Notes on analytic number theory (updated 28 October 2015)

Kiran S. Kedlaya

Department of Mathematics, University of California, San Diego
E-mail address: kedlaya@ucsd.edu

Contents

Preface

This text is a lightly edited version of the lecture notes of a course on analytic number theory (18.785) that I gave at MIT in the spring of 2017. This course was an introduction to analytic number theory, including the use of zeta functions, L-functions, and sieving methods to prove distribution results concerning prime numbers (e.g., the prime number theorem in arithmetic progressions). The announced prerequisites for the course were undergraduate courses in elementary number theory and complex analysis. The primary references were Davenport, *Multiplicative Number Theory* (for zeta functions and L-functions) and Iwaniec–Kowalski, *Analytic Number Theory* (for sieving methods).

The principal goal of the course was to present the following theorem of Goldston, Pintz, and Yıldırım (2005): if p_n denotes the n-th prime number, then

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

In other words, there are infinitely many pairs of consecutive primes closer together than any fixed multiple of the average spacing predicting by the prime number theorem. Starting in 2013, these ideas were refined by Zhang, Maynard, Tao, the Polymath project, et al. to prove *bounded gaps between primes*:

$$\liminf_{n\to\infty}(p_{n+1} - p_n) \leq C$$

for various explicit values of C; while it is my intention to update the notes to reflect these more recent developments, I have not yet done so. (The value $C = 2$ would yield the twin prime conjecture, but it is not expected that the techniques under discussion here can establish such a strong bound.)

As this document is not yet in a final state, corrections and comments are welcome.

Part 1

First steps

CHAPTER 1

Introduction to the course

Welcome to 18.785! This course is meant to be an introduction to analytic number theory; this handout provides an overview of what we will be talking about in the course. It also fixes some notation that I'll be using throughout.

1. An overview of the course

The fundamental questions in analytic number theory, and the ones which we focus on in this course, concern the interplay between the additive and multiplicative structures on the integers. Specifically, it is quite natural to ask questions of an additive nature about constructions which are intrinsically multiplicative. In rare cases, these questions lead us to interesting algebraic structures; for instance, the fact (due to Fermat) that every prime $p \equiv 1 \pmod 4$ can be written uniquely as the sum of two squares leads to the study of the ring of Gaussian integers, and the fact (due to Lagrange) that every positive integer can be written as the sum of four squares ties in nicely to quaternions. However, most additive questions about multiplicative structures admit insufficiently useful algebraic structure; for instance, one cannot use algebraic techniques alone to determine which primes can be written as the sum of two cubes.

We thus turn instead to techniques from analysis; that is, we apply *continuous* techniques to study *discrete* phenomena. This tends to be most successful when proving *average* statements; for instance, one cannot give an exact formula for the number of primes in an interval $[1, x]$, but we can establish an *asymptotic* formula, and give some upper bounds for the discrepancy between the exact and asymptotic formulas.

Although this methodology turns out to be unexpectedly powerful, we must remain humbled by the fact that it is comically easy to pose open and probably extremely hard questions about prime numbers, including the following old chestnuts.

- (Twin primes problem) Are there infinitely many pairs of consecutive primes which differ by 2?
- (Sophie Germain problem) Are there infinitely many pairs of primes p, q such that $q = 2p - 1$?
- (Goldbach problem) Is every even integer $n > 2$ equal to the sum of two primes?

2. Basic structure of the course

In the first part of the course, our use of analysis will mainly involve the theory of complex functions, specifically the notions of analytic (holomorphic) and meromorphic functions. (One can argue that one is really using properties of *real*

harmonic functions, since the real and imaginary parts of a holomorphic function have that property, and in other situations one gets number-theoretic information by considering harmonic functions in a setting where there is no complex structure. Indeed, there is a lot of research in this direction to back up this point of view, but I am completely unqualified to talk about it!)

In the second part of the course, we will draw on a second set of ideas, related to the notion of *sieving*. I will give an appropriate introduction to that idea in due course; in the interim, you should have in mind the Sieve of Erastothenes as a technique for isolating the primes among all positive integers. You may also keep in mind the target application: the Bombieri-Vinogradov theorem, which gives a quantitative statement to the effect that if one looks at all of the arithmetic progressions of a single modulus which contain any primes at all, then the primes tend to distribute themselves uniformly among these.

In the third part of the course, we will prove a very explicit theorem about the distribution of primes, due to Goldston, Pintz, and Yıldırım (sic) from 2005. It states the following: if p_n denotes the n-th prime, then

$$\inf_n \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

The proof combines the Bombieri-Vinogradov theorem, from the second part of the course, with some estimates on divisor sums using techniques of complex analysis, as in the first part.

If there is time for a fourth part (which I expect there will be), we will consider some classes of "nonabelian" L-functions, and see how to use analyticity properties of these L-functions (which I will not be proving, as they are much deeper than anything I plan to discuss) to prove some equidistribution results in the spirit of Dirichlet's theorem. One class of examples is the Artin L-functions, leading to the Chebotarev density theorem: for L a number field which is Galois over $\mathbb{Q}$ with Galois group G, this theorem predicts (among other things) the density of primes p for which the prime ideal (p) in $\mathbb{Z}$ factors in a given way in the ring of integers of L. A second class of examples is the L-functions associated to elliptic curves, leading to the Sato-Tate conjecture: for E an elliptic curve over $\mathbb{Q}$, this theorem predicts the distribution of the number of points on the reduction of E modulo p, as the prime p varies. (The latter is the subject of a recent breakthrough by Clozel, Harris, and Taylor.)

3. Notations

I want to try to keep my notation consistent throughout the semester. Here are a few conventions I have in mind; I may add more later.

Basics. Throughout this course, $\mathbb{N}$ denotes the set of *positive* integers. Whether $\mathbb{N}$ should include 0 is a matter of some controversy, but in this course it will be more convenient to omit 0. I might write $\mathbb{N}_0$ for the nonnegative integers.

We reserve the letter p for a prime number, and a sum or product over p without further explanation means p runs over all prime numbers. (If a condition is imposed, like $p \equiv 1 \pmod 4$, instead take all primes obeying that condition.)

Asymptotics. Suppose we are interested in limiting behavior of some functions of x as x tends to some limit. (If otherwise unspecified we will mean $x \to \infty$, but it should be clear from context.) We write $f(x) \sim g(x)$ to mean that

$\lim f(x)/g(x) = 1$. We write $O(f(x))$ to denote any function $g(x)$ such that $\limsup g(x)/f(x) < \infty$. We write $o(f(n))$ to denote any function $g(x)$ such that $\limsup g(x)/f(x) = 0$.

Beware that sometimes we talk about limiting behavior in one variable of functions that also depend on other variables. Unless otherwise specified, you should assume the limits are *not* uniform in the other variables. When they are, I will make that more clear.

Miscellaneous. It may happen sometimes during a proof that there are a number of auxiliary constants whose values I don't care about. I may use a single letter (like c) to refer to every such constant; if I do this, I'll make this abundantly clear beforehand.

CHAPTER 2

The prime number theorem

Most of my handouts will come with exercises attached; see the web site for the due dates. (For example, these are due February 14.)

There are likely to be typos in all of my handouts; it would be helpful if you could report these by email (including ones I point out in class).

Thanks to Ben Brubaker for filling in for me; I will be back February 12.

1. Euler's idea: revisiting the infinitude of primes

To begin our story, we turn to Euler's viewpoint on the fact, originally due to Euclid, that there are infinitely many prime numbers. Euclid's original proof was quite simple, and entirely algebraic: assume there are only finitely many primes, multiply them together, add 1, then factor the result.

Euler realized instead that a basic fact from analysis also leads to the infinitude of primes. This fact is the divergence of the harmonic series

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots,$$

which follows for instance from the fact that

$$\sum_{n=1}^{N} \frac{1}{n} \geq \sum_{n=1}^{N} \frac{1}{2^{\lceil \log_2 n \rceil}} \geq \frac{1}{2} \lfloor \log_2 N \rfloor$$

and the right side tends to ∞ as $N \to \infty$. (We will usually want a more precise estimate; see the exercises.) On the other hand, if there were only finitely many primes, then unique factorization of positive integers into prime powers would imply that

$$\sum_{n=1}^{\infty} \frac{1}{n} = \prod_{p} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right) = \prod_{p} \left(1 - \frac{1}{p}\right)^{-1},$$

which would give the equality between a divergent series and a finite quantity. Contradiction.

Euler's idea turns out to be quite fruitful: the introduction of analysis into the study of prime numbers allows us to prove distribution statements about primes in a much more flexible fashion than is allowed by algebraic techniques. For instance, we will see in an upcoming unit how Dirichlet adapted this idea to prove that every arithmetic progression whose terms do not all share a common factor contains infinitely many primes.

2. Riemann's zeta function

For the moment, however, let us turn to Riemann's one paper in number theory, in which he fleshes out Euler's idea and fits it into the theory of complex functions

of one variable. He considered the series

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

for all $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$. Note that for $\mathrm{Re}(s) > 1$, the series is absolutely convergent; moreover, it converges uniformly in any region of the form $\mathrm{Re}(s) \geq 1+\epsilon$ for $\epsilon > 0$. Consequently, it gives rise to an analytic function in the half-plane $\mathrm{Re}(s) > 1$. The boundary $\mathrm{Re}(s) = 1$ is sometimes called the *critical line*.

In the domain of absolute convergence, we can also write

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

and this product converges absolutely and uniformly for $\mathrm{Re}(s) \geq 1 + \epsilon$ for $\epsilon > 0$. (Reminder: a product $\prod_i (1+a_i)$ converges absolutely if and only if $\sum_i a_i$ converges absolutely.) It follows that $\zeta(s) \neq 0$ for $\mathrm{Re}(s) > 1$.

For future reference, we note that the product representation is sometimes more useful in the form

$$\begin{aligned} \log \zeta(s) &= \sum_p -\log(1 - p^{-s}) \\ &= \sum_p \sum_{n=1}^{\infty} \frac{p^{-ns}}{n}. \end{aligned}$$

We now show that ζ extends somewhat beyond the domain of absolute convergence of the original series.

THEOREM 2.1. *The function $f(s) = \zeta(s) - \frac{s}{s-1}$ on the domain $\mathrm{Re}(s) > 1$ extends (uniquely) to a holomorphic function on the domain $\mathrm{Re}(s) > 0$. Consequently, $\zeta(s)$ is meromorphic on $\mathrm{Re}(s) > 0$, with a simple pole at $s = 1$ of residue 1 and no other poles.*

PROOF. This is an easy application of one of the basic tools in this subject, Abel's method of *partial summation* (or *summation by parts*, as in integration by parts). Namely,

$$\sum_{n=1}^{N} a_n b_n = a_{N+1} B_N - \sum_{n=1}^{N} (a_{n+1} - a_n) B_n, \qquad B_n = \sum_{i=1}^{n} b_i.$$

We apply partial summation to $\zeta(s)$ by taking $a_n = n^{-s}$ and $b_n = 1$, so that $B_n = n$. Rather, we apply partial summation to the truncated sum $\sum_{n=1}^{N} n^{-s}$, and note that the error term $a_{N+1} B_N = (N+1)^{-s} N$ tends to 0 for $\mathrm{Re}(s) > 1$. (Warning: in general, I am not going to be nearly so verbose when applying partial summation. So make sure you understand this example!)

With that said, we have

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

We can thus write

$$f(s) = -s\int_{i=1}^{\infty} \{x\} x^{-s-1}\,dx,$$

for $\{x\}$ the fractional part of x; the integral converges absolutely for $\mathrm{Re}(s) > 0$, and uniformly for $\mathrm{Re}(s) \geq \epsilon$ for any $\epsilon > 0$. This proves the claim. □

We already know that $\zeta(s)$ cannot vanish for $\mathrm{Re}(s) > 1$; to prove the prime number theorem, we need to also exclude zeroes on the boundary of that half-plane.

THEOREM 2.2 (Hadamard, de la Vallée-Poussin). *The function $\zeta(s)$ has no zero on the line* $\mathrm{Re}(s) = 1$.

PROOF (MERTENS). See exercises. □

We will return to Riemann's memoir, establishing more detailed properties of ζ, in a subsequent unit.

3. Towards the prime number theorem

Using the aforementioned properties of the zeta function, Hadamard and de la Vallée-Poussin independently established the prime number theorem in 1897. We'll follow here an argument due to D.J. Newman; our presentation is liberally plagiarized from D. Zagier, Newman's short proof of the Prime Number Theorem, *American Mathematical Monthly* **104** (1997), 705–708.

For $x \in \mathbb{R}$, write

$$\begin{aligned}\pi(x) &= \sum_{p \leq x} 1 \\ \vartheta(x) &= \sum_{p \leq x} \log p.\end{aligned}$$

The prime number theorem then asserts that

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

This is equivalent to

$$\vartheta(x) \sim x,$$

because for any $\epsilon > 0$,

$$\begin{aligned}\vartheta(x) &\leq \sum_{p \leq x} \log x = \pi(x) \log x \\ \vartheta(x) &\geq \sum_{x^{1-\epsilon} \leq p \leq x} \log x^{1-\epsilon} = (1-\epsilon)(\pi(x) + O(x^{1-\epsilon})) \log x.\end{aligned}$$

What we will prove is that the improper integral

$$\int_1^\infty \frac{\vartheta(x) - x}{x^2}\, dx \tag{1}$$

converges; remember that this means that for every $\epsilon > 0$, there exists N such that for $y, z \geq N$,

$$\left| \int_y^z \frac{\vartheta(x) - x}{x^2}\, dx \right| < \epsilon.$$

(It is much easier to prove that these integrals are bounded; see exercises.) To then deduce $\vartheta(x) \sim x$, suppose that there exists $\lambda > 1$ such that $\vartheta(x) \geq \lambda x$ for arbitrarily large x. Since ϑ is nondecreasing, it then follows that for any such x,

$$\int_x^{\lambda x} \frac{\vartheta(t) - t}{t^2} dt \geq \int_x^{\lambda x} \frac{\lambda x - t}{t^2} dt = \int_1^\lambda \frac{\lambda - t}{t^2} dt > 0,$$

contradiction. Likewise, if there exists $\lambda < 1$ such that $\vartheta(x) \leq \lambda x$ for arbitrarily large x, then such x satisfy

$$\int_{\lambda x}^x \frac{\vartheta(t) - t}{t^2} dt \leq \int_{\lambda x}^x \frac{\lambda x - t}{t^2} dt = \int_\lambda^1 \frac{\lambda - t}{t^2} dt < 0,$$

contradiction.

4. The Tauberian argument

We have thus reduced the prime number theorem to the convergence of the integral (1); we turn to this next. Consider the function $\Phi(s) = -\zeta'(s)/\zeta(s)$; from the log-product representation for ζ, using partial summation as in Theorem 2.1, and substituting $x = e^t$, we find

$$\begin{aligned}
\Phi(s) &= \sum_p (\log p) p^{-s} + \sum_p \sum_{n=2}^\infty (\log p) p^{-ns} \\
&= s \int_1^\infty \vartheta(x) x^{-s-1}\, dx + s \int_1^\infty \vartheta(x) \left(\sum_{n=2}^\infty n x^{-ns-1} \right) dx \\
&= s \int_0^\infty e^{-st} \vartheta(e^t)\, dt + s \int_0^\infty \frac{2e^{-2st} - e^{-3st}}{(1 - e^{-st})^2} \vartheta(e^t)\, dt
\end{aligned}$$

Define the functions

$$\begin{aligned}
f(t) &= \vartheta(e^t) e^{-t} - 1 \\
g(z) &= \frac{\Phi(z+1)}{z+1} - \frac{1}{z};
\end{aligned}$$

by the above,

$$g(z) = \int_0^\infty f(t) e^{-zt}\, dt + \int_0^\infty \frac{2e^{-2(z+1)t} - e^{-3(z+1)t}}{(1 - e^{-(z+1)t})^2} \vartheta(e^t)\, dt.$$

Right now, we know that the integral defining $g(z)$ makes sense for $\mathrm{Re}(z) > 0$, but we will deduce (1) (after substituting $x = e^t$) and hence the prime number theorem if we can obtain convergence of $g(z)$ in the case $z = 0$. (Note that the second term converges absolutely for $z = 0$, so we only have to worry about the first term.)

The idea is to do this by leveraging complex function-theoretic information about Φ; this sort of operation is known as a *Tauberian argument*. To be precise,

by what we know about ζ, $\Phi(s)$ is meromorphic on $\mathrm{Re}(s) > 0$, with a simple pole at $s = 1$ of residue 1 and no other poles in $\mathrm{Re}(s) \geq 1$. It follows that f and g satisfy the conditions of the following theorem.

THEOREM 2.3 (Newman). *Let $f : [0, +\infty) \to \mathbb{R}$ be a bounded, locally integrable function, and define $g(z) = \int_0^\infty f(t)e^{-zt}\,dt$; note that this integral converges absolutely uniformly for $\mathrm{Re}(z) \geq \epsilon$ for any $\epsilon > 0$. Suppose that $g(z)$ extends to a holomorphic function on a neighborhood of $\mathrm{Re}(z) \geq 0$. Then $\int_0^\infty f(t)\,dt$ exists and equals $g(0)$.*

PROOF (ZAGIER, AFTER NEWMAN). For $T > 0$, put $g_T(z) = \int_0^T f(t)e^{-zt}\,dt$; each function g_T is entire, and we want $\lim_{T\to\infty} g_T(0) = g(0)$.

For R large (but fixed until further notice), let C be the boundary of the region

$$\{z \in \mathbb{C} : |z| \leq R, \quad \mathrm{Re}(z) \geq -\delta\}$$

for some $\delta = \delta(R) > 0$ chosen small enough that C lies inside the domain on which g is holomorphic. By the Cauchy integral theorem,

$$g(0) - g_T(0) = \frac{1}{2\pi i}\int_C (g(z) - g_T(z))e^{zT}\left(1 + \frac{z^2}{R^2}\right)\frac{dz}{z}; \tag{2}$$

namely, the only pole of the integrand is a simple pole at $z = 0$, so we simply pop out the residue there.

To bound the right side of (2), we separate the contour of integration C into

$$C_+ = C \cap \{z \in \mathbb{C} : \mathrm{Re}(z) \geq 0\}$$
$$C_- = C \cap \{z \in \mathbb{C} : \mathrm{Re}(z) \leq 0\}.$$

Remember that we assumed f is bounded; choose $B > 0$ so that $|f(t)| \leq B$ for all t. For $\mathrm{Re}(z) > 0$ with $|z| = R$, we have

$$\begin{aligned} |g(z) - g_T(z)| &= \left|\int_T^\infty f(t)e^{-zt}\,dt\right| \\ &\leq B\int_T^\infty |e^{-zt}|\,dt \\ &= \frac{Be^{-\mathrm{Re}(z)T}}{\mathrm{Re}(z)} \end{aligned}$$

and

$$\left|e^{zT}\left(1 + \frac{z^2}{R^2}\right)\frac{1}{z}\right| = e^{\mathrm{Re}(z)T}\frac{2\,\mathrm{Re}(z)}{R^2}.$$

Since the length of the contour is at most $2\pi R$, the contribution over C_+ to (2) is bounded in absolute value by

$$\frac{1}{2\pi}(2\pi R)\frac{Be^{-\mathrm{Re}(z)T}}{\mathrm{Re}(z)}e^{\mathrm{Re}(z)T}\frac{2\,\mathrm{Re}(z)}{R^2} = \frac{2B}{R}.$$

Over C_-, we separate the integral into integrals involving g and g_T. Since g_T is entire, its integral over C_- can instead be calculated over the semicircle

$C'_- = \{z \in \mathbb{C} : |z| = R, \mathrm{Re}(z) \leq 0\}$. Since for $\mathrm{Re}(z) < 0$ we have

$$\begin{aligned} |g_T(z)| &= \left| \int_0^T f(t) e^{-zt}\, dt \right| \\ &\leq B \int_{-\infty}^T |e^{-zt}|\, dt \\ &= \frac{B e^{-\mathrm{Re}(z)T}}{|\mathrm{Re}(z)|}, \end{aligned}$$

as above we bound this contribution to (2) by $2B/R$.

Finally, we consider the contribution to (2) from g over C_-; we are going to show that this contribution tends to 0 as $T \to \infty$. By parametrizing the contour, we can write

$$\frac{1}{2\pi i} \int_{C_-} g(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} = \int_0^1 a(u) e^{b(u)T}\, du,$$

where $a(u)$ and $b(u)$ are continuous, and $\mathrm{Re}(b(u)) < 0$ for $0 < u < 1$; the key point is that a does not depend on T, so as $T \to \infty$ the integrand tends to 0 pointwise except at the endpoints. Since the integrands are all bounded, Lebesgue's dominated convergence theorem implies that the integral tends to 0 as $T \to \infty$. (Again, I'm being more explicit with the analysis than I will be in general.)

We conclude that

$$\limsup_{T \to \infty} |g(0) - g_T(0)| \leq \frac{4B}{R};$$

since R can be chosen arbitrarily large, this yields the desired result. □

You might be thinking at this point that if one knew g extended to a holomorphic function on a region a bit larger than $\mathrm{Re}(s) \geq 0$, then maybe one could prove something about the rate of convergence of the integral $\int_0^\infty f(t)\, dt$. In particular, if one can exclude zeroes of ζ in some region beyond the line $\mathrm{Re}(s) = 1$, one should correspondingly get a prime number theorem with an improved error term. We will see that this is correct in a subsequent unit, at least if we replace the approximation $\pi(x) \sim x/(\log x)$ with Gauss's approximation $\pi(x) \sim \mathrm{li}(x)$ (see exercises).

Historical aside: the Erdős-Selberg method

About 40 years after the original proof, Erdős and Selberg gave so-called elementary proofs of the prime number theorem, which do not use any complex analysis. The key step in Selberg's proof is to give an elementary proof of the bound

$$|R(x)| \leq \frac{1}{\log x} \int_1^x |R(x/t)|\, dt + O\left(x \frac{\log\log x}{\log x}\right), \tag{3}$$

where $R(x) = \vartheta(x) - x$; I will probably say something about this result in the section on sieving.

Using (3) and the fact that

$$\int_1^x \frac{R(t)}{t^2} dt = O(1) \tag{4}$$

(much easier than the convergence of the integral; see exercises), one then produces $0 < c < 1$ such that if there exists $\alpha > 0$ such that $|R(x)| < \alpha x$ for x large, then

also $|R(x)| < \alpha cx$ for x large. I find this step somewhat unenlightening; if you must know the details, see A. Selberg, An elementary proof of the prime number theorem, *Annals of Math.* **50** (1949), 305–313. Or see Chapter XXII of Hardy-Wright, or Nathanson's *Elementary Methods in Number Theory.*

Exercises

(1) Prove that there exists a positive constant γ such that

$$\sum_{i=1}^{n} \frac{1}{i} - \log n = \gamma + O(n^{-1}),$$

by comparing the sum to a Riemann sum for $\int_1^n \frac{1}{x}\,dx$. The number γ is called *Euler's constant*, and it is one of the most basic constants in analytic number theory. However, since it is defined purely analytically, we remain astonishingly ignorant about it; for instance, γ is most likely irrational (even transcendental) but no proof is known.

(2) Let $d(n)$ denote the number of divisors of $n \in \mathbb{N}$. Prove that

$$\sum_{i=1}^{n} d(i) = n \log n + (2\gamma - 1)n + O(n^{1/2}),$$

by estimating the number of lattice points in the first quadrant under the curve $xy = n$.

(3) (Mertens) Fix $t \in \mathbb{R}$ nonzero. Prove that the function

$$Z(s) = \zeta(s)^3 \zeta(s+it)^4 \zeta(s+2it)$$

extends to a meromorphic function on $\mathrm{Re}(s) > 0$. Then show that if $s \in \mathbb{R}$ and $s > 1$, then $\log |Z(s)| = \mathrm{Re}(\log Z(s))$ can be written as a series of nonnegative terms, so $|Z(s)| \geq 1$.

(4) Use the previous exercise to prove that $\zeta(s)$ has no zeroes on the line $\mathrm{Re}(s) = 1$.

(5) (Chebyshev) Prove that

$$\prod_{n < p \leq 2n} p \leq 2^{2n}$$

by considering the central binomial coefficient $\binom{2n}{n}$. Then deduce that $\vartheta(x) = O(x)$.

(6) Let k be a positive integer. Prove that for any $c > 0$, if we write C_R for the straight contour from $c - iR$ to $c + iR$, then

$$\lim_{R \to \infty} \frac{1}{2\pi i} \int_{C_R} \frac{x^s\,ds}{s(s+1)\cdots(s+k)} = \begin{cases} \frac{1}{k!}\left(1 - \frac{1}{x}\right)^k & x \geq 1 \\ 0 & 0 \leq x \leq 1. \end{cases}$$

(Hint: use a contour-shifting argument.)

(7) (Gauss) Define the *logarithmic integral function*

$$\mathrm{li}(x) = \int_2^x \frac{dt}{\log t}.$$

(Warning: there is some disagreement in the literature about what lower limit of integration to use.) Prove that $\mathrm{li}(x) \sim x/(\log x)$, so that the prime number theorem is equivalent to $\pi(x) \sim \mathrm{li}(x)$. In fact, Gauss noticed

empirically, and we will prove later, that $\mathrm{li}(x)$ gives a somewhat better approximation to $\pi(x)$ than $x/(\log x)$.

(8) Using the identity

$$\sum_{n\leq x} \log n = \sum_{i=1}^{\infty} \sum_{p:p^i\leq x} \left\lfloor \frac{x}{p^i} \right\rfloor \log p,$$

prove that

$$\sum_{p\leq x} \frac{\log p}{p} = \log x + O(1),$$

then deduce (4) by partial summation.

Part 2

Zeta functions and L-functions

CHAPTER 3

Dirichlet series and arithmetic functions

1. Dirichlet series

The Riemann zeta function ζ is a special example of a type of series we will be considering often in this course. A *Dirichlet series* is a formal series of the form $\sum_{n=1}^{\infty} a_n n^{-s}$ with $a_n \in \mathbb{C}$. You should think of these as a number-theoretic analogue of formal power series; indeed, our first order of business is to understand when such a series converges absolutely.

LEMMA 3.1. *There is an extended real number $L \in \mathbb{R} \cup \{\pm\infty\}$ with the following property: the Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ converges absolutely for $\mathrm{Re}(s) > L$, but not for $\mathrm{Re}(s) < L$. Moreover, for any $\epsilon > 0$, the convergence is uniform on $\mathrm{Re}(s) \geq L + \epsilon$, so the series represents a holomorphic function on all of $\mathrm{Re}(s) > L$.*

PROOF. Exercise. □

The quantity L is called the *abscissa of absolute convergence* of the Dirichlet series; it is an analogue of the radius of convergence of a power series. (In fact, if you fix a prime p, and only allow a_n to be nonzero when n is a power of p, then you get an ordinary power series in p^{-s}. So in some sense, Dirichlet series are a strict generalization of ordinary power series.)

Recall that an ordinary power series in a complex variable must have a singularity at the boundary of its radius of convergence. For Dirichlet series with *nonnegative real* coefficients, we have the following analogous fact.

THEOREM 3.2 (Landau). *Let $f(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ be a Dirichlet series with* nonnegative real *coefficients. Suppose $L \in \mathbb{R}$ is the abscissa of absolute convergence for $f(s)$. Then f cannot be extended to a holomorphic function on a neighborhood of $s = L$.*

PROOF. Suppose on the contrary that f extends to a holomorphic function on the disc $|s - L| < \epsilon$. Pick a real number $c \in (L, L + \epsilon/2)$, and write

$$\begin{aligned} f(s) &= \sum_{n=1}^{\infty} a_n n^{-c} n^{c-s} \\ &= \sum_{n=1}^{\infty} a_n n^{-c} \exp((c-s)\log n) \\ &= \sum_{n=1}^{\infty} \sum_{i=0}^{\infty} \frac{a_n n^{-c} (\log n)^i}{i!} (c-s)^i. \end{aligned}$$

Since all coefficients in this double series are nonnegative, everything must converge absolutely in the disc $|s - c| < \epsilon/2$. In particular, when viewed as a power series in

$c-s$, this must give the Taylor series for f around $s=c$. Since f is holomorphic in the disc $|s-c|<\epsilon/2$, the Taylor series converges there; in particular, it converges for some real number $L'<L$. But now we can run the argument backwards to deduce that the original Dirichlet series converges absolutely for $s=L'$, which implies that the abscissa of absolute convergence is at most L'. This contradicts the definition of L. □

2. Euler products

Remember that among Dirichlet series, the Riemann zeta function had the unusual property that one could factor the Dirichlet series as a product over primes:

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

In fact, a number of natural Dirichlet series admit such factorizations; they are the ones corresponding to multiplicative functions.

We define an *arithmetic function* to simply be a function $f:\mathbb{N}\to\mathbb{C}$. Besides the obvious operations of addition and multiplication, another useful operation on arithmetic functions is the *(Dirichlet) convolution* $f*g$, defined by

$$(f*g)(n)=\sum_{d|n} f(d)g(n/d).$$

Just as one can think of formal power series as the generating functions for ordinary sequences, we may think of a formal Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ as the "arithmetic generating function" for the multiplicative function $n\mapsto a_n$. In this way of thinking, convolution of multiplicative functions corresponds to ordinary multiplication of Dirichlet series:

$$\sum_{n=1}^{\infty}(f*g)(n)n^{-s}=\left(\sum_{n=1}^{\infty} f(n)n^{-s}\right)\left(\sum_{n=1}^{\infty} g(n)n^{-s}\right).$$

In particular, convolution is a commutative and associative operation, under which the arithmetic functions taking the value 1 at $n=1$ form a group. The arithmetic functions taking all integer values (with the value 1 at $n=1$) form a subgroup (see exercises).

We say f is a *multiplicative function* if $f(1)=1$, and $f(mn)=f(m)f(n)$ whenever $m,n\in\mathbb{N}$ are coprime. Note that an arithmetic function f is multiplicative if and only if its Dirichlet series factors as a product (called an *Euler product*):

$$\sum_{n=1}^{\infty} f(n)n^{-s}=\prod_{p}\left(\sum_{i=0}^{\infty} f(p^i)p^{-is}\right).$$

In particular, the property of being multiplicative is clearly stable under convolution, and under taking the convolution inverse.

We say f is *completely multiplicative* if $f(1)=1$, and $f(mn)=f(m)f(n)$ for any $m,n\in\mathbb{N}$. Note that an arithmetic function f is multiplicative if and only if its Dirichlet series factors in a very special way:

$$\sum_{n=1}^{\infty} f(n)n^{-s}=\prod_{p}(1-f(p)p^{-s})^{-1}.$$

In particular, the property of being completely multiplicative is *not* stable under convolution.

3. Examples of multiplicative functions

Here are some examples of multiplicative functions, some of which you may already be familiar with. All assertions in this section are left as exercises.

