17 The functional equation

In the previous lecture we proved that the Riemann zeta function $\zeta(s)$ has an Euler product and an analytic continuation to the right half-plane Re(s) > 0. In this lecture we complete the picture by deriving a functional equation that relates the values of $\zeta(s)$ to those of $\zeta(1-s)$. This will then also allow us to extend $\zeta(s)$ to a meromorphic function on \mathbb{C} that is holomorphic except for a simple pole at s=1.

17.1 Fourier transforms and Poisson summation

A key tool we will use to derive the functional equation is the *Poisson summation formula*, a result from harmonic analysis that we now recall.

Definition 17.1. A Schwartz function on \mathbb{R} is a complex-valued C^{∞} function $f: \mathbb{R} \to \mathbb{C}$ that decays rapidly to zero: for all $m, n \in \mathbb{Z}_{\geq 0}$ we have

$$\sup_{x \in \mathbb{R}} \left| x^m f^{(n)}(x) \right| < \infty,$$

where $f^{(n)}$ denotes the *n*th derivative of f. The Schwartz space $\mathcal{S}(\mathbb{R})$ of all Schwartz functions on \mathbb{R} is a \mathbb{C} -vector space dense in $L^p(\mathbb{R})$ for any $p \in \mathbb{R}_{\geq 1}$ (under the L^p -norm). It is closed under differentiation, sums and products, linear change of variable, and convolution: for any $f, g \in \mathcal{S}(\mathbb{R})$ the function

$$(f * g)(x) := \int_{\mathbb{D}} f(y)g(x - y)dy$$

is also in $\mathcal{S}(\mathbb{R})$. Convolution is commutative, associative, and bilinear.

Example 17.2. All compactly supported functions C^{∞} functions are Schwartz functions, as is the Gaussian $g(x) := e^{-\pi x^2}$. Non-examples include any function that does not tend to zero as $x \to \pm \infty$ (so all nonzero polynomials), as well as functions like $(1 + x^{2n})^{-1}$.

Definition 17.3. The Fourier transform of a Schwartz function $f \in \mathcal{S}(\mathbb{R})$ is the function

$$\hat{f}(y) := \int_{\mathbb{R}} f(x)e^{-2\pi ixy}dx,$$

which is also a Schwartz function. The Fourier transform is an invertible linear operator on the vector space $\mathcal{S}(\mathbb{R})$; we can recover f(x) from $\hat{f}(y)$ via the inverse transform

$$f(x) := \int_{\mathbb{R}} \hat{f}(y)e^{+2\pi ixy}dy.$$

The fact that the map $f \mapsto \hat{f}$ is invertible on $\mathcal{S}(\mathbb{R})$ is a standard result that we won't prove. It is in some sense baked into the definition of $\mathcal{S}(\mathbb{R})$: inside $L^1(\mathbb{R})$ (the largest space of functions for which our definition of the Fourier transform makes sense), the Fourier transform of a smooth function decays rapidly to zero, and the Fourier transform of a function that decays rapidly to zero is smooth; this leads one to define $\mathcal{S}(\mathbb{R})$ as we have.

¹The Schwartz space has its on metric topology, not induced by any L^p -norm, in which it is complete (and even a Fréchet space), but this topology will not concern us here. These comments also apply to $\mathcal{S}(\mathbb{R}^n)$.

The Fourier transform changes convolutions into products, and vice versa. We have

$$\widehat{f * g} = \widehat{f}\widehat{g}$$
 and $\widehat{fg} = \widehat{f} * \widehat{g}$,

for all $f, g \in \mathcal{S}(\mathbb{R})$ (see Problem Set 8). One can thus view the Fourier transform as an isomorphism of (non-unital) \mathbb{C} -algebras that sends $(\mathcal{S}(\mathbb{R}), +, \times)$ to $(\mathcal{S}(\mathbb{R}), +, *)$.

Lemma 17.4. For $f \in \mathcal{S}(\mathbb{R})$, we have $\widehat{f(ax)}(y) = \frac{1}{a}\widehat{f}(\frac{y}{a})$.

