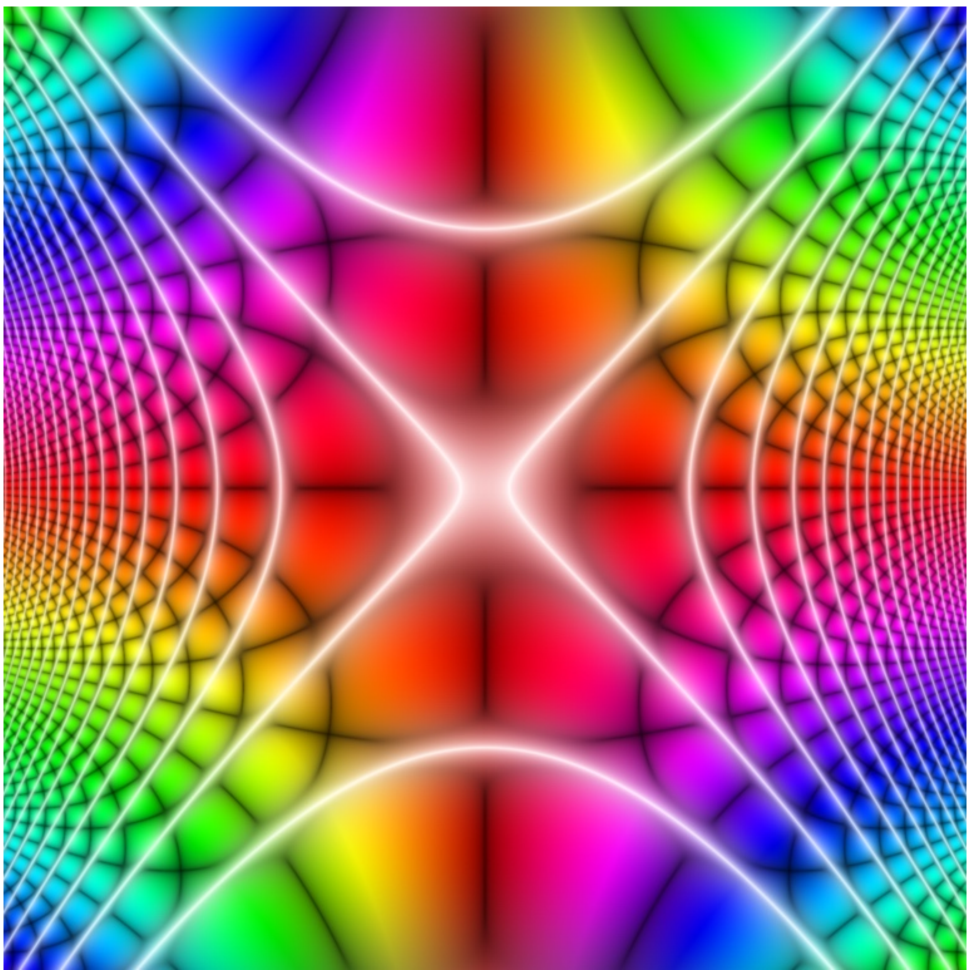
Complex Analysis Lecture Notes

Dan Romik



About this document. These notes were created for use as primary reading material for the graduate course *Math 205A: Complex Analysis* at UC Davis. The notes were not heavily vetted for accuracy and may contain minor typos or errors. You can help me continue to improve them by emailing me with any comments or corrections you have.

The current 2020 revision (dated 2020/05/14) updates my earlier version of the notes from 2018. With some exceptions, the exposition follows the textbook *Complex Analysis* by E. M. Stein and R. Shakarchi (Princeton University Press, 2003).

Acknowledgements. I am grateful to Christopher Alexander, Jennifer Brown, Brynn Caddel, Keith Conrad, Bo Long, Anthony Nguyen, Jianping Pan, and Brad Velasquez for comments that helped me improved the notes. Figure 5 on page 21 was created by Jennifer Brown and is used with her permission.

Complex Analysis Lecture Notes

Document version: May 14, 2020

Email comments and feedback to `romik@math.ucdavis.edu`

Cover figure: a heat map plot of the entire function $z \mapsto z(z-1)\pi^{-z/2}\Gamma(z/2)\zeta(z)$.

Created with Mathematica using code by Simon Woods, available at

http://mathematica.stackexchange.com/questions/7275/how-can-i-generate-this-domain-coloring-plot

Contents

1 Introduction: why study complex analysis?

These notes are about *complex analysis*, the area of mathematics that studies analytic functions of a complex variable and their properties. While this may sound a bit specialized, there are (at least) two excellent reasons why all mathematicians should learn about complex analysis. First, it is, in my humble opinion, one of the most beautiful areas of mathematics. One way of putting it that has occurred to me is that complex analysis has a very high ratio of theorems to definitions (i.e., a very low "entropy"): you get a lot more as "output" than you put in as "input."

The second reason is complex analysis has a large number of applications (in both the pure math and applied math senses of the word) to things that seem like they ought to have little to do with complex numbers. For example:

- Solving polynomial equations: historically, this was the motivation for introducing complex numbers by Cardano, who published the famous formula for solving cubic equations in 1543, after learning of the solution found earlier by Scipione del Ferro. An important point to keep in mind is that Cardano's formula sometimes requires taking operations in the complex plane as an intermediate step to get to the final answer, even when the cubic equation being solved has only real roots.

  Example 1. Using Cardano's formula, it can be found that the solutions to the cubic equation

  $$z^3 + 6z^2 + 9z + 3 = 0$$

  are

  $$\begin{aligned} z_1 &= 2\cos(2\pi/9) - 2, \\ z_2 &= 2\cos(8\pi/9) - 2, \\ z_3 &= 2\sin(\pi/18) - 2. \end{aligned}$$

- Proving Stirling's formula: $n! \sim \sqrt{2\pi n}(n/e)^n$. Here, $a_n \sim b_n$ is the standard "asymptotic to" relation, defined to mean $\lim_{n\to\infty} a_n/b_n = 1$.
- Proving the prime number theorem: $\pi(n) \sim \frac{n}{\log n}$, where $\pi(n)$ denotes the number of primes less than or equal to n (the prime-counting function).
- Proving many other asymptotic formulas in number theory and combinatorics, e.g. (to name one other of my favorite examples), the Hardy-Ramanujan formula

  $$p(n) \sim \frac{1}{4\sqrt{3}n} e^{\pi\sqrt{2n/3}},$$

  where $p(n)$ is the number of integer partitions of n.
- Evaluation of complicated definite integrals, for example

  $$\int_0^\infty \sin(t^2)\,dt = \frac{1}{2}\sqrt{\frac{\pi}{2}}.$$

  (This application is strongly emphasized in older textbooks, and has been known to result in a mild case of post-traumatic stress disorder.)

- Solving physics problems in hydrodynamics, heat conduction, electrostatics and more.
- Analyzing alternating current electrical networks by extending Ohm's law to electrical impedance. Complex analysis also has many other important applications in electrical engineering, signals processing and control theory.
- Probability and combinatorics, e.g., the Cardy-Smirnov formula in percolation theory and the connective constant for self-avoiding walks on the hexagonal lattice.
- It was proved in 2016 that the optimal densities for sphere packing in 8 and 24 dimensions are $\pi^4/384$ and $\pi^{12}/12!$, respectively. The proofs make spectacular use of complex analysis (and more specifically, a part of complex analysis that studies certain special functions known as modular forms).
- Nature uses complex numbers in Schrödinger's equation and quantum field theory. This is not a mere mathematical convenience or sleight-of-hand, but in fact appears to be a built-in feature of the very equations describing our physical universe. Why? No one knows.
- Conformal maps, which come up in purely geometric applications where the algebraic or analytic structure of complex numbers seems irrelevant, are in fact deeply tied to complex analysis. Conformal maps were used by the Dutch artist M.C. Escher (though he had no mathematical training) to create amazing art, and used by others to better understand, and even to improve on, Escher's work. See Fig. 1, and see [10] for more on the connection of Escher's work to mathematics.

Figure 1: *Print Gallery*, a lithograph by M.C. Escher which was discovered to be based on a mathematical structure related to a complex function $z \mapsto z^\alpha$ for a certain complex number α, although it was constructed by Escher purely using geometric intuition. See the paper [8] and this website, which has animated versions of Escher's lithograph brought to life using the mathematics of complex analysis.

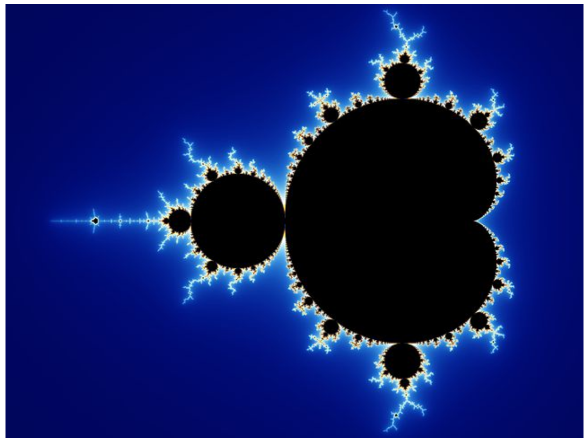


Figure 2: The Mandelbrot set. [Source: Wikipedia]

- Complex dynamics, e.g., the iconic Mandelbrot set. See Fig. 2.

There are many other applications and beautiful connections of complex analysis to other areas of mathematics. (If you run across some interesting ones, please let me know!)

In the next section I will begin our journey into the subject by illustrating a few beautiful ideas and along the way begin to review the concepts from undergraduate complex analysis.

2 The fundamental theorem of algebra

One of the most famous theorems in complex analysis is the not-very-aptly named Fundamental Theorem of Algebra. This seems like a fitting place to start our journey into the theory.

Theorem 1 (The Fundamental Theorem of Algebra.)**.** *Every nonconstant polynomial $p(z)$ over the complex numbers has a root.*

The fundamental theorem of algebra is a subtle result that has many beautiful proofs. I will show you three of them. Let me know if you see any "algebra"...

First proof: analytic proof. Let

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0$$

be a polynomial of degree $n \geq 1$, and consider where $|p(z)|$ attains its infimum.

First, note that it can't happen as $|z| \to \infty$, since

$$|p(z)| = |z|^n \cdot (|a_n + a_{n-1} z^{-1} + a_{n-2} z^{-2} + \ldots + a_0 z^{-n}|),$$

and in particular $\lim_{|z|\to\infty} \frac{|p(z)|}{|z|^n} = |a_n|$, so for large $|z|$ it is guaranteed that $|p(z)| \geq |p(0)| = |a_0|$. Fixing some radius $R > 0$ for which $|z| > R$ implies $|p(z)| \geq |a_0|$, we therefore have that

$$m_0 := \inf_{z \in \mathbb{C}} |p(z)| = \inf_{|z| \leq R} |p(z)| = \min_{|z| \leq R} |p(z)| = |p(z_0)|$$

where $z_0 = \arg\min\limits_{|z| \leq R} |p(z)|$, and the minimum exists because $p(z)$ is a continuous function on the disc $D_R(0)$.

Denote $w_0 = p(z_0)$, so that $m_0 = |w_0|$. We now claim that $m_0 = 0$. Assume by contradiction that it doesn't, and examine the local behavior of $p(z)$ around z_0; more precisely, expanding $p(z)$ in powers of $z - z_0$ we can write

$$p(z) = w_0 + \sum_{j=1}^{n} c_j (z - z_0)^j = w_0 + c_k (z - z_0)^k + \ldots + c_n (z - z_0)^n,$$

where k is the minimal positive index for which $c_j \neq 0$. (Exercise: why can we expand $p(z)$ in this way?) Now imagine starting with $z = z_0$ and traveling away from z_0 in some direction $e^{i\theta}$. What happens to $p(z)$? Well, the expansion gives

$$p(z_0 + re^{i\theta}) = w_0 + c_k r^k e^{ik\theta} + c_{k+1} r^{k+1} e^{i(k+1)\theta} + \ldots + c_n r^n e^{in\theta}.$$

When r is very small, the power r^k dominates the other terms r^j with $k < j \leq n$, i.e.,

$$\begin{aligned} p(z_0 + re^{i\theta}) &= w_0 + r^k (c_k e^{ik\theta} + c_{k+1} r e^{i(k+1)\theta} + \ldots + c_n r^{n-k} e^{in\theta}) \\ &= w_0 + c_k r^k e^{ik\theta} (1 + g(r, \theta)), \end{aligned}$$

where $\lim_{r \to 0} |g(r, \theta)| = 0$. To reach a contradiction, it is now enough to choose θ so that the vector $c_k r^k e^{ik\theta}$ "points in the opposite direction" from w_0, that is, such that

$$\frac{c_k r^k e^{ik\theta}}{w_0} \in (-\infty, 0).$$

Obviously this is possible: take $\theta = \frac{1}{k}(\arg w_0 - \arg(c_k) + \pi)$. It follows that, for r small enough,

$$|w_0 + c_k r^k e^{ik\theta}| < |w_0|$$

and for r small enough (possibly even smaller than the previous small r)

$$|p(z_0 + re^{i\theta})| = |w_0 + c_k r^k e^{ik\theta} (1 + g(r, \theta))| < |w_0|,$$

a contradiction. This completes the proof. $\square$

Exercise 1. Complete the last details of the proof (for which r are the inequalities valid, and why?) Note that "complex analysis" is part of "analysis" — you need to develop facility with such estimates until they become second nature.

Second proof: topological proof. Let $w_0 = p(0)$. If $w_0 = 0$, we are done. Otherwise consider the image under p of the circle $|z| = r$. Specifically:

1. For r very small the image is contained in a neighborhood of w_0, so it cannot "go around" the origin.

2. For r very large we have

$$\begin{aligned} p(re^{i\theta}) &= a_n r^n e^{in\theta}\left(1 + \frac{a_{n-1}}{a_n} r^{-1} e^{-i\theta} + \ldots + \frac{a_0}{a_n} r^{-n} e^{-in\theta}\right) \\ &= a_n r^n e^{in\theta}(1 + h(r,\theta)) \end{aligned}$$

where $\lim_{r\to\infty} h(r,\theta) = 0$ (uniformly in θ). As θ goes from 0 to 2π, this is a closed curve that goes around the origin n times (approximately in a circular path, that becomes closer and closer to a circle as $r \to \infty$).

As we gradually increase r from 0 to a very large number, in order to transition from a curve that doesn't go around the origin to a curve that goes around the origin n times, there has to be a value of r for which the curve crosses 0. That means the circle $|z| = r$ contains a point such that $p(z) = 0$, which was the claim. □

Remark 1. The argument presented in the topological proof is imprecise. It can be made rigorous in a couple of ways — one way we will see a bit later is using Rouché's theorem and the argument principle. This already gives a hint as to the importance of subtle topological arguments in complex analysis.

Remark 2. The topological proof should be compared to the standard calculus proof that any odd-degree polynomial over the reals has a real root. That argument is also "topological" (based on the mean value theorem), although much more trivial.

Third proof: standard textbook proof (or: "hocus-pocus" proof). Recall:

Theorem 2 (Liouville's theorem.)**.** *A bounded entire function is constant.*

Assuming this result, if $p(z)$ is a polynomial with no root, then $1/p(z)$ is an entire function. Moreover, it is bounded, since as we noted before $\lim_{|z|\to\infty} \frac{|p(z)|}{|z|^n} = |a_n|$, so $\lim_{|z|\to\infty} 1/p(z) = 0$. It follows that $1/p(z)$ is a constant, which then has to be 0, which is a contradiction. □

To summarize this section, we saw three proofs of the fundamental theorem of algebra. They are all beautiful — the "hocus-pocus" proof certainly packs a punch, which is why it is a favorite of complex analysis textbooks — but personally I like the first one best since it is elementary and doesn't use Cauchy's theorem or any of its consequences, or subtle topological concepts. Moreover, it is a "local" argument that is based on understanding how a polynomial behaves locally, where by contrast the other two proofs can be characterized as "global." It is a general philosophical principle in analysis (that has analogies in other areas, such as number theory and graph theory) that local arguments are easier than global ones.

3 Analyticity, conformality and the Cauchy-Riemann equations

In this section we begin to build the theory by laying the most basic cornerstone of the theory, the definition of analyticity, along with some of the useful ways to think about this fundamental concept.

3.1 Definition and basic meanings of analyticity

Definition 1 (analyticity)**.** A function $f(z)$ of a complex variable is *holomorphic* (a.k.a. *complex-differentiable*, *analytic*[1]) at z if the limit

$$f'(z) := \lim_{h\to 0} \frac{f(z+h)-f(z)}{h}$$

exists. In this case we call $f'(z)$ the *derivative of* f *at* z.

In the case when $f'(z) \neq 0$, the existence of the derivative has a geometric meaning: if we write the polar decomposition $f'(z) = re^{i\theta}$ of the derivative, then for points w that are close to z, we will have the approximate equality

$$\frac{f(w)-f(z)}{w-z} \approx f'(z) = re^{i\theta},$$

or equivalently

$$f(w) \approx f(z) + re^{i\theta}(w-z) + [\text{lower order terms}],$$

where "lower order terms" refers to a quantity that is much smaller in magnitude that $|w-z|$. Geometrically, this means that to compute $f(w)$, we start from $f(z)$, and move by a vector that results by taking the displacement vector $w-z$, rotating it by an angle of θ, and then scaling it by a factor of r (which corresponds to a magnification if $r > 1$, a shrinking if $0 < r < 1$, or doing nothing if $r = 1$). This idea can be summarized by the slogan:

"*Analytic functions behave locally as a rotation composed with a scaling.*"

The local behavior of analytic functions in the case $f'(z) = 0$ is more subtle; we will consider that a bit later.

A further interpretation of the meaning of analyticity is that analytic functions are *conformal mappings* where their derivatives don't vanish. More precisely, if γ_1, γ_2 are two differentiable planar curves such that $\gamma_1(0) = \gamma_2(0) = z$, f is differentiable at z and $f'(z) \neq 0$, then, denoting $v_1 = \gamma_1'(0)$, $v_2 = \gamma_2'(0)$, $w_1 = (f \circ \gamma_1)'(0)$, $w_2 = (f \circ \gamma_2)'(0)$, we can write the inner products (in the ordinary sense of vector geometry) between the complex number pairs v_1, v_2 and w_1, w_2 as

$$\begin{aligned}
\langle v_1, v_2 \rangle &= \operatorname{Re}(v_1 \overline{v_2}), \\
\langle w_1, w_2 \rangle &= \langle (f'(\gamma_1(0))\gamma_1'(0)), (f'(\gamma_2(0))\gamma_2'(0)) \rangle \\
&= f'(z)\overline{f'(z)}\langle v_1, v_2 \rangle = |f'(z)|^2 \langle v_1, v_2 \rangle.
\end{aligned}$$

If we denote by θ (resp. φ) the angle between v_1, v_2 (resp. w_1, w_2), it then follows that

$$\cos\varphi = \frac{\langle w_1, w_2 \rangle}{|w_1|\,|w_2|} = \frac{|f'(z)|^2 \langle v_1, v_2 \rangle}{|f'(z)v_1|\,|f'(z)v_2|} = \frac{\langle v_1, v_2 \rangle}{|v_1|\,|v_2|} = \cos\theta.$$

That is, the function f maps two curves meeting at an angle θ at z to two curves that meet at the same angle at $f(z)$. A function with this property is said to be *conformal* at z.

Conversely, if f is conformal in a neighborhood of z then (under some additional mild assumptions) it is analytic — we will prove this below after discussing the Cauchy-Riemann equations. Thus the theory of

[1] Note: some people use "analytic" and "holomorphic" with two a priori different definitions that are then proved to be equivalent; I find this needlessly confusing so I may use these two terms interchangeably.

analytic functions contains the theory of planar conformal maps as a special (and largely equivalent) case, although this is by no means obvious from the purely geometric definition of conformality.

Let us briefly review some properties of derivatives.

Lemma 1. *Under appropriate assumptions, we have the relations*

$$\begin{aligned}(f+g)'(z) &= f'(z)+g'(z),\\ (fg)(z) &= f'(z)g(z)+f(z)g'(z),\\ \left(\frac{1}{f}\right)' &= -\frac{f'(z)}{f(z)^2},\\ \left(\frac{f}{g}\right)' &= \frac{f'(z)g(z)-f(z)g'(z)}{g(z)^2},\\ (f\circ g)'(z) &= f'(g(z))g'(z).\end{aligned}$$

Exercise 2. Explain precisely what the assumptions in the lemma are (see Proposition 2.2 on page 10 of [11]).

3.2 The Cauchy-Riemann equations

In addition to the geometric picture associated with the definition of the complex derivative, there is yet another quite different but also extremely useful way to think about analyticity, that provides a bridge between complex analysis and ordinary multivariate calculus. Remembering that complex numbers are vectors that have real and imaginary components, we can denote $z = x+iy$, where x and y will denote the real and imaginary parts of the complex number z, and $f = u+iv$, where u and v are real-valued functions of z (or equivalently of x and y) that return the real and imaginary parts, respectively, of f. Now, if f is analytic at z then

$$\begin{aligned}f'(z) &= \lim_{h\to 0}\frac{f(z+h)-f(z)}{h}\\ &= \lim_{h\to 0,\ h\in\mathbb{R}}\frac{u(x+h+iy)-u(x+iy)}{h}+i\frac{v(x+h+iy)-v(x+iy)}{h}\\ &= \frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x}.\end{aligned}$$

On the other hand also

$$\begin{aligned}f'(z) &= \lim_{h\to 0}\frac{f(z+h)-f(z)}{h}\\ &= \lim_{h\to 0,\ h\in i\mathbb{R}}\frac{u(x+h+iy)-u(x+iy)}{h}+i\frac{v(x+h+iy)-v(x+iy)}{h}\\ &= \lim_{h\to 0,\ h\in\mathbb{R}}\frac{u(x+iy+ih)-u(x+iy)}{ih}+i\frac{v(x+iy+ih)-v(x+iy)}{ih}\\ &= -i\frac{\partial u}{\partial y}-i\cdot i\frac{\partial v}{\partial y}=\frac{\partial v}{\partial y}-i\frac{\partial u}{\partial y}.\end{aligned}$$

Since these limits are equal, by equating their real and imaginary parts we get a famous system of coupled partial differential equations, the *Cauchy-Riemann equations*:

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y},\qquad \frac{\partial v}{\partial x}=-\frac{\partial u}{\partial y}.$$

We have proved that if f is analytic at $z = x + iy$ then the components u, v of f satisfy the Cauchy-Riemann equations. Conversely, we now claim if $f = u + iv$ is continuously differentiable at $z = x + iy$ (in the sense that each of u and v is a continuously differentiable function of x, y as defined in ordinary real analysis) and satisfies the Cauchy-Riemann equations there, f is analytic at z.

Proof. The assumption implies that f has a differential at z, i.e., in the notation of vector calculus, denoting $f = (u, v)$, $z = (x, y)^\top$, $\Delta z = (h_1, h_2)^\top$, we have

$$f(z + \Delta z) = \begin{pmatrix} u(z) \\ v(z) \end{pmatrix} + \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + E(h_1, h_2),$$

where $E(h_1, h_2) = o(|\Delta z|)$ as $|\Delta z| \to 0$. Now, by the assumption that the Cauchy-Riemann equations hold, we also have

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x} h_1 + \frac{\partial u}{\partial y} h_2 \\ -\frac{\partial u}{\partial y} h_1 + \frac{\partial u}{\partial x} h_2 \end{pmatrix},$$

which is the vector calculus notation for the complex number

$$\left(\frac{\partial u}{\partial x} - i \frac{\partial u}{\partial y} \right) (h_1 + i h_2) = \left(\frac{\partial u}{\partial x} - i \frac{\partial u}{\partial y} \right) \Delta z.$$

So, we have shown that (again, in complex analysis notation)

$$\lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} = \lim_{\Delta z \to 0} \left(\frac{\partial u}{\partial x} - i \frac{\partial u}{\partial y} + \frac{E(\Delta z)}{\Delta z} \right) = \frac{\partial u}{\partial x} - i \frac{\partial u}{\partial y}.$$

This proves that f is holomorphic at z with derivative given by $f'(z) = \frac{\partial u}{\partial x} - i \frac{\partial u}{\partial y}$. □

With the help of the Cauchy-Riemann equations, we can now prove our earlier claim that conformality implies analyticity.

Theorem 3. *If $f = u + iv$ is conformal at z, continuously differentiable in the real analysis sense, and satisfies $\det J_f > 0$ (i.e., f preserves orientation as a planar map), then f is holomorphic at z.*

Proof. In the notation of the proof above, we have as before that

$$f(z + \Delta z) = \begin{pmatrix} u(z) \\ v(z) \end{pmatrix} + \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + E(h_1, h_2),$$

where $E(h_1, h_2) = o(|\Delta z|)$ as $|\Delta z| \to 0$. The assumption is that the differential map

$$J_f = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

preserves orientation and is conformal; the conclusion is that the Cauchy-Riemann equations are satisfied (which would imply that f is holomorphic at z by the result shown above). So the theorem will follow once we prove the simple claim about 2×2 matrices contained in Lemma 2 below. □

Lemma 2 (Conformality lemma.). *Assume that* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *is a* 2×2 *real matrix. The following are equivalent:*

(a) *A preserves orientation (that is,* $\det A > 0$*) and is conformal, that is*

$$\frac{\langle Aw_1, Aw_2 \rangle}{|Aw_1|\,|Aw_2|} = \frac{\langle w_1, w_2 \rangle}{|w_1|\,|w_2|}$$

for all $w_1, w_2 \in \mathbb{R}^2$.

(b) *A takes the form* $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ *for some* $a, b \in \mathbb{R}$ *with* $a^2 + b^2 > 0$.

(c) *A takes the form* $A = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ *for some* $r > 0$ *and* $\theta \in \mathbb{R}$. *(That is, geometrically A acts by a rotation followed by a scaling.)*

Proof that (a) $\Longrightarrow$ *(b).* Note that both columns of A are nonzero vectors by the assumption that $\det A > 0$. Now applying the conformality assumption with $w_1 = (1, 0)^\top$, $w_2 = (0, 1)^\top$ yields that $(a, c) \perp (b, d)$, so that $(b, d) = \kappa(-c, a)$ for some $\kappa \in \mathbb{R} \setminus \{0\}$. On the other hand, applying the conformality assumption with $w_1 = (1, 1)^\top$ and $w_2 = (1, -1)^\top$ yields that $(a+b, c+d) \perp (a-b, c-d)$, which is easily seen to be equivalent to $a^2 + c^2 = b^2 + d^2$. Together with the previous relation that implies that $\kappa = \pm 1$. So A is of one of the two forms $\begin{pmatrix} a & -c \\ c & a \end{pmatrix}$ or $\begin{pmatrix} a & c \\ c & -a \end{pmatrix}$. Finally, the assumption that $\det A > 0$ means it is the first of those two possibilities that must occur. ☐

Exercise 3. Complete the proof of the lemma above by showing the implications (b) $\Longleftrightarrow$ (c) and that (b) $\Longrightarrow$ (a).

Another curious consequence of the Cauchy-Riemann equations, which gives an alternative geometric picture to that of conformality, is that analyticity implies the orthogonality of the level curves of u and of v. That is, if $f = u + iv$ is analytic then

$$\langle \nabla u, \nabla v \rangle = (u_x, u_y) \perp (v_x, v_y) = u_x v_x + u_y v_y = v_y v_x - v_x v_y = 0.$$

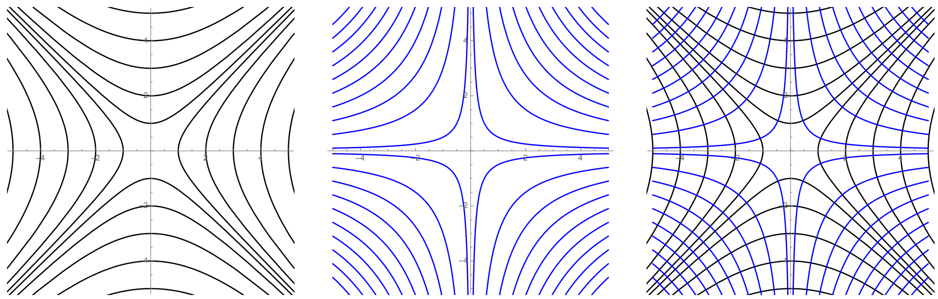
Since ∇u (resp. ∇v) is orthogonal to the level curve $\{u = c\}$ (resp. the level curve $\{v = d\}$, this proves that the level curves $\{u = c\}$, $\{v = d\}$ meet at right angles whenever they intersect.

Yet another important and remarkable consequence of the Cauchy-Riemann equations is that, at least under mild assumptions (which we will see later can be removed) the functions u, v are *harmonic* functions. Assume that f is analytic at z and *twice* continuously differentiable there. Then

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} &= \frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right) \\ &= \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial v}{\partial x}\right) = \frac{\partial^2 v}{\partial x \partial y} - \frac{\partial^2 v}{\partial y \partial x} = 0, \end{aligned}$$

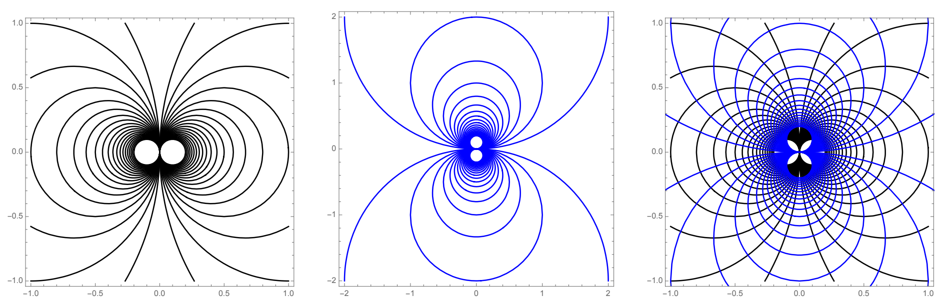
i.e., u satisfies Laplace's equation

$$\triangle u = 0,$$



(a) (b) (c)

Figure 3: The level curves for the (a) real and (b) imaginary parts of $z^2 = (x^2 - y^2) + i(2xy)$. (c) shows the superposition of both families of level curves.



(a) (b) (c)

Figure 4: The level curves for the real and imaginary parts of $z^{-1} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}$.

where $\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the two-dimensional Laplacian operator. Similarly (check), v also satisfies

$$\triangle v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

That is, we have shown that u and v are harmonic functions. This is an extremely important connection between complex analysis and the theory of partial differential equations, which also relates to many other areas of real analysis.

We will later see that the assumption of twice continuous differentiability is unnecessary, but proving this requires some subtle complex-analytic ideas.

A final remark related to analyticity and the Cauchy-Riemann equation is the observation that if $f = u + iv$ is analytic then its Jacobian (in the sense of multivariate calculus when we consider it as a map from $\mathbb{R}^2$ to $\mathbb{R}^2$) is given by

$$J_f = \det \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} = u_x v_y - u_y v_x = u_x^2 + v_x^2 = |u_x + iv_x| = |f'(z)|^2.$$

This can also be understood geometrically. (Exercise: how?)

4 Power series

Until now we have not discussed any specific examples of functions of a complex variable. Of course, there are the standard functions that you probably encountered already in your undergraduate studies: polynomials, rational functions, e^z, the trigonometric functions, etc. But aside from these examples, it would be useful to have a general way to construct a large family of functions. Of course, there is such a way: *power series*, which—non-obviously—turn out to be essentially as general a family of functions as one could hope for.

To make things precise, a power series is a function of a complex variable z that is defined by

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

where $(a_n)_{n=0}^{\infty}$ is a sequence of complex numbers, or more generally by

$$g(z) = f(z - z_0) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

where $(a_n)_{n=0}^{\infty}$ is again a a sequence and z_0 is some fixed complex number. These functions are defined wherever the respective series converge.

For which values of z does this formula make sense? It is not hard to see that it converges absolutely precisely for $0 \le |z| < R$, where the value of R is given by

$$R = \left(\limsup_{n\to\infty} |a_n|^{1/n} \right)^{-1}.$$

R is called the radius of convergence of the power series.

Proof. Assume $0 < R < \infty$ (the edge cases $R = 0$ and $R = \infty$ are left as an exercise). The defining property of R is that for all $\epsilon > 0$, we have that $|a_n| < \left(\frac{1}{R} + \epsilon\right)^n$ if n is large enough, and R is the minimal number with that property. Let $z \in D_R(0)$. Since $|z| < R$, we have $|z| \left(\frac{1}{R} + \epsilon\right) < 1$ for some fixed $\epsilon > 0$ chosen small enough. That implies that for $n > N$ (for some large enough N as a function of ϵ),

$$\sum_{n=N}^{\infty} |a_n z^n| < \sum_{n=N}^{\infty} \left[\left(\frac{1}{R} + \epsilon \right) |z| \right]^n,$$

so the series is dominated by a convergent geometric series, and hence converges.

Conversely, if $|z| > R$, then, $|z| \left(\frac{1}{R} - \epsilon\right) > 1$ for some small enough fixed $\epsilon > 0$. Taking a subsequence $(a_{n_k})_{k=1}^{\infty}$ for which $|a_{n_k}| > \left(\frac{1}{R} - \epsilon\right)^{n_k}$ (guaranteed to exist by the definition of R), we see that

$$\sum_{n=0}^{\infty} |a_n z^n| \ge \sum_{k=1}^{\infty} \left[|z| \left(\frac{1}{R} - \epsilon \right) \right]^{n_k} = \infty,$$

so the power series diverges. □

Exercise 4. Complete the argument in the extreme cases $R = 0, \infty$.

