

Introduction

The Riemann zeta function is the function $\zeta(s)$, defined for $\operatorname{Re} s > 1$ by

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

It has a meromorphic continuation to the complex plane, with a simple pole at $s = 1$ with residue 1. Completing this function with the factor for the archimedean valuation, $\zeta_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2)$, one obtains the function

$$\zeta_{\mathbb{Z}}(s) = \zeta_{\mathbb{R}}(s)\zeta(s). \quad (1)$$

The function $\zeta_{\mathbb{Z}}$ is meromorphic on \mathbb{C} with simple poles at 0 and 1, and it satisfies the functional equation $\zeta_{\mathbb{Z}}(1-s) = \zeta_{\mathbb{Z}}(s)$. This is proved by Riemann [Ri1] using the “Riemann–Roch” formula¹

$$\theta(t^{-1}) = t\theta(t), \quad (2)$$

where $\theta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t^2}$ is closely related to the so-called theta-function.² For $\operatorname{Re} s > 1$, the zeta function satisfies the Euler product

$$\zeta_{\mathbb{Z}}(s) = \zeta_{\mathbb{R}}(s) \prod_p \frac{1}{1 - p^{-s}}, \quad (3)$$

where the product is taken over all prime numbers. It follows that the zeros of $\zeta_{\mathbb{Z}}$ all lie in the vertical strip $0 \leq \operatorname{Re} s \leq 1$.³ The Riemann

¹ Formula (2) goes back to Cauchy and is called “Riemann–Roch Theorem” in Tate’s thesis.

² The classical theta function is defined as $\vartheta(z, \tau) = \sum_{n=-\infty}^{\infty} e^{\pi i n^2 \tau} e^{2\pi i n z}$. Its relation to θ is $\theta(t) = \vartheta(0, it^2)$.

³ It is also known that the zeros do not lie on the boundary of this “critical strip” [In, Theorem 19, p. 58] and [vF1, Theorem 2.4].

hypothesis states that these zeros all lie on the line $\operatorname{Re} s = 1/2$:

Riemann hypothesis: $\zeta_{\mathbb{Z}}(s) = 0$ implies $\operatorname{Re} s = 1/2$.

See [Ri1] and [Ed, Har2, Lap-vF1, Lap-vF2, Ta] for more information about the Riemann and other zeta functions.

In this exposition, we prove the Riemann hypothesis for the zeta function of a nonsingular curve over a finite field. Let q be a power of a prime number p , and let $m(T, X)$ be a polynomial in two variables with coefficients in \mathbb{F}_q , the finite field with q elements. The equation $m(T, X) = 0$ defines a curve \mathcal{C} over \mathbb{F}_q , which we assume to be nonsingular. Let $N_{\mathcal{C}}(n)$ be the number of solutions of the equation $m(t, x) = 0$ in the finite set $\mathbb{F}_{q^n} \times \mathbb{F}_{q^n}$. Thus $N_{\mathcal{C}}(n)$ is the number of points on \mathcal{C} with coordinates in \mathbb{F}_{q^n} . A famous theorem of F. K. Schmidt in 1931 (see [fSch] and [Has1, Has2, Tr]) says that there exist an integer g , the genus of \mathcal{C} , and algebraic numbers $\omega_1, \dots, \omega_{2g}$, such that⁴

$$N_{\mathcal{C}}(n) = q^n - \sum_{\nu=1}^{2g} \omega_{\nu}^n + 1.$$

We also define $N_{\mathcal{C}}(0) = 2 - 2g$. From the formula for $N_{\mathcal{C}}(n)$, the Mellin transform (generating function)

$$\mathcal{M}N_{\mathcal{C}}(s) = \sum_{n=0}^{\infty} N_{\mathcal{C}}(n)q^{-ns}$$

can be computed as a rational function of q^{-s} ,

$$\mathcal{M}N_{\mathcal{C}}(s) = \frac{1}{1 - q^{1-s}} - \sum_{\nu=1}^{2g} \frac{1}{1 - \omega_{\nu}q^{-s}} + \frac{1}{1 - q^{-s}}.$$

