^s} - \frac{1}{5^s} + - \cdots = \sum_{n=1}^{\infty} \chi_3(n) n^{-s},$$

Where χ_3 is the multiplicative function defined by

$$\chi_3(n) = \begin{cases} +1, & \text{if } n \equiv +1 \bmod 3; \\ -1, & \text{if } n \equiv -1 \bmod 3; \\ 0, & \text{if } 3|n. \end{cases}$$

With a tad more work we can deal with $q = 8$. Let $\chi_8(n)$ be $+1$ if $n \equiv \pm 1 \bmod 8$, -1 if $n \equiv \pm 3 \bmod 8$, and 0 if n is even. This is a multiplicative function, as is $\chi_4\chi_8$; the resulting L-functions

$$\begin{aligned} L(s, \chi_8) &= \sum_{n=1}^{\infty} \chi_8(n) n^{-s} = 1 - \frac{1}{3^s} - \frac{1}{5^s} + \frac{1}{7^s} + - - + \cdots, \\ L(s, \chi_4\chi_8) &= \sum_{n=1}^{\infty} \chi_4\chi_8(n) n^{-s} = 1 + \frac{1}{3^s} - \frac{1}{5^s} - \frac{1}{7^s} + + - - \cdots \end{aligned}$$

have Euler products for $s > 1$ and are positive for $s > 0$ (to prove this for $L(s, \chi_8)$, group the terms in fours rather than pairs and use the convexity of the function $n \mapsto n^{-s}$). We deduce that

$$\sum_p \chi_8(p) p^{-s} = O(1) \quad \text{and} \quad \sum_p \chi_4\chi_8(p) p^{-s} = O(1)$$

for $s \in (1, 2]$, which combined with previous results yields Dirichlet's theorem for $q = 8$. Similarly we can handle $q = 12$, and with some more effort even $q = 24$.

What about $q = 5$? We have the "quadratic character", which takes n to $+1$ or -1 if $x \equiv \pm 1$ or $\pm 2 \bmod 5$ (and to 0 if $5|n$), but this only lets us separate quadratic from non-quadratic residues mod 5. We need a new idea to

get at the individual nonzero residue classes mod 5. (Recall that $\{5k+2\}$ and $\{5k-2\}$ are the first cases of arithmetic progressions that we could not prove contain infinitely many primes using variations of Euclid's proof.) Let χ be the multiplicative function from $\mathbf{Z}$ to the *complex* numbers which takes $n \equiv 0, 1, 2, 3, 4 \bmod 5$ to $0, 1, i, -i, -1$. Another such function is the complex conjugate $\overline{\chi} = \chi^3$, while χ^2 is the quadratic character and χ^4 is the "trivial character" taking n to 0 or 1 according as $5|n$ or not. The resulting L-functions $\sum_n \chi(n)n^{-s}$, $\sum_n \overline{\chi}(n)n^{-s}$ then take complex values, but still have Euler products and extend to continuous functions on $s > 0$. Moreover, these functions never vanish on $s > 0$; indeed their real and imaginary parts are both nonzero, as we see by combining the real terms into $(5k+1, 5k+4)$ pairs and the imaginary terms into $(5k+2, 5k+3)$ pairs. Likewise the L-function associated to the quadratic character χ^2 has an Euler product and is positive for $s > 0$ by convexity of n^{-s}. We conclude as before that $\sum_p \chi^j(p)p^{-s} = O(1)$ as $s \to 1+$ for each $j = 1, 2, 3$, and recover Dirichlet's theorem for $q = 5$ by taking linear combinations of these sums and $\sum_p p^{-s} = \log \frac{1}{s-1} + O(1)$.

For general q, we proceed analogously, using linear combinations of *Dirichlet characters*, whose definition follows.

Definition. For a positive integer q, a Dirichlet character mod q is a function $\chi : \mathbf{Z} \to \mathbf{C}$ which is

- q-periodic: $n \equiv n' \bmod q \Rightarrow \chi(n) = \chi(n')$;
- supported on the integers coprime to q and on no smaller subset of $\mathbf{Z}$: $(n, q) = 1 \Leftrightarrow \chi(n) \neq 0$; and
- multiplicative: $\chi(m)\chi(n) = \chi(mn)$ for all integers m, n.

To such a character is associated the *Dirichlet L-series*

$$L(s, \chi) := \sum_{n=1}^{\infty} \chi(n)n^{-s} = \prod_p \frac{1}{1 - \chi(p)p^{-s}} \qquad (s > 1). \tag{5}$$

Examples: The *trivial character* χ_0 mod q is defined by $\chi(n) = 1$ if $(n, q) = 1$ and $\chi(n) = 0$ otherwise. Its associated L-series is

$$L(s, \chi_0) = \prod_{p|q} (1 - p^{-s}) \cdot \zeta(s). \tag{6}$$

If l is prime then the *Legendre symbol* $(\cdot/l)$, defined by $(n/l) = 0, 1, -1$ according as n is zero, a nonzero square, or not a square mod l, is a character mod l. If χ is a Dirichlet character mod q then so is its complex conjugate $\overline{\chi}$ (defined of course by $\overline{\chi}(n) = \overline{\chi(n)}$), with $L(s, \overline{\chi}) = \overline{L(s, \chi)}$ for $s > 1$. If χ, χ' are characters mod q, q' then $\chi\chi'$ is a character mod $\mathrm{lcm}(q, q')$. In particular, we have:

Lemma: *For each q, the characters mod q constitute a group under pointwise multiplication, with identity χ_0 and inverse $\chi^{-1} = \overline{\chi}$.*

What is this group? A Dirichlet character mod q is just a homomorphism from $(\mathbf{Z}/q)^*$ to the unit circle, extended by zero to a function on $\mathbf{Z}/q$ and lifted to $\mathbf{Z}$. Therefore the group of such characters is the *Pontrjagin dual* of $(\mathbf{Z}/q)^*$. Pontrjagin duality for *finite* abelian groups like $(\mathbf{Z}/q)^*$ is easy, since it is equivalent to the theory of the discrete Fourier transform. We next recall the basic facts.

For any finite abelian group G, let $\hat{G}$ be its Pontrjagin dual, defined as the group of homomorphisms from G to the unit circle in $\mathbf{C}$. Then the dual of $G \times H$ is $\hat{G} \times \hat{H}$, and the dual of $\mathbf{Z}/m$ is a cyclic group of order m. Since any finite abelian group is a product of cyclic groups, it follows that $\hat{G}$ is isomorphic with G. This isomorphism is not in general canonical,[3] but there is a canonical isomorphism from G to the dual of $\hat{G}$ (the second dual of G), namely the map taking an $g \in G$ to the homomorphism $\chi \mapsto \chi(g)$. That this is an isomorphism can be checked directly for cyclic groups, and then deduced for any finite abelian G because all such G are direct sums of abelian groups.

The characters of G are *orthogonal*:

$$\sum_{g \in G} \chi_1(g)\,\overline{\chi_2(g)} = \begin{cases} |G|, & \text{if } \chi_1 = \chi_2; \\ 0, & \text{if } \chi_1 \neq \chi_2. \end{cases}$$

In particular, they are linearly independent; since there are $|G|$ of them, they form a basis for the vector space of complex-valued functions on G. The decomposition of an arbitrary such function $f : G \to \mathbf{C}$ is a linear combination of characters is achieved by the *inverse Fourier transform*:

$$f = \sum_{\chi \in \hat{G}} f_\chi \chi, \quad \text{where} \quad f_\chi := \frac{1}{|G|} \sum_{g \in G} \overline{\chi}(g) f(g).$$

In particular, the characteristic function of any $g_0 \in G$ is $|G|^{-1} \sum_\chi \overline{\chi}(g_0)\chi$.

What does all this tell us about Dirichlet L-functions and distribution of primes mod q? First, that if we define $\pi(\cdot, \chi)$ by

$$\pi(x, \chi) := \sum_{a \bmod q} \chi(a)\pi(x, a \bmod q) = \sum_{p < x} \chi(p)$$

then, for all a coprime to q,

$$\pi(x, a \bmod q) = \frac{1}{\varphi(q)} \sum_{\chi \bmod q} \overline{\chi}(a)\pi(x, \chi).$$

Second, that

$$s \int_1^\infty \pi(y, \chi) y^{-1-s}\, dy = \sum_p \chi(p) p^{-s} = \log L(s, \chi) + O(1) \tag{7}$$

[3] For instance, if G is cyclic of order 5, there can be no canonical nondegenerate pairing $\langle \cdot, \cdot \rangle : G \times G \to \mathbf{C}^*$, because such a pairing would have to be invariant under $\mathrm{Aut}(G) = (\mathbf{Z}/5)^*$, but $\langle g^2, g^2 \rangle = \langle g, g \rangle^4 \neq \langle g, g \rangle$.

For $1 < s \le 2$. This is again obtained by taking logarithms in the Euler product (5). The Euler product shows that $L(s,\chi) \ne 0$ for $s > 1$; if χ is complex, "$\log L(s,\chi)$" means the branch of the logarithm that approaches 0 as $s \to \infty$.

For the behavior of $L(s,\chi)$ near $s = 1$, we have:

Lemma. *i) If $\chi = \chi_0$ then $\log L(s,\chi) = \log(1/(s-1)) + O(1)$ as $s \to 1+$.*
ii) For nontrivial χ, the sum defining $L(s,\chi)$ converges for $s > 0$ and defines a continuous function on the positive reals.

Proof: (i) follows from (6), together with our estimate on $\zeta(s)$ for $s \to 1+$. As for (ii), as a special case of character orthogonality we have $\sum_{a \bmod q} \chi(a) = 0$, so $S_\chi(x) := \sum_{0<n<x} \chi(n)$ is a bounded function of x. Hence (for large $M, N \notin \mathbf{Z}$)[4]

$$\sum_{M<n<N} \frac{\chi(n)}{n^s} = \int_M^N y^{-s}\, dS_\chi(y) = S_\chi(y) y^{-s}\Big|_M^N + s\int_M^N y^{-1-s} S_\chi(y)\, dy$$

$$\ll_\chi M^{-s} + N^{-s},$$

which for fixed $s > 0$ tends to zero as $M, N \to \infty$. Thus the sum $\sum_{n=1}^\infty \chi(n) n^{-s}$ converges. Moreover, for any $s_0 > 0$, the convergence is uniform in $s \ge s_0$. Hence $\sum_{n=1}^\infty \chi(n) n^{-s}$ is the uniform limit of continuous functions $\sum_{n=1}^N \chi(n) n^{-s}$, and is therefore a continuous function on $(0,\infty)$, as claimed. $\square$

From (7) we see that the crucial question is whether $L(1,\chi)$ is nonzero: the right-hand side is $O(1)$ if $L(1,\chi) \ne 0$ but $\le -\log(1/(s-1)) + O(1)$ if $L(1,\chi) = 0$ (since $L(s,\chi)$ is differentiable at $s = 1$). Our experience with small q, and our expectation that the primes should not favor one congruence class in $(\mathbf{Z}/q)^*$ to another, both suggest that $L(1,\chi)$ will not vanish. This is true, and can be checked in any given case by a finite computation; but our methods thus far do not let us prove it in general (try doing it for $\chi = (\cdot/67)$ or $(\cdot/163)$!). For the time being, then, we can only obtain a conditional result:

Proposition. *Assume that $L(1,\chi) \ne 0$ for all nontrivial characters $\chi \bmod q$. Then Dirichlet's theorem holds for all arithmetic progressions mod q.*

Proof: For each $a \in (\mathbf{Z}/q)^*$, multiply (7) by $\overline{\chi}(a)$, and average over χ to obtain

$$\sum_{p \equiv a \bmod q} p^{-s} = \frac{1}{\varphi(q)} \sum_\chi \overline{\chi}(a) \log L(s,\chi) + O(1) = \frac{1}{\varphi(q)} \log \frac{1}{s-1} + O(1)$$

for $1 < s \le 2$, since χ_0 contributes $\chi_0(a) \log \zeta(s) + O(1) = \log \frac{1}{s-1} + O(1)$ to the sum, while the other terms remain bounded by hypothesis. Thus the primes congruent to $a \bmod q$ have logarithmic density $1/\varphi(q)$, as claimed.

In fact the nonvanishing of $L(1,\chi)$ was proved by Dirichlet, who thus established his celebrated theorem on primes in arithmetic progressions. At least three

[4] We require that $M, N \notin \mathbf{Z}$ to avoid the distraction of whether the Riemann-Stieltjes integral $\int_M^N y^{-s}\, dS_\chi(y)$ includes the terms with $n = M$ or $n = N$ in the sum.

proofs are now known. These three proofs all start with the product of the L-functions associated to all $\varphi(q)$ Dirichlet characters mod q:

$$\prod_{\chi \bmod q} L(s,\chi) = \prod_p \left(\prod_{\chi \bmod q} (1-\chi(p)p^{-s}) \right)^{-1}.$$

The inner product can be evaluated with the following cyclotomic identity:

Let G be a finite abelian group and $g \in G$ an element of order m. Then

$$\prod_{\chi \in \hat{G}} (1-\chi(g)z) = (1-z^m)^{|G|/m} \tag{8}$$

hold identically for all z.

The identity is an easy consequence of the factorization of $1-z^m$ together with the fact that any character of a subgroup $H \subseteq G$ extends in $[G:H]$ ways to a character of G (in our case H will be the cyclic subgroup generated by g).

Let m_p, then, be the multiplicative order of p mod q (for all but the finitely many primes p dividing q). Then we get

$$\prod_{\chi \bmod q} L(s,\chi) = \prod_{p \nmid q} (1-p^{-m_p s})^{-\varphi(q)/m_p}. \tag{9}$$

The left-hand side contains the factor $L(s,\chi_0)$, which is $C/(s-1)+O(1)$ as $s \to 1+$ for some $C > 0$ [in fact $C = \varphi(q)/q$]. Since the remaining factors are differentiable at $s=1$, if any of them were to vanish there the product would remain bounded as $s \to 1+$. So we must show that this cannot happen.

Dirichlet's original approach was to observe that (9) is, up to a few factors $1-n^{-s}$ with $n|q$, the "zeta function of the cyclotomic number field $\mathbf{Q}(e^{2\pi i/q})$". He then proved that the zeta function $\zeta_K(s)$ of *any* number field K is $\sim C/(s-1)$ as $s \to 1+$ for some positive constant C (and gave an exact formula for C, which includes the class number of K and is thus called the "Dirichlet class number formula"). That is undoubtedly the best way to go about it — but it requires more algebraic number theory than I want to assume here. Fortunately there are at least two ad-hoc simplifications available.

The first is that we need only worry about real characters. If $L(1,\chi) = 0$ then also $L(1,\overline{\chi}) = 0$. Hence if $\chi \neq \overline{\chi}$ but $L(1,\chi) = 0$ then there are at least *two* factors in the left-hand side of (9) that vanish at $s=1$; since they are differentiable there, the product would be not only bounded as $s \to 1+$, but approach zero there — which is impossible because the right-hand side is > 1 for all $s > 1$. But if χ is a real character then $L(s,\chi_0)L(s,\chi)$ is (again within a few factors $1-n^{-s}$ of) the L-function of a quadratic number field. Developing the algebraic number theory of quadratic number fields takes considerably less work than is needed for the full Dirichlet class number formula, and if we only want to get unboundedness as $s \to 1+$ it is even easier — for instance, if $\chi(-1) = -1$

then the right-hand side of (9) is dominated by the zeta function of a binary quadratic form, which is easily seen to be $\gg 1/(s-1)$. However, even this easier proof is beyond the scope of what I want to assume or fully develop in this class.

Fortunately there is a way to circumvent any ζ_K beyond $K = \mathbf{Q}$, using the fact that the right-hand side of (9) also dominates the series $\zeta(\varphi(q) \cdot s)$, which diverges not at $s = 1$ but at $s = 1/\varphi(q)$. Since this s is still positive, we can still get a proof of $L(1, \chi) \neq 0$ from it, but only by appealing to the magic of complex analysis. We thus defer the proof until we have considered $\zeta(s)$ and more generally $L(s, \chi)$ as functions of a *complex* variable s, which we shall have to do anyway to obtain the Prime Number Theorem and results on the density (not just logarithmic density) of primes in arithmetic progressions.

Remarks

Let K be any number field (finite algebraic extension of $\mathbf{Q}$), and O_K its ring of algebraic integers. The "zeta function" $\zeta_K(s)$ is $\sum_I |I|^{-s}$, where I ranges over nonzero ideals of O_K and $|I| = [O_K : I]$ is the norm of I. For instance, $\zeta(s) = \zeta_{\mathbf{Q}}(s)$, and if $K = \mathbf{Q}[i]$ then $\zeta_K(s) = \frac{1}{4}\sum(m^2+n^2)^{-s}$, the sum extending over all $(m, n) \in \mathbf{Z}^2$ other than $(0, 0)$. The relation between the product (9) and the zeta function of $\mathbf{Q}(e^{2\pi i/q})$ can be made more precise: if we replace each χ by its underlying primitive character (see the Exercises), the product is exactly the zeta function of that cyclotomic number field. Similarly, for any quadratic field K there is a primitive Dirichlet character χ such that $\zeta_K(s) = \zeta(s)L(s, \chi)$. These are the prototypical examples of the factorization of a zeta function as a product of Artin L-functions; the fact that the "Artin L-functions" for 1-dimensional representations are Dirichlet series is a prototype for class field theory. Dirichlet's theorem in turn generalizes to the Čebotarev density theorem. These theorems all require more algebraic machinery than the results we shall obtain using only the Riemann zeta and Dirichlet L-functions, but much the same analytic methods. Therefore we shall not develop them further in Math 259.

Exercises

Concerning density:

1. If P is an infinite set of integers, the *(natural) density* of any subset $S \subseteq P$ is

$$\lim_{x\to\infty} \#\{n \in S : n < x\}/\#\{n \in P : n < x\},$$

if the limit exists. Check that this satisfies the same properties we noted for the logarithmic density (density of subsets, disjoint unions, etc.). Show that if $\sum_{n\in P} 1/n$ diverges and $S \subset P$ has density δ in S then it also has logarithmic density δ in S. (Use partial summation to write $\sum_{n\in P} n^{-s}$ as an integral involving $\#\{n \in S : n < x\}$.) If P is the set of natural numbers and S_d $(d = 1, 2, \ldots, 9)$ is the subset consisting of integers whose first decimal digit is d, show that S_d has logarithmic density $\log_{10}(1 + \frac{1}{d})$ in P but no natural density. Does every set of natural numbers have a logarithmic density?

While not every set with a logarithmic density has a natural density, we shall see that the primes congruent to a mod q do have natural density $1/\varphi(q)$ in the primes. As for the sets S_d, their logarithmic densities account for "Benford's Law", the observation that in many naturally occurring "random numbers" the initial digit d occurs with frequency $\log_{10}(1+\frac{1}{d})$, rather than $1/9$ as one might expect.

Concerning Euler products:

2. One may associate to any sequence $(a_1, a_2, a_3, \ldots)$ of complex numbers an L-series $L(s) = \sum_{n=1}^{\infty} a_n n^{-s}$, which converges absolutely in some right half-plane $s > s_0$ if $a_n \ll n^{s_0-1}$. Show that $L(s)$ has an Euler product

$$L(s) = \prod_p \left(\sum_{c=0}^{\infty} \frac{a_{p^c}}{p^{cs}} \right)$$

if and only if $a_{mn} = a_m a_n$ for any m, n such that $\gcd(m,n) = 1$. (Such functions $n \mapsto a_n$ are called "multiplicative". Note that necessarily $a_1 = 1$ if $\{a_n\}$ is multiplicative.)

3. Let $f(s)$ be the sum of n^{-s} over squarefree positive integers n. Express $f(s)$ in terms of the zeta function, and evaluate $f(2)$. What are the a_n such that $1/\zeta(s) = \sum_{n=1}^{\infty} a_n n^{-s}$? Given k, what is the coefficient of n^{-s} in $\zeta(s-k)$, or $\zeta(s)\zeta(s-k)$?

4. Find $a_1, a_2, \ldots$ such that $\sum_p p^{-s} = \sum_{n=1}^{\infty} a_n \log \zeta(ns)$ for all $s > 1$. Use this (and a computer package that knows about $\zeta(2n)$ and high-precision arithmetic) to calculate that

$$\sum_p \frac{1}{p^2} = 0.45224742004106549850654336483224793417323\ldots$$

Note that this is much greater accuracy than we could reasonably expect to reach by summing the series directly. We shall see that this trick can be adapted to efficiently compute $\sum_p f(p)$ for many natural choices of f.

Concerning Pontrjagin duality:

5. Show that to any homomorphism $\alpha : H \to G$ between finite abelian groups there is a canonically associated homomorphism $\hat{\alpha} : \hat{G} \to \hat{H}$ in the opposite direction between their Pontrjagin duals. Check that α is the dual of $\hat{\alpha}$ (under the canonical identification of G and H with the duals of $\hat{G}, \hat{H}$), and that if β is a homomorphism from G to a finite abelian group K then the dual of the composite homomorphism $\beta \circ \alpha : H \to K$ is $\hat{\alpha} \circ \hat{\beta}$. Prove that $\mathrm{im}(\alpha) = \ker(\beta)$ if and only if $\mathrm{im}(\hat{\beta}) = \ker(\hat{\alpha})$.

In particular, if $H \to G$ is an injection, it follows (by taking β to be the quotient map $G \to G/\alpha(H)$) that the restriction map $\hat{\alpha} : \hat{G} \to \hat{H}$ is a surjection; this was used to prove the cyclotomic identity (8). An adherent of the categorical imperative would summarize this exercise, together with the easy observations that $\widehat{\mathrm{id}} = \mathrm{id}$ (when $G = H$) and $\hat{0} = 0$, by saying that Pontrjagin duality is an "exact contravariant functor on the category of finite abelian groups".

Concerning Dirichlet characters:

6. Show that the integers q modulo which all the Dirichlet characters are *real* (take on only the values $0, \pm 1$) are precisely 24 and its factors. Show that every real Dirichlet character is of the form $\chi_0 \psi \prod_{l \in S} (\cdot/l)$, where χ_0 is the trivial character, $\psi = \chi_4^{\epsilon_4} \chi_8^{\epsilon_8}$ for some $\epsilon_4, \epsilon_8 \in \{0, 1\}$, and S is a (possibly empty) finite set S of odd primes.

7. Let χ_0 be the trivial character mod q, and let q_1 be some factor of q. For any character $\chi_1 \bmod q_1$ there is a character $\chi \bmod q$ defined by $\chi = \chi_0 \chi_1$. Express $L(s, \chi)$ in terms of $L(s, \chi_1)$. Conclude that $L(1, \chi) \neq 0$ if and only if $L(1, \chi_1) \neq 0$.

8. A character mod q that cannot be obtained in this way from any character mod a *proper* factor $q_1 | q$ (a factor other than q itself) is called *primitive*. Show that any Dirichlet character χ comes from a unique primitive character χ_1. [The modulus of this χ_1 is called the *conductor* of χ.] Show that the number of primitive characters mod n is $n \prod_{p|n} \alpha_p$, where $\alpha_p = ((p-1)/p)^2$ if $p^2 | n$ and $(p-2)/p$ if $p \| n$. NB there are no primitive characters mod n when $2 \| n$.

The notation $p^f \| n$ means that p^f divides n "exactly"; that is, $p^f | n$ but p^{f+1} does not divide n. Equivalently, the p-valuation of n is f.

9. Deduce the fact that for any q there is at most one nontrivial character χ mod q such that $L(1, \chi) = 0$, as a consequence of (7) together with the fact that $\pi(x, a \bmod q) \geq 0$ for all x, a, q. [In the final analysis, this is not much different from our proof using the product of L-series.] Using either this approach or the one based on (9), prove that there is at most one *primitive* Dirichlet character of any modulus whose L-function vanishes at $s = 1$. [Assume there were two, and obtain two different imprimitive characters to the same modulus whose L-functions both vanish at $s = 1$, which we've already shown impossible. We shall encounter this trick again when we come to Siegel's ineffective lower bound on $L(1, \chi)$.]

Concerning L-series:

10. Show that if χ is a nontrivial character then $L(s, \chi)$ is infinitely differentiable on $s \in (0, \infty)$, and its m-th derivative is given by the convergent sum $\sum_{n=1}^{\infty} (-\log n)^m \chi(n) n^{-s}$ $(m = 1, 2, 3, \ldots)$.

11. i) Prove that if s has real part $\sigma > 1$ then $\zeta(2\sigma)/\zeta(\sigma) < |L(s, \chi)| \leq \zeta(\sigma)$ for all Dirichlet characters χ.
ii) Prove that these bounds are sharp by showing that for all $\sigma > 1$ and $\epsilon > 0$ there exist infinitely many χ such that $L(\sigma, \chi) > \zeta(\sigma) - \epsilon$ and infinitely many χ such that $L(\sigma, \chi) < \zeta(2\sigma)/\zeta(\sigma) + \epsilon$.

We shall show that the bounds are also sharp for individual Dirichlet characters: for each χ, σ, ϵ there exist s of real part σ such that $|L(s, \chi)|$ is arbitrarily close to $\zeta(\sigma)$, and s of real part σ such that $|L(s, \chi)|$ is arbitrarily close to $\zeta(2\sigma)/\zeta(\sigma)$.

12. [Zeta function of a quadratic form] For some positive integer r, let Q :

$\mathbf{R}^r \to \mathbf{R}$ be a positive-definite quadratic form. Show that

$$\zeta_Q(s) := \sum_{\substack{n \in \mathbf{Z}^r \\ n \neq 0}} \frac{1}{Q(n)^s}$$

converges absolutely if and only if s has real part $> r/2$, and determine the limit of $(s - (r/2))\zeta_Q(s)$ as $s \to r/2$ from above. [Use partial summation with respect to $\#\{n \in \mathbf{Z}^r : Q(n) \leq x\}$. Check that your answer is consistent with the answer for $r = 1$, when $Q(n) = an^2$ and $\zeta_Q(s) = a^{-s}\zeta(2s)$.] If Q is the standard quadratic form $Q(n) = n_1^2 + n_2^2$, prove that $\zeta_Q(s) = 4\zeta(s)L(s, \chi_4)$. (This is an example of the relation between (9) and the zeta function of a number field. Check that it is consistent with your formula for the growth of $\zeta_Q(s)$ as $s \to r/2$.) Obtain a similar formula for $Q(n) = n_1^2 + n_1 n_2 + n_2^2$. What other Q can you find for which $\zeta_Q(s)$ is proportional to a product of Dirichlet L-functions?

Math 259: Introduction to Analytic Number Theory

Čebyšev (and von Mangoldt and Stirling)

Before investigating $\zeta(s)$ and $L(s,\chi)$ as functions of a complex variable, we give another elementary approach to estimating $\pi(x)$, due to Čebyšev. This method, unlike Euler's, produces upper and lower bounds on $\pi(x)$ that remain within a small constant factor as $x \to \infty$. These bounds $x/\log x \ll \pi(x) \ll x/\log x$ are sufficient for many theoretical and practical applications, which thus do not require the more advanced and subtle techniques that enter into the proof of the Prime Number Theorem. (The bounds are also close enough to let Čebyšev prove "Bertrand's Postulate": every interval $(x, 2x)$ with $x > 1$ contains a prime. See [HW 1996, p.343–4] for Erdös's simplification of Čebyšev's proof; this simplified proof is also on the Web: http://forum.swarthmore.edu/dr.math/problems/kuropatwa.4.3.97.html .) For us Čebyšev's method also has the advantage of introducing the von Mangoldt function and the Stirling approximation to $x!$, both of which will figure prominently in our future analysis.

It is well known[1] that for any prime p and positive integer x the exponent of p in $x!$ (a.k.a. the p-valuation of $x!$) is

$$c_p(x) := \left\lfloor \frac{x}{p} \right\rfloor + \left\lfloor \frac{x}{p^2} \right\rfloor + \left\lfloor \frac{x}{p^3} \right\rfloor + \cdots = \sum_{k=1}^{\infty} \left\lfloor \frac{x}{p^k} \right\rfloor,$$

the sum being finite because eventually $p^k > x$. It was Čebyšev's insight that one could extract information about $\pi(\cdot)$ from the resulting formula

$$x! = \prod_p p^{c_p(x)},$$

or equivalently

$$\log x! = \sum_p c_p(x) \log(p) = \sum_{n=1}^{\infty} \left\lfloor \frac{x}{n} \right\rfloor \Lambda(n), \tag{1}$$

where $\Lambda(n)$ is the *von Mangoldt function*

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

To make use of (1) we need to estimate

$$\log x! = \sum_{n=1}^{x} \log n$$

[1] If only thanks to the perennial problems along the lines of "how many zeros end 2003! ?".

for large x. We do this by in effect applying the first few steps of symmetrized Euler-Maclaurin summation, to find:

Lemma. *There exists a constant C such that*

$$\log x! = (x+\frac{1}{2})\log x - x + C + O(1/x) \tag{2}$$

holds for all positive integers x.

Proof: For any $\mathcal{C}^2$ function f we have (by integrating by parts twice)

$$\int_{-1/2}^{1/2} f(y)\,dy = f(0) + \frac{1}{2}\left[\int_{-1/2}^{0} f''(y)\big(y+\frac{1}{2}\big)^2\,dy + \int_{0}^{1/2} f''(y)\big(y-\frac{1}{2}\big)^2\,dy\right]$$
$$= f(0) + \frac{1}{2}\int_{-1/2}^{1/2} f''(y)\big\|y+\frac{1}{2}\big\|^2\,dy,$$

where $\|z\|$ is the distance from z to the nearest integer. Thus

$$\sum_{k=1}^{N} f(k) = \int_{1/2}^{N+\frac{1}{2}} f(y)\,dy + \frac{1}{2}\int_{\frac{1}{2}}^{N+\frac{1}{2}} f''(y)\big\|y+\frac{1}{2}\big\|^2\,dy.$$

Taking $f(y) = \log(y)$ and $N = x$ we thus have

$$\log x! = (x+\frac{1}{2})\log(x+\frac{1}{2}) + \frac{1}{2}\log 2 - x - \frac{1}{2}\int_{\frac{1}{2}}^{x+\frac{1}{2}} \big\|y+\frac{1}{2}\big\|^2\frac{dy}{y^2}.$$

The integral is

$$-\frac{1}{2}\int_{\frac{1}{2}}^{\infty} \big\|y+\frac{1}{2}\big\|^2\frac{dy}{y^2} \;+\; O(1/x),$$

and the other terms are

$$(x+\frac{1}{2})\log x - x + \frac{1}{2}(\log 2 + 1) + O(1/x),$$

from which (2) follows. □

[Stirling also determined the value of C (which turns out to be $\frac{1}{2}\log(2\pi)$, as we shall soon see), and extended (2) to an asymptotic series for $x!/((x/e)^x\sqrt{2\pi x}\,)$ in inverse powers of x. But for our purposes $\log x! = (x+\frac{1}{2})\log x - x + O(1)$ is more than enough. In fact, since for the time being we're really dealing with $\log\lfloor x\rfloor!$ and not $\log\Gamma(x+1)$, the best honest error term we can use is $O(\log x)$.]

Now let

$$\psi(x) := \sum_{1\le n\le x} \Lambda(n).$$

Then from (1) and (2) we have

$$\sum_{k=1}^{\infty} \psi(x/k) = (x+\frac{1}{2})\log x - x + O(1).$$

This certainly suggests that $\psi(x) \sim x$, and lets us prove upper and lower bounds on $\psi(x)$ proportional to x. For instance, since $x \geq 1 + \sum_{m=1}^{\infty} \lfloor x/2^m \rfloor$ for all $x \geq 1$, we have

$$\psi(x) \leq \log x! - \sum_{m=1}^{\infty} \log \left\lfloor \frac{x}{2^m} \right\rfloor !,$$

which yields

$$\psi(x) \leq \left[\sum_{m=1}^{\infty} \frac{m}{2^m} \log 2 \right] x + O(\log^2 x) = (2 \log 2)x + O(\log^2 x).$$

For a lower bound we can use the inequality

$$\psi(x) \geq \sum_{k=1}^{\infty} (-1)^{k-1} \psi(x/k) = \log \frac{x!}{(x/2)!^2} = (\log 2)x + O(\log x)$$

for an even integer $x = 2n$; This is essentially the same tactic of factoring $\binom{2n}{n}$ that Čebyšev used to prove $\pi(2x) > \pi(x)$.

It is true that we're ultimately interested in $\pi(x)$, not $\psi(x)$. But it is easy to get from one to the other. For one thing, the contribution to $\psi(x)$ of prime powers p^k with $k > 1$ is negligible — certainly less than $\sum_{k=2}^{\log_2 x} \lfloor x^{1/k} \rfloor \log x \ll x^{1/2} \log x$. The remaining sum, $\sum_{p \leq x} \log p$, can be expressed in terms of $\pi(x)$ and vice versa using partial summation, and we find:

$$\psi(x) = \log(x)\pi(x) - \int_2^x \pi(y) \frac{dy}{y} + O(x^{1/2} \log x),$$

$$\pi(x) = \frac{\psi(x)}{\log x} + \int_2^x \psi(y) \frac{dy}{y \log^2 y} + O(x^{1/2}).$$

It follows that the Prime Number Theorem $\pi(x) \sim x/\log x$ holds if and only if $\psi(x) \sim x$, and good error terms on one side imply good error terms on the other. It turns out that we can more readily get at $\psi(x)$ than at $\pi(x)$; for instance, $\psi(x)$ is quite well approximated by x, while the "right" estimate for $\pi(x)$ is not $x/\log x$ but $(x/\log x) + \int^x dy/\log^2 y$, i.e., the "logarithmic integral" $\int^x dy/\log y$. It is in the form $\psi(x) \sim x$ that we'll actually prove the Prime Number Theorem.

Exercises

On Čebyšev's method:

1. How many consecutive 0's are there at the end of the base-12 expansion of 2006!? Why did I choose 12 rather than any smaller base (including the default 10), and what other bases less than 100 would serve the same purpose?

2. Since our upper and lower asymptotic bounds $\log 2, \log 4$ on $\psi(x)/x$ are within a factor of 2 of each other, they do not quite suffice to prove Bertrand's Postulate. But any improvement *would* prove that $\pi(2x) > \pi(x)$ for sufficiently

large x, from which the proof for all x follows by exhibiting a few suitably spaced primes. It turns out that better bounds are available starting from (1). For instance, show that $\psi(x) < (\frac{1}{2}\log 12)x + O(\log^2 x)$. Can you obtain Čebyšev's bounds of 0.9 and 1.1? In fact it is known that the upper and lower bounds can be brought arbitrarily close to 1, but alas the only known proof of that fact depends on the Prime Number Theorem!

To recover Bertrand's Postulate, one needs for once to convert all the $O(\cdot)$'s to explicit error estimates. One then obtains an explicit x_0 such that $\pi(2x) > \pi(x)$ for all $x \geq x_0$, which reduces Bertrand's Postulate to the finite computation of verifying $\pi(2x) > \pi(x)$ for each $x \in (1, x_0)$. This can be done by calculating a sequence of $O(\log x_0)$ primes $2, 3, 5, 7, 13, 23, \ldots, p$, each less than twice the previous prime, and with $p > x_0$. Once we prove the Prime Number Theorem it will follow that for each $\epsilon > 0$ there exists x_0 such that $\pi((1+\epsilon)x) > \pi(x)$ for all $x \geq x_0$.

3. Estimate $\log \prod (m^2 + n^2)$, where the product extends over all $(m, n) \in \mathbf{Z}^2$ such that $0 < m^2 + n^2 \leq x$. What is the exponent of a prime $p \leq x$ in this product? Using this information, how close can you come to the asymptotic formula $\pi(x, 1 \bmod 4) \sim \frac{1}{2}x/\log x$?

Bernoulli polynomials, Euler-Maclaurin summation, and efficient computation of $\zeta(s)$ and $L(s, \chi)$:

4. The *Bernoulli polynomials* $\mathrm{B}_n(x)$ are defined for $n = 0, 1, 2, 3, \ldots$ by the generating function

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} \mathrm{B}_n(x)\frac{t^n}{n!}.$$

The *Bernoulli numbers* B_n are the rational numbers $\mathrm{B}_n(0)$, with generating function $t/(e^t - 1) = \sum_{n=0}^{\infty} B_n t^n/n!$. The first few Bernoulli polynomials are

$$\mathrm{B}_0(x) = 1, \quad \mathrm{B}_1(x) = x - \frac{1}{2}, \quad \mathrm{B}_2(x) = x^2 - x + \frac{1}{6},$$

$$\mathrm{B}_3(x) = x^3 - \frac{3}{2}x^2 + \frac{1}{2}x, \quad \mathrm{B}_4(x) = x^4 - 2x^3 + x^2 - \frac{1}{30}.$$

Show that in general $\mathrm{B}_n(x) = \sum_{k=0}^{n} \binom{n}{k} B_k x^{n-k}$ (= "$(B+x)^{[n]}$" mnemonically), that $\mathrm{B}_n'(x) = n\mathrm{B}_{n-1}(x)$, and that B_n $(n = 1, 2, 3, \ldots)$ is the unique polynomial such that $\mathrm{B}_n(x+1) - \mathrm{B}_n(x) = nx^{n-1}$ and $\int_0^1 \mathrm{B}_n(x)\,dx = 0$. Show that the Bernoulli number B_n vanishes for odd $n > 1$. What is $\mathrm{B}_n(x) + \mathrm{B}_n(x + \frac{1}{2})$?

5. Now let f be a $\mathcal{C}^n$ function on $[t, t+1]$. Prove that

$$f(t) = \int_t^{t+1} f(x)\,dx + \sum_{m=1}^{n} \frac{B_m}{m!}\big(f^{(m-1)}(t+1) - f^{(m-1)}(t)\big)$$

$$+ (-1)^{n+1} \int_t^{t+1} f^{(n)}(x)\frac{\mathrm{B}_n(x-t)}{n!}\,dx.$$

Therefore, if f is a $\mathcal{C}^n$ function on $[M,N]$ for some integers M,N then

$$\sum_{n=M}^{N-1} f(n) = \int_M^N f(x)\,dx + \sum_{m=1}^{n} \frac{B_m}{m!}\big(f^{(m-1)}(N) - f^{(m-1)}(M)\big) + (-1)^{n+1}\int_M^N f^{(n)}(x)\frac{\mathrm{B}_n(x-\lfloor x\rfloor)}{n!}\,dx;$$

and if f is $\mathcal{C}^n$ on $[M,\infty)$ then

$$\sum_{n=M}^{\infty} f(n) = \int_M^\infty f(x)\,dx - \sum_{m=1}^{n} \frac{B_m}{m!} f^{(m-1)}(M) + (-1)^{n+1}\int_M^\infty f^{(n)}(x)\frac{\mathrm{B}_n(x-\lfloor x\rfloor)}{n!}\,dx, \tag{3}$$

provided the integrals converge and each $f^{(m-1)}(N)\to 0$ as $N\to\infty$. This is a rigorous form of the "Euler-Maclaurin formula"

$$\sum_{n=M}^{\infty} f(n) = \int_M^\infty f(x)\,dx - \sum_{m=1}^{\infty} \frac{B_m}{m!} f^{(m-1)}(M),$$

which rarely converges (can you find any nonzero f for which it *does* converge?), but is often useful as an asymptotic series. For instance, show that for any $s>1$ one can efficiently compute $\zeta(s)$ to within $\exp(-N)$ in time $N^{O(1)}$ by taking $f(x)=x^{-s}$ in (3) and choosing M,n appropriately. Do the same for $L(s,\chi)$ where χ is any nontrivial Dirichlet character and $s>0$. For instance, one can compute Catalan's constant

$$G = L(2,\chi_4) = 1 - \frac{1}{3^2} + \frac{1}{5^2} - \frac{1}{7^2} + - \cdots = .915965594177219015054 6\ldots$$

in this way.

We could also use (3) to obtain the analytic continuation of $\zeta(s)$ and $L(s,\chi)$ to the half-plane $\sigma > 1-n$, and thus to the whole complex plane since n is arbitrary. But this is a less satisfactory approach than using the functional equation which relates $L(s,\chi)$ to $L(1-s,\overline{\chi})$ and thus achieves the analytic continuation to $\mathbf{C}$ in one step.

More about $\psi(x)$:

6. Show that

$$\sum_{p\le x}\log p = \psi(x) - \psi(x^{1/2}) - \psi(x^{1/3}) - \psi(x^{1/5}) + \psi(x^{1/6}) \cdots = \sum_{k=1}^{\infty}\mu(k)\psi(x^{1/k}),$$

where μ is the *Möbius function* taking the product of $r\ge 0$ *distinct* primes to $(-1)^r$ and any non-square-free integer to 0.

Finally, another elementary approach to estimating $\pi(x)$ that gets within a constant of the Prime Number Theorem:

7. Let $P(u)$ be any nonzero polynomial of degree d with integer coefficients; then

$$\int_0^1 f(u)^{2n}\,du \geq 1/\mathrm{lcm}(1,2,\ldots,2dn+1) = \exp(-\psi(2dn+1)).$$

Thus

$$\psi(2dn+1) < 2n \log \min_{0<u<1} 1/|P(u)|.$$

For instance, taking $f(u) = u - u^2$ we find (at least for $4|x$) that $\psi(x) < x \log 4$. This is essentially the same (why?) as Čebyšev's trick of factoring $\binom{2n}{n}$, but suggests different sources of improvement; try $f(u) = (u-u^2)(1-2u)$ for example. [Unfortunately here the upper bound cannot be brought down to $1+\epsilon$; see [Montgomery 1994, Chapter 10] — thanks to Madhav Nori for bringing this to my attention.]

References

[HW 1996] Hardy, G.H., Wright, E.M.: *An Introduction to the Theory of Numbers*, 5th ed. Oxford: Clarendon Press, 1988 [AB 9.88.10 / QA241.H37].

[Montgomery 1994] Montgomery, H.L.: *Ten lectures on the interface between analytic number theory and harmonic analysis.* Providence: AMS, 1994 [AB 9.94.9].

Erdos' Proof

Here's Erdos's proof of Bertrand's Postulate, paraphrased from Hardy and Wright, "An Introduction to the Theory of Numbers."

The proof of Bertand's Postulate uses some simple properties of the function theta(x), defined for x >= 0 by

theta(x) = sum(log p: p is prime and 0 < p <= x)

We show that

theta(x) < 2x log(2)

(I use log(x) always to mean the natural log of x.)

It is enough to show this when x is an integer. We're going to prove this by induction.

The trick is to look at the binomial coefficient C(2m+1, m), which is

(2m+1)!/m!(m+1)!

Call this M for short.

Let p be a prime such that m+1 < p <= 2m+1. Then p divides the numerator of M but not the denominator, so p divides M. So the product of all such primes divides M, and

sum (log p: m+1 < p <= 2m+1) < log M

or in terms of the function theta(x)

theta(2m+1) - theta(m+1) < log M.

On the other hand, the binomial expansion of (1 + 1)^(2m+1) has two terms equal to M, so

2M < 2^(2m+1)

M < 2^2m

log M < 2m log(2)

so

theta(2m+1) - theta(m+1) < 2m log(2)

We're going to use this formula in the induction step of our proof that theta(x) < 2x log(2)

For x = 1, we have
theta(1) = 0 < 2 log(2)

and for x = 2, we have
theta(2) = log 2 < 4 log(2)

Suppose the inequality is true for x < n. Let us prove it for x = n.

If n is even and > 2, then it is certainly not prime, so

theta(n) = theta(n-1) < 2(n-1) log(2) < 2n log(2).

If n is odd, let n = 2m + 1. Then by what we proved above, we have

theta(2m+1) - theta(m+1) < 2m log(2)

theta(2m+1) < theta(m+1) + 2m log(2)

< 2(m+1) log(2) + 2m log(2)

= (3m + 1) log(2)

< (4m + 2) log(2)

= 2n log(2).

This completes the proof that

theta(x) < 2x log(2).

Let's catch our breath.

The next thing we're going to do is to look at the highest power of p that divides n!, where p is any prime. We call this number j(n, p).

We use the notation [x] for the largest integer <= x.

Every p'th number is a multiple of p, so we get [n/p] factors of p in n!. But every p^2'th number is a multiple not just of p but of p squares, and [n/p] doesn't count these, so we need to add [n/p^2} for these extra factors of p. Similarly every p^3'th number is a multiple of p^3 which we have not counted yet. So the highest power of p that divides n! is the sum of all the

[n/p^m]

for m >= 1. Of course [n/p^m] = 0 as soon as p^m > n: that is, for m > log(n)/log(p).

Now we're going to suppose that Bertran's Postulate is false, and that there is no prime p such that n < p < 2n, for some n.

We're going to look at another binomial coefficient. This one is

C(2n,n) = (2n!)/(n!)^2

which we'll call N for short.

By our assumption, all the primes that divide N are <= n. Now using the notation above, we have

N = (2n)!/(n!)^2 =

product(p^j(2n, p): p <= 2n)
---
product(p^2j(n, p): p <= n)

but there aren't any primes between n and 2n by assumption, so the "p <= 2n" in the numerator can be replaced by "p < = n" and we get

N = product (p^(j(2n, p) - 2j(n, p)): p <= n).

Let's call j(2n, p) - 2j(n, p) k(p) for short. Taking logs on both sides, we get

log N = sum(k(p) log(p): p <= n).

Notice that k(p) is a sum of terms of the form [2x] - 2[x}. [2x] - 2[x] is always either 0 or 1. If [2x] is even, [2x] - 2[x] is 0; otherwise it is 1.

We show first that k(p) = 0 for p > 2n/3. For in that case,

2n/3 < p <= n

or 2 <= 2n/p < 3

and [2n/p] = 2, so [2n/p] - 2[n/p] = 0.

p^2 > (4/9)n^2 > 2n as long as n > 4,

and we can certainly assume n is > 4, since we are assuming there is no prime between n and 2n, and 5, for example, is between 4 and 8.

So there are no terms involving higher powers of p.

Next we show that terms with k(p) >= 2 don't contribute very much.

To get such a term we have to have p^2 < 2n or p < sqrt(2n), so the number of such terms is at most sqrt(2n). k(p), on the other

hand, is a sum of terms [2n/p^m] - 2[n/p^m], which is certainly
0 if p^m > 2n, or m > log(2n)/log(p), so k(p) is at most log(2n)/
log(p), and k(p) log(p) <= log(2n), so

sum(k(p) log(p) : k(p) >= 2) <= sqrt(2n) log(2n), taking the maximum possible number of such primes p and a number bigger than any of the k(p) log(p).

For the terms with k(p) = 1, we have at most

sum(log(p): p <= 2n/3) = theta(2n/3) < (4n/3) log(2)

by what we proved way back when.

Putting together what we've got so far gives us

log N < (4n/3) log(2) + sqrt(2n) log(2n).

Time for another breather before we close in for the kill.

Looking back at the definition of N, we have

2^(2n) = 2 + C(2n, 1) + C(2n, 2) + ... + C(2n, 2n-1)

(Binomial Theorem with first and last terms combined).

This is a sum of 2n terms, the largest of which is C(2n, n) or N.

So

2^(2n) < 2nN

or

2n log(2) < log(2n) + log(N)
<= log(2n) + (4n/3) log(2) + sqrt(2n) log(2n)

by what we proved just before the breather.

Now for large values of n, the only term that counts on the right side is the 4n/3 log(2), which is smaller than the 2n log(2). So what we're going to do is figure out how big n needs to be to make this inequality false, and then just prove the postulate directly for smaller values of n.

Take n >= 2^9 and note that log(2n) = log(2^10) = 10 log(2). Divide the inequality by log(2) to get

2^10 < 10 + 2^10(2/3) + (2^5) 10

or

2^10 (1 - 2/3) < 10 (2^5 + 1)
2^10 (1/3) < 10 (2^5 + 1)
2^10 < 30 (2^5 + 1) < 31 (2^5 + 1) = (2^5 - 1) (2^5 + 1)
= 2^10 - 1

which is false!!

So the assumptions that Bertrand's Postulate is false for n and that n >= 2^9 lead to a contradiction. All that remains is to verify the postulate for n < 2^9 = 512.

Here we can just look at the sequence of primes

2, 3, 5, 7, 13, 23, 43, 83, 163, 317, 631

each of which is less that twice the one before.

-Doctor Wilkinson, The Math Forum
Check out our web site! http://mathforum.org/dr.math/

Math 259: Introduction to Analytic Number Theory

The contour integral formula for $\psi(x)$

We now have several examples of Dirichlet series, that is, series of the form[1]

$$F(s) = \sum_{n=1}^{\infty} a_n n^{-s} \tag{1}$$

from which we want to extract information about the growth of $\sum_{n<x} a_n$ as $x \to \infty$. The key to this is a contour integral. We regard $F(s)$ as a function of a *complex* variable $s = \sigma + it$. For real $y > 0$ we have seen already that $|y^{-s}| = y^{-\sigma}$. Thus if the sum (1) converges absolutely[2] for some real σ_0, then it converges uniformly and absolutely to an analytic function on the half-plane $\mathrm{Re}(s) \geq \sigma_0$; and if the sum converges absolutely for all real $s > \sigma_0$, then it converges absolutely to an analytic function on the half-plane $\mathrm{Re}(s) > \sigma_0$. Now for $y > 0$ and $c > 0$ we have

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^s \frac{ds}{s} = \begin{cases} 1, & \text{if } y > 1; \\ \frac{1}{2}, & \text{if } y = 1; \\ 0, & \text{if } y < 1, \end{cases} \tag{2}$$

in the following sense: the contour of integration is the vertical line $\mathrm{Re}(s) = c$, and since the integral is then not absolutely convergent it is regarded as a principal value:

$$\int_{c-i\infty}^{c+i\infty} f(s)\, ds := \lim_{T \to \infty} \int_{c-iT}^{c+iT} f(s)\, ds.$$

Thus interpreted, (2) is an easy exercise in contour integration for $y \neq 1$, and an elementary manipulation of $\log s$ for $y = 1$. So we expect that if (1) converges absolutely in $\mathrm{Re}(s) > \sigma_0$ then

$$\sum_{n<x} a_n = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^s F(s) \frac{ds}{s} \tag{3}$$

for any $c > \sigma_0$, using the principal value of the integral and adding $a_x/2$ to the sum if x happens to be an integer. But getting from (1) and (2) to (3) involves interchanging an infinite sum with a conditionally convergent integral, which is not in general legitimate. Thus we replace $\int_{c-i\infty}^{c+i\infty}$ by $\int_{c-iT}^{c+iT}$, which legitimizes the manipulation but introduces an error term into (2). We estimate this error term as follows:

Lemma. *For $y, c, T > 0$ we have*

$$\frac{1}{2\pi i} \int_{c-iT}^{c+iT} y^s \frac{ds}{s} = \begin{cases} 1 + O(y^c \min(1, \frac{1}{T|\log y|})), & \text{if } y \geq 1; \\ O(y^c \min(1, \frac{1}{T|\log y|})), & \text{if } y \leq 1, \end{cases} \tag{4}$$

[1] As noted by Serre, everything works just as well with "Dirichlet series" $\sum_{k=0}^{\infty} a_k n_k^{-s}$, where n_k are positive reals such that $n_k \to \infty$ as $k \to \infty$. In that more general setting we would seek to estimate $\sum_{n_k<x} a_k$ as $x \to \infty$.

[2] We shall see later that the same results hold if absolute convergence is replaced by conditional convergence throughout. For example, for every nonprincipal character χ the series for $L(s, \chi)$ converges uniformly in the half-plane $\mathrm{Re}(s) > \sigma_0$ for each positive σ_0.

the implied O-constant being effective and uniform in y, c, T.

(In fact the error's magnitude is less than both y^c and $y^c/\pi T|\log y|$. Of course if y *equals* 1 then the error term is regarded as $O(1)$ and is valid for both approximations $0, 1$ to the integral.)

Proof: Complete the contour of integration to a rectangle extending to real part $-M$ if $y \geq 1$ or $+M$ if $y \leq 1$. The resulting contour integral is 1 or 0 respectively by the residue theorem. We may let $M \to \infty$ and bound the horizontal integrals by $(\pi T)^{-1} \int_0^\infty y^{c \pm r} dr$; this gives the estimate $y^c/\pi T|\log y|$. Using a circular arc centered at the origin instead of a rectangle yields the same residue with a remainder of absolute value $< y^c$. $\square$

This Lemma will let us approximate $\sum_{n<x} a_n$ by $(2\pi i)^{-1} \int_{c-iT}^{c+iT} x^s F(s)\, ds/s$. We shall eventually choose some T and exploit the analytic continuation of F to shift the contour of integration past the region of absolute convergence to obtain nontrivial estimates.

The next question is, which F should we choose? Consider for instance $\zeta(s)$. We have in effect seen already that if we take $F(s) = \log \zeta(s)$ then the sum of the resulting a_n over $n < x$ closely approximates $\pi(x)$. Unfortunately, while $\zeta(s)$ continues meromorphically to $\sigma \leq 1$, its logarithm does not: it has essential logarithmic singularities at the pole $s = 1$, and at zeros of $\zeta(s)$ to be described later. So we use the *logarithmic derivative* of $\zeta(s)$ instead, which at each pole or zero of ζ has a simple pole with a known residue and thus a predictable effect on our contour integral.

What are the coefficients a_n for this logarithmic derivative? It is convenient to use not ζ'/ζ but $-\zeta'/\zeta$, which has positive coefficients. Using the Euler product we find

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_p \frac{d}{ds} \log(1 - p^{-s}) = \sum_p \log p \frac{p^{-s}}{1 - p^{-s}} = \sum_p \log p \sum_{k=1}^{\infty} p^{-ks}.$$

That is,

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

So the coefficient of n^{-s} is none other than the von Mangoldt function which arose in the factorization of $x!$. Hence our contour integral

$$\frac{1}{2\pi i} \int_{c-iT}^{c+iT} -\frac{\zeta'}{\zeta}(s)\, x^s \frac{ds}{s} \quad (c > 1)$$

approximates $\psi(x)$. The error can be estimated by our Lemma (4): since $|\Lambda(n)| \leq \log n$, the error is of order at most

$$\sum_{n=1}^{\infty} (x/n)^c \log n \cdot \min(1, \frac{1}{T|\log(x/n)|})$$

which is $O(T^{-1} x^c \log^2 x)$ provided $1 < T < x$. (See the Exercises below.)

Taking $c = 1 + A/\log x$, so that $x^c \ll x$, we find:

$$\psi(x) = \frac{1}{2\pi i}\int_{1+\frac{A}{\log x}-iT}^{1+\frac{A}{\log x}+iT} -\frac{\zeta'}{\zeta}(s)\, x^s \frac{ds}{s} + O_A\left(\frac{x\log^2 x}{T}\right). \tag{5}$$

Similarly for any Dirichlet character χ we obtain a formula for

$$\psi(x,\chi) := \sum_{n<x}\chi(n)\Lambda(n)$$

by replacing $\zeta(s)$ in (5) by $L(s,\chi)$.

