6 Two examples – the Gamma function and the Riemann ζ-function

6.1 The Gamma function

A natural question to ask is whether the factorial function $n!$ is actually giving the values at nonnegative integers of some nice function with a larger domain. This problem had certainly been raised as early as the 1720s by Daniel Bernoulli and Christian Goldbach. As we shall show, one can actually fit a meromorphic function Γ to this data.

Euler was the first to solve Bernoulli and Goldbach's problem. In 1729 Euler had sent the formula

$$m! = \prod_{n=1}^{\infty} \frac{\left(1+\frac{1}{n}\right)^m}{1+\frac{m}{n}}$$

to Goldbach ... without fussing too much about whether anything converges! In 1730 he wrote again to Goldbach that he had shown that

$$m! = \int_0^1 (-\ln s)^m \, ds.$$

The way that the Gamma function is usually introduced is as a function of a real variable, defined via an integral

$$\Gamma(x) = \int_0^{\infty} e^{-t}\, t^{x-1}\, dt \qquad x > 0.$$

The notation and formula, which are due to Legendre in 1811, just comes from the previous formula by a change of variables. (Note that $\Gamma(x+1) = x!$.) As we saw, this integral converges if you replace x with z in the right half-plane, extending $\Gamma(z)$ for $\operatorname{Re} z > 0$.

Via a series of messy of analytic continuations one can extend the definition to give a function analytic on all of $\mathbb{C}$ except at 0 and the negative integers (where the function has simple poles). Rather than do this, we'll follow Euler's original idea and define Γ in one go via infinite products. For this we need to know that Γ should have poles at $0, -1, -2, \ldots,$ and that $\Gamma(1)$ should be $0! = 1$.

Let $a_n = -n$, for $n = 0, 1, 2, \ldots$. Then the rank of the canonical product corresponding to $\{a_n\}$ is 1 and so the canonical product is

$$Q(z) = z \prod_{n=1}^{\infty} E_1\left(\frac{-z}{n}\right) = z \prod_{n=1}^{\infty} \left(1+\frac{z}{n}\right) e^{-\frac{z}{n}}.$$

By Weierstrass's theorem, Q is entire with simple zeros at $0, -1, -2, -3, \ldots$. Hence, for any constant c, the function

$$f(z) = \frac{e^{cz}}{Q(z)}$$

is a *meromorphic function* with simple poles at $z = 0, -1, -2, -3, \ldots$. In order to defined Γ we want to choose c in order to make $f(1) = 1$. Now, at $z = 1$ all the terms in the infinite product are positive reals, so there is no problem taking logarithms.

$$\begin{aligned}
\operatorname{Log} f(1) = \ln f(1) &= c - \ln 1 - \lim_{N\to\infty} \sum_{n=1}^{N} \ln\left(\left(1+\frac{1}{n}\right)e^{-\frac{1}{n}}\right) \\
&= c - \lim_{N\to\infty} \sum_{n=1}^{N} \ln\left(\frac{n+1}{n}\right) - \frac{1}{n} \\
&= c + \lim_{N\to\infty} \sum_{n=1}^{N} \left(\frac{1}{n} - \ln(n+1) + \ln n\right) \\
&= c + \lim_{N\to\infty} \left(\sum_{n=1}^{N} \frac{1}{n}\right) - \ln(N+1)
\end{aligned}$$

since most of the log terms cancel. This requires just a small tweak, using the fact that $\lim_{N\to\infty} \ln(N+1) - \ln N = 0$:

$$\begin{aligned}
\operatorname{Log} f(1) &= c + \lim_{N\to\infty} \left(\left(\sum_{n=1}^{N} \frac{1}{n} - \ln N\right) + (\ln N - \ln(N+1))\right) \\
&= c + \lim_{N\to\infty} \left(\sum_{n=1}^{N} \frac{1}{n} - \ln N\right).
\end{aligned}$$

This last term is called *Euler's constant* γ:

$$\gamma = \lim_{N\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N} - \ln N\right)$$

Remark: γ is a fundamental constant of mathematics as are π and e. It is conjectured that γ is transcendental — it is not even known whether γ is irrational.

Since we want $\operatorname{Log} f(1) = 0$, this means that we need to choose $c = -\gamma$.