We define the *zeta function of \mathcal{C}* by

$$\zeta_{\mathcal{C}}(s) = q^{s(g-1)} \frac{\prod_{\nu=1}^{2g} (1 - \omega_{\nu}q^{-s})}{(1 - q^{1-s})(1 - q^{-s})},$$

so that $\mathcal{M}N_{\mathcal{C}}$ is recovered as its logarithmic derivative,

$$-\frac{\zeta'_{\mathcal{C}}(s)}{\zeta_{\mathcal{C}}(s)} = (\mathcal{M}N_{\mathcal{C}}(s) + g - 1) \log q.$$

⁴ In $N_{\mathcal{C}}(n)$, also the finitely many points “at infinity” need to be counted. See Chapter 5.

The function $\zeta_{\mathcal{C}}$ satisfies the functional equation $\zeta_{\mathcal{C}}(1-s) = \zeta_{\mathcal{C}}(s)$. This functional equation can be proved using the Riemann–Roch theorem

$$l(D) = \deg D + 1 - g + l(\mathcal{K} - D), \quad (4)$$

which is the analogue of (2) above. This theorem is proved in Chapter 3.

Since the zeta function of \mathcal{C} is a rational function of q^{-s} , it is periodic with period $2\pi i / \log q$. It has two simple poles at $s = 0$ and 1 , and $2g$ zeros at $s = \log_q \omega_\nu$, all repeated modulo $2\pi i / \log q$. It satisfies an Euler product, analogous to (3), which converges for $\operatorname{Re} s > 1$,

$$\zeta_{\mathcal{C}}(s) = q^{s(g-1)} \prod_v \frac{1}{1 - q^{-s \deg v}},$$

where the product is taken over all orbits of the Frobenius flow on \mathcal{C} , and $\deg v$ is the length of an orbit. It follows that $1 \leq |\omega_\nu| \leq q$. Artin conjectured⁵ that $\zeta_{\mathcal{C}}(s)$ has its zeros on the line $\operatorname{Re} s = 1/2$. In terms of the exponentials of the zeros, the numbers ω_ν , this means that

$$|\omega_\nu| = \sqrt{q} \quad \text{for every } \nu = 1, \dots, 2g.$$

This is the analogue of the Riemann hypothesis for $\zeta_{\mathcal{C}}$. It is trivially verified for $\mathcal{C} = \mathbb{P}^1$, when $g = 0$ and $\zeta_{\mathcal{C}}$ does not have any zeros. It was proved by H. Hasse in the case of elliptic curves ($g = 1$), and first in full generality by A. Weil.⁶ Later proofs, based on one of Weil’s proofs, have been given by P. Roquette [Roq] and others. Weil uses the intersection of divisors with the graph of Frobenius in $\mathcal{C} \times \mathcal{C}$, and his second proof uses the action of Frobenius on the embedding of \mathcal{C} into its Jacobian (see [Ros, Appendix]). There have been some attempts to translate the first and second proof to the Riemann zeta function, when \mathcal{C} is the “curve” $\operatorname{spec} \mathbb{Z}$, but so far without success, one of the obstacles being that in the category of schemes, $\operatorname{spec} \mathbb{Z} \times \operatorname{spec} \mathbb{Z}$ is one-dimensional and not two-dimensional as the surface $\mathcal{C} \times \mathcal{C}$ (see [Har1]).

A completely new technique was discovered by Stepanov [Ste], initially only for hyperelliptic curves. W. M. Schmidt [wSch] used his method to reprove the Riemann hypothesis for $\zeta_{\mathcal{C}}$, and a simplified proof was given by Bombieri [Bom1, Bom2]. Bombieri’s proof uses the graph of Frobenius

⁵ In his thesis [Art1], Artin considers only quadratic extensions of $\mathbb{F}_p(T)$, that is, hyperelliptic curves over \mathbb{F}_p . Moreover, in his zeta functions, the Euler factors corresponding to the points at infinity are missing. Later, F. K. Schmidt introduced the zeta function of a general projective curve over an arbitrary finite field [fSch].

