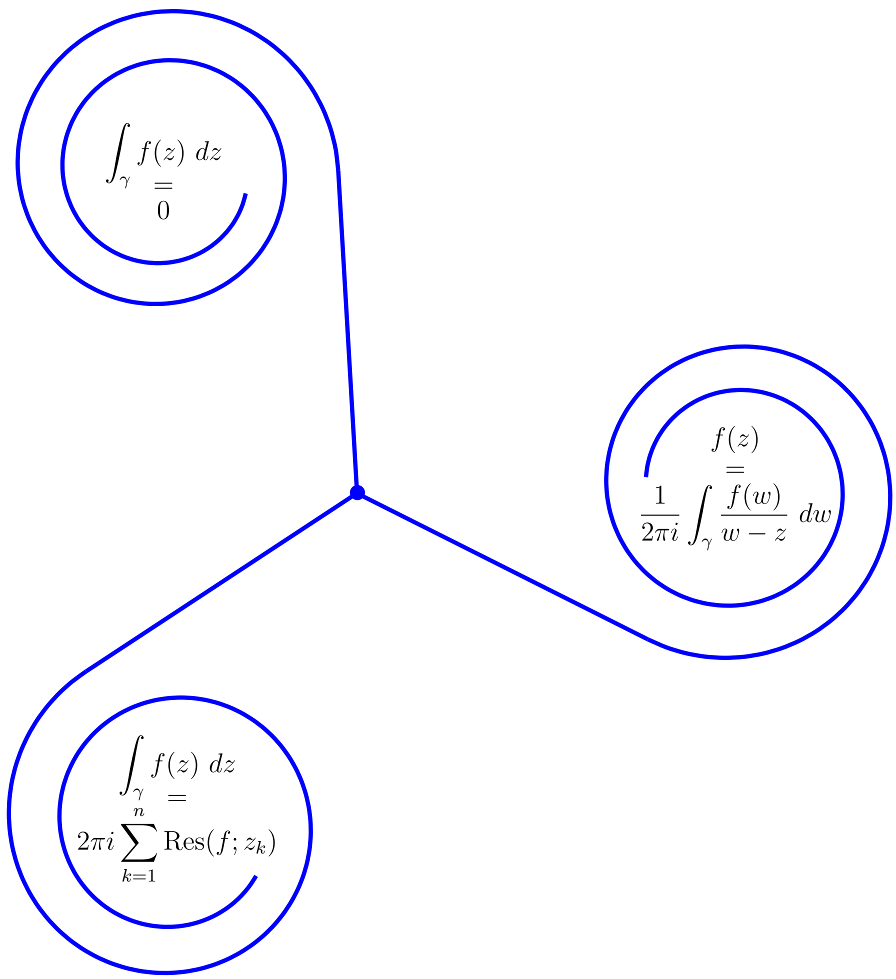
COMPLEX ANALYSIS

Timothy E. Faver

April 28, 2024



CONTENTS

OVERVIEW OF NOTES

These are lecture notes for a course in complex analysis. The prerequisite is multivariable calculus, not real analysis.

The notes contain three classes of problems.

(!) Problems marked (!) are meant to be attempted *immediately*. They will directly address or reinforce something that we covered in class. It will be to your great benefit to pause and work (!)-problems as you encounter them.

($\star$) Problems marked ($\star$) are intentionally more challenging and deeper than (!)-problems. The ($\star$)-problems will summarize and generalize ideas that we have discussed in class and give you broader, possibly more abstract perspectives. You should attempt the ($\star$)-problems on a second rereading of the lecture notes, after you have completed the (!)-problems and required problems from the textbook. As you prepare for an exam, you should definitely attempt all ($\star$)-problems in sections that will appear on that exam.

(+) Problems marked (+) are candidates for the portfolio project. These are meant to be more challenging than the (!)- and ($\star$)-problems and will take you deeper into calculations and proofs and make connections to concepts across and beyond the course. Some (+)-problems do presume knowledge of other classes (e.g., linear algebra, differential equations, real analysis, topology), but the majority do not. It is not necessary to do all (+)-problems in preparation for an exam; instead, you should look out for (+)-problems that you find interesting and exciting, as that will make the portfolio project more meaningful (and palatable) for you.

INTRODUCTION

The fundamental object of our study in this course is the **COMPLEX NUMBER**: an expression of the form $x + iy$, where x and y are real numbers and the symbol i satisfies $i^2 = -1$. We denote by $\mathbb{R}$ the set of all real numbers and by $\mathbb{C}$ the set of all complex numbers, so

$$\mathbb{C} = \left\{ x + iy \;\middle|\; x,\ y \in \mathbb{R},\ i^2 = -1 \right\}.$$

There are at least three algebraic and set-theoretic problems with this attempt at defining complex numbers.

Problem 1. What exactly does "expression" mean? This is (probably) not a formally defined mathematical term.

Problem 2. Why should there exist an "object" (for lack of a better word right now) i such that $i^2 = -1$? Certainly i cannot be a real number, as $x^2 \geq 0$ for any $x \in \mathbb{R}$.

Problem 3. If $y \in \mathbb{R}$ and i satisfies $i^2 = -1$ (whatever i is...), what does the "multiplication" iy mean? How is this operation defined? Likewise, how do we add x and iy if the only addition that we know works only on real numbers?

We defer any rigorous treatment of these questions (and there are several such approaches) to Appendix B.2. For now, we address a more immediate question: who cares?

We fundamentally care about complex numbers because they inherently arise in problems that ostensibly contain only real numbers. Perhaps the canonical example of such a problem is the quadratic equation.

Example. **(i)** The quadratic equation that most directly gives rise to complex numbers is surely

$$x^2 + 1 = 0.$$

Symbol-pushing of the follow-one's-nose sort yields the following:

$$x^2 + 1 = 0 \Longrightarrow x^2 = -1 \Longrightarrow x = \pm\sqrt{-1} \Longrightarrow x = \pm i.$$

However, our experience with real numbers tells us that only *nonnegative* real numbers have square roots, so $\sqrt{-1}$ is something new. Moreover, the complaint that we raised above appears here: if $x \in \mathbb{R}$, then $x^2 \geq 0$, so no real x can satisfy $x^2 = -1$.

(ii) More generally, any quadratic equation of the form

$$ax^2 + bx + c = 0$$

will have complex, nonreal roots if $b^2 - 4ac < 0$.

(iii) Quadratic equations with complex roots play a central role in two related applications. First, the **CHARACTERISTIC EQUATION** of the ordinary differential equation

$$my'' + by' + \kappa y = 0$$

is the quadratic equation

$$m\lambda^2 + b\lambda + \kappa = 0.$$

The existence of complex, nonreal roots for this quadratic equation generates oscillatory (sinusoidal) solutions for the ODE above. Note that physically m, b, and κ are often taken as nonnegative parameters (with m and κ usually *positive*), and so the "data" of this differential equation is very much real.

(iv) Second, the **characteristic equation** (same name as above, different setting) of the matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

is the quadratic equation

$$\lambda^2 - (a+d)\lambda + (ad - bc) = 0.$$

The roots of this quadratic equation are the "eigenvalues" of the matrix above, and the minimal data of eigenvalues control a wide variety of useful properties of this matrix.

Complex numbers also enter ostensibly real-valued problems via external instruments (which, at first glance, might appear a bit artificial). Fourier analysis offers a bevy of such instruments.

Example. Suppose that for $y \in \mathbb{R}$, we define the **complex exponential** e^{iy} to be

$$e^{iy} := \cos(y) + i\sin(y).$$

There are many good, natural reasons for doing this, which we will discuss later—starting from the very definition of what an exponential is, or should *do*.

(i) For a "suitably nice" function $f\colon \mathbb{R} \to \mathbb{R}$, we can represent f as an integral via the **Fourier inversion identity**:

$$f(x) = \int_{-\infty}^{\infty} \mathfrak{F}[f](k)e^{ikx}\,dk, \qquad \mathfrak{F}[f](k) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(t)e^{-ikt}\,dt,$$

with $\mathfrak{F}[f]$ as the **Fourier transform** of f. These improper integrals may look much worse than just f, but they turn out to be valuable means of encoding actions on and properties of f, in particular how f behaves in differential equations and certain measures of "size" of f. For example, if we assume that we can "differentiate under the integral" (which, like "suitably nice," can be made rigorous), then

$$f'(x) = \frac{d}{dx}\int_{-\infty}^{\infty} \mathfrak{F}[f](k)e^{ikx}\,dk = \int_{-\infty}^{\infty} \frac{\partial}{\partial x}[\mathfrak{F}[f](k)e^{ikx}]\,dk = \int_{-\infty}^{\infty} ik\mathfrak{F}[f](k)e^{ikx}\,dk,$$

and so "on the Fourier side," differentiating (an analytic process involving limits) corresponds to multiplying $\mathfrak{F}[f](k)$ by ik (an algebraic process that hopefully is simpler than taking limits).

(ii) For a "suitably nice" 2π-periodic function $f\colon [-\pi,\pi] \to \mathbb{R}$, we can expand f as a **Fourier series**:

$$f(x) = \sum_{k=-\infty}^{\infty} \widehat{f}(k)e^{ikx} := \lim_{n\to\infty} \sum_{k=-n}^{n} \widehat{f}(k)e^{ikx}, \qquad \widehat{f}(k) := \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t)e^{-ikt}\, dt,$$

with $\widehat{f}(k)$ as the kth **Fourier coefficient** of f for integers k. This morally has much in common with Fourier inversion and transforms of functions on $\mathbb{R}$, but one advantage here is that essentially all important properties of f can be recovered from the *countable* collection of Fourier coefficients $\widehat{f}(k)$, which is much less data than having to manage every single value of f on the uncountable set $[-\pi,\pi]$. Another advantage is that, term-by-term, complex exponentials are extremely easy to manipulate algebraically and analytically.

Any control over either Fourier transforms or series must rest on a firm understanding of the complex exponential.

Internally, problems that are posed with real numbers can sometimes be "extended" to involve complex numbers. Such extensions often elicit features of these problems that are "invisible" when considered only from the real point of view.

Example. **(i)** The function

$$f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto \frac{1}{1+x^2}$$

is **infinitely differentiable** in the sense that its kth derivative $f^{(k)}$ exists for any integer $k \geq 0$. Moreover, f is **real analytic** on the interval $(-1,1)$ in the sense that its **Taylor series** centered at 0, which is

$$S(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k = \sum_{j=0}^{\infty} (-1)^j x^{2j},$$

converges to $f(x)$ for any $|x| < 1$, i.e., $f(x) = S(x)$ on $(-1,1)$. However, this Taylor series diverges for $|x| \geq 1$. In light of the apparently good behavior of f on $\mathbb{R}$, this should seem surprising. What deeper property of or mechanism within f, beyond routine applications of series convergence tests from calculus, restricts the convergence of this Taylor series to just $(-1,1)$?

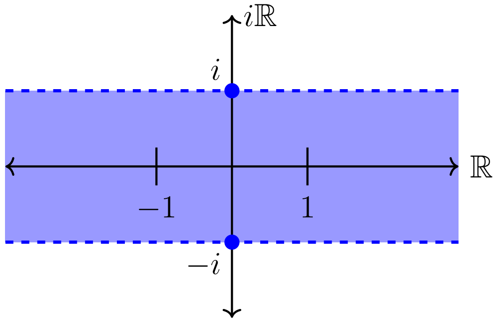
"Extending" f to $\mathbb{C}$, or as much of $\mathbb{C}$ as possible, suggests an answer. Define

$$\widetilde{f}\colon \mathbb{C} \setminus \{\pm i\} \to \mathbb{C}\colon z \mapsto \frac{1}{1+z^2}$$

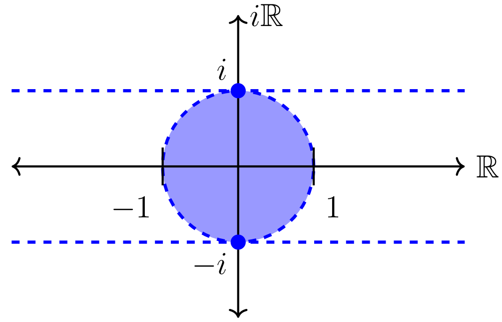
We have $\widetilde{f}(x) = f(x)$ for all $x \in \mathbb{R}$, and so $\widetilde{f}$ "extends" f; we also expect that $\widetilde{S}$ is the Taylor series of $\widetilde{f}$ centered at 0. (Strictly speaking, we have not yet discussed how to divide complex numbers, so this is a bit premature; likewise, we have not discussed what the convergence of a series of complex numbers should mean.)

We will eventually show that if $S(r)$ converges for some real number $r > 0$, then $\widetilde{f}$ must be differentiable on the "ball" $\{x + iy \mid x^2 + y^2 < r^2\}$, whatever "differentiable" means for a function of a complex variable. But surely there is no way to make $\widetilde{f}$ differentiable at $\pm i$ due to the division by 0 there, and so $S(r)$ must diverge for $r > 1$. This shows that the radius of convergence of S is 1 *without* using any convergence tests but *by* using deeper properties of f itself.

The faults of $\widetilde{f}$ lie not in its behavior at ± 1, or, indeed, anywhere on $\mathbb{R}$, or even on any "strip" of the form $\{x + iy \mid x \in \mathbb{R},\ |y| < b\}$ for $0 < b < 1$. Rather, the problems—really, the "poles"—appear at $\pm i$, which are, of course, invisible from a real perspective. When z is "close" to these poles, the values $f(z)$ "blow up" in a way that we will later make precise and that destroys differentiability there.



$\widetilde{f}$ is well-behaved on this "strip"



The Taylor series converges only on this "ball."

(ii) In somewhat the opposite direction from the previous situation, the function

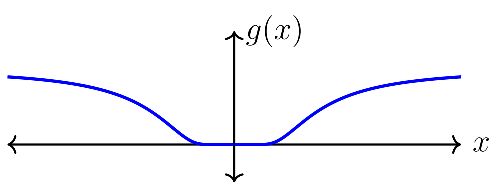
$$g\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto \begin{cases} e^{-1/x^2}, & x \neq 0 \\ 0, & x = 0 \end{cases}$$

is infinitely differentiable, but one can, with some effort, calculate that $g^{(k)}(0) = 0$ for all k. Thus the Taylor series for g converges on $\mathbb{R}$ (always to 0) but only to $g(x)$ at $x = 0$.

When we extend g to $\mathbb{C}$ as

$$\widetilde{g}\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto \begin{cases} e^{-1/z^2}, & z \neq 0 \\ 0, & z = 0, \end{cases}$$

it is possible to "approach" 0 from multiple directions in the two-dimensional complex plane and along the way observe much more "nervous" and "erratic" behavior in $\widetilde{g}$ than the superbly tame graph of g just on $\mathbb{R}$ below suggests.



(iii) The much-touted Fourier transform from the previous example requires computing improper integrals like

$$\int_{-\infty}^{\infty} f(x)e^{-ikx}\ dx = \int_{-\infty}^{\infty} f(x)\cos(kx)\ dx + i\int_{-\infty}^{\infty} f(x)\sin(kx)\ dx.$$

Quite frequently, integrands of this form have no antiderivatives in terms of elementary functions, and so we cannot use the fundamental theorem of calculus and the definition of the improper integral to evaluate them. For example, to calculate the Fourier transform of the function $f(x) = 1/(1+x^2)$ that we just discussed, we would need to evaluate

$$\int_{-\infty}^{\infty} \frac{\cos(kx)}{1+x^2}\ dx,$$

since the integral involving the sine vanishes due to the oddness of the integrand. Nothing from calculus prepares us to calculate this improper integral (at least for $k \neq 0$).

It turns out that we can relate the one-dimensional improper integral above to a two-dimensional "line integral" over a "curve" lying in a "strip" of width less than 1 centered on the real line (i.e., like the strip drawn above, just narrower). Then methods of complex analysis allow us to evaluate this line integral and extract the original improper integral from that.

(iv) The integrals that appear in the Fourier transform are **OSCILLATORY INTEGRALS**: the integrands involve the product of a given function against a complex exponential. It may be desirable to estimate those integrals if we cannot evaluate them precisely (and often even if we can—estimates are usually better than inequalities in analysis). If $f\colon \mathbb{R} \to \mathbb{R}$ and its first n derivatives "vanish at infinity" in the sense that $\lim_{x\to\pm\infty} f^{(k)}(x) = 0$ "sufficiently fast" for $0 \le k \le n$, and if ϵ, $\omega > 0$, then one can integrate by parts and prove an estimate like

$$\left|\int_{-\infty}^{\infty} f(x)e^{i\omega x/\epsilon}\ dx\right| \le C_n\left(\frac{\epsilon}{\omega}\right)^n.$$

This says that the oscillatory integral on the left is quite small in the parameter ϵ.

However, if we can extend f to a strip like $\{x+iy \mid x,\ y \in \mathbb{R},\ |y| < b\}$ for some $b > 0$, and if the extension of f is "particularly nice" on this strip, then some further line integral techniques enable us to eke out a vastly better estimate of the form

$$\left|\int_{-\infty}^{\infty} f(x)e^{i\omega x/\epsilon}\ dx\right| \le Ce^{-\omega b/\epsilon}.$$

(Try graphing $e^{-1/\epsilon}$ and ϵ^n for different integers n. Which gets smaller faster?)

This is where we finished on Monday, January 8, 2024.

While we will touch on some of these applications of complex numbers in this course (in particular, solving certain polynomial equations and developing some specialized properties of transform theory), our primary goal will be to learn *how calculus works* when the functions involved are defined for complex inputs and allowed to have complex outputs. In short, it works very well. To do this, we will proceed (at a somewhat accelerated pace) through the same journey that we took learning real-variable calculus. We will begin with the precalculus of arithmetic, geometry, and algebra and build a small bestiary of functions; then we will treat the differential calculus, touching on limits and continuity as needed; and finally, gloriously we will study the integral calculus, perhaps to a depth that we never plumbed in real-variable calculus. After that, a multiverse of possibilities opens, but throughout we will return to the three complex leitmotifs of **ALGEBRA**, **GEOMETRY**, and **ANALYSIS**:

1. Algebra: the complex numbers $\mathbb{C}$ contain an element i such that $i^2 = -1$. In equally technical language, *the additive inverse,* -1*, of the multiplicative identity,* 1*, has a square root.* This is simply not true in $\mathbb{R}$.

2. Geometry: the complex numbers are inherently two-dimensional, and while all real numbers are complex numbers, the best and most versatile results in the course will come when we work with subsets of $\mathbb{C}$ that contain nonreal elements and possess certain nice geometric properties (all of which can be stated in excruciatingly precise set-theoretic detail but drawn using kindergarten-accessible shapes).

3. Analysis: it is possible to define and compute limits of functions that are defined on subsets of $\mathbb{C}$ just as we did for functions defined on subsets of $\mathbb{R}$. Limits give us everything in calculus—continuity, derivatives, integrals, and much more. Many symbolic results are the same (and often the proofs are the same—just replace x with z) until they are not: frequently, the two-dimensional geometry of $\mathbb{C}$ introduces something surprising.

1. PRECALCULUS

1.1. Arithmetic and geometry.

1.1.1. The (un)definition of complex numbers.

The logically correct way to discuss complex numbers would be to prove that they exist—if an object does not exist, how can we do math with it? However, such a proof probably would not convince us of anything that we do not already believe. So, we just spell out here the fundamental assumptions and conventions that we will use in everyday life in this course, and we defer to Appendix B.2 a more rigorous construction of the complex numbers (which itself hinges on some further fundamental assumptions and conventions about real numbers that we do not prove).

1.1.1 Undefinition. **(i)** *A* **COMPLEX NUMBER** *is an expression of the form $z = x + iy$, where x, $y \in \mathbb{R}$ and i satisfies $i^2 = -1$.*

(ii) *We denote the set of all complex numbers by $\mathbb{C}$.*

(iii) *The* **REAL PART** *of z is $\mathrm{Re}(z) = \mathrm{Re}(x + iy) := x$ and the* **IMAGINARY PART** *of z is $\mathrm{Im}(z) = \mathrm{Im}(x + iy) := y$.*

(iv) *If $z \in \mathbb{C}$ with $\mathrm{Im}(z) = 0$, then we may say that z is* **PURELY IMAGINARY**, *and we write*

$$i\mathbb{R} := \{iy \in \mathbb{C} \mid y \in \mathbb{R}\}.$$

If $z \in \mathbb{C}$ with $\mathrm{Im}(z) \neq 0$, then we may say that z is **NONREAL**.

(v) *Two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ are equal if and only if $x_1 = x_2$ and $y_1 = y_2$. That is, $z_1 = z_2$ if and only if $\mathrm{Re}(z_1) = \mathrm{Re}(z_2)$ and $\mathrm{Im}(z_1) = \mathrm{Im}(z_2)$.*

Previously we identified several problems with this kind of definition—the uncertainty surrounding the existence of an "object" i such that $i^2 = -1$ and the ambiguities of what the operations of addition and multiplication in the symbol $x + iy$ mean, if all we know is arithmetic in $\mathbb{R}$. However, more positively, there are at least three successful aspects of this definition.

Success 1. Every real number is a complex number, since any $x \in \mathbb{R}$ can be written as

$$x = x + 0 = x + (i \cdot 0),$$

if, again we interpret arithmetic as we expect.

Success 2. The object i itself is a complex number, if we interpret arithmetic once more in the natural way, since

$$i = 0 + i = 0 + (i \cdot 1).$$

Success 3. Every object that we have met in our prior lives that has purported itself to be a complex number *is* a complex number, per this definition, since those numbers probably looked like 42, or $3i$, or $1+2i$.

Here are some concrete, and hopefully wholly unsurprising, calculations that deploy the preceding notation.

1.1.2 Example. **(i)** $\mathrm{Re}(1+2i) = 1$ and $\mathrm{Im}(1+2i) = 2$.

(ii) $\mathrm{Re}(1) = 1$ and $\mathrm{Im}(1) = 0$.

(iii) $\mathrm{Re}(2i) = 0$ and $\mathrm{Im}(2i) = 2$.

Going forward, we will reserve letters like z and w for complex numbers (which may be real numbers); the letters x, y, u, and v will typically appear in conjunction with the real or imaginary parts of a complex number.

1.1.2. Addition and multiplication.

We should view the following calculations as purely formal, "follow our noses" exercises that operate under a fundamental assumption.

1.1.3 Hypothesis. *All arithmetic works exactly as it should if all quantities were real numbers, with the exception that the symbol i always satisfies $i^2 = -1$. (More precisely, arithmetic in $\mathbb{C}$ satisfies the field axioms ($\mathbb{R}$1).)*

1.1.4 Example. Let $z = 1+2i$ and $w = 3+4i$. Then the following computations should be valid.

(i) We group like terms to find

$$z + w = (1+2i) + (3+4i) = (1+3) + (2i+4i) = 4+6i.$$

Note that $\mathrm{Re}(z+w) = \mathrm{Re}(z) + \mathrm{Re}(w)$ and $\mathrm{Im}(z+w) = \mathrm{Im}(z) + \mathrm{Im}(w)$.

(ii) We distribute multiplication over addition to find

$$zw = (1+2i)(3+4i) = (1+2i)3 + (1+2i)4i = 3+6i+4i+8i^2 = 3+10i-8 = -5+10i.$$

Note that $\mathrm{Re}(zw) \neq \mathrm{Re}(z)\,\mathrm{Re}(w)$ and $\mathrm{Im}(zw) \neq \mathrm{Im}(z)\,\mathrm{Im}(w)$.

(iii) We distribute division ($=$ multiplication by the reciprocal) over addition to find

$$\frac{z}{5} = \frac{1+2i}{5} = \frac{1}{5} + \frac{2i}{5} = \frac{1}{5} + \left(\frac{2}{5}\right)i.$$

1.1.5 Problem ($\star$). Let z, $w \in \mathbb{C}$ and $a \in \mathbb{R}$. Extract from the preceding example the following general rules and formulas.

(i) $\mathrm{Re}(z+w) = \mathrm{Re}(z) + \mathrm{Re}(w)$ and $\mathrm{Im}(z+w) = \mathrm{Im}(z) + \mathrm{Im}(w)$.

(ii) $\mathrm{Re}(az) = a\,\mathrm{Re}(z)$ and $\mathrm{Im}(az) = a\,\mathrm{Im}(z)$.

(iii) Express $\mathrm{Re}(iaz)$ and $\mathrm{Im}(iaz)$ in terms of $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$.

(iv) Express $\mathrm{Re}(zw)$ and $\mathrm{Im}(zw)$ in terms of the real and imaginary parts of z and w.

The work above shows that if $z = x + iy \in \mathbb{C}$ and $a \in \mathbb{R} \setminus \{0\}$, then we expect

$$\frac{z}{a} = \frac{x}{a} + i\left(\frac{y}{a}\right).$$

What, however, does should the symbol z/w mean for $w \in \mathbb{C} \setminus \{0\}$? This is a situation that Hypothesis 1.1.3 does not fully cover.

For example, what meaning should we give to the expression

$$\frac{1+2i}{3+4i}?$$

Certainly this should be the same as

$$(1+2i)\left(\frac{1}{3+4i}\right).$$

So, what does

$$\frac{1}{3+4i}$$

mean? It should satisfy

$$\left(\frac{1}{3+4i}\right)(3+4i) = 1.$$

That is, if $w \in \mathbb{C}\setminus\{0\}$, then the symbol $1/w$ should denote the complex number satisfying

$$\left(\frac{1}{w}\right)w = 1,$$

just as it does when $w \in \mathbb{R} \setminus \{0\}$. Since complex numbers are uniquely determined by their real and imaginary parts, can we compute the real and imaginary parts of $1/w$ directly from the real and imaginary parts of w?

1.1.6 Problem (+). Yes. Here is the brute-force approach. (*Brute force is the best force.*) Let $w \in \mathbb{C} \setminus \{0\}$ and write $w = w_1 + iw_2$ and $1/w = m_1 + im_2$ for w_1, w_2, m_1, $m_2 \in \mathbb{R}$.

Show that $(m_1 + im_2)(w_1 + iw_2) = 1$ if and only if

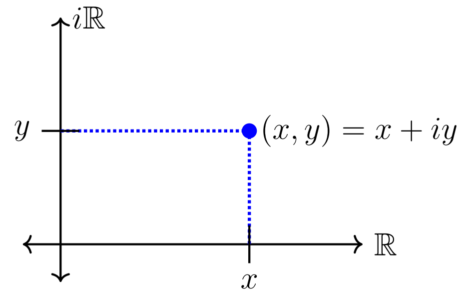
$$\begin{cases} w_1m_1 - w_2m_2 = 1 \\ w_2m_1 + m_2w_1 = 0. \end{cases}$$

With w_1 and w_2 given, this is a system of linear equations for m_1 and m_2. Solve this system.

Computing $1/w$ in terms of w will be more efficient with a new tool. To develop that tool, we will change our focus from arithmetic to geometry.

1.1.3. The modulus.

We started by (un)defining complex numbers as expressions of the form $x + iy$ for $x, y \in R$. This suggests identifying complex numbers with ordered pairs $(x, y) \in \mathbb{R}^2$. Set-theoretically, this leads to a number of messes, but it works out spectacularly for visualizations. Here is the key picture.



We will draw such pictures frequently and always call the horizontal axis the **REAL AXIS** and the vertical axis the **IMAGINARY AXIS**. Such a picture suggests that we can impute a notion of "size" or "length" to a complex number by thinking about the length of the line segment from the origin $(0, 0)$ to the ordered pair (x, y), which is, of course, $\sqrt{x^2 + y^2}$.

1.1.7 Definition. *Let $z = x + iy \in \mathbb{C}$. The* **MODULUS** *of z is*

$$|z| = |x + iy| := \sqrt{x^2 + y^2}.$$

1.1.8 Example. **(i)** $|1 + 2i| = \sqrt{1^2 + 2^2} = \sqrt{1 + 4} = \sqrt{5}$.

(ii) $|2i| = |0 + 2i| = \sqrt{0^2 + 2^2} = \sqrt{4} = 2$.

1.1.9 Remark. **(i)** *Throughout this course, we will assume that all nonnegative real numbers have unique square roots. That is, given $t \in \mathbb{R}$ with $t \geq 0$, there is a unique nonnegative number $s \geq 0$ such that $t = s^2$. We write $s = \sqrt{t}$. We will not write $s = t^{1/2}$; we will later show how, if the square root is taken for granted, we can construct nth roots of all nonzero complex numbers for any integer n.*

(ii) *Recall that the* **ABSOLUTE VALUE** *of $t \in \mathbb{R}$ is the real number*

$$|t| := \begin{cases} t, & t \geq 0 \\ -t, & t < 0. \end{cases}$$

It may appear that we are overworking our notation in using the absolute value symbol for the modulus. We are not, for the modulus of a real number $t = t + (i \cdot 0)$ is

$$|t| = |t + (i \cdot 0)| = \sqrt{t^2 + 0^2} = \sqrt{t^2}.$$

As stated just above, $\sqrt{t^2}$ is the unique nonnegative number s such that $s^2 = t^2$. Certainly if $t \geq 0$, then $s = t$ works; if $t < 0$, then $s = -t$ works. In either case, we have $\sqrt{t^2} = |t|$, where now by $|t|$ we mean the absolute value. Thus the absolute value and the modulus of a real number are the same.

1.1.10 Problem ($\star$). **(i)** Show that the modulus is "multiplicative" in the sense that

$$|zw| = |z||w|$$

for all z, $w \in \mathbb{C}$. [Hint: *compute the squares of both sides.*]

(ii) Show that $|\operatorname{Re}(z)| \leq |z|$ and $|\operatorname{Im}(z)| \leq |z|$ for all $z \in \mathbb{C}$. [Hint: *use the fact that the square root function is increasing on* $[0, \infty)$.] Draw a picture and interpret these inequalities geometrically; use the words "triangle" and "hypotenuse" in your interpretation.

The modulus $|z|$ of $z \in \mathbb{C}$ captures the distance from a point z to the origin 0; in particular, $|z| = |z - 0|$. More generally, since

$$|z - w| = \sqrt{[\operatorname{Re}(z) - \operatorname{Re}(w)]^2 + [\operatorname{Im}(z) - \operatorname{Im}(w)]^2},$$

the difference $|z - w|$ captures the distance between points z, $w \in \mathbb{C}$.

We claim that for $z_0 \in \mathbb{C}$ and $r > 0$, the set

$$\{z \in \mathbb{C} \mid |z - z_0| = r\}$$

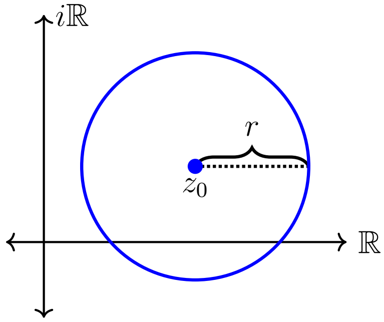
is the circle centered at z_0 of radius r; by the preceding discussion, this is the set of all points that lie at a distance $r > 0$ from z_0. Recall that in our prior lives, we probably wrote the equation of the circle centered at (a, b) of radius r as

$$(x - a)^2 + (y - b)^2 = r^2.$$

Here, we have

$$r^2 = |z - z_0|^2 = [\operatorname{Re}(z) - \operatorname{Re}(z_0)]^2 + [\operatorname{Im}(z) - \operatorname{Im}(z_0)]^2,$$

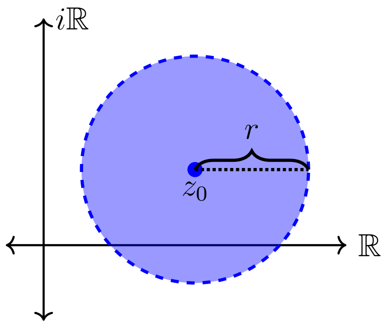
which is our prior equation for a circle.