To make use of this we'll want to shift the line of integration to the left, where $|x^s|$ is smaller. As we do so we shall encounter poles at $s = 1$ and at zeros of $\zeta(s)$ (or $L(s,\chi)$), and will have to estimate $|\zeta'/\zeta|$ (or $|L'(s,\chi)/L(s,\chi)|$) over the resulting contour. **This is why we are interested in the analytic continuation of $\zeta(s)$ and likewise $L(s,\chi)$ and in their zeros.** We investigate these matters next.

Remarks

We can already surmise that $\psi(x)$ will be approximated by $x - \sum_\rho x^\rho/\rho$, the sum running over zeros ρ of $\zeta(s)$ counted with multiplicity, and thus that the Prime Number Theorem is tantamount to the nonvanishing of $\zeta(s)$ on $\mathrm{Re}(s) = 1$. The fact that $\zeta(1+it) \neq 0$ is also the key step in various "elementary" proofs or the Prime Number Theorem such as [Newman 1980] (see also [Zagier 1997]). Likewise for $L(1+it,\chi)$ and the asymptotic formula for $\pi(x, a \bmod q)$.

The formula for $\psi(x)$ as a contour integral can be viewed as an instance of the inverse Mellin transform. Suppose $F(s)$ is a generalized Dirichlet series $\sum_{k=0}^\infty a_k n_k^{-s}$, converging for $\mathrm{Re}(s) > \sigma_0$. Let $A(x) = \sum_{n_k<x} a_k$, and assume that $A(x)\to\infty$ as $x\to\infty$. In particular, $\sigma_0 \geq 0$. Now

$$F(s) = \int_0^\infty x^{-s} dA(x) = s\int_0^\infty x^{-s}A(x)\frac{dx}{x},$$

so $F(s)/s$ is the Mellin transform of $A(x)$. Thus we expect that

$$A(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^s F(s)\frac{ds}{s}$$

for $c > \sigma_0$. Due to the discontinuities of $A(x)$ at $x = n_k$, this integral cannot converge absolutely, but its principal value does equal $A(x)$ at all $x \notin \{n_k\}$.

Exercises

1. Verify that the error

$$\sum_{n=1}^\infty (x/n)^c \log n \cdot \min(1, \frac{1}{T|\log(x/n)|})$$

in our approximation of $\psi(x)$ is $O(T^{-1}x^c\log^2 x)$ provided $1 < T < x$. Explain why the bound need not hold if T is large compared to x.

2. Use (4) to show that nevertheless $\psi(x)$ is given by the principal value integral

$$\psi(x) = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} -\frac{\zeta'}{\zeta}(s)\, x^s \frac{ds}{s} \tag{6}$$

for all $x, c > 1$.

3. Show that $\sum_{n=1}^{\infty} \mu(n) n^{-s} = 1/\zeta(s)$, with μ being the Möbius function defined in the previous set of exercises. Deduce an integral formula for $\sum_{n<x} \mu(n)$ analogous to (6), and an approximate integral formula analogous to (5) but with error only $O(T^{-1} x \log x)$ instead of $O(T^{-1} x \log^2 x)$.

References

[Newman 1980] Newman, D.J.: Simple Analytic Proof of the Prime Number Theorem, *Amer. Math. Monthly* **87** (1980), 693–696.

[Zagier 1997] Zagier, D.: Newman's Short Proof of the Prime Number Theorem, *Amer. Math. Monthly* **104** (1994), 705–708.

Math 259: Introduction to Analytic Number Theory

The Riemann zeta function and its functional equation
(and a review of the Gamma function and Poisson summation)

Recall Euler's identity:

$$[\zeta(s) :=] \sum_{n=1}^{\infty} n^{-s} = \prod_{p \text{ prime}} \left(\sum_{c_p=1}^{\infty} p^{-c_p s} \right) = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}. \tag{1}$$

We showed that this holds as an identity between absolutely convergent sums and products for real $s > 1$. Riemann's insight was to consider (1) as an identity between functions of a *complex* variable s. We follow the curious but nearly universal convention of writing the real and imaginary parts of s as σ and t, so

$$s = \sigma + it.$$

We already observed that for all real $n > 0$ we have $|n^{-s}| = n^{-\sigma}$, because

$$n^{-s} = \exp(-s \log n) = n^{-\sigma} e^{it \log n}$$

and $e^{it \log n}$ has absolute value 1; and that both sides of (1) converge absolutely in the half-plane $\sigma > 1$, and are equal there either by analytic continuation from the real ray $t = 0$ or by the same proof we used for the real case. Riemann showed that the function $\zeta(s)$ extends from that half-plane to a meromorphic function on all of $\mathbf{C}$ (the "Riemann zeta function"), analytic except for a simple pole at $s = 1$. The continuation to $\sigma > 0$ is readily obtained from our formula

$$\zeta(s) - \frac{1}{s-1} = \sum_{n=1}^{\infty} \left[n^{-s} - \int_n^{n+1} x^{-s}\, dx \right] = \sum_{n=1}^{\infty} \int_n^{n+1} (n^{-s} - x^{-s})\, dx,$$

since for $x \in [n, n+1]$ $(n \geq 1)$ and $\sigma > 0$ we have

$$|n^{-s} - x^{-s}| = \left| s \int_n^x y^{-1-s}\, dy \right| \leq |s| n^{-1-\sigma}$$

so the formula for $\zeta(s) - (1/(s-1))$ is a sum of analytic functions converging absolutely in compact subsets of $\{\sigma + it : \sigma > 0\}$ and thus gives an analytic function there. (See also the first Exercise below.) Using the Euler-Maclaurin summation formula with remainder, we could proceed in this fashion, extending ζ to $\sigma > -1$, $\sigma > -2$, etc. However, once we have defined $\zeta(s)$ on $\sigma > 0$ we can obtain the entire analytic continuation at once from Riemann's *functional equation* relating $\zeta(s)$ with $\zeta(1-s)$. This equation is most nicely stated by introducing the meromorphic function $\xi(s)$ defined by[1]

$$\xi(s) := \pi^{-s/2} \Gamma(s/2) \zeta(s)$$

[1] Warning: occasionally one still sees $\xi(s)$ defined as what we would call $(s^2 - s)\xi(s)$ or $(s^2 - s)\xi(s)/2$, as in [GR 1980, 9.561]. The factor of $(s^2 - s)$ makes the function entire, and does not affect the functional equation since it is symmetric under $s \leftrightarrow 1 - s$. However, for most uses it turns out to be better to leave this factor out and tolerate the poles at $s = 0, 1$.

for $\sigma > 0$. Then we have:

Theorem (Riemann)**.** *The function ξ extends to a meromorphic function on $\mathbf{C}$, regular except for simple poles at $s = 0, 1$, which satisfies the functional equation*

$$\xi(s) = \xi(1 - s). \tag{2}$$

It follows that ζ also extends to a meromorphic function on $\mathbf{C}$, which is regular except for a simple pole at $s = 1$, and that this analytic continuation of ζ has simple zeros at the negative even integers $-2, -4, -6, \ldots$, and no other zeros outside the closed critical strip *$0 \le \sigma \le 1$.*

[The zeros $-2, -4, -6, \ldots$ of ζ outside the critical strip are called the *trivial zeros* of the Riemann zeta function.]

The *proof* has two ingredients: properties of $\Gamma(s)$ as a meromorphic function of $s \in \mathbf{C}$, and the *Poisson summation formula*. We next review these two topics.

The *Gamma function* was defined for real $s > 0$ by Euler[2] as the integral

$$\Gamma(s) := \int_0^\infty x^s e^{-x} \frac{dx}{x}. \tag{3}$$

We have $\Gamma(1) = \int_0^\infty e^{-x}\, dx = 1$ and, integrating by parts,

$$s\Gamma(s) = \int_0^\infty e^{-x} d(x^s) = -\int_0^\infty x^s d(e^{-x}) = \Gamma(s+1) \qquad (s > 0),$$

so by induction $\Gamma(n) = (n-1)!$ for positive integers n. Since $|x^s| = x^\sigma$, the integral (3) defines an analytic function on $\sigma > 0$, which still satisfies the recursion $s\Gamma(s) = \Gamma(s+1)$ (proved either by repeating the integration by parts or by analytic continuation from the positive real axis). That recursion then extends Γ to a meromorphic function on $\mathbf{C}$, analytic except for simple poles at $0, -1, -2, -3, \ldots$. (What are the residues at those poles?) For s, s' in the right half-plane $\sigma > 0$ the *Beta function*[3] $\mathrm{B}(s, s')$, defined by the integral

$$\mathrm{B}(s, s') := \int_0^1 x^{s-1}(1-x)^{s'-1}\, dx,$$

is related with Γ by

$$\Gamma(s + s')\mathrm{B}(s, s') = \Gamma(s)\Gamma(s'). \tag{4}$$

(This is proved by Euler's trick of calculating $\int_0^\infty\int_0^\infty x^{s-1}y^{s'-1}e^{-(x+y)}\, dx\, dy$ in two different ways.) Since $\Gamma(s) > 0$ for real positive s, it readily follows that Γ has no zeros in $\sigma > 0$, and therefore none in the complex plane.

This is enough to derive the poles and trivial zeros of ζ from the functional equation (2). [Don't take my word for it — do it!] But where does the functional equation come from? There are several known ways to prove it; we give

[2] Actually Euler used $\Pi(s-1)$ for what we call $\Gamma(s)$; thus $\Pi(n) = n!$ for $n = 0, 1, 2, \ldots$.
[3] a.k.a. "Euler's first integral", with (3) being "Euler's second integral".

Riemann's original method, which generalizes to $L(s,\chi)$, and further to L-series associated to modular forms.

Riemann expresses $\xi(s)$ as a Mellin integral involving the *theta function*[4]

$$\theta(u) := \sum_{n=-\infty}^{\infty} e^{-\pi n^2 u} = 1 + 2(e^{-\pi u} + e^{-4\pi u} + e^{-9\pi u} + \ldots),$$

the sum converging absolutely to an analytic function on the upper half-plane $\mathrm{Re}(u) > 0$. Integrating termwise we find:

$$2\xi(s) = \int_0^\infty (\theta(u) - 1) u^{s/2} \frac{du}{u} \qquad (\sigma > 0).$$

(That is, $\xi(-2s)$ is the Mellin transform of $(\theta(u) - 1)/2$.) But we shall see:

Lemma. *The function $\theta(u)$ satisfies the identity*

$$\theta(1/u) = u^{1/2}\theta(u). \tag{5}$$

Assume this for the time being. We then rewrite our integral for $2\xi(s)$ as

$$\int_0^1 (\theta(u) - 1) u^{s/2} \frac{du}{u} + \int_1^\infty (\theta(u) - 1) u^{s/2} \frac{du}{u}$$
$$= -\frac{2}{s} + \int_0^1 \theta(u) u^{s/2} \frac{du}{u} + \int_1^\infty (\theta(u) - 1) u^{s/2} \frac{du}{u},$$

and use the change of variable $u \leftrightarrow 1/u$ to find

$$\int_0^1 \theta(u) u^{s/2} \frac{du}{u} = \int_1^\infty \theta(u^{-1}) u^{-s/2} \frac{du}{u}$$
$$= \int_1^\infty \theta(u) u^{(1-s)/2} \frac{du}{u} = \frac{2}{s-1} + \int_1^\infty (\theta(u) - 1) u^{(1-s)/2} \frac{du}{u}$$

if also $\sigma < 1$. Therefore

$$\xi(s) + \frac{1}{s} + \frac{1}{1-s} = \frac{1}{2} \int_1^\infty (\theta(u) - 1)(u^{s/2} + u^{(1-s)/2}) \frac{du}{u},$$

which is manifestly symmetrical under $s \leftrightarrow 1 - s$, and analytic since $\theta(u)$ decreases exponentially as $u \to \infty$. This concludes the proof of the functional equation and analytic continuation of ξ, assuming our lemma (5).

This lemma, in turn, is the special case $f(x) = e^{-\pi u x^2}$ of the Poisson summation formula:

[4] Jacobi introduced four "theta functions" of two variables; in his notation, our $\theta(u)$ would be $\theta_3(0, e^{-\pi u})$. We can call this $\theta(u)$ because we shall not use $\theta_1, \theta_2, \theta_4$, nor $\theta_3(z,q)$ for $z \neq 0$.

Theorem. *Let $f : \mathbf{R} \to \mathbf{C}$ be a $\mathcal{C}^2$ function such that $(|x|^r + 1)(|f(x)| + |f''(x)|)$ is bounded for some $r > 1$, and let $\hat{f}$ be its Fourier transform*

$$\hat{f}(y) = \int_{-\infty}^{+\infty} e^{2\pi i x y} f(x)\, dx.$$

Then

$$\sum_{m=-\infty}^{\infty} f(m) = \sum_{n=-\infty}^{\infty} \hat{f}(n), \tag{6}$$

the sums converging absolutely.

[The hypotheses on f can be weakened, but this formulation of Poisson summation is more than enough for our purposes.]

Proof: Define $F : \mathbf{R}/\mathbf{Z} \to \mathbf{C}$ by

$$F(x) := \sum_{m=-\infty}^{\infty} f(x + m),$$

the sum converging absolutely to a $\mathcal{C}^2$ function by the assumption on f. Thus the Fourier series of F converges absolutely to F, so in particular

$$F(0) = \sum_{n=-\infty}^{\infty} \int_0^1 e^{2\pi i n x} F(x)\, dx.$$

But $F(0)$ is just the left-hand side of (6), and the integral is

$$\sum_{m \in \mathbf{Z}} \int_0^1 e^{2\pi i n x} f(x + m)\, dx = \sum_{m \in \mathbf{Z}} \int_m^{m+1} e^{2\pi i n x} f(x)\, dx = \int_{-\infty}^{\infty} e^{2\pi i n x} f(x)\, dx$$

which is just $\hat{f}(n)$, so its sum over $n \in \mathbf{Z}$ yields the right-hand side of (6). □

Now let $f(x) = e^{-\pi u x^2}$. The hypotheses are handily satisfied for any r, so (6) holds. The left-hand side is just $\theta(u)$. To evaluate the right-hand side, we need the Fourier transform of f, which is $u^{-1/2} e^{-\pi u^{-1} y^2}$. [Contour integration reduces this claim to $\int_{-\infty}^{\infty} e^{-\pi u x^2}\, dx = u^{-1/2}$, which is the well-known Gauss integral — see the Exercises.] Thus the right-hand side is $u^{-1/2}\theta(1/u)$. Multiplying both sides by $u^{1/2}$ we then obtain (5), and finally complete the proof of the analytic continuation and functional equation for $\xi(s)$.

Remarks. We noted already that to each number field K there corresponds a zeta function

$$\zeta_K(s) := \sum_I |I|^{-s} = \prod_{\wp} (1 - |\wp|^{-s})^{-1} \qquad (\sigma > 1),$$

in which $|I|$ is the norm of an ideal I, the sum and product extend respectively over ideals I and prime ideals $\wp$ of the ring of integers O_K, and their equality

expresses unique factorization. As in our case of $K = \mathbf{Q}$, this zeta function extends to a meromorphic function on $\mathbf{C}$, regular except for a simple pole at $s = 1$. Moreover it satisfies a functional equation $\xi_K(s) = \xi_K(1-s)$, where

$$\xi_K(s) := \Gamma(s/2)^{r_1}\Gamma(s)^{r_2}(4^{-r_2}\pi^{-n}|d|)^{s/2}\zeta_K(s),$$

in which $n = r_1 + 2r_2 = [K : \mathbf{Q}]$, the exponents r_1, r_2 are the numbers of real and complex embeddings of K, and d is the discriminant of $K/\mathbf{Q}$. The factors $\Gamma(s/2)^{r_1}, \Gamma(s)^{r_2}$ may be regarded as factors corresponding to the "archimedean places" of K, as the factor $(1 - |\wp|^{-s})^{-1}$ corresponds to the finite place $\wp$. The functional equation can be obtained from generalized Poisson summation as in [Tate 1950]. Most of our results for $\zeta = \zeta_{\mathbf{Q}}$ carry over to these ζ_K, and yield a Prime Number Theorem for primes of K; L-series generalize too, though the proper generalization requires some thought when the class and unit groups need no longer be trivial and finite as they are for $\mathbf{Q}$. See for instance H.Heilbronn's "Zeta-Functions and L-Functions", Chapter VIII of [CF 1967].

Exercises

Concerning the analytic continuation of $\zeta(s)$:

1. Show that if $\alpha : \mathbf{Z} \to \mathbf{C}$ is a function such that $\sum_{m=1}^{n} \alpha(m) = O(1)$ (for instance, if α is a nontrivial Dirichlet character) then $\sum_{n=1}^{\infty} \alpha(n)n^{-s}$ converges uniformly, albeit not absolutely, in compact subsets of $\{\sigma + it : \sigma > 0\}$, and thus defines an analytic function on that half-plane. Apply this to

$$(1 - 2^{1-s})\zeta(s) = 1 - \frac{1}{2^s} + \frac{1}{3^s} - \frac{1}{4^s} + - \cdots$$

(with $\alpha(n) = (-1)^{n-1}$) and to $(1 - 3^{1-s})\zeta(s)$ to obtain a different proof of the analytic continuation of ζ to $\sigma > 0$.

2. Prove that the Bernoulli polynomials B_n $(n > 0)$ have the Fourier expansion

$$\mathrm{B}_n(x) = -n! \sum_k{}' \frac{e^{2k\pi i x}}{(2k\pi i)^n} \tag{7}$$

for $0 < x < 1$, in which $\sum_k'$ is the sum over nonzero integers k. Deduce that

$$\zeta(n) = \frac{1}{2}(2\pi)^n \frac{|B_n|}{n!} \quad (n = 2, 4, 6, 8, \ldots),$$

and thus that $\zeta(1-n) = -B_n/n$ for all integers $n > 1$. For example, $\zeta(-1) = -1/12$. What is $\zeta(0)$?

It is known that in general $\zeta_K(-m) \in \mathbf{Q}$ $(m = 0, 1, 2, \ldots)$ for any number field K. In fact the functional equation for ζ_K indicates that once $[K : \mathbf{Q}] > 1$ all the $\zeta_K(-m)$ vanish unless K is totally real and m is odd, in which case the rationality of $\zeta_K(-m)$ was obtained in [Siegel 1969].

A further application of (7):

3. Prove that $\sum\sum_{k,k'=0}^{\infty}(kk'(k+k'))^{-n}$ is a rational multiple of π^{3n} for each $n = 2, 4, 6, 8, \ldots$; for instance,

$$\sum_{k=0}^{\infty}\sum_{k'=0}^{\infty}\frac{1}{(kk'(k+k')^2)} = \frac{4\pi^6}{3}\int_0^1 \mathrm{B}_2(x)^3\,dx = \frac{\pi^6}{2835}.$$

Concerning the Gamma function:

4. If you've never seen it yet, or did it once but forgot, prove (4) by starting from the integral representation of the right-hand side as

$$\int_0^\infty\int_0^\infty x^{s-1}y^{s'-1}e^{-(x+y)}\,dx\,dy$$

and applying the change of variable $(x, y) = (uz, (1-u)z)$.

We will have little use for the Beta function in Math 259, but an analogous transformation will arise later in the formula relating Gauss and Jacobi sums.

5. Now take $s = s' = 1/2$ to prove that $\Gamma(1/2) = \sqrt{\pi}$, and thus to obtain the Gauss integral

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

Then take $s' = s$ and use the change of variable $u = (1-2x)^2$ in the integral defining $\mathrm{B}(s, s)$ to obtain $\mathrm{B}(s, s) = 2^{1-2s}\mathrm{B}(s, 1/2)$, and thus the *duplication formula*

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

Concerning functional equations:

6. Use the duplication formula and the identity $\mathrm{B}(s, 1-s) = \pi/\sin(\pi s)$ to write (2) in the equivalent form

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

This asymmetrical formulation of the functional equation has the advantage of showing the trivial zeros of $\zeta(s)$ more clearly (given the fact that $\zeta(s)$ has a simple pole at $s = 1$ and no other poles or zeros on the positive real axis).

7. Let χ_8 be the Dirichlet character mod 8 defined by $\chi_8(\pm 1) = 1$, $\chi_8(\pm 3) = -1$. Show that if f is a function satisfying the hypotheses of Poisson summation then

$$\sum_{m=-\infty}^{\infty}\chi_8(m)f(m) = 8^{-1/2}\sum_{n=-\infty}^{\infty}\chi_8(n)\hat{f}(n/8).$$

Letting $f(x) = e^{-\pi u x^2}$, obtain an identity analogous to (5), and deduce a functional equation for $L(s, \chi_8)$.

8. Now let χ_4 be the character mod 4 defined by $\chi_4(\pm 1) = \pm 1$. Show that, again under the Poisson hypotheses,

$$\sum_{m=-\infty}^{\infty} \chi_4(m) f(m) = \frac{1}{2} \sum_{n=-\infty}^{\infty} \chi_4(n) \hat{f}(n/4).$$

This time, taking $f(x) = e^{-\pi u x^2}$ does not accomplish much! Use $f(x) = x e^{-\pi u x^2}$ instead to find a functional equation for $L(s, \chi_4)$.

We shall see that the L-function associated to any primitive Dirichlet character χ satisfies a similar functional equation, with the Gamma factor depending on whether $\chi(-1) = +1$ or $\chi(-1) = -1$.

9. For light relief after all this hard work, differentiate the identity (5) with respect to u, set $u = 1$, and conclude that $e^\pi > 8\pi - 2$. What is the approximate size of the difference?

Further applications of Poisson summation:

10. Use Poisson summation to evaluate $\sum_{n=1}^{\infty} 1/(n^2+c^2)$ for $c > 0$. [The Fourier transform of $1/(x^2 + c^2)$ is a standard exercise in contour integration.] Verify that your answer approaches $\zeta(2) = \pi^2/6$ as $c \to 0$.

11. [Higher-dimensional Poisson, and more on zeta functions of quadratic forms] Let A be a real positive-definite symmetric matrix of order r, and $Q : \mathbf{R}^r \to \mathbf{R}$ the associated quadratic form $Q(x) = (x, Ax)$. The *theta function of* Q is

$$\theta_Q(u) := \sum_{n \in \mathbf{Z}^r} \exp(-\pi Q(n) u).$$

For instance, if $r = 1$ and $A = 1$ then $\theta_Q(u)$ is just $\theta(u)$. More generally, show that if A is the identity matrix I_r (so $Q(x) = \sum_{j=1}^r x_j^2$) then $\theta_Q(u) = \theta(u)^r$. Prove an r-dimensional generalization of the Poisson summation formula, and use it to obtain a generalization of (5) that relates $\theta_Q(u)$ with $\theta_{Q^*}(1/u)$, where Q^* is the quadratic form associated to A^{-1}. Using this formula, and a Mellin integral formula for

$$\zeta_Q(s) = \sum_{\substack{n \in \mathbf{Z}^r \\ n \neq 0}} \frac{1}{Q(n)^s},$$

conclude that ζ_Q extends to a meromorphic function on $\mathbf{C}$ that satisfies a functional equation relating ζ_Q with ζ_{Q^*}. Verify that when $r = 2$ and $A = I_2$ your functional equation is consistent with the identity $\zeta_Q(s) = 4\zeta(s)L(s, \chi_4)$ and the functional equations for $\zeta(s)$ and $L(s, \chi_4)$.

References

[CF 1967] Cassels, J.W.S., Fröhlich, A., eds.: *Algebraic Number Theory.* London: Academic Press 1967. [AB 9.67.2 / QA 241.A42]

[GR 1980] Gradshteyn, I.S., Ryzhik, I.M.: *Table of Integrals, Series, and Products.* New York: Academic Press 1980. [D 9.80.1 / BASEMENT REFERENCE QA55.G6613]

[Siegel 1969] Siegel, C.L.: Berechnung von Zetafunktionen an ganzzahligen Stellen, *Gött. Nach.* **10** (1969), 87–102.

[Tate 1950] Tate, J.T.: *Fourier Analysis in Number Fields and Hecke's Zeta-Functions.* Thesis, 1950; Chapter XV of [CF 1967].

Math 259: Introduction to Analytic Number Theory

More about the Gamma function

We collect some more facts about $\Gamma(s)$ as a function of a complex variable that will figure in our treatment of $\zeta(s)$ and $L(s,\chi)$. All of these, and most of the Exercises, are standard textbook fare; one basic reference is Ch. XII (pp. 235–264) of [WW 1940]. One reason for not just citing Whittaker & Watson is that some of the results concerning Euler's integrals B and Γ have close analogues in the Gauss and Jacobi sums associated to Dirichlet characters, and we shall need these analogues before long.

The product formula for $\Gamma(s)$. Recall that $\Gamma(s)$ has simple poles at $s = 0, -1, -2, \ldots$ and no zeros. We readily concoct a product that has the same behavior: let

$$g(s) := \frac{1}{s} \prod_{k=1}^{\infty} e^{s/k} \Big/ \Big(1 + \frac{s}{k}\Big) ,$$

the product converging uniformly in compact subsets of $\mathbf{C} - \{0, -1, -2, \ldots\}$ because $e^x/(1+x) = 1 + O(x^2)$ for small x. Then Γ/g is an entire function with neither poles nor zeros, so it can be written as $\exp \alpha(s)$ for some entire function α. We show that $\alpha(s) = -\boldsymbol{\gamma} s$, where $\boldsymbol{\gamma} = 0.57721566490\ldots$ is *Euler's constant*:

$$\boldsymbol{\gamma} := \lim_{N\to\infty} \Big(-\log N + \sum_{k=1}^{N} \frac{1}{k} \Big).$$

That is, we show:

Lemma. *The Gamma function has the product formulas*

$$\Gamma(s) = e^{-\boldsymbol{\gamma} s} g(s) = \frac{e^{-\boldsymbol{\gamma} s}}{s} \prod_{k=1}^{\infty} e^{s/k} \Big/ \Big(1 + \frac{s}{k}\Big) = \frac{1}{s} \lim_{N\to\infty} \left(N^s \prod_{k=1}^{N} \frac{k}{s+k} \right). \quad (1)$$

Proof: For $s \neq 0, -1, -2, \ldots$, the quotient $g(s+1)/g(s)$ is the limit as $N\to\infty$ of

$$\begin{aligned} \frac{s}{s+1} \prod_{k=1}^{N} e^{1/k} \frac{1 + \frac{s}{k}}{1 + \frac{s+1}{k}} &= \frac{s}{s+1} \left(\exp \sum_{k=1}^{N} \frac{1}{k} \right) \prod_{k=1}^{N} \frac{k+s}{k+s+1} \\ &= s \cdot \frac{N}{N+s+1} \cdot \exp \left(-\log N + \sum_{k=1}^{N} \frac{1}{k} \right). \end{aligned}$$

Now the factor $N/(N+s+1)$ approaches 1, while $-\log N + \sum_{k=1}^{N} \frac{1}{k} \to \boldsymbol{\gamma}$. Thus $g(s+1) = s e^{\boldsymbol{\gamma}} g(s)$, and if we define $\Gamma^?(s) := e^{-\boldsymbol{\gamma} s} g(s)$ then $\Gamma^?$ satisfies the same functional equation $\Gamma^?(s+1) = s\Gamma^?(s)$ satisfied by Γ. We are claiming that in fact $\Gamma^? = \Gamma$.

Consider $q := \Gamma/\Gamma^?$, an entire function of period 1. Thus it is an analytic function of $e^{2\pi i s} \in \mathbf{C}^*$. We wish to show that $q = 1$ identically. By the definition of g we have $\lim_{s\to 0} sg(s) = 1$; hence

$$\lim_{s\to 0} s\Gamma^?(s) = \lim_{s\to 0} sg(s) = 1 = \lim_{s\to 0} s\Gamma(s),$$

and $q(0) = 1$. We claim that there exists a constant C such that

$$|q(\sigma + it)| \le Ce^{\pi|t|/2} \tag{2}$$

for all real σ, t; since the coefficient $\pi/2$ in the exponent is less than 2π, it will follow that q is constant, and thus that $\Gamma^? = \Gamma$ as claimed.

Since q is periodic, we need only prove (2) for $s = \sigma + it$ with $\sigma \in [1, 2]$. For such s, we have $|\Gamma(\sigma + it)| \le \Gamma(\sigma)$ by the integral formula and

$$\left|\frac{\Gamma^?(\sigma + it)}{\Gamma^?(\sigma)}\right| = \prod_{k=0}^{\infty} \frac{\sigma + k}{|\sigma + k + it|} = \exp -\frac{1}{2}\sum_{k=0}^{\infty} \log\left(1 + \frac{t^2}{(\sigma + k)^2}\right).$$

The summand is a decreasing function of k, so the sum is

$$\le \int_0^{\infty} \log\bigl(1 + (t/x)^2\bigr)\, dx = |t| \int_0^{\infty} \log\bigl(1 + (1/x)^2\bigr)\, dx,$$

which on integration by parts becomes $2|t| \int_0^{\infty} dx/(x^2 + 1) = \pi|t|$. This proves (2) with $C = \sup_{1\le\sigma\le 2} q(\sigma)$, and completes the proof of (1). $\square$

Consequences of the product formula. Our most important application of the product formula for $\Gamma(s)$ is the *Stirling approximation*[1] to $\log\Gamma(s)$. Fix $\epsilon > 0$ and let R_ϵ be the region

$$\{s \in \mathbf{C}^* : |\mathrm{Im}(\log s)| < \pi - \epsilon\}.$$

Then R_ϵ is a simply-connected region containing none of the poles of Γ, so there is an analytic function $\log\Gamma$ on R_ϵ, real on $R_\epsilon \cap \mathbf{R}$, and given by the above product formula:

$$\log\Gamma(s) = \lim_{N\to\infty} \Bigl(s \log N + \log N! - \sum_{k=0}^{N} \log(s + k)\Bigr). \tag{3}$$

We prove:

Lemma. *The approximation*

$$\log\Gamma(s) = (s - \frac{1}{2})\log s - s + \frac{1}{2}\log(2\pi) + O_\epsilon(|s|^{-1}) \tag{4}$$

holds for all s in R_ϵ.

[1] Originally only for $n! = \Gamma(n + 1)$, but we need it for complex s as well.

Proof: The estimate holds for small s, say $|s| < 1$, because $O_\epsilon(|s|^{-1})$ well exceeds all the other terms. We thus assume $|s| \geq 1$, and estimate the sum in (3) as we did for $\log x!$ in obtaining the original form of Stirling's approximation. The sum differs from

$$\int_{-\frac{1}{2}}^{N+\frac{1}{2}} \log(s+x)\,dx = (N+\frac{1}{2}+s)\log(N+\frac{1}{2}+s) - (s-\frac{1}{2})\log(s-\frac{1}{2}) - N - 1$$

$$= (N+\frac{1}{2}+s)\log N + (N+\frac{1}{2}+s)\log(1+\frac{1}{N}(s+\frac{1}{2})) - (s-\frac{1}{2})\log(s-\frac{1}{2}) - N - 1$$

by

$$\frac{1}{2}\int_{-\frac{1}{2}}^{N+\frac{1}{2}} \frac{\|x+\frac{1}{2}\|^2}{(s+x)^2}\,dx \ll_\epsilon |s|^{-1}.$$

We already know that $\log N! = (N+1/2)\log N - N + A + O(N^{-1})$ for some constant A. The estimate (4) follows upon taking $N\to\infty$, except for the value $\frac{1}{2}\log(2\pi)$ of the constant term. This constant can be obtained by letting $s\to\infty$ in the duplication formula $\Gamma(2s) = \pi^{-1/2}2^{2s-1}\Gamma(s)\Gamma(s+\frac{1}{2})$. $\square$

One can go on to expand the $O_\epsilon(|s|^{-1})$ error in an asymptotic series in inverse powers of s (see the Exercises), but (4) is already more than sufficient for our purposes, in that we do not need the identification of the constant term with $\frac{1}{2}\log 2\pi$.

The logarithmic derivative of our product formula for $\Gamma(s)$ is

$$\frac{\Gamma'(s)}{\Gamma(s)} = -\gamma - \frac{1}{s} + \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{s+k}\right) = \lim_{N\to\infty}\left[\log N - \sum_{k=0}^{N}\frac{1}{s+k}\right].$$

Either by differentiating[2] (4) or by applying the same Euler-Maclaurin step to $\sum_0^N 1/(s+k)$ we find that

$$\frac{\Gamma'(s)}{\Gamma(s)} = \log s - \frac{1}{2s} + O_\epsilon(|s|^{-2}). \tag{5}$$

Remark

The product formula for $\Gamma(s)$ can also be obtained for real s by elementary means, starting from the characterization of Γ as the unique logarithmically convex function satisfying the recursion $\Gamma(s+1) = s\Gamma(s)$ and normalized by $\Gamma(1) = 1$ (the Bohr-Mollerup theorem, see for instance [Rudin 1976, p.193]. The theorem for complex s can then be obtained by analytic continuation. The method used here, though less elegant, generalizes to a construction of product formulas for a much more general class of functions, as we shall see next.

[2]While real asymptotic series cannot in general be differentiated (why?), complex ones can, thanks to Cauchy's integral formula for the derivative. The logarithmic derivative of $\Gamma(s)$ is often called $\psi(s)$ in the literature, but alas we cannot use this notation because it conflicts with $\psi(x) = \sum_{n<x}\Lambda(n)\ldots$

Exercises

On the product formula:

1. Verify that the duplication formula for $\Gamma(2s)$ yields the correct constant term in (4). Apply Euler-Maclaurin to the sum in (3) to show that the $O_\epsilon(|s|^{-1})$ error can be expanded in an asymptotic series in inverse powers of s.

2. Use (1) to obtain a product formula for $\Gamma(s)\Gamma(-s)$, and deduce that

$$\Gamma(s)\Gamma(1-s) = \pi/\sin \pi s. \tag{6}$$

(This can also be obtained from $\Gamma(s)\Gamma(1-s) = \mathrm{B}(s, 1-s)$ by using the change of variable $x = y/(y-1)$ in the Beta integral and evaluating the resulting expression by contour integration.) Use this together with the duplication formula and Riemann's formula for $\zeta(1-s)$ to obtain the equivalent asymmetrical form

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

of the functional equation for $\zeta(s)$. Note that the duplication formula, and its generalization

$$\Gamma(ns) = (2\pi)^{\frac{1-n}{2}} n^{ns-\frac{1}{2}} \prod_{k=0}^{n-1} \Gamma\left(s + \frac{k}{n}\right),$$

can also be obtained from (1).

3. Show that $\log \Gamma(s)$ has the Taylor expansion

$$\log \Gamma(s) = -\boldsymbol{\gamma}(s-1) + \sum_{n=2}^{\infty} \frac{(-1)^n}{n} \zeta(n)(s-1)^n$$

about $s = 1$. Recover from this the Laurent expansion

$$\Gamma(s) = \frac{1}{s} - \boldsymbol{\gamma} + \big(\boldsymbol{\gamma}^2 + \frac{\pi^2}{6}\big)\frac{s}{2} + O(s^2)$$

of $\Gamma(s)$ about $s = 0$.

Behavior of $\Gamma(s)$ on vertical lines:

4. Deduce from (4) that for fixed $\sigma \in \mathbf{R}$

$$\mathrm{Re}\big(\log \Gamma(\sigma + it)\big) = (\sigma - \frac{1}{2}) \log |t| - \frac{\pi}{2}|t| + C_\sigma + O_\sigma(|t|^{-1})$$

as $|t| \to \infty$. Check that for $\sigma = 0, 1/2$ this agrees with the exact formulas

$$|\Gamma(it)|^2 = \frac{\pi}{t \sinh \pi t}, \quad |\Gamma(1/2 + it)|^2 = \frac{\pi}{\cosh \pi t}$$

obtained from (6).

5. For $a, b, c > 0$, determine the Fourier transform of $f(x) = \exp(ax - be^{cx})$, and check your answer by using contour integration to calculate the Fourier transform of $\hat{f}$. Now apply Poisson summation, let $a \to 0$ and $C = e^c > 1$, and describe the behavior of $\sum_{n=0}^{\infty} z^{C^n}$ as $z \to 1$ from below. What does

$$\sum_{n=0}^{\infty} (-1)^n z^{2^n} = z - z^2 + z^4 - z^8 + z^{16} - + \cdots$$

do as $z \to 1$? Use this to prove that $\mathbf{Z} \cap \bigcup_{m=0}^{\infty} \left[2^{2^{2m}}, 2^{2^{2m+1}}\right)$ is an explicit example of a set of integers that does not have a logarithmic density.

An alternative proof of the functional equation for $\zeta(s)$:

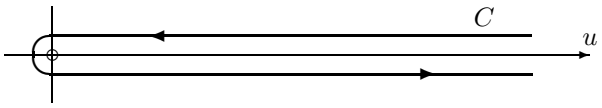
6. Prove that

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} u^{s-1} \frac{du}{e^u - 1}$$

for $\sigma > 1$, and that when s is not a positive integer an equivalent formula is

$$\zeta(s) = -\frac{e^{-\pi i s}}{2\pi i} \Gamma(1-s) \int_C u^{s-1} \frac{du}{e^u - 1}$$

where C is a contour coming from $+\infty$, going counterclockwise around $u = 0$, and returning to $+\infty$:



Show that this gives the analytic continuation of ζ to a meromorphic function on $\mathbf{C}$; shift the line of integration to the left to obtain the functional equation relating $\zeta(s)$ to $\zeta(1-s)$ for $\sigma < 0$, and thus for all s by analytic continuation.

References

[WW 1940] Whittaker, E.T., Watson, G.N.: *A Course of Modern Analysis...*[3] (fourth edition). Cambridge University Press, 1940 (reprinted 1963). [HA 9.40 / QA295.W38]

[Rudin 1976] Rudin, W.: *Principles of Mathematical Analysis* (3rd edition). New York: McGraw-Hill, 1976.

[3]The full title is 26 words long, which was not out of line when the book first appeared in 1902. You can find the title in Hollis.

Math 259: Introduction to Analytic Number Theory

Functions of finite order: product formula and logarithmic derivative

This chapter is another review of standard material in complex analysis. See for instance Chapter 11 of [Davenport 1967], keeping in mind that Davenport uses "integral function" for what we call an "entire function"; Davenport treats only the case of order (at most) 1, which is all that we need, but it is scarcely harder to deal with any finite order as we do here.

The *order* of an entire function $f(\cdot)$ is the smallest $\alpha \in [0, +\infty]$ such that $f(z) \ll_\epsilon \exp|z|^{\alpha+\epsilon}$ for all $\epsilon > 0$. Hadamard showed that entire functions of finite order are given by nice product formulas. We have seen already the cases of $\sin z$ and $1/\Gamma(z)$, both of order 1. As we shall see, $(s^2 - s)\xi(s)$ also has order 1 (as do analogous functions that we'll obtain from Dirichlet L-series). From the product formula for $\xi(s)$ we shall obtain a partial-fraction decomposition of $\zeta'(s)/\zeta(s)$, and will use it to manipulate the contour-integral formula for $\psi(x)$.

Hadamard's product formula for a general entire function of finite order is given by the following result.

Theorem. *Let f be an entire function of order $\alpha < \infty$. Assume that f does not vanish identically on $\mathbf{C}$. Then f has a product formula*

$$f(z) = z^r e^{g(z)} \prod_{k=1}^{\infty} (1 - \frac{z}{z_k}) \exp \sum_{m=1}^{a} \frac{1}{m} \left(\frac{z}{z_k}\right)^m, \tag{1}$$

where $a = \lfloor \alpha \rfloor$, the integer r is the order of vanishing of f at $z = 0$, the z_k are the other zeros of f with multiplicity, g is a polynomial of degree at most a, and the product converges uniformly in bounded subsets of $\mathbf{C}$. Moreover, for $R > 1$ we have

$$\#\{k : |z_k| < R\} \ll_\epsilon R^{\alpha+\epsilon}. \tag{2}$$

Conversely, suppose r is any nonnegative integer, that g is a polynomial of degree at most $a = \lfloor \alpha \rfloor$, and that z_k are nonzero complex numbers such that $|z_k| < R$ for at most $O_\epsilon(R^{\alpha+\epsilon})$ choices of k. Then the right-hand side of (1) defines an entire function of order at most α.

To prove this, we first show:

Lemma. *A function f has finite order and no zeros if and only if $f = e^g$ for some polynomial g.*

Proof: Clearly e^g satisfies the hypotheses if g is a polynomial. Conversely, f is an entire function with no zeros if and only if $f = e^g$ for some entire function g; we shall show that if also $|f| \ll_\epsilon \exp|z|^{\alpha+\epsilon}$ then g is a polynomial. Indeed the real part of g is $< O(|z|^{\alpha+\epsilon})$ for large z. But then the same is true of $|g(z)|$, as the following argument shows. Let $h = g - g(0)$, so $h(0) = 0$; and let $M = \sup_{|z| \le 2R} \operatorname{Re} h(z)$. By assumption $M \ll R^{\alpha+\epsilon}$ for large R. Then $h_1 := h/(2M - h)$ is analytic in the closed disc $D := \{z \in \mathbf{C} : |z| \le 2R\}$,

with $h_1(0) = 0$ and $|h_1(z)| \le 1$ in D. Consider now the analytic function $\phi(z) := 2Rh_1(z)/z$ on D. On the boundary of that disc, $|\phi(z)| \le 1$. Thus by the maximum principle the same is true for all $z \in D$. In particular, if $|z| \le R$ then $|h_1(z)| \le 1/2$. But then $|h(z)| \le 2M$. Hence $|g(z)| \le 2M + g(0) \ll |z|^{\alpha+\epsilon}$ for large $|z|$, and g is a polynomial in z as claimed. Moreover, the degree of that polynomial is just the order of f. $\square$

We shall reduce the Theorem to this Lemma by dividing a given function f of finite order by a product $P(z)$ whose zeros match those of f. To show that this product converges, we first need to obtain the bound (2) on the number of zeros of f in a disc. We shall deduce this bound from *Jensen's inequality* for the function $f_0 = f/z^r$. This inequality states: if f_0 is an analytic function on the disc $|z| \le R$ then

$$|f_0(0)| \le \prod_{\zeta} \frac{|\zeta|}{R} \cdot \sup_{|z|=R} |f(z)|, \tag{3}$$

where the product ranges over the zeros ζ of f_0 in the disc, counted with multiplicity.

Let $z_1, z_2, \ldots$ be the zeros of f_0, listed with the correct multiplicity in non-decreasing order of $|z_k|$:

$$0 < |z_1| \le |z_2| \le |z_3| \le \cdots.$$

For $R > 0$, let $n(R)$ be the left-hand side of (2), which is the number of k such that $|z_k| < R$. Thus $n(R) = k$ if and only if $|z_k| < R < |z_{k+1}|$. Consider first f_0 in $|z| < 1$. Let $\phi(z)$ be the *Blaschke product* $\prod_{k=1}^{n(1)}(z - z_k)/(1 - \bar{z}_k z)$. This is a rational function designed to have the same zeros as f_0 in the unit disc but with $|\phi(z)| = 1$ on $|z| = 1$. Then $f_1 := f_0/\phi$ is analytic on $|z| \le 1$, and $|f(z)| = |f_0(z)| = |f_1(z)|$ on the boundary $|z| = 1$. Therefore by the maximum principle $|f_1(0)| \le \max_{|z|=1} |f(z)|$, so

$$|f_0(0)| = |\phi(0)f_1(0)| = \prod_{k=1}^{n(1)} |z_k| \cdot |f_1(0)| \le \prod_{k=1}^{n(1)} |z_k| \cdot \max_{|z|=1} |f(z)|.$$

Applying this to the function $f_0(Rz)$, whose zeros in the unit disc are z_k/R for $k \le n(R)$, we obtain Jensen's inequality (3). Taking logarithms, we find

$$\begin{aligned} \log \max_{|z|=R} |f(z)| &\ge r \log R + \log |f_0(0)| + \sum_{k=1}^{n(R)} \log \frac{R}{|z_k|} \\ &= r \log R + \log |f_0(0)| + \int_0^R n(r) \frac{dr}{r}. \end{aligned}$$

If f has order at most $\alpha < \infty$ then $\log \max_{|z|=R} |f(z)| \ll_\epsilon R^{\alpha+\epsilon}$, and we conclude that

$$n(R) = \int_R^{eR} n(R) \frac{dr}{r} \le \int_0^{eR} n(r) \frac{dr}{r} \ll_\epsilon R^{\alpha+\epsilon}.$$

We have thus proved (2). It follows that $\sum_{k=1}^{\infty} |z_k|^{-\beta}$ converges if $\beta > \alpha$, since the sum is

$$\int_0^{\infty} r^{-\beta} dn(r) = \beta \int_{|z_1|}^{\infty} r^{-\beta-1} n(r)\, dr \ll \int_{|z_1|}^{\infty} r^{\alpha+\epsilon-\beta-1}\, dr < \infty$$

for any positive $\epsilon < \beta - \alpha$. Therefore the product

$$P(z) := z^r \prod_{k=1}^{\infty} (1 - \frac{z}{z_k}) \exp \sum_{m=1}^{a} \frac{1}{m} \left(\frac{z}{z_k}\right)^m \tag{4}$$

converges for all $z \in \mathbf{C}$, and is not affected by any permutation of the zeros z_k. Moreover, the convergence is uniform in bounded subsets of $\mathbf{C}$, because on $|z| \le R$ we have

$$\log(1 - z/z_k) + \sum_{m=1}^{a} (z/z_k)^m/m \ll (z/z_k)^{a+1} \ll z_k^{-a-1} \tag{5}$$

uniformly once $k > n(2R)$. Therefore $P(z)$ is an entire function, with the same zeros and multiplicities as f.

It follows that f/P is an entire function without zeros. We claim that it too has order at most α, and is thus $\exp g(z)$ for some polynomial g of degree at most a. This would be clear if it were true that

$$\frac{1}{P(z)} \ll_\epsilon \exp |z|^{\alpha+\epsilon},$$

but such an inequality cannot hold for all z due to the zeros of P. But it is enough to show that for each $R > 0$ a bound

$$\frac{1}{P(z)} \ll_\epsilon \exp R^{\alpha+\epsilon}, \tag{6}$$

holds on the circle $|z| = r$ for some $r \in (R, 2R)$, because then we would have $|f(z)/P(z)| \ll_\epsilon \exp R^{\alpha+\epsilon}$ for all z on that circle, and thus also on $|z| = R$ by the maximum principle. We do this next.

Write $P = z^r P_1 P_2$, with P_1, P_2 being the product in (4) over $k \le n(4R)$ and $k > n(4R)$ respectively. We may ignore the factor z^r, whose norm exceeds 1 once $R > 1$. The k-th factor of $P_2(z)$ is $\exp O(|z/z_k|^{a+1})$ by (5), so

$$\log |P_2(z)| \ll R^{a+1} \sum_{k > n(4R)} |z_k|^{-a-1} \ll R^{a+1} \int_{4R}^{\infty} r^{-a-1}\, dn(r) \ll_\epsilon R^{\alpha+\epsilon},$$

using integration by parts and $n(r) \ll_\epsilon r^{\alpha+\epsilon}$ in the last step (check this!). As to P_1, it is a finite product, which we write as $e^{h(z)} \prod_{k \le n(4R)} (1 - z/z_k)$, where $h(z)$ is the polynomial

$$h(z) = \sum_{k=1}^{n(4R)} \sum_{m=1}^{a} \frac{1}{m} \left(\frac{z}{z_k}\right)^m$$

of degree at most a. Thus $h(z) \ll R^a \sum_{k \le n(4R)} |z_k|^{-a}$, which readily yields $h(z) \ll R^{\alpha+\epsilon}$. (Again you should check this by carrying out the required partial summation and estimates; note too that the upper bounds on the *absolute value* of $\log|P_2(z)|$ and $h(z)$ yield lower as well as upper bounds on $|P_2(z)|$ and $|\exp h(z)|$.) So far, our lower bounds on the factors of $P(z)$ hold for all z in the annulus $R < |z| < 2R$, but we cannot expect the same for $P_3(z) := \prod_{k \le n(4R)}(1 - z/z_k)$, since it may vanish at some points of the annulus. However, we can prove that *some* r works by estimating the average[1]

$$-\frac{1}{R}\int_R^{2R} \min_{|z|=r} \log|P_3(z)|\, dr \le -\sum_{k=1}^{n(4R)} \frac{1}{R}\int_R^{2R} \log\Big|1 - \frac{r}{|z_k|}\Big|\, dr.$$

The integral is elementary, if not pretty, and at the end we conclude that the average is again $\ll R^{\alpha+\epsilon}$. This shows that for some $r \in (R, 2R)$ the desired lower bound holds, and we have finally proved the product formula (1).

To complete the proof of our Theorem we need only show the converse: (1) converges to an entire function of order at most α under the stated hypotheses on r, g, z_k. The convergence was proved already, and the upper bound on $|f(z)|$ follows readily from (5). □□

Taking logarithmic derivatives in (1), we deduce

$$\frac{f'}{f}(z) = g'(z) + \frac{P'}{P}(z) = g'(z) + \frac{r}{z} + \sum_{k=1}^{\infty}\Big[\frac{1}{z - z_k} + \sum_{m=1}^{a} \frac{z^{m-1}}{z_k^m}\Big]$$

$$= g'(z) + \frac{r}{z} + \sum_{k=1}^{\infty} \frac{(z/z_k)^a}{z - z_k}\,.$$

We note too that if $\alpha > 0$ and $\sum_k |z_k|^{-\alpha} < \infty$ then there exists a constant C such that $f(z) \ll \exp C|z|^\alpha$. This follows from the existence of a constant C_α such that

$$\Big|(1-w)\exp\sum_{m=1}^{a} w^m/m\Big| \ll \exp C_\alpha |w|^\alpha$$

for all $w \in \mathbf{C}$. Contrapositively, if $f(z)$ is a function of order α which grows faster than $\exp C|z|^\alpha$ for all C then $\sum_k |z_k|^{-\alpha}$ diverges. For instance this happens for $f(s) = 1/\Gamma(s)$. [This approach may appear circular because it is proved from the product formula for $\Gamma(s)$, but it need not be; see Exercise 6 below.] As we shall see, the same is true for $f(s) = (s^2 - s)\xi(s)$; it will follow that ξ, and thus ζ, has infinitely many nontrivial zeros ρ with real part in $[0, 1]$, and in fact that $\sum_\rho |\rho|^{-1}$ diverges.

[1] This averaging trick is a useful technique that we'll encounter again several times; it is closely related to the "probabilistic method" in combinatorics, in which an object with some property is proved to exist by showing that the property holds with positive probability.

Exercises

1. The bound $f(z) \ll \exp C|z|^\alpha$ was proved under the hypothesis $\alpha > 0$. Is this hypothesis necessary?

2. Find an entire function $f(z)$ of order 1 such that $|f(z)| \ll \exp O(|z|)$ but $\sum_{k=1}^\infty |z_k^{-1}| = \infty$. [Hint: you don't have to look very far.]

3. Supply the missing steps in our proof of (1).

4. Suppose z_k $(k = 1, 2, 3, \ldots)$ are distinct complex numbers with $0 < |z_k| < 1$, and m_k are some positive integers. Prove that $\prod_k |z_k|^{m_k} > 0$ if and only if there exists a *bounded* nonzero analytic function $f \not\equiv 0$ on the open disc $|z| < 1$ with a root at each z_k of multiplicity m_k.

5. Prove Jensen's formula: if f is an analytic function on $|z| \le R$ such that $f(0) \ne 0$ then $(2\pi)^{-1} \int_0^{2\pi} \log|f(Re^{i\theta})| d\theta = \log|f(0)| + \sum_k \log(R/|z_k|)$, where the z_k are the zeros of f in $|z| \le R$ with the correct multiplicities. What is $(2\pi)^{-1} \int_0^{2\pi} \log|f(Re^{i\theta})| d\theta$ if $f(0) = 0$ but f does not vanish identically?

6. Show that $1/\Gamma(s)$ is an entire function of order 1, using only the following tools available to Euler: the integral formulas for $\Gamma(s)$ and $\mathrm{B}(s, s')$, and the identities $B(s, s') = \Gamma(s)\Gamma(s')/\Gamma(s+s')$ and $\Gamma(s)\Gamma(1-s) = \pi/\sin \pi s$. [The hard part is getting an upper bound for $1/|\Gamma(s)|$ on a vertical strip; remember how we showed that $\Gamma(s) \ne 0$, and use the formula for $|\Gamma(1/2 + it)|^2$ to get a better lower bound on $|\Gamma(s)|$.] Use this to recover the product formula for $\Gamma(s)$, up to a factor e^{A+Bs} which may be determined from the behavior of $\Gamma(s)$ at $s = 0, 1$.

7. Prove that if $f(z)$ is an entire function of order $\alpha > 0$ then

$$\iint_{|z|<r} |f'(z)/f(z)| \, dx \, dy \ll r^{\alpha+1+\epsilon} \qquad (z = x + iy)$$

as $r \to \infty$. [Note that the integral is improper (except in the trivial case that f has no zeros) but still converges: if ϕ is a meromorphic function on a region $U \subset \mathbf{C}$ with simple but no higher-order poles then $|\phi|$ is integrable on compact subsets $K \subset U$, even K that contain poles of ϕ.]

Reference

[Davenport 1967] Davenport, H.: *Multiplicative Number Theory.* Chicago: Markham, 1967; New York: Springer-Verlag, 1980 (GTM 74). [9.67.6 & 9.80.6 / QA 241.D32]

Math 259: Introduction to Analytic Number Theory

The product formula for $\xi(s)$ and $\zeta(s)$; vertical distribution of zeros

Behavior on vertical lines. We next show that $(s^2 - s)\xi(s)$ is an entire function of order 1; more precisely:

Lemma. *There exists a constant C such that $(s^2 - s)\xi(s) \ll \exp(C|s| \log |s|)$, but no constant C' such that $(s^2 - s)\xi(s) \ll \exp(C'|s|)$.*

Proof: By the functional equation $\xi(s) = \xi(1 - s)$, it is enough to consider $s = \sigma + it$ with $\sigma \geq 1/2$. From Stirling it follows that for fixed $\sigma \in \mathbf{R}$

$$\mathrm{Re}\big(\log \Gamma(\sigma + it)\big) = (\sigma - \frac{1}{2}) \log |t| - \frac{\pi}{2}|t| + C_\sigma + O_\sigma(|t|^{-1}).$$

For $\sigma > 1$, the Euler product for $\zeta(s)$ shows that $\log |\zeta(\sigma + it)| = O_\sigma(1)$; indeed we have the upper and lower bounds

$$\zeta(\sigma) \geq |\zeta(\sigma + it)| > \prod_p (1 + p^{-s})^{-1} = \zeta(2\sigma)/\zeta(\sigma).$$

Hence $|\xi(\sigma + it)|$ is within a constant factor of $|t|^{(\sigma-1)/2} e^{-\pi|t|/4}$ for large $|t|$. This estimate on $|\xi(\sigma + it)|$ already proves that $|(s^2 - s)\xi(s)|$ grows faster than $\exp(C'|s|)$ for any C'; together with the functional equation, it also shows that for each $\sigma < 0$ there exists C_σ such that $|\zeta(\sigma + it)|$ is within a factor of C_σ of $|t|^{1/2-\sigma}$ for large $|t|$.