Another important theorem is:

Theorem 4. *Power series are holomorphic functions in the interior of the disc of convergence and can be differentiated termwise.*

Proof. Denote

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} a_n z^n = S_N(z) + E_N(z), \\ S_N(z) &= \sum_{n=0}^{N} a_n z^n, \\ E_N(z) &= \sum_{n=N+1}^{\infty} a_n z^n, \\ g(z) &= \sum_{n=1}^{\infty} n a_n z^{n-1}. \end{aligned}$$

The claim is that f is differentiable on the disc of convergence and its derivative is the power series g. Since $n^{1/n} \to 1$ as $n \to \infty$, it is easy to see that $f(z)$ and $g(z)$ have the same radius of convergence. Fix z_0 with $|z| < r < R$. We wish to show that $\frac{f(z_0+h)-f(z_0)}{h}$ converges to $g(z_0)$ as $h \to 0$. Observe that

$$\begin{aligned} \frac{f(z_0+h)-f(z_0)}{h} - g(z_0) = &\left(\frac{S_N(z_0+h)-S_N(z_0)}{h} - S_N'(z_0) \right) \\ &+ \frac{E_N(z_0+h)-E_N(z_0)}{h} + (S_N'(z_0) - g(z_0)) \end{aligned}$$

The first term converges to 0 as $h \to 0$ for any fixed N. To bound the second term, fix some $\epsilon > 0$, and note that, if we assume that not only $|z_0| < r$ but also $|z_0 + h| < r$ (an assumption that's clearly satisfied for h close enough to 0) then

$$\begin{aligned} \left| \frac{E_N(z_0+h)-E_N(z_0)}{h} \right| &\le \sum_{n=N+1}^{\infty} |a_n| \left| \frac{(z_0+h)^n - z_0^n}{h} \right| \\ &= \sum_{n=N+1}^{\infty} |a_n| \left| \frac{h \sum_{k=0}^{n-1} h^k (z_0+h)^{n-1-k}}{h} \right| \\ &\le \sum_{n=N+1}^{\infty} |a_n| n r^{n-1}, \end{aligned}$$

where we use the algebraic identity

$$a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \ldots + ab^{n-2} + b^{n-1}).$$

The last expression in this chain of inequalities is the tail of an absolutely convergent series, so can be made $< \epsilon$ be taking N large enough (before taking the limit as $h \to 0$).

Third, when choosing N also make sure it is chosen so that $|S_N'(z_0) - g(z_0)| < \epsilon$, which of course is possible since $S_N'(z_0) \to g(z_0)$ as $N \to \infty$. Finally, having thus chosen N, we get that

$$\limsup_{h \to 0} \left| \frac{f(z_0+h)-f(z_0)}{h} - g(z_0) \right| \le 0 + \epsilon + \epsilon = 2\epsilon.$$

Since ϵ was an arbitrary positive number, this shows that $\frac{f(z_0+h)-f(z_0)}{h} \to g(z_0)$ as $h \to 0$, as claimed. $\square$

The proof above can be thought of as a special case of the following more conceptual result: if g_n is a sequence of holomorphic functions on a region Ω, and $g_n \to g$ *uniformly on closed discs* in Ω, $g_n' \to h$ uniformly on closed discs on Ω, and h is continuous, then g is holomorphic and $g' = h$ on Ω. (Exercise: prove this, and explain the connection to the previous result.)

Corollary 1. *Analytic functions defined as power series are (complex-)differentiable infinitely many times in the disc of convergence.*

Corollary 2. *For a power series $g(z) = \sum_{n=0}^\infty a_n(z-z_0)^n$ with a positive radius of convergence, we have*

$$a_n = \frac{g^{(n)}(z_0)}{n!}.$$

In other words $g(z)$ satisfies Taylor's formula

$$g(z) = \sum_{n=0}^\infty \frac{g^{(n)}(z_0)}{n!}(z-z_0)^n.$$

5 Contour integrals

We now introduce *contour integrals*, which are another fundamental building block of the theory.

Contour integrals, like many other types of integrals, take as input a function to be integrated and a "thing" (or "place") over which the function is integrated. In the case of contour integrals, the "thing" is a *contour*, which is (for our current purposes at least) a kind of planar curve. We start by developing some terminology to discuss such objects. First, there is the notion of a *parametrized curve*, which is simply a continuous function $\gamma : [a,b] \to \mathbb{C}$. The value $\gamma(a)$ is called the starting point and $\gamma(b)$ is called the ending point. Two curves $\gamma_1 : [a,b] \to \mathbb{C}$, $\gamma_2 : [c,d] \to \mathbb{C}$ are called *equivalent*, which is denoted $\gamma_1 \sim \gamma_2$, if $\gamma_2(t) = \gamma_1(I(t))$ where $I : [c,d] \to [a,b]$ is a continuous, one-to-one, onto, increasing function. A "curve" is an equivalence class of parametrized curves with respect to this equivalence relation.

In practice, we will usually refer to parametrized curves as "curves", which is the usual abuse of terminology that one sees in various places in mathematics, in which one blurs the distinction between equivalence classes and their members, remembering that various definitions, notation, and proof arguments need to "respect the equivalence" in the sense that they do not depend of the choice of member. (Meta-exercise: think of 2–3 other examples of this phenomenon.)

For our present context of developing the theory of complex analysis, we shall assume all our curves are piecewise continuously differentiable. More generally, one can assume them to be rectifiable, but we will not bother to develop that theory. There are yet more general contexts in which allowing curves to be merely continuous is beneficial (and indeed some of the ideas we will develop in a complex-analytic context can be carried over to that more general setting), but we will not pursue such distractions either.

You probably encountered curves and parametrized curves in your earlier studies of multivariate calculus, where they were used to define the notion of *line integrals* of vector and scalar fields. Recall that there are two types of line integrals, which are referred to as line integrals of the first and second kind. The line integral of the first kind of a scalar (usually real-valued) function $u(z)$ over a curve γ is defined as

$$\int_\gamma u(z)ds = \lim_{\max_j \Delta s_j \to 0} \sum_{j=1}^n u(z_j)\Delta s_j \qquad \text{(line integral of the first kind)},$$

where the limit is a limit of Riemann sums with respect to a family of partitions of the interval $[a, b]$ over which the curve γ is defined, as the norm of the partitions shrinks to 0. Here the partition points are $a = t_0 < t_1 < \ldots < t_n = b$, the points $z_j = f(t_j)$ are their images on the curve γ, and the symbols Δs_j refer to finite line elements, namely $\Delta s_j = |z_j - z_{j-1}|$.

The line integral of the second kind is defined for a vector field $\mathbf{F} = (P, Q)$ (the more traditional notation from calculus for what we would denote in the current context as the complex-valued function $F = P + iQ$) by

$$\int_\gamma \mathbf{F} \cdot d\mathbf{s} = \int_\gamma P\,dx + Q\,dy = \lim_{\max_j \Delta s_j \to 0} \sum_{j=1}^n P(z_j)\Delta x_j + Q(z_j)\Delta y_j,$$

where z_j are as before and $x_j = \mathrm{Re}(z_j)$, $y_j = \mathrm{Im}(z_j)$.

It is well-known from calculus that line integrals can be expressed in terms of ordinary (single-variable) Riemann integrals. Take a couple of minutes to remind yourself of why the following formulas are true (assuming all the functions involved are piecewise continuously differentiable):

$$\int_\gamma u(z)\,ds = \int_a^b u(\gamma(t))|\gamma'(t)|\,dt,$$

$$\int_\gamma \mathbf{F} \cdot d\mathbf{s} = \int_a^b \mathbf{F}(\gamma(\mathbf{t})) \cdot \gamma'(t)\,dt.$$

As a further reminder, the basic result known as the *fundamental theorem of calculus for line integrals* states that if $F = \nabla u$ then

$$\int_\gamma \mathbf{F} \cdot d\mathbf{s} = u(\gamma(b)) - u(\gamma(a)).$$

Definition 2 (contour integrals and arc length intervals)**.** For a function $f = u + iv$ of a complex variable z and a curve γ, define

$$\begin{aligned}
\int_\gamma f(z)\,dz &= \text{``}\int_\gamma (u + iv)(dx + idy)\text{''} \\
&= \left(\int_\gamma u\,dx - v\,dy\right) + i\left(\int_\gamma v\,dx + u\,dy\right) \\
&= \int_a^b f(\gamma(t))\gamma'(t)\,dt \qquad \text{(contour integral)}, \\
\int_\gamma f(z)\,|dz| &= \int_\gamma f(z)\,ds = \int_\gamma u\,ds + i\int_\gamma v\,ds \qquad \text{(arc length integral)}.
\end{aligned}$$

If γ is a *closed* curve (the two endpoints are the same, i.e., it satisfies $\gamma(a) = \gamma(b)$), we denote the contour integral as $\oint_\gamma f(z)\,dz$, and similarly $\oint_\gamma f(z)\,|dz|$ for the arc length integral.

A special case of an arc length integral is the length of the curve, defined as the integral of the constant function 1:

$$\mathrm{len}(\gamma) = \int_\gamma |dz| = \int_a^b |\gamma'(t)|\,dt.$$

As mentioned above, our convention of mildly abusing terminology puts on us the burden of having to remeber to check that these definitions do not depend on the parametrization of the curve. Indeed: if $\gamma_1 \sim \gamma_2$ are representatives of the same equivalence class of parametrized curves, that is, $\gamma_2(t) = \gamma_1(I(t))$ for some nicely-behaved function, then using a standard change of variables in single-variable integrals we see that

$$\begin{aligned}\int_{\gamma_2} f(z)\,dz &= \int_c^d f(\gamma_2(t))\gamma_2'(t)dt = \int_c^d f(\gamma_1(I(t)))(\gamma_1 \circ I)'(t)\,dt \\ &= \int_c^d f(\gamma_1(I(t)))\gamma_1'(I(t))I'(t)\,dt = \int_a^b f(\gamma_1(\tau))\gamma_1'(\tau)\,d\tau \\ &= \int_{\gamma_1} f(z)\,dz.\end{aligned}$$

Exercise 5. Show that the integral with respect to arc length similarly does not depend on the parametrization.

Contour integrals have many surprising properties, but the ones on the following list of basic properties are not of the surprising kind:

Proposition 1 (properties of contour integrals)**.** *Contour integrals satisfy the following properties:*

(a) Linearity as an operator on functions: $\int_\gamma (\alpha f(z) + \beta g(z))\,dz = \alpha \int_\gamma f(z)\,dz + \beta \int_\gamma g(z)\,dz.$

(b) Linearity as an operator on curves: if a contour Γ *is a "composition" of two contours* γ_1 *and* γ_2 *(in a sense that is easy to define graphically, but tedious to write down precisely), then*

$$\int_\Gamma f(z)\,dz = \int_{\gamma_1} f(z)\,dz + \int_{\gamma_2} f(z)\,dz.$$

Similarly, if γ_2 *is the "reverse" contour of* γ_1*, then*

$$\int_{\gamma_2} f(z)\,dz = -\int_{\gamma_1} f(z)\,dz.$$

(c) Triangle inequality:

$$\left|\int_\gamma f(z)\,dz\right| \le \int |f(z)|\,|dz| \le \operatorname{len}(\gamma)\cdot \sup_{z\in\gamma} |f(z)|.$$

Exercise 6. Prove this result (part of the exercise is to define precisely the notions of "composition of curves" and "reverse curve").

Contour integrals have their own version of the fundamental theorem of calculus.

Theorem 5 (The fundamental theorem of calculus for contour integrals.)**.** *If* γ *is a curve connecting two points* w_1, w_2 *in a region* Ω *on which a function* F *is holomorphic, then*

$$\int_\gamma F'(z)\,dz = F(w_2) - F(w_1).$$

Equivalently, the theorem says that to compute a general contour integral $\int_\gamma f(z)\,dz$, we try to find a *primitive* (a.k.a. *anti-derivative*) of f, that is, a function F such that $F'(z) = f(z)$ on all of Ω. If we found such a primitive then the contour integral $\int_\gamma f(z)\,dz$ is given by $F(w_2) - F(w_1)$.

Proof. For smooth curves, an easy application of the chain rule gives

$$\int_\gamma F'(z)\,dz = \int_a^b F'(\gamma(t))\gamma'(t)\,dt = \int_a^b (F\circ\gamma)'(t)\,dt = (F\circ\gamma)(t)|_{t=a}^{t=b}$$
$$= F(\gamma(b)) - F(\gamma(a)) = F(w_2) - F(w_1).$$

For piecewise smooth curves, this is a trivial extension that is left as an exercise. ☐

Many of our discussions of contour integrals will involve the behavior of integrals over *closed* contours, and the interplay between the properties of such integrals and integrals over general contours. As an example of this interplay, the above result has an easy — but important — consequence for integrals over closed contours.

Corollary 3. *If $f = F'$ where F is holomorphic on a region Ω (in that case we say that f* ***has a primitive****), γ is a closed curve in Ω, then*

$$\oint_\gamma f(z)\,dz = 0.$$

This last result has a partial converse:

Proposition 2. *if $f : \Omega \to \mathbb{C}$ is a continuous function on a region Ω such that*

$$\oint_\gamma f(z)\,dz = 0$$

holds for any closed contour in Ω, then f has a primitive.

Proof. Fix some $z_0 \in \Omega$. For any $z \in \Omega$, there is some path $\gamma(z_0, z)$ connecting z_0 and z (since Ω is connected and open, hence pathwise-connected — a standard exercise in topology, see the exercises in Chapter 1 of [11]). Define

$$F(z) = \int_{\gamma(z_0,z)} f(w)\,dw.$$

By the assumption, this integral does not depend on which contour $\gamma(z_0, z)$ connecting z_0 and z was chosen, so $F(z)$ is well-defined. We now claim that F is holomorphic and its derivative is equal to f. To see this, note that

$$\begin{aligned} &\frac{F(z+h)-F(z)}{h} - f(z) \\ &\quad= \frac{1}{h}\left(\int_{\gamma(z_0,z+h)} f(w)\,dw - \int_{\gamma(z_0,z)} f(w)\,dw\right) - f(z) \\ &\quad= \frac{1}{h}\int_{\gamma(z,z+h)} f(w)\,dw - f(z) = \frac{1}{h}\int_{\gamma(z,z+h)} (f(w)-f(z))\,dw \end{aligned}$$

where $\gamma(z, z+h)$ denotes a contour connecting z and $z+h$. When $|h|$ is sufficiently small so that the disc $D(z, h)$ is contained in Ω, one can take $\gamma(z, z+h)$ as the straight line segment connecting z and $z+h$. For such h we get that

$$\left|\frac{F(z+h)-F(z)}{h} - f(z)\right| \le \frac{1}{h}\operatorname{len}(\gamma(z,z+h)) \sup_{w\in D(z,h)} |f(w)-f(z)|$$
$$= \sup_{w\in D(z,h)} |f(w)-f(z)| \xrightarrow[h\to 0]{} 0,$$

by continuity of f. ☐

Remark 3. Note that with the last result, if we knew that holomorphic functions are differentiable infinitely many times (the so-called regularity theorem), we could conclude that a function that satisfies the assumption that all its contour integrals on closed contours were 0 is holomorphic. This is in fact true, and is called **Morera's theorem** (and is an important fact in complex analysis), but we won't be able to prove it until we've proved Cauchy's theorem.

Example 2. Compute $\oint_{|z|=1} z^n\,dz$ for $n \in \mathbb{Z}$. What do we learn from the fact that the integral is not zero for $n = -1$? (Hint: something; but what?) And what do we learn from the fact that it's 0 when $n \neq -1$? (Hint: nothing; but why?)

Lemma 3. *If f is holomorphic on Ω and $f' \equiv 0$ then f is a constant.*

Proof. Fix some $z_0 \in \Omega$. For any $z \in \Omega$, as we discussed above there is a path $\gamma(z_0, z)$ connecting z_0 and z. Then

$$f(z) - f(z_0) = \int_{\gamma(z_0,z)} f'(w)\,dw = 0,$$

hence $f(z) \equiv f(z_0)$, so f is constant. □

6 Cauchy's theorem

One of the central results in complex analysis is *Cauchy's theorem.*

Theorem 6 (Cauchy's theorem.)**.** *If f is holomorphic on a* simply-connected *region Ω, then for any closed curve in Ω we have*

$$\oint_\gamma f(z)\,dz = 0.$$

The challenges facing us are: first, to prove Cauchy's theorem for curves and regions that are relatively simple (where we do not have to deal with subtle topological considerations); second, to define what simply-connected means; third, which will take a bit longer and we won't do immediately, to extend the theorem to the most general setting.

Two other theorems that are closely related to Cauchy's theorem are *Goursat's theorem*, a relatively easy special case of Cauchy's theorem, and *Morera's theorem* which is a kind of converse to Cauchy's theorem.

Theorem 7 (Goursat's theorem)**.** *If f is holomorphic on a region Ω, and T is a triangle contained in Ω, then $\oint_{\partial T} f(z)\,dz = 0$ (where T refers to the "full" triangle, and ∂T refers to its boundary considered as a curve oriented in the usual positive direction).*

Theorem 8 (Morera's theorem)**.** *If $f : \Omega \to \mathbb{C}$ is a continuous function on a region Ω such that*

$$\oint_\gamma f(z)\,dz = 0$$

holds for any closed contour in Ω, then f is holomorphic on Ω.

Proof of Goursat's theorem. The proof can be summarized with a slogan: "localize the damage." Namely, try to translate a global statement about the integral around the triangle to a local statement about behavior near a specific point inside the triangle, which would become manageable since we have a good understanding of the local behavior of a holomorphic function near a point. If something goes wrong with the global integral, something has to go wrong at the local level, and we will show that can't happen (although technically the proof is not a proof by contradiction, conceptually I find this a helpful way to think about it).

The idea can be made more precise using *triangle subdivision*. Specifically, let $T^{(0)} = T$, and define a hierarchy of subdivided triangles

$$\begin{aligned}
\text{order 0 triangle:} \quad & T^{(0)}, \\
\text{order 1 triangles:} \quad & T_j^{(1)}, 1 \le j \le 4, \\
\text{order 2 triangles:} \quad & T_{j,k}^{(2)}, 1 \le j, k \le 4 \\
\text{order 3 triangles:} \quad & T_{j,k,\ell}^{(3)}, 1 \le j, k, \ell \le 4, \\
\vdots \quad & \\
\text{order } n \text{ triangles:} \quad & T_{j_1,\ldots,j_n}^{(n)}, 1 \le j_1, \ldots, j_n \le 4. \\
\vdots \quad &
\end{aligned}$$

Here, the triangles $T_{j_1,\ldots,j_n}^{(n)}$ for $j_n = 1, 2, 3, 4$ are obtained by subdividing the order $n-1$ triangle $T_{j_1,\ldots,j_{n-1}}^{(n-1)}$ into 4 subtriangles whose vertices are the vertices and/or edge bisectors of $T_{j_1,\ldots,j_{n-1}}^{(n-1)}$ (see Figure 1 on page 35 of [11]).

Now, given the way this subdivision was done, it is clear that we have the equality

$$\oint_{\partial T_{j_1,\ldots,j_{n-1}}^{(n-1)}} f(z)\,dz = \sum_{j_n=1}^{4} \oint_{\partial T_{j_1,\ldots,j_n}^{(n)}} f(z)\,dz$$

due to cancellation along the internal edges, and hence

$$\oint_{\partial T^{(0)}} f(z)\,dz = \sum_{j_1,\ldots,j_n=1}^{4} \oint_{\partial T_{j_1,\ldots,j_n}^{(n)}} f(z)\,dz.$$

That is, the integral along the boundary of the original triangle is equal to the sum of the integrals over all 4^n triangles of order n. Now, the crucial observation is that *one of these integrals has to have a modulus that is at least as big as the average.* That is, we have

$$\left| \oint_{\partial T^{(0)}} f(z)\,dz \right| \le \sum_{j_1,\ldots,j_n=1}^{4} \left| \oint_{\partial T_{j_1,\ldots,j_n}^{(n)}} f(z)\,dz \right| \le 4^n \left| \oint_{\partial T_{\mathbf{j}(n)}^{(n)}} f(z)\,dz \right|$$

where $\mathbf{j}(n) = (j_1^{(n)}, \ldots, j_n^{(n)})$ is some n-tuple chosen such that the second inequality holds. Moreover, we can choose $\mathbf{j}(n)$ inductively in such a way that the triangles $T_{\mathbf{j}(n)}^{(n)}$ are nested — that is, $T_{\mathbf{j}(n)}^{(n)} \subset T_{\mathbf{j}(n-1)}^{(n-1)}$ for $n \ge 1$, or equivalently $\mathbf{j}(n) = (j_1^{(n-1)}, \ldots, j_{n-1}^{(n-1)}, k)$ for some $1 \le k \le 4$ — to make this happen, choose k to be such that $\left| \oint_{\partial T_{(\mathbf{j}(n-1),k)}^{(n)}} f(z)\,dz \right|$ is greater than (or equal to) the average

$$\frac{1}{4} \sum_{d=1}^{4} \left| \oint_{\partial T_{(\mathbf{j}(n-1),d)}^{(n)}} f(z)\,dz \right|,$$

which in turn is (by induction) greater than or equal to

$$\left|\frac{1}{4}\sum_{d=1}^{4}\oint_{\partial T^{(n)}_{(\mathbf{j}(n-1),d)}} f(z)\,dz\right| = \left|\oint_{\partial T^{(n-1)}_{\mathbf{j}(n-1)}} f(z)\,dz\right| \geq 4^{-(n-1)}\oint_{\partial T} f(z)\,dz.$$

Now observe that the sequence of nested triangles shrinks to a single point. That is, we have

$$\bigcap_{n=0}^{\infty} T^{(n)}_{\mathbf{j}(n)} = \{z_0\}$$

for some point $z_0 \in T$. This is true because the diameter of the triangles goes to 0 as $n \to \infty$, so certainly there can't be two distinct points in the intersection; whereas, on the other hand, the intersection cannot be empty, since the sequence $(z_n)_{n=0}^{\infty}$ of centers (in some obvious sense, e.g., intersection of the angle bisectors) of each of the triangles is easily seen to be a Cauchy sequence (and hence a convergent sequence, by the completeness property of the complex numbers), whose limit must be an element of the intersection.

Having defined z_0, write $f(z)$ for z near z_0 as

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \psi(z)(z - z_0),$$

where

$$\psi(z) = \frac{f(z) - f(z_0)}{z - z_0} - f'(z_0).$$

The holomorphicity of f at z_0 implies that $\psi(z) \to 0$ as $z \to z_0$. Denote by $d^{(n)}$ the diameter of $T^{(n)}_{\mathbf{j}(n)}$ and by $p^{(n)}$ its perimeter. Each subdivision shrinks both the diameter and perimeter by a factor of 2, so we have

$$d^{(n)} = 2^{-n} d^{(0)}, \qquad p^{(n)} = 2^{-n} p^{(0)}.$$

It follows that

$$\begin{aligned}\left|\int_{\partial T^{(n)}_{\mathbf{j}(n)}} f(z)\,dz\right| &= \left|\int_{\partial T^{(n)}_{\mathbf{j}(n)}} f(z_0) + f'(z_0)(z - z_0) + \psi(z)(z - z_0)\,dz\right| \\ &= \left|\int_{\partial T^{(n)}_{\mathbf{j}(n)}} \psi(z)(z - z_0)\,dz\right| \leq p^{(n)} d^{(n)} \sup_{z \in T^{(n)}_{\mathbf{j}(n)}} |\psi(z)|\end{aligned}$$

Finally, combining this with the relationship between $\left|\oint_{\partial T^{(0)}} f(z)\,dz\right|$ and $|\int_{\partial T^{(n)}_{\mathbf{j}(n)}} f(z)\,dz|$, we get that

$$\left|\int_{\partial T^{(0)}} f(z)\,dz\right| \leq p^{(0)} d^{(0)} \sup_{z \in T^{(n)}_{\mathbf{j}(n)}} |\psi(z)| \xrightarrow[n \to \infty]{} 0,$$

which finishes the proof. $\square$

In the next section we will explore some of the amazing consequences of this ultimately quite simple result. But first, let us note a few not-very-amazing consequences..

Corollary 4 (Goursat's theorem for rectangles.)**.** *Theorem 7 is also true when we replace the word "triangle" with "rectangle."*

Proof. Obvious: a rectangle can be decomposed as the union of two triangles, with the contour integral around the rectangle being the sum of the integrals around the two triangles due to cancellation of the integrals going in both directions along the diagonal (see Prop 1(b)). □

Corollary 5 (existence of a primitive for a holomorphic function on a disc)**.** *If f is holomorphic on a disc D, then $f = F'$ for some holomorphic function F on D.*

Proof. The idea is similar to the proof of Proposition 2 above. If we knew that all contour integrals of f around closed contours vanished, that result would give us what we want. As it is, we know this is true but only for triangular contours. How can we make use of that information? The textbook [11] gives a clever approach in which the contour $\gamma(z_0, z)$ is comprised of a horizontal line segment followed by a vertical line segment. Then one shows in three steps, each involving a use of Goursat's theorem (see Figure 4 on page 38 of [11]), that $F(z_0+h) - F(z_0)$ is precisely the contour integral over the line segment connecting z_0 and $z_0 + h$. From there the theorem proceeds in exactly the same way as before. □

Corollary 6 (Cauchy's theorem for a disc.)**.** *If f is holomorphic on a disc, then $\oint_\gamma f\,dz = 0$ for any closed contour γ in the disc.*

Proof. f has a primitive, and we saw that that implies the claimed consequence. □

Theorem 9 (Cauchy's theorem for a region enclosed by a "toy contour")**.** *The statement $\int_\gamma f(z)dz = 0$ is also true for a function that's analytic in a region enclosed by a contour that is simple enough that the method of proof used for the disc above can be extended to it.*[2]

Proof. Repeat the same ideas, going from Goursat's theorem, to the fact that the function has a primitive, to the fact that its contour integrals along closed curves vanish. The difficulty as the toy contour gets more complicated is to make sure that the geometry works out when proving the existence of the primitive — see for example the (incomplete) discussion of the case of "keyhole contours" on pages 40–41 of [11]. □

7 Consequences of Cauchy's theorem

Theorem 10 (Cauchy's integral formula)**.** *If f is holomorphic on a region Ω, and $C = \partial D$ is a circular contour contained in Ω, then*

$$\frac{1}{2\pi i}\oint_C \frac{f(w)}{w-z}\,dw = \begin{cases} f(z) & \text{if } z \in D, \\ 0 & \text{if } z \in \Omega \setminus \overline{D}, \\ \text{undefined} & \text{if } z \in C. \end{cases}$$

Proof. The case when $z \notin \overline{D}$ is covered by Cauchy's theorem in a disc, since in that case the function $w \mapsto f(w)/(w-z)$ is holomorphic in an open set containing $\overline{D}$. It remains to deal with the case $z \in D$. In this case, denote by z_0 the center of the circle C. The idea is now to consider instead the integral

$$\oint_{\Gamma_{\epsilon,\delta}} F_z(w)\,dw = \oint_{\Gamma_{\epsilon,\delta}} \frac{f(w)}{w-z}\,dw,$$

[2]The book [11] calls such a contour a "*toy contour*", leaving the term as a somewhat vaguely defined meta-concept.

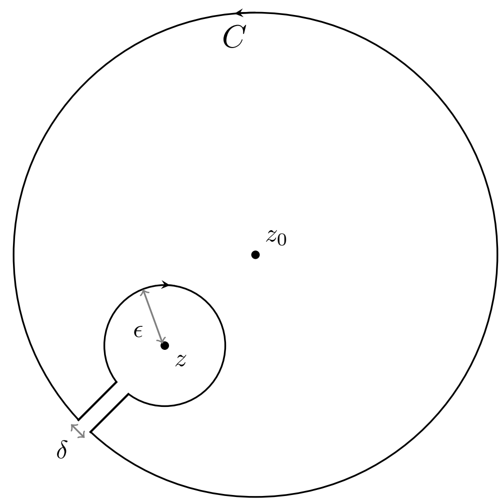


Figure 5: The keyhole contour used in the proof of Cauchy's integral formula.

where $\Gamma_{\epsilon,\delta}$ is a so-called *keyhole contour*, namely a contour comprised of a large circular arc around z_0 that is a subset of the circle C, and another smaller circular arc of radius ϵ centered at z, with two straight line segments connecting the two circular arcs to form a closed curve, such that the width of the "neck" of the keyhole is δ (think of δ as being much smaller than ϵ); see Fig. 5. Note that the function $F_z(w)$ is holomorphic inside the region enclosed by $\Gamma_{\epsilon,\delta}$, so Cauchy's theorem for toy contours (assuming you can be convinced that the keyhole contour is a toy contour) gives that

$$\oint_{\Gamma_{\epsilon,\delta}} F_z(w)\,dw = 0.$$

As $\delta \to 0$, the two parts of the integral along the "neck" of the contour $\Gamma_{\epsilon,\delta}$ cancel out in the limit because F_z is continuous, and hence uniformly continuous, on the compact set $\overline{D} \setminus D(z,\epsilon)$. So we can conclude that

$$\oint_C F_z(w)\,dw = \oint_{|w-z|=\epsilon} F_z(w)\,dw.$$

The next, and final, step, is to take the limit as $\epsilon \to 0$ of the right-hand side of this equation, after first decomposing $F_z(w)$ as

$$F_z(w) = \frac{f(w)-f(z)}{w-z} + f(z)\cdot\frac{1}{w-z},$$

Integrating each term separately, we have for the first term

$$\begin{aligned}\left|\oint_C \frac{f(w)-f(z)}{w-z}\,dw\right| &\le 2\pi\epsilon \cdot \sup_{|w-z|=\epsilon} \frac{|f(w)-f(z)|}{\epsilon} \\ &= 2\pi \sup_{|w-z|=\epsilon} |f(w)-f(z)| \xrightarrow[\epsilon\to 0]{} 0,\end{aligned}$$

by continuity of f; and for the second term,

$$\oint_{|w-z|=\epsilon} f(z)\cdot\frac{1}{w-z}\,dw = f(z)\oint_{|w-z|=\epsilon}\frac{1}{w-z}\,dw = 2\pi i f(z)$$

(by a standard calculation, see Example 2 above). Putting everything together gives $\oint_C \frac{1}{2\pi i} F_z(w)\,dw = f(z)$, which was the formula to be proved. $\square$

Example 3. in the case when z is the *center* of the circle $C = \{w : |w-z| = r\}$, Cauchy's formula gives that

$$f(z) = \frac{1}{2\pi} \oint_{|w-z|=r} f(w)\,\frac{dw}{i(w-z)} = \frac{1}{2\pi}\int_0^{2\pi} f(z+re^{it})dt.$$

In other words, we have proved:

Theorem 11 (the mean value property for holomorphic functions)**.** *The value of a holomorphic function f at z is equal to the average of its values around a circle $|w-z| = r$ (assuming it is holomorphic on an open set containing the disc $|w-z| \le r$).*

Considering what the mean value property means for the real and imaginary parts of $f = u + iv$, which are harmonic functions, we see that they in turn also satisfy a similar mean value property:

$$u(x,y) = \frac{1}{2\pi}\int_0^{2\pi} u(x + r\cos t, y + r\sin t)\,dt.$$

This is in fact true for all harmonic functions — a fact, known as the *mean value property for harmonic functions*, that can be proved separately using PDE/real analysis methods, or derived from the above considerations by proving that every harmonic function in a disc is the real part of a holomorphic function.

Theorem 12 (Cauchy's integral formula, extended version)**.** *Under the same assumptions as in Theorem 10, f is infinitely differentiable, and for $z \in D$ its derivatives $f^{(n)}(z)$ are given by*

$$f^{(n)}(z) = \frac{n!}{2\pi i}\oint_C \frac{f(w)}{(w-z)^{n+1}}\,dw.$$

The fact that holomorphic functions are differentiable infinitely many times is referred to by [11] as the *regularity theorem.*

Proof. The key observation is that the expression on the right-hand side of Cauchy's integral formula for $f(z)$ (which is the case $n = 0$ of the "extended" version) can be differentiated under the integral sign. To make this precise, let $n \ge 1$, and assume inductively that we proved

$$f^{(n-1)}(z) = \frac{(n-1)!}{2\pi i}\oint_C \frac{f(w)}{(w-z)^n}\,dw.$$

Then

$$\begin{aligned}&\frac{f^{(n-1)}(z+h) - f^{(n-1)}(z)}{h} \\ &\qquad = \frac{(n-1)!}{2\pi i}\oint_C f(w)\cdot\frac{1}{h}\left(\frac{1}{(w-z-h)^n} - \frac{1}{(w-z)^n}\right)dw.\end{aligned}$$

It is easily seen that as $h \to 0$, the divided difference $\frac{(w-z-h)^{-n} - (w-z)^{-n}}{h}$ converges to $n(w-z)^{-n-1}$, *uniformly over $w \in C$*. (The same claim without the uniformity is just the rule for differentiation of a power function; to get the uniformity one needs to "go back to basics" and repeat the elementary algebraic calculation that was originally used to derive this power rule — an illustration of the idea that in mathematics it is important not just to understand results but also the techniques used to derive them.) It follows that we can go to the limit $h \to 0$ in the above integral representation, to get

$$f^{(n)}(z) = \frac{(n-1)!}{2\pi i} \oint_C f(w) n (w-z)^{-n-1}\, dz,$$

which is precisely the nth case of Cauchy's integral formula. □

Proof of Morera's theorem. We already proved that if f is a function all of whose contour integrals over closed curves vanish, then f has a primitive F. The regularity property now implies that the derivative $F' = f$ is also holomorphic, hence f is holomorphic, which was the claim of Morera's theorem. □

As another immediate corollary to Cauchy's integral formula, we now get an extremely useful family of inequalities that bounds a function $f(z)$ and its derivatives at some specific point $z \in C$ in terms of the values of the function on the boundary of a circle centered at z.

Theorem 13 (Cauchy inequalities)**.** *For f holomorphic in a region Ω that contains the closed disc $\overline{D_R(z)}$, we have*

$$|f^{(n)}(z)| \le n! R^{-n} \sup_{z \in \partial D_R(z)} |f(z)|$$

(where $\partial D_R(z)$ refers to the circle of radius R around z).

Theorem 14 (Analyticity of holomorphic functions)**.** *If f is holomorphic in a region Ω that contains a closed disc $\overline{D_R(z_0)}$, then f has a power series expansion at z_0*

$$f(z) = \sum_{n=0}^{\infty} a_n (z-z_0)^n,$$

that is convergent for all $z \in D_R(z_0)$, where (of course) $a_n = f^{(n)}(z_0)/n!$.

Proof. The idea is that Cauchy's integral formula gives us a representation of $f(z)$ as a weighted "sum" (=an integral, which is a limit of sums) of functions of the form $z \mapsto (w-z)^{-1}$. Each such function has a power series expansion since it is, more or less, a geometric series, so the sum also has a power series expansion.