Definition: The *Gamma function* is defined as

$$\Gamma(z) := \frac{e^{-\gamma z}}{Q(z)} = \frac{e^{-\gamma z}}{z} \prod_{n=1}^{\infty} \left(1 + \frac{z}{n}\right)^{-1} e^{\frac{z}{n}}.$$

Remark: This formula for the Gamma function is due to Weierstrass (1815–1897). It is what motivated him to develop the Weierstrass Factorization Theorem.

Let's heuristically expand out Euler's original product:

$$
\begin{aligned}
m! &= \prod_{n=1}^{\infty} \frac{\left(1+\frac{1}{n}\right)^m}{1+\frac{m}{n}} \\
&= \prod_{n=1}^{\infty} \frac{(n+1)^m}{n^m\left(1+\frac{m}{n}\right)} \\
&= \left(\frac{2^m}{1+m}\right)\left(\frac{3^m}{2^{m-1}(2+m)}\right)\left(\frac{4^m}{3^{m-1}(3+m)}\right)\cdots \\
&= \frac{1\cdot 2\cdot 3\cdot 4\cdots}{(1+m)\cdot(2+m)\cdot(3+m)\cdots} ???
\end{aligned}
$$

Gauss in fact rewrote this taking a little more care about how the limit should be taken (although still not checking convergence!).

Lemma 29. *(Gauss's formula)* *For $z \neq 0, -1, -2, \ldots$,*

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!\, n^z}{z(z+1)\cdots(z+n)}.$$

Remark: The term n^z should be interpreted as the principal value, that is

$$\text{p.v.}\, n^z = \exp(z \operatorname{Log} n).$$

Proof. For any fixed z we have

$$
\begin{aligned}
\Gamma(z) &= \frac{e^{-\gamma z}}{z} \lim_{n\to\infty} \prod_{k=1}^{n} \left(\frac{k}{k+z}\right) e^{\frac{z}{k}} \\
&= \lim_{n\to\infty} \frac{n!}{z(z+1)\cdots(z+n)} \exp\left(-\gamma z + \sum_{1}^{n} \frac{z}{k}\right) \\
&= \lim_{n\to\infty} \frac{n!}{z(z+1)\cdots(z+n)} \cdot \exp(z\log n)\cdot \exp\left(z\left(-\gamma + \sum_{1}^{n}\frac{1}{k} - \log n\right)\right) \\
&= \lim_{n\to\infty} \frac{n!}{z(z+1)\cdots(z+n)} \cdot n^z \cdot \lim_{n\to\infty} \exp\left(z\left(-\gamma + \sum_{1}^{n}\frac{1}{k} - \log n\right)\right) \\
&= \lim_{n\to\infty} \frac{n! n^z}{z(z+1)\cdots(z+n)}
\end{aligned}
$$

∎

Example: Take $z = 3$. The right-hand side is then

$$\lim_{n\to\infty} \frac{n!\, n^3}{3\cdot 4\cdot 5\cdots(n+3)} = \lim_{n\to\infty} \frac{1\cdot 2\cdot n^3}{(n+1)(n+2)(n+3)} = 2 = (z-1)!$$

Gauss's formula gives the following

Lemma 30. *(The Functional Equation)* *For* $z \neq 0, -1, -2, \dots$

$$\Gamma(z+1) = z\Gamma(z).$$

Proof. For any fixed z,

$$\begin{aligned}
\Gamma(z+1) &= \lim_{n\to\infty} \frac{n!\, n^{z+1}}{(z+1)(z+2)\cdots(n+z+1)} \\
&= z \lim_{n\to\infty} \frac{n!\, n^{z+1}}{z(z+1)(z+2)\cdots(n+z+1)} \\
&= z \lim_{n\to\infty} \frac{(n+1)!\,(n+1)^{z}}{z(z+1)(z+2)\cdots(n+z+1)} \cdot \lim_{n\to\infty} \left(\frac{n}{n+1}\right)^{z+1} \\
&= z\,\Gamma(z)\cdot 1
\end{aligned}$$

where we have used Gauss' formula, as well as the continuity of $w \mapsto w^{z+1}$ at $w = 1$. ∎

We now have a function Γ which has poles at the rights places and for which $\Gamma(m+1) = m!$ for any positive integer m. We would like to show that it equals the integral formula that one usually starts with.