To prove our Lemma, it remains to bound $\zeta(s)$ for s in or near the critical strip. Generalizing our formula for analytically continuing $\zeta(s)$, we find for $\sigma > 0$

$$\zeta(s) = \sum_{n=1}^{N-1} n^{-s} + \frac{N^{1-s}}{s-1} + \sum_{n=N}^{\infty} \int_n^{n+1} (n^{-s} - x^{-s})\, dx,$$

which for large t, N is $\ll N^{1-\sigma} + |t|N^{-\sigma}$, uniformly at least for $\sigma \geq 1/2$. Taking $N = |t| + O(1)$ we find $\zeta(\sigma + it) \ll |t|^{1-\sigma}$ for $\sigma \geq 1/2$, $|t| > 1$. Together with Stirling's approximation, this completes the proof of our Lemma. $\square$

A remark about our choice of $N \sim |t|$ in the bound $\zeta(\sigma + it) \ll N^{1-\sigma} + |t|N^{-\sigma}$: of course we wanted to choose N to make the bound as good as possible, i.e., to minimize $N^{1-\sigma} + |t|N^{-\sigma}$. In calculus we learned to do this by setting the derivative equal to zero. That would give N proportional to $|t|$, but we arbitrarily set the constant of proportionality to 1 even though another choice would make $N^{1-\sigma} + |t|N^{-\sigma}$ slightly smaller. In general when we bound some quantity by a sum $O(f(N) + g(N))$ of an increasing and a decreasing function of some parameter N, we shall simply choose N so that $f(N) = g(N)$ (or, if N is constrained to be an integer, so that $f(N), g(N)$ are nearly equal). This is much simpler and less error-prone than fumbling with derivatives, and is sure to give the minimum to within a factor of 2, which is good enough when we're dealing with $O(\cdots)$ bounds.

Product and logarithmic-derivative formulas. By our general product formula for an entire function of finite order we know that $\xi(s)$ has a product expansion:

$$\xi(s) = \frac{e^{A+Bs}}{s^2 - s} \prod_{\rho} (1 - s/\rho) e^{s/\rho}, \tag{1}$$

for some constants A, B, with the product ranging over zeros ρ of ξ (that is, the nontrivial zeros of ζ) listed with multiplicity. Moreover, $\sum_\rho |\rho|^{-1-\epsilon} < \infty$ for all $\epsilon > 0$ but $\sum_\rho |\rho|^{-1} = \infty$. The logarithmic derivative of (1) is

$$\frac{\xi'}{\xi}(s) = B - \frac{1}{s} - \frac{1}{s-1} + \sum_{\rho} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right); \tag{2}$$

since $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ we also get a product formula for $\zeta(s)$, and a partial-fraction expansion of its logarithmic derivative:

$$\frac{\zeta'}{\zeta}(s) = B - \frac{1}{s-1} + \frac{1}{2}\log\pi - \frac{1}{2}\frac{\Gamma'}{\Gamma}(\frac{s}{2} + 1) + \sum_{\rho} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right). \tag{3}$$

(We have shifted from $\Gamma(s/2)$ to $\Gamma(s/2+1)$ to absorb the term $-1/s$; note that $\zeta(s)$ does not have a pole or zero at $s = 0$.)

Vertical distribution of zeros. Since the zeros ρ of $\xi(s)$ are limited to a strip we can find much more precise information about the distribution of their sizes than the convergence and divergence of $\sum_\rho |\rho|^{-1-\epsilon}$ and $\sum_\rho |\rho|^{-1}$. Let $N(T)$ be the number of zeros in the rectangle $\sigma \in [0,1]$, $t \in [0,T]$ — which is very nearly half of what we would call $n(T)$ in the context of the general product formula for $(s^2 - s)\xi(s)$.

Theorem (von Mangoldt)**.** *As $T \to \infty$,*

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T). \tag{4}$$

Proof: We follow chapter 15 of [Davenport 1967], keeping track of the fact that Davenport's ξ and ours differ by a factor of $(s^2 - s)/2$.

We may assume that T does not equal the imaginary part of any zero of $\zeta(s)$. Then

$$2N(T) - 2 = \frac{1}{2\pi i}\oint_{C_R} \frac{\xi'}{\xi}(s)\, ds = \frac{1}{2\pi i}\oint_{C_R} d(\log\xi(s)) = \frac{1}{2\pi}\oint_{C_R} d(\operatorname{Im}\log\xi(s)),$$

where C_R is the boundary of the rectangle $\sigma \in [-1,2]$, $t \in [-T,T]$. Since $\xi(s) = \xi(1-s) = \overline{\xi(\bar{s})}$, we may by symmetry evaluate the last integral by integrating over a quarter of C_R and multiplying by 4. We use the top right quarter, going from 2 to $2 + iT$ to $1/2 + iT$. At $s = 2$, $\log\xi(s)$ is real, so we have

$$\pi(N(T)-1) = \operatorname{Im}\log\xi(\frac{1}{2}+iT) = \operatorname{Im}(\log\Gamma(\frac{1}{4}+\frac{iT}{2})) - \frac{T}{2}\log\pi + \operatorname{Im}(\log\zeta(\frac{1}{2}+iT)).$$

By Stirling, the first term is within $O(T^{-1})$ of

$$\operatorname{Im}\left((\frac{iT}{2}-\frac{1}{4})\log(\frac{iT}{2}+\frac{1}{4})\right)-\frac{T}{2}$$

$$=\frac{T}{2}\log\left|\frac{iT}{2}+\frac{1}{4}\right|-\frac{1}{4}\operatorname{Im}\log(\frac{iT}{2}+\frac{1}{4})-\frac{T}{2}=\frac{T}{2}\big(\log\frac{T}{2}-1\big)+O(1).$$

Thus (4) is equivalent to

$$\operatorname{Im}\log\zeta(\frac{1}{2}+iT)\ll\log T. \tag{5}$$

We shall show that for $s=\sigma+it$ with $\sigma\in[-1,2]$, $|t|>1$ we have

$$\frac{\zeta'}{\zeta}(s)=\sum_{|\operatorname{Im}(s-\rho)|<1}\frac{1}{s-\rho}+O(\log|t|), \tag{6}$$

and that the sum comprises at most $O(\log|t|)$ terms, from which our desired estimate will follow by integrating from $s=2+iT$ to $s=1/2+iT$. We start by taking $s=2+it$ in (3). At that point the LHS is uniformly bounded (use the Euler product) and the RHS is

$$\sum_\rho\left(\frac{1}{2+it-\rho}+\frac{1}{\rho}\right)+O(\log|t|)$$

by Stirling. Thus the sum, and in particular its real part, is $O(\log|t|)$. But each summand has positive real part, which is at least $1/(4+(t-\operatorname{Im}\rho)^2)$. Our second claim, that $|t-\operatorname{Im}\rho|<1$ holds for at most $O(\log|t|)$ zeros ρ, follows immediately. It also follows that

$$\sum_{|\operatorname{Im}(s-\rho)|\geq1}\frac{1}{\operatorname{Im}(s-\rho)^2}\ll\log|t|.$$

Now by (3) we have

$$\frac{\zeta'}{\zeta}(s)-\frac{\zeta'}{\zeta}(2+it)=\sum_\rho\Big(\frac{1}{s-\rho}-\frac{1}{2+it-\rho}\Big)+O(1).$$

The LHS differs from that of (6) by $O(1)$, as noted already; the RHS summed over zeros with $|\operatorname{Im}(s-\rho)|<1$ is within $O(\log|t|)$ of the RHS of (6); and the remaining terms are

$$(2-\sigma)\sum_{|\operatorname{Im}(s-\rho)|\geq1}\frac{1}{(s-\rho)(2+it-\rho)}\ll\sum_{|\operatorname{Im}(s-\rho)|\geq1}\frac{1}{\operatorname{Im}(s-\rho)^2}\ll\log|t|.$$

This proves (6) and thus also (5); von Mangoldt's theorem (4) follows. □

For much more about the vertical distribution of the nontrivial zeros ρ of $\zeta(s)$ see [Titchmarsh 1951], Chapter 9.

Remarks

In our proof of the product formula for $\xi(s)$ we showed that for each σ there exists ν such that $|\zeta(\sigma+it)| \ll |t|^\nu$ as $|t|\to\infty$. This was more than enough to prove that $(s^2-s)\xi(s)$ has order 1, but one may naturally ask how small ν can become. Let $\mu(\sigma)$ be the infimum of all such ν; that is,

$$\mu(\sigma) := \limsup_{|t|\to\infty} \frac{\log|\zeta(\sigma+it)|}{\log|t|}.$$

We have seen that $\mu(\sigma) = 0$ for $\sigma > 1$, that $\mu(1-\sigma) = \mu(\sigma) + \sigma - \frac{1}{2}$ by the functional equation (so in particular $\mu(\sigma) = \frac{1}{2} - \sigma$ for $\sigma < 0$), and that $\mu(\sigma) \le 1-\sigma$ for $\sigma < 1$. For $\sigma \in (0,1)$ one can improve on these bounds using the "approximate functional equation" for $\zeta(s)$ (usually attributed to Siegel, but now known to have been used by Riemann himself) to show that $\mu(\sigma) \le (1-\sigma)/2$; this result, and the fact that $\mu(\sigma) \ge 0$ for all σ, also follows from general results in complex analysis, which indicate that since $\mu(\sigma)$ is finite for all σ, the function $\mu(\cdot)$ must be convex. For example, $\mu(1/2) \le 1/4$, so $|\zeta(\frac{1}{2}+it)| \ll_\epsilon |t|^{\frac{1}{4}+\epsilon}$.

The value of $\mu(\sigma)$ is not known for any $\sigma \in (0,1)$. The *Lindelöf conjecture* asserts that $\mu(1/2) = 0$, from which it would follow that $\mu(\sigma) = 0$ for all $\sigma \ge 1/2$ while $\mu(\sigma) = \frac{1}{2} - \sigma$ for all $\sigma \le 1/2$. Equivalently, the Lindelöf conjecture asserts that $\zeta(\sigma+it) \ll_\epsilon |t|^\epsilon$ for all $\sigma \ge 1/2$ (excluding a neighborhood of the pole $s=1$), and thus by the functional equation that also $\zeta(\sigma+it) \ll_\epsilon |t|^{1/2-\sigma+\epsilon}$ for all $\sigma \le 1/2$. We shall see that this conjecture is implied by the Riemann hypothesis, and also that it holds on average in the sense that $\int_0^T |\zeta(\frac{1}{2}+it)|^2\,dt \ll T^{1+\epsilon}$. However, the best upper bound currently proved on $\mu(1/2)$ is only a bit smaller than $1/6$; when we get to exponential sums later this term we shall derive the upper bound of $1/6$.

Exercises

1. Show that in the product formula (1) we may take $A = 0$. Prove the formula

$$\gamma = \lim_{s\to 1}\Big(\zeta(s) - \frac{1}{s-1}\Big)$$

for Euler's constant, and use it to compute

$$B = \lim_{s\to 0}\Big(\frac{\xi'}{\xi}(s) + \frac{1}{s}\Big) = \lim_{s\to 1}\Big(\frac{\xi'}{\xi}(s) + \frac{1}{1-s}\Big)$$

$$= \frac{1}{2}\log 4\pi - 1 - \frac{\gamma}{2} = -0.0230957\ldots.$$

Show also (starting by pairing the ρ and $\bar\rho$ terms in the infinite product) that

$$B = -\sum_\rho \mathrm{Re}(\rho)/|\rho|^2,$$

and thus that $|\operatorname{Im}(\rho)| > 6$ for every nontrivial zero ρ of $\zeta(s)$. [From [Davenport 1967], Chapter 12. It is known that in fact the smallest zeros have (real part 1/2 and) imaginary part $\pm 14.134725\ldots$]

2. Prove the alternative infinite product

$$\xi(s) = \frac{\xi(1/2)}{4(s-s^2)} \prod_{\rho}{}^{+} \left[1 - \left(\frac{s-1/2}{\rho-1/2}\right)^2\right],$$

the product extending over zeros ρ of ξ whose imaginary part is positive.

3. Let f be any analytic function on the vertical strip $a < \sigma < b$ such that

$$M_f(\sigma) := \limsup_{|t|\to\infty} \frac{\log|f(\sigma+it)|}{\log|t|}$$

is finite for all $\sigma \in (a,b)$. Prove that M_f is a convex function on that interval. [Hint: Apply the maximum principle to αf for suitable analytic functions $\alpha(s)$.]

It follows in particular that M_f is continuous on (a,b). While $\zeta(s)$ is not analytic on vertical strips that contain $s = 1$, we can still deduce the convexity of $\mu : \mathbf{R}\to\mathbf{R}$ from $\mu(\sigma) = M_f(\sigma)$ for $f(s) = \zeta(s) - (1/(s-1))$.

Much the same argument proves the "three lines theorem": if f is actually bounded on the strip then $\log\sup_t |f(\sigma+it)|$ is a convex function of σ. The name of this theorem alludes to the equivalent formulation: if $a < \sigma_1 < \sigma_2 < \sigma_3 < b$ then the supremum of $|f(s)|$ on the line $s = \sigma_2 + it$ is bounded by a weighted geometric mean of its suprema on the lines $s = \sigma_1 + it$ and $s = \sigma_3 + it$.

Reference

[Titchmarsh 1951] Titchmarsh, E.C.: *The Theory of the Riemann Zeta-Function.* Oxford: Clarendon, 1951. [HA 9.51.14 / QA351.T49; 2nd ed. revised by D.R. Heath-Brown 1986, QA246.T44]

Math 259: Introduction to Analytic Number Theory

A zero-free region for $\zeta(s)$

We first show, as promised, that $\zeta(s)$ does not vanish on $\sigma = 1$. As usual nowadays, we give Mertens' elegant version of the original arguments of Hadamard and (independently) de la Vallée Poussin. Recall that

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

has a simple pole at $s = 1$ with residue $+1$. If $\zeta(s)$ were to vanish at some $1 + it$ then $-\zeta'/\zeta$ would have a simple pole with residue -1 (or $-2, -3, \ldots$) there. The idea is that $\sum_n \Lambda(n)/n^s$ converges for $\sigma > 1$, and as s approaches 1 from the right all the terms contribute towards the positive-residue pole. As $\sigma \to 1 + it$ from the right, the corresponding terms have the same magnitude but are multiplied by n^{-it}, so a pole with residue -1 would force "almost all" the phases n^{-it} to be near -1. But then near $1 + 2it$ the phases n^{-2it} would again approximate $(-1)^2 = +1$, yielding a pole of positive residue, which is not possible because then ζ would have another pole besides $s = 1$.

To make precise the idea that if $n^{-it} \approx -1$ then $n^{-2it} \approx +1$, we use the identity

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

from which it follows that the right-hand side is positive. Thus if $\theta = t \log n$ we have

$$3 + 4\,\mathrm{Re}(n^{-it}) + \mathrm{Re}(n^{-2it}) \geq 0.$$

Multiplying by $\Lambda(n)/n^\sigma$ and summing over n we find

$$3\left[-\frac{\zeta'}{\zeta}(\sigma)\right] + 4\,\mathrm{Re}\left[-\frac{\zeta'}{\zeta}(\sigma + it)\right] + \mathrm{Re}\left[-\frac{\zeta'}{\zeta}(\sigma + 2it)\right] \geq 0 \qquad (1)$$

for all $\sigma > 1$ and $t \in \mathbf{R}$. Fix $t \neq 0$. As $\sigma \to 1+$, the first term in the LHS of this inequality is $3/(\sigma - 1) + O(1)$, and the remaining terms are bounded below. If ζ had a zero of order $r > 0$ at $1 + it$, the second term would be $-4r/(\sigma - 1) + O(1)$. Thus the inequality yields $4r \leq 3$. Since r is an integer, this is impossible, and the proof is complete.

We next use (1), together with the partial-fraction formula

$$-\frac{\zeta'}{\zeta}(s) = \frac{1}{s-1} + B_1 + \frac{1}{2}\frac{\Gamma'}{\Gamma}(\frac{s}{2} + 1) - \sum_\rho \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right),$$

to show that even the existence of a zero close to $1 + it$ is not possible. How close depends on t; specifically, we show:[1]

[1] See for instance Chapter 13 of Davenport's book [Davenport 1967] cited earlier. This classical bound has been improved; the current record of $1 - \sigma \ll \log^{-2/3-\epsilon} |t|$, due to Korobov and perhaps Vinogradov, has stood for 40 years. See [Walfisz 1963] or [Montgomery 1971, Chapter 11].

Theorem. *There is a constant $c > 0$ such that if $|t| > 2$ and $\zeta(\sigma + it) = 0$ then*

$$\sigma < 1 - \frac{c}{\log |t|}. \tag{2}$$

Proof: Let $\sigma \in [1, 2]$ and[2] $|t| \geq 2$ in the partial-fraction formula. Then the B_1 and Γ'/Γ terms are $O(\log |t|)$, and each of the terms $1/(s - \rho)$, $1/\rho$ has positive real part as noted in connection with von Mangoldt's theorem on $N(T)$. Therefore[3]

$$-\operatorname{Re} \frac{\zeta'}{\zeta}(\sigma + 2it) < O(\log |t|),$$

and if some $\rho = 1 - \delta + it$ then

$$-\operatorname{Re} \frac{\zeta'}{\zeta}(\sigma + 2it) < O(\log |t|) - \frac{1}{\sigma + \delta - 1}.$$

Thus (1) yields

$$\frac{4}{\sigma + \delta - 1} < \frac{3}{\sigma - 1} + O(\log |t|).$$

In particular, taking[4] $\sigma = 1 + 4\delta$ yields $1/20\delta < O(\log |t|)$. Hence $\delta \gg (\log |t|)^{-1}$, and our claim (2) follows. $\square$

Once we obtain the functional equation and partial-fraction decomposition for Dirichlet L-functions $L(s, \chi)$, the same argument will show that (2) also gives a zero-free region for $L(s, \chi)$, though with the implied constant depending on χ.

Remarks

The only properties of $\Lambda(n)$ that we used in the proof of $\zeta(1 + it) \neq 0$ are that facts that $\Lambda(n) \geq 0$ for all n and that $\sum_n \Lambda(n)/n^s$ has an analytic continuation with a simple pole at $s = 1$ and no other poles of real part ≥ 1. Thus the same argument exactly will show that $\prod_{\chi \bmod q} L(s, \chi)$, and thus each of the factors $L(s, \chi)$, has no zero on the line $\sigma = 1$.

The $3 + 4\cos\theta + \cos 2\theta$ trick is worth remembering, since it has been adapted to other uses. For instance we shall revisit and generalize it when we develop the Drinfeld-Vlăduţ upper bounds on points of a curve over a finite field and the Odlyzko-Stark lower bounds on discriminants of number fields. See also the following Exercises.

[2] Any lower bound > 1 would do — and the only reason we cannot go lower is that our bounds are in terms of $\log |t|$ so we do not want to allow $\log |t| = 0$.

[3] Note that we write $< O(\log |t|)$, not $= O(\log |t|)$, to allow the possibility of an arbitrarily large *negative* multiple of $|\log t|$.

[4] $1 + \alpha\delta$ will do for any $\alpha > 3$. This requires that $\alpha\delta \leq 1$, for instance $\delta \leq 1/4$ for our choice of $\alpha = 4$, else $\sigma > 2$; but we're concerned only with δ near zero, so this does not matter.

Exercises

1. Use the inequality $3 + 4\cos\theta + \cos 2\theta \geq 0$ to give an alternative proof that $L(1, \chi) \neq 0$ when χ is a complex Dirichlet character (a character such that $\chi \neq \overline{\chi}$).

2. Show that for each $\alpha > 2$ there exists $t \in \mathbf{R}$ such that

$$\int_{-\infty}^{\infty} \exp(-|x|^{\alpha} + itx)\, dx < 0.$$

(Yes, this is related to the present topic; see [EOR 1991, p.633]. The integral is known to be positive for all $t \in \mathbf{R}$ when $\alpha \in (0, 2]$; see for instance [EOR 1991, Lemma 5].)

References

[EOR 1991] Elkies, N.D., Odlyzko, A., Rush, J.A.: On the packing densities of superballs and other bodies, Invent. Math. 105 (1991), 613–639.

[Montgomery 1971] Montgomery, H.L.: *Topics in Multiplicative Number Theory.* Berlin: Springer, 1971. [LNM **227** / QA3.L28 #227]

[Walfisz 1963] Walfisz, A.: *Weylsche Exponentialsummen in der neueren Zahlentheorie.* Berlin: Deutscher Verlag der Wissenschaften, 1963. [AB 9.63.5 / Sci 885.110(15,16)]

Math 259: Introduction to Analytic Number Theory

Proof of the Prime Number Theorem;
the Riemann Hypothesis

We finally have all the ingredients that we need to assemble a proof of the Prime Number Theorem with an explicit error bound. We shall give an upper bound on $|(\psi(x)/x) - 1|$ that decreases faster than any power of $1/\log x$ as $x \to \infty$, though slower than any positive power of $1/x$. Specifically, we show:

Theorem. *There exists an effective constant $C > 0$ such that*

$$\psi(x) = x + O(x \exp(-C\sqrt{\log x})) \tag{1}$$

for all $x \geq 1$.

Proof: There is no difficulty with small x, so we may and shall assume that $x \geq e$, so $\log x \geq 1$. We use our integral approximation

$$\psi(x) = \frac{1}{2\pi i} \int_{1+\frac{1}{\log x}-iT}^{1+\frac{1}{\log x}+iT} -\frac{\zeta'}{\zeta}(s)\, x^s \frac{ds}{s} + O\left(\frac{x \log^2 x}{T}\right) \qquad (T \in [1, x]) \tag{2}$$

to $\psi(x)$. Assume that $T \geq e$, and that T does not coincide with the imaginary part of any ρ. Shifting the line of integration leftwards, say to real part -1, yields

$$\psi(x) - \left(x - \sum_{|\operatorname{Im}(\rho)|<T} \frac{x^\rho}{\rho}\right) = I_1 + I_2 - \frac{\zeta'}{\zeta}(0) + O\left(\frac{x \log^2 x}{T}\right), \tag{3}$$

in which I_1, I_2 are the integrals of $-(\zeta'(s)/\zeta(s))x^s\, ds/s$ over the vertical line $\sigma = -1, |t| < T$ and the horizontal lines $\sigma \in [-1, 1 + 1/\log x]$, $t = \pm T$ respectively. We next show that I_1 is small, and that I_2 can be made small by adding $O(1)$ to T. The vertical integral I_1 is clearly

$$\ll \frac{\log T}{x} \sup_{|t|<T} \left|\frac{\zeta'}{\zeta}(-1+it)\right| \ll \frac{\log^2 T}{x}.$$

The horizontal integrals in I_2 are

$$\ll \frac{1}{T} \int_{-1}^{1+\frac{1}{\log x}} x^\sigma\, d\sigma \cdot \sup_{\sigma \in [-1,2]} \left|\frac{\zeta'}{\zeta}(\sigma + iT)\right|.$$

The σ integral is $\ll x/\log x$. We have seen already that for $s = \sigma + iT$ and $-1 \leq \sigma \leq 2$ we have

$$\zeta'(s)/\zeta(s) = \sum_{|T - \operatorname{Im}\rho|<1} \frac{1}{s-\rho} + O(\log T),$$

in which the sum has $O(\log T)$ terms. Since the number of $\operatorname{Im}\rho$ in the interval $[T-1, T+1]$ is $\ll \log T$, some point in the middle half of that interval is at distance $\gg 1/\log T$ from all of them; choosing that as our new value of T, we see that each term is $\ll \log T$, and thus that the sum is $\ll \log^2 T$. In conclusion, then,

$$I_2 \ll x \log^2 T/T \log x.$$

Better estimates can be obtained (we could save a factor of $\log T$ by averaging over $[T-\frac{1}{2}, T+\frac{1}{2}]$), but are not necessary because $x\log^2 T/T\log x$ is already less than the error $(x\log^2 x)/T$ in (2).

Thus the RHS of (3) may be absorbed into the $O((x\log^2 x)/T)$ error. In the LHS, we use our zero-free region, that is, the lower bound

$$1-\sigma > c/\log|t|, \tag{4}$$

to find that

$$|x^\rho| = x^{\operatorname{Re}(\rho)} \ll x^{1-\frac{c}{\log T}} = x\exp\Big(-c\,\frac{\log x}{\log T}\Big).$$

Since[1]

$$\sum_{|\operatorname{Im}(\rho)|<T} \frac{1}{|\rho|} < \sum_{|\operatorname{Im}(\rho)|<T} \frac{1}{|\operatorname{Im}\rho|}$$

$$= 2\int_1^T \frac{dN(t)}{t} = \frac{2N(T)}{T} + 2\int_1^T \frac{N(t)\,dt}{t^2} \ll \log T + \int_1^T \frac{\log t\,dt}{t} \ll \log^2 T,$$

we thus have

$$\sum_{|\rho|<T} \frac{x^\rho}{\rho} \ll x\log^2 T \exp\Big(-c\,\frac{\log x}{\log T}\Big).$$

Therefore

$$\left|\frac{\psi(x)}{x} - 1\right| \ll \left(\frac{1}{T} + \exp\big(-c\frac{\log x}{\log T}\big)\right)\log^2 x.$$

We choose T so that the logarithms $-\log T$, $-\log x/\log T$ of the two terms $1/T$, $\exp(-c\log x/\log T)$ are equal. That is, we take $T = \exp\sqrt{\log x}$. Then both terms are $O(\exp(-C_1\log^{1/2} x))$ for some $C_1 > 0$. We then absorb the factor $\log^2 x$ into this estimate by changing C_1 to any positive $C < C_1$, and at last complete the proof of (1). □□

The equivalent result for $\pi(x)$ follows by partial summation:

Corollary. *There exists an effective constant $C > 0$ such that*

$$\pi(x) = \operatorname{li}(x) + O(x\exp(-C\sqrt{\log x})).$$

for all $x \geq 1$.

[1] We can use $\int_1^T$ because we have shown that there are no complex zeros ρ with $|\operatorname{Im}(\rho)| \leq 1$. If there were such zeros, we could absorb their terms x^ρ/ρ into the error estimate. We shall do this in the proof of the corresponding estimates on $\psi(x,\chi)$.

[Recall that $\mathrm{li}(x)$ is the principal value of $\int_0^x dy/\log y$, whence

$$\mathrm{li}(x) = \int_2^x dy/\log y + O(1) = x/\log x + O(x/\log^2 x).]$$

Proof: We have seen already that

$$\pi(x) = \frac{\psi(x)}{\log x} + \int_2^x \psi(y)\frac{dy}{y\log^2 y} + O(x^{1/2}). \tag{5}$$

On the other hand, integration by parts yields

$$\mathrm{li}(x) = \frac{x}{\log x} - \int_2^x y\, d(1/\log y) + O(1) = \frac{x}{\log x} + \int_2^x \frac{dy}{\log^2 y} + O(1).$$

The Corollary now follows from (1). □

The Riemann Hypothesis and some consequences

The error estimate in (1), while sufficient to prove the Prime Number Theorem, is not nearly as strong as one might wish. The growth rate of $|\psi(x) - x|$ and $|\pi(x) - \mathrm{li}(x)|$ hinges on the *Riemann Hypothesis (RH)*, which we introduce next.

The RH and its generalizations are arguably the most important open problems in mathematics. We shall see and explore some of these generalizations later. The original RH is Riemann's inspired guess that *all the nontrivial zeros of* $\zeta(s)$ *have real part equal to* $1/2$, i.e., lie on the critical line $\sigma = 1/2$ at the center of the critical strip. At the time there was scant evidence for the conjecture: the symmetry of the zeros with respect to the critical line, and also numerical computations of the first few zeros (not reported in Riemann's memoir but found among his papers after his death). The conjecture is now supported by a wealth of numerical evidence, as well as compelling analogies with "geometrical" zeta functions for which the conjecture has been proved — notably the zeta functions of varieties over finite fields, for which the RH was proved by Hasse [1936] (elliptic curves), Weil [1940, 1941, 1948] (arbitrary curves and abelian varieties), and Deligne (the general case). These analogies also suggest that proving the "arithmetical" RH and its generalizations will involve fundamental new insights in number theory, quite beyond the immediate applications to the distribution of primes and related arithmetical functions. For now we content ourselves with the most direct connections between the RH and the error estimate in the Prime Number Theorem.

If the RH holds then we may take $T = x$ in (3) to find $\psi(x) = x + O(x^{1/2}\log^2 x)$. More generally:

Proposition. *Suppose there exists* θ *with* $1/2 \le \theta < 1$ *such that* $\mathrm{Re}\,\rho \le \theta$ *for all zeros* ρ *of* ζ. *Then* $\psi(x) = x + O(x^\theta \log^2 x)$ *and* $\pi(x) = \mathrm{li}(x) + O(x^\theta \log x)$ *for large* x.

Proof: Take $T = x + O(1)$ in (3). By our bounds on I_1, I_2, the right-hand side is $O(\log^2 x)$. By hypothesis, each of the terms x^ρ/ρ has absolute value at most

$x^\theta/|\rho| < x^\theta/|\operatorname{Im}\rho|$. Hence

$$\left|\sum_{|\operatorname{Im}(\rho)|<T} \frac{x^\rho}{\rho}\right| < 2x^\theta \sum_{0<\operatorname{Im}(\rho)<T} \frac{1}{\operatorname{Im}\rho}.$$

We have seen already that the last sum is $O(\log^2 T)$; here $T = x + O(1)$, so we conclude that

$$\psi(x) - x = O(x^\theta \log^2 T) + O(\log^2 x) = O(x^\theta \log^2 x),$$

as claimed. The corresponding estimate on $\pi(x) - \operatorname{li}(x)$ then follows from (5), since $\theta \geq 1/2$. $\square$

A converse implication also holds:

Proposition. *Suppose there exists θ with $1/2 \leq \theta < 1$ such that $\psi(x) = x + O_\epsilon(x^{\theta+\epsilon})$ for all $\epsilon > 0$. Then $\zeta(s)$ has no zeros of real part $> \theta$. The same conclusion holds if $\pi(x) = \operatorname{li}(x) + O_\epsilon(x^{\theta+\epsilon})$.*

(So, for instance, RH is *equivalent* to the assertion that $\pi(x) = \operatorname{li} x + O(x^{1/2}\log x)$. The hypotheses on $\pi(x)$ and $\psi(x)$ are equivalent, again by (5).)

Proof: Write $-\zeta'(s)/\zeta(s) = \sum_n \Lambda(n) n^{-s}$ as a Stieltjes integral and integrate by parts to find

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

If $\psi(x) - x \ll_\epsilon x^{\theta+\epsilon}$ then the resulting integral for $s/(s-1) + \zeta'(s)/\zeta(s)$ extends to an analytic function on $\sigma > \theta$, whence that half-plane contains no zeros of $\zeta(s)$. $\square$

Note the amusing consequence that an estimate $\psi(x) = x + O_\epsilon(x^{\theta+\epsilon})$ would automatically improve to $\psi(x) = x + O(x^\theta \log^2 x)$, and similarly for $\pi(x)$.

Remarks

One may naturally ask whether $\psi(x)$ tends to be larger or smaller than its approximation x, and likewise whether $\pi(x)$ tends to be larger or smaller than $\operatorname{li}(x)$. For the former question, our formula (3) suggests that $\psi(x)$ can as easily be larger or smaller than x: the terms x^ρ/ρ in the formula (3) for $x - \psi(x)$ oscillate as x increases, and if we choose $\log x$ uniformly from $[1, U]$ then the phase of each term tends to uniform distribution on the circle as $U \to \infty$. It may be surprising then that $\pi(x)$ behaves quite differently: it is very hard to find any x such that $\pi(x) > \operatorname{li}(x)$. This is because $\pi(x)$ is expressed as a Stieltjes integral involving not $\psi(x)$ but $\sum_{p<x} \log p$, and

$$\psi(x) - \sum_{p<x} \log p \sim \psi(x^{1/2}) \sim x^{1/2}.$$

Under the Riemann Hypothesis, $x^{1/2}$ is exactly of the same asymptotic order as each of the terms x^ρ/ρ in (3), and much larger than each single term because

$|\rho|^{-1} < 1/14$. For large x, we can imagine the terms x^ρ/ρ ($\operatorname{Im}\rho > 0$) as random complex numbers z_ρ drawn independently from the circle $|z| = x^{1/2}/\rho$.[2] Then $\sum_\rho x^\rho/\rho = 2\operatorname{Re}\sum_{(\operatorname{Im}\rho)>0} z_\rho$. Since $\sum_\rho 1/|\rho|^2 < \infty$, this heuristic suggests that for "random large x" the scaled error $x^{-1/2}(\psi(x) - x)$ is drawn from a distribution symmetric about the origin, and thus that $x^{-1/2}(\sum_{p<x} \log p - x)$ is drawn from a distribution symmetric about -1. Since $\sum_\rho 1/|\rho| = +\infty$, it is possible for $-2\operatorname{Re}\sum_{(\operatorname{Im}\rho)>0} z_\rho$ to exceed x, and thus for $\sum_{p<x}\log p$ to exceed x and likewise for $\pi(x)$ to exceed $\operatorname{li}(x)$. But this does not happen routinely, and indeed it was once thought that $\operatorname{li}(x)$ might always exceed $\pi(x)$.

Littlewood first showed that the difference changes sign infinitely often. In particular, there exist x such that $\pi(x) > \operatorname{li}(x)$. But none has been found yet. The earliest explicit upper bound on the smallest such x was the (in)famously astronomical "Skewes' number" [Skewes 1933]. That bound has since fallen, but still stands at several hundred digits, too large to reach directly even with the best algorithms known for computing $\pi(x)$ — algorithms that themselves depend on the analytical formulas such as (2); see [LO 1982].

Exercises

1. Use the partial-fraction decomposition of ζ'/ζ to get the following exact formula:

$$\psi(x) = x - \sum_\rho \frac{x^\rho}{\rho} - \frac{\zeta'}{\zeta}(0) - \frac{1}{2}\log(1 - x^{-2}).$$

Here $\sum_\rho$ is taken to mean $\lim_{T\to\infty}\sum_{|\rho|<T}$; and if $x = p^k$, so that $\psi(x)$ is discontinuous at x, then we interpret $\psi(x)$ as $(\psi(x-\epsilon)+\psi(x+\epsilon))/2$. Note that $-\frac{1}{2}\log(1-x^{-2})$ is the sum of $-x^r/r$ over the trivial zeros $r = -2, -4, -6, \ldots$ See [Davenport 1967, Chapter 17].

2. Show that the improvement $1-\sigma > c_\epsilon/\log^{(2/3)+\epsilon}|t|$ on (4) yields an estimate $O\big(x\exp(-C_\epsilon \log^{(3/5)-\epsilon} x)\big)$ on the error in the Prime Number Theorem.

3. Prove that

$$\lim_{x\to\infty}\left(\log x - \sum_{n=1}^{x}\frac{\Lambda(n)}{n}\right) = \boldsymbol{\gamma},$$

and give an error bound both unconditionally and under the Riemann Hypothesis. Deduce that $\log x - \sum_{p<x}\log p/p$ and $\log\log x - \sum_{p<x} 1/p$ approach finite limits as $x\to\infty$. (The last of these refines Euler's theorem that $\sum_p 1/p$ diverges.)

4. [A theorem of Mertens; see for instance [Titchmarsh 1951], pages 38–39.] Prove that

$$\lim_{x\to\infty}\left(\log\log x - \sum_{n=1}^{x}\frac{\Lambda(n)}{n\log n}\right) = -\boldsymbol{\gamma},$$

[2] We shall later make this heuristic more precise, and show that it is equivalent to the conjecture that the numbers $\gamma > 0$ such that $\zeta(\frac{1}{2} + i\gamma) = 0$ are $\mathbf{Q}$-linearly independent. This conjecture is almost certainly true and extremely difficult to prove. See [RS 1994] and [BFHR 2001] for more information.

(Warning: this requires a contour integral involving $\log\big((s-1)\zeta(s)\big)$, which cannot be pushed past the zero-free region.) Deduce that

$$\lim_{x\to\infty}\left(\log x\prod_{p<x}\frac{p-1}{p}\right)=e^{-\gamma}.$$

As in the previous exercise, and give error bounds both unconditionally and under the Riemann Hypothesis.

In the last exercise, we illustrate the power of the method we used to prove the Prime Number Theorem by applying it to different kind of asymptotic averaging problem. We'll address a special case posed as an open problem in [Rawsthorne 1984]:

> Set $a_0 = 1$ and for $n \geq 1$, $a_n = a_{n'} + a_{n''} + a_{n'''}$ where $n' = \lfloor n/2 \rfloor$, $n'' = \lfloor n/3 \rfloor$, $n''' = \lfloor n/6 \rfloor$. Find $\lim_{n\to\infty} a_n/n$.

(It is not immediately obvious even that the limit exists.) The general problem can be solved in much the same way, though one usually gets somewhat less precise estimates on the vertical distribution of the zeros than are available for our special case. Only two solutions were received (see *Math. Magazine* **58**, 51–52): the solution outlined here, and a solution by Erdős, Odylzko, Hildebrand, Pudaite, and Reznick, which they subsequentely generalized in [EHOPR 1987]. Their method corresponds to one of the "elementary proofs" of the Prime Number Theorem. The sequence $\{a(n)\}$ of Rawsthorne's problem is now #A007731 in Sloane's *On-Line Encyclopedia of Integer Sequences*.

5. i) Let $f(s) = 1 - 2^{-s} - 3^{-s} - 6^{-s}$. Note that f has a simple zero at $s = 1$. Prove that all its other zeros lie in the strip $|\sigma| < 1$, and that f has $\frac{\log 6}{2\pi}T + O(1)$ zeros ρ with $0 < \operatorname{Im}\rho < T$; more precisely, that each rectangle

$$\{\sigma + it : |\sigma| \leq 1,\ \left|\frac{\log 6}{2\pi}t - n\right| < 1/2\}$$

($n \in \mathbf{Z}$) contains a unique zero of f (so in particular the zeros are all simple). NB Unlike the case of $\zeta(s)$, here there is no functional equation, nor a "Riemann Hypothesis"; indeed, it can be shown that some complex zeros have real parts arbitrarily close to 1, as well as zeros whose real parts are arbitrarily close to -1.
ii) Let a_n be the coefficients of the Dirichlet series $\sum_{n=1}^{\infty} a_n/n^s = 1/f(s)$. Show that $a_n \geq 0$, with equality unless $n = 2^a 3^b$ for some integers a, b. Find a constant C such that $\sum_{n<x} a_n \sim Cx$ as $x\to\infty$. Can you give an explicit error bound?
iii) Solve Rawsthorne's problem above. How far can you generalize it? (Warning: for more general recursions of this kind you may have to contend with multiple poles, or simple poles that nearly coincide and have large residues.)

References

[BFHR 2001] Bays, C., Ford, K., Hudson, R.H., Rubinstein, M.: Zeros of Dirichlet L-functions near the Real Axis and Chebyshev's Bias. *J. Number Th.* **87** (2001) #1, 54–76.

[Hasse 1936] Hasse, H.: Zur Theorie der abstrakten elliptischen Funktionkörper I, II, III, *J. reine angew. Math.* **175** (1936), 55–62, 69–88, and 193–208.

[EHOPR 1987] Erdős, P., Hildebrand, A., Odlyzko, A., Pudaite, P., Reznick, B.: The asymptotic behavior of a family of sequence, *Pacific J. Math.* **126** (1987), 227–241.

[LO 1982] Lagarias, J.C., Odlyzko, A.M.: New algorithms for computing $\pi(x)$. Pages 176–193 in *Number Theory: New York 1982* (D.V. and G.V. Chudnovsky, H. Cohn, M.B. Nathanson, eds.; Berlin: Springer 1984, LNM **1052**).

[Rawsthorne 1984] Rawsthorne, D.A.: Problem 1185, *Math. Magazine* **57** (1984), p.42.

[RS 1994] Rubinstein, M. Sarnak, P.: Chebyshev's Bias. *Exp. Math.* **3** (1994) #3, 173–197.

[Skewes 1933] Skewes, S.: On the difference $\pi(x) - \mathrm{li}(x)$ (I). *J. London Math. Soc.* (1st ser.) **8** (1933), 277–283.

[Weil 1940] Weil, A.: Sur les fonctions algébriques à corps de constantes fini, *C.R. Acad. Sci. Paris* **210** (1940), 592–594.

[Weil 1941] Weil, A.: On the Riemann hypothesis in functionfields, *Proc. Nat. Acad. Sci. USA* **27** (1941), 345–347.

[Weil 1948] Weil, A.: *Variétés abéliennes et courbes algébriques*, Paris: Hermann 1948.

Math 259: Introduction to Analytic Number Theory

$L(s, \chi)$ as an entire function; Gauss sums

We first give, as promised, the analytic proof of the nonvanishing of $L(1, \chi)$ for a Dirichlet character χ mod q; this will complete our proof of Dirichlet's theorem that there are infinitely primes in the arithmetic progression $\{mq+a : m \in \mathbf{Z}_{>0}\}$ whenever $(a, q) = 1$, and that the logarithmic density of such primes is $1/\varphi(q)$.[1]

We follow [Serre 1973, Ch. VI §2]. Functions such as $\zeta(s)$, $L(s, \chi)$ and their products are special cases of what Serre calls "Dirichlet series": functions

$$f(s) := \sum_{n=1}^{\infty} a_n e^{-\lambda_n s} \tag{1}$$

with $a_n \in \mathbf{C}$ and $0 \leq \lambda_n < \lambda_{n+1} \to \infty$. [For instance, $L(s, \chi)$ is of this form with $\lambda_n = \log n$ and $a_n = \chi(n)$.] We assume that the a_n are small enough that the sum in (1) converges for some $s \in \mathbf{C}$. We are particularly interested in series such as

$$\zeta_q(s) := \prod_{\chi \bmod q} L(s, \chi)$$

whose coefficients a_n are nonnegative. Then if (1) converges at some real σ_0, it converges uniformly on $\sigma \geq \sigma_0$, and $f(s)$ is analytic on $\sigma > \sigma_0$. Thus a series (1) has a maximal open half-plane of convergence (if we agree to regard $\mathbf{C}$ itself as an open half-plane for this purpose), namely $\sigma > \sigma_0$ where σ_0 is the infimum of the real parts of $s \in \mathbf{C}$ at which (1) converges. This σ_0 is then called the "abscissa of convergence"[2] of (1).

We claim that if σ_0 is finite then it is a singularity of f; that is:

Proposition. *Suppose that the series (1) has positive coefficients a_n, and that there exists $\rho \in \mathbf{R}$ such that the series converges in the half-plane $\sigma > \rho$ and extends to an analytic function in a neighborhood of ρ. Then the abscissa of convergence of the series is strictly smaller than ρ.*

Proof: Since $f(s - \rho)$ is again of the form (1) with nonnegative coefficients $e^{\lambda_n \rho} a_n$, it is enough to prove the Proposition for $\rho = 0$. Since f is then analytic in $\sigma > 0$ and also in $|s| < \delta$ for some $\delta > 0$, it is analytic in $|s - 1| \leq 1 + \epsilon$ for sufficiently small ϵ, specifically any $\epsilon < \sqrt{1 + \delta^2} - 1$. Expand f in a Taylor series about $s = 1$. Since (1) converges uniformly in a neighborhood of that point, it may be differentiated termwise, and we find that its m-th derivative there is

$$f^{(m)}(1) = \sum_{n=1}^{\infty} (-\lambda_n)^m a_n e^{-\lambda_n}.$$

[1] Davenport gives a simpler proof of Dirichlet's theorem, also involving L-functions but not yet obtaining even the logarithmic density, in Chapter 4, attributing the basic idea to Landau 1905.

[2] The quaint word "abscissa" for "x-coordinate" is still sometimes encountered in analytic geometry, alongside "ordinate" (a.k.a. "y-coordinate").

Taking $s = -\epsilon$, we obtain the convergent sum

$$f(-\epsilon) = \sum_{m=0}^{\infty} \frac{(-1-\epsilon)^m}{m!} f^{(m)}(1) = \sum_{m=0}^{\infty} \frac{(1+\epsilon)^m}{m!} \left[\sum_{n=1}^{\infty} (+\lambda_n)^m a_n e^{-\lambda_n} \right].$$

Since all the terms in the sum are nonnegative, the sum converges absolutely, and may be summed in reverse order. Therefore

$$f(-\epsilon) = \sum_{n=1}^{\infty} a_n \left[\sum_{m=0}^{\infty} e^{-\lambda_n} \frac{(1+\epsilon)^m}{m!} \lambda_n^m \right].$$

But the new inner sum is just a Taylor series for $e^{\lambda_n \epsilon}$. So we have shown that the series (1) converges at $s = -\epsilon$, and thus has abscissa of convergence $\sigma_0 \leq -\epsilon < 0 = \rho$. $\square$

We can now prove:

Theorem. *Let χ be a nontrivial character mod q. Then $L(1, \chi) \neq 0$.*

Proof: We know already that $L(s, \chi)$ extends to a function on $\sigma > 0$ analytic except for the simple pole of $L(s, \chi_0)$ at $s = 1$. If any $L(s, \chi)$ vanished at $s = 1$ then

$$\zeta_q(s) := \prod_{\chi \bmod q} L(s, \chi)$$

would extend to an analytic function on $\sigma > 0$. But we observed already that $\zeta_q(s)$ is a Dirichlet series $\sum_n a_n n^{-s}$ with nonnegative coefficients that converges at least in $\sigma > 1$. By our Proposition, this series would thus converge in $\sigma > 0$. But we also have $a_n \geq 1$ if $n = k^{\varphi(q)}$ for some k coprime to q. Therefore $\sum_n a_n n^{-\sigma}$ diverges for $\sigma \leq 1/\varphi(q)$. This contradiction proves that no $L(1, \chi)$ vanish. $\square$

We have thus established Dirichlet's theorem on the infinitude and logarithmic density of primes $qm + a$. But we want more than logarithmic density, namely asymptotics of $\pi(x, a \bmod q)$, or equivalently of $\pi(x, \chi)$. As with the Prime Number Theorem, it will be enough to estimate

$$\psi(x, \chi) := \sum_{n<x} \chi(n) \Lambda(n),$$

for which we have an integral approximation

$$\psi(x, \chi) = \frac{1}{2\pi i} \int_{1+\frac{1}{\log x}-iT}^{1+\frac{1}{\log x}+iT} -\frac{L'}{L}(s, \chi)\, x^s \frac{ds}{s} + O\left(\frac{x \log^2 x}{T}\right) \qquad (T \in [1, x]).$$

We therefore seek a partial-fraction decomposition for L'/L, which in turn leads us to prove an analytic continuation and functional equation for $L(s, \chi)$.

Our key tool in proving the functional equation for $\zeta(s)$ was the Poisson summation formula, which we recovered from the Fourier series of

$$F(x) := \sum_{m=-\infty}^{\infty} f(x+m)$$

by setting $x = 0$. We now need this Fourier series

$$F(x) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{-2\pi i n x}$$

for fractional x. (Here $\hat{f}$ is the Fourier transform of f, defined by

$$\hat{f}(y) = \int_{-\infty}^{+\infty} e^{2\pi i x y} f(x)\, dx. \tag{2}$$

as before.) Let $a \mapsto c(a)$ be any function from $\mathbf{Z}/q\mathbf{Z}$ to $\mathbf{C}$. Then we have

$$\sum_{m=-\infty}^{\infty} c(m) f(m/q) = \sum_{a \bmod q} c(a) F(a/q) = \sum_{n=-\infty}^{\infty} \hat{c}(-n)\hat{f}(n), \tag{3}$$

where $\hat{c}$ is the *discrete Fourier transform* of c, defined by

$$\hat{c}(n) := \sum_{a \bmod q} c(a) e^{2\pi i n a/q}. \tag{4}$$

Now suppose c is a *primitive* character $\chi \bmod q$. We use the notation τ_n for its discrete Fourier transform; that is,

$$\tau_n(\chi) := \sum_{a \bmod q} \chi(a) e^{2\pi i n a/q}.$$

We claim:

Lemma. *Assume that χ is a primitive character mod q. Then*

$$\tau_n(\chi) = \overline{\chi}(n)\tau_1(\chi) \tag{5}$$

holds for all $n \in \mathbf{Z}$. That is, the discrete Fourier transform of a primitive character χ is $\tau_1(\chi)\overline{\chi}$. If n is coprime to q then (5) holds for all characters $\chi \bmod q$, primitive or not.

Proof: If $\gcd(n, q) = 1$ then we may replace a by $n^{-1}a$, from which $\tau_n(\chi) = \overline{\chi}(n)\tau_1(n)$ follows. If $(n, q) > 1$ then $\overline{\chi}(n) = 0$, so we want to show $\tau_n(\chi) = 0$. Let $d = (n, q)$ and $q_0 = q/d$, and rearrange the $\tau_n(\chi)$ sum according to $a \bmod q_0$:

$$\tau_n(\chi) = \sum_{a_0 \bmod q_0} \sum_{\substack{a \bmod q \\ a \equiv a_0 \bmod q_0}} \chi(a) e^{2\pi i n a/q} = \sum_{a_0 \bmod q_0} e^{2\pi i n a_0/q} \left[\sum_{a \equiv a_0 \bmod q_0} \chi(a) \right].$$

We claim that the inner sum vanishes. This is clear unless $\gcd(a_0, q_0) = 1$. In that case the inner sum is

$$\chi(a_1) \sum_{\substack{a \bmod q \\ a \equiv 1 \bmod q_0}} \chi(a),$$

for any $a_1 \equiv a_0 \bmod q_0$. But this last sum is the sum of a character on the group of units mod q congruent to 1 mod q_0, and so vanishes unless that character is trivial — and if $\chi(a) = 1$ whenever $a \equiv 1 \bmod q_0$ then χ comes from a character mod q_0 (why?) and is thus not primitive. This completes the proof. □

We generally abbreviate $\tau_1(\chi)$ as $\tau(\chi)$, and call that number

$$\tau(\chi) := \sum_{a \bmod q} \chi(a) e^{2\pi i a/q} \tag{6}$$

the *Gauss sum* of the character χ. We then have:

Theorem. *Let $f : \mathbf{R} \to \mathbf{C}$ be any function satisfying the hypotheses of Poisson summation. Then for any primitive character $\chi \bmod q$ we have*

$$\sum_{m=-\infty}^{\infty} \chi(m) f(m/q) = \tau(\chi) \sum_{n=-\infty}^{\infty} \overline{\chi}(-n) \hat{f}(n). \tag{7}$$

Proof: Substitute the formula (5) of our Lemma into (3). □

This may be regarded as the "twist by χ" of the Poisson summation formula.

For even characters χ, we know what to do next: take $f(x) = e^{-\pi u (qx)^2}$ in (7), and apply the Mellin transform to the resulting identity. This is actually easier than our proof of the functional equation for $\zeta(s)$, because we do not need to split the integral in two. (Ultimately this is because, unlike $\zeta(\cdot)$, the L-function of a nontrivial primitive character has no poles.) Let[3]

$$\theta_\chi(u) := \sum_{n=-\infty}^{\infty} \chi(n) e^{-\pi n^2 u}.$$

By (7), together with the fact that the Fourier transform of $f(x) = e^{-\pi u (qx)^2}$ is $\hat{f}(y) = u^{-1/2} q^{-1} e^{-\pi u^{-1} (y/q)^2}$, we obtain

$$\theta_\chi(u) = \frac{\tau(\chi)}{q u^{1/2}} \sum_{n=-\infty}^{\infty} \overline{\chi}(-n) e^{-\pi u^{-1} (n/q)^2} = \frac{\tau(\chi)}{q u^{1/2}} \theta_{\overline{\chi}}(1/q^2 u). \tag{8}$$

Note that, unless $q = 1$, it follows that $\theta_\chi(u)$ is rapidly decreasing as $u \to 0+$, because $\chi(0) = 0$. Integrating termwise, we find

$$2\pi^{-s/2} \Gamma(s/2) L(s, \chi) = \int_0^\infty \theta_\chi(u) u^{s/2} \, \frac{du}{u}$$

for $\mathrm{Re}(s) > 1$. Since $\theta_\chi(u) \ll \exp(-\pi/q^2 u)$ as $u \to 0+$, the integral converges for all s, and gives the analytic continuation of $L(s, \chi)$ to an entire function with

[3] Our $\theta_\chi(u)$ is called $\psi(qu, \chi)$ in [Davenport 1967, Ch.9].

zeros at the poles $s = 0, -2, -4, -6, \ldots$ of $\Gamma(s/2)$. Moreover, by (8) the integral is also

$$\frac{\tau(\chi)}{q}\int_0^\infty \theta_{\overline{\chi}}(1/q^2u)u^{(s-1)/2}\,\frac{du}{u} = \frac{\tau(\chi)}{q}\int_0^\infty \theta_{\overline{\chi}}(u)(q^2u)^{(1-s)/2}\,\frac{du}{u}$$

$$= \frac{\tau(\chi)}{q^s}\int_0^\infty \theta_{\overline{\chi}}(u)u^{(1-s)/2}\,\frac{du}{u}.$$

This last integral is $2\tau(\chi)q^{-s}\Gamma(\frac{1}{2}(1-s))\pi^{(s-1)/2}L(1-s,\overline{\chi})$ for $\sigma \in (0,1)$, and thus by analytic continuation for all $s \in \mathbf{C}$. We can write the functional equation symmetrically by setting

$$\xi(s,\chi) := (\pi/q)^{-s/2}\Gamma(s/2)L(s,\chi),$$

which is now an entire function: $\xi(s,\chi)$ is related with $\xi(s,\overline{\chi})$ by

$$\xi(s,\chi) = \frac{\tau(\chi)}{\sqrt{q}}\xi(1-s,\overline{\chi}). \tag{9}$$

What about odd χ? The same definition of θ_χ would yield zero. We already indicated (in the exercises on the functional equation for ζ and ξ) the correct approach: we apply (7) not to the Gaussian $e^{-\pi u(qx)^2}$ but to its derivative, which is proportional to $xe^{-\pi u(qx)^2}$. Using the general fact that the Fourier transform of f' is $2\pi iy\hat{f}(y)$ (integrate by parts in the definition (2) of $\widehat{f'}$) we see that the Fourier transform of $xe^{-\pi u(qx)^2}$ is $(iy/(u^{1/2}q)^3)e^{-\pi u^{-1}(y/q)^2}$. So, if we define[4]

$$\vartheta_\chi(u) := \sum_{n=-\infty}^{\infty} n\chi(n)e^{-\pi n^2 u},$$

we find

$$\vartheta_\chi(u) = \frac{\tau(\chi)}{iq^2u^{3/2}}\vartheta_{\overline{\chi}}(1/q^2u). \tag{10}$$

This time we must multiply ϑ_χ by $u^{(s+1)/2}\,du/u$ to cancel the extra factor of n. We obtain the integral formula

$$2\pi^{-(s+1)/2}\Gamma((s+1)/2)L(s,\chi) = \int_0^\infty \vartheta_\chi(u)u^{(s+1)/2}\,\frac{du}{u}$$

for $L(s,\chi)$. Again (10) together with $\chi(0) = 0$ tells us that $\vartheta_\chi(u)$ vanishes rapidly as $u\to 0+$, and thus that our integral extends to an entire function of s; note however that the resulting trivial zeros of $L(s,\chi)$ are at the negative *odd* integers. The functional equation (10) again gives us a relation between $L(s,\chi)$ and $L(1-s,\overline{\chi})$, which this time has the symmetrical form

$$\xi(s,\chi) = \frac{\tau(\chi)}{i\sqrt{q}}\xi(1-s,\overline{\chi}), \tag{11}$$

[4]Our $\vartheta_\chi(u)$ is Davenport's $\psi_1(qu,\chi)$.

with

$$\xi(s,\chi) := (\pi/q)^{-(s+1)/2}\Gamma((s+1)/2)L(s,\chi).$$

We may combine (9) with (11) by introducing an integer a depending on χ:

$$\mathsf{a} := \begin{cases} 0, & \text{if } \chi(-1) = +1; \\ 1, & \text{if } \chi(-1) = -1. \end{cases}$$

That is, $\mathsf{a} = 0$ or 1 according as χ is even or odd. We then have:

Theorem. *Let χ be any primitive character mod q, and $\mathsf{a} = 0$ or 1 as above. Define*

$$\xi(s,\chi) := (\pi/q)^{-(s+\mathsf{a})/2}\Gamma((s+\mathsf{a})/2)L(s,\chi).$$

Then the ξ functions of χ and $\overline{\chi}$ are related by the functional equation

$$\xi(s,\chi) = \frac{\tau(\chi)}{i^{\mathsf{a}}\sqrt{q}}\xi(1-s,\overline{\chi}). \tag{12}$$

Note that this holds even for the case $q = 1$, in which $L(s,\chi)$ and $\xi(s,\chi)$ reduce to $\zeta(s)$ and $\xi(s)$. In that case, we concluded that $(s^2-s)\xi(s)$ is an entire function of order 1. For $q > 1$, the function $\xi(s,\chi)$ has no poles, and we find in the same way that $\xi(s,\chi)$ is an entire function of order 1. We shall develop its product formula and deduce the asymptotics of $\psi(x,\chi)$ in the next lecture notes.