We can also consider the "interior" or "inside" of the circle centered at $z_0 \in \mathbb{C}$ of radius $r > 0$. Since points *on* this circle lie at a distance r from z_0, points *inside* this circle should lie at a distance less than r from z_0. We define this precisely and introduce some convenient notation.

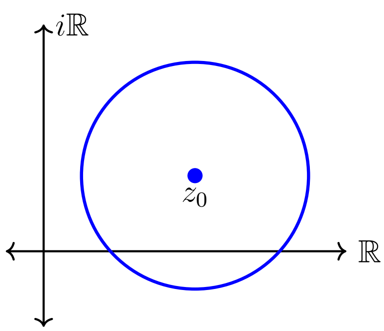
1.1.11 Definition. **(i)** *The* OPEN BALL *of radius* $r > 0$ *centered at* $z_0 \in \mathbb{C}$ *is*

$$\mathcal{B}(z_0; r) := \{z \in \mathbb{C} \mid |z - z_0| < r\}.$$



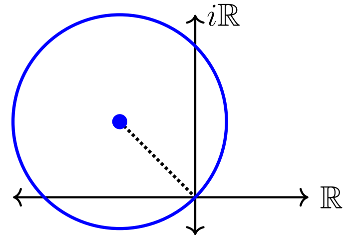
(ii) *The* CIRCLE *of radius* $r > 0$ *centered at* $z_0 \in \mathbb{C}$ *is*

$$\partial\mathcal{B}(z_0; r) := \{z \in \mathbb{C} \mid |z - z_0| = r\}.$$



The notational choice of $\partial\mathcal{B}(z_0; r)$ is meant to reflect the more general topological concepts of boundary, which we will not discuss in this course.

1.1.12 Example. Let $z_0 = -1 + i$, so $|z_0| = \sqrt{2}$. We draw $\mathcal{B}(-1+i; |-1+i|)$ and $\partial\mathcal{B}(-1+i; |-1+i|)$ below.



Since $|-1+i| > 1 = \mathrm{Im}(-1+i)$, the circle and the open ball extend into Quadrant III; likewise, since $|-1+i| > 1 = |\mathrm{Re}(-1+i)|$, the circle and the ball extend into Quadrant I. However, neither extends into Quadrant IV. Indeed, for the circle, suppose that $|z - (-1+i)| = |-1+i|$ with $z = x + iy$, $x > 0$, and $y < 0$. Then

$$\begin{aligned}\sqrt{2} = |(x+iy) - (-1+i)| = |(x+1) + i(y-1)| &= \sqrt{(x+1)^2 + (y-1)^2} \\ &= \sqrt{(x+1)^2 + (|y|+1)^2} > \sqrt{1+1} = \sqrt{2}.\end{aligned}$$

The second inequality holds because $x > 0$ and $|y| > 0$, and so we have a contradiction. This analytically justifies the dotted line in our sketches: the closest point on the circle to Quadrant IV is the origin.

One of the most important tasks that we will undertake repeatedly in this course will be to show that two quantities are the same, with the hope being that one expression is either simpler or more informative than the other. That is, we will have two complex numbers z_1 and z_2, and we will want to show that $z_1 = z_2$. By Undefinition 1.1.1, this is the same as showing $\mathrm{Re}(z_1) = \mathrm{Re}(z_2)$ and $\mathrm{Im}(z_1) = \mathrm{Im}(z_2)$, but this is not always efficient. Here is a better way.

1.1.13 Problem ($\star$). Let z_1, $z_2 \in \mathbb{C}$. Show that $z_1 = z_2$ (as undefined in Undefinition 1.1.1 if and only if $|z_1 - z_2| = 0$. [Hint: *explain why* $|w| = 0$ *if and only if* $w = 0$.]

1.1.4. The conjugate (and the modulus, continued).

Symmetries and "reflections" are key tools throughout mathematics. For example, we can view multiplying a complex number by -1 as "reflecting across the origin."

1.1.14 Problem (!). Explain this. Draw pictures.

It turns out that "reflecting across the real axis" is a very useful tool, too. If $z = x + iy$, then its reflection across the real axis should be the number $x - iy$. This number has a name.

1.1.15 Definition. *The* **COMPLEX CONJUGATE** *of* $z = x + iy$ *is*

$$\overline{z} = \overline{x + iy} := x - iy.$$

1.1.16 Example. **(i)** $\overline{1 + 2i} = 1 - 2i$.

(ii) $\overline{2i} = \overline{0 + 2i} = 0 - 2i = -2i$.

1.1.17 Problem (!). Let z, $w \in \mathbb{C}$. Show that

$$\overline{z + w} = \overline{z} + \overline{w} \quad \text{and} \quad \overline{zw} = \overline{z}\,\overline{w}.$$

1.1.18 Problem (!). Draw a picture with the following elements. Plot some $z = x + iy$ in the plane with $x > 0$ and $y > 0$. Then plot $-z$ and $\overline{z}$ and check that these points are the reflections that are claimed above. Finally, develop a formula for the reflection of a point across the *imaginary* axis and plot that, too.

1.1.19 Problem (+). **(i)** Check that

$$\text{Re}(z) = \frac{z + \overline{z}}{2} \quad \text{and} \quad \text{Im}(z) = \frac{z - \overline{z}}{2i}$$

for all $z \in \mathbb{C}$.

(ii) (Presumes knowledge of linear algebra.) Show that

$$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}\begin{pmatrix} z \\ \overline{z} \end{pmatrix} = \begin{pmatrix} \text{Re}(z) \\ \text{Im}(z) \end{pmatrix}.$$

Invert this matrix and use that inverse to solve for z and $\overline{z}$ in terms of $\text{Re}(z)$ and $\text{Im}(z)$. Do you get what you expected?

Our first significant use of the conjugate will be to give a new formula for the modulus.

1.1.20 Theorem. *Let* $z \in \mathbb{C}$. *Then*

$$|z|^2 = z\overline{z} \quad \textit{and} \quad |z| = \sqrt{z\overline{z}}.$$

Proof. Write $z = x + iy$, so $|z| = \sqrt{x^2 + y^2}$ and $|z|^2 = x^2 + y^2$. We compute directly

$$\begin{aligned} z\overline{z} = (x+iy)\big(\overline{x+iy}\big) = (x+iy)(x-iy) = (x+iy)x + (x+iy)(-iy) &= x^2 + iyx - ixy - i^2y^2 \\ &= x^2 - (-1)y^2 = x^2 + y^2 = |z|^2. \end{aligned}$$

Since $|z| \geq 0$, this shows that $z\overline{z}$ is real and nonnegative, so we may take its square root and find $|z| = \sqrt{z\overline{z}}$. ■

1.1.21 Problem (⋆). Use Theorem 1.1.20 to give a quick proof that $|zw| = |z||w|$ for all z, $w \in \mathbb{C}$. This proof should not use the real and imaginary parts of z and w.

1.1.22 Problem (!). Now use the definition of the conjugate and the modulus to give a (probably) longer proof that

$$\overline{z+w} = \overline{z} + \overline{w} \quad \text{and} \quad |\overline{z}| = |z|$$

for all z, $w \in \mathbb{C}$. This proof will probably need to use the real and imaginary parts of z and w.

Much of analysis hinges on carefully estimating quantities "from above" or "from below." When performing estimates in this course, we will almost always use one of the following two inequalities.

1.1.23 Theorem (Triangle inequality). *Let z, $w \in \mathbb{C}$. Then*

$$|z+w| \le |z| + |w| \quad \textit{and} \quad \big||z| - |w|\big| \le |z-w|.$$

The first inequality above is usually called the **TRIANGLE INEQUALITY**, *while the second inequality is the* **REVERSE TRIANGLE INEQUALITY**.

1.1.24 Problem (⋆). Let $z = 1 + 2i$ and $w = 3 + 4i$. Check that the triangle inequality holds by computing $|z+w|$, $|z|$, and $|w|$. Plot z, w, and $z+w$ and see if you can visualize the triangle inequality. Where does a "triangle" enter the picture? [Hint: *consider the triangles whose vertices are first the origin, z, and $z+w$ and next the origin, w, and $z+w$.*]

1.1.25 Problem (+). Prove Theorem 1.1.23 as follows.

(i) Explain why it suffices to show

$$|z+w|^2 \le \big(|z| + |w|\big)^2.$$

(ii) Compute

$$|z+w|^2 = |z|^2 + |w|^2 + w\overline{z} + z\overline{w}$$

and explain why it now suffices just to show

$$w\overline{z} + z\overline{w} \le 2|z||w|. \tag{1.1.1}$$

(iii) For A, $B \in \mathbb{R}$, prove the auxiliary inequality

$$2AB \le A^2 + B^2. \tag{1.1.2}$$

[Hint: *what is $A^2 - 2AB + B^2$?*]

(iv) Suppose $z = x + iy$ and $w = a + ib$ with x, y, a, $b \in \mathbb{R}$. Compute

$$w\overline{z} + z\overline{w} = 2(ax + by).$$

Since $ax + by \leq |a||x| + |b||y|$, explain why to obtain (1.1.1), it now suffices to show just

$$|a||x| + |b||y| \leq |z||w|. \tag{1.1.3}$$

(v) Prove (1.1.3). [Hint: *square both sides of the inequality.*]

(vi) To prove the reverse triangle inequality, first argue that it is equivalent to

$$-|z - w| \leq |z| - |w| \leq |z - w|. \tag{1.1.4}$$

To prove the second inequality in (1.1.4), "add zero" by writing

$$|z| = |z + 0| = |z - w + w|$$

and then use the ordinary triangle inequality to conclude

$$|z| - |w| \leq |z - w|. \tag{1.1.5}$$

(vii) Explain why (1.1.5) implies

$$|w| - |z| \leq |w - z| = |z - w|, \tag{1.1.6}$$

and from (1.1.6) conclude the first inequality in (1.1.4).

1.1.5. Division.

The union of the conjugate and the modulus is the tool that we need to develop an effective notion of division of complex numbers. Recall that if $z \in \mathbb{C} \setminus \{0\}$, then the **RECIPROCAL** of z should be the complex number $1/z$ such that

$$\left(\frac{1}{z}\right) z = 1.$$

And recall that we have the identity $|z|^2 = z\overline{z}$. Formally manipulating the symbols, we arrive at

$$\frac{1}{z} = \left(\frac{1}{|z|^2}\right)\overline{z} = \frac{\text{Re}(z)}{|z|^2} + i\left(-\frac{\text{Im}(z)}{|z|^2}\right). \tag{1.1.7}$$

Note that on the right we have division by real numbers, i.e., multiplication by the reciprocals of real numbers.

Now we check that (1.1.7) really gives the reciprocal of z:

$$\left[\left(\frac{1}{|z|^2}\right)\overline{z}\right] z = \frac{1}{|z|^2}(\overline{z}z) = \frac{1}{|z|^2}|z|^2 = 1.$$

The first equality above was the associativity of multiplication (which we are assuming is true for complex numbers by Hypothesis 1.1.3) and the second was Theorem 1.1.20.

This is where we finished on Wednesday, January 10, 2024.

1.1.26 Example. We return to our prior problem of computing $z = (1+2i)/(3+4i)$. To be clear, by "computing" we mean that we want to give simple expressions for $\mathrm{Re}(z)$ and $\mathrm{Im}(z)$.

First, we have

$$\frac{1}{3+4i} = \frac{\overline{3+4i}}{|3+4i|^2} = \frac{3-4i}{9+16} = \frac{3-4i}{25}.$$

Then

$$\frac{1+2i}{3+4i} = (1+2i)\left(\frac{1}{3+4i}\right) = (1+2i)\left(\frac{3-4i}{25}\right) = \frac{(1+2i)(3-4i)}{25} = \frac{3-4i+6i-8i^2}{25} = \frac{11+2i}{25}.$$

From this, we can plainly see that $\mathrm{Re}(z) = 11/25$ and $\mathrm{Im}(z) = 2/25$, which probably was not at all obvious from the starting expression of $z = (1+2i)/(3+4i)$.

Of course, in practice when computing Now that we have an adequate notion of division of complex numbers, we can define integer powers.

1.1.27 Definition. *Let $z \in \mathbb{C}$.*

(i) *We define $z^0 := 1$.*

(ii) *Let $k \geq 1$ be an integer. We define z^k recursively by*

$$z^k := \begin{cases} z, & k = 1 \\ z^{k-1}z, & k \geq 2. \end{cases}$$

(iii) *Suppose $z \neq 0$. Then we define*

$$z^{-1} := \frac{1}{z} = \frac{\overline{z}}{|z|^2}.$$

If $k \leq -2$, we define

$$z^k := (z^{|k|})^{-1}.$$

From this definition, one can show that the usual rules

$$z^{m+n} = z^m z^n \quad \text{and} \quad (z^m)^n = z^{mn}$$

hold for m, $n \in \mathbb{Z}$. We will not attempt to define fractional or rational (let alone *irrational*) powers of complex numbers for quite some time; indeed, they behave rather strangely.

1.1.28 Example. We have

$$i^0 = 1, \qquad i^1 = i, \qquad i^2 = -1, \qquad i^3 = i^2 i = -i, \qquad \text{and} \qquad i^4 = i^2 i^2 = (-1)^2 = 1.$$

Also,

$$i^{-1} = \frac{1}{i} = \frac{i}{i^2} = \frac{i}{-1} = -i.$$

Indeed,

$$(-i)i = -i^2 = -(-1) = 1,$$

so $-i$ is the multiplicative inverse of i.

1.1.29 Problem (!). **(i)** Let $k \in \mathbb{Z}$. Explain how the values of i^k are "4-periodic" in k, i.e., $i^k = i^{k+4}$ for all $k \in \mathbb{Z}$.

(ii) Compute i^{1977}, i^{1980}, and i^{1983}.

1.2. Functions.

This course is really about functions, so much so that complex analysis classes are sometimes (archaically) titled "Functions of a Complex Variable," and the whole subject is sometimes called (again, archaically) "function theory." What is a function? Our instinct might be to equate "function" with "formula"—surely the object $f(z) = 2z$ is a function. This string of symbols pairs each $z \in \mathbb{C}$ with its double $2z$. But there are plenty of other "pairings" of numbers that do not have such transparent formulas—for $z \in \mathbb{C}$, let $g(z)$ be the smallest integer that is greater than or equal to $\text{Re}(z)$. For example, $g(1/2) = 1$ and $g(2i) = 0$, but we do not really have an "algebraic" formula for the value of $g(z)$ in general.

The right definition of function hinges on pairings, not formulas. And not just any pairing: we know that each element in the domain of a function must be paired with exactly one output. We cannot omit elements of the domain from the pairings, and we cannot pair the same element of the domain with two outputs. So, here is a first stab at a definition of function.

1.2.1 Undefinition. *Let A and B be sets. A* FUNCTION FROM A TO B *is a rule that pairs each element of A with exactly one element of B.*

This morally resembles our first effort at defining complex numbers in Undefinition 1.1.1. It is exactly the right idea, and the one that everyone uses on a day-to-day basis, but it lacks clarity. What is a "rule"? What do we mean by "pairing"? As with complex numbers, we can clean things up by introducing the language of ordered pairs.

1.2.2 Definition. *Let A and B be sets. A* **FUNCTION** f **FROM** A **TO** B *is a set of ordered pairs with the following properties.*

(i) *If $(z, w) \in f$, then $z \in A$ and $w \in B$.*

(ii) *For each $z \in A$, there is a unique $w \in B$ such that $(z, w) \in f$.*

We often use the notation $f\colon A \to B$ to mean that f is a function from A to B; strictly speaking, we may want to think of a function from A to B as an ordered triple (f, A, B), where f satisfies the two properties above.

If $(z, w) \in f$, then we write $w = f(z)$. The set A is the **DOMAIN** *of f, and the set B is the* **CODOMAIN** *of f. The* **IMAGE** *or* **RANGE** *of f is the set*

$$f(A) := \{f(z) \mid z \in A\}.$$

More generally, if $E \subseteq A$, then the **IMAGE OF** E **UNDER** f *is*

$$f(E) := \{f(z) \mid z \in E\}.$$

1.2.3 Example. Let $f := \{(1, i), (2, -1), (3, -i), (4, i)\}$.

(i) Then f is a function from $A = \{1, 2, 3, 4\}$ to $B = \{i, -i, 1, -1\}$, and the range of f is B.

(ii) But f is also a function from A to $\mathbb{C}$; this indicates that while the range of a function is always fixed, the codomain can change depending on our desired point of view.

(iii) However, f is not a function from A to $\mathbb{R}$, since $i \notin \mathbb{R}$.

(iv) Likewise, f is not a function from $\{1, 2, 3, 4, 5\}$ to B, since there is no $w \in B$ such that $(5, w) \in f$.

(v) And f is not a function from $\{1, 2, 3\}$ to B, since $(4, 1) \in f$, but $4 \notin \{1, 2, 3\}$.

1.2.4 Problem ($\star$). Suppose that f, $g\colon A \to B$ are functions. Show that $f = g$ if and only if $f(z) = g(z)$ for all $z \in A$. [Hint: *the equality $f = g$ means that f and g are equal as sets of ordered pairs.*]

We do not need formulas to define functions. Example 1.2.3 is a good initial illustration of this, as it gives us a perfectly good function just defined as a set of ordered pairs. Our transcendence of formulas will eventually reach its zenith when we define functions by integrals *without* evaluating those integrals as we eventually always did in calculus. We might summarize our attitude toward formulas in the following profession.

1.2.5 Hypothesis (Analyst's creed). *Having a formula for something is not the same as understanding that thing.*

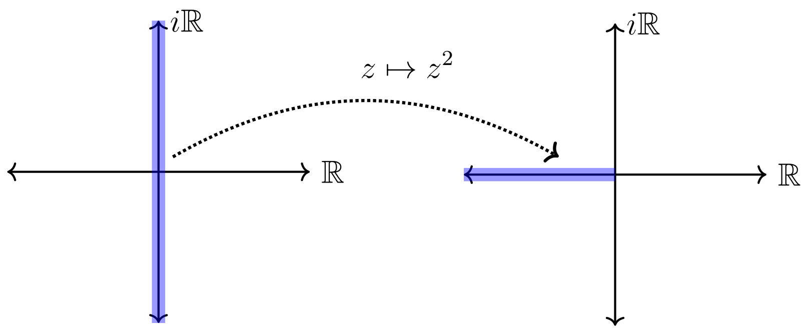
Nonetheless, we will enjoy certainly enjoy formulas when we have them. The function in Example 1.2.3 is really raising i to integer powers. When we have such a transparent formula, we might write our functions in the following way:

$$f\colon \{1, 2, 3, 4\} \to \mathbb{C}\colon k \mapsto i^k. \tag{1.2.1}$$

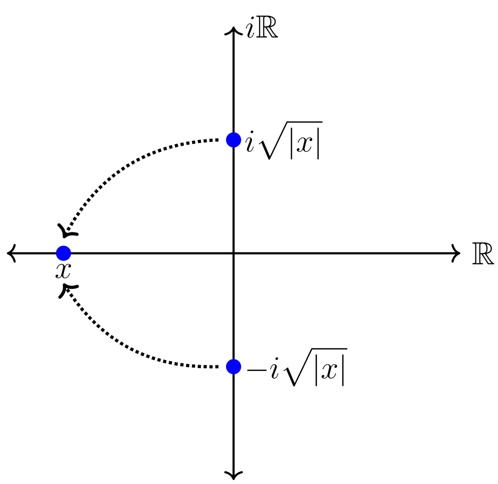
We should view the string of symbols in (1.2.1) as another way of writing the function $\{(1, i), (2, -1), (3, -i), (4, 1)\}$.

We cannot really graph functions from subsets of $\mathbb{C}$ to subsets of $\mathbb{C}$ as we would real-valued functions of a real variable; such graphs would need to exist in four dimensions! What we sometimes do is graph the domain and range separately on two pairs of two-dimensional axes.

1.2.6 Example. Define $f\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto z^2$. We claim that $f(i\mathbb{R}) = (-\infty, 0]$. Here is how we might illustrate the action of f on $i\mathbb{R}$.



Now we prove the claim. First, if $z = iy$, then $z^2 = (iy)^2 = -y^2 \le 0$, so $f(z) \in (-\infty, 0]$. Conversely, let $x < 0$, so $x = -|x| = i^2(\sqrt{|x|})^2 = f(i\sqrt{|x|})$. Note that $f(-i\sqrt{|x|}) = x$, too, and so f is not "one-to-one."



1.2.7 Problem (⋆). Let $\mathcal{D} \subseteq \mathbb{C}$ be such that if $z \in \mathcal{D}$, then $-z \in \mathcal{D}$. (We might call such a set $\mathcal{D}$ "symmetric about the origin.") A function $g\colon \mathcal{D} \to \mathbb{C}$ is **EVEN** if $g(-z) = g(z)$ for all $z \in \mathcal{D}$, while $h\colon \mathcal{D} \to \mathbb{C}$ is **ODD** if $h(-z) = -h(z)$ for all $z \in \mathcal{D}$. The **PARITY** of a function (the function's state of being even, odd, or neither) often encodes useful symmetries in mathematical problems.

Let $f\colon \mathcal{D} \to \mathbb{C}$ be a function and define

$$f_{\mathrm{e}}(z) := \frac{f(z) + f(-z)}{2} \quad \text{and} \quad f_{\mathrm{o}}(z) := \frac{f(z) - f(-z)}{2}.$$

Show that f_{e} is even, f_{o} is odd, and $f = f_{\mathrm{e}} + f_{\mathrm{o}}$. That is, any function whose domain is symmetric about the origin can be written as the sum of an even function and an odd function. Does this remind you of Problem 1.1.19?

Two of the friendliest kinds of functions are polynomials (sums of nonnegative integer powers) and rational functions (quotients of polynomials). To discuss them conveniently, we first introduce sigma notation. Let $z_0, \ldots, z_n \in \mathbb{C}$, where $n \geq 0$ is an integer, and put

$$\sum_{k=0}^{n} z_k := \begin{cases} z_0, & k = 0 \\ z_n + \displaystyle\sum_{k=0}^{n-1} z_k, & n \geq 1. \end{cases} \tag{1.2.2}$$

More generally, if for each $k \in \mathbb{Z}$ we have $z_k \in \mathbb{C}$, and if m, $n \in \mathbb{Z}$, we could define

$$\sum_{k=m}^{n} z_k := \begin{cases} 0, & m > n \\ z_m, & m = n \\ z_n + \displaystyle\sum_{k=m}^{n-1} z_k, & n \geq m + 1. \end{cases}$$

If $m > n$, then we define the "empty sum" $\sum_{k=m}^{n} z_k$ to be 0; for example,

$$\sum_{k=5}^{1} z_k = 0,$$

regardless of what these z_k are individually.

1.2.8 Example. $\displaystyle\sum_{k=1}^{3} k = \sum_{k=1}^{2} k + 3 = \sum_{k=1}^{1} k + 2 + 3 = 1 + 2 + 3 = 6.$

1.2.9 Definition. **(i)** *Let $n \geq 0$ be an integer and $a_0, \ldots, a_n \in \mathbb{C}$. A* **POLYNOMIAL** *is a* *function*

$$p\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto \sum_{k=0}^{n} a_k z^k.$$

If $n \geq 1$ and $a_n \neq 0$, then we say that p is a polynomial of degree n.

(ii) *A* **RATIONAL FUNCTION** *is a function of the form*

$$r\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto \frac{p(z)}{q(z)},$$

where p and q be polynomials and $\mathcal{D} = \{z \in \mathbb{C} \mid q(z) \neq 0\}$.

This is where we finished on Friday, January 12, 2024.

1.2.10 Example. The functions $f(z) = 1$ and $g(z) = z^2 + 1$ are polynomials, while

$$h(z) = \frac{2}{z^2 + 1}$$

is a rational function. Implicitly, the domains of f and g are $\mathbb{C}$, while the domain of h is $\mathbb{C} \setminus \{i, -i\}$.

In the previous example, the domains given for f, g, and h were the largest subsets of $\mathbb{C}$ on which those functions could be defined via the given formulas. Sometimes we may want to consider a smaller domain; this certainly arises if we start with a function defined on (a subset of) $\mathbb{R}$ and want to extend it to (a subset of) $\mathbb{C}$.

1.2.11 Example. The functions

$$f\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto t^2 \quad \text{and} \quad g\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto z^2$$

certainly "do the same thing": they take a number and square it. But f works only with real numbers, while g includes nonreal numbers.

We can make this more precise set-theoretically:

$$f = \left\{(t, t^2) \;\middle|\; t \in \mathbb{R}\right\} \quad \text{and} \quad g = \left\{(z, z^2) \;\middle|\; z \in \mathbb{C}\right\},$$

so $f \subseteq g$, but $f \neq g$, since, for example, $(i, -1) \in g$ but $(i, -1) \notin f$.

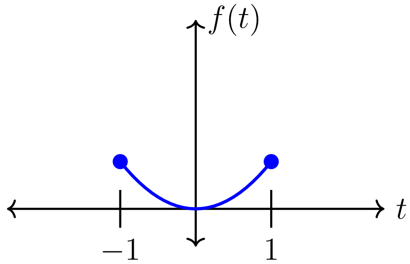
The following definition makes precise the relationship between f and g in the preceding example.

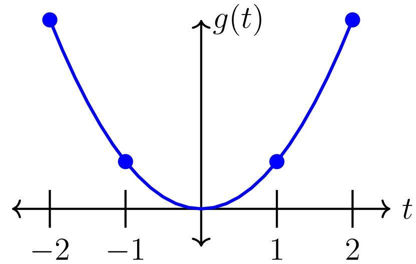
1.2.12 Definition. *Let $f\colon A \to B$ and $g\colon C \to D$ be functions. Suppose that $A \subseteq C$ and $f(z) = g(z)$ for all $z \in A$. Then f is the* **RESTRICTION OF** g **TO** A*, and we write $f = g\big|_A$. Conversely, g is an* **EXTENSION OF** f **TO** C.

To a considerable degree, much of our work in this course involves extending functions defined on (subsets of) $\mathbb{R}$ to (subsets of) $\mathbb{C}$, determining what properties the extensions inherit from the original functions of a real variable, and divining how new knowledge of the extensions enlightens us about the original functions.

1.2.13 Example. Define $f\colon [-1,1] \to \mathbb{R}\colon t \mapsto t^2$.

(i) The function $g\colon [-2,2] \to \mathbb{R}\colon t \mapsto t^2$ is an extension of f. Note that the graph of g (which we draw on a pair of real axes, unlike the drawings in Example 1.2.6) literally "extends" the graph of f from $[-1,1]$ to $[-2,2]$, or, equivalently, the graph of f is the graph of g "restricted to" the interval $[-1,1]$.

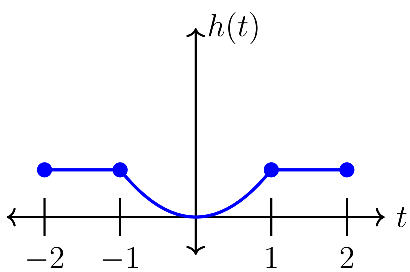




(ii) The function

$$h\colon [-2,2] \to \mathbb{R}\colon t \mapsto \begin{cases} -1, & -2 \le t < -1 \\ t^2, & -1 \le t \le 1 \\ 1, & t > 1 \end{cases}$$

is also an extension of f to $[-2,2]$. In general, extensions need not be unique.



1.2.14 Problem (!). Let $f = \{(1,i),(2,-1),(3,-i),(4,1)\}$, as in Example 1.2.3.

(i) What is $f\big|_{\{1,3\}}$?

(ii) Give an example of an extension $\widetilde{f}$ of f to $C := \{1, 2, 3, 4, 5, 6\}$ such that $\widetilde{f}(C) = f(\{1, 2, 3, 4\})$.

These are all the essential tools of functions that we will need for now. We give some further details on composition of functions in Appendix A.2.

The functions that we have encountered so far have been fairly pedestrian—polynomials and rationals—or just toy examples of sets of ordered pairs. Before we can really take up the calculus—which involves at every step the study of classes of functions united by deeper properties than their formulas—we must build a better bestiary of functions. A major task for our course will be extending "familiar" functions from $\mathbb{R}$ to (subsets of) $\mathbb{C}$. Can we, for example, assign meaning to e^z for any $z \in \mathbb{C}$? Or $\sin(z)$? And, in doing so, can we preserve the famous properties of these functions on $\mathbb{R}$? Will we have $e^{z+w} = e^z e^w$? And is there only one way to extend a function from $\mathbb{R}$ to $\mathbb{C}$? That is, are extensions unique?

1.3. Sequences.

Intuitively, a sequence should connote an "ordered list." First something, then something else, then something else, and so on. Mathematically, we want to develop sequences as *indexed* lists of numbers.

We will find sequences to be very useful tools for at least two reasons. First, sequences will help us, in a variety of diverse contexts, to turn "continuous" problems into "discrete" ones and in the process provide valuable "tests." Which is better: trying to think about the behavior of a function f at *all* numbers close to a given point z_0, or just what f does to a countable family z_k for $k \in \mathbb{N}$? Second, sequences help us define *series*, and all good functions in complex analysis are ultimately series.

1.3.1 Definition. *A* SEQUENCE *in* $\mathbb{C}$ *is a function* $f \colon \mathbb{N} \cup \{0\} \to \mathbb{C}$. *If* $z_k := f(k)$ *for* $k \ge 0$, *then we write* $(z_k) := f$. *That is,* $(z_k) = \{(k, z_k) \mid k \in \mathbb{N} \cup \{0\}\}$. *The number* z_k *is the* kTH TERM *of* (z_k).

Some sources denote what we call the sequence (z_k) by $\{z_k\}$ or $\{z_k\}_{k=0}^{\infty}$. This is perilous, as the latter notations more universally suggest *sets* of complex numbers, not *functions* (and functions are *ordered pairs* of complex numbers).

1.3.2 Example. Abbreviate $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and define $f \colon \mathbb{N}_0 \to \mathbb{C} \colon k \mapsto i^k$. Then

$$f = \left\{(k, i^k) \;\middle|\; k \in \mathbb{N}_0\right\} \qquad \text{and also} \qquad f = (i^k).$$

The range of the sequence (i^k) is the set $\left\{i^k \;\middle|\; k \in \mathbb{N}_0\right\} = \{i, -1, -i, 1\}$.

The starting index of a sequence is typically irrelevant; only the "end behavior" of a sequence usually matters. As a generalization of the definition above, we could say that a sequence is a function $f \colon [m, \infty) \cap \mathbb{Z} \to \mathbb{C}$ for some $m \in \mathbb{Z}$; in this case, if $f(k) = z_k$, then we might want to write $f = (z_k)_{k \ge m}$ to indicate where the domain starts. However, we will not usually do this and instead interpret the domain of a sequence to be the largest set of nonnegative integers on which it is defined; for example, we would, unless otherwise instructed, take the domain of (i^k/k) to be $\mathbb{N}$, not $\mathbb{N}_0$.

1.3.1. Analytic and geometric notions of sequential convergence.

The most common property of a sequence that we will study is its convergence to a given complex number. If (z_k) is a sequence and $L \in \mathbb{C}$, then we want to say that the limit of (z_k) as $k \to \infty$ equals L if we can make the terms z_k arbitrarily close to L by taking k to be sufficiently large.

First, two complex numbers are "close" if the "distance" between them is "small." We can measure distance by subtracting and taking the modulus. Thus z_k and L will be "close" if the nonnegative number $|z_k - L|$ is small, and we can safely conclude that k is "large" if $k \geq N$ for some known integer $N \in \mathbb{N}$. Let $\epsilon > 0$; this will be the threshold for measuring how small $|z_k - L|$ is. We want to force $|z_k - L| < \epsilon$ by taking $k \geq N$ for N large enough, and with N allowed to be dependent on ϵ.

We can distill these ideas into a definition.