Lemma 31. *For* $x > 0$, $\Gamma(x) = \int_0^\infty e^{-t}\, t^{x-1}\, dt$. *Indeed,*

$$\Gamma(z) = \int_0^\infty e^{-t}\, t^{z-1}\, dt \qquad \text{for } \operatorname{Re} z > 0.$$

Proof. Suppose that $x > 0$ and that n is a nonnegative integer. First observe that

$$\begin{aligned}
\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1}\, dt &= \frac{1}{x} \int_0^n \left(1 - \frac{t}{n}\right)^{n-1} t^{x}\, dt \\
&= \frac{n-1}{n}\, \frac{1}{x(x+1)} \int_0^n \left(1 - \frac{t}{n}\right)^{n-2} t^{x+1}\, dt \\
&= \frac{n!}{n^n}\, \frac{1}{x(x+1)\cdots(x+n-1)} \int_0^n t^{x+n-1}\, dt \\
&= \frac{n!}{n^n}\, \frac{n^{x+n}}{x(x+1)\cdots(x+n-1)(x+n)}.
\end{aligned}$$

That is,

$$\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1}\, dt = \frac{n! n^x}{x(x+1)\cdots(x+n)}.$$

If we can show that $\int_0^n (1 - \frac{t}{n})^n t^{x-1}\, dt \to \int_0^\infty e^{-t} t^{x-1}\, dt$ as $n \to \infty$ then we have our result from Gauss' Lemma.

This is a little delicate as we are dealing with

- an infinite interval $[0, \infty)$;
- potentially unbounded functions (at $t = 0$);
- n appearing in the upper limit of the integral.

Fix $\epsilon > 0$. Now for $0 \le t \le 1$ and any n, $0 \le \left(1 - \frac{t}{n}\right)^n \le 1$ and $0 \le e^{-t} \le 1$. Since $\int_0^1 t^{x-1}\,dx$ converges there exists $\alpha > 0$ such that $\int_0^\alpha t^{x-1}\,dx < \epsilon/5$, and hence

$$\left|\int_0^\alpha \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt\right| < \frac{\epsilon}{5}, \qquad \left|\int_0^\alpha e^{-t}\,t^{x-1}\,dt\right| < \frac{\epsilon}{5}.$$

Also, there exists R such that

$$\left|\int_R^\infty e^{-t}\,t^{x-1}\,dt\right| < \frac{\epsilon}{5}.$$

If $R \le t \le n$ then

$$0 \le \left(1 - \frac{t}{n}\right)^n < e^{-t}$$

and so

$$\left|\int_R^n \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt\right| < \left|\int_R^n e^{-t}\,t^{x-1}\,dt < \right|\int_R^\infty e^{-t}\,t^{x-1}\,dt\right| < \frac{\epsilon}{5}.$$

Now on the compact subset $[\alpha, R]$

$$\left(1 - \frac{t}{n}\right)^n t^{x-1} \to e^{-t}\,t^{x-1}$$

uniformly in t as $n \to \infty$, and so for large enough n,

$$\left|\int_\alpha^R \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt - \int_\alpha^R e^{-t}\,t^{x-1}\,dt\right| < \frac{\epsilon}{5}$$

too. Putting this all together, for large enough n,

$$\begin{aligned}
&\left|\int_0^n \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt - \int_0^\infty e^{-t}\,t^{x-1}\,dt\right| \\
&\quad = \left|\int_0^\alpha \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt + \int_\alpha^R \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt + \int_R^n \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt\right. \\
&\qquad\quad \left. - \int_0^\alpha e^{-t}\,t^{x-1}\,dt - \int_\alpha^R e^{-t}\,t^{x-1}\,dt - \int_R^\infty e^{-t}\,t^{x-1}\,dt\right| \\
&\quad \le \left|\int_0^\alpha \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt\right| + \left|\int_R^n \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt\right| + \left|\int_0^\alpha e^{-t}\,t^{x-1}\,dt\right| \\
&\qquad\quad + \left|\int_R^\infty e^{-t}\,t^{x-1}\,dt\right| + \left|\int_\alpha^R \left(1 - \frac{t}{n}\right)^n t^{x-1}\,dt - \int_\alpha^R e^{-t}\,t^{x-1}\,dt\right| \\
&\quad < \epsilon.
\end{aligned}$$

Exercise: Show that $\int_0^\infty t^{z-1}e^{-t}\,dt$ is analytic for $\operatorname{Re} z > 0$ and therefore by analytic continuation coincides with $\Gamma(z)$ in that region. ■

The Gamma function satisfies the following useful functional equation.