To make this precise, write

$$\begin{aligned} \frac{1}{w-z} &= \frac{1}{(w-z_0)-(z-z_0)} = \frac{1}{w-z_0} \cdot \frac{1}{1 - \left(\frac{z-z_0}{w-z_0}\right)} \\ &= \frac{1}{w-z_0} \sum_{n=0}^{\infty} \left(\frac{z-z_0}{w-z_0}\right)^n = \sum_{n=0}^{\infty} (w-z_0)^{-n-1} (z-z_0)^n. \end{aligned}$$

This is a power series in $z - z_0$ that, assuming $w \in C_R(z_0)$, converges absolutely for all z such that $|z-z_0| < R$ (that is, for all $z \in D_R(z_0)$). Moreover the convergence is clearly uniform in $w \in C_R(z_0)$. Since infinite summations that are absolutely and uniformly convergent can be interchanged with integration operations, we then get, using the extended version of Cauchy's integral formula, that

$$\begin{aligned}
f(z) &= \frac{1}{2\pi i}\oint_{C_R(z_0)} \frac{f(w)}{w-z}\,dw \\
&= \frac{1}{2\pi i}\oint_{C_R(z_0)} f(w)\sum_{n=0}^{\infty}(w-z_0)^{-n-1}(z-z_0)^n\,dw \\
&= \sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\oint_{C_R(z_0)} f(w)(w-z_0)^{n-1}\,dw\right)(z-z_0)^n \\
&= \sum_{n=0}^{\infty}\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n,
\end{aligned}$$

which is precisely the expansion we were after. □

Remark 4. In the above proof, if we only knew the simple ($n = 0$) case of Cauchy's integral formula (and in particular didn't know the regularity theorem that follows from the extended case of this formula), we would still conclude from the penultimate expression in the above chain of equalities that $f(z)$ has a power series expansion of the form $\sum_n a_n(z-z_0)^n$, with $a_n = (2\pi i)^{-1}\int_{C_R(z_0)} f(w)(w-z)^{-n-1}$. It would then follow from earlier results we proved that $f(z)$ is differentiable infinitely many times, and that $a_n = f^{(n)}(z_0)/n!$, which would again give the extended version of Cauchy's integral formula.

Theorem 15 (Liouville's theorem)**.** *A bounded entire function is constant.*

Proof. An easy application of the (case $n = 1$ of the) Cauchy inequalities gives upon taking the limit $R \to \infty$ that $f'(z) = 0$ for all z, hence, as we already proved, f must be constant. □

Theorem 16. *If f is holomorphic on a region Ω, and $f = 0$ for z in a set containing a limit point in Ω, then f is identically zero on Ω.*

Proof. If the limit point is $z_0 \in \Omega$, that means there is a sequence $(w_k)_{k=0}^{\infty}$ of points in Ω such that $f(w_k) = 0$ for all n, $w_k \to z$, and $w_k \neq z_0$ for all k. We know that in a neighborhood of z_0, f has a convergent power series expansion. If we assume that f is not identically zero in a neighborhood of z_0, then we can write the power series expansion as

$$\begin{aligned}
f(z) &= \sum_{n=0} a_n(z-z_0)^k = \sum_{n=m}^{\infty} a_n(z-z_0)^{\infty} \\
&= a_m(z-z_0)^m\sum_{n=0}^{\infty}\frac{a_{n+m}}{a_m}(z-z_0)^n = a_m(z-z_0)^m(1+g(z)),
\end{aligned}$$

where m is the smallest index such that $a_m \neq 0$, and where $g(z) = \sum_{n=1}^{\infty}\frac{a_{n+m}}{a_m}(z-z_0)^n$ is a holomorphic function in the neighborhood of z_0 that satisfies $g(z_0) = 0$. It follows that for all k,

$$a_m(w_k-z_0)^m(1+g(w_k)) = f(w_k) = 0,$$

but for large enough k this is impossible, since $w_k - z_0 \neq 0$ for all k and $g(w_k) \to g(z_0) = 0$ as $k \to \infty$.

The conclusion is that f is identically zero at least in a neighborhood of z_0. But now we claim that that also implies that f is identically zero on all of Ω, because Ω is a region (open and connected). More precisely, denote by U the set of points $z \in \Omega$ such that f is equal to 0 in a neighborhood of z. It is obvious that U is open, hence also relatively open in Ω since Ω itself is open; U is also closed, by the argument above; and U is nonempty (it contains z_0, again by what we showed above). It follows that $U = \Omega$ by the well-known characterization of a connected topological space as a topological space that has no "'clopen" (closed and open) sets other than the empty set and the entire space.

An alternative way to finish the proof is the following. For every point $z \in \Omega$, let $r(z)$ be the radius of convergence of the power series expansion of f around z. Thus the discs $\{D_{r(z)}(z) : z \in \Omega\}$ form an open covering of Ω. Take $w \in \Omega$ (with z_0 being as above), and take a path $\gamma : [a, b] \to \Omega$ connecting z_0 and w (it exists because Ω is open and connected, hence pathwise-connected). The open covering of Ω by discs is also an open covering of the compact set $\gamma[a, b]$ (the range of γ). By the Heine-Borel property of compact sets, it has a finite subcovering $\{D_{r(z_j)}(z_j) : j = 0, \ldots, m\}$ (where we take $w = z_{m+1}$. The proof above shows that f is identically zero on $D_{r(z_0)}(z_0)$, and also shows that if we know f is zero on $D_{r(z_j)}(z_j)$ then we can conclude that it is zero on the next disc $D_{r(z_{j+1})}(z_{j+1})$. It follows that we can get all the way to the last disc $D_{r(w)}(w)$. In particular, $f(w) = 0$, as claimed. □

Remark 5. The above result is also sometimes described under the heading **zeros of holomorphic functions are isolated**, since it can be formulated as the following statement: if f is holomorphic on Ω, is not identically zero on Ω, and $f(z_0) = 0$ for $z_0 \in \Omega$, then for some $\epsilon > 0$, the punctured neighborhood $D_\epsilon(z_0) \setminus \{z_0\}$ of z_0 contains no zeros of f. In other words, the set of zeros of f contains only isolated points.

Remark 6. The condition that the limit point z_0 be in Ω is needed. Note that it *is* possible to have a sequence $z_n \to z_0$ of points in Ω such that $f(z_n) = 0$ for all n. For example, consider the function $e^{1/z} - 1$ — it has zeros in every neighborhood of $z_0 = 0$.

Corollary 7. *If f, g are holomorphic on a region Ω, and $f(z) = g(z)$ for z in a set with limit point in Ω (e.g., an open disc, or even a sequence of points z_n converging to some $z \in \Omega$), then $f \equiv g$ everywhere in Ω.*

Proof. Apply the previous result to $f - g$. □

The previous result is usually reformulated slightly as the following conceptually important result:

Theorem 17 (Principle of analytic continuation)**.** *If f is holomorphic on a region Ω, and f_+ is holomorphic on a bigger region $\Omega_+ \supset \Omega$ and satisfies $f(z) = f_+(z)$ for all $z \in \Omega$, then f_+ is the* unique *such extension, in the sense that if $\tilde{f_+}$ is another function with the same properties then $f_+(z) = \tilde{f_+}(z)$ for all $z \in \Omega_+$.*

Example 4. In real analysis, we learn that "formulas" such as

$$1 - 1 + 1 - 1 + 1 - 1 + \ldots = \frac{1}{2},$$

$$1 + 2 + 4 + 8 + 16 + 32 + \ldots = -1$$

don't have any meaning. However, in the context of complex analysis one can in fact make perfect sense of such identities, using the principle of analytic continuation! Do you see how? We will also learn later in the course about additional such amusing identities, the most famous of which being

$$1 + 2 + 3 + 4 + \ldots = -\frac{1}{12},$$

$$1 - 2 + 3 - 4 + \ldots = \frac{1}{4}.$$

Such supposedly "astounding" formulas have attracted a lot of attention in recent years, being the subject of a popular Numberphile video, a New York Times article, a discussion on the popular math blog by Terry Tao, a Wikipedia article, a discussion on Mathematics StackExchange, and more.

A "toy" (but still very interesting) example of analytic continuation is the case of *removable singularities*. A point $z_0 \in \Omega$ is called a removable singularity of a function $f : \Omega \to \mathbb{C} \cup \{\text{undefined}\}$ if f is holomorphic in a punctured neighborhood of Ω, is not holomorphic at z_0, but its value at z_0 can be redefined so as to make it holomorphic at z_0.

Theorem 18 (Riemann's removable singularities theorem)**.** *If f is holomorphic in Ω except at a point $z_0 \in \Omega$ (where it may be undefined, or be defined but not known to be holomorphic or even continuous). Assume that f is bounded in a punctured neighborhood $D_r(z_0) \setminus \{z_0\}$ of z_0. Then f can be extended to a holomorphic function $\tilde{f}$ on all of Ω by defining (or redefining) its value at z_0 appropriately.*

Proof. Fix some disc $D = D_R(z_0)$ around z_0 whose closure is contained in Ω. The idea is to prove that the Cauchy integral representation formula

$$f(z) = \frac{1}{2\pi i} \oint_{C_R(z_0)} \frac{f(w)}{w-z}\, dw =: \tilde{f}(z)$$

is satisfied for all $z \in D \setminus \{z_0\}$. Once we show this, we will set $\tilde{f}(z_0)$ to be defined by the same integral representation, and it will be easy to see that that gives the desired extension.

To prove that the representation above holds, consider a "double keyhole" contour $\Gamma_{\epsilon,\delta}$ that surrounds most of circle $C = \partial D$ but makes diversions to avoid the points z_0 and z, circling them in the negative direction around most of a circle of radius ϵ. After applying a limiting argument similar to the one used in the proof of Cauchy's integral formula, we get that

$$\frac{1}{2\pi i} \oint_C \frac{f(w)}{w-z} = \frac{1}{2\pi i} \oint_{C_\epsilon(z)} \frac{f(w)}{w-z} + \frac{1}{2\pi i} \oint_{C_\epsilon(z_0)} \frac{f(w)}{w-z}.$$

On the right-hand side, the first term is $f(z)$ by Cauchy's integral formula (since f is known to be holomorphic on an open set containing $\overline{D_\epsilon(z)}$). The second term can be bounded in magnitude using the assumption that f is bounded in a neighborhood of z_0; more precisely, we have

$$\left| \oint_{C_\epsilon(z_0)} \frac{f(w)}{w-z} \right| \le 2\pi\epsilon \sup_{w \in C_\epsilon(z_0)} |f(w)| \cdot \frac{1}{|z-z_0| - \epsilon} \xrightarrow[\epsilon \to 0]{} 0.$$

Thus by taking the limit as $\epsilon \to 0$ we obtain precisely the desired representation for f.

Finally, once we have the integral representation $\tilde{f}$ (defined only in terms of the values of $f(w)$ for $w \in C_R(z_0)$), the fact that this defines a holomorphic function on all of $D_R(z_0)$ is easy to see, and is something we implicitly were aware of already. For example, the relevant argument (involving a direct manipulation of the divided differences $\frac{1}{h}(\tilde{f}(z+h) - \tilde{f}(z))$) appeared in the proof of the extended version of Cauchy's integral formula. Another approach is to show that integrating $\tilde{f}$ over closed contours gives 0 (which requires interchanging the order of two integration operations, which will not be hard to justify) and then use Morera's theorem. The details are left as an exercise. ∎

Definition 3 (Uniform convergence on compact subsets)**.** If $f, (f_n)_{n=0}^\infty$ are holomorphic functions on a region Ω, we say that the sequence f_n converges to f **uniformly on compact subsets** if for any compact set $K \subset \Omega$, $f_n(z) \to f(z)$ uniformly on K.

Theorem 19. *If $f_n \to f$ uniformly on compact subsets in Ω and f_n are holomorphic, then f is holomorphic, and $f_n' \to f'$ uniformly on compact subsets in Ω.*

Proof. The fact that f is holomorphic is an easy consequence of a combination of Cauchy's theorem and Morera's theorem. More precisely, note that for each closed disc $\overline{D} = \overline{D_r(z_0)} \subset \Omega$ we have $f_n(z) \to f$ uniformly on $\overline{D}$. In particular, for each curve γ whose image is contained in the open disc $D = D_r(z_0)$,

$$\int_\gamma f_n(z)\,dz \xrightarrow[n\to\infty]{} \int_\gamma f(z)\,dz.$$

By Cauchy's theorem, the integrals in this sequence are all 0, so $\int_\gamma f(z)\,dz$ is also zero. Since this is true for all such γ, by Morera's theorem f is holomorphic on D. This was true for any disc in Ω, and holomorphicity is a local property, so in other words f is holomorphic on all of Ω.

Next, let $D = D_r(z_0)$ be a disc whose closure $\overline{D}$ satisfies $\overline{D} \subset \Omega$. for $z \in D$ we have by Cauchy's integral formula that

$$\begin{aligned} f_n'(z) - f'(z) &= \frac{1}{2\pi i}\oint_{\partial D} \frac{f_n(w)}{(w-z)^2}\,dw - \frac{1}{2\pi i}\oint_{\partial D} \frac{f(w)}{(w-z)^2}\,dw \\ &= \frac{1}{2\pi i}\oint_{\partial D} \frac{f_n(w)-f(w)}{(w-z)^2}\,dw. \end{aligned}$$

It is easy to see therefore that $f_n'(z) \to f'(z)$ as $n \to \infty$, *uniformly as z ranges on the disc $D_{r/2}(z_0)$*, since $f_n(w) \to f(w)$ uniformly for $w \in \partial D \subset \overline{D}$, and $|w-z|^{-2} \le (r/2)^{-2}$ for $z \in D_{r/2}(z_0)$, $w \in \partial D$.

Finally, let $K \subset D$ be compact. For each $z \in K$ let $r(z)$ be the radius of a disc $D_{r(z)}(z)$ around z whose closure is contained in Ω. The family of discs $\{D_z = D_{r(z)/2}(z) : z \in \Omega\}$ is an open covering of K, so by the Heine-Borel property of compact sets it has a finite subcovering $D_{z_1}, \ldots, D_{z_n}$. We showed that $f_n'(z) \to f'(z)$ uniformly on every D_{z_j}, so we also have uniform convergence on their union, which contains K, so we get that $f_n' \to f'$ uniformly on K, as claimed. ☐

8 Zeros, poles, and the residue theorem

Definition 4 (zeros)**.** z_0 is a zero of a holomorphic function f if $f(z_0) = 0$.

Definition(+lemma) 5. If f is a holomorphic function on a region Ω that is not identically zero and z_0 is a zero of f, then f can be represented in the form

$$f(z) = (z-z_0)^m g(z)$$

in some neighborhood of z_0, where $m \ge 1$ and g is a holomorphic function in that neighborhood such that $g(z) \ne 0$. The number m is determined uniquely and is called the *order* of the zero z_0, i.e, z_0 will be described as a *zero of order m*.

A zero of order 1 is called a *simple zero.*

Remark 7. In the case when z_0 is *not* a zero of f, the same representation holds with $m = 0$ (and $g = f$), so in certain contexts one may occasionally say that z_0 is a zero of order 0.

Proof of the definition-lemma. Power series expansions: this is similar to the argument used in the proof that zeros of holomorphic functions are isolated. That is, write the power series expansion (known to converge in a neighborhood of z_0)

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} a_n (z-z_0)^n = \sum_{n=m}^{\infty} a_n (z-z_0)^n \\ &= (z-z_0)^m \sum_{n=0}^{\infty} a_{m+n} (z-z_0)^n =: (z-z_0)^m g(z), \end{aligned}$$

where m is the smallest index ≥ 0 such that $a_m \neq 0$. This gives the desired representation. On the other hand, given a representation of this form, expanding $g(z)$ as a power series shows that m has to be the smallest index of a nonzero coefficient in the power series expansion of $f(z)$, which proves the uniqueness claim. □

Definition 6 (poles)**.** If f is defined and holomorphic in a punctured neighborhood of a point z_0, we say that it has a *pole of order* m at z_0 if the function $h(z) = 1/f(z)$ (defined to be 0 at z_0) has a zero of order m at z_0. A pole of order 1 is called a *simple pole*.

Remark 8. As with the case of zeros, one can extend this definition in an obvious way to define a notion of a "pole of order 0". If $f(z)$ is actually holomorphic and nonzero at z_0 (or has a removable singularity at z_0 and can be made holomorphic and nonzero by defining its value at z_0 appropriately), we define the order of the pole as 0 and consider f to have a pole of order 0 at z_0.

Lemma 4. *f has a pole of order m at z_0 if and only if it can be represented in the form*

$$f(z) = (z-z_0)^{-m} g(z)$$

in a punctured neighborhood of z_0, where g is holomorphic in a neighborhood of z_0 and satisfies $g(z_0) \neq 0$.

Proof. Apply the previous lemma to $1/f(z)$. □

Theorem 20. *If f has a pole of order m at z_0, then it can be represented uniquely as*

$$f(z) = \frac{a_{-m}}{(z-z_0)^m} + \frac{a_{-m+1}}{(z-z_0)^{m-1}} + \ldots + \frac{a_{-1}}{z-z_0} + G(z)$$

where G is holomorphic in a neighborhood of z_0.

Proof. This follows immediately on expressing $f(z)$ as $(z-z_0)^{-m} g(z)$ as in the previous lemma and separating the power series expansion of $g(z)$ into the powers $(z-z_0)^k$ with $0 \leq k \leq m-1$ and the powers with $k \geq m$. □

Definition 7 (principal part, residue)**.** The expansion $\frac{a_{-m}}{(z-z_0)^m} + \frac{a_{-m+1}}{(z-z_0)^{m-1}} + \ldots + \frac{a_{-1}}{z-z_0}$ in the above representation is called the *principal part* of f at the pole z_0. The coefficient a_{-1} is called the *residue of f at z_0* and denoted $\mathrm{Res}_{z_0}(f)$.

Exercise 7. The definitions of the order of a zero and a pole can be consistently unified into a single definition of the *(generalized) order of a zero*, where if f has a pole of order m at z_0 then we say instead that f has a zero of order $-m$. Denote the order of a zero of f at z_0 by $\mathrm{ord}_{z_0}(f)$. With these definitions, prove that

$$\mathrm{ord}_{z_0}(f+g) \geq \min\left(\mathrm{ord}_{z_0}(f), \mathrm{ord}_{z_0}(g)\right)$$

(can you give a useful condition when equality holds?), and that

$$\mathrm{ord}_{z_0}(fg) = \mathrm{ord}_{z_0}(f) + \mathrm{ord}_{z_0}(g).$$

Theorem 21 (The residue theorem (simple version))**.** *Assume that f is holomorphic in a region containing a closed disc $\overline{D}$, except for a pole at $z_0 \in D$. Then*

$$\oint_{\partial D} f(z)\,dz = 2\pi i \,\mathrm{Res}_{z_0}(f).$$

Proof. By the standard argument involving a keyhole contour, we see that the circle $C = \partial D$ in the integral can be replaced with a circle $C_\epsilon = C_\epsilon(z_0)$ of a small radius $\epsilon > 0$ around z_0, that is , we have

$$\oint_{\partial D} f(z)\,dz = \oint_{C_\epsilon} f(z)\,dz.$$

When ϵ is small enough, inside C_ϵ we can use the decomposition

$$f(z) = \sum_{k=-m}^{-1} a_k (z-z_0)^k + G(z)$$

for f into its principal part and a remaining holomorphic part. Integrating termwise gives 0 for the integral of $G(z)$, by Cauchy's theorem; 0 for the integral powers $(z-z_0)^k$ with $-m \leq k \leq -2$, by a standard computation; and $2\pi i a_{-1} = 2\pi i \,\mathrm{Res}_{z_0}(f)$ for the integral of $r(z-z_0)^{-1}$, by the same standard computation. This gives the result. □

Theorem 22 (The residue theorem (extended version))**.** *Assume that f is holomorphic in a region containing a closed disc $\overline{D}$, except for a finite number of poles at $z_1, \ldots, z_N \in D$. Then*

$$\oint_{\partial D} f(z)\,dz = 2\pi i \sum_{k=1}^{N} \mathrm{Res}_{z_k}(f).$$

Proof. The idea is the same, except one now uses a contour with multiple keyholes to deduce after a limiting argument that

$$\oint_{\partial D} f(z)\,dz = \sum_{k=1}^{N} \oint_{C_\epsilon(z_k)} f(z)\,dz$$

for a small enough ϵ, and then proceeds as before.

(Note: The above argument seems slightly dishonest to me, since it relies on the assertion that a multiple keyhole contour with arbitrary many keyholes is a "toy contour"; while this is intuitively plausible, it will be undoubtedly quite difficult to think of, and write, a detailed proof of this argument.) □

Theorem 23 (The residue theorem (version for general toy contours).). *Assume that f is holomorphic in a region containing a toy contour γ (oriented in the positive direction) and the region R_γ enclosed by it, except for poles at the points $z_1, \ldots, z_N \in R_\gamma$. Then*

$$\oint_\gamma f(z)\,dz = 2\pi i \sum_{k=1}^N \mathrm{Res}_{z_k}(f).$$

Proof. Again, construct a multiple keyhole version of the same contour γ (assuming that one can believably argue that the resulting contour is still a toy contour), and then use a limiting argument to conclude that

$$\oint_\gamma f(z)\,dz = \sum_{k=1}^N \oint_{C_\epsilon(z_k)} f(z)\,dz,$$

for a small enough ϵ. Then proceed as before. □

9 Meromorphic functions, holomorphicity at ∞ and the Riemann sphere

Definition 8 (meromorphic function). A meromorphic function on a region Ω is a function $f : \Omega \to \mathbb{C} \cup \{\text{undefined}\}$ such that f is holomorphic except for an isolated set of poles.

Definition 9 (holomorphicity at ∞). Let $U \subset \mathbb{C}$ be an open set containing the complement $\mathbb{C} \setminus \overline{D_R(0)}$ of a closed disc around 0. A function $f : U \to \mathbb{C}$ is **holomorphic at** ∞ if $g(z) = f(1/z)$ (defined on a neighborhood $D_{1/R}(0)$ of 0) has a removable singularity at 0. In that case we define $f(\infty) = g(0)$ (the value that makes g holomorphic at 0).

Definition 10 (order of a zero/pole at ∞). Let $U \subset \mathbb{C}$ be an open set containing the complement $\mathbb{C} \setminus \overline{D_R(0)}$ of a closed disc around 0. We say that a function $f : U \to \mathbb{C}$ has a zero (resp. pole) of order m at ∞ if $g(z) = f(1/z)$ has a zero (resp. pole) at $z = 0$, after appropriately defining the value of g at 0.

Conceptually, it is useful to think of meromorphic functions as holomorphic functions with range in the Riemann sphere $\widehat{\mathbb{C}}$. Let's define what that means.

Definition 11 (Riemann sphere). The Riemann sphere (a.k.a. the *extended complex numbers*) is the set $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, equipped with the following additional structure:

- *Topologically*, we think of $\widehat{\mathbb{C}}$ as the *one-point compactification* of $\mathbb{C}$; that is, we add to $\mathbb{C}$ an additional element ∞ (called "the point at infinity") and say that the neighborhoods of ∞ are the complements of compact sets in $\mathbb{C}$. This turns $\widehat{\mathbb{C}}$ into a topological space in a simple way.
- *Geometrically*, we can identify $\widehat{\mathbb{C}}$ with an actual sphere embedded in $\mathbb{R}^3$, namely

$$S^2 = \left\{ (x,y,z) \in \mathbb{R}^3 \,:\, x^2 + y^2 + \left(z - \tfrac{1}{2}\right)^2 = \tfrac{1}{2} \right\}.$$

  The identification is via *stereographic projection*, given explicitly by the formula

$$(X,Y,Z) \in S^2 \longmapsto \begin{cases} x + iy = \frac{X}{1-Z} + i\frac{Y}{1-Z} & \text{if } (X,Y,Z) \neq (0,0,1), \\ \infty & \text{if } (X,Y,Z) = (0,0,1). \end{cases}$$

See page 88 in [11] for a more detailed explanation. One can check that this geometric identification is a homeomorphism between S^2 (equipped with the obvious topology inherited from $\mathbb{R}^3$) and $\widehat{\mathbb{C}}$ (with the one-point compactification topology defined above).

- *Analytically*, the above definition of what it means for a function on a neighborhood of ∞ to be holomorphic at ∞ provides a way of giving $\widehat{\mathbb{C}}$ the structure of a **Riemann surface** (the simplest case of a manifold with a complex-analytic structure). The details can be found in many textbooks and online resources, and we will not discuss them in this course.

With the above definitions, the concept of a meromorphic function $f : \Omega \to \mathbb{C} \cup \{\text{undefined}\}$ can be seen to coincide with the notion of a holomorphic function $f : \Omega \to \widehat{\mathbb{C}}$ — that is, the underlying concept of the definition is still holomorphicity, but it pertains to functions taking values in $\widehat{\mathbb{C}}$, a different Riemann surface, instead of $\mathbb{C}$.

Definition 12 (classification of singularities)**.** If a function $f : \Omega \to \widehat{\mathbb{C}} \cup \{\text{undefined}\}$ is holomorphic in a punctured neighborhood $D_r(z_0) \setminus \{z_0\}$ of z_0, we say that f has a *singularity* at z_0 if f is not holomorphic at z_0. We classify singularities into three types, two of which we already defined:

- *Removable singularities*: when f can be made holomorphic at z_0 by defining or redefining its value at z_0;
- *poles*;
- any singularity that is not removable or a pole is called an *essential singularity.*

For a function defined on a neighborhood of ∞ that is not holomorphic at ∞, we say that f has a singularity at ∞, and classify the singularity as a removable singularity, a pole, or an essential singularity, according to the type of singularity that $z \mapsto f(1/z)$ has at $z = 0$.

Theorem 24 (Casorati-Weierstrass theorem on essential singularities)**.** *If f is holomorphic in a punctured neighborhood $D_r(z_0) \setminus \{z_0\}$ of z_0 and has an essential singularity at z_0, the image $f(D_r(z_0) \setminus \{z_0\})$ of the punctured neighborhood under f is dense in $\mathbb{C}$.*

Proof. Assume that the closure $\overline{f(D_r(z_0) \setminus \{z_0\})}$ does not contain a point $w \in \mathbb{C}$. Then $g(z) = \frac{1}{f(z)-w}$ is a function that's holomorphic and bounded in $D_r(z_0) \setminus \{z_0\}$. By Riemann's removable singularity theorem, its singularity at z_0 is removable, so we can assume it is holomorphic at z_0 after defining its value there. It then follows that

$$f(z) = w + \frac{1}{g(z)}$$

has either a pole or a removable singularity at z_0, depending on whether $g(z_0) = 0$ or not. $\square$

10 The argument principle

Definition 13. The **logarithmic derivative** of a holomorphic function $f(z)$ is $f'(z)/f(z)$.

Lemma 5. *The logarithmic derivative of a product is the sum of the logarithmic derivatives. That is,*

$$\frac{\left(\prod_{k=1}^n f_k\right)'}{\prod_{k=1}^n f_k} = \sum_{k=1}^n \frac{f_k'(z)}{f_k(z)}.$$

Proof. Show this for $n = 2$ and proceed by induction. $\square$

Theorem 25 (the argument principle)**.** *Assume that f is meromorphic in a region Ω containing a closed disc $\overline{D}$, such that f has no poles on the circle ∂D. Denote its zeros and poles inside D by $z_1, \ldots, z_n$, where z_k is a zero of order $m_k = \operatorname{ord}_{z_k}(f)$ (in the sense mentioned above, where $m_k = m$ is a positive integer if z_k is a zero of order m, and $m_k = -m$ is a negative integer if z_k is a pole of order m). Then*

$$\begin{aligned}\frac{1}{2\pi i}\oint_{\partial D}\frac{f'(z)}{f(z)}\,dz &= \sum_{k=1}^{n} m_k \\ &= \textit{[total number of zeros of } f \textit{ inside } D\textit{]} \\ &\quad - \textit{[total number of poles of } f \textit{ inside } D\textit{]}.\end{aligned} \tag{1}$$

Proof. Define

$$g(z) = \prod_{k=1}^{n} (z - z_k)^{-m_k} f(z).$$

Then $g(z)$ is meromorphic on Ω, has no singularities zeros on ∂D, and inside D it has no poles or zeros, only removable singularities at $z_1, \ldots, z_n$ (so after redefining its values at these points we can assume it is holomorphic on D). It follows that

$$f(z) = \prod_{k=1}^{n} (z - z_k)^{m_k} g(z).$$

Taking the logarithmic derivative of this equation gives that

$$\frac{f'(z)}{f(z)} = \sum_{k=1}^{n} \frac{m_k}{z - z_k} + \frac{g'(z)}{g(z)}.$$

The result now follows by integrating this equation and using the residue theorem (the term $g'(z)/g(z)$ is holomorpic on D so does not contribute anything to the integral). $\square$

Remark 9. By similar reasoning, the theorem also holds when the circle is replaced by a toy contour γ.

The proof above hides, as some slick mathematical proofs have a way of doing, the fact that the formula (1) has a fairly simple intuitive explanation. Start by noting that the integral in the argument principle can be represented as

$$\begin{aligned}\frac{1}{2\pi i}\oint_{\gamma}\frac{f'(z)}{f(z)}\,dz &= \frac{1}{2\pi i}\int_a^b \frac{f'(\gamma(t))\gamma'(t)}{f(\gamma(t)}\,dt = \frac{1}{2\pi i}\int_a^b \frac{(f\circ\gamma)'(t)}{(f\circ\gamma)(t)}\,dt \\ &= \frac{1}{2\pi i}\int_{f\circ\gamma}\frac{1}{w}\,dw,\end{aligned}$$

that is, an integral of dw/w over the contour $f \circ \gamma$ — the image of γ under f. Now note that the differential form dw/w has a special geometric meaning in complex analysis, namely we have

$$\frac{dw}{w} = \text{“}d\,(\log w)\text{”} = \text{“}d\,(\log|w| + i\arg w)\text{”}.$$

We put these expressions in quotes since the logarithm and argument are not single-valued functions so it needs to be explained what such formulas mean. However, at least $\log|w|$ is well-defined for a curve that does not cross 0, so when integrating over the closed curve $f \circ \gamma$, the real part is zero by the fundamental theorem of calculus. The imaginary part (which becomes real after dividing by $2\pi i$) can be interpreted intuitively as the *change in the argument over the curve* — that is, initially at time $t = a$ one fixes a specific value of $\arg w = \arg \gamma(a)$; then as t increases from $t = a$ to $t = b$, one tracks the increase or decrease in the argument as one travels along the curve $\gamma(t)$; if this is done correctly (i.e., in a continuous fashion), at the end the argument must have a well-defined value. Since the curve is closed, the total change in the argument must be an integer multiple of 2π, so the division by $2\pi i$ turns it into an integer.

Of course, this explanation also explains the name "the argument principle," which may sound arbitrary and uninformative when you first hear it.

Connection to winding numbers. What the above reasoning shows is that in general, an integral of the form

$$\frac{1}{2\pi i} \oint_\gamma \frac{f(w)}{w}\, dw$$

over a closed curve γ that does not cross 0 carries the meaning of "the total number of times the curve γ goes around the origin," with the number being positive if the curve goes in the positive direction around the origin; negative if the curve goes in the negative direction around the origin; or zero if there is no net change in the argument. This number is more properly called the **winding number** of f around $w = 0$ (also sometimes referred to as the **index** of the curve around 0), and denoted

$$\operatorname{Ind}_0(f) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z}\, dz.$$

More generally, one can define the winding number at $z = z_0$ as the number of times a curve γ winds around an arbitrary point z_0, which (it is easy to see) will be given by

$$\operatorname{Ind}_{z_0}(f) = \frac{1}{2\pi i} \oint_\gamma \frac{f(z)}{z - z_0}\, dz,$$

assuming that γ does not cross z_0.

Note that winding number is a *topological* concept of planar geometry that can be considered and studied without any reference to complex analysis; indeed, in my opinion that is the correct approach. It is possible, and not especially difficult, to define it in purely topological terms without mentioning contour integrals, and then show that the complex analytic and topological definitions coincide. Try to think what such a definition might look like.

Theorem 26 (Rouché's theorem)**.** *Assume that f, g are holomorphic on a region Ω containing a circle $\gamma = C$ and its interior (or, more generally, a toy contour γ and the region U enclosed by it). If $|f(z)| > |g(z)|$ for all $z \in \gamma$ then f and $f + g$ have the same number of zeros inside the region U.*

Proof. Define $f_t(z) = f(z) + tg(z)$ for $t \in [0, 1]$, and note that $f_0 = f$ and $f_1 = f + g$, and that the condition $|f(z)| > |g(z)|$ on γ implies that f_t has no zeros on γ for any $t \in [0, 1]$. Denote

$$n_t = \frac{1}{2\pi i} \oint_\gamma \frac{f_t'(z)}{f_t(z)}\, dz,$$

which by the argument principle is the number of "generalized zeros" (zeros or poles, counting multiplicities) of f_t in U. In particular, the function $t \mapsto n_t$ is integer-valued. If we also knew that it was continuous, then it would have to be constant (by the easy exercise: any integer-valued continuous function on an interval $[a, b]$ is constant), so in particular we would get the desired conclusion that $n_1 = n_0$.

To prove continuity of n_t, note that the function $g(t, z) = f_t'(z)/f_t(z)$ is continuous, hence also uniformly continuous, on the compact set $[0, 1] \times \gamma$. For $s, t \in [0, 1]$ satisfying $|t - s| < \delta$, we can write

$$|n_t - n_s| \le \frac{1}{2\pi i} \oint_\gamma |g(t, z) - g(s, z)| \cdot |dz|$$
$$\le \frac{1}{2\pi i} \operatorname{len}(\gamma) \sup\{|g(u, z) - g(v, z)| \,:\, z \in \gamma,\ u, v \in [0, 1],\ |u - v| < \delta\}.$$

Given $\epsilon > 0$, we can choose δ that ensures that this expression is $< \epsilon$ if $|t-s| < \delta$, by the uniform continuity. This is precisely what is needed to show that $t \mapsto n_t$ is continuous. □

As with the argument principle, Rouché's theorem also has a rather amusing intuitive explanation, which I learned from the book *Visual Complex Analysis* by Tristan Needham. The slogan to remember is "walking the dog". Imagine that you are walking in a large empty park containing at some "origin" point 0 a large pole (in the English sense of a metal post sticking out of the ground, not the complex analysis sense). You start at some point X and go for a walk along some curve, ending back at the same starting point X. Let N denote your winding number around the pole at the origin — that is, the total number of times you went around the pole, with the appropriate sign.