Meanwhile, we prove some basic results concerning Gauss sums. First we find $|\tau(\chi)|$:

Proposition. *The Gauss sum $\tau(\chi)$ of any primitive Dirichlet character χ mod q has absolute value $q^{1/2}$.*

Proof: We obtain this as a special case of the Parseval identity for discrete Fourier transforms:

$$\sum_{n \bmod q} |\hat{c}(n)|^2 = q \sum_{a \bmod q} |c(a)|^2 \tag{13}$$

for any function $a \mapsto c(a)$ from $\mathbf{Z}/q\mathbf{Z}$ to $\mathbf{C}$. The identity (13) can be proved either directly or by observing that the functions $a \mapsto e^{2\pi i n a/q}$ are orthogonal with constant norm q. Now take $c(a) = \chi(a)$. We have seen that $\hat{c}(n) = \tau(\chi)\overline{\chi}(n)$. Therefore (13) becomes $|\tau(\chi)|^2\phi(q) = q\varphi(q)$, whence $|\tau(\chi)|^2 = q$ as claimed. □

The Gauss sum may be regarded as a discrete analogue of the Gamma integral: the factors x^{s-1} and e^{-x} in the integral $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\,dx$ are a varying homomorphism from $\mathbf{R}^*_{>0}$ to $\mathbf{C}^*$ and a fixed homomorphism from the additive group $\mathbf{R}$ to $\mathbf{C}^*$; in the Gauss sum (6), these are replaced by the varying homomorphism χ from $(\mathbf{Z}/q\mathbf{Z})^*$ and the fixed homomorphism $a \mapsto e^{2\pi i a/q}$ from $(\mathbf{Z}/q\mathbf{Z}, +)$. The analogy is particularly close when q is prime, in which case $\mathbf{Z}/q\mathbf{Z}$, like $\mathbf{R}$, is a field. In this case the Beta integral has a corresponding

analogue in the *Jacobi sums*

$$J(\chi,\chi') := \sum_{c \bmod q} \chi(c)\chi'(1-c).$$

We do not require that χ and χ' be primitive. Note that unlike $\tau(\chi)$, which may involve both q-th and $(q-1)$st roots of unity, the Jacobi sum $J(\chi,\chi')$ involves only $(q-1)$st roots. Nevertheless it can be evaluated in terms of Gauss sums:

Proposition. *Let χ,χ' be Dirichlet characters mod q. Then*

$$J(\chi,\chi') = \frac{\tau(\chi)\tau(\chi')}{\tau(\chi\chi')}, \tag{14}$$

provided that none of $\chi,\chi',\chi\chi'$ is the trivial character χ_0. If $\chi=\chi'=\chi_0$ then $J(\chi,\chi')=q-2$; if exactly one of χ and χ' is trivial then $J(\chi,\chi')=-1$; and if χ,χ' are nontrivial but $\chi\chi'=\chi_0$ then $J(\chi,\chi')=-\chi(-1)$.

Proof: It is clear that $J(\chi_0,\chi_0)=q-2$, so we henceforth assume that χ,χ' are not both trivial. As in our evaluation of $\mathrm{B}(s,s')$ in terms of the Gamma function, we consider the double sum

$$\tau(\chi)\tau(\chi') = \sum_{a,a' \bmod q}\sum \chi(a)\chi'(a')e^{2\pi i(a+a')/q}.$$

Let $b=a+a'$. The terms with $b=0$ sum to

$$\chi'(-1)\sum_{a \bmod q}\chi\chi'(a) = \begin{cases}\chi(-1)(q-1), & \text{if } \chi'=\overline{\chi};\\ 0, & \text{otherwise.}\end{cases}$$

To sum the terms for fixed nonzero b, let $a=cb$ and $a'=(1-c)b$ to find

$$e^{2\pi ib/q}\chi\chi'(b)\sum_{c \bmod q}\chi(c)\chi'(1-c) = e^{2\pi ib/q}\chi\chi'(b)J(\chi,\chi'). \tag{15}$$

Hence if $\chi\chi'=\chi_0$ (that is, if $\chi'=\overline{\chi}$), we have

$$\tau(\chi)\tau(\chi') = \chi(-1)(q-1) - J(\chi,\chi').$$

But

$$\tau(\overline{\chi}) = \sum_{a \bmod q}\overline{\chi}(a)e^{2\pi ia/q} = \overline{\sum_{a \bmod q}\chi(a)e^{-2\pi ia/q}} = \chi(-1)\overline{\tau(\chi)}, \tag{16}$$

so

$$J(\chi,\overline{\chi}) = \chi(-1)(q-1) - \tau(\chi)\tau(\chi') = \chi(-1)(q-1-|\tau(\chi)|^2) = -\chi(-1).$$

(We could also have obtained this directly from $\chi(c)\overline{\chi}(1-c)=\overline{\chi}(c^{-1}-1)$, which in turn yields an alternative proof of $|\tau(\chi)|=q^{1/2}$ in the prime case.) Otherwise (15) yields (14). □

Corollary. *The Jacobi sum $J(\chi, \chi')$ has absolute value $q^{1/2}$ if each of $\chi, \chi', \chi\chi'$ is nontrivial.*

The formula (14) is the beginning of a long and intricate chapter of the arithmetic of cyclotomic number fields; it can also be used to count solutions of certain Diophantine equations mod q, showing for instance that if $q \equiv 1 \bmod 3$ then there are $q/9 + O(\sqrt{q})$ values of $c \neq 0, 1$ in $\mathbf{Z}/q\mathbf{Z}$ such that both c and $1 - c$ are cubes. See the Exercises for more details and examples.

Remarks

Our extended Poisson identity (3) also has a generalization to locally compact abelian groups. Let G be such a group, H a closed subgroup, and K a closed subgroup of H. Then the annihilators $H^\perp, K^\perp$ in $\hat{G}$ are closed subgroups, with $H^\perp \subseteq K^\perp$. Moreover, $K^\perp/H^\perp$ is canonically identified with the Pontrjagin dual of H/K. Then one can choose Haar measures on G, H, H/K, and $K^\perp$ such that

$$\int_{x \in H} c(x + K) f(x) = \int_{y \in K^\perp} \hat{c}(-y + H^\perp) \hat{f}(y).$$

under suitable hypothesis on the functions $c : H/K \to \mathbf{C}$ and $f : H \to \mathbf{C}$. The formula (3) is the special case $G = \hat{G} = \mathbf{R}$, $K = K^\perp = \mathbf{Z}$, $H = q^{-1}\mathbf{Z}$, $H^\perp = q\mathbf{Z}$.

The formula for $\xi(s, \chi)$, and the distinction between even and odd characters χ, can be interpreted structurally as follows. Let χ be a primitive character mod q. We mentioned already that $L(s, \chi)$ is a factor in a product formula for $\zeta_K(s)$, the zeta function of the cyclotomic number field $K = \mathbf{Q}(e^{2\pi i/q})$; and that for any number field K, the Euler product for ζ_K can be "completed" to a function $\xi_K(s)$ that satisfies a functional equation $\xi_K(s) = \xi_K(1 - s)$. The additional factors come from the discriminant and archimedean valuations of K. When K is cyclotomic, we can also give such an interpretation of $\xi(s, \chi)$. We may regard χ as a character of the group $(\mathbf{Z}/q\mathbf{Z})^*$ which is canonically isomorphic with the Galois group $G = \mathrm{Gal}(K/\mathbf{Q})$. For each prime $p \nmid q$, the number $\chi(p)$ that appears in the local factor $(1 - \chi(p)p^{-s})^{-1}$ is then the image under χ of the p-Frobenius element in G. That is, this factor records the p-adic behavior of the Galois extension $K/\mathbf{Q}$. Just as the factor $\Gamma(s/2)$ in $\xi(s) = \xi_{\mathbf{Q}}(s)$ was regarded as a local factor at the archimedean place of $\mathbf{Q}$, the factor $\Gamma((s + \mathsf{a})/2)$ in $\xi(s, \chi)$ can be regarded as an archimedean local factor. Instead of Frobenius, the archimedean place is associated with *complex conjugation*, whose image in $\mathrm{Gal}(K/\mathbf{Q})$ is identified with $-1 \in (\mathbf{Z}/q\mathbf{Z})^*$ under the isomorphism from $(\mathbf{Z}/q\mathbf{Z})$ to G. The factor $\Gamma((s + \mathsf{a})/2)$, which depends on $\chi(-1)$, thus records the image under χ of complex conjugation.

We could also have obtained the identity $|\tau(\chi)|^2 = q$ indirectly from the twisted Poisson formula (7). If f is a function such that both $f, \hat{f}$ satisfy the Poisson hypotheses, we may apply (7) twice to find that either

$$\tau(\chi)\tau(\overline{\chi}) = \chi(-1)q \tag{17}$$

or $\sum_{m \in \mathbf{Z}} \chi(m) f(m/q) = 0$. Since the latter possibility cannot hold for all f

(for instance, consider $f(x) = \exp(-C(x-1/q)^2)$ for large C), we have deduced (17). But (17) is equivalent to $|\tau(\chi)|^2 = q$ by the formula (16) which relates $\tau(\chi)$ and $\tau(\overline{\chi})$.

Exercises

On general Dirichlet series:

1. Suppose the λ_n are closed under addition.
i) Show that for any right half-plane $H = \{s \in \mathbf{C} : \mathrm{Re}(s) \geq \sigma_0\}$ or $H = \{s \in \mathbf{C} : \mathrm{Re}(s) > \sigma_0\}$ the space of Dirichlet series (1) that converge absolutely in H is closed under multiplication. (We allow $\sigma = -\infty$, in which case $H = \mathbf{C}$, as well as $\sigma = +\infty$, in which case we are dealing with formal Dirichlet series.)
ii) If $f(s) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n s}$ and $g(s) = \sum_{n=1}^{\infty} b_n e^{-\lambda_n s}$ are two such functions then the sequence of coefficients of $fg(s)$ is the *convolution* $a*b$ of the sequences a_n, b_n. Prove that convolution is associative: $(a*b)*c = a*(b*c)$. [NB: While this is suggested by part (i), it is not quite an immediate consequence, unless you exclude formal Dirichlet series and show that two Dirichlet series that are equal on H have the same coefficients.] Note that when $\lambda_n = n\lambda_1$ this recovers the usual (additive) convolution $(a*b)_n = \sum_{j+k=n} a_j b_k$.
iii) When $\lambda_n = \log n$, we have $(a*b)_n = \sum_{jk=n} a_j b_k$. Show that the result of part (ii) includes *Möbius inversion*: if $c_n = \sum_{d|n} a_d$ then $a_n = \sum_{d|n} \mu(n/d) c_d$.

2. Consider a Dirichlet series (1) in which a_n need not be positive reals. Clearly this series still has an abscissa of absolute convergence. Less obvious, but still true, is that it also has an abscissa of ordinary convergence. Show that if the sum (1) converges in the usual sense of $\lim_{N\to\infty} \sum_1^N$ at some s_0 then it converges also in $\sigma \geq \mathrm{Re}(s_0)$, the convergence being uniform in $\arg(s - s_0) \leq \alpha$ for each $\alpha < \pi/2$. Deduce that (1) defines an analytic function on $\sigma \geq \mathrm{Re}(s_0)$. [Since $f(s - s_0)$ is again of the form (1), it is enough to prove this claim for $s_0 = 0$. Assume then that $\sum_{n=1}^{\infty} a_n$ converges, and let $A(x) = -\sum_{\lambda_n > x} a_n$; by hypothesis $A(x) \to 0$ as $x \to \infty$. For large M, N with $M \leq N$, write

$$\sum_{n=M}^{N} a_n e^{-\lambda_n s} = \int_{\lambda_M}^{\lambda_N} e^{-\lambda s}\, dA(\lambda),$$

etc. This is equivalent to the route taken by Serre, but probably more transparent to us. Note that the convergence of the Dirichlet series for $L(s, \chi)$ on $\sigma > 0$ now follows automatically from its convergence for positive real s.]

3. Are the Dirichlet series (1) that converge (but might not converge absolutely) on a given right half-plane H closed under multiplication? (Hint: use the method of the previous Exercise to show that any such Dirichlet series f satisfies $f(s) \ll 1 + |s|$ on H.)

On L-functions:

4. Complete the missing steps in the proof of (11).

5. Suppose χ is a real character. Then (12) relates $\xi(s, \chi)$ with $\xi(s, 1 - \chi)$.

Deduce that $L(s,\chi)$ has a zero of even or odd multiplicity at $s = 1/2$ according as $\tau(\chi) = +i^{\mathfrak{a}}\sqrt{q}$ or $\tau(\chi) = -i^{\mathfrak{a}}\sqrt{q}$. In particular, in the minus case $L(1/2,\chi) = 0$.

But it is known that in fact the minus case never occurs:

$$\tau(\chi) = +i^{\,\mathfrak{a}}\sqrt{q} \tag{18}$$

holds for all primitive real χ mod q. This was first proved by Gauss after much work. (Davenport proves this in the special case of prime q in Chapter 2, using a later method of Dirichlet that relies on Poisson summation; we outline another proof in the next few Exercises.) It follows that the order of vanishing of $L(s,\chi)$ at $s = 1/2$ is even; it is conjectured, but not proved, that in fact $L(1/2,\chi) > 0$ for all Dirichlet characters χ. More complicated number fields are known whose zeta functions do vanish (always to even order) at $s = 1/2$.

On Gauss sums:

6. Suppose χ is a character mod $q = q_1q_2$ with q_1, q_2 coprime. Then $\chi = \chi_1\chi_2$ for some characters χ_i mod q_i, and χ is primitive if and only if both χ_1 and χ_2 are primitive. (You have shown this in the course of enumerating primitive characters.) In this case, express $\tau(\chi)$ in terms of the q_i, χ_i, and $\tau(\chi_i)$.

7. Suppose further that χ is a real character. Then the same is true of χ_1 and χ_2 (why?). Use your result in the previous Exercise, together with Quadratic Reciprocity, to verify that (18) holds for χ if it holds for each of χ_1 and χ_2. (In the opposite direction, if one proves in some other way that (18) holds for all primitive real Dirichlet characters then one can deduce Quadratic Reciprocity.) Conclude that (18) holds for all real characters if it holds for real characters mod q where q is either 4, 8, or an odd prime. Verify the three cases of even q, and show that if χ is the primitive real character modulo an odd prime q then

$$\tau(\chi) = \sum_{n \bmod q} e^{2\pi i n^2/q}. \tag{19}$$

8. Observe that $\sum_{n \bmod q} e^{2\pi i n^2/q}$ is the trace of the operator $T : \mathbf{C}^q \to \mathbf{C}^q$ that takes a complex-valued function c on $\mathbf{Z}/q\mathbf{Z}$ to its discrete Fourier transform $\hat{c}$. Show that $\hat{\hat{c}}(a) = q\,c(-a)$, and thus that each of the q eigenvalues λ of T is one of $\pm q^{1/2}$ or $\pm iq^{1/2}$. Thus we can evaluate $\tau(q) = \sum_\lambda \lambda$ by determining the multiplicity of each of these four eigenvalues. We already know that $\tau(\chi) = \pm q^{1/2}$ or $\pm iq^{1/2}$ according as $q \equiv 1$ or $-1 \bmod 4$. Check that this reduces the determination of $\tau(\chi)$ to computing $\det T$. Compute this determinant (Hint: T is represented by a Vandermonde matrix, and only the phase of $\det T$ is at issue because $|\det T| = \prod_\lambda |\lambda| = q^{q/2}$) to complete the proof of the formula (18) for the sign of $\tau(\chi)$.

The trace description of the sum in (19) does not depend on the primality of q, and yields $N^{-1/2}\sum_{n \bmod N} e^{2\pi i n^2/N} \in \mathbf{Z}[i]$ for all integers $N \geq 1$. It is known that in fact

$$N^{-1/2} \sum_{n \bmod N} e^{2\pi i n^2/N} = \frac{1 + (-i)^N}{1 - i};$$

that is, $1+i$, 1, 0, or i according as $N \equiv 0$, 1, 2 or 3 mod 4.

9. Can you find a τ analog of the duplication formula for the Gamma function?

On Jacobi sums:

10. For characters $\chi_1, \ldots, \chi_n$ modulo a prime q, define the generalized Jacobi sum $J(\chi_1, \ldots, \chi_n)$ by

$$J(\chi_1, \ldots, \chi_n) := \sum \cdots \sum \chi_1(a_1) \cdots \chi_n(a_n),$$

where the sum extends over all $(a_1, \ldots, a_n)$ mod q such that $a_1 + \cdots + a_n = 1$. Evaluate $J(\chi_1, \ldots, \chi_n)$ in terms of Gauss sums under suitable hypotheses on the χ_i. What is the analogous formula for definite integrals?

11. Let χ be the Legendre symbol modulo an odd prime q. Evaluate $\tau(\chi)^n$ in two ways to count the number of solutions mod q of $x_1^2 + \cdots + x_n^2 = 1$. If n is also an odd prime, use your formula to recover Quadratic Reciprocity (Hint: how many solutions are fixed under cyclic permutation of the x_i?). Can you modify this proof to also obtain the supplementary formula $(2/q) = \chi_8(q)$?

This proof of Quadratic Reciprocity can be modified to avoid explicit use of $\tau(\chi)$, because the solutions of $x_1^2 + \cdots + x_n^2 = 1$ can also be enumerated inductively by elementary means starting from results such as the parametrization of Pythagorean triples.

The usual way to recover Quadratic Reciprocity from the fact that $\tau(\chi)^2 = q^* = \pm q$ is to compare $\tau(\chi)$ with $\tau_n(\chi)$ mod n. On the one hand, they differ by a factor $\chi(n)$. On the other hand, if n is prime then $\tau_n(\chi) \equiv \tau(\chi)^n$ mod n, and thus equals $\tau(\chi)$ or $-\tau(\chi)$ according as q^* is a square mod n or not.

12. [Jacobi sums and Fermat curves mod q.] Let q be a prime, n a positive integer, and G the group of Dirichlet characters χ mod q such that $\chi^n = 1$. This is a cyclic group of order $\gcd(n, q-1)$ (why?). Prove that $\sum_{\chi,\chi' \in G} J(\chi, \chi')$ is the number of solutions of $x^n + y^n = 1$ in nonzero $x, y \in \mathbf{Z}/q\mathbf{Z}$. Conclude that this number is $q + O(n^2 q^{1/2})$, and thus that if "Fermat's Last Theorem" holds in $\mathbf{Z}/q\mathbf{Z}$ then $q \ll n^4$.

More precisely, the "Fermat curve" $F_n : x^n + y^n = z^n$ in the projective plane over $\mathbf{Z}/q\mathbf{Z}$ has $q + 1 - \sum_i \lambda_i$ points, where each λ_i is $-J(\chi, \chi')$ for one of the $(|G|-1)(|G|-2)$ choices of $\chi, \chi' \in G$ such that $\chi, \chi', \chi\chi'$ are all nontrivial. In particular, if $q \equiv 1$ mod n then $(|G|-1)(|G|-2) = (n-1)(n-2) = 2g$ where g is the genus of F_n. In this case the λ_i are the "eigenvalues of Frobenius" of F_n, and the fact that they all have norm $q^{1/2}$ is a special case of the Riemann hypothesis for the function field of an algebraic curve over a finite field. (If q is coprime to n but not 1 mod n then F_n still has $(n-1)(n-2)$ eigenvalues of Frobenius, which include the Jacobi sums but also other eigenvalues whose effect on the point counts of F_n appears only over finite extensions of $\mathbf{Z}/q\mathbf{Z}$.)

The last Exercise develops these ideas further for two curves of genus 1: the cubic Fermat curve F_3, and a quotient of F_4 whose $\mathbf{Q}$-rational points were determined by Fermat.

13. i) Suppose q is a prime congruent to 1 mod 3. It is known that q can be written uniquely as $(a^2 + 27b^2)/4$ for some positive integers a, b. Show that the number of solutions of $x^3 + y^3 = 1$ with $x, y \in \mathbf{Z}/q\mathbf{Z}$ is $q - 2 \pm a$, and determine the correct sign. Conclude that 2 is a cube mod q if and only if $2|a$, i.e., if and only if $q = m^2 + 27n^2$.
ii) Suppose q is an odd prime. How many solutions mod q do the equations $y^2 = x^4 - 1$ and $y^2 = x^3 - x$ have? (This should be easy if $q \equiv -1$ mod 4; for $q \equiv +1$ mod 4, cf. the previous $1\frac{1}{2}$ Exercises.)

These enumerations of rational points on $x^3 + y^3 = 1$ and $y^2 = x^3 - x$ mod q are now known to be special cases of the arithmetic of elliptic curves of complex multiplication; see for instance [Silverman 1986].

References

[Davenport 1967] Davenport, H.: *Multiplicative Number Theory.* Chicago: Markham, 1967; New York: Springer-Verlag, 1980 (GTM 74). [9.67.6 & 9.80.6 / QA 241.D32]

[Serre 1973] Serre, J.-P.: *A Course in Arithmetic.* New York: Springer, 1973 (GTM **7**). [AB 9.70.4 (reserve case) / QA243.S4713]

[Silverman 1986] Silverman, J.H.: *The Arithmetic of Elliptic Curves.* New York: Springer, 1986. [AB 9.86.1 (reserve case) / QA567.S44]

Math 259: Introduction to Analytic Number Theory

The asymptotic formula for primes in arithmetic progressions;
the Extended Riemann Hypothesis, concerning the zeros of $L(s,\chi)$

Now that we have the functional equation for $L(s,\chi)$, the asymptotics for $\psi(x,\chi)$, and thus also for $\psi(x, a \bmod q)$ and $\pi(x, a \bmod q)$, follow just as they did for $\psi(x)$ and $\pi(x)$ — at least if we are not very concerned with how the implied constants depend on q. We also state the Extended Riemann Hypothesis and relate it with conjectural improvements of the error estimates. The proofs are similar enough that we relegate most of the details to the Exercises. We shall soon see how to better control the dependence of our estimates on q. But the relatively crude bounds below are still of interest because these bounds, but not the later improvements for $L(s,\chi)$, generalize to other Dirichlet series such as the zeta functions of number fields; see the final Exercise.

Let χ be a primitive character mod $q > 1$. We readily adapt our argument showing that $(s^2 - s)\xi(s)$ is an entire function of order 1 to show that $\xi(s,\chi)$ is an entire function of order 1, and thus has a Hadamard product

$$\xi(s,\chi) = Ae^{Bs}s^{1-\mathsf{a}}\prod_{\rho}(1 - s/\rho)e^{s/\rho}. \tag{1}$$

Here $A = \xi(0,\chi)$ or $\xi'(0,\chi)$ according as χ is odd or even. The product ranges over zeros of $\xi(s,\chi)$, counted with multiplicity and excluding the simple zero at the origin if χ is even; that is, ρ ranges over the "nontrivial zeros" of $L(s,\chi)$, those with $\sigma \in [0,1]$. Thus

$$\frac{\xi'}{\xi}(s,\chi) = B + \frac{1-\mathsf{a}}{s} + \sum_{\rho}\left(\frac{1}{s-\rho} + \frac{1}{\rho}\right). \tag{2}$$

Note that B depends on χ; see Exercise 1 below. Fortunately it will usually cancel out from our formulas. It follows that

$$\frac{L'}{L}(s,\chi) = B - \frac{1}{2}\log\frac{q}{\pi} - \frac{1}{2}\frac{\Gamma'}{\Gamma}((s+\mathsf{a})/2) + \frac{1-\mathsf{a}}{s} + \sum_{\rho}\left(\frac{1}{s-\rho} + \frac{1}{\rho}\right). \tag{3}$$

How are these zeros ρ distributed? We noted already that their real parts lie in $[0,1]$. If $L(\rho,\chi) = 0$ then by the functional equation $0 = L(1-\rho,\overline{\chi}) = L(1-\bar{\rho},\chi)$. Thus the zeros are symmetrical about the line $\sigma = 1/2$, but not (unless χ is real) about the real axis. So the proper analog of $N(T)$ is half of $N(T,\chi)$, where $N(T,\chi)$ is defined as the number of zeros of $L(s,\chi)$ in $\sigma \in (0,1)$, $|t| < T$, counted with multiplicity. [NB this excludes the trivial zero at $s = 0$, which occurs for even χ.] Again we evaluate this by integrating ξ'/ξ around a rectangle. The new factor $q^{s/2}$ in $\xi(s,\chi)$ introduces an extra term of $(T/2\pi)\log q$ into the formula for $N(T,\chi)/2$. That factor is also responsible for the new term $-\frac{1}{2}\log q$ in (3),

which forces us to subtract $O(\log q)$ from our lower bound on the real part of $(L'/L)(s,\chi)$. This bound now becomes

$$\frac{L'}{L}(s,\chi) = \sum_{|\operatorname{Im}(s-\rho)|<1} \frac{1}{s-\rho} + O(\log|qt|)$$

($\sigma \in [-1,2]$), the sum comprising $O(\log|qt|)$ terms. We conclude:

Theorem. *The estimate*

$$\frac{1}{2}N(T,\chi) = \frac{T}{2\pi}\log\frac{qT}{2\pi} - \frac{T}{2\pi} + O(\log qT) \tag{4}$$

holds for all $T \geq 2$, with an implied constant independent of q.

(The lower bound on T could be replaced by any $T_0 > 1$, possibly changing the implied constant if $q = 1$ and T_0 is too close to 1.)

Proof: See Exercise 2. □

To isolate the primes in arithmetic progressions mod q, we need also characters that are not primitive, such as χ_0. Let χ_1 be the primitive character mod $q_1|q$ underlying a nonprimitive χ mod q. Then

$$L(s,\chi) = \prod_{p|q}\bigl(1-\chi_1(p)p^{-s}\bigr)\cdot L(s,\chi_1).$$

The elementary factor $\prod_{p|q}(1-\chi_1(p)p^{-s})$ has, for each p dividing q but not q_1, a total of $(T/\pi)\log p + O(1)$ purely imaginary zeros of absolute value $< T$. This, together with the estimate (4) for $N(T,\chi_1)$, shows that the RHS of (4) is an upper bound on $\frac{1}{2}N(T,\chi)$, even when χ is not principal.

The horizontal distribution of ρ is subtler. We noted already that the logarithmic derivative of

$$\zeta_q(s) := \prod_{\chi \bmod q} L(s,\chi)$$

is a Dirichlet series $-\sum_n \Lambda_q(n)n^{-s}$ with $\Lambda_q(n) \geq 0$ for all n, and deduced that the $3+4\cos\theta+\cos 2\theta$ trick shows that ζ_q, and thus each factors $L(s,\chi)$, does not vanish at $s = 1+it$. We can then adapt the proof of the classical zero-free region for $\zeta(s)$; since, however, $\zeta_q(s)$ is the product of $\varphi(q)$ L-series, each of which contributes $O(\log|qt|)$ to the bound on $(\zeta_q'/\zeta_q)(\sigma+it)$, the resulting zero-free region is not $1-\sigma < c/\log|t|$ or even $1-\sigma < c/\log|qt|$ but $1-\sigma < c/(\varphi(q)\log|qt|)$. Moreover, the fact that this only holds for say $|t| > 2$ is newly pertinent: unlike $\zeta(s)$, the L-series might have zeros of small imaginary part. [Indeed it is known that there are Dirichlet L-series that vanish on points arbitrarily close to the real axis.] Still, for every q there are only finitely many zeros with $|t| \leq 2$. So our formula

$$\psi(x,\chi) = \frac{1}{2\pi i}\int_{1+\frac{1}{\log x}-iT}^{1+\frac{1}{\log x}+iT} -\frac{L'}{L}(s,\chi)\,x^s\frac{ds}{s} + O\left(\frac{x\log^2 x}{T}\right) \qquad (T \in [1,x])$$

yields an estimate as before, with only the difference that when $\chi \neq \chi_0$ there is no "main term" coming from a pole at $s = 1$. We thus find

$$\psi(x, \chi) \ll_\chi x \exp(-C_\chi \sqrt{\log x}) \tag{5}$$

for some constant $C_\chi > 0$. Multiplying by $\overline{\chi}(a)$ and averaging over χ (including χ_0, for which $\psi(x, \chi_0) = x + O(x \exp(-C\sqrt{\log x}))$ instead of (5), we obtain

$$\psi(x, a \bmod q) = \frac{1}{\varphi(q)} x + O_q(x \exp(-C_q \sqrt{\log x})), \tag{6}$$

and thus

$$\pi(x, a \bmod q) = \frac{1}{\varphi(q)} \operatorname{li}(x) + O_q(x \exp(-C_q \sqrt{\log x})). \tag{7}$$

Note however that the dependence of the error terms on q is unpredictable. The zero-free region depends explicitly on q (though as we shall see it need not shrink nearly as fast as $1/(\varphi(q) \log q)$, a factor which alone would make C_q proportional to $(\varphi(q) \log q)^{-1/2})$, but it excludes a neighborhood of the real axis. It would then seem that to specify C_χ and $\ll_\chi$ we would have to compute for each χ the largest $\operatorname{Re}(\rho)$. There's also the matter of the contribution of the B's from (3).

Consider, by comparison, the consequences of the *Extended Riemann Hypothesis* (ERH), which is the conjecture that each nontrivial zero ρ of an L-series associated to a primitive Dirichlet character χ has real part $1/2$.[1] Our analysis of $\psi(x)$ under RH then carries over almost verbatim to show that $\psi(x, \chi) \ll x^{1/2} \log^2 x$ as long as $q < x$ with an absolute and effective implied constant, and thus that

$$\psi(x, a \bmod q) = \frac{x}{\varphi(q)} + O(x^{1/2} \log^2 x) \tag{8}$$

holds if $L(s, \chi)$ satisfies ERH for all Dirichlet characters mod q_1 with $q_1 | q$, again with the O-constant effective and independent of q. It would also follow that

$$\pi(x, a \bmod q) = \frac{\operatorname{li}(x)}{\varphi(q)} + O(x^{1/2} \log x). \tag{9}$$

Again there are some remarkable effects of the difference between $\psi(x, a \bmod q)$ and the sum of $\log p$ over the primes counted in $\pi(x, a \bmod q)$. Most strikingly, for nontrivial real characters χ, $\pi(x, \chi)$ tends to be negative, because the contribution of $\sum_{n<x} \chi(n^2)\Lambda(n^2)$ to $\psi(x, \chi)$ is asymptotic to $+\sqrt{x}$. For instance, $\pi(x, 1 \bmod 4) < \pi(x, 3 \bmod 4)$ for "most" x, where one might intuitively expect that $\pi(x, \chi_4)$ could as easily be positive as negative. This is the "Chebyshev's Bias" of the title of [RS 1994], and [BFHR 2001], to which we refer for more precise statements, as well as subtler effects of this kind and numerical computations of the theoretical and "experimental" sizes of these biases.

[1] Attributed by Davenport to "Piltz in 1884" (page 129). We distinguish ERH from the *Generalized Riemann Hypothesis* (GRH), which pertains to much more general Dirichlet series, such as a zeta function of a number field or the L-series attached to a modular form.

Exercises

1. Show that the real part of the term B of (2) is $-\sum_\rho \mathrm{Re}(1/\rho)$. Conclude that $\mathrm{Re}(B) < 0$. [Davenport 1967, page 85.]

In the next three Exercises you will fill in the proofs outlined above of formulas for $N(T,\chi)$, $\psi(x,\chi)$ $\psi(x, a \bmod q)$, and $\pi(x, a \bmod q)$. Again the complete proofs may be found in [Davenport 1967] and elsewhere.

2. Complete the missing steps in the proof of (4).

3. Complete the missing steps in the proof of (5), (6), and (7).

4. Verify that, under the relevant ERH, the O-constant in (9) does not depend on q. Obtain an analogous estimate on the weaker assumption that ζ_q has no zeros of real part $> \theta$ for some $\theta \in (\frac{1}{2}, 1)$. Show that if for some q we have $\pi(x, a \bmod q) \ll_\epsilon x^{\theta+\epsilon}$ for all $a \in (\mathbf{Z}/q)^*$ then all the $L(s,\chi)$ for Dirichlet characters $\chi \bmod q$ are nonzero on $\sigma > \theta$.

5. [The Prime Number Theorem for number fields] Let K be a number field of degree $n = r_1 + 2r_2$. We already defined the zeta function

$$\zeta_K(s) := \sum_I |I|^{-s} = \prod_\wp (1 - |\wp|^{-s})^{-1} \qquad (\sigma > 1),$$

in which $|I|$ is the norm of an ideal I, and the sum and product extend respectively over nonzero ideals I and prime ideals $\wp$ of the ring of integers O_K. We also reported that ζ_K is known to extend to a meromorphic function on $\mathbf{C}$, regular except for a simple pole at $s = 1$, that satisfies a functional equation $\xi_K(s) = \xi_K(1-s)$, where

$$\xi_K(s) := \Gamma(s/2)^{r_1} \Gamma(s)^{r_2} (4^{-r_2} \pi^{-n} |d|)^{s/2} \zeta_K(s)$$

and d is the discriminant of K. In particular, $(s^2 - s)\xi_K(s)$ is an entire function; it is also known that it is an entire function of order 1. Use this to obtain an approximation of the number of nontrivial zeros ρ of ζ_K such that $|\mathrm{Im}(\rho)| \le T$, with an error estimate depending explicitly on n and $|d|$. Obtain a zero-free region for ζ_K, and deduce that

$$\#\{\wp : |\wp| \le x\} = \mathrm{li}(x) + O(x \exp(-C_K \sqrt{\log x}))$$

for some constant $C_K > 0$.

Math 259: Introduction to Analytic Number Theory

A nearly zero-free region for $L(s,\chi)$, and Siegel's theorem

We used positivity of the logarithmic derivative of ζ_q to get a crude zero-free region for $L(s,\chi)$. Better zero-free regions can be obtained with some more effort by working with the $L(s,\chi)$ individually. The situation is most satisfactory for complex χ, that is, for characters with $\chi^2 \neq \chi_0$. (Recall that real χ were also the characters that gave us the most difficulty in the proof of $L(1,\chi) \neq 0$; it is again in the neighborhood of $s = 1$ that it is hard to find a good zero-free region for the L-function of a real character.)

To obtain the zero-free region for $\zeta(s)$, we started with the expansion of the logarithmic derivative

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

and applied the inequality

$$0 \leq \frac{1}{2}(z + 2 + \bar{z})^2 = \mathrm{Re}(z^2 + 4z + 3) = 3 + 4\cos\theta + \cos 2\theta \qquad (z = e^{i\theta})$$

to the phases $z = n^{-it} = e^{i\theta}$ of the terms n^{-s}. To apply the same inequality to

$$-\frac{L'}{L}(s,\chi) = \sum_{n=1}^{\infty} \chi(n)\Lambda(n) n^{-s},$$

we must use $z = \chi(n)n^{-it}$ instead of n^{-it}, obtaining

$$0 \leq \mathrm{Re}\big(3 + 4\chi(n)n^{-it} + \chi^2(n)n^{-2it}\big).$$

Multiplying by $\Lambda(n)n^{-\sigma}$ and summing over n yields

$$0 \leq 3\left[-\frac{L'}{L}(\sigma,\chi_0)\right] + 4\,\mathrm{Re}\left[-\frac{L'}{L}(\sigma+it,\chi)\right] + \mathrm{Re}\left[-\frac{L'}{L}(\sigma+2it,\chi^2)\right]. \qquad (1)$$

Now we see why the case $\chi^2 = \chi_0$ will give us trouble near $s = 1$: for such χ the last term in (1) is within $O(1)$ of $\mathrm{Re}(-(\zeta'/\zeta)(\sigma + 2it))$, so the pole of $(\zeta'/\zeta)(s)$ at $s = 1$ will undo us for small $|t|$.

Let us see how far (1) does take us. Our bounds will involve $\log q$ for small $|t|$, and $\log q|t|$ for large $|t|$. To cover both ranges, and also to accommodate the case $q = 1$ (the Riemann zeta function), we use the convenient and conventional abbreviation

$$\mathscr{L} := \log q(|t| + 2).$$

We shall prove:

Theorem. *There is a constant $c > 0$ such that if $L(\sigma + it, \chi) = 0$ for some primitive complex Dirichlet character $\chi \bmod q$ then*

$$\sigma < 1 - \frac{c}{\mathscr{L}}\,. \tag{2}$$

If χ is a real primitive character then (2) holds for all zeros of $L(s, \chi)$ with at most one exception. The exceptional zero, if it exists, is real and simple.

Proof: We again apply (1) for suitable σ, t with $1 < \sigma \le 2$. The first term is

$$\le 3\left[-\frac{\zeta'}{\zeta}(\sigma)\right] < \frac{3}{\sigma - 1} + O(1).$$

For the remaining terms, we use the partial-fraction expansion

$$-\frac{L'}{L}(s,\chi) = \frac{1}{2}\log\frac{q}{\pi} + \frac{1}{2}\frac{\Gamma'}{\Gamma}((s+\mathsf{a})/2) - B_\chi - \frac{1-\mathsf{a}}{s} - \sum_\rho\left(\frac{1}{s-\rho} + \frac{1}{\rho}\right).$$

To eliminate the contributions of B_χ and $\sum_\rho 1/\rho$ we use this formula to evaluate $(L'/L)(2,\chi) - (L'/L)(s,\chi)$. By the Euler product we have $(L'/L)(2,\chi) = O(1)$. Since also $1/2 - 1/s = O(1)$, we have

$$-\frac{L'}{L}(s,\chi) = \frac{1}{2}\frac{\Gamma'}{\Gamma}((s+\mathsf{a})/2) - \sum_\rho\left(\frac{1}{s-\rho} - \frac{1}{2-\rho}\right) + O(1).$$

Next take real parts. For ρ of real part in $[0,1]$ we have $\mathrm{Re}(1/(2-\rho)) \ll |2-\rho|^{-2}$. To estimate the sum of this over all ρ, we may apply Jensen's theorem to $\xi(2+s,\chi)$, finding that the number (with multiplicity) of $|\rho|$ at distance at most r from 2 is $O(r\log qr)$, and thus by partial summation that $\sum_\rho |2-\rho|^{-2} \ll \log q$. We estimate the real part of the Γ'/Γ term by Stirling as usual, and find

$$\mathrm{Re}\left[-\frac{L'}{L}(s,\chi)\right] < O(\mathscr{L}) - \sum_\rho \mathrm{Re}\,\frac{1}{s-\rho}.$$

Again each of the $\mathrm{Re}(1/(s-\rho))$ is nonnegative, so the estimate remains true if we include only some of the zeros ρ, or none of them.

In particular it follows that

$$\mathrm{Re}\left[-\frac{L'}{L}(\sigma + 2it, \chi^2)\right] < O(\mathscr{L}), \tag{3}$$

at least when χ^2 is a primitive character. If χ^2 is not primitive, but still not the trivial character χ_0, then (3) holds when χ^2 is replaced by its corresponding primitive character; but the error thus introduced is at most

$$\sum_{p|q}\frac{p^{-\sigma}}{1-p^{-\sigma}}\log p < \sum_{p|q}\log p \le \log q < \mathscr{L},$$

so can be absorbed into the $O(\mathscr{L})$ error. But when $\chi^2 = \chi_0$ the partial-fraction expansion of its $-L'/L$ has a term $+1/(s-1)$ which cannot be discarded, and can absorbed into $O(\mathscr{L})$ only if s is far enough from 1. We thus conclude that (3) holds unless $\chi^2 = \chi_0$ and $|t| < c/\log q$, the implied constant in (3) depending on c. (Equivalently, we could change $O(\mathscr{L})$ to $1/|t| + O(\mathscr{L})$ when $\chi^2 = \chi_0$.)

The endgame is the same as we have seen for the classical zero-free region for $\zeta(s)$: if there is a zero $\rho = 1 - \delta + it$ with δ small, use its imaginary part t in (1) and find from the partial-fraction expansion that

$$\operatorname{Re}\left[-\frac{L'}{L}(\sigma + it, \chi)\right] < O(\mathscr{L}) - \frac{1}{\sigma - \operatorname{Re}(\rho)}.$$

Combining this with our previous estimates yields

$$\frac{4}{\sigma + \delta - 1} < \frac{3}{\sigma - 1} + O(\mathscr{L});$$

choosing $\sigma = 1 + 4\delta$ as before yields $1/\delta \ll O(\mathscr{L})$ under the hypotheses of (3), completing the proof of (2) with the possible exception of real χ and zeros of imaginary part $\ll 1/\log q$.

Next suppose that χ is a real character and fix some $\delta > 0$, to be chosen later. We have a zero-free region for $|t| \geq \delta/\log q$. To deal with zeros of small imaginary part, let $s = \sigma$ in (1) — or, more simply, use the inequality[1] $1 + \operatorname{Re}(e^{i\theta}) \geq 0$ — to find

$$\sum_{|\operatorname{Im}(\rho)| < \delta/\log q} \operatorname{Re}(1/(\sigma - \rho)) < \frac{1}{\sigma - 1} + O(\log q),$$

the implied O-constant not depending on δ. Each term $\operatorname{Re}(1/(\sigma-\rho))$ equals $\operatorname{Re}(\sigma-\rho)/|\sigma-\rho|^2$. Choosing $\sigma = 1+(2\delta/\log q)$ we find that $|\operatorname{Im}(\rho)| < \frac{1}{2}(\sigma-1) < \frac{1}{2}\operatorname{Re}(\sigma-\rho)$, and thus that $|\sigma-\rho|^2 < \frac{5}{4}\operatorname{Re}(\sigma-\rho)^2$. Therefore $\operatorname{Re}(1/(\sigma-\rho)) > \frac{4}{5}/\operatorname{Re}(\sigma-\rho)$. So,

$$\frac{4}{5} \sum_{|\operatorname{Im}(\rho)| < \delta/\log q} \left(1 - \operatorname{Re}(\rho) + \frac{2\delta}{\log q}\right)^{-1} < \frac{\log q}{2\delta} + A \log q,$$

for some constant A independent of the choice of δ. Therefore, if δ is small enough, we can find $c > 0$ such that at most one ρ can have real part greater than $1 - c/\log q$. (Specifically, we may choose any $\delta < 3/10A$, and take $c = 2\delta(3 - 10A\delta)/5(1 + 2A\delta)$.) Since ρ's are counted with multiplicity and come in complex conjugate pairs, it follows that this exceptional zero, if it exists, is real and simple. $\square$

This exceptional zero is usually denoted by β. Of course we expect, by the Extended Riemann Hypothesis, that there is no such β. The nonexistence of β, though much weaker than ERH, has yet to be proved; but we can still obtain

[1] That is, use the positivity of $-\zeta_\chi'/\zeta_\chi$, where $\zeta_\chi(s) = \zeta(s)L(s,\chi)$ is the zeta function of the quadratic number field corresponding to χ.

some strong restrictions on how β can vary with q and χ. We begin by showing that at most one of the Dirichlet characters mod q can have an L-series with an exceptional zero, and deduce a stronger estimate on the error terms in our approximate formulas for $\psi(x, a \bmod q)$ and $\pi(x, a \bmod q)$. Since χ need not be primitive, it follows that in fact β cannot occur even for characters of different moduli if we set the threshold low enough:

Theorem. [Landau 1918] *There is a constant $c > 0$ such that, for any distinct primitive real characters χ_1, χ_2 to (not necessarily distinct) moduli q_1, q_2 at most one of $L(s, \chi_1)$ and $L(s, \chi_2)$ has an exceptional zero $\beta > 1 - c/\log q_1 q_2$.*

Proof: Since χ_1, χ_2 are distinct primitive real characters, their product $\chi_1\chi_2$, while not necessarily primitive, is also a nontrivial Dirichlet character, with modulus at most $q_1 q_2$. Hence $-(L'/L)(\sigma, \chi_1\chi_2) < O(\log q_1 q_2)$ for $\sigma > 1$. Let[2]

$$F(s) = \zeta(s)L(s, \chi_1)L(s, \chi_2)L(s, \chi_1\chi_2) \tag{4}$$

Then $-F'/F$ is the sum of the negative logarithmic derivatives of $\zeta(s)$, $L(s, \chi_1)$, $L(s, \chi_2)$, and $L(s, \chi_1\chi_2)$, which is the positive Dirichlet series

$$\sum_{n=1}^{\infty} (1 + \chi_1(n))(1 + \chi_2(n))\Lambda(n)n^{-s}.$$

In particular, this series is positive for real $s > 1$. Arguing as before, we find that if β_i are exceptional zeros of $L(s, \chi_i)$ then

$$\frac{1}{\sigma - \beta_1} + \frac{1}{\sigma - \beta_2} < \frac{1}{\sigma - 1} + O(\log q_1 q_2);$$

if $\beta_i > 1 - \delta$ then we may take $\sigma = 1 + 2\delta$ to find $1/6\delta < O(\log q_1 q_2)$, whence $\delta \gg 1/\log q_1 q_2$ as claimed. $\square$

In particular, for each q there is at most one real character mod q whose L-series has an exceptional zero $\beta > 1 - (c/\log q)$. This lets us obtain error terms that depend explicitly on q and (if it exists) β in the asymptotic formulas for $\psi(x, a \bmod q)$ and $\pi(x, a \bmod q)$. For instance (see e.g. Chapter 20 of [Davenport 1967]), we have:

Theorem. *For every $C > 0$ there exists $c > 0$ such that whenever $\gcd(a, q) = 1$ we have*

$$\psi(x, a \bmod q) = \left(1 + O(\exp -c\sqrt{\log x})\right) \frac{x}{\varphi(q)} \tag{5}$$

and

$$\pi(x, a \bmod q) = \left(1 + O(\exp -c\sqrt{\log x})\right) \frac{\mathrm{li}(x)}{\varphi(q)} \tag{6}$$

[2] If $\chi_1\chi_2$ is itself primitive then $F(s)$ is the zeta function of a biquadratic number field K, namely the compositum of the quadratic fields corresponding to χ_1 and χ_2. In general $\zeta_K(s) = \zeta(s)L(s, \chi_1)L(s, \chi_2)L(s, \chi_3)$ where χ_3 is the primitive character underlying $\chi_1\chi_2$; thus $F(s)$ always equals $\zeta_K(s)$ multiplied by a finite Euler product.

for all $x > \exp(C \log^2 q)$, unless there is a Dirichlet character $\chi \bmod q$ for which $L(s,\chi)$ has an exceptional zero β, in which case

$$\psi(x, a \bmod q) = \left(1 - \frac{\chi(a) x^{\beta-1}}{\beta} + O(\exp -c\sqrt{\log x})\right) \frac{x}{\varphi(q)} \tag{7}$$

and

$$\pi(x, a \bmod q) = \frac{1}{\varphi(q)} \left(\mathrm{li}(x) - \mathrm{li}(x^\beta) + O(x \exp -c\sqrt{\log x})\right) \tag{8}$$

for all $x > \exp(C \log^2 q)$. The constant implicit in each $O(\cdot)$ may depend on C but not on q.

Proof: See the Exercises. □

Just how close can this β come to 1? We first show that very small $1-\beta$ imply small $L(1,\chi)$. Since $L(1,\chi) = \int_\beta^1 L'(\sigma,\chi)\, d\sigma$, it is enough to prove an upper bound on $|L'(\sigma,\chi)|$ for σ near 1. We show:

Lemma. *There exists an absolute constant C such that $|L'(\sigma,\chi)| < C \log^2 q$ for any nontrivial Dirichlet character $\chi \bmod q$ and any $\sigma \le 1$ such that $1 - \sigma \le 1/\log q$.*

Proof: We may assume $q > 2$, so that the series $\sum_{n=1}^\infty \chi(n)(\log n) n^{-\sigma}$ for $-L'(\sigma,\chi)$ converges if $1 - \sigma \le 1/\log q$. Split this sum into $\sum_{n \le q} + \sum_{n > q}$. The first sum is $O(\log^2 q)$, because the n-th term has abolute value at most

$$\frac{n^{1-\sigma}}{n} \log n \le \frac{q^{1-\sigma}}{n} \log n \le \frac{e}{n} \log n.$$

The sum over $n > q$ can be bounded by partial summation together with the crude estimate $|\sum_q^N \chi(n)| < q$, yielding an upper bound $e \log q$, which is again $O(\log^2 q)$. □

Corollary. *If for some Dirichlet character $\chi \bmod q$ the L-series $L(s,\chi)$ has a zero $\beta > 1 - (1/\log q)$ then $L(1,\chi) < C(1-\beta)\log^2 q$.*

But the Dirichlet class number formula for the quadratic number field corresponding to χ gives $L(1,\chi) \gg q^{-1/2}$. (We shall soon prove this directly.) Therefore

$$1 - \beta \gg \frac{1}{q^{1/2} \log^2 q}.$$

Siegel [Siegel 1935] proved a much better inequality:

Theorem. *For each $\epsilon > 0$ there exists $C_\epsilon > 0$ such that*

$$L(1,\chi) > C_\epsilon q^{-\epsilon}$$

holds for all real Dirichlet characters $\chi \bmod q$. Hence there exists $C'_\epsilon > 0$ such that any zero β of $L(s,\chi)$ satisfies

$$1 - \beta > C'_\epsilon q^{-\epsilon}.$$

Proof: Let χ_1, χ_2 be different primitive real characters to moduli $q_1, q_2 > 1$, and let

$$\lambda = L(1,\chi_1)L(1,\chi_2)L(1,\chi_1\chi_2) = (s-1)F(s)\big|_{s=1},$$

with $F(s)$ as in (4). We shall prove that there exist universal constants $\theta < 1$ and $A, B, C > 0$ such that

$$F(s) > A - \frac{B\lambda}{1-s}(q_1 q_2)^{C(1-s)} \tag{9}$$

holds for all $s \in (\theta, 1)$. (Specifically, we can use $\theta = 9/10$ and $A = 1/2$, $C = 8$.) Assume (9) for the time being. Since $F(s)$ is positive for $s > 1$ and has a simple pole at $s = 1$, we have $F(\beta) \le 0$ for any $\beta \in (\theta, 1)$ such that F has no zero in $(\theta, 1)$. Of course $F(\beta) \le 0$ also holds if β is a zero of F. For such β we have

$$\lambda > \frac{A}{B}(1-\beta)(q_1 q_2)^{-C(1-\beta)}. \tag{10}$$

We shall fix χ_1 and β and use (10) to deduce a lower bound on $L(1, \chi_2)$ for all $\chi_2 \bmod q_2$ such that $q_2 > q_1$. If there is some real χ_1 such that $L(\beta_1, \chi_1) = 0$ for some $\beta_1 > 1 - (\epsilon/2C)$ then we use that character for χ_1 and the zero β_1 for β. Otherwise $F(s)$ never has a zero in $(1-(\epsilon/2C), 1)$, so we choose χ_1 arbitrarily and β subject to $0 < 1 - \beta < \epsilon/2C$. Then for any primitive $\chi_2 \bmod q_2 > q_1$ we use (10), together with the upper bound $L(1,\chi) \ll \log q$ (see the Exercises), to find that

$$L(1,\chi_2) > c\, q_2^{-C(1-\beta)} / \log q_2,$$

with c depending only (but ineffectively!) on ϵ via χ_1 and β. Since $C(1-\beta) < \epsilon/2$, Siegel's theorem follows.

It remains to prove (9). Siegel originally showed this using class field theory; we follow the more direct approach of [Estermann 1948]. (See also [Chowla 1950] for another direct proof.)