- The unit function ε: $\varepsilon(1) = 1$ and $\varepsilon(n) = 0$ for $n > 1$. This is the identity under $*$.
- The constant function 1: $1(n) = 1$.
- The *Möbius function* μ: if n is squarefree with d distinct prime factors, then $\mu(n) = (-1)^d$, otherwise $\mu(n) = 0$. This is the inverse of 1 under $*$.
- The identity function id: $\mathrm{id}(n) = n$.
- The k-th power function id^k: $\mathrm{id}^k(n) = n^k$.
- The *Euler totient function* ϕ: $\phi(n)$ counts the number of integers in $\{1, \dots, n\}$ coprime to n. Note that $1 * \phi = \mathrm{id}$, so $\mathrm{id} * \mu = \phi$.
- The divisor function d (or τ): $d(n)$ counts the number of integers in $\{1, \dots, n\}$ dividing n. Note that $1 * 1 = d$.
- The divisor sum function σ: $\sigma(n)$ is the sum of the divisors of n. Note that $1 * \mathrm{id} = d * \phi = \sigma$.
- The divisor power sum functions σ_k: $\sigma_k(n) = \sum_{d|n} d^k$. Note that $\sigma_0 = d$ and $\sigma_1 = \sigma$. Also note that $1 * \mathrm{id}^k = \sigma_k$.

Of these, only $\varepsilon, 1, \mathrm{id}, \mathrm{id}^k$ are completely multiplicative. We will deal with some more completely multiplicative functions, the Dirichlet characters, in a subsequent unit.

Note that all of the Dirichlet series corresponding to the aforementioned functions can be written explicitly in terms of the Riemann zeta function; see exercises. An important non-multiplicative function with the same property is the *von Mangoldt function* $\Lambda = \mu * \log$; see exercises.

Exercises

(1) Prove Lemma 3.1. Then exhibit examples to show that a Dirichlet series with some abscissa of absolute convergence $L \in \mathbb{R}$ may or may not converge absolutely on $\mathrm{Re}(s) = L$.
(2) Give a counterexample to Theorem 3.2 in case the series need not have nonnegative real coefficients. (Optional, and I don't know the answer: must a Dirichlet series have a singularity *somewhere* on the abscissa of absolute convergence?)
(3) Let $f : \mathbb{N} \to \mathbb{Z}$ be an arithmetic function with $f(1) = 1$. Prove that the convolution inverse of f also has values in $\mathbb{Z}$; deduce that the set of such f forms a group under convolution. (Likewise with $\mathbb{Z}$ replaced by any subring of $\mathbb{C}$, e.g., the integers in an algebraic number field.)
(4) Prove the assertions involving $*$ in Section 3. Then use them to write the Dirichlet series for all of the functions introduced there in terms of the Riemann zeta function.
(5) Here is a non-obvious example of a multiplicative function. Let $r_2(n)$ be the number of pairs (a, b) of integers such that $a^2 + b^2 = n$. Prove that

$r_2(n)/4$ is multiplicative, using facts you know from elementary number theory.

(6) We defined the *von Mangoldt function* as the arithmetic function $\Lambda = \mu * \log$. Prove that

$$\Lambda(n) = \begin{cases} \log(p) & n = p^i, i \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

and that the Dirichlet series for Λ is $-\zeta'/\zeta$.

(7) For t a fixed positive real number, verify that the function

$$Z(s) = \zeta^2(s)\zeta(s+it)\zeta(s-it)$$

is represented by a Dirichlet series with nonnegative coefficients which does not converge everywhere. (Hint: check $s = 0$.)

(8) Assuming that $\zeta(s) - s/(s-1)$ extends to an entire function (we'll prove this in a subsequent unit), use the previous exercise to give a second proof that $\zeta(s)$ has no zeroes on the line $\mathrm{Re}(s) = 1$.

(9) (Dirichlet's hyperbola method) Suppose f, g, h are arithmetic functions with $f = g * h$, and write

$$G(x) = \sum_{n \leq x} g(n), \qquad H(x) = \sum_{n \leq x} h(n).$$

Prove that (generalizing a previous exercise)

$$\sum_{n \leq x} f(n) = \left(\sum_{d \leq y} g(d) H(x/d) \right) + \left(\sum_{d \leq x/y} h(d) G(x/d) \right) - G(y) H(x/y).$$

(10) Prove that the abscissa of absolute convergence L of a Dirichlet series $\sum_{n=1}^{\infty} a_n n^{-s}$ satisfies the inequality

$$L \leq \limsup_{n \to \infty} \left(1 + \frac{\log |a_n|}{\log n} \right)$$

(where $\log 0 = -\infty$), with equality if the $|a_n|$ are bounded away from 0. Then exhibit an example where the inequality is strict. (Thanks to Sawyer for pointing this out.) Optional (I don't know the answer): is there a formula that computes the abscissa of absolute convergence in general? Dani proposed

$$\limsup_{n \to \infty} \frac{\log \sum_{m \leq n} |a_m|}{\log n}$$

but Sawyer found a counterexample to this too.

CHAPTER 4

Dirichlet characters and L-functions

In this unit, we introduce some special multiplicative functions, the Dirichlet characters, and study their corresponding Dirichlet series. We will use these in a subsequent unit to prove Dirichlet's theorem on primes in arithmetic progressions, and the prime number theorem in arithmetic progressions.

1. Dirichlet characters

For N a positive integer, a *Dirichlet character of level* N is an arithmetic function χ which factors through a homomorphism $(\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}$ on integers $n \in \mathbb{N}$ coprime to N, and is zero on integers not coprime to N; such a function is completely multiplicative. Note that the nonzero values must all be N-th roots of unity, and that the characters of level N form a group under termwise multiplication.

For each level N, there is a Dirichlet character taking the value 1 at all n coprime to N; it is called the *principal* (or *trivial*) *character of level* N. A non-principal Dirichlet character of level N is given by the Legendre-Jacobi symbol

$$\chi(n) = \left(\frac{n}{N}\right).$$

LEMMA 4.1. *If χ is nonprincipal of level N, then*

$$\chi(1) + \cdots + \chi(N) = 0.$$

PROOF. The sum is invariant under multiplication by $\chi(m)$ for any $m \in \mathbb{N}$ coprime to N, but if χ is nonprincipal, then we can choose m with $\chi(m) \neq 1$. □

Sometimes a Dirichlet character of level N can be written as the termwise product of the principal character of level N with a character of some level $N' < N$ (of course N' must divide N). We say the character is *imprimitive* in this case and *primitive* otherwise.

2. L-series

The Dirichlet series associated to a Dirichlet character χ of level N is called a *Dirichlet L-series* (or *Dirichlet L-function*) *of level* N, denoted $L(s, \chi)$. (It may also be denoted $L_\chi(s)$, so that one can refer to L_χ as a function without explicitly naming the variable.) Since χ is completely multiplicative, $L(s, \chi)$ formally factors as

$$\prod_p (1 - \chi(p)p^{-s})^{-1}. \tag{5}$$

In particular, if χ is imprimitive corresponding to the character χ' of level N', then

$$L(s, \chi) = L(s, \chi') \prod_{p|N, p\nmid N'} (1 - \chi'(p)p^{-s}). \tag{6}$$

(Note that (6) reduces to $L(s,\chi) = L(s,\chi')$ if N and N' have the same prime factors, e.g., if N' is prime and $N = (N')^2$.) In particular, the abscissa of absolute convergence of the principal character of level N, and hence of each of the characters of level N, is 1, and the product representation (5) is valid for $\mathrm{Re}(s) > 1$. In particular, $L(s,\chi) \neq 0$ for $\mathrm{Re}(s) > 1$.

THEOREM 4.2. *Let χ be a Dirichlet character of level N. Then $L(s,\chi)$ extends to a meromorphic function on $\mathrm{Re}(s) > 0$ with no poles away from $s = 1$. If χ is principal, then $L(s,\chi)$ has a simple pole at $s = 1$ of residue $\prod_{p|N}(1 - p^{-1})$; otherwise, $L(s,\chi)$ is holomorphic also at $s = 1$.*

PROOF. If χ is principal, then by (6),

$$L(s,\chi) = \zeta(s) \prod_{p|N} (1 - p^{-s}),$$

and the claims about $L(s,\chi)$ follow from what we already know about ζ. So assume hereafter that χ is nonprincipal. By partial summation, we can write

$$L(s,\chi) = \sum_{n=1}^{\infty} (\chi(1) + \cdots + \chi(n))(n^{-s} - (n+1)^{-s}). \tag{7}$$

Since $\chi(1) + \cdots + \chi(N) = 0$ by Lemma 4.1, the quantities $\chi(1) + \cdots + \chi(n)$ are bounded for all n. Meanwhile,

$$\begin{aligned} n^{-s} - (n+1)^{-s} &= n^{-s}(1 - (1 + 1/n)^{-s}) \\ &= sn^{-s-1} + O(n^{-s-2}), \end{aligned}$$

where the implied constant in the big O can be taken uniform over s in a compact set. Consequently, the sum representation for $L(s,\chi)$ given by (7) converges uniformly for $\mathrm{Re}(s) \geq \epsilon$ for any $\epsilon > 0$. This yields the claim. □

3. Nonvanishing of L-functions on $\mathrm{Re}(s) = 1$

Much as we used nonvanishing of ζ on the line $\mathrm{Re}(s) = 1$ to study the prime number theorem, we will use nonvanishing of L-functions on that line to study the prime number theorem in arithmetic progressions. An additional wrinkle, though, is that we have to do some extra work to understand what is going on at $s = 1$ itself; see next section.

LEMMA 4.3. *Let $f(s)$ be a meromorphic function on a neighborhood of $\mathrm{Re}(s) \geq L$, with at worst a simple pole at $s = L$ and no other poles. Suppose that $\log f(s)$ is represented by a Dirichlet series with abscissa of convergence $\leq L$ and* nonnegative real *coefficients. Then $f(s) \neq 0$ for $\mathrm{Re}(s) \geq L$.*

PROOF. See exercises. □

THEOREM 4.4. *Let N be a positive integer. Let $f_N(s)$ be the product of all of the Dirichlet L-series of level N. Then $f_N(s) \neq 0$ for $s \in \mathbb{C}$ with $\mathrm{Re}(s) = 1$.*

PROOF. Note that for $\mathrm{Re}(s) > 1$, we have

$$\log f_N(s) = \sum_{p:(p,N)=1} \sum_{n=1}^{\infty} \left(\sum_{\chi} \chi(p^n) \right) p^{-ns}, \tag{8}$$

which is a Dirichlet series with nonnegative real coefficients. (The sum over χ is invariant under multiplication by $\chi(p^n)$ for any single χ, so either the sum is zero or all of the summands are equal to 1.) We may thus apply Lemma 4.3. □

This tells us a lot about nonvanishing of individual L-functions, but not quite everything.

THEOREM 4.5. *For any Dirichlet character χ, $L(s, \chi) \neq 0$ when $\mathrm{Re}(s) = 1$ and $s \neq 1$.*

PROOF. Let N be the level of χ. Then $f_N(s)$ is a product of functions, one of which is $L(s, \chi)$, all of which are holomorphic at s. By Theorem 4.4, $f_N(s)$ has no zero at s, so none of the factors can either. □

It will take a bit more work to deal with $s = 1$; see next section.

4. Nonvanishing for L-functions at $s = 1$

At $s = 1$ (the so-called *critical point* for Dirichlet L-functions), life is a bit more complicated; to deduce that none of the $L(1, \chi)$ vanish, I would need to know that the function $f_N(s)$ in Theorem 4.4 has a simple pole, rather than being holomorphic, at $s = 1$.

We say a Dirichlet character is *real* if it takes values in ± 1, and *nonreal* (or *complex*) otherwise.

THEOREM 4.6. *For any nonreal Dirichlet character χ, $L(1, \chi) \neq 0$.*

PROOF. Let N be the level of χ. If $L(1, \chi) = 0$, then also $L(1, \overline{\chi}) = 0$, where $\overline{\chi}$ denotes the complex conjugate character. But then $f_N(s)$ is the product of one factor with a simple pole at $s = 1$ (coming from the principal character), two factors with zeroes at $s = 1$ (coming from χ and $\overline{\chi}$, and a bunch of factors which are holomorphic at $s = 1$. This would force $f_N(s)$ to have a zero at $s = 1$, contradicting Theorem 4.4. □

For the real characters, the above argument fails because $\overline{\chi}$ and χ are the same character, so they don't give two different contributions to $f_N(s)$. Instead, we use a different trick. (There are a number of proofs of this result; see exercises for a second approach.)

THEOREM 4.7. *For any real nonprincipal Dirichlet character χ, $L(1, \chi) \neq 0$.*

PROOF. Assume on the contrary that $L(1, \chi) = 0$. Define

$$\psi(s) = \frac{L(s, \chi) L(s, \chi_0)}{L(2s, \chi_0)},$$

where χ_0 is the principal character of level N. Then the numerator of ψ is holomorphic for $\mathrm{Re}(s) > 0$, because $L(s, \chi)$ counterbalances the simple pole of $L(s, \chi_0)$ at $s = 1$. On the other hand, the denominator of ψ is holomorphic and nonzero for $\mathrm{Re}(s) > 1/2$; moreover, it extends meromorphically to a neighborhood of $s = 1/2$ with a simple pole at $s = 1/2$. Therefore ψ is holomorphic for $\mathrm{Re}(s) > 1/2$, and extends holomorphically to a neighborhood of $1/2$ with a simple zero at $s = 1/2$.

However, $\psi(s)$ admits the formal factorization

$$\psi(s) = \prod_{p : \chi(p) = 1} \left(\frac{1 + p^{-s}}{1 - p^{-s}} \right)$$

and so expands as a Dirichlet series with nonnegative real coefficients and constant coefficient 1. The product factorization converges absolutely for $\mathrm{Re}(s) > 1$, so the Dirichlet series does too. But ψ is holomorphic for $\mathrm{Re}(s) > 1/2$, so Landau's theorem implies that the Dirichlet series converges absolutely on $\mathrm{Re}(s) > 1/2$.

This yields $\psi(s) \geq 1$ for $s > 1/2$, whereas $\lim_{s \to (1/2)^+} \psi(s) = 0$, contradiction. □

The proofs above have the merit that one could rewrite them without using complex analysis, in order to obtain a complex analysis-free proof of Dirichlet's theorem. (Dirichlet was working before the properties of complex analytic functions were completely understood, so his proofs tend to only involve real s.) However, Dani and Sawyer pointed out that you can also argue more directly as follows. Suppose *any* of the $L(s, \chi)$ had a zero at $s = 1$; then $f_N(s)$ would be holomorphic on $\mathrm{Re}(s) > 0$. Since $\log f_N(s)$ has nonnegative real coefficients, so does $f_N(s)$ by formal exponentiation. Landau's theorem would then imply that the Dirichlet series for $f_N(s)$ converges absolutely for $\mathrm{Re}(s) > 0$. However, since the Dirichlet series for $\log f_N(s)$ diverges for $s = 1 - 1/\phi(N)$ (exercise), so does the series for $f_N(s)$, contradiction.

5. Historical aside: Dirichlet's class number formula

Dirichlet introduced the Dirichlet L-series before Riemann had introduced complex function theory into the picture, and so did not have access to such simple arguments in order to prove $L(1, \chi) \neq 0$. However, the workaround he found is quite important in its own right; he was able to express the value $L(1, \chi)$ in terms of a important numerical invariant, the *class number* of binary quadratic forms of a given discriminant. That number evidently being positive, he obtained nonvanishing of $L(1, \chi)$, and even determined its sign.

Nowadays, one typically expresses this in the language of algebraic number theory. (If you are not familiar with this language, feel free to ignore the rest of this section.) Let K be a quadratic number field, and let χ_K be the character such that

$$\chi_K(p) = \begin{cases} 1 & p \text{ is unramified and split in } K \\ -1 & p \text{ is unramified and inert in } K \\ 0 & p \text{ is ramified in } K. \end{cases}$$

One then proves that $L(1, \chi_K)$ is equal to the class number h_K of K times the regulator R_K. (The latter equals 1 if K is imaginary quadratic, and otherwise is equal to a fixed normalization factor times the logarithm of the fundamental unit of K.)

The point here is that $L(1, \chi_0)L(1, \chi_K)$ is (up to multiplication by Euler factors for the ramified primes) equal to the Dedekind zeta function ζ_K of K, defined by

$$\zeta_K(s) = \sum_{\mathfrak{a}} \mathrm{Norm}(\mathfrak{a})^{-s},$$

for $\mathfrak{a}$ running over nonzero ideals of the ring of integers $\mathfrak{o}_K$. For a general number field K, ζ_K has a simple pole at 1, whose residue is the class number of K times the regulator of K times a normalization factor (determined by the number of real and complex places of K); the point is that each factor in this product is visibly nonzero.

One sometimes turns this around and tries to use analytic information about ζ_K to get information about the product $h_K R_K$. It is quite difficult to separate the two factors in this expression; indeed, one can make a good case that they really are simply two separate factors in the computation of the volume of a certain compact topological group, the *Arakelov class group* of R, whose group of components is isomorphic to the usual class group.

Notable exception: there is no regulator for an imaginary quadratic field, so you can get good bounds in this case. For instance, the Brauer-Siegel theorem says that the class number of an imaginary quadratic field of discriminant D is at least $c_\epsilon D^{1/2-\epsilon}$ for any $\epsilon > 0$, though unfortunately the constant c_ϵ cannot be effectively determined from ϵ. The best effective results are due to Goldfeld, who proves an effective lower bound which is polynomial in $\log(D)$; this is a far cry from the truth, but is for instance enough to solve Gauss's class number 1 problem (there are exactly nine imaginary quadratic fields of class number 1).

Exercises

(1) Prove Lemma 4.3. (Hint: recall how you proved the special case $f = \zeta$ earlier.)

(2) Let f be a meromorphic function on some neighborhood of $\mathrm{Re}(s) \geq L$, with a pole of order $e > 0$ at $s = L$ and no other poles. Suppose that $\log f(s)$ is represented by a Dirichlet series with nonnegative real coefficients and abscissa of absolute convergence $\leq L$. Prove that every zero of f on the line $\mathrm{Re}(s) = L$ has multiplicity $\leq e/2$. (For $e = 1$, this implies Lemma 4.3.)

(3) Prove directly that the Dirichlet series in (8) does not converge for $s = 1 - 1/\phi(N)$. (Hint: for every Dirichlet character χ of order N, $\chi^{\phi(N)}$ is principal.)

(4) Let χ be a real nonprincipal character. Use Dirichlet's hyperbola method (from a prior homework) to show that

$$\sum_{n \leq x} f(n) n^{-1/2} = 2L(1, \chi) x^{1/2} + O(1),$$

for $f(n) = \sum_{d|n} \chi(d)$.

(5) Use the previous exercise to show that $L(1, \chi) > 0$, giving a second proof of Theorem 4.7. (Hint: prove that $f(n) \geq 1$ if n is a perfect square, and $f(n) \geq 0$ otherwise.)

CHAPTER 5

Primes in arithmetic progressions

In this unit, we first prove Dirichlet's theorem on primes in arithmetic progressions. We then prove the prime number theorem in arithmetic progressions, modulo some exercises.

1. Dirichlet's theorem

For short, I will say an arithmetic progression is *eligible* if it has the form $m, m+N, m+2N, \ldots$ where $\gcd(m, N) = 1$; it is equivalent to ask that any two consecutive terms are relatively prime.

THEOREM 5.1 (Dirichlet). *Any eligible arithmetic progression of positive integers contains infinitely many primes.*

There are a few special cases where one can prove this directly, but otherwise algebraic methods cannot touch this problem. Dirichlet's idea was to prove, in some appropriate quantitative sense, that the primes distribute themselves equally among the eligible arithmetic progressions with a particular difference; this goes back to Euler's proof of the infinitude of primes using the Riemann zeta function.

2. Asymptotic density and Dirichlet density

In order to speak quantitatively about the distribution of certain types of primes, or integers in general, we need some sort of measure theory on the set of primes or the set of integers. Note that Lebesgue-type measure theory is not an option for countable sets: we can only hope to make finitely additive measures.

For $S \subseteq T$ two sets of positive integers, with T infinite, the *upper natural density* and *lower natural density* of S in T are defined as

$$\limsup_{N\to\infty} \frac{\#\{n \in S : n \leq N\}}{\#\{n \in T : n \leq N\}}, \qquad \liminf_{N\to\infty} \frac{\#\{n \in S : n \leq N\}}{\#\{n \in T : n \leq N\}}.$$

Of course the upper density is never less than the lower density. If they coincide, we call the common value the *natural density* (or *asymptotic density*) of S in T.

Many interesting sets fail to have a natural density (e.g., see exercises). We get a less restrictive notion of density by using Dirichlet series.

For $S \subseteq T$ two sets of positive integers, with $\sum_{n\in T} n^{-1}$ divergent, the *upper Dirichlet density* and *lower Dirichlet density* of S in T are defined as

$$\limsup_{s\to 1^+} \frac{\sum_{n\in S} n^{-s}}{\sum_{n\in T} n^{-s}}, \qquad \liminf_{s\to 1^+} \frac{\sum_{n\in S} n^{-s}}{\sum_{n\in T} n^{-s}}.$$

If they coincide, we call the common value the *Dirichlet density* of S in T.

Let us make this explicit in two cases of interest. Recall that $\zeta(s)$ has a simple pole of residue 1 at $s = 1$, so as $s \to 1^+$,

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

Hence if $T = \mathbb{N}$, then the Dirichlet density of S is given by

$$\lim_{s\to 1^+} (s-1)\sum_{n\in S} n^{-s}$$

if the limit exists.

Taking logarithms, we see that as $s \to 1^+$,

$$\log\zeta(s) + \log(s-1) = \log(s-1) + \sum_p \sum_{n=1}^{\infty} p^{-ns} = O(1).$$

Moreover, $\sum_p \sum_{n=2}^{\infty} p^{-ns} = O(1)$, so

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

Hence if T is the set of primes, then the Dirichlet density of S is given by

$$\lim_{s\to 1^+} \frac{\sum_{p\in S} p^{-s}}{-\log(s-1)}$$

if the limit exists.

Here are some easy facts about density. (If I don't specify natural vs. Dirichlet, I mean that the statement holds if you make a choice and use it consistently throughout the statement.)

- Any finite set has density 0 in any infinite set.
- Density is a finitely additive measure: if $S_1, \ldots, S_m$ are disjoint subsets of T with densities $\delta_1, \ldots, \delta_m$ in T, then their union has density $\delta_1 + \cdots + \delta_m$ in T. Corollary: two subsets of T whose combined density exceeds 1 must have infinite intersection.
- If S has density δ in $\mathbb{N}$, then for any positive integer n, $nS = \{ns : s \in S\}$ has density δ/n.

I can't help mentioning a fun example of the additivity of densities. Let α, β be positive irrational numbers with $1/\alpha + 1/\beta = 1$. Put

$$S_\alpha = \{\lfloor n\alpha \rfloor : n \in \mathbb{N}\}$$
$$S_\beta = \{\lfloor n\beta \rfloor : n \in \mathbb{N}\}.$$

Then S_α, S_β have natural densities $1/\alpha, 1/\beta$. The fact that these add up to 1 is explained by the beautiful result (Beatty's theorem) that S_α, S_β are disjoint and their union is $\mathbb{N}$! (If you've never seen this before, I recommend this as an amusing exercise.)

LEMMA 5.2. *Let $S \subseteq T$ be subsets of $\mathbb{N}$ such that S has natural density δ in T. Then S also has Dirichlet density δ in T.*

PROOF. See exercises. (The converse is false; also see exercises.) □

To prove Theorem 5.1, we will prove the following.

THEOREM 5.3. *For any positive integers m, N with $\gcd(m, N) = 1$, the set of primes congruent to m modulo N has Dirichlet density $1/\phi(N)$ in the set of all primes (hence is infinite).*

3. L-functions and discrete Fourier analysis

For χ a Dirichlet character of level N, we can write

$$\log L(s, \chi) = \sum_p \sum_{n=1}^{\infty} \chi(p^n) p^{-ns}$$

for $\mathrm{Re}(s) > 1$; as $s \to 1^+$, we have

$$\log L(s, \chi) = \sum_p \chi(p) p^{-s} + O(1).$$

For χ nonprincipal, we know that $L(s, \chi)$ is holomorphic and nonvanishing at $s = 1$, so

$$\sum_p \chi(p) p^{-s} = O(1),$$

whereas for χ_0 the principal conductor of level N, we saw above that

$$\sum_p \chi_0(p) p^{-s} = -\log(s-1) + O(1).$$

At this point it may be clear how to proceed: form a certain linear combination of the $\log L(s, \chi)$ to isolate $\sum_{p \equiv m\,(N)} p^{-s}$, and compare the asymptotic contributions of $-\log(s-1)$.

The fact that we can do this amounts to what is sometimes called *discrete Fourier analysis*; if you prefer, it is the representation theory of the finite abelian group $(\mathbb{Z}/N\mathbb{Z})^*$.

THEOREM 5.4 (Discrete abelian Fourier analysis). *Let G be a finite abelian group, and let $\hat{G}$ be the* character group *(or* dual group*) of G, i.e., the group of homomorphisms $G \to \mathbb{C}^*$.*

(a) *The order of $\hat{G}$ is equal to the order of G.*
(b) *(Orthogonality of characters) If $\chi_1, \chi_2 \in \hat{G}$, then*

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

(c) *If $g_1, g_2 \in G$, then*

$$\sum_{\chi \in \hat{G}} \chi(g_1) \overline{\chi(g_2)} = \begin{cases} |G| & g_1 = g_2 \\ 0 & g_1 \neq g_2. \end{cases}$$

PROOF. (a) If $G = G_1 \times G_2$, then clearly $\hat{G} = \hat{G_1} \times \hat{G_2}$. Since every finite abelian group G is a product of cyclic groups, we may reduce to the case where G is cyclic, and then the result is clear. (For $G = (\mathbb{Z}/N\mathbb{Z})^*$, we can make this more explicit: we can use the Chinese remainder theorem to split N into distinct prime-power factors, then use the fact that $(\mathbb{Z}/p^n\mathbb{Z})^*$ is cyclic unless $p = 2$, in which case it is $\{\pm 1\}$ times a cyclic group.)

(b) We saw this argument once before, but here it goes again: the left side is invariant under multiplication by $\chi_1(h)\overline{\chi_2(h)}$ for any $h \in G$, because there is no difference between summing over g or over gh. If $\chi_1 \neq \chi_2$, then we can make that multiplier different from 1 by choosing suitable h. So the sum vanishes if $\chi_1 \neq \chi_2$. If $\chi_1 = \chi_2$, each summand is equal to 1 because characters of finite groups takes values which are roots of unity.
(c) See exercises.

□

So now it is clear what to do: given a choice of m coprime to N, taking sums of χ over all Dirichlet characters of level N, we obtain

$$\begin{aligned}\frac{1}{\phi(N)}\sum_{\chi}\overline{\chi(m)}\log L(s,\chi) &= \frac{1}{\phi(N)}\sum_{\chi}\sum_{p}\chi(p)\overline{\chi(m)}p^{-s} + O(1)\\ &= \sum_{p\equiv m\ (N)} p^{-s} + O(1)\end{aligned}$$

as $s \to 1^+$. On the other hand,

$$\begin{aligned}\frac{1}{\phi(N)}\sum_{\chi}\overline{\chi(m)}\log L(s,\chi) &= \frac{1}{\phi(N)}\log L(s,\chi_0) + \frac{1}{\phi(N)}\sum_{\chi\neq\chi_0}\overline{\chi(m)}\log L(s,\chi)\\ &= -\frac{1}{\phi(N)}\log(s-1) + O(1).\end{aligned}$$

This yields Theorem 5.3.

4. The prime number theorem in arithmetic progressions

The proof of Dirichlet's theorem only uses information about the behavior of the $L(s,\chi)$ near $s = 1$. Using the fact that $L(s,\chi) \neq 0$ on the entire line $\mathrm{Re}(s) = 1$, we can prove a much stronger result.

THEOREM 5.5 (Prime number theorem in arithmetic progressions). *For m, N positive integers with $\gcd(m,N)$, the set of primes congruent to m modulo N has* natural *density $1/\phi(N)$. In other words, the number of primes $p \leq x$ with $p \equiv m \pmod N$ is asymptotic to $\frac{1}{\phi(N)}\frac{x}{\log x}$ as $x \to \infty$.*

PROOF. Given what we now know, this is a straightforward adaptation of our proof of the prime number theorem. For χ a Dirichlet character of level N, define

$$\vartheta_\chi(x) = \sum_{p\leq x}\chi(p)\log p.$$

Given a choice of m coprime to N, put

$$\begin{aligned}\vartheta_m(x) &= \frac{1}{\phi(N)}\sum_{\chi}\overline{\chi(m)}\vartheta_\chi(x)\\ &= \sum_{p\leq x: p\equiv m\ (N)}\log p.\end{aligned}$$

As in the proof of the prime number theorem, if we prove that the improper integral

$$\int_1^\infty \frac{\phi(N)\vartheta_m(x) - x}{x^2}\,dx$$

converges, we may then deduce that $\vartheta_m(x) \sim \frac{1}{\phi(N)}x$ as desired.

It suffices in turn to check that for χ principal,

$$\int_1^\infty \frac{\vartheta_\chi(x) - x}{x^2}\,dx$$

converges, and for χ nonprincipal,

$$\int_1^\infty \frac{\vartheta_\chi(x)}{x^2}\,dx$$

converges. The former is an immediate consequence of the corresponding fact for ϑ (which we proved in the unit on the prime number theorem), since ϑ and ϑ_χ differ in only finitely many terms. For the latter, see exercises. □

As with the proof of the prime number theorem, we get very little information about the error term, i.e., the difference between the actual number of primes $p \leq x$ with $p \equiv m \pmod N$ and the asympotic count $\frac{1}{\phi(N)}\frac{x}{\log x}$. That becomes a problem if, for instance, we want to know how long it takes to find *one* prime in an arithmetic progression. To address this, we must first get better results on zero-free regions for the $L(s,\chi)$, then make a better analytic argument to take advantage of the improved analytic information. We turn to this in the next few lectures.

Exercises

(1) Prove Lemma 5.2. (Hint: use partial summation.)

(2) Let S be the set of positive integers which have first digit 1 when written in base 10.
   (a) Compute the upper and lower natural density of S, and verify that S does not have a natural density.
   (b) Prove that S has a natural Dirichlet density, and compute it.

   Optional (not to be turned in): generalize to an arbitrary base $b \geq 3$. Even more optional: prove the analogous result for the set of primes with first digit 1 in base b.

(3) Prove that there exists a constant c such that

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + c + o(1).$$

   (Hint: you established asymptotics for $\sum_{p\leq x}\frac{\log p}{p}$ on a previous homework. Apply partial summation.)

(4) (a result of Mertens; tricky, optional) In the previous exercise, prove that

$$c = \gamma + \sum_p \left(\log(1-p^{-1}) + p^{-1}\right),$$

   where γ is Euler's constant. Then deduce that

$$\prod_{p\leq x}(1-p^{-1}) \sim \frac{e^{-\gamma}}{\log x}.$$

(5) Deduce point (c) of Theorem 5.4 from points (a) and (b). (Hint: form the matrix A with rows indexed by $g \in G$, columns indexed by $\chi \in \hat{G}$,

and entries $\chi(g)$. Then compare AA^* with A^*A, where $*$ denotes conjugate transpose. Or if you prefer, prove that the dual of $\hat{G}$ is canonically isomorphic to G.)

(6) Prove that $\int_1^\infty \frac{\vartheta_\chi(t)}{t^2}\,dt$ converges for χ nonprincipal, by applying the Tauberian theorem from the unit on the prime number theorem. (Hint: use the fact that $L(s,\chi) \neq 0$ for $\mathrm{Re}(s) \geq 1$ to argue that $-L'(s,\chi)/L(s,\chi)$ is holomorphic in a neighborhood of $\mathrm{Re}(s) \geq 1$. There will be an extra term to deal with, just as there was a term I neglected in the original notes from the prime number theorem unit; see the corrected notes online.)