Proof. Applying the substitution t = ax yields

$$\widehat{f(ax)}(y) = \int_{\mathbb{R}} f(ax)e^{-2\pi ixy}dx = \frac{1}{a} \int_{\mathbb{R}} f(t)e^{-2\pi ity/a}dt = \frac{1}{a}\widehat{f}\left(\frac{y}{a}\right).$$

Lemma 17.5. For $f \in \mathcal{S}(\mathbb{R})$ we have $\frac{d}{dy}\hat{f}(y) = -2\pi i \widehat{xf(x)}(y)$ and $\widehat{\frac{d}{dx}f(x)}(y) = 2\pi i y \hat{f}(y)$.

Proof. Noting that $xf \in \mathcal{S}(\mathbb{R})$, the first identity follows from

$$\frac{d}{dy}\widehat{f}(y) = \frac{d}{dy}\left(\int_{\mathbb{R}} f(x)e^{-2\pi ixy}dx\right) = \int_{\mathbb{R}} f(x)(-2\pi ix)e^{-2\pi ixy}dx = -2\pi i \widehat{xf(x)}(y),$$

since we may differentiate under the integral via dominated convergence. For the second, we note that $\lim_{x\to\pm\infty} f(x) = 0$, so integration by parts yields

$$\widehat{\frac{d}{dx}f(x)}(y) = \int_{\mathbb{R}} f'(x)e^{-2\pi ixy}dx = 0 - \int_{\mathbb{R}} f(x)(-2\pi iy)e^{-2\pi ixy}dx = 2\pi iy\widehat{f}(y). \qquad \Box$$

The Fourier transform is compatible with the inner product $\langle f,g\rangle:=\int_{\mathbb{R}}f(x)\overline{g(x)}dx$. Indeed, we can easily derive Parseval's identity:

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x) \overline{g(x)} dx = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(y) \overline{g(x)} e^{+2\pi i x y} dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(y) \overline{\hat{g}(y)} dy = \langle \hat{f}, \hat{g} \rangle,$$

which when applied to g = f yields Plancherel's identity:

$$||f||_2 = \langle f, f \rangle = \langle \hat{f}, \hat{f} \rangle = ||\hat{f}||_2,$$

where $||f||_2 = \int_{\mathbb{R}} |f(x)|^2 dx = \int_{\mathbb{R}} f(x) \bar{f}(x) dx$ is the L^2 -norm. For number-theoretic applications there is an analogous result due to Poisson.

Theorem 17.6 (Poisson Summation Formula). For all $f \in \mathcal{S}(\mathbb{R})$ we have the identity

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(n).$$

Proof. We first note that both sums are well defined; the rapid decay property of Schwartz functions guarantees absolute convergence. Let $F(x) := \sum_{n \in \mathbb{Z}} f(x+n)$. Then F is a periodic C^{∞} -function, so it has a Fourier series expansion

$$F(x) = \sum_{n \in \mathbb{Z}} c_n e^{2\pi i n x},$$

with Fourier coefficients

$$c_n = \int_0^1 F(x)e^{-2\pi i n t} dt = \int_0^1 \sum_{m \in \mathbb{Z}} f(x+m)e^{-2\pi i n y} dy = \int_{\mathbb{R}} f(x)e^{-2\pi i n y} dy = \hat{f}(n).$$

We then note that

$$\sum_{n\in\mathbb{Z}} f(n) = F(0) = \sum_{n\in\mathbb{Z}} \hat{f}(n).$$

Finally, we note that the Gaussian function $e^{-\pi x^2}$ is its own Fourier transform.

Lemma 17.7. Let $g(x) := e^{-\pi x^2}$. Then $\hat{g}(y) = g(y)$.