Since $F(s)$ has a nonnegative Dirichlet series, its Taylor series about $s = 2$ is

$$F(s) = \sum_{m=0}^{\infty} b_m (2-s)^m$$

with $b_0 = F(2) > 1$ and all $b_m > 0$. Since F is entire except for a simple pole of residue λ at $s = 1$, we have the Taylor expansion

$$F(s) - \frac{\lambda}{s-1} = \sum_{m=0}^{\infty} (b_m - \lambda)(2-s)^m,$$

valid for all $s \in \mathbf{C}$. Consider this on $|s-2| = 3/2$. We have there the crude bounds $L(s,\chi_1) \ll q_1$, $L(s,\chi_2) \ll q_2$, $L(s,\chi_1\chi_2) \ll q_1 q_2$, and of course $\zeta(s)$ is

bounded on $|s-2| = 3/2$. So, $F(s) \ll (q_1 q_2)^2$ on this circle, and thus the same is true of $F(s) - \lambda/(s-1)$. Hence

$$|b_m - \lambda| \ll (2/3)^m (q_1 q_2)^2.$$

For any fixed $\theta \in (1/2, 1)$ it follows that

$$\sum_{m=M}^{\infty} |b_m - \lambda| \, (2-s)^m \ll (q_1 q_2)^2 \Big(\frac{2}{3}(2-\theta)\Big)^M$$

holds for all $s \in (\theta, 1)$. Since $b_0 > 1$ and $b_m \geq 0$, we thus have

$$F(s) - \frac{\lambda}{s-1} \geq 1 - \lambda \frac{(2-s)^M - 1}{1-s} - O(q_1 q_2)^2 \Big(\frac{2}{3}(2-\theta)\Big)^M.$$

Let M be the largest integer such that the error estimate $O(q_1 q_2)^2((4-2\theta)/3)^M$ is $< 1/2$. Then

$$F(s) > \frac{1}{2} - \frac{\lambda}{1-s}(2-s)^M.$$

But

$$(2-s)^M = \exp(M \log(2-s)) < \exp M(1-s),$$

and $\exp M \ll (q_1 q_2)^{O(1)}$, which completes the proof of (9) and thus of Siegel's theorem.

Remarks

Unfortunately the constants C_ϵ and C'_ϵ in Siegel's theorem, unlike all such constants that we have encountered so far, are ineffective for every $\epsilon < 1/2$, and remain ineffective almost seventy years later. This is because we need more than one counterexample to reach a contradiction. What can be obtained effectively are constants C_ϵ such that $1 - \beta > C_\epsilon q^{-\epsilon}$ holds for every q and all real primitive characters mod q, with at most a single exception (q, χ, β). This β, if it exists, is called the "Siegel zero" or "Siegel-Landau zero". Note that a zero β of some $L(s, \chi)$ that violates the ERH may or may not qualify as a Siegel(-Landau) zero depending on the choice of ϵ and C_ϵ.

Dirichlet's class number formula relates $L(1, \chi)$ for real characters χ with the class number of quadratic number fields. Siegel's theorem and its refinements thus yield information on these class numbers. For instance, if χ is an odd character then the imaginary quadratic field $K = \mathbf{Q}(\sqrt{-q}\,)$ has a zeta function $\zeta_K(s)$ that factors into $\zeta(s)L(s, \chi)$. Let $h(K)$ be the class number of K. Siegel's theorem, together with Dirichlet's formula, yields the estimate $h(K) \gg_\epsilon q^{1/2-\epsilon}$. In particular, $h(K) > 1$ for all but finitely many K. But this does not reduce the determination of all such K to a finite computation because the implied constant cannot be made effective. Gauss had already conjectured in 1801 (in terms of binary quadratic forms, see *Disq. Arith.*, §303) that $h(K) = 1$ only for $K = \mathbf{Q}(\sqrt{-q}\,)$ with

$$q = 3,\ 4,\ 7,\ 8,\ 11,\ 19,\ 43,\ 67,\ 163.$$

But the closest that Siegel's method can bring us to this conjecture is the theorem that there is at most one further such q. Heilbronn and Linfoot showed this a year before Siegel's theorem [HL 1934] using a closely related method [Heilbronn 1934].

It was only with much further effort that Heegner ([1952], corrected by Deuring [1968]), and later Baker [1966] and Stark [1967] (working independently and using different approaches), proved Gauss's conjecture, at last exorcising the "tenth discriminant". None of these approaches yields an effective lower bound on $h(K)$ that grows without limit as $q \to \infty$. Such a bound was finally obtained by Goldfeld, Gross, and Zagier ([Goldfeld 1976], [GZ 1986]), but it grows much more slowly than $q^{1/2-\epsilon}$; namely, $h(K) > c \log q / \log\log q$ (and $h(K) > c' \log q$ for prime q). Even this was a major breakthrough that combined difficult algebraic and analytic techniques. See [Goldfeld 1986] for an overview.

Exercises

1. Complete the derivation of (5,6,7,8) from our (nearly) zero-free region for Dirichlet L-functions.

2. Show that for each $\epsilon > 0$ there exists C such that whenever $\gcd(a, q) = 1$ we have

$$|\varphi(q)\, \psi(x, a \bmod q) - x| < \epsilon x$$

for all $x > q^C$, unless there is a Dirichlet character $\chi \bmod q$ for which $L(s, \chi)$ has an exceptional zero β, in which case

$$|\varphi(q)\, \psi(x, a \bmod q) - x - (x^\beta/\beta)| < \epsilon x$$

for all $x > q^C$. [NB $q^C = \exp(C \log q)$.] Use this argument, together with the fact that $\psi(x, a \bmod q) \geq 0$, to give an alternative proof of Landau's theorem (at most one exceptional zero for any character mod q). Also, obtain the corresponding estimates for $\pi(x, a \bmod q)$, and deduce that there is an absolute constant C such that if there is no exceptional zero then there exists a prime $p \equiv a \bmod q$ with $p < q^C$. (Linnik [1944] showed this unconditionally, by showing that very small values of $1 - \beta$ force the other zeros farther away from the line $\sigma = 1$.)

3. Check that $|L'(\sigma, \chi)| \ll_A \log^2 q$ holds in any interval $0 \leq 1 - \sigma \leq A/\log q$. How does the implied constant depend on A? Use the same method to show that also $|L(\sigma, \chi)| \ll \log q$ in the same interval, and in particular at $\sigma = 1$. What bound can you obtain on the higher derivatives of $L(\sigma, \chi)$ for σ near 1?

4. Prove that for every $\epsilon > 0$ there exist positive constants A, c such that if $L(1, \chi) < Aq^{-\epsilon}$ for some primitive Dirichlet character $\chi \bmod q$ then $L(\beta, \chi) = 0$ for some β such that $1 - \beta < c/\log q$. [Use the fact that

$$L(1, \chi) = L(\sigma, \chi)\, \exp \int_1^\sigma -\frac{L'}{L}(s, \chi)\, ds,$$

or work directly with the product formula for $L(s, \chi)$. This Exercise, which

shows that small $L(1,\chi)$ implies small $1-\beta$, may be regarded as a qualitative converse of our computation showing that small $1-\beta$ implies small $L(1,\chi)$.]

References

[Baker 1966] Baker, A.: Linear forms in the logarithms of algebraic numbers, *Mathematika* **13** (1966), 204–216.

[Chowla 1950] Chowla, S.: A new proof of a theorem of Siegel. *Annals of Math.* (2) **51** (1950), 120–122.

[Deuring 1968] Deuring, M.: Imaginäre quadratische Zahlkörper mit der Klassenzahl Eins, *Invent. Math.* **5** (1968), 169–179.

[Estermann 1948] Estermann, T.: On Dirichlet's L functions, *J. London Math. Soc.* **23** (1948), 275–279.

[Goldfeld 1976] Goldfeld, D.M.: The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3** (1976) #4, 624–663.

[Goldfeld 1986] Goldfeld, D.M.: Gauss's class number problem for imaginary quadratic fields, *Bull. Amer. Math. Soc.* (N.S.) **13** (1985) #1, 23–37.

[GZ 1986] Gross, B.H., Zagier, D.B.: Heegner points and derivatives of L-series, *Invent. Math.* **84** (1986) #2, 225–320.

[Heegner 1952] Heegner, K.: Diophantische Analysis und Modulfunktionen, *Math. Z.* **56** (1952), 227–253.

[Heilbronn 1934] Heilbronn, H.: On the Class-number in Imaginary Quadratic Fields, *Quarterly J. Math.* **5** (1934), 150–160.

[HL 1934] Heilbronn, H., Linfoot, E.H.: On the imaginary quadratic corpora of class-number one, *Quarterly J. Math.* **5** (1934), 293–301;

[Landau 1918] Landau, E: Über die Classenzahl imaginär-quadratischer Zahlkörper, *Göttinger Nachrichten* 1918, 285–295. [= pages 150–160 of *Collected Works* Vol.VII, Essen: Thales, 1985.] [QA3.L24]

[Linnik 1944] Linnik, U.V.: On the Least Prime in an Arithmetic Progression. I. The Basic Theorem; II. The Deuring-Heilbronn Phenomenon. *Mat. Sbornik* (N.S.) **15 (57)** (1944), 139–178 and 347–368.

[Siegel 1935] Siegel, C.L.: Über die Classenzahl quadratischer Zahlkörper, *Acta Arith.* **1** (1935), 83–86 [= pages 406–409 (#21) of *Gesammelte Abhandlungen* (Collected Works) Vol.I, Berlin: Springer, 1966.] [O 9.66.2 / QA3.S56]

[Stark 1967] Stark, H.M.: A complete determination of the complex quadratic fields of class-number one, *Michigan Math. J.* **14** (1967), 1–27.

Math 259: Introduction to Analytic Number Theory

Formulas for $L(1,\chi)$

Let χ be a primitive character mod $q > 1$. We shall obtain a finite closed form for $L(1,\chi)$. As with several of our other formulas involving $L(s,\chi)$, this one will have one shape if χ is even ($\chi(-1) = +1$), another if the character is odd ($\chi(-1) = -1$).

Recall our formula

$$\chi(n) = \frac{1}{\tau(\overline{\chi})} \sum_{a \bmod q} \overline{\chi}(a) e^{2\pi i n a/q}.$$

This yields

$$L(1,\chi) = \frac{1}{\tau(\overline{\chi})} \sum_{a \bmod q} \overline{\chi}(a) \sum_{n=1}^{\infty} \frac{1}{n} e^{2\pi i a n/q}, \tag{1}$$

the implied interchange of sums being justified if the inner sum converges for each a mod q coprime with q. But this convergence follows by partial summation from the boundedness of the partial sum $\sum_{n=1}^{M} e^{2\pi i a n/q}$ for all nonzero a mod q. In fact we recognize it as the Taylor series for

$$-\log(1 - e^{2\pi i a/q}) = -\log\left(2\sin\frac{a\pi}{q}\right) + \frac{i\pi}{2}\left(1 - \frac{2a}{q}\right)$$

(if we choose the representative of a mod q with $0 < a < q$). Either the real or the imaginary part will disappear depending on whether χ is odd or even.

Assume first that χ is even. Then the terms $(1 - 2a/q)$ cancel in $(a, q-a)$ pairs. Moreover, the terms $\overline{\chi}(a)\log 2$ sum to zero, and we have

$$L(1,\chi) = -\frac{1}{\tau(\overline{\chi})} \sum_{a \bmod q} \overline{\chi}(a) \log\sin\frac{a\pi}{q}. \tag{2}$$

For example, if χ is a real character then

$$\sqrt{q} L(1,\chi) = 2\log\epsilon$$

where

$$\epsilon = \prod_{a=1}^{\lfloor q/2 \rfloor} \sin^{\chi(a)}\frac{a\pi}{q}$$

is a *cyclotomic unit* of $\mathbf{Q}(\sqrt{q}\,)$. The Dirichlet class number formula then asserts in effect that $\epsilon = \epsilon_0^h$ where ϵ_0 is the fundamental unit of that real quadratic field and h is its class number.

If on the other hand χ is odd then it is the logarithm terms that cancel in symmetrical pairs. Using again that fact that $\sum_{a \bmod q} \overline{\chi}(a) = 0$ we simplify (1) to

$$L(1,\chi) = -\frac{i\pi}{q\tau(\overline{\chi})} \sum_{a=1}^{q-1} a\overline{\chi}(a) \tag{3}$$

In particular if χ is real then (again using the sign of $\tau(\chi)$ for real characters)

$$L(1,\chi) = -\pi q^{-3/2} \sum_{a=1}^{q-1} a\chi(a).$$

Thus $\sum_{a=1}^{q-1} a\chi(a)$ is negative, and by Dirichlet equals $-q$ times the class number of the imaginary quadratic field $\mathbf{Q}(\sqrt{-q}\,)$, except for $q = 3, 4$ when that field has extra roots of unity.

Let us concentrate on the case of real characters to prime modulus $q \equiv -1 \bmod 4$. The inequality $\sum_{a=1}^{q-1} a\chi(a) < 0$ suggests that the quadratic residues mod q tend to be more numerous in the interval $[1, q/2]$ than in $[q/2, q]$. We can prove this by evaluating the sum

$$S_\chi(N) := \sum_{n=1}^{N} \chi(n)$$

at $N = q/2$. We noted already that for any nontrivial character $\chi \bmod q$ we have $S_\chi(q) = 0$ and thus $|S_\chi(N)| < q$ for all N. In fact, using the Gauss-sum formula for $\chi(n)$ we have

$$S_\chi(N) = \frac{1}{\tau(\overline{\chi})} \sum_{a \bmod q} \overline{\chi}(a) \sum_{n=1}^{N} e^{2\pi i n a/q} = \frac{-1}{\tau(\overline{\chi})} \sum_{a \bmod q} \overline{\chi}(a) \frac{1 - e^{2\pi i N a/q}}{1 - e^{-2\pi i a/q}}. \tag{4}$$

We note in passing that this formula quickly yields:

Lemma ([Pólya 1918], [Vinogradov 1918])**.** *There exists an absolute constant A such that*

$$|S_\chi(N)| < A q^{1/2} \log q$$

for all primitive Dirichlet characters $\chi \bmod q$ $(q > 1)$ and all $N \in \mathbf{Z}$.

Proof: We have $(1 - e^{2\pi i N a/q})/(1 - e^{-2\pi i a/q}) \ll \max(q/a, q/(q-a))$. We already saw that $|\tau(\chi)| = q^{1/2}$. Therefore (4) yields

$$S_\chi(N) \ll q^{1/2} \sum_{a=1}^{\lfloor q/2 \rfloor} \frac{1}{a} \ll q^{1/2} \log q. \quad \square$$

Now let χ be the quadratic character modulo a prime $q \equiv -1 \bmod 4$ and let $N = (q-1)/2$. (What would happen for $q \equiv +1 \bmod 4$?) Then (4) becomes

$$S_\chi((q-1)/2) = \sum_{n=1}^{q-1} \chi(n)\phi(n/q)$$

where $\phi(x)$ is the periodic function defined by

$$\phi(x) = \begin{cases} 0, & \text{if } 2x \in \mathbf{Z}; \\ +1/2, & \text{if } 0 < x - \lfloor x \rfloor < 1/2; \\ -1/2, & \text{otherwise} \end{cases}$$

("square wave"). This has the Fourier series

$$\phi(x) = \frac{2}{\pi}\Big(\sin 2\pi x + \frac{1}{3}\sin 6\pi x + \frac{1}{5}\sin 10\pi x + \frac{1}{7}\sin 14\pi x + \cdots\Big).$$

We thus have

$$S_\chi((q-1)/2) = \frac{1}{i\pi}\sum_{\substack{m=1 \\ m \text{ odd}}}^{\infty} \frac{1}{m}\sum_{a=1}^{q-1}\chi(a)(e^{2\pi i m a/q} - e^{-2\pi i m a/q})$$

The inner sum is

$$\tau(\chi)(\chi(m) - \overline{\chi}(-m)) = 2i\sqrt{q}\,\chi(m).$$

Thus our final formula for $S_\chi((q-1)/2)$ is

$$\frac{2\sqrt{q}}{\pi}\sum_{m \text{ odd}}\frac{\chi(m)}{m} = \frac{(2-\chi(2))\sqrt{q}}{\pi}L(1,\chi).$$

It follows, as claimed, that there are more quadratic residues than nonresidues in $[1, q/2]$; in fact, once $q > 3$ the difference between the counts is either h or $3h$ according as $\chi(2) = 1$ or -1, that is, according as q is 7 or 3 mod 8. Even the positivity of $S_\chi((q-1)/2)$ has yet to be proved without resort to such analytic methods!

Exercises

1. Show directly that if χ is a primitive, odd, real character mod $q > 4$ then $\sum_{a=1}^{q-1} a\chi(a)$ is a multiple of q, at least when q is prime.

2. Suppose χ is a primitive character mod q, and n is a positive integer such that $(-1)^n = \chi(-1)$. Prove that $q^{1/2}\pi^{-n}L(n,\chi)$ is a rational number by finding a closed form that generalizes our formula for $n = 1$.

For instance, if $\chi = \chi_4$ we have $\pi^{-n}L(n,\chi) = (-1)^n E_{n-1}/(2^{n+1}(2n-1)!)$, where the integer E_{n-1} is the $(n-1)$-st Euler number.

3. Using the functional equation, conclude that $L(n,\chi) \in \mathbf{Q}$ for all real Dirichlet characters χ (possibly trivial and/or non-primitive) and integers $n \le 0$.

4. What can you say of $S_\chi(\lfloor q/4 \rfloor)$? What about the sums $\sum_{a=1}^{q-1} a^m\chi(a)$ for $m = 2, 3, \ldots$? (See [ACW 1967], [TW 1999].)

References

[ACW 1967] Ayoub, R., Chowla, S., Walum, H.: On sums involving quadratic characters, *J. London Math. Soc.* **42** (1967), 152–154.

[Pólya 1918] Pólya, G.: Über die Verteilung der quadratische Reste und Nichtreste, *Königl. Ges. Wiss. Göttingen Nachr.* (1918), 21–29.

[Vinogradov 1918] Vinogradov, I.M.: On the distribution of power residues and non-residues, *J. Phys. Math. Soc. Perm Univ.* **1** (1918), 94–98 [= *Selected Works* (Berlin, 1985), 271–296.

[TW 1999] Teske, E., Williams, H.C.: A Problem Concerning a Character Sum, *Experimental Math.* **8** (1999) #1, 63–72.

Math 259: Introduction to Analytic Number Theory

The Selberg (quadratic) sieve and some applications

An elementary and indeed naïve approach to the distribution of primes is the following argument: an integer n is prime if and only if it is not divisible by the primes $\leq \sqrt{n}$; but half the integers are odd, 2/3 are not multiples of 3, 4/5 not multiples of 5, etc., and divisibility by any prime is independent of divisibility by finitely many other primes, so... Moreover, if n is restricted to an arithmetic progression $a \bmod q$ with $(a, q) = 1$ then the same factors $(p-1)/p$ arise except those for which $l|q$, from which we recover the factor $\prod_{p|q} p/(p-1) = q/\phi(q)$ in the asymptotic formula for $\pi(qx, a \bmod q)$.

The problem with estimating $\pi(x)$ etc. this way is that the divisibilities aren't quite independent. This is already implicit in our trial-division test for primality: if n is known to contain no primes $\leq \sqrt{n}$, the conditional probability that it be a multiple of some other prime $p \in (\sqrt{n}, n)$ is not $1/p$ but zero. Already for small p, the number of $n < x$ divisible by p is not quite x/p but $x/p + O(1)$, and similarly for n divisible by a product of distinct primes; so if we try to use the principle of inclusion and exclusion to recover the number of primes $n < x$, or even of n not divisible by r primes $p_1, \ldots, p_r$, we get an estimate of $x \prod_{i=1}^{r} (1 - \frac{1}{p_i})$ as expected, but with an "error term" $O(2^r)$ that swamps the estimate long before r can get usefully large.

This quandary is prototypical of "sieve" situations, in which we have a set S of A integers such that $\#(S \cap D\mathbf{Z})/A$ is approximated by a multiplicative function $\alpha(D)$ of the squarefree integer D, and are interested in the number $A(\prod_{p \in P} p)$ of $n \in S$ not divisible by any of the primes p in a given set P. (For instance, if S is an interval then $\alpha(D) = 1/D$; in general, α must be multiplicative for the divisibility of a random $n \in S$ by a prime p to be approximately independent of its divisibility by any other primes.) Several methods are now known for deriving "sieve inequalities", which are nontrivial upper bounds on $A(\prod_{p \in P} p)$. These inequalities use a variety of methods, but curiously the resulting bounds are similar in many important contexts, and often exceed the expected number by a factor asymptotic to 2. We shall develop one of the most general such inequalities, due to Selberg, and give some typical examples of its use in analytic number theory. While we state Selberg's sieve in the context of divisibility, in fact all that we are using is that each prime p sifts out a subset of S and that the events that a random $n \in S$ survives these tests for different p are approximately independent. Thus Selberg's sieve has a counterpart in the context of probability theory, for which see the final Exercise. Selberg's and many other sieves are collected in [Selberg 1969]; nice applications of sieve inequalities to other kinds of problems in number theory are interspersed throughout [Serre 1992].

Assume, then, that a_n ($n \in \mathbf{Z}$) are nonnegative real numbers with $\sum_{n \in \mathbf{Z}} a_n = A < \infty$, and that α is a multiplicative function satisfying $0 \leq \alpha(d) \leq 1$ for

each d (equivalently, for each prime d). For each squarefree $d > 0$ let

$$A_d := \sum_{m \in \mathbf{Z}} a_{md} = A\,\alpha(d) + r(d);$$

in any application, the $r(d)$ must be small compared to A. Let P be a finite set of primes, and D the squarefree integer $\prod_{p \in P} p$. We are interested in

$$A(D) := \sum_{(n,D)=1} a_n,$$

which is the number of $n \in A$ not divisible by any of the primes in P. We hope that $A(D)$ is approximately $A \prod_{p|D}(1 - \alpha(p))$, with an error that is usefully small if the $r(d)$ are. What we can show is:

Theorem (Selberg): *For each $z \geq 1$ we have*

$$A(D) \leq \frac{A}{S(D,z)} + R(D,z), \tag{1}$$

where S, R are defined by

$$S(D,z) := \sum_{\substack{d|D \\ d \leq z}} \prod_{p|d} \frac{\alpha(p)}{1-\alpha(p)}, \qquad R(D,z) := \sum_{\substack{d|D \\ d \leq z^2}} 3^{\omega(d)} |r(d)|$$

and $\omega(d) := \sum_{p|d} 1$, the number of distinct prime factors of d.

Remark: Given D and α, the series $S(D,z)$ and $R(D,z)$ are increasing functions of z because they accumulate more positive terms as z grows. For $z = 1$ we have $S(D,z) = 1$ and $R(D,z)$=0, so (1) is the trivial inequality $A(D) \leq A(1) = A$. For $z \geq D$ we have

$$S(D,z) = S(D,D) = \sum_{d|D} \prod_{p|d} \frac{\alpha(p)}{1-\alpha(p)} = \prod_{p|D} \left(1 + \frac{\alpha(p)}{1-\alpha(p)}\right) = \prod_{p|D} \frac{1}{1-\alpha(p)},$$

so $1/S(D,z)$ is the expected factor $\prod_{p|D}(1-\alpha(p))$, and (1) is implied by the inclusion-exclusion estimate $\big|A(D) - A\prod_{d|D}(1-\alpha(p))\big| \leq \sum_{d|D} |r(d)|$. Thus Selberg's inequality may be regarded as an interpolation between inclusion-exclusion and the trivial $A(D) \leq A$. Note that (1) is only an upper bound: we do *not* claim that $|A(D) - A/S(D,z)| \ll R(D,z)$.

Proof: Let λ_d $(d|D)$ be arbitrary real parameters with $\lambda_1 = 1$ (and eventually $\lambda_d = 0$ once $d > z$). Then

$$A(D) \leq \sum_n a_n \Big(\sum_{d|(n,D)} \lambda_d \Big)^2 = \sum_{d_1,d_2|D} \sum \lambda_{d_1} \lambda_{d_2} \sum_{[d_1,d_2]|n} a_n,$$

where $[d_1,d_2] := \mathrm{lcm}(d_1,d_2)$. The inner sum is just $A_{[d_1,d_2]}$, so we have

$$A(D) \leq \sum_{d_1,d_2|D} \sum \lambda_{d_1}\lambda_{d_2} \big(A\alpha([d_1,d_2]) + r([d_1,d_2]) \big) \leq AQ + R,$$

where Q is the quadratic form

$$Q := \sum_{d_1,d_2|D}\sum \alpha([d_1,d_2])\,\lambda_{d_1}\lambda_{d_2}$$

in the λ_d, and

$$R := \sum_{d_1,d_2|D}\sum \big|\lambda_{d_1}\lambda_{d_2} r([d_1,d_2])\big|.$$

Now for $d|D$ the number of pairs d_1, d_2 such that $d = [d_1, d_2]$ is $3^{\omega(d)}$ (why?); thus (1) will follow from the following

Lemma: *The minimum of the quadratic form Q subject to the conditions $\lambda_1 = 1$ and $d > z \Rightarrow \lambda_d = 0$ is $1/S(D,z)$, and is attained by λ_d with $|\lambda_d| \le 1$.*

Proof of Lemma: By continuity we may assume that $0 < \alpha(p) < 1$ for all $p \in P$. (In fact, for our purpose we can exclude from the start the possibilities $\alpha(p) = 0$ or 1 — do you see why?) Since $[d_1,d_2]\gcd(d_1,d_2) = d_1 d_2$ and α is multiplicative, we have

$$Q = \sum_{d_1,d_2|D}\sum \frac{\alpha(d_1)\lambda_{d_1}\cdot\alpha(d_2)\lambda_{d_2}}{\alpha(\gcd(d_1,d_2))}.$$

Selberg's key insight is that this quadratic form is diagonalized by introducing coefficients $\delta(e)$ for $e|D$, determined by

$$\frac{1}{\alpha(d)} = \sum_{e|d}\delta(e).$$

Then

$$Q = \sum_{e|D}\delta(e)\Big[\sum_{e|d}\alpha(d)\lambda_d\Big]^2.$$

Let $x(e)$, then, be defined by

$$x(e) := \sum_{e|d}\lambda_d\alpha(d).$$

By Möbius inversion we find

$$\delta(e) = \prod_{p|e}\frac{1-\alpha(p)}{\alpha(p)}, \qquad \lambda_d = \frac{1}{\alpha(d)}\sum_{d|e}\mu(e/d)x(e).$$

Our conditions on the λ_d then become

$$\sum_{e|D}\mu(e)x(e) = \alpha(1)\lambda_1 = 1, \qquad e > z \Rightarrow x(e) = 0.$$

By the Schwarz inequality, the minimum of Q subject to these conditions is

$$\Big[\sum_{e|D,\,e\le z}\frac{1}{\delta(e)}\Big]^{-1} = 1/S(D,z),$$

and is attained at $x(e) = \mu(e)/(\delta(e)S(D,z))$. This yields

$$S(D,z)\lambda_d = \frac{\mu(d)}{\alpha(d)} \sum_{d|e\le z} \frac{1}{\delta(e)} = \frac{\mu(d)}{\alpha(d)\delta(d)} \sum_{\substack{f|(D/d)\\ f\le z/d}} \frac{1}{\delta(f)}.$$

But we have

$$\frac{1}{\alpha(d)\delta(d)} = \sum_{e|d} \frac{1}{\delta(e)} :$$

both sides of the equation are multiplicative functions of the squarefree integer d (since α, δ are both multiplicative), and the equation holds for prime d by our above formula for $\delta(e)$. Thus we have

$$S(D,z)\lambda_d = \mu(d) \sum\sum_{e,f} \frac{1}{\delta(ef)},$$

with each $ef \le z$ and no ef values repeated. Thus the sum has absolute value at most $S(D,z)$, so $|\lambda_d| \le 1$ as claimed. This concludes the proof of the Lemma, and thus also of Selberg's inequality (1). □□

Typically we will let $D = D(y) = \prod_{p\le y} p$. For instance, we show:[1]

Corollary. *Fix q. For all a, x_0, A such that $\gcd(a,q) = 1$, we have*

$$\pi(x_0 + Aq, a \bmod q) - \pi(x_0, a \bmod q) < \Big(\frac{2q}{\phi(q)} + O(\frac{\log\log A}{\log A})\Big)\frac{A}{\log A}. \qquad (2)$$

Proof: Let a_n be the characteristic function of the arithmetic progression

$$\{n | n \equiv a \bmod q,\ 0 < n - x_0 < Aq\}.$$

Then $A(D(y))$ is an upper bound on $\pi(x_0+Aq, a \bmod q) - \pi(x_0, a \bmod q) - \pi(y)$. We take $\alpha(n) = 1/n$ if $\gcd(n,q) = 1$ and $\alpha(n) = 0$ otherwise. Then $|r(d)| \le 1$ for each d, and so $R(D,z)$ is bounded by the sum of the n^{-s} coefficients of $\zeta^3(s)$ for $n \le z^2$, so is $\ll (z\log z)^2$. [An equivalent and more elementary way to handle $\sum_{n\le x} 3^{\omega(n)}$ is to note that $3^{\omega(n)}$ is at most the number of representations $n = n_1n_2n_3$ of n as a product of three positive integers.] As to $S(D,z)$, we take $z = y$ and expand $\alpha/(1-\alpha)$ in a geometric series to find

$$S(D,z) > \sum_{\substack{n\le z\\ (n,q)=1}} \frac{1}{n} = \frac{\phi(q)}{q}\log z + O(1). \qquad (3)$$

Thus Selberg's bound (1) is $(q/\phi(q))A/\log z + O(z^2\log^2 z)$. We choose $y = z = A^{1/2}/\log^2 A$, and deduce the upper bound (2), absorbing the correction $\pi(y)$ into the error term since $\pi(y) < y < A^{1/2}$. □

[1] This result in fact predates Selberg, because this choice of a_n is regular enough to be treated with earlier sieve inequalities.

In particular, we may to obtain an elementary upper bound on $\pi(Aq, a \bmod q)$ by taking $x_0 = 0$. The implied O-constant in (2) depends on q, but tractably and effectively so, without invoking zeros of L-functions and the like. The only issue is the dependence on q of the $O(1)$ error in (3). We may write

$$\sum_{\substack{n \le z \\ (n,q)=1}} \frac{1}{n} = \sum_{d|q} \mu(q/d) \sum_{n=1}^{\lfloor z/d \rfloor} \frac{1}{dn} = \sum_{d|q} \frac{\mu(q/d)}{d} (\log z + O(1 + \log d)).$$

Thus the error in (3) is bounded by

$$\sum_{d|q} |\mu(q/d)| \frac{1 + \log d}{d}.$$

For instance, we readily deduce that for all ϵ there exists an effective $q_0(\epsilon)$ such that if $q > q_0(\epsilon)$ then

$$\pi(x_0 + Aq, a \bmod q) - \pi(x_0, a \bmod q) < (2 + \epsilon) \frac{q}{\phi(q)} \frac{A}{\log A}$$

for all A, x_0, a with $A > q$. If the coefficient 2 were any smaller, this upper bound would be enough to banish the Siegel-Landau zero!

Exercises

1. What are the λ_d if $z \ge D$? Explain.

2. Complete the two proofs outlined above that $\sum_{n \le x} 3^{\omega(n)} \ll x(\log x)^2$ (one by comparison with the coefficients of ζ^3, the other by counting solutions of $n_1 n_2 n_3 \le x$). Can you prove that in fact $\sum_{n \le x} 3^{\omega(n)} \sim Cx(\log x)^2$ for some constant $C > 0$, and numerically compute C?

3. Prove that for each integer $n > 0$ the number of primes $p < x$ such that $p + 2n$ is also prime is $O_n(x/\log^2 x)$. In particular, conclude that the sum

$$\frac{1}{3} + \frac{1}{5} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \frac{1}{13} + \frac{1}{17} + \frac{1}{19} + \frac{1}{29} + \frac{1}{31} + \frac{1}{41} + \frac{1}{43} + \cdots$$

of the reciprocals of twin primes converges. (This result was first obtained by Brun [1919] using his less powerful sieve inequality. The sum may be considered "convergent" if it is finite, that is, if the twin-prime conjecture is false.)

4. Prove that for all $\epsilon > 0$ there exists an effective constant $x_0(\epsilon)$ such that, for each $x > x_0(\epsilon)$, there are at most $((8/\pi) + \epsilon)x/\log x$ integers $n < x$ such that $n^2 + 1$ is a prime. Generalize.

It is of course a famous open problem to find a similar *lower* bound on the number of such n, or even to prove that it is unbounded as $x \to \infty$ — that is, to prove that there are infinitely many primes of the form $n^2 + 1$. More generally, one conjectures that for every irreducible polynomial $P \in \mathbf{Z}[X]$ there exist infinitely many integers n such that $P(n)$ is prime, provided that for each prime p there exists at least one $n \in \mathbf{Z}$ such that $P(n) \not\equiv 0 \bmod p$. This, in turn, is the special case $k = 1$ of "Hypothesis H" of Schinzel

and Sierpiński, which asserts that for irreducible polynomials $P_1, \ldots, P_k \in \mathbf{Z}[X]$ there are infinitely many $n \in \mathbf{Z}$ such that each $P_i(n)$ is prime, provided that for each prime p there exists at least one $n \in \mathbf{Z}$ such that $P_i(n) \not\equiv 0 \bmod p$ for each i. Hypothesis H is also a generalization of a conjecture of Dickson on the simultaneous primality of $a_i n + b_i$, which itself generalizes the twin prime conjecture. In each case one expects that in fact the number of $n < x$ such that each $P_i(n)$ is prime is asymptotic to $cx/(\log x)^k$ for some constant c given by an infinite product over p depending on the polynomials P_i (assuming that $P_i \neq \pm P_j$ for distinct i, j; this is the Bateman-Horn conjecture). For more information on these various conjectures, see Chapter 6 of [Ribenboim 1996], particularly pages 372, 391, and 409. The only case of any of these conjectures that has been proved is the case of a single linear polynomial, which is Dirichlet's theorem. Sieve methods, including Selberg's, yield an upper bound with the same asymptotic behavior but a larger c.

As usual, one can formulate analogous problems over polynomial rings such as $\mathbf{F}_q[T]$ in place of $\mathbf{Z}$. For instance, fix $P \in \mathbf{F}_q[T, X]$, and ask whether there exist infinitely many polynomials $n(T)$ such that $P(T, n(T))$ is irreducible. A necessary condition is that P be irreducible as a polynomial of two variables and that for each nonconstant $p \in \mathbf{F}_q[T]$ there exist $n \in \mathbf{F}_q[T]$ such that $P(T, n(T))$ is not a multiple of p. One might be tempted to conjecture that again this necessary condition is also sufficient; but here this conjecture is false! Explicit families of counterexamples are given in [CCG 2003]. In any event, one can use the same sieve inequalities to give an upper bound $O(q^d/d)$ on the number of $n(T)$ of degree at most d for which $P(T, n(T))$ is irreducible.

5. Let p_i ($i \in [m] := \{1, 2, \ldots, m\}$) be probabilities, i.e., real numbers in $[0, 1]$; and let $E_1, \ldots, E_m$ be events approximating independent events with those probabilities, i.e., such that for each $I \subseteq [m]$ the probability that E_i occurs for all $i \in I$ is $\prod_{i \in I} p_i + r(I)$. Obtain upper bounds on the probability that *none* of the E_i occurs, bounds which correspond to and/or generalize Selberg's (1). (See for instance [Chow 1998], where an even further generalization is proposed.)

References

[Brun 1919] Brun, V.: La série $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + \ldots$, où les dénominateurs sont nombres premiers jumeaux est convergente ou finie, *Bull. des Sci. Math.* **43** (1919), 100–104 and 124–128.

[Chow 1998] Chow, T.: The Combinatorics behind Number-Theoretic Sieves, *Adv. in Math.* **138** (1998), 293–305.

[CCG 2003] Conrad, B., Conrad, K., Gross, R.: Hardy-Littlewood Conjecture for function fields. Preprint, 2003.

[Ribenboim 1996] Ribenboim, P.: *The New Book of Prime Number Records*, New York: Springer 1996.

[Selberg 1969] Selberg, A.: Lectures on Sieves, pages 66–247 of his *Collected Papers II* [O 9.89.2 (II)]

[Serre 1992] Serre, J.-P.: *Topics in Galois Theory.* Boston: Jones and Bartlett 1992. [BB 9.92.12 / QA214.S47]

Math 259: Introduction to Analytic Number Theory

Introduction to exponential sums; Weyl equidistribution

The "exponential" in question is the complex exponential, which we normalize with a factor of 2π and abbreviate by $e(\cdot)$:

$$e(x) := e^{2\pi i x}$$

(with $x \in \mathbf{R}$ in most cases). On occasion we also use the notation

$$e_m(x) := e(mx) = e^{2\pi i m x};$$

note that $e_1(x) = e(x)$ and $e_0(x) = 1$ for all x. An "exponential sum" is a sum of the form $\sum_{n=1}^{N} e(x_n)$ for some real numbers x_n, or more generally $\sum_{n=1}^{N} \chi(a_n) e(x_n)$ for some real x_n, integral a_n, and character χ. (We have already seen the examples of Gauss and Jacobi sums.) The general problem is to find a nontrivial estimate on such a sum, which usually means an upper bound significantly smaller than N on its absolute value. Such problems are ubiquitous in number theory, analytic and otherwise, and occasionally arise in other branches of mathematics (we mentioned [CEP 1996] in the Introduction). Sometimes these sums arise directly or nearly so; for instance, the Lindelöf conjecture concerns the size of

$$\zeta(1/2 + it) = \sum_{n=1}^{N} n^{-1/2-it} + \frac{N^{1/2-it}}{it - 1/2} + O(tN^{-1/2}),$$

so it would follow from a proof of

$$\sum_{n=1}^{\lfloor t^2 \rfloor} n^{-1/2-it} \ll |t|^{\epsilon},$$

which in turn would follow by partial summation from good estimates on

$$\sum_{n=1}^{M} n^{-it} = \sum_{n=1}^{M} e\left(\frac{t \log n}{2\pi}\right).$$

Likewise the Lindelöf conjecture for a Dirichlet L-series $L(s, \chi)$ hinges on upper bounds on $\sum_{n=1}^{M} \chi(n) e(t \log n/(2\pi))$. Often the translation of a problem to estimating exponential sums takes more work. We have already seen one example, the Pólya-Vinogradov estimate on $\sum_{n=1}^{N} \chi(n)$ (which is already an "exponential sum" as we have defined the term, with all $x_n = 0$, but whose analysis required the Gauss exponential sums). Our next example is Weyl's criterion for equidistribution mod 1.

A sequence $c_1, c_2, c_3, \ldots$ of real numbers is said to be *equidistributed mod* 1 if the fractional parts $\langle c_n \rangle$ cover each interval in $\mathbf{R}/\mathbf{Z}$ in proportion to its length; that is, if

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N : a \le \langle c_n \rangle \le b\} = b - a \tag{1}$$

for all a, b such that $0 \le a \le b \le 1$. This is connected with exponential sums via a famous result of Weyl [1914]:

Theorem. *For a sequence $\{c_n\}_{n=1}^{\infty}$ in $\mathbf{R}$ (or equivalently in $\mathbf{R}/\mathbf{Z}$), the following are equivalent:*

(i) Condition (1) holds for all a, b such that $0 \le a \le b \le 1$;

(ii) For any continuous function $f : (\mathbf{R}/\mathbf{Z}) \to \mathbf{C}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(c_n) = \int_0^1 f(t)\, dt; \tag{2}$$

(iii) For each $m \in \mathbf{Z}$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e_m(c_n) = \delta_m \left[= \int_0^1 e_m(t)\, dt \right]. \tag{3}$$

Note that (iii) is precisely the problem of nontrivially estimating an exponential sum.

Proof: (i)⇒(ii) Condition (i) means that (ii) holds when f is the characteristic function of an interval (NB such a function is not generally continuous, but it is integrable, which is enough for the sequel); also both sides of (2) are linear in f, so (ii) holds for finite linear combinations of such characteristic functions, a.k.a. step functions. If $|f(t)| < \epsilon$ for all $t \in \mathbf{R}/\mathbf{Z}$ then both sides of (2) are bounded by ϵ for all N. Thus (ii) holds for any function on $\mathbf{R}/\mathbf{Z}$ uniformly approximable by step functions. But this includes all continuous functions.

(ii)⇒(i) Estimate the characteristic function of $[a, b]$ from below and above by continuous functions whose integral differs from $b - a$ by at most ϵ.

(ii)⇒(iii) is clear because (iii) is a special case of (ii).

(iii)⇒(ii) follows from Fejér's theorem: every continuous function on $\mathbf{R}/\mathbf{Z}$ is uniformly approximated by a finite linear combination of the functions e_m. □

[NB the approximation is in general *not* an initial segment of the Fourier series for f. See [Körner 1988], chapters 1–3 (pages 3–13). The existence of uniform approximations is also a special case of the Stone-Weierstrass theorem.]

Interlude on the "little oh" notation $o(\cdot)$. We have gotten this far without explicitly using the "little oh" notation; this is as good a place as any to

introduce it. The notation $f = o(g)$ means that[1] ($g > 0$ and) $(f/g) \to 0$. This begs the question "approaches zero as what?", whose answer should usually be clear from context if it is not stated explicitly. Thus Weyl's theorem states that $\{c_n\}$ is equidistributed mod 1 if and only if $\sum_{n=1}^{N} e_m(c_n) = o(N)$ as $N \to \infty$ for each nonzero $m \in \mathbf{Z}$; that is, if and only if for each $m \neq 0$ we can improve on the trivial bound $|\sum_{n=1}^{N} e_m(c_n)| \leq N$ by a factor that tends to ∞ with N. For instance, we have Weyl's first application of this theorem: *For $r \in \mathbf{R}$ the sequence $\{nr\}$ is equidistributed* mod 1 *if and only if $r \notin \mathbf{Q}$*. Indeed if r is rational then $\langle nr \rangle$ takes only finitely many values; but if r is irrational then for each m we have $e_m(r) \neq 1$ and thus

$$\sum_{n=1}^{N} e_m(nr) = \frac{e_m((N+1)r) - e_m(r)}{e_m(r) - 1} = O_m(1) = o_m(N).$$

(As with $O_m(\cdot)$, the subscript in $o_m(\cdot)$ emphasizes that the convergence to 0 may not be uniform in m.) In general, we cannot reasonably hope that $\sum_{n=1}^{N} e_m(c_n)$ is bounded for each m, but we will be often able to show that the sum is $o(N)$, which suffices to prove equidistribution. For instance, we'll see that if $P \in \mathbf{R}[x]$ is a polynomial at least one of whose nonconstant coefficients is irrational then $\{P(n)\}$ is equidistributed mod 1. (This was Weyl's main application of his theorem in [Weyl 1914]; the example of $\{nr\}$ is the special case of linear polynomials.) We'll also show this for $\{\log_{10}(n!)\}$ and thus obtain the distribution of the first d digits of $n!$ for each d.

Exercises

1. (An easy variation on Weyl's theorem.) Let $A_n \subset \mathbf{R}$ be finite subsets with $\#(A_n) \to \infty$, and say that A_n is *asymptotically equidistributed modulo* 1 if

$$\lim_{n \to \infty} \frac{\#\{t \in A_n : a \leq \langle t \rangle \leq b\}}{\#(A_n)} = b - a$$

for all a, b such that $0 \leq a \leq b \leq 1$. Prove that this is the case if and only if

$$\lim_{n \to \infty} \frac{1}{\#(A_n)} \sum_{t \in A_n} e_m(t) = \delta_m.$$

Show that this condition is satisfied by A_n constructed as follows: let e_n be some positive integers, $q_n = c_n e_n + 1$ be primes such that $q_n / e_n^2 \to \infty$, and a_n arbitrary elements of $(\mathbf{Z}/q_n\mathbf{Z})^*$; and let A_n be the set of $a_n r / q_n$ for representatives r of the c_n residue classes of nonzero e_n-th powers mod q_n.

Presumably such A_n remain asymptotically equidistributed mod 1 if we require only that $q_n \gg e_n^{\theta}$ for some $\theta > 1$, but this is much harder to prove.

[1] Sometimes $g = 0$ is allowed, in which case $f = o(g)$ means that $(f/g) \to 0$, except at points where $g = 0$, at which f must also vanish. Equivalently, for all $\epsilon > 0$ it is true that eventually $|f| \leq \epsilon g$. For instance, we could use this notation to write the definition of the derivative as follows: a function F is differentiable at x if there exists $F'(x)$ such that $F(y) = F(x) + F'(x)(y - x) + o(y - x)$ as $y \to x$.

2. (Recognizing other distributions mod 1.) In Weyl's theorem suppose condition (iii) holds for all nonzero $m \neq \pm 1$, but

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e_{\pm 1}(c_n) = 1/2.$$

What can you conclude about the limits in (i) and (ii)? Generalize.

3. (Weyl in higher dimensions.) What should it mean for a sequence of vectors in $\mathbf{R}^k$ to be equidistributed mod $\mathbf{Z}^k$? Generalize Weyl's theorem to give a necessary and sufficient condition for equidistribution of a sequence in $(\mathbf{R}/\mathbf{Z})^k$. Deduce a condition on the entries of a vector $r \in \mathbf{R}^k$ that is necessary and sufficient for $\{nr\}_{n=1}^{\infty}$ to be equidistributed mod $\mathbf{Z}^k$.

4. (An application of equidistribution mod $\mathbf{Z}^k$.) Prove that $\inf_{t\in\mathbf{R}} |\zeta(\sigma + it)| = \zeta(2\sigma)/\zeta(\sigma)$ for each $\sigma > 1$, and indeed that

$$\liminf_{|t|\to\infty} |\zeta(\sigma + it)| = \zeta(2\sigma)/\zeta(\sigma), \qquad \limsup_{|t|\to\infty} |\zeta(\sigma + it)| = \zeta(\sigma).$$

What can you say about the behavior of $\log \zeta(\sigma + it)$, or more generally of $\log L(\sigma + it, \chi)$, for fixed $\sigma > 1$ and Dirichlet character χ?

5. (Basic properties of $o(\cdot)$.) If $f = o(g)$ then $f = O(g)$. If $f = o(g)$ and $g = O(h)$, or $f \ll g$ and $g = o(h)$, then $f = o(h)$ (assuming that the same implied limit is taken in both premises). If $f_1 = o(g_1)$ and $f_2 = O(g_2)$ then $f_1 f_2 = o(g_1 g_2)$; if moreover $f_2 = o(g_2)$ then $f_1 + f_2 = o(g_1 + g_2) = o(\max(g_1, g_2))$. Given a positive function g, the functions f such that $f = o(g)$ constitute a vector space.

6. (Effective and ineffective $o(\cdot)$.) An estimate $f = o(g)$ is said to be *effective* if for each $\epsilon > 0$ we can compute a specific point past which $|f| < \epsilon g$ (or $|f| \leq \epsilon g$ if $g = 0$ is allowed); otherwise it is *ineffective*. Show that the transformations in the previous exercise preserve effectivity. Give an example of an ineffective $o(\cdot)$.

References

[Körner 1988] Körner, T.W.: *Fourier Analysis.* Cambridge, England: Cambridge University Press, 1988. [HA 9.88.14 / QA403.5.K67]

[Weyl 1914] Weyl, H.: Über ein Problem aus dem Gebiete der diophantischen Approximationen. *Ges. Abh.* I (Springer: Berlin 1968), 487–497. [O 9.68.1]

Math 259: Introduction to Analytic Number Theory

Exponential sums II: the Kuzmin and Montgomery-Vaughan estimates

[Blurb on algebraic vs. analytical bounds on exponential sums goes here]

While proving that an arithmetic progression with irrational step size is equidistributed mod 1, we encountered the estimate

$$\left|\sum_{n=1}^{N} e(cn)\right| \leq \frac{2}{|1-e(c)|} = 1/|\sin \pi c| \ll \|c\|^{-1},$$

where $\|c\|$ is the distance from c to the nearest integer. Kuzmin (1927) obtained a much more general estimate of this kind:

Proposition. *Let c_n $(0 \leq n \leq N)$ be a sequence of real numbers whose sequence of differences $\delta_n := c_n - c_{n-1}$ $(1 \leq n \leq N)$ is monotonic and contained in $[k+\lambda, k+1-\lambda]$ for some $k \in \mathbf{Z}$ and $\lambda > 0$. Then*

$$\left|\sum_{n=0}^{N} e(c_n)\right| \leq \cot \frac{\pi\lambda}{2} \ll \lambda^{-1}.$$

Proof: Let

$$\zeta_n = \frac{1}{1-e(\delta_n)} = \frac{e(c_{n-1})}{e(c_{n-1})-e(c_n)}.$$

Note that the ζ_n are collinear:

$$\zeta_n = (1 + i \cot \pi\delta_n)/2;$$

since the sequence $\{\delta_n\}$ is monotonic, the ζ_n are positioned consecutively on the vertical line $\mathrm{Re}(\zeta) = 1/2$. Now our exponential sum is

$$\sum_{n=0}^{N} e(c_n) = e(c_N) + \sum_{n=1}^{N} (e(c_{n-1}) - e(c_n))\zeta_n$$
$$= (1-\zeta_N)e(c_N) + \zeta_1 e(c_0) + \sum_{n=1}^{N-1} e(c_n)(\zeta_{n+1} - \zeta_n).$$

Thus

$$\left|\sum_{n=0}^{N} e(c_n)\right| \leq |\zeta_1| + \sum_{n=1}^{N-1} |\zeta_{n+1} - \zeta_n| + |1-\zeta_N| = |\zeta_1| + |\zeta_N - \zeta_1| + |\zeta_N|,$$

where in the last step we used the monotonicity of $\mathrm{Im}(\zeta_n)$ and the fact that $\mathrm{Re}(\zeta_n) = 1/2$. The conclusion of the proof,

$$|\zeta_1| + |\zeta_N - \zeta_1| + |\zeta_N| \leq \frac{1}{\sin \pi\lambda} + \frac{1}{\tan \pi\lambda} = \cot \frac{\pi\lambda}{2},$$

is an exercise in trigonometry. $\square$

For instance, it follows that for $t/\pi < N_1 < N_2$ we have

$$\sum_{n=N_1}^{N_2} n^{-it} \ll N_2/t,$$

since we are dealing with $c_n = -(t \log n)/2\pi$ and thus $\delta_n \sim -t/2n\pi$. By partial summation it follows that

$$\sum_{n=N_1}^{N_2} n^{-1/2-it} \ll \frac{1}{t} \int_{N_1}^{N_2} n^{-3/2} \cdot n \, dn \ll t^{-1} N_2^{1/2},$$

and thus

$$\zeta(1/2+it) = \sum_{n=1}^{N} n^{-1/2-it} + O(1),$$

uniformly for all N, t with $|t|/\pi < N \ll t^2$.

With some more work, we can (and soon will) push the upper limit of the sum further down, but not (yet?) all the way to t^ϵ; as n decreases, the phase $e((t \log n)/2\pi)$ varies more erratically, making the sum harder to control. Still, if we sum random complex numbers of norm c_n, the variance of the sum is $\sum_n |c_n|^2$, so we expect that the sum would grow as the square root of that, which for $\zeta(1/2+it)$ would make it $\log^{1/2} |t|$ "on average". We shall prove this as an application of one of a series of general mean-square results of this kind, in which the summands are not independent variables but complex exponentials with different frequencies:

$$f(t) = \sum_{\mu \in A} c_\mu e(\mu t)$$

for some finite set $A \subset \mathbf{R}$ and coefficients $c_\mu \in \mathbf{C}$. For example, to estimate $\int_{T_1}^{T_2} |\zeta(\sigma+it)|^2 \, dt$ for some nonnegative T_1, T_2, we will take $A = \{(2\pi)^{-1} \log n : \pi n < T_2\}$ and $c_\mu = n^{-\sigma}$ for each $\mu = (2\pi)^{-1} \log n \in A$.

To begin with, if we fix A and c_μ then clearly

$$\int_{T_1}^{T_2} |f(t)|^2 \, dt = (T_2 - T_1) \sum_{\mu \in A} |c_\mu|^2 + O(1).$$

How does the "$O(1)$" depend on A, c_μ, T_1, T_2? Consider first the special case that A is contained in an arithmetic progression $\{\mu_0 + n\delta : n \in \mathbf{Z}\}$ with common difference $\delta > 0$. Then $e(-\mu_0 t) f(t)$ is a periodic function of period δ^{-1}, and $\int_{T_1}^{T_2} |f(t)|^2 \, dt = (T_2 - T_1) \sum_{\mu \in A} |c_\mu|^2$ holds exactly if $T_2 - T_1 \in \delta^{-1}\mathbf{Z}$. It follows that for any T_1, T_2 we have

$$\left| \int_{T_1}^{T_2} |f(t)|^2 \, dt - (T_2 - T_1) \sum_{\mu \in A} |c_\mu|^2 \right| < \delta^{-1} \sum_{\mu \in A} |c_\mu|^2. \tag{1}$$

Remarkably the same inequality can be proved under the much weaker hypothesis that $|\mu - \nu| \geq \delta$ for all distinct $\mu, \nu \in A$. We follow [Vaaler 1985], who attributes the argument to Selberg in 1974.

Lemma. *Let χ_I be the characteristic function of the interval $I = [T_1, T_2]$. Suppose $\beta_-, \beta_+ : \mathbf{R} \to \mathbf{R}$ are functions such that:*
i) $\beta_-(z) \leq \chi_I(z) \leq \beta_+(z)$ for all real z;
ii) $\int_{-\infty}^{\infty} \beta_-(z)\, dz$ and $\int_{-\infty}^{\infty} \beta_+(z)\, dz$ converge, say to B_- and B_+ respectively;
iii) The Fourier transforms

$$\hat{\beta}_-(r) = \int_{-\infty}^{\infty} \beta_-(z)\, e(rz)\, dz, \qquad \hat{\beta}_+(r) = \int_{-\infty}^{\infty} \beta_+(z)\, e(rz)\, dz$$

vanish for all real r with $|r| \geq \delta$.
Then for every finite set $A \subset \mathbf{R}$ such that $|\mu - \nu| \geq \delta$ for all distinct $\mu, \nu \in A$, and any $c_\mu \in \mathbf{C}$, we have

$$B_- \sum_{\mu \in A} |c_\mu|^2 \leq \int_{T_1}^{T_2} |f(t)|^2\, dt \leq B_+ \sum_{\mu \in A} |c_\mu|^2$$

where $f(t) = \sum_{\mu \in A} c_\mu e(\mu t)$.