Lemma 32. *For $z \in \mathbb{C} \setminus \mathbb{Z}$,*

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

Proof.

$$\begin{aligned}
&\Gamma(z)\,\Gamma(1-z)\\
&= \lim_{n\to\infty} \frac{(n!)^2\, n^z\, n^{1-z}}{z(z+1)\cdots(z+n)(1-z)(2-z)\cdots(n+1-z)}\\
&= \lim_{n\to\infty} \frac{1}{z}\left\{(1+z)\left(1+\frac{z}{2}\right)\cdots\left(1+\frac{z}{n}\right)(1-z)\left(1-\frac{z}{2}\right)\cdots\left(1-\frac{z}{n}\right)\right\}^{-1}\cdot\frac{n}{n+1-z}\\
&= \lim_{n\to\infty}\frac{1}{z}\prod_{k=1}^{n}\left(1-\frac{z^2}{k^2}\right)^{-1} = \frac{1}{z}\,\frac{\pi z}{\sin \pi z} = \frac{\pi}{\sin \pi z}.
\end{aligned}$$

■

Example: Taking $z = \pi/2$, $\Gamma\left(\frac{1}{2}\right)^2 = \pi$.

6.2 The Riemann ζ-function

The **Basel Problem** was posed by Pietro Mengoli in 1644: compute $\sum_{n=1}^{\infty}\frac{1}{n^2}$. This was finally solved by Euler in 1735. As we saw

$$\frac{\sin(z)}{z} = \prod_{n=1}^{\infty}\left(1-\frac{z^2}{\pi^2 n^2}\right).$$

If you are bold, you multiply out the brackets:

$$\frac{\sin(z)}{z} = 1 - \frac{1}{\pi^2}\left(\sum_{n=1}^{\infty}\frac{1}{n^2}\right)z^2 + (\dots)z^4 + \dots$$

On the other hand, writing our the Taylor series

$$\frac{\sin(z)}{z} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \dots$$

and so, on comparing coefficients, Euler concluded that

$$\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{1}{6!}.$$

Not surprisingly, Euler then considered the function $s \mapsto \sum_{n=1}^{\infty} \frac{1}{n^s}$ for other integer values of $s > 1$. Chebyshev (1821–1894) then extended this to real $s > 1$.

In 1859 Riemann extended this further to complex values and provided our current notation. For $\operatorname{Re} z > 1$, let

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

Given any $\delta > 0$, the series is uniformly and absolutely convergent on $\operatorname{Re} z \geq 1 + \delta$ because then $|n^{-z}| \leq n^{-1-\delta}$. Since each term $n^{-z} = \exp(-z \operatorname{Log} n)$ is an entire function it follows that ζ is analytic on $\operatorname{Re} z > 1$.

One reason why the ζ-function is so central in number theory stems from the fact that it contains information about primes.

Lemma 33. *(Euler's formula)* *For* $\operatorname{Re} z > 1$,

$$\zeta(z) = \prod_{n=1}^{\infty} \left(1 - \frac{1}{p_n^z}\right)^{-1} = \prod_{n=1}^{\infty} \left(1 - p_n^{-z}\right)^{-1}.$$

where p_n *is the* n*th prime.*

Proof. This is just about a statement that the Sieve of Eratosthenes removes all the primes! To see this

$$\zeta(z) = 1 + \frac{1}{2^z} + \frac{1}{3^z} + \frac{1}{4^z} + \frac{1}{5^z} + \dots$$

so

$$\frac{1}{2^z}\zeta(z) = \frac{1}{2^z} + \frac{1}{4^z} + \frac{1}{6^z} + \frac{1}{8^z} + \dots.$$

Subtracting we get

$$\left(1 - \frac{1}{2^z}\right)\zeta(z) = 1 + \frac{1}{3^z} + \frac{1}{5^z} + \frac{1}{7^z} + \frac{1}{9^z} + \dots$$

Repeating

$$\left(1 - \frac{1}{3^z}\right)\left(1 - \frac{1}{2^z}\right)\zeta(z) = 1 + \frac{1}{5^z} + \frac{1}{7^z} + \frac{1}{11^z} + \dots$$

gives the series with all the terms corresponding to multiples of 2 and 3 removed. Indeed for any N

$$\zeta(z) \prod_{n=1}^{N} \left(1 - \frac{1}{p_n^z}\right) = 1 + \frac{1}{p_{n+1}^z} + \dots \tag{6}$$

having removed all the multiples of 2 up to p_n.