CHAPTER 6

The functional equation for the Riemann zeta function

In this unit, we establish the functional equation property for the Riemann zeta function, which will imply its meromorphic continuation to the entire complex plane. We will do likewise for Dirichlet L-functions in the next unit.

1. The functional equation for ζ

A "random" Dirichlet series $\sum_n a_n n^{-s}$ will not exhibit very interesting analytic behavior beyond its abscissa of absolute convergence. However, we already know that ζ is atypical in this regard, in that we can extend it at least as far as $\mathrm{Re}(s) > 0$ if we allow the simple pole at $s = 1$. One of Riemann's key observations is that in the strip $0 < \mathrm{Re}(s) < 1$, ζ obeys a symmetry property relating $\zeta(s)$ to $\zeta(1-s)$; once we prove this, we will then be able to extend ζ all the way across the complex plane. (This is essentially Riemann's original proof; several others are possible.)

We first recall the definition and basic properties of the Γ function. We may define

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

for $\mathrm{Re}(s) > 0$. It is then straightforward to check (by integration by parts) that

$$\Gamma(s+1) = s\Gamma(s) \qquad (\mathrm{Re}(s) > 0). \tag{9}$$

Using (9), we may extend Γ to a meromorphic function on all of $\mathbb{C}$, with simple poles at $s = 0, -1, \ldots$. Since $\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1$, we have that for n a nonnegative integer,

$$\Gamma(n+1) = n!.$$

Substituting $t = \pi n^2 x$ in the definition of Γ, we have

$$\pi^{-s/2}\Gamma(s/2)n^{-s} = \int_0^\infty x^{s/2-1} e^{-n^2\pi x}\, dx \qquad \mathrm{Re}(s) > 0.$$

If we sum over n, we can interchange the sum and integral for $\mathrm{Re}(s) > 1$ because the sum-integral converges absolutely. Hence

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \int_0^\infty x^{s/2-1}\omega(x)\, dx \qquad \mathrm{Re}(s) > 1$$

for

$$\omega(x) = \sum_{n=1}^\infty e^{-n^2\pi x}.$$

It is slightly more convenient to work with the function θ defined by

$$\theta(x) = \sum_{n=-\infty}^{\infty} e^{-n^2 \pi x},$$

which clearly satisfies $2\omega(x) = \theta(x) - 1$.

At this point, Riemann recognized θ as a function of the sort considered by Jacobi in the late 19th century; from that work, Riemann knew about the identity

$$\theta(x^{-1}) = x^{1/2}\theta(x) \qquad x > 0. \tag{10}$$

We will return to the proof of (10) in the next section; for the moment, let's see how we use this to get a functional equation for ζ.

Returning to

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \int_0^\infty x^{s/2-1}\omega(x)\,dx,$$

we take the natural step of splitting the integral at $x = 1$, then substituting $1/x$ for x in the integral from 0 to 1. This yields

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \int_1^\infty x^{s/2-1}\omega(x)\,dx + \int_1^\infty x^{-s/2-1}\omega(1/x)\,dx.$$

From (10), we deduce

$$\omega(x^{-1}) = -\frac{1}{2} + \frac{1}{2}x^{1/2} + x^{1/2}\omega(x),$$

yielding

$$\int_1^\infty x^{-s/2-1}\omega(x^{-1})\,dx = -\frac{1}{s} + \frac{1}{s-1} + \int_1^\infty x^{-s/2-1/2}\omega(x)\,dx$$

and so

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = -\frac{1}{s(1-s)} + \int_1^\infty (x^{s/2-1} + x^{(1-s)/2-1})\omega(x)\,dx, \tag{11}$$

at least for $\mathrm{Re}(s) > 1$.

But now notice that the left side of (11) represents a meromorphic function on $\mathrm{Re}(s) > 0$, whereas the right side of (11) represents a meromorphic function on all of $\mathbb{C}$, because the integral converges absolutely for all z. (That's because $\omega(x) = O(e^{-\pi x})$ as $x \to +\infty$.)

This has tons of consequences. First, (11) is also valid for $\mathrm{Re}(s) > 0$. Second, we can use (11) to *define* $\zeta(s)$ as a meromorphic function on all of $\mathbb{C}$. Third, the right side of (11) is invariant under the substitution $s \mapsto 1 - s$, so we obtain a functional equation for ζ. One often writes this by defining

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

and then the functional equation is $\xi(1-s) = \xi(s)$. Fourth, the function $\xi(s)$ is actually entire for $\mathrm{Re}(s) > 0$ because the factor of $s - 1$ counters the pole of ζ at $s = 1$; by the functional equation, ξ is entire everywhere.

Remember that $\zeta(s)$ has no zeroes in the region $\mathrm{Re}(s) \geq 1$. By the functional equation, in the region $\mathrm{Re}(s) \leq 0$, the only zeroes of ζ occur at the poles of $\Gamma(s/2)$ (except for $s = 0$, where the factor of s counters the pole), i.e., at negative even integers. These are called the *trivial zeroes* of ζ; the other zeroes, which are forced to lie in the range $0 < \mathrm{Re}(s) < 1$, are much more interesting!

2. The θ function and the Fourier transform

I still owe you the functional equation (10) for the θ function. It is usually deduced from the Poisson summation formula for Fourier transforms, which I'll now recall.

Let $f : \mathbb{R} \to \mathbb{C}$ be an integrable (L^1) function. The *Fourier transform* of f is then defined as the function $\hat{f} : \mathbb{R} \to \mathbb{C}$ given by

$$\hat{f}(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t} f(t)\, dt;$$

it is uniformly continuous.

It is convenient to restrict attention to a smaller class of functions. We say $f : \mathbb{R} \to \mathbb{C}$ is a *Schwarz function* if f is infinitely differentiable and, for each nonnegative integer n and each $c \in \mathbb{R}$, $|f^{(n)}(t)| = o(|t|^c)$ as $t \to \pm\infty$.

LEMMA 6.1. *Let $f, g : \mathbb{R} \to \mathbb{C}$ be Schwarz functions.*

(a) *The functions $\hat{f}, \hat{g}$ are again Schwarz functions.*

(a) *We have $\hat{\hat{f}}(t) = f(-t)$.*

(b) *If we define the convolution $f \star g$ by*

$$(f \star g)(t) = \int_{-\infty}^{\infty} f(t-u) g(u)\, du,$$

then $\widehat{f \star g}(s) = \hat{f}(s)\hat{g}(s)$.

PROOF. Exercise. □

THEOREM 6.2 (Poisson summation formula). *Let $f : \mathbb{R} \to \mathbb{C}$ be a Schwarz function. Then*

$$\sum_{m \in \mathbb{Z}} f(m) = \sum_{n \in \mathbb{Z}} \hat{f}(n); \tag{12}$$

in particular, the sum on the right converges.

SKETCH OF PROOF. One can give a simple proof using Fourier series; see exercises. Here I'll sketch a more direct approach which has some significance in analytic number theory; it is a very simple version of the Hardy-Littlewood "circle method".

Write

$$\begin{aligned}
\sum_{n=-N}^{N} \hat{f}(n) &= \int_{-\infty}^{+\infty} \sum_{n=-N}^{n} e^{-2\pi i n t} f(t)\, dt \\
&= \sum_{m=-\infty}^{\infty} \int_{m-1/N}^{m+1/N} \frac{e^{-2\pi i N t} - e^{2\pi i (N+1) t}}{1 - e^{2\pi i t}} f(t)\, dt \\
&\qquad + \sum_{m=-\infty}^{\infty} \int_{m+1/N}^{m+1-1/N} \frac{e^{-2\pi i N t} - e^{2\pi i (N+1) t}}{1 - e^{2\pi i t}} f(t)\, dt.
\end{aligned}$$

Then check that the summand in the first sum converges (uniformly on m) to $f(m)$, while the second summand converges (uniformly on m) to zero. (If you prefer, first use a partition of unity to reduce to the case where f is supported on an interval like $[-2/3, 2/3]$, so that the sums over m become finite.) □

The Poisson summation formula immediately yields (10) as soon as one checks that $f(t) = e^{-\pi t^2}$ is invariant under the Fourier transform (see exercises): it then follows that the Fourier transform of $f(t) = e^{-\pi x t^2}$ is $\hat{f}(s) = x^{-1/2} e^{-\pi x s^2}$, and Poisson summation gives (10).

2.1. Asides. Our study of θ merely grazes the top of a very large iceberg. Here are three comments to this effect.

A much more general version of θ was considered by Jacobi, in which he considered a quadratic form $Q(x_1, \ldots, x_m)$ and formed the sum

$$\theta_Q(x) = \sum_{n_1, \ldots, n_m \in \mathbb{Z}} e^{-Q(n_1, \ldots, n_m)\pi x};$$

if Q is positive definite, this again converges rapidly for all x.

One can also think of θ as a example of a special sort of complex function called a *modular form*. Nowadays, modular forms are central not just to analytic number theory, but a lot of algebraic number theory as well. For instance, the "modularity of elliptic curves" is central to the proof of Fermat's last theorem; I may say a bit about this later in the course.

The Fourier transform is a typical example of an *integral transform*; the function $s, t \mapsto e^{-2\pi i s t}$ is the *kernel* of this transform. Another important integral transform in analytic number theory is the *Mellin transform*: for a function $f : [0, \infty) \to \mathbb{C}$, the Mellin transform $M(f)$ is given by

$$M(f)(s) = \int_0^\infty f(t) t^{s-1}\, dt.$$

For instance, $\Gamma(s)$ is the Mellin transform of e^{-t}.

Exercises

(1) Rewrite the functional equation directly in terms of $\zeta(s)$ and $\zeta(1-s)$.
(2) What is the residue of the pole of Γ at a nonpositive integer s?
(3) Prove Lemma 6.1.
(4) Prove the Poisson summation formula either by completing the sketch given above, or by considering the Fourier series of the function

$$F(s) = \sum_{m \in \mathbb{Z}} f(s + m).$$

(5) Prove that the function $f(t) = e^{-\pi t^2}$ is its own Fourier transform.

CHAPTER 7

Functional equations for Dirichlet L-functions

In this unit, we establish the functional equation property for Dirichlet L-functions. Much of the work is left as exercises.

1. Even characters

Let χ be a Dirichlet character of level N. We say χ is *even* if $\chi(-1) = 1$ and *odd* if $\chi(-1) = -1$.

For χ even, we can derive a functional equation for $L(s, \chi)$ by imitating the argument we used for ζ. Start with

$$\chi(n)\pi^{-s/2}N^{s/2}\Gamma(s/2)n^{-s} = \int_0^\infty \chi(n)e^{-\pi n^2 x/N}x^{s/2-1}\,dx$$

and sum over n to obtain

$$\pi^{-s/2}N^{s/2}\Gamma(s/2)L(s,\chi) = \frac{1}{2}\int_0^\infty x^{s/2-1}\theta(x,\chi)\,dx \tag{13}$$

for

$$\theta(x,\chi) = \sum_{n=-\infty}^{\infty} \chi(n)e^{-\pi n^2 x/N}.$$

(Notice there is no additive constant because $\chi(0) = 0$.)

Applying the Poisson summation formula to $\theta(x, \chi)$ looks problematic, because $\chi(n)$ does not extend nicely to a function on all of $\mathbb{R}$. Fortunately we can avoid this by doing a bit more Fourier analysis, but this time on the *additive* group $\mathbb{Z}/N\mathbb{Z}$: write

$$\chi(n) = \sum_{m=1}^{N} c_{\chi,m}e^{2\pi imn/N}$$

with

$$c_{\chi,m} = \frac{1}{N}\sum_{l=1}^{N}\chi(l)e^{-2\pi ilm/N}.$$

I'll come back to what this quantity $c_{\chi,m}$ actually is in a moment. In the meantime, let's see what happens when we use this new expression for $\chi(n)$. Or rather, I'll let you see what happens as an exercise; you should get

$$\theta(x,\chi) = (N/x)^{1/2}\sum_{m=1}^{N} c_{\chi,m}\sum_{n=-\infty}^{\infty} e^{-\pi(nN+m)^2/(xN)}. \tag{14}$$

To get further, we need some description of the $c_{\chi,m}$ which is somehow uniform in m. Here it is: if m is coprime to N, then

$$\begin{aligned} c_{\chi,m} &= \frac{1}{N}\sum_{l=1}^{N}\chi(l)e^{-2\pi ilm/N} \\ &= \frac{1}{N}\sum_{l=1}^{N}\overline{\chi(m)}\chi(lm)e^{-2\pi ilm/N} \\ &= \overline{\chi}(m)c_{\chi,1}. \end{aligned}$$

For m not coprime to N, we must assume χ is primitive, and then again

$$c_{\chi,m} = \overline{\chi}(m)c_{\chi,1} \tag{15}$$

but this is not so obvious; see exercises.

This gives us

$$\theta(x,\chi) = (N/x)^{1/2}c_{\chi,1}\theta(x^{-1},\overline{\chi}),$$

and now we are home free: again split the integral (13) at 1 and substitute $x \mapsto x^{-1}$ in one term, to obtain

$$\begin{aligned} \pi^{-s/2}N^{s/2}\Gamma(s/2)L(s,\chi) &= \frac{1}{2}\int_1^\infty x^{s/2-1}\theta(x,\chi)\,dx + \frac{1}{2}\int_1^\infty x^{-s/2-1}\theta(x^{-1},\chi)\,dx \\ &= \frac{1}{2}\int_1^\infty x^{s/2-1}\theta(x,\chi)\,dx + \frac{1}{2}N^{1/2}c_{\chi,1}\int_1^\infty x^{(1-s)/2-1}\theta(x,\overline{\chi})\,dx. \end{aligned}$$

Similarly,

$$\pi^{-s/2}N^{s/2}\Gamma(s/2)L(s,\overline{\chi}) = \frac{1}{2}\int_1^\infty x^{s/2-1}\theta(x,\overline{\chi})\,dx + \frac{1}{2}N^{1/2}c_{\overline{\chi},1}\int_1^\infty x^{(1-s)/2-1}\theta(x,\chi)\,dx.$$

It is elementary to check that $c_{\chi,1}c_{\overline{\chi},1} = N^{-1}$ (see exercises); we thus obtain

$$\pi^{-(1-s)/2}N^{(1-s)/2}\Gamma((1-s)/2)L(1-s,\overline{\chi}) = N^{1/2}c_{\overline{\chi},1}\pi^{-s}N^{s/2}\Gamma(s/2)L(s,\chi). \tag{16}$$

Again, the extra factors of π, N, Γ should be thought of as an "extra Euler factor" coming from the "prime at infinity".

Pay close attention to the fact that unless $\chi = \overline{\chi}$, the functional equation 16 relates two *different* L-functions. In a few circumstances, this makes it less useful than if it related a single $L(s,\chi)$ to itself, but so be it.

Also note that quantity $c_{\chi,1}$ is related to the more commonly introduced *Gauss sum* associated to χ:

$$\tau(\chi) = Nc_{\overline{\chi},1} = \sum_{l=1}^{N}\chi(l)e^{2\pi il/N}.$$

For more about Gauss sums, see the exercises.

2. Odd characters

We have to do something different if $\chi(-1) = -1$, as then the function $\theta(x,\chi)$ as defined above is identically zero. Instead we use

$$\theta_1(x,\chi) = \sum_{n=\infty}^{\infty} n\chi(n)e^{-n^2\pi x/N}$$

and shift s around a bit. Namely,

$$\pi^{-(s+1)/2}N^{(s+1)/2}\Gamma((s+1)/2)L(s,\chi) = \frac{1}{2}\int_0^\infty \theta_1(x,\chi)x^{(s+1)/2-1}\,dx.$$

Again you split the integral at $x = 1$ and use an inversion formula; this time the right identity is

$$\sum_{n=-\infty}^{\infty} ne^{-n^2\pi x/N + 2\pi imn/N} = i(N/x)^{3/2}\sum_{n=-\infty}^{\infty}\left(n+\frac{m}{N}\right)e^{-\pi(n+m/N)^2N/x}. \tag{17}$$

You should end up with the functional equation

$$\pi^{-(2-s)/2}N^{(2-s)/2}\Gamma((2-s)/2)L(1-s,\overline{\chi}) = -i\tau(\overline{\chi})N^{-1/2}\pi^{-(1+s)/2}N^{(1+s)/2}\Gamma((1+s)/2)L(s,\chi). \tag{18}$$

Since this is now the third time through this manner of argument, I leave further details to the exercises.

Exercises

(1) Prove the following functional equations for Γ:

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

Then use these to give a simplified functional equation for ζ of the form "$\zeta(1-s)$ equals $\zeta(s)$ times some explicit function".

(2) Prove that (15) holds for χ primitive whether or not m is coprime to N.

(3) Prove that for χ primitive, $\tau(\chi)\overline{\tau(\chi)} = N$. (Warning: the value of $\tau(\chi)\tau(\overline{\chi})$ depends on whether χ is even or odd.) Then exhibit an example where this fails if χ is imprimitive.

(4) For χ a Dirichlet character of level N, based on the functional equation, where does $L(s,\chi)$ have zeroes and poles in the region $\mathrm{Re}(s) \le 0$?

(5) Prove that

$$\sum_{n=-\infty}^{\infty} e^{-(n+\alpha)^2\pi/x} = x^{1/2}\sum_{n=-\infty}^{\infty} e^{-n^2\pi x + 2\pi i n\alpha} \qquad (\alpha \in \mathbb{R}, x > 0);$$

then prove (17) by the same method (namely Poisson summation).

(6) Use the previous exercise to deduce (14).

(7) Prove the functional equation (18).

(8) Pick an example of a nonprincipal nonprimitive character χ, and write out the functional equation for $L(s,\chi)$.

(9) (Dirichlet) For $a, b \in \mathbb{Z}$ and $f : \mathbb{R} \to \mathbb{C}$ a function obtained by taking a continuous function on $[a,b]$ and setting its other values to 0, the Poisson summation formula still holds if interpreted as

$$\frac{1}{2}f(a) + f(a+1) + \cdots + f(b-1) + \frac{1}{2}f(b) = \sum_{n=-\infty}^{\infty}\hat{f}(n)$$

(you don't have to prove this). Apply this to the function

$$f(t) = \begin{cases} e^{2\pi i t^2/N} & t \in [0,N] \\ 0 & \text{otherwise} \end{cases}$$

in order to evaluate $\sum_{n=1}^{N} e^{2\pi i n^2/N}$ for N a positive integer. Then use this to compute $G(\chi)$ for χ the quadratic character $\chi(m) = \left(\frac{m}{p}\right)$. (Optional, not to be turned in: give a more elementary computation of $G(\chi)^2$.)

CHAPTER 8

Error bounds in the prime number theorem

In this unit, we introduce (without proof for now) a formula which relates the distribution of primes to the zeroes of the Riemann zeta function. Given a suitable zero-free region for $\zeta(s)$ in the critical strip, this can be used to prove the prime number theorem with an estimate for the error term.

1. Zeta zeroes and prime numbers

For $x \notin \mathbb{N}$, define the counting function

$$\psi(x) = \sum_{n \leq x} \Lambda(n),$$

where $\Lambda : \mathbb{N} \to \mathbb{R}$ is the von Mangoldt function

$$\Lambda(n) = \begin{cases} \log p & n = p^a, a \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $x \in \mathbb{N}$, it is convenient to modify the definition to

$$\psi(x) = \sum_{n < x} \Lambda(n) + \frac{1}{2}\Lambda(x).$$

Note that for the function ϑ we defined earlier as

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

we have

$$\psi(x) - \vartheta(x) = O(x^{1/2} \log x) \qquad (x \to \infty)$$

so the prime number theorem is equivalent to

$$\psi(x) \sim x \qquad (x \to \infty).$$

The formula of von Mangoldt expresses the difference $\psi(x) - x$ in terms of the zeroes of $\zeta(s)$. We will prove this formula in a later unit.

THEOREM 8.1 (von Mangoldt's formula). *For $x \geq 2$ and $T > 0$,*

$$\psi(x) - x = -\sum_{\rho : |\operatorname{Im}(\rho)| < T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log(1 - x^{-2}) + R(x, T)$$

with ρ running over the zeroes of $\zeta(s)$ in the region $\operatorname{Re}(s) \in [0, 1]$, and

$$R(x, T) = O\left(\frac{x \log^2(xT)}{T} + (\log x) \min\left\{1, \frac{x}{T\langle x \rangle}\right\}\right).$$

Here $\langle x \rangle$ denotes the distance from x to the nearest prime power other than possibly x itself.

The region $\mathrm{Re}(s) \in [0,1]$ is called the *critical strip* for ζ, because we can account for all of the zeroes outside this strip: they are the trivial zeroes $s = -2, -4, \ldots$ forced by the functional equation and the fact that $\Gamma(s/2)$ has poles at nonpositive even integers. In fact, the last term in the formula is merely $-\sum_\rho \frac{x^\rho}{\rho}$ for ρ running over the trivial zeroes.

Incidentally, one can check by a numerical calculation that there are no real zeroes of ζ in the critical strip, by numerically approximating the integral representation of $\xi(s)$. This raises an interesting point: in general, direct numerical approximation can be used to prove that an analytic function does not vanish in a region, but not that it does vanish at a particular point. The best one can do is use a zero-counting formula to prove that there must be a zero near the proposed vanishing point.

Note that for x fixed, $R(x,T) = o(1)$ as $T \to \infty$, so we have

$$\psi(x) - x = -\sum_\rho \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log(1 - x^{-2})$$

as long as we interpret the sum over ρ to mean the limit of the partial sums over $|\mathrm{Im}(\rho)| < T$ as $T \to \infty$. This formula, while pretty, is not as useful in practice as the form with remainder; we will use the remainder form by taking T to be some (preferably large) function of x as $x \to \infty$.

2. How to use von Mangoldt's formula

In order to use von Mangoldt's formula to bound $\psi(x) - x$, we need to give an upper bound on the sum $\sum_\rho x^\rho/\rho$ for ρ running over nontrivial zeroes of ζ in the region $|\mathrm{Im}(s)| \leq T$.

Put $\beta = \mathrm{Re}(\rho), \gamma = \mathrm{Im}(\rho)$. Suppose we can prove that $\beta < 1 - f(|\gamma|)$ for some nonincreasing function $f : [0,\infty) \to (0, 1/2)$; then

$$|x^\rho| = x^\beta < x^{1-f(|\gamma|)} < x^{1-f(T)}$$

and $|\rho| \geq |\gamma|$. We thus have

$$\left| \sum_{\rho : |\gamma| < T} \frac{x^\rho}{\rho} \right| \leq x^{1-f(T)} \sum_{\rho : |\gamma| < T} \frac{1}{\gamma}.$$

Let $N(T)$ be the number of zeroes in the critical strip with $|\gamma| \leq T$. Then

$$\sum_{\rho : 0 < |\gamma| < T} \frac{1}{\gamma} = \int_0^T t^{-1} dN(t) = \frac{N(T)}{T} + \int_0^T t^{-2} N(t)\, dt.$$

At this point we need some information about $N(T)$; again, we will prove this (and a bit more) later.

THEOREM 8.2 (Hadamard). *We have $N(T) = O(T \log T)$ as $T \to \infty$.*

This implies that

$$\left| \sum_{\rho : |\gamma| < T} \frac{1}{\gamma} \right| = O(\log^2 T),$$

so

$$\left| \sum_{\rho:|\gamma|<T} \frac{x^\rho}{\rho} \right| = O(x^{1-f(T)} \log^2 T).$$

For x an integer, we now take $T = T(x)$ to be a suitable function of x, and invoke von Mangoldt's formula with remainder to deduce that

$$\psi(x) - x = O\left(x^{1-f(T)} \log^2 T(x) + \frac{x \log^2 x}{T(x)} + \frac{x \log^2 T(x)}{T(x)} \right). \tag{19}$$

3. The Riemann Hypothesis

Riemann calculated a few of the zeroes of ζ and, based on this evidence, made the following remarkable conjecture (whose resolution is worth $1,000,000 from the Clay Mathematics Institute).

CONJECTURE 8.3 (Riemann Hypothesis). *The nontrivial zeroes of ζ all lie on the line* $\mathrm{Re}(s) = \frac{1}{2}$.

This is a best-case scenario in terms of deducing error bounds on $\psi(x) - x$. Namely, suppose every nontrivial zero ρ of ζ satisfies $c \leq \mathrm{Re}(\rho) \leq 1 - c$ for some $c \in (0, 1/2)$; then we can take $f(T) = c$ in (19), yielding

$$\psi(x) - x = O\left(x^{1-c} \log^2 T(x) + \frac{x \log^2 x}{T(x)} + \frac{x \log^2 T(x)}{T(x)} \right).$$

By taking $T(x) = x$, we obtain

$$\psi(x) - x = O(x^{1-c} \log^2 x).$$

If I can take c to be any value less than $1/2$, that means

$$\psi(x) - x = O(x^{1/2+\epsilon}) \qquad (\epsilon > 0),$$

and similarly one gets a strong estimate on $\pi(x)$ (see exercises).

Unfortunately, for *no* value of $c > 0$ are we able at present to prove that every nontrivial zero ρ satisfies $\mathrm{Re}(\rho) \leq 1 - c$. We will give a much smaller zero-free region in a later unit.

4. Variants for L-functions

For χ a Dirichlet character, define

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

where again we multiply the $n = x$ term by $1/2$ if it is present.

THEOREM 8.4. *For χ a nonprincipal Dirichlet character of level N,*

$$\psi(x, \chi) = -\sum_{\rho:|\gamma|<T} \frac{x^\rho}{\rho} - (1-a) \log x - b(\chi) + \sum_{m=1}^{\infty} \frac{x^{a-2m}}{2m-a} + R(x, T),$$

where $b(\chi)$ is an explicit constant, $a = 1$ for χ even and $a = 0$ for χ odd, and

$$R(x, T) = O\left(\frac{x \log^2(NxT)}{T} + (\log x) \min\left\{ 1, \frac{x}{T\langle x \rangle} \right\} \right).$$

For a fixed N, one can use this formula together with a zero-free region for all of the $L(s,\chi)$ with χ of level N, to obtain a prime number theorem for arithmetic progressions of difference N with an estimate for the error term.

However, one would also like to be able to establish a prime number theorem with error term for arithmetic progressions where the difference is allowed to vary. In this case, one of course must have a zero-free region for all of the relevant characters. But there are two extra complications.

- One must understand how the constant $b(\chi)$ varies with χ.
- One must deal with possible roots of $L(s,\chi)$ that are very close to $s=0$ or $s=1$ (so-called *Siegel zeroes*).

Dealing with these goes beyond the level of detail I have in mind for this course; see Davenport §14–22 for a systematic exposition.

Exercises

(1) Assume that $\psi(x) = x + o(x^{1-\epsilon})$ for some given $\epsilon \in (0, 1/2)$. Deduce a corresponding upper bound for $\pi(x) - \mathrm{li}(x)$, where $\mathrm{li}(x)$ is the logarithmic integral function

$$\mathrm{li}(x) = \int_2^x \frac{dt}{\log t}.$$

Then deduce that

$$\pi(x) - \frac{x}{\log x} \neq o(x^{1-\delta})$$

for any $\delta > 0$. (This last statement can be proved unconditionally, but don't worry about that for now.) This is the sense in which $\mathrm{li}(x)$ is a better approximation than $x/(\log x)$ of the count of primes.

CHAPTER 9

More on the zeroes of zeta

In this unit, we derive some results about the location of the zeroes of the Riemann zeta function, including a small zero-free region inside the critical strip.

1. Order of an entire function

For $\alpha > 0$, an entire function $f : \mathbb{C} \to \mathbb{C}$ is said to have *order* $\leq \alpha$ if for all $\beta > \alpha$,

$$f(z) = O(\exp |z|^\beta) \qquad (|z| \to \infty).$$

We say f has order α if it has order $\leq \alpha$ but not order $\leq \beta$ for any $\beta < \alpha$.

LEMMA 9.1. *The function*

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

satisfies

$$|\xi(s)| < \exp(C|s| \log |s|) \qquad (|s| \to \infty),$$

and so is of order ≤ 1. *(An analogue is true for L-functions, but that is too easy even to give as an exercise.)*

PROOF. By the functional equation $\xi(s) = \xi(1-s)$, it suffices to check for $|\operatorname{Re}(s)| \geq 1/2$, in which case

$$\left| \frac{1}{2} s(s-1)\pi^{-s/2} \right| < \exp(C_1|s|)$$

$$|\Gamma(s/2)| < \exp(C_2|s| \log |s|)$$

(see exercises for the second estimate). For ζ, we use the integral representation from the first lecture:

$$\zeta(s) = \frac{s}{s-1} - s \int_1^\infty (x - \lfloor x \rfloor) x^{-s-1}\, dx \qquad (\operatorname{Re}(s) > 0).$$

For $\operatorname{Re}(s) \geq 1/2$, the integral is bounded, so $|\zeta(s)| < C_3|s|$. This yields the claim. □

There is a rich theory of integral functions of finite order due to Hadamard (which I believe was introduced originally for the very purpose of studying ζ). The basic idea is to generalize the fact that a polynomial can be written as a product of linear factors (the Fundamental Theorem of Algebra), to write an entire function as a product of one factor for each zero times an exponential.

To do this, one must first control the number of zeroes of f in a disc. There is no harm in assuming that $f(0) \neq 0$, since otherwise we just divide by a suitable power of z. Then recall the following fact from complex analysis.

THEOREM 9.2 (Jensen's formula). *If $f(0) \neq 0$ and f has no zeroes on the circle $|z| = R$, then*

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

where ρ runs over the zeroes of f in the disc $|z| < R$ counted with multiplicity.

PROOF. Write $f(z) = (z-\rho_1)\cdots(z-\rho_n)g(z)$, where g is nonzero on the disc $|z| \leq R$, and check the equality for each factor individually. For $z - \rho_i$, this is an easy exercise; for g, apply the Cauchy residue formula to the contour integral $\int \log(g(z))\frac{dz}{z}$ around the circle $|z| = R$, then take real parts. □

The right side is also

$$\log|f(0)| + \int_0^R \#\{\rho : |\rho| < r\}\frac{dr}{r}.$$

If $\log|f(z)| < r(|z|)$ for some function r, then the left side of Jensen's formula is bounded by $2r(R)$, whereas the right side is at least

$$\log|f(0)| + \log(2)\#\{\rho : |\rho| \leq R/2\}.$$

Consequently, if $r(R) = O(R^\alpha)$, then the number of roots of f in the disc $|\rho| \leq R$ is also $O(R^\alpha)$. Similarly, the fact that $\log|\xi(s)| = O(|s|\log|s|)$ implies that the number of zeroes of ζ with $|\operatorname{Im}(s)| \leq T$ is $O(T\log T)$, which I claimed without proof in the previous unit.