Proof. The function g(x) satisfies the first order ordinary differential equation

$$g' + 2\pi x g = 0, (1)$$

with initial value g(0) = 1. Multiplying both sides by i and taking Fourier transforms yields

$$i(\hat{g}' + 2\pi \hat{x}\hat{g}) = i(2\pi i x \hat{g} - i\hat{g}') = \hat{g}' + 2\pi x \hat{g} = 0,$$

via Lemma 17.5. So \hat{g} also satisfies (1), and $\hat{g}(0) = \int_{\mathbb{R}} e^{-\pi x^2} dx = 1$, so $\hat{g} = g$.

17.1.1 Jacobi's theta function

We now define the theta function²

$$\Theta(\tau) := \sum_{n \in \mathbb{Z}} e^{\pi i n^2 \tau}.$$

The sum is absolutely convergent for $\operatorname{Im} \tau > 0$ and thus defines a holomorphic function on the upper half plane. It is easy to see that $\Theta(\tau)$ is periodic modulo 2, that is,

$$\Theta(\tau + 2) = \Theta(\tau),$$

but it it also satisfies another functional equation.

Lemma 17.8. For all $a \in \mathbb{R}_{>0}$ we have $\Theta(ia) = \Theta(i/a)/\sqrt{a}$.

Proof. Put $g(x) := e^{-\pi x^2}$ and $h(x) := g(\sqrt{a}x) = e^{-\pi x^2 a}$. Lemmas 17.4 and 17.7 imply

$$\hat{h}(y) = \widehat{g(\sqrt{a}x)}(y) = \hat{g}(y/\sqrt{a})/\sqrt{a} = g(y/\sqrt{a})/\sqrt{a}.$$

Plugging $\tau = ia$ into $\Theta(\tau)$ and applying Poisson summation (Theorem 17.6) yields

$$\Theta(ia) = \sum_{n \in \mathbb{Z}} e^{-\pi n^2 a} = \sum_{n \in \mathbb{Z}} h(n) = \sum_{n \in \mathbb{Z}} \hat{h}(n) = \sum_{n \in \mathbb{Z}} g\left(n/\sqrt{a}\right)/\sqrt{a} = \Theta(i/a)/\sqrt{a} \qquad \Box$$

17.1.2 Euler's gamma function

You are probably familiar with the gamma function $\Gamma(s)$, which plays a key role in the functional equation of not only the Riemann zeta function but many of the more general zeta functions and L-series we wish to consider. Here we recall some of its analytic properties. We begin with the definition of $\Gamma(s)$ as a Mellin transform.

Definition 17.9. The *Mellin transform* of a function $f: \mathbb{R}_{>0} \to \mathbb{C}$ is the complex function defined by

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

whenever this integral converges. It is holomorphic on Re $s \in (a,b)$ for any interval (a,b) in which the integral $\int_0^\infty |f(t)| t^{\sigma-1} dt$ converges for all $\sigma \in (a,b)$.

²The function $\Theta(\tau)$ we define here is a special case of one of four parameterized families of theta functions $\Theta_i(z:\tau)$ originally defined by Jacobi for i=0,1,2,3, which play an important role in the theory of elliptic functions and modular forms; in terms of Jacobi's notation, $\Theta(\tau) = \Theta_3(0;\tau)$.

Definition 17.10. The Gamma function

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

is the Mellin transform of e^{-t} . Since $\int_0^\infty |e^{-t}| t^{\sigma-1} dt$ converges for all $\sigma > 0$, the integral defines a holomorphic function on Re(s) > 0.

Integration by parts yields

$$\Gamma(s) = \frac{t^s e^{-t}}{s} \bigg|_0^\infty + \frac{1}{s} \int_0^\infty e^{-t} t^s dt = \frac{\Gamma(s+1)}{s},$$

thus $\Gamma(s)$ has a simple pole at s=0 with residue 1 (since $\Gamma(1)=\int_0^\infty e^{-t}dt=1$), and

$$\Gamma(s+1) = s\Gamma(s) \tag{2}$$

for Re(s) > 0. Equation (2) allows us to extend $\Gamma(s)$ to a meromorphic function on \mathbb{C} with simple poles at $s = 0, -1, -2, \ldots$, and no other poles.