For any z with $\operatorname{Re} z > 1$, the convergence of $\zeta(z)$ implies that the right-hand side of (6) converges to 1, and hence

$$\prod_{n=1}^{\infty}\left(1-\frac{1}{p_n^z}\right)=\frac{1}{\zeta(z)}$$

which gives the required result. ■

Another proof can be found by writing each term $\left(1-p_n^{-z}\right)^{-1}$ out as a geometric series.

The Riemann zeta function is related to the Gamma function via the following identity.

Theorem 34. *For* $\operatorname{Re} z > 1$,

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

Proof. For t near ∞ the term $(e^t-1)^{-1}$ behaves like e^{-t} whilst for t near 0 it behaves like t^{-1}. Consequently it is easy to check that the two improper integrals $\int_0^1(e^t-1)^{-1}t^{z-1}\,dt$ and $\int_1^{\infty}(e^t-1)^{-1}t^{z-1}\,dt$ converge uniformly on compact sets in $\operatorname{Re} z > 1$. Consequently, $\int_0^{\infty}(e^t-1)^{-1}t^{z-1}\,dt$ is analytic on $\operatorname{Re} z > 1$ and therefore, by analytic continuation, it is enough to check the above equality for $x > 1$.

For such x,

$$\begin{aligned}
\zeta(x)\Gamma(x) &= \lim_{N\to\infty}\sum_{n=1}^{N}\frac{1}{n^x}\cdot\int_0^{\infty}e^{-u}u^{x-1}\,du\\
&= \lim_{N\to\infty}\sum_{n=1}^{N}\int_0^{\infty}e^{-u}\left(\frac{u}{n}\right)^{x-1}\frac{du}{n}\\
&\qquad\text{(let } u=nt\text{)}\\
&= \lim_{N\to\infty}\sum_{n=1}^{N}\int_0^{\infty}e^{-nt}t^{x-1}\,dt\\
&= \lim_{N\to\infty}\int_0^{\infty}\sum_{n=1}^{N}\left(e^{-t}\right)^n t^{x-1}\,dt\\
&= \lim_{N\to\infty}\int_0^{\infty}\frac{e^{-t}-e^{-(N+1)t}}{1-e^{-t}}\,t^{x-1}\,dt.
\end{aligned}$$

At this point let

$$g_N(t)=\frac{e^{-t}-e^{-(N+1)t}}{1-e^{-t}}\,t^{x-1},\qquad g(t)=\frac{e^{-t}}{1-e^{-t}}\,t^{x-1}.$$

Then $0 \le g_N(t) \le g(t)$. Since $\int_0^\infty g(t)\,dt$ converges and $g_N \to g$ pointwise on $(0,\infty)$, the Dominated Convergence Theorem tells us that $\int_0^\infty g_N(t)\,dt \to \int_0^\infty g(t)\,dt$. That is

$$\begin{aligned}\zeta(x)\Gamma(x) &= \int_0^\infty \frac{e^{-t}}{1-e^{-t}}\,t^{x-1}\,dt\\ &= \int_0^\infty (e^t-1)^{-1}t^{z-1}\,dt\end{aligned}$$

as claimed. ■

Introduced as we did above, one still has the issue of extending ζ to be a meromorphic function on the plane. We could try to proceed as we did for the Gamma function, and use a product expansion. Let us label the zeros of the zeta function as $\{a_n\}$. The following expansion is due to Hadamard.

Theorem 35. *If* $\operatorname{Re} z > 1$, *then*

$$\zeta(z) = \sum_{n=1}^\infty \frac{1}{n^z} = \frac{e^{(\ln(2\pi)-1-\gamma/2)z}}{2(z-1)\Gamma(1+z/2)} \prod_{n=1}^\infty \left(1-\frac{z}{a_n}\right)e^{z/a_n}.$$

Actually if you group the terms the right way you can also make sense of writing

$$\zeta(z) = \sum_{n=1}^\infty \frac{1}{n^z} = \frac{\pi^{z/2}}{2(z-1)\Gamma(1+z/2)} \prod_{n=1}^\infty \left(1-\frac{z}{a_n}\right).$$

The problem with these expansions is that (unlike the case for Γ) we don't know precisely what the zeros of ζ are. One is therefore stuck with the rather involved series of analytic extensions that we described earlier. Each of these steps requires showing that a certain formula

(a) defines an analytic function

(b) agrees with the previous definition on some reasonable sized set.