Now imagine that you also have a dog that is walking alongside you in some erratic path that is sometimes close to you, sometimes less close. As you traverse your curve C_1, the dog walks along on its own curve C_2, which also begins and ends in the same place. Let M denote the *dog's* winding number around the pole at the origin. Can we say that $N = M$? The answer is: *yes, we can, provided that we know the dog's distance to you was always less than your distance to the pole.* To see this, imagine that you had the dog on a leash of variable length; if the distance condition was not satisfied, it would be possible for the dog to reach the pole and go in a short tour around it while you were still far away and not turning around the pole, causing an entanglement of the leash with the pole.

Amazingly, the above scenario maps in a precise way to Rouché's theorem, using the following dictionary: the curve $f \circ \gamma$ represents your path; the curve $(f + g) \circ \gamma$ represents the dog's path; $g \circ \gamma$ represents the vector pointing from you to the dog; the condition $|f| > |g|$ along γ is precisely the correct condition that the dog stays closer to you than your distance to the pole; and the conclusion that the two winding numbers are the same is precisely the theorem's assertion that f and $f + g$ have the same number of generalized zeros in the region U enclosed by γ (see the discussion above regarding the connection between the integral $(2\pi i)^{-1} \oint_\gamma f'/f\, dz$ and the winding number of $f \circ \gamma$ around 0).

Exercise 8. Spend a few minutes thinking about the above correspondence and make sure you understand it. You will probably forget the technical details of the proof of Rouché's theorem in a few weeks or months, but I hope you will remember this intuitive explanation for a long time.

As another small cryptic remark to think about, the proof of Rouché's theorem given above can be thought of as an argument about the invariance of a certain integral under the homotopy between two curves. Can you see how?

11 Applications of Rouché's theorem

Rouché's theorem is an important tool in estimating the numbers of roots of polynomials and other functions in regions of interests (see Exercises 28–29 on page 84). The next results show how Rouché's theorem can also be used to deduce interesting theoretical results.

Theorem 27 (The open mapping theorem)**.** *Holomorphic functions are* ***open mappings****, that is, they map open sets to open sets.*

Proof. Let f be holomorphic in a region Ω, $z_0 \in \Omega$, and denote $w_0 = f(z_0)$. What we need to show is that the image of any neighborhood $D_\epsilon(z_0)$ for $\epsilon > 0$ contains a neighborhood $D_\delta(z_0)$ of w_0 for some $\delta > 0$. Fixing w (visualized as being near w_0), denote

$$h(z) = f(z) - w = (f(z) - w_0) + (w_0 - w) =: F(z) + G(z).$$

The idea is now to apply Rouché's theorem to $F(z)$ and $G(z)$. Fix $\epsilon > 0$ small enough so that the disc $D_\epsilon(z_0)$ is contained in Ω and does not contain solutions of the equation $f(z) = w_0$ other than z_0 (this is possible, by the property that zeros of holomorphic functions are isolated). Defining

$$\delta = \inf\{|f(z) - w_0| \,:\, z \in \overline{D_\epsilon(z_0)}\},$$

we therefore have that $\delta > 0$ and $|f(z) - w_0| \geq \delta$ for z on the circle $|z - z_0| = \epsilon$. That means that under the assumption that $|w - w_0| < \delta$ (i.e., if w is assumed to be close enough to w_0), the condition $|F(z)| > |G(z)|$ in Rouché's theorem will be satisfied for $z \in D_\epsilon(z_0)$. The conclusion is that the equation $h(z) = 0$ (or equivalently $f(z) = w$) has the same number in solutions (in particular, at least one solution) as the equation $f(z) = w_0$ in the disc $D_\epsilon(z_0)$. This was precisely the claim to be proved. □

Corollary 8 (the maximum modulus principle)**.** *If f is a non-constant holomorphic function on a region Ω, then $|f|$ cannot attain a maximum on Ω.*

Proof. Trivial exercise. □

At the beginning of the course we discussed several proofs of the fundamental theorem of algebra. One of them, the topological proof, was only sketched. As a final demonstration of the power of Rouché's theorem, problem 30 on page 84 asks you to use the theorem to make precise the idea of the topological proof.

12 Simply-connected regions and the general version of Cauchy's theorem

Definition 14 (homotopy of curves)**.** Given a region $\Omega \subset \mathbb{C}$, two parametrized curves $\gamma_1, \gamma_2 : [0,1] \to \Omega$ (assumed for simplicity of notation to be defined on $[0,1]$) are said to be **homotopic (with fixed endpoints)** if $\gamma_1(0) = \gamma_2(0)$, $\gamma_1(1) = \gamma_2(1)$, and there exists a function $F : [0,1] \times [0,1] \to \Omega$ such that

i) F is continuous.

ii) $F(0,t) = \gamma_1(t)$ for all $t \in [0,1]$.

iii) $F(1,t) = \gamma_2(t)$ for all $t \in [0,1]$.

iv) $F(s,0) = \gamma_1(0)$ for all $s \in [0,1]$.

v) $F(s,1) = \gamma_1(1)$ for all $s \in [0,1]$.

The map F is called a homotopy between γ_1 and γ_2. Intuitively, for each $s \in [0,1]$ the function $t \mapsto F(s,t)$ defines a curve connecting the two endpoints $\gamma_1(0)$, $\gamma_1(1)$. As s grows from 0 to 1, this family of curves transitions in a continuous way between the curve γ_1 and γ_2, with the endpoints being fixed in place.

Exercise 9. Prove that the relation of being homotopic is an equivalence relation.

Definition 15 (simply-connected region)**.** A region Ω is called **simply-connected** if any two curves γ_1, γ_2 in Ω with the same endpoints are homotopic.

Remark 10. A common alternative way to define the notion of homotopy of curves is for closed curves, where the endpoints are not fixed but the homotopy must keep the curves closed as it is deforming them. The definition of a simply-connected region then becomes a region in which any two closed curves are homotopic. It is not hard to show that those two definitions are equivalent.

Theorem 28. *If f is a holomorphic function on a region Ω, and γ_0,γ_1 are two curves on Ω with the same endpoints that are homotopic, then*

$$\int_{\gamma_0} f(z)\,dz = \int_{\gamma_1} f(z)\,dz.$$

Proof. This proof is based on the idea of translating the global statement about the equality of the two contour integrals into a local statement, similarly to the proof of Goursat's theorem but in a more general setting. (See also pages 93–95 in [11] for a variant of the proof presented below.)

Denote by $F : [0,1] \times [0,1] \to \Omega$ the homotopy between γ_0 and γ_1, and for any $s \in [0,1]$ denote by $\gamma_s : [0,1] \to \mathbb{C}$ the curve $\gamma_s(t) = F(s,t)$. The strategy of the proof is to show that there are values $0 = s_0 < s_1 < s_2 < \ldots < s_n = 1$ such that

$$\int_{\gamma_{s_0}} f(z)\,dz = \int_{\gamma_{s_1}} f(z)\,dz = \ldots = \int_{\gamma_{s_{n-1}}} f(z)\,dz = \int_{\gamma_{s_n}} f(z)\,dz.$$

That is, we will show that a slight perturbation of the s parameter does not change the value of the integral. To this end, for two fixed values $0 \le s < s' \le 1$ that are close to each other (in a sense we will make precise shortly), we break up the t-interval $[0,1]$ over which the curves s, s' are defined into very small segments by fixing points $0 = t_0 < t_1 < \ldots < t_m = 1$ that are very close together (in a sense that, again, we will need to make precise below once we understand what is needed to make the argument work), and then write

$$\int_{\gamma_s} f(z)\,dz = \sum_{j=1}^{m} \int_{\gamma_s([t_{j-1},t_j])} f(z)\,dz, \qquad \int_{\gamma_{s'}} f(z)\,dz = \sum_{j=1}^{m} \int_{\gamma_{s'}([t_{j-1},t_j])} f(z)\,dz.$$

We will now show that these two integrals are equal by exploiting our knowledge of local properties of f that follow from its analyticity. Specifically, assume that for each $1 \le j \le m$ we know that there exists an open disk D_j containing the convex hull of the union of the two curve segments $\gamma_s([t_{j-1},t_j])$ and $\gamma_{s'}([t_{j-1},t_j])$. For each $0 \le j \le m$, let η_j denote a straight line segment (considered as a parametrized curve) from $\gamma_s(t_j)$ to $\gamma_{s'}(t_j)$, and for each $1 \le j \le m$ let Γ_j denote the closed curve $\gamma_s([t_{j-1},t_j]) + \eta_j - \gamma_{s'}([t_{j-1},t_j]) - \eta_{j-1}$

(the concatenation of the four curves $\gamma_s([t_{j-1},t_j])$, "the reverse of η_j", $\gamma_{s'}([t_{j-1},t_j])$, and "the reverse of η_{j-1}"). By Cauchy's theorem on a disc, we have

$$\oint_{\Gamma_j} f(z)dz = 0,$$

or, more explicitly,

$$\int_{\gamma_s([t_{j-1},t_j])} f(z)\,dz - \int_{\gamma_{s'}([t_{j-1},t_j])} f(z)\,dz = \int_{\eta_{j-1}} f(z)\,dz - \int_{\eta_j} f(z)\,dz.$$

Summing this relation over j, we get that

$$\begin{aligned}\int_{\gamma_s} f(z)\,dz - \int_{\gamma_{s'}} f(z)\,dz &= \sum_{j=1}^{m}\left(\int_{\gamma_s([t_{j-1},t_j])} f(z)\,dz - \int_{\gamma_{s'}([t_{j-1},t_j])} f(z)\,dz\right)\\ &= \sum_{j=1}^{m}\left(\int_{\eta_{j-1}} f(z)\,dz - \int_{\eta_j} f(z)\,dz\right) = \int_{\eta_0} f(z)\,dz - \int_{\eta_m} f(z)\,dz = 0,\end{aligned}$$

since in the next-to-last step the sum is telescoping, and in the last step we note that η_0 and η_m are both degenerate curves each of which simply stays at a single point ($\gamma_s(0) = F(s,0) = \gamma_{s'}(0)$, and $\gamma_s(1) = F(s,1) = \gamma_{s'}(1)$, respectively). This is precisely the equality we wanted.

We still need to justify the assumption about the discs D_j. This can be made to work if s and s' are sufficiently close to each other and the points $0 < t_0 < t_1 < \ldots < t_m = 1$ are sufficiently densely spaced, using an argument involving continuity and compactness. Here is one way to make the argument: fix $0 \le s \le 1$. At each $0 \le t \le 1$, by continuity of the homotopy function F there exists a number $\delta > 0$ and a disc $D_{s,t}$ centered at $\gamma_s(t) = F(s,t)$ such that for any $s', t' \in [0,1]$ with $|s'-s| < \epsilon$, $|t'-t| < \delta$, we have $\gamma_{s'}(t') \in D_{s,t}$. The family of discs $D_{s,t}$ for $0 \le t \le 1$ are an open cover of the curve γ_s (or more precisely of its image $\gamma_s([0,1])$, which is a compact set), so by the Heine-Borel property we will have a finite subcovering $D_{s,t_0}, \ldots, D_{s,t_m}$. This is enough to prove the assumption for s' sufficiently close to s.

Finally, for each s denote by $\delta(s)$ the value of δ chosen above as a function of s. The collection of open intervals $\{(s-\delta(s), s+\delta(s)) : s \in [0,1]\}$ form an open covering of the interval $[0,1]$, so again using the Heine-Borel property, we can extract a finite subcovering. This enables us to find a sequence $0 = s_0 < s_1 < \ldots < s_n = 1$ that we claimed exist at the beginning of the proof, namely where the relation

$$\int_{\gamma_{s_{j-1}}} f(z)\,dz = \int_{\gamma_{s_j}} f(z)\,dz$$

holds for each $j = 1, \ldots, n$ (with s_{j-1} playing the role of s and s_j playing the role of s' in the discussion above). □

Theorem 29 (Cauchy's theorem (general version))**.** *If f is holomorphic on a simply-connected region Ω, then for any closed curve in Ω we have*

$$\oint_\gamma f(z)\,dz = 0.$$

Proof. Assume for simplicity that γ is parametrized as a curve on $[0,1]$. Then it can be thought of as the concatenation of two curves γ_1 and $-\gamma_2$, where $\gamma_1 = \gamma_{|[0,1/2]}$ and γ_2 is the "reverse" of the curve $\gamma_{|[1/2,1]}$. Note that γ_1 and γ_2 have the same endpoints. By the invariance property of contour integrals under homotopy proved above, we have

$$\int_\gamma f(z)\,dz = \int_{\gamma_1-\gamma_2} f(z)\,dz = \int_{\gamma_1} f(z)\,dz - \int_{\gamma_2} f(z)\,dz = 0.$$

$\square$

Corollary 9. *Any holomorphic function on a simply-connected region has a primitive.*

Exercise 10. The proof of Theorem 28 above still involves a minor amount of what I call "dishonesty"; that is, the proof is not actually formally correct as written but contains certain inconsistencies between what the assumptions of the theorem are and what we end up actually using in the body of the proof. Can you identify those inconsistencies? What additional work might be needed to fix these problems? And why do you think the author of these notes, and the authors of the textbook [11], chose to present things in this way rather than treat the subject in a completely rigorous manner devoid of any inaccuracies? (The last question is a very general one about mathematical pedagogy; coming up with a good answer might help to demystify for you a lot of similar decisions that textbook authors and course instructors make in the teaching of advanced material, and make the study of such topics a bit less confusing in the future.)

13 The logarithm function

The logarithm function can be defined as

$$\log z = \log|z| + i \arg z$$

on any region Ω that does not contain 0 and where one can make a consistent, smoothly varying choice of $\arg z$ as z ranges over Ω. It is easy to see that this formula gives an inverse to the exponential function.

For example, if

$$\Omega = \mathbb{C} \setminus (-\infty, 0]$$

(the "slit complex plane" with the negative real axis removed), we can set

$$\operatorname{Log} z = \log|z| + i \operatorname{Arg} z$$

where $\operatorname{Arg} z$ is set to take values in $(-\pi, \pi)$. This is called the *principal branch of the logarithm* — basically a kind of standard version of the log function that complex analysts have agreed to use whenever this is convenient (or not too inconvenient). However, sometimes we may want to consider the logarithm function on more strange or complicated regions. When can this be made to work? The answer is: precisely when Ω is simply-connected.

Theorem 30. *Assume that Ω is a simply-connected region with $0 \notin \Omega$, $1 \in \Omega$. Then there exists a function $F(z) = \log_\Omega(z)$ with the properties:*

i) F is holomorphic in Ω.

ii) $e^{F(z)} = z$ for all $z \in \Omega$.

iii) $F(r) = \log r$ (the usual logarithm for real numbers) for all real numbers $r \in \Omega$ sufficiently close to 1.

Proof. We define F as a primitive function of the function $z \mapsto 1/z$, that is, as

$$F(z) = \int_1^z \frac{dw}{w},$$

where the integral is computed along a curve γ connecting 1 to z. By the general version of Cauchy's theorem for simply-connected regions, this integral is independent of the choice of curve. As we have already seen, this function is holomorphic and satisfies $F'(z) = 1/z$ for all $z \in \Omega$. It follows that

$$\frac{d}{dz}\left(ze^{-F(z)}\right) = e^{-F(z)} - zF'(z)e^{-F(z)} = e^{-F(z)}(1 - z/z) = 0,$$

so $ze^{-F(z)}$ is a constant function. Since its value at $z = 1$ is 1, we see that $e^{F(z)} = z$, as required. Finally, for real r close to 1 we have that $F(z) = \int_1^r \frac{dw}{w} = \log r$, which can be seen by taking the integral to be along the straight line segment connecting 1 and r. □

Exercise 11. Prove that the principal branch of the logarithm has the Taylor series expansion

$$\operatorname{Log} z = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(z-1)^n \qquad (|z| < 1).$$

Exercise 12. Modify the proof above to prove the existence of a branch of the logarithm function in any simply-connected region Ω not containing 0, without the assumption that $1 \in \Omega$. In what way is the conclusion weakened in that case?

Exercise 13. Explain in what sense the logarithm functions $F(z) = \log_\Omega(z)$ satisfying the properties proved in the theorem above (and its generalization described in the previous exercise) are unique.

Exercise 14. Prove the following generalization of the logarithm construction above: if f is a holomorphic function on a simply-connected region Ω, and $f \neq 0$ on Ω, then there exists a holomorphic function g on Ω, referred to as a branch of the logarithm of f, satisfying

$$e^{g(z)} = f(z).$$

Definition 16 (power functions and nth roots)**.** On a simply-connected region Ω we can now define the power function $z \mapsto z^\alpha$ for an arbitrary $\alpha \in \mathbb{C}$ by setting

$$z^\alpha = e^{\alpha \log z}.$$

In the special case $\alpha = 1/n$ this has the meaning of the nth root function $z \mapsto z^{1/n}$, which satisfies

$$(z^{1/n})^n = \left(e^{\frac{1}{n}\log z}\right)^n = e^{n\frac{1}{n}\log z} = e^{\log z} = z.$$

Note that if $f(z) = z^{1/n}$ is one choice of an nth root function, then for any $0 \leq k \leq n-1$, the function $g(z) = e^{2\pi i k/n} f(z)$ will be another function satisfying $g(z)^n = z$. Conversely, it is easy to see that those are precisely the possible choices for an nth root function.

14 The Euler gamma function

The *Euler gamma function* (often referred to simply as the gamma function) is one of the most important special functions in mathematics. It has applications to many areas, such as combinatorics, number theory, differential equations, probability, and more, and is probably the most ubiquitous transcendental function after the "elementary" transcendental functions (the exponential function, logarithms, trigonometric functions and their inverses) that one learns about in calculus. It is a natural meromorphic function of a complex variable that extends the factorial function to non-integer values. In complex analysis it is particularly important in connection with the theory of the Mellin transform (a version of the Fourier transform associated with the multiplicative group of positive real numbers in the same way that the ordinary Fourier transform is associated with the additive group of the real numbers).

Most textbooks define the gamma function in one way and proceed to prove several other equivalent representations of it. However, the truth is that none of the representations of the gamma function is more fundamental or "natural" than the others. So, it seems more logical to start by simply listing the various formulas and properties associated with it, and then proving that the different representations are equivalent and that the claimed properties hold.

Theorem 31 (the Euler gamma function)**.** *There exists a unique function $\Gamma(s)$ of a complex variable s that has the following properties:*

1. *$\Gamma(s)$ is a meromorphic function on $\mathbb{C}$.*
2. ***Connection to factorials:** $\Gamma(n+1) = n!$ for $n = 0, 1, 2, \ldots$.*
3. ***Important special value:** $\Gamma(1/2) = \sqrt{\pi}$.*
4. ***Integral representation:***
$$\Gamma(s) = \int_0^\infty e^{-x} x^{s-1}\, dx \qquad (\operatorname{Re} s > 0).$$
5. ***Hybrid series-integral representation:***
$$\Gamma(s) = \sum_{n=0}^\infty \frac{(-1)^n}{n!(n+s)} + \int_1^\infty e^{-x} x^{s-1}\, dx \qquad (s \in \mathbb{C}).$$
6. ***Infinite product representation:***
$$\Gamma(s)^{-1} = s e^{\gamma s} \prod_{n=1}^\infty \left(1 + \frac{s}{n}\right) e^{-s/n} \qquad (s \in \mathbb{C}),$$
where $\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} - \log n\right) \doteq 0.577215$ is the Euler-Mascheroni constant.
7. ***Limit of finite products representation:***
$$\Gamma(s) = \lim_{n\to\infty} \frac{n!\, n^s}{s(s+1)\cdots(s+n)} \qquad (s \in \mathbb{C}).$$
8. ***Zeros:** the gamma function has no zeros (so $\Gamma(s)^{-1}$ is an entire function).*
9. ***Poles:** the gamma function has poles precisely at the non-positive integers $s = 0, -1, -2, \ldots$, and is holomorphic everywhere else. The pole at $s = -n$ is a simple pole with residue*
$$\operatorname{Res}_{s=-n}(\Gamma) = \frac{(-1)^n}{n!} \qquad (n = 0, 1, 2, \ldots).$$

Functional equation:

$$\Gamma(s+1) = s\,\Gamma(s) \qquad (s \in \mathbb{C}).$$

The reflection formula *(a surprising connection to trigonometry):*

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)} \qquad (s \in \mathbb{C}).$$

To begin the proofs, we do have to define the function we are claiming exists *somehow*, so let's take the formula

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

as our working definition of $\Gamma(s)$. This improper integral is easily seen to converge absolutely for $\mathrm{Re}(s) > 0$, since

$$\left| \int_0^\infty e^{-x} x^{s-1}\,dx \right| \le \int_0^\infty e^{-x} |x^{s-1}|\,dx = \int_0^\infty e^{-x} x^{\mathrm{Re}(s)-1}\,dx.$$

I leave it as an exercise to check (or read the easy explanation in [11]) that the function it defines is holomorphic in that region.

Next, perform an integration by parts, to get that, again for $\mathrm{Re}(s) > 0$, we have

$$\Gamma(s+1) = \int_0^\infty e^{-x} x^s\,dx = -e^{-x}x^s\Big|_{x=0}^{x=\infty} + \int_0^\infty e^{-x} s x^{s-1}\,dx = s\,\Gamma(s),$$

which is the functional equation.

Combining the trivial evaluation $\Gamma(1) = \int_0^\infty e^{-x}\,dx = 1$ with the functional equation shows by induction that $\Gamma(n+1) = n!$.

The special value $\Gamma(1/2) = \sqrt{\pi}$ follows immediately by a change of variable $x = u^2$ in the integral and an appeal to the standard Gaussian integral $\int_{-\infty}^\infty e^{-u^2}\,du = \sqrt{\pi}$:

$$\Gamma(1/2) = \int_0^\infty e^{-x} x^{-1/2}\,dx = \int_0^\infty e^{-u^2}\,2\,du = \int_{-\infty}^\infty e^{-u^2}\,du = \sqrt{\pi}.$$

The functional equation can now be used to perform an analytic continuation of $\Gamma(s)$ to a meromorphic function on $\mathbb{C}$: for example, we can define

$$\Gamma_1(s) = \frac{\Gamma(s+1)}{s},$$

which is a function that is holomorphic on $\mathrm{Re}(s) > -1, s \neq 0$ and coincides with $\gamma(s)$ for $\mathrm{Re}(s) > 0$. By the principle of analytic continuation this provides a unique extension of $\Gamma(s)$ to the region $\mathrm{Re}(s) > -1$. Because of the factor $1/s$ and the fact that $\Gamma(1) = 1$ we also see that $\Gamma_1(s)$ has a simple pole at $s = 0$ with residue 1.

Next, for $\mathrm{Re}(s) > -2$ we define

$$\Gamma_2(s) = \frac{\Gamma_1(s+1)}{s} = \frac{\Gamma(s+2)}{s(s+1)},$$

a function that is holomorphic on $\mathrm{Re}(s) > -2$, $s \neq 0, -1$, and coincides with $\Gamma_1(s)$ for $\mathrm{Re}(s) > -1$, $s \neq 0$. Again, this provides an analytic continuation of $\Gamma(s)$ to that region. The factors $1/s(s+1)$ show that $\Gamma_2(s)$ has a simple pole at $s = -1$ with residue -1.

Continuing by induction, having defined an analytic continuation $\Gamma_{n-1}(s)$ of $\Gamma(s)$ to the region $\mathrm{Re}(s) > -n+1$, $s \neq 0, -1, -2, \ldots, -n+2$, we now define

$$\Gamma_n(s) = \frac{\Gamma_{n-1}(s+1)}{s} = \ldots = \frac{\Gamma(s+n)}{s(s+1)\cdots(s+n-1)}.$$

By inspection we see that this gives a meromorphic function in $\mathrm{Re}(s) > -n$ whose poles are precisely at $s = -n+1, \ldots, 0$ and have the claimed residues.

An alternative way to perform the analytic continuation is to separate the integral defining $\Gamma(s)$ into

$$\Gamma(s) = \int_0^1 e^{-x} x^{s-1}\,dx + \int_1^\infty e^{-x} x^{s-1}\,dx$$

and to note that the integral over $[1, \infty)$ converges (and defines a holomorphic function of s) for *all* $s \in \mathbb{C}$, and the integral over $[0,1]$ can be computed by expanding e^{-x} as a power series in x and integrating term by term. That is, for $\mathrm{Re}(s) > 0$ we have

$$\begin{aligned}\int_0^1 e^{-x} x^{s-1}\,dx &= \int_0^1 \sum_{n=0}^\infty \frac{(-1)^n}{n!} x^{n+s-1}\,dx = \sum_{n=0}^\infty \frac{(-1)^n}{n!} \int_0^1 x^{n+s-1}\,dx \\ &= \sum_{n=0}^\infty \frac{(-1)^n}{n!(n+s)}\end{aligned}$$

The justification for interchanging the summation and integration operations is easy and is left as an exercise. Thus, we have obtained not just an alternative proof for the meromorphic continuation of $\Gamma(s)$, but a proof of the hybrid series-integral representation of $\Gamma(s)$, which also clearly shows where the poles of $\Gamma(s)$ are and that they are simple poles with the correct residues.

Lemma 6. *For* $\mathrm{Re}(s) > 0$ *we have*

$$\Gamma(s) = \lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n x^{s-1}\,dx.$$

Proof. As $n \to \infty$, the integrand converges to $e^{-x}x^{s-1}$ pointwise. Furthermore, the factor $\left(1 - \frac{x}{n}\right)^n$ is bounded from above by the function e^{-x} (because of the elementary inequality $1 - t \leq e^{-t}$ that holds for all real t). The claim therefore follows from the dominated convergence theorem. $\square$

Lemma 7. *For* $\mathrm{Re}(s) > 0$ *we have*

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

Proof. For $n = 1$, the claim is that

$$\int_0^1 (1-x)\,x^{s-1}\,dx = \frac{1}{s(s+1)},$$

which is easy to verify directly. For the general claim, using a linear change of variables and an integration by parts we see that

$$\begin{aligned}\int_0^n \left(1-\frac{x}{n}\right)^n x^{s-1}\,dx &= n^s \int_0^1 (1-t)^n t^{s-1}\,dt \\ &= n^s \left[(1-t)^n \frac{t^s}{s}\Big|_{t=0}^{t=1} + \int_0^1 n(1-t)^{n-1}\frac{t^s}{s}\,dt \right] \\ &= n^s \cdot \frac{n}{s}\int_0^1 (1-t)^{n-1} t^{(s+1)-1}\,dt,\end{aligned}$$

so the claim follows by induction on n. □

Corollary 10. *For* $\mathrm{Re}(s) > 0$ *we have*

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

Proof of the infinite product representation for $\Gamma(s)$. For $\mathrm{Re}(s) > 0$ we have

$$\begin{aligned}\Gamma(s)^{-1} &= \lim_{n\to\infty} \frac{s(s+1)\cdots(s+n)}{n!\,n^s} \\ &= s \lim_{n\to\infty} e^{-s\log n}\left(1+\frac{s}{1}\right)\left(1+\frac{s}{2}\right)\cdots\left(1+\frac{s}{n}\right) \\ &= s \lim_{n\to\infty} e^{s(\sum_{k=1}^n \frac{1}{k} - \log n)} \prod_{k=1}^n \left(1+\frac{s}{k}\right)e^{-s/k} \\ &= s e^{\gamma s} \prod_{n=1}^\infty \left(1+\frac{s}{n}\right)e^{-s/n}.\end{aligned}$$

□

We now check that the infinite product actually converges absolutely and uniformly on compact subsets in all of $\mathbb{C}$, so defines an entire function. Let's start with some preliminary elementary observations on infinite products.

Recall that for a sequence of complex numbers c_n, the infinite product $\prod_{n=1}^\infty c_n$ is defined as the limit of finite (partial) products $\lim_{n\to\infty}\prod_{k=1}^n c_k$, if the limit exists. This is analogous to the notation for infinite series. One point of terminology that differs slightly from the corresponding terminology for infinite series, is that we say the product $\prod_{n=1}^\infty c_n$ *converges* if the limit of the finite products exists *and is non-zero.*

Lemma 8. *For a sequence of complex numbers* $(a_n)_{n=1}^\infty$*, if* $a_n \neq -1$ *for all* n *and* $\sum_{n=1}^\infty |a_n| < \infty$ *then the infinite product* $\prod_{n=1}^\infty (1+a_n)$ *converges.*

Proof. Under the assumption, there exists some large enough $N_0 \geq 1$ such that $|a_n| < 1/2$ for all $n \geq N_0$, which in particular implies that $1 + a_n = \exp(\mathrm{Log}(1+a_n))$, where $\mathrm{Log}(z)$ is the principal branch of the logarithm function. Now, by the assumption, clearly it is enough to prove the convergence of the infinite product $\prod_{n=N_0}^\infty (1+a_n)$, and this can be written as

$$\prod_{n=N_0}^\infty (1+a_n) = \lim_{n\to\infty}\prod_{k=N_0}^n (1+a_k) = \lim_{n\to\infty}\prod_{k=N_0}^n \exp\left(\mathrm{Log}(1+a_k)\right) = \lim_{n\to\infty}\exp\left(\sum_{k=N_0}^n \mathrm{Log}(1+a_k)\right).$$

If we knew that the infinite series $\sum_{n=N_0}^{\infty} \operatorname{Log}(1+a_n) = \lim_{n\to\infty} \sum_{k=N_0}^{n} \operatorname{Log}(1+a_k)$ converged, we could continue the above chain of equalities as

$$= \exp\left(\lim_{n\to\infty} \sum_{k=N_0}^{n} \operatorname{Log}(1+a_k)\right) = \exp\left(\sum_{n=N_0}^{\infty} \operatorname{Log}(1+a_n)\right),$$

and since e^z is never 0, that would mean that the product $\prod_{n=N_0}^{\infty}(1+a_n)$ exists and is non-zero (that is, it converges), hence as was mentioned above the infinite product $\prod_{n=1}^{\infty}(1+a_n)$ also converges and the claim would be proved.

To see the convergence of the series $\sum_{n=N_0}^{\infty} \operatorname{Log}(1+a_n)$, recall that the function $z \mapsto \operatorname{Log}(z)$ has the convergent Taylor expansion

$$\operatorname{Log}(z) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m}(z-1)^m \qquad (|z-1|<1).$$

In particular, there is some constant $C>0$ such that

$$|\operatorname{Log}(1+w)| \le |w| + C|w|^2 \qquad \text{if } |w| < 1/2.$$

Now using the assumption that $\sum_{n=1}^{\infty} |a_n| < \infty$ it follows that

$$\sum_{n=N_0}^{\infty} |\operatorname{Log}(1+a_n)| \le \sum_{n=N_0}^{\infty} |a_n| + C \sum_{n=N_0}^{\infty} |a_n|^2 \le \sum_{n=N_0}^{\infty} |a_n| + C\left(\sum_{n=N_0}^{\infty} |a_n|\right)^2 < \infty,$$

which is what we needed. $\square$

Lemma 9. *Let $(f_n)_{n=1}^{\infty}$ be a sequence of functions on some region Ω such that the functions $1+f_n$ are everywhere holomorphic and nonzero. If the series $\sum_{n=1}^{\infty} |f_n|$ converges uniformly on compacts in Ω, the infinite product $\prod_{n=1}^{\infty}(1+f_n)$ also converges uniformly on compact subsets in Ω to a nonzero holomorphic function.*

Proof. Lemma 8 above implies that the infinite product $\prod_{n=1}^{\infty}(1+f_n(z))$ converges to a nonzero limit for any $z \in \Omega$. By repeating the same estimates in the proof of that lemma in the context of z being allowed to range on a compact subset $K \subset \Omega$, one sees that the sequence of partial products $\prod_{k=1}^{n}(1+f_n)$ actually converges uniformly on compacts, so the limiting function is holomorphic. $\square$

Proof that $\prod_{n=1}^{\infty}\left(1+\frac{z}{n}\right)e^{-z/n}$ is an entire function.

$$\sum_{n=1}^{\infty}\left|\left(1+\frac{z}{n}\right)e^{-z/n} - 1\right| = \sum_{n=1}^{\infty}\left|\left(1+\frac{z}{n}\right)\left(1-\frac{z}{n}+O\left(\frac{z^2}{n^2}\right)\right) - 1\right| = \sum_{n=1}^{\infty}\left|O\left(\frac{z^2}{n^2}\right)\right| < \infty$$

(where the big-O notation hides a universal constant — the dependence on z is encapsulated in the z^2 factor). In particular, the convergence is uniform on compacts on $\mathbb{C}$. So we are almost in the setting of Lemma 9, except that in order to apply that result, which requires the functions participating in the product to be nonzero, one needs to be a bit more careful and separate out the zeros: for a fixed disc $D_{N+1/2}(0)$ of radius $N+1/2$ around 0, consider only the product starting at $n=N+1$ — those functions are nonzero in the disc so the previous result applies to give a function that's holomorphic and nonzero in $D_N(0)$. Then separately the factors $(1+z/n)$, $n=1,\dots,N$ contribute simple zeros at $z=-1,\dots,-N$. $\square$

Corollary 11 (the reflection formula)**.** $\Gamma(s)\Gamma(1-s) = \dfrac{\pi}{\sin \pi s}$.