⁶ Weil announced his ideas in 1940 [W1, W2] and explained them in 1942 in a letter to Artin [W3]. But the complete proof (see [W5, W7]) had to await the completion of his *Foundations* [W4]. See also [Ray].

in $\mathcal{C} \times \mathcal{C}$.⁷ It relies on the Riemann–Roch formula and also on the action of Frobenius. So his proof uses less geometry that cannot be translated to $\text{spec } \mathbb{Z}$. We present this proof in Chapter 5.

Around the same time, P. Deligne proved the Weil conjectures [Del, FreiK, K2]. His proof, applied in the one-dimensional case, gives yet another proof of the Riemann hypothesis for curves over finite fields. This proof relies on detailed information about the action of Frobenius on the étale cohomology groups of the variety (see [K1, K3] for more information). As with the other proofs, it is unclear how to translate this proof to the number field case.

Recently, Alain Connes found a completely new method again, based on the work of Shai Haran [Har1] (and then extended by Haran in the papers [Har2, Har3]), using harmonic analysis on the ring of adèles. Connes establishes the positivity of the trace of a certain shift operator, thus proving the Riemann hypothesis for $\text{spec } \mathbb{Z}$ (and for all L-functions associated with Grössencharacters), up to a “lemma” about a noncommutative space. In Chapter 6, we adapt Haran’s approach to obtain a proof of the Riemann hypothesis for $\zeta_{\mathcal{C}}$. This proof does not use $\mathcal{C} \times \mathcal{C}$ or Frobenius. Indeed, the only difficulty in translating it to a proof for $\text{spec } \mathbb{Z}$ is to provide a suitable local analysis at the real component of the adèles. Connes does this by using special functions on $[-c, c]$, the Fourier transforms of which are also almost supported on $[-c, c]$. In our case, working with the curve \mathcal{C} , we have no archimedean valuations to deal with, and the local analysis at the primes “at infinity” is not different from the local analysis at the other points of \mathcal{C} .

It is interesting to see the development in these proofs. Gradually, more geometry that cannot be translated to the number field case has been taken out. The question arises what exactly is needed to prove the Riemann hypothesis for curves, and what can we learn from this for the Riemann hypothesis for $\text{spec } \mathbb{Z}$.

Weil’s first proof uses the geometry of $\mathcal{C} \times \mathcal{C}$, and, in particular, the intersection of the graph of Frobenius with the diagonal. There is some reason to believe that no analogue will ever exist for number fields, or, at least, that constructing an analogue is harder than establishing the Riemann hypothesis. His second proof uses the Jacobian of \mathcal{C} , and, again, no direct analogue may ever be constructed for the integers.

Deligne’s proof works for higher-dimensional varieties and uses very detailed information about the geometry, along with sophisticated tools

⁷ Bombieri does not mention $\mathcal{C} \times \mathcal{C}$, but he uses $\mathbb{F}_q^a(\mathcal{C}) \otimes \mathbb{F}_q^a(\mathcal{C})$, which is the ring of functions on $\mathcal{C} \times \mathcal{C}$.

to make deductions from this information. It is not impossible that an adequate analogue of these cohomological theories exists for the integers (see [Den1, Den2]).

Bombieri's proof uses very little of the geometry of $\mathcal{C} \times \mathcal{C}$, but it uses the action of Frobenius and Riemann–Roch. Since Tate's thesis, it is known that the Riemann–Roch equality translates into formula (2). Bombieri's proof naturally divides into two steps. In the first step, he uses the action of Frobenius to obtain a discrete flow on the curve, which is analyzed to obtain an upper bound for the number of points on the curve, which is a weak form of the prime number theorem for the curve (see Table 5.1 in Section 5.5). One sees that the horizontal coordinate of $\mathcal{C} \times \mathcal{C}$ plays an “arithmetic” role, and the vertical coordinate plays a “geometric” role (see Remark 5.4.7). In the second step, the Riemann hypothesis for \mathbb{P}^1 is used, along with the fact that the Frobenius automorphism generates the local Galois group (decomposition group) at a point on the curve, to obtain a lower bound for the number of points on the curve. Combining the two steps, one first obtains the analogue of the prime number theorem with a good error term, and from this it is a small step to deduce the Riemann hypothesis. Therefore, one might conclude that the right approach to the Riemann hypothesis is to first prove the prime number theorem with a good bound for the error.