Rather than head in this direction, we'll look (without proof) at a few of the properties of ζ. One thing that Hadamard's formula does show is that ζ is a meromorphic function. With a little more analysis one can show that the product has a single simple pole at $z = 1$ and is analytic elsewhere.

The final stage in the continuation process uses the following.

Theorem 36. *(Riemann's functional equation) For* $z \ne 0, 1, 2, \ldots$

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

This equation allows us to read off some of the properties of ζ. Suppose, for example that $z = -2k$ $(k = 1, 2, \dots)$. Then $\zeta(z) = 0$ since the sin term vanishes and the rest of the terms are finite. These are called the trivial zeros of ζ. A good deal of modern number theory concerns identifying where the other zeros are. By Euler's product formula, ζ has no zeros with $\operatorname{Re} z > 1$. If $\operatorname{Re} z < 0$ then $\Gamma(1-z)$ and $\zeta(1-z)$ are therefore nonzero, and so there are no nontrivial zeros there either. Thus, every nontrivial zero of ζ lies in the strip $0 \le \operatorname{Re} z \le 1$.

The statement that none of the zeros lies on the line $\operatorname{Re} z = 1$ is equivalent to the Prime Number Theorem, that is, if $\pi(x)$ denotes the number of primes less than or equal to x, then

$$\lim_{x\to\infty} \frac{\pi(x)}{x \ln x} = 1.$$

(See H.G. Diamond, Elementary methods in the study of the distribution of prime numbers. *Bull. Amer. Math. Soc. (N.S.)* **7** (1982), 553–589.)

A more precise version of the Prime Number Theorem says that

$$\pi(x) = \operatorname{Li}(x) + O(xe^{-c\sqrt{\ln x}})$$

where $\operatorname{Li}(x) = \displaystyle\int_2^x \frac{1}{\ln t}\, dt = \operatorname{li}(x) - \operatorname{li}(x)$ is the offset logarithmic integral function.

The Riemann hypothesis says that in fact all the nontrivial zeros lie on the line $\operatorname{Re} z = \frac{1}{2}$. Assuming this, one can get a better estimate for $\pi(x)$:

$$|\pi(x) - \operatorname{li}(x)| \le \frac{\sqrt{x}\ln x}{8\pi}.$$

7 The Open Mapping Principle and the Inverse Function Theorem

[Chapter 8 of Brown and Churchill, Seventh Edition, has a good coverage of mappings by elementary functions and in addition gives a nice introduction to Riemann surfaces. Chapter 9 covers conformal mappings.]

One of the great challenges in complex analysis is to get a good geometric feeling for what analytic functions are doing. One no longer has a graph to look at. The best that one can often do is to try to understand what the function does to certain subsets of $\mathbb{C}$.

Example: Let $f(z) = \dfrac{z+1}{z-1}$. This is an example of a Möbius transformation and you saw in second year that it maps (lines and circles) to (lines and circles) in the (extended) complex plane. Furthermore, Möbius transformation are invertible. A consequence of this is that the boundary of a region Ω gets mapped to the boundary of $f(\Omega)$. In this case, if we want to identify $f(\mathbb{D})$ we can see that the unit circle maps to the line through $f(-1) = 0$ and $f(i) = \frac{i+1}{i-1} = -i$, that is, the imaginary axis. This is the boundary of $f(\mathbb{D})$. The point 0 maps to -1 and so $f(\mathbb{D})$ must be the left half-plane.

Example: . Let $f(z) = \dfrac{4z}{(z+1)^2}$. This is NOT a Möbius transformation! One can show (exercise) that

$$f(e^{i\theta}) = \frac{2}{\cos\theta + 1}$$

and so the image of the unit circle is the ray $[1, \infty)$. So where does $\mathbb{D}$ map to? This is already quite complicated. For example, the interval $(-1, 1]$ maps to $(-\infty, 1]$, while the interval from $-i$ to i maps to the circle centred at 1 of radius 1.