Proof.

$$\begin{aligned}
\frac{1}{\Gamma(s)\Gamma(1-s)} &= \Gamma(s)^{-1}(-s)^{-1}\Gamma(-s)^{-1} \\
&= \frac{-1}{s} \cdot s e^{\gamma s} \prod_{n=1}^{\infty} \left(1+\frac{s}{n}\right) e^{-s/n} \cdot (-s) e^{-\gamma s} \prod_{n=1}^{\infty} \left(1-\frac{s}{n}\right) e^{s/n} \\
&= s \prod_{n=1}^{\infty} \left(1-\frac{s^2}{n^2}\right) = s\frac{\sin(\pi s)}{\pi s} = \frac{\sin(\pi s)}{\pi},
\end{aligned}$$

where we used the product representation $\sin(\pi z) = \pi z \prod_{n=1}^{\infty}(1-z^2/n^2)$ for the sine function derived in a homework problem. ☐

Alternative derivation of the reflection formula ([11], page 164). By analytic continuation, it is enough to prove the formula for real s in $(0,1)$. For such s we have

$$\begin{aligned}
\Gamma(s)\Gamma(1-s) &= \int_0^{\infty} e^{-t} t^{-s} \Gamma(s)\, dt \\
&= \int_0^{\infty} e^{-t} t^{-s} \left(t \int_0^{\infty} e^{-vt} (vt)^{s-1}\, dv \right) dt \\
&= \int_0^{\infty} \int_0^{\infty} e^{-t(1+v)} v^{s-1}\, dv\, dt = \int_0^{\infty} \left(\int_0^{\infty} e^{-t(1+v)}\, dt \right) v^{s-1}\, dv \\
&= \int_0^{\infty} \frac{v^{s-1}}{1+v}\, dv = \int_{-\infty}^{\infty} \frac{e^{sx}}{1+e^x}\, dx \quad \text{(by setting } v = e^x\text{)}.
\end{aligned}$$

So it is enough to prove that for $0 < s < 1$ we have

$$\int_{-\infty}^{\infty} \frac{e^{sx}}{1+e^x}\, dx = \frac{\pi}{\sin(\pi s)}.$$

This integral can be evaluated using residue calculus; see Example 2 in Section 2.1, Chapter 3, pages 79–81 of [11] for the details. ☐

Note that by combining the alternative derivation of the reflection formula given above with the infinite product representation for the gamma function, we get a new proof of the infinite product representation for $\sin(\pi z)$.

15 The Riemann zeta function

15.1 Definition and basic properties

The *Riemann zeta function* (often referred to simply as the zeta function when there is no risk of confusion), like the gamma function is considered one of the most important special functions in “higher” mathematics. However, the Riemann zeta function is a lot more mysterious than the gamma function, and remains the subject of many famous open problems, including the most famous of them all: the **Riemann hypothesis**, considered by many (including myself) as the most important open problem in mathematics.

The main reason for the zeta function's importance is its connection with prime numbers and other concepts and quantities from number theory. Its study, and in particular the attempts to prove the Riemann hypothesis, have also stimulated an unusually large number of important developments in many areas of mathematics.

As with the gamma function, the Riemann zeta function is usually defined on only part of the complex plane and its definition is then extended by analytic continuation. Again, I will formulate this as a theorem asserting the existence of the zeta function and its various properties.

Theorem 32 (Riemann zeta function)**.** *There exists a unique function, denoted $\zeta(s)$, of a complex variable s, having the following properties:*

1. *$\zeta(s)$ is a meromorphic function on $\mathbb{C}$.*
2. *For $\mathrm{Re}(s) > 1$, $\zeta(s)$ is given by the series*
$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \ldots.$$
3. ***Euler product formula:** for $\mathrm{Re}(s) > 1$, $\zeta(s)$ also has an infinite product representation*
$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}},$$
where the product ranges over the prime numbers $p = 2, 3, 5, 7, 11, \ldots$.
4. *$\zeta(s)$ has no zeros in the region $\mathrm{Re}(s) > 1$.*
5. *$\zeta(s)$ has no zeros on the line $\mathrm{Re}(s) = 1$ (this requires a separate proof from the previous claim).*
6. ***The "trivial" zeros:** the zeros of $\zeta(s)$ in the region $\mathrm{Re}(s) \le 0$ are precisely at $s = -2, -4, -6, \ldots$.*
7. *$\zeta(s)$ has a unique pole located at $s = 1$. It is a simple pole with residue 1.*
8. ***The "Basel problem" and its generalizations:** the values of $\zeta(s)$ at even positive integers are given by Euler's formula*
$$\zeta(2n) = \frac{(-1)^{n-1}(2\pi)^{2n}}{2(2n)!} B_{2n} \qquad (n = 1, 2, \ldots),$$
where $(B_m)_{m=0}^{\infty}$ are the Bernoulli numbers, defined as the coefficients in the Taylor expansion
$$\frac{z}{e^z - 1} = \sum_{m=0}^{\infty} \frac{B_m}{m!} z^m.$$
Many of the properties of these amazing numbers were discussed in our homework problem sets.
9. ***Values at negative integers:** we have*
$$\zeta(-n) = -\frac{B_{n+1}}{n+1} \qquad (n = 1, 2, 3, \ldots).$$
(Note that for negative even integers, this coincides with the property stated above about the trivial zeros at $s = -2, -4, -6, \ldots$, since it is an easy fact that the Bernoulli numbers satisfy $B_{2k+1} = 0$ for integer $k \ge 1$. But this formula adds information about the values of $\zeta(s)$ at negative odd integers.)

10. ***Functional equation:*** *the zeta function satisfies*

$$\zeta^*(1-s) = \zeta^*(s),$$

where we denote by $\zeta^*(s)$ the ***symmetrized zeta function***

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

11. ***Mellin transform representation:*** *an expression for $\zeta(s)$ valid for all $s \in \mathbb{C}$ is*

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) \\ = -\frac{1}{1-s} - \frac{1}{s} + \frac{1}{2}\int_1^\infty \left(t^{-\frac{s+1}{2}} + t^{\frac{s-2}{2}}\right)(\vartheta(t)-1)\,dt,$$

where the function $\vartheta(t)$ is one of ***Jacobi theta series****, defined as*

$$\vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} = 1 + 2\sum_{n=1}^{\infty} e^{-\pi n^2 t}.$$

12. ***Contour integral representation:*** *another expression for $\zeta(s)$ valid for all $s \in \mathbb{C}$ is*

$$\zeta(s) = \frac{\Gamma(1-s)}{2\pi i}\int_C \frac{(-x)^s}{e^x - 1}\frac{dx}{x},$$

where C is a keyhole contour coming from $+\infty$ to 0 slightly above the positive x-axis, then circling the origin in a counterclockwise direction around a circle of small radius, then going back to $+\infty$ slightly below the positive x-axis.

13. ***Connection to prime number enumeration — the "explicit formula of number theory":*** *define Von Mangoldt's weighted prime counting function*

$$\psi(x) = \sum_{p^k \le x} \log p,$$

where the sum is over all prime powers less than or equal to x. Then for non-integer $x > 1$,

$$\psi(x) = x - \sum_\rho \frac{x^\rho}{\rho} - \log(2\pi),$$

where the sum ranges over all zeros ρ of the Riemann zeta function counted with their respective multiplicities. (In most textbooks the sum is separated into two sums, one ranging over the trivial zeros which can be evaluated explicitly, and the other ranging over the much less trivial zeros in the strip $0 < \mathrm{Re}(s) < 1$. Also the sum is only conditionally convergent; refer to a book on analytic number theory for the proper way to interpret it to get a convergent sum.)

The explicit formula of number theory illustrates that knowing where the zeros of $\zeta(s)$ are has important consequences for prime number enumeration. In particular, proving that $\mathrm{Re}(s)$ has no zeros in $\mathrm{Re}(s) \geq 1$ will enable us to prove one of the most famous theorems in mathematics.

Theorem 33 (Prime number theorem)**.** *Let $\pi(x)$ denote the number of prime numbers less than or equal to x. Then we have*

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

Conjecture 1 (The Riemann hypothesis)**.** *All the nontrivial zeros of $\zeta(s)$ are on the "critical line"* $\mathrm{Re}(s) = 1/2$.

To begin the proof of Theorem 32, again, let's take as the definition of $\zeta(s)$ the standard representation

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

Since $\sum_n |n^{-s}| = \sum_n n^{-\mathrm{Re}(s)}$, we see that the series converges absolutely precisely when $\mathrm{Re}(s) > 1$, and that the convergence is uniform on any half-plane of the form $\mathrm{Re}(s) > \alpha$ where $\alpha > 1$. In particular, it is uniform on compact subsets, so $\zeta(s)$ is holomorphic in this region.

Similarly, the Euler product $Z(s) := \prod_p (1 - p^{-s})^{-1}$ converges absolutely if and only if the series $\sum_p |p^{-s}| = \sum_p p^{-\mathrm{Re}(s)}$ converges, and in particular if $\mathrm{Re}(s) > 1$. It follows that $Z(s)$ is well-defined, holomorphic and nonzero for $\mathrm{Re}(s) > 1$.

Proof of the Euler product formula. We now prove that $Z(s) = \zeta(s)$. This can be done by manipulating the partial products associated with the infinite product defining $Z(s)$, as follows:

$$\begin{aligned}\zeta_N(s) := \prod_{p \le N} \frac{1}{1 - p^{-s}} &= \prod_{p \le N} (1 + p^{-s} + p^{-2s} + p^{-3s} + \ldots) \\ &= \sum_{\substack{n = p_1^{j_1} \cdots p_k^{j_k} \\ p_1, \ldots, p_k \text{ primes } \le N}} \frac{1}{n^s},\end{aligned}$$

where the last equality follows from the fundamental theorem of arithmetic, together with the fact that when multiplying two (or a finite number of) infinitely convergent series, the summands can be rearranged and summed in any order we desire. So, we have represented $\zeta_N(s)$ as a series of a similar form as the series defining $\zeta(s)$, but involving terms of the form n^{-s} only for those positive integers n whose prime factorization contains only primes $\le N$. It follows that

$$|\zeta(s) - \zeta_N(s)| \le \sum_{n > N} \frac{1}{n^s}.$$

Taking the limit as $N \to \infty$ shows that $Z(s) = \lim_{N \to \infty} \zeta_N(s) = \zeta(s)$. This proves the validity of the Euler product formula. □

Corollary 12. *$\zeta(s)$ has no zeros in the region* $\mathrm{Re}(s) > 1$.

Proof. The Euler product formula gives a convergent product for $\zeta(s)$ in this region where each factor $(1 - p^{-s})^{-1}$ has no zeros. □

To prove the functional equation, we need a somewhat powerful tool from harmonic analysis, the Poisson summation formula.

Theorem 34 (the Poisson summation formula)**.** *For a sufficiently well-behaved function $f : \mathbb{R} \to \mathbb{C}$, we have*

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{k=-\infty}^{\infty} \hat{f}(k),$$

where

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

is the Fourier transform of f.

Proof. Define a function $g : [0,1] \to \mathbb{C}$ by

$$g(x) = \sum_{n=-\infty}^{\infty} f(x+n),$$

the "periodiciziation" of f. Assume that $f(x)$ is sufficiently well-behaved (i.e., decays fast enough as $x \to \pm\infty$ so that $g(x)$ is in turn well-behaved, and has reasonable smoothness properties). In that case, a standard result from harmonic analysis states that $g(x)$ will have a convergent Fourier expansion of the form

$$g(x) = \sum_{k=-\infty}^{\infty} \hat{g}(k) e^{2\pi i k x},$$

where the Fourier coefficients $\hat{g}(k)$ can be computed as

$$\hat{g}(k) = \int_0^1 g(x) e^{-2\pi i k x}\, dx.$$

In particular, setting $x = 0$ in the formula for $g(x)$ gives the basic fact that

$$g(0) = \sum_{k=-\infty}^{\infty} \hat{g}(k).$$

However, note that $g(0) = \sum_{n=-\infty}^{\infty} f(n)$, the quantity on the left-hand side of the Poisson summation formula. On the other hand, the Fourier coefficient $\hat{g}(k)$ can be expressed in terms of the Fourier coefficients of the original function $f(x)$:

$$\begin{aligned}
\hat{g}(k) &= \int_0^1 g(x) e^{-2\pi i k x}\, dx = \int_0^1 \sum_{n=-\infty}^{\infty} f(x+n) e^{-2\pi i k x}\, dx \\
&= \sum_{n=-\infty}^{\infty} \int_0^1 f(x+n) e^{-2\pi i k x}\, dx = \sum_{n=-\infty}^{\infty} \int_n^{n+1} f(u) e^{-2\pi i k u}\, du \\
&= \int_{-\infty}^{\infty} f(u) e^{-2\pi i k u}\, du = \hat{f}(k).
\end{aligned}$$

Combining these observations gives the result, modulo a few details we've glossed over concerning the precise assumptions that need to be made about $f(x)$ (we will only apply the Poisson summation formula for one extremely well-behaved function, so I will not bother discussing those details). ☐

Theorem 35. *The Jacobi theta function $\vartheta(t)$ satisfies the functional equation*

$$\vartheta(t) = \frac{1}{\sqrt{t}} \vartheta(1/t) \qquad (t > 0).$$

Remark 11. Equations of this form are studied in the theory of *modular forms*, an area of mathematics combining number theory, complex analysis and algebra in a very surprising and beautiful way.

Proof. The idea is to apply the Poisson summation formula to the function

$$f(x) = e^{-\pi t x^2},$$

for which it can be checked that

$$\hat{f}(k) = t^{-1/2} e^{-\pi k^2/t},$$

using a simple change of variables from the standard integral evaluation

$$\int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x u}\, du = e^{-\pi u^2}$$

(that is, the fact that the function $e^{-\pi x^2}$ is its own Fourier transform); this evaluation appears in Example 1, Chapter 2, pages 42–44 in [11]. With the above substitution for $f(x)$ and $\hat{f}(k)$, the Poisson summation formula becomes precisely the functional equation for $\vartheta(t)$. □

Exercise 15. (a) Use the residue theorem to evaluate the contour integral

$$\oint_{\gamma_N} \frac{e^{-\pi z^2 t}}{e^{2\pi i z} - 1}\, dz,$$

where γ_N is the rectangle with vertices $\pm(N + 1/2) \pm i$ (with N a positive integer), then take the limit as $N \to \infty$ to derive the integral representation

$$\vartheta(t) = \int_{-\infty - i}^{\infty - i} \frac{e^{-\pi z^2 t}}{e^{2\pi i z} - 1}\, dz - \int_{-\infty + i}^{\infty + i} \frac{e^{-\pi z^2 t}}{e^{2\pi i z} - 1}\, dz$$

for the Jacobi theta function.

(b) In this representation, expand the factor $(e^{2\pi i z} - 1)^{-1}$ as a geometric series in $e^{-2\pi i z}$ (for the first integral) and as a geometric series in $e^{2\pi i z}$ (for the second integral). Evaluate the resulting infinite series, rigorously justifying all steps, to obtain an alternative proof of the functional equation for $\vartheta(t)$.

Lemma 10. *The asymptotic behavior of $\vartheta(t)$ near $t = 0$ and $t = +\infty$ is given by*

$$\vartheta(t) = O\left(\frac{1}{\sqrt{t}}\right) \qquad (t \to 0+),$$

$$\vartheta(t) = 1 + O(e^{-\pi t}) \qquad (t \to \infty).$$

Proof. The asymptotics as $t \to \infty$ is immediate from

$$\vartheta(t) - 1 = 2\sum_{n=1}^{\infty} e^{-\pi n^2 t} \le 2\sum_{n=1}^{\infty} e^{-\pi n t} = \frac{2e^{-\pi t}}{1 - e^{-\pi t}},$$

which is bounded by $Ce^{-\pi t}$ if $t > 10$. Using the functional equation now gives that $\vartheta(t) = t^{-1/2}(1 + O(e^{-\pi/t})) = O(t^{-1/2})$ as $t \to 0+$. □

Proof of the analytic continuation of $\zeta(s)$. Start with the formula

$$\Gamma\left(\frac{s}{2}\right) = \int_0^{\infty} e^{-x} x^{s/2-1}\, dx,$$

valid for $\mathrm{Re}(s) > 0$. A linear change of variable $x = \pi n^2 t$ brings this to the form

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right) n^{-s} = \int_0^\infty e^{-\pi n^2 t} t^{s/2-1}\, dt.$$

Summing the left-hand side over $n = 1, 2, \ldots$ gives $\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s)$ — the function we denoted $\zeta^*(s)$ — adding the stronger assumption that $\mathrm{Re}(s) > 1$. For the right-hand side we have that

$$\begin{aligned}\sum_{n=1}^\infty \int_0^\infty e^{-\pi n^2 t} t^{s/2-1}\, dt. &= \int_0^\infty \left(\sum_{n=1}^\infty e^{-\pi n^2 t}\right) t^{s/2-1}\, dt \\ &= \int_0^\infty \frac{\vartheta(t)-1}{2} t^{s/2-1}\, dt,\end{aligned}$$

where the estimates in the lemma are needed to justify interchanging the order of the summation and integration, and show that the integral converges for $\mathrm{Re}(s) > 1$. Thus we have obtained the representation

$$\zeta^*(s) = \frac{1}{2}\int_0^\infty (\vartheta(t)-1) t^{s/2-1}\, dt = \int_0^\infty \varphi(t) t^{s/2-1}\, dt,$$

where we denote $\varphi(t) = \frac{1}{2}(\vartheta(t)-1)$. Next, the idea is to use the functional equation for $\vartheta(t)$ to bring this to a new form that can be seen to be well-defined for all $s \in \mathbb{C}$ except $s = 1$. Specifically, we note that the functional equation for can be expressed in the form

$$\varphi(t) = t^{-1/2}\varphi(1/t) + \tfrac{1}{2}t^{-1/2} - \tfrac{1}{2}.$$

We can therefore write

$$\begin{aligned}\zeta^*(s) &= \int_0^1 \varphi(t) t^{s/2-1}\, dt + \int_1^\infty \varphi(t) t^{s/2-1}\, dt \\ &= \int_0^1 \left(t^{-1/2}\varphi(1/t) + \tfrac{1}{2}t^{-1/2} - \tfrac{1}{2}\right) t^{s/2-1}\, dt + \int_1^\infty \varphi(t) t^{s/2-1}\, dt \\ &= -\frac{1}{1-s} - \frac{1}{s} + \int_1^\infty \left(t^{-s/2-1/2} + t^{s/2-1}\right)\varphi(t)\, dt.\end{aligned}$$

We have derived a formula for $\zeta^*(s)$ (one of the formulas claimed in the main theorem above) that is now seen to define a meromorphic function on all of $\mathbb{C}$ — the integrand decays rapidly as $t \to \infty$ so actually defines an entire function, so the only poles are due to the two terms $-1/s$ and $1/(s-1)$. We have therefore proved that $\zeta(s)$ can be analytically continued to a meromorphic function on $\mathbb{C}$. □

Corollary 13. *The zeta function satisfies the functional equation*

$$\zeta^*(1-s) = \zeta^*(s).$$

Equivalently, because of the reflection formula satisfied by the gamma function, it is easy to check that the functional equation can be rewritten in the form

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

Proof. The representation we derived for $\zeta^*(s)$ is manifestly symmetric with respect to replacing each occurrence of s by $1-s$. □

Corollary 14. *The only pole of $\zeta(s)$ is a simple pole at $s = 1$ with residue* 1.

Proof. Our representation for $\zeta^*(s)$ expresses it as a sum of $-\frac{1}{s}$, $\frac{1}{s-1}$, and an entire function. Thus the poles of $\zeta^*(s)$ are simple poles at $s = 0, 1$ with residues -1 and 1, respectively. It follows that

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

has a pole at $s = 1$ with residue $\pi^{1/2}\Gamma(1/2)^{-1} = 1$, and a pole (that turns out to be a removable singularity) at $s = 0$ with residue $\pi^0\Gamma(0)^{-1} = 0$. (That is, the pole of $\zeta^*(s)$ at $s = 0$ is cancelled out by the zero of $\Gamma(s/2)$.) $\square$

Corollary 15. $\zeta(-n) = -B_{n+1}/(n+1)$ *for* $n = 1, 2, 3, \ldots$.

Proof. Using the functional equation, we have that

$$\begin{aligned}\zeta(-n) &= 2^{-n}\pi^{-n-1}\sin(-\pi n/2)\Gamma(n+1)\zeta(n+1)\\ &= 2^{-n}\pi^{-n-1}\sin(-\pi n/2)n!\zeta(n+1).\end{aligned}$$

If $n = 2k$ is even, then $\sin(-\pi n/2) = \sin(-\pi k) = 0$, so we get that $\zeta(-2k) = 0$ (that is, $n = 2k$ is one of the so-called "trivial zeros"). We also know that $B_{2k+1} = 0$ for $k = 1, 2, 3, \ldots$, so the formula $\zeta(-n) = B_{n+1}/(n+1)$ is satisfied in this case.

If on the other hand $n = 2k-1$ is odd, then $\sin(-\pi(2k-1)/2) = (-1)^k$, and therefore we get, using the formula expressing $\zeta(2k)$ in terms of the Bernoulli numbers (derived in the homework and in the textbook), that

$$\begin{aligned}\zeta(-n) &= (-1)^k 2^{-2k+1}\pi^{-2k}(2k-1)!\zeta(2k)\\ &= (-1)^k 2^{-2k+1}\pi^{-2k}(2k-1)!\frac{(-1)^{k-1}(2\pi)^{2k}}{2(2k)!}B_{2k}\\ &= -\frac{B_{2k}}{2k} = -\frac{B_{n+1}}{n+1},\end{aligned}$$

so again the formula is satisfied. $\square$

Corollary 16. *The zeros of $\zeta(s)$ in the region* $\mathrm{Re}(s) < 0$ *are precisely the trivial zeros* $s = -2, -4, -6, \ldots$.

Proof. We already established the existence of the trivial zeros. The fact that there are no other zeros also follows easily from the functional equation and is left as an exercise. $\square$

Remark 12 (alternative approaches to the analytic continuation of $\zeta(s)$)**.** There is a more "down-to-earth" approach to the analytic continuation of $\zeta(s)$ based on the standard idea from numerical analysis of approximating an integral by a sum (or in this case going in the other direction, approximating a sum by an integral). The technical name for this procedure, when it is done in a more systematic way, is

Euler-Maclaurin summation.

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

This representation is certainly valid for $\mathrm{Re}(s) > 1$. However, note that we have the bound

$$\left| x^{-s} - \lfloor x \rfloor^{-s} \right| \le |s| \cdot \lfloor x \rfloor^{-\mathrm{Re}(s)-1} \qquad (x \ge 1)$$

by the mean value theorem. Thus, the integral is actually an absolutely convergent integral in the larger region $\mathrm{Re}(s) > 0$, and the representation we derived gives an analytic continuation of $\zeta(s)$ to a meromorphic function on $\mathrm{Re}(s) > 0$, which has a single pole at $s = 1$ (a simple pole with residue 1) and is holomorphic everywhere else.

An elaboration of this idea using what is known as the Euler-Maclaurin summation formula can be used to perform the analytic continuation of $\zeta(s)$ to a meromorphic function on $\mathbb{C}$ by extending it inductively from each region $\mathrm{Re}(s) > -n$ to $\mathrm{Re}(s) > -n-1$, as we saw could be done for the gamma function. Another approach is to use the analytic continuation for $\mathrm{Re}(s) > 0$ shown above, then prove that the functional equation $\zeta^(1-s) = \zeta^*(s)$ holds in the region $0 < \mathrm{Re}(s) < 1$, and then use the functional equation to analytically continue $\zeta(s)$ to $\mathrm{Re}(s) \le 0$ (which is the reflection of the region $\mathrm{Re}(s) \ge 1$ under the transformation $s \mapsto 1-s$).

15.2 A theorem on the zeros of the Riemann zeta function

Next, we prove a nontrivial and very important fact about the zeta function that will play a critical role in our proof of the prime number theorem.

Theorem 36. *$\zeta(s)$ has no zeros on the line* $\mathrm{Re}(s) = 1$.

This theorem can also be thought of as a "toy" version of the Riemann hypothesis. If you ever want to try solving this famous open problem, getting a good understanding of its toy version seems like a good idea...

Proof. For this proof, denote $s = \sigma + it$, where we assume $\sigma > 1$ and t is real and nonzero. The proof is based on investigating simultaneously the behavior of $\zeta(\sigma + it)$, $\zeta(\sigma + 2it)$, and $\zeta(\sigma)$, for fixed t as $\sigma \searrow 1$. Consider the following somewhat mysterious quantity

$$X = \log \left| \zeta(\sigma)^3 \zeta(\sigma + it)^4 \zeta(\sigma + 2it) \right|.$$

We can evaluate “X” as

$$
\begin{aligned}
&\log|\zeta(\sigma)^3\zeta(\sigma+it)^4\zeta(\sigma+2it)| \\
&\quad = 3\log|\zeta(\sigma)| + 4\log|\zeta(\sigma+it)| + \log|\zeta(\sigma+2it)| \\
&\quad = 3\log\left(\prod_p |1-p^{-\sigma}|^{-1}\right) + 4\log\left(\prod_p |1-p^{-\sigma-it}|^{-1}\right) \\
&\qquad\quad + \log\left(\prod_p |1-p^{-\sigma-2it}|^{-1}\right) \\
&\quad = \sum_p \Big(-3\log|1-p^{-\sigma}| - 4\log|1-p^{-\sigma-it}| - \log|1-p^{-\sigma-2it}|\Big) \\
&\quad = \sum_p \Big(-3\,\mathrm{Re}\left[\mathrm{Log}(1-p^{-\sigma})\right] - 4\,\mathrm{Re}\left[\mathrm{Log}(1-p^{-\sigma-it})\right] \\
&\qquad\qquad\qquad - \mathrm{Re}\,\mathrm{Log}\left[1-p^{-\sigma-2it}\right]\Big),
\end{aligned}
$$

where $\mathrm{Log}(\cdot)$ denotes the principal branch of the logarithm function. Now note that for $z = a + ib$ with $a > 1$ and p prime we have $|p^{-z}| = p^{-a} < 1$, so

$$
-\mathrm{Log}(1-p^{-z}) = \sum_{m=1}^{\infty} \frac{p^{-mz}}{m},
$$

and

$$
\begin{aligned}
-\mathrm{Re}\Big[\mathrm{Log}(1-p^{-z})\Big] &= \sum_{m=1}^{\infty} \frac{p^{-ma}}{m}\,\mathrm{Re}\Big[\cos(mb\log p) + i\sin(mb\log p)\Big] \\
&= \sum_{m=1}^{\infty} \frac{p^{-ma}}{m}\cos(mb\log p).
\end{aligned}
$$

So we can rewrite X as

$$
X = \sum_{n=1}^{\infty} c_n n^{-\sigma}(3 + 4\cos\theta_n + \cos(2\theta_n))
$$

where $\theta_n = t\log n$ and $c_n = 1/m$ if $n = p^m$ for some prime p. We can now use a simple trigonometric identity

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

to rewrite X yet again as

$$
X = 2\sum_{n=1}^{\infty} c_n n^{-\sigma}(1+\cos\theta_n)^2.
$$

We have proved a crucial fact, namely that $X \geq 0$, or equivalently that

$$
e^X = |\zeta(\sigma)^3\zeta(\sigma+it)^4\zeta(\sigma+2it)| \geq 1.
$$

We now claim that this innocent-looking inequality is incompatible with the existence of a zero of $\zeta(s)$ on the line $\mathrm{Re}(s) = 1$. Indeed, assume by contradiction that $\zeta(1 + it) = 0$ for some real $t \neq 0$. Then the three quantities $\zeta(\sigma)$, $\zeta(\sigma + it)$ and $\zeta(\sigma + 2it)$ have the following asymptotic behavior as $\sigma \searrow 1$:

$$\begin{aligned}
|\zeta(\sigma)| &= \frac{1}{\sigma - 1} + O(1) && \text{(since } \zeta(s) \text{ has a pole at } s = 1\text{)}, \\
|\zeta(\sigma + it)| &= O(\sigma - 1) && \text{(since } \zeta(s) \text{ has a zero at } s = 1 + it\text{)}, \\
|\zeta(\sigma + 2it)| &= O(1) && \text{(since } \zeta(s) \text{ is holomorphic at } s = 1 + 2it\text{)}.
\end{aligned}$$

Combining these results we have that

$$e^X = |\zeta(\sigma)^3 \zeta(\sigma + it)^4 \zeta(\sigma + 2it)| = O((\sigma - 1)^{-3}(\sigma - 1)^4) = O(\sigma - 1).$$

In particular, $e^X \to 0$ as $\sigma \searrow 1$, in contradiction to the result we proved above that $e^X \geq 1$. This proves the claim that $\zeta(s)$ cannot have a zero on the line $\mathrm{Re}(s) = 1$. $\square$

Exercise 16. The above proof that $e^X \geq 1$ (which immediately implied the claim of the theorem) relied on showing that for any prime number p, the corresponding factors in the Euler product formula satisfy the inequality

$$(1 - p^{-\sigma})^{-3} |1 - p^{-\sigma} p^{-it}|^{-4} |1 - p^{-\sigma} p^{-2it}|^{-1} \geq 1,$$

and this was proved by taking the logarithm of the left hand-side, expanding in a power series and using the elementary trigonometric identity $3 + 4\cos\theta + \cos 2\theta = 2(1 + \cos\theta)^2$. However, one can imagine a more direct approach that starts as follows: denote $x = p^{-\sigma}$ and $z = p^{-it} = e^{-it\log p}$. Then the inequality reduces to the claim that

$$(1 - x)^3 |1 - zx|^4 |1 - z^2 x| \leq 1$$

for all $x \in [0, 1]$ and z satisfying $|z| = 1$. Since this is an elementary inequality, it seems like it ought to have an elementary proof (i.e., a proof that does not involve logarithms and power series expansions). Can you find such a proof?

16 The prime number theorem

The prime number theorem was proved in 1896 by Jacques Hadamard and independently by Charles Jean de la Vallée Poussin, using the groundbreaking ideas from Riemann's famous 1859 paper in which he introduced the use of the Riemann zeta function as a tool for counting prime numbers. (This was the only number theory paper Riemann wrote in his career!) The history (including all the technical details) of these developments is described extremely well in the classic textbook [4], which I highly recommend.

The original proofs of the prime number theorem were very complicated and relied on the "explicit formula of number theory" (that I mentioned in the previous section) and some of its variants. Throughout the 20th century, mathematicians worked hard to find simpler ways to derive the prime number theorem. This resulted in several important developments (such as the Wiener tauberian theorem and the Hardy-Littlewood tauberian theorem) that advanced not just the state of analytic number theory but also complex analysis, harmonic analysis and functional analysis. Despite all the efforts and the discovery of several paths to a proof that were simpler than the original approach, all proofs remained quite difficult... until the year 1980, when the mathematician Donald Newman discovered a wonderfully simple way to derive the theorem using a completely elementary use of complex analysis. It is Newman's proof (as presented in the short paper [13] by D. Zagier) that I present here.

Define the weighted prime counting functions

$$\pi(x) = \#\{p \text{ prime} : p \le x\} = \sum_{p \le x} 1,$$
$$\psi(x) = \sum_{p^k \le x} \log p = \sum_{p \le x} \log p \left\lfloor \frac{\log x}{\log p} \right\rfloor,$$

with the convention that the symbol p in a summation denotes a prime number, and p^k denotes a prime power, so that summation over $p \le x$ denotes summation over all primes $\le x$, and the summation over p^k denotes summation over all prime powers $\le x$. Another customary way to write the function $\psi(x)$ is as

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

where the function $\Lambda(n)$, called the von Mangoldt function, is defined by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k,\ p \text{ prime}, \\ 0 & \text{otherwise}. \end{cases}$$

Lemma 11. *The prime number theorem $\pi(x) \sim \frac{x}{\log x}$ is equivalent to the statement that $\psi(x) \sim x$.*

Proof. Note the inequality

$$\psi(x) = \sum_{p \le x} \log p \left\lfloor \frac{\log x}{\log p} \right\rfloor \le \sum_{p \le x} \log p \frac{\log x}{\log p} = \sum_{p \le x} \log x = \log x \cdot \pi(x).$$

In the opposite direction, we have a similar (but slightly less elegant) inequality, namely that for any $0 < \epsilon < 1$ and $x \ge 2$,

$$\begin{aligned} \psi(x) &\ge \sum_{p \le x} \log p \ge \sum_{x^{1-\epsilon} < p \le x} \log p \ge \sum_{x^{1-\epsilon} < p \le x} \log\left(x^{1-\epsilon}\right) \\ &= (1-\epsilon)\log x \Big(\pi(x) - \pi(x^{1-\epsilon})\Big) \ge (1-\epsilon)\log x \Big(\pi(x) - x^{1-\epsilon}\Big). \end{aligned}$$

Now assume that $\psi(x) \sim x$ as $x \to \infty$. Then the first of the two bounds above implies that

$$\pi(x) \ge \frac{\psi(x)}{\log x},$$

so

$$\liminf_{x \to \infty} \pi(x) / \left(\frac{x}{\log x}\right) \ge 1.$$

On the other hand, the second of the two bounds implies that

$$\pi(x) \le \frac{1}{1-\epsilon} \cdot \frac{\psi(x)}{\log x} + x^{1-\epsilon},$$

which implies that $\limsup_{x \to \infty} \pi(x) / \left(\frac{x}{\log x}\right) \le \frac{1}{1-\epsilon} + \limsup_{x \to \infty} \frac{\log x}{x^{\epsilon}} = \frac{1}{1-\epsilon}$. Since ϵ was an arbitrary number in $(0,1)$, it follows that

$$\limsup_{x \to \infty} \pi(x) / \left(\frac{x}{\log x}\right) \le 1.$$

Combining the two results about the lim inf and the lim sup gives that $\pi(x) \sim x/\log x$.