1.3.3 Definition. *Let (z_k) be a sequence in $\mathbb{C}$ and let $L \in \mathbb{C}$. Then the* **LIMIT** *of (z_k) as $k \to \infty$ equals L if for all $\epsilon > 0$, there is an integer $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k - L| < \epsilon$. In this case, we write $\lim_{k\to\infty} z_k = L$ or $z_k \to L$. We also say that the sequence (z_k)* **CONVERGES** *to L. If there is no $L \in \mathbb{C}$ such that (z_k) converges to L, then we say that (z_k)* **DIVERGES**.

The definite article in "*the* limit of a sequence" needs justification.

1.3.4 Theorem. *Limits are unique in the sense that if (z_k) is a sequence and L_1, $L_2 \in \mathbb{C}$ both satisfy the condition in Definition 1.3.3 to be the limit of (z_k), then $L_1 = L_2$. More precisely, if (z_k) is a sequence and L_1, $L_2 \in \mathbb{C}$ satisfy*

$$\forall \epsilon > 0\ \exists N \in \mathbb{N} : k \geq N \Longrightarrow |z_k - L_1| < \epsilon \quad \text{and} \quad \forall \epsilon > 0\ \exists N \in \mathbb{N} : k \geq N \Longrightarrow |z_k - L_2| < \epsilon,$$

then $L_1 = L_2$.

Proof. We have $L_1 = L_2$ if and only if $|L_1 - L_2| = 0$, and, by Problem B.1.5, we will have $|L_1 - L_2| = 0$ if and only if $|L_1 - L_2| < \epsilon$ for all $\epsilon > 0$. This is what we now prove. Fix $\epsilon > 0$. Definition 1.3.3 allows us to relate L_1 and L_2 to ϵ by approximating L_1 and L_2 with elements of the sequence (z_k). Specifically, we may choose integers N_1, $N_2 \in \mathbb{N}$ such that if $k \geq N_1$, then $|z_k - L_1| < \epsilon/2$, and if $k \geq N_2$, then $|z_k - L_2| < \epsilon/2$. Let $N = \max\{N_1, N_2\}$.

$$|L_1 - L_2| = |L_1 - z_N + z_N - L_2| \leq |L_1 - z_N| + |z_N - L_2| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad \blacksquare$$

1.3.5 Example. For $k \geq 1$, let

$$z_k = \frac{i^k}{k}.$$

The presence of k in the denominator might suggest that $z_k \to 0$ as $k \to \infty$, and so we try to check this according to the definition. First, we calculate

$$|z_k - 0| = \left|\frac{i^k}{k} - 0\right| = \left|\frac{i^k}{k}\right| = \frac{|i^k|}{|k|} = \frac{|i|^k}{k} = \frac{1^k}{k} = \frac{1}{k}.$$

Now, given $\epsilon > 0$, it suffices to find $N \in \mathbb{N}$ such that if $k \geq N$, then $1/k < \epsilon$. But this inequality is equivalent to $1/\epsilon < k$.

The remarks above are the preparatory "scratchwork" of a convergence proof. Here is the slick, rigorous proof (which involves no algebraic hemming and hawing, and also no insight).

Given $\epsilon > 0$, take $N \in \mathbb{N}$ to satisfy $1/\epsilon < N$. Then if $k \geq N$, we have $k > 1/\epsilon$, and so

$$|z_k - 0| = \frac{1}{k} < \epsilon.$$

1.3.6 Example. The sequence (i^k) diverges. This should be intuitively obvious, as the terms of the sequence alternate around four different values and take those values infinitely many times without approaching any one exclusively, but we can give a rigorous proof of divergence using the definition. The idea is to look at the behavior of "subsequences": consider the sequence $(i^{4k}) = (1)$ and $(i^{4k+1}) = (i)$. If the whole sequence (i^k) converges, then those subsequences should converge to the same number, which would force $1 = i$.

However, to avoid developing a theory of subsequences (an otherwise worthwhile task), here is a slightly different argument (which is really a subsequence argument anyway). Suppose that (i^k) converges to some $L \in \mathbb{C}$. Let $\epsilon > 0$ and take $N \in \mathbb{N}$ such that if $k \geq N$, then $|i^k - L| < \epsilon/2$. Since $4N \geq N$ and $4N + 1 \geq N$, we have

$$|1 - L| = |i^{4N} - L| < \frac{\epsilon}{2} \quad \text{and} \quad |1 - i| = |i^{4N+1} - L| < \frac{\epsilon}{2}.$$

Then the triangle inequality gives

$$|1 - i| = |(1 - L) + (L - i)| \leq |1 - L| + |L - i| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since $\epsilon > 0$ was arbitrary, Problem B.1.5 implies $1 = i$.

This is where we finished on Wednesday, January 17, 2023.

We have defined sequential convergence *analytically*, i.e., through (in)equalities, but there is also an equivalent, and very helpful, geometric perspective on sequential convergence. Recall from Definition 1.1.11 that the open ball centered at $z \in \mathbb{C}$ of radius $\epsilon > 0$ is the set

$$\mathcal{B}(z; \epsilon) = \{w \in \mathbb{C} \mid |w - z| < \epsilon\}.$$

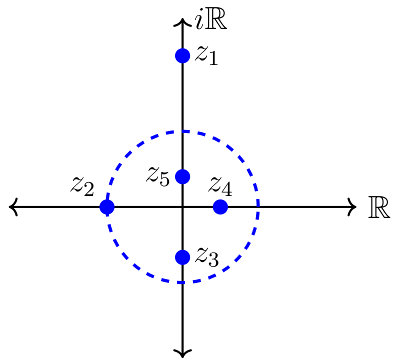
The primary utility of balls is that they give an efficient geometric mechanism for describing sequential convergence (and, later, functional limits). Because a sequence (z_k) converges to

$z \in \mathbb{C}$ if and only if for all $\epsilon > 0$, there is $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k - z| < \epsilon$, we can rephrase the definition of $z_k \to z$ using balls as follows. We can phrase the conclusion of this if-then statement with balls: $|w_k - w| < \epsilon$ if and only if $w_k \in \mathcal{B}(w;\epsilon)$. This proves the following theorem.

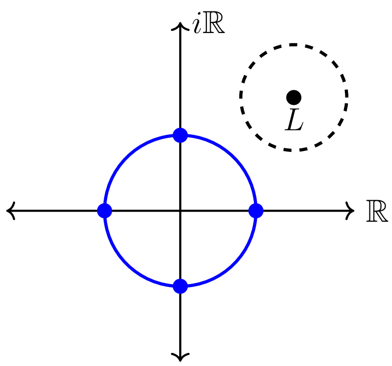
1.3.7 Theorem. *Let (z_k) be a sequence in $\mathbb{C}$ and $z \in \mathbb{C}$. Then $z_k \to z$ if and only if for all $\epsilon > 0$, there is $N \in \mathbb{N}$ such that if $k \geq N$, then $z_k \in \mathcal{B}(z;\epsilon)$.*

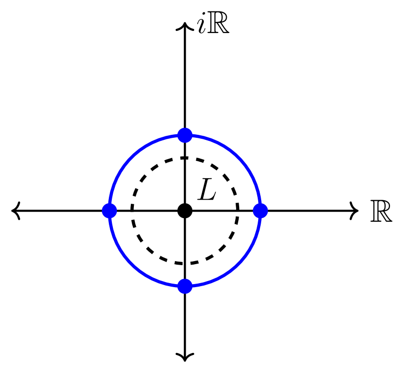
In other words, if $z_k \to z$, then no matter how small a ball around z we draw, "eventually" the terms of (z_k) must lie in this ball

1.3.8 Example. **(i)** Below we draw the ball $\mathcal{B}(0;1/2)$ and plot the first five terms of the sequence (i^k/k), which converges to 0 by Example 1.3.5. The first term i does not belong to this ball; the second term $i/2$ belongs to the circle of radius 1/2 (and therefore not to the ball); but the terms starting with $k = 3$ do belong to the ball (and they start to bunch up together quite quickly).



(ii) All terms of the sequence (i^k), which diverges by Example 1.3.6, are one of the four numbers 1, -1, i, or $-i$, and all of these numbers lie on the unit circle. Consequently, given $L \in \mathbb{C}$, we should be able to find $\epsilon > 0$ sufficiently small so that there is no $N \in \mathbb{N}$ such that if $k \geq N$, then $i^k \in \mathcal{B}(L;\epsilon)$. The "bad" ϵ might depend on L.





These pictures are illustrations of the negation of the definition of the limit. With a few more quantifiers than in Definition 1.3.3, the sequence (z_k) converges if

$$\exists L \in \mathbb{C}\ \forall \epsilon > 0\ \exists N \in \mathbb{N}\ \forall k \geq N\colon |z_k - L| < \epsilon,$$

and the negation of this statement is

$$\forall L \in \mathbb{C}\ \exists \epsilon > 0\ \forall N \in \mathbb{N}\ \exists k \geq N\colon |z_k - L| \geq \epsilon.$$

1.3.2. Properties of convergent sequences.

First, we can often ignore a finite (but possibly quite large) amount of data in a sequence. Specifically, we can change finitely many terms of a sequence without affecting its convergence.

1.3.9 Problem (!). Suppose that (z_k) and (w_k) are sequences in $\mathbb{C}$ such that $\lim_{k\to\infty} z_k = L$ for some $L \in \mathbb{C}$. If there is $N \in \mathbb{N}$ such that $z_k = w_k$ for $k \geq N$, show that $\lim_{k\to\infty} w_k = L$ as well.

Fortunately, we do not often need to use the analytic (or geoemtric) definition of sequential convergence, as all of the algebraic properties of convergence that we expect to be true are true. A few that we perhaps did not anticipate are also true.

1.3.10 Theorem (Algebra of sequences). *Let (z_k) and (w_k) be sequences in $\mathbb{C}$ such that*

$$z_k \to L_1 \quad \text{and} \quad w_k \to L_2$$

for some z, $w \in \mathbb{C}$. Then the following hold.

(i) $(z_k + w_k) \to L_1 + L_2$.

(ii) $\alpha z_k \to \alpha L_1$ *for any* $\alpha \in \mathbb{C}$.

(iii) $z_k w_k \to L_1 L_2$.

(iv) *If $L_2 \neq 0$, then*

$$\frac{z_k}{w_k} \to \frac{L_1}{L_2}.$$

(v) $\overline{z_k} \to \overline{L_1}$.

(vi) $|z_k| \to |L_1|$.

Proof. We prove only part (v), as the proofs of the other parts are virtually identical to those in real-variable calculus (in general, replace x with z). Here we are assuming $z_k \to L_1$ and we want to show $\overline{z_k} \to \overline{L_1}$. That is, for all $\epsilon > 0$, we know there is $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k - L_1| < \epsilon$. We want to show that for all $\epsilon > 0$, there is $M \in \mathbb{N}$ (here we are writing M, not N, so as not to overwork our notation) such that if $k \geq M$, then $|\overline{z_k} - \overline{L_1}| < \epsilon$. Here we need two fundamental properties of the conjugate (from Problem 1.1.22): that

$$\overline{z + w} = \overline{z} + \overline{w} \quad \text{and} \quad |\overline{z}| = z$$

for all $z, w \in \mathbb{C}$. Then

$$|\overline{z_k} - \overline{L_1}| = |\overline{z_k - L_1}| = |z_k - L_1|.$$

So, given $\epsilon > 0$, we take $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k - L_1| < \epsilon$. The calculation above therefore shows that if $k \geq N$, then

$$|\overline{z_k} - \overline{L_1}| = |z_k - L_1| < \epsilon.$$

In other words, we took $M = N$ in the sentence beginning "We want to show" above. ■

It is also possible to deduce convergence of a sequence of complex numbers just from the behavior of its real and imaginary parts. This is not a phenomenon in real single-variable calculus (although it is morally the same as deducing the behavior of a *vector* from that of its components), so we discuss part of its proof.

1.3.11 Theorem. *Let (z_k) be a sequence in $\mathbb{C}$ and $L \in \mathbb{C}$. Then $z_k \to L$ if and only if both $\text{Re}(z_k) \to \text{Re}(L)$ and $\text{Im}(z_k) \to \text{Im}(L)$.*

Proof. We prove the forward implication as an illustration of some further complex mechanics. Suppose that $z_k \to L$; we want to show that $\text{Re}(z_k) \to \text{Re}(L)$. The proof for the imaginary part is the same, so we omit it.

We know that for all $\epsilon > 0$, there is $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k - L| < \epsilon$. We want to show that for all $\epsilon > 0$, there is $M \in \mathbb{N}$ such that if $k \geq M$, then $|\text{Re}(z_k) - \text{Re}(L)| < \epsilon$. Here we need two auxiliary facts: that

$$\text{Re}(z) + \text{Re}(w) = \text{Re}(z + w) \quad \text{and} \quad |\text{Re}(z)| \leq |z| \tag{1.3.1}$$

for all $z \in \mathbb{C}$. We use these facts to compute

$$|\text{Re}(z_k) - \text{Re}(L)| = |\text{Re}(z_k - L)| \leq |z_k - L|.$$

Therefore, given $\epsilon > 0$, we take $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k - L| < \epsilon$. Then if $k \geq N$, we have

$$|\text{Re}(z_k) - \text{Re}(L)| \leq |z_k - L| < \epsilon,$$

and so $\text{Re}(z_k) \to \text{Re}(L)$. Again, we have taken $M = N$ in the sentence beginning "We want to show." ■

1.3.12 Problem (!). Prove those auxiliary facts (1.3.1).

Unfortunately, while knowledge of a sequence (z_k) translates to knowledge about the "modulated" sequence $(|z_k|)$—see part (vi) of Theorem 1.3.10, the reverse is not true.

1.3.13 Example. Let $z_k = i^k$. We saw in Example 1.3.6 that (z_k) diverges. However, $|z_k| = |i^k| = 1$ for all k, and so $|z_k| \to 1$.

Nonetheless, there is a useful situation in which knowledge of the original sequence and the "modulated" sequence lead to the same conclusion.

1.3.14 Theorem. *Let (z_k) be a sequence. Then $z_k \to 0$ if and only if $|z_k| \to 0$.*

Proof. ($\Longrightarrow$) Suppose that $z_k \to 0$ and let $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that if $k \geq N$, then $|z_k| = |z_k - 0| < \epsilon$. Consequently, $\big||z_k| - 0\big| = |z_k| < \epsilon$ for $k \geq N$, and so $|z_k| \to 0$.

($\Longleftarrow$) Suppose that $|z_k| \to 0$ and let $\epsilon > 0$. Then there is $N \in \mathbb{N}$ such that if $k \geq N$, then $\big||z_k| - 0\big| < \epsilon$. But $\big||z_k| - 0\big| = |z_k|$ as above, and so $|z_k - 0| = |z_k| < \epsilon$ for $k \geq N$. Thus $z_k \to 0$. ■

We will use results of this flavor constantly: to see that a certain quantity is small, pop a modulus on it and estimate away.

1.3.15 Example. Example 1.3.5 had us study the sequence (i^k/k), but what really mattered was the behavior of the sequence $(1/k)$. Let $z_k = i^k/k$. Then $|z_k| = 1/k$; if we have already shown that $1/k \to 0$, then Theorem 1.3.14 tells us more quickly than Example 1.3.5 that $z_k \to 0$.

1.3.16 Problem ($\star$). Suppose that (z_k) is a sequence and $L \in \mathbb{C}$ with $z_k \to L$. Justify each identity and inequality below to prove that for some integer $N \geq 0$, if $k \geq N$, then $|z_k| \geq |L|/2$. In particular, conclude that if $z_k \to L$ with $L \neq 0$, then $|z_k| > 0$ for all $k \geq N$.

$$|z_k| = |L - (L - z_k)| \geq |L| - |L - z_k| \geq |L| - \frac{|L|}{2} = \frac{|L|}{2}.$$

When working with sequences of real numbers, we have the order structure of $\mathbb{R}$ to help us, and there we enjoyed the following.

1.3.17 Theorem (Squeeze theorem for sequences in $\mathbb{R}$). *Let (a_k), (b_k), and (c_k) be sequences of real numbers with $a_k \leq b_k \leq c_k$ and $a_k \to L$ and $c_k \to L$ for some $L \in \mathbb{R}$. Then $b_k \to L$.*

A variation on Theorem 1.3.14 is the only version of the squeeze theorem that we can obtain for complex sequences.

1.3.18 Theorem (Squeeze theorem for sequences in $\mathbb{C}$). *Let (z_k) and (w_k) be sequences with the following properties.*

(i) *There is $C > 0$ such that $|z_k| \leq C|w_k|$ for all k,*

(ii) *$w_k \to 0$.*

Then $z_k \to 0$.

Proof. While we could give a direct proof using the definition of $z_k \to 0$, we use the squeeze theorem for real sequences and Theorem 1.3.14. We have $0 \leq |z_k l \leq C|w_k|$. With $a_k = 0$,

$b_k = |z_k|$, and $c_k = C|w_k|$, we have $a_k \to 0$ and $c_k \to 0$, so the real squeeze theorem implies $b_k \to 0$. Then Theorem 1.3.14 implies $z_k \to 0$. ■

1.3.19 Example. Let

$$z_k = \frac{1+i^k}{k+1}.$$

We could rewrite a_k as the sum

$$z_k = \frac{1}{k+1} + \frac{i^k}{k+1}$$

with

$$\frac{1}{k+1} \to 0 \quad \text{and} \quad \frac{i^k}{k+1} \to 0,$$

but it may be faster to use the triangle inequality and the squeeze theorem.

We estimate

$$|z_k| = \left|\frac{1+i^k}{k+1}\right| = \frac{|1+i^k|}{k+1} \le \frac{|1|+|i^k|}{k+1} = \frac{2}{k+1}.$$

This is true for all $k \ge 0$. In the notation of the squeeze theorem, we are using $C = 2$ and $w_k = 1/(k+1)$, or $C = 1$ and $w_k = 2/(k+1)$.

As we build our bestiary of complex functions and develop sequences whose terms are not so easily algebraically manipulated as in the example above, the squeeze theorem and its descendants will become even more valuable.

1.4. Series.

The tool of series formalizes the notion of "adding infinitely many numbers together." It turns out that many familiar functions are best defined by series, after a fashion, and that some of the nicest functions are inherently series.

1.4.1 Definition. *Let $(z_k)_{k\ge 0}$ be a sequence in $\mathbb{C}$.*

(i) *The* **SERIES** $\sum_{k=0}^{\infty} z_k$ *is the sequence of* n**TH PARTIAL SUMS**, *which are* $\sum_{k=0}^{n} z_k$. *That is,*

$$\sum_{k=0}^{\infty} z_k := \left(\sum_{k=0}^{n} z_k\right),$$

and so $\sum_{k=0}^{\infty} z_k$ *is the function*

$$\sum_{k=0}^{\infty} z_k : \mathbb{N}_0 \to \mathbb{C} : n \mapsto \sum_{k=0}^{n} z_k.$$

(ii) *Additionally, if the sequence of nth partial sums converges, then* $\sum_{k=0}^{\infty} z_k$ *also denotes* *that limit and is called the* **SUM** *of the series. That is, if* $\lim_{n\to\infty} \sum_{k=0}^{n} z_k$ *exists, then*

$$\sum_{k=0}^{\infty} z_k := \lim_{n\to\infty} \sum_{k=0}^{n} z_k.$$

If this limit exists, we say that the series **CONVERGES**, *and otherwise the series* **DIVERGES**. *The numbers* z_k *are the* **TERMS** *of the series* $\sum_{k=0}^{\infty} z_k$.

Thus the symbol $\sum_{k=0}^{\infty} z_k$ may have two very different meanings; it is always a sequence, and it may be the limit of that sequence, if that limit exists. Context will make clear the intended meaning of $\sum_{k=0}^{\infty} z_k$. Also, we certainly do not have to start the series at $k = 0$; if $(z_k)_{k\geq m}$ is a sequence in $\mathbb{C}$ with $m \in \mathbb{Z}$, then

$$\sum_{k=m}^{\infty} z_k = \left(\sum_{k=m}^{n} z_k\right)_{n\geq m} \quad \text{and} \quad \sum_{k=m}^{\infty} z_k = \lim_{n\to\infty} \sum_{k=m}^{n} z_k \text{ if this limit exists.}$$

Here the partial sum $\sum_{k=m}^{n} z_k$ can be defined recursively as in (1.2.2). For convenience, we will typically assume that $m = 0$, and this will not affect the proofs of any results that we state. Finally, we can change finitely many terms of a series without affecting its convergence.

1.4.2 Problem (!). As in Problem 1.3.9, rephrase this last sentence more precisely and prove it.

1.4.3 Remark. *We might say something like "Let* (z_k) *be a sequence in* $\mathbb{C}$ *and suppose* $\sum_{k=0}^{\infty} z_k$ *converges with* $S = \sum_{k=0}^{\infty} z_k$*." The first occurrence of the symbol* $\sum_{k=0}^{\infty} z_k$ *in the previous sentence represents the sequence* $\left(\sum_{k=0}^{n} z_k\right)$, *while the second represents the limit* $\lim_{n\to\infty} \sum_{k=0}^{n} z_k$*." Thus we might paraphrase the first sentence as "Let* (z_k) *be a sequence in* $\mathbb{C}$ *and suppose that the sequence* $\left(\sum_{k=0}^{n} z_k\right)$ *converges with* $S = \lim_{n\to\infty} \sum_{k=0}^{n} z_k$*."*

Series behave algebraically much the way we (should) expect. The first three results below can be deduced from the corresponding results for sequences in Theorem 1.3.10.

1.4.4 Theorem (Algebra of series). *Let* (z_k) *and* (w_k) *be sequences in* $\mathbb{C}$.

(i) *If* $\sum_{k=0}^{\infty} z_k$ *and* $\sum_{k=0}^{\infty} w_k$ *converge, then* $\sum_{k=0}^{\infty} (z_k + w_k)$ *also converges, and*

$$\sum_{k=0}^{\infty} (z_k + w_k) = \sum_{k=0}^{\infty} z_k + \sum_{k=0}^{\infty} w_k.$$

(ii) *If* $\alpha \in \mathbb{C}$ *and* $\sum_{k=0}^{\infty} z_k$ *converges, then* $\sum_{k=0}^{\infty} \alpha z_k$ *also converges, and*

$$\sum_{k=0}^{\infty} \alpha z_k = \alpha \sum_{k=0}^{\infty} z_k.$$

(iii) *If* $\sum_{k=0}^{\infty} z_k$ *converges, then*

$$\overline{\sum_{k=0}^{\infty} z_k} = \sum_{k=0}^{\infty} \overline{z_k}.$$

(iv) *The series* $\sum_{k=0}^{\infty} z_k$ *converges if and only if the series* $\sum_{k=m}^{\infty} z_k$ *converges for any* $m \geq 1$ *as well, and*

$$\sum_{k=0}^{\infty} z_k = \sum_{k=0}^{m-1} z_k + \sum_{k=m}^{\infty} z_k.$$

That is, the limit $\lim_{n\to\infty} \sum_{k=0}^{n} z_k$ *exists if and only if the limit* $\lim_{n\to\infty} \sum_{k=m}^{n} z_k$ *exists for all* $m \geq 1$*, in which case*

$$\lim_{n\to\infty} \sum_{k=0}^{n} z_k = \sum_{k=0}^{m-1} z_k + \lim_{n\to\infty} \sum_{k=m}^{n} z_k.$$

Moreover, the REINDEXING *identity*

$$\sum_{k=m}^{\infty} z_k = \sum_{j=0}^{\infty} z_{j+m}$$

holds.

1.4.5 Problem (!). In part (i) of Theorem 1.4.4, the symbol $\sum$ appears six times. In three of those instances, this symbol represents a series that should be interpreted as a sequence of partial sums. In the other three, it represents a series that should be interpreted as a limit (i.e., a complex number). Which is which?

1.4.6 Problem ($\star$). Prove part (iv) of Theorem 1.4.4.

We will now state a number of useful results about series. We will not, however, prove most of them; the proofs are excellent applications of techniques in analysis (estimates, convergence arguments) that probably would not teach us much specifically about *complex* analysis. (In general, such proofs follow by replacing x from real analysis with z.)

1.4.7 Theorem (Test for divergence). *Let* (z_k) *be a sequence in* $\mathbb{C}$.

(i) *If* $\sum_{k=0}^{\infty} z_k$ *converges, then* $\lim_{k\to\infty} z_k = 0$.

(ii) *If* $\lim_{k\to\infty} z_k \neq 0$*, or if* $\lim_{k\to\infty} z_k$ *does not exist, then* $\sum_{k=0}^{\infty} z_k$ *diverges.*

1.4.8 Example. (i) We have seen (Example 1.3.6) that the sequence (i^k) diverges, and so the series

$$\sum_{k=0}^{\infty} i^k$$

diverges by the test for divergence. Here the only meaning that we can assign to the symbol $\sum_{k=0}^{\infty} i^k$ is that it is the sequence of partial sums

$$\sum_{k=0}^{\infty} i^k = \left(\sum_{k=0}^{n} i^k\right).$$

This is where we finished on Friday, January 19, 2024.

(ii) Since $2^k \to \infty$, the series

$$\sum_{k=0}^{\infty} 2^k$$

diverges. Here we have $\sum_{k=0}^{\infty} 2^k = \left(\sum_{k=0}^{n} 2^k\right)$, although we could also reasonably say $\sum_{k=0}^{\infty} 2^k = \infty$.

1.4.9 Problem ($\star$). Prove the first part of the test for divergence by assuming that $S = \sum_{k=0}^{\infty} z_k$ for some $S \in \mathbb{C}$, writing

$$|z_k| = \left|\left(\sum_{j=0}^{k} z_j - S\right) + \left(S - \sum_{j=0}^{k-1} z_j\right)\right|,$$

and using the triangle inequality. Prove the second part by contrapositive.

Many series tests in calculus require nonnegative or positive terms. Since complex numbers are, in general, neither positive nor negative, it may seem impossible to import those tests to the complex plane. Happily, this is not so, thanks to the following concept and result.

1.4.10 Definition. *Let (z_k) be a sequence in $\mathbb{C}$. The series $\sum_{k=0}^{\infty} z_k$* **CONVERGES ABSOLUTELY** *if $\sum_{k=0}^{\infty} |z_k|$ converges.*

1.4.11 Theorem. *Let (z_k) be a sequence in $\mathbb{C}$. If $\sum_{k=0}^{\infty} z_k$ converges absolutely, then $\sum_{k=0}^{\infty} z_k$ converges, and the* **TRIANGLE INEQUALITY FOR SERIES** *holds:*

$$\left|\sum_{k=0}^{\infty} z_k\right| \le \sum_{k=0}^{\infty} |z_k|. \tag{1.4.1}$$

Our strategy going forward when given a series $\sum_{k=0}^{\infty} z_k$ will frequently be to study the "modulated" series $\sum_{k=0}^{\infty} |z_k|$. To do that, we will need more tests from calculus, and to use those tests, it will be helpful to know some series that actually converge.

One of the two most important series in the course, and possibly in all of mathematics, is the geometric series. It is one of the few series whose sum is always explicitly known and, in the final analysis, not terribly difficult to prove.

1.4.12 Theorem (Geometric series). *Let $z \in \mathbb{C}$. Then the* GEOMETRIC SERIES

$$\sum_{k=0}^{\infty} z^k$$

converges absolutely if $|z| < 1$ and diverges if $|z| > 1$. In particular, for $|z| < 1$,

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}.$$

Proof. First we show divergence for $|z| \geq 1$. If $|z| = 1$, then $|z^k| = |z|^k = 1$ as well, and so $\lim_{k\to\infty} |z^k| = 1$. But then $\lim_{k\to\infty} z^k \neq 0$ by Theorem 1.3.14, so the test for divergence implies that $\sum_{k=0}^{\infty} z^k$ diverges. Similarly, if $|z| > 1$, then $\lim_{k\to\infty} |z^k| = \lim_{k\to\infty} |z|^k = \infty$, and so $\lim_{k\to\infty} z^k \neq 0$ once again.

Now suppose $|z| < 1$. We claim that

$$\sum_{k=0}^{n} w^k = \frac{1-w^{n+1}}{1-w} \tag{1.4.2}$$

for any $n \geq 0$ and any $w \in \mathbb{C} \setminus \{1\}$.

Assuming the claim and taking $w = |z| < 1$ in (1.4.2) gives

$$\lim_{n\to\infty} \sum_{k=0}^{n} |z|^k = \lim_{n\to\infty} \frac{1-|z|^{n+1}}{1-|z|} = \frac{1}{1-|z|},$$

and so $\sum_{k=0}^{\infty} |z|^k$ converges. Therefore $\sum_{k=0}^{\infty} z^k$ converges absolutely. Moreover, taking $w = z$ in (1.4.2) implies the sum formula (1.4.2).

We conclude by proving the claim (1.4.2). If $n = 0$, then the result is immediate, as both sides of (1.4.2) equal 1 in this case. Otherwise, for $n \geq 1$, the identity (1.4.2) is equivalent to

$$(1-z)\sum_{k=0}^{n} z^k = 1 - z^{n+1}.$$

We compute on the left

$$(1-z)\sum_{k=0}^{n} z^k = \sum_{k=0}^{n} z^k - \sum_{k=0}^{n} z^{k+1} = \sum_{k=0}^{n} z^k - \sum_{k=1}^{n+1} z^k = 1 + \sum_{k=1}^{n} z^k - \sum_{k=1}^{n} z^k - z^{n+1} = 1 - z^{n+1},$$

and this proves (1.4.2). ∎

1.4.13 Problem (!). Let $m \geq 0$ be an integer and $z \in \mathbb{C}$ with $|z| < 1$. Show that

$$\sum_{k=m}^{\infty} z^k = \frac{z^m}{1-z}$$

Most of the familiar series tests from real-variable calculus require the terms in the series to be nonnegative. Since most complex numbers are neither negative nor positive, we may think that those tests will no longer be helpful. This is not the case, as we can usually invoke those tests by first passing to a "comparison" of series.

1.4.14 Theorem (Comparison test). *Let (z_k) and (w_k) be sequences such that $|z_k| \leq |w_k|$ for all k and $\sum_{k=0}^{\infty}|w_k|$ converges. Then $\sum_{k=0}^{\infty} z_k$ converges absolutely, and*

$$\sum_{k=0}^{\infty} |z_k| \leq \sum_{k=0}^{\infty} |w_k|. \tag{1.4.3}$$

By absolute convergence, the comparison test thus reduces testing the convergence of the series $\sum_{k=0}^{\infty} z_k$ to that of the "comparator" series $\sum_{k=0}^{\infty}|w_k|$, and for the latter we need other tests. We also obtain a bound on the modulus of the sum $\sum_{k=0}^{\infty} z_k$ by combining the estimates (1.4.1) and (1.4.3).

1.4.15 Example. It is not likely that we could find a formula for the sum of the series

$$\sum_{k=0}^{\infty} \frac{i^k \operatorname{Re}(i^{k+1})}{2^k},$$

if it converges. But we can determine its convergence by estimating

$$\left|\frac{i^k \operatorname{Re}(i^{k+1})}{2^k}\right| = \frac{|i|^k |\operatorname{Re}(i^{k+1})|}{2^k} \leq \frac{1}{2^k}$$

and using the comparison test in conjunction with geometric series. The inequality holds because $|i|^k = 1$ while $|\operatorname{Re}(i^{k+1})| \leq |i^{k+1}| = 1$. This inequality can be strict, such as when $k = 2$ and $\operatorname{Re}(i^3) = \operatorname{Re}(-i) = 0$.