Actually, in this case, for $z \neq 0$,

$$f(1/z) = \frac{4/z}{(1/z+1)^2} = f(z)$$

so in particular, the image of the punctured disk $0 < |z| < 1$ is exacly the same as the image of the outside. Indeed, for all $w \neq 0$, the equation $f(z) = w$ is a quadratic equation in z which has two solutions

$$z = \frac{1 - 2\,w \pm \sqrt{1 - 4\,w}}{2w}.$$

Thus f is an onto function. The image of the open unit disk is $\mathbb{C} \setminus [1, \infty)$.

Example: What about $f(z) = \sin z$. This is clearly not one-to-one. Even working out the image of the unit circle is not at all easy: writing

$$f(e^{i\theta}) = \frac{e^{ie^{i\theta}} - e^{-ie^{i\theta}}}{2i}$$

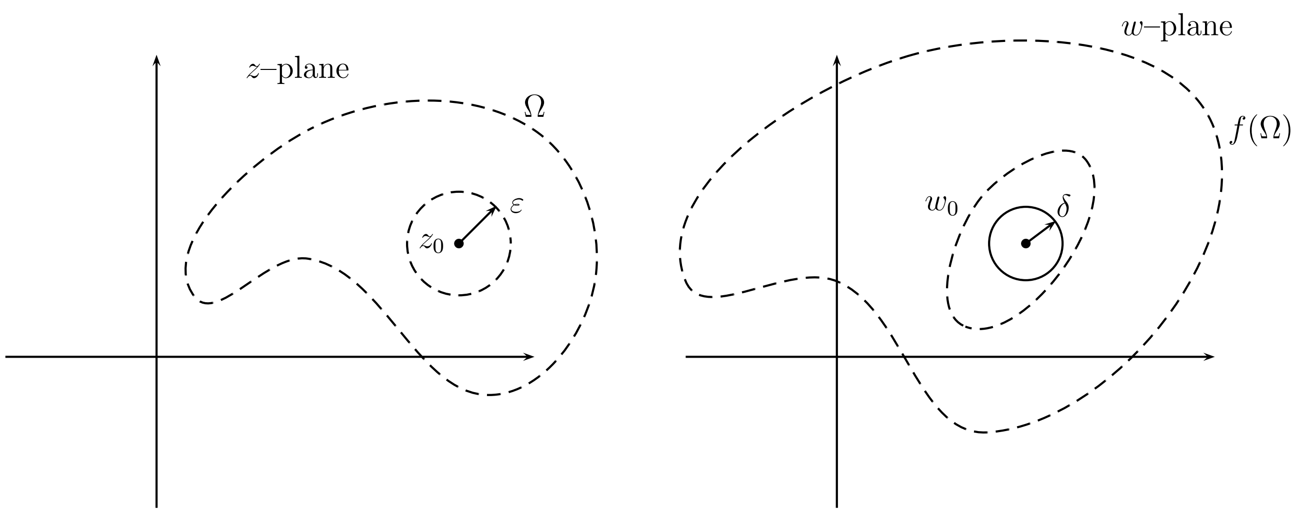
doesn't look at that promising! If you look at the image of the circle of radius 2, it has a couple of loops in it. This doesn't happen with small circles however (heuristically because $\sin z \approx z$ for small z.

We need some extra facts about analytic maps to progress.

7.1 Open mappings

Our first result says that a nonconstant analytic function never squashes open balls flat!

Theorem 37. *(The Open Mapping Principle) If f is analytic in a region Ω, and is not constant, then $w = f(z)$ maps open sets of Ω into open sets in the w-plane. More specifically, if z_0 is in Ω and if $w_0 = f(z_0)$ then for every $\epsilon > 0$ there is a $\delta > 0$ such that the image of $|z - z_0| < \epsilon$ contains the ball $|w - w_0| < \delta$.*



Proof. The function $g(z) = f(z) - w_0$ clearly has a zero at z_0, of order n say. Since nonconstant analytic functions have isolated zeros there exists $0 < \rho < \epsilon$ such that $f(z) - w_0$ does not vanish for $0 < |\, z - z_0 \,| \leq \rho$. Let δ be the minimum of $|f(z) - w_0|$ on the circle $|z - z_0| = \rho$. By compactness, $\delta > 0$. We are going to show that $D(w_0, \delta) \subseteq f(D(z_0, \rho)) \subseteq f(D(z_0, \epsilon))$.