An immediate consequence of (2) is that for integers n > 0 we have

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = n(n-1)(n-2)\cdots 2\cdot 1\cdot \Gamma(1) = n!,$$

thus the gamma function can be viewed as an extension of the factorial function. The gamma function satisfies many useful identities in addition to (2), including the following.

Theorem 17.11 (EULER'S REFLECTION FORMULA). We have

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

as meromorphic functions on \mathbb{C} with simple poles at each integer $s \in \mathbb{Z}$.

Proof. Let $f(s) := \Gamma(s)\Gamma(1-s)\sin(\pi s)$. The function $\Gamma(s)\Gamma(1-s)$ has a simple pole at each $s \in \mathbb{Z}$ and no other poles, while $\sin(\pi s)$ has a zero at each $s \in \mathbb{Z}$ and no poles, so f(s) is holomorphic on \mathbb{C} . We now note that

$$f(s+1) = \Gamma(s+1)\Gamma(-s)\sin(\pi s + \pi) = -s\Gamma(s)\Gamma(-s)\sin(\pi s) = \Gamma(s)\Gamma(1-s)\sin(\pi s) = f(s),$$

so f is periodic (with period 1). Using the substitution $u = e^t$ we obtain

$$|\Gamma(s)| \le \int_0^\infty |e^{-t}t^{s-1}|dt = \int_{-\infty}^\infty |e^{-e^u}e^{u(s-1)}|e^udu = \int_{-\infty}^\infty e^{u\operatorname{Re}(s)-e^u}du.$$

This implies $|\Gamma(s)|$ is bounded on $\operatorname{Re}(s) \in [1,2]$, hence on $\operatorname{Re}(s) \in [0,1] \cap \operatorname{Im}(s) \geq 1$, via (2). It follows that in the strip $\operatorname{Re}(s) \in [0,1]$ we have

$$|f(s)| = |\Gamma(s)||\Gamma(1-s)||\sin(\pi s)| = O(e^{\text{Im}(s)})$$

as $\text{Im}(s) \to \infty$, since $|\sin(\pi s)| = \frac{1}{2}|e^{is} - e^{i\bar{s}}| = O(e^{\text{Im}(s)})$. By Lemma 17.12 below, f(s) is constant. To determine the constant, as $s \to 0$ we have $\Gamma(s) \sim \frac{1}{s}$ and $\sin(\pi s) \sim \pi s$, thus

$$f(0) = \lim_{s \to 0} f(s) = \lim_{s \to 0} \Gamma(s) \Gamma(1 - s) \sin(\pi s) = \lim_{s \to 0} \frac{1}{s} \cdot 1 \cdot \pi s = \pi,$$

and the theorem follows.

Lemma 17.12. Let f(s) be a holomorphic function on \mathbb{C} such that f(s+1) = f(s) and $|f(s)| = O(e^{\operatorname{Im}(s)})$ as $\operatorname{Im}(s) \to \infty$ in the vertical strip $\operatorname{Re}(s) \in [0,1]$. Then f is constant.

Proof. The function

$$g(s) = \frac{f(s) - f(a)}{\sin(\pi(s - a))}$$

is holomorphic on \mathbb{C} , since f(s)-f(a) is holomorphic and vanishes at the zeros $a+\mathbb{Z}$ of $\sin(\pi(s-a))$ (all of which are simple). We also have g(s+1)=g(s), and |g(s)| is bounded on $\text{Re}(s) \in [0,1]$, since as $\text{Im}(s) \to \infty$ we have $|f(s)-f(a)|=O(e^{\text{Im}(s)})$ and $|\sin(\pi(s-a))| \sim e^{\pi \text{Im}(s)}$. It follows that g(s) is bounded on \mathbb{C} , hence constant, by Liouville's theorem. We must have g=0, since $|g(s)|=O(e^{(1-\pi)\text{Im}(s)})=o(1)$ as $\text{Im}(s)\to\infty$, and this implies f(s)=f(a) for all $s\in\mathbb{C}$.