Proof: By (i) we have

$$\int_{-\infty}^{\infty} |f(t)|^2 \beta_-(t)\, dt \leq \int_{T_1}^{T_2} |f(t)|^2\, dt \leq \int_{-\infty}^{\infty} |f(t)|^2 \beta_+(t)\, dt.$$

We expand $|f(t)|^2$ into $\sum\sum_{\mu,\nu \in A} c_\mu \bar{c}_\nu e((\mu - \nu)t)$, and find that the lower and upper bounds are $\sum\sum_{\mu,\nu \in A} c_\mu \bar{c}_\nu \hat{\beta}_\pm(\mu - \nu)$. By (ii), the main terms (with $\mu = \nu$) sum to $B_\pm \sum_{\mu \in A} |c_\mu|^2$; by (iii), the cross terms (with $\mu \neq \nu$) vanish. $\square$

It is not at all obvious that any functions $\beta_\pm$ can be found that satisfy all three conditions of the Lemma. Note that we must have $\beta_\pm(z) = \int_{-\delta}^{\delta} \hat{\beta}_\pm(r)\, e(-rz)\, dr$ by condition (iii) and the inversion formula for Fourier transforms, so in particular the $\beta_\pm(z)$ must extend to entire functions of $z = x + iy$ with $\beta_\pm(x + iy) \ll \exp 2\pi\delta|y|$. We construct suitable $\beta\pm$ as follows.

The Beurling function $B(z)$ is defined by

$$B(z) := \left(\frac{\sin \pi z}{\pi}\right)^2 \left[\frac{2}{z} + \sum_{n=0}^{\infty} \frac{1}{(n-z)^2} - \sum_{m=1}^{\infty} \frac{1}{(m+z)^2}\right]. \tag{2}$$

This is an entire function of z, because the double zeros of $\sin^2 \pi z$ at $z \in \mathbf{Z}$ cancel the double poles of $1/(n-z)^2$ and $1/(m+z)^2$. Beurling [1938] proved:

Proposition. *i) $0 \leq B(z) - \operatorname{sgn}(z) < 2/(\pi z)^2$ for all $z \in \mathbf{R}$, with $B(z) = \operatorname{sgn}(z)$ if and only if z is a nonzero integer.*
ii) $\int_{-\infty}^{\infty} (B(z) - \operatorname{sgn}(z))\, dz = 1$.

iii) For $z = x + iy \in \mathbf{C}$ we have $B(z) - \operatorname{sgn}(x) \ll (1 + |z|)^{-2} \exp 2\pi|y|$; in particular, $B(z) \ll \exp 2\pi|\operatorname{Im}(z)|$ for all $z \in \mathbf{C}$.

Here $\operatorname{sgn}(z)$ is the sign (a.k.a. signum) of the real number z, equal to 1, -1, or 0 according as z is positive, negative, or zero.

Assuming this Proposition, consider the functions $\beta_\pm$ defined by

$$\beta_-(z) = -\frac{1}{2}\Big[B\big(\delta(T_1 - z)\big) + B\big(\delta(z - T_2)\big)\Big],$$

$$\beta_+(z) = +\frac{1}{2}\Big[B\big(\delta(z - T_1)\big) + B\big(\delta(T_2 - z)\big)\Big].$$

By (i), together with the observation that $\chi_I(z) = \big(\operatorname{sgn}(z-T_1)+\operatorname{sgn}(T_2-z)\big)/2$, we have $\beta_-(z) \le \chi_I(z) \le \beta_+(z)$ for all real z. The same observation together with (ii) yields

$$\int_{-\infty}^{\infty} \beta_\pm(z)\,dz = \pm\frac{1}{\delta} + \int_{-\infty}^{\infty} \chi_I(z)\,dz = T_2 - T_1 \pm \frac{1}{\delta}\,.$$

Finally, by (iii) the $\beta_\pm(z)$ are analytic with $\beta_\pm(z) \ll z^{-2}\exp 2\pi\delta|\operatorname{Im}(z)|$. Thus for $|r| \ge \delta$ we can prove $\int_{-\infty}^{\infty} \beta_\pm(z)\,e(rz)\,dz = 0$ by contour integration, moving the path of integration up if $r \ge \delta$ and down if $r \le -\delta$. This together with the preceding Lemma establishes the inequality (1) whenever $|\mu - \nu| \ge \delta$ for all distinct $\mu, \nu \in A$.

It remains to prove the Proposition on Beurling's function.

Proof: We use the well-known partial-fraction decomposition $(\pi/\sin\pi z)^2 = \sum_{n=-\infty}^{\infty} 1/(z-n)^2$, which we shall write as

$$\Big(\frac{\pi}{\sin\pi z}\Big)^2 = \sum_{n=0}^{\infty} \frac{1}{(n-z)^2} + \sum_{m=1}^{\infty} \frac{1}{(m+z)^2}\,. \tag{3}$$

i) We have $B(0) = 1 > \operatorname{sgn}(0)$. For $z > 0$ we use (3) to write

$$B(z) - 1 = 2\left(\frac{\sin\pi z}{\pi}\right)^2 \left[\frac{1}{z} - \sum_{m=1}^{\infty} \frac{1}{(m+z)^2}\right].$$

Since $1/(t+z)^2$ is a decreasing function of t on $[0,\infty)$, we have

$$\int_1^{\infty} \frac{dt}{(t+z)^2} < \sum_{m=1}^{\infty} \frac{1}{(m+z)^2} < \int_0^{\infty} \frac{dt}{(t+z)^2}.$$

Thus $\sum_{m=1}^{\infty} \in (1/(z+1), 1/z)$, so $B(z) \ge 1$ with equality if and only if $z \in \mathbf{Z}$, and $B(z) - 1 \le 2(\sin\pi z/\pi z)^2 \le 2/(\pi z)^2$. Likewise, for $z < 0$ we use

$$B(z) + 1 = 2\left(\frac{\sin\pi z}{\pi}\right)^2 \left[\frac{1}{z} + \sum_{n=0}^{\infty} \frac{1}{(n-z)^2}\right].$$

Since $1/(t-z)^2$ is a decreasing function of t on $[0,\infty)$, we have

$$\sum_{n=0}^{\infty} \frac{1}{(n-z)^2} > \int_0^{\infty} \frac{dt}{(t-z)^2} = \frac{1}{-z},$$

so $B(z) \geq -1$, again with equality if and only if $z \in \mathbf{Z}$. On the other hand,

$$\sum_{n=0}^{\infty} \frac{1}{(n-z)^2} = \frac{1}{z^2} + \sum_{n=1}^{\infty} \frac{1}{(n-z)^2} < \frac{1}{z^2} + \int_0^{\infty} \frac{dt}{(t-z)^2} = \frac{1}{-z} + \frac{1}{z^2},$$

from which $B(z)+1 \leq 2(\sin \pi z/\pi z)^2 \leq 2/(\pi z)^2$. We have proved the claimed inequality whether z is zero, positive, or negative.

ii) By part (i), the integral converges, and thus equals $\frac{1}{2}\int_{-\infty}^{\infty}(B(z)+B(-z))\,dz$. But $(B(z)+B(-z))/2$ is simply $((\sin \pi z)/\pi z)^2$, and it is well-known that

$$\int_{-\infty}^{\infty} \left(\frac{\sin \pi z}{\pi z}\right)^2 dz = 1.$$

(For instance, one may calculate that $((\sin \pi z)/\pi z)^2 = \int_{-1}^{1}(1-|r|)\,e(rz)\,dz$, and then use Fourier inversion.) Hence $\int_{-\infty}^{\infty}(B(z)-\mathrm{sgn}(z))\,dz = 1$ as claimed.

iii) We may assume $|z| > 1$. Again we use our formulas for $B(z)-1$ or $B(z)+1$ according as $x \geq 0$ or $x \leq 0$. In the former case,

$$\begin{aligned} \frac{1}{z} - \sum_{m=1}^{\infty} \frac{1}{(m+z)^2} &= \sum_{m=1}^{\infty} \int_{m-1}^{m} \left(\frac{1}{(t+z)^2} - \frac{1}{(m+z)^2}\right) dt \\ &\ll \sum_{m=1}^{\infty} \frac{1}{|m+z|^3} \ll \frac{1}{z^2}. \end{aligned}$$

Likewise if $x \leq 0$ we have $z^{-1} + \sum_{n=0}^{\infty}(n-z)^{-2} \ll 1/z^2$. Since $\sin^2 \pi z \ll \exp 2\pi|y|$, the claimed inequality follows. $\square$

The inequality (1) is not quite enough for us to prove directly that

$$\int_0^T |\zeta(\frac{1}{2}+it)|^2\,dt \sim T \log T, \tag{4}$$

because so far we must approximate $\zeta(\frac{1}{2}+it)$ by a partial sum of length $N \gg |t|$ to assure an $O(1)$ error, and thus must take $\delta \ll 1/N \ll 1/T$ and get an error term in (1) proportional to the main term $T \log T$. We can nevertheless obtain (4) in two ways, by improving our estimates on exponential sums either individually or in mean square. (See also the Exercises.) For now we continue with the mean-square approach.

Again we set $f(t) = \sum_{\mu \in A} c_\mu e(\mu t)$, and integrate $|f(t)|^2$ termwise. This time we write the result as

$$\int_{T_1}^{T_2} |f(t)|^2\,dt - (T_2 - T_1)\sum_{\mu \in A} |c_\mu|^2 = Q_A(\vec{c}_2) - Q_A(\vec{c}_1),$$

where Q_A is the sesquilinear form on $\mathbf{C}^A$ defined by

$$Q_A(\vec{x}) = \frac{1}{2\pi i} \sum\sum_{\substack{\mu,\nu\in A\\ \mu\neq\nu}} \frac{x_\mu \bar{x}_\nu}{\mu-\nu}$$

and $c_j \in \mathbf{C}^A$ $(j = 1, 2)$ are the vectors with μ coordinate $c_\mu e(\mu T_j)$. The termwise estimate $|Q_A(\vec{x})| \le \pi^{-1} \sum\sum_{\mu>\nu} |x_\mu x_\nu|/(\mu-\nu)$ is already sufficient to prove $T^{-1}\int_0^T |\zeta(1/2+it)|^2\,dt \ll \log^2 T$. But remarkably a tighter estimate holds in this general setting. Let

$$\delta(\mu) = \min_\nu |\nu - \mu|,$$

the minimum taken over all $\nu \in A$ other than μ itself. We shall show:

Theorem (Montgomery-Vaughan Hilbert Inequality). *For any finite set $A \subset \mathbf{R}$ and any $\vec{c} \in \mathbf{C}^A$ we have*

$$|Q_A(\vec{c})| \ll \sum_{\mu\in A} \frac{|c_\mu|^2}{\delta(\mu)},$$

and thus

$$\int_{T_1}^{T_2} \Big|\sum_{\mu\in A} c_\mu e(\mu t)\Big|^2\,dt = \sum_{\mu\in A} \Big[T_2 - T_1 + \frac{\theta}{\delta(\mu)}\Big]\,|c_\mu|^2$$

with $\theta \ll 1$.

Why "Hilbert Inequality" and not simply "Inequality"? Because this is a grand generalization of the original Hilbert inequality, which is the special case $A = \{\mu \in \mathbf{Z} : |\mu| < M\}$. In that case our function $f(t)$ is $\mathbf{Z}$-periodic, and as Schur observed the inequality $|Q_A(\vec{c})| < (1/2)\sum_\mu |c_\mu|^2$ follows from the integral formula $Q_A(\vec{c}) = i\int_0^1 (t-\frac{1}{2})|f(t)|^2\,dt$ (though as we've seen in the periodic case the resulting estimate on $\int_{T_1}^{T_2}|f(t)|^2\,dt$ is even easier than the upper bound on $|Q_A(\vec{c})|$).

The Montgomery-Vaughan inequality does not have as precise an error bound as (1), but it has the advantage that the coefficient of $|c_\mu|^2$ is smaller when the distance from μ to the rest of A greatly exceeds $\delta = \min_{\mu\in A}\delta(\mu)$. For example, the formula (4) for the second moment of $\zeta(\frac{1}{2}+it)$ follows quickly from Montgomery-Vaughan: take $A = \{\log n/2\pi : n = 1, 2, 3, \dots, N\}$ to find that

$$\int_{T_1}^{T_2} \Big|\sum_{n=1}^N c_n n^{it}\Big|^2\,dt = \sum_{n=1}^N (T_2 - T_1 + O(n))|c_n|^2$$

for any T_1, T_2, c_n; then choose $(T_1, T_2) = (-T, -T/2)$ and $c_n = n^{-1/2}$ to find

$$\int_{T/2}^{T} \Big|\sum_{n=1}^N n^{-1/2-it}\Big|^2\,dt = \frac{1}{2}T\log N + O(T+N),$$

and conclude that

$$\int_{T/2}^{T} |\zeta(1/2+it)|^2\,dt = \frac{1}{2}T\log T + O(T\sqrt{\log T}),$$

from which (4) follows.

Proof of the Montgomery-Vaughan Hilbert inequality: Consider $\mathbf{C}^A$ as a finite-dimensional complex Hilbert space with inner product

$$\langle \vec{c}, \vec{c}' \rangle := \sum_{\mu\in A} c_\mu \bar{c}'_\mu/\delta(\mu).$$

Then $Q_A(\vec{x}) = \langle \vec{x}, L\vec{x}\rangle$ where L is the Hermitian operator taking $\vec{x}$ to the vector with μ coordinate $(2\pi i)^{-1}\delta(\mu)\sum_{\nu\neq\mu} x_\nu/(\mu-\nu)$, and we want to show that $\langle \vec{c}, L\vec{c}\rangle \ll \langle \vec{c}, \vec{c}\rangle$ for all $\vec{c}\in\mathbf{C}^A$. But this is equivalent to the condition that L have norm $O(1)$ as an operator on that Hilbert space, and since the operator is Hermitian it is enough to check that $[-Q_A(\vec{c}) =]\langle \vec{c}, L\vec{c}\rangle \ll 1$ holds when $\vec{c}$ is a normalized eigenvector. Thus it is enough to prove that $Q_A(\vec{c}) \ll 1$ for all $A, \vec{c}$ such that

$$\sum_{\mu\in A} |c_\mu|^2/\delta(\mu) = 1$$

and there exists some $\lambda\in\mathbf{R}$ such that

$$\delta(\mu)\sum_{\nu\neq\mu} c_\nu/(\mu-\nu) = i\lambda c_\mu$$

for each $\mu\in A$, in which case $\lambda = 2\pi Q_A(\vec{c})$.

Now for any $\vec{c}$ we have

$$|2\pi Q(\vec{c})|^2 = \Big|\sum_\nu \bar{c}_\nu \sum_{\mu\neq\nu} \frac{c_\mu}{\mu-\nu}\Big|^2 \le \Big(\sum_\nu \frac{|c_\nu|^2}{\delta(\nu)}\Big)\Big(\sum_\nu \delta(\nu)\Big|\sum_{\mu\neq\nu}\frac{c_\mu}{\mu-\nu}\Big|^2\Big).$$

By assumption $\sum_\nu |c_\nu|^2/\delta(\nu) = 1$. For the other factor, we expand

$$\Big|\sum_{\mu\neq\nu}\frac{c_\mu}{\mu-\nu}\Big|^2 = \sum_{\mu\neq\nu}\Big|\frac{c_\mu}{\mu-\nu}\Big|^2 + \sum\sum_{\mu_1\neq\mu_2}\frac{c_{\mu_1}\bar{c}_{\mu_2}}{(\mu_1-\nu)(\mu_2-\nu)}.$$

The single sum contributes

$$\sum_\mu |c_\mu|^2 \sum_{\nu\neq\mu}\frac{\delta(\nu)}{(\mu-\nu)^2}$$

to $\sum_\nu \delta(\nu)\big|\sum_{\mu\neq\nu} c_\mu/(\mu-\nu)\big|^2$; let T_μ be the inner sum $\sum_{\nu\neq\mu}\delta(\nu)/(\mu-\nu)^2$, so the above contribution is $\sum_\mu |c_\mu|^2 T_\mu$. The double sum contributes

$$\sum\sum_{\mu_1\neq\mu_2} c_{\mu_1}\bar{c}_{\mu_2}\sum_{\nu\neq\mu_1,\mu_2}\frac{\delta(\nu)}{(\mu_1-\nu)(\mu_2-\nu)}.$$

The key trick is now to use the partial fraction decomposition

$$\frac{1}{(\mu_1-\nu)(\mu_2-\nu)} = \frac{1}{\mu_2-\mu_1}\left(\frac{1}{\mu_1-\nu} - \frac{1}{\mu_2-\nu}\right)$$

to rewrite this last triple sum as

$$\sum_{\mu_1\neq\mu_2}\sum \frac{c_{\mu_1}\bar{c}_{\mu_2}}{\mu_2-\mu_1}\Big[\sum_{\nu\neq\mu_1,\mu_2}\Big(\frac{\delta(\nu)}{(\mu_1-\nu)} - \frac{\delta(\nu)}{(\mu_2-\nu)}\Big)\Big].$$

The point is that the first part of the inner sum is almost independent of μ_2, while the second half is almost independent of μ_1: the other μ enters only as a single excluded ν. That is, the triple sum is

$$\sum_{\mu_1\neq\mu_2}\sum \frac{c_{\mu_1}\bar{c}_{\mu_2}}{\mu_2-\mu_1}\left[\left(S(\mu_1) - \frac{\delta(\mu_2)}{\mu_1-\mu_2}\right) - \left(S(\mu_2) - \frac{\delta(\mu_1)}{\mu_2-\mu_1}\right)\right]$$

where

$$S(\mu) := \sum_{\nu\neq\mu}\frac{\delta(\nu)}{\mu-\nu}.$$

And now we get to use the eigenvalue hypothesis to show that the $S(\mu_j)$ terms cancel each other. Indeed we have

$$\sum_{\mu_1\neq\mu_2}\sum \frac{c_{\mu_1}\bar{c}_{\mu_2}}{\mu_2-\mu_1}S(\mu_1) = \sum_{\mu_1} c_{\mu_1}S(\mu_1)\sum_{\mu_2\neq\mu_1}\frac{\bar{c}_{\mu_2}}{\mu_2-\mu_1}$$

and the inner sum is just $i\lambda\bar{c}_{\mu_1}/\delta(\mu_1)$, so

$$\sum_{\mu_1\neq\mu_2}\sum \frac{c_{\mu_1}\bar{c}_{\mu_2}}{\mu_2-\mu_1}S(\mu_1) = i\lambda\sum_{\mu} S(\mu)\frac{|c_\mu|^2}{\delta(\mu)}.$$

The same computation shows that

$$\sum_{\mu_1\neq\mu_2}\sum \frac{c_{\mu_1}\bar{c}_{\mu_2}}{\mu_2-\mu_1}S(\mu_2) = i\lambda\sum_{\mu} S(\mu)\frac{|c_\mu|^2}{\delta(\mu)},$$

so the $S(\mu_j)$ terms indeed drop out! Collecting the surviving terms, we are thus left with

$$|2\pi Q(\vec{c})|^2 \leq \sum_{\mu\in A}|c_\mu|^2 T_\mu + \sum_{\mu_1\neq\mu_2}\sum c_{\mu_1}\bar{c}_{\mu_2}\frac{\delta(\mu_1)+\delta(\mu_2)}{(\mu_2-\mu_1)^2}. \tag{5}$$

By now all the coefficients are positive, so we will have no further magic cancellations and will have to just estimate how big things can get. We'll need some lemmas (which are the only place we actually use the definition of $\delta(\mu)$!): first, for each $k = 2, 3, \ldots$,

$$\mu\in A \Rightarrow \sum_{\nu\neq\mu}\frac{\delta(\nu)}{(\mu-\nu)^k} \ll_k \delta(\mu)^{1-k}; \tag{6}$$

second,

$$\mu_1, \mu_2 \in A \Rightarrow \sum_{\nu \neq \mu_1, \mu_2} \frac{\delta(\nu)}{(\mu_1 - \nu)^2 (\mu_2 - \nu)^2} \ll \frac{[\delta(\mu_1)^{-1}] + [\delta(\mu_2)^{-1}]}{(\mu_1 - \mu_2)^2}. \quad (7)$$

Now the first sum in (5) is $O(1)$ because

$$T_\mu = \sum_{\nu \neq \mu} \frac{\delta(\nu)}{(\mu - \nu)^2} \ll \frac{1}{\delta(\mu)}$$

by the case $k = 2$ of (6). The second sum will be bounded by Cauchy-Schwarz. That sum is bounded by twice

$$B := \sum\sum_{\mu_1 \neq \mu_2} |c_{\mu_1} \bar{c}_{\mu_2}| \frac{\delta(\mu_1)}{(\mu_2 - \mu_1)^2} = \sum\sum_{\mu \neq \nu} |c_\mu \bar{c}_\nu| \frac{\delta(\mu)}{(\mu - \nu)^2}.$$

Since $\sum_\mu |c_\mu|^2/\delta(\mu) = 1$, we have

$$|B|^2 \leq \sum_\nu \delta(\nu) \Big(\sum_{\mu \neq \nu} \frac{|c_\mu| \delta(\mu)}{(\mu - \nu)^2} \Big)^2.$$

Expanding and switching $\sum$'s we rewrite this as

$$|B|^2 \leq \sum\sum_{\mu_1, \mu_2} |c_{\mu_1} c_{\mu_2}| \delta(\mu_1) \delta(\mu_2) \Big(\sum_{\nu \neq \mu_1, \mu_2} \frac{\delta(\nu)}{(\mu_1 - \nu)^2 (\mu_2 - \nu)^2} \Big).$$

When $\mu_1 = \mu_2$, the inner sum is $\ll \delta(\mu)^{-3}$ (by (6) with $k = 4$), so the contribution of those terms is $\ll \sum_\mu |c_\mu|^2/\delta(\mu) = 1$. When $\mu_1 \neq \mu_2$ we apply (7), and the resulting estimate on the sum of the cross-terms is twice the double sum defining B! So, we've shown (modulo the proofs of (6, 7)) that $B^2 \ll 1 + B$. Thus $B \ll 1$ and we're finally done.

Exercises

On the Kuzmin inequality:

1. Prove (4) in yet another way as follows. Write

$$\int_0^T \Big| \sum_{n=1}^N n^{-1/2 - it} \Big|^2 dt - T \sum_{n=1}^N \frac{1}{N} + 2 \sum\sum_{0 < n < n' \leq N} (nn')^{-1/2} \operatorname{Im} \frac{(n'/n)^{iT}}{\log(n'/n)}.$$

For each $j = 1, 2, 3, \ldots$ use Kuzmin to obtain nontrivial bounds on

$$\sum_{n' = n + j} (nn')^{-1/2} (n'/n)^{iT} / \log(n'/n)^{iT}.$$

(This is closely related to the van der Corput bounds that will be our next topic.)

On the Beurling function:

2. Show that $B(z) = B(z-1) - 2\pi^{-2}\sin^2 \pi z/(z^3 - z^2)$. Explain how this can be used to efficiently compute $B(z)$ to high accuracy. (There are at least two approaches, one of which works also for large and/or complex z.)

3. Prove that the constant δ^{-1} in (1) is best possible. Use this to show that the Beurling function minimizes $\int_{-\infty}^{\infty}(f(z) - \mathrm{sgn}(z))\,dz$ over all entire functions f satisfying $f(z) \geq \mathrm{sgn}(z)$ for $z \in \mathbf{R}$ and $f(z) \ll \exp 2\pi|\,\mathrm{Im}(z)|$ for all $z \in \mathbf{C}$.

Beurling showed that in fact $B(z)$ is the unique minimizing function. The same argument shows that if $T_2 - T_1$ is a positive integer then our $\beta_\pm$ are optimal; but they are not unique: see [GV 1981, p.289] for Selberg's description of all the optimal $\beta_\pm$. When $T_2 - T_1 \notin \mathbf{Z}$, the best $\beta_\pm$ are slightly better than those constructed from Beurling's function; Logan [1977] found the optimal $\beta_\pm$ in this case and proved their uniqueness.

4. (A further application of β_+ to mean-square bounds on exponential sums; look up "Large Sieve" in [Selberg 1969] for the context)
i) Suppose $T_2, T_1 \in \mathbf{Z}\delta$. Prove that there exists an entire function f such that $\beta_+(z) = |f(z)|^2$ for all $z \in \mathbf{R}$ and $f(z) \ll |z|^{-1}\exp \pi\delta|\,\mathrm{Im}(z)|$ for all $z \in \mathbf{C}$, and thus that $f|_{\mathbf{R}}$ is an L_2 function with $\hat{f}$ supported on $|r| \leq \delta/2$. [Hint: a polynomial $P \in \mathbf{R}[x]$ is nonnegative for all real x if and only if $P = |Q|^2$ for some $Q \in \mathbf{C}[x]$. According to [Vaaler 1985], the result holds even without the hypothesis that $T_2 - T_1 \in \mathbf{Z}\delta$, using a theorem of Fejér.]
ii) Now let $S(x)$ be a trigonometric polynomial of the form $\sum_{n=T_1}^{T_2} c_n e(nx)$ for some complex numbers c_n ($T_1 \leq n \leq T_2$), and set $S^*(x) = \sum_{T_1}^{T_2} f(n)^{-1}c_n e(nx)$. Then S is the convolution of S^* with $\hat{f}$, so

$$\begin{aligned} |S(x)|^2 &\leq \left(\int_{-\delta/2}^{\delta/2} |\hat{f}(u)|^2\,du\right)\left(\int_{-\delta/2}^{\delta/2} |S^*(x+u)|^2\,du\right) \\ &= (T_2 - T_1 + \delta^{-1})\int_{-\delta/2}^{\delta/2} |S^*(x+u)|^2\,du. \end{aligned}$$

Conclude that if ($\delta \leq 1$ and) $A \subset \mathbf{R}/\mathbf{Z}$ is any finite set such that $\|x - x'\| \geq \delta$ for all distinct $x, x' \in A$ then

$$\sum_{x\in A} |S(x)|^2 < (T_2 - T_1 + \delta^{-1})\sum_{n=T_1}^{T_2} |c_n|^2.$$

Explain why this is elementary when $A \subseteq x_0 + (\mathbf{Z}R^{-1}/\mathbf{Z})$ for some integer $R \geq 1$.

On the generalized Hilbert inequality:

5. Prove that the constant π in the original Hilbert inequality is best possible, and show that it holds even if c_μ is allowed to be nonzero for every integer μ (this is in fact what Hilbert originally proved).

6. More generally, prove that the norm of Q_A relative to the standard inner product $(\vec{c}, \vec{c}') = \sum_{\mu \in A} c_\mu \bar{c}'_\mu$ is less than $(2\delta)^{-1}$.

7. Deduce the second-moment estimate (1), albeit with a slightly worse error bound than we proved using Montgomery-Vaughan, from the generalized Hilbert inequality (1), as follows: write $\sum_{n=1}^N n^{-1/2-it} = f_1(t) + f_2(t)$, where $f_1(t) = \sum_{n=1}^A n^{-1/2-it}$ and $f_2(t) = \sum_{n=A+1}^N n^{-1/2-it}$; use (1) to estimate $\int_{T/2}^T |f_1(t)|^2\,dt$ and $\int_{T/2}^T |f_2(t)|^2\,dt$; and then use $\|f_1 + f_2\|_2 = \|f_1\|_2 + O(\|f_2\|_2)$ (triangle inequality in $L_2(T/2, T)$).

On the Montgomery-Vaughan inequality:

8. Complete the proof by verifying the inequalities (6,7).

9. Let χ be a character (primitive or not) mod q. Obtain an asymptotic formula for $\int_0^T |L(1/2 + it, \chi)|^2\,dt$. How does the error term depend on q? (It is conjectured that $L(1/2 + it, \chi) \ll_\epsilon (q|t|)^\epsilon$; naturally this problem is still wide open: it has the Lindelöf conjecture as a special case.)

References

[Beurling 1938] Beurling, A.: Sur les intégrales de Fourier absolument convergentes et leur application à fonctionelle, *Neuvième Congrès des Mathématiciens Scandinaves*, Helsinfgors, 1938.

[GV 1981] Graham, S.W., Vaaler, J.D.: A class of extremal functions for the Fourier transform, *Trans. Amer. Math. Soc.* **265** (1981), 283–302.

[HLP] Hardy, H.G., Littlewood, J.E., Pólya, G.: *Inequalities*. Cambridge Univ. Press, 1967 (2nd ed.).

[Logan 1977] Logan, B.F.: Bandlimited functions bounded below over an interval, *Notices Amer. Math. Soc.* **24** (1977), A-331.

[MV 1974] Montgomery, H.L., Vaughan, R.C.: Hilbert's Inequality, *J. London Math. Soc.* (2) **8** (1974), 73–81.

[Vaaler 1985] Vaaler, J.D.: Some extremal functions in Fourier analysis, *Bull. Amer. Math. Soc.* (N.S.) **12** (1985) #2, 183–216.

Math 259: Introduction to Analytic Number Theory

Exponential sums III: the van der Corput inequalities

Let $f(x)$ be a sufficiently differentiable function, and $S = \sum_{n=1}^{N} e(f(n))$. The Kuzmin inequality tells us in effect that

I If $f'(x)$ is monotonic and $\lambda_1 < \{f'(x)\} < 1-\lambda_1$ for $x \in [1, N]$ then $S \ll 1/\lambda_1$.

We shall use this to deduce van der Corput's estimates on S in terms of N and higher derivatives of f. In each case the inequality is useful only if f has a derivative $f^{(k)}$ of constant sign which is significantly smaller than 1.

II If there are constants c, C with $0 < c < C$ such hat $c\lambda_2 < f'' < C\lambda_2$ for all $x \in [1, N]$ then

$$S \ll_{c,C} N\lambda_2^{1/2} + \lambda_2^{-1/2}.$$

III If there are constants c, C with $0 < c < C$ such hat $c\lambda_3 < f''' < C\lambda_3$ for all $x \in [1, N]$ then

$$S \ll_{c,C} N\lambda_3^{1/6} + N^{1/2}\lambda_3^{-1/6}.$$

In general there is a k-th inequality

$$S \ll_{c,C} N\lambda_k^{1/(2^k-2)} + N^{1-2^{2-k}}\lambda_k^{-1/(2^k-2)}$$

when $c\lambda_k < f^{(k)} < C\lambda_k$ for all $x \in [1, N]$, but we'll make use only of van der Corput **II** and **III**.

Here is a typical application, due to van der Corput: $\zeta(1/2+it) \ll |t|^{1/6}\log|t|$. We have seen that

$$\zeta(1/2+it) = \sum_{n=1}^{\lfloor |t|/\pi \rfloor} n^{-1/2-it} + O(1).$$

We break up the sum into segments $\sum_{n=N}^{N_1}$ with $N < N_1 \le 2N$, and use $f(x) = (t\log x)/2\pi$, so $\lambda_k = t/N^k$. Then **II** and **III** give

$$\sum_{n=N}^{N'} n^{it} \ll |t|^{1/2} + N/|t|^{1/2}, \qquad \sum_{n=N}^{N'} n^{it} \ll N^{1/2}|t|^{1/6} + N/|t|^{1/6}$$

for $N < N' < N_1$. By partial summation of $\overline{S}$, it follows that

$$\sum_{n=N}^{N'} n^{-1/2-it} \ll (|t|/N)^{1/2} + (N/|t|)^{1/2}, \qquad \sum_{n=N}^{N'} n^{-1/2-it} \ll |t|^{1/6} + N^{1/2}/|t|^{1/6}$$

Choosing the first estimate for $N \gg |t|^{2/3}$ and the second for $N \ll |t|^{2/3}$ we find that the sum is $\ll |t|^{1/6}$ in either case. Since the total number of $[N, N']$ segments is $O(\log|t|)$, the inequality $\zeta(1/2+it) \ll |t|^{1/6}\log|t|$ follows.

The inequality **II** is an easy consequence of Kuzmin's **I**. [NB the following is not van der Corput's original proof, for which see for instance Lecture 3 of [Montgomery 1994]. The proof we give is much more elementary, but does not as readily yield the small further reductions of the exponents that are available with the original method.] We may assume that $f''(x) < 1/4$ on $[1, N]$, else $\lambda_2 \gg 1$ and the inequality is trivial. Split $[1, N]$ into $O(N\lambda_2 + 1)$ intervals on which $\lfloor f' \rfloor$ is constant. Let λ_1 be a small positive number to be determined later, and take out $O(N\lambda_2 + 1)$ subintervals of length $O(\lambda_1/\lambda_2 + 1)$ on which f' is within λ_1 of an integer. On each excised interval, estimate the sum trivially by its length; on the remaining intervals, use Kuzmin. This yields

$$S \ll (N\lambda_2 + 1)(\lambda_1^{-1} + \lambda_1/\lambda_2 + 1).$$

Now take $\lambda_1 = \lambda_2^{1/2}$ to get

$$S \ll (N\lambda_2 + 1)(\lambda_2^{-1/2} + 1).$$

But by assumption $\lambda_2 \ll 1$, so the second factor is $\ll \lambda_2^{-1/2}$. This completes the proof of **II**.

For **III** and higher van der Corput bounds, we shall follow Weyl by showing that

$$S \ll \left\{ \frac{N}{H} \sum_{h=0}^{H} \left| \sum_{n=1}^{N-h} e(f(n+h) - f(n)) \right| \right\}^{1/2}. \tag{1}$$

for $H \leq N$. If $f(x)$ has small positive k-th derivative then each $f(x+h) - f(x)$ has small $(k-1)$-st derivative, which is positive except for $h = 0$ when the inner sum is N. This will let us prove **III** from **II**, and so on by induction (see the first Exercise below).

To prove (1), define z_n for $n \in \mathbf{Z}$ by $z_n = e(f(n))$ for $1 \leq n \leq N$ and $z_n = 0$ otherwise. Then

$$S = \sum_{n=-\infty}^{\infty} z_n = \frac{1}{H} \sum_{n=-\infty}^{\infty} \left(\sum_{h=1}^{H} z_{n+h} \right),$$

in which fewer than $N + H$ of the inner sums are nonzero. Thus by the (Cauchy-)Schwarz inequality,

$$|S|^2 \leq \frac{N+H}{H^2} \sum_{n=-\infty}^{\infty} \left| \sum_{n=1}^{H} z_{n+h} \right|^2 \ll \frac{N}{H^2} \sum_{h_1,h_2=1}^{H} \left| \sum_{n \in \mathbf{Z}} z_{n+h_1} \overline{z_{n+h_2}} \right|.$$

But the inner sum depends only on $|h_1 - h_2|$, and each possible $h := h_1 - h_2$ occurs at most H times. So,

$$|S|^2 \ll \frac{N}{H} \sum_{h=0}^{H} \left| \sum_{n \in \mathbf{Z}} z_{n+h} \overline{z_n} \right|,$$

from which (1) follows.

Now to prove **III**: we may assume $N^{-3} < \lambda_3 < 1$, else the inequality is trivial. Apply (1), and to each of the inner sums with $h \neq 0$ apply **II** with $\lambda_2 = h\lambda_3$. This yields

$$\begin{aligned} |S|^2 &\ll \frac{N^2}{H} + \frac{N}{H}\sum_{h=1}^{H}[N(h\lambda_3)^{1/2} + (h\lambda_3)^{-1/2}] \\ &= N^2\big((H\lambda_3)^{1/2} + H^{-1}\big) + N/(H\lambda_3)^{1/2}. \end{aligned}$$

Now make the first two terms equal by taking $H = \lfloor \lambda_3^{-1/3} \rfloor$:

$$|S|^2 \ll N^2\lambda_3^{1/3} + N\lambda_3^{-1/3}.$$

Extracting square roots yields **III**.

Exercises

1. Prove the van der Corput estimates **IV**, **V**, etc. by induction.

2. Prove that $\{\log_b n!\}_{n=0}^{\infty}$ is equidistributed mod 1 for any $b > 1$.

3. Use (1) to prove the equidistribution of $\{nP(n)\}$ mod 1 for any polynomial $P(x)$ with an irrational coefficient (which was Weyl's original application of (1)). Give necessary and sufficient conditions on polynomials $P_1, P_2, \ldots, P_k \in \mathbf{R}[x]$ for the sequence of vectors $(P_1(n), P_2(n), \ldots, P_k(n))$ to be equidistributed mod $\mathbf{Z}^k$.

Reference

[Montgomery 1994] Montgomery, H.L.: *Ten lectures on the interface between analytic number theory and harmonic analysis.* Providence: AMS, 1994 [AB 9.94.9].

X-RAY OF RIEMANN'S ZETA-FUNCTION

J. ARIAS-DE-REYNA

1. Introduction

This paper is the result of the effort to give the students of the subject *Analytic Number Theory* an idea of the complexity of the behaviour of the Riemann zeta-function. I tried to make them **see** with their own eyes the mystery contained in its apparently simple definition.

There are precedents for the figures we are about to present. In the tables of Jahnke-Emde [9] we can find pictures of the zeta-function and some other graphs in which we can see some of the lines we draw. In the dissertation of A. Utzinger [21], directed by Speiser, the lines $\operatorname{Re}\zeta(s) = 0$ and $\operatorname{Im}\zeta(s) = 0$ are drawn on the rectangle $(-9, 10) \times (0, 29)$.

Besides, Speiser's paper contains some very interesting ideas. He proves that the Riemann Hypothesis is equivalent to the fact that the non trivial zeros of $\zeta'(s)$ are on the right of the critical line. He proves this claim using an entirely geometric reasoning which is on the borderline between the proved and the admissible. Afterwards rigorous proofs of this statement have been given.

Our figures arise from a simple idea. If $f(z) = u(z) + iv(z)$ is a meromorphic function, then the curves $u = 0$ and $v = 0$ meet precisely at the zeros and poles of the function. That is the reason why we mark the curves where the function is real or the curves where it is imaginary on the z-plane. In order to distinguish one from the other, we will draw with thick lines the curves where the function is real and with thin lines the curves where the function is imaginary.

When I distributed the first figure (the X ray of the zeta function) to my students, I was surprised at the amount of things one could see in the graphic. I spent a whole hour commenting on this figure.

Afterwards, I have kept thinking about these graphics. I believe they can be used to systematize the knowledge which today is scattered. The graphics give it a coherency which makes it easier to remember.

Date: Submitted to a journal on September 13, 2002.

1991 *Mathematics Subject Classification.* Primary: 11M06, 11M26.

Key words and phrases. Zeta-function, Riemann-Siegel formula, Gram Points, Gram's Law, Rosser's rule .

Research supported by the Spanish Governement under grant BFM2000-0514.

2. The X ray

2.1. **Remarkable Points.** Recall that the Riemann zeta-function is defined for $\operatorname{Re}(s) > 1$ by the series

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

but it is possible to extend it to the whole plane as a meromorphic function with a single pole in $s = 1$, which is simple.

We already have the X ray on sight. The main thing to take into account is that the thick lines are formed by those points s in which $\zeta(s)$ is real, and the thin lines by those in which $\zeta(s)$ is imaginary.

In the figure we see that, in fact, the lines have a simpler behaviour on the right of the line $s = 1$, that is, on the right of the grey strip which represents the so called **critical strip**: $0 \le \operatorname{Re}(s) \le 1$.

In the figure **the real axis**, **the oval** and a thick line that surrounds the oval stand specially out.

We can see that the real axis cuts lots of thin lines, first the oval in the pole $s = 1$, and in $s = -2$, which is a zero of the function, later, a line in $s = -4$, another in $s = -6$, ..., which are the so called **trivial zeros** of the zeta-function. In the figure we can see these zeros up to $s = -28$, because the graph represents the rectangle $(-30, 10) \times (-10, 40)$. These zeros situated in the negative part of the real axis seem to draw the spine of this X ray.

In the critical strip we see that thick and thin lines meet, that is, we detect the existence of non trivial zeros. With two decimal digits they are $0.5+i14.13$, $0.5+i21.02$, $0.5+i25.01$, $0.5+i30.42$, $0.5+i32.93$, $0.5+i37.58$. They seem to have a real part equal to $\frac{1}{2}$. In fact, it is possible to prove it.

The next remarkable points are the points in which the real axis meets a thick line. These are points at which the derivative $\zeta'(s) = 0$. Since the function $\zeta(s)$ is real when s is real and since it has zeros in the points $-2n$, Rolle's theorem from elemental calculus tells us that these zeros have to exist, one between each pair of consecutive even numbers.

For every meromorphic function $f(s) = u(s) + iv(s)$, the lines $u = 0$ (and the lines $v = 0$) are smooth curves, except in the points where the derivative vanishes, in which case two or more curves meet.

An important property which is valid in the general case of a meromorphic function is the **monotony** of the function along the curves.

Let us think of a thick line $v = 0$. In this curve, the function f is real. At each point of this curve, the derivative of v in the direction of the curve vanishes. If the derivative of f does not vanish in any point of the curve, neither will the derivative of $u = f$.

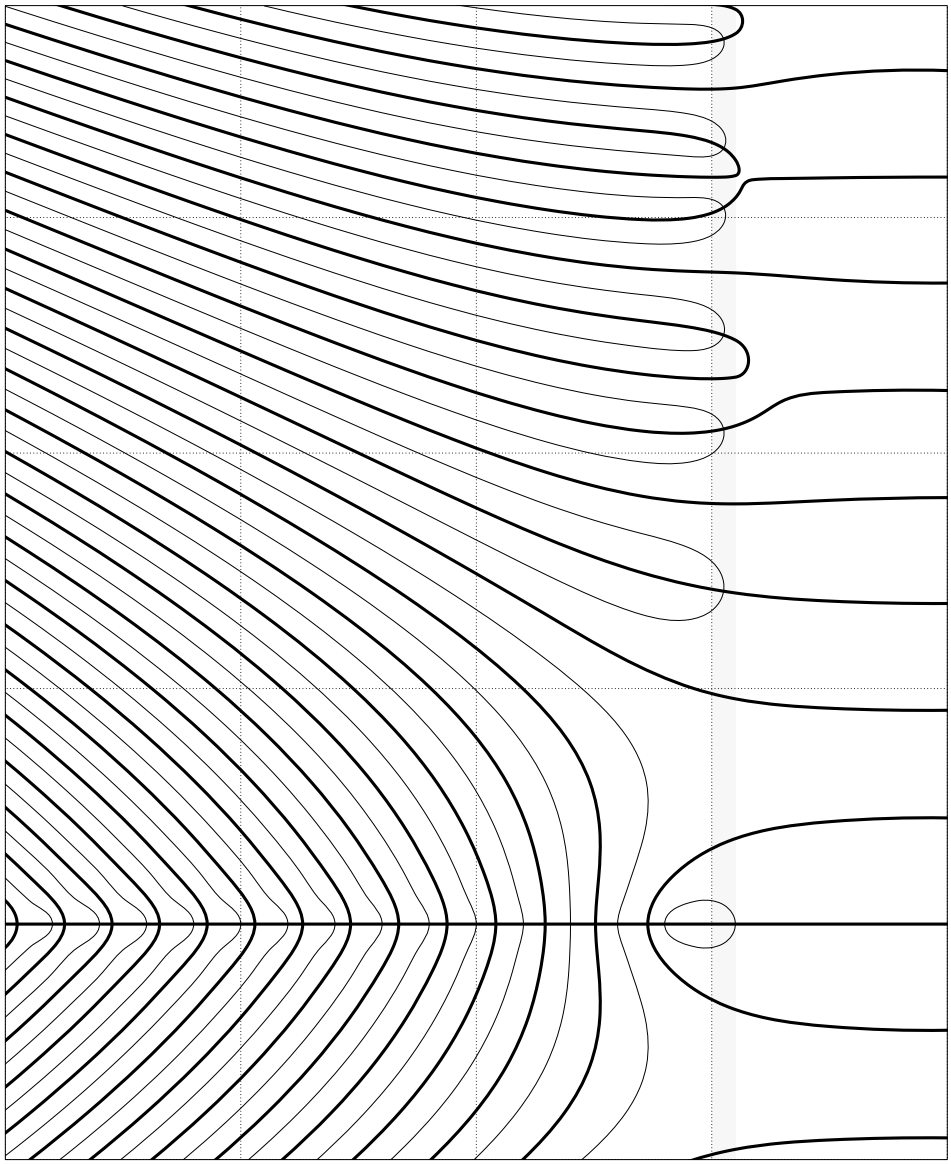


FIGURE 1. X ray of $\zeta(s)$

So, if on a portion of a curve the derivative of f does not vanish, then the function varies monotonously as we moves along the curve. In fact, the derivative of u in the direction of the curve can not vanish on that portion of the curve, so that it will keep having a constant sign.

When we are dealing with the $u = 0$ curves, the monotonous function is $v = f/i$.

We would like to know, for each line and each direction, whether the function increases or decreases. But before tackling this point, we must give each line a name.

2.2. **The numbering of the lines.** There are three special lines which we will not number: the oval, the one that surrounds the oval, and the real axis.

The rest of the lines can be numbered with integer numbers. In the first place, there are thin lines which pass through the zeros in -4, -6, -8, ... Line $-2n$ is the line which cuts the real axis at the point $-2n$.

Between these lines we can find thick lines. For example, there is one line which goes between the lines numbered -4 and -6. This one does not go through the point -5, but it cuts the real axis on a point which is between -4 and -6. We will say that this one is line -5. In the same way lines $-(2n+1)$, $n = 2, 3, \ldots$, are defined. These are thick lines which cut the real axis between $-2n$ and $-2n-2$.

In this way, we now know which are the lines numbered n for $n = -4$, -5, -6, ... If we consider now how these lines cut the left border of the figure (line $x = -30$), we see that the lines already numbered are followed by other lines, so that it seems natural to call line -3 the thick line which runs parallel to line -4, above it, until the latter turns to cut the real axis while the thick line goes on to the right, crossing the critical strip at a height near 10.

So, the even lines are always lines in which the function $\zeta(s)$ is purely imaginary, while the odd lines are thick lines in which $\zeta(s)$ is real. We must also add that, in many cases, two lines join in a zero. For example, lines -2 and 0 join in the first non trivial zero of the zeta-function. Lines 5 and 7 also join in the third zero.

This numbering does not include the symmetrical lines below the real axis.

Now we can speak of concrete lines. Do you see how line 11 turns? It seems it intended to continue with line 13 or 15, but in the end it follows another path, cutting lines 10 and 12.

2.3. **Lines on the right of the critical strip.** The behaviour on the right of the critical strip is governed by the fact that there are no thin lines. This is easy to understand. For $\operatorname{Re}(s) = \sigma \to +\infty$ we have $\zeta(s) \to 1$ uniformly. So, there exists an abscissa σ_0 so that for $\sigma > \sigma_0$ it holds $\operatorname{Re}\zeta(s) > 0$. It follows that the function does not take purely imaginary values in this half-plane. It is not difficult to prove that there are no thin lines passing through the half-plane $\sigma > 1.63622\ldots$.

On the right, the only thick lines that exist are essentially parallel to the x axis, and they are equally spaced. In order to understand the reason,

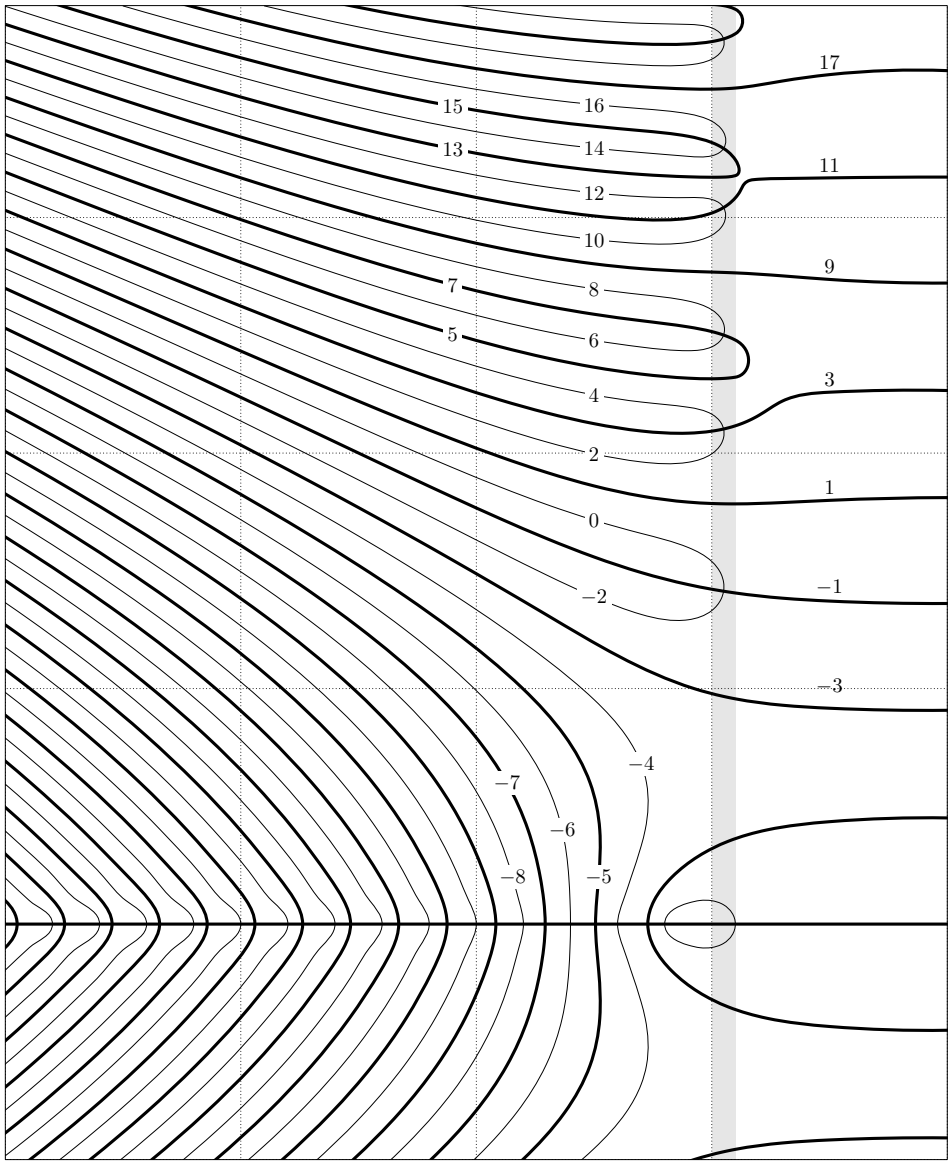


FIGURE 2. Numbering the lines

we start with

$$\operatorname{Im}\zeta(s) = -\sum_{n=2}^{\infty} \frac{\sin(t\log n)}{n^{\sigma}}.$$

For big enough values of σ, the first term of this sum dominates the rest of them. It is clear that this first term vanishes for $t = n\pi/\log 2$. So,

some parallel lines, separated by a distance which is approximately equal to $\pi/\log 2 \approx 4.53236014\ldots$, exist.

We can see in the figure that these parallel lines, when crossing the critical strip, alternatively, contain a non trivial zero, or not. It is easy to explain. The derivative of $\zeta(s)$ along the line which goes at a height $n\pi/\log 2$, when $\sigma \gg 0$, is given by

$$\frac{\partial}{\partial\sigma}\zeta(s) = \frac{\partial}{\partial\sigma}\Bigl(1 + \frac{\cos(t\log 2)}{2^\sigma} + \cdots\Bigr) \approx -\frac{\cos(t\log 2)}{2^\sigma}\log 2 \approx -\frac{(-1)^n}{2^\sigma}\log 2.$$

Thus, when the function $\zeta(s)$ runs along this curve from right to left, (parting from 1), it is increasing for even values of n. So, for even n, $\zeta(s)$ will take on this line the values in $(1, +\infty)$. On the other hand, for odd n it will take the values in $(1, -\infty)$, and, in particular, it will vanish.

Of course we are using the fact that the function is monotonous on the lines, which depends on the fact that the derivative $\zeta'(s)$ does not vanish on these lines. This is what happens to the line surrounding the oval, which ought to contain a zero but does not. So we shall call one of these lines *parallel* only when $\zeta'(s) \neq 0$ on it.

2.4. **Orientating ourselves.** It holds $\lim_{\sigma\to+\infty}\zeta(\sigma + it) = 1.$ uniformly in t. Let us suppose that we go along line -1, starting form the zero on the critical line, and going to the right. The function will take values in the interval $(0, 1)$, starting from zero and tending to 1. The points we leave on our left, that is, the ones which are a little above line -1, will turn into points on the left of the segment $(0, 1)$, that means, points from the first quadrant.

If, on the contrary, we start from the zero and move to the left following line -1 between lines -2 and 0, the zeta-function will take negative values, in the interval $(-\infty, 0)$. The points between lines -1 and 0 will turn into points from the second quadrant.

If we situate ourselves at the zero with our arms outstretched, the right arm to the right and the left one to the left, we see in front of us a thin line (line 0), on which the function will take values ix with $x > 0$. Behind us we have a line (line -2) in which it will take values ix with $x < 0$. On the northeast we have points s which turn into values $\zeta(s)$ situated on the first quadrant. On the northwest we have a region which turns into the second quadrant, etc.

In a line such as line -3, which comes from the right and does not contain any zeros, the zeta-function takes real values greater than 1, precisely all the points in the interval $(1, +\infty)$, because it is known that, on the left, the modulus of the zeta-function tends to infinity.

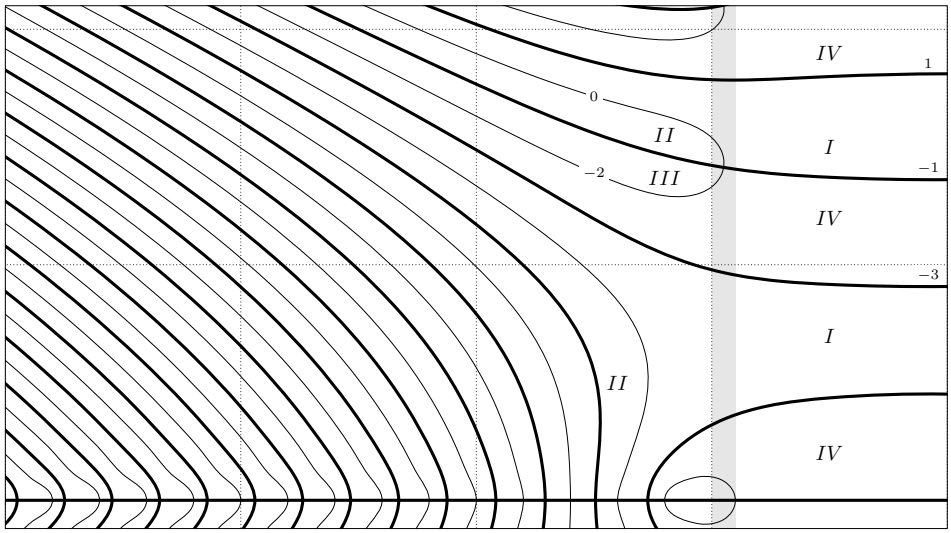


FIGURE 3. Each region transform into a quadrant.