1.4.16 Problem (!). Use (1.4.1), (1.4.3), and the formula for the sum of a geometric series to estimate

$$\left|\sum_{k=0}^{\infty} \frac{i^k \operatorname{Re}(i^{k+1})}{2^k}\right| \leq 2.$$

1.4.17 Problem ($\star$). Show that the series

$$\sum_{k=0}^{\infty} \frac{1}{k+i}$$

diverges. [Hint: *by considering real and imaginary parts, show that it suffices for the series* $\sum_{k=0}^{\infty} k(k^2+1)^{-1}$ *to diverge. Find* $C > 0$ *such that* $k^{-1} \le Ck(k^2+1)^{-1}$ *and use the divergence of the harmonic series to conclude the divergence of* $\sum_{k=0}^{\infty} k(k^2+1)^{-1}$*. Another option would be to use the integral test.*]

The preceding example is a prototype of how we often prove convergence of a series of complex numbers: first compare the given series to a series of nonnegative terms, and then use a test from real-variable calculus on that second series.

1.4.18 Theorem (Ratio test). *Let* (z_k) *be a sequence in* $\mathbb{C} \setminus \{0\}$ *and suppose that the limit*

$$L := \lim_{k \to \infty} \frac{|z_{k+1}|}{|z_k|} \tag{1.4.4}$$

exists (as a possibly extended-nonnegative number in $[0, \infty]$*). Then the series* $\sum_{k=0}^{\infty} z_k$ *converges absolutely if* $L < 1$ *and diverges if* $L > 1$*.*

The ratio test gives no information when $L = 1$, and there are both convergent and divergent series for which the limit (1.4.4) exists and equals 1. Likewise, the failure of the limit (1.4.4) to exist says nothing about the convergence or divergence of the series.

1.4.19 Problem ($\star$). **(i)** Give an example of a convergent series $\sum_{k=0}^{\infty} z_k$ and a divergent series $\sum_{k=0}^{\infty} w_k$ such that the ratio limit (1.4.4) for both series is 1; this illustrates that further analysis beyond the ratio test is sometimes necessary. [Hint: *try p-series.*]

(ii) Use the comparison test to show that the series $\sum_{k=0}^{\infty} z_k$ with

$$z_k := \begin{cases} 1/2^{k+1}, & k \text{ odd} \\ 1/2^k, & k \text{ even} \end{cases}$$

converges but the ratio limit (1.4.4) does not exist.

(iii) Give an example of a divergent series $\sum_{k=0}^{\infty} z_k$ such that the ratio limit (1.4.4) does not exist. [Hint: *such a series can be constructed by allowing* z_k *to take only two different values.*]

1.4.20 Problem ($\star$). Use the ratio test to discuss the convergence of the geometric series $\sum_{k=0}^{\infty} z^k$. Recover the results of Theorem 1.4.12 except for divergence when $|z| = 1$.

Some of the most interesting series in complex analysis depend on a complex number z

as an auxiliary parameter. The convergence of such series often hinges on the values of z, and these series are really *functions* of z. The geometric series $\sum_{k=0}^{\infty} z^k$ is one such series; here is another.

1.4.21 Example. Let $z \in \mathbb{C}$ and consider the series

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

Of course, this series should define the exponential. For $z = 0$, all of the terms for $k \geq 1$ are 0, so

$$\sum_{k=0}^{\infty} \frac{0^k}{k!} = \frac{0^0}{0!} = 0.$$

For $z \neq 0$, we establish convergence via the ratio test:

$$\frac{|z^{k+1}|}{(k+1)!}\left(\frac{k!}{|z^k|}\right) = \frac{|z|^k z k!}{(k+1)k!|z|^k} = \frac{|z|}{k+1} \to 0 \text{ as } k \to \infty.$$

This convergence is true regardless of what $z \in \mathbb{C} \setminus \{0\}$ that we use, and so the ratio test proves the (absolute) convergence of the series.

This is where we finished on Monday, January 22, 2024.

1.5. The exponential and trigonometric functions.

We showed in Example 1.4.21 that the series $\sum_{k=0}^{\infty} z^k/k!$ converges absolutely for all $z \in \mathbb{C}$, and this series, of course, defines the exponential; along with the geometric series, is the most important series that we will study. We first develop some properties of the exponential by itself, but it turns out that a trigonometric viewpoint will be even more enriching, and so we will take up trigonometric in short order, too.

1.5.1. The exponential.

The exponential is the primordial transcendental function, and all good things come from it.

1.5.1 Definition. *Let $z \in \mathbb{C}$. The* EXPONENTIAL *of z is the series*

$$\exp(z) := \sum_{k=0}^{\infty} \frac{z^k}{k!}.$$

Of course, we will eventually write $\exp(z) = e^z$ with $e := \exp(1)$, but for now we prefer to keep the notation exp to emphasize that the exponential is really a *function* on $\mathbb{C}$. The exponential has all of the properties that we expect (and a few that we probably do not);

remarkably, we can develop all of them from just a handful of fundamentals, and most of those fundamentals are very easy to prove from the series definition of the exponential. Here are those fundamentals.

1.5.2 Theorem. *Let $z, w \in \mathbb{C}$.*

(exp1) **[Functional equation]** $\exp(z+w) = \exp(z)\exp(w)$.

(exp2) $\overline{\exp(z)} = \exp(\overline{z})$.

(exp3) $\exp(0) = 1$.

(exp4) $\exp(t_1) < \exp(t_2)$ *if* $t_1, t_2 \in \mathbb{R}$ *with* $0 \le t_1 < t_2$.

(exp5) *For each* $s \in \mathbb{R}$ *with* $s > 0$*, there exists* $t \ge 0$ *such that* $\exp(t) = s$.

Properties (exp2), (exp3), and (exp4) are easy to prove from the definition of the exponential and properties of series.

1.5.3 Problem (!). Prove them.

The functional equation (exp1) has a more involved proof involving multiplication of series, which we will not give here. Likewise, property (exp5) is also more involved and requires some calculus; since we have not taken up calculus yet, and since the calculus that leads to property (exp5) would not teach us much about complex numbers, we will not prove this property, either. Instead, we can use the five properties of the exponential in Theorem 1.5.2 to obtain all other familiar and necessary features that the exponential enjoys. In doing so, as much as possible we no longer use explicitly the definition of the exponential as a series; rather, it is a function on $\mathbb{C}$ that satisfies these five properties. In fact, we will see that if $f \colon \mathbb{C} \to \mathbb{C}$ is a function that satisfies the functional equation $f(z+w) = f(z)f(w)$ and the initial condition $f(0) = 1$, then f must be the exponential!

1.5.4 Theorem. *The exponential has the following additional properties.*

(i) $\exp(z) \ne 0$ *for all* $z \in \mathbb{C}$.

(ii) $\exp(-z) = 1/\exp(z)$ *for all* $z \in \mathbb{C}$.

(iii) *If* $t \in \mathbb{R}$*, then* $\exp(t) > 0$.

(iv) *If* $t_1, t_2 \in \mathbb{R}$ *with* $t_1 < t_2$*, then* $\exp(t_1) < \exp(t_2)$.

(v) *If* $t \in \mathbb{R}$*, then* $|\exp(it)| = 1$.

Proof. **(i)** For any $z \in \mathbb{C}$, we use the functional equation and property (exp3) to compute

$$1 = \exp(0) = \exp(z - z) = \exp(z)\exp(-z). \tag{1.5.1}$$

If $\exp(z) = 0$, then $1 = 0$, so $\exp(z) \neq 0$.

(ii) The calculation (1.5.1) tells us $\exp(z)\exp(-z) = 1$, and so by definition of reciprocal we have $\exp(-z) = 1/\exp(z)$.

(iii) Property (exp4) tells us $\exp(t) > 0$ when $t > 0$, and property (exp3) extends that to $t = 0$. Now let $t < 0$, so $t = -|t|$. Then part (ii) implies

$$\exp(t) = \exp(-|t|) = \frac{1}{\exp(|t|)} > 0.$$

(iv) This is true when $0 \leq t_1 < t_2$ by property (exp4). We therefore need to consider the cases $t_1 < t_2 < 0$ and $t_1 < 0 < t_2$. In the first case, we have $0 < -t_2 < -t_1$, so property (exp4) gives $0 < \exp(-t_2) < \exp(-t_1)$. Then $1/\exp(-t_1) < 1/\exp(-t_2)$, and so we can use part (ii) to obtain $\exp(t_1) < \exp(t_2)$.

Second, if $t_1 < 0 < t_2$, we know $1 = \exp(0) < \exp(t_2)$, so we just need to show $\exp(t_1) < 1$. We have $0 < -t_1$, and so $1 = \exp(0) < \exp(-t_1)$, thus $1/\exp(-t_1) < 1/\exp(0) = 1$, and therefore, by part (ii) again, we have $\exp(t_1) < 1$, as desired.

(v) We use properties (exp2) and (exp3) and the functional equation to compute

$$|\exp(it)|^2 = \exp(it)\exp(-it) = \exp(it - it) = \exp(0) = 1.$$ ■

1.5.5 Problem (!). Show that $|\exp(z)| = \exp(\mathrm{Re}(z))$ for all $z \in \mathbb{C}$.

1.5.6 Problem ($\star$). (Requires induction, probably.) Let $k \in \mathbb{Z}$ and $z \in \mathbb{C}$. Prove that $\exp(kz) = [\exp(z)]^k$, where the latter is defined via Definition 1.1.27.

1.5.2. The sine and the cosine.

To get anywhere much further with the exponential in particular (and complex analysis in general), we need to introduce the trigonometric functions. While there are many such functions, they all arise from the exponential.

1.5.7 Definition. *Let $z \in \mathbb{C}$. The* **COSINE** *of z is*

$$\cos(z) := \frac{\exp(iz) + \exp(-iz)}{2}$$

and the **SINE** *of z is*

$$\sin(z) := \frac{\exp(iz) - \exp(-iz)}{2i}.$$

1.5.8 Remark. *Here is our rationale for introducing the sine and cosine as above. This is largely a matter of personal taste. First, we might ask what are the sine and cosine of a real number t. Any rigorous answer will boil down to either a statement about the solutions to certain second-order initial value problems (which requires a decent knowledge of differential* *equations, or a willingness to accept certain facts about differential equations—and we have come nowhere close to discussing the derivative in this course on complex analysis), or power series. Specifically, we might define*

$$\cos(t) = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!} \quad \textit{and} \quad \sin(t) = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!}.$$

If we manipulate the exponentials in Definition 1.5.7, this is exactly the power series expansions that we obtain for the sine and cosine of complex numbers.

1.5.9 Problem ($\star$). **(i)** Show that if $t \in \mathbb{R}$, then

$$\cos(t) = \mathrm{Re}[\exp(it)] \quad \text{and} \quad \sin(t) = \mathrm{Im}[\exp(it)].$$

In particular, the sine and cosine are real-valued on $\mathbb{R}$. Conclude for all $t \in \mathbb{R}$ the familiar estimate

$$|\cos(t)| \le 1 \quad \text{and} \quad |\sin(t)| \le 1.$$

(ii) Conclude **Euler's formula** for $t \in \mathbb{R}$:

$$\exp(it) = \cos(t) + i\sin(t).$$

For $z = x + iy \in \mathbb{C}$, generalize Euler's formula to

$$\exp(x + iy) = \exp(x)\big[\cos(y) + i\sin(y)\big].$$

(iii) Prove the **Pythagorean identity** for all $z \in \mathbb{C}$:

$$[\sin(z)]^2 + [\cos(z)]^2 = 1.$$

(iv) Write $\exp(z)$ in terms of $\sin(iz)$ and $\cos(iz)$. That is, if the values of $\sin(iz)$ and $\cos(iz)$ are known, how can the value of $\exp(z)$ be recovered? Casting this as the matrix-vector problem

$$\frac{1}{2}\begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} \begin{pmatrix} \exp(z) \\ \exp(-z) \end{pmatrix} = \begin{pmatrix} \cos(iz) \\ \sin(iz) \end{pmatrix}$$

may be helpful (but is not, strictly speaking, necessary).

(v) Show that the cosine is even and the sine is odd. (Recall from Problem 1.2.7 that a function $f\colon \mathbb{C} \to \mathbb{C}$ is **even** if $f(-z) = f(z)$ for all z, while a function $g\colon \mathbb{C} \to \mathbb{C}$ is **odd** if $g(-z) = -g(z)$ for all $z \in \mathbb{C}$.)

Any deeper discussion of trigonometry must mention that most marvelous number π. As with properties (exp1) and (exp5) of the exponential, we will take the following for granted and build up all the other necessary properties of π out of the subsequent few.

1.5.10 Theorem. *There exists a real number $\pi > 0$ such that the following hold.*

(π1) $\cos(\pi/2) = 0$.

(π2) $0 < \cos(t_2) < \cos(t_1) < 1$ *for* $0 < t_1 < t_2 < \pi/2$.

(π3) $\sin(\pi/2) = 1$.

(π4) *If $z \in \mathbb{C}$ with $|z| = 1$, then there is $t \in (-\pi, \pi]$ such that* $\exp(it) = z$.

1.5.11 Problem (!). Show that if $p \in \mathbb{R}$ satisfies $\cos(p/2) = 0$, then $\sin^2(p/2) = 1$. Consequently, property (π3) is really just a specification of the sign of a sine. [Hint: *what is* $|\exp(ip/2)|^2$*?*]

Property (π4) of π is just a formalization of our intuition that the coordinates $(\cos(t), \sin(t))$ parametrize the unit circle when t ranges from $-\pi$ to π. We can strengthen this property to say that this $t \in (-\pi, \pi]$ is unique; see Problem 1.5.17. This property is the foundation of polar coordinates in $\mathbb{R}^2$, and thus in $\mathbb{C}$, a highly useful representation of ordered pairs and complex numbers that we will frequently exploit (and confuse) in the future. Of course, in the past we usually took the parametrization to be over the interval $[0, 2\pi]$, but later we will see some distinct advantages to working on the interval $(-\pi, \pi]$.

The definitions of cosine and sine, their properties (π1), (π2), and (π3), and a few more properties that we will soon develop allow us to sketch their graphs, albeit crudely, as real-valued functions on the interval $[-\pi, \pi]$. These graphs are of course exactly what we know from trigonometry. The point here is that cosine and sine take the expected values at certain multiples of π and, more generally, that they are positive and negative on the usual subintervals of $[-\pi, \pi]$.

We develop this in the following. First we prove the expected periodicity of cosine and sine via the perhaps *unexpected* periodicity of the exponential.

1.5.12 Theorem. *The exponential is $2\pi i$-periodic:*

$$\exp(z + 2\pi i) = \exp(z).$$

for all $z \in \mathbb{C}$. In particular,

$$\exp\left(\frac{\pi i}{2}\right) = i, \qquad \exp(\pi i) = -1, \qquad \exp\left(\frac{3\pi i}{2}\right) = -i, \qquad \textit{and} \qquad \exp(2\pi i) = 1.$$

Proof. We build this up from repeated applications of the functional equation. First, Euler's formula tells us

$$\exp\left(\frac{\pi i}{2}\right) = \cos\left(\frac{\pi}{2}\right) + i \sin\left(\frac{\pi}{2}\right) = i.$$

The second equality just above follows from properties (π1) and (π3). Next,

$$\exp(\pi i) = \exp\left(\frac{\pi i}{2} + \frac{\pi i}{2}\right) = \exp\left(\frac{\pi i}{2}\right)\exp\left(\frac{\pi i}{2}\right) = i^2 = -1.$$

Then

$$\exp(2\pi i) = \exp(\pi i + \pi i) = \exp(\pi i)\exp(\pi i) = (-1)^2 = 1.$$

At last, we have

$$\exp(z + 2\pi i) = \exp(z)\exp(2\pi i) = \exp(z).$$

The formula for $\exp(3\pi i/2)$ follows from a similar application of the functional equation. ■

1.5.13 Problem (!). Perform that similar application of the functional equation to show $\exp(3\pi i/2) = -i$. [Hint: $\exp(2\pi i) = 1$ *and* $\exp(\pi i/2) = i$.] Also, using only the definition of the sine and cosine, and maybe results from Problem 1.5.9 and Theorem 1.5.12, show that

$$\sin(0) = \sin(\pi) = 0, \qquad \cos(0) = 1, \quad \text{and} \quad \cos(\pi) = -1.$$

1.5.14 Problem (!). Show that the sine and cosine are also $2\pi i$-periodic. Conclude that

$$\sin(k\pi) = 0 \quad \text{and} \quad \cos\left(\frac{(2k+1)\pi}{2}\right) = 0$$

for all $k \in \mathbb{Z}$. These are the familiar roots of the sine and cosine.

1.5.15 Problem (!). Show that the cosine is "odd about $\pi/2$" in the sense that

$$\cos\left(\frac{\pi}{2} + t\right) = -\cos\left(\frac{\pi}{2} - t\right)$$

for any $t \in \mathbb{R}$. [Hint: *use the definition and* $\exp(i\pi/2) = -1$.]

1.5.16 Problem ($\star$). Here is how we establish the identities

$$\sin\left(\frac{\pi}{4}\right) = \cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}.$$

(i) Use the definition of the cosine to prove the **HALF-ANGLE IDENTITY**

$$\left[\cos\left(\frac{z}{2}\right)\right]^2 = \frac{1 + \cos(z)}{2}, \ z \in \mathbb{C}.$$

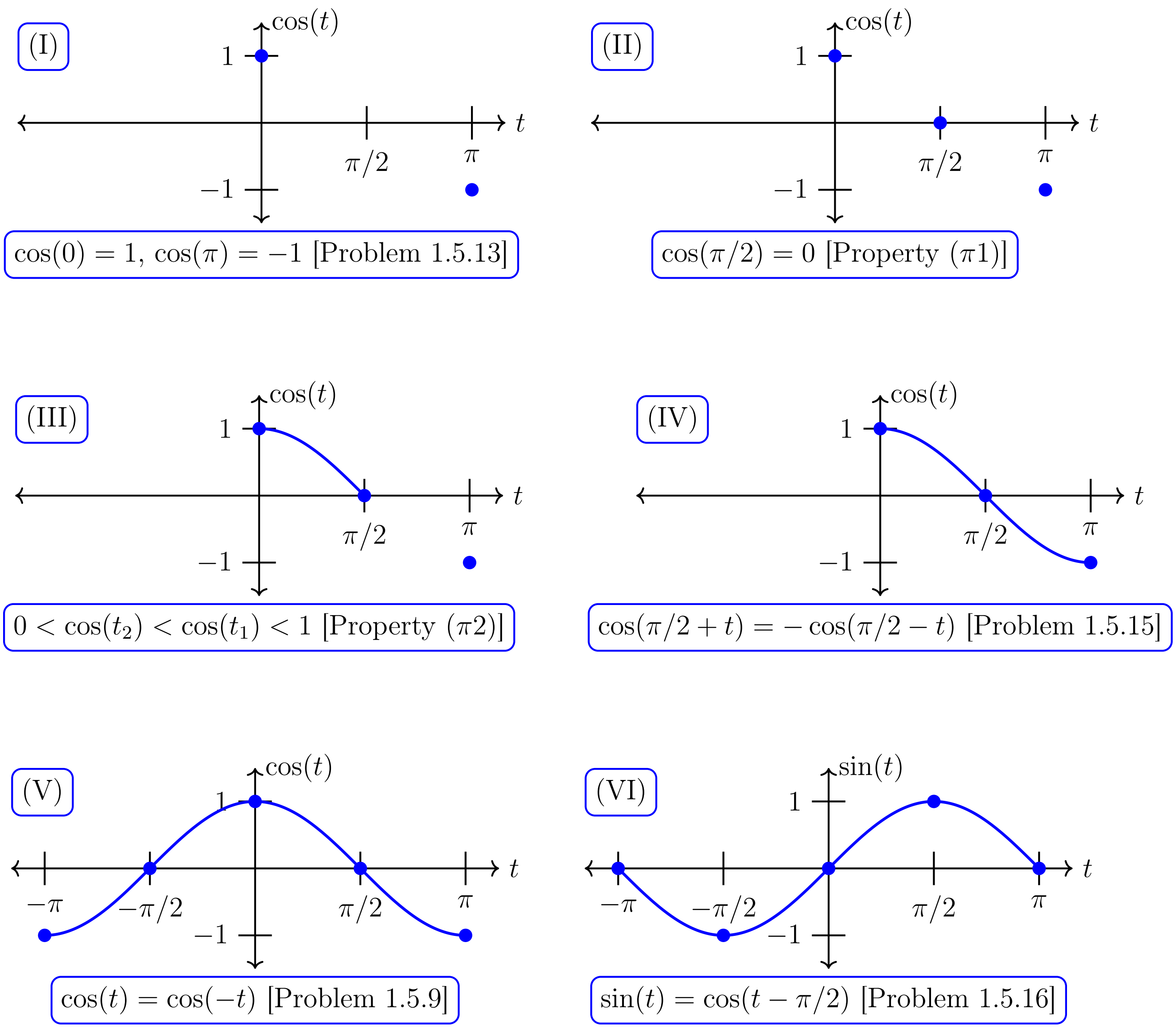
(ii) Use the half-angle identity to compute $\cos(\pi/4)$ and recall that $\cos(t) > 0$ for $0 < t < \pi/2$.

(iii) Use the definitions of sine and cosine to show that

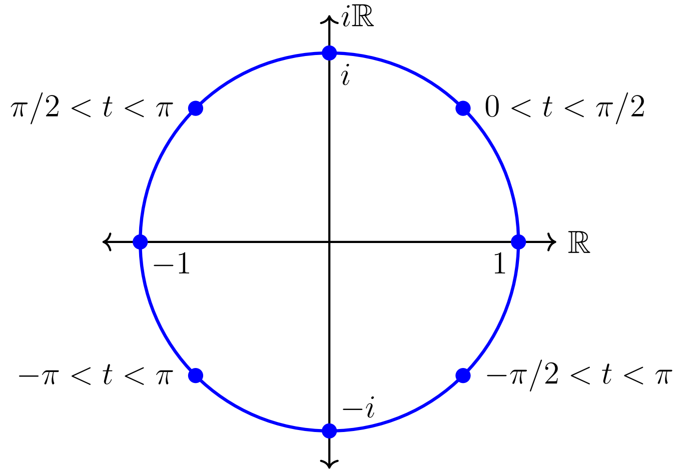
$$\sin(z) = \cos\left(z - \frac{\pi}{2}\right), \ z \in \mathbb{C}. \tag{1.5.2}$$

(iv) Compute $\sin(\pi/4)$.

Now we can graph the cosine and sine and check that the parametrization certain arcs of the unit circle corresponds to the usual subintervals of $[-\pi,\pi]$. We develop the graph of cosine first and refer to the essential prior results that help us sketch it.



These familiar graphs of cosine and sine, combined with Euler's formula $\exp(it) = \cos(t) + i\sin(t)$, show that the arcs of the unit circle in Quadrants I through IV correspond to the expected subintervals of $[-\pi,\pi]$. For example, if $z \in \mathbb{C}$ with $|z| = 1$, $\text{Re}(z) > 0$ and $\text{Im}(z) > 0$, i.e., z lies on the portion of the unit circle in Quadrant I, then Euler's formula shows that with $z = \exp(it)$, $-\pi < t \le \pi$, we need $\cos(t) > 0$ and $\sin(t) > 0$, and then the graphs above show $0 < t < \pi/2$, as expected.



1.5.17 Problem ($\star$). **(i)** Let $t_1,\ t_2 \in (-\pi, \pi]$. Show that if $\exp(it_1) = \exp(it_2)$, then $\exp(i(t_1 - t_2)) = 1$. Conclude that, to show $\exp(it_1) \neq \exp(it_2)$ for all distinct $t_1,\ t_2 \in (-\pi, \pi]$, it suffices to show $\exp(it) \neq 1$ for all $t \in (-\pi, \pi]$ with $t \neq 0$.

(ii) Explain why if $\cos(t) \neq 1$ for $t \in (-\pi, \pi]$ with $t \neq 0$, then $\exp(it) \neq 1$. [Hint: *that* $\cos(t) \neq 1$ *for* $t \in (-\pi, \pi]$ *is suggested by the graph of* $\cos$ *that we have developed above. Refresh your memory of this process by using the referenced results in graphs (III), (IV), and (V) to give a rigorous, step-by-step proof that* $\cos(t) \neq 1$ *for* $t \in (-\pi, \pi]$ *with* $t \neq 0$.]

1.5.18 Problem ($+$). A deeper result than Problem 1.5.14 is to establish that extending the sine and cosine to $\mathbb{C}$ introduces no new roots or periods beyond what we know from the real case. We will show that if $p \in \mathbb{C}$ with $\sin(z + p) = \sin(z)$ for all $z \in \mathbb{C}$, then $p = 2\pi ik$ for some $k \in \mathbb{Z}$, and likewise for the cosine. Also, we will show that $\sin(z) = 0$ if and only if $z = k\pi$ for some $k \in \mathbb{Z}$, and likewise that $\cos(z) = 0$ if and only if $z = (2k + 1)\pi/2$ for some $k \in \mathbb{Z}$. Finally, we will show that $\exp(z + p) = \exp(z)$ for all $z \in \mathbb{C}$ if and only if $p = 2\pi ik$ for some $k \in \mathbb{Z}$. We do so in the following steps.

(i) [Roots of the cosine] Suppose that $\cos(z) = 0$ for some $z \in \mathbb{C}$. Use the definition of the cosine to obtain $\exp(2iz) = -1$. Take the modulus and conclude $\exp(\operatorname{Re}(2iz)) = 1$. Since the exponential is strictly increasing on $\mathbb{R}$, conclude $\operatorname{Re}(2iz) = 0$ and therefore $\operatorname{Im}(z) = 0$. Write $z = x$ for $x \in \mathbb{R}$, so $\cos(x) = 0$; we may assume $x = \widehat{x} + 2k\pi$ for some $k \in \mathbb{Z}$ and $\widehat{x} \in (-\pi, \pi)$. Establish the identity $\cos(z + \pi/2) = -\cos(\pi/2 - z)$, valid for all $z \in \mathbb{C}$. If $\widehat{x} > 0$, explain why write $\widehat{x} = \pi/2 + \theta$ for some $\theta \in [0, \pi/2)$ and conclude $-\cos(\pi/2 - \theta) = 0$. Why does this force $\theta = 0$? If $\widehat{x} < 0$, study $-\widehat{x}$ and conclude $\widehat{x} = -\pi/2$. Thus all roots of the cosine have the form $\pm\pi/2 + 2\pi k$ for some $k \in \mathbb{Z}$; show that these are precisely the numbers of the form $(2j + 1)\pi/2$ for some $j \in \mathbb{Z}$.

(ii) [Periodicity of the cosine] Suppose that $p \in \mathbb{C}$ satisfies $\cos(z + p) = \cos(z)$ for all $z \in \mathbb{C}$. Take $z = \pi/2$ and conclude that $p = k\pi$ for some $k \in \mathbb{Z}$. To see that k must be even, show that $\cos(2k\pi) = \cos(k\pi)$ and consider what goes wrong if k is odd.

(iii) [Periodicity and roots of the sine] Use the identity $\sin(z) = \cos(z - \pi/2)$ to deduce the desired results about the periodicity and roots of the sine.

(iv) [Periodicity of the exponential] Suppose that $p \in \mathbb{C}$ satisfies $\exp(z+p) = \exp(z)$ for all $z \in \mathbb{C}$. Show that $\cos(z + p/i) = \cos(z)$ for all $z \in \mathbb{C}$ and conclude $p/i = 2\pi k$ for some $k \in \mathbb{Z}$.

(v) [Nonreality of the exponential] To confirm what we should already be expecting from the unit circle, show that $\exp(it) \notin \mathbb{R}$ for $t \in (-\pi, \pi]$ with $t \neq 0, \pi$.

1.5.19 Example. The results of Problem 1.5.18 may lull us into a false sense of security: yes, the exponential is periodic and not one-to-one on $\mathbb{C}$, but at least the periodicity and root structure of the sine and cosine do not change on the plane. However, the sine and cosine are not bounded on $\mathbb{C}$, unlike on $\mathbb{R}$. That is, there does not exist $M > 0$ such that $|\sin(z)| \leq M$ or $|\cos(z)| \leq M$ for all $z \in \mathbb{C}$.

Consider the cosine at purely imaginary values:

$$\cos(iy) = \frac{\exp(i(iy) + \exp(-i(iy)|}{2} = \frac{\exp(-y) + \exp(y)}{2}.$$

Our intuition says that $\exp(-y) \to 0$ as $y \to \infty$ and $\exp(y) \to \infty$ as $y \to \infty$, thus $|\cos(iy)| \to \infty$ as $y \to \infty$. Since we have not yet introduced limit structures, we should temper our intuition and instead, given integers $k \geq 1$, summon up $y_k \in \mathbb{R}$ such that $\exp(y_k) = k$, which part (exp5) of Theorem 1.5.2 permits. Then $\cos(iy_k) = (k + 1/k)/2$; given $M > 0$, we can choose $k > 2M$, and we will then have $|\cos(iy_k)| > M$.

1.5.20 Problem (!). Use the strategy of the preceding example to prove that the sine is unbounded on $\mathbb{C}$.

The following result will be very helpful in the future, and it is a nice opportunity to use a variety of techniques for the exponential simultaneously.

1.5.21 Theorem. $\exp(z) = 1$ *if and only if* $z = 2\pi i k$ *for some* $k \in \mathbb{Z}$.

Proof. ($\Longleftarrow$) This direction is slightly easier, so we do it first. It is really a direct calculation: since exp is $2\pi i$-periodic, we expect

$$\begin{aligned}\exp(2\pi i k) = \exp(2\pi i(k-1) + 2\pi i) = \exp(2\pi i(k-1)) = \exp(2\pi i(k-2) + 2\pi i) \\ = \exp(2\pi i(k-2)) = \cdots = \exp(2\pi i) = 1.\end{aligned}$$

As with most proofs involving "$\cdots$," we could make this more rigorous by induction. Indeed, another, quicker proof of this direction uses Problem 1.5.6 (whose proof probably needs induction anyway):

$$\exp(2\pi i k) = [\exp(2\pi i)]^k = 1^k = 1.$$

($\Longrightarrow$) There are multiple ways of proceeding; here is one. If $\exp(z) = 1$, then for all $w \in \mathbb{C}$, the functional equation tells us that

$$\exp(w+z) = \exp(w)\exp(w) = \exp(w) \cdot 1 = \exp(w),$$

and so z is a period of the exponential. Consequently (and this needs to be checked), iz is a period of the sine. Problem 1.5.18 assures us that $iz = 2\pi k$ for some $k \in \mathbb{Z}$, and so $z = -2\pi i k = 2\pi i(-k)$. ■

The product $2\pi i$ is quite special in complex analysis, and we will see its happy roles in many places in the future.

1.5.22 Problem (+). **(i)** Show that the exponential is one-to-one on strips of width 2π. That is, if $q \in \mathbb{R}$ and

$$\Sigma_q := \{z \in \mathbb{C} \mid q < \operatorname{Im}(z) \le q + 2\pi\},$$

then if $\exp(z_1) = \exp(z_2)$ for some z_1, $z_2 \in \Sigma_q$, it must be the case that $z_1 = z_2$.

(ii) With Σ_q as defined above, show that the cosine is bounded on Σ_q. That is, find $M > 0$ such that $|\cos(z)| \le M$ for all $z \in \Sigma_q$.

This is where we finished on Wednesday, January 24, 2024.