Now let f be entire of order ≤ 1. Let $\rho_1, \rho_2, \ldots$ be the zeroes of f sorted so that $|\rho_1| \leq |\rho_2| \leq \cdots$, and put

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

Note that this converges uniformly on any disc, because the multiplicand is

$$1 + \frac{1}{2}\left(\frac{z}{\rho_n}\right)^2 + O\left(\left(\frac{z}{\rho_n}\right)^3\right)$$

and the fact that the number of roots of norm $\leq R$ is $O(R^{1+\epsilon})$ implies that $\sum 1/\rho_n^2$ converges (by partial summation). By a somewhat intricate argument (see Davenport §11 or Ahlfors), it can be shown that f/h is also of order ≤ 1. Since f/h has no zeroes, the function $g(z) = \log(f(z)/h(z))$ is entire and satisfies $|g(z)| = O(|z|^{1+\epsilon})$. Consequently,

$$g_2(z) = \frac{g(z) - g(0) - g'(0)z}{z^2}$$

is entire and bounded, hence constant by Liouville's theorem. This yields the following.

THEOREM 9.3 (Hadamard). *Let $f(z)$ be an entire function of order ≤ 1. Then*

$$f(z) = e^{A+Bz}\prod_{n=1}^\infty (1 - z/\rho_n)e^{z/\rho_n}$$

for some constants A, B.

2. A zero-free region for ζ

We now use the product representation for ξ to obtain a zero-free region for ζ. The idea (due to de la Vallée Poussin (1899)) is to squeeze a bit of extra information out of the proof we used for nonvanishing on the line $\operatorname{Re}(s) = 1$. One way to phrase that argument: since

$$\operatorname{Re}(\log(\zeta(s)) = \sum_p \sum_{n=1}^{\infty} \frac{1}{n} \cos(\operatorname{Im}(s) \log p^n) p^{-n \operatorname{Re}(s)}$$

and

$$3 + 4\cos\theta + \cos 2\theta \geq 0,$$

we have

$$3 \operatorname{Re}(\log \zeta(\sigma)) + 4 \operatorname{Re}(\log \zeta(\sigma + it)) + \operatorname{Re}(\log \zeta(\sigma + 2it)) \geq 0 \qquad (\sigma > 1, t \in \mathbb{R})$$

whereas if $\zeta(1 + it)$ vanished, then the sum would tend to $-\infty$ as $\sigma \to 1^+$ (because $4 > 3$).

We can apply the same argument with $\log \zeta$ replaced by its negative derivative

$$-\operatorname{Re} \zeta'(s)/\zeta(s) = \sum_{n=1}^{\infty} \Lambda(n) n^{-\operatorname{Re}(s)} \cos(\operatorname{Im}(s) \log n)$$

to obtain an analogous inequality

$$-3 \operatorname{Re} \frac{\zeta'(\sigma)}{\zeta(\sigma)} - 4 \operatorname{Re} \frac{\zeta'(\sigma + it)}{\zeta(\sigma + it)} - \operatorname{Re} \frac{\zeta'(\sigma + 2it)}{\zeta(\sigma + 2it)} \geq 0 \qquad (\sigma > 1, t \in \mathbb{R}). \tag{20}$$

Let's see how to use (20) to get some information about zeroes just past the line $\operatorname{Re}(s) = 1$. We do this by bounding above each term on the left side of (20) for σ slightly bigger than 1. For starters, since ζ has a simple pole at $s = 1$,

$$-\frac{\zeta'(\sigma)}{\zeta(\sigma)} < \frac{1}{\sigma - 1} + *$$

where every $*$ in this argument is a positive constant, but no two need be the same.

Applying Hadamard's theorem and taking a logarithmic derivative, we get

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

Adjusting to get rid of the gamma factors, we get

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

For $1 \leq \operatorname{Re}(s) \leq 2$ and $|\operatorname{Im}(s)| \geq 1$, everything on the right side aside from the sum over ρ is dominated by $* \log |\operatorname{Im}(s)|$. Hence taking real parts, we obtain

$$-\operatorname{Re} \frac{\zeta'(s)}{\zeta(s)} < * \log |\operatorname{Im}(s)| - \sum_{\rho} \operatorname{Re} \left(\frac{1}{s - \rho} + \frac{1}{\rho} \right).$$

Since $\operatorname{Re}(\rho) > 0$ and $\operatorname{Re}(s - \rho) > 0$, we also have $\operatorname{Re}(1/\rho) > 0$ and $\operatorname{Re}(1/(s - \rho)) > 0$, so the sum over ρ is positive. Hence

$$-\operatorname{Re} \frac{\zeta'(s)}{\zeta(s)} < * \log |\operatorname{Im}(s)|;$$

this is the estimate I'll use for $s = \sigma + 2it$.

Let t be the imaginary part of a zero ρ of ζ; I will bound $-\operatorname{Re}\frac{\zeta'(s)}{\zeta(s)}$ for $s=\sigma+it$ by keeping only the summand corresponding to ρ. Namely, if $\rho=\beta+it$, then I get

$$-\operatorname{Re}\frac{\zeta'(s)}{\zeta(s)} < *\log|t| - \frac{1}{\sigma-\beta}.$$

From (20), I now deduce

$$\frac{4}{\sigma-\beta} < \frac{3}{\sigma-1} + *\log|t|.$$

For $\sigma = 1 + */(\log|t|)$, I can deduce

$$\beta < 1 - \frac{*}{\log|t|}.$$

In other words:

THEOREM 9.4. *There exists a constant $c>0$ such that there is no zero of ζ in the region* $\operatorname{Re}(s) \geq 1 - c/\log\operatorname{Im}(s), \operatorname{Im}(s) \geq 1$.

By von Mangoldt's formula (presented in the previous unit, with proof still to follow), this yields a nontrivial error bound in the prime number theorem, namely

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

(exercise).

3. What about L-functions?

The previous argument goes through more or less unchanged for L-functions. But there is a new complication: remember that we only looked at zeroes whose imaginary part was not too small. We took $|\operatorname{Im}(s)| \geq 1$, but the lower bound could have been any *fixed* positive constant.

The real issue is that while we can check once and for all that $\zeta(s)$ has no zeroes on the real line, we cannot rule this out for L-functions. But $L(s,\chi)$ could in principle have a real zero; such a hypothetical zero is called a *Siegel zero*. These can only occur for real nonprincipal characters.

Exercises

(1) Prove that $1/\Gamma$ is entire of order ≤ 1. Then prove that

$$\frac{1}{s\Gamma(s)} = e^{\gamma s}\prod_{n=1}^{\infty}(1+s/n)e^{-s/n} \qquad (s \neq 0,-1,-2,\dots),$$

where γ is Euler's constant, by applying Hadamard's theorem.

(2) Prove that

$$\frac{\Gamma'(s)}{\Gamma(s)} = \log(s) + O(|s|^{-1}) \qquad (|s|\to\infty, \operatorname{Re}(s)\geq 1/2).$$

(Hint: use the previous exercise.)

(3) Derive the estimate

$$|\Gamma(s/2)| < \exp(C_2|s|\log|s|) \qquad (\operatorname{Re}(s)\geq 1/2)$$

by first proving a suitably strong version of Stirling's formula, e.g.,

$$\log\Gamma(s) = \left(s-\frac{1}{2}\right)\log s - s + \frac{1}{2}\log 2\pi + O(|s|^{-1}) \qquad (|s|\to\infty, \operatorname{Re}(s)\geq 1/2).$$

(4) Prove that a function of order $\leq \alpha$ need not satisfy $|f(z)| = O(\exp(|z|^\alpha)$. (Hint: look at ζ on the positive real axis.)
(5) Find the constants A and B in the product representation for ξ given by Hadamard's theorem. Then deduce as a corollary that $\frac{\zeta'(0)}{\zeta(0)} = \log 2\pi$.
(6) Use the zero-free region and von Mangoldt's formula to prove that for some $c > 0$,

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

(By contrast, the leading term is $x \exp(-\log\log x)$.)

CHAPTER 10

von Mangoldt's formula

In this unit, we derive von Mangoldt's formula estimating $\psi(x) - x$ in terms of the critical zeroes of the Riemann zeta function. This finishes the derivation of a form of the prime number theorem with error bounds. It also serves as another good example of how to use contour integration to derive bounds on number-theoretic quantities; we will return to this strategy in the context of the work of Goldston-Pintz-Yıldırım.

1. The formula

First, let me recall the formula I want to prove. Again, ψ is the function

$$\psi(x) = \sum_{n<x} \Lambda(n) + \frac{1}{2}\Lambda(x),$$

where Λ is the von Mangoldt function (equaling $\log p$ if $n > 1$ is a power of the prime p, and zero otherwise).

THEOREM 10.1 (von Mangoldt's formula). *For $x \geq 2$ and $T > 0$,*

$$\psi(x) - x = -\sum_{\rho:|\operatorname{Im}(\rho)|<T} \frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)} - \frac{1}{2}\log(1 - x^{-2}) + R(x,T)$$

with ρ running over the zeroes of $\zeta(s)$ in the region $\operatorname{Re}(s) \in [0,1]$, and

$$R(x,T) = O\left(\frac{x\log^2(xT)}{T} + (\log x)\min\left\{1, \frac{x}{T\langle x\rangle}\right\}\right).$$

Here $\langle x\rangle$ denotes the distance from x to the nearest prime power other than possibly x itself.

2. Truncating a Dirichlet series

The basic idea is due to Riemann; it is to apply the following lemma to the Dirichlet series

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

(We will deduce this from Lemma 10.3 later.)

LEMMA 10.2. *For any $c > 0$,*

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

where the contour integral is taken along the line $\operatorname{Re}(s) = c$.

To pick out the terms with $n \leq x$, use the integral from Lemma 10.2 with $y = x/n$; this gives

$$\psi(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} -\frac{\zeta'(s)}{\zeta(s)} \frac{x^s}{s}\, ds.$$

What we want to do is shift the contour of integration to the left, to pick up the residues at the poles of the integrand. Remember that for f meromorphic, $\frac{1}{2\pi i}\frac{f'}{f}$ has a simple pole at each s which is a zero or pole of f, and the residue is the order of vanishing (positive for a zero, negative for a pole) of f at s. In particular, the integrand we are looking at has only simple poles: the only pole of x^s/s is at $s = 0$, which is not a zero or pole of ζ.

We now compute residues. The pole of ζ at $s = 1$ contributes x, and every zero ρ of ζ (counted with multiplicity) contributes $-x^\rho/\rho$. This includes the trivial zeroes, whose contributions add up to

$$\sum_{n=1}^{\infty} -\frac{x^{-2n}}{(-2n)} = -\frac{1}{2}\log(1 - x^{-2}).$$

The only pole of x^s/s is at $s = 0$, and it contributes $-\zeta'(0)/\zeta(0)$.

We thus pick up all of the main terms in von Mangoldt's formula by shifting from the straight contour $c - iT \to c + iT$ to the rectangular contour $c - iT \to -U - iT \to -U + iT \to c + iT$, then taking the limit as $U \to \infty$. (We do have to make sure that the new contour does not itself does not itself pass through any poles of the integrand!) To prove the formula, it thus suffices to prove that:

- the discrepancy between the integral $c - iT \to c + iT$ and the full vertical integral $c - i\infty \to c + i\infty$,
- the horizontal integrals $c \pm iT \to -\infty \pm iT$, and
- the limit as $U \to -\infty$ of the vertical integral $-U - iT \to -U + iT$

are all subsumed by the proposed bound on the error term $R(x, T)$.

3. Truncating the vertical integral

We first replace the infinite vertical integral in Lemma 10.2 with a finite integral, and estimate the error term.

LEMMA 10.3. *For $c, y, T > 0$, put*

$$I(y, T) = \frac{1}{2\pi i} \int_{c-iT}^{c+iT} y^s \frac{ds}{s},$$

with the integral taken along the straight contour, and

$$\delta(y) = \begin{cases} 0 & 0 < y < 1 \\ \frac{1}{2} & y = 1 \\ 1 & y > 1. \end{cases}$$

Then

$$|I(y, T) - \delta(y)| < \begin{cases} y^c \min\{1, T^{-1}|\log y|^{-1}\} & y \neq 1 \\ cT^{-1} & y = 1. \end{cases}$$

PROOF. I'll do the case $0 < y < 1$ to illustrate, and leave the others for you. Note that there are two separate inequalities to prove; we establish them using two different contours.

Since y^s/s has no poles in $\mathrm{Re}(s) > 0$, for any $d > 0$, we can write

$$\int_{c-iT}^{c+iT} y^s \frac{ds}{s} = \int_{c-iT}^{d-iT} y^s \frac{ds}{s} - \int_{c+iT}^{d+iT} y^s \frac{ds}{s} + \int_{d-iT}^{d+iT} y^s \frac{ds}{s},$$

in which each contour is straight. As $d \to \infty$, the integrand in the third integral converges uniformly to 0. We can thus write

$$\int_{c-iT}^{c+iT} y^s \frac{ds}{s} = \int_{c-iT}^{\infty-iT} y^s \frac{ds}{s} - \int_{c+iT}^{\infty+iT} y^s \frac{ds}{s}$$

and each of the two terms is dominated by

$$\frac{1}{T}\int_c^\infty y^t\, dt = y^c T^{-1} |\log y|^{-1}.$$

Since we must then divide by $2\pi > 2$, we get one of the claimed inequalities.

Now go back and replace the original straight contour with a minor arc of a circle centered at the origin. This arc has radius $R = \sqrt{c^2 + T^2}$, and on the arc the integrand y^s/s is dominated by y^c/R because $y < 1$. Thus the integral is dominated by $\pi R(y^c/R)$, and dividing by 2π yields the other claimed inequality. □

We will use Lemma 10.3 to show that

$$\int_{c-i\infty}^{c+i\infty} -\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}\, ds - \int_{c-iT}^{c+iT} -\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}\, ds = O\left(\frac{x(\log x)^2}{T} + (\log x)\min\left\{1, \frac{x}{T\langle x\rangle}\right\}\right).$$

By the lemma (applied with $y = x/n$), the left side is dominated by

$$\sum_{n=1, n\neq x}^{\infty} \Lambda(n)\left(\frac{x}{n}\right)^c \min\{1, T^{-1}|\log(n/x)|^{-1}\} + cT^{-1}\Lambda(x).$$

We get to choose any convenient value of c; it keeps the notation simple to take $c = 1 + (\log x)^{-1}$. Note that then $x^c = ex = O(x)$.

To estimate the summand, it helps to distinguish between terms where $\log(n/x)$ is close to zero, and those where it is bounded away from zero. For the latter, the quantity $|\log(n/x)|^{-1}$ is bounded above; so the summands with, say, $|n/x - 1| \geq 1/4$, are dominated by

$$O\left(xT^{-1}\left(-\frac{\zeta'(c)}{\zeta(c)}\right)\right) = O(xT^{-1}\log x).$$

For the former, consider values n with $3/4 < n/x < 1$ (the values with $1 < n/x < 5/4$ are treated similarly, and $n/x = 1$ contributes $O(\log x)$). Let x' be the largest prime power strictly less than x; then the summands $x' < n < x$ all vanish. In particular, it is harmless to assume $x' > 3x/4$, since otherwise the summands we want to bound all vanish.

We now separately consider the summand $n = x'$, and all of the summands with $3/4 < n < x'$. The former contributes

$$O\left(\log(x)\min\left\{1, \frac{x}{T(x-x')}\right\}\right).$$

For each term of the latter form, we can write $n = x' - m$ with $0 < m < x/4$, and

$$\log \frac{x}{n} \geq -\log\left(1 - \frac{m}{x'}\right) \geq \frac{m}{x'},$$

so these terms contribute

$$O\left(xT^{-1}(\log x)^2\right).$$

4. Shifting the contour

It remains to rewrite the integral

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

by shifting the contour and picking up residues. The new contour will be the three sides of the rectangle joining $c - iT, -U - iT, -U + iT, U + iT$ in that order, for suitable T and U.

We should choose U to be large and positive, so as to keep the vertical segment away from the trivial zeroes of ζ. Since those occur at negative even integers, we may simply take U to be a large *odd* positive integer.

It is a bit trickier to pick T. Note that we were actually given a value of T in the hypotheses of the theorem, but that T might be very close to the imaginary part of a zero of ζ. However, there is no harm in shifting T by a bounded amount: the sum over zeroes may change by the presence or absence of $O(\log T)$ terms each of size $O(xT^{-1}\log T)$, but we are allowing the error term to be as big as $O(xT^{-1}\log^2 T)$.

We now need to know how far away we can make T from the nearest zero, given that we can only shift by a bounded amount. This requires a slightly more refined count of zeroes than the one we gave before; see exercises.

LEMMA 10.4. *The number of zeroes of ζ with imaginary part in $[T, T+1]$ is $O(\log T)$.*

This means we can shift T so that the difference between it and the imaginary part of any zero of ζ is at least some constant times $(\log T)^{-1}$.

We will also need a truncated version of the product representation of ζ'/ζ; see exercises.

LEMMA 10.5. *For $s = \sigma + it$ with $-1 \leq \sigma \leq 2$ and t not equal to the imaginary part of any zero of ζ,*

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

where ρ runs over critical zeroes of ζ.

Putting these two lemmas together, we deduce that (after shifting T by a bounded amount) for s on the contour with $\mathrm{Re}(s) \in [-1, 2]$,

$$\frac{\zeta'(s)}{\zeta(s)} = O(\log^2 T).$$

Thus the integrals over the horizontal contours $c - iT \to -1 - iT$ and $-1 + iT \to c + iT$ are

$$O\left(\log^2 T \int_{-1}^{c} |x^s/s|\,ds\right) \leq O\left(\frac{x\log^2 T}{T\log x}\right),$$

which is subsumed by our proposed error bound.

It remains to bound the integrals over the rectangular contour $-1 - iT \to -U - iT \to -U + iT \to -1 + iT$. For this, we use the functional equation for ζ, in the form

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

Using a classical identity (one of Legendre's duplication formulas for Γ), we can rewrite this as

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

We want to bound the log derivative of the left side; it is equal to the sum of the log derivatives of the various factors on the right side. The first two factors give constants. The third gives a constant times $\tan(\pi s/2)$, which is bounded if we keep s at a bounded distance from any odd integer. The fourth gives $\Gamma'(s)/\Gamma(s)$, which we proved in a previous exercise is $O(\log |s|)$ as $|s| \to \infty$ if $\mathrm{Re}(s) \geq 1/2$. The fifth gives $\zeta'(s)/\zeta(s)$, which is bounded as $|s| \to \infty$ if $\mathrm{Re}(s) \geq 2$.

Putting it all together, we deduce that if s is kept at a bounded distance from any negative even integer, we have

$$\frac{\zeta'(s)}{\zeta(s)} = O(\log |s|) \qquad (|s| \to \infty, \mathrm{Re}(s) \leq -1).$$

Applying this along the remaining rectangular contour, we bound the horizontal contributions by

$$O\left(\int_1^\infty (\log s + \log T) x^{-s}/T \, ds\right) \leq O\left(\frac{1}{Tx \log^2 x} + \frac{\log T}{Tx \log x}\right),$$

which is subsumed by our error bound. We bound the vertical contribution in the limit as $U \to 0$ by

$$O\left(\frac{T \log U}{U x^U}\right),$$

which tends to zero. We are done!

Exercises

(1) Prove that for $T > 0$,

$$\sum_\rho \frac{1}{1 + (T - \mathrm{Im}(\rho))^2} = O(\log T),$$

where ρ runs over nontrivial zeroes of ζ. (Hint: this should have been on the previous handout. Go back to the proof of the zero-free region for ζ.)

(2) Deduce Lemma 10.4 from the previous exercise.

(3) Prove Lemma 10.5. (Hint: use the product representation for $\zeta'(s)/\zeta(s)$ evaluated at $s = \sigma + it$, then at $2 + it$, and subtract the two; everything left but the sum over ρ should be $O(\log |t|)$. Then use exercise 1 to control the contribution from the zeroes with $|t - \mathrm{Im}(\rho)| \geq 1$.) This can be used to derive a precise asymptotic for the number of zeroes of ζ in the critical strip with imaginary part in $(0, T)$:

$$\frac{T}{2\pi} \log \frac{T}{2\pi} - \frac{T}{2\pi} + O(\log T);$$

but I won't do so here. (See Davenport §15.)

(4) Check the remaining cases of Lemma 10.3. (You should do $y = 1$ by a direct calculation. In the case $y > 1$, you should shift contours in the opposite direction, picking up the pole at $s = 0$.)

CHAPTER 11

Error bounds in the prime number theorem in arithmetic progressions

In this unit, we summarize how to derive a form of the prime number theorem in arithmetic progressions with an appropriate uniformity in the modulus; many proofs are missing, and will not be included in this course. We will revisit this uniformity later in the Bombieri-Vinogradov theorem.

1. Uniformity in the explicit formula

Let χ be a primitive Dirichlet character of level N.

LEMMA 11.1. *The number of critical zeroes of $L(s,\chi)$ with imaginary part in $[T, T+1]$ is $O(\log(NT))$, where the implied constant is absolute (i.e., it does not depend on N).*

Put

$$\psi(x,\chi) = \sum_{n \leq x} \chi(n)\Lambda(n).$$

Then as long as x is an integer and $T \leq x$, the explicit formula for $\psi(x,\chi)$ has the form

$$\psi(x,\chi) = -\sum_{\rho : |\operatorname{Im}(\rho)| < T} \frac{x^\rho}{\rho} - b(\chi) + O(xT^{-1}\log^2(Nx)),$$

where $b(\chi)$ is defined by

$$\frac{L'(s,\chi)}{L(s,\chi)} = \begin{cases} s^{-1} + b + O(s) & \chi(-1) = 1 \\ b + O(s) & \chi(-1) = -1. \end{cases}$$

The proof is as for von Mangoldt's formula; I will not redo it here. The point is that everything is uniform in N *except* the constant $b(\chi)$.

So to get uniform estimates, one must control $b(\chi)$, which we can do by expressing it in terms of zeroes of $L(s,\chi)$. For this we go back to the product expansion:

$$\frac{L'(s,\chi)}{L(s,\chi)} = -\frac{1}{2}\log(N/\pi) - \frac{1}{2}\frac{\Gamma'(s/2+a/2)}{\Gamma(s/2+a/2)} + B(\chi) + \sum_\rho \left(\frac{1}{s-\rho} + \frac{1}{\rho}\right),$$

where $a = 0$ if χ is even and $a = 1$ if χ is odd. The constant B here is not the same as b (it includes the contribution from the exponential part of the Hadamard product expansion), but no matter; we can eliminate it by comparing a given s with $s = 2$. Hence

$$\frac{L'(s,\chi)}{L(s,\chi)} = O(1) - \frac{1}{2}\frac{\Gamma'(s/2+a/2)}{\Gamma(s/2+a/2)} + \sum_\rho \left(\frac{1}{s-\rho} - \frac{1}{2-\rho}\right).$$

If $a = 1$, everything is regular near $s = 0$; if $a = 0$, the two log derivatives both have a simple pole at $s = 0$, and the residues match. We can thus equate the constant terms of the expansions around $s = 0$, to obtain

$$b(\chi) = O(1) - \sum_{\rho} \left(\frac{1}{\rho} + \frac{1}{2-\rho} \right).$$

As happened with ζ, we can bound the contribution of the zeroes with $|\operatorname{Im}(\rho)| \geq 1$ by $O(\log N)$. The same goes for the term $1/(2-\rho)$ when $|\operatorname{Im}(\rho)| \leq 1$.

This means that we have

$$\psi(x, \chi) = - \sum_{\rho : |\operatorname{Im}(\rho)| < T} \frac{x^\rho}{\rho} + \sum_{\rho : |\operatorname{Im}(\rho)| < 1} \frac{1}{\rho} + O(xT^{-1} \log^2(Nx)).$$

2. Controlling the exceptional zeroes

The question that now remains is: how big is the contribution to $\psi(x, \chi)$ from the sum of $1/\rho$ over critical zeroes $L(s, \chi)$ in the range $|\operatorname{Im}(\rho)| < 1$?

To answer this, we need a uniform zero-free region near the real axis. Here's what happens when you try to produce this.

THEOREM 11.2. *There is a constant $c > 0$ such that there is no zero of $L(s, \chi)$ with*

$$|\operatorname{Im}(\rho)| < 1, \qquad \operatorname{Re}(\rho) > 1 - \frac{c}{\log N}$$

except *perhaps if χ is real, in which case there may be* one *such zero (counting multiplicity), necessarily real.*

The Riemann Hypothesis for $L(s, \chi)$ implies that no such oddball zero exists, but to prove unconditional results we must allow for it. In particular, it helps to have a label for the hypothetical zero: we call it an *exceptional zero*, or *Siegel zero*, of $L(s, \chi)$. (Note that the criterion for being a Siegel zero depends on the choice of the cutoff parameter c.)

If we exclude any exceptional zero β and also its mirror image $1 - \beta$, then the sum of $1/\rho$ over the remaining zeroes in the range $|\operatorname{Im}(\rho)| < 1$ is $O((\log N)^2)$, since there are $O(\log N)$ such zeroes and each term contributes $O(\log N)$ to the sum. We then have

$$\psi(x, \chi) = - \sum_{\rho : |\operatorname{Im}(\rho)| < T}^{\sim} \frac{x^\rho}{\rho} - \frac{x^\beta}{\beta} - \frac{x^{1-\beta} - 1}{1-\beta} + O(xT^{-1} \log^2(Nx)),$$

where the tilde means don't count β and $1-\beta$ as zeroes. The term $(x^{1-\beta}-1)/(1-\beta)$ is $O(x^c \log x)$, but controlling the term x^β/β requires preventing the exceptional zero from getting too close to 1. Here's one way to do that.

THEOREM 11.3 (Siegel). *For any $\epsilon > 0$, there exists a constant $c = c(\epsilon)$ with the following property: for any real primitive Dirichlet character χ of level N, every real zero β of $L(s, \chi)$ satisfies*

$$\beta \leq 1 - cN^{-\epsilon}.$$

(The proof of this uses the previous theorem; the idea is to show that if you have an exceptional zero for one real character, it "repels" real zeroes for other characters.)

This is enough to get the following form of the prime number theorem in arithmetic progressions with error term, called the Siegel-Walfisz theorem. For N a positive integer and a an integer coprime to N, put

$$\pi(x, N, a) = \sum_{p \leq x, p \equiv a(N)} 1$$
$$\psi(x, N, a) = \sum_{n \leq x, n \equiv a(N)} \Lambda(n).$$

THEOREM 11.4. *Fix $\epsilon > 0$ and $A > 0$. Then*

$$\pi(x, N, a) = \frac{\mathrm{li}(x)}{\phi(N)} + O(x \log^{-A} x)$$
$$\psi(x, N, a) = \frac{x}{\phi(N)} + O(x \log^{-A} x)$$

where the implied and explicit constants depend only on A and ϵ, not on N.

One can improve this error bound a bit unconditionally, but not much. On the other hand, under the *Generalized Riemann Hypothesis* (i.e., the critical zeroes of every $L(s, \chi)$ lie on the line $\mathrm{Re}(s) = 1/2$), you get errors of $O(x^{1/2+\epsilon})$.

With the error bound as is, the theorem only has content for N no bigger than a fixed power of $\log x$. You can prove much better results, say for N up to x^C for a fixed $c < 1/2$, if you are willing to accept an *average* statement about the error bounds. More on this when we discuss the Bombieri-Vinogradov theorem.

3. Why the exceptional zero?

One can see where the possibility of an exceptional zero arises by beginning to imitate for $L(s, \chi)$ the proof we gave of the zero-free region for ζ. We have

$$-\frac{L'(s, \chi)}{L(s, \chi)} = \sum_{n=1}^{\infty} \Lambda(n) n^{-\mathrm{Re}(s)} \chi(n) e^{-i \,\mathrm{Im}(s) \log n},$$

and using the trigonometric inequality, we have for $\sigma > 1$

$$-3\frac{L'(\sigma, \chi_0)}{L(\sigma, \chi_0)} - 4\,\mathrm{Re}\,\frac{L'(\sigma + it, \chi)}{L(\sigma + it, \chi)} - \mathrm{Re}\,\frac{L'(\sigma + 2it, \chi^2)}{L(\sigma + 2it, \chi)} \geq 0.$$

Here χ_0 is the principal character of the same level as χ.

The argument to exclude zeroes close to the edge of the critical strip proceeds as before if $\mathrm{Im}(\rho)$ is bounded away from 0, say $|\,\mathrm{Im}(\rho)| > c/(\log N)$. For χ nonreal, you do better: χ^2 is nonprincipal and so $L(\sigma + 2it, \chi^2)$ stays bounded as $\sigma \to 1^+$. So you get an inequality of the form

$$\frac{4}{\sigma - \mathrm{Im}(\rho)} < \frac{3}{\sigma - 1} + O(\log N + \log(|\,\mathrm{Im}(\rho)| + 1)),$$

and that gives you a zero-free region all the way down to the real line.

Unfortunately, if χ is real, then $L(\sigma + 2it, \chi^2)$ blows up at $\chi = 1$, and our present methods cannot exclude a single zero very close to 1: you only end up with an inequality of the form

$$\frac{4}{\sigma - \mathrm{Im}(\rho)} < \frac{3}{\sigma - 1} + \mathrm{Re}\left(\frac{1}{\sigma - 1 + 2i\,\mathrm{Im}(\rho)}\right) + O(\log N + \log(|\,\mathrm{Im}(\rho)| + 1)).$$

However, we can exclude two such zeroes ρ_1, ρ_2, by writing

$$-\frac{L'(s,\chi)}{L(s,\chi)} < -\frac{1}{\sigma-\rho_1} - \frac{1}{\sigma-\rho_2} + O(\log N + \log(|\operatorname{Im}(\rho)|+1))$$

and so on.

Part 3

Sieve methods

CHAPTER 12

Revisiting the sieve of Eratosthenes

This unit begins the second part of the course, in which we will investigate a class of methods in analytic number theory known as *sieves*. (For non-native speakers of English: in ordinary life, a *sieve* is a device through which you pour a powder, like flour, to filter out large impurities.) Whereas the first part of the course leaned heavily on methods from complex analysis, here the emphasis will be more combinatorial.

1. The Sieve of Eratosthenes

The original sieve is of course the Sieve of Eratosthenes for finding prime numbers. To find the prime numbers in $\{2, \ldots, n\}$, you repeat the following operation as long as there are unmarked numbers: find the first unmarked number p, mark it as prime, then mark $2p, 3p, \ldots$ as composite until you get to a number greater than n.

Of course, one need only sift out multiples of primes up to $n^{1/2}$ in order to leave only primes behind. More generally, if one is only able to sift out multiples of primes up to n^α, what remain are numbers with no prime factors less than n^α. In particular, any such number has at most $\lfloor \alpha^{-1} \rfloor$ prime factors, and so is in some sense "nearly prime".

Of course, in the process of sieving, many numbers will be sifted out more than once. If one wants to draw any sort of quantitative conclusion from this process, one must keep track of the multiple counting; this suggests using inclusion-exclusion.

2. The principle of inclusion-exclusion

Let S be a finite set, and let $P_1, \ldots, P_n$ be subsets of S. Think of each P_i as containing the elements of S with a certain property.

Suppose we have some way to count the number of elements in the intersection of any subcollection of the P_i, but what we really want is to count the complement of the union of all of the P_i. The formula that computes this is:

$$\#(S \setminus (P_1 \cup \cdots \cup P_n)) = \sum_{T \subseteq \{1,\ldots,n\}} (-1)^{\#T} \# \left(\bigcap_{t \in T} P_t \right).$$

Proof: if $s \in S$ belongs to m of the subsets, then the number of times it gets counted on the right side is

$$\binom{m}{0} - \binom{m}{1} + \cdots$$

which equals 1 if $m = 0$ and vanishes otherwise.