2.5. **The regions of the plane.** The lines in the graphics divide the plane into regions. It is easy to realize that the points of one of these regions have to turn by the function $\zeta(s)$ into points of the same quadrant.

If the points of a region Ω turn into the first quadrant, for example, and crossing a thin line we reach another region, then the points of this new region will turn into points of the second quadrant; if, on the contrary, we leave Ω crossing a thick line, then the new region will turn into the fourth quadrant.

We have various ways to know into which quadrant will a given region turn. For example, if we situate ourselves at a non trivial zero and we orientate in the way described above, then the region in the northeast will turn into the first quadrant, the one situated in the northwest in the second, and so on.

Another way to see it could be to take into account that, if we walk along a thick line, so that the function $\zeta(s)$, which is real, increases, then we know that on our left we have a region which turns into the first quadrant and on our right one which turns into the fourth quadrant. For example, let us consider the thick line number 1 (this is the line which comes from the right and goes between the first and the second non trivial zeros of the zeta-function). and let us walk through it to the left. That is, we walk through this line coming from $+\infty$. The values of the zeta-function on the points of this line are greater than 1, and, as we walk through it, the values on its points are growing. Then, the region on our right will turn into the first quadrant.

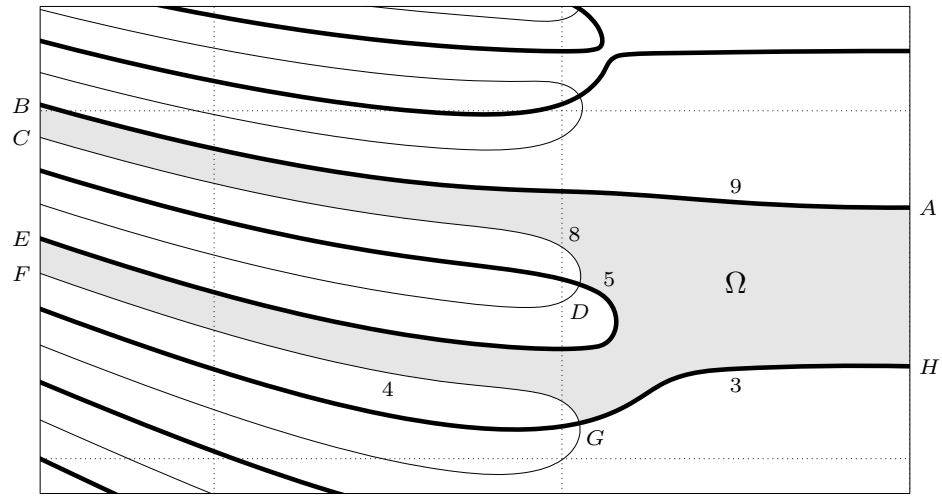


FIGURE 4. The region Ω.

As an exercise, we frame this question: Into what do the points of the oval turn?

2.6. **The equation $\zeta(s) = a$.** Let us consider an equation of the form $\zeta(s) = a$. Where are its solutions? How are they distributed?

If a is situated in a determined quadrant, say the first, then there will not be any solution in the region bordered by lines -2 and -1. In fact, the points of this region turn into points from the third quadrant. Thus, the solutions have to be situated in the regions we know to turn into the first quadrant. We can say how many solutions are there in each of these regions.

For example, let us consider the region Ω, bordered by lines 9, 8, 5, 4 and part of line 3.

As s runs through the border of this region, ABCDEFGHA, the image $\zeta(s)$ runs through a closed path, according to the following table

s	$A \to B$	$C \to D$	$D \to E$	$F \to G$	$G \to H$
$\zeta(s)$	$1 \to +\infty$	$+i\infty \to 0$	$0 \to +\infty$	$+i\infty \to 0$	$0 \to 1$

so that $\zeta(s)$ goes round the first quadrant twice. Thus, if a is a point from the first quadrant, the equation $\zeta(s) = a$ will have exactly two solutions when $s \in \Omega$, according to the argument principle.

It is clear that, under the former conditions, the equation $\zeta'(s) = 0$ must have a solution in the region Ω. There are other solutions of $\zeta'(s) = 0$, one for each of the lines -3, -5, -7, ... precisely in the points at which these lines meet the real axis.

In the X ray figure we thus see 14 solutions of $\zeta'(s) = 0$, situated on the real axis, and, besides, the existence of two more can be inferred, one in the region called Ω and the other in the region situated between lines 11 and 17.

2.7. **The construction of the graphics.** The first one who calculated zeros of the function $\zeta(s)$ was Riemann himself. Gram [7] calculated the first ten zeros and proved that they were the only ones satisfying $0 < \mathrm{Im}(\rho) < 50$. The following step in this direction were the works of Backlund [2], who managed to prove that the zeros $\alpha + i\beta$, with $\alpha > 0$ and $0 < \beta < 200$ where exactly 79, every one of them with real part $\alpha = 1/2$. Later, Hutchinson [8] managed to extend these calculations up to 300. This task has been followed by many mathematicians like Titchmarsh, Comrie, Turing, Lehmer, Rosser, and so on.

We have used the same technics they used to localize certain values of σ, the solutions of $\mathrm{Re}\,\zeta(\sigma + it) = 0$ or $\mathrm{Im}\,\zeta(\sigma + it) = 0$. Afterwards, we have written other programs, based on these, ones which situate the points belonging to the same line in order.

These graphics owe much to the program Metafont, created by Knuth to design the fonts used by TEX, and to its modification MetaPost, by J. D. Hobby, which allows one to obtain PostScript graphics.

For $|\sigma - 1/2| > 5/2$ I have proved that the curves follow the path I have drawn here, but for the area near the critical strip I just followed the method until I was convinced that the curves follow the path I have drawn. This has required, in some cases, the calculation of numerous points on each line. So they can not be considered proved.

2.8. **The existence of non trivial zeros.** One of the problems which have arised in drawing these graphics was to know the number of a line. Applying the Stirling series for $\log\Gamma(s)$ and the functional equation of $\zeta(s)$ we can prove the following:

Theorem 1. *The number of a line passing through the point $-1 + it$ (with $t > 5$) is the integer number which is nearer to*

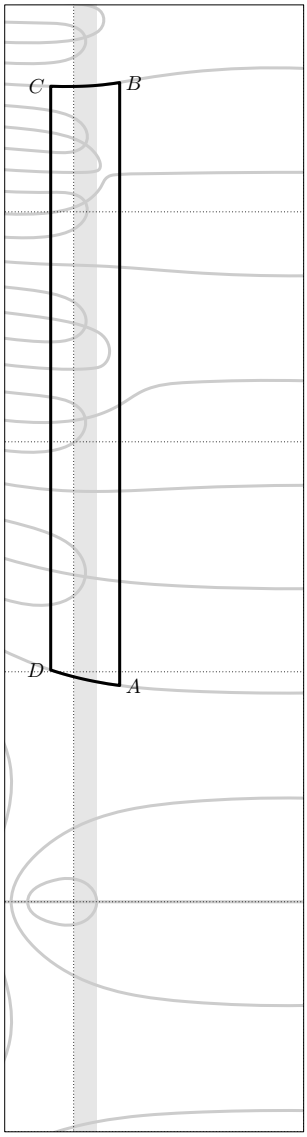
$$\frac{2t}{\pi}\log\frac{t}{2\pi} - \frac{2t}{\pi} + \frac{1}{2}. \tag{2.1}$$

Logically, the role of line $\sigma = -1$ is not important as long as we are not near the critical strip.

If we watch all the lines coming to the critical strip from the left and we know that the thin lines can not surpass a certain point, we see they must come back and they can not do it unless they cut the thick lines which accompany them. This little rigorous reasoning can be turned into a proof, by means of the argument principle, of the following theorem.

Theorem 2. *Let us suppose that $-1+iT$ is on one of the parallel lines which do not contain any zero. The number of non trivial zeros $\rho=\beta+i\gamma$ below this line and satisfying $\beta>0$ is equal to the integer number nearest to*

$$\frac{T}{2\pi}\log\frac{T}{2\pi}-\frac{T}{2\pi}+\frac{7}{8}. \tag{2.2}$$



Proof. In order to prove it, let us consider the region bordered by the lines $\sigma=-1$, $\sigma=2$, line number 3 and the line which is referred to in the statement. In the figure we have represented the case when this line is line number 17. The number of zeros we are looking for is equal to the variation of the argument of $\zeta(s)$ along this curve. But, along the segment AB, there is no variation, because all the segment turns into points of the first and the fourth quadrant, starting and finishing on the real axis. In the portions of the curves CB and DA there is also no variation because on them the function $\zeta(s)$ is real. Thus, the variation of the argument is exactly the one which takes place in the segment CD. This segment cuts several lines. As it goes from one to the other, the argument of $\zeta(s)$ varies precisely in $\pi/2$. This way, the variation of the argument will be $(N+3)\pi/2$ if the highest line is line number N. By the previous theorem N is the integer number nearest to 2.1. It follows easily that the number of zeros is equal to $(N+3)/4$, and thus it is the integer number nearest to 2.2. □

If the line we are dealing with is one of the parallel line which do contain zeros, an analogous reasoning proves that the number of zeros below this line, counting the one which is on line N on, is $(N+5)/4$, and so it is the integer number which is nearest to

$$\frac{T}{2\pi}\log\frac{T}{2\pi}-\frac{T}{2\pi}+\frac{11}{8}. \tag{2.3}$$

So, the parallel lines which do not contain zeros have a number $N\equiv 1$ and the ones which do contain zeros, on the contrary, satisfy $N\equiv 3 \pmod 4$.

3. Technics to calculate the function $\zeta(s)$.

In order to construct the graphics the possibility of calculating the function is essential. There are two basic ways to calculate the function $\zeta(s)$.

3.1. **Euler-MacLaurin Formula.** It is the following

$$\zeta(s) = \sum_{n=1}^{N-1} \frac{1}{n^s} + \frac{1}{2}\frac{1}{N^s} + \frac{N^{1-s}}{s-1} + \sum_{k=1}^{M} T_k + R(N,M),$$

where

$$T_k = \frac{B_{2k}}{(2k)!} N^{1-s-2k} \prod_{j=0}^{2k-2} (s+j),$$

B_n are the Bernoulli numbers, and there are good known bounds for the error term $R(N,M)$.

Choosing N and M conviniently as a function of s allows one to calculate $\zeta(s)$ for any s and with arbitrary precission.

3.2. **Riemann-Siegel Formula.** Before explaining what does it consist of, we should define Hardy's function (called so though it was known by Riemann)

$$Z(t) = e^{i\theta(t)}\zeta\Big(\frac{1}{2} + i\,t\Big)$$

with

$$\theta(t) = \mathrm{Im}\Big(\log\Gamma\Big(\frac{1}{4} + i\,\frac{t}{2}\Big)\Big) - \frac{t}{2}\log\pi.$$

The functional equation implies that the function $Z(t)$ is real for real values of t. This allows one to locate zeros $\rho = \beta + i\gamma$ with $\beta = 1/2$, because a change in the sign of $Z(t)$ implies the existence of a zero with abscissa exactly $1/2$.

The Riemann-Siegel formula allows one to calculate the function for any point s, but with a limited precission depending on who s is. We will only write it for a point of the form $s = 1/2 + it$. It is the following:

$$Z(t) = 2\sum_{n=1}^{m} \frac{\cos(\theta(t) - t\log n)}{\sqrt{n}} + g(t) + R,$$

where

$$\begin{aligned} m &= \lfloor\sqrt{t/2\pi}\rfloor \\ g(t) &= (-1)^{m-1}\Big(\frac{t}{2\pi}\Big)^{-1/4} h(\xi) \\ h(\xi) &= (\sec 2\pi\xi)\cos 2\pi\phi \end{aligned} \qquad \begin{aligned} \xi &= \Big(\frac{t}{2\pi}\Big)^{1/2} - m \\ \phi &= \xi - \xi^2 + 1/16 \end{aligned}$$

The error is $\boldsymbol{\mathcal{O}}(t^{-3/4})$. The Riemann-Siegel formula requires, in order to achieve a fixed precission, the calculation of approximately $\sqrt{t}$ terms of the first sum, while the Euler-MacLaurin formula requires a number of terms of the order of t. When $t \approx 10^9$, this difference turns out to be very significant.

This formula has an interesting story. In a letter directed to Weierstrass, Riemann claims that:

> the two theorems I have just stated here: that between 0 and T there exist approximately $\frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi}$ real roots of the equation $\xi(\alpha) = 0$.
>
> that the series $\sum_\alpha \bigl(\mathrm{Li}(x^{1/2+i\alpha}) + \mathrm{Li}(x^{1/2-i\alpha})\bigr)$, when its terms are ordered according to the incresing growth of α, tends to the same limit than the expression
>
> $$\frac{1}{2\pi i \log x} \int_{a-bi}^{a+bi} \frac{d\frac{1}{s}\log\frac{\xi\left((s-\frac{1}{2})i\right)}{\xi(0)}}{ds} x^s\, ds$$
>
> when b grows beyond boundaries.
>
> *are consecuences of a new expansion ξ which I have not been able to simplify enough to comunicate it.*

Riemann considered proved[1] all his other claims contained in [14].

Because of this, it was clear that, among Riemann's papers should be the expansion he was talking about. Some 70 years after his death, C. L. Siegel [16] managed to figure out these papers. It would be worth for the reader to have a look at the photocopy of the sheets of Riemann's manuscript which contain the famous expansion, which can be found in Edward's book [6], to realize the difficulty of Siegel's task.

4. Graphics of the zeta-function in the critical line.

In the following pages we present some graphics corresponding to the values of t between 0 and 560. We have just drawn the strip $-1 \le \sigma \le 2$. The part which is not represented has nothing new and our imagination can supply it without any trouble. On the left we have pointed the values of t, each time it increases in twenty unities, and we have written the number of some lines (the parallel ones which do not contain any zeros). We leave till later the explanations about some points we have marked in the graphics with little circles, situated on the critical line.

[1] It seems that the Riemann's capability to prove his claims has been doubted more than once without proper support. To get more information on this point you can consult my paper[1].

In the usual proof about the number of non trivial zeros up till a given height T it is proved that a horizontal segment cutting the critical strip between $\sigma = -1$ and $\sigma = 2$, at a height of, say, t, can only cut a number of lines of the order $\log t$. This, and the fact that the lines coming from the right can not cut themselves, because that would generate a region bordered by points where the function $\zeta(s)$ is real and bounded, allows Speiser to infer that the lines coming from the right have to cross the critical strip and move away to the right towards the infinity.

In fact, if one of these lines went up towards the infinity through the critical strip it would force the others to go above it, also through the critical line, to infinity. As there are a number of them of the order of t, we would reach a contradiction with the results already proved about the horizontal segment in the critical strip.

As the number of lines on the left has an order of $t \log t$, we see that it is necessary for some of them to go back to the right, joining others. Speiser calls these figures, formed by two lines, when the function is real, *sheets*.

The graphics give rise to some integer sequences, the most obvious is the sequence formed by the number of lines which escape to the right. That is,

$$-3, -1, 1, 3, 9, 11, 17, 23, 29, 35, 41, 47, 53, 59, 69, 75, 81, 91, 97, 103, 113,$$

$$123, 129, 135, 145, 155, 161, 171, 181, 187, 197, 207, 217, 223, 237, 247, 253$$

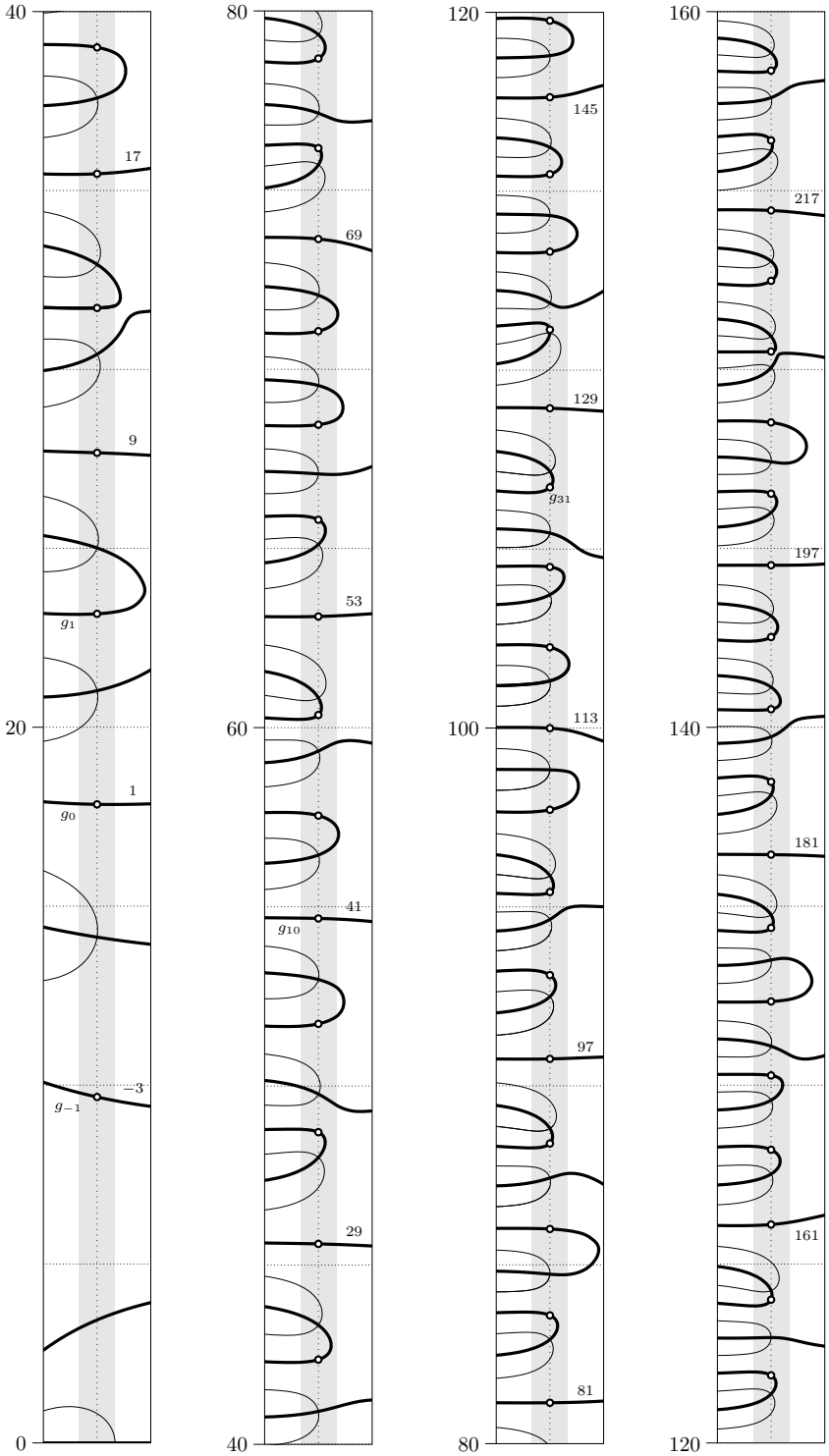
$$263, 273, 283, 293, 307, 313, 323, 329, 343, 353, 359, 373, 383, 393, 403, 417$$

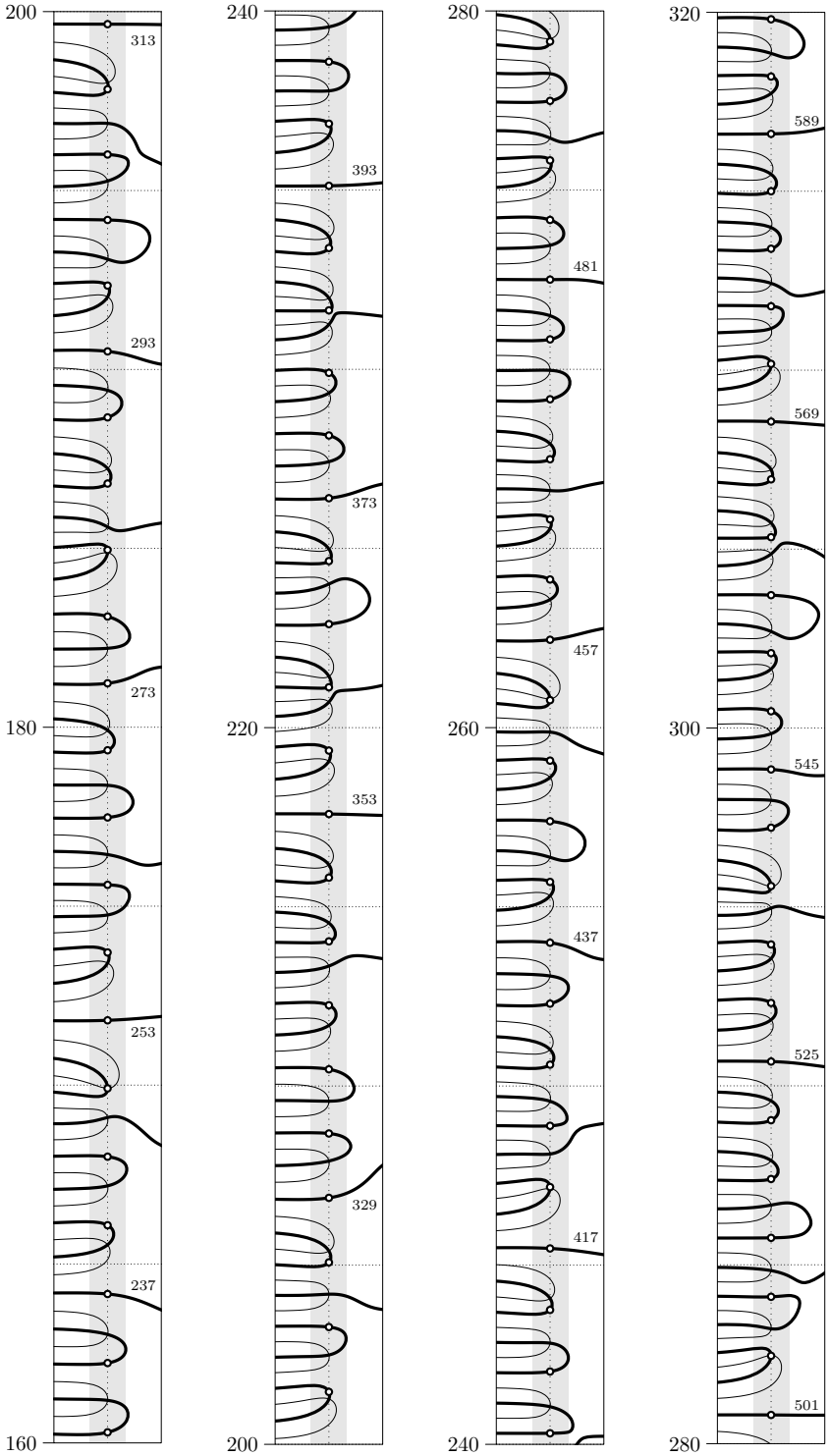
$$423, 437, 451, 457, 467, 481, 491, 501, 511, 525, 535, 545, 559, 569, 579, \ldots$$

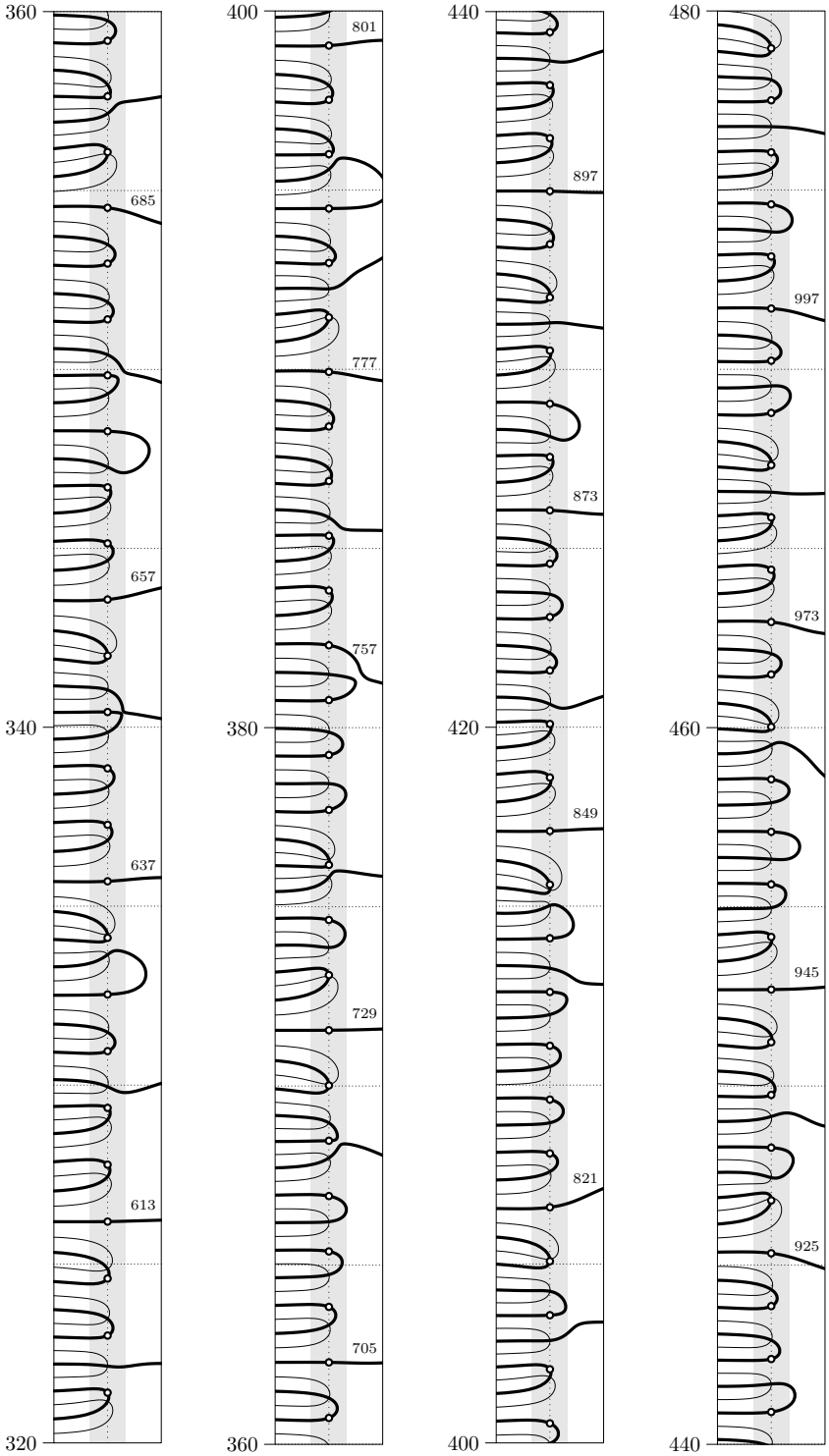
Speiser suggests that it could be connected with the distribution of the prime numbers. I could not see any conection.

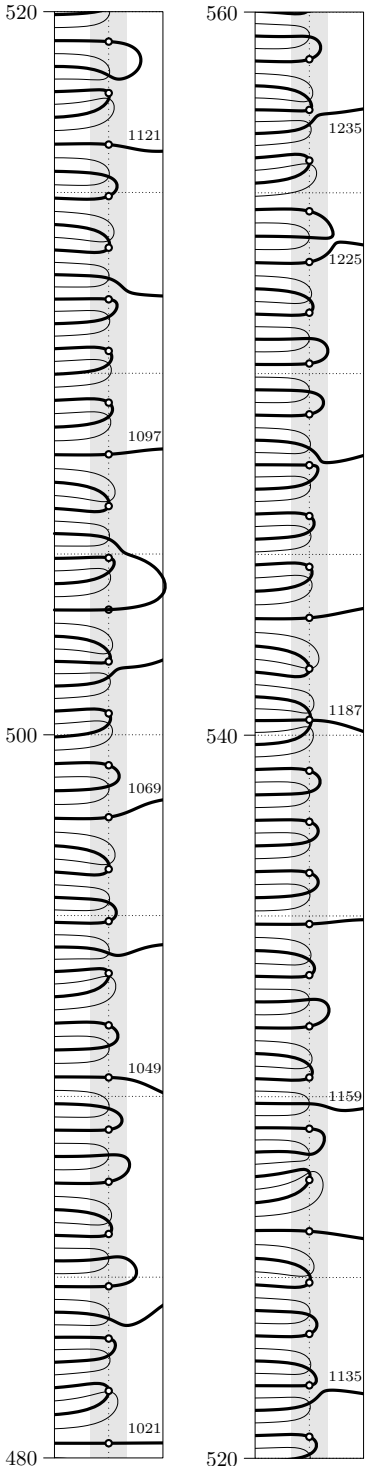
Looking at these graphics, a general scheme seems to appear. Between two parallel lines which do not contain zeros there are an even number of thin lines, joining by pairs, a thick line coming parallel from the right cuts one of these loops and the rest are inserted with loops of thick lines. Later we will see how these simple ideas about the function break.

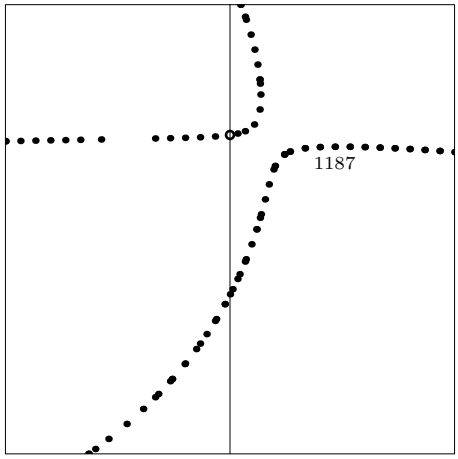
The thin lines almost do not cross over the critical line, while the thick lines sometimes reach $\mathrm{Re}(s) = 2$. In fact, we see that at a high height they do cross over this line, this happens for the first time in the sheet formed by lines 789 and 791. But what is really surprising happens in line 1085. This forms a sheet with line 1091. So this sheet surrounds completely the one formed by lines 1087 and 1089. For higher values of t this becomes quite frequent.


40
20
0
17
9
g_1
1
g_0
-3
g_{-1}
80
60
40
69
53
41
g_{10}
29
120
100
80
145
129
g_{31}
113
97
81
160
140
120
217
197
181
161


200
313
293
273
180
253
237
160
240
393
373
353
220
329
200
280
481
457
260
437
417
240
320
589
569
545
300
525
501
280


360
340
320
685
657
637
613
400
380
360
801
777
757
729
705
440
420
400
897
873
849
821
480
460
440
997
973
945
925





Sometimes, as in line 1187, can happen that the path followed by the line is not clear. In such cases we draw the confusing area, for example, in this case we can see here a square which represents the rectangle $(0.2, 0.8) \times (540, 540.8)$, marking the points where the function $\zeta(s)$ is real. This way it is clear that line 1187 turns down when it gets to the critical line, and the pair of lines 1189 and 1191 form a sheet.

The monotony of the function on the curves is valid except in the poles, provided we define suitably the prolongation of the curve in the points where the derivative vanishes. In fact, in one of these points several curves meet, and there are as many on which the function increases as on which the function decreases, thus it always is possible, when we reach a zero of order n of the derivative, to go back with an angle of $2\pi/n$, so that the monotony is maintained all through the curve.

Another particularity of the curves we would like to emphasize is that, except possibly the ones which go through the poles, the curves have to leave every compact. Thus they come from and go to the infinity.

5. First Theorem of Speiser

Theorem 3 (Speiser)**.** *The Riemann Hypothesis is equivalent to the fact that all the sheets meet the critical line.*

Proof. Let us recall that the sheets are the thick lines in which $\zeta(s)$ is real and which, coming form the left, go back to the left joining another thick line.

In the first place, on each sheet, the zeta-function takes all the real values. In fact, the fuction is real and monotonous and it tends to infinity as $\sigma \to -\infty$. Thus in a sheet there is always one (and only one) zero of the function.

If some sheet would not touch the critical line, the corresponding zero would be an exception to the Riemann Hypothesis.

The proof of the other implication is a little more intricate. In the first place, we must notice that if a real line crosses the critical line through a point α, then it holds that $\zeta(s) = 0$ or else the modulus $|\zeta(s)|$ decreases as the curve goes through this point from left to right. To see this, we notice that, since the function of Hardy $Z(t)$ is real for real values of t, it follows that $\theta(t) + \arg\zeta(1/2+it) = \text{cte}$, except at one zero of the zeta-funcion. Since $\theta(t)$ is an increasing function it follows that $\partial_t \arg\zeta(1/2+it) < 0$ unless $1/2+it$ is a zero of the zeta-function. Let us consider then the analytic function

$$\log\zeta(1/2+it) = \log|\zeta(1/2+it)| + i\arg\zeta(1/2+it).$$

From the Cauchy-Riemann equations it now follows that the derivative of $|\zeta(s)|$ with respect to σ in the point $\sigma = 1/2$ must be negative.

If a line (of a sheet) surpasses the critical line from left to right it must go through it again in order to come back to the left. If the zero which is on the sheet is not one of the points where the line cuts the critical line, we can apply the results obtained above to both of them. It follows that in one of the points it must be satisfied that $\zeta(s) > 0$, and in the other, $\zeta(s') < 0$, so that the absolute value can be decreasing in both. Since the function is monotonous along the sheet, it follows that the zero is situated on the right of the critical line.

Thus there are two kinds of sheets cutting the critical line. In some of them, the zero is located in one of the two cuts between the critical line and the sheet. In others, the zero is situated on the right of the critical line.

Consecuently, if the Riemann Hypothesis is false, there would be a zero situated on the left of the critical line, which can be situated only on a sheet which does not cut the critical line. If the Riemann Hypothesis is true, then all the sheets are of the kind which have a zero exactly on the line and so they cross it (or are tangent at it).

This way, we see that if the Riemann Hypothesis is true all the sheets would meet the critical line. □

In the figures above we can see that, in fact, all the sheets do contain a zero and another point on which it cuts the critical line. We have marked the latter with a

little circle. These points are the so called Gram points, and we will see some interesting results about these points below.

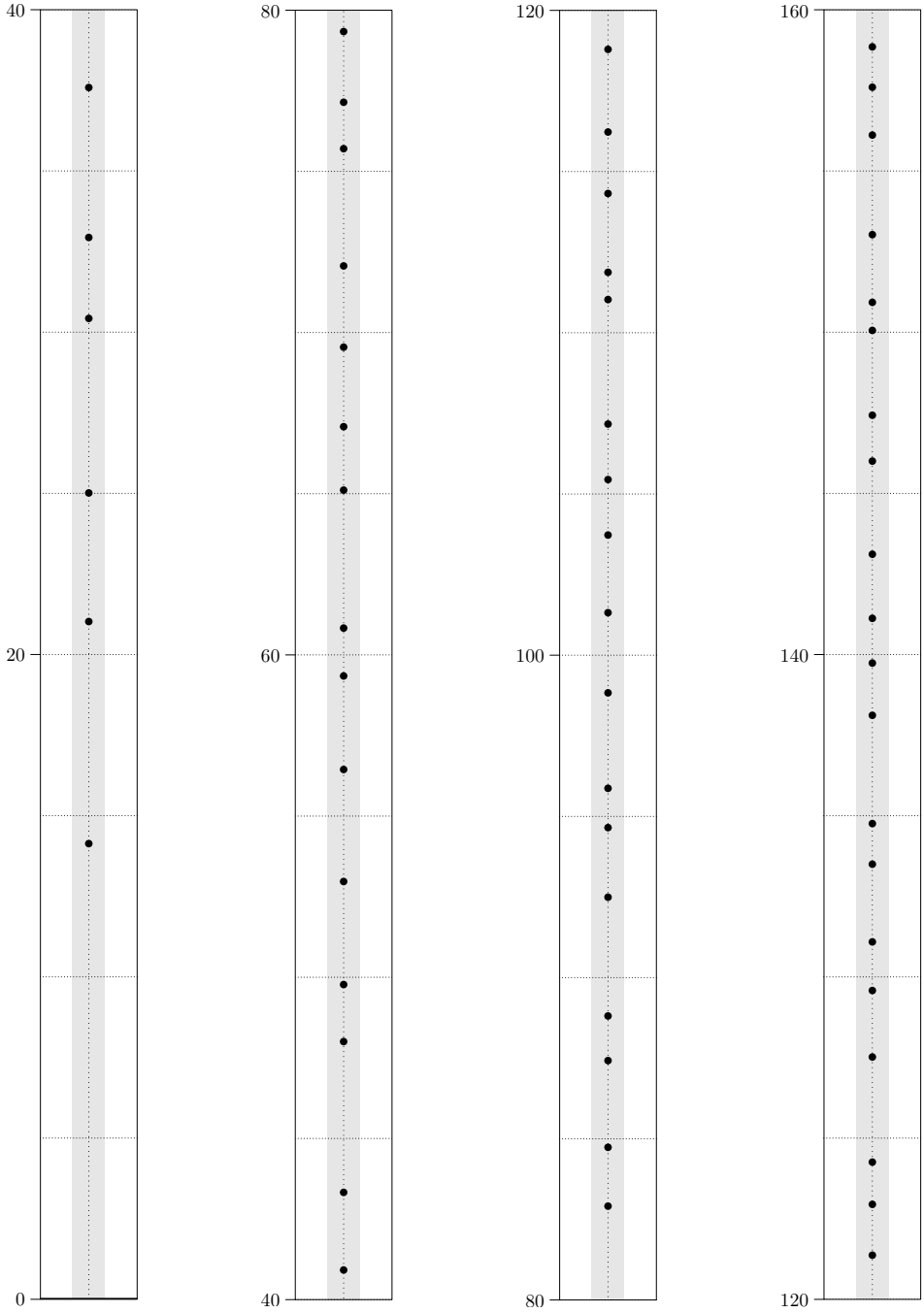


FIGURE 12. Zeros of $\zeta(s)$

6. Separating the zeros. Gram points.

In the last page we marked the zeros of the function $\zeta(s)$. We see that there is some randomness in their distribution.

To try to understand their distribution the Gram points are brought in. The Gram points, which we have marked with a little circle in our figures, are points on the critical line where the function $\zeta(s)$ is real and does not vanish. In the graphics of the following page we have marked Gram points. It is remarkable how regularly are these points distributed.

The Gram point g_n is defined as the solution of the following equation:

$$\theta(g_n) = n\pi$$

Since Hardy's function is real,

$$\zeta(1/2 + i\, g_n) = e^{-i\theta(g_n)} Z(g_n) = (-1)^n Z(g_n)$$

is real. Thus these points are situated on thick lines.

Looking at the preceding figures, we verify that, in most of the cases represented there, it holds:

$$\zeta(1/2 + i\, g_n) > 0.$$

We will see later that this inequality has exceptions for greater values of t.

In the Riemann-Siegel formula, the first term, which is the most important, has a value of $(-1)^n$ for $t = g_n$, which partly explains the tendency of $Z(g_n)$ to have the sign $(-1)^n$.

An important consequence of the preceding is that $Z(g_n)$ and $Z(g_{n+1})$ have oposed signs, so that the function $\zeta(1/2+it)$ will vanish in the interval (g_n, g_{n+1}). In Figure 14 we can see this graphically.

When one intends to calculate the real zeros of a polynomial, a first task is to separate the zeros, that is, to find an increasing sequence of values for t in which the polynomial takes alternatively values of different signs. An analogous aim is achieved with Gram points.

Gram noticed that these points seemed to separate the zeros of the zeta-function and claimed that this would be true for not too high values of t. Hutchinson named the fact that in each interval (g_n, g_{n+1}) would be a zero of the zeta-function Gram's law.

Titchmarsh uses this idea to prove that there is an infinity of zeros on the critical line, proving that the mean value of $Z(g_{2n})$ is positive and that of $Z(g_{2n+1})$ is negative.

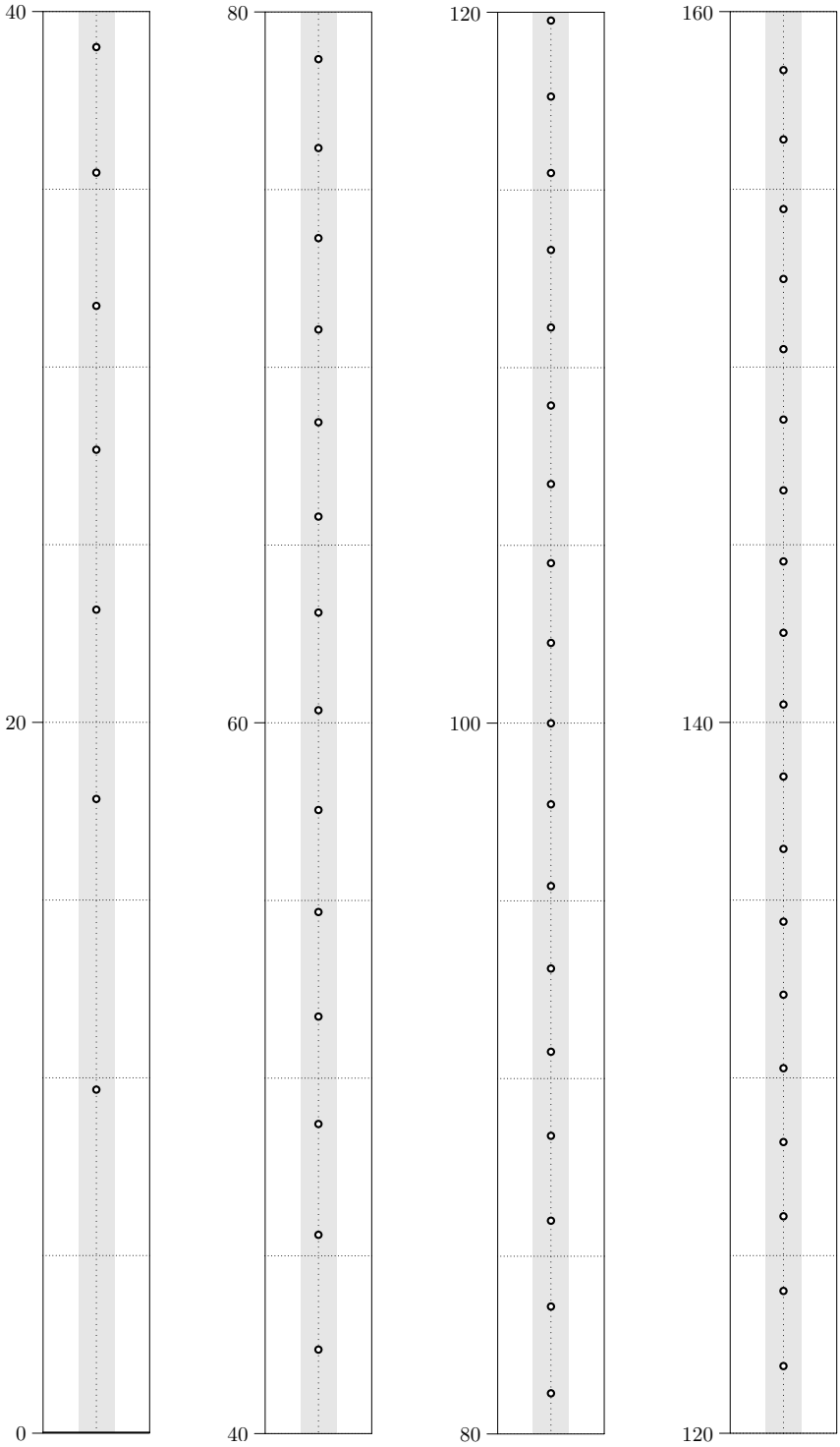


FIGURE 13. Gram Points.

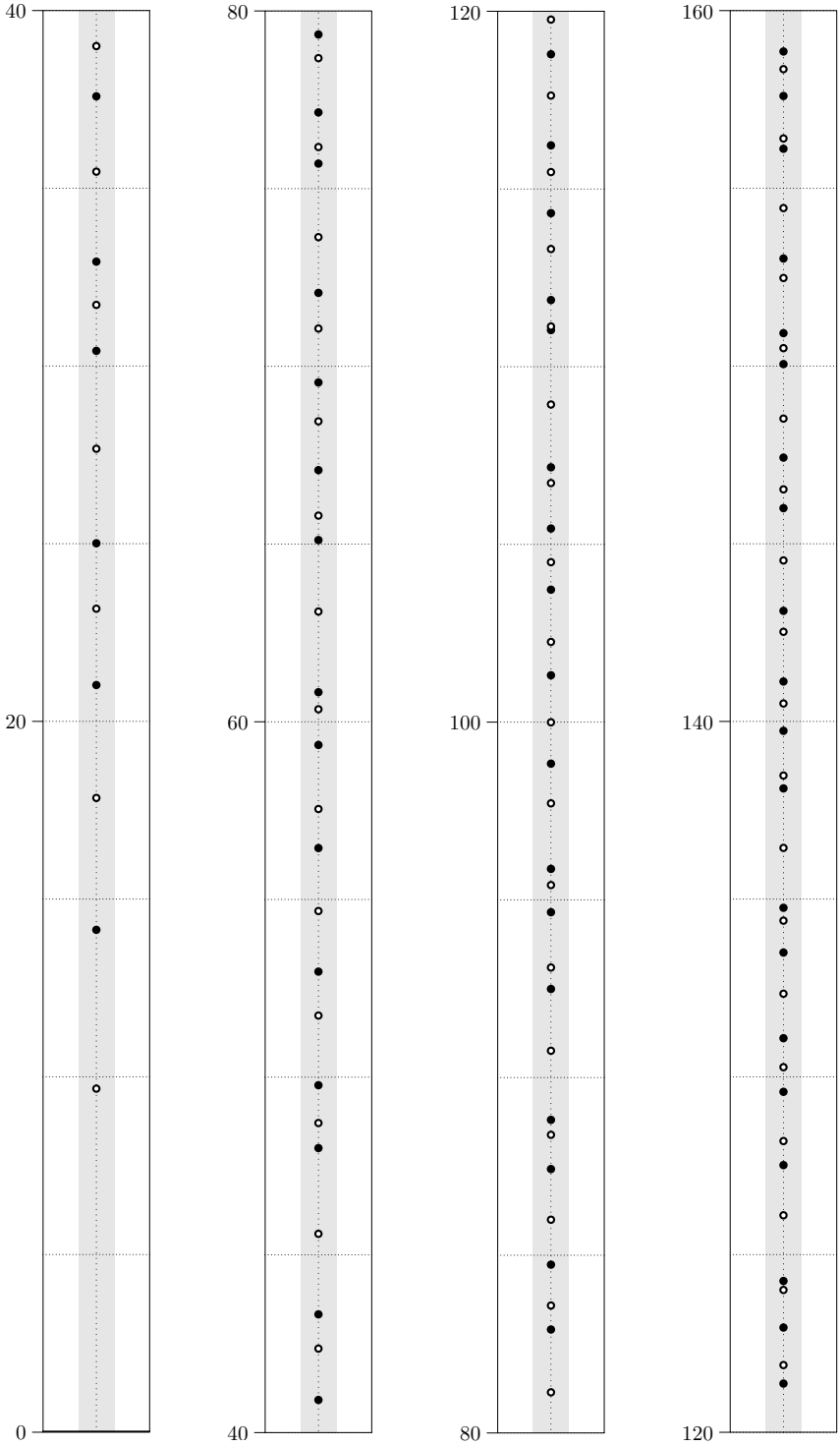


FIGURE 14. Gram's points separate the zeros.

7. Second Theorem of Speiser.

We present here Speiser's proof, and, as we have already said, his methods are between the proved and the acceptable. Everybody quotes him but nobody reproduces his theorems. His methods do not seem convincing to me, either, though I think his proof is essentially sound. We present it more like a challenge: to turn it into a proof, filling its gaps. In any case, a flawless proof of a stronger result can be found in Levinson and Montgomery [12].

Theorem 4 (Speiser)**.** *The Riemann Hypothesis is equivalent to the fact that the non trivial zeros of the derivative $\zeta'(s)$ have a real part $\geq 1/2$, that is, that they are on the right of the critical line.*

Proof. Let us assume that there is a zero a of $\zeta(s)$ on the left of the critical line. Let us consider the lines of constant argument $\arg\zeta(s) = \text{cte}$ which come from the point a. In all of them the modulus $|\zeta(s)|$ is increasing. Thus, these lines can not cross the critical line, because at crossing it from left to right, the absolute value of $\zeta(s)$ ought to decrease.

They also, can not be tangent at the critical line, because the tangency point would be a point in the critical line where $\arg\zeta(1/2+it)$ would be stationary, and this is possible only if it is a zero. But, in the line, $|\zeta(s)|$ increases, starting from zero, so it can not go through a zero of the function.

Thus all these lines come back to the left. Some of them go back leaving the point a below, others leave it above. The line which separates both kinds of lines must reach a zero of the derivative, which will allow it to go back. This would be a zero of the derivative on the left of the critical line.

Consequently, if the Riemann Hypothesis is false, we see that there must exist a zero of the derivative on the left of the critical line.

Now let us suppose that $\zeta'(a) = 0$, where $\mathrm{Re}(a) < 1/2$. We have to find a zero of the function on the left of the critical line. We can assume that $\zeta(a) \neq 0$, because if this were the case we would have already finished. Since the derivative vanishes, there exist two opposite lines, of constant argument and along which $|\zeta(s)|$ decreases. We follow these two lines, and we must reach a zero, because $|\zeta(s)|$ decreases. If it is on the left of the critical line, we have finished, while, in other case, it is clear that we will reach the critical line.

Our two paths, till they reach the critical line, and the segment from the critical line they determine, enclose a region Ω. From the point a, two opposite paths also set off, along which $|\zeta(s)|$ increases, and the argument of $\zeta(s)$ is constant. One of them enters our region Ω, (because in the point a, the borderline of the region has a well-defined tangent).

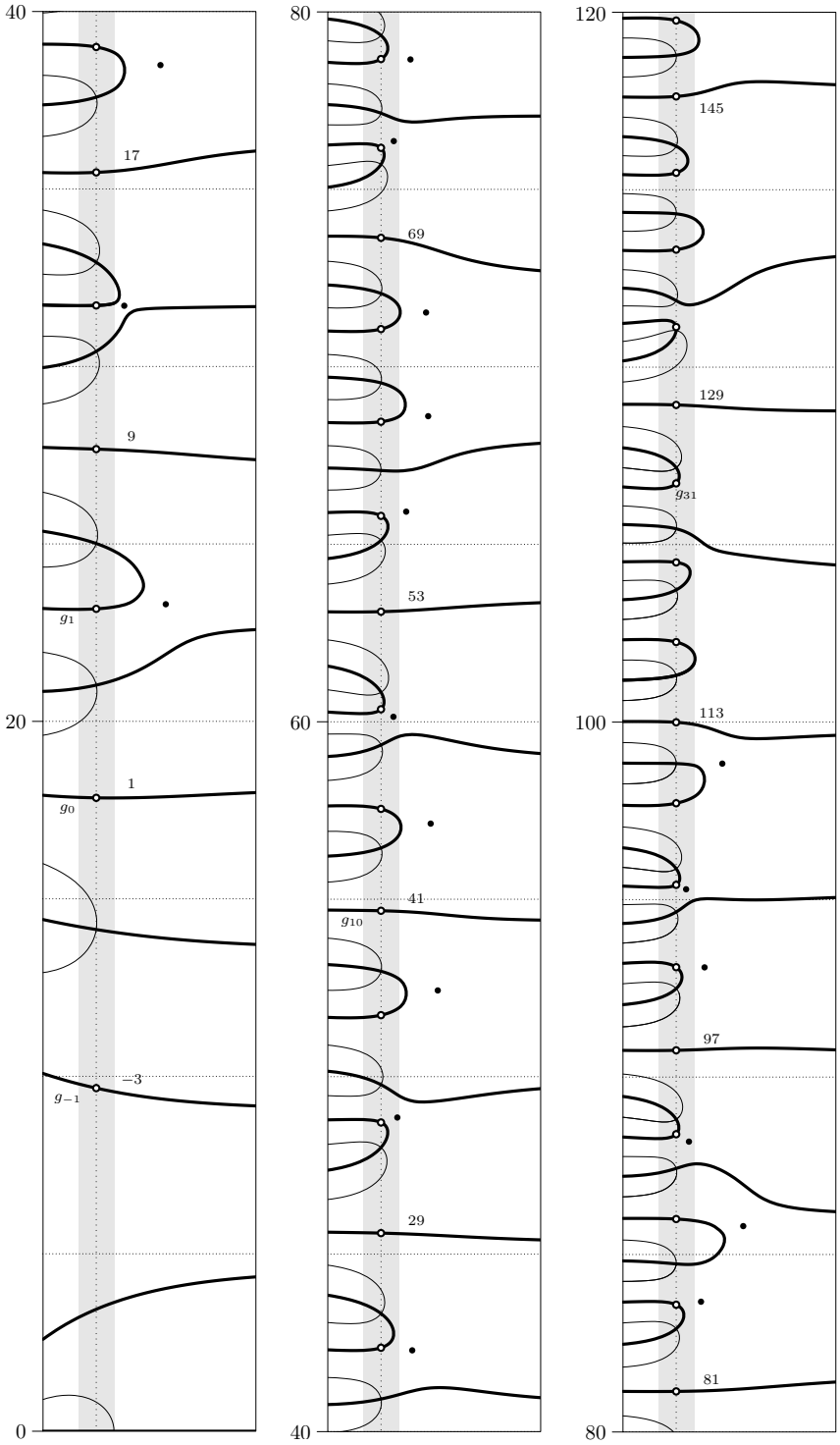


FIGURE 15. Zeros $\zeta'(s)$.

We will follow this path. Since $|\zeta(s)|$ increases to infinity along this curve, it must leave the region Ω, but it can not do it across the curves which we used to define it since along them $|\zeta(s)| < |\zeta(a)|$, and we are considering a curve where the values of $|\zeta(s)|$ are greater than $|\zeta(a)|$. It also can not leave Ω across the segment from the critical line, because to cross it from left to right it ought to do it with $|\zeta(s)|$ decreasing.

Thus supposing that there was no zero of $\zeta(s)$ on the left of the critical line has lead us to a contradiction. □

Speiser's theorem makes the zeros of $\zeta'(s)$ far more interesting. R. Spira who has given a complete proof of half of Speiser Theorem [18] has calculated the first ones (those which have an abscissa less than 100), which are represented in the figure.

We see that the real curves (thick ones) seem to be attracted by the zeros, and each sheet seem to have one zero associated with it, which would justify their crossing the critical line to approach their corresponding zero. In this way we have an insight about the place where the zeros which are beyond line 113 (that we have not draw in the figure) are situated.

We see that a zero of the derivative does explain the behaviour of line 11. If the derivative vanished at a point in which the function is $\zeta(s)$ is real, at this point two thick lines would meet perpendiculary. What happens here is that the function is almost real in the zero and the curves resemble the meeting we have described. This is also what happens in line 1187. If we see the graphics with the dots we had to do in order to see the path this line follows, we can verify that the derivative at this point has a zero with an abscissa slightly greater than 1/2.