Now assume that $\pi(x) \sim \frac{x}{\log x}$, and apply the inequalities we derived above in the opposite direction from before. That is, we have

$$\psi(x) \le \log x \cdot \pi(x),$$

so

$$\limsup_{x\to\infty} \psi(x)/x \le 1.$$

On the other hand,

$$\psi(x) \ge (1-\epsilon)\log x(\pi(x) - x^{1-\epsilon})$$

implies that

$$\liminf_{x\to\infty} \psi(x)/x \ge \lim_{x\to\infty} (1-\epsilon)\left(1 - \frac{\log x}{x^\epsilon}\right) = 1-\epsilon.$$

Again, since $\epsilon \in (0,1)$ was arbitrary, it follows that $\liminf_{x\to\infty} \frac{\psi(x)}{x} = 1$. Combining the two results about the lim inf and lim sup proves that $\lim_{x\to\infty} \frac{\psi(x)}{x} = 1$, as claimed. □

Lemma 12. *For* $\mathrm{Re}(s) > 1$ *we have*

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

Proof. Using the Euler product formula and taking the logarithmic derivative (which is an operation that works as it should when applied to infinite products of holomorphic functions that are uniformly convergent on compact subsets), we have

$$\begin{aligned} -\frac{\zeta'(s)}{\zeta(s)} &= \sum_p \frac{\frac{d}{ds}(1-p^{-s})}{1-p^{-s}} = \sum_p \frac{\log p \cdot p^{-s}}{1-p^{-s}} \\ &= \sum_p \log p\,(p^{-s} + p^{-2s} + p^{-3s} + \ldots) = \sum_{p \text{ prime}} \sum_{k=1}^\infty \log p \cdot p^{-ks} \\ &= \sum_{n=1}^\infty \Lambda(n) n^{-s}. \end{aligned}$$

□

Lemma 13. *There is a constant* $C > 0$ *such that* $\psi(x) < Cx$ *for all* $x \ge 1$.

Proof. The idea of the proof is that the binomial coefficient $\binom{2n}{n}$ is not too large on the one hand, but is divisible by many primes (all primes between n and $2n$) on the other hand — hence it follows that there cannot be too many primes, and in particular the weighted prime-counting function $\psi(x)$ can be easily bounded from above using such an argument. Specifically, we have that

$$\begin{aligned} 2^{2n} = (1+1)^{2n} &= \sum_{k=0}^{2n} \binom{2n}{k} > \binom{2n}{n} \ge \prod_{n<p\le 2n} p = \exp\left(\sum_{n<p\le 2n} \log p\right) \\ &= \exp\Big(\psi(2n) - \psi(n) - \sum_{n<p^k\le 2n,\ k>1} \log p\Big). \\ &\ge \exp\Big(\psi(2n) - \psi(n) - O(\sqrt{n}\log^2 n)\Big). \end{aligned}$$

(The estimate $O(\sqrt{n}\log^2 n)$ for the sum of $\log p$ for prime powers higher than 1 is easy and is left as an exercise.) Taking the logarithm of both sides, this gives the bound

$$\psi(2n) - \psi(n) \le 2n\log 2 + C_1\sqrt{n}\log n \le C_2 n,$$

valid for all $n \ge 1$ with some constant $C_2 > 0$. It follows that

$$\begin{aligned}\psi(2^m) &= (\psi(2^m) - \psi(2^{m-1})) \\ &\qquad + (\psi(2^{m-1}) - \psi(2^{m-2})) + \ldots + (\psi(2^1) - \psi(2^0)) \\ &\le C_2(2^{m-1} + \ldots + 2^0) \le C_2 2^m,\end{aligned}$$

so the inequality $\psi(x) \le C_2 x$ is satisfied for $x = 2^m$. It is now easy to see that this implies the result also for general x, since for $x = 2^m + \ell$ with $0 \le \ell < 2^m$ we have

$$\psi(x) = \psi(2^m + \ell) \le \psi(2^{m+1}) \le C_2 2^{m+1} \le 2C_2 2^m \le 2C_2 x.$$

□

Theorem 37 (Newman's tauberian theorem)**.** *Let $f : [0,\infty) \to \mathbb{R}$ be a bounded function that is integrable on compact intervals. Define a function $g(z)$ of a complex variable z by*

$$g(z) = \int_0^\infty f(t)e^{-zt}\,dt$$

(g is known as the Laplace transform *of f). Clearly $g(z)$ is defined and holomorphic in the open half-plane* $\mathrm{Re}(z) > 0$*. Assume that $g(z)$ has an analytic continuation to an open region Ω containing the* closed *half-plane* $\mathrm{Re}(z) \ge 0$*. Then $\int_0^\infty f(t)\,dt$ exists and is equal to $g(0)$ (the value at $z = 0$ of the analytic continuation of g).*

Proof. Define a truncated version of the integral defining $g(z)$, namely

$$g_T(z) = \int_0^T f(t)e^{-zt}\,dt$$

for $T > 0$, which for any T is an entire function of z. Our goal is to show that $\lim_{T\to\infty} g_T(0) = g(0)$. This can be achieved using a clever application of Cauchy's integral formula. Fix some large $R > 0$ and a small $\delta > 0$ (which depends on R in a way that will be explained shortly), and consider the contour C consisting of the part of the circle $|z| = R$ that lies in the half-plane $\mathrm{Re}(z) \ge -\delta$, together with the straight line segment along the line $\mathrm{Re}(z) = -\delta$ connecting the top and bottom intersection points of this circle with the line (see Fig. 6(a)). Assume that δ is small enough so that $g(z)$ (which extends analytically at least slightly to the right of $\mathrm{Re}(z) = 0$) is holomorphic in an open set containing C and the region enclosed by it. Then by Cauchy's integral formula we have

$$\begin{aligned}g(0) - g_T(0) &= \frac{1}{2\pi i}\int_C (g(z) - g_T(z))e^{Tz}\left(1 + \frac{z^2}{R^2}\right)\frac{dz}{z} \\ &= \frac{1}{2\pi i}\left(\int_{C_+} + \int_{C_-}\right)(g(z) - g_T(z))e^{Tz}\left(1 + \frac{z^2}{R^2}\right)\frac{dz}{z},\end{aligned}$$

where we separate the contour into two parts, a semicircular arc C_+ that lies in the half-plane $\mathrm{Re}(z) > 0$,

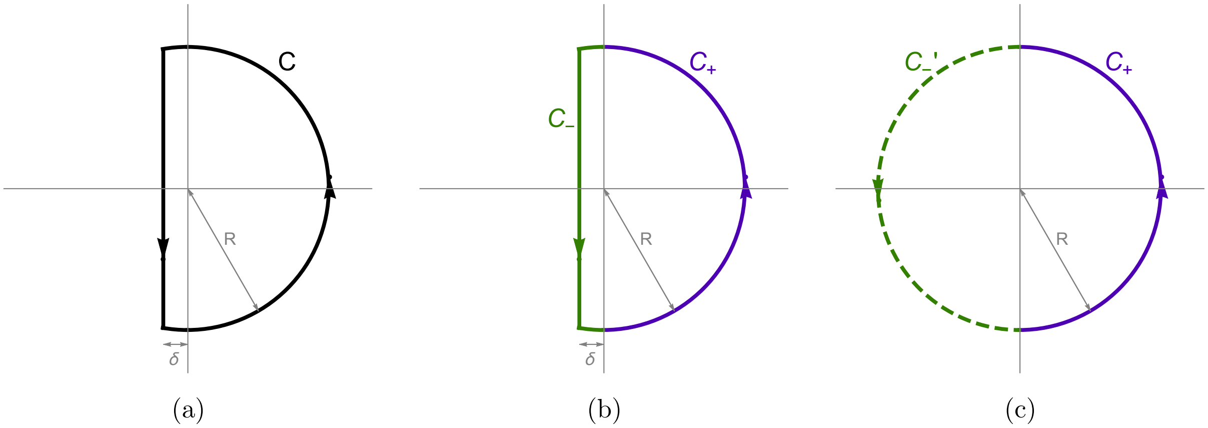


Figure 6: The contours C, C_+, C_- and C'_-.

and the remaining part C_- in the half-plane $\mathrm{Re}(z) < 0$ (Fig. 6(b)). We now bound the integral separately on C_+ and on C_-. First, for z lying on C_+ we have

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

where $B = \sup_{t\ge 0} |f(t)|$, and

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

(by the trivial identity $|1 + e^{it}|^2 = |e^{it}(e^{it} + e^{it})|^2 = 2\cos(t)$, valid for $t \in \mathbb{R}$). So in combination we have

$$\left| \frac{1}{2\pi i} \int_{C_+} (g(z) - g_T(z))e^{Tz}\left(1 + \frac{z^2}{R^2}\right)\frac{dz}{z} \right| \le (\pi R)\frac{2B}{2\pi R^2} = \frac{B}{R}.$$

Next, for C_-, we bound the integral by bounding the contributions from $g(z)$ and $g_T(z)$ separately. In the case of $g_T(z)$, the function is entire, so we can deform the contour, replacing it with the semicircular arc $C'_- = \{|z| = R, \mathrm{Re}(z) < 0\}$ (Fig. 6(c)). On this contour we have the estimate

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

which leads using a similar calculation as before to the estimate

$$\frac{1}{2\pi i} \int_{C'_-} \left| g_T(z)e^{Tz}\left(1 + \frac{z^2}{R^2}\right) \right| \frac{|dz|}{|z|} \le \frac{B}{R}.$$

The remaining integral

$$\frac{1}{2\pi i} \int_{C_-} \left| g(z)e^{Tz}\left(1 + \frac{z^2}{R^2}\right) \right| \frac{|dz|}{|z|}$$

tends to 0 as $T \to \infty$, since the dependence on T is only through the factor e^{Tz}, which converges to 0 uniformly on compact sets in $\mathrm{Re}(z) < 0$ as $T \to \infty$.

Combining the above estimates, we have shown that

$$\limsup_{T\to\infty} |g(0) - g_T(0)| \le \frac{2B}{R}.$$

Since R was an arbitrary positive number, the lim sup must be 0, and the theorem is proved. □

Consider now a very specific application of Newman's theorem: take

$$f(t) = \psi(e^t)e^{-t} - 1 \qquad (t \ge 0),$$

which is bounded by the lemma we proved above, as our function $f(t)$. The associated function $g(z)$ is then

$$\begin{aligned} g(z) &= \int_0^\infty (\psi(e^t)e^{-t} - 1)e^{-zt}\,dt = \int_1^\infty \left(\frac{\psi(x)}{x} - 1\right) x^{-z-1}\,dx \\ &= \int_1^\infty \psi(x)x^{-z-2}\,dx - \frac{1}{z} = \int_1^\infty \left(\sum_{n\le x} \Lambda(n)\right) x^{-z-2}\,dx - \frac{1}{z} \\ &= \sum_{n=1}^\infty \Lambda(n) \left(\int_n^\infty x^{-z-2}\,dx\right) - \frac{1}{z} = \sum_{n=1}^\infty \Lambda(n)\ \frac{x^{-z-1}}{-z-1}\bigg|_n^\infty - \frac{1}{z} \\ &= \frac{1}{z+1}\sum_{n=1}^\infty \Lambda(n)n^{-z-1} - \frac{1}{z} = -\frac{1}{z+1}\cdot\frac{\zeta'(z+1)}{\zeta(z+1)} - \frac{1}{z} \qquad (\mathrm{Re}(z) > 0). \end{aligned}$$

Recall that $-\zeta'(s)/\zeta(s)$ has a simple pole at $s = 1$ with residue 1 (because $\zeta(s)$ has a simple pole at $s = 1$; it is useful to remember the more general fact that if a holomorphic function $h(z)$ has a zero of order k at $z = z_0$ then the logarithmic derivative $h'(z)/h(z)$ has a simple pole at $z = z_0$ with residue k). So $-\frac{1}{z+1}\cdot\frac{\zeta'(z+1)}{\zeta(z+1)}$ has a simple pole with residue 1 at $z = 0$, and therefore $-\frac{1}{z+1}\cdot\frac{\zeta'(z+1)}{\zeta(z+1)} - \frac{1}{z}$ has a *removable* singularity at $z = 0$. Thus, the identity $g(z) = -\frac{1}{z+1}\cdot\frac{\zeta'(z+1)}{\zeta(z+1)} - \frac{1}{z}$ shows that $g(z)$ extends analytically to a holomorphic function in the set

$$\{z \in \mathbb{C} : \zeta(z+1) \ne 0\}.$$

By the "toy Riemann Hypothesis" — the theorem we proved according to which $\zeta(s)$ has no zeros on the line $\mathrm{Re}(s) = 1$, $g(z)$ in particular extends holomorphically to an open set containing the half-plane $\mathrm{Re}(z) \ge 0$. Thus, $f(t)$ satisfies the assumption of Newman's theorem. We conclude from the theorem that the integral

$$\begin{aligned} \int_0^\infty f(t)\,dt &= \int_0^\infty (\psi(e^t)e^{-t} - 1)dt = \int_1^\infty \left(\frac{\psi(x)}{x} - 1\right)\frac{dx}{x} \\ &= \int_1^\infty \frac{\psi(x) - x}{x^2}\,dx \end{aligned}$$

converges.

Proof of the prime number theorem. We will prove that $\psi(x) \sim x$, which we already showed is equivalent to the prime number theorem. Assume by contradiction that $\limsup_{x\to\infty} \frac{\psi(x)}{x} > 1$ or $\liminf_{x\to\infty} \frac{\psi(x)}{x} < 1$. In the first case, that means there exists a number $\lambda > 1$ such that $\psi(x) \ge \lambda x$ for arbitrarily large x. For such values of x it then follows that

$$\int_x^{\lambda x} \frac{\psi(t) - t}{t^2}\,dt \ge \int_x^{\lambda x} \frac{\lambda x - t}{t^2}\,dt = \int_1^\lambda \frac{\lambda - t}{t^2}\,dt =: A > 0,$$

but this is inconsistent with the fact that the integral $\int_1^\infty (\psi(x) - x)x^{-2}\,dx$ converges.

Similarly, in the event that $\liminf_{x\to\infty} \frac{\psi(x)}{x} < 1$, that means that there exists a $\mu < 1$ such that $\psi(x) \le \mu x$ for arbitrarily large x, in which case we have that

$$\int_{\lambda x}^{x} \frac{\psi(t) - t}{t^2}\,dt \le \int_{\lambda x}^{x} \frac{\lambda x - t}{t^2}\,dt = \int_{\lambda}^{1} \frac{\lambda - t}{t^2}\,dt =: B < 0,$$

again giving a contradiction to the convergence of the integral. □

17 Introduction to asymptotic analysis

In this section we'll learn how to use complex analysis to prove asymptotic formulas such as

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \qquad \text{(Stirling's formula)}, \tag{2}$$

$$p(n) \sim \frac{1}{4\sqrt{3}n} e^{\pi\sqrt{2n/3}} \qquad \text{(the Hardy-Ramanujan formula)}, \tag{3}$$

$$\mathrm{Ai}(x) \sim \frac{1}{2\sqrt{\pi}} x^{-1/4} \exp\left(-\frac{2}{3}x^{3/2}\right) \qquad \text{(asymptotics for the Airy function)}, \tag{4}$$

and more. At the heart of many such results is an important technique known as the *saddle point method*. Some related techniques (that are all minor variations on the same theme) are *Laplace's method*, the *steepest descent method* and the *stationary phase method*.

17.1 First example: Stirling's formula

Our goal in this subsection is to prove a version of Stirling's approximation (2) for the factorial function $n!$. Let us start by simply having a bit of seemingly aimless fun and asking ourselves, what are some really easy things we can say about the magnitude of $n!$ for large n?[3] An obvious upper bound is $n! \le n^n$. As for a lower bound, one can say similarly trivial things such as $n! \ge (n/2)^{n/2}$, but that is quite far from the upper bound. We can do a bit better by making the simple observation that for any real number $x > 0$, we have the relations

$$\frac{x^n}{n!} \le \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x,$$

and this gives a lower bound

$$n! \ge e^{-x}x^n.$$

It now makes sense to try to get the best lower bound possible by looking for the x where the lower-bounding function is maximal. This happens when

$$0 = \frac{d}{dx}\left(e^{-x}x^n\right) = e^{-x}\left(-x^n + nx^{n-1}\right) = e^{-x}x^{n-1}(-x+n),$$

i.e., when $x = n$. Plugging this value into the inequality gives the bound

$$n! \ge (n/e)^n \qquad (n \ge 1).$$

[3] As a general rule of problem-solving, it's often helpful to start attacking a problem by thinking about really easy things you can say about it before moving on to advanced techniques. It's a great way to develop your intuition, and sometimes you discover that the easy techniques solve the problem outright.

This is of course a standard and very easy result, but it's intriguing to note that it's brought us quite close to the "true" asymptotics given by (2). The point of this trivial calculation is that, as we shall see below, there is something special about the value $x = n$ that resulted from this maximization operation; when interpreted in the context of complex analysis, it corresponds to a so-called "saddle point," since it is a local minimum of e^x/x^n as one moves along the real axis, but it will be a local maximum when one moves in the orthogonal direction parallel to the imaginary axis. Thus, the trivial approach has revealed the kernel of a deeper, much more powerful one.

Now let's move on to the more powerful approach. Our "toy" approach was based purely on real analysis, making use of the Taylor expansion of the function $x \mapsto e^x$; it turns out we could do better by thinking complex-analytically. Start with the power series expansion

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

As we know very well from our study of Cauchy's integral formula and the residue theorem, the nth Taylor coefficient can be extracted from the function by contour integration, that is, by writing

$$\frac{1}{n!} = \frac{1}{2\pi i} \oint_{|z|=r} \frac{e^z}{z^{n+1}} dz,$$

where the radius r of the circle chosen as the contour of integration is an arbitrary positive number. It turns out that some values of r are better than others when one is trying to do asymptotics. We select $r = n$ (I'll explain later where that seemingly inspired choice comes from), to get

$$\begin{aligned}
\frac{1}{n!} &= \frac{1}{2\pi i} \oint_{|z|=n} \frac{e^z}{z^{n+1}} dz = \frac{1}{2\pi i} \int_{-\pi}^{\pi} \exp\left(ne^{it}\right) n^{-n} e^{-int} i\, dt \\
&= \frac{1}{2\pi\, n^n} \int_{-\pi}^{\pi} \exp\left(n(e^{it} - it)\right)\, dt \\
&= \frac{e^n}{2\pi\, n^n} \int_{-\pi}^{\pi} \exp\left(n(e^{it} - 1 - it)\right)\, dt,
\end{aligned}$$

where we have strategically massaged the integrand (by pulling out the factor e^n) to cancel out a term in the Taylor expansion of e^{it}, in addition to a term that was already canceled out. For convenience, rewrite this as

$$\frac{n^n}{e^n n!} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(n(e^{it} - 1 - it)\right)\, dt.$$

Now noting that

$$n\left(e^{it} - 1 - it\right) = -\frac{nt^2}{2} + O(nt^3) = \frac{(\sqrt{n}t)^2}{2} + O\left(\frac{(\sqrt{n}t)^3}{\sqrt{n}}\right),$$

for $|t|$ small, we see that a change of variable $u = \sqrt{n}t$ in the integral will enable us to rewrite this as

$$n\left(e^{iu/\sqrt{n}} - 1 - \frac{iu}{\sqrt{n}}\right) = -\frac{u^2}{2} + O\left(\frac{u^3}{\sqrt{n}}\right).$$

Performing the change of variable and moving a factor of $\sqrt{n}$ to the left-hand side, the integral then becomes

$$\frac{\sqrt{n}n^n}{e^n n!} = \frac{1}{2\pi} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \exp\left(n\left(e^{iu/\sqrt{n}} - 1 - \frac{iu}{\sqrt{n}}\right)\right) du.$$

The integrand converges pointwise to $e^{-u^2/2}$ (for u fixed and $n \to \infty$), so it's reasonable to guess that the integral should converge to $\int_{-\infty}^{\infty} e^{-u^2/2}\, du = \sqrt{2\pi}$, which would lead to the formula

$$\frac{\sqrt{n}n^n}{e^n n!} \approx \frac{1}{\sqrt{2\pi}},$$

or

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,$$

which is precisely Stirling's formula. However, note that the $O(u^3/\sqrt{n})$ estimate holds whenever $t = u/\sqrt{n}$ is in a neighborhood of 0, and since u actually ranges in $[-\pi\sqrt{n}, \pi\sqrt{n}]$, we need to be more careful to get a precise asymptotic result. To proceed, it makes sense to divide the integral into two parts. Denote $M = n^{1/10}$, and let

$$\begin{aligned}
I &= \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \exp\left(n\left(e^{iu/\sqrt{n}} - 1 - \frac{iu}{\sqrt{n}}\right)\right) du = I_1 + I_2, \\
I_1 &= \int_{-M}^{M} \exp\left(n\left(e^{iu/\sqrt{n}} - 1 - \frac{iu}{\sqrt{n}}\right)\right) du, \\
I_2 &= \int_{[-\pi\sqrt{n},\pi\sqrt{n}]\setminus[-M,M]} \exp\left(n\left(e^{iu/\sqrt{n}} - 1 - \frac{iu}{\sqrt{n}}\right)\right) du.
\end{aligned}$$

We now estimate each of I_1 and I_2 separately. For I_1, we have

$$\begin{aligned}
I_1 &= \int_{-M}^{M} \exp\left(-\frac{u^2}{2} + O\left(\frac{u^3}{\sqrt{n}}\right)\right) du \\
&= \int_{-M}^{M} e^{-u^2/2} \exp\left(O\left(\frac{u^3}{\sqrt{n}}\right)\right) du \\
&= \int_{-M}^{M} \left(1 + O\left(\frac{u^3}{\sqrt{n}}\right)\right) e^{-u^2/2} du = \left(1 + O(n^{-1/5})\right) \int_{-M}^{M} e^{-u^2/2} du \\
&= \left(1 + O(n^{-1/5})\right) \left(\int_{-\infty}^{\infty} - 2\int_{M}^{\infty}\right) e^{-u^2/2} du \\
&= \left(1 + O(n^{-1/5})\right) \left(\sqrt{2\pi} - O\left(\exp\left(-n^{-1/5}\right)\right)\right) \\
&= \left(1 + O(n^{-1/5})\right) \sqrt{2\pi}.
\end{aligned}$$

For I_2, we have

$$\begin{aligned}
|I_2| &\le 2\int_{M}^{\pi\sqrt{n}} \left|\exp\left(n\left(e^{iu/\sqrt{n}} - 1 - \frac{iu}{\sqrt{n}}\right)\right)\right| du \\
&= 2\int_{M}^{\pi\sqrt{n}} \exp\left(n \operatorname{Re}\left(e^{iu/\sqrt{n}} - 1\right)\right) du \\
&= 2\int_{M}^{\pi\sqrt{n}} \exp\left[n\left(\cos\left(\frac{u}{\sqrt{n}}\right) - 1\right)\right] du
\end{aligned}$$

Now use the elementary fact that $\cos(t) \le 1 - t^2/8$ for $x \in [-\pi, \pi]$ (see Fig. 7) to infer further that

$$|I_2| \le 2\int_{M}^{\pi\sqrt{n}} \exp\left(-u^2/8\right) du \le 2\pi\sqrt{n}\exp\left(-n^{1/5}\right) = O(n^{-1/5}).$$

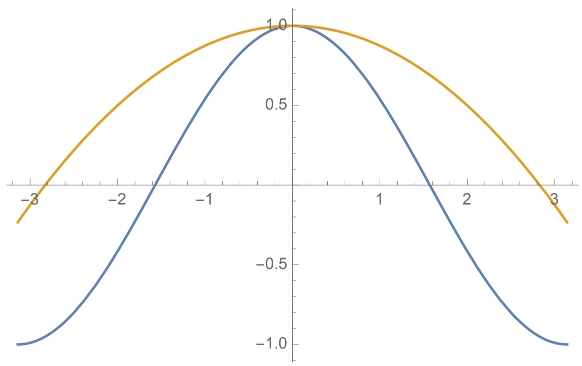


Figure 7: Illustration of the inequality $\cos(t) \le 1 - t^2/8$.

Combining the above results, we have proved the following version of Stirling's formula with a quantitative (though suboptimal) bound:

Theorem 38 (Stirling's approximation for $n!$)**.** *The asymptotic relation*

$$n! = \left(1 + O(n^{-1/5})\right) \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$$

holds as $n \to \infty$.

17.2 Second example: the central binomial coefficient

Let $a_n = \binom{2n}{n} = \frac{(2n)!}{(n!)^2}$. A standard way to find the asymptotic behavior for a_n as $n \to \infty$ is to use Stirling's formula. This easily gives that

$$\binom{2n}{n} = (1 + o(1)) \frac{4^n}{\sqrt{\pi n}}.$$

(Note that this is not too far from the trivial upper bound $\binom{2n}{n} \le (1+1)^{2n} = 2^{2n}$.) It is instructive to rederive this result using the saddle-point method, starting from the expansion

$$(1+z)^{2n} = \sum_{k=0}^{2n} \binom{2n}{k} z^n,$$

which in particular gives the contour integral representation

$$\binom{2n}{n} = \frac{1}{2\pi i} \oint_{|z|=r} \frac{(1+z)^{2n}}{z^{n+1}} dz.$$

By the same trivial method for deriving upper bounds that we used in the case of the Taylor coefficients $1/n!$ of the function e^z, we have that for each $x > 0$,

$$\binom{2n}{n} \le (1+x)^{2n}/x^n = \exp\left(\log(1+x) - n \log x\right).$$

We optimize over x by differentiating the expression $\log(1+x) - n\log x$ inside the exponent and setting the derivative equal to 0. This gives $x = 1$, the location of the saddle point. For this value of x, we again recover the trivial inequality $\binom{2n}{n} \le 2^{2n}$.

Next, equipped with the knowledge of the saddle point, we set $r = 1$ in the contour integral formula, to get

$$\begin{aligned}\binom{2n}{n} &= \frac{1}{2\pi i}\oint_{|z|=r} \frac{(1+z)^{2n}}{z^{n+1}}\,dz = \frac{1}{2\pi}\int_{-\pi}^{\pi} (1+e^{it})^{2n} e^{-int}\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(n\left(2\log(1+e^{it}) - t\right)\right)dt.\end{aligned}$$

Now note that the expression in the exponent has the Taylor expansion

$$n(2\log(1+e^{it}) - t) = 2\log 2 - \frac{1}{4}nt^2 + O(nt^4) \qquad \text{as } t \to 0.$$

Again, we see that a change of variables $u = t/\sqrt{n}$ will bring the integrand to an asymptotically scale-free form. More precisely, we have

$$\begin{aligned}\binom{2n}{n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(n\left(2\log 2 - \frac{1}{4}nt^2 + O(nt^4)\right)\right)dt \\ &= \frac{4^n}{2\pi\sqrt{n}}\int_{-\pi}^{\pi} \exp\left(-\frac{1}{4}u^2 + O\left(\frac{u^4}{n}\right)\right)du.\end{aligned}$$

It is now reasonable to guess that in the limit as $n \to \infty$, the pointwise limit of the integrands translates to a limit of the integrals, so that we get the approximation

$$\binom{2n}{n} \approx \frac{4^n}{2\pi\sqrt{n}}\int_{-\infty}^{\infty} e^{-u^2/4}\,du = \frac{4^n}{2\pi\sqrt{n}}2\sqrt{\pi} = \frac{4^n}{\sqrt{\pi n}},$$

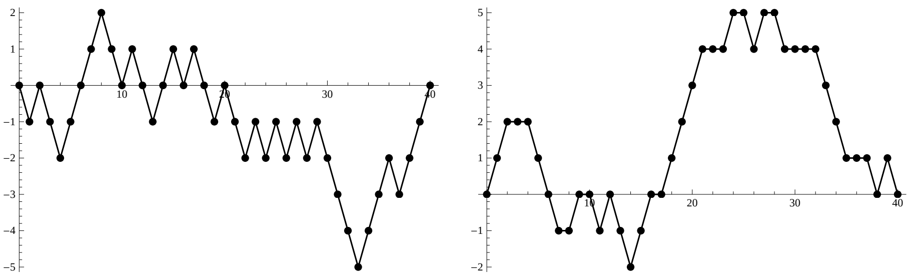
as required. Indeed, this is correct, but it remains to make this argument precise by breaking up the integral into two parts, a "central part" where the $O(u^4/n)$ error term can be shown to be small, and the remaining part that has to be bounded separately.

Exercise 17. Complete this analysis to give a rigorous proof using this method of the asymptotic formula $\binom{2n}{n} = (1+o(1))4^n/\sqrt{\pi n}$.

Exercise 18. Repeat this analysis for the sequence $(b_n)_{n=1}^{\infty}$ of central *trinomial* coefficients, where b_n is defined as the coefficient of x^n in the expansion of $(1+x+x^2)^n$, a definition that immediately gives rise to the contour integral representation

$$b_n = \frac{1}{2\pi i}\oint_{|z|=r} \frac{(1+z+z^2)^n}{z^{n+1}}\,dz.$$

Like their more famous cousins the central binomial coefficients, these coefficients are important in combinatorics and probability theory. Specifically, a_n and b_n correspond to the numbers of random walks on $\mathbb{Z}$ that start and end at 0 and have n steps, where in the case of the central binomial coefficients the allowed steps of the walk are -1 or $+1$, and in the case of the central trinomial coefficients the allowed steps are -1, 0 or 1; see Fig. 8.



(a) (b)

Figure 8: An illustration (with $n = 40$) of the random walks enumerated by (a) the central binomial coefficients and (b) the central trinomial coefficients.

Using a saddle point analysis, show that the asymptotic behavior of b_n as $n \to \infty$ is given by

$$b_n \sim \frac{\sqrt{3} \cdot 3^n}{\sqrt{\pi n}}.$$

17.3 A conceptual explanation

In both the examples of Stirling's formula and the central binomial coefficient we analyzed above, we made what looked like ad hoc choices regarding how to "massage" the integrals, what value r to use for the radius of the contour of integration, what change of variables to make in the integral, etc. Now let us think more conceptually and see if we can generalize these ideas. Note that the quantities we were trying to estimate took a particular form, where for some function $g(z)$ our sequence of numbers could be represented in the form

$$\begin{aligned} a(n) &= \frac{1}{2\pi i} \oint_{|z|=r} \frac{e^{-ng(z)}}{z^n} \frac{dz}{z} = \frac{1}{2\pi i} \oint_{|z|=r} \exp\Big(- n(g(z) + \log z)\Big) \frac{dz}{z} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ng(re^{it})} r^{-n} e^{-int}\, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\Big(- n(g(re^{it}) + it - \log r)\Big) dt. \end{aligned}$$

(Sometimes $g(z)$ would actually be $g_n(z)$, a sequence of functions that depends on n.) The idea that is key to making the method work turns out to be to choose the contour radius r as the solution to the equation

$$\frac{d}{dz} \left(g(z) + \log z\right) = g'(z) + \frac{1}{z} = 0.$$

This causes the first-order term in the Taylor expansion of $g(z) + \log z$ around $z = r$ to disappear. One is then left with a constant term, that can be pulled outside of the integral; a second order term, which (in favorable circumstances where this technique actually works) causes the integrand to be well-approximated by a Gaussian density function $e^{-u^2/2}$ near $z = r$; and lower-order terms which can be shown to be asymptotically negligible.

Geometrically, if one plots the graph of $|g(z) + 1/z|$ then one finds the emergence of a saddle point at $z = r$, and this is the origin of the term "saddle point method." This phenomenon is illustrated with many beautiful examples and graphical figures in the lecture slides [6] prepared by Flajolet and Sedgewick as an online resource to accompany their excellent textbook Analytic Combinatorics [5].

Exercise 19. It is instructive to see an example where the saddle point analysis *fails* if applied mindlessly without checking that the part of the integral that is usually assumed to make a negligible contribution actually behaves that way. A simple example illustrating what can go wrong is the function

$$f(z) = e^{z^2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{n!} = \sum_{n=0}^{\infty} b_n z^n,$$

where the Taylor coefficients are

$$b_n = \begin{cases} \frac{1}{(n/2)!} & n \text{ even}, \\ 0 & \text{otherwise}. \end{cases}$$

Clearly any analysis, asymptotic or not, needs to address and take into account the fact that b_n behaves differently according to whether n is even or odd. Try to apply the method we developed to derive an asymptotic formula for b_n. The method fails, but the failure can easily be turned into a success by noting that there are actually *two* saddle points, each of which makes a contribution to the integral, in such a way that for odd n the contributions cancel and for even n they reinforce each other. This shows that *periodicities* are one common pitfall to look out for when doing asymptotic analysis.

Exercise 20. As another amusing example, apply the saddle point method to the function $f(z) = 1/(1-z) = \sum_{n=0}^{\infty} d_n z^n$, for which the Taylor coefficients $d_n = 1$ are all equal to 1. Can you succeed in deriving an asymptotic formula for the constant function 1?

17.4 Third example: Stirling's formula for the gamma function.

Our next goal is to prove a stronger version of Stirling's formula that gives an asymptotic formula for $\Gamma(t)$, the extension of the factorial function to non-integer arguments. Specifically, we will prove.