More generally, if $f : S \to \mathbb{C}$ is some function, and we want to compute the sum of f over the complement of the P_i, we have

$$\sum_{s \in S \setminus (P_1 \cup \cdots \cup P_n)} f(s) = \sum_{T \subseteq \{1,\ldots,n\}} (-1)^{\#T} \left(\sum_{s \in \cap_{t \in T} P_t} f(s) \right).$$

In number theory, we are often taking $S = \{1, \ldots, N\}$ and taking the sets $P_1, P_2, \ldots$ to be the sets of multiples of certain small primes. It is convenient to rewrite the principle of inclusion-exclusion in terms of the arithmetic function μ, the Möbius function:

$$\mu(n) = \begin{cases} (-1)^d & n = p_1 \cdots p_d \quad (p_1, \ldots, p_d \text{ distinct}, d \geq 0) \\ 0 & \text{otherwise.} \end{cases}$$

3. Smooth numbers

Before proceeding, I need a quick lemma concerning smooth numbers. A natural number is *z-smooth* if its prime factors are all less than or equal to z.

LEMMA 12.1 (Rankin). *Let $\Phi(x, z)$ be the number of z-smooth numbers less than or equal to x. Then for any $\delta > 0$,*

$$\Phi(x, z) \leq x^\delta \prod_{p \leq z} (1 - p^{-\delta})^{-1}.$$

PROOF. If we expand the right side as a product of geometric series, we get a term $(x/n)^\delta \geq 1$ for each z-smooth number $n \leq x$ (among other terms). This yields the claim. □

4. Back to Eratosthenes

Here is a modern version of the Sieve of Eratosthenes, following Murty and Saradha. Let A be a set of natural numbers, and let P be a set of primes; also set

$$P(z) = \prod_{p \in P, p \leq z} p.$$

For each $p \in P$, choose a set R_p consisting of some number $\omega(p)$ of residue classes modulo p, and let A_p be the subset of A whose elements belong to the chosen residue classes. Put

$$W(z) = \prod_{p | P(z)} \left(1 - \frac{\omega(p)}{p}\right),$$

For d squarefree with all prime factors in P, put $\omega(d) = \prod_{p|d} \omega(p)$ and $A_d = \cap_{p|d} A_p$.

We wish to estimate $S(A, P, z)$, the number of elements of A not belonging to A_p for any $p \leq z$. For this, we must assume some good properties about the chosen residue classes. For starters, we want that for some $\kappa > 0$,

$$\sum_{p \leq z, p \in P} \frac{\omega(p) \log p}{p} \leq \kappa \log z + O(1), \tag{21}$$

where the big-O bound is for $z \to \infty$ and the constant depends on P, R_p, κ.

LEMMA 12.2. *Assuming* (21), *we have*

$$\sum_{d<t,d|P(z)} \omega(d) = O\left(t(\log z)^\kappa \exp\left(-\frac{\log t}{\log z}\right)\right),$$

where the big-O bound is for $z \to \infty$ and the constant depends on P, R_p, κ.

PROOF. Exercise. □

LEMMA 12.3. *Fix $C > 0$. Assuming* (21), *we have*

$$\sum_{d>Cx,d|P(z)} \frac{\omega(d)}{d} = O\left((\log z)^{\kappa+1} \exp\left(-\frac{\log x}{\log z}\right)\right),$$

where the big-O bound is for $z \to \infty$ and the constant depends on P, R_p, κ, C.

PROOF. Put $F_\omega(t,z) = \sum_{d<t,d|P(z)} \omega(d)$. Then

$$\sum_{d>Cx,d|P(z)} \frac{\omega(d)}{d} \leq \int_{Cx}^{\infty} \frac{F_\omega(t,z)}{t^2} dt \tag{22}$$

(exercise), so the result follows from Lemma 12.2. □

THEOREM 12.4. *Fix P, R_p, κ satisfying* (21), *and also fix $C, c > 0$. Then for any set A and any $X, x > 0$ such that*

$$\left|\#A_d - \frac{\omega(d)}{d} X\right| \leq c\omega(d)$$

and $\#A_d = 0$ for $d > Cx$, we have

$$S(A,P,z) = XW(z) + O\left(x \log^{\kappa+1} z \exp\left(-\frac{\log x}{\log z}\right)\right),$$

where the big-O bound is for $z \to \infty$, uniformly in A, x, X.

PROOF. Exercise. □

5. Motivation: the twin prime conjecture

The *twin prime conjecture* states that there are infinitely many primes p such that $p+2$ is also prime. One can even guess the correct asymptotic up to a constant factor, by a very simple argument: since the probability of a random number in $[1, \dots, N]$ being prime is asymptotically $1/\log N$, the number of twin primes in $[1, \dots, N]$ should be asymptotic to $N/\log^2 N$. (Getting the constant right is a bit trickier; I won't deal with that just now.)

As a corollary of Theorem 12.4, we obtain the following result of Brun (with a slightly simpler proof).

THEOREM 12.5. *The number of primes $p \leq x$ such that $p+2$ is also prime is $O(x(\log\log x)^2/(\log x)^2)$.*

PROOF. We will apply Theorem 12.4 with $A = \{1, \dots, x\}$ and $P = \{p : 2 < p \leq z\}$. For each $p \in P$, let R_p consist of the residue classes of $0, -2$, so that $\omega(p) = 2$. For d odd squarefree, $\omega(d) = 2^{\nu(d)}$ for $\nu(d)$ the number of prime factors of d. One checks easily (exercise) that

$$\left|\#A_d - x\frac{\omega(d)}{d}\right| \leq 2^{\nu(d)}. \tag{23}$$

Since

$$\sum_{p \leq z} \frac{\log p}{p} = O(\log z)$$

from a prior homework, we can take $\kappa = 2$ in Theorem 12.4. This yields

$$S(A, P, z) = xW(z) + O\left(x \log^3 z \exp\left(-\frac{\log x}{\log z}\right)\right),$$

where the big-O constant does not depend on x or z. We now take

$$\log z = \frac{\log x}{A \log \log x}$$

for a suitable constant A. Since

$$W(z) \leq \prod_{3 \leq p \leq z} \left(1 - \frac{1}{p}\right)^2 = O((\log z)^{-2})$$

by a prior homework exercise, we deduce that $S(A, P, z) = O(x(\log \log x)^2/(\log x)^2)$.

To conclude, note that $S(A, P, z)$ includes all primes $z + 2 \leq p \leq x$ such that $p + 2$ is also prime. The number of twin primes up to x that we missed is at most $z = x^{1/(A \log \log x)}$, so this doesn't affect the claim. □

We will get a sharper result using Selberg's sieve in a subsequent lecture.

Exercises

(1) Prove Lemma 12.2 using Rankin's trick.
(2) Prove (22).
(3) Prove Theorem 12.4.
(4) Prove (23).
(5) (Brun) Prove that the sum of the reciprocals of the twin primes converges.
(6) Prove that

$$\Phi(x, z) = O\left(x \log z \exp\left(-\frac{\log x}{\log z}\right)\right)$$

where the big-O bound is for $z \to \infty$, uniformly in x. (Hint: apply Rankin's lemma with $\delta = 1 - (\log z)^{-1}$.)

(7) Prove that the number of squarefree integers in $\{1, \ldots, N\}$ is

$$\frac{6}{\pi^2} N + O(N^{1-\epsilon})$$

for some explicit value of ϵ. (Hint: this is much easier than sieving over primes! Just make sure to round round no more than $O(N^{1-\epsilon})$ fractions off to the nearest integer. Also, don't forget that $6/\pi^2 = 1/\zeta(2) = \prod_p (1 - 1/p^2)$.)

CHAPTER 13

Brun's combinatorial sieve

In this unit, we describe a more intricate version of the sieve of Eratosthenes, introduced by Viggo Brun in order to study the Goldbach conjecture and the twin prime conjecture. It is most useful for providing lower bounds; for upper bounds, the Selberg sieve (to be introduced in the following unit) is much less painful.

1. Sieve setup

Let $f : \mathbb{N} \to \mathbb{C}$ be an arithmetic function, and suppose we want to estimate the sum of f over primes. More precisely, let P be a set of primes, and put

$$P(z) = \prod_{p \leq z, p \in P} p.$$

If we define

$$S(x,z) = \sum_{n \leq x, (n, P(z))=1} f(n),$$

$$A_d(x) = \sum_{n \leq x, n \equiv 0 \pmod{d}} f(n)$$

(with the dependence on P and f suppressed from the notation), we have

$$S(x,z) = \sum_{d | P(z)} \mu(d) A_d(x).$$

As before, suppose there is a multiplicative function g such that for d squarefree with all prime factors in P,

$$A_d(x) = g(d)X + r_d(x),$$

with $X = X(x)$ independent of d, and the error term $r_d(x)$ small when d is small relative to x (in a sense to be made precise later). Suppose further that

$$g(p) \in [0,1) \quad (p \in P); \qquad g(p) = 0 \quad (p \notin P). \tag{24}$$

(If we need to take $g(p) = 1$, then we cannot expect to get much of a contribution from numbers not divisible by p; we should resign ourselves to this, and instead remove p from P.) Then we can rewrite

$$S(x,z) = V(z)X + R(x,z)$$

$$V(z) = \prod_{p | P(z)} (1 - g(p))$$

$$R(x,z) = \sum_{d | P(z)} \mu(d) r_d(x).$$

If z is small relative to x, which in practice will mean $z < x^\alpha$ for some cutoff $\alpha \in (0,1)$, we may be able to show that the main term $V(z)X$ dominates the error term $R(x,z)$. Again, the main term is what you would predict from the heuristic that if an integer is chosen randomly, its divisibilities by different primes should act like independent random events.

For instance, if P is the set of all primes and $z \geq x^{1/2}$, then $S(x,z) = \sum_{p \leq x} f(p)$. If f is the function

$$f(n) = \begin{cases} 1 & n-2 \text{ prime} \\ 0 & \text{otherwise,} \end{cases}$$

then by the error term in the prime number theorem for arithmetic progressions,

$$r_d(x) = O(x \log^{-A} x)$$

for any fixed $A > 0$. (It is now important that we have that bound uniformly in d!) Also, $S(x, x^{1/2})$ counts twin primes up to x, whereas $S(x, x^{1/(N+1)})$ counts primes p such that $p+2$ has no prime factor less than $x^{1/(N+1)}$, and hence has at most N prime factors.

2. Brun's combinatorial sieve

We would like somewhat finer control than was provided by the sieve of Eratosthenes; the trouble is that $R(x,z)$ has too many terms for us to be able to control it.

Brun's approach to get aronud this is to truncate the Möbius function by restricting it to suitable subsets D^+ and D^-, subject to the restriction that for n a product of primes in P, the incomplete convolutions

$$\delta^+(n) = \sum_{d|n, d \in D^+} \mu(d), \qquad \delta^-(n) = \sum_{d|n, d \in D^-} \mu(d)$$

satisfy

$$\delta^-(n) \leq \delta(n) \leq \delta^+(n) \tag{25}$$

for

$$\delta(n) = \sum_{d|n} \mu(d) = \begin{cases} 1 & n = 1 \\ 0 & n > 1. \end{cases}$$

One such choice would be to take D^+ and D^- to consist of all squarefree numbers whose number of distinct prime factors is even or odd, respectively. This choice is much too crude; we should instead make a choice that allows some cancellation in δ^- and δ^+ without messing up the inequality (25). Moreover, we want to restrict D^+ and D^- to be subsets of $\{1, \dots, y\}$ for some y which is not too large compared to x.

Let $\lambda^+(d)$ and $\lambda^-(d)$ denote the functions which agree with μ on D^+ and D^-, respectively, and are zero elsewhere. Put

$$V^\pm(z) = \sum_{d|P(z)} \lambda^\pm(d) g(d)$$

$$R^\pm(x,z) = \sum_{d|P(z)} \lambda^\pm(d) r_d(x).$$

Then by virtue of (25), we have

$$V^-(z)X + R^-(x,z) \le S(x,z) \le V^+(z)X + R^+(x,z). \tag{26}$$

It is not at all obvious how one can usefully arrange for D^+, D^- to satisfy (25); here is Brun's choice. For d a squarefree positive integer, write $d = p_1 \cdots p_r$ with $p_1 > \cdots > p_r$. Set

$$D^+ = \{d = p_1 \cdots p_r : p_m < y_m \quad m \text{ odd}\}$$
$$D^- = \{d = p_1 \cdots p_r : p_m < y_m \quad m \text{ even}\},$$

where $y_1, y_2, \ldots$ are certain parameters which may depend on d. (By convention, $1 \in D^{\pm}$.) We then have the following.

LEMMA 13.1. *With notation as above, let $V_n(z)$ be the sum of $g(p_1 \cdots p_n)V(p_n)$ over sequences $p_1 > \cdots > p_n$ of primes such that:*

(a) $p_1 < z$;
(b) $p_n \ge y_n$;
(c) $p_m < y_m$ *for* $m < n$ *with* $m \equiv n \pmod 2$.

Then

$$V(z) = V^+(z) - \sum_{n \equiv 1\,(2)} V_n(z)$$
$$V(z) = V^-(z) + \sum_{n \equiv 0\,(2)} V_n(z)$$

and so

$$V^-(z) \le V(z) \le V^+(z). \tag{27}$$

PROOF. Exercise. □

In particular, for a given n, we deduce (25) from (27) by rigging up the set P so that $P(z) = n$ and putting $g(d) = 1$ for all d.

The functions λ^+ and λ^- given above are together called the *combinatorial sieve* with parameters $y_1, y_2, \ldots$. To use it, one must bound

$$R(x,y) = \sum_{d<y, d|P(z)} |r_d(x)|,$$

for y such that $D^{\pm} \subset \{1, \ldots, y\}$; in this case $R(x,y) \ge |R^{\pm}(x,z)|$, giving error bounds in (26). One must also bound $V^{\pm}(z)$.

3. Setting some parameters

To turn this into an actual numerical theorem, we must set the sieve parameters; we do this following Iwaniec-Kowalski. Remember that we may allow the y_i to depend on d.

Write $d = p_1 \cdots p_r$ with $p_1 > \cdots > p_r$; we now take

$$y_m = (y/(p_1 \cdots p_m))^{1/\beta},$$

where $\beta > 1$ will be specified later. This makes it clear that all elements of $D^+ \cup D^-$ belong to $\{1, \ldots, y\}$ except possibly for single primes in D^-. We can remedy this by requiring $z \le y$; more precisely, we will take $z = y^{1/s}$ for some $s \ge \beta$.

We will also need to make some restriction on the multiplicative function g. Namely, we assume that for some $K > 1$ and $\kappa > 0$, we have for all w, z,

$$\prod_{w \le p < z} (1 - g(p))^{-1} \le K \left(\frac{\log z}{\log w} \right)^{\kappa}. \tag{28}$$

We refer to κ as a *sieve dimension* of the function g. This number is quite critical; it will determine how large we can make z compared to y, which determines how many small primes we can use for sieving.

4. Bounding the main term

We need an upper bound on $V^+(z)$ and a lower bound on $V^-(z)$; we get both of these by getting an upper bound on $V_n(z)$. First, let us simplify the sum by relaxing the summation conditions. We claim that for any tuple $p_1, \dots, p_n$ appearing in the sum defining $V_n(z)$, and any $m < n$,

$$p_1 \cdots p_{m-1} p_m^{\beta} < y. \tag{29}$$

Namely, if $m \equiv n \pmod 2$, we have the stronger inequality

$$p_1 \cdots p_{m-1} p_m^{1+\beta} < y.$$

If $m > 1$ and $m \not\equiv n \pmod 2$, we have

$$p_1 \cdots p_{m-1} p_m^{\beta} < p_1 \cdots p_{m-2} p_{m-1}^{1+\beta} < y.$$

Finally, if $m = 1$ and $m \not\equiv n \pmod 2$, we have

$$p_1^{\beta} < z^{\beta} = y^{\beta/s} \le y.$$

From (29), we deduce by induction on m that

$$p_1 \cdots p_m < y^{1-(1-\beta^{-1})^m} \qquad (m = 1, \dots, n-1).$$

In particular,

$$p_n \ge (y/(p_1 \cdots p_{n-1}))^{1/(\beta+1)} \ge y^{\frac{1}{\beta+1}(1-\beta^{-1})^{n-1}} \ge y^{\frac{1}{\beta}(1-\beta^{-1})^n} \ge z_n$$

if we put

$$z_n = z^{(1-\beta^{-1})^n}.$$

We will now retain only the conditions $z > p_1 > \cdots > p_n \ge z_n$ on the primes, which will make the sum bigger because every summand is nonnegative. That is,

$$\begin{aligned} V_n(z) &\le \sum_{z > p_1 > \cdots > p_n \ge z_n} g(p_1 \cdots p_n) V(p_n) \\ &\le \frac{1}{n!} V(z_n) \left(\sum_{z_n \le p < z} g(p) \right)^n \\ &\le \frac{1}{n!} V(z_n) \left(\log \frac{V(z_n)}{V(z)} \right)^n. \end{aligned}$$

Here is where we need the assumption (28) about the sieve dimension. It implies

$$\frac{V(z_n)}{V(z)} \le K(1 + (\beta - 1)^{-1})^{\kappa n} < K e^{n/b}$$

for $\beta = \kappa b + 1$ (using the bound $1 + x \le e^x$ for $x = (\beta - 1)^{-1} = 1/(\kappa b)$), which gives us

$$\begin{aligned} V_n(z) &< \frac{K}{n!}\left(\frac{n}{b} + \log K\right)^n e^{n/b} V(z) \\ &\le \frac{K}{n!}\left(\frac{n}{b} e^{1/b}\right)^n K^b V(z) \end{aligned}$$

(using the bound $1 + x \le e^x$ for $x = b(\log K)/n$). Since $n! \ge e(n/e)^n$ (by taking logs and comparing integrals), we obtain

$$V_n(z) < e^{-1} a^n K^{b+1} V(z)$$

for $a = b^{-1} e^{1+b^{-1}}$.

To conclude, we clean things up a bit. Remember that we were at liberty to choose $\beta > 1$, which is equivalent to choosing $b > 0$. By taking b sufficiently large, we can force $a < 1$; for instance, we could take $b = 9$ to get $a < e^{-1}$. Note also that because

$$p_1 > \cdots > p_n \ge y_n = (y/(p_1 \cdots p_n))^{1/\beta},$$

we have $p_1^{n+\beta} > y$. Since we also have $p_1 < z = y^{1/s}$, we deduce that $V_n(z) = 0$ unless $n + \beta > s$. Therefore

$$\sum_{n>0} V_n(z) = \sum_{n > s-\beta} V_n(z) < \frac{a^{s-\beta}}{e(1-a)} K^{b+1} V(z).$$

To conclude, we have the following bound (Theorem 6.1 in Iwaniec-Kowalski).

THEOREM 13.2. *In the combinatorial sieve with parameters $y_1, y_2, \ldots$ as above, and $\beta = 9\kappa + 1$, for any multiplicative function $g(d)$ satisfying* (24) *and* (28) *for a given K, and any $s \ge \beta$, for $z = y^{1/s}$ we have*

$$\begin{aligned} V^+(z) &< (1 + e^{\beta - s} K^{10}) V(z) \\ V^-(z) &> (1 + e^{\beta - s} K^{10}) V(z). \end{aligned}$$

Consequently,

$$(1 - e^{\beta-s} K^{10}) V(z) X - R(x, z^s) \le S(x, z) \le (1 + e^{\beta-s} K^{10}) V(z) X + R(x, z^s).$$

5. Consequences for twin almost-primes

Again consider the example

$$f(n) = \begin{cases} 1 & n - 2 \text{ prime} \\ 0 & \text{otherwise.} \end{cases}$$

By applying the combinatorial sieve, we may deduce the following (see exercises).

THEOREM 13.3. *There are infinitely many primes p such that $p + 2$ is the product of at most twenty distinct primes.*

By refinements of the sieving method, Chen was able to prove the following.

THEOREM 13.4. *There are infinitely many primes p such that $p + 2$ is the product of at most two distinct primes.*

This is tantalizingly close to the twin prime conjecture, but it seems that sieving methods fall short of delivering that particular prize.

One can also use the combinatorial sieve to deduce that the number of twin primes $\leq x$ is $O(x/\log^2 x)$; however, since this is a question about an upper bound rather than a lower bound, we will be able to derive this much less painfully using the Selberg sieve.

Exercises

(1) Prove Lemma 13.1. (Hint: use the identity

$$V(z) = 1 - \sum_{p<z} g(p)V(p)$$

plus inclusion-exclusion.)

(2) Apply the combinatorial sieve to show that the number of integers less than or equal to x with no prime factors less than $x^{1/20}$ is at least $cx/\log^2 x$ for some $c > 0$. (You will need the prime number theorem in arithmetic progressions with error term, in order to control the error term $R(x, z)$.) Then deduce Theorem 13.3.

CHAPTER 14

The Selberg sieve

1. Review of notation

Let $f : \mathbb{N} \to \mathbb{C}$ be an arithmetic function, and suppose we want to estimate the sum of f over primes. More precisely, let P be a set of primes, and put

$$P(z) = \prod_{p \leq z, p \in P} p.$$

If we define

$$S(x,z) = \sum_{n \leq x, (n, P(z))=1} f(n),$$
$$A_d(x) = \sum_{n \leq x, n \equiv 0 \pmod{d}} f(n)$$

(with the dependence on P and f suppressed from the notation), we have

$$S(x,z) = \sum_{d | P(z)} \mu(d) A_d(x).$$

Let $g(d)$ be a multiplicative function with

$$g(p) \in [0,1) \qquad (p \in P);$$
$$g(p) = 0 \qquad (p \notin P),$$

and write

$$A_d(x) = g(d)x + r_d(x).$$

Then

$$S(x,z) = V(z)x + R(x,z)$$
$$V(z) = \prod_{p | P(z)} (1 - g(p))$$
$$R(x,z) = \sum_{d | P(z)} r_d(x).$$

2. The Selberg upper bound sieve

In the previous unit, we used the combinatorial sieve to construct an arithmetic function $\lambda^+ : \mathbb{N} \to \mathbb{R}$ such that

$$\lambda^+(1) = 1$$
$$\sum_{d | n} \lambda^+(d) \geq 0 \quad (n > 1).$$

By setting

$$V^+(z) = \sum_{d|P(z)} \lambda^+(d) g(d)$$
$$R^+(x,z) = \sum_{d|P(z)} \lambda^+(d) r_d(x),$$

we were able to obtain the bound

$$V^-(z)x + R^-(x,z) \leq S(x,z) \leq V^+(z)x + R^+(x,z), \tag{30}$$

but controlling V^+ and R^+ was rather painful.

Selberg proposed instead to construct an arithmetic function $\rho : \mathbb{N} \to \mathbb{R}$ with $\rho(1) = 1$ and

$$\sum_{d|n} \lambda^+(n) = \left(\sum_{d|n} \rho(d) \right)^2.$$

In other words, let ρ be any arithmetic function with $\rho(1) = 1$, and put

$$\lambda^+(n) = \sum_{d_1,d_2 : \mathrm{lcm}(d_1,d_2)=n} \rho(d_1)\rho(d_2).$$

We will typically want $\lambda^+(d) = 0$ for $d \geq y$, for some prespecified number y; to enforce this, we may insist that $\rho(n) = 0$ for $n \geq \sqrt{y}$. We call the resulting λ^+ an *L^2-sieve of level y*, or more commonly a *Selberg (upper bound) sieve of level y*.

Let us drop x from consideration by agreeing to only consider functions f with finite support. (That is, we replace f by the function vanishing above x.) If we again set

$$S(z) = \sum_{(n,P(z))=1} f(n)$$
$$V^+(z) = \sum_{d|P(z)} \lambda^+(d) g(d)$$
$$= \sum_{d_1,d_2|P(z)} \rho(d_1)\rho(d_2) g(\mathrm{lcm}(d_1,d_2))$$

$$R^+(z) = \sum_{d|P(z)} \lambda^+(d) r_d(x)$$
$$= \sum_{d_1,d_2|P(z)} \rho(d_1)\rho(d_2) r_{\mathrm{lcm}(d_1,d_2)}(x),$$

we again have

$$S(z) \leq V^+(z)x + R^+(z). \tag{31}$$

Ignoring the error term $R^+(z)$ for the moment, one can ask about optimizing the main term $V^+(z)x$ in the bound (31). This amounts to viewing $V^+(z)$ as a quadratic form and then minimizing it.

For simplicity, we will assume that $g(p) \in (0,1)$ for $p \in P$, and $g(p) = 0$ for $p \notin P$. (Before we only wanted $g(p) \in [0,1)$ for $p \in P$, but there is no harm in

adding to P those primes p for which $g(p) = 0$ into P.) Let h be a multiplicative function with

$$h(p) = \frac{g(p)}{1 - g(p)}.$$

We can then diagonalize the quadratic form as follows: first, put $c = \gcd(d_1, d_2)$, $a = d_1/c$, $b = d_2/c$ to obtain

$$\begin{aligned} V^+(z) &= \sum_{a,b,c:abc|P(z)} \rho(ac)\rho(bc)g(abc) \\ &= \sum_{c|P(z)} g(c)^{-1} \sum_{a,b:abc|P(z)} (g(ac)\rho(ac))(g(bc)\rho(bc)). \end{aligned}$$

Note that since $P(z)$ is squarefree, the condition $abc|P(z)$ forces $\gcd(a, b) = 1$. We now perform inclusion-exclusion on $\gcd(a, b)$ to obtain

$$\begin{aligned} V^+(z) &= \sum_{c|P(z)} g(c)^{-1} \sum_{d|P(z)/c} \mu(d) \left(\sum_{m|P(z)/(cd)} g(cdm)\rho(cdm) \right)^2 \\ &= \sum_{c|P(z)} g(c)^{-1} \sum_{d|P(z)/c} \mu(d) \left(\sum_{m|P(z):m\equiv 0\,(cd)} g(m)\rho(m) \right)^2. \end{aligned}$$

We next substitute $e, f/e$ in for c, d, and reorder the sum:

$$\begin{aligned} V^+(z) &= \sum_{f|P(z)} \sum_{e|f} \mu(f/e)g(e)^{-1} \left(\sum_{m|P(z):m\equiv 0\,(f)} g(m)\rho(m) \right)^2 \\ &= \sum_{f|P(z)} h(f)^{-1} \left(\sum_{m|P(z):m\equiv 0\,(f)} g(m)\rho(m) \right)^2. \end{aligned}$$

Let's put

$$\xi(d) = \mu(d) \sum_{m|P(z):m\equiv 0\,(d)} g(m)\rho(m),$$

so that we have

$$V^+(z) = \sum_{d|P(z)} h(d)^{-1}\xi(d)^2.$$

Before we can minimize this quadratic form, we must first reexpress in terms of ξ the conditions we imposed on ρ. Namely, by Möbius inversion,

$$\rho(n) = \frac{\mu(n)}{g(n)} \sum_{d|P(z):d\equiv 0\,(n)} \xi(d),$$

so the condition $\rho(1) = 1$ is equivalent to

$$\sum_{d|P(z)} \xi(d) = 1,$$

and the condition $\rho(d) = 0$ for $d \geq \sqrt{y}$ is equivalent to

$$\xi(d) = 0 \qquad (d \geq \sqrt{y}).$$

That is, ξ is restricted to a hyperplane.

Here's where the L^2 part comes in. By the Cauchy-Schwartz inequality,

$$V^+(z) \geq H^{-1}, \qquad H = \sum_{d<\sqrt{y}, d|P(z)} h(d)$$

and equality holds for

$$\xi(d) = h(d)H^{-1} \qquad (d < \sqrt{y}).$$

Backing up, we get

$$\rho(d) = \mu(d)\frac{h(d)}{g(d)}H^{-1} \sum_{n<\sqrt{y}/d:\gcd(d,n)=1} h(n).$$

Putting this together, we obtain the following.

THEOREM 14.1 (Selberg). *Let $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be an arithmetic function with finite support. Let P be a set of primes, and put $P(z) = \prod_{p\leq z, p\in P} p$. For $d|P(z)$, write*

$$A_d = \sum_{n\equiv 0\,(d)} f(n) = g(d)X + r_d(z)$$

for $X > 0$ and g a multiplicative function with $0 < g(p) < 1$ for all $p \in P$. Let $h(d)$ be a multiplicative function with $h(p) = g(p)(1 - g(p))^{-1}$ for all $p \in P$, and put

$$H = \sum_{d<\sqrt{y}, d|P(z)} h(d)$$

for some $y > 1$. Then

$$S(z) = \sum_{(n,P(z))=1} f(n) \leq XH^{-1} + \sum_{d|P(z)} \lambda^+(d) r_d(z), \tag{32}$$

for

$$\lambda^+(n) = \sum_{d_1,d_2:\mathrm{lcm}(d_1,d_2)=n} \rho(d_1)\rho(d_2)$$

$$\rho(d) = \mu(d)\frac{h(d)}{g(d)}H^{-1} \sum_{n<\sqrt{y}/d:\gcd(d,n)=1} h(n).$$

As a somewhat miraculous corollary (due to van Lint and Richert), we obtain

$$0 \leq \mu(d)\rho(d) \leq 1 \tag{33}$$

(exercise); this makes it easy to estimate the error term in (32), e.g., by

$$|\lambda^+(d)| \leq d^{(\log 3)/(\log 2)} \tag{34}$$

(exercise).

Exercises

(1) Prove (33). (Hint: group terms in the definition of H according to the common divisor of d with some fixed number e.)
(2) Deduce (34) from (33), by proving that $|\lambda^+(d)| \leq 3^{\nu(d)}$, for $\nu(d)$ equal to the number of prime factors of d.

(3) In the Selberg sieve, prove that if we extend g to a completely multiplicative function, then
$$H \geq \sum_{n<\sqrt{y}} g(n).$$
(4) Prove that for some $c > 0$,
$$\sum_{n\leq x} \frac{2^{\nu(n)}}{n} \geq c \log^2 x \qquad (x \geq 1).$$
(Hint: an elementary proof is possible, but one can also use analytic arguments on the Dirichlet series $\zeta^2(s)/\zeta(2s) = \sum_{n=1}^{\infty} 2^{\nu(n)} n^{-s}$.)
(5) Let $d(n)$ denote the number of divisors of the positive integer n. Prove that
$$\sum_{n\leq x} d(n) \sim x \log x.$$
(This is needed for the next problem.)
(6) Use the Selberg sieve to prove that the number of twin primes $p \leq x$ is $O(x/\log^2 x)$. (Hint: put $f(n) = 1$ if $n = m(m+2)$ for some m and $f(n) = 0$ otherwise, then apply the Selberg sieve with $z = x^{1/4}$. You may need some of the earlier exercises as well.)
(7) (Brun-Titchmarsh theorem) Prove that for any $\epsilon > 0$, there exists $x_0 = x_0(\epsilon)$ with the following property: for any positive integers m, N with $\gcd(m, N) = 1$, and any $x \geq \max\{N, x_0(\epsilon)\}$, the number of primes $p \leq x$ with $p \equiv m \pmod N$ is at most
$$\frac{(2+\epsilon)x}{\phi(N)\log(2x/N)}.$$
This is one of several problems in which the Selberg sieve applies to give you a result which is off by a factor of 2 from the expected best result.
(8) Prove that
$$\sum_{n\leq x} \frac{n}{\phi(n)} = O(x),$$
then deduce by partial summation that
$$\sum_{n\leq x} \frac{1}{\phi(n)} = O(\log x),$$
(Hint: first prove that the sum $\sum_n 1/(n\gamma(n))$ converges, where $\gamma(n) = \prod_{p|n} p$.)
(9) Use the previous two exercises to deduce that
$$\sum_{p\leq x} d(p-1) = O(x),$$
where $d(n)$ denotes the number of divisors of n.