Example 17.13. Putting $s = \frac{1}{2}$ in the reflection formula yields $\Gamma(\frac{1}{2})^2 = \pi$, so $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Corollary 17.14. *The function* $\Gamma(s)$ *has no zeros on* \mathbb{C} .

Proof. Suppose $\Gamma(s_0) = 0$. The RHS of the reflection formula $\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$ is never zero, since $\sin(\pi s)$ has no poles, so $\Gamma(1-s)$ must have a pole at s_0 . Therefore $1-s_0 \in \mathbb{Z}_{\leq 0}$, equivalently, $s_0 \in \mathbb{Z}_{\geq 1}$, but then $\Gamma(s_0) = (s_0-1)! \neq 0$, a contradiction. \square

17.1.3 Completing the zeta function

Let us now consider the function

$$F(s) := \pi^{-s} \Gamma(s) \zeta(2s),$$

which is a meromorphic on \mathbb{C} and holomorphic on $\mathrm{Re}(s) > 1/2$. In the region $\mathrm{Re}(s) > 1/2$ we have an absolutely convergent sum

$$F(s) = \pi^{-s}\Gamma(s)\sum_{n>1}n^{-2s} = \sum_{n>1}(\pi n^2)^{-s}\Gamma(s) = \sum_{n>1}\int_0^\infty (\pi n^2)^{-s}t^{s-1}e^{-t}dt,$$

and the substitution $t = \pi n^2 y$ with $dt = \pi n^2 dy$ yields

$$F(s) = \sum_{n \ge 1} \int_0^\infty (\pi n^2)^{-s} (\pi n^2 y)^{s-1} e^{-\pi n^2 y} \pi n^2 dy = \sum_{n \ge 1} \int_0^\infty y^{s-1} e^{-\pi n^2 y} dy.$$

By the Fubini-Tonelli theorem, we can swap the sum and the integral to obtain

$$F(s) = \int_0^\infty y^{s-1} \sum_{n>1} e^{-\pi n^2 y} dy.$$

We have $\Theta(iy) = \sum_{n \in \mathbb{Z}} e^{-\pi n^2 y} = 1 + 2 \sum_{n \geq 1} e^{-\pi n^2 y}$, thus

$$F(s) = \frac{1}{2} \int_0^\infty y^{s-1} (\Theta(iy) - 1) dy$$
$$= \frac{1}{2} \left(\int_0^1 y^{s-1} \Theta(iy) dy - \frac{1}{s} + \int_1^\infty y^{s-1} (\Theta(iy) - 1) dy \right)$$

We now focus on the first integral on the RHS. The change of variable $t = \frac{1}{u}$ yields

$$\int_0^1 y^{s-1} \Theta(iy) dy = \int_\infty^1 t^{1-s} \Theta(i/t) (-t^{-2}) dt = \int_1^\infty t^{-s-1} \Theta(i/t) dt.$$

By Lemma 17.8, $\Theta(i/t) = \sqrt{t}\Theta(it)$, and adding $-\int_1^\infty t^{-s-1/2}dt + \int_1^\infty t^{-s-1/2}dt = 0$ yields

$$\begin{split} &= \int_{1}^{\infty} t^{-s-1/2} \big(\Theta(it) - 1 \big) dt + \int_{1}^{\infty} t^{-s-1/2} dt \\ &= \int_{1}^{\infty} t^{-s-1/2} \big(\Theta(it) - 1 \big) dt - \frac{1}{1/2 - s}. \end{split}$$