Looking at the first step of Bombieri's proof, one might even guess that the key to a prime number theorem with a good error bound is to construct a function (possibly a Fourier or Dirichlet polynomial, in the spirit of the methods of Baker, Gelfond, and Schneider) that vanishes at the first N primes to a high order. If one could bound the degree of this polynomial, then one would obtain an upper bound for the number of primes. This would already imply the Riemann hypothesis, so the second step becomes unnecessary. According to Deninger [Den1, Den2], the analogue of the Frobenius flow of the first step might be provided by the shift on the real line.⁸

Connes, in turn, does not use the geometry of $\mathcal{C} \times \mathcal{C}$ or Frobenius. Instead, he uses Fourier analysis on the ring of adèles and the diagonal embedding of the global field (the field of rational numbers). Even though he does not use the Riemann–Roch theorem, this theorem is a

⁸ According to Bombieri, the correct philosophy is as follows: since Frobenius in characteristic p is based on the fact that the binomial coefficients $\binom{p}{k}$ are divisible by p if $k \neq 0, p$, an understanding of the archimedean Frobenius should come from an understanding of the size of $\binom{n}{k}$. This leads to the Gaussian $e^{-\pi x^2}$ and probabilities [Bom2] (personal communication, October 2008).

natural consequence of the formalism that he sets up, much as in Tate’s thesis. The only problem is that on the real line, the Fourier transform of a compactly supported function is not compactly supported.⁹ To complete the proof of the Riemann hypothesis for $\text{spec } \mathbb{Z}$, one might try to construct a suitable substitute for this requirement.

The fact that Connes does not use the action of Frobenius should make one suspicious about his approach. However, Connes uses the action of the idele class group on the space of adèle classes. As Connes points out [Conn1, Remark c, p. 72], this action is the counterpart of the action of Frobenius on the curve. Indeed, by class field theory, there exists an isomorphism between the exact sequences

$$\begin{array}{ccccc}
 \ker & \longrightarrow & \mathbb{A}^*/K^* & \xrightarrow{|\cdot|} & q^{\mathbb{Z}} \\
 \parallel & & \parallel & & \parallel \\
 \pi_{\text{ab}}(\mathcal{C}) & \longrightarrow & G_{\text{ab}} & \longrightarrow & \langle \phi \rangle
 \end{array} \tag{5}$$

The vertical isomorphisms come from class field theory. The upper sequence gives the norm from the idele class group \mathbb{A}^*/K^* to the group of powers of q . The kernel of this map corresponds to the group $\pi_{\text{ab}}(\mathcal{C})$ of abelian covers of the curve \mathcal{C} inside the Galois group of abelian extensions

$$G_{\text{ab}} = \text{Gal}(\mathbb{F}_q(\mathcal{C})^{\text{ab}}, \mathbb{F}_q(\mathcal{C})).$$

On the level of Galois theory (the lower sequence in (5)), the second arrow maps G_{ab} onto the group generated by the Frobenius automorphism ϕ . This automorphism acts on constant field extensions of the function field of \mathcal{C} , which corresponds to the Frobenius flow on \mathcal{C} .

The counterpart for \mathbb{Z} is the diagram of exact sequences

$$\begin{array}{ccccc}
 \mathbb{A}^*/\mathbb{Q}^*\mathbb{R}^+ & \longrightarrow & \mathbb{A}^*/\mathbb{Q}^* & \xrightarrow{|\cdot|} & \mathbb{R}^+ \\
 \parallel & & \parallel & & \parallel \\
 \text{Gal}(\mathbb{Q}^{\text{ab}}, \mathbb{Q}) & \longrightarrow & ? & \longrightarrow & ?
 \end{array} \tag{6}$$

where the left vertical map is the isomorphism from class field theory between the group $\mathbb{A}^*/\mathbb{Q}^*\mathbb{R}^+$ and the Galois group of the maximal abelian extension of \mathbb{Q} . For the other vertical equalities, no counterpart is known

⁹ In the field of p -adic numbers, and in general in any nonarchimedean field, the Fourier transform of a function that is locally constant is compactly supported. This is not true for the archimedean fields \mathbb{R} and \mathbb{C} . See Section 3.1.