CHAPTER 15

Applying the Selberg sieve

Here are some suggestions about how to apply the Selberg sieve; this should help with some of the exercises on the previous handout (the bound on twin primes, and the Brun-Titchmarsh inequality).

1. Review of the setup

Recall the setup.

THEOREM 15.1 (Selberg). *Let $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ be an arithmetic function with finite support. Let P be a set of primes, and put $P(z) = \prod_{p \leq z, p \in P} p$. For $d|P(z)$, write*

$$A_d = \sum_{n \equiv 0 \, (d)} f(n) = g(d)X + r_d(z)$$

for $X > 0$ and g a multiplicative function with $0 < g(p) < 1$ for all $p \in P$. Let $h(d)$ be a multiplicative function with $h(p) = g(p)(1 - g(p))^{-1}$ for all $p \in P$, and put

$$H = \sum_{d < \sqrt{y}, d|P(z)} h(d)$$

for some $y > 1$. Then

$$S(z) = \sum_{(n,P(z))=1} f(n) \leq XH^{-1} + \sum_{d|P(z)} \lambda^+(d) r_d(z), \tag{35}$$

for

$$\lambda^+(n) = \sum_{d_1, d_2 : \mathrm{lcm}(d_1,d_2)=n} \rho(d_1)\rho(d_2)$$

$$\rho(d) = \mu(d)\frac{h(d)}{g(d)} H^{-1} \sum_{n < \sqrt{y}/d : \gcd(d,n)=1} h(n).$$

Also recall that we could bound $\lambda^+(d)$ by $\tau_3(d)$, the number of ways to write d as a product of 3 positive integers.

2. Interlude: bounding sums of multiplicative functions

Let f be a multiplicative function, for which we want to bound $\sum_{n \leq x} f(n)$. Here is an argument that does this for us (due to Wirsing), assuming some control over the values of f at prime powers.

To be specific, let e be the arithmetic function defined by the following identity of formal Dirichlet series:

$$\sum_{n=1}^{\infty} e(n) n^{-s} = -\frac{d}{ds} \log \sum_{n=1}^{\infty} f(n) n^{-s}.$$

We will impose the condition that for some $\kappa > 0$,

$$\sum_{n \le x} e(n) = \kappa \log x + O(1) \tag{36}$$

and

$$\sum_{n \le x} |f(n)| = O(\log^{|\kappa|} x). \tag{37}$$

(The superfluous absolute value in (37) is included because it actually suffices to take $\kappa > -1/2$, but we won't use this.)

Define

$$M_f(x) = \sum_{n \le x} f(n),$$

which is what we want to estimate. We first obtain

$$(\kappa + 1) \sum_{n \le x} f(n) \log n = \kappa M_f(x) \log x + O(\log^{\kappa} x) \tag{38}$$

(exercise). Since

$$\sum_{n \le x} f(n) \log(x/n) = \int_1^x M_f(y) y^{-1}\, dy,$$

we obtain

$$\Delta(x) = M_f(x) \log x - (\kappa + 1) \int_2^x M_f(y) y^{-1}\, dy = O(\log^{\kappa} x).$$

We next derive the following identity:

$$M_f(x) = \log^{\kappa} x \int_2^x -\Delta(y) d(\log y)^{-\kappa - 1} + \Delta(x) \log^{-1} x \tag{39}$$

(exercise). This implies

$$M_f(x) = c_f \log^{\kappa} x + O(\log^{\kappa - 1} x)$$

for

$$c_f = -\int_2^{\infty} \Delta(y) d(\log y)^{-\kappa - 1},$$

but it would be nice to be able to describe c_f more explicitly. Fortunately this is possible: we have

$$c_f = \frac{1}{\Gamma(\kappa + 1)} \prod_p (1 - p^{-1})^{\kappa} (1 + f(p) + f(p^2) + \cdots) \tag{40}$$

(exercise).

3. Bounding the main term

To get an upper bound on the main term XH^{-1}, we need a lower bound on H. A simple example occurs when $g(d) = d^{-1}$; see exercises.

A more generic example occurs when we have

$$\sum_{p \le x} g(p) \log p = \kappa \log x + O(1)$$

for some $\kappa > 0$, and

$$\sum_p g(p)^2 \log p < \infty.$$

For instance, this holds if $g(p) = c/p$. By Wirsing's bound, we get

$$H = c \log^{\kappa} \sqrt{y}(1 + O(\log^{-1} y))$$
$$c = \frac{1}{\Gamma(\kappa+1)} \prod_p (1 - g(p))^{-1}(1 - p^{-1})^{\kappa}.$$

This can be more usefully written as

$$H^{-1} = 2^{\kappa}\Gamma(\kappa+1) H_g \log^{-\kappa} y(1 + O(\log^{-1} y)), \tag{41}$$

where

$$H_g = \prod_p (1 - g(p))(1 - p^{-1})^{-\kappa}.$$

4. Bounding the error term

Suppose our function g satisfies the conditions

$$g(d)d \geq 1 \qquad (d|P(z)) \tag{42}$$

and

$$\sum_{y \leq p \leq x} g(p) \log p = O(\log(2x/y)). \tag{43}$$

Suppose also that the individual error terms r_d are not too large:

$$|r_d(z)| \leq g(d)d \qquad (d|P(z)). \tag{44}$$

Then it is straightforward to derive the bound

$$\left| \sum_{d|P(z)} \lambda^+(d) r_d(z) \right| \leq y \log^{-2} y \tag{45}$$

(exercise).

Exercises

(1) In the Selberg sieve, prove that
$$H > \log \sqrt{y}.$$
Moreover, if we instead take $g(d) = d^{-1}$ and P to be the set of all primes, then
$$H > (\log \sqrt{y}) \prod_{p|q} (1 - g(p)).$$

(2) Prove (38).

(3) Prove (39).

(4) Prove (40). (Hint: write $\sum_{n=1}^{\infty} f(n) n^{-s}$ in terms of c_f by partial summation, then multiply by $\zeta(s+1)^{\kappa}$ and compare to the Euler product.)

(5) Prove (45). (Hint: first bound the sum on the left by
$$\left(\sum_{d < \sqrt{y}} |\rho_d| g(d) d \right)^2 \leq \left(\frac{1}{H} \sum_{n < \sqrt{y}} h(n)\sigma(n) \right)^2,$$
where σ is the usual sum-of-divisors function. Then apply the prime number theorem plus partial summation to control this.)

CHAPTER 16

Introduction to large sieve inequalities

In this unit, we consider a relatively simple example of a large sieve inequality, of the sort introduced by Linnik. This is a setup for the multiplicative large sieve inequality we will need for Bombieri-Vinogradov.

1. Overview

The purpose of a "large sieve" is to allow sieving over a range of primes not possible with the traditional sieve methods we considered earlier. The price to be paid is that one only gets results of an aggregate nature. For instance, in the Bombieri-Vinogradov theorem, we will consider the error terms in the prime number theorem in arithmetic progression for *all* moduli in some range, and show that the sum of the errors cannot be too large.

This said, a "large sieve inequality" does not itself involve a sieve, at least not the way we look at these things nowadays; the sieves only appear in the application. The general *large sieve problem*: given a finite set V of vectors $v \in \mathbb{C}^n$, find the smallest constant $C = C(V)$ such that for any vector $x \in \mathbb{C}^n$,

$$\sum_{v \in V} |v \cdot \overline{x}|^2 \leq C x \cdot \overline{x}. \tag{46}$$

(Note that $v \cdot \overline{x}$ is the usual Hermitian inner product.) Of course one has $C \leq \sum_{v \in V} v \cdot \overline{v}$ by Cauchy-Schwarz term by term, but this is nowhere near optimal if the vectors v are pointing in all different directions, as then the vector x cannot simultaneously be nearly parallel to all of them. A trivial example is given by an orthonormal set of vectors, in which case $C = 1$; see the exercises for another simple example.

In number theory applications, we tend to view the same setup as follows. Given a finite set X of complex-valued sequences, and a cutoff N, find a constant $C = C(X, N)$ such that for any $a_n \in \mathbb{C}$,

$$\sum_{x \in X} \left| \sum_{n \leq N} a_n x(n) \right|^2 \leq C \sum_{n \leq N} |a_n|^2.$$

2. An additive large sieve

In the additive large sieve, we take the sequences $x \in X$ to be of the form $\exp(2\pi i \alpha n)$ for some $\alpha = \alpha_x \in \mathbb{R}$ (or better, in $\mathbb{R}/\mathbb{Z}$). In order for these to be "not too parallel", we insist that the corresponding α_x be δ-*spaced* for some $\delta > 0$, i.e., if $x, y \in X$ are distinct, then $\alpha_x - \alpha_y$ must have distance at least δ from the nearest integer. The following inequality is due independently to Selberg, and to Montgomery and Vaughan; it refines a result of Davenport and Halberstam.

THEOREM 16.1. *Fix $\delta \in (0, 1/2]$. Let $S \subset \mathbb{R}$ be a δ-spaced set (necessarily finite). Then for any $a_n \in \mathbb{C}$ for $M < n \leq M + N$,*

$$\sum_{\alpha \in S} \left| \sum_{M < n \leq M+N} a_n \exp(2\pi i \alpha n) \right|^2 \leq (\delta^{-1} + N - 1) \sum_{M < n \leq M+N} |a_n|^2.$$

The key input is the following inequality, a variation of a classic inequality of Hilbert.

LEMMA 16.2. *Let $\lambda_1, \ldots, \lambda_n$ be real numbers with $|\lambda_i - \lambda_j| \geq \delta$ whenever $i \neq j$. Then for any $z_1, \ldots, z_n \in \mathbb{C}$,*

$$\left| \sum_{i \neq j} \frac{z_i \overline{z_j}}{\lambda_i - \lambda_j} \right| \leq \frac{\pi}{\delta} \sum_{i=1}^{n} |z_i|^2.$$

PROOF. Exercise. □

COROLLARY 16.3. *For $S = \{\alpha_1, \ldots, \alpha_n\}$ a δ-spaced set and $z_1, \ldots, z_n \in \mathbb{C}$,*

$$\left| \sum_{i \neq j} \frac{z_i \overline{z_j}}{\sin \pi(\alpha_i - \alpha_j)} \right| \leq \delta^{-1} \sum_{i=1}^{n} |z_i|^2.$$

PROOF. Let K be a large positive integer. By the previous lemma applied to the set of $M + \alpha_i$ and the numbers $(-1)^M z_i$ for $i = 1, \ldots, n$ and $M = 1, \ldots, K$, we get

$$\left| \sum_{(i,M) \neq (j,N)} (-1)^{M-N} \frac{z_i \overline{z_j}}{M - N + \alpha_i - \alpha_j} \right| \leq \frac{\pi K}{\delta} \sum_{i=1}^{n} |z_i|^2.$$

It changes nothing to run the sum over pairs of pairs in which only $i \neq j$, since the terms $(i, M), (i, N)$ and $(i, N), (i, M)$ cancel each other. Put $k = M - N$ and divide by K to obtain

$$\left| \sum_{i \neq j} z_i \overline{z_j} \sum_{k=-K}^{K} \left(1 - \frac{|k|}{K}\right) \frac{(-1)^k}{k + \alpha_i - \alpha_j} \right| \leq \frac{\pi}{\delta} \sum_{i=1}^{n} |z_i|^2.$$

Taking $K \to \infty$ and recalling that

$$\frac{1}{\alpha} + \sum_{k=1}^{\infty} \left(\frac{(-1)^k}{k + \alpha} + \frac{(-1)^{-k}}{-k + \alpha} \right) = \frac{\pi}{\sin \pi \alpha}$$

yields the claim. □

COROLLARY 16.4. *With notation as in the previous corollary, for any $x \in \mathbb{R}$,*

$$\left| \sum_{i \neq j} z_i \overline{z_j} \frac{\sin 2\pi x(\alpha_i - \alpha_j)}{\sin \pi(\alpha_i - \alpha_j)} \right| \leq \delta^{-1} \sum_{i=1}^{n} |z_i|^2.$$

PROOF. Apply the previous corollary twice, multiplying z_i by $\exp(\pm 2\pi i x \alpha_i)$. □

We also need the following "duality" lemma.

LEMMA 16.5 (Duality). *Let $A_{m,n} \in \mathbb{C}$ and $C \in \mathbb{R}$ be constants such that for any $\beta_n \in \mathbb{C}$,*

$$\sum_m \left| \sum_n \beta_n A_{m,n} \right|^2 \leq C \sum_n |\beta_n|^2.$$

Then for any $\alpha_m \in \mathbb{C}$,

$$\sum_n \left| \sum_m \alpha_m A_{m,n} \right|^2 \leq C \sum_m |\alpha_m|^2.$$

PROOF. Exercise. □

PROOF OF THEOREM 16.1. We prove here only the bound with the factor $\delta^{-1} + N - 1$ replaced by $\delta^{-1} + N$; there is a fun trick to pick up the extra -1 (see exercises).

By duality, we may reduce to showing that for any $z_\alpha \in \mathbb{C}$,

$$\sum_{M < n \leq M+N} \left| \sum_{\alpha \in S} z_\alpha \exp(2\pi i n \alpha) \right|^2 \leq (\delta^{-1} + N) \sum_{\alpha \in S} |z_\alpha|^2.$$

When we expand the square on the left side, the diagonal terms contribute $N \sum_\alpha |z_\alpha|^2$. The off-diagonal terms give

$$\sum_{\alpha \neq \beta} z_\alpha \overline{z_\beta} \exp(2\pi i K(\alpha - \beta)) \frac{\sin \pi N(\alpha - \beta)}{\sin \pi(\alpha - \beta)}$$

for $K = M + (N+1)/2$. By Corollary 16.4, this is bounded by $\delta^{-1} \sum_\alpha |z_\alpha|^2$. □

Exercises

(1) Find the optimal constant in the large sieve inequality (46) when the vectors in V are taken to be unit vectors forming the corners of a regular simplex in $\mathbb{R}^n$ with center at the origin. (Hint: it may simply matters to view the situation inside an $(n+1)$-dimensional space.)

(2) Prove Lemma 16.2. (Hint: by Cauchy-Schwarz, it is enough to prove

$$\sum_{i=1}^n \left| \sum_{j \neq i} \frac{z_j}{\lambda_i - \lambda_j} \right|^2 \leq \frac{\pi^2}{\delta^2} \sum_{i=1}^n |z_i|^2.$$

Do this by extremizing an appropriate Hermitian (quadratic) form, and noting that the extremal vector must be an eigenvector.)

(3) Prove Lemma 16.5.

(4) (Cohen) Prove Theorem 16.1 as stated, assuming the version in which the factor $\delta^{-1} + N - 1$ is replaced by $\delta^{-1} + N$. (Hint: apply the weak version to the δK-spaced points $(\alpha + k)/K$ for α running over S and k running over $\{1, \ldots, K\}$, and the values b_m being related to the original a_n via

$$\sum_m b_m \exp(2\pi i \alpha m) = \sum_n a_n \exp(2\pi i K \alpha n).$$

Then take the limit as $K \to \infty$.)

CHAPTER 17

A multiplicative large sieve inequality

In this unit, we convert the additive large sieve inequality from the previous unit, which concerned characters of the additive group, into a result about Dirichlet characters.

1. Review of the additive large sieve

The additive large sieve inequality from last time stated the following.

THEOREM 17.1. *Fix $\delta \in (0, 1/2]$. Let $S \subset \mathbb{R}$ be a δ-spaced set (necessarily finite). Then for any $a_n \in \mathbb{C}$ for $M < n \leq M + N$,*

$$\sum_{\alpha \in S} \left| \sum_{M < n \leq M+N} a_n \exp(2\pi i \alpha n) \right|^2 \leq (\delta^{-1} + N - 1) \sum_{M < n \leq M+N} |a_n|^2.$$

We will need in particular the special case

$$S = \{a/q : 1 \leq q \leq Q, 0 \leq a < q, \gcd(a, q) = 1\}.$$

Note that if $a/q, a'/q' \in S$ are distinct and $m \in \mathbb{Z}$, then

$$\left| \frac{a}{q} - \frac{a'}{q'} - n \right| = \left| \frac{*}{qq'} \right| \geq Q^{-2}.$$

That is, S is δ-spaced for $\delta = Q^{-2}$. We thus obtain the following from the large sieve inequality.

THEOREM 17.2. *Let N be a positive integer, and choose $a_n \in \mathbb{C}$ for $M < n \leq M + N$. Then*

$$\sum_{1 \leq q \leq Q} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left| \sum_{M < n \leq M+N} a_n \exp(2\pi i a n/q) \right|^2 \leq (Q^2 + N - 1) \sum_{M < n \leq M+N} |a_n|^2.$$

2. The Bombieri-Davenport inequality

We now ask the question: what if we replace the exponentials in the large sieve by the primitive Dirichlet characters of all moduli $q \leq Q$?

THEOREM 17.3 (Bombieri-Davenport). *Fix positive integers Q, N. For any $a_n \in \mathbb{C}$ for $M < n \leq M + N$, we have*

$$\sum_{1 \leq q \leq Q} \frac{q}{\phi(q)} \sum_{\chi} \left| \sum_{M < n \leq M+N} a_n \chi(n) \right|^2 \leq (Q^2 + N - 1) \sum_{M < n \leq M+N} |a_n|^2. \tag{47}$$

One can prove a stronger inequality in which you allow also some terms corresponding to imprimitive characters, but I won't need this.

PROOF. As in the proof of the functional equation for Dirichlet L-functions, we use the expansion of primitive Dirichlet characters in terms of Gauss sums:

$$\chi(n) = \tau(\overline{\chi})^{-1} \sum_{a \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(a) \exp(2\pi i a n/q),$$

where

$$\tau(\chi) = \sum_{b \in \mathbb{Z}/q\mathbb{Z}} \chi(b) \exp(2\pi i b/q)$$

has the property that

$$|\tau(\chi)| = \sqrt{q}.$$

If we put

$$S(\alpha) = \sum_{M < n \leq M+N} a_n \exp(2\pi i \alpha n),$$

we can then write

$$\frac{q}{\phi(q)} \left| \sum_{M < n \leq M+N} a_n \chi(n) \right|^2 = \frac{1}{\phi(q)} \left| \sum_{a \in \mathbb{Z}/q\mathbb{Z}} \overline{\chi}(a) S(a/q) \right|^2.$$

Summing over $1 \leq q \leq Q$ and χ primitive gives the left side of (47). I can get an upper bound by summing over $1 \leq q \leq Q$ and *all* χ, primitive or not. By orthogonality of characters for the group $(\mathbb{Z}/q\mathbb{Z})^*$, this yields

$$\sum_{1 \leq q \leq Q} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |S(a/q)|^2 = \sum_{1 \leq q \leq Q} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left| \sum_{M < m \leq M+N} a_n \exp(2\pi i a n/q) \right|^2$$

as an upper bound for the left side of (47). Applying Theorem 17.2 gives the right side of (47), completing the proof. □

3. An application of the large sieve

We will use the large sieve crucially in the Bombieri-Vinogradov theorem, but first let us illustrate its use with one of its original applications, due to Linnik.

The setup here is as in the sieve of Eratosthenes: I have a sequence of complex numbers a_n with finite support, a set of primes P, and for each $p \in P$, I wish to exclude a set of residue classes Ω_p of size $\omega(p)$. That is, I wish to compute Z, the sum of a_n over thoes n which do not reduce to a class in Ω_p for any $p \in P$. However, I'm not going to require $\omega(p)$ to be as small as I did before; that's what makes this a "large sieve".

THEOREM 17.4. *Suppose the support of a_n belongs to an interval of length N, and that $\omega(p) < p$ for all $p \in P$. Let h be the multiplicative function with $h(q) = 0$ for q not squarefree and*

$$h(p) = \frac{\omega(p)}{p - \omega(p)}.$$

Then for any $Q \geq 1$,

$$|Z|^2 \leq \frac{N + Q^2}{H} \sum_n |a_n|^2,$$

where H is the sum of $h(q)$ over $q \le Q$ squarefree. In particular, if $a_n \in \{0,1\}$ for all n, then

$$Z \le \frac{N+Q^2}{H}.$$

The proof will be immediate from Theorem 17.3 plus the following lemma (summed over q).

LEMMA 17.5. *Put $S(\alpha) = \sum_n a_n \exp(2\pi i \alpha n)$. For any positive squarefree integer q,*

$$h(q)|S(0)|^2 \le \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left| S\left(\frac{a}{q}\right)\right|^2.$$

PROOF. We first reduce to the case where q is prime. Suppose $q = q_1 q_2$ and we know the desired result for both q_1 and q_2. By the Chinese remainder theorem,

$$\begin{aligned}\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left| S\left(\frac{a}{q}\right)\right|^2 &= \sum_{a_1 \in (\mathbb{Z}/q_1\mathbb{Z})^*} \sum_{a_2 \in (\mathbb{Z}/q_2\mathbb{Z})^*} \left| S\left(\frac{a_1}{q_1 + \frac{a_2}{q_2}}\right)\right|^2 \\ &\ge h(q_2) \sum_{a_1 \in (\mathbb{Z}/q_1\mathbb{Z})^*} \left| S\left(\frac{a_1}{q_1}\right)\right|^2 \\ &\ge h(q_1)h(q_2)|S(0)|^2 = h(q)|S(0)|^2.\end{aligned}$$

It remains to prove the case where q is prime; we leave this case as an exercise. □

Here is Linnik's application of the large sieve. For p prime, let $q(p)$ be the least positive integer which is not a quadratic residue modulo p. It is conjectured that $q(p) = O(p^\epsilon)$ for any $\epsilon > 0$, but unconditionally this is only known for $\epsilon > e^{-1/2}/4 \cong 0.152$. On the other hand, under GRH, one can do much better: one proves $q(p) = O(\log^2 p)$.

THEOREM 17.6 (Linnik). *For any fixed $\epsilon > 0$, there exists $c = c(\epsilon)$ such that for any N, there are at most c primes $p \le N$ such that $q(p) > N^\epsilon$.*

PROOF. For convenience, we will prove instead that for some $c = c(\epsilon)$, for any N there are at most c primes $p \le \sqrt{N}$ with $q(p) > N^\epsilon$. Let P be the set of primes $p \le \sqrt{N}$ such that $\left(\frac{n}{p}\right) = 1$ for all $n \le N^\epsilon$, and let Ω_p be the classes of quadratic nonresidues mod p. (This is indeed a large sieve, because $\omega(p) = (p-1)/2$, so $h(p) = (p-1)/(p+1) \sim 1/2$ as $p \to \infty$, whereas in our earlier examples $\omega(p)$ was bounded.)

We will now sieve on the set $\{1, \dots, N\}$, i.e., take $a_n = 1$ for $1 \le n \le N$ and $a_n = 0$ otherwise. The resulting sifted set includes all $n \le N$ with no prime divisors greater than N^ϵ; if we let Z_ϵ be the number of these, then Theorem 17.4 applied with $Q = \sqrt{N}$ yields

$$Z_\epsilon \le 2NH^{-1}.$$

On the other hand, if we let X_ϵ be the number of primes $p \le \sqrt{N}$ with $q(p) > N^\epsilon$, then because $h(p) \ge 1/3$ for all p,

$$\frac{1}{3} X_\epsilon \le \sum_{p \le \sqrt{N}, q(p) \ge N^\epsilon} h(p) \le H.$$

Hence $X_\epsilon Z_\epsilon \leq 6N$.

To conclude, we need to show that $Z_\epsilon \geq cN$ for some $c > 0$. In fact it can be shown that $Z_\epsilon \sim cN$ for some N, but as we don't care about the particular constant, it will suffice to exhibit a special class of numbers being counted by Z_ϵ which are sufficiently numerous. Namely, take $n = mp_1 \cdot p_k \leq N$ with $N^{\epsilon-\epsilon^2} < p_j < N^\epsilon$ for $j = 1, \ldots, k = \epsilon^{-1}$; then

$$Z_\epsilon \geq \sum_{p_1,\ldots,p_k} \left\lfloor \frac{N}{p_1 \cdots p_k} \right\rfloor \geq cN,$$

completing the proof. □

Exercises

(1) Prove the following multivariate version of the additive large sieve inequality (but without optimizing the constant). Fix $\delta > 0$ and $d \geq 1$, and let $\alpha_i = (\alpha_{i,1}, \ldots, \alpha_{i,d}) \in \mathbb{R}^d/\mathbb{Z}^d$ be points which are δ-spaced, in the sense that the distance from each $\alpha_{i,k} - \alpha_{j,k}$ to the nearest integer is at least δ (whenever $i \neq j$ and $1 \leq k \leq d$). Prove that there exists $c = c(d)$ (independent of δ and the α_i) such that for any $a_n \in \mathbb{C}$ with n running over $\{1, \ldots, N\}^d$,

$$\sum_i \left| \sum_n a_n \exp(2\pi i (n \cdot \alpha_i)) \right|^2 \leq c(\delta^{-d} + N^d) \sum_n |a_n|^2.$$

(2) Prove directly (by expanding the squares) that if we take take *all* characters, not just the primitive ones, of a single modulus q, then the large sieve inequality holds with the constant $q + N$. (This is not very useful in practice.)

(3) Prove Lemma 17.5 in the case that q is prime. (Hint: there is no loss of generality in assuming that there is at most one n in each residue class modulo p, and none in the classes in Ω_p, such that $a_n \neq 0$. Then use orthogonality of characters on $\mathbb{Z}/q\mathbb{Z}$.)

CHAPTER 18

The Bombieri-Vinogradov theorem (statement)

In this unit, we state the Bombieri-Vinogradov theorem, which is a surprisingly strong control on the error terms in the prime number theorem in arithmetic progressions. We also mention some related theorems and conjectures. To attack these (which we will do in the next unit), we will need to bring to bear everything we have studied in the course so far!

1. Statement of the theorem

For m, N coprime positive integers, put

$$\psi(x; N, m) = \sum_{n \leq x, n \equiv m \pmod N} \Lambda(n).$$

Recall that the prime number theorem in arithmetic progressions says $\psi(x; N, m) \sim x/\phi(N)$, and that unconditionally we could get an error term

$$\psi(x; N, m) = \frac{x}{\phi(N)} + O(x(\log x)^{-A})$$

for any fixed $A > 0$. This is only meaningful if $N = O((\log x)^A)$. However, under GRH (for the Dirichlet characters of modulus N),

$$\psi(x; N, m) = \frac{x}{\phi(N)} + O(x^{1/2}(\log x)^2),$$

and this is meaningful for $N = O(x^{1/2}(\log x)^{-2})$.

The Bombieri-Vinogradov theorem is an amazingly strong unconditional replacement for the GRH bound. It says that if you pick out the worst error term modulo N for *each* N up to about $x^{1/2}$, and add these up, you get roughly what GRH predicts you should get.

THEOREM 18.1 (Bombieri-Vinogradov). *For any fixed $A > 0$, there exist constants $c = c(A)$ and $B = B(A)$ such that*

$$\sum_{N \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} \left| \psi(x; N, m) - \frac{x}{\phi(N)} \right| \leq cx(\log x)^{-A}$$

for $Q = x^{1/2}(\log x)^{-B}$.

It is expected that one can do better than this.

CONJECTURE 18.2 (Elliott-Halberstam). *For any fixed $A > 0$ and $\epsilon > 0$, there exists $c > 0$ such that*

$$\sum_{N \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} \left| \psi(x; N, m) - \frac{x}{\phi(N)} \right| \leq cx(\log x)^{-A}$$

for $Q = x^{1-\epsilon}$.

This conjecture appears to be extremely hard; for instance, it is not known to follow from GRH. One of the results of Goldston-Pintz-Yıldırım is that the Elliott-Halberstam almost implies the twin primes conjecture: it implies that there are infinitely many pairs of primes at distance ≤ 16. In fact, this (with 16 replaced by some other constant, depending on ϵ) would follow if we could prove the weaker version of Elliott-Halberstam in which $Q = x^{1/2+\epsilon}$, for any fixed $\epsilon > 0$. (Even that does not follow from GRH.)

Note that in the Bombieri-Vinogradov theorem, for each modulus N we look at the worst error term among arithmetic progressions of that modulus. If we instead average over the progressions, we should be able to take Q larger, and in fact that is what happens. (Note: there is a typo in the statement of the theorem in Iwaniec-Kowalski.)

THEOREM 18.3 (Barban, Davenport, Halberstam). *For any fixed $A > 0$, there exist constants $c = c(A)$ and $B = B(A)$ such that*

$$\sum_{N \leq Q} \sum_{m \in (\mathbb{Z}/N\mathbb{Z})^*} \left(\psi(x; N, m) - \frac{x}{\phi(N)}\right)^2 \leq cx^2 (\log x)^{-A}$$

for $Q = x(\log x)^{-B}$.

Finally, we note that Bombieri proved a slightly stronger result, which I will not be proving in this course. (See Davenport §28 for a proof by Montgomery.)

THEOREM 18.4. *For any fixed $A > 0$, there exists $c > 0$ such that*

$$\sum_{N \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} \left|\psi(x; N, m) - \frac{x}{\phi(N)}\right| \leq cx^{1/2} Q (\log x)^5$$

for $x^{1/2}(\log x)^{-A} \leq Q \leq x^{1/2}$.

Exercises

(1) (a) Check that Theorem 18.4 implies Theorem 18.1.
    (b) Check that Conjecture 18.2 implies a slightly weakened version of Theorem 18.3, in which we take $Q = x^{1-\epsilon}$.

(2) Use the Bombieri-Vinogradov theorem, plus the strong Brun-Titchmarsh inequality

$$\pi(x + y; N, m) - \pi(x; N, m) < \frac{2y}{\phi(N) \log(y/N)} + O\left(\frac{y}{N \log^2(y/N)}\right)$$

(where $\pi(x; N, m)$ is the number of primes $p \leq x$ with $p \equiv m \pmod N$) to prove that

$$\sum_{p \leq x} \tau(p - 1) \sim \frac{\zeta(2)\zeta(3)}{\zeta(6)} x,$$

where $\tau(n)$ counts the number of divisors of n. (Hint: use Dirichlet's hyperbola method to reduce to counting primes $\equiv 1 \pmod d$ over $d \leq \sqrt{x}$.)

CHAPTER 19

The Bombieri-Vinogradov theorem (proof)

In this unit, we prove the Bombieri-Vinogradov theorem, in the form stated in the previous unit.

1. Bounding character sums

For f an arithmetic function, put

$$D_f(x;N,m) = \sum_{n\leq x, n\equiv m \pmod N} f(n) - \frac{1}{\phi(N)} \sum_{n\leq x, n\in(\mathbb{Z}/N\mathbb{Z})^*} f(n);$$

that is, $D_f(x;N,m)$ measures the deviation between the sum of f on an arithmetic progression, and the sum on all arithmetic progressions of the same modulus. The following lemma tells us that bounding this deviation allows us to control the sum of f twisted by a Dirichlet character.