1.6. Geometry revisited: polar coordinates.

Part (π4) of Theorem 1.5.10 and part (ii) of Problem 1.5.17 tell us that for each $z \in \mathbb{C}$ with $|z| = 1$, there is a unique $t \in (-\pi, \pi]$ such that $\exp(it) = z$. Given $z \in \mathbb{C} \setminus \{0\}$, we then have

$$\left|\frac{z}{|z|}\right| = 1,$$

and we may therefore write

$$\frac{z}{|z|} = \exp(it)$$

for some unique $t \in (-\pi, \pi]$, thus

$$z = |z|\exp(it).$$

1.6.1 Definition. *Let $z \in \mathbb{C} \setminus \{0\}$.*

(i) *The* PRINCIPAL ARGUMENT *of z is the unique number $t \in (-\pi, \pi]$ such that $z = |z|\exp(it)$. We denote it by $t = \operatorname{Arg}(z)$.*

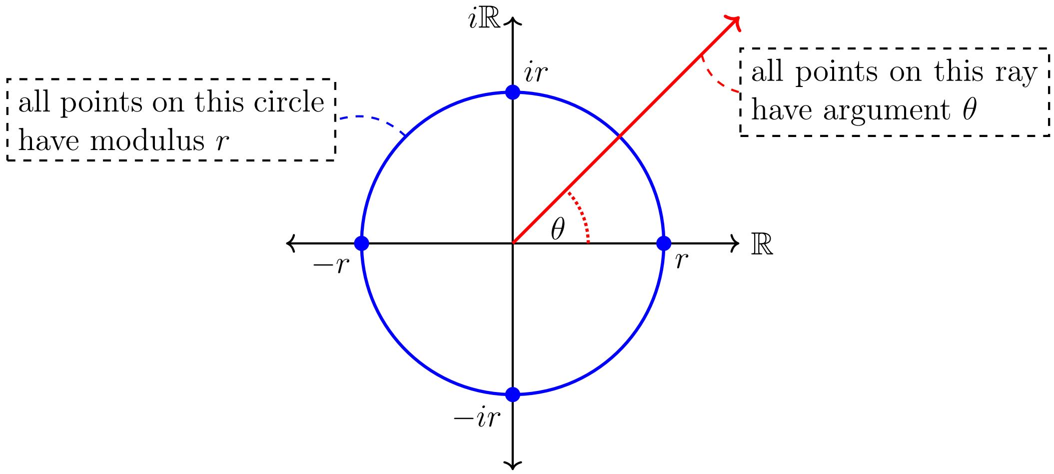
(ii) *An* ARGUMENT *of z is any $\theta \in \mathbb{R}$ such that $z = |z|\exp(i\theta)$. We denote the set of all arguments of z by $\arg(z)$. That is,*

$$\arg(z) = \{\theta \in \mathbb{R} \mid z = |z|\exp(i\theta)\}.$$

(iii) *A* **RAY** *is a set of the form* $\{z \in \mathbb{C} \mid \operatorname{Arg}(z) = \theta\}$ *for some given* $\theta \in (-\pi, \pi]$.

Whenever we have written a complex number z *in the form* $z = |z|\exp(i\theta)$ *for some* $\theta \in \mathbb{R}$, *we will refer to this representation as* **POLAR COORDINATES** *for* z.

Polar coordinates $z = |z|\exp(i\theta)$ allow us to decouple the behavior of the complex number z into the "modular" component $|z|$ and the "oscillatory" component $\exp(i\theta)$. Experience will show the utility, and frustration, of this decomposition. The following cartoon encapsulates the previous definition.



1.6.2 Example. **(i)** Let $-\pi < t \leq \pi$ and put $z = \exp(it)$. Then $|z| = 1$, so $z = 1 \cdot \exp(it) = |z|\exp(it)$. Since $t \in (-\pi, \pi]$, by definition we have $t = \operatorname{Arg}(z)$. That is, $\operatorname{Arg}(\exp(it)) = t$ when $-\pi < t \leq \pi$.

(ii) More generally, if $r > 0$ and $-\pi < t \leq \pi$, then $\operatorname{Arg}(r\exp(it)) = t$. This, too, follows from the definition of Arg, since putting $z = r\exp(it)$ gives $|z| = r$, thus $z = |z|\exp(it)$.

(iii) In particular, if $x > 0$, then $x = |x| = |x|\exp(i \cdot 0)$, and so $\operatorname{Arg}(x) = 0$. But if $x < 0$, then $x = -|x| = |x|\exp(i\pi)$, and so $\operatorname{Arg}(x) = \pi$.

1.6.3 Problem (!). **(i)** Let $z \in \mathbb{C} \setminus \{0\}$. Show that

$$\arg(z) = \{\operatorname{Arg}(z) + 2\pi k \mid k \in \mathbb{Z}\}.$$

(ii) Show that any ray $\mathcal{R}$ can be written in the form

$$\mathcal{R} = \{r\exp(i\phi) \mid r > 0\}$$

for some $\phi \in \mathbb{R}$.

(iii) Let $z \in \mathbb{C}$. Show that $\operatorname{Im}(z) \in \arg(\exp(z))$, but give an example to show that we need not always have $\operatorname{Arg}(\exp(z)) = \operatorname{Im}(z)$.

1.6.4 Problem (!). Let $z \in \mathbb{C} \setminus \{0\}$ and $\theta \in \arg(z)$. Show that

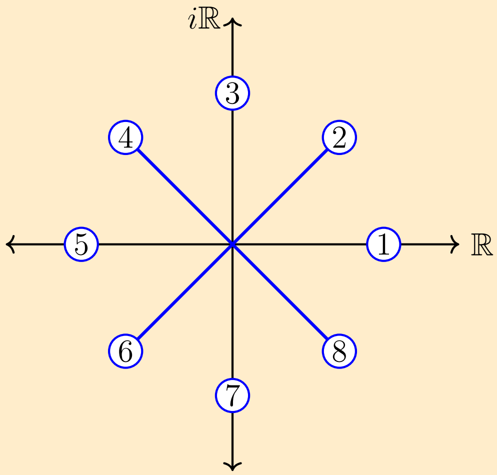
$$|z| = |\operatorname{Re}(\exp(-i\theta)z)|.$$

[Hint: *use the polar coordinates of* z *and explain why* $\exp(-i\theta)z \in \mathbb{R}$.]

Polar coordinates are the cause of and solution to most of life's problems in complex analysis. Their chief advantage is that they represent complex numbers in a geometrically transparent way that also lends itself to facile algebraic manipulations. Their chief disadvantage is ambiguity: arguments are 2π-periodic.

There are very few arguments that we can calculate explicitly, and there are even fewer that we will need.

1.6.5 Problem (!). Use familiar data from the unit circle (which you do not have to prove—but see Theorems 1.5.10 and 1.5.12 and Problems 1.5.13 and 1.5.16) to compute principal argument of each kind of point labeled in the plane below. These are the kinds of arguments that we will use most often.

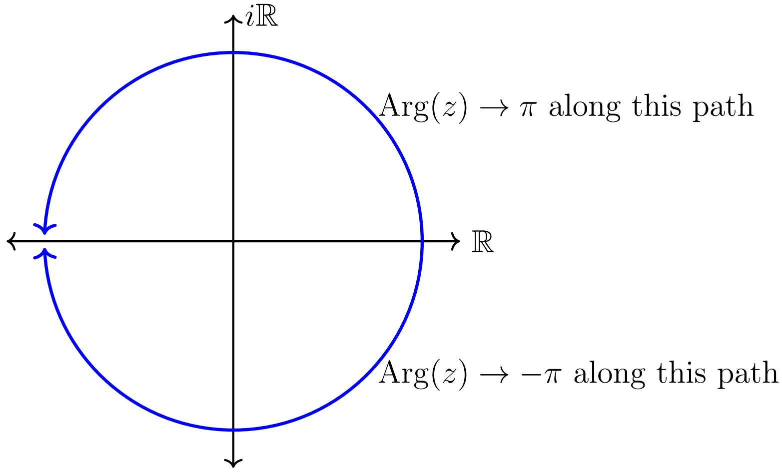


① $\operatorname{Re}(z) > 0$, $\operatorname{Im}(z) = 0$
② $\operatorname{Re}(z) = \operatorname{Im}(z) > 0$
③ $\operatorname{Re}(z) = 0$, $\operatorname{Im}(z) > 0$
④ $\operatorname{Re}(z) = -\operatorname{Im}(z)$, $\operatorname{Im}(z) > 0$
⑤ $\operatorname{Re}(z) < 0$, $\operatorname{Im}(z) = 0$
⑥ $\operatorname{Im}(z) = \operatorname{Re}(z)$, $\operatorname{Re}(z) < 0$
⑦ $\operatorname{Re}(z) = 0$, $\operatorname{Im}(z) < 0$
⑧ $\operatorname{Im}(z) = -\operatorname{Re}(z)$, $\operatorname{Re}(z) > 0$

1.6.6 Problem (!). Translate the sentence "Complex numbers are multiplied by multiplying their moduli as real numbers and adding their arguments" into precise mathematical notation. Then explain why this sentence is true.

The choice of the range $(-\pi, \pi]$ for the principal argument may seem strange, especially given that our prior experience is likely to parameterize the unit circle over $[0, 2\pi]$. This choice of range is largely a matter of convention; once fixed in an interval of the form $(\alpha, \alpha + 2\pi]$ for some $\alpha \in \mathbb{R}$, all resulting theory would flow just as well as for $\alpha = -\pi$. One (possibly superficial) advantage of the interval $(-\pi, \pi]$ over $[0, 2\pi]$ or $[0, 2\pi)$ is that $(-\pi, \pi]$ is symmetric about the origin. A deeper advantage has to do with continuity.

1.6.7 Example. While we have yet to define continuity rigorously for functions of a complex variable, our intuition with the unit circle and our calculus background should suggest to us that the following picture is true.



It looks like the values of Arg tend to both π and $-\pi$ as we approach the negative real axis, and so Arg should be discontinuous on $(-\infty, 0)$. Indeed, Arg will be continuous on $\mathbb{C} \setminus (-\infty, 0]$ and in particular continuous on $(0, \infty)$; this, in turn, will imply that a certain extension of the natural logarithm is continuous on $(0, \infty)$, just like the natural logarithm. In short, there is some payoff at the calculus level for this definition of Arg.

There will also be times when it will be worthwhile to move the (putative) discontinuity of the argument to a ray of our choosing that is not necessarily the negative real axis.

1.6.8 Lemma. *Let $\alpha \in \mathbb{R}$. Then for each $z \in \mathbb{C} \setminus \{0\}$, there is a unique $t \in (\alpha, \alpha + 2\pi]$ such that $z = |z| \exp(it)$, and we write $t = \arg_\alpha(z)$. The resulting function $\arg_\alpha \colon \mathbb{C} \setminus \{0\} \to (\alpha, \alpha + 2\pi]$ is the* αTH BRANCH OF THE ARGUMENT. *The ray $\{z \in \mathbb{C} \setminus \{0\} \mid \arg_\alpha(z) = \alpha + 2\pi\}$ is the* BRANCH CUT FOR $\arg_\alpha$.

A good proof of the lemma hopefully uses the idea that we can start with Arg(z) and then add/subtract an integer multiple of 2π to Arg(z) and eventually arrive at $\arg_\alpha(z)$.

1.6.9 Problem (+). Give a good proof of this lemma. [Hint: *to make it "good," use the fact that, for $\alpha \in \mathbb{R}$ fixed, we can write $\mathbb{R}$ as the disjoint union of intervals of the form $(\alpha + 2\pi k, \alpha + 2\pi(k + 1)]$ for $k \in \mathbb{Z}$.*]

1.6.10 Problem (!). Let $\alpha \in \mathbb{R}$. Explain why $\{z \in \mathbb{C} \setminus \{0\} \mid \arg_\alpha(z) = \alpha\} = \varnothing$.

1.6.11 Problem (!). For what $\alpha \in \mathbb{R}$ do we have $\text{Arg} = \arg_\alpha$?

The following characterization of the αth branch of the argument will be fundamental in many future calculations.

1.6.12 Theorem. *Let $\alpha \in \mathbb{R}$, $r > 0$, and $\alpha < t \le \alpha + 2\pi$. Then*

$$\arg_\alpha(r\exp(it)) = t.$$

1.6.13 Problem (!). Prove this theorem. [Hint: *reread and adapt Example 1.6.2.*]

1.6.14 Example. We study the function $\arg_{\pi/2}$.

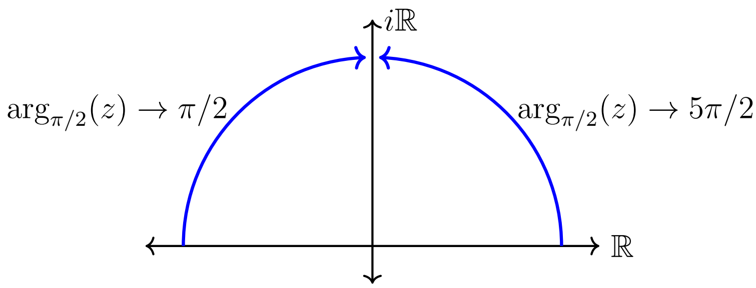
(i) We compute $\arg_{\pi/2}(1)$. With $t = \arg_{\pi/2}(1)$, we want $\exp(it) = 1$ and $\pi/2 < t \le 5\pi/2$. We know that t needs to be an integer multiple of 2π, and we cannot use $\operatorname{Arg}(1) = 0$. But $\pi/2 < 2\pi < 5\pi/2$, so $\arg_{\pi/2}(1) = 2\pi$.

(ii) We compute $\arg_{\pi/2}(-1)$. With $t = \arg_{\pi/2}(-1)$, we want $\exp(it) = -1$ and $\pi/2 < t \le 5\pi/2$. We know that $\exp(i\pi) = -1$, and we also have $\pi/2 < \pi < 5\pi/2$, so $\arg_{\pi/2}(-1) = \pi = \operatorname{Arg}(-1)$. In this case, the argument did not change with the branch.

(iii) We compute $\arg_{\pi/2}(i)$. With $t = \arg_{\pi/2}(i)$, we want $\exp(it) = i$ and $\pi/2 < t \le 5\pi/2$. We know $\exp(i\pi/2) = i$, and we know $\exp(i\pi/2 + 2\pi i) = i$. And $\pi/2 + 2\pi = 5\pi/2$. Thus $\arg_{\pi/2}(i) = 5\pi/2$.

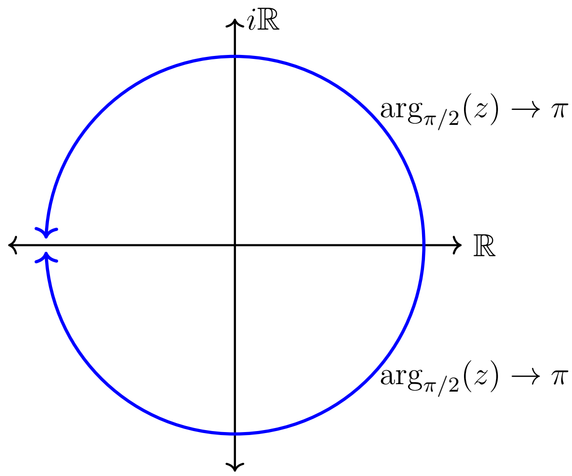
(iv) The (purportedly) bad continuity behavior of Arg on the negative real axis gets "rotated" to the positive imaginary axis for $\arg_{\pi/2}$. Fix z with $\operatorname{Re}(z) > 0$ and $\operatorname{Im}(z) > 0$ (so z is in "Quadrant I"). Then we expect $0 < \operatorname{Arg}(z) < \pi/2$, and so $2\pi < \operatorname{Arg}(z) + 2\pi < 5\pi/2$. Since $z = |z|\exp(i(\operatorname{Arg}(z) + 2\pi))$, we have $\arg_{\pi/2}(z) = \operatorname{Arg}(z) + 2\pi$. Thus $2\pi < \arg_{\pi/2}(z) < 5\pi/2$ for z with $\operatorname{Re}(z) > 0$ and $\operatorname{Im}(z) > 0$, and so we expect $\arg_{\pi/2}(z) \to 5\pi/2$ as z approaches the positive imaginary axis but remains with $\operatorname{Re}(z) > 0$ and $\operatorname{Im}(z) > 0$.

Similarly, for z with $\operatorname{Re}(z) < 0$ and $\operatorname{Im}(z) > 0$ (i.e., z is in "Quadrant II"), we have $\pi/2 < \operatorname{Arg}(z) < \pi$. This is within the range of $\arg_{\pi/2}$, so we have $\pi/2 < \arg_{\pi/2}(z) < \pi$ for z with $\operatorname{Re}(z) < 0$ and $\operatorname{Im}(z) > 0$. Thus we expect $\arg_{\pi/2}(z) \to \pi/2$ as z approaches the positive imaginary axis but remains with $\operatorname{Re}(z) < 0$ and $\operatorname{Im}(z) > 0$.



The reasoning of the preceding paragraph also suggests that $\arg_{\pi/2}(z) \to \pi$ as z approaches the negative real axis regardless of whether this approach is in Quadrant II or III. (This is not what happened with Arg in Example 1.6.7.) Indeed, if we approach the negative real axis in Quadrant II, i.e., with $\operatorname{Re}(z) < 0$ and $\operatorname{Im}(z) > 0$, then since such z satisfy

$\pi/2 < \text{Arg}(z) < \pi$, we have $\text{Arg}(z) = \arg_{\pi/2}(z)$. We therefore expect that $\arg_{\pi/2}(z) \to \pi$ as z approaches the negative real axis from within Quadrant II. And if z approaches the negative real axis from within Quadrant III, i.e., with $\text{Re}(z) < 0$ and $\text{Im}(z) < 0$, then $-\pi < \text{Arg}(z) < -\pi/2$, so $\pi < \text{Arg}(z) + 2\pi < 3\pi/2$, and so $\arg_{\pi/2}(z) = \text{Arg}(z) + 2\pi$. We therefore expect that since $\text{Arg}(z) + 2\pi \to -\pi + 2\pi = \pi$ as z approaches the negative real axis from within Quadrant III, we should have $\arg_{\pi/2}(z) \to \pi$ there, as well.



Our expectation, then, is that $\arg_{\pi/2}$ has a discontinuity along its branch cut $\left\{z \in \mathbb{C} \;\middle|\; \arg_{\pi/2}(z) = 2\pi + \pi/2\right\} = \{z \in \mathbb{C} \mid \text{Arg}(z) = \pi/2\}$, i.e., the positive imaginary axis, but, unlike Arg, it should be the case that $\arg_{\pi/2}$ is continuous on the negative real axis. This is our prior claim: that the (purportedly) bad continuity behavior of Arg on the negative real axis gets “rotated” to the positive imaginary axis for $\arg_{\pi/2}$.

1.6.15 Problem (!). Redo Example 1.6.14 for $\arg_{\pi/4}$. That is, calculate $\arg_{\pi/4}(z)$ for $z = 1$, -1, and i, and argue informally that $\arg_{\pi/4}$ should be discontinuous on its branch cut but continuous on the negative real axis.

1.6.16 Problem ($\star$). Most calculations with $\arg_\alpha$ can be accomplished by thinking about Arg and adding enough multiples of 2π to get to α. For this reason, it is worthwhile to develop some relationships between Arg and an arbitrary branch $\arg_\alpha$.

(i) For which $\alpha \in \mathbb{R}$ do we have $\text{Arg}(z) = \arg_\alpha(z)$ for all $z \in \mathbb{C} \setminus \{0\}$?

(ii) Fix $z \in \mathbb{C} \setminus \{0\}$. For which $\alpha \in \mathbb{R}$ do we have $\text{Arg}(z) = \arg_\alpha(z)$? (This is not the same as part (i)—your answer here will depend on the given z.)

(iii) Let $\alpha \in \mathbb{R}$ and $z \in \mathbb{C} \setminus \{0\}$. What relationship is there between $\arg_\alpha(z)$ and $\arg_{\alpha+2\pi}(z)$?

1.6.17 Problem ($\star$). A **SECTOR** is a set of the form

$$\{z \in \mathbb{C} \mid \alpha \leq \operatorname{Arg}(z) \leq \beta \text{ or } z = 0\}$$

for some α, $\beta \in \mathbb{R}$ satisfying $-\pi < \alpha < \beta \leq \pi$.

(i) Sketch the sector

$$\{z \in \mathbb{C} \mid \pi/4 \leq \operatorname{Arg}(z) \leq 3\pi/4 \text{ or } z = 0\}.$$

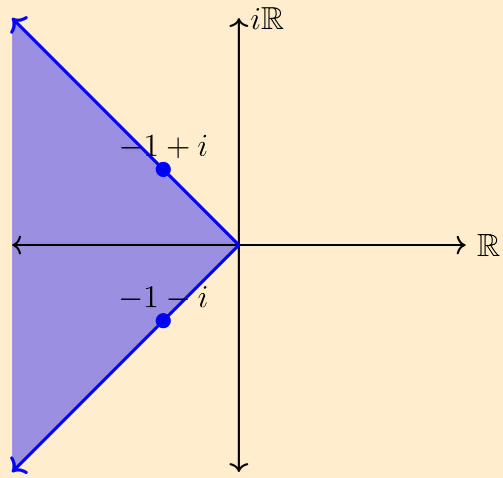
(ii) Let $0 < \omega_1 < \omega_2 < \pi$. How do the sectors

$$\{z \in \mathbb{C} \mid |\operatorname{Arg}(z)| \leq \omega_1 \text{ or } z = 0\} \quad \text{and} \quad \{z \in \mathbb{C} \mid |\operatorname{Arg}(z)| \leq \omega_2 \text{ or } z = 0\}$$

compare to each other?

(iii) Find α, $\beta \in \mathbb{R}$ such that the sector below is the set

$$\left\{z \in \mathbb{C} \;\middle|\; \alpha \leq \arg_{-\pi/2}(z) \leq \beta \text{ or } z = 0\right\}.$$



This is where we finished on Friday, January 26, 2024.

1.7. Logarithms and powers.

We now have the tools we need to invert the exponential; of course we will call its inverse the logarithm. But we will quickly see that the article "the" in "*the* logarithm" is too optimistic—we will find many logarithms! They will serve as valuable (and annoying) examples and tools in our subsequent development of the calculus. One immediate application of logarithms will be the rigorous construction of noninteger powers of complex numbers.

1.7.1. The natural logarithm.

First, we recall the original logarithm, the natural logarithm. The function $\exp\big|_{\mathbb{R}} \colon \mathbb{R} \to (0, \infty)$ is one-to-one and onto: for each $s \in (0, \infty)$, there is a unique $t \in \mathbb{R}$ such that $\exp(t) = s$. The existence of t is property (exp5) of the exponential from Theorem 1.5.2; the uniqueness follows from part (iv) of Theorem 1.5.4. Consequently, by Theorem A.2.4, there is a unique function $\ln \colon (0, \infty) \to \mathbb{R}$ such that

$$\ln(\exp\big|_{\mathbb{R}}(t)) = t \text{ for all } t \in \mathbb{R} \quad \text{and} \quad \exp\big|_{\mathbb{R}}(\ln(s)) = s \text{ for all } s \in (0, \infty). \tag{1.7.1}$$

We call this function ln the **natural logarithm**.

Unfortunately, this development does not provide us with an explicit formula for the natural log (like the power series definition of the exponential) except when evaluating very special numbers in $(0, \infty)$. We will eventually use calculus to get several very transparent formulas for the natural log, but we can get a lot just from (1.7.1).

1.7.1 Problem. Use only (1.7.1) and previously proved properties of exp to establish the following properties of ln.

(i) $\ln(1) = 0$.

(ii) $\ln(s) < 0$ for $0 < s < 1$ and $\ln(s) > 0$ for $1 < s$.

(iii) If s_1, $s_2 \in (0, \infty)$, then $\ln(s_1 s_2) = \ln(s_1) + \ln(s_2)$. [Hint: *what is* $\exp(\cdot)$ *evaluated at each side of the desired equality?*]

1.7.2. Complex logarithms.

Equipped with the natural logarithm, we can (try to) invert the exponential by solving $\exp(w) = z$ for w given $z \in \mathbb{C} \setminus \{0\}$. Experience probably teaches us that it is easiest to solve exponential equations when there are exponentials on both sides of the equation. So, we write z in polar coordinates (which we may do, since $z \neq 0$): suppose

$$z = |z| \exp(i\theta) \quad \text{with} \quad \theta = \operatorname{Arg}(z).$$

To obtain more control over w, write it as $w = x + iy$ for x, $y \in \mathbb{R}$. (Note that we are not going to write w in polar coordinates here, since there is already an exponential on the left side of $\exp(w) = z$.) Then we have

$$\exp(x)\exp(iy) = |z|\exp(i\theta) \tag{1.7.2}$$

This is only one equation, but we have two unknowns (x and y)—not a recipe for success, usually.

We can eliminate one of the variables in (1.7.2) by taking the modulus of both sides. Since $\exp(x) > 0$ and since $|\exp(iy)| = |\exp(i\theta)| = 1$, we obtain

$$\exp(x) = |z|.$$

And since $|w| > 0$, we have $x = \ln(|z|)$.

If we substitute $x = \ln(|z|)$ back into (1.7.2), we can divide $|z|$ from both sides to find

$$\exp(iy) = \exp(i\theta),$$

and therefore

$$\exp(i(y-\theta)) = 1.$$

Example 1.5.21 tells us that $i(y-\theta) = 2\pi ik$ for some $k \in \mathbb{Z}$, and so

$$y = \theta + 2\pi k.$$

We have therefore proved the following theorem.

1.7.2 Theorem. *Let $w \in \mathbb{C}$ and $z \in \mathbb{C} \setminus \{0\}$. Then $\exp(w) = z$ if and only if*

$$w = \ln(|z|) + i\,\mathrm{Arg}(z) + 2\pi ik$$

for some $k \in \mathbb{Z}$.

This result motivates the following definition.

1.7.3 Definition. **(i)** *The* LOGARITHM *of $z \in \mathbb{C} \setminus \{0\}$ is the set*

$$\log(z) := \{\ln(|z|) + i\,\mathrm{Arg}(z) + 2\pi ik \mid k \in \mathbb{Z}\} = \{\ln(|z|) + i\theta \mid \theta \in \arg(z)\}.$$

(ii) *The* PRINCIPAL LOGARITHM *is the function*

$$\mathrm{Log}\colon \mathbb{C} \setminus \{0\} \to \mathbb{C}\colon z \mapsto \ln(|z|) + i\,\mathrm{Arg}(z).$$

1.7.4 Problem (!). Show that $\exp(\mathbb{C}) = \mathbb{C} \setminus \{0\}$.

1.7.5 Problem (!). Let $t \in \mathbb{R}$ with $t > 0$. Show that $\mathrm{Log}(t) = \ln(t)$. That is,

$$\mathrm{Log}\big|_{(0,\infty)} = \ln.$$

We will later see that defining Log via Arg, and in turn requiring Arg to take values in $(-\pi, \pi]$, preserves some of the best "calculus" properties of ln in Log.

1.7.6 Example. **(i)** We do what we have probably wanted to do since high school and take the logarithm of a negative number. Specifically, we have

$$\log(-1) = \{\ln(|-1|) + i\,\mathrm{Arg}(-1) + 2\pi ik \mid k \in \mathbb{Z}\}.$$

Since $\ln(|-1|) = \ln(1) = 0$ and $\mathrm{Arg}(-1) = \pi$, we find

$$\log(-1) = \{i\pi + 2\pi ik \mid k \in \mathbb{Z}\} = \{(2k+1)\pi i \mid k \in \mathbb{Z}\}.$$

In particular,

$$\text{Log}(-1) = \ln(|-1|) + i\,\text{Arg}(-1) = 0 + i\pi = i\pi.$$

This is exactly what we expect, since

$$\exp((2k+1)\pi i) = \exp(2k\pi i + \pi i) = \exp(2k\pi i)\exp(\pi i) = \exp(\pi i) = -1.$$

(ii) Now we go purely imaginary and compute

$$\log(i) = \{\ln(|i|) + i\,\text{Arg}(i) + 2\pi ik \mid k \in \mathbb{Z}\} = \left\{\frac{i\pi}{2} + 2\pi ik \;\middle|\; k \in \mathbb{Z}\right\}$$

and

$$\text{Log}(i) = \ln(|i|) + i\,\text{Arg}(i) = \frac{i\pi}{2}.$$

This is exactly what we expect, since

$$\exp\left(\frac{i\pi}{2} + 2\pi ik\right) = \exp\left(\frac{i\pi}{2}\right)\exp(2\pi ik) = \exp\left(\frac{i\pi}{2}\right) = i.$$

1.7.7 Remark. *The object* $\log(z)$ *as we have defined it in Definition 1.7.3 is sometimes called a "set-valued" or "multi-valued" function. Of course,* $\log$ *cannot be a function from* $\mathbb{C} \setminus \{0\}$ *to* $\mathbb{C}$*. The infinite number of values that* $\log(z)$ *can take is ultimately an artifact of the* $2\pi i$*-periodicity of the exponential.*

In practice, we frequently dispense with the set-builder notation and just write something like

$$\log(z) = \ln(|z|) + i\,\text{Arg}(z) + 2\pi ik,$$

where we understand the sum on the right above really to be an element of a set indexed by $k \in \mathbb{Z}$*.*

1.7.8 Problem (⋆). Let $z \in \mathbb{C}$. Show that $\text{Log}(\exp(z)) = \exp(z)$ if and only if $-\pi < \text{Im}(z) \le \pi$. [Hint: $\exp(z) = \exp(\text{Re}(z))\exp(i\,\text{Im}(z))$.]

We have already intuited that the principal argument will suffer a discontinuity on the negative real axis, and the principal logarithm will likely inherit that behavior. It will be worthwhile to have a tool that (1) inverts the exponential (otherwise known as a logarithm) and (2) will be discontinuous on a ray of our choosing. This is possible by our modification of the principal argument into its branches, which takes care of property (2).

1.7.9 Definition. *Let* $\alpha \in \mathbb{R}$*. The* α**TH BRANCH OF THE LOGARITHM** *is the function*

$$\log_\alpha \colon \mathbb{C} \setminus \{0\} \to \mathbb{C} \colon z \mapsto \ln(|z|) + i\arg_\alpha(z).$$

The ray $\{z \in \mathbb{C} \setminus \{0\} \mid \arg_\alpha(z) = \alpha + 2\pi\}$ *is the* **BRANCH CUT FOR** $\log_\alpha$*.*

1.7.10 Example. Recall from Example 1.6.14 that $\arg_{\pi/2}(1) = 2\pi$. Thus

$$\log_{\pi/2}(1) = \ln(|1|) + i\arg_{\pi/2}(1) = 2\pi.$$

This contrasts with the familiar result $\mathrm{Log}(1) = \ln(1) = 0$ but agrees with $\exp(2\pi i) = 1$.

1.7.11 Problem (!). Let $z \in \mathbb{C} \setminus \{0\}$.

(i) Prove that $\log(z) = \{\log_\alpha(z) \mid \alpha \in \mathbb{R}\}$.

(ii) Suppose that $\mathrm{Log}(z) = \log_\alpha(z)$ for some $\alpha \in \mathbb{R}$ and $z \in \mathbb{C} \setminus \{0\}$. How are α and z related?

1.7.12 Problem (⋆). This problem examines the precise relationship between exp and the different species of log that we have developed.

(i) Show that $\exp(\log_\alpha(z)) = z$ for all $\alpha \in \mathbb{R}$ and all $z \in \mathbb{C} \setminus \{0\}$.

(ii) Let $z \in \mathbb{C}$. Describe all elements of the set $\log(\exp(z))$. More generally, describe all elements of the set $\log_\alpha(\exp(z))$.

(iii) For what $z \in \mathbb{C}$ do we have $\mathrm{Log}(\exp(z)) = z$? More generally, for what $z \in \mathbb{C}$ do we have $\log_\alpha(\exp(z)) = z$?

1.7.3. Powers.

Let z, $a \in \mathbb{C}$. What should the symbol z^a mean? Better, what should the symbol z^a *do*?

At some point in life, we probably learned that

$$\ln(x^a) = a\ln(x)$$

when a, $x > 0$. If this is true, then we can exponentiate both sides to find

$$x^a = \exp(\ln(x^a)) = \exp(a\ln(x)). \tag{1.7.3}$$

Note that we already know what exp is (and we even have an explicit formula for it as a power series), and we know what ln is at an existential level (it inverts exp). And so the identity (1.7.3) really *defines* x^a.