Plugging this back into our equation for F(s) we obtain the identity

$$F(s) = \frac{1}{2} \int_{1}^{\infty} (y^{s-1} + y^{-s-1/2}) (\Theta(iy) - 1) dy - \frac{1}{2s} - \frac{1}{1 - 2s},$$

valid on Re(s) > 1/2. We now observe that $F(s) = F(\frac{1}{2} - s)$ for $s \neq 0, \frac{1}{2}$, which allows us to analytically extend F(s) to a meromorphic function on $\mathbb C$ with poles only at $s = 0, \frac{1}{2}$. Replacing s with s/2 leads us to define the *completed zeta function*

$$Z(s) := \pi^{-s/2} \Gamma(s/2) \zeta(s), \tag{3}$$

which is meromorphic on \mathbb{C} and satisfies the functional equation

$$Z(1-s) = Z(s). (4)$$

It has simple poles at 0 and 1 (and no other poles). The only zeros of Z(s) on Re(s) > 0 are the zeros of $\zeta(s)$, since by Corollary 17.14, the gamma function $\Gamma(s)$ has no zeros (and neither does $\pi^{-s/2}$). Thus the zeros of Z(s) on $\mathbb C$ all lie in the critical strip 0 < Re(s) < 1.

The functional equation also allows us to extend $\zeta(s)$ to a meromorphic function on \mathbb{C} . It has no poles other than the simple pole at s=1, since $\pi^{-s/2}\Gamma(s)$ has no zeros and the simple pole of Z(s) at 0 corresponds to the simple pole of $\Gamma(s/2)$ at zero. Notice that $\Gamma(s/2)$ has poles at $0, -2, -4, \ldots$, so our extended $\zeta(s)$ must have zeros at $-2, -4, \ldots$ (but not at 0). These are the *trivial zeros* of $\zeta(s)$; all the interesting zeros lie in the critical strip, and under the Riemann hypothesis, on the critical line $\operatorname{Re}(s) = 1/2$, the axis of symmetry in the functional equation.

We can compute $\zeta(0)$ using the functional equation. From (3) and (4) we have

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

We know that $\zeta(s)$ has a simple pole with residue 1 at s=1, thus

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

In the limit the denominator on the RHS is 1, since $\Gamma(1/2) = \pi^{1/2}$, and in the numerator we have $\pi^{-(1-s)/2} = 1$. Using $\Gamma(z) = \frac{1}{z}\Gamma(z+1)$ to shift the gamma factor in the numerator,

$$1 = \lim_{s \to 1^+} (s-1) \frac{2}{1-s} \Gamma\left(\frac{3-s}{2}\right) \zeta(1-s) = -2\Gamma(1)\zeta(0) = -2\zeta(0),$$

thus $\zeta(0) = -1/2$.

If we write out the Euler product for the completed zeta function, we have

$$Z(s) = \pi^{-s/2} \Gamma(s/2) \prod_{p} (1 - p^{-s})^{-1}.$$

One should think of this as a product over the places of the field \mathbb{Q} ; the leading factor

$$\Gamma_{\mathbb{R}}(s) := \pi^{-2/s} \Gamma(s/2)$$

that distinguishes the completed zeta function Z(s) from $\zeta(s)$ corresponds to the real archimedean place of \mathbb{Q} . When we discuss Dedekind zeta functions in a later lecture we will see that there are gamma factors $\Gamma_{\mathbb{R}}$ and $\Gamma_{\mathbb{C}}$ associated to each of the real and complex places of a number field.

If we incorporate an additional factor of $\binom{s}{2} := \frac{s(s-1)}{2}$ in Z(s) we can remove the poles at 0 and 1, yielding a function $\xi(s)$ holomorphic on \mathbb{C} . This yields Riemann's seminal result.

Theorem 17.15 (ANALYTIC CONTINUATION II). The function

$$\xi(s) := \binom{s}{2} \Gamma_{\mathbb{R}}(s) \zeta(s)$$

is holomorphic on \mathbb{C} and satisfies the functional equation

$$\xi(1-s) = \xi(s).$$

The zeros of $\xi(s)$ all lie in the critical strip 0 < Re(s) < 1.

Remark 17.16. We will usually work with Z(s) and deal with the poles rather than making it holomorphic by introducing additional factors; some authors use $\xi(s)$ to denote our Z(s).