LEMMA 19.1. *Let f be an arithmetic function with support in $\{1,\dots,x\}$, and put $|f|_2 = (\sum_n |f(n)|^2)^{1/2}$. Suppose that for some $\Delta \in (0,1]$, we have*

$$|D_f(x;N,m)| \leq x^{1/2}\Delta^9 |f|_2 \tag{48}$$

whenever $m \in (\mathbb{Z}/N\mathbb{Z})^*$. Then for any nonprincipal character χ of modulus r, and any positive integer s,

$$\left| \sum_{n\in(\mathbb{Z}/s\mathbb{Z})^*} f(n)\chi(n) \right| \leq x^{1/2}\Delta^3 r\tau(s)|f|_2.$$

PROOF. By Möbius inversion, we can write

$$\sum_{n\in(\mathbb{Z}/s\mathbb{Z})^*} f(n)\chi(n) = \sum_{k|s} \mu(k) \sum_{n\equiv 0\,(k)} f(n)\chi(n).$$

We split this sum on k at $K = \Delta^{-6}$. We bound the sum for each fixed $k > K$ by Cauchy-Schwarz; the total is thus dominated by

$$\sum_{k|s, k>K} |f|_2 (x/k)^{1/2} \leq |f|_2 x^{1/2} K^{-1/2} \tau(s).$$

For the terms $k \leq K$, we write the sum as (using Möbius inversion again)

$$\sum_{k|s, k\leq K} \mu(k) \sum_{\ell|k} \mu(\ell) \sum_{n\in(\mathbb{Z}/\ell\mathbb{Z})^*} f(n)\chi(n).$$

We split the inside sum over classes modulo ℓr; on each class, we apply (48). Since we are summing over all residue classes, and χ is nonprincipal, the main terms

cancel out; the sum is thus dominated by

$$|f|x^{1/2}\Delta^9 \sum_{k|s,k\le K} \sum_{\ell|k} |\mu(\ell)|\phi(\ell r) \le |f|_2 x^{1/2}\Delta^9 K\phi(r)\tau(s).$$

Since $K = \Delta^{-6}$, we may add the two bounds to give the desired inequality. □

Using the large sieve inequality, we obtain the following.

THEOREM 19.2. *There exists an absolute constant $c > 0$ with the following property. Let f be an arithmetic function with support in $\{1, \dots, x\}$ satisfying* (48)*. Let g be an arithmetic function with support in $\{1, \dots, y\}$, and let $h = f \star g$ be the Dirichlet convolution. Then*

$$\sum_{N\le Q} \max_{m\in(\mathbb{Z}/N\mathbb{Z})^*} |D_h(xy; N, m)| \le c|f|_2|g|_2(\Delta(xy)^{1/2} + x^{1/2} + y^{1/2} + Q)\log^2 Q.$$

PROOF. We have

$$D_h(xy; N, a) = \frac{1}{\phi(N)} \sum_{\chi\ne\chi_0} \overline{\chi}(a) \left(\sum_m f(m)\chi(m)\right)\left(\sum_n g(n)\chi(n)\right),$$

with χ running over Dirichlet characters of modulus N. Rewriting this as a sum only over primitive characters (factoring $N = rs$, where r is the "primitive modulus"), and using the fact that $\phi(rs) \ge \phi(r)\phi(s)$ for all r, s, we can bound the left side of the desired inequality by

$$\sum_{s\le Q} \frac{1}{\phi(s)} \sum_{1<r\le Q} \frac{1}{\phi(r)} \sum_{\chi} \left|\sum_{(m,s)=1} f(m)\chi(m)\right| \left|\sum_{(n,s)=1} g(n)\chi(n)\right|, \tag{49}$$

with χ now running over primitive characters of level r.

We now split the sum over r at $R = \Delta^{-1}$. For $r \le R$, we apply Lemma 19.1; those terms are dominated by

$$|f||g|y^{1/2}\Delta^3 \sum_{s\le Q} \frac{\tau(s)}{\phi(s)} \sum_{r\le R} r \le c|f||g|y^{1/2}\Delta^3 R^2 \log^2 Q.$$

(Note: we are not doing anything to the g terms other than bounding the whole sum by $|g|$ and pulling it out. We apply the lemma to the f terms.) For $r > R$, we split the sum further into ranges like $P < r \le 2P$ and apply the multiplicative large sieve inequality in each range. Rather, we apply it twice: once with the f sum to obtain

$$\sum_{P<r\le 2P} \frac{1}{\phi(r)} \sum_{\chi} \left|\sum_{(m,s)=1} f(m)\chi(m)\right|^2 \le \frac{1}{P}(4P^2 + x - 1)|f|_2^2,$$

and again with the g sum. Putting together with Cauchy-Schwarz, we get a bound

$$\sum_{P<r\le 2P} \frac{1}{\phi(r)} \sum_{\chi} \left|\sum_{m\in(\mathbb{Z}/s\mathbb{Z})^*} f(m)\chi(m)\right| \left|\sum_{n\in(\mathbb{Z}/s\mathbb{Z})^*} g(n)\chi(n)\right| \le \frac{1}{P}(4P^2+x)^{1/2}(4P^2+y)^{1/2}|f|_2|g|_2.$$

Now summing, over $P = R, 2R, \dots$ until $P > Q$, we get a bound on the sum over r in (49) of

$$c|f|_2|g|_2(Q + x^{1/2} + y^{1/2} + x^{1/2}y^{1/2}R^{-1}).$$

(That R^{-1} is the reason we had to limit this argument to r large.) The sum over s throws on another two factors of $\log Q$, yielding the claim. □

2. Proof of the theorem

We now proceed to the proof of the Bombieri-Vinogradov theorem. First, we mention an identity of Vaughan that will be useful: for any $y, z \geq 1$ and $n > z$,

$$\Lambda(n) = \sum_{b \leq y, b|n} \mu(b) \log \frac{n}{b} - \sum_{b \leq y, c \leq z, bc|n} \mu(b)\Lambda(c) + \sum_{b > y, c > z, bc|n} \mu(b)\Lambda(c). \tag{50}$$

Given x, define the incomplete logarithm

$$\lambda(\ell) = \log \ell - \sum_{k \leq x^{1/5}, k|\ell} \Lambda(k);$$

then (50) with $y = z = x^{1/5}$ implies that for $x^{1/5} < n \leq x$,

$$\Lambda(n) = \sum_{\ell m = n, m \leq x^{1/5}} \lambda(\ell)\mu(m) + \sum_{\ell m = n, x^{1/5} < m \leq x^{4/5}} \lambda(\ell)\mu(m). \tag{51}$$

Let $\Lambda_0(n)$ and $\Lambda_1(n)$ denote the two sums on the right side of (51). Then

$$D_\Lambda(x; N, m) = D_{\Lambda_0}(x; N, m) + D_{\Lambda_1}(x; N, m) + O(x^{1/5} \log x),$$

with the error term coming from terms with $n < x^{1/5}$.

It is straightforward to prove that

$$\sum_{N \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} |D_{\Lambda_0}(x; N, m)| = O(Qx^{2/5} \log x), \tag{52}$$

so we concentrate on the contribution from Λ_1. We want to apply Theorem 19.2, but we cannot write the sum $\Lambda_1(n)$ as a convolution because of the restriction $n \leq x$.

To get around this, we cut the interval $1 \leq n \leq x$ into $O(\delta^{-1})$ subintervals of the form $y < n \leq (1+\delta)y$, where $x^{1/5} < \delta \leq 1$ is a parameter we will set later. We cover the summation range

$$\ell m = n, x^{1/5} < m \leq x$$

by ranges

$$\ell m = n, L < \ell \leq (1+\delta)L, M < m \leq (1+\delta)M$$

with L, M taking values $(1+\delta)^j$. We run L, M over the ranges $x^{1/5} < L, M < x^{4/5}$ with $LM = x$; the only trouble is that we do not properly cover the areas $n < x^{1/5}$ and $(1+\delta)^{-1}x < n < (1+\delta)x$. The contribution from the error regions is $O(\delta N^{-1} x \log x)$.

What remains is the sum over L, M of

$$D(L, M; N, m) = \sum_{l, m \equiv m \pmod N} \lambda(\ell)\mu(m) - \frac{1}{\phi(N)} \sum_{lm \in (\mathbb{Z}/N\mathbb{Z})^*},$$

where l, m run over $L < \ell \leq (1+\delta)L, M < m \leq (1+\delta)M$. For each L, M, we may apply Theorem 19.2 with $\Delta = (\log x)^{-A}$; the hypothesis (48) is satisfied by the Siegel-Walfisz theorem (the error bound on the prime number theorem in arithmetic progressions). If we take $Q = \Delta x^{1/2}$, we get

$$\sum_{N \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} |D(L, M; N, m)| = O(\delta \Delta x (\log x)^3).$$

Summing over L, M, we obtain

$$\sum_{N \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} |D_{\Lambda_1}(x; N, m)| = O((\delta^{-1}x + \Delta)x(\log x)^3.$$

We now choose $\delta = \Delta^{1/2}$, so this bound becomes $\Delta^{1/2}x(\log x)^3$. Adding back in (52) gives

$$\sum_{N \leq \Delta x^{1/2}} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} \left|\psi(x; N, m) - \frac{\psi(x)}{\phi(N)}\right| = O(\Delta^{1/2}x(\log x)^3).$$

Using the prime number theorem with error term, we can take $\psi(x) = x + O(\delta x)$. This gives the Bombieri-Vinogradov theorem with $B(A) = 2A + 6$.

3. The Barban-Davenport-Halberstam theorem

We leave the proof of the Barban-Davenport-Halberstam theorem to the reader; it is actually somewhat simpler than Bombieri-Vinogradov. Here is the key step.

THEOREM 19.3. *There exists an absolute constant $c > 0$ with the following property. Let f be an arithmetic function with support in $\{1, \dots, x\}$ satisfying* (48)*. Then*

$$\sum_{N \leq Q} \sum_{a \in (\mathbb{Z}/N\mathbb{Z})^*} |D_f(x; N, m)|^2 \leq c|f|^2(\Delta x + Q)(\log Q)^2.$$

We note in passing the following corollary.

COROLLARY 19.4. *With conditions as in Theorem 19.2, for $ab \neq 0$, we have*

$$\sum_{N \leq Q, (ab,N)=1} \left| \sum_{m,n: am \equiv bn\ (N), (mn,N)=1} f(m)g(n) - \frac{1}{\phi(N)} \left(\sum_{(m,N)=1} f(m) \right) \left(\sum_{(n,N)=1} g(n) \right) \right|$$
$$\leq c|f||g|(x + Q)^{1/2}(\Delta y + Q)^{1/2} \log^2 Q.$$

Exercises

(1) Prove (50).
(2) Use (50) to deduce (51).
(3) Prove (52).
(4) Prove Theorem 19.3, by imitating the proof of Theorem 19.2.
(5) Deduce Corollary 19.4 from Theorem 19.3. (Hint: rewrite the difference in terms of D_f and D_g.)

Part 4

Gaps between primes

CHAPTER 20

Prime k-tuples

This unit begins the third part of the course, in which we apply the results gathered in the first two parts in order to say something about the extent to which primes cluster together in short intervals.

Reference cited below: P.X. Gallagher, On the distribution of primes in short interval, *Mathematika* **23** (1976), 4–9; corrigendum, *ibid.* **28** (1981), 86. (I couldn't find this online.)

1. The Hardy-Littlewood k-tuples conjecture

Let $\mathcal{H}$ denote a k-tuple of distinct integers. What does one expect about the distribution of the integers n such that $n+h$ is prime for each $h \in \mathcal{H}$?

Here is a rather simple-minded guess. The prime number theorem suggests that if one chooses a random integer of size x, it will be prime with probability $1/(\log x)$. If one then chooses k distinct integers of size x, and there is no obvious reason why they cannot all be prime, then one might expect them to be simultaneously prime with probability $\log^{-k} x$, and the number of such tuples with terms bounded by x should be asymptotic to $x \log^{-k} x$, with the constant 1.

However, this turns out not to be the correct constant, as is easily verified against experimental evidence in the case of twin primes. The reason is perhaps obvious: the facts that the different $n+h$ are coprime to a fixed prime p are not independent, and one needs to account for this. Here is the recipe for doing so proposed by Hardy-Littlewood (and mentioned by Ben Green in his guest lecture).

Fix a prime p. The probability that k randomly chosen integers are all not divisible by p is $(1-1/p)^k$. On the other hand, the probability that the $n+h$ are all coprime to p is $1-v_{\mathcal{H}}(p)/p$, where $v_{\mathcal{H}}(p)$ is the number of residue classes modulo p represented by elements of $\mathcal{H}$. We thus set

$$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{v_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}$$

(called the "singular series", because it occurs in the Hardy-Littlewood circle method as a series summed over singularities of some integral) and conjecture as follows.

Conjecture 20.1 (Hardy-Littlewood). *Suppose that $v_{\mathcal{H}}(p) < p$ for all p. Then the number of integers $n \le x$ such that $n+h$ is prime for each $h \in \mathcal{H}$ is asymptotic to $\mathfrak{S}(\mathcal{H}) x \log^{-k} x$.*

Of course if $v_{\mathcal{H}}(p) = 0$ for some p, then there is a trivial obstruction created by divisibility mod p, so you only get finitely many prime k-tuples of that shape. On the other hand, if $v_{\mathcal{H}}(p) < p$ for all p, then the product converges absolutely and so $\mathfrak{S}(\mathcal{H}) > 0$.

Convention: it will be convenient later to take the same definition for $\mathfrak{S}(\mathcal{H})$ even if $\mathcal{H}$ does not have distinct entries.

2. k-tuples and prime gaps

If one is only interested in looking for primes which are close together, without specifying exactly what the gaps are, one could go back to the probabilistic model (attributed to Cramér, more famous for his rule for solving linear systems). It suggests that the distribution of $\pi(n+h)-\pi(n)$, for $n \leq N$ and $h \sim \lambda \log N$ with λ fixed, should approach a Poisson distribution with parameter λ as $N \to \infty$. The fact that this follows from a suitably uniform version of the k-tuples conjecture is due to Gallagher; the main part of the argument is the following result, which we will need later.

THEOREM 20.2 (Gallagher). *We have*

$$\sum_{\mathcal{H} \in \{1,\dots,x\}^k} \mathfrak{S}(\mathcal{H}) \sim x^k.$$

In other words, the fudge factor $\mathfrak{S}(\mathcal{H})$ between the probabilistic model and the Hardy-Littlewood prediction averages out to 1, so the prediction based on the probabilistic model is consistent with Hardy-Littlewood. (Note: the contribution from tuples not having distinct entries is $O(x^{k-1})$, so it doesn't matter whether we include them or not.)

Here is a sketch of Gallagher's proof, with the missing details left as exercises. (Throughout, keep k fixed.) Put

$$a(p,m) = \left(1-\frac{m}{p}\right)\left(1-\frac{1}{p}\right)^{-k} - 1$$
$$a_{\mathcal{H}}(p) = a(p, v_{\mathcal{H}}(p))$$

so that

$$\mathfrak{S}(\mathcal{H}) = \prod_p (1 + a_{\mathcal{H}}(p)).$$

Extend a by multiplicativity to squarefree arguments d, so that

$$\mathfrak{S}(\mathcal{H}) = \sum_d a_{\mathcal{H}}(d)$$

with the sum on the right being absolutely convergent.

We can truncate the sum over d by showing that for each fixed $\epsilon > 0$,

$$\sum_{\mathcal{H} \in \{1,\dots,x\}^k} \mathfrak{S}(\mathcal{H}) = \sum_{d \leq y} \sum_{\mathcal{H}} a_{\mathcal{H}}(d) + O(x^k (xy)^\epsilon / y) \tag{53}$$

with the constant depending only on k, ϵ and not on x, y (exercise).

For any given d, we can rewrite the inner sum of (53) as a sum

$$\sum_v \left(\prod_{p|d} a(p, v(p)) \right) f_d(x, v),$$

where v runs over vectors indexed by the prime factors of d, with $v(p) \in \{1, \dots, p\}$ for each $p|d$, and $f_d(x,v)$ counts k-tuples $\mathcal{H} \in \{1,\dots,x\}^r$ which occupy exactly $v(p)$ residue classes modulo p for each $p|d$.

Write $\left\{ {a \atop b} \right\}$ for the number of partitions of an a-element set into b unordered parts (Stirling number of the second kind). If we set

$$A(d) = \sum_v \prod_{p|d} a(p, v(p)) \binom{p}{v(p)} v(p)! \left\{ {k \atop v(p)} \right\}$$

$$B(d) = \sum_v \prod_{p|d} |a(p, v(p))| \binom{p}{v(p)} v(p)! \left\{ {k \atop v(p)} \right\}$$

$$C(d) = \sum_v \prod_{p|d} |a(p, v(p))|,$$

then

$$\sum_{\mathcal{H}} a_{\mathcal{H}}(d) = (x/d)^k A(d) + O((x/d)^{k-1} B(d)) + O(x^{k-1} C(d)). \tag{54}$$

From this, plus the identities

$$\sum_{v=1}^{p} \binom{p}{v} v! \left\{ {k \atop v} \right\} = p^k \tag{55}$$

$$\sum_{v=1}^{p} v \binom{p}{v} v! \left\{ {k \atop v} \right\} = p^{k+1} - (p-1)^k p, \tag{56}$$

it is not difficult to deduce Theorem 20.2.

Exercises

(1) Prove that for k a positive integer,

$$\int_2^x \log^{-k} t \, dt \sim x \log^{-k} x.$$

(2) Prove (53). (If you get stuck, see the hint for problem 5.)

(3) Prove (54). (Hint: it might help to think in terms of counting lattice points.)

(4) Prove the identities (55), (56).

(5) Complete the proof of Theorem 20.2 from (53) and (54). (Hint: first use the Stirling number identities to calculate $A(d)$. Then estimate $B(d)$ and $C(d)$, using the bound

$$|a(p, m)| \leq \begin{cases} c(k)(p-1)^{-2} & m = k \\ c(k)(p-1)^{-1} & m < k. \end{cases}$$

That is, the constant $c(k)$ depends on k but not on p or m. Finally, take $y = x^{1/2}$ in (53).)

CHAPTER 21

Small gaps between primes (after Goldston-Pintz-Yıldırım)

In this section, we introduce the strategy initiated by Goldston-Yıldırım, and carried out by them and Pintz, for proving new results on the existence of short gaps between primes. Some calculations are postponed to a later unit.

References: much of this unit is liberally plagiarized from K. Soundararajan, Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım, *Bull. Amer. Math. Soc.* **44** (2007), 1–18. For more details (which will be plagiarized later), see D.A. Goldston, Y. Motodashi, J. Pintz, and C. Yıldırım, Small gaps between primes exist, *Proc. Japan Acad. Ser. A Math. Sci.* **82** (2006), 61–65. (Both references are available online, e.g., via MathSciNet.)

1. The target theorem

Let p_n denote the n-th prime. As noted in the previous unit, we can use a probabilistic model to make plausible predictions about the ratio $(p_{n+1} - p_n)/(\log p_n)$, by supposing that $\pi(x + y) - \pi(x)$, for x large and $y \sim \lambda \log x$, obeys a Poisson distribution with parameter λ.

What we will prove is a rather crude assertion consistent with this model. (Before this work, this was only known for $\epsilon \approx 0.24$.)

THEOREM 21.1 (GPY). *For any $\epsilon > 0$, there exist infinitely many p_n such that $p_{n+1} - p_n < \epsilon \log p_n$.*

Goldston et al also get a quantitative version of this, and they can even do this with $p_{n+1} - p_n < (\log p_n)^{1-\epsilon}$ for some specific $\epsilon > 0$. For simplicity, I won't get into these improvements. But I will discuss the following, whose proof is a good setup for the proof of Theorem 21.1.

THEOREM 21.2 (GPY). *Assume the Elliott-Halberstam conjecture for $Q = x^\theta$ with any fixed $\theta > 1/2$. Then there exists $c = c(\theta)$ such that there exist infinitely many p_n such that $p_{n+1} - p_n < c$. (If $\theta > 20/21$, one has $c(\theta) = 20$.)*

2. The approach

Fix a positive integer k; the basic idea is to try to prove a weak version of the Hardy-Littlewood k-tuples conjecture, for a k-tuple $\mathcal{H} = (h_1, \dots, h_k)$ of distinct integers. Namely, we'll try to prove that there are infinitely many n such that *at least two* of $n + h_1, \dots, n + h_k$ are prime; this would imply that there are infinitely many prime gaps no greater than $\max \mathcal{H} - \min \mathcal{H}$.

To do this, we will try to find an arithmetic function $a(n)$ with nonnegative values, such that for $j = 1, \dots, k$, we can establish

$$\sum_{x<n\leq 2x, n+h_j=p} a(n) > \frac{1}{k} \sum_{x<n\leq 2x} a(n). \tag{57}$$

If we had such a function, we could sum over j to obtain

$$\sum_{x<n\leq 2x} \#\{1 \leq j \leq k : n + h_j \text{ prime}\} \cdot a(n) > \sum_{x<n\leq 2x} a(n),$$

which would immediately imply that for some $x < n \leq 2x$, at least two of $n + h_1, \dots, n + h_k$ are prime. (Note: this strategy is poorly adapted to look for three or more primes in the same tuple. In fact, no satisfactory alternative has been proposed!)

Note that if $a(n)$ is supported only on those n for which $n + h_1, \dots, n + h_k$ is prime, then the k-tuples conjecture would imply (57), but we have no hope of proving (57) directly. Instead, we make a transition that is directly inspired by the transition from the combinatorial sieve to the Selberg sieve.

3. Selberg revisited

Namely, we pick a cutoff parameter R (which will ultimately depend on k and x), and choose $a(n)$ of the form

$$a(n) = \left(\sum_{d|(n+h_1)\cdots(n+h_k)} \rho(d) \right)^2$$

for some arithmetic function ρ with $\rho(1) = 1$ and support in $\{1, \dots, R\}$. As in Selberg's sieve, we have built the nonnegativity requirement into the construction, and we are now free to vary the values of ρ in order to maximize the ratio between the two sides of (57).

Unfortunately, we are not in as simple a situation as in Selberg's sieve, where we could simply diagonalize a quadratic form to find the desired minimum. In our case, we are comparing two different quadratic forms, which cannot be simultaneously diagonalized, and hitting the situation with Lagrange multipliers creates a mess. The best we can hope to do is to pick ρ of a special form with at least one parameter left in, run the calculation, and then optimize the choice of the parameter(s).

In Selberg's sieve, the optimal choice would have been

$$\rho(d) \approx \mu(d) \left(\frac{\log R/d}{\log R} \right)^k \qquad (d \leq R).$$

In our setting, we will instead put

$$\rho(d) = \mu(d) \left(\frac{\log R/d}{\log R} \right)^{k+\ell} \qquad (d \leq R) \tag{58}$$

for ℓ a nonnegative integer depending on k, in a fashion to be specified later.

4. Comparing the two sides

With this choice, one can calculate the two sides of (57) using the sorts of techniques we used in the first section of this course. I will postpone those calculations to a later unit, so that I can continue giving an overview of the method. First, here is what one gets for the right side of (57).

LEMMA 21.3. *With notation as above, there exist $C, c > 0$ depending on k, ℓ, such that for $R \leq x^{1/2}/(\log x)^C$,*

$$\sum_{x<n\leq 2x} a(n) = \frac{\mathfrak{S}(\mathcal{H})(k+\ell)!^2}{(k+2\ell)!(\log R)^{2k+2\ell}}\binom{2\ell}{\ell}x(\log R)^{k+2\ell} + O\left(\frac{x(\log x)^{k+2\ell-1}(\log\log x)^c}{(\log R)^{2k+2\ell}}\right).$$

The left side of (57) is more complicated, because of the extra restriction that $n+h_j$ must be prime. It is on this side that the arithmetic subtleties will creep in. Expanding the square, we get

$$\sum_{d_1,d_2\leq R} \rho(d_1)\rho(d_2)\#\{x < n \leq 2x : [d_1,d_2]|(n+h_1)\cdots(n+h_k), n+h_j \text{ prime}\}. \tag{59}$$

The count on the right side involves first pinning n down among some number of arithmetic progressions modulo the lcm $[d_1, d_2]$, then looking for primes in that arithmetic progression. Thus one expects to be able to approximate (59) by

$$\frac{x}{\log x}\sum_{d_1,d_2\leq R} \rho(d_1)\rho(d_2)\frac{g([d_1,d_2])}{\phi([d_1,d_2])}, \tag{60}$$

where g is the multiplicative function with $g(p) = v_{\mathcal{H}}(p) - 1$.

LEMMA 21.4. *With notation as above, there exist $C, c > 0$ depending on k, ℓ, such that for $R \leq x^{1/2}/(\log x)^C$, we have the following.*

(a) *For $h \notin \mathcal{H}$,* (60) *equals*

$$\frac{\mathfrak{S}(\mathcal{H},h)}{(\log R)^{2k+2\ell}}\frac{(k+\ell)!^2}{(k+2\ell)!}\binom{2\ell}{\ell}\frac{x}{\log x}(\log R)^{k+2\ell} + O\left(\frac{x(\log x)^{k+2\ell-2}(\log\log x)^c}{(\log R)^{2k+2\ell}}\right).$$

(b) *For $h \in \mathcal{H}$,* (60) *equals*

$$\frac{\mathfrak{S}(\mathcal{H})}{(\log R)^{2k+2\ell}}\frac{(k+\ell)!^2}{(k+2\ell+1)!}\binom{2(\ell+1)}{\ell+1}\frac{x}{\log x}(\log R)^{k+2\ell+1} + O\left(\frac{x(\log x)^{k+2\ell-1}(\log\log x)^c}{(\log R)^{2k+2\ell}}\right).$$

Crunching the numbers, we see that the ratio between (60) and the right side of (57) is asymptotic to

$$\frac{\log R}{\log x}\frac{2k(2\ell+1)}{(\ell+1)(k+2\ell+1)}. \tag{61}$$

The second fraction is always less than 4, but it tends to 4 as $k, \ell \to \infty$. Thus if we can safely approximate (59) by (60) in the range $R \leq x^{1/2-\epsilon}$, or even $R \leq x^{1/4+\epsilon}$, we get bounded gaps between primes. (Here's where we get stuck looking for three primes in one tuple: we can't hope to get past $R = x^{1/2-\epsilon}$ because of our earlier errors.) For instance, if we could take $R = x^{1/2-\epsilon}$, then already we get (60) with $k = 7, \ell = 1$. Using the 7-tuple $\mathcal{H} = \{11, 13, 17, 19, 23, 29, 31\}$, one then deduces that there are infinitely many prime gaps of size at most 20.

One can tweak the above argument by changing (58) to allow a polynomial $P(\log(R/d)/(\log R))$ instead of just a power. That polynomial must satisfy $P(1) = 1$ and must vanish to order at least k at 0. The quantity analogous to (61) is

$$\frac{\log R}{\log x} k \frac{\int_0^1 \frac{y^{k-2}}{(k-2)!} P^{(k-1)}(1-y)^2\,dy}{\int_0^1 \frac{y^{k-1}}{(k-1)!} P^{(k)}(1-y)^2\,dy}. \tag{62}$$

If we can take $R = x^{1/2-\epsilon}$, then one can get this ratio over 1 already with $k = 6$, so one gets infinitely many prime gaps bounded by 16 (using $\mathcal{H} = \{7, 11, 13, 17, 19, 23\}$) rather than 20. But even with the flexibility of choosing P, one can never get the second factor (excluding $(\log R)/(\log x)$) over 4 (exercise)!

5. The error terms, first attempt

None of the above matters unless we can control the discrepancy between (59) and (60). This discrepancy is spawned by error terms in the prime number theorem with moduli of the form $[d_1, d_2]$ for $d_1, d_2 \leq R$, so the moduli can run up to R^2.

Now recall what we know about these discrepancies. Let $\pi(x; N, m)$ be the number of primes $p \leq x$ congruent to m modulo N. If we allow Elliott-Halberstam, then for any fixed $A > 0$ and $\epsilon > 0$, there exists $c > 0$ such that

$$\sum_{q \leq Q} \max_{m \in (\mathbb{Z}/N\mathbb{Z})^*} \left| \pi(2x; N, m) - \pi(x; N, m) - \frac{x}{\phi(N) \log x} \right| \leq cx(\log x)^{-A}$$

for $Q = x^{1-\epsilon}$. This would allow taking $R = x^{1/2-\epsilon}$; we thus deduce Theorem 21.2.

Unconditionally, Bombieri-Vinogradov only allows $Q = x^{1/2-\epsilon}$. This looks like a disaster: we must take $R = x^{1/4-\epsilon}$, and so we can never get (57)! What now?

6. The error terms, second attempt

Remember that Theorem 21.1 is a much weaker assertion than the existence of infinitely many bounded gaps between primes; there is thus no need to insist on establishing (57) for any particular tuple $\mathcal{H}$. Instead, we are free to aggregate over all $\mathcal{H}$ in a certain range; to clarify, write $a(n; \mathcal{H})$ instead of $a(n)$ to indicate the dependence on $\mathcal{H}$.

Fix $\delta > 0$ for which we want infinitely many n with $p_{n+1} - p_n \leq H = \delta \log x$. We will now try to prove the inequality

$$\sum_{\mathcal{H} \in \{1,\dots,H\}^k} \sum_{1 \leq h \leq H, n+h=p} a(n, \mathcal{H}, h) > \frac{1}{h} \sum_{\mathcal{H} \in \{1,\dots,H\}^k} \sum_{1 \leq h \leq H, n+h=p} \sum_{x < n \leq 2x} a(n, \mathcal{H}), \tag{63}$$

which again is enough: now we get an n such that at least two of $n+1, \dots, n+h$ are prime.

For the right side of (63), Gallagher's result from the previous unit gives us the same asymptotics as before, except with $\mathfrak{S}(\mathcal{H})$ replaced by 1 and slightly worse error terms. We get an improvement on the left side, which we separate into terms with $h \notin \mathcal{H}$ and terms with $h \in \mathcal{H}$. We estimate the latter termse exactly as before; for the former terms, we note that if $n + h$ is prime, then

$$a(n; \mathcal{H}) = a(n; \mathcal{H}, h).$$

Namely, the difference comes from summands d which divide $(n+h_1)\cdots(n+h_k)(n+h)$ but not $(n+h_1)\cdots(n+h_k)$; those are all multiples of $n+h > x > R$, so $\rho(d) = 0$ for such d.

Thus we can simply appeal back to Lemma 21.3 with k replaced by $k+1$ and $\mathcal{H}$ replaced by $\mathcal{H}, h$. If we now compare the ratio of the two sides of (63), the contribution in the numerator from $h \in \mathcal{H}$ is exactly (61), to which we add the contribution $H/(\log x) = \delta$ from the terms with $h \notin \mathcal{H}$. As noted earlier, that's just enough to get over 1 with $R = x^{1/2-\epsilon}$ and k, ℓ sufficiently large. This yields Theorem 21.1.

Exercises

(1) Use the Poisson distribution model to compute a predicted distribution for the ratio $(p_{n+1} - p_n)/(\log p_n)$.
(2) Say we want to produce *large* gaps between primes. Take N to be the product of the primes up to m, and consider $N+2, \ldots, N+m$. For what function f does this imply $p_{n+1} - p_n > f(p_n)$ for infinitely many n?
(3) Let P be a polynomial with $P(1) = 1$ vanishing to order at least k at 0. Prove that the quantity (62) sans the factor $(\log R)/(\log x)$ is at most 4.

CHAPTER 22

Small gaps between primes (proofs)

Here are the missing calculations from the Goldston-Pintz-Yıldırım theorems. The reference is the article by Goldston et al cited in the previous unit.