within Galois theory. Indeed, such a counterpart must lie outside of Galois theory, since any continuous map $\mathbb{R}^+ \rightarrow \text{Gal}(\mathbb{Q}^a/\mathbb{Q})$ is constant, the latter group being profinite. Thus, only the “geometric part” of diagram (5), the abelian fundamental group $\pi_{\text{ab}}(\mathcal{C})$, has its counterpart in number theory, while the “arithmetic part,” corresponding to constant field extensions, remains a mystery in number theory.¹⁰ However, in [Conn2], Connes defines the noncommutative space \mathbb{A}/\mathbb{Q}^* of the adèles modulo the multiplicative action of the diagonally embedded rationals. This space has a natural multiplicative action of the idele class group by shift operators, thus completing the question marks in (6). In particular, the question mark under \mathbb{R}^+ corresponds to the multiplicative action of \mathbb{R}^+ on \mathbb{A}/\mathbb{Q}^* by shifts, which, by diagram (5), corresponds to the action of Frobenius. It seems that everything is in place to prove the Riemann hypothesis for $\text{spec } \mathbb{Z}$.

This seems to be the first time since the formulation of the Riemann hypothesis in 1859 that we have a serious heuristic argument for its truth. It is surprising that such an easily stated problem, either as “all nonreal zeros of $\zeta(s)$ have real part $1/2$ ” or as “the prime number theorem has an error term of order $x^{1/2+\varepsilon}$,” should be so hard to solve. Indeed, this was not immediately appreciated at the time, since Barnes assigned the Riemann hypothesis to Littlewood as a thesis problem (see [Conr]). But from the solution of the Riemann hypothesis for curves over finite fields, it seems that there may never be a proof using only methods from analytic number theory.¹¹

¹⁰ Every extension of \mathbb{Q} (abelian or not) is ramified at some primes, hence should be considered as a geometric cover of curves. On the other hand, every abelian extension of \mathbb{Q} is cyclotomic, hence could be considered as a constant field extension. It seems that the first two Galois groups in (5) have collapsed into the one group $\text{Gal}(\mathbb{Q}^{\text{ab}}, \mathbb{Q})$, and the group $\langle \phi \rangle$ is missing.

¹¹ For a possible approach using the theory of fractal strings, see [vF1, Remark 4.5].

1

Valuations

Valuations correspond to orbits of points on a curve: every orbit of Frobenius gives a valuation, and every valuation gives an orbit. This will be elaborated in Chapter 5. In this chapter, we study how valuations distinguish constants from nonconstant functions and how valuations can ramify.

We first develop this theory for an arbitrary extension L/K . In Section 1.4, we apply our theory to the situation that is the subject of this book, where L/K is a finite extension K/\mathbf{q} of the field of rational functions $\mathbf{q} = \mathbb{F}_q(T)$ over the finite field of constants \mathbb{F}_q .

The embedding $\mathbf{q} \hookrightarrow K$ corresponds to the projection $\mathcal{C} \rightarrow \mathbb{P}^1$, where a point on the curve \mathcal{C} projects to the value of the function T at this point. This will be explained in Chapter 5.

1.1 Trace and norm

Let L/K be a finite extension of fields. We can find a polynomial

$$m(X) = X^n + m_1X^{n-1} + \cdots + m_n,$$

irreducible over K , such that $L = K[X]/(m)$.

For $x \in L$, we consider the map $M_x: L \rightarrow L$ of multiplication by x ,

$$M_x(y) = xy \quad (x, y \in L).$$

This map is linear in y and consequently has a determinant and a trace, defined independently of a choice of a basis for L as a K -vector space.

Definition 1.1.1 Let L/K be a finite separable extension, and $x \in L$. The *trace* of x over K is the trace of M_x ,

$$\text{Tr}_{L/K}(x) = \text{Tr}(M_x).$$

The *norm* of x over K is the determinant of M_x , $N_{L/K}(x) = \det(M_x)$.