Theorem 39 (Stirling's approximation for $\Gamma(t)$)**.** *For a real-valued argument t, the gamma function satisfies the asymptotic formula*

$$\Gamma(t) = \left(1 + O(t^{-1/5})\right) \sqrt{\frac{2\pi}{t}} \left(\frac{t}{e}\right)^t \qquad (t \to \infty).$$

Proof. We use a method called *Laplace's method*, which is a variant of the saddle-point method adapted to estimating real integrals instead of contour integrals around a circle. Start with the integral formula

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

Performing the change of variables $x = tu$ in the integral gives that

$$\begin{aligned} \Gamma(t) &= t^t \int_0^{\infty} e^{-tu} u^{t-1}\, du = t^t e^{-t} \int_0^{\infty} e^{-tu+t} u^{t-1}\, du \\ &= t^t e^{-t} \int_0^{\infty} e^{-tu+t} u^{t-1}\, du = t^t e^{-t} \int_0^{\infty} e^{-t\Phi(u)} \frac{du}{u} = \left(\frac{t}{e}\right)^t I(t), \end{aligned}$$

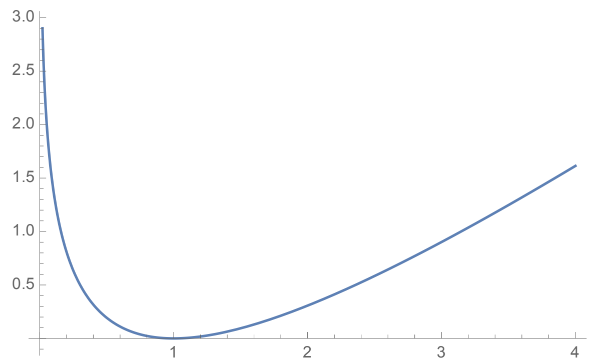


Figure 9: The function $\Phi(u) = u - 1 - \log u$.

where we define

$$\Phi(u) = u - 1 - \log u,$$
$$I(t) = \int_0^\infty e^{-t\Phi(u)} \frac{du}{u}.$$

(Again, note that we massaged the integrand to cancel the Taylor expansion of $-\log u$ around $u = 1$ up to the first order.) Our goal is to prove that

$$I(t) = \sqrt{\frac{2\pi}{t}} + O(t^{-7/10}) \qquad \text{as } t \to \infty.$$

As before, this will be done by splitting the integral into a main term and error terms. The idea is that for large t, the bulk of the contribution to the integral comes from a region very near the point where $\Phi(u)$ takes its minimum. It is easy to check by differentiation that this minimum is obtained at $u = 1$, and that we have

$$\Phi(1) = 0, \quad \Phi'(1) = 0, \quad \Phi''(1) = 1,$$

and $\Phi(u) \geq 0$ for all $u \geq 0$. See Fig. 9. Denote

$$I_1 = \int_0^{1/2} e^{-t\Phi(u)} \frac{du}{u},$$
$$I_2 = \int_{1/2}^{2} e^{-t\Phi(u)} \frac{du}{u},$$
$$I_3 = \int_2^{\infty} e^{-t\Phi(u)} \frac{du}{u},$$

so that $I(t) = I_1 + I_2 + I_3$. The main contribution will come from I_2, the part of the integral that contains the critical point $u = 1$, so let us examine that term first. Expanding $\Phi(u)$ in a Taylor series around $u = 1$, we have

$$\Phi(u) = \frac{(u-1)^2}{2} + O((u-1)^3)$$

for $u \in [1/2, 2]$ (in fact the explicit bound $\left|\Phi(u) - \frac{(u-1)^2}{2}\right| \le (u-1)^3$ on this interval can be easily checked). As before, noting that

$$t\left[\frac{(u-1)^2}{2} + O((u-1)^3)\right] = \frac{1}{2}(\sqrt{t}(u-1))^2 + O\left(\frac{(\sqrt{t}(u-1))^3}{\sqrt{t}}\right),$$

we see that it is natural to apply a linear change of variables $v = \sqrt{t}(u-1)$ to bring the integrand to a scale-free, centered form. This results in

$$\begin{aligned} I_2 &= \frac{1}{\sqrt{t}} \int_{-\frac{1}{2}\sqrt{t}}^{\sqrt{t}} \exp\left(-t\Phi\left(1 + \frac{v}{\sqrt{t}}\right)\right) \frac{1}{1 + v/\sqrt{t}}\, dv \\ &= \frac{1}{\sqrt{t}} \int_{-\frac{1}{2}\sqrt{t}}^{\sqrt{t}} \exp\left(-\frac{v^2}{2} + O\left(\frac{v^3}{\sqrt{t}}\right)\right)\left(1 + O\left(\frac{t}{\sqrt{t}}\right)\right) dv. \end{aligned}$$

As before, we actually need to split up this integral into two parts to take into account the fact that the $O(v^3/\sqrt{t})$ term can blow up when v is large enough. Let $M = t^{1/10}$, and denote

$$\begin{aligned} J_1 &= \frac{1}{\sqrt{t}} \int_{-M}^{M} \exp\left(-t\Phi\left(1 + \frac{v}{\sqrt{t}}\right)\right) \frac{1}{1 + v/\sqrt{t}}\, dv, \\ J_2 &= \frac{1}{\sqrt{t}} \int_{[-\frac{1}{2}\sqrt{t}, \sqrt{t}] \setminus [-M, M]} \exp\left(-t\Phi\left(1 + \frac{v}{\sqrt{t}}\right)\right) \frac{1}{1 + v/\sqrt{t}}\, dv, \end{aligned}$$

so that $I_2 = J_1 + J_2$. For J_1 we have

$$\begin{aligned} J_1 &= \frac{1}{\sqrt{t}} \int_{-M}^{M} \exp\left(-t\Phi\left(1 + \frac{v}{\sqrt{t}}\right)\right) \frac{1}{1 + v/\sqrt{t}}\, dv \\ &= \frac{1}{\sqrt{t}} \int_{-M}^{M} e^{-v^2/2}\left(1 + O\left(\frac{v^3}{\sqrt{t}}\right)\right)\left(1 + O\left(\frac{v}{\sqrt{t}}\right)\right) dv \\ &= \frac{1}{\sqrt{t}}\left(1 + O(t^{-1/5})\right) \int_{-M}^{M} e^{-v^2/2} dv = \sqrt{\frac{2\pi}{t}}\left(1 + O(t^{-1/5})\right), \end{aligned}$$

in the last step using a similar estimate as the one we used in our proof of Stirling's approximation for $n!$. Next, for J_2 we use the elementary inequality (prove it as an exercise)

$$\Phi(u) \ge \frac{(u-1)^2}{2} \qquad (0 \le u \le 1),$$

and the more obvious fact that $1/(1 + v/\sqrt{t}) \le 2$ for $v \in [-\frac{1}{2}\sqrt{t}, \sqrt{t}]$ to get that

$$\begin{aligned} J_2 &\le \frac{2}{\sqrt{t}} \int_{[-\frac{1}{2}\sqrt{t}, \sqrt{t}] \setminus [-M, M]} e^{-v^2/2}\, dv \le \frac{4}{\sqrt{t}} \int_{M}^{\infty} e^{-v^2/2}\, dv \\ &= O(e^{-M}) = \frac{1}{\sqrt{t}} O(t^{-1/5}). \end{aligned}$$

as in our earlier proof. Combining the above results, we have shown that

$$I_2 = \left(1 + O(t^{-1/5})\right)\sqrt{\frac{2\pi}{t}}.$$

Next, we bound I_1. Here we use a different method since there is a different source of potential trouble near the left end $u = 0$ of the integration interval. Considering first a truncated integral over $[\varepsilon, 1/2]$ and performing an integration by parts, we have

$$\begin{aligned}\int_\varepsilon^{1/2} e^{-t\Phi(u)} \frac{du}{u} &= -\frac{1}{t}\int_\varepsilon^{1/2} \frac{d}{du}\left(e^{-t\Phi(u)}\right)\frac{1}{\Phi'(u)u}\,du \\ &= -\frac{1}{t}\left[\frac{e^{-t\Phi(u)}}{u-1}\right]_{u=\varepsilon}^{u=1/2} - \frac{1}{t}\int_\varepsilon^{1/2} e^{-t\Phi(u)}\frac{du}{(u-1)^2}.\end{aligned}$$

Taking the limit as $\varepsilon \to 0$ (and noting that $\Phi(\varepsilon) \to +\infty$ in this limit) yields the formula

$$I_1 = \frac{2}{t}e^{-t\Phi(1/2)} - \frac{1}{t}\int_0^{1/2} e^{-t\Phi(u)}\frac{du}{(u-1)^2} = O\left(\frac{1}{t}\right) \qquad \text{as } t \to \infty.$$

Finally, I leave it as an exercise to obtain a similar estimate $I_3 = O(1/t)$ for the remaining integral on $[2, \infty)$. Combining the various estimates yields the claimed result that

$$I(t) = I_1 + I_2 + I_3 = \left(1 + O(t^{-1/5})\right)\sqrt{\frac{2\pi}{t}}.$$

□

The proof above is a simplified version of the analysis in Appendix A of [11]. The more detailed analysis there shows that the asymptotic formula we proved for $\Gamma(t)$ remains valid for complex t. Specifically, they prove that for complex s in the "Pac-Man shaped" region

$$S_\delta = \{z \in \mathbb{C} : |\arg z| \geq \pi - \delta\}$$

(for each fixed $0 < \delta < \pi$) the gamma function satisfies

$$\Gamma(s) = \left(1 + O(|s|^{-1/2})\right)\sqrt{2\pi}s^{s-1/2}e^{-s} \qquad \text{as } |s| \to \infty,\ s \in S_\delta.$$

Here, $s^{s-1/2}$ is defined as $\exp((s - 1/2)\operatorname{Log} s)$, where Log denotes as usual the principal branch of the logarithm function. This sort of approximation is important in certain applications of complex analysis, for example to analytic number theory.

Problems

1. Below is a list of basic formulas in complex analysis (think of them as "formulas you need to know like the back of your hand"). Review each of them, making sure that you understand what it says and why it is true—that is, if it is a theorem, prove it, or if it is a definition, make sure you understand that that is the case.

   In the formulas below, a, b, c, d, t, x, y denote arbitrary real numbers; w, z denote arbitrary complex numbers.

   a. $(a+bi)(c+di) = (ac-bd)+(ad+bc)i$
   b. $i^2 = -1$
   c. $\dfrac{1}{i} = -i$
   d. $z = \mathrm{Re}(z) + i\,\mathrm{Im}(z)$
   e. $\overline{z} = \mathrm{Re}(z) - i\,\mathrm{Im}(z)$
   f. $\mathrm{Re}(z) = \dfrac{z+\overline{z}}{2}$
   g. $\mathrm{Im}(z) = \dfrac{z-\overline{z}}{2i}$
   h. $|z|^2 = z\overline{z}$
   i. $\dfrac{1}{z} = \dfrac{\overline{z}}{|z|^2}$
   j. $\dfrac{1}{x+iy} = \dfrac{x-iy}{x^2+y^2}$
   k. $\overline{w\cdot z} = \overline{w}\cdot\overline{z}$
   l. $|wz| = |w|\cdot|z|$
   m. $\Big||w|-|z|\Big| \le |w+z| \le |w|+|z|$
   n. $e^{x+iy} = e^x(\cos(y)+i\sin(y))$
   o. $|e^z| = e^{\mathrm{Re}(z)}$
   p. $|e^z| \le e^{|z|}$
   q. $e^{it} = \cos(t)+i\sin(t)$
   r. $|e^{it}| = 1$
   s. $\cos(t) = \dfrac{e^{it}+e^{-it}}{2}$
   t. $\sin(t) = \dfrac{e^{it}-e^{-it}}{2i}$
   u. $e^{\pi i} = -1$
   v. $e^{\pm\pi i/2} = \pm i$
   w. $e^{2\pi i} = 1$

2. Remind yourself of the definitions of the following terms in complex analysis, referring to the textbook [11] (in particular, sections 1.2, 1.3, 2.1, 2.2) or online sources if necessary. Try to spend some time thinking about the answers yourself before looking them up.

   a. real part
   b. imaginary part
   c. complex conjugate
   d. modulus
   e. argument
   f. open set (in $\mathbb{C}$)
   g. closed set
   h. connected set
   i. bounded set
   j. compact set
   k. region
   l. convergent sequence
   m. Cauchy sequence
   n. limit point
   o. accumulation point
   p. continuous function
   q. differentiable function (of a complex variable)
   r. holomorphic function
   s. analytic function

t. entire function

u. meromorphic function

v. harmonic function (of two variables)

3. For each of the following functions, determine for which z it is analytic

   a. $f(z) = z$

   b. $f(z) = \mathrm{Re}(z)$

   c. $f(z) = |z|$

   d. $f(z) = |z|^2$

   e. $f(z) = \overline{z}$

   f. $f(z) = 1/z$

4. For each of the following functions $u(x, y)$, determine if there exists a function $v(x, y)$ such that $f(x + iy) = u(x, y) + iv(x, y)$ is an entire function, and if so, find it, and try to find a formula for $f(z)$ directly in terms of z rather than in terms of its real and imaginary parts. (**Hint**[4])

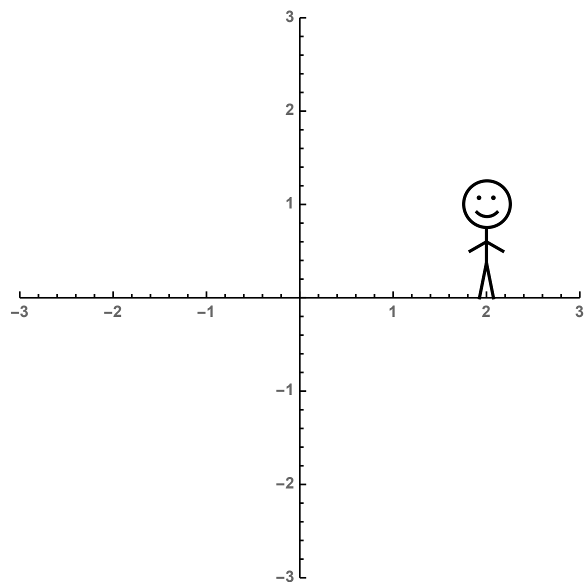
   a. $u(x, y) = x^2 - y^2$

   b. $u(x, y) = y^3$

   c. $u(x, y) = x^4 - 6x^2y^2 + 3x + y^4 - 2$

   d. $u(x, y) = \cos x \cosh y$

5. Draw (approximately, or with as much precision as you can) the image in the w-plane of the following figure in the z-plane



   under each of the following maps $w = f(z)$:

[4] Use the well-known equations named after two pioneers of complex analysis.

a. $w = \frac{1}{2}z$

b. $w = iz$

c. $w = \overline{z}$

d. $w = (2+i)z - 3$

e. $w = 1/z$

f. $w = z^2 - 1$

6. Prove that the complex numbers a, b, c form the vertices of an equilateral triangle if and only if $a^2 + b^2 + c^2 = ab + ac + bc$.
7. Illustrate the claim from page 9 regarding the orthogonality of the level curves of the real and imaginary parts an analytic functions by drawing (by hand after working out the relevant equations, or using a computer) the level curves of $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ for $f = z^2$, $f = e^z$.
8. An immediate corollary of the Fundamental Theorem of Algebra (together with standard properties of polynomials, namely the fact that c is a root of $p(z)$ if and only if $p(z)$ is divisible by the linear factor $z - c$) is that any complex polynomial

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0,$$

(where $a_0, \ldots, a_n \in \mathbb{C}$ and $a_n \neq 0$), can be factored as

$$p(z) = a_n \prod_{k=1}^{n} (z - z_k)$$

for some $z_1, \ldots, z_n \in \mathbb{C}$ (these are the roots of $p(z)$ counted with multiplicities). Use this to prove that any such polynomial where the coefficients $a_0, \ldots, a_n$ are *real* has a factorization

$$p(z) = a_n Q_1(z) Q_2(z) \ldots Q_m(z)$$

where each $Q_k(z)$ is a linear or quadratic monic polynomial (i.e., is of one of the forms $z - c$ or $z^2 + bz + c$) with real coefficients.
9. Let $p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0$ be a complex polynomial of degree n (that is, $a_0, \ldots, a_n \in \mathbb{C}$ and $a_n \neq 0$), which as mentioned in question 1 above can be factored as

$$p(z) = a_n \prod_{k=1}^{n} (z - z_k)$$

where $z_1, \ldots, z_n$ are the roots of $p(z)$ counted with multiplicities. Assuming that $n \geq 2$, the derivative $p'(z)$ can be similarly factored as

$$p'(z) = n a_n \prod_{k=1}^{n-1} (z - w_k)$$

where $w_1, \ldots, w_{n-1}$ denote the roots of $p'(z)$. Prove that $w_1, \ldots, w_{n-1}$ all lie in the convex hull of $z_1, \ldots, z_n$ (see Figure 1 for an illustration). That is, each w_k can be expressed as a convex combination

$$w_k = \alpha_1 z_1 + \alpha_2 z_2 + \ldots + \alpha_n z_n,$$

where $\alpha_1, \ldots, \alpha_n$ are nonnegative real numbers and $\sum_j \alpha_j = 1$. (To be clear, there are different coefficients for each k.)

Hint. A complex number z is a root of $p'(z)$ that is not also a root of $p(z)$ if and only if $p'(z)/p(z) = 0$. (Note: the expression p'/p is known as the logarithmic derivative of p.) Find a way to make more explicit what this equation says.

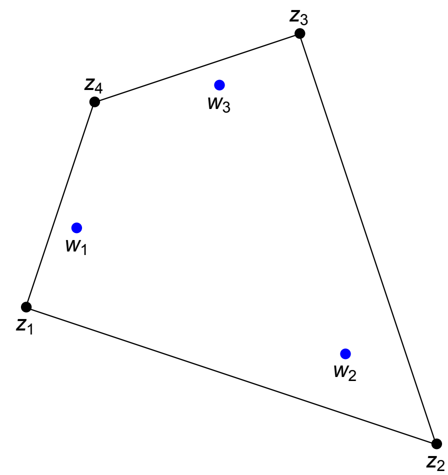


Figure 10: An example of the roots of a complex polynomial and of its derivative. Here $z_1 = 0$, $z_2 = 3 - i$, $z_3 = 2 + 2i$, $z_4 = \frac{1+3i}{2}$ and $w_1 \doteq 0.375 + 0.586i$, $w_2 \doteq 2.336 - 0.335i$, $w_3 \doteq 1.414 + 1.624i$.

10. **Cardano's method for solving cubic equations.** Let $p(z) = az^3 + bz^2 + cz + d$, with $a, b, c, d \in \mathbb{C}$, $a \neq 0$. We wish to solve the equation $p(z) = 0$, i.e., find the roots of the cubic polynomial $p(z)$.

   (a) Show that the substitution $w = z - \frac{b}{3a}$ brings the equation to the simpler form
   $$w^3 + pw + q = 0 \tag{5}$$
   for some values of p, q (find them!) given as functions of a, b, c, d.

   (b) Show that assuming a solution to (5) of the form $w = u + v$, the equation (5) for w can be solved by finding a pair u, v of complex numbers such that the equations
   $$p = -3uv, \tag{6}$$
   $$q = -(u^3 + v^3) \tag{7}$$
   are satisfied.

   (c) Explain why, in order to solve the pair of equations (6)–(7), one can alternatively solve
   $$\frac{p^3}{27} = -RS, \tag{8}$$
   $$q = -(R + S), \tag{9}$$
   where we now denote new unknowns R, S defined by $R = u^3$, $S = v^3$. More precisely, any solution of (6)–(7) can be obtained from *some* (easily determined) solution of (8)–(9).

   (d) Explain why the problem of solving (8)–(9) in the unknowns R, S is equivalent to solving the quadratic equation
   $$t^2 + qt - \frac{p^3}{27} = 0 \tag{10}$$

in a (complex) unknown variable t.

(e) Using the above reductions, show that the three solutions of the simplified cubic (5) can be expressed as

$$\begin{aligned} w_1 &= u + v, \\ w_2 &= \zeta\, u + \overline{\zeta}\, v, \\ w_3 &= \overline{\zeta}\, u + \zeta\, v, \end{aligned}$$

where $\zeta = e^{2\pi i/3} = \frac{1}{2}(-1 + i\sqrt{3})$ (a cube root of unity) and u, v are properly chosen cube roots of R, S obtained as solutions to (10).

(f) Illustrate the above procedure by applying it to get formulas for the roots of the cubic equation

$$z^3 + 6z^2 + 9z + 3 = 0.$$

Bring the formulas to a form that makes it clear that the roots are real numbers.

11. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a 2×2 real matrix. Prove the "conformality lemma" from page 9, which asserts the equivalence of the following three conditions:

(a) A as a linear map preserves orientation (that is, $\det A > 0$) and is conformal, that is

$$\frac{\langle Aw_1, Aw_2 \rangle}{|Aw_1|\,|Aw_2|} = \frac{\langle w_1, w_2 \rangle}{|w_1|\,|w_2|}$$

for all $w_1, w_2 \in \mathbb{R}^2$. (Here $\langle w_1, w_2 \rangle$ denotes the standard inner product in $\mathbb{R}^2$, and $|w| = \langle w, w \rangle^{1/2}$ is the usual two-dimensional norm of a vector in $\mathbb{R}^2$.)

(b) A takes the form $A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ for some $a, b \in \mathbb{R}$.

(c) A takes the form $A = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ for some $r > 0$ and $\theta \in \mathbb{R}$. (That is, geometrically A acts by a rotation followed by a scaling.)

12. A function $f = u + iv$ of a complex variable $z = x + iy$ is traditionally thought of as a function of the two coordinates x and y. However, if we think of the equations

$$z = x + iy, \qquad \overline{z} = x - iy$$

as representing a formal change of variables from the "real coordinates" (x, y) to the "complex conjugate coordinates" $(z, \overline{z})$, then it may make sense to think of f as a function of the two variables z and $\overline{z}$ (pretending that those are two independent variables). Thus we may suggestively write $u = u(z, \overline{z})$ and $v = v(z, \overline{z})$, and consider operations such as taking the partial derivatives of f, u, v with respect to z and $\overline{z}$.

Show that, from this somewhat strange point of view, the Cauchy-Riemann equations

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

can be rewritten in the more concise equivalent form

$$\frac{\partial f}{\partial \bar{z}} = 0,$$

assuming that it is okay to apply the chain rule from multivariable calculus; and moreover, that in this notation we also have the identity

$$f'(z) = \frac{\partial f}{\partial z}.$$

13. Let $f : \Omega \to \mathbb{C}$ be a function defined on a region Ω such that both the functions $f(z)$ and $zf(z)$ have real and imaginary parts that are harmonic functions (i.e., satisfy the Laplace equation $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$). Prove that $f(z)$ is holomorphic on Ω.

14. **Uniform convergence on compact subsets.** Given a sequence of functions $f_n : \Omega \to \mathbb{C}$, $n \geq 1$, defined on a region $\Omega \subset \mathbb{C}$, we say that f_n *converges uniformly on compact subsets* to a limiting function $f : \Omega \to \mathbb{C}$ if for any compact subset $K \subset \Omega$, $f_n(z) \to f(z)$ as $n \to \infty$, uniformly on $z \in K$.

    (a) (Warm-up) Write this definition more precisely in ϵ-δ language.

    (b) Prove that if f_n are holomorphic functions, $f_n \to f$ uniformly on compact subsets, $f_n' \to g$ uniformly on compact subsets, and g is continuous, then f is holomorphic and $f' = g$.

    Hint. Fix some $z_0 \in \mathbb{C}$. Start by proving that for z in a sufficiently small neighborhood of z we have the two identities

$$f_n(z) = f_n(z_0) + \int_{z_0}^{z} f_n'(w)\,dw,$$
$$f(z) = f(z_0) + \int_{z_0}^{z} g(w)\,dw,$$

    where the integral is over the line segment connecting z_0 to z.

    (c) Prove that a power series $\sum_{n=0}^{\infty} a_n z^n$ converges uniformly on compact subsets in its disk of convergence, and that the function it defines is continuous.

    Hint. It is enough (explain why) to prove uniform convergence on any closed disk of the form $\overline{D_r(0)}$ where $0 < r < R$ and R is the radius of convergence of the series.

    (d) Deduce that power series are holomorphic functions that can be differentiated termwise (a fact we already proved in class in a more direct way; the above approach provides an alternative proof).

Remark. This problem requires only elementary arguments, but using more advanced material we will prove as a consequence of Cauchy's theorem that in part (b) above the assumption that the sequence of derivatives f_n' converges to a limit can be dropped; that is, if a sequence of holomorphic functions converges uniformly on compact subsets, then the limiting function is automatically a holomorphic function whose derivative is the limit (in the sense of uniform convergence on compacts) of the sequence of derivatives of the original sequence. This is a surprising and nontrivial fact, as illustrated for example by the observation that the analogous statement in real analysis is false (e.g., by the Weierstrass approximation theorem, any continuous function on a closed interval is the uniform limit of a sequence of polynomials).

15. **Cauchy's theorem and irrotational vector fields.** Recall from vector calculus that a planar vector field $\mathbf{F} = (P, Q)$ defined on some region $\Omega \subset \mathbb{C} = \mathbb{R}^2$ is called *conservative* if it is of the form $\mathbf{F} = \nabla g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right)$ (the gradient of g) for some scalar function $g : \Omega \to \mathbb{R}$. By the fundamental theorem of calculus for line integrals, for such a vector field we have

$$\oint_\gamma \mathbf{F} \cdot \mathbf{ds} = 0$$

for any *closed* curve γ. Recall also that (as is easy to check) any conservative vector field is *irrotational*, that is, it satisfies

$$\operatorname{curl} \mathbf{F} = 0$$

(where in the context of two-dimensional vector fields, the curl is simply $\operatorname{curl} \mathbf{F} = \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$). The converse also holds under suitable conditions: if the region Ω is *simply-connected* (a concept we will discuss later in the course), then a theorem in vector calculus says that an irrotational vector field is also conservative.

Use these background results to show that if $f = u + iv$ is holomorphic on a simply-connected domain Ω, then

$$\oint_\gamma f(z)\, dz = 0$$

for any closed curve γ in Ω. (This is, of course, Cauchy's theorem.)

16. **The Bernoulli numbers.** Define the function

$$f(z) = \begin{cases} \dfrac{z}{e^z - 1} & \text{if } z \neq 0, \\ 1 & \text{if } z = 0. \end{cases}$$

(a) Convince yourself that $f(z)$ is analytic in a neighborhood of 0. Where else is it analytic? In particular, find the maximal radius R such that $f(z)$ is analytic on the disk $D_R(0)$.

(b) One of the basic complex analysis theorems we will discuss is that analytic functions have a power series expansion. The *Bernoulli numbers* are the numbers $(B_n)_{n=0}^\infty$ defined by the power series expansion

$$\sum_{n=0}^\infty \frac{B_n}{n!} z^n = f(z).$$

For example, the first three Bernoulli numbers are $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$. Prove that the Bernoulli numbers satisfy the following identities:

i. $B_{2k+1} = 0$ for $k = 1, 2, \ldots$ (but not for $k = 0$).

Hint. A function $g(z) = \sum_{n=0}^\infty a_n z^n$ satisfies $a_1 = a_3 = a_5 = \ldots = 0$ if and only if $g(z) = g(-z)$, i.e., $g(z)$ is an even function.

ii. $(n+1)B_n = -\sum_{k=0}^{n-1} \binom{n+1}{k} B_k, \ \ (n \geq 2)$.

iii. $(2n+1)B_{2n} = -\sum_{k=1}^{n-1} \binom{2n}{2k} B_{2k} B_{2n-2k}, \ \ (n \geq 2)$.

Hint. Show that the function $g(z) = f(z) + z/2$ satisfies the equation

$$g(z) - z g'(z) = g(z)^2 - z^2/4.$$

iv. $\dfrac{z}{2}\coth\left(\dfrac{z}{2}\right) = \displaystyle\sum_{n=0}^{\infty} \frac{B_{2n}}{(2n)!} z^{2n}$.

(c) Another general result is that the radius of convergence of the power series of an analytic function around $z = z_0$ is precisely the radius of the maximal disk around z_0 where f is analytic. Assuming this, deduce that

$$\limsup_{n\to\infty} \left| \frac{B_n}{n!} \right|^{1/n} = 1/R,$$

where R is the number you found in part (a). (Note: in a later problem — problem 26 on pages 82–83 — we will derive a much more precise estimate for the asymptotic rate of growth of the Bernoulli numbers.)

17. **Bessel functions.** The *Bessel functions* are a family of functions $(J_n)_{n=-\infty}^{\infty}$ of a complex variable,defined by

$$J_n(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!(k+n)!} \left(\frac{z}{2}\right)^{2k+n}.$$

(For example, note that $J_0(-2\sqrt{x}) = \sum_{k=0}^{\infty} \frac{x^k}{(k!)^2}$, which is reminiscent of the exponential function and already seems like a fairly natural function to study.) Find the radius of convergence of the series defining $J_n(z)$, and prove that the Bessel functions satisfy the following properties:

i. $J_{-n}(z) = (-1)^n J_n(z)$.

ii. Recurrence relation: $\quad J_{n+1}(z) = \dfrac{2n}{z} J_n(z) - J_{n-1}(z)$.

iii. Bessel's differential equation: $\; z^2 J_n''(z) + z J_n'(z) + (z^2 - n^2) J_n(z) = 0$.

iv. Summation identity: $\quad \displaystyle\sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{z}{2}\right)^n J_n(z) = 1$.

v.* Other miscellaneous identities (for those who enjoy this sort of thing—feel free to skip if you find these sorts of computations uninteresting):

$$\exp\left[\frac{z}{2}\left(t - \frac{1}{t}\right)\right] = \sum_{n=-\infty}^{\infty} J_n(z) t^n,$$

$$\cos(z \sin t) = J_0(z) + 2\sum_{n=1}^{\infty} J_{2n}(z)\cos(2nt),$$

$$\sin(z \sin t) = 2\sum_{n=0}^{\infty} J_{2n+1}(z)\sin((2n+1)t),$$

$$\cos(z \cos t) = J_0(z) + 2\sum_{n=1}^{\infty} (-1)^n J_{2n}(z)\cos(2nt),$$

$$\sin(z \cos t) = 2\sum_{n=0}^{\infty} (-1)^n J_{2n+1}(z)\sin((2n+1)t),$$

$$J_n(z) = \frac{1}{\pi}\int_0^{\pi} \cos\left(z \sin t - nt\right)\, dt.$$

Hint for the last equation: $\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$.

Remark. The Bessel functions are very important functions in mathematical physics, and appear naturally in connection with various problems in diffusion, heat conduction, electrodynamics, quantum mechanics, Brownian motion, probability, and more. More recently they played an important role in some problems in combinatorics related to longest increasing subsequences (a subject I wrote a book about, available to download from my home page). Their properties as analytic functions of a complex variable are also a classical, though no longer very fashionable, topic of study.

18. Show that Liouville's theorem ("a bounded entire function is constant") can be proved directly using the "simple" ($n = 0$) case of Cauchy's integral formula, instead of using the case $n = 1$ of the extended formula as we did in the lecture.

    Hint. For an arbitrary pair of complex numbers $z_1, z_2 \in \mathbb{C}$, show that $|f(z_1) - f(z_2)| = 0$.

19. Show that Liouville's theorem can in fact be deduced even just from the mean value property of holomorphic functions, which is the special case of Cauchy's integral formula in which z is taken as the center of the circle around which the integration is performed.

    Hint. Here it makes sense to consider a modified version of the mean value property (that follows easily from the original version) that says that $f(z)$ is the average value of $f(w)$ over a *disc* $D_R(z)$ (instead of a circle $C_R(z)$). That is,

    $$f(z) = \frac{1}{\pi R^2} \iint_{D_R(z)} f(x + iy)\, dx\, dy,$$

    where the integral is an ordinary two-dimensional Riemann integral. Explain why this formula holds, then use it to again bound $|f(z_1) - f(z_2)|$ from above by a quantity that goes to 0 as $R \to \infty$.

20. Prove the following generalization of Liouville's theorem: let f be an entire function that satisfies for all $z \in \mathbb{C}$ the inequality

    $$|f(z)| \leq A + B|z|^n$$

    for some constants $A, B > 0$ and integer $n \geq 0$. Then f is a polynomial of degree $\leq n$.

21. If $p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + \ldots + a_0$ is a polynomial of degree n such that

    $$|a_n| > \sum_{j=0}^{n-1} |a_j|,$$

    prove that $p(z)$ has exactly n zeros (counting multiplicities) in the unit disc $|z| < 1$.

    Hint. Use the fundamental theorem of algebra.

    Note. This is a special case of a less elementary fact that can be proved using Rouché's theorem; see problem 28

22. The Cauchy integral formula is intimately connected to an important formula from the theory of the Laplace equation and harmonic functions called the *Poisson integral formula.* Solve exercises 11–12 (pages 66–67) in Chapter 2 of [Stein-Shakarchi], which explore this connection, and more generally the connection between holomorphic and harmonic functions.

23. Spend at least 5–10 minutes thinking about the concept of a *toy contour*. Specifically, for the case of a *keyhole contour* we discussed in the context of the proof of Cauchy's integral formula, think carefully about the steps that are needed to get a proof of Cauchy's theorem for the region enclosed by such a contour. Even better, sketch a proof of the key result that a function holomorphic in such a region (and therefore having the property that its contour integral along triangles and rectangles vanish) has a primitive.

24. **Characterizing some important families of holomorphic functions on $\mathbb{C}$ and $\widehat{\mathbb{C}}$.** For the problems below, denote

$$
\begin{aligned}
\widehat{\mathbb{C}} &= \mathbb{C} \cup \{\infty\} = \text{the Riemann sphere}, \\
\mathcal{K} &= \text{the set of constant functions } z \mapsto c \in \mathbb{C}, \\
\mathcal{L} &= \text{the set of linear functions } z \mapsto az + b, \quad a, b \in \mathbb{C}, \\
\mathcal{P} &= \text{the set of complex polynomials } z \mapsto \sum_{k=0}^{n} a_k z^k, \\
\mathcal{R} &= \text{the set of rational functions } z \mapsto \frac{p(z)}{q(z)}, \quad p, q \in \mathcal{P}, \\
\mathcal{M} &= \text{the set of Möbius transformations } z \mapsto \frac{az+b}{cz+d}, \quad a, b, c, d \in \mathbb{C}.
\end{aligned}
$$

Note the containment relations $\mathcal{K} \subset \mathcal{L} \subset \mathcal{P} \subset \mathcal{R} \supset \mathcal{M} \supset \mathcal{L} \supset \mathcal{K}$.

(a) (Warm-up problem) Prove that an entire function has a removable singularity at ∞ if and only if it is constant.

(b) Prove that the set of entire functions $f : \mathbb{C} \to \mathbb{C}$ that have a nonessential singularity at ∞ is $\mathcal{P}$, the polynomials.

(c) Prove that the set of meromorphic functions $f : \mathbb{C} \to \widehat{\mathbb{C}}$ that have a nonessential singularity at ∞ is $\mathcal{R}$, the rational functions.

(d) Prove that the set of meromorphic, one-to-one and onto functions $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is $\mathcal{M} \setminus \mathcal{K}$, the set of nonconstant Möbius transformations.

Hint. Use the characterization in problem 24c above. Specifically, show that a rational function $f(z) = p(z)/q(z)$ that is one-to-one must be a Möbius transformation. For example (I'm not sure if this is the simplest argument): argue that if z_0 is a complex number such that $q(z_0) \neq 0$ and such that $p'(z_0)q(z_0) - p(z_0)q'(z_0) \neq 0$, and $w_0 = f(z_0)$, then the equation $f(z) = w_0$ must have more than one solution in z, unless $p(z), q(z)$ are linear functions.

(e) Prove that set of entire functions $f : \mathbb{C} \to \mathbb{C}$ that are one-to-one and onto is precisely $\mathcal{L} \setminus \mathcal{K}$, the set of nonconstant linear functions.