Because of Speiser's reasoning we know that a line which comes parallel from the right at a height T does not rise or fall of level until it crosses the critical strip at a height greater than $\boldsymbol{\mathcal{O}}(\log T)$. Thus, the number of parallel lines below it is $T\log 2/\pi + \boldsymbol{\mathcal{O}}(\log T)$. This lines, alternatively, contain a zero of $\zeta(s)$ which is not associated with a sheet, or do not contain it. Thus the zeros which are not associated with a sheet up till a height T is approximately equal to $T\log 2/2\pi$. According to this, the number of sheets below a height T is

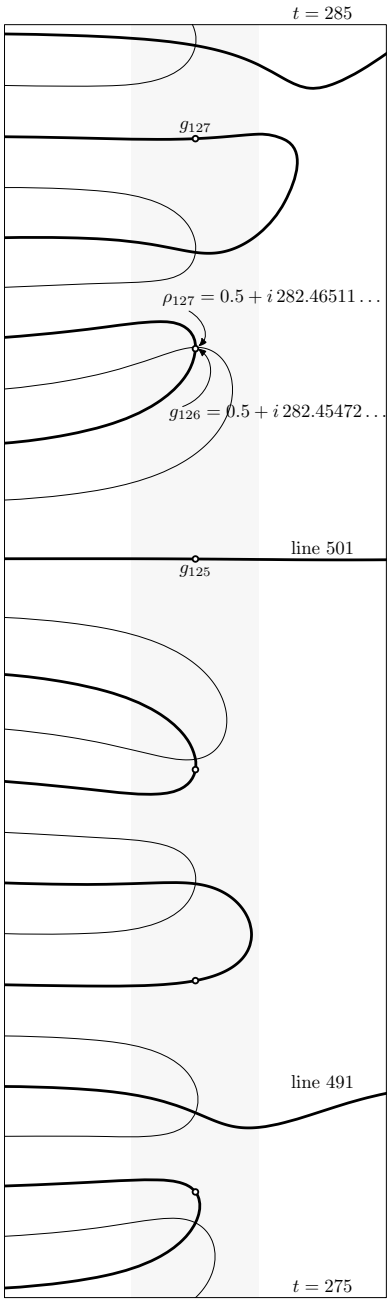
$$\frac{T}{2\pi}\log\frac{T}{2\pi} - \frac{T}{2\pi} + \frac{7}{8} - \Big(\frac{T\log 2}{2\pi} - \frac{1}{2}\Big) = \frac{T}{2\pi}\log\frac{T}{4\pi} - \frac{T}{2\pi} + \frac{11}{8},$$

with an error of the order of $\log T$.

If the remarks we have made concerning the zeros of the derivative $\zeta'(s)$ are true this is the number of zeros of the derivative up till a height T. Spira conjectured this claim and Berndt [3] has proved it.

8. Looking higher. A Counterexample to Gram's Law

We will now see that most of the regularities in the behaviour of the function break at a great enough height.



Gram's law claims that between every two consecutive Gram points there exists a zero of the function $\zeta(s)$. The remark is Gram's [7], but it was Hutchinson who named it *Gram's law*, although it was him who found the first counterexample, which can be seen in the figure. Gram [7] only claimed that this would happen for the first values of n. The interval (g_{125}, g_{126}) does not contain zeros of the zeta-function. On the contrary, the next interval contains two zeros thus reestablishing the total count.

Later Lehmer [11] finds out that the exceptions grow more frequent as n increases. He also notices that, in general, these exceptions consist of a Gram interval in which there are no zeros, next to another which has two.

Possibly the only valid rule was the one formulated by Speiser: The number of thick lines crossing the line $\sigma = -1$ below a height T is

$$\frac{T}{\pi}\log\frac{T}{2\pi} - \frac{T}{\pi} + \frac{1}{4}$$

and the number of Gram points below this height is only half this number. A thick line which crosses the critical line has to do it through a Gram point or a zero. The parallel lines, alternatively, contain a Gram point or a zero. Speiser believes that each sheet uses a Gram point and a zero to enter and exit the area on the right of the critical line.

This is true, there exists a bijectiv map between the zeros and Gram points, but it is the one established by the fact that they are on the same sheet. It may be convenient to call also sheet to a parallel line which does not contain zeros and the parallel line inmediately above it.

This way, another sequence of natural numbers associated with the graphics of the zeta-function arises. In fact, Gram points g_{-1}, g_0, g_1, ... are associated with the zeros numbered

$$1, 2, 3, 4, 5, 7, 6, 8, 10, 9, 11, 13, 12, 14, 16, 15, 17, 18, 20, 19, 21, 23, 24, 22, 26,$$
$$25, 27, 28, 30, 31, 29, 32, 34, 33, 35, 36, \ldots$$

It is a permutation σ of the natural numbers, so that $|\sigma(n) - n| \leq C \log n$. But, actually, it is well defined only if the Riemann Hypothesis is valid.

9. Almost Counterexample to Riemann Hypothesis (Lehmer)

An important landmark in the numerical study of the zeros of the zeta-function is Lehmer's paper [11] in the year 1956. In it, he proves that the first 10000 zeros of the function have a real part exactly equal to 1/2, so that the Riemann Hypothesis is valid at least for $t \leq 9878.910$.

He establishes that, at this height, one out of ten Gram interval does not satisfy Gram's law. The number of exceptions increases continuously. He also finds out that in many occasions, in order to separate a zero, he must turn to Euler-MacLaurin formula because Riemann-Siegel's one does not have enough precission. This happens because the zeros of the zeta-function are very close. He studies specifically a particularly difficult case: it is an area near $t = 1114,89$, situated in the Gram interval (g_{6707}, g_{6708}), where the function has two extremely close zeros,

$$\frac{1}{2} + i\ 7005.0629 \qquad \frac{1}{2} + i\ 7005.1006.$$

We will not repeat the graphics that Lehmer made about the behaviour of the function $Z(t)$ in a neighborhood of these points. Between these two zeros Hardy's function has the lowest relative maximum. This maximum is only 0.0039675 and it occurs at the point $t = 7005.0819$. Looking at the terms of $Z(t)$ in this point we see that a few of the first terms quickly increasing are counteracted by conspiracy of lots of small terms which sum up, thus the maximum turns out almost negative. A negative relative maximum would imply, it can be proved, a counterexample to the Riemann Hypothesis. So we call this situation an almost counterexample to the Riemann Hypothesis.

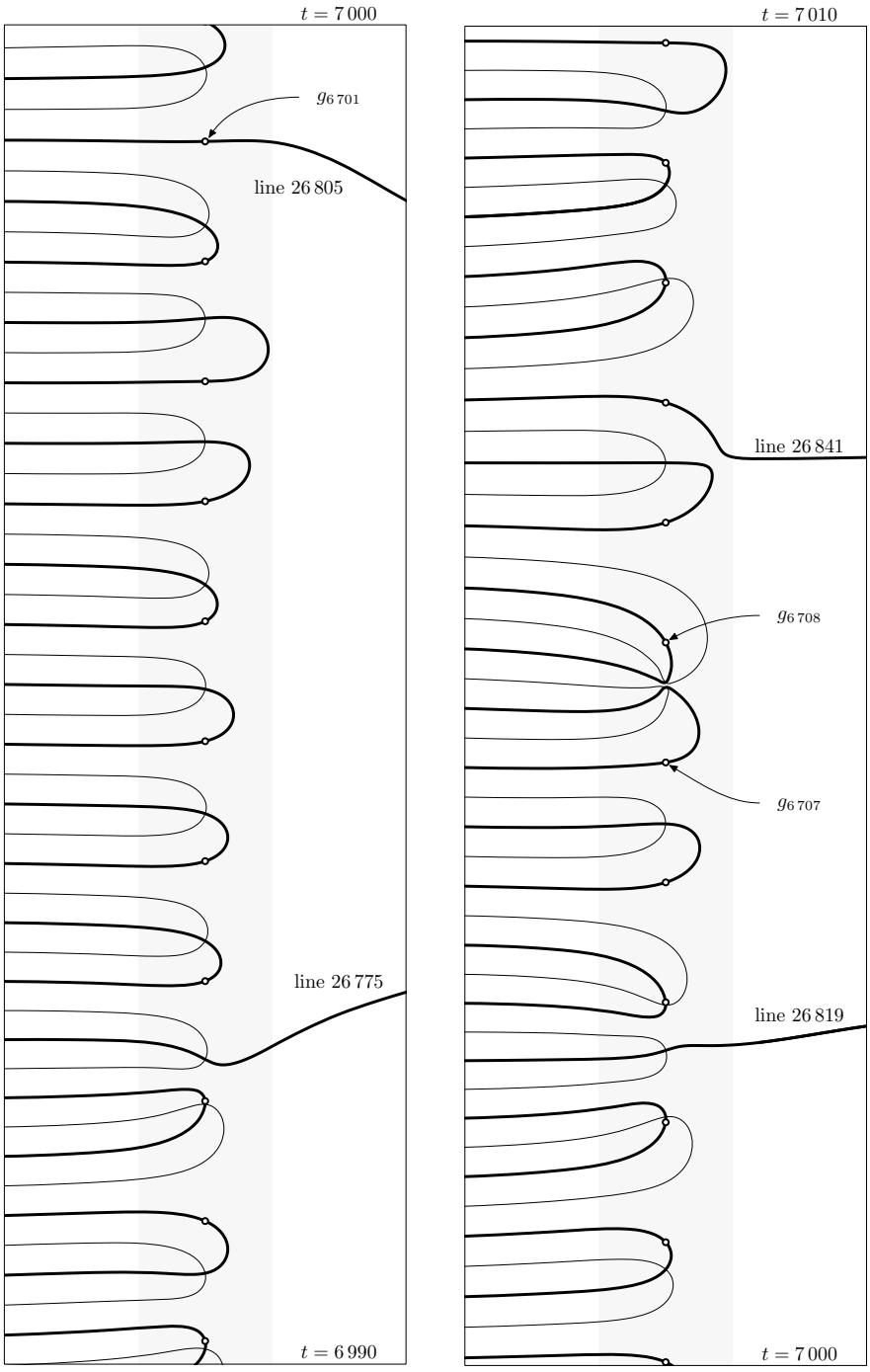


FIGURE 17. Almost counterexample of Lehmer

In the figures included in this page we can see the case analyzed by Lehmer. We notice that, from our point of view, it consists of two very close zeros, so that, viewing it from a far distance, it seems to be a double zero. We see that the lines seem to be continued more smoothly by the ones which are not actually joining them. Instead, the lines turn abruptly. Every time this is the case, that is, there are drastic changes in the direction of the lines, there is a zero of the derivative hanging about.

It would be easy to modify the lines artificially so that the two lines containing the Gram points would join and the other two thick lines would join each other too. So, slightly modifying the path of the thin lines, we could generate two zeros outside the critical line, which are symmetrical with respect to each other.

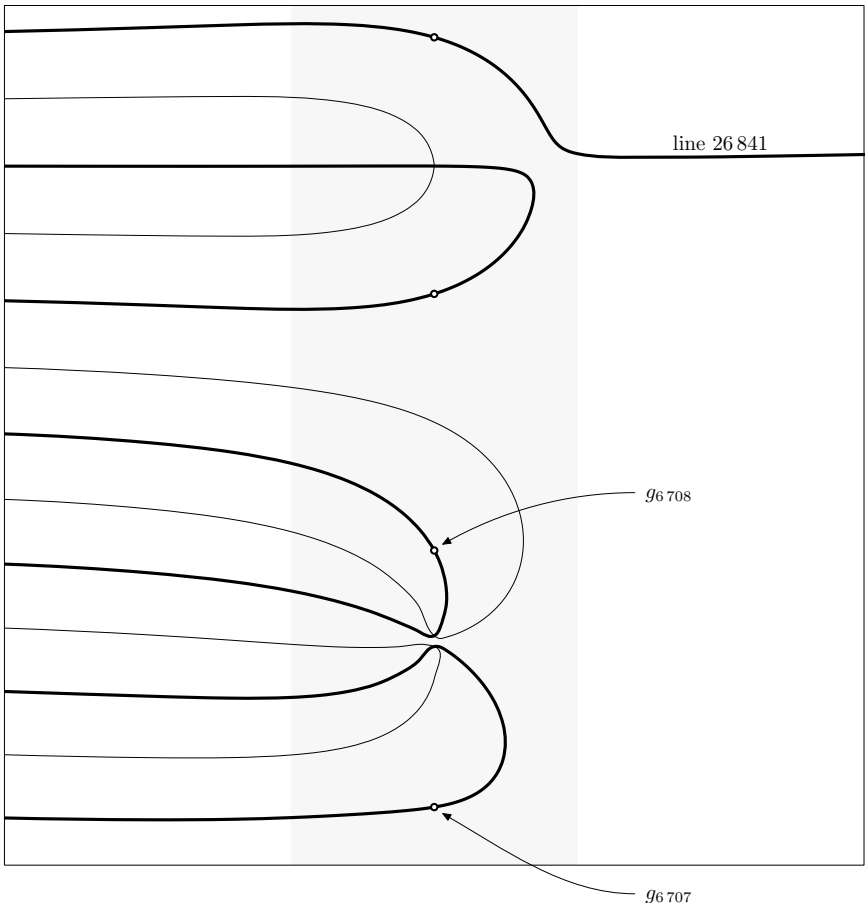


FIGURE 18. Detail of the last figure.

10. Rosser Law

Rosser, Yohe and Schoenfeld (1968) expand Lehmer's calculations and prove that the first 3 500 000 zeros are simple and situated on the critical line. These authors find out a certain regularity in the failures of Gram's law.

They distinguish between *good* and *bad* Gram points. Good points are those at which $\zeta(1/2+ig_n)>0$ holds, and bad ones are the rest of them. We know how to tell these points apart with the nacked eye in the X ray. Bad Gram points are surrounded by a thin line.

They call **Gram block** to a consecutive set of bad Gram points surrounded by two good ones. For example, in the precedent figure points g_{6707}, g_{6708} and g_{6709} form a Gram block.

Rosser's law claims that in a Gram block there are as many zeros as the number of Gram's intervals.

10.1. **The function $S(t)$.** Let $N(T)$ be the number of zeros $\rho=\beta+i\gamma$ with $0\le\gamma\le T$. An approximation to $N(T)$ is $\pi^{-1}\theta(T)+1$, so that

$$N(T)=\pi^{-1}\theta(T)+1+S(T).$$

The value $\pi S(T)$ is also the variation of the argument of $\zeta(s)$ when s goes from $+\infty+iT$ to $1/2+iT$.

Von Mangoldt proved that, as Riemann says, $S(T)=\boldsymbol{\mathcal{O}}(\log T)$. Later Littlewood proved

$$\int_0^T S(t)\,dt=\boldsymbol{\mathcal{O}}(\log T).$$

Selberg proves that

$$S(t)=\Omega_\pm\bigl((\log t)^{1/3}(\log\log t)^{-7/3}\bigr).$$

Thus, there exist values of t at which $S(t)$ is as high as we want it to be.

10.2. **First counterexample to Rosser's law.** The first counterexample to Rosser's law (see Figure 9) is in the Gram block

$$(g_{13999525},g_{13999527})$$

in which there is no zero of the function. In the following interval $J=(g_{13999527},g_{13999528})$ there are three zeros which balance the total count.

In the graphics we see how the function $S(T)$ takes a value greater than 2 in a point which is situated between point $g_{13999527}$ and the first zero contained in the interval J.

This is a general rule: Gram's law is satisfied as long as $|S|<1$ and Rosser's law as long as $|S|<2$.

We have already said that there exist points on which S takes values as high as desired.

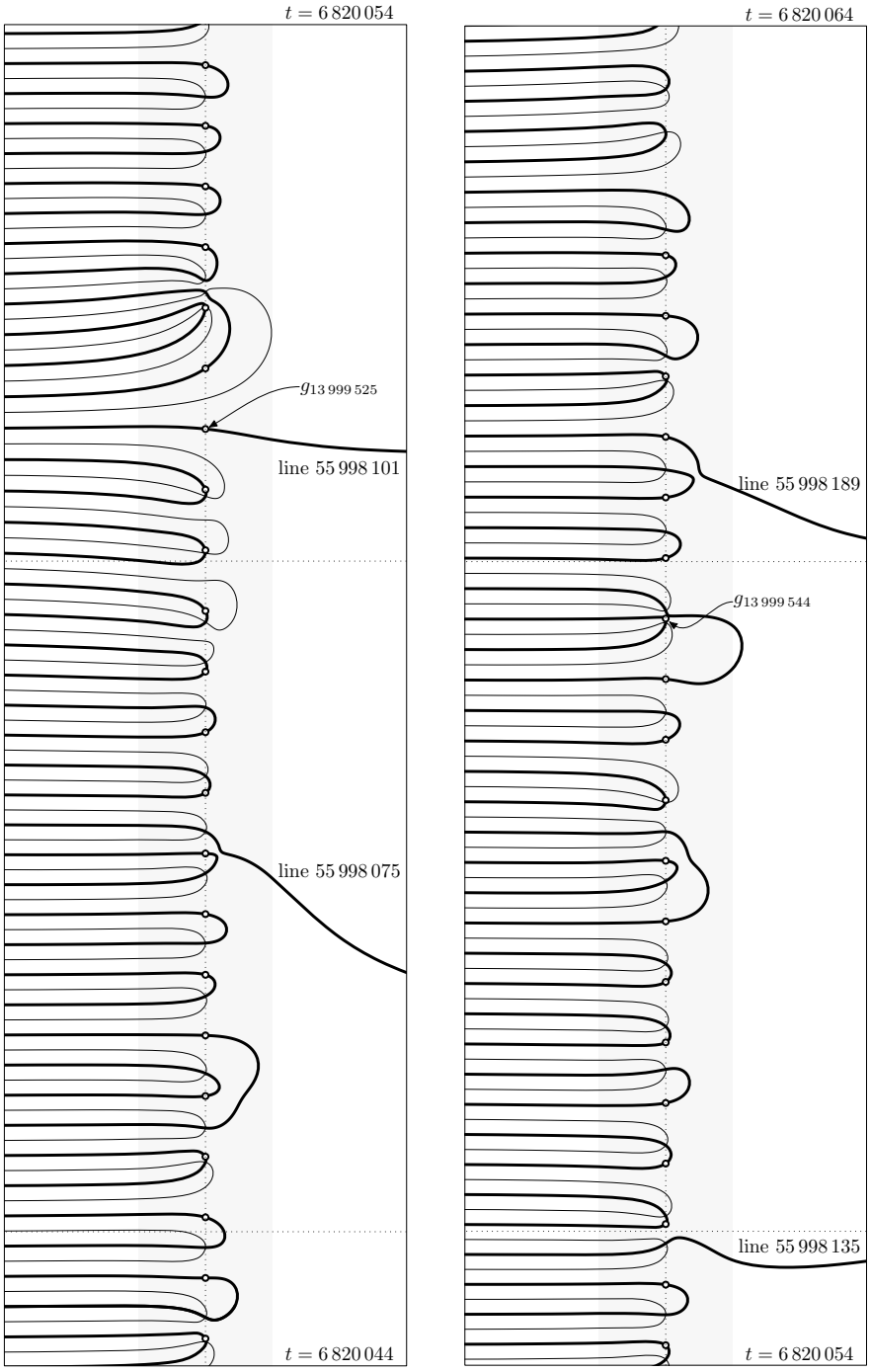


FIGURE 19. First counterexample to Rosser's law.

A high value of $S(t)$ corresponds to a point in the critical line, $1/2+it$, such that the segment from $1/2+it$ to $+\infty+it$ meets a high number of our lines.

In the tables from [4] and [5] we see that the extreme values of S which are known hardly surpass an absolute value of 2.

In Brent's table is already quoted the first counterexample to Rosser's law associated to the Gram point number 13999525, where $S(t)$ reaches a value of -2.004138. In fact, we see that there is a point $1/2+it_0$ imperceptibly above $g_{13999527}$, but before the next zero of $\zeta(s)$, such that the segment from this point to $1+it_0$ meets firstly a thick line, then a thin line, followed by another thick and another thin line. In the point $+\infty+it_0$ we start from a value equal to 1, enter the fourth quadrant, cross a thin line and thus we reach the third quadrant, cross a thick line entering the second quadrant, cross another thin line and find ourselves in the first quadrant, and finally another thick line and so we end up in the fourth quadrant. Consequently, the argument has changed in a quantity between 2π and $5\pi/2$, that, as we can see, agrees with the value given by Brent.

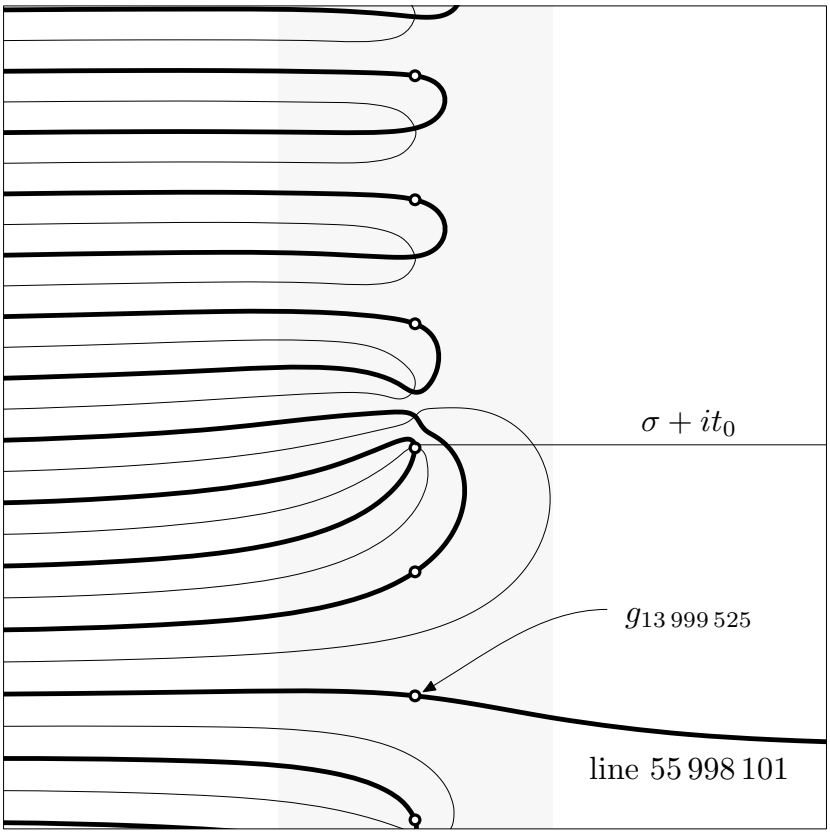


FIGURE 20. Detail of the last figure

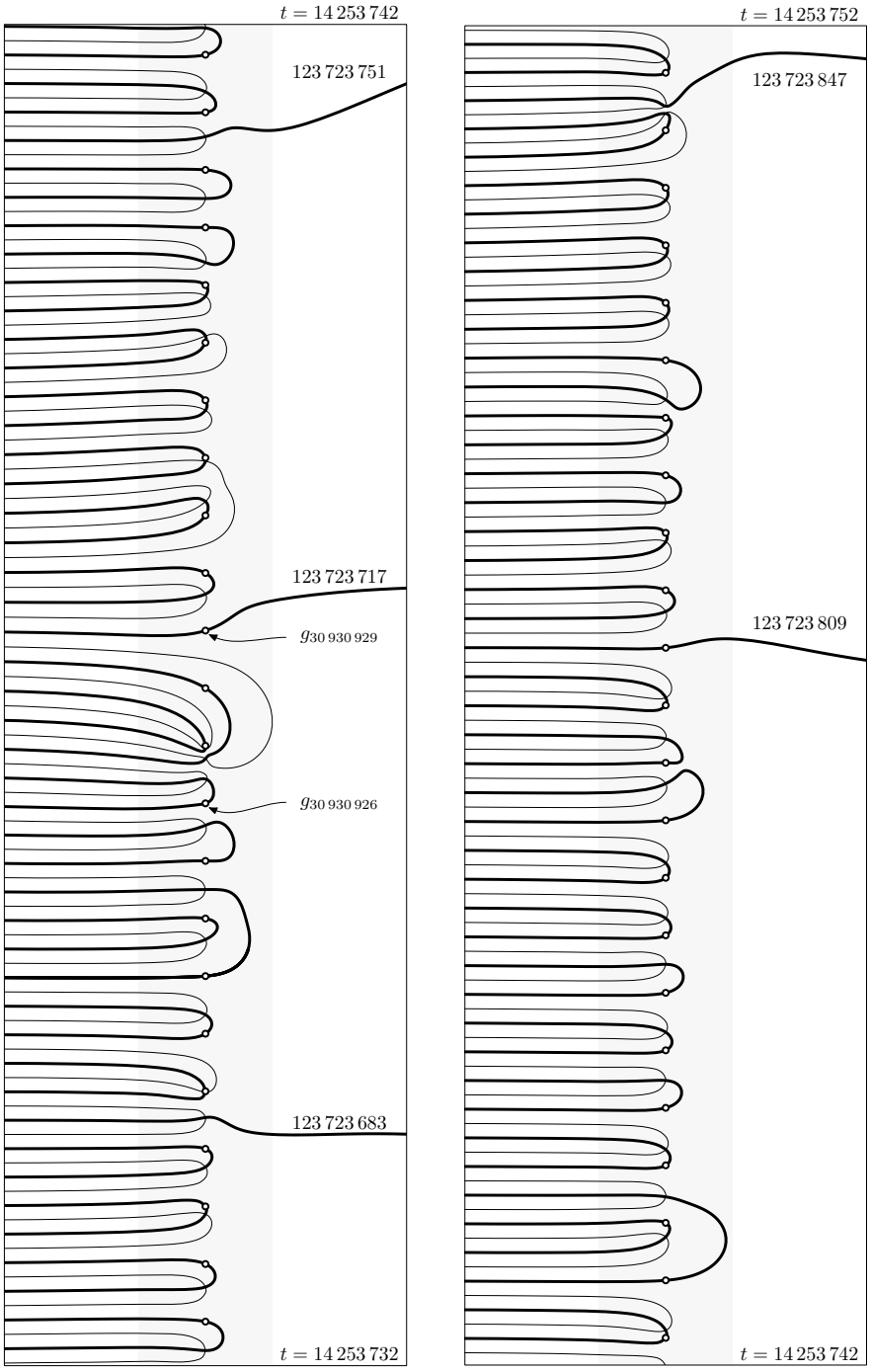


FIGURE 21. Another counterexample to Rosser's law.

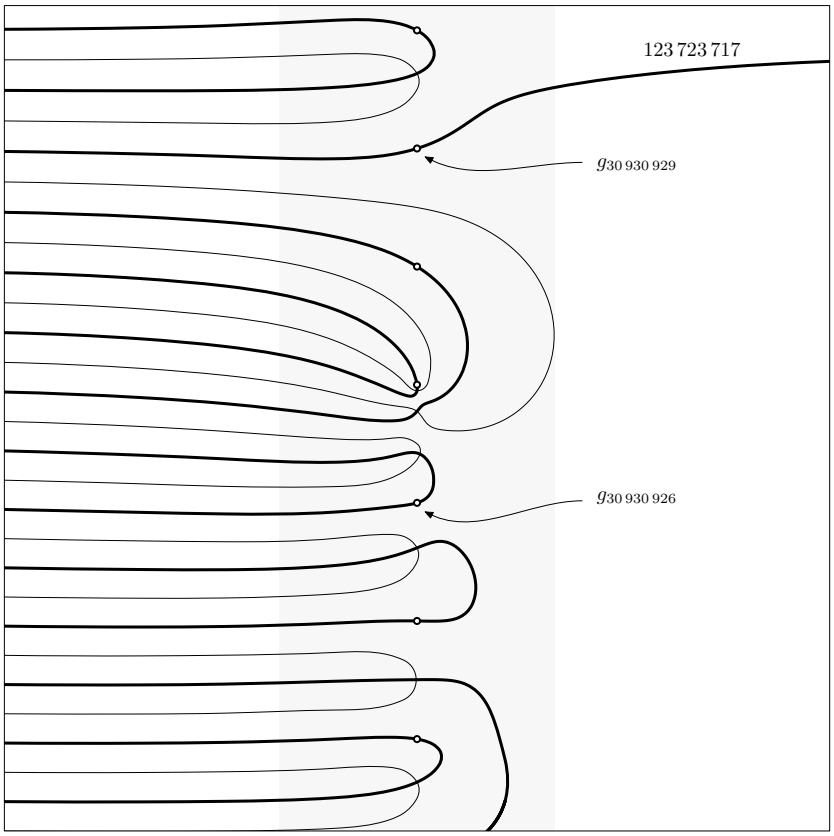


FIGURE 22. Detail of the last figure.

In the example presented in this page, S reaches a value 2.0506 and so the existence of the point $1/2 + it_0$ is a little clearer. In this case, S is positive, while in the precedent it was negative.

The previously quoted result of Selberg assures us that there are points at which $S(t)$ is as high as desired. Thus we can assume that, in a higher level, we will see coils which are analogous to Figures 22 and 20, but in which an arbitary number of lines gets involved.

Watching the preceding figures one may wonder if the thin lines do not cross the line $\sigma = 0$. This is not true, by a theorem of Bohr (see [19] p. 300) the function $\zeta(s)$ takes every value $\neq 0$ infinitely often in the halfplane $\operatorname{Re}(s) > 1$. In fact Van de Lune [13] has shown that $\sigma_0 = \sup\{\sigma \in \mathbf{R} : \operatorname{Re}\zeta(\sigma + it) < 0 \text{ for some } t \in \mathbf{R}\}$ is given by the unique solution of the equation $\sum_p \arcsin(p^{-\sigma}) = \pi/2$, $\sigma > 1$. Brent and Van de Lune have computed $\sigma_0 = 1.1923473371861\ldots$ with more than 400 decimal digits.

We finish with three figures of the zeta-function near a thousand millions, to show a ramdomly chosen area.

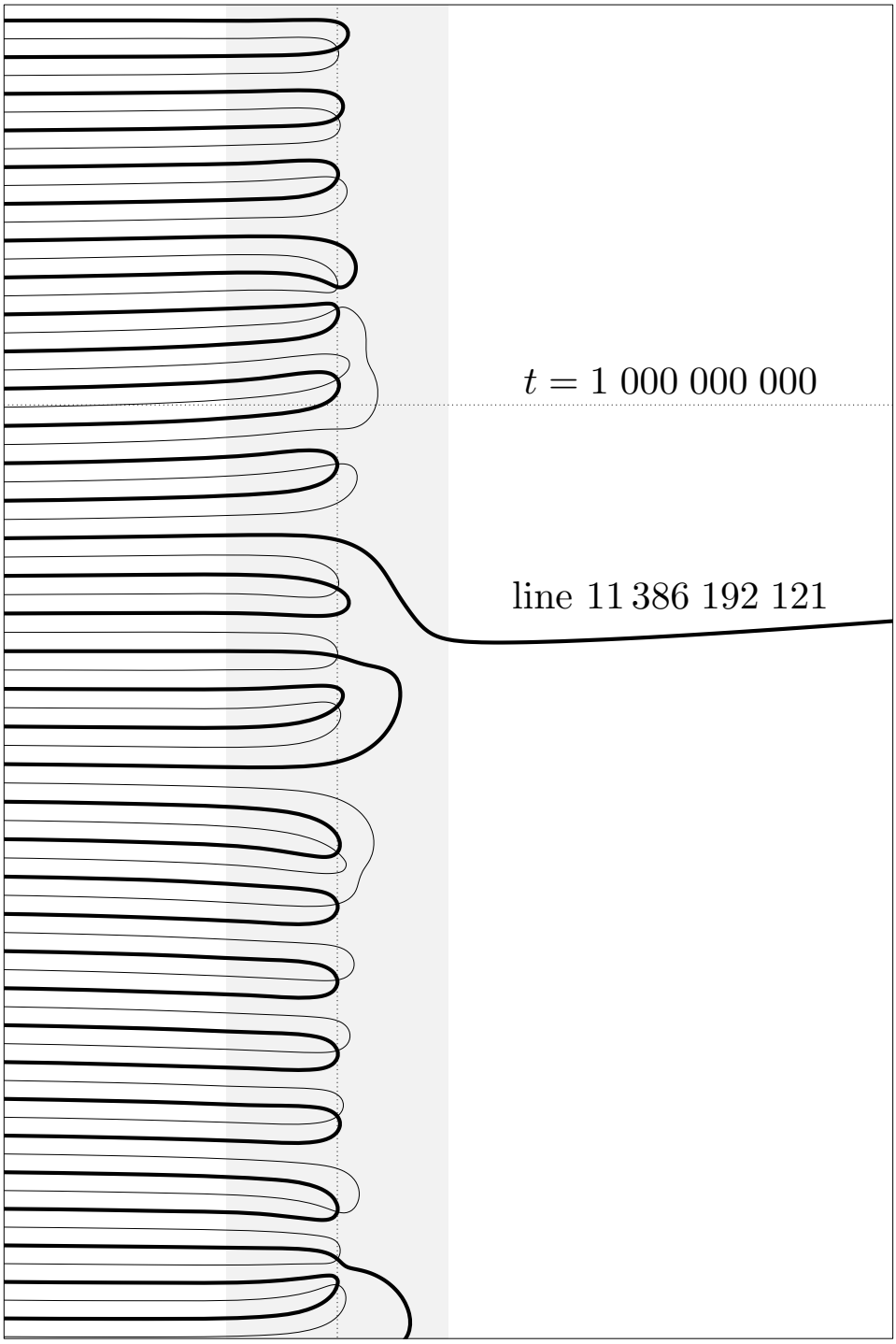


FIGURE 23. $\zeta(s)$ near $t = 1\,000\,000\,000$.

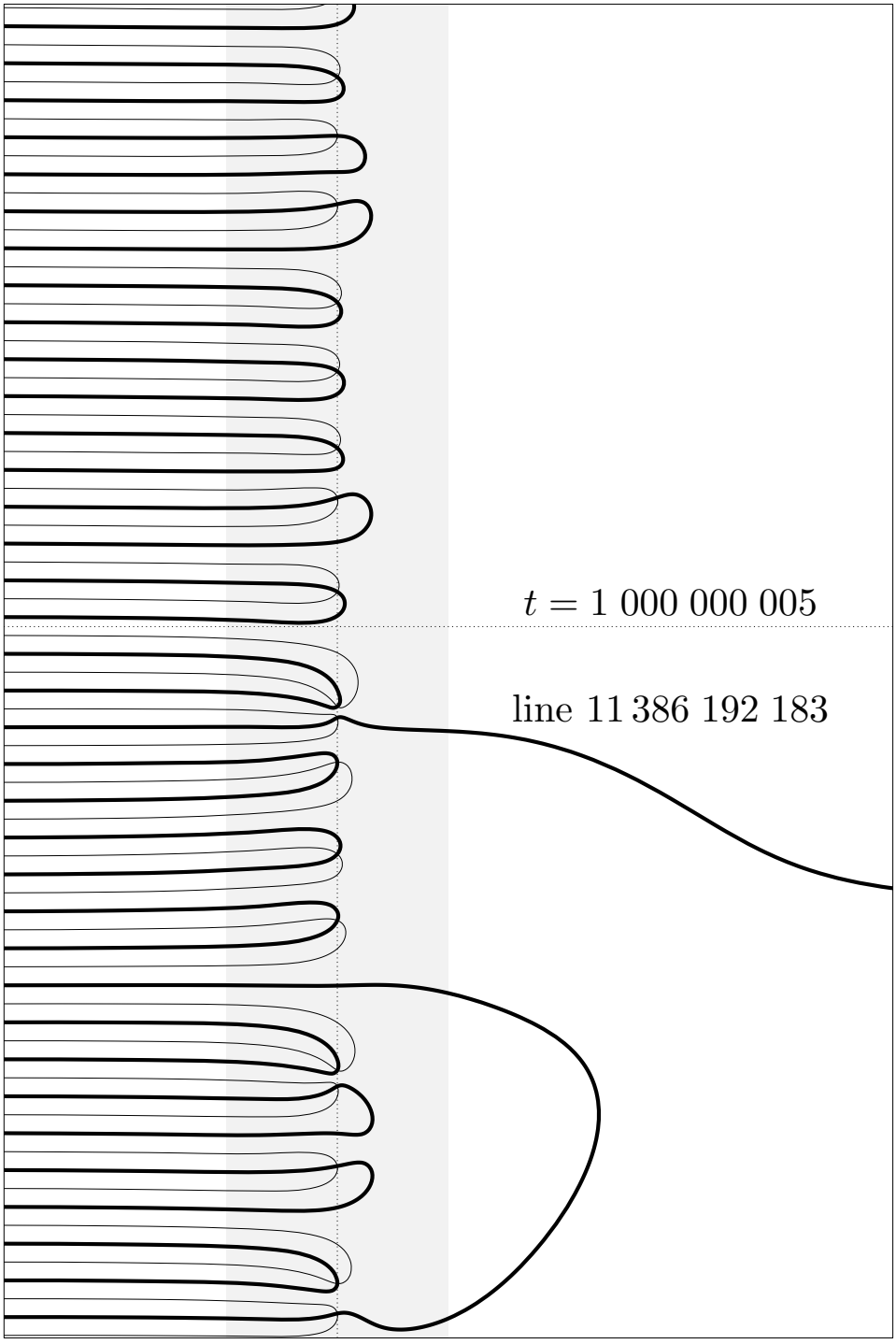


FIGURE 24 . $\zeta(s)$ near $t = 1\,000\,000\,000$.

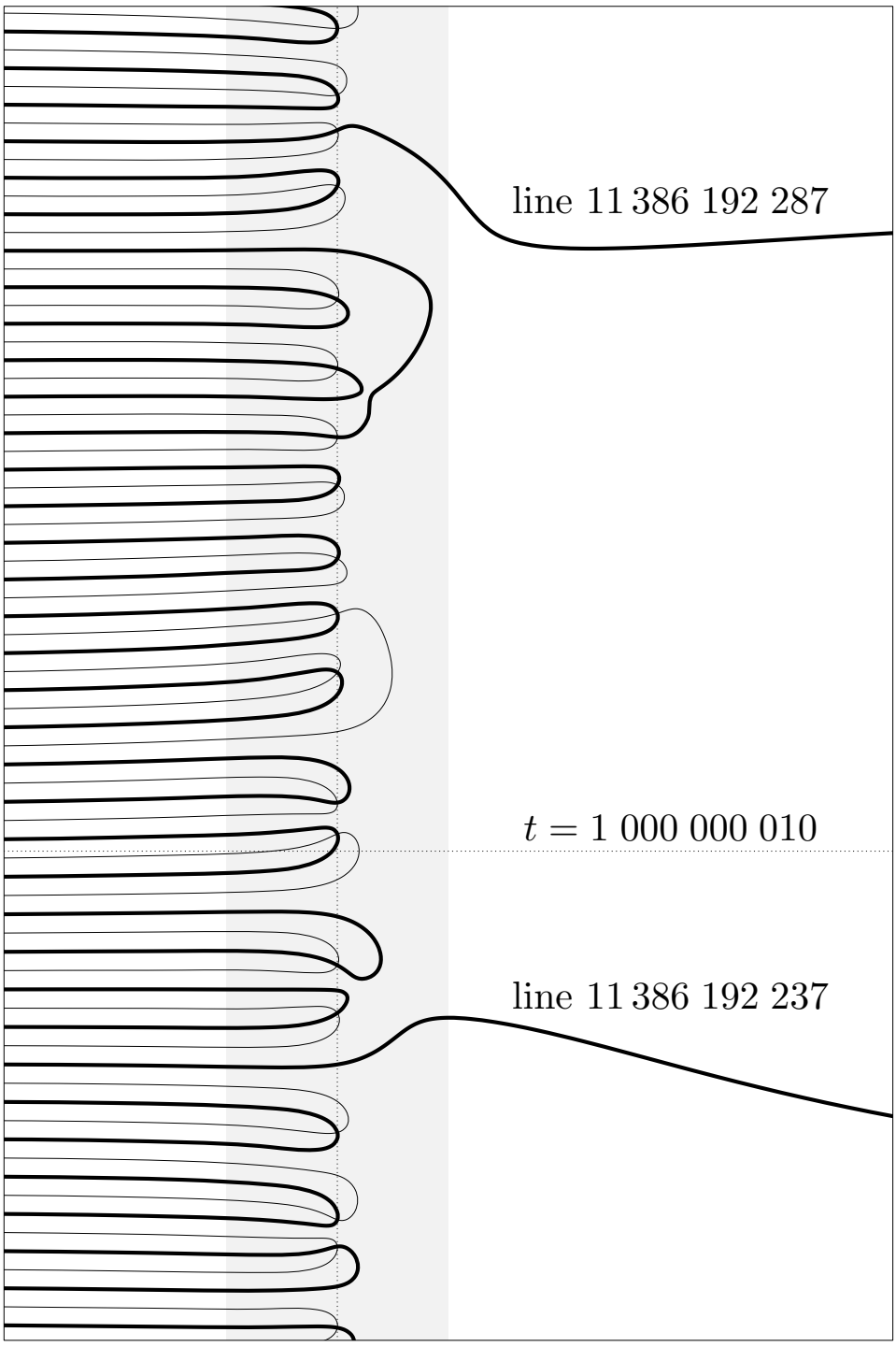


FIGURE 25. $\zeta(s)$ near $t = 1\,000\,000\,000$.

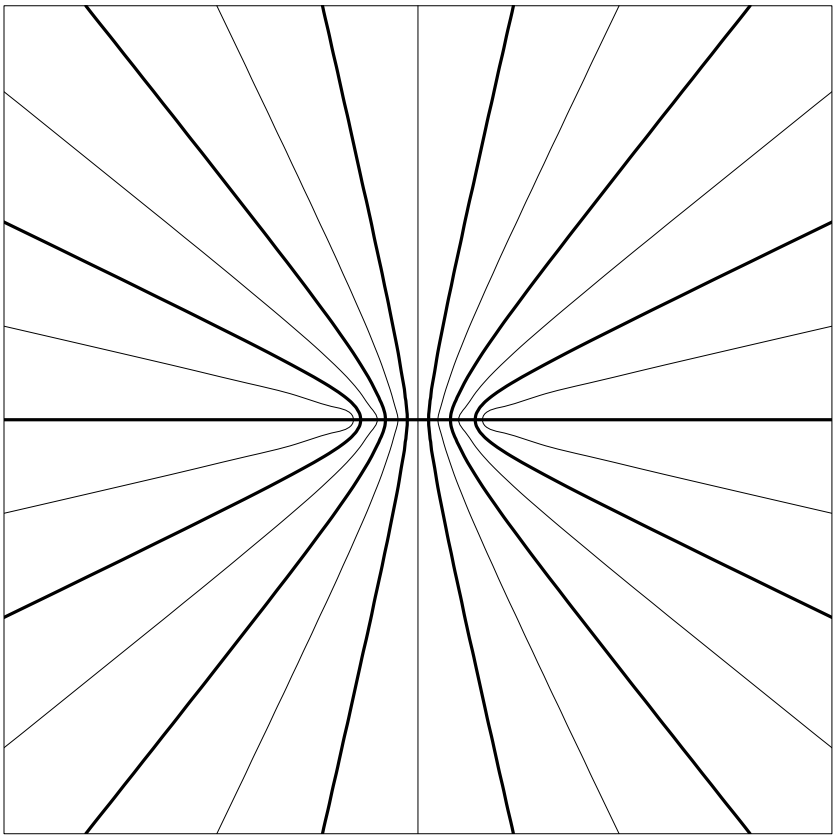


FIGURE 26. Hermite polynomial $H_7(z)$

To show how boring life can be outside Number Theory, we include the graphics of some functions.

The first one is Hermite's polynomial $H_7(z)$, that is,

$$128z^7 - 1344z^5 + 3360z^3 - 1680z$$

It has a degree equal to seven and all its roots are real. Here we represent it in the rectangle $(-17, 17)^2$, which is enough to get a clear idea of how the graphic is.

We can see the seven zeros of the function and the six ceros of the derivative.

The graphics of all Hermite's polynomials are analogous. It also looks like the graphics of other orthogonal polynomial families. But we must point out that a general polynomial can have a very complicated graphics. The regularity in this case is due to the fact that it is a very particular polynomial.

In this page we have the X rays corresponding to the Bessel function $J_7(z)$ and the Airy function $Ai(z)$ defined by

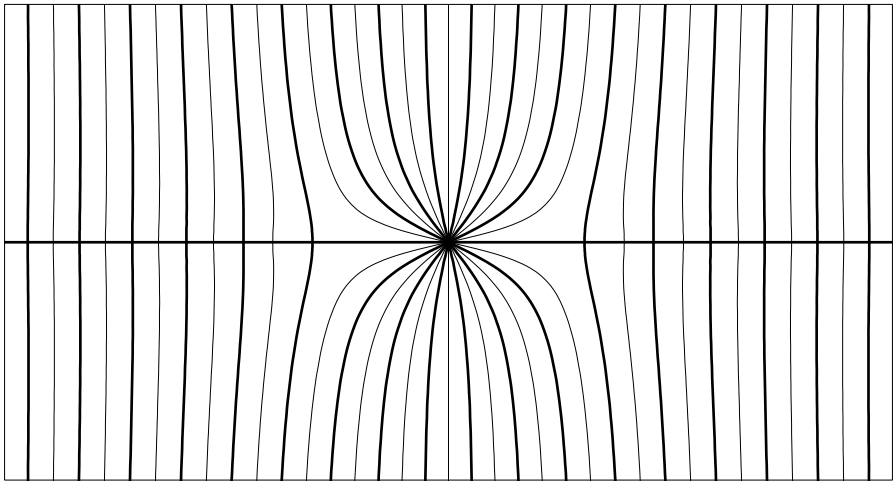


FIGURE 27. Bessel function $J_7(z)$

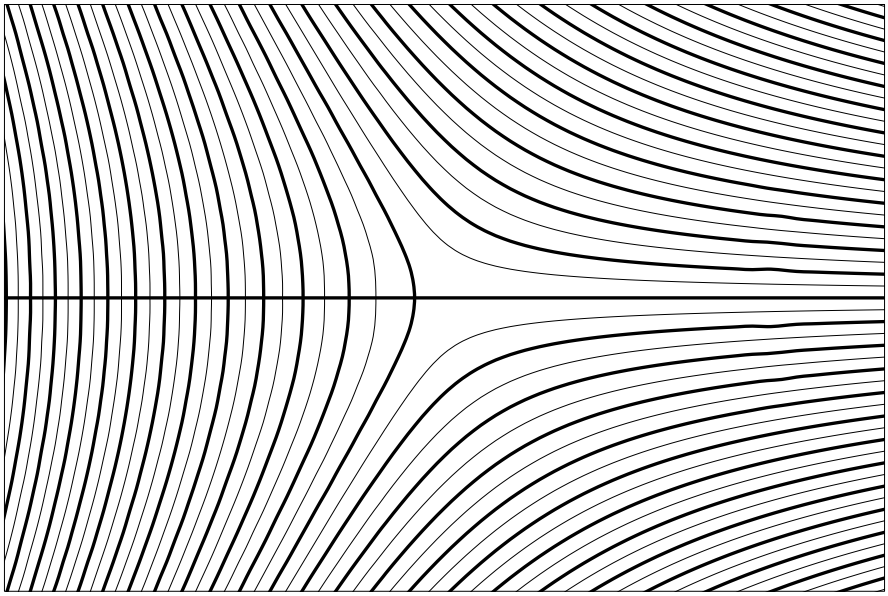


FIGURE 28. Airy function $Ai(z)$

$$J_7(z) = \Big(\frac{z}{2}\Big)^7 \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^k}{k!(k+7)!}, Ai(z) = \frac{3^{-2/3}}{\pi} \sum_{k=0}^{\infty} \frac{\Gamma\big((k+1)/3\big) \sin\frac{2\pi}{3}(k+1)}{k!} (3^{1/3}z)^k$$

The Bessel function is represented on the rectangle $(-28, 28)\times(-20, 20)$ and the Airy function on $(-15, 15) \times (-10, 10)$.

The Bessel function has a zero of order 7 in the origen. Its other zeros are real and, appart from the obvious zeros of the derivative, which are real, the derivative does not vanish.

The Airy function is surprising because of the likeness of its X ray with that of the Gamma function.

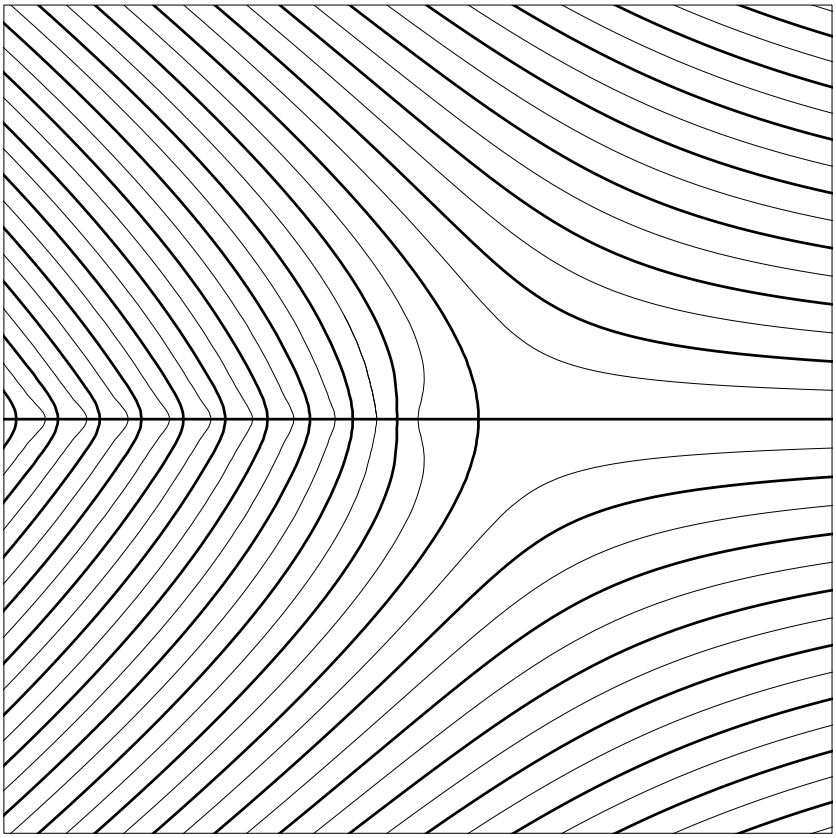


FIGURE 29. Function $\Gamma(s)$

The graphic of the function $\Gamma(s)$ shows that its derivative vanishes only at the obvious zeros. The graphics of a function and its inverse do always coincide. The figure shows the rectangle $(-10, 10)^2$.

REFERENCES

[1] Arias de Reyna, J., *Riemann's fragment on limit values of elliptic modular functions*, Ramanujan J. (to appear).

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

[3] Berndt, B., The nunber of zeros of $\zeta^{(k)}(s)$. J. London Math. Soc. (2) **2** (1970), 577–580.

[4] Brent, R. P., *On the zeros of the Riemann Zeta Function in the Critical Strip*, Math. of Computation **33** (1979), 1361–1372.

[5] Brent, R. P. & van de Lune, J. & te Riele, H. J. J., & Winter, D. T. *On the Zeros of the Riemann Zeta Function in the Critical Strip*, Math. of Computation **39** (1982), 681–688.

[6] Edwards, H. M., *Riemann's Zeta Function*, Academic Press, New York, (1974).
[7] Gram, J. P., *Sur les ZÈros de la Fonction $\zeta(s)$ de Riemann*, Acta Math. **27** (1903), 289–304.
[8] Hutchinson, J. I., *On the roots of the Riemann zeta-function*, Trans. Amer. Math. Soc. **27** (1925), 49–60.
[9] Jahnke, E. & Emde, F. *Tables of Higher Functions*, sixth edition, revised by F. Lˆsch, McGraw-Hill Book Co., New York (1960).
[10] Lehman, R. S., *On the distribution of zeros of the Riemann Zeta-function*, Proc. London Math. Soc. [3] **20** (1970), 303–320.
[11] Lehmer, D. H., *On the roots of the Riemann Zeta-function*, Acta Math. **95** (1956), 291–298.
[12] Levinson, N., & Montgomery, H. L., *Zeros of the Derivatives of the Riemann Zeta-Function*, Acta Math. **133**, (1974), 49–65.
[13] Van de Lune, J., *Some observations concerning the zero-curves of the real and imaginary parts of Riemann's zeta function*, Afdeling Zuivere Wiskunde, **201**, Mathematisch Centrum, Amsterdam (1983), i+25.
[14] B. Riemann, *Ueber die Anzahl der Primzahlen unter einer gegebenen Grˆsse*, Monatsber. Kˆnig. Preuss. Akad. Wiss. Berlin (1859), 671–680; Mathematische Werke, 2nd Auf., Dover, New York, 1953, pp. 145–153.
[15] Rosser, J. B. & Yohe, J. M. & Schoenfeld, L., *Rigorous computation and the zeros of the Riemann zeta-function*, Cong. Proc. Int. Federation Information Process. (1968), 70–76, Spartan, Washington, D. C. and Macmillan, New York, 1969
[16] Siegel, C. L., *Über Riemanns Nachlaß zur analytischen Zahlentheorie* Quellen Studien zur Geschichte der Math. Astron. und Phys. Abt. B: Studien **2**, (1932), 45–80. (Also in *Gesammelte Abhandlungen* Vol. 1, Springer-Verlag, Berlin and New York 1966).
[17] Speiser, A., *Geometrisches zur Riemannschen Zetafunktion*, Math. Ann. **110** (1934), 514–521.
[18] Spira, R., *Zero-free regions of $\zeta^{(k)}(s)$*, J. London Math. Soc. **40** (1965), 677–682.
[19] Titchmarsh, E. C. & Heath-Brown, D. R., *The Theory of the Riemann Zeta-function*, Second Edition, Oxford University Press, Oxford, (1986).
[20] Turing, A. M., *Some Calculations of the Riemann Zeta-function*, Proc. London Math. Soc. [3] **3** (1953), 99–117.
[21] Utzinger, A., *Über die reellen Z¸ge der Riemannschen Zetafunktion*, Zurich, (1934).
[22] H. Weber(editor), *Collected Works of Bernhard Riemann*, Second edition and the supplement, Dover, New York (1953).

This paper was presented on a lecture on Analytic Number Theory on May 29, 2002

Facultad de Matem·ticas, Universidad de Sevilla, P. O. box 1160, 41080-Seville, Spain
E-mail address: arias@us.es