In particular, let $x = X + (m)$ be the root of m in L . The matrix of M_x on the basis $(1, x, \dots, x^{n-1})$ of L over K is given by

$$M_x = \begin{pmatrix} 0 & O & -m_n \\ 1 & \ddots & \vdots \\ & \ddots & 0 & -m_2 \\ O & & 1 & -m_1 \end{pmatrix}.$$

The characteristic polynomial of this matrix is $m(X)$. Assume that L/K is separable, i.e., all roots of m are different. Then, over a splitting field of m , the matrix M_x diagonalizes. Denoting by $\sigma_1 x, \dots, \sigma_n x$ the images of x in a splitting field F under the n embeddings $\sigma_1, \dots, \sigma_n$ of L into F , we find that this matrix diagonalizes as

$$M_x = \begin{pmatrix} \sigma_1 x & & O \\ & \ddots & \\ O & & \sigma_n x \end{pmatrix} \quad (x = X + (m)), \quad (1.1)$$

on a suitable basis for F^n .

Exercise 1.1.2 Show that $\det(\lambda I - M_x) = m(\lambda)$.

Exercise 1.1.3 L is n -dimensional over K , and F^n is n -dimensional over F . Show that $K[X]/(m) \otimes F \cong F[X]/(m) \cong F \oplus \dots \oplus F$.

Clearly, for any two elements $x, y \in L$, we have $M_{x+y} = M_x + M_y$ and $M_{xy} = M_x M_y$. Hence $M_{f(x)} = f(M_x)$ for every polynomial f over K . An element $y \in L$ can be written as $y = f(X) + (m)$ for a polynomial f over K . Therefore, $M_y = f(M_x)$ for M_x as in (1.1), and the matrix of M_y diagonalizes as

$$M_y = \begin{pmatrix} f(\sigma_1 x) & & O \\ & \ddots & \\ O & & f(\sigma_n x) \end{pmatrix} = \begin{pmatrix} \sigma_1 y & & O \\ & \ddots & \\ O & & \sigma_n y \end{pmatrix},$$

since $f(\sigma_i x) = \sigma_i f(x) = \sigma_i y$. We deduce the following proposition and corollary:

Proposition 1.1.4 *Let L/K be a finite separable extension of fields, and let K^a be an algebraic closure of K . Then*

$$\text{Tr}_{L/K}(x) = \sum_{\sigma:L \rightarrow K^a} \sigma x,$$

where the summation extends over all embeddings of L into K^a .

For a finite separable extension M of L , each embedding $\sigma : L \rightarrow K^a$ extends to an embedding of M into K^a in $[M : L]$ different ways. Thus we obtain the following corollary:

Corollary 1.1.5 *$\text{Tr}_{M/K} = \text{Tr}_{L/K} \circ \text{Tr}_{M/L}$ in a tower $M/L/K$ of finite separable extensions.*

This property of the trace will be needed in Chapter 3 to define the additive character of a field.

Exercise 1.1.6 Formulate and prove the analogue of Proposition 1.1.4 and Corollary 1.1.5 for the norm.

1.1.1 The canonical pairing

We have a K -valued pairing between elements of L , given by

$$(x, y) \mapsto \text{Tr}_{L/K}(xy).$$

Writing Tr for $\text{Tr}_{L/K}$, the matrix of this pairing on a basis $(\epsilon_1, \dots, \epsilon_n)$ for L as a K -vector space is

$$D_\epsilon = \begin{pmatrix} \text{Tr}(\epsilon_1^2) & \text{Tr}(\epsilon_1\epsilon_2) & \dots & \text{Tr}(\epsilon_1\epsilon_n) \\ \text{Tr}(\epsilon_2\epsilon_1) & \text{Tr}(\epsilon_2^2) & \dots & \text{Tr}(\epsilon_2\epsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Tr}(\epsilon_n\epsilon_1) & \text{Tr}(\epsilon_n\epsilon_2) & \dots & \text{Tr}(\epsilon_n^2) \end{pmatrix}.$$

The pairing of two elements x and y can be computed using D_ϵ as follows. Write $x = x_1\epsilon_1 + \dots + x_n\epsilon_n$ and $y = y_1\epsilon_1 + \dots + y_n\epsilon_n$ with coefficients x_i and y_i in K . Then

$$\text{Tr}_{L/K}(xy) = (x_1 \dots x_n) D_\epsilon \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$