In turn, this motivates the following definition.

1.7.13 Definition. *Let $a \in \mathbb{C}$ and $z \in \mathbb{C} \setminus \{0\}$. Then*

$$z^a := \{\exp(aw) \mid w \in \log(z)\} = \left\{\exp\left(a[\ln(|z|) + i\,\mathrm{Arg}(z) + 2\pi i k]\right) \;\middle|\; k \in \mathbb{Z}\right\}.$$

This is where we finished on Monday, January 29, 2024.

1.7.14 Example. We have

$$\begin{aligned}1^i &= \{\exp(iw) \mid w \in \log(1)\} \\ &= \left\{\exp\left(i(\ln(|1|) + i\operatorname{Arg}(1) + 2\pi i k)\right) \,\middle|\, k \in \mathbb{Z}\right\} \\ &= \left\{\exp(2\pi i^2 k) \,\middle|\, k \in \mathbb{Z}\right\} \\ &= \{\exp(-2\pi k) \mid k \in \mathbb{Z}\} \\ &= \{\exp(2\pi k) \mid k \in \mathbb{Z}\}.\end{aligned}$$

From real-variable calculus, we expect that $1^x = 1$ for any real number x; this is no longer the case with our new interpretation of powers, but (taking $k = 0$ above), it is at least the case that $1 \in 1^i$.

1.7.15 Example. We have previously used the symbol z^a when a was an integer; see Definition 1.1.27. It would be unfortunate if Definition 1.7.13 gave a different output for z^k when $k \in \mathbb{Z}$. To check this, we take $w \in \log(z)$ and write $w = \ln(|z|) + i\operatorname{Arg}(z) + 2\pi i j$ for some $j \in \mathbb{Z}$. Then

$$\exp(kw) = \exp\left(k(\ln(|z|) + i\operatorname{Arg}(z) + 2\pi i j)\right) = \exp\left(k(\ln(|z|) + i\operatorname{Arg}(z))\right)\exp(2\pi i j k).$$

Since $jk \in \mathbb{Z}$, we have $\exp(2\pi i j k) = 1$. Now we use Problem 1.5.6 to compute

$$\exp\left(k(\ln(|z|) + i\operatorname{Arg}(z))\right) = \left[\exp(\ln(|z| + i\operatorname{Arg}(z)\right]^k = z^k.$$

Thus $\exp(kw) = z^k$, and so

$$\{\exp(kw) \mid k \in \mathbb{Z}\} = \{z^k\}.$$

Up to the fact that Definition 1.7.13 returns a set (not a complex number), we see that z^k is unambiguously defined for integers k.

Unfortunately, one of our other favorite powers is not so unambiguously defined. Until now, we have always, and intentionally, written the exponential as $\exp(z)$ and not e^z. In fact, we never defined the number e.

1.7.16 Definition. $e := \exp(1)$.

However, if we use Definition 1.7.13 to evaluate e^z, we will typically obtain an infinite set.

1.7.17 Problem (+). For which $z \in \mathbb{C}$ is e^z infinite (where e^z is interpreted according to Definition 1.7.13)? Finite? If the set e^z is finite, how many elements can it have?

For this reason, and to spare ourselves the burden of writing $\exp(z)$ all the time, we will agree that

$$e^z := \exp(z).$$

In particular, we have the functional equation

$$e^{z+w} = e^z e^w,$$

the useful property

$$e^z = 1 \iff z = 2\pi i k, \ k \in \mathbb{Z},$$

and the logarithmic identity

$$e^{\log_\alpha(z)} = z$$

for all $z \in \mathbb{C} \setminus \{0\}$.

As with the logarithm, in practice we often dispense with the set-builder notation surrounding z^a and just write

$$z^a = e^{a(\ln(|z|)+i\operatorname{Arg}(z)+2\pi i k)}, \ k \in \mathbb{Z}.$$

We can certainly fix a branch of the logarithm and decide that the symbol z^a will have the value $z^a = \exp(a\log_\alpha(z))$. However, the symbol z^a does not lend itself easily to incorporating dependence on α, and so if we want to specify a branch of the logarithm when working with powers, we will need to do so "in words" beforehand.

1.7.18 Problem (+). Let a, $b \in \mathbb{C}$ and $z \in \mathbb{C} \setminus \{0\}$. What possible meaning(s) could you give to the symbol $(z^a)^b$, and why is that meaning probably not the same as the meaning of z^{ab}?

1.7.19 Problem ($\star$). Lars Ahlfors claims, in his magisterial *Complex Analysis*, that "there is essentially only one elementary transcendental function" (p. 48). Recall that a TRANSCENDENTAL function is not ALGEBRAIC, i.e., it does not satisfy an algebraic (polynomial) equation. Based on our constructions of the trigonometric, logarithmic, and power functions, justify Ahlfors's claim and discuss the role of that "one" transcendental function in developing all the others. Then discuss the key differences that have appeared when extending the transcendental functions from (subsets of) $\mathbb{R}$ to (subsets of) $\mathbb{C}$.

1.8. Algebra: solving $z^n = w$.

Our first exposure to complex numbers was probably through the failure of real numbers to solve polynomial equations like $t^2 + 1 = 0$. As an illustration of the power of complex numbers and polar coordinates, here we solve the problem $z^n = w$, where n is a positive integer, $w \in \mathbb{C}$ is given, and z is the unknown. We will take $w \neq 0$, as otherwise the only solution is $z = 0$.

Our instinct is probably to say that if $z^n = w$, then $z = w^{1/n}$, except we know that the symbol $w^{1/n}$ should be a set of, probably, multiple elements. Instead, we might want to say that $z^n = w$ if and only if $z \in w^{1/n}$, where $w^{1/n}$ is defined as in Definition 1.7.13. Given $z \in w^{1/n}$ as defined there, we could compute directly that $z^n = w$.

1.8.1 Problem (!). Compute this directly.

In the following, we check that if $z^n = w$, then $z \in w^{1/n}$, and we give a simpler formula for the elements of $w^{1/n}$ than Definition 1.7.13.

We start by writing z and w in polar coordinates as $z = |z|e^{i\theta}$ and $w = |w|e^{i\phi}$. We view $|z|$ and θ as our two unknowns and $|w|$ and ϕ as given numbers. (This is morally similar to how we solved $\exp(z) = w$ and constructed the logarithm, except now we are using polar coordinates to represent both z and w.) Then we want

$$|z|^n e^{in\theta} = |w|e^{i\phi}. \tag{1.8.1}$$

This is one equation, and there are two unknowns; as with constructing the logarithm, we can eliminate one unknown temporarily by taking the modulus of both sides of (1.8.1):

$$|z|^n = |w|. \tag{1.8.2}$$

What is important here is that both the known quantity $|w|$ and the unknown $|z|$ are positive real numbers, and so we expect that (1.8.2) has a unique solution, i.e., that $|w|$ has a unique nth root. Previously (Remark 1.1.9) we assumed that any positive real number has a unique square root, but we did not discuss nth roots.

1.8.2 Theorem. *Let $n \ge 1$ be an integer and define*

$$\sqrt[n]{\cdot} \colon [0,\infty) \to \mathbb{R} \colon t \mapsto \begin{cases} 0, & t = 0 \\ e^{\ln(t)/n}, & t > 0. \end{cases}$$

This **PRINCIPAL nTH ROOT** *function satisfies the following.*

(i) $\left(\sqrt[n]{t}\right)^n = t$ *for all* $t \ge 0$.

(ii) *If $t \ge 0$, then the unique nonnegative solution to $s^n = t$ is $s = \sqrt[n]{t}$.*

1.8.3 Problem (!). Prove this theorem. [Hint: *use Problem 1.5.6 for the first part and the fact that $f \colon [0,\infty) \to [0,\infty) \colon s \mapsto s^n$ is strictly increasing for the second part.*]

Thus if (1.8.2) holds, then

$$|z| = \sqrt[n]{|w|}.$$

We evaluate (1.8.1) with this value for $|z|$, divide both sides by $|w|$, and conclude that θ must satisfy

$$e^{in\theta} = e^{i\phi}.$$

That is,

$$e^{i(n\theta - \phi)} = 1,$$

and so

$$n\theta - \phi = 2\pi k$$

for some $k \in \mathbb{Z}$. We rearrange and find

$$\theta = \frac{\phi + 2\pi k}{n}$$

for some $k \in \mathbb{Z}$.

We arrive, more or less, at the following theorem.

1.8.4 Theorem. *Let $w \in \mathbb{C} \setminus \{0\}$ and let $\phi \in \arg(w)$. Let $n \geq 1$ be an integer. Then $z^n = w$ if and only if*

$$z = \sqrt[n]{|w|}e^{(\phi+2\pi ik)/n}, \ 1 \leq k \leq n.$$

In particular, the equation $z^n = w$ has exactly n distinct solutions.

1.8.5 Problem ($\star$). Here is the "more or less" aspect of our arrival. The work preceding the statement of this theorem shows that if $z^n = w$, then $z = \sqrt[n]{|w|}e^{(\phi+2\pi ik)/n}$ for some $k \in \mathbb{Z}$.

(i) Check that $\left[\sqrt[n]{|w|}e^{(\phi+2\pi ik)/n}\right]^n = w$, assuming $\phi \in \arg(w)$.

(ii) Show that for any $k \in \mathbb{Z}$, there is a positive integer j satisfying $1 \leq j \leq n$ and

$$e^{(\phi+2\pi ik)/n} = e^{(\phi+2\pi ij)/n}.$$

(iii) Show that if $1 \leq j < k \leq n$, then

$$e^{(\phi+2\pi ik)/n} \neq e^{(\phi+2\pi ij)/n}.$$

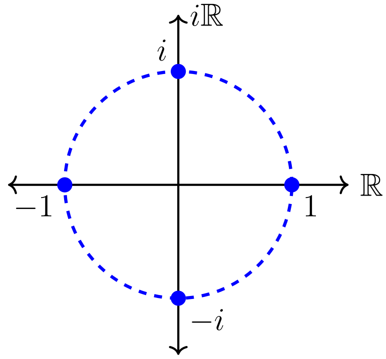
This justifies the statement in the theorem that $z^n = w$ has n *distinct* solutions.

1.8.6 Remark. *The* **FUNDAMENTAL THEOREM OF ALGEBRA** *says that if p is a polynomial of degree n with complex coefficients, i.e., $p(z) = \sum_{k=0}^{n} a_k z^k$ with $a_k \in \mathbb{C}$ and $a_n \neq 0$, then p has n roots in $\mathbb{C}$, "counting multiplicities." We will prove a version of this theorem later (and define rather precisely "multiplicities"), but for now we can interpret Theorem 1.8.4 as a fundamental theorem of algebra for the special polynomial $p(z) = z^n - w$ with $w \in \mathbb{C}$ fixed. In particular, we get n distinct roots, not just n roots "counting multiplicities."*

1.8.7 Example. We solve $z^4 = 1$. We expect that $z = 1$ is a solution (and it is, of course), but there should be three others. Since $\operatorname{Arg}(1) = 0$, we know that the four solutions to $z^4 = 1$ are

$$\begin{aligned}
z_1 &:= e^{2\pi i/4} = e^{\pi i/2} = i \\
z_2 &:= e^{4\pi i/4} = e^{\pi i} = -1 \\
z_3 &:= e^{6\pi i/4} = e^{3\pi i/2} = -i \\
z_4 &:= e^{8\pi i/4} = e^{2\pi i} = 1.
\end{aligned}$$

It is instructive to see how these four solutions fall out on the unit circle.



The four solutions to $z^4 = 1$ are all spaced $\pi/2$ radians apart on the unit circle; in particular, we can plot them by first plotting 1 and then marking points on the unit circle in increments of $\pi/2$ radians from 1.

1.8.8 Definition. *A complex number z such that $z^n = 1$ for some integer $n \geq 1$ is an* n**TH ROOT OF UNITY**.

1.8.9 Problem (!). Generalize the observations in Example 1.8.7 about the positioning of the solutions of $z^4 = 1$ on the unit circle to the positioning of the solutions of $z^n = 1$ on the unit circle.

1.8.10 Example. We expect that the only two complex numbers to satisfy $z^2 = -1$ are $z = \pm i$. Theorem 1.8.4 validates this rigorously, as it tells us that the only solutions to $z^2 = -1$ are

$$z_1 := e^{(\pi+2\pi i)/2} = e^{3\pi i2} = -i \quad \text{and} \quad z_2 := e^{(\pi+4\pi i)/2} = e^{\pi i/2+2\pi i} = i.$$

2. DIFFERENTIAL CALCULUS

2.1. Functions (briefly revisited).

We now have a rich bestiary of functions to manipulate and study. So far, as is typical in precalculus, we have considered classes of functions largely separately from each other—yes, the exponential is the source of most interesting functions, but we have been considering the properties of exponentials, trig functions, logs, and powers in turn and not necessarily seeing what they have in common (beyond, of course, the exponential). This changes with calculus, which considers the deeper properties that functions share beyond their cosmetic (formulaic) differences.

Recall that the notation $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ means that f is a complex-valued function with domain $\mathcal{D}$; the range of f is $f(\mathcal{D}) = \{f(z) \mid z \in \mathcal{D}\}$. The notation allows $\mathcal{D} \subseteq \mathbb{R}$ and $f(\mathcal{D}) \subseteq \mathbb{R}$, too, and we will see that such real restrictions on the domain and/or range lead to distinct conclusions about the properties of f. Unsurprisingly, much of calculus hinges on algebraic operations on functions.

2.1.1 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ and let $f, g\colon \mathcal{D} \to \mathbb{C}$ be functions. We define*

$$f + g\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto f(z) + g(z) \quad \text{and} \quad fg\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto f(z)g(z).$$

Additionally, we put

$$|f|\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto |f(z)| \quad \text{and} \quad \overline{f}\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto \overline{f(z)}.$$

2.1.2 Remark. *The symbol $+$ in the preceding definition has two meanings. First, there is the addition of the complex numbers $f(z)$ and $g(z)$, denoted by $f(z) + g(z)$. Second, there is the new function from $\mathcal{D}$ to $\mathbb{C}$ whose range consists of these sums $f(z) + g(z)$; we call this function $f + g$. We should remember that $f(z) + g(z)$ is, given z, a single complex number, while $f + g$ is a function, i.e., a set of ordered pairs of complex numbers. Likewise, the juxtaposition $f(z)g(z)$ is the product of the two complex numbers $f(z)$ and $g(z)$, while fg is the function whose range consists of these products $f(z)g(z)$.*

Any complex number $z \in \mathbb{C}$ is determined by its real and imaginary parts $\text{Re}(z), \text{Im}(z) \in \mathbb{R}$, and knowledge of the real and imaginary parts separately usually amounts to full knowledge of z via the identity $z = \text{Re}(z) + i\,\text{Im}(z)$; recall, for example, Theorem 1.3.11 on divining the convergence of a sequence via the convergence of its real and imaginary parts. We can, of course, consider the real and imaginary parts of a function f; for $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$, define

$$\text{Re}[f]\colon \mathcal{D} \to \mathbb{R}\colon z \mapsto \text{Re}[f(z)] \quad \text{and} \quad \text{Im}[f]\colon \mathcal{D} \to \mathbb{R}\colon z \mapsto \text{Im}[f(z)]. \tag{2.1.1}$$

Then since

$$f(z) = \text{Re}[f(z)] + i\,\text{Im}[f(z)], \tag{2.1.2}$$

we have

$$f = \text{Re}[f] + i\,\text{Im}[f]. \tag{2.1.3}$$

In the spirit of Remark 2.1.2, the identity (2.1.2) is an equality of complex numbers, whereas (2.1.3) is an equality of functions; given $z \in \mathbb{C}$, $\mathrm{Re}[f(z)]$ is a single complex number but $\mathrm{Re}[f]$ is a function from a subset of $\mathbb{C}$ to $\mathbb{R}$.

It is frequently helpful to see how the real and imaginary parts of f depend explicitly on the real and imaginary parts of the independent variable of f. If we put $z = x + iy$ with x, $y \in \mathbb{R}$, then we can set

$$u(x,y) := \mathrm{Re}[f(x+iy)] = \mathrm{Re}[f](x+iy) \quad \text{and} \quad v(x,y) := \mathrm{Im}[f(x+iy)] = \mathrm{Im}[f](x+iy)$$

to find

$$f(x+iy) = u(x,y) + iv(x,y).$$

Here, if the domain of f is the set $\mathcal{D}$ of complex numbers, then u and v are functions of the *ordered pair* of real variables (x,y) in the set $\widetilde{\mathcal{D}} := \big\{(x,y) \in \mathbb{R}^2 \;\big|\; x + iy \in \mathcal{D}\big\}$. (Recall, strictly speaking, in this course that a complex number is *not* an ordered pair of real numbers—see Appendix B.2.)

2.1.3 Example. Define $f \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto z^2$. Then

$$f(x+iy) = (x+iy)^2 = x^2 + 2iy + i^2y^2 = (x^2 - y^2) + i(2xy).$$

So, if we set

$$u(x,y) := x^2 - y^2 \quad \text{and} \quad v(x,y) := 2xy,$$

then we have

$$u(x,y) = \mathrm{Re}[f(x+iy)], \qquad v(x,y) = \mathrm{Im}[f(x+iy)], \qquad \text{and} \qquad f(x+iy) = u(x,y)+iv(x,y).$$

Going forward, one of our major questions will be how the "calculus properties" of the real and imaginary parts of f (usually in the sense of familiar multivariable calculus on $\mathbb{R}^2$) affect the "calculus properties" of f itself as a function of a complex variable.

This is where we finished on Wednesday, January 31, 2024.

2.2. Limits.

Limits describe how the of outputs functions behave as their inputs approach certain values. Let $\mathcal{D} \subseteq \mathbb{C}$, $f \colon \mathcal{D} \to \mathbb{C}$, and a, $L \in \mathbb{C}$. We want to say that *the limit of f as z approaches a equals L*, written $\lim_{z \to a} f(z) = L$, if we can make $f(z)$ and L arbitrarily close by taking $z \in \mathcal{D}$ and a to be sufficiently close. The symbol $\lim_{z \to a} f(z) = L$ is an abbreviation for the previous sentence in italics.

2.2.1 Example. (i) Define

$$f \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto \begin{cases} 1, & z \neq 0 \\ i, & z = 0. \end{cases}$$

Since f stays at 1 for all but a single point in its domain, we probably want to say

$$\lim_{z \to 0} f(z) = 1,$$

even though $f(0) \neq 1$.

(ii) Define

$$g \colon \mathbb{C} \setminus \{0\} \to \mathbb{C} \colon z \mapsto 1.$$

Again, we probably want to say

$$\lim_{z \to 0} g(z) = 1,$$

even though g is not defined at 0. (Note, by the way, that $g = f\big|_{\mathbb{C} \setminus \{0\}}$.)

2.2.1. *The (correct) definition of limit.*

Let $f \colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ and let a, $L \in \mathbb{C}$. The statement $\lim_{z \to a} f(z) = L$ needs to capture the idea that we can make $f(z)$ as close to L as we want by taking z sufficiently close to a. Example 2.2.1 reminds us that we do not want to require that a belong to the domain of f, nor that $f(a) = L$ even if a is in the domain of f. There are several ways of defining limits rigorously; our approach here is to exploit our prior hard work with sequences so we can "get to the good stuff" of calculus quickly. Earlier we said that sequences have two chief virtues in calculus: they help us define series (which in turn give many interesting functions), and they help us "test" or "measure" concepts that are inherently *continuous* in a conveniently *discrete* way. A limit is a *continuous* concept, as it involves the behavior of a function at *all* values approaching a certain point. A sequence, however, is *discrete*, as it takes only *countably many* values.

With this in mind, we define limits of functions via limits of sequences. We want the values of $f(z)$ to become close to L when z is close to a. One way to test "close" is with convergent sequences. Suppose that (z_k) is a sequence in $\mathcal{D}$ with $z_k \to a$. Then the values of (z_k) are certainly becoming very close to a! If the values of $f(z)$ are becoming very close to L when z is close to a, we should hope, then, that $f(z_k) \to L$.

We want a certain arbitrariness with the inputs to f in the definition of the limit: no matter what $z \in \mathcal{D}$ we choose, as long as z is close to a, we will have $f(z)$ close to L. So, we expand our test of "closeness" from one sequence (z_k) in $\mathcal{D}$ with $z_k \to a$ to *all* sequences (z_k) in $\mathcal{D}$ with $z_k \to a$. Our first stab at a definition of limit is then

$$\lim_{z \to a} f(z) = L \iff \big[(z_k) \text{ is a sequence in } \mathcal{D} \text{ and } z_k \to a \Longrightarrow f(z_k) \to L\big]. \tag{2.2.1}$$

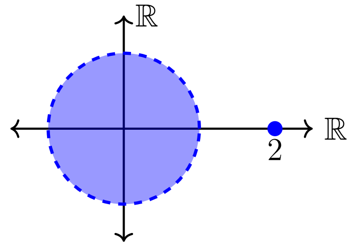
There are two problems with this definition. One is easily fixed. Recall that we do not want to say anything about whether or not $a \in \mathcal{D}$, nor about the value $f(a)$ if, indeed, $a \in \mathcal{D}$. So, when testing "closeness" via sequences, we do not want to risk a being among the terms of the sequence and confusing our measurements. We therefore amend (2.2.1) to

$$\lim_{z \to a} f(z) = L \iff \big[(z_k) \text{ is a sequence in } \mathcal{D} \setminus \{a\} \text{ and } z_k \to a \Longrightarrow f(z_k) \to L\big]. \tag{2.2.2}$$

The remaining problem with (2.2.2) is subtler. The right side of the $\iff$ presumes that there is a sequence (z_k) in $\mathcal{D} \setminus \{a\}$ such that $z_k \to a$. If there is no such sequence, then the if-then statement on the right has a false hypothesis and therefore is vacuously true. It would therefore be the case that $\lim_{z \to a} f(z) = L$ for any $L \in \mathbb{C}$, and surely this violates the intuitive notion that limits are unique. This situation with a can easily occur.

2.2.2 Example. Let

$$\mathcal{D} = \{z \in \mathbb{C} \mid |z| < 1 \text{ or } z = 2\} \quad \text{and} \quad f \colon \mathcal{D} \to \mathbb{C} \colon z \mapsto \begin{cases} z, & |z| < 1 \\ 2i, & z = 2. \end{cases} \tag{2.2.3}$$



The sketch above should make it clear that there is no sequence (z_k) in $\mathcal{D} \setminus \{2\}$ such that $z_k \to 2$. Indeed, such a sequence would need to satisfy $|z_k| < 1$, and then we would have

$$2 = \lim_{k \to \infty} |z_k| < 1,$$

which is impossible. Trying to compute $\lim_{z \to 2} f(z)$ is therefore pointless: there is no sensible way to measure the behavior of f as z becomes "close" to (but not equal to) 2.

For this reason, we only want to consider limits at points a that can be "reached" by sequences in $\mathcal{D}$ not consisting of a. Here is the behavior of a that we need.

2.2.3 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$. A point $a \in \mathbb{C}$ is a* **ACCUMULATION POINT** *or* **LIMIT POINT** *of $\mathcal{D}$ if there is a sequence of distinct points (z_k) in $\mathcal{D} \setminus \{a\}$ such that $z_k \to a$.*

The key restriction in the definition of accumulation point is that the terms of the sequence (z_k) cannot be a. This ensures that other elements of $\mathcal{D}$ "approach" or "cluster around" a sufficiently. An accumulation point of $\mathcal{D}$ need not be an element of $\mathcal{D}$ but will belong to the "boundary" of $\mathcal{D}$ (in a way that we could make topologically precise but will not).

2.2.4 Example. Let $\mathcal{D}$ be defined as in (2.2.3).

(i) The point 0 is an accumulation point of $\mathcal{D}$. Take $z_k := 1/(k+1)$ for $k \geq 1$, so $z_k \neq 0$, $|z_k| < 1$ (and thus $z_k \in \mathcal{D} \setminus \{1\}$), and $z_k \to 0$. Thus $z_k \in \mathcal{D} \setminus \{1\}$.

(ii) The point 2 is not an accumulation point of $\mathcal{D}$. We prove this by contradiction: if (z_k) is a sequence in $\mathcal{D} \setminus \{2\}$ with $z_k \to 2$, then by definition of $\mathcal{D}$, it must be the case that $|z_k| < 1$. And so

$$1 = \lim_{k \to \infty} |z_k| < 2,$$

which is impossible. This is exemplifies the slogan "Membership in $\mathcal{D}$ says nothing about being an accumulation point of $\mathcal{D}$."

2.2.5 Problem (!). Show that 0 is an accumulation point of $\mathcal{D} := \{z \in \mathbb{C} \mid \operatorname{Re}(z) > 0\}$ but -1 is not.

2.2.6 Problem ($\star$). Let $I \subseteq \mathbb{R}$ be an interval. Show that every $a \in I$ is an accumulation point of I. [Hint: *since I is an interval, if t_1, $t_2 \in I$ with $t_1 < t_2$, and if $t_1 < t < t_2$, then $t \in I$.*]

Embiggened with the concept of accumulation point, we can finally define limits.

2.2.7 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ and $f\colon \mathcal{D} \to \mathbb{C}$. Let $L \in \mathbb{C}$ and let $a \in \mathbb{C}$ be an accumulation point of $\mathcal{D}$. Then the limit of f as z approaches a equals L, written $\lim_{z\to a} f(z) = L$, if for any sequence (z_k) in $\mathcal{D} \setminus \{a\}$ with $z_k \to a$, we also have $f(z_k) \to L$.*

Before proceeding, we should check that limits as defined above really are unique, so that we can speak of "the" limit.

2.2.8 Theorem. *Let $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ with $a \in \mathbb{C}$ an accumulation point of $\mathcal{D}$. Let L_1, $L_2 \in \mathbb{C}$ with $\lim_{z\to a} f(z) = L_1$ and $\lim_{z\to a} f(z) = L_2$. Then $L_1 = L_2$.*

Proof. Since a is an accumulation point of $\mathcal{D}$, there is a sequence (z_k) in $\mathcal{D} \setminus \{a\}$ such that $z_k \to a$. Since $\lim_{z\to a} f(z) = L_1$, we have $f(z_k) \to L_1$, and since $\lim_{z\to a} f(z) = L_2$, we have $f(z_k) \to L_2$. That is, the sequence $(f(z_k))$ converges to both L_1 and L_2, so, by uniqueness of limits of sequences, we have $L_1 = L_2$. ■

2.2.9 Problem (!). Use only Definition 2.2.7 to prove that $\lim_{z\to a} f(z) = L$ if and only if $\lim_{z\to a}(f(z) - L) = 0$.

2.2.2. *Algebraic properties of limits.*

Our work with sequences, specifically Theorem 1.3.10, helps us prove all of algebraic properties of limits that we expect from calculus.

2.2.10 Theorem (Algebra of limits). *Let $\mathcal{D} \subseteq \mathbb{C}$, let $a \in \mathbb{C}$ be an accumulation point of $\mathcal{D}$, and let f, $g\colon \mathcal{D} \to \mathbb{C}$ with*

$$\lim_{z\to a} f(z) = L_1 \quad \text{and} \quad \lim_{z\to a} g(z) = L_2$$

for some L_1, $L_2 \in \mathbb{C}$. Then the following hold.

(i) $\lim_{z\to a} \big(f(z) + g(z)\big) = L_1 + L_2$.

(ii) $\lim_{z\to a} \alpha f(z) = \alpha L_1$ *for any* $\alpha \in \mathbb{C}$.

(iii) $\lim_{z\to a} f(z)g(z) = L_1 L_2$

(iv) *If* $L_2 \neq 0$, *then*

$$\lim_{z\to a} \frac{f(z)}{g(z)} = \frac{L_1}{L_2}.$$

(v) $\lim_{z\to a} \overline{f(z)} = \overline{L_1}$.

(vi) $\lim_{z\to a} |f(z)| = |L_1|$

Frequently showing a zero limit is easier than showing any other limit, and so we are fortunate to have the equivalence

$$\lim_{z\to a} f(z) = L \iff \lim_{z\to a} \big(f(z) - L\big) = 0. \tag{2.2.4}$$

To prove this, just take $g(z) = L$ and use the algebra of limits.

2.2.11 Example. We can iterate the algebraic rules for limits (or, more precisely, induct) to show that polynomials and rational functions are well-behaved under limits:

$$\lim_{z\to a} \sum_{k=0}^{n} c_k z^k = \sum_{k=0}^{n} c_k a^k$$

for all $a \in \mathbb{C}$, integers $n \geq 0$, and $c_0, \ldots, c_n \in \mathbb{C}$. Consequently, if p and q are polynomials and $q(a) \neq 0$, we also have

$$\lim_{z\to a} \frac{p(z)}{q(z)} = \frac{p(a)}{q(a)}.$$

As in Theorem 1.3.11, a function's limiting behavior is equivalent to the simultaneous limits of its real and imaginary parts.

2.2.12 Theorem. *Let* $\mathcal{D} \subseteq \mathbb{C}$, *let* $a \in \mathbb{C}$ *be an accumulation point of* $\mathcal{D}$, *let* $L \in \mathbb{C}$, *and let* $f\colon \mathcal{D} \to \mathbb{C}$. *Then* $\lim_{z\to a} f(z) = L$ *if and only if both* $\lim_{z\to a} \mathrm{Re}[f(z)] = \mathrm{Re}(L)$ *and* $\lim_{z\to a} \mathrm{Im}[f(z)] = \mathrm{Im}(L)$.

And as in Theorem 1.3.14, there is a close relationship between the zero limit of a function and the zero limit of its modulus.

2.2.13 Problem (!). Show that $\lim_{z\to a} f(z) = 0$ if and only if $\lim_{z\to a} |f(z)| = 0$. [Hint: *use the definition of limit and Theorem 1.3.14.*] If $\lim_{z\to a} |f(z)|$ exists, does that imply anything about $\lim_{z\to a} f(z)$? [Hint: *a counterexample would be nice.*]

The squeeze theorem also has a highly useful counterpart for functions. However, the phrasing of this squeeze theorem in complex analysis is more restrictive than its (hopefully) familiar phrasing in real-valued calculus, as we cannot compare outputs of complex, nonreal-valued functions using inequalities.

2.2.14 Theorem (Squeeze theorem for functions). *Suppose that* $f, g \colon \mathcal{D} \to \mathbb{C}$ *and* $a \in \mathbb{C}$ *is an accumulation point of* $\mathcal{D}$. *Suppose also that* $|f(z)| \le |g(z)|$ *for all* $z \in \mathcal{D} \setminus \{a\}$ *and* $\lim_{z \to a} g(z) = 0$. *Then* $\lim_{z \to a} f(z) = 0$ *as well.*

2.2.15 Problem (!). Use the squeeze theorem for sequences to prove the squeeze theorem for functions.

2.2.16 Example. The exponential is well-behaved under limits: $\lim_{z \to a} e^z = e^a$ for all $a \in \mathbb{C}$. (This is really a statement about the continuity of the exponential, of course.) By (2.2.4), this is equivalent to

$$\lim_{z \to a}(e^z - e^a) = 0.$$

Most of our good results for the exponential come from the functional equation, so the trick here is to use the functional equation to expose the difference $z - a$ lurking within $e^z - e^a$. Specifically, we rewrite

$$e^z - e^a = e^{z+a-a} - e^a = e^{z-a}e^a - e^a = e^a\big(e^{z-a} - 1\big). \tag{2.2.5}$$

We claim the existence of $C > 0$ such that if $|w| \le 1$, then

$$|e^w - 1| \le C|w|. \tag{2.2.6}$$

Thus

$$|e^z - e^a| = |e^a||e^{z-a} - 1| \le Ce^a|z - a|.$$

Since $\lim_{z \to a}(z - a) = 0$, the squeeze theorem implies $\lim_{z \to a} |e^z - e^a| = 0$, thus $\lim_{z \to a}(e^z - e^a) = 0$, and so $\lim_{z \to a} e^z = e^a$.