Suppose now that $w = w_0 + \alpha$ is an arbitrary point in the disk centred at w_0 of radius δ (so $|\alpha| < \delta$. Then $h(z) = f(z) - w_0 - \alpha$ has the same number of zeros in the disk $|z - z_0| < \rho$ by Rouché's Theorem. That is, h has n zeros in this disk if counted according to multiplicity. In particular, there exists at least one point z in this disk at which $f(z) = w$. Thus $D(w_0, \delta)$ is contained in $f(D(z_0, \rho))$. ■

Corollary 38. *$f(\Omega)$ is a region (if Ω is a region).*

Proof. We have just proved that Ω is open. Since f is continuous $f(\Omega)$ is connected ("connectedness" is a topological concept). ■

Remark: While connectedness is preserved under continuous maps, simple connectedness is not. For example, let $f(z) = z^2$ and let $\Omega = \{re^{i\theta} : 1 < r < 2, \quad 0 < \theta < 5\}$.

For a differentiable function $f : (a, b) \to \mathbb{R}$ we know that if $f'(x)$ is never zero on (a, b) then f is one-to-one on (a, b). This is a simple consequence of Rolle's Theorem. Unfortunately, for complex functions there is no Rolle's Theorem, and the fact that $f'(z)$ does not vanish on some region Ω is not enough to deduce that f is one-to-one on Ω — see the example above! The best that one could hope for is local invertibility.

Recall that if $F : \mathbb{R}^2 \to \mathbb{R}^2$,

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u(x, y) \\ v(x, y) \end{pmatrix}$$

is continuously differentiable at (x_0, y_0), and

$$DF = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}$$

is invertible there, then F is one-to-one on some neighbourhood of (x_0, y_0). Of course every analytic function $f : \Omega \to \mathbb{C}$ can be thought of as corresponding to such a function, with some particular restrictions on the components u and v.

Theorem 39. *Suppose f is analytic on Ω, and that $z_0 \in \Omega$. If $f'(z_0) \neq 0$ then f is one-to-one on some neighbourhood of z_0.*

Proof. Let $f = u + iv$ and let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the corresponding function as above. Note that

$$\det(DF) = u_x v_y - v_x u_y = u_x^2 + u_y^2 = v_x^2 + v_y^2$$

so $\det(DF)(x, y) = 0$ if and only if $f'(x + iy) = 0$.

In particular, if $f'(z_0) = f'(x_0 + iy_0) \neq 0$ then F and hence f is one-to-one/invertible on some open neighbourhood of z_0. ■

What is perhaps surprising is that the converse of this result holds. Note that the corresponding result for $F : \mathbb{R}^2 \to \mathbb{R}^2$ is not true. One can take, for example

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^3 \\ y \end{pmatrix}$$

which is a bijection from $\mathbb{R}^2$ to $\mathbb{R}^2$, but DF is not invertible at $(0, 0)$.

Theorem 40. *Suppose f is analytic on Ω, and that $z_0 \in \Omega$. If $f'(z_0) = 0$ then f is not one-to-one on any neighbourhood of z_0.*

Proof. As in the proof of the Open Mapping Principle, let $w_0 = f(z_0)$ and let $g(z) = f(z) - w_0$. If $f'(z_0) = 0$ then the series expansion for f around z_0 looks like

$$f(z) = f(z_0) + \frac{f''(z_0)}{2}(z - z_0)^2 + \dots$$

and so g has a zero of order at least 2 at z_0. Proceeding as in the proof of the OMP we see that for any small $\epsilon > 0$ there will be a $\delta > 0$ such that for all $w \in D(w_0, \delta)$ there are at least two solutions (by multiplicity) in $D(z_0, \epsilon)$ to $f(z) = w$.

If f were one-to-one on such a neighbourhood this would require that these multiple solutions were all repeated roots. This would imply that $f'(z) = 0$ for all $z \in f^{-1}(D(w_0, \delta))$. Since f is continuous, this inverse image is open and hence contains an open disk $D(z_0, r)$. But if $f'(z) = 0$ on an open disk around z_0 it must be constant on that disk which would contradict that f were one-to-one. ■

Example: Let $f(z) = z^2$ (or z^3 or $\dots$). Think about what f does to a little ball around $z_0 \neq 0$. Then think about what it does to any little ball around $z_0 = 0$. The above discussion is just making a little more precise the statement that locally an analytic function behaves like its affine approximation $g(z) = f(z_0) + f'(z_0)(z - z_0)$.