1. Review of notation

Fix once and for all positive integers k, ℓ. Let x be a parameter tending to ∞. Let $\mathcal{H} = (h_1, \dots, h_k)$ be a k-tuple of distinct integers in the range $1, \dots, H$, where $H \leq \lambda \log x$ for some fixed λ. For p prime, we set

$$\Omega(p) = \text{image}(-\mathcal{H} \to \mathbb{Z}/p\mathbb{Z})$$

and $v_{\mathcal{H}}(p) = \#\Omega(p)$. Extend both of these by multiplicativity to squarefree d. We set

$$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{v_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

For the GPY method, we set

$$a(n) = \left(\sum_{d|(n+h_1)\cdots(n+h_k)} \rho(d)\right)^2$$

for

$$\rho(d) = \mu(d)\left(\frac{\log R/d}{\log R}\right)^{k+\ell} \qquad (d \leq R) \tag{64}$$

with $R \leq x^{1/2}/(\log x)^C$, where C is a constant depending on k, ℓ to be specified later. It will be more convenient to renormalize

$$\rho'(d) = \frac{(\log R)^{k+\ell}}{(k+\ell)!}\rho(d) = \mu(d)\frac{1}{(k+\ell)!}(\log R/d)^{k+\ell} \qquad (d \leq R).$$

2. The main calculation

LEMMA 22.1. *With notation as above, there exist $c > 0$ depending on k, ℓ, such that for C sufficiently large (depending on k, ℓ),*

$$\sum_{x < n \leq 2x} a(n) = \frac{\mathfrak{S}(\mathcal{H})(k+\ell)!^2}{(k+2\ell)!(\log R)^{2k+2\ell}}\binom{2\ell}{\ell}x(\log R)^{k+2\ell} + O\left(\frac{x(\log x)^{k+2\ell-1}(\log\log x)^c}{(\log R)^{2k+2\ell}}\right).$$

PROOF. Expanding the square in the definition of $a(n)$, we get the sum over d_1, d_2 of $\rho(d_1)\rho(d_2)$ times the number of $x < n \le 2x$ with $n \in \Omega(d_1), \Omega(d_2)$. That gives

$$\sum_{x<n\le 2x} a(n) = x\frac{(k+\ell)!^2}{(\log R)^{2k+2\ell}}\mathcal{T} + O\left(\left(\sum_d |v_{\mathcal{H}}(d)\rho(d)|\right)^2\right)$$

$$\mathcal{T} = \sum_{d_1,d_2} \frac{v_{\mathcal{H}}([d_1,d_2])}{[d_1,d_2]}\rho'(d_1)\rho'(d_2)$$

since $|\Omega(d)| \le \tau_k(d)$, we can replace the error term with $O(R^2(\log R)^c)$.

We now convert over to a problem in complex analysis, as in the first section of the course. The key formula is

$$\rho'(d) = \frac{\mu(d)}{2\pi i}\int_{(1)} (R/d)^s \frac{ds}{s^{k+\ell+1}},$$

where (α) denotes the vertical contour $\alpha - i\infty \to \alpha + i\infty$. This gives

$$\mathcal{T} = \frac{1}{(2\pi i)^2}\int_{(1)}\int_{(1)} F(s_1,s_2,\mathcal{H})\frac{R^{s_1+s_2}}{(s_1s_2)^{k+\ell+1}}ds_1\,ds_2,$$

with

$$\begin{aligned} F(s_1,s_2,\mathcal{H}) &= \sum_{d_1,d_2}\mu(d_1)\mu(d_2)\frac{v_{\mathcal{H}}([d_1,d_2])}{[d_1,d_2]d_1^{s_1}d_2^{s_2}} \\ &= \prod_p \left(1 - \frac{v_{\mathcal{H}}(p)}{p}(p^{-s_1}+p^{-s_2}-p^{-s_1-s_2})\right) \end{aligned}$$

in the region of absolute convergence.

Now put

$$G(s_1,s_2,\mathcal{H}) = F(s_1,s_2,\mathcal{H})\left(\frac{\zeta(s_1+1)\zeta(s_2+1)}{\zeta(s_1+s_2+1)}\right)^k.$$

Since $v_{\mathcal{H}}(p) = k$ for almost all p, this function is holomorphic and bounded for $\mathrm{Re}(s_1), \mathrm{Re}(s_2) > -c$. In particular, we recover the singular series as

$$\mathfrak{S}(\mathcal{H}) = G(0,0,\mathcal{H}).$$

Now note that from the Euler product expansion, we see that for $\min\{\mathrm{Re}(s_1), \mathrm{Re}(s_2), 0\} = \sigma \ge -c$, we have

$$G(s_1,s_2,\mathcal{H}) = O(\exp(c(\log x)^{-2\sigma}\log\log\log x)). \tag{65}$$

(More specifically, we can uniformly bound the Euler products over $p \le k^2$ and $p > H$; we get the quoted estimate from the range $k^2 < p \le H$.)

We use (65) to truncate the infinite integral, but first we shift the contours. Put $U = \exp(\sqrt{\log x})$. We shift the s_1-contour to $L_1 = (\log U)^{-1} + it$, and the s_2-contour to $L_2 = (2\log U)^{-1} + it$. If we now truncate to $|t| \le U$ and $|t| \le U/2$, respectively, we have

$$\begin{aligned} \mathcal{T} = \frac{1}{(2\pi i)^2}\int_{L_2}\int_{L_1} G(s_1,s_2,\mathcal{H})\left(\frac{\zeta(s_1+s_2+1)}{\zeta(s_1+1)\zeta(s_2+1)}\right)^k \frac{R^{s_1+s_2}}{(s_1s_2)^{k+\ell+1}}ds_1\,ds_2 \\ + O(\exp(-c\sqrt{\log x})). \end{aligned}$$

We now shift the s_1-contour again, this time to $L_1' = -(\log U)^{-1} + it$ with $|t| \leq U$; we pick up residues at $s_1 = 0$ and $s_1 = -s_2$. Again using (65), we get

$$\mathcal{T} = \frac{1}{2\pi i} \int_{L_2} (\mathrm{Res}_{s_1=0} + \mathrm{Res}_{s_1=-s_2})\ ds_2 + O(\exp(-c\sqrt{\log x})).$$

We wish to show that the residue at $s_1 = -s_2$ may be neglected, by rewriting it in terms of the integral over the circle $|s_1 + s_2| = (\log x)^{-1}$. In this integral, $G(s_1, s_2, \Omega) = O((\log\log x)^c)$, $R^{s_1+s_2} = O(1)$, $\zeta(s_1 + s_2 + 1) = O(\log x)$. We also have

$$(s_1\zeta(s_1 + 1))^{-1} = O((|s_2| + 1)^{-1} \log(|s_2| + 2)).$$

Putting this together,

$$\mathrm{Res}_{s_1=-s_2} \leq O\left((\log x)^{k-1}(\log\log x)^c \left(\frac{\log(|s_2| + 2)}{|s_2| + 1}\right)^{2k} |s_2|^{-2\ell-2}\right),$$

so

$$\mathcal{T} = \frac{1}{2\pi i} \int_{L_2} \mathrm{Res}_{s_1=0}\ ds_2 + O((\log x)^{k+\ell}). \tag{66}$$

It remains to deal with $\mathrm{Res}_{s_1=0}$; note that the pole has order $\ell + 1$. If I put

$$Z(s_1, s_2, \mathcal{H}) = G(s_1, s_2, \mathcal{H}) \left(\frac{(s_1 + s_2)\zeta(s_1 + s_2 + 1)}{s_1\zeta(s_1 + 1)s_2\zeta(s_2 + 1)}\right)^k,$$

then $Z(s_1, s_2, \mathcal{H})$ is holomorphic near $(0, 0)$, and

$$\mathrm{Res}_{s_1=0} = \frac{R^{s_2}}{\ell! s_2^{\ell+1}} \left(\frac{\partial}{\partial s_1}\right)^{\ell}_{s_1=0} \left(\frac{Z(s_1, s_2, \mathcal{H})}{(s_1 + s_2)^k} R^{s_1}\right).$$

We now stuff this into (66) and repeat the operation: that is, we shift the s_2-contour to $L_2' : -(2\log U)^{-1} + it$ for $|t| \leq U/2$. Again, the new integral is $O(\exp(-c\sqrt{\log x}))$, so all that is left is the residue at $s_2 = 0$. In other words,

$$\mathcal{T} = \mathrm{Res}_{s_2=0} \mathrm{Res}_{s_1=0} + O((\log N)^{k+\ell}).$$

This constitutes success: we have isolated the integral at the point $(0, 0)$, so now we will have no trouble evaluating it.

Fix some $\rho > 0$ small, let C_1 be the circle $|s_1| = \rho$, and let C_2 be the circle $|s_2| = 2\rho$. Then

$$\mathcal{T} = \frac{1}{(2\pi i)^2} \int_{C_2} \int_{C_1} \frac{Z(s_1, s_2, \mathcal{H}) R^{s_1+s_2}}{(s_1 + s_2)^k (s_1 s_2)^{\ell+1}} ds_1\, ds_2 + O((\log x)^{k+\ell}).$$

We now change variables to s, ξ where $s_1 = s$ and $s_2 = s\xi$, over the contours $C : |s| = \rho$ and $C' : |\xi| = 2$. By the same argument as in the runup to (66) (applied to s), this gives

$$\mathcal{T} = \frac{Z(0, 0)}{2\pi i (k + 2\ell)!} (\log R)^{k+2\ell} \int_{C'} \frac{(\xi + 1)^{2\ell}}{\xi^{\ell+1}}\, d\xi + O((\log x)^{k+2\ell-1}(\log\log x)^c).$$

We can now read off the residue of the integrand as $\binom{2\ell}{\ell}$, completing the proof. $\square$

3. Twisting with primes

The second estimate proceeds mostly the same way, so I will skip most details. Note that the translation trick from last time means we don't have to worry about case (b): if $h \in \mathcal{H}$ and $n+h$ is prime, then $a(n,\mathcal{H}) = a(n,\mathcal{H},h)$.

LEMMA 22.2. *With notation as above, there exist $c > 0$ depending on k, ℓ, such that for C sufficiently large (depending on k, ℓ), we have the following.*

(a) *For $h \notin \mathcal{H}$, the quantity*

$$\frac{x}{\log x} \sum_{d_1,d_2 \leq R} \rho(d_1)\rho(d_2) \frac{g([d_1,d_2])}{\phi([d_1,d_2])}, \tag{67}$$

where g is the multiplicative function with $g(p) = v_{\mathcal{H}}(p) - 1$, equals

$$\frac{\mathfrak{S}(\mathcal{H},h)}{(\log R)^{2k+2\ell}} \frac{(k+\ell)!^2}{(k+2\ell)!} \binom{2\ell}{\ell} \frac{x}{\log x} (\log R)^{k+2\ell} + O\left(\frac{x(\log x)^{k+2\ell-2}(\log\log x)^c}{(\log R)^{2k+2\ell}} \right).$$

(b) *For $h \in \mathcal{H}$, (67) equals*

$$\frac{\mathfrak{S}(\mathcal{H})}{(\log R)^{2k+2\ell}} \frac{(k+\ell)!^2}{(k+2\ell+1)!} \binom{2(\ell+1)}{\ell+1} \frac{x}{\log x} (\log R)^{k+2\ell+1} + O\left(\frac{x(\log x)^{k+2\ell-1}(\log\log x)^c}{(\log R)^{2k+2\ell}} \right).$$

PROOF. It is a bit more convenient to multiply both sides by $\log x$, pull $\log x$ into the summand, then replace it by $\Lambda(n)$; by the prime number theorem (and the fact that I'm working in a dyadic range), this does not affect the outcome.

In a similar fashion as above, we end up dealing with the exrpession

$$\mathcal{T}' = \frac{1}{(2\pi i)^2} \int_{(1)} \int_{(1)} \prod_p \left(1 - \frac{v_{\mathcal{H},h}(p) - 1}{p-1} (p^{-s_1} + p^{-s_2} - p^{-s_1-s_2}) \right) \frac{R^{s_1+s_2}}{(s_1 s_2)^{k+\ell+1}} \, ds_1 \, ds_2.$$

Everything proceeds as before *unless* $v_{\mathcal{H},h}(p) = p$ for some p. In that case, the Euler product above vanishes at one of $s_1 = 0$ or $s_2 = 0$, to order equal to the number of primes for which $v_{\mathcal{H},h}(p) = 0$. But this can only happen for $p \leq k+1$, and so we can still proceed as above: all that changes is that now the main term vanishes, consistent with $\mathfrak{S}(\mathcal{H} \cup \{h\}) = 0$. □

Part 5

Additional topics

CHAPTER 23

Artin L-functions and the Chebotarev density theorem

This unit begins the fourth and final part of the course. In this part, we describe some other types of L-functions that are used for arithmetic purposes. This merely scratches the surface of what is now a rather vast theory of L-functions; §5 of Iwaniec-Kowalski gives a somewhat less narrow account.

Some of this discussion will only make sense if you have studied some algebraic number theory. The book I used to teach 18.786 last year is a reasonable place to start: it is Janusz, *Algebraic Number Fields*. I'm also presuming you are happy with representation theory of finite groups at the level of 18.702.

1. Frobenius elements of Galois groups

Let K be a finite Galois extension of $\mathbb{Q}$, and put $G = \mathrm{Gal}(K/\mathbb{Q})$. Let $\mathfrak{o}_K$ be the ring of integers of K; that is, $\alpha \in \mathfrak{o}_K$ if and only if α is a root of a monic polynomial with coefficients in $\mathbb{Z}$.

A prime p is said to be *ramified (in K)* if the ring $\mathfrak{o}_K/p\mathfrak{o}_K$ is not reduced (i.e., has nilpotent elements). For instance, if $K = \mathbb{Q}(i)$, then $\mathfrak{o}_K = \mathbb{Z}[i]$, and the only ramified prime is $p = 2$. In general, only finitely many primes are ramified; they are the ones dividing the discriminant of $K/\mathbb{Q}$.

On $\mathfrak{o}_K/p\mathfrak{o}_K$, one has both an action of G and a Frobenius map $x \mapsto x^p$.

LEMMA 23.1. *There exists $g \in G$ such that $x^p = x^g$ has a nonzero solution $x \in \mathfrak{o}_K/p\mathfrak{o}_K$. Moreover, if p is unramified, then the set of such $g \in G$ forms a single conjugacy class.*

Any such g is called a *Frobenius element* for the prime p.

One can also define Frobenius elements for the infinite place: given an embedding of K into $\mathbb{C}$, complex conjugation on $\mathbb{C}$ induces an automorphism of K. Any such automorphism is called a *Frobenius element* for the infinite place, or more simply a *complex conjugation* on K.

2. Linear representations and L-functions

Let $\rho : G \to \mathrm{GL}_n(\mathbb{C})$ be a linear representation, with character $\chi : G \to \mathbb{C}$; that is, $\chi(g) = \mathrm{Trace}\,\rho(g)$. We define the *(incomplete) Artin L-function* associated to ρ as the function

$$L(\rho, s) = \prod_p \det(1 - \rho(\mathrm{Frob}_p)p^{-s})^{-1},$$

where we only allow p to run over unramified primes, and Frob_p means any Frobenius element of p; it doesn't matter which one because they are all conjugate.

(There is a correct way to put in the ramified primes: you only look at the determinant of the action of ρ on the invariants under an inertia group corresponding to p. If you don't know what that means, never mind.)

For example, if we take $\rho : G \to \mathrm{GL}_1(\mathbb{C})$ to be the trivial representation, then $L(\rho, s)$ equals the Riemann zeta function with a few Euler factors missing. Note also that

$$L(\rho_1 \oplus \rho_2, s) = L(\rho_1, s)L(\rho_2, s).$$

Also, if K' is another Galois extension of $\mathbb{Q}$ contained in K, ρ factors through $\mathrm{Gal}(K'/\mathbb{Q})$, then the L-functions computed in terms of K and K' agree, in the sense that for each p which appears in both products (which is all but finitely many), the Euler factor is the same.

Note that

$$\left|1 - \det(1 - \rho(\mathrm{Frob}_p)p^{-s})^{-1}\right| = O(p^{-s}),$$

where the implied constant depends on the dimension of ρ. Consequently, the Euler product converges absolutely for $\mathrm{Re}(s) > 1$, uniformly for $\mathrm{Re}(s) \geq 1 + \epsilon$, and never vanishes in this region.

3. Artin's conjecture

The following is one of the deepest conjectures in modern number theory.

Conjecture 23.2 (Artin). *The function $L(\rho, s)$ extends to a meromorphic function on all of $\mathbb{C}$, with no poles away from $s = 1$, and order of pole at $s = 1$ equal to the number of copies of the trivial representation contained in ρ (or equivalently, $1/|G| \sum_{g \in G} \chi(g)$).*

There are various stronger versions of this conjecture. For instance, there is also supposed to be a functional equation relating $L(\rho, s)$ with $L(\overline{\rho}, 1 - s)$. More comprehensively, there should be some sort of analogue for $L(\rho, s)$ of the function θ that we used for the proof of the functional equation of the Riemann zeta function. (Such a thing would be an example of a *modular form*.)

Here are some results about Artin's conjecture.

(1) It holds for ρ trivial, by reducing to the Riemann zeta function.
(2) It holds for ρ of dimension 1: by the Kronecker-Weber theorem, any such ρ is a Dirichlet character, and so we get a Dirichlet L-function.
(3) It holds if ρ is induced by a permutation representation (e.g., the regular representation). In this case, this follows from the analytic properties of Dedekind ζ-functions. More generally, it holds if ρ is obtained by induction from a one-dimensional representation; see below.
(4) If ρ has dimension 2, then either the image of ρ is solvable, in which case the conjecture is a theorem of Langlands and Tunnell, or the image is the icosahedral group A_5. In the latter case, the conjecture is known when ρ is *odd* (any complex conjugation has determinant -1) by recent results from the theory of modular forms (resolution of Serre's conjecture).

4. Induced representations

Let H be a subgroup of G and let $\sigma : H \to \mathrm{GL}_m(\mathbb{C})$ be a linear representation. Let V be the set of functions $f : G \to \mathbb{C}^m$ such that $h(f(g)) = f(hg)$ for all $h \in H$. This forms a representation of G notated $\mathrm{Ind}_H^G(\sigma)$. For instance, if σ is the trivial

representation, then $\mathrm{Ind}_H^G(\sigma)$ is the linear representation given by the permutation representation of G on the cosets of H.

THEOREM 23.3 (Artin). *Every character of G is a $\mathbb{Q}$-linear combination of characters of induced representations from cyclic subgroups.*

PROOF. By orthogonality of characters, it suffices to check that for each conjugacy class in G, one can construct a linear combination of induced representations whose character is nonvanishing except on that class. Let g be an element of the class, generating the cyclic subgroup H. Then there is a linear combination of one-dimensional representations of H whose character is nonzero only on g; inducing those to G gives the linear combination we seek. □

5. Chebotarev's density theorem

Here is a weaker form of Artin's conjecture that can be proved, but even this requires some heavy machinery.

THEOREM 23.4. *For any ρ, the L-function $L(\rho, s)$ extends to a meromorphic function on a neighborhood of $\mathrm{Re}(s) \geq 1$. Moreover, on the line $\mathrm{Re}(s) = 1$, $L(\rho, s)$ is nonvanishing for $s \neq 1$, and for $s = 1$, the order of vanishing of $L(\rho, s)$ is $-1/|G| \sum_{g \in G} \chi(g)$. (In other words, there is a pole at $s = 1$ of order equal to the multiplicity of the trivial representation in ρ.)*

SKETCH OF PROOF. One first proves the claim for $\rho = \mathrm{Ind}_H^G \sigma$ for any abelian subgroup H of G and any one-dimensional representation $\sigma : H \to \mathrm{GL}_1(\mathbb{C})$. This requires class field theory in general, because one has to first write σ as a ray class character. See Janusz's book for details.

By Artin's theorem, for any given ρ, we deduce the claim for $\rho^{\oplus m}$. To deduce the claim for ρ, we may reduce to the case where ρ has no trivial subrepresentations; then $L(\rho, s)^m$ extends holomorphically to a neighborhood of $\mathrm{Re}(s) \geq 1$, without vanishing anywhere. Consequently, we can take the m-th root in a neighborhood of any s with $\mathrm{Re}(s) = 1$; by choosing the right root, we get a function that patches together with the function defined on $\mathrm{Re}(s) > 1$. □

By imitating the proof of Dirichlet's theorem, we deduce the following theorem of Chebotarev, which can be considered a nonabelian generalization of Dirichlet's theorem.

THEOREM 23.5 (Chebotarev). *For any conjugacy class C of G, the set of primes p for which $\mathrm{Frob}_p \in C$ has natural density $\#C/\#G$.*

It is also possible to prove this without using class field theory, as in Chebotarev's original work. (In fact, this result was one of the original impetuses for class field theory to be developed!) There is a nice explanation of this by Lenstra and Stevenhagen, available online at:

`http://www.math.leidenuniv.nl/~hwl/papers/cheb.pdf`

6. Exercises (optional)

Remember, there are no more problem sets, so don't bother turning these in.

(1) Suppose that G is a group in which any two elements that generate the same cyclic subgroup are conjugate (e.g., S_n). Prove that every character of G is a $\mathbb{Q}$-linear combination of permutation representations. For such G, you don't need class field theory to prove Chebotarev as above, just the analytic properties of Dedekind zeta functions.
(2) Let P be a polynomial with Galois group G. Use Chebotarev's theorem to compute the density of primes modulo which P factors into irreducibles of degrees $d_1, \ldots, d_k$. (This was proved by Frobenius, who then made a conjecture that became Chebotarev's theorem.)

CHAPTER 24

Elliptic curves and their L-functions

The standard book on elliptic curves is Silverman's *The Arithmetic of Elliptic Curves.*

1. Elliptic curves and their L-functions

An *elliptic curve* over a field K is a nonsingular cubic plane curve. If K has characteristic > 2, any such curve can be put in the form $y^2 = P(x)$, where $P(x) = x^3 + ax^2 + bx + c$ is a polynomimal with no repeated roots. (This is a pretty *ad hoc* definition; see Silverman's book for a proper definition.)

If $E : y^2 = P(x)$ is an elliptic curve over $\mathbb{Q}$, then for all but finitely many primes p, the reduction of E modulo p is a nonsingular cubic, and hence elliptic curve over $\mathbb{F}_p$. (Warning: the finite set of bad primes depends on the choice of the equation $y^2 = P(x)$, not just on the isomorphism class of E. There is an optimal choice of the defining equation, but we won't use that here.) We define the *L-function* of E as the product

$$L(E, s) = \prod_p (1 - a_p p^{-s} + p^{1-2s})^{-1}$$

over all of the nonexceptional primes, where $p + 1 - a_p$ is the number of points on E modulo $\mathbb{F}_p$. (Remember that E is being drawn in the projective plane, so you have to count the one point $[0 : 1 : 0]$ at infinity.)

It's obvious that there are at most $2p + 1$ points on E, two for each possible x-coordinate plus one point at infinity; that implies that $L(E, s)$ converges absolutely for $\mathrm{Re}(s) > 2$. Actually one can do better.

LEMMA 24.1 (Hasse). *We have $|a_p| \leq 2\sqrt{p}$ for all p.*

This means that $L(E, s)$ actually converges absolutely for $\mathrm{Re}(s) > 3/2$.

In a few cases, the a_p exhibit predictable behavior. For instance, if E is the curve $x^3 + y^3 = 1$, then for $p \equiv 2 \pmod 3$, we have $a_p = 0$, whereas for $p \equiv 1 \pmod 3$, we can write a_p in terms of integers A, B for which $A^2 + 3B^2 = p$ (this was first observed by Gauss). In most cases, however, no such easy formula exists.

By contrast, suppose E is a nonsingular conic curve passing through at least one $\mathbb{Q}$-rational point; then $\#E(\mathbb{F}_p) = p + 1$ always. (The points on the curve are in bijection with the lines through the given point.)

THEOREM 24.2. *The function $L(E, s)$ extends to a holomorphic function on $\mathbb{C}$, satisfying a functional equation between s and $2 - s$.*

This is a consequence of the *modularity of elliptic curves.* This theorem is the result of work of Wiles, Taylor-Wiles, Diamond, Fujiwara, Conrad-Diamond-Taylor, and Breuil-Conrad-Diamond-Taylor. (Whew!) When combined with a theorem of

Ribet (part of Serre's conjecture), the modularity of elliptic curves (actually just the special case proved by Wiles) resolves the Fermat problem.

There is also an amazing conjecture relating the L-function to the $\mathbb{Q}$-rational points of E. (The Mordell-Weil theorem states that $E(\mathbb{Q})$ is a finitely generated abelian group.)

CONJECTURE 24.3 (Birch, Swinnerton-Dyer). *The order of vanishing of $L(E, s)$ at $s = 1$ equals the rank of the finitely generated abelian group $E(\mathbb{Q})$.*

This is known by work of Kolyvagin, Kato, Gross-Zagier, etc. in case the order of vanishing is 0 or 1.

CHAPTER 25

The Sato-Tate distribution

Source: Serre's book *Abelian ℓ-adic representations and elliptic curves*, appendix to chapter 1.

1. Equidistribution on compact groups

Let X be a compact topological space. Let $C(X)$ be the space of continuous functions $X \to \mathbb{C}$; this is a Banach space under the supremum norm. Let μ be a measure on X, i.e., a continuous linear map $C(X) \to \mathbb{C}$ which is nonnegative (i.e., the integral of a function taking nonnegative real values is nonnegative) and of total measure 1.

A sequence $x_1, x_2, \ldots$ of elements of X is *equidistributed* with respect to μ if for any continuous function f,

$$\int_X f \, d\mu = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(x_i).$$

2. Topological groups

The key example for us is when X is a compact Lie group (e.g., a finite group), and K is the space of conjugacy classes of X (viewed with the quotient topology from G). In this case, K has a unique translation-invariant measure with total measure 1, called the *Haar measure*; we use this measure on X and on K.

THEOREM 25.1 (Peter-Weyl). *With notation as above, the sequence $x_1, x_2, \ldots$ is equidistributed with respect to the Haar measure μ if and only if for any irreducible character $\chi : G \to \mathbb{C}$ of G,*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \chi(x_i) = \int_X \chi \, d\mu.$$

Note that the integral on the right is 1 if χ is the trivial character and 0 otherwise (orthogonality of characters).

3. L-functions and equidistribution

Here is a big generalization of our approach to Chebotarev's density theorem. Take K and X as in the previous example. Let $x_1, x_2, \ldots$ be a sequence of elements of X, and let $x_i \to N(x_i)$ be a function whose values are all integers at least 2. We make the following additional hypotheses.

(i) Assume that the Euler product

$$\prod_i (1 - N(x_i)^{-s})^{-1}$$

converges absolutely for $\mathrm{Re}(s) > 1$, and extends to a meromorphic function on a neighborhood of $\mathrm{Re}(s) \geq 1$ with no zeroes or poles in $\mathrm{Re}(s) \geq 1$ except for a simple pole at $s = 1$.

(ii) Let ρ be any irreducible representation of K with character χ. Put

$$L(s, \rho) = \prod_i \det(1 - \rho(x_i)N(x_i)^{-s})^{-1}.$$

(Note that $\rho(x_i)$ is only defined up to conjugation.) Then $L(s, \rho)$ converges absolutely for $\mathrm{Re}(s) > 1$, and extends to a meromorphic function on a neighborhood of $\mathrm{Re}(s) \geq 1$ with no zeroes or poles in $\mathrm{Re}(s) \geq 1$ except possibly at $s = 1$.

THEOREM 25.2. *The number of x_i with $N(x_i) \leq n$ is asymptotic to $n/\log n$ as $n \to \infty$. Moreover, for any irreducible character χ of G,*

$$\sum_{i:N(x_i)\leq n} \chi(x_i) = c(\chi)n/\log n + o(n/\log n),$$

where $-c(\chi)$ is the order of vanishing of $L(s, \rho)$ at $s = 1$.

PROOF. Yet another straightforward generalization of our original proof of the prime number theorem. □

COROLLARY 25.3. *Assume that there exists c such that for any $n \in \mathbb{Z}$, there are at most c values of i with $N(x_i) \leq c$. Then the x_i are equidistributed for Haar measure if and only if $c(\chi) = 0$ for every nontrivial irreducible character at χ.*

This reproduces the Chebotarev density theorem from the previous unit.

4. The Sato-Tate conjecture

The following is a rather nonobvious application of the above formalism.

CONJECTURE 25.4 (Sato-Tate). *Suppose E does not have complex multiplication. Let α_p be the root of $x^2 - a_p x + p$ with nonnegative imaginary part. Then $\arg(\alpha_p/\sqrt{p})$ is equidistributed in $[0, \pi]$ for the measure $\frac{2}{\pi}\sin^2\theta d\theta$.*

What does the condition that E does not have complex multiplication mean? The points of E naturally form an abelian group, in which three points add to 0 if and only if they are collinear. We say E has *complex multiplication* if the only endomorphisms of E as an algebraic group are multiplication by integers. (Over $\mathbb{C}$, E forms a Riemann surface which looks like the quotient of $\mathbb{C}$ by a lattice; an endomorphism of E corresponds to a complex number which multiplies the lattice into itself.)

THEOREM 25.5 (Clozel, Harris, Taylor). *The Sato-Tate conjecture holds if $j(E) \notin \mathbb{Z}$. (This implies that E does not have complex multiplication.)*

I'll skip the definition of the j-invariant E for now; see Silverman's book.

5. Equidistribution and Sato-Tate

How does the elliptic curve example relate to Sato-Tate? Put $K = SU(2)$, the group of 2×2 unitary matrices of determinant 1. Any class in X contains a unique matrix of the form

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \qquad 0 \leq \theta \leq \pi.$$

Thus we may use the α_p's to generate elements x_p of X by taking $\theta = \arg(\alpha_p/\sqrt{p})$. The Haar measure on X is precisely the Sato-Tate measure, so we are reduced to asking whether the x_p are equidistributed.

The irreducible representations of K are just the symmetric powers of the standard 2-dimensional representation. Hence Sato-Tate reduces to the following, which is the real hard content in the work of Clozel-Harris-Taylor. (Note that you have to shift the abscissa of absolute convergence by 1/2.)

THEOREM 25.6. *Let $P_n(T)$ be the polynomial with constant coefficient 1 and roots $\alpha_p^n, \alpha_p^{n-1}\overline{\alpha_p}, \ldots, \overline{\alpha_p}^n$. If $j(E) \notin \mathbb{Z}$, then the Euler product*

$$\prod_p P_n(p^{-s})^{-1}$$

extends to a holomorphic function on $\mathbb{C}$. (Since the Euler product converges absolutely for $\mathrm{Re}(s) > 3/2$, the product cannot vanish for $\mathrm{Re}(s) \geq 3/2$.)

Exercises (optional)

(1) Let $\alpha_1, \ldots, \alpha_m$ be real numbers such that $1, \alpha_1, \ldots, \alpha_m$ are linearly independent over $\mathbb{Q}$. Apply Weyl's criterion to prove that the sequence $x_n = (n\alpha_1, \ldots, n\alpha_m) \in (\mathbb{R}/\mathbb{Z})^m$ is equidistributed for the usual measure.

(2) Prove that the sequence $\log n$ is not uniformly distributed for *any* measure on $\mathbb{R}/\mathbb{Z}$.