Specifically, to invoke the squeeze theorem, we might put $\mathcal{D} = \mathcal{B}(a; 1)$, $f(z) = e^z - e^a$, and $g(z) = C(z - a)$. Then if $z \in \mathcal{D}$, we have $|z - a| < 1$, so with $w = z - a$, the estimate (2.2.6) gives $|f(z)| \le |g(z)|$.

2.2.17 Problem (+). Prove the estimate (2.2.6). [Hint: *use the power series definition of the exponential to rewrite*

$$e^w - 1 = w\sum_{j=0}^{\infty} \frac{w^j}{(j+1)!}.$$

Then use the comparison test (and maybe the ratio test) to show that the series on the right is uniformly bounded by some $C > 0$ *when* $w| \le 1$.]

2.2.3. Limits and geometry.

So far, we have used various calculus techniques to ensure that limits exist and to compute them. A standard way to "break" limits is to approach a point from two different directions and show that the limits "along those directions" exist but are not equal. In other words, when approaching a point from one direction, the function tends to a certain value, but along a different approach the function has different behavior. There are only two directions of approach (left and right) for functions on $\mathbb{R}$, but in $\mathbb{C}$ there are infinitely many, thanks to the two-dimensional geometry of $\mathbb{C}$. As in multivariable calculus, this makes it harder for limits to exist in $\mathbb{C}$ and, conversely, adds some "strength" to limits when they do exist.

2.2.18 Example. Define

$$f\colon \mathbb{C}\setminus\{0\} \to \mathbb{C}\colon z \mapsto \frac{\overline{z}}{z}.$$

Then 0 is an accumulation point of $\mathbb{C}\setminus\{0\}$, and f is not defined at 0. We show that $\lim_{z\to 0} f(z)$ does not exist by approaching 0 along the real and imaginary axes.

Put

$$z_k := \frac{1}{k} \quad \text{and} \quad w_k := \frac{i}{k}$$

for $k \ge 1$, so both $z_k \to 0$ and $w_k \to 0$ with $z_k \in \mathbb{R}$ and $w_k \in i\mathbb{R}$. Then

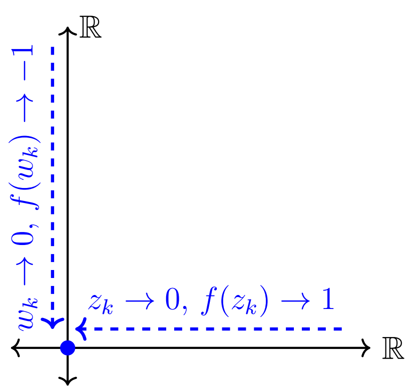
$$f(z_k) = \frac{\overline{z_k}}{z_k} = \frac{z_k}{z_k} = 1,$$

since z_k is real, and so $f(z_k) \to 1$. But

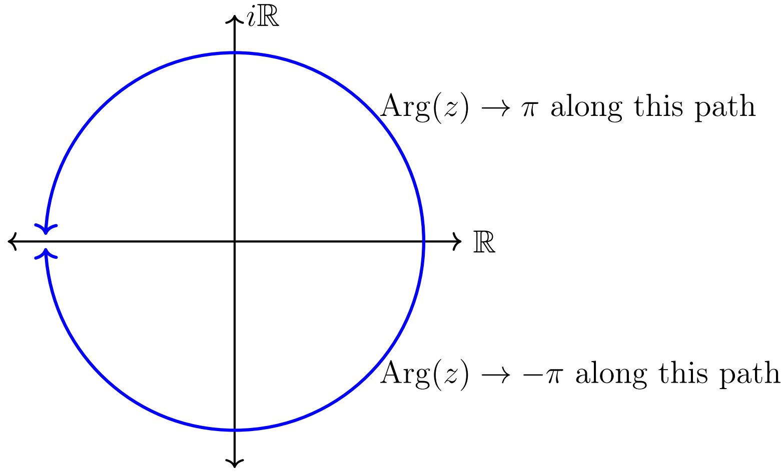
$$f(w_k) = \frac{\overline{w_k}}{w_k} = \frac{\overline{i/k}}{i/k} = -\frac{i/k}{i/k} = -1,$$

and so $f(w_k) \to -1$. Consequently, $\lim_{z\to 0} f(z)$ cannot exist by the definition of limit.

Here is a picture of how we approached 0 in different directions and got different behaviors of f.



2.2.19 Example. The reasoning in Example 1.6.7 suggests that $\lim_{z\to -1} \operatorname{Arg}(z)$ does not exist (or, more broadly, that the limit as $z \to -x$ with $x > 0$ does not exist). We can formalize this using the sequential characterization of limits, but all the key ideas come from the expectations of the picture in that example, which we redraw here.



We want to find two sequences (z_k) and (w_k) such that $z_k \to -1$ and $w_k \to -1$, but $\operatorname{Arg}(z_k) \to \pi$ and $\operatorname{Arg}(w_k) \to -\pi$. One way to do this is to put

$$z_k := e^{i(\pi-1/k)} \quad \text{and} \quad w_k := e^{i(-\pi+1/k)}$$

for $k \ge 1$. Then $0 < \pi - 1/k \le \pi$, and so $\operatorname{Arg}(e^{i(\pi-1/k)}) = \pi - 1/k$. And $-\pi < -\pi + 1/k < 0$, and so $\operatorname{Arg}(e^{i(-\pi+1/k)}) = -\pi + 1/k$. Thus z_k and w_k have the desired behavior. (By the way, $w_k = -\overline{z_k}$.)

2.2.20 Problem (!). Adapt the reasoning of Example 2.2.19 to show that $\lim_{z \to -x} \operatorname{Arg}(z)$ does not exist for any $x > 0$.

This is where we finished on Friday, February 2, 2024.

2.2.21 Example. We revisit Example 1.6.14, in which we intuited that $\lim_{z \to i} \arg_{\pi/2}(z)$ would not exist. The idea in that example was that approaching i from Quadrant I versus Quadrant II would lead to different behaviors of $\arg_{\pi/2}$. Specifically, we saw that if z is in Quadrant I ($\operatorname{Re}(z) > 0$ and $\operatorname{Im}(z) > 0$), then $2\pi < \arg_{\pi/2}(z) < \pi/2 + 2\pi$, but if z is in Quadrant II ($\operatorname{Re}(z) < 0$ and $\operatorname{Im}(z) > 0$), then $\pi/2 < \arg_{\pi/2}(z) < \pi$. So, if z is in Quadrant I, then $\arg_{\pi/2}(z)$ is at least 2π, but if z is in Quadrant II, then $\arg_{\pi/2}(z)$ is at most π. How, then, can $\arg_{\pi/2}$ approach a fixed value as z becomes arbitrarily close to i? (It cannot.)

All of the important reasoning to show that $\lim_{z \to i} \arg_{\pi/2}(z)$ does not exist is contained in the previous paragraph and discussed (and drawn) in more detail in Example 1.6.14. All that we do now is check that the formal definition of the limit does not hold. Our strategy is to construct two sequences (z_k) and (w_k) in $\mathbb{C} \setminus \{i, 0\}$ (note that the domain of $\arg_{\pi/2}$ is $\mathbb{C} \setminus \{0\}$) such that $z_k \to i$, $w_k \to i$, but $\lim_{k\to\infty} \arg_{\pi/2}(z_k) \ne \lim_{k\to\infty} \arg_{\pi/2}(w_k)$. The easiest choice is to note that $i = e^{i\pi/2}$ and then "perturb" from $\pi/2$ to send the points of the sequences into either Quadrant I or II. So, put

$$z_k := e^{i(\pi/2-1/k)} \quad \text{and} \quad w_k := e^{i(\pi/2+1/k)},$$

so z_k is in Quadrant I and w_k is in Quadrant II. Now, recall that $\arg_{\pi/2}$ is characterized by the properties that

$$t = \arg_{\pi/2}(z) \iff z = |z|e^{it} \text{ and } \frac{\pi}{2} < t < \frac{\pi}{2} + 2\pi.$$

For $k \in \mathbb{N}$, we have $\pi/2 - 1/k < \pi/2$, so $\arg_{\pi/2}(z_k) \neq \pi/2 - 1/k$. But, instead, $z_k = e^{i(\pi/2-1/k+2\pi)}$ and $\pi/2 < \pi/2-1/k+2\pi < \pi/2+2\pi$, so $\arg_{\pi/2}(z_k) = \pi/2-1/k+2\pi \to \pi/2+2\pi$. However, $\pi/2 < \pi/2+1/k < \pi/2+2\pi$, so $\arg_{\pi/2}(w_k) = \pi/2+1/k \to \pi/2$. This allows us to conclude that $\lim_{z \to i} \arg_{\pi/2}(z)$ does not exist.

2.2.22 Problem (!). Revisit the reasoning of Problem 1.6.15 and show that $\lim_{z \to 1+i} \arg_{\pi/4}(z)$ does not exist.

2.2.23 Problem (+). Let $\alpha \in \mathbb{R}$. Adapt the reasoning of the previous example and problem to show that $\lim_{z \to z_\star} \arg_\alpha(z)$ does not exist for any $z_\star \in \mathbb{C}$ with $\arg_\alpha(z_\star) = \alpha + 2\pi$. To what extent does the branch cut for an argument remind you of the International Date Line?

2.2.24 Problem ($\star$). As the examples above indicate, we often show that a limit fails to exist by approaching the point in question along two different directions, and often those directions are the real and imaginary axes or two arcs of a circle. Here is a situation where we should approach the point along a line that is not an axis.

Let $f(z) := (z/\overline{z})^2$. Let $w \in \mathbb{C} \setminus \{0\}$ and define a sequence (z_k) by $z_k = w/k$. What are $\lim_{k\to\infty} z_k$ and $\lim_{k\to\infty} f(z_k)$? How can you choose w to show that $\lim_{z\to 0} f(z)$ does not exist? Would approaching 0 along just the real and imaginary axes help here, or do you have to consider a third direction of approach?

2.2.4. Limits in $\mathbb{R}$.

For a complex-valued function of a real variable, we can say both more and less about the geometry of its limits. Less, because its independent variable is real, and we can approach a point in $\mathbb{R}$ from (at most) two directions: left and right. More, because we can *only* approach from the left and right, not from the many possibilities of two-dimensional space.

Here is one special case. Let $a, b \in \mathbb{R}$ with $a < b$, and consider a function $f\colon (a,b] \to \mathbb{C}$. If $\lim_{t\to a} f(t) = L$ for some $L \in \mathbb{C}$, then Definition 2.2.7 (with $\mathcal{D} = (a,b]$) says that for all sequences (t_k) in $(a,b]$ such that $t_k \to a$, we have $f(t_k) \to L$. (Here we are following the usual custom in this course of denoting real numbers by t, not z.) Since (t_k) is a sequence in $(a,b]$, we must have $a < t_k \le b$ for all k. That is, we approach a only "from the right."

We generalize this situation into one-sided limits for functions defined on closed, bounded intervals (which are, primarily, the only situations for limits of functions of a real variable that we will consider in any detail).

2.2.25 Definition. *Let a, b, $c \in \mathbb{R}$ with $a \le c \le b$ and let $f \colon [a,b] \setminus \{c\} \to \mathbb{C}$. Let $L \in \mathbb{C}$.*

(i) *We say that $\lim_{t \to c^+} f(t) = L$ if whenever (t_k) is a sequence in $(c,b]$ with $t_k \to c$, then $f(t_k) \to L$.*

(ii) *Let $t \in (a,b]$. We say that $\lim_{t \to c^-} f(t) = L$ if whenever (t_k) is a sequence in $[a,c)$ with $t_k \to c$, then $f(t_k) \to L$.*

2.2.26 Problem (!). Assume the hypotheses of Definition 2.2.25 and show that

$$\lim_{t \to c^+} f(t) = L \iff \lim_{t \to c} f\big|_{(c,b]}(t) = L \quad \text{and} \quad \lim_{t \to c^-} f(t) = L \iff \lim_{t \to c} f\big|_{[a,c)}(t) = L.$$

This is mostly an exercise in reading that definition and in thinking about restrictions and Definition 2.2.7. Note that some statements are vacuously true if $c = a$ or $c = b$.

Of course, limits from the left and the right would be useless if they did not talk to each other correctly.

2.2.27 Theorem. *Let a, b, $c \in \mathbb{R}$ with $a \le c \le b$ and let $f \colon [a,b] \setminus \{c\} \to \mathbb{C}$. Let $L \in \mathbb{C}$.*

(i) *If $c = a$, then $\lim_{t \to a} f(t) = L$ if and only if $\lim_{t \to a^+} f(t) = L$.*

(ii) *If $c = b$, then $\lim_{t \to b} f(t) = L$ if and only if $\lim_{t \to b^-} f(t) = L$.*

(iii) *If $c \in (a,b)$, then $\lim_{t \to c} f(t) = L$ if and only if both $\lim_{t \to c^-} f(t) = L$ and $\lim_{t \to c^+} f(t) = L$.*

Proof. **(i)** Here it is important to remember the definition of limit: for $f \colon [a,b] \setminus \{a\} \subseteq \mathbb{C} \to \mathbb{C}$, we have $\lim_{t \to a} f(t) = L$ if and only if whenever (t_k) is a sequence in $[a,b] \setminus \{a\}$ with $t_k \to t$, we also have $f(t_k) \to L$. Since $[a,b] \setminus \{a\} = (a,b]$, the definitions of $\lim_{t \to a} f(t)$ and $\lim_{t \to a^+} f(t)$ are equivalent.

(ii) This is the same as the above, except now we use $[a,b] \setminus \{b\} = [a,b)$.

(iii) That $\lim_{t \to t} f(t) = L$ implies both $\lim_{t \to t^-} f(t) = L$ and $\lim_{t \to t^+} f(t) = L$ is a direct consequence of the definitions. The converse requires more work.

Let (t_k) be a sequence in $[a,b] \setminus \{t\}$ such that $t_k \to c$. First suppose $t_k < c$ for all but finitely many k. Then we can delete those t_k from the sequence and obtain (a not-relabeled) sequence (t_k) in $[a,c)$ such that $t_k \to c$. Since $\lim_{t \to t^-} f(t) = L$, we have $f(t_k) \to L$. The same can be done if $t_k > t$ for all but finitely many k.

So, it remains to consider the case in which $t_k < t$ for infinitely many k and also $t_k > t$ for infinitely many k. Let (ℓ_j) be the sequence of integers such that $t_{\ell_j} < c$ and let (m_j) be the sequence of integers such that $t_{m_j} > c$. Since $t_k \ne c$ for all k, we have $\{t_k\}_{k=1}^{\infty} =$

$\{t_{\ell_j}\}_{j=1}^{\infty} \cup \{t_{m_j}\}_{j=1}^{\infty}$. It follows that $t_{\ell_j} \to c$ and $t_{m_j} \to c$, so by the existence of the left and right limits $f(t_{\ell_j}) \to L$ and $f(t_{m_j}) \to L$. And from this it follows that $f(t_k) \to L$. ■

2.2.28 Problem (+). Justify more carefully the last two sentences of the preceding proof (the ones using the weaselly phrase "it follows"). Your justification should involve the definition of the limit of a sequence and the letter ϵ.

2.2.29 Problem (!). We know that $\lim_{z \to -1} \operatorname{Arg}(z)$ does not exist, but how about $\lim_{t \to -1} \operatorname{Arg}\big|_{(-\infty,0)}(t)$? This is a good opportunity to reflect on how the domain $\mathcal{D}$ of a function $f \colon \mathcal{D} \to \mathbb{C}$ plays a critical, but sometimes understated, role in the limit properties of that function. Here we restate the symbolic definition (2.3.1) of a limit with the appearance of $\mathcal{D}$ highlighted more prominently:

$$\lim_{z \to a} f(z) = L \iff \big(\forall \epsilon > 0 \ \exists \delta > 0 : 0 < |z - a| < \delta \text{ and } z \in \mathcal{D} \Longrightarrow |f(z) - L| < \epsilon\big).$$

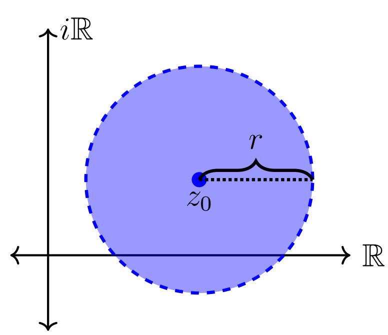
The different limit behaviors of the function Arg, with domain $\mathbb{C} \setminus \{0\}$, and $\operatorname{Arg}\big|_{(-\infty,0)}$, with domain $(-\infty, 0)$, illustrate how changing the domain of a function can change the limit behavior of that function! (Of course, if we change the domain of a function, then we get a new function, as the baroque "ordered triple" remarks in Definition 1.2.2 emphasize.)

2.3. Limits and topology.

While we have used geometry to help us compute limits (or, rather, disprove their existence), our *definition* of limit has been strictly algebraic—if one sequence converges, so does another. There is a more geometric, dynamic perspective on limits that we now present. To describe this perspective, we need an extremely useful species of subset of $\mathbb{C}$ that will accompany us for the rest of the course. We already met the first and last members of this species in Definition 1.1.11.

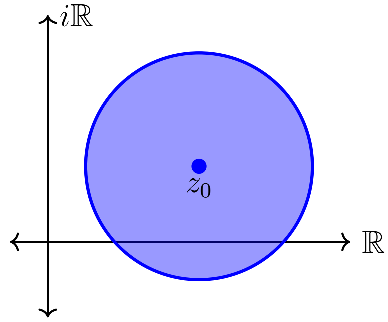
2.3.1 Definition. **(i)** *The* OPEN BALL *of radius $r > 0$ centered at $z_0 \in \mathbb{C}$ is*

$$\mathcal{B}(z_0; r) := \{z \in \mathbb{C} \mid |z - z_0| < r\}.$$



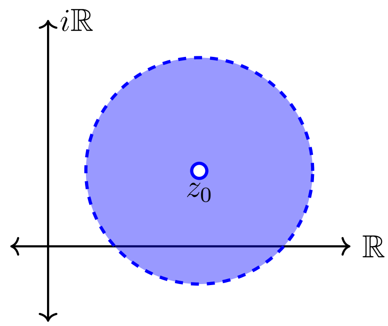
(ii) *The* **CLOSED BALL** *of radius $r > 0$ centered at $z_0 \in \mathbb{C}$ is*

$$\overline{\mathcal{B}}(z_0; r) := \{z \in \mathbb{C} \mid |z - z_0| \leq r\}\,.$$



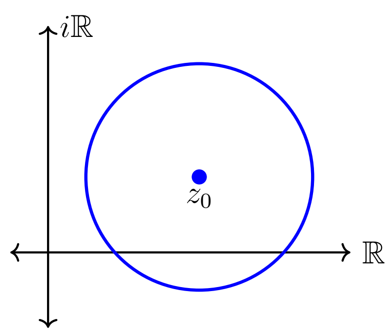
(iii) *The* **PUNCTURED (OPEN) BALL** *of radius $r > 0$ centered at $z_0 \in \mathbb{C}$ is*

$$\mathcal{B}^*(z_0; r) := \{z \in \mathbb{C} \mid 0 < |z - z_0| < r\}\,.$$



(iv) *The* **CIRCLE** *of radius $r > 0$ centered at $z_0 \in \mathbb{C}$ is*

$$\partial\mathcal{B}(z_0; r) := \{z \in \mathbb{C} \mid |z - z_0| = r\}\,.$$



Like $\partial\mathcal{B}(z_0; r)$, the notational choice of $\overline{\mathcal{B}}(z_0; r)$ is meant to reflect the more general topological concept of closure, which we will not discuss in this course.

2.3.2 Example. The drawings above should indicate that we can construct the open ball of radius r centered at z_0 by removing the circle of radius r centered at z_0 from the closed ball of radius r centered at z_0. That is, we expect

$$\mathcal{B}(z_0; r) = \overline{\mathcal{B}}(z_0; r) \setminus \partial\mathcal{B}(z_0; r).$$

This is indeed the case, as we now show. We have $z \in \mathcal{B}(z_0; r)$ if and only if $|z - z_0| < r$. The inequality $|z - z_0| < r$ is true if and only if $|z - z_0| \le r$ and $|z - z_0| \ne r$. In turn, $|z - z_0| \le r$ if and only if $z \in \overline{\mathcal{B}}(z_0; r)$, while $|z - z_0| \ne r$ if and only if $z \notin \partial\mathcal{B}(z_0; r)$. So, we have

$$z \in \mathcal{B}(z_0; r) \iff z \in \overline{\mathcal{B}}(z_0; r) \setminus \partial\mathcal{B}(z_0; r),$$

and this establishes the desired set equality.

The following problems offer lots of practice with ball notation and mechanics, and we will call upon many of these results in future technical steps.

2.3.3 Problem (!). Let $z_0 \in \mathbb{C}$ and $r > 0$. Prove the following using only Definition 2.3.1.

(i) $\overline{\mathcal{B}}(z_0; r) = \mathcal{B}(z_0; r) \cup \partial\mathcal{B}(z_0; r)$.

(ii) $\mathcal{B}^*(z_0; r) = \mathcal{B}(z_0; r) \setminus \{z_0\}$.

(iii) $\mathcal{B}(z_0; r) \subseteq \overline{\mathcal{B}}(z_0; r)$.

(iv) $\overline{\mathcal{B}}(z_0; r) \subseteq \mathcal{B}(z_0; R)$ if $r \le R$. Draw a picture illustrating this phenomenon when $r < R$.

2.3.4 Problem (⋆). It will sometimes be helpful to take a "polar" perspective on balls and circles. Let $z_0 \in \mathbb{C}$ and $r > 0$.

(i) Prove that $\mathcal{B}(z_0; r) = \{z_0 + \rho e^{i\theta} \mid 0 \le \rho < r,\ 0 \le \theta \le 2\pi\}$.

(ii) Give similar "polar" descriptions of $\overline{\mathcal{B}}(z_0; r)$ and $\mathcal{B}^*(z_0; r)$.

2.3.5 Problem (⋆). Let $z_0 \in \mathbb{C}$ and $r > 0$. Prove that $a \in \mathbb{C}$ is an accumulation point of $\mathcal{B}(z_0; r)$ if and only if $a \in \overline{\mathcal{B}}(z_0; r)$. [Hint: ($\Longrightarrow$) *If $z_n \to a$, then $|a - z_0| = \lim_{n\to\infty} |z_n - z_0|$. If $|z_n - z_0| < r$, what does this imply about $|a - z_0|$? ($\Longleftarrow$) Write $a = z_0 + re^{i\theta}$ and consider $z_n = z_0 + \rho_n e^{i\theta}$ with ρ_n suitably chosen.*]

2.3.6 Problem (⋆). Let x_0, y_0, x_1, $y_1 \in \mathbb{R}$ with $x_0 < x_1$ and $y_0 < y_1$. Let $r > 0$. Suppose that $x_1 + iy_1 \in \mathcal{B}(x_0 + iy_0; r)$. Prove that if $x \in [x_0, x_1]$ and $y \in [y_0, y_1]$, then $x + iy \in \mathcal{B}(x_0 + iy_0; r)$. Draw a picture, too.

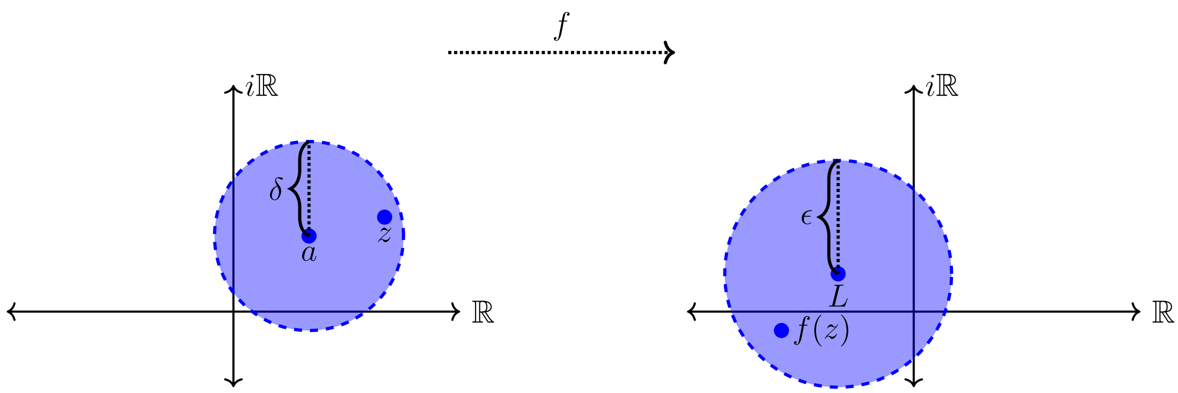
Now we use balls to develop a useful geometric counterpart to Theorem 1.3.7 for functional limits that does not involve sequences. For a function $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$, an accumulation point $a \in \mathbb{C}$ of $\mathcal{D}$, and a point $L \in \mathbb{C}$, the intuitive meaning of the sentence $\lim_{z\to a} f(z) = L$ is that we can make $f(z)$ and L arbitrarily close by taking z and a sufficiently close (but not necessarily equal). We quantify "arbitrarily close" by desiring that $|f(z) - L| < \epsilon$ for some $\epsilon > 0$. Then we quantify "sufficiently close" by hoping that taking $z \in \mathcal{D}$ with $0 < |z-a| < \delta$ for some $\delta > 0$ will force $|f(z) - L| < \epsilon$. The lower bound $0 < |z - a|$ is necessary to ensure that we are not assuming $z = a$. In symbols, we are hoping

$$\lim_{z\to a} f(z) = L \iff \big(\forall \epsilon > 0\ \exists \delta > 0 : z \in \mathcal{D} \text{ and } 0 < |z-a| < \delta \Longrightarrow |f(z)-L| < \epsilon\big). \quad (2.3.1)$$

Now we translate these inequalities into balls. We have $0 < |z - a| < \delta$ if $z \in \mathcal{B}^*(a;\delta)$, and so we have $z \in \mathcal{D}$ with $0 < |z-a| < \delta$ if $z \in \mathcal{D} \cap \mathcal{B}^*(a;\delta)$. Next, we have $|f(z) - L| < \epsilon$ if $f(z) \in \mathcal{B}(L;\epsilon)$. And so the symbolic counterpart to (2.3.1) in terms of balls is

$$\lim_{z\to a} f(z) = L \iff \big(\forall \epsilon > 0\ \exists \delta > 0 : z \in \mathcal{D} \cap \mathcal{B}^*(a;\delta) \Longrightarrow f(z) \in \mathcal{B}(L;\epsilon)\big). \quad (2.3.2)$$

Here is a cartoon of the if-then statement on the right of the if-and-only-if statement above (assuming, for convenience, $\mathcal{D} = \mathbb{C}$).



Of course, both (2.3.1) and (2.3.2) turn out to be true, and, indeed, we could have started with either of them as the definition of limit instead of Definition 2.2.7 with sequences. We took the approach of defining limits via sequences because sequences are so helpful in breaking limits geometrically in complex analysis, but this is far from the only approach. Here is the formal conclusion from the work above.

2.3.7 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ and $f\colon \mathcal{D} \to \mathbb{C}$. Let $L \in \mathbb{C}$ and let $a \in \mathbb{C}$ be an accumulation point of $\mathcal{D}$. Then $\lim_{z\to a} f(z) = L$ if and only if for any $\epsilon > 0$, there is $\delta > 0$ such that if $z \in \mathcal{D}$ and $0 < |z - a| < \delta$, then $|f(z) - L| < \epsilon$.*

2.3.8 Problem (+). Prove this theorem. [Hint: *for the reverse direction, show that if $z_k \to a$, then $f(z_k) \to L$. How do ϵ, δ, and k all talk to each other? For the forward direction, prove the contrapositive: let P be the statement "$\lim_{z\to a} f(z) = L$" and let Q be the statement "$\forall \epsilon > 0\ \exists \delta > 0 : z \in \mathcal{D}$ and $0 < |z - a| < \delta \Longrightarrow |f(z) - L| < \epsilon$." Assume* *that the negation of Q is true and prove the negation of P. This involves negating many quantifiers; note that P contains various quantifiers, per Definitions 2.2.7 and 1.3.3. How do all of these quantifiers interact, and, again, how do ϵ and δ interact with the index k of the sequence in play?*]

Going further, we can use the language of functional images to say

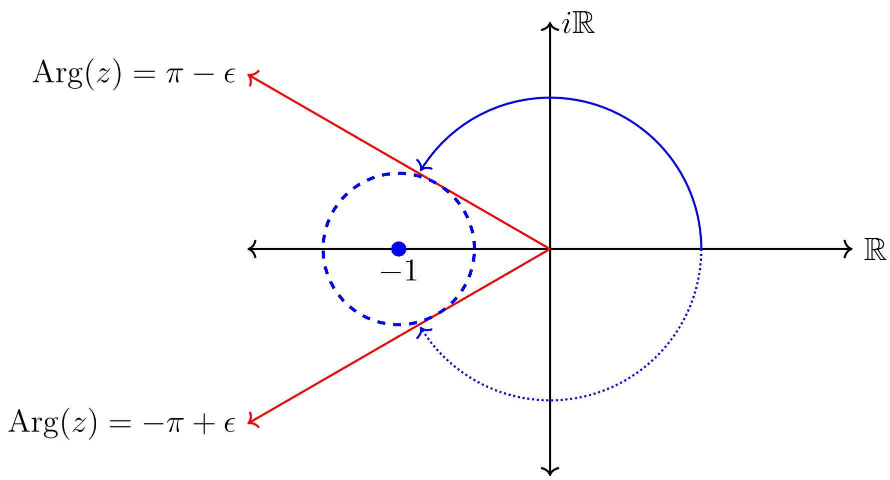
$$\lim_{z \to a} f(z) = L \iff \Big(\forall \epsilon > 0 \ \exists \delta > 0 : f\big(\mathcal{B}^*(a;\delta) \cap \mathcal{D}\big) \subseteq \mathcal{B}(L;\epsilon)\Big). \qquad (2.3.3)$$

2.3.9 Problem (!). Go further and verify (2.3.3).

2.3.10 Example. In Example 2.2.19, we saw that $\lim_{z \to -1} \operatorname{Arg}(z)$ does not exist for any $x > 0$. That is, the principal argument does not have a limit at any point on the negative real axis. However, consider the ameliorating effect of taking the modulus. As z approaches -1, we expect that $\operatorname{Arg}(z)$ will become close to either π or $-\pi$, but depending on the direction of the approach, $\operatorname{Arg}(z)$ does not have to approach π or $-\pi$ exclusively. But if we take the modulus, we should find that $|\operatorname{Arg}(z)|$ gets close to just π. That is, we expect

$$\lim_{z \to -1} |\operatorname{Arg}(z)| = \pi.$$

Here is one way to show this more rigorously. Given $\epsilon > 0$, we can draw a small ball around -1 that is "wedged" between the rays $\operatorname{Arg}(z) = \pi - \epsilon$ and $\operatorname{Arg}(z) = -\pi + \epsilon$. The radius of this ball will be the δ that we use in Theorem 2.3.7.



Then every point z in this ball (indeed, in this wedge) will satisfy either

$$\pi - \epsilon < \operatorname{Arg}(z) \le \pi \text{ if } \operatorname{Im}(z) \ge 0 \quad \text{or} \quad -\pi \le \operatorname{Arg}(z) < -\pi + \epsilon \text{ if } \operatorname{Im}(z) \le 0.$$

In either case, we can manipulate the inequalities above to show $\big||\operatorname{Arg}(z)| - \pi\big| \le \epsilon$. That is, given $\epsilon > 0$, with δ as the radius of the ball above, we have $\big||\operatorname{Arg}(z)| - \pi\big| \le \epsilon$, and so Theorem 2.3.7 implies that $\lim_{z \to -1} |\operatorname{Arg}(z)| = \pi$.