Hint. This is a slightly more advanced problem since it relies on both the Casorati-Weierstrass theorem and the open mapping theorem (although a more elementary solution may exist). See the guidance for exercise 14 on page 105 of [Stein-Shakarchi].

Remarks. Given a region $\Omega \subset \mathbb{C}$, or more generally a Riemann surface Σ, complex analysts are interested in understanding the structure of its set of holomorphic functions ($\mathbb{C}$-valued holomorphic functions on Σ); its set of meromorphic functions ($\widehat{\mathbb{C}}$-valued holomorphic functions on Σ); and its set of holomorphic automorphisms (holomorphic, one-to-one and onto mappings from Σ to itself). Although we won't get into the general theory of Riemann surfaces, once one defines these concepts it easy to see that the above exercises essentially prove the following conceptually important results:

(i) The constant functions are the only holomorphic functions on $\widehat{\mathbb{C}}$.

(ii) The rational functions are the meromorphic functions on $\widehat{\mathbb{C}}$.

(iii) The nonconstant linear functions are the holomorphic automorphisms of $\mathbb{C}$.

(iv) The nonconstant Möbius transformations are the holomorphic automorphisms of $\widehat{\mathbb{C}}$.

Another related result that is not very difficult to prove is:

(v) The holomorphic automorphisms of the upper half-plane $\mathbb{H} = \{z : \operatorname{Im} z > 0\}$ are the Möbius transformations $z \mapsto \frac{az+b}{cz+d}$ with $a, b, c, d \in \mathbb{R}$ and $ad - bc > 0$. (Try to prove that any such map is indeed an automorphism of $\mathbb{H}$; the reverse implication that all automorphisms of $\mathbb{H}$ are of this form is a bit more difficult and requires a result known as the Schwarz lemma, which will likely be covered at some point in MAT205A/B.)

Note that the set of holomorphic functions on $\mathbb{C}$ (a.k.a. entire functions) and the set of meromorphic functions on $\mathbb{C}$ are much larger families of functions that do not have such a simple description as the functions in the relatively small families $\mathcal{L}, \mathcal{P}, \mathcal{R}, \mathcal{M}$. This is related to the fact that $\mathbb{C}$ is a non-compact Riemann surface.

25. **The partial fraction expansion of the cotangent function.** In this multipart question you are asked to prove a well-known infinite summation identity (equation (12) below), known as the partial fraction expansion of the cotangent function. In the next problem we will explore some of the consequences of this important result.

(a) Let $z \in \mathbb{C} \setminus \mathbb{Z}$. Use the residue theorem to evaluate the contour integral

$$I_N := \oint_{\gamma_N} \frac{\pi \cot(\pi w)}{(w+z)^2}\, dw$$

over the contour γ_N going in the positive direction around the rectangle with the four vertices $(\pm(N + 1/2), \pm N)$. Take the limit as $N \to \infty$ to deduce the identity

$$\frac{\pi^2}{(\sin \pi z)^2} = \sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^2} \qquad (z \in \mathbb{C} \setminus \mathbb{Z}). \tag{11}$$

Guidance. This is not a trivial exercise, but is not very difficult when broken down into the following elementary substeps:

i. Start by identifying the location of the singularities of the function $w \mapsto f_z(w) = \frac{\pi \cot(\pi w)}{(w+z)^2}$ (considered as a function of w for a fixed z whose value is not an integer), and their residues. This provides some good practice with residue computations.

ii. Use the residue theorem to obtain an expression for the contour integral I_N defined above.

iii. Separately, obtain estimates for I_N that can be used to show that $I_N \to 0$ as $N \to \infty$. Specifically, show using elementary manipulations that

$$|\sin(x+iy)|^2 = \sin^2 x + \sinh^2 y, \qquad |\cos(x+iy)|^2 = \cos^2 x + \sinh^2 y,$$

use this to conclude that when $x = \pi(N+1/2)$ and y is arbitrary,

$$|\cot(x+iy)| = \frac{\sinh^2 y}{1+\sinh^2 y} \le 1,$$

and that when $y = N$ and x is arbitrary,

$$|\cot(x+iy)| \le \frac{1+\sinh^2 N}{\sinh^2 N} \le 2 \quad (\text{if } N > 10);$$

then use these estimates to bound the integral.

iv. By comparing the two results about I_N, deduce (11).

(b) Integrate the identity (11) to deduce (using some additional fairly easy reasoning) the formulas

$$\pi\cot(\pi z) = \lim_{N\to\infty} \sum_{n=-N}^{N} \frac{1}{z+n} = \frac{1}{z} + \sum_{n=1}^{\infty} \frac{2z}{z^2-n^2} \qquad (z \in \mathbb{C}\setminus\mathbb{Z}). \tag{12}$$

26. **Consequences of the partial fraction expansion of the cotangent function.**

(a) Show that (12) implies the following infinite-product representation for the sine function:

$$\sin(\pi z) = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) \qquad (z \in \mathbb{C}). \tag{13}$$

Note that the function on the right-hand side is (or can be easily checked to be) an entire function of z with a simple zero at any integer $z = n \in \mathbb{Z}$, and whose Taylor expansion around $z = 0$ starts with $\pi z + O(z^3)$; thus it is a natural guess for an infinite product expansion of $\sin(\pi z)$, although the fact that this guess is correct is far from obvious; for example one can multiply the right-hand side by an arbitrary function of the form $e^{g(z)}$ and still have an entire function with the same set of zeros.

Hint. Compute the logarithmic derivatives of both sides of (13). You may want to review some basic properties of infinite products, as discussed for example on pages 140–142 of [Stein-Shakarchi]. (Spoiler alert: pages 142–144 contain a solution to this subexercise, starting with an independent proof of (12) and proceeding with a derivation of (13) along the same lines as I described above.)

(b) By specializing the value of z in (13) to an appropriate specific value, obtain the following infinite product formula for π, known as *Wallis' product* (first proved by John Wallis in 1655):

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdot\frac{8}{7}\cdot\frac{8}{9}\cdots.$$

(c) By comparing the first terms in the Taylor expansion around $z = 0$ of both sides of (13), derive the well-known identities

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}, \qquad \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}.$$

(d) More generally, one can use (13), or more conveniently (12), to obtain closed formulas for all the series

$$\zeta(2k) = \sum_{n=1}^{\infty} \frac{1}{n^{2k}} \qquad (k = 1, 2, \ldots) = 1 + \frac{1}{2^{2k}} + \frac{1}{3^{2k}} + \frac{1}{4^{2k}} + \ldots,$$

that is, the special values of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ at the positive even integers. To see this, first, rewrite (12) as

$$\pi \cot(\pi z) = \frac{1}{z} + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \left(\frac{1}{z+n} - \frac{1}{n} \right) \qquad (z \in \mathbb{C} \setminus \mathbb{Z}). \tag{14}$$

Expand both sides of (14) in a Taylor series around $z = 0$, making use of the expansion

$$\frac{z}{2} \coth\left(\frac{z}{2}\right) = \sum_{n=0}^{\infty} \frac{B_{2n}}{(2n)!} z^{2n}$$

we proved in an earlier homework exercise (where $(B_n)_{n=0}^{\infty}$ are the Bernoulli numbers). Compare coefficients and simplify to get the formula

$$\zeta(2k) = \frac{(-1)^{k+1}(2\pi)^{2k}}{2(2k)!} B_{2k}.$$

For example, using the first few values $B_2 = \frac{1}{6}$, $B_4 = -\frac{1}{30}$, $B_6 = \frac{1}{42}$, $B_8 = -\frac{1}{30}$, we get

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6},$$
$$\zeta(4) = \sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90},$$
$$\zeta(6) = \sum_{n=1}^{\infty} \frac{1}{n^6} = \frac{\pi^6}{945},$$
$$\zeta(8) = \sum_{n=1}^{\infty} \frac{1}{n^8} = \frac{\pi^8}{9450},$$

where of course the first two values coincide with those found earlier.

(e) Show that $\zeta(2k) = 1 + O(2^{-2k})$ as $k \to \infty$, and deduce that the asymptotic behavior of the Bernoulli numbers is given by

$$B_{2k} = (1 + O(2^{-2k}))(-1)^{k+1} \frac{2(2k)!}{(2\pi)^{2k}}, \qquad k \to \infty.$$

Note that this is consistent with our earlier (and much weaker) result that

$$\limsup_{k \to \infty} \left| \frac{B_{2k}}{(2k)!} \right|^{1/2k} = \frac{1}{2\pi}.$$

27. Let $f(z) = p(z)/q(z)$ be a rational function such that $\deg q \geq \deg p + 2$ (where $\deg p$ denotes the degree of a polynomial p). Prove that the sum of the residues of $f(z)$ over all its poles is equal to 0.

28. (A generalization of the result from problem 21) If $p(z) = a_n z^n + a_{n-1}z^{n-1} + \ldots + \ldots + a_0$ is a polynomial of degree n such that for some $0 \le k \le n$ we have

$$|a_k| > \sum_{\substack{0\le j\le n\\ j\neq k}} |a_j|,$$

prove that $p(z)$ has exactly k zeros (counting multiplicities) in the unit disk $|z| < 1$.

29. Suggested reading: go to the Mathematics Stack Exchange website (`https://math.stackexchange.com`) and enter "Rouche" into the search box, to get an amusing list of questions and exercises involving applications of Rouché's theorem to count zeros of polynomials and other analytic functions.

30. Show how Rouché's theorem can be used to give yet another proof of the fundamental theorem of algebra. This proof is one way to make precise the intuitively compelling "topological" proof idea we discussed at the beginning of the course.

31. (a) Draw a simply-connected region $\Omega \subset \mathbb{C}$ such that $0 \notin \Omega$, $1, 2 \in \Omega$, and such that there exists a branch $F(z)$ of the logarithm function on Ω satisfying

$$F(1) = 0, \qquad F(2) = \log 2 + 2\pi i$$

(where $\log 2 = 0.69314\ldots$ is the ordinary logarithm of 2 in the usual sense of real analysis).

(b) More generally, let $k \in \mathbb{Z}$. If we were to replace the above condition $F(2) = \log 2 + 2\pi i$ with the more general condition $F(2) = \log 2 + 2\pi i k$ but keep all the other conditions, would an appropriate simply-connected region $\Omega = \Omega(k)$ exist to make that possible? If so, what would this region look like, roughly, as a function of k?

32. Prove the following properties satisfied by the gamma function:

    i. Values at half-integers:

    $$\Gamma\left(n + \tfrac{1}{2}\right) = \frac{(2n)!}{4^n n!}\sqrt{\pi} \qquad (n = 0, 1, 2, \ldots).$$

    ii. The duplication formula:

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

    iii.* The multiplication theorem:

    $$\Gamma\left(s\right)\Gamma\left(s + \tfrac{1}{k}\right)\Gamma\left(s + \tfrac{2}{k}\right)\cdots\Gamma\left(s + \tfrac{k-1}{k}\right) = (2\pi)^{(k-1)/2}k^{1/2-ks}\Gamma(ks).$$

33. For $n \ge 1$, let V_n denote the volume of the unit ball in $\mathbb{R}^n$. By evaluating the n-dimensional integral

$$A_n = \iint\!\ldots\!\int_{\mathbb{R}^n} \exp\left(-\tfrac{1}{2}\sum_{j=1}^{n} x_j^2\right) dx_1\, dx_2 \ldots dx_n$$

in two ways, prove the well-known formula

$$V_n = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2} + 1\right)}.$$

Note. This problem requires applying a small amount of geometric intuition (or, alternatively, having some technical knowledge of spherical coordinates in $\mathbb{R}^n$). The solution can be found on this Wikipedia page.

34. The *beta function* is a function $B(s,t)$ of two complex variables, defined for $\mathrm{Re}(s), \mathrm{Re}(t) > 0$ by

$$B(s,t) = \int_0^1 x^{s-1}(1-x)^{t-1}\,dx.$$

(a) (Warm-up) Convince yourself that the improper integral defining $B(s,t)$ converges if and only if $\mathrm{Re}(s), \mathrm{Re}(t) > 0$.

(b) Show that $B(s,t)$ can be expressed in terms of the gamma function as

$$B(s,t) = \frac{\Gamma(s)\Gamma(t)}{\Gamma(s+t)}.$$

Hint. Start by writing $\Gamma(s)\Gamma(t)$ as a double integral on the positive quadrant $[0,\infty)^2$ of $\mathbb{R}^2$ (with integration variables, say, x and y); then make the change of variables $u = x+y$, $v = x/(x+y)$ and use the change of variables formula for two-dimensional integrals to show that the integral evaluates as $\Gamma(s+t)B(s,t)$.

Remark. Note the similarity of the identity relating the gamma and beta functions to the formula $\binom{n}{k} = \frac{n}{k!(n-k)!}$; indeed, using the relation $\Gamma(m+1) = m!$ and the functional equation $\Gamma(s+1) = s\,\Gamma(s)$, we see using the above relation that for nonnegative, integer-valued arguments we have

$$B(n,m)^{-1} = \frac{nm}{n+m} \cdot \frac{\Gamma(n+m+1)}{\Gamma(n+1)\Gamma(m+1)} = \frac{nm}{n+m}\binom{n+m}{n}.$$

In other words, except for the correction factor $\frac{nm}{n+m}$, the inverse of the beta function can be thought of as a natural extention of binomial coefficients to real-valued arguments.

35. The *digamma function* $\psi(s)$ is the logarithmic derivative

$$\psi(s) = \frac{\Gamma'(s)}{\Gamma(s)}$$

of the gamma function, also considered as a somewhat important special function in its own right.

(a) Show that $\psi(s)$ has the convergent series expansions

$$\begin{aligned}\psi(s) &= -\gamma - \frac{1}{s} + \sum_{n=1}^{\infty} \frac{s}{n(n+s)} \\ &= -\gamma + \sum_{n=0}^{\infty}\left(\frac{1}{n+1} - \frac{1}{n+s}\right) \qquad (s \neq 0, -1, -2, \ldots).\end{aligned}$$

where γ is the Euler-Mascheroni constant.

(b) Equivalently, show that $\psi(s)$ can be expressed as

$$\psi(s) = -\lim_{n\to\infty}\left(\sum_{k=1}^{n} \frac{1}{k+s} - \log n\right).$$

(c) Show that $\psi(s)$ satisfies the functional equation

$$\psi(s+1) = \psi(s) + \frac{1}{s}.$$

(d) Show that

$$\psi(n+1) = -\gamma + \sum_{k=1}^{n} \frac{1}{k} \qquad (n = 0, 1, 2, \ldots).$$

That is, $\psi(x) + \gamma$ can be thought of as extending the definition of the **harmonic numbers** $H_n = \sum_{k=1}^{n} \frac{1}{k}$ to non-integer arguments.

(e) Show that $\psi(s)$ satisfies the reflection formula

$$\psi(1-s) - \psi(s) = \pi \cot(\pi s).$$

(f)* Here is an amusing application of the digamma function. Consider the sequence of polynomials

$$P_n(x) = x(x-1)\ldots(x-n) \qquad (n = 0, 1, 2, \ldots)$$

and their derivatives

$$Q_n(x) = P_n'(x).$$

Note that by Rolle's theorem, $Q_n(x)$ has precisely one root in each interval $(k, k+1)$ for $0 \le k \le n-1$. Denote this root by $k + \alpha_{n,k}$, so that the numbers $\alpha_{n,k}$ (the fractional parts of the roots of $Q_n(x)$) are in $(0, 1)$.

A curious phenomenon can now be observed by plotting the points $\alpha_{n,k}$, $k = 0, \ldots, n-1$ numerically, say for $n = 50$ (Figure 11(a)). It appears that for large n they approximate some smooth limiting curve. This is correct, and in fact the following precise statement can be proved.

Theorem. Let $t \in (0, 1)$. Let $k = k(n)$ be a sequence such that $0 \le k(n) \le n-1$, $k(n) \to \infty$ as $n \to \infty$, $n - k(n) \to \infty$ as $n \to \infty$, and $k(n)/n \to t$ as $n \to \infty$. Then we have

$$\lim_{n\to\infty} \alpha_{n,k(n)} = R(t) := \frac{1}{\pi} \operatorname{arccot}\left(\frac{1}{\pi} \log\left(\frac{1-t}{t} \right) \right).$$

In the above formula, $\operatorname{arccot}(\cdot)$ refers to the branch of the inverse cotangent function taking values between 0 and π. The limiting function $R(t)$ is shown in Figure 11(b).

Prove this.

Guidance. Take the logarithmic derivative of $P_n(x)$ to see when the equation $Q_n(x)/P_n(x) = 0$ (which is equivalent to $Q_n(x) = 0$) holds. This will give an equation with a sum of terms. Find a way to separate them into two groups such that the sum in each group can be related, in an asymptotic sense as $n \to \infty$, to the digamma function evaluated at a certain argument (using property (b) above). Take the limit as $n \to \infty$, then simplify using the reflection formula (part (c)).

36. Given two integrable functions $f, g : \mathbb{R} \to \mathbb{C}$ (of a *real* variable), their *convolution* is the new function $h = f * g$ defined by the formula

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(t) g(x-t)\, dt \qquad (x \in \mathbb{R}).$$

As you might be aware, the convolution operation is extremely important in harmonic analysis, since it corresponds to a simple multiplication operation in the Fourier domain; in probability theory, where it

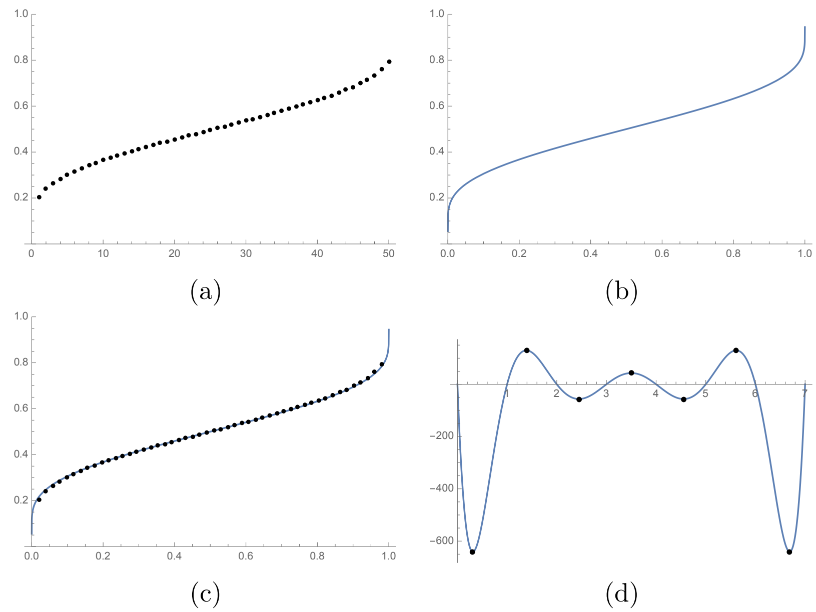


Figure 11: (a) A plot of the fractional parts of the roots of $Q_n(x)$ for $n = 50$. (b) The limiting function $R(t)$. (c) The two previous plots combined. (d) The polynomial $P_7(x)$. Note that the roots of $Q_7(x)$ correspond to the local minima and maxima of $P_7(x)$, which are highlighted.

corresponds to the addition of independent random variables; and in many other areas of mathematics, science and engineering.

For $\alpha > 0$ define the *gamma density* with parameter α, denoted $\gamma_\alpha : \mathbb{R} \to \mathbb{R}$, to be the function

$$\gamma_\alpha(x) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1} 1_{[0,\infty)}(x) \qquad (x \in \mathbb{R})$$

(where $1_A(x)$ denotes the characteristic function of a set $A \subset \mathbb{R}$, equal to 1 on the set and 0 outside it). Note that $\gamma_\alpha(x)$ is the nonnegative function whose integral equals $\Gamma(\alpha)$, except that it is divided by $\Gamma(\alpha)$ so that it becomes a probability density function. See Figure 12 for an illustration.

Show that for each $\alpha, \beta > 0$ we have

$$\gamma_\alpha * \gamma_\beta = \gamma_{\alpha+\beta}.$$

That is, the family of density functions $(\gamma_\alpha)_{\alpha>0}$ is closed under the convolution operation. This fact is one of the reasons why the family of gamma densities plays a very important role in probability theory and appears in many real-life applications.

37. (a) Show that the Laurent expansion of $\Gamma(s)$ around $s = 0$ is of the form

$$\Gamma(s) = \frac{1}{s} - \gamma + O(s)$$

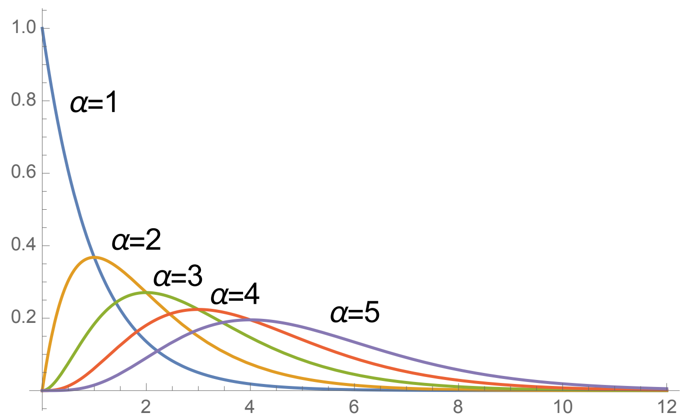


Figure 12: The gamma densities $\gamma_\alpha(x)$ for $\alpha = 1, 2, 3, 4, 5$.

(where γ is the Euler-Mascheroni constant). If you're feeling especially energetic, derive the more detailed expansion

$$\Gamma(s) = \frac{1}{s} - \gamma + \left(\frac{\gamma^2}{2} + \frac{\pi^2}{12}\right) s + O(s^2)$$

and proceed to derive (by hand, or if you prefer using a symbolic math software application such as SageMath or Mathematica) as many additional terms in the expansion as you have the patience to do.

(b) Show that the Laurent expansion of $\zeta(s)$ around $s = 1$ is of the form

$$\zeta(s) = \frac{1}{s-1} + \gamma + O(s-1).$$

38. Show that the symmetric version of the functional equation for the zeta function

$$\zeta^*(1-s) = \zeta^*(s),$$

where $\zeta^*(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$, can be rewritten in the equivalent form

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

39. Show that the Taylor expansion of the digamma function $\psi(s) = \frac{\Gamma'(s)}{\Gamma(s)}$ around $s = 1$ is given by

$$\psi(s) = -\gamma - \sum_{n=1}^{\infty} (-1)^{n-1} \zeta(n+1)(s-1)^n \qquad (|s-1| < 1).$$

40. Define a function $D(s)$ of a complex variable s by

$$D(s) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} = 1 - \frac{1}{2^s} + \frac{1}{3^s} - \frac{1}{4^s} + \ldots.$$

(a) Prove that the series defining $D(s)$ converges uniformly on any half-plane of the form $\mathrm{Re}(s) \ge \alpha$ where $\alpha > 0$, and conclude that $D(s)$ is defined and holomorphic in the half-plane $\mathrm{Re}(s) > 0$.

(b) Show that $D(s)$ is related to the Riemann zeta function by the formula

$$D(s) = (1 - 2^{1-s})\zeta(s) \qquad (\mathrm{Re}(s) > 1).$$

(c) Using this relation, deduce a new proof that the zeta function can be analytically continued to a meromorphic function on $\mathrm{Re}(s) > 0$ that has a simple pole at $s = 1$ with residue 1 and is holomorphic everywhere else in the region.

41. Let $\psi(x) = \sum_{p^k \le x} \log p$ denote von Mangoldt's weighted prime counting function. Show that $\psi(n) = \log \mathrm{lcm}(1, 2, \ldots, n)$, where for integers $a_1, \ldots, a_k$, the notation $\mathrm{lcm}(a_1, \ldots, a_k)$ denotes the least common multiple of $a_1, \ldots, a_k$.

Note that this implies that an equivalent formulation of the prime number theorem is the interesting statement that

$$\mathrm{lcm}(1, \ldots, n) = e^{(1+o(1))n} \qquad \text{as } n \to \infty.$$

42. (a) Prove that for all $x \ge 1$,

$$\prod_{p \le x} \frac{1}{1 - \frac{1}{p}} \ge \log x$$

(where the product is over all prime numbers p that are $\le x$).

(b) Pass to the logarithm and deduce that for some constant $K > 0$ we have the bound

$$\sum_{p \le x} \frac{1}{p} \ge \log \log x - K \qquad (x \ge 1).$$

That is, the *harmonic series of primes* $\sum_p \frac{1}{p}$ diverges as $\log \log x$, in contrast to the usual harmonic series which diverges as $\log x$.

43. **An alternative proof of the functional equation of the Jacobi theta function.** Recall that in Subsection 15.1 we defined the Jacobi theta function by

$$\vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t} \qquad (t > 0),$$

and showed that it satisfies the functional equation

$$\vartheta\left(\frac{1}{t}\right) = \sqrt{t}\vartheta(t). \tag{15}$$

(a) Use the residue theorem to evaluate the contour integral

$$\oint_{\gamma_N} \frac{e^{-\pi z^2 t}}{e^{2\pi i z} - 1}\, dz,$$

where γ_N is the rectangle with vertices $\pm(N+1/2) \pm i$ (with N a positive integer), then take the limit as $N \to \infty$ to derive the integral representation

$$\vartheta(t) = \int_{-\infty-i}^{\infty-i} \frac{e^{-\pi z^2 t}}{e^{2\pi i z} - 1}\, dz - \int_{-\infty+i}^{\infty+i} \frac{e^{-\pi z^2 t}}{e^{2\pi i z} - 1}\, dz$$

for the function $\vartheta(t)$.

(b) In this representation, expand the factor $(e^{2\pi iz}-1)^{-1}$ as a geometric series in $e^{-2\pi iz}$ (for the first integral) and as a geometric series in $e^{2\pi iz}$ (for the second integral). Evaluate the resulting infinite series, rigorously justifying all steps, to obtain an alternative proof of the functional equation (15).

44. (a) Reprove Theorem 36 (the "toy Riemann hypothesis" — the result that the Riemann zeta function has no zeros on the line $\mathrm{Re}(s)=1$) by considering the behavior of

$$Y = \mathrm{Re}\left[-3\frac{\zeta'(\sigma)}{\zeta(\sigma)} - 4\frac{\zeta'(\sigma+it)}{\zeta(\sigma+it)} - \frac{\zeta'(\sigma+2it)}{\zeta(\sigma+2it)}\right]$$

for $t \in \mathbb{R}\setminus\{0\}$ fixed and $\sigma \searrow 1$, instead of the quantity

$$X = \log|\zeta(\sigma)^3\zeta(\sigma+it)^4\zeta(\sigma+2it)|.$$

Use the series expansion

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

where $\Lambda(n)$ is von Mangoldt's function (equal to $\log p$ if $n=p^k$ is a prime power, and 0 otherwise).

(b) Try to reprove the same theorem in yet a third way by considering

$$Z = \log|\zeta(\sigma)^{10}\zeta(\sigma+it)^{15}\zeta(\sigma+2it)^6\zeta(\sigma+3it)|,$$

and attempting to repeat the argument involving expanding the logarithm in a power series and deducing that $Z \geq 0$. Does this give a proof of the theorem? If not, what goes wrong?

Hint. $(a+b)^6 = a^6+6a^5b+10a^4b^2+15a^3b^3+10a^2b^4+6ab^5+b^6$.

45. Define arithmetic functions taking an integer argument n, as follows:

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n = p_1p_2\cdots p_k \text{ is a product of } k \text{ distinct primes,} \\ 0 & \text{otherwise,} \end{cases}$$

(the Möbius μ-function),

$$d(n) = \sum_{d|n} 1, \qquad \text{(the number of divisors function),}$$

$$\sigma(n) = \sum_{d|n} d, \qquad \text{(the sum of divisors function),}$$

$$\phi(n) = \#\{1 \leq k \leq n-1 \,:\, \gcd(k,n)=1\}, \quad \text{(the Euler totient function),}$$

$$\Lambda(n) = \begin{cases} \log p & \text{if } n=p^k,\ p \text{ prime,} \\ 0 & \text{otherwise,} \end{cases} \qquad \text{(the von Mangoldt } \Lambda\text{-function).}$$

We saw that the zeta function and its logarithmic derivative have the Dirichlet series representations

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

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

Use the Euler product formula for the zeta function or other elementary manipulations to prove the following identities (valid for $\mathrm{Re}(s) > 1$):

$$\zeta'(s) = -\sum_{n=1}^{\infty} \log n \cdot n^{-s},$$

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

$$\frac{\zeta(s)}{\zeta(2s)} = \sum_{n=1}^{\infty} |\mu(n)| n^{-s}.$$

Other famous Dirichlet series representations you may want to think about or look up are

$$\zeta(s)^2 = \sum_{n=1}^{\infty} d(n) n^{-s},$$

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

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

46. Sendov's conjecture, an elementary statement in complex analysis proposed by the mathematician Blagovest Sendov in 1959 and still open today, is the claim that if $p(z) = (z - z_1) \ldots (z - z_n)$ is a complex polynomial whose roots z_j, $j = 1, \ldots, n$ all lie in the closed unit disc $|z| \leq 1$, then for each root z_j there is a root α of the derivative $p'(z)$ for which $|z_j - \alpha| \leq 1$.

(a) Prove the conjecture for the case $n = 2$ of quadratic polynomials.

(b) Prove that in the inequality $|z_j - \alpha| \leq 1$, if the number 1 is replaced by any smaller number then the claim is false.

(c) Prove the conjecture for the case $n = 3$ of cubic polynomials. (This is not a trivial result; for one possible proof, see the paper [2].

Suggested topics for course projects

(a) **Complex analysis and self-avoiding random walks.** Let a_n denote the number of *self-avoiding walks* of length n in the square lattice $\mathbb{Z}^2$ (that is, lattice paths with n steps) starting from $(0,0)$. It is known that $a_n^{1/n} \xrightarrow[n\to\infty]{} \mu$ for some number μ, approximately equal to 2.638. No closed form formula is known, or even suspected to exist, for μ. However, for the hexagonal lattice, a precise analogous result is known: if b_n denotes the number of self-avoiding walks of length n starting at the origin, then $b_n \xrightarrow[n\to\infty]{} \sqrt{2+\sqrt{2}} \approx 1.8477$. This is proved in the paper [3]. The proof is elementary and uses ideas from complex analysis in a crucial way. See https://en.wikipedia.org/wiki/Connective_constant for more details.

(b) **Complex analysis and the art of M. C. Escher.** See Fig. 1 on page 2 and the references [8], [10].

(c) **Elliptic functions and the characterization of meromorphic maps on the complex torus.** In Problem 24 we were interested in classifying the holomorphic functions $f : M \to N$ when M, N are Riemann surfaces; we answered this question for several interesting pairs of surfaces (such as $M = \mathbb{C}$, $N = \hat{C}$, etc.). The case when N is the Riemann sphere $\hat{\mathbb{C}}$ and M is a complex torus $\mathbb{C}/\Lambda$ where $\Lambda = \mathbb{Z} + \tau\mathbb{Z}$ for some complex number $\tau \in \mathbb{C} \setminus \mathbb{R}$ is an especially interesting example of such a question: in that case the relevant family turns out to be the family of elliptic functions (also known as doubly periodic functions). Some places where you can read about this topic are [11, Ch. 9] and [1, Ch. 1].

(d) **Sendov's conjecture.** This conjecture is described in Problem 46, and can be a suitable topic for a project.

(e) **Advanced topic in analytic combinatorics.** In the area of analytic combinatorics, one uses asymptotic methods in complex analysis (as well as real and harmonic analysis) to study the asymptotic behavior of integer sequences (or more generally integer arrays, probability distributions and other quantities) arising in combinatorics, similarly to the discussion in Section 17. The book [5] is an excellent reference for this subject (see also [6]), and provides many topics suitable for a project.

(f) **The Jordan curve theorem.** The Jordan curve theorem states that a continuous, simple planar curve separates the plane into two connected components, precisely one of which is unbounded. The theorem plays an important role in complex analysis, discussed in Appendix B of [11]. Despite its intuitive nature it is not trivial to prove, and has several different proofs described in different sources, for example [11, Appendix B], [9], and [12]. A project could present one of those proofs, and discuss additional related topics such as the significance of the theorem for complex analysis, various generalizations of it (for example its higher-dimensional analogues), and related theorems and open problems in topology. For more details, see https://en.wikipedia.org/wiki/Jordan_curve_theorem, and this MathOverflow discussion.

References

[1] T. Apostol. Modular Functions and Dirichlet Series in Number Theory. Springer, 1990.

[2] G. L. Cohen, G. H. Smith.. A simple verification of Ilieff's conjecture for polynomials with three zeros. *Amer. Math. Monthly* 95 (1988), 734–737.

[3] H. Duminil-Copin, S. Smirnov. The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$. *Ann. Math.* 175 (2012), 1653–1665.

[4] H. M. Edwards. Riemann's Zeta Function. Dover Publications, 2001.

[5] P. Flajolet, R. Sedgewick. Analytic Combinatorics. Cambridge University Press, 2009.

[6] P. Flajolet, R. Sedgewick. Analytic Combinatorics — Online Course Materials. Online resource: https://ac.cs.princeton.edu/online/. Accessed March 20, 2020.

[7] A. Ivić. The Riemann Zeta-Function: Theory and Applications. Dover Publications, 2003.

[8] B. de Smit, H. W. Lenstra Jr. *Artful mathematics: the heritage of M.C. Escher*. *Notices Amer. Math. Soc.* **50** (2003), 446–457.

[9] R. Maehara. The Jordan curve theorem via the Brouwer fixed point theorem.. *Amer. Math. Monthly* 91 (1984), 641–643.

[10] D. Schattsschneider. The mathematical side of M. C. Escher. *Notices Amer. Math. Soc.* 57 (2010), 706–718.

[11] E. M. Stein, R. Shakarchi, Complex Analysis. Princeton University Press, 2003.

[12] H. Tverberg. A proof of the Jordan curve theorem. *Bull. London Math. Soc.* 12 (1980), 34–38.

[13] D. Zagier. Newman's short proof of the prime number theorem. *Amer. Math. Monthly* 104 (1997), 705–708.