2.3.11 Problem ($\star$). **(i)** In the previous example, find a formula for δ in terms of ϵ.

(ii) Prove that if $\pi - \epsilon < \operatorname{Arg}(z) \le \pi$ if $\operatorname{Im}(z) \ge 0$ or $-\pi \le \operatorname{Arg}(z) < -\pi + \epsilon$, then $\big||\operatorname{Arg}(z)| - \pi\big| \le \epsilon$.

2.3.12 Problem ($+$). We have seen (Theorem 2.2.12) that the limit behavior of a function reduces to the *real-valued* limit behavior of its real and imaginary parts. How does the limit behavior of the real and imaginary parts of that function's *independent variable* interact with the limit behavior of that function?

(i) Suppose that $f\colon \mathcal{D} \to \mathbb{C}$ with $a \in \mathbb{C}$ an accumulation point of $\mathcal{D}$ and $L \in \mathbb{C}$ satisfies $\lim_{z\to a} f(z) = \mathcal{D}$. Let $I := \{t \in \mathbb{R} \setminus \{0\} \mid t + i\operatorname{Im}(a) \in \mathcal{D}\}$. First show that $\operatorname{Re}(a)$ is an accumulation point of I. Then define

$$\widetilde{f}\colon I \to \mathbb{C}\colon t \mapsto f(t + i\operatorname{Im}(a)).$$

Show that $\lim_{t\to\operatorname{Re}(a)} \widetilde{f}(t) = L$. The same can be done replacing $\operatorname{Re}(a)$ with $\operatorname{Im}(a)$ throughout. Thus the existence of the limit in the independent variable z implies the existence of the "limit in the real and imaginary parts of the independent variable."

(ii) Take $\mathcal{D} = \mathbb{C}\setminus\{0\}$, $a = 0$, and $f(z) = \overline{z}/z$. Let $I = \mathbb{R}\setminus\{0\}$ and define $\widetilde{f}_1(t) := f(t)$ and $\widetilde{f}_2(t) := f(it)$ for $t \in I$. Show that $\lim_{t\to 0}\widetilde{f}_1(t)$ and $\lim_{t\to 0}\widetilde{f}_2(t)$ both exist but $\lim_{z\to 0} f(z)$ does not (all limits were done in an earlier example). Thus the existence of the "limits in the real and imaginary parts of the independent variable" does not imply the existence of the limit in the independent variable z.

2.4. Continuity.

Now that we have a robust knowledge of limits, our treatment of continuity can proceed mostly as it did in calculus.

2.4.1. The definition of continuity and examples.

First, we define continuity exactly as we (probably) met it in calculus.

2.4.1 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$. A function $f\colon \mathcal{D} \to \mathbb{C}$ is* CONTINUOUS AT *$a \in \mathcal{D}$ if a is an accumulation point of $\mathcal{D}$ and if $\lim_{z\to a} f(z) = f(a)$. If f is not continuous at $a \in \mathcal{D}$, or if $a \in \mathbb{C}\setminus\mathcal{D}$, then f is* DISCONTINUOUS *at a.*

It is worthwhile (over)emphasizing that f can fail to be continuous at $a \in \mathbb{C}$ for three reasons.

1. The point a is not in the domain of f, or a is in the domain of f but is not an accumulation point of the domain. (In either of these situations, it is worthwhile pondering if it is even fair to *ask* if f is continuous at a, since we are basically setting ourselves up for failure.)

2. The limit $\lim_{z \to a} f(z)$ does not exist, whether or not a belongs to the domain of f.

3. The point a is in the domain of f and $\lim_{z \to a} f(z)$ exists but does not equal $f(a)$.

2.4.2 Problem (!). Let $\mathcal{D} \subseteq \mathbb{C}$ and $a \in \mathcal{D}$ be an accumulation point of $\mathcal{D}$. Prove that $f: \mathcal{D} \to \mathbb{C}$ is continuous at a if

$$\forall \epsilon > 0 \ \exists \delta > 0 : z \in \mathcal{B}(a; \delta) \Longrightarrow f(z) \in \mathcal{B}(f(a); \epsilon).$$

All of the algebraic rules for continuity that we expect to be true are true. Specifically, the limits in Theorem 2.2.10 carry over to continuity rules. Composition also interacts well with continuity.

2.4.3 Theorem. **(i)** *Let $\mathcal{D} \subseteq \mathbb{C}$ and let $a \in \mathcal{D}$ be an accumulation point of $\mathcal{D}$. Let f, $g: \mathcal{D} \to \mathbb{C}$ be continuous at a. Then $f + g$ and fg are continuous at a; so is f/g if $g(a) \neq 0$, and so is αf for any $\alpha \in \mathbb{C}$. Likewise, $\overline{f}$ and $|f|$ are continuous at a.*

(ii) *Let $\mathcal{D}_1$, $\mathcal{D}_2 \subseteq \mathbb{C}$ and let $f: \mathcal{D}_1 \to \mathbb{C}$ and $g: \mathcal{D}_2 \to \mathbb{C}$ with $f(\mathcal{D}_1) \subseteq \mathcal{D}_2$. If $a \in \mathcal{D}_1$ is an accumulation point of $\mathcal{D}_1$ and $f(a)$ is an accumulation point of $\mathcal{D}_2$, and if f is continuous at a and g is continuous at $f(a)$, then $g \circ f$ is continuous at a.*

This is where we finished on Monday, February 5, 2024.

Additionally, Theorem 2.2.12 allows us to characterize continuity of a function in terms of the continuity of its real and imaginary parts. This is a straightforward, but important, part of our quest to see how the calculus properties of the real-valued real and imaginary parts of a function interact with the calculus properties of the whole function.

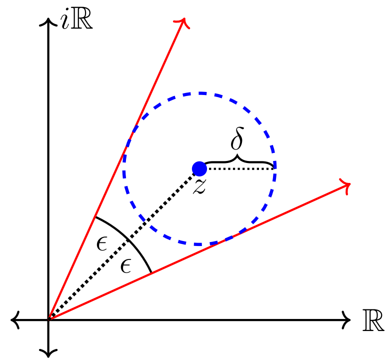
2.4.4 Theorem. *Let $f: \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ and let $a \in \mathcal{D}$ be an accumulation point of $\mathcal{D}$. Then f is continuous at a if and only if both $\text{Re}[f]$ and $\text{Im}[f]$ are continuous at a.*

2.4.5 Example. **(i)** Example 2.2.16 shows that the exponential is continuous on $\mathbb{C}$, and Example 2.2.11 shows that polynomials are continuous on $\mathbb{C}$ and rational functions are continuous except at the roots of their denominators.

(ii) Example 2.2.19 shows that $\lim_{z \to -x} \text{Arg}(z)$ does not exist for any $x > 0$, so Arg is discontinuous at each point in $(-\infty, 0)$. Since Arg is not defined at 0, Arg is also discontinuous there.

The techniques that prove the existence of the principal argument in $(\pi 4)$ of Theorem 1.5.10, which we did not present, could also show that Arg is continuous on $\mathbb{C} \setminus (-\infty, 0]$. We will not make such a formal argument, but hopefully a picture should make the continuity of Arg reasonable.

Fix $z \in \mathbb{C} \setminus (-\infty, 0]$ and $\epsilon > 0$ and draw a small ball of radius δ around z, as we did in Example 2.3.10. If we take δ to be small enough, then every point in this ball should have principal argument within $\pm\epsilon$ of $\operatorname{Arg}(z)$.



(iii) In Example **??**, we showed that $\lim_{z\to a} \arg_{\pi/4}(z)$ does not exist for any $a \in \mathbb{C}$ with $\operatorname{Arg}(a) = \pi/4$, i.e., for any a on the branch cut of $\arg_{\pi/4}$. Consequently, $\arg_{\pi/4}$ is discontinuous on its branch cut.

(iv) For $z \in \mathbb{C} \setminus \{0\}$, we have $\operatorname{Log}(z) = \ln(|z|) + i\operatorname{Arg}(z)$. We know that Arg is continuous on $\mathbb{C} \setminus (-\infty, 0]$. Algebraic rules for limits tell us that the mapping $z \mapsto |z|$ is continuous. Finally, it is possible to prove, using deeper techniques of analysis, that the map $\ln\colon (0,\infty) \to \mathbb{R}$ satisfying $e^{\ln(t)} = t$ for all $t > 0$ is continuous. (This is an "inverse function theorem" sort of argument.) Since $\operatorname{Re}[\operatorname{Log}(z)] = \ln(|z|)$ and $\operatorname{Im}[\operatorname{Log}(z)] = \operatorname{Arg}(z)$, and since these real and imaginary parts are continuous on $\mathbb{C} \setminus (-\infty, 0]$, we conclude that Log is continuous on $\mathbb{C}\setminus(-\infty, 0]$, too. And since Arg is discontinuous on $(-\infty, 0]$, we know that Log is discontinuous on $(-\infty, 0]$, too.

2.4.6 Remark. *At this point, it is a perfectly natural, normal thing to feel wholly unsatisfied with our development of logarithms and arguments. We have made some major assumptions about their existence, and all subsequent proofs rely strongly on those unproven assumptions. However, it is possible to develop more or less from scratch their existence and fundamental properties. Specifically, we will use integrals and calculus to develop a map* $L\colon \mathbb{C}\setminus\{0\} \to \mathbb{C}$ *such that* $e^{L(z)} = z$ *for all* $z \in \mathbb{C}$ *and* $L(e^t) = t$ *for all* $t \in \mathbb{R}$*; then we will develop a map* $A\colon \mathbb{C} \setminus \{0\} \to (-\pi, \pi]$ *such that* $z = |z|e^{iA(z)}$ *for all* $z \in \mathbb{C} \setminus \{0\}$*. Along the way, we will prove the continuity (and differentiability) of* L *and* A *on* $\mathbb{C} \setminus (-\infty, 0]$.

2.4.7 Problem (!). **(i)** Explain why the function $\operatorname{Arg}\colon \mathbb{C} \setminus \{0\} \to \mathbb{R}$ is discontinuous at each point in $(-\infty, 0)$, but the restriction $\operatorname{Arg}\big|_{(-\infty,0)}\colon (-\infty, 0) \to \mathbb{R}$ is continuous. [Hint: *think carefully about the role of* $\mathcal{D}$ *in Definition 2.4.1.*]

(ii) Use Example 2.3.10 to argue that the function $f\colon \mathbb{C} \setminus \{0\} \to \mathbb{R}\colon z \mapsto |\operatorname{Arg}(z)|$ is continuous at each point in $(-\infty, 0)$.

2.4.8 Problem ($\star$). Let a, $b \in \mathbb{R}$ with $a \leq b$ and let $f\colon [a, b] \to \mathbb{C}$. Prove the following.

(i) f is continuous at a if and only if $\lim_{\tau \to a^+} f(\tau) = f(a)$.

(ii) f is continuous at b if and only if $\lim_{\tau \to b^-} f(\tau) = f(b)$.

(iii) f is continuous at $t \in (a, b)$ if and only if

$$\lim_{\tau \to t^-} f(\tau) = f(t) \quad \text{and} \quad \lim_{\tau \to t^+} f(\tau) = f(t).$$

2.4.9 Problem (!). Reread Example 1.6.14. Then, for any $\alpha \in \mathbb{R}$, make a conjecture about where $\arg_\alpha$ is discontinuous. Do not try to prove your conjecture but instead discuss the process of how you made it.

2.4.10 Problem (+). This problem outlines a proof that no "argument function" can be continuous on all of $\mathbb{C} \setminus \{0\}$.

(i) Let $I \subseteq \mathbb{R}$ be an interval. Show that if $f \colon I \to \mathbb{Z}$ is continuous, then f is constant. [Hint: *suppose that f is not constant and use the intermediate value theorem to derive a contradiction.*]

(ii) Suppose that $\Theta \colon \mathbb{C} \setminus \{0\} \to \mathbb{R}$ satisfies $z = |z|e^{i\Theta(z)}$ for all $z \in \mathbb{C} \setminus \{0\}$. If Θ is continuous, deduce the existence of $k_0 \in \mathbb{Z}$ such that $t = \Theta(e^{it}) + 2\pi k_0$ for all $t \in \mathbb{R}$. Obtain from this a contradiction.

2.4.2. Removable discontinuities.

Sometimes a function fails to be continuous at a point (possibly because the function is not defined there), but the failure of continuity is "tame" enough that the discontinuity can be "removed."

2.4.11 Example. Consider the functions that we studied in Example 2.2.1.

(i) The piecewise function

$$f \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto \begin{cases} 1, & z \neq 0 \\ i, & z = 0 \end{cases}$$

is certainly not continuous at 0, because

$$\lim_{z \to 0} f(z) = 1 \neq i = f(0).$$

However, we could put

$$\widetilde{f} \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto 1,$$

so that $\widetilde{f}$ is continuous on $\mathbb{C}$ and $\widetilde{f}(z) = f(z)$ for all $z \neq 0$. We have "removed" the discontinuity of f at 0 with our new definition of $\widetilde{f}$.

(ii) The function
$$g\colon \mathbb{C}\setminus\{0\} \to \mathbb{C}\colon z \mapsto 1$$
is not continuous at 0 since g is not defined there. However, we could put
$$\widetilde{g}\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto 1,$$
so that $\widetilde{g}$ is continuous on $\mathbb{C}$ and $\widetilde{g}(z) = g(z)$ for all $z \neq 0$. Again, we have "removed" the discontinuity of g at 0 by extending g in an appropriate way. Indeed, this is the *only* extension of g that is continuous on $\mathbb{C}$, for if $h\colon \mathbb{C} \to \mathbb{C}$ satisfies $h\big|_{\mathbb{C}\setminus\{0\}} = g$, then we expect
$$h(0) = \lim_{z\to 0} h(z) = \lim_{z\to 0} h\big|_{\mathbb{C}\setminus\{0\}}(z) = \lim_{z\to 0} g(z) = 1.$$

This example suggests that if $\lim_{z\to a} f(z)$ exists but does not equal $f(a)$, or if the limit exists but f is not defined at a, we can probably redefine f to be continuous at a. However, if the limit fails to exist, there is probably no hope of redefining f to be continuous at a.

2.4.12 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ and let $a \in \mathbb{C}$ be an accumulation point of $\mathcal{D}$. Suppose that $f\colon \mathcal{D} \to \mathbb{C}$ is a function that is discontinuous at a. If $\lim_{z\to a} f(z)$ exists, then f has a* **REMOVABLE DISCONTINUITY** *at a. If $\lim_{z\to a} f(z)$ does not exist, then f has a* **NONREMOVABLE DISCONTINUITY** *at a.*

Note that this definition allows for $a \notin \mathcal{D}$, and so f may be discontinuous at a because it is undefined at a.

Now we generalize Example 2.4.11. First, we formalize the (hopefully expected) notion that two functions have the same limit if both functions agree at all points near but not equal to the point of approach.

2.4.13 Lemma. *Let $\mathcal{D} \subseteq \mathbb{C}$ and let $a \in \mathbb{C}$ be an accumulation point of $\mathcal{D}$. For a function $f\colon \mathcal{D} \to \mathbb{C}$, the limit $\lim_{z\to a} f(z)$ exists if and only if $\lim_{z\to a} f\big|_{\mathcal{D}\setminus\{a\}}(z)$ exists, in which case the limits are equal.*

2.4.14 Problem ($\star$). Prove this. [Hint: *when all else fails, give up and go back to the definition—here, Definition 2.2.7.*]

2.4.15 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ and let $a \in \mathbb{C}$ be an accumulation point of $\mathcal{D}$. Suppose that the function $f\colon \mathcal{D} \to \mathbb{C}$ has a removable discontinuity at $a \in \mathbb{C}$ with $L := \lim_{z\to a} f(z)$ and define*
$$\widetilde{f}\colon \mathcal{D}\cup\{a\} \to \mathbb{C}\colon z \mapsto:= \begin{cases} f(z), & z \in \mathcal{D}\setminus\{a\} \\ L, & z = a. \end{cases}$$
Then $\widetilde{f}$ is continuous at a.

Proof. First, note that we do not specify whether $a \in \mathcal{D}$ or not, and so maybe $\mathcal{D} \cup \{a\} = \mathcal{D}$, or not, and maybe $\mathcal{D} \setminus \{a\} = \mathcal{D}$, or not. We are trying to be as general as possible here with how a relates to $\mathcal{D}$.

In any case, we have $\widetilde{f}(z) = f(z)$ for $z \in \mathcal{D} \setminus \{a\}$. That is, $\widetilde{f} = f\big|_{\mathcal{D}\setminus\{a\}}$, and since $\lim_{z \to a} f(z)$ exists, the limit $\lim_{z \to a} \widetilde{f}(z)$ also exists and equals $\lim_{z \to a} f(z)$, by Lemma 2.4.13. That is,

$$\lim_{z \to a} \widetilde{f}(z) = \lim_{z \to a} f(z) = L = \widetilde{f}(a),$$

and so $\widetilde{f}$ is continuous at a. ■

We emphasize that this theorem does not even presume that f is defined at a in the first place, since we were not assuming $a \in \mathcal{D}$.

2.4.16 Example. **(i)** No discontinuity of Arg is removable. First, we know that Arg is discontinuous on $(-\infty, 0)$ because $\lim_{z \to -x} \operatorname{Arg}(z)$ does not exist for any $x > 0$. This means that every point in $(-\infty, 0)$ is a nonremovable discontinuity of Arg. Second, Arg is discontinuous at 0 because Arg is not defined at 0. However, it is also the case that $\lim_{z \to 0} \operatorname{Arg}(z)$ does not exist; this can be seen by approaching 0 along the coordinate axes.

(ii) Define

$$f \colon \mathbb{C} \setminus \{0\} \to \mathbb{C} \colon z \mapsto z \operatorname{Arg}(z)$$

Part (ii) of Example 2.4.5 tells us that f is continuous. However, since $\lim_{z \to 0} \operatorname{Arg}(z)$ does not exist, we cannot compute $\lim_{z \to 0} f(z)$ using algebraic properties of limits; we might be tempted to say

$$\lim_{z \to 0} f(z) = \left(\lim_{z \to 0} z\right)\left(\lim_{z \to 0} \operatorname{Arg}(z)\right) = 0 \cdot \lim_{z \to 0} \operatorname{Arg}(z) = 0,$$

but this is meaningless, since the second limit in the product does not exist.

Nonetheless, we know that $-\pi < \operatorname{Arg}(z) \le \pi$, and so $|\operatorname{Arg}(z)| \le \pi$, thus $|f(z)| \le \pi|z|$. Then the squeeze theorem says that $\lim_{z \to 0} f(z) = 0$, and so f has a removable discontinuity at 0. We can therefore extend f to $\mathbb{C}$ by setting

$$\widetilde{f} \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto \begin{cases} z \operatorname{Arg}(z), & z \ne 0 \\ 0, & z = 0. \end{cases}$$

This function $\widetilde{f}$ is continuous on $\mathbb{C}$, and $\widetilde{f}\big|_{\mathbb{C}\setminus\{0\}} = f$.

2.4.17 Problem (!). Show that $\lim_{z \to 0} \operatorname{Arg}(z)$ does not exist. [Hint: *consider any two of the four sequences* $(1/k)$, $(-1/k)$, (i/k), *or* $(-i/k)$.]

2.4.18 Example. We have shown that $\lim_{z \to -1} \operatorname{Arg}(z)$ does not exist. Can we redefine Arg at -1 to force continuity there? Suppose that we could define a function $f \colon \mathbb{C} \setminus \{0\} \to \mathbb{C}$

such that $f(z) = \operatorname{Arg}(z)$ for $z \neq -1$ and $\lim_{z \to -1} f(z)$ exists. Then (by Lemma 2.4.13), since $f(z) = \operatorname{Arg}(z)$ except at -1, the limit $\lim_{z \to -1} \operatorname{Arg}(z)$ would also exist. This is impossible.

2.4.3. The extreme value theorem.

Recall that if a, $b \in \mathbb{R}$ with $a < b$ and $f \colon [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ is continuous, then f has extreme values on $[a,b]$. Specifically, f has an absolute maximum and an absolute minimum on $[a,b]$: there are t_m, $t_M \in [a,b]$ such that

$$f(t_m) \le f(t) \le f(t_M)$$

for all $t \in [a,b]$. Of course, we should not expect such a result to be true for complex-valued functions, because inequalities do not make sense for complex, nonreal numbers. However, if we incorporate the modulus, we do get an extreme value theorem over the complex plane's analogue of a closed, bounded interval.

2.4.19 Theorem (Extreme value). *Let $z_0 \in \mathbb{C}$ and $r > 0$, and suppose that $f \colon \overline{\mathcal{B}}(z_0; r) \to \mathbb{C}$ is continuous. Then there are $z_{\min}$, $z_{\max} \in \overline{\mathcal{B}}(z_0; r)$ such that*

$$|f(z_{\min})| \le |f(z)| \le |f(z_{\max})| \text{ for all } z \in \overline{\mathcal{B}}(z_0; r).$$

We will not prove this theorem, as it relies on some deeper topological machinery than we care to develop. We mention that it holds for a much broader class of subsets of $\mathbb{C}$ than just closed balls, but we will never need a richer version than this.

2.4.20 Example. Fix $z_0 \in \mathbb{C}$ and $r > 0$. By continuity, there is $z_{\max} \in \overline{\mathcal{B}}(z_0; r)$ such that $|e^z| \le |e^{z_{\max}}|$ for all $z \in \overline{\mathcal{B}}(z_0; r)$. However, we can be more precise than this existential result. First, $|e^z| = e^{\operatorname{Re}(z)}$. So, what point in $\overline{\mathcal{B}}(z_0; r)$ has the largest real part? If we draw a picture, hopefully we see that this point is $z_0 + r$; this could be verified more precisely, of course. Thus the maximum of f on $\overline{\mathcal{B}}(z_0; r)$ is $e^{\operatorname{Re}(z_0) + r}$.

2.4.21 Problem (!). Prove the claim in the previous example: if $z \in \overline{\mathcal{B}}(z_0; r)$, then $\operatorname{Re}(z) \le \operatorname{Re}(z_0) + r$.

This is where we finished on Wednesday, February 7, 2024.

2.5. Differentiation.

Most of the rest of this course will really study *differentiable* functions. Our immediate goal will be to see how differentiability on $\mathbb{C}$ *superficially* resembles differentiability on $\mathbb{R}$ in the sense that the formulas for the definition of the derivative and differentiation rules (e.g., product, quotient, chain) are exactly the same but the *true nature* of differentiable functions on $\mathbb{C}$ is vastly distinct from that of differentiable functions on $\mathbb{R}$. Later (and it will take some time to reach this), we will use the twin pillars of complex algebra (the fact that $i^2 = -1$) and complex geometry (the fact that limits move in a two-dimensional world) to see just how different complex derivatives are from what we have seen in the real-variable case.

2.5.1. The definition of the derivative.

We begin with the good news: we are not changing the definition of the derivative.

2.5.1 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$. A function $f : \mathcal{D} \to \mathbb{C}$ is* **DIFFERENTIABLE AT** *$a \in \mathcal{D}$ if a is an accumulation point of $\mathcal{D}$ and if the limit*

$$\lim_{z \to a} \frac{f(z) - f(a)}{z - a} \tag{2.5.1}$$

exists. If so, we call this limit the **DERIVATIVE OF** *f* **AT** *a and denote it by $f'(a)$. We say that f is* **DIFFERENTIABLE ON** *$\mathcal{D}$ if f is differentiable at each $a \in \mathcal{D}$. A differentiable function $f : \mathbb{C} \to \mathbb{C}$ is called* **ENTIRE**.

Of course, there is another limit formula for the derivative that we will use interchangeably with the original definition.

2.5.2 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ and $f : \mathcal{D} \to \mathbb{C}$. Suppose that $a \in \mathcal{D}$ is an accumulation point of $\mathcal{D}$. Then f is differentiable at a if and only if the limit*

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{2.5.2}$$

exists. If this limit exists, then it equals $f'(a)$ as defined by (2.5.1).

2.5.3 Problem (+). Why do you think this theorem is true? [Hint: *what symbolic similarities do you see between* (2.5.1) *and* (2.5.2)*?*] Prove that it is true, using the formal definition (Definition 2.2.7) of the limit.

2.5.4 Remark. *We have seen that the existence of limits depends greatly on the domain of the function under consideration. The derivative limits in* (2.5.1) *and* (2.5.2) *are not limits of f but rather of a "difference quotient" function constructed from f. What values of z and h are allowed in those limits? In the following, assume that $f : \mathcal{D} \to \mathbb{C}$ is differentiable at $a \in \mathbb{C}$.*

(i) *Define*

$$\phi : \mathcal{D} \setminus \{a\} \to \mathbb{C} : z \mapsto \frac{f(z) - f(a)}{z - a}.$$

Then f is differentiable at a if and only if $\lim_{z \to a} \phi(z)$ exists, in which case $f'(a) = \lim_{z \to a} \phi(z)$, of course. So, in the limit (2.5.1)*, we must allow z to be any element of*

$\mathcal{D} \setminus \{a\}$.

(ii) *Let*

$$\mathcal{Z}_a := \{h \in \mathbb{C} \setminus \{0\} \mid a + h \in \mathcal{D}\}. \tag{2.5.3}$$

We claim that 0 *is an accumulation point of* $\mathcal{Z}_a$*; see Problem 2.5.5. Put*

$$\varphi \colon \mathcal{Z}_a \to \mathbb{C} \colon h \mapsto \frac{f(a+h) - f(a)}{h}.$$

Since 0 *is an accumulation point of* $\mathcal{Z}_a$*, it makes sense to discuss* $\lim_{h \to 0} \varphi(h)$*, and* f *is differentiable at* a *if and only if this limit exists, in which case* $f'(a) = \lim_{h \to 0} \varphi(h)$*. So, in the limit* (2.5.2)*, we must allow* h *to be any element of* $\mathcal{Z}_a$*.*

2.5.5 Problem (!). With $\mathcal{Z}_a$ defined in (2.5.3), show that 0 is an accumulation point of $\mathcal{Z}_a$. [Hint: *since* a *is an accumulation point of* $\mathcal{D}$*, there is a sequence* (z_k) *in* $\mathcal{D} \setminus \{a\}$ *such that* $z_k \to a$*. What do you know about the sequence* $(z_k - a)$*?*]

2.5.6 Example. We show that exp is differentiable and $\exp' = \exp$; once again, the functional equation comes to the rescue. We want to manipulate the difference quotient

$$\frac{\exp(a+h) - \exp(a)}{h} = \exp(a) \left(\frac{\exp(h) - \exp(1)}{h} \right).$$

We claim that

$$\lim_{h \to 0} \frac{\exp(h) - 1}{h} = 1. \tag{2.5.4}$$

Assuming this to be true, we have

$$\lim_{h \to 0} \frac{\exp(a+h) - \exp(a)}{h} = \lim_{h \to 0} \exp(a) \left(\frac{\exp(h) - \exp(1)}{h} \right) = \exp(a) \lim_{h \to 0} \frac{\exp(h) - 1}{h}$$
$$= \exp(a).$$

Consequently, exp is differentiable, and $\exp'(a) = \exp(a)$.

2.5.7 Problem (+). Show that the limit (2.5.4) is true. [Hint: *use the definition of the exponential as a power series to compute, for* $h \in \mathbb{C} \setminus \{0\}$*,*

$$\frac{\exp(h) - 1}{h} = 1 + h \sum_{j=0}^{\infty} \frac{h^j}{(j+2)!}.$$

Resist the urge to say

$$\lim_{h \to 0} \sum_{j=0}^{\infty} \frac{h^j}{(j+2)!} = \sum_{j=0}^{\infty} \lim_{h \to 0} \frac{h^j}{(h+2)!} = 1.$$

While this turns out to be true, interchanging the limit in h and the series (which is really a limit as $j \to \infty$) requires justification. Instead, show the existence of $C > 0$ such that if $|h| \le 1$, then

$$\left|\sum_{j=0}^{\infty} \frac{h^j}{(j+2)!}\right| \le C.$$

Then use the squeeze theorem to conclude (2.5.4).]

2.5.8 Example. We claim that $f \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto \overline{z}$ is nowhere differentiable. Again we manipulate the difference quotient:

$$\frac{f(z+h) - f(z)}{h} = \frac{\overline{z+h} - \overline{z}}{h} = \frac{\overline{z} + \overline{h} - \overline{z}}{h} = \frac{\overline{h}}{h}.$$

Long ago in Example 2.2.18, we saw that $\lim_{h \to 0} \overline{h}/h$ does not exist, and so f cannot be differentiable at any point in $\mathbb{C}$. We will see other proofs of this fact later.

2.5.2. Local linearity.

A function is **LINEAR** if it is a first-degree polynomial, i.e., f is linear if $f(z) = az + b$ for some a, $b \in \mathbb{C}$. (Strictly speaking, such a function might better be called **AFFINE**, as this is *not* linear in the sense of linear algebra's linear transformations unless $b = 0$.) Linear functions are among the most transparent to handle algebraically and analytically.

Perhaps the next best thing to linearity is the local linearity of differentiable functions. A differentiable function is **LOCALLY LINEAR** in the sense that if f is differentiable at z_0, then we can expose the leading-order linear terms "in" f via the expansion

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \rho(z), \tag{2.5.5}$$

where we implicitly define

$$\rho(z) = f(z) - [f(z_0) + f'(z_0)(z - z_0)].$$

This "remainder" term vanishes very quickly as $z \to z_0$ in the sense that

$$\lim_{z \to z_0} \frac{\rho(z)}{z - z_0} = 0. \tag{2.5.6}$$

Since $\lim_{z \to z_0}(z - z_0) = 0$, when z is close to z_0, the numerator $\rho(z)$ in (2.5.6) must be *very* small—even smaller than the denominator—for the quotient in (2.5.6) to have a zero limit as $z \to z_0$. We could then rewrite the expansion (2.5.5) as

$$f(z) = f(z_0) + \underbrace{f'(z_0)(z - z_0)}_{\text{"small"}} + \underbrace{\widetilde{\rho}(z)(z - z_0)}_{\text{"smaller"}},$$

where

$$\widetilde{\rho}(z) := \begin{cases} \dfrac{\rho(z)}{z - z_0}, & z \ne z_0 \\ 0, & z = z_0 \end{cases} \qquad \text{and therefore} \qquad \lim_{z \to z_0} \widetilde{\rho}(z) = 0$$

to isolate the leading-order behavior of f near z_0 precisely.

2.5.9 Problem (!). Use the definition of the derivative to prove (2.5.6).

A LOCAL LINEAR APPROXIMATION to f at z_0 should just be a linear function that takes the value $f(z_0)$ at z_0, i.e., a function of the form $\ell_a(z) := f(z_0) + a(z - z_0)$ for some $a \in \mathbb{C}$. A function's derivative is its best local linear approximation in the sense that $\ell_{f'(z_0)}$ will always "beat" any other ℓ_a as an approximation to f for z sufficiently close to z_0 (where "sufficiently close" depends on a).

2.5.10 Theorem. *Suppose that $f\colon \mathcal{D} \to \mathbb{C}$ is differentiable at $z_0 \in \mathcal{D}$ and let $a \in \mathbb{C}$. Then there is $\delta > 0$ such that*

$$|f(z) - [f(z_0) + f'(z_0)(z - z_0)]| < |f(z) - [f(z_0) + A(z - z_0)]| \tag{2.5.7}$$

whenever $|z - z_0| < \delta$.

Proof. Certainly this is true if $f'(z_0) = A$, so suppose $f'(z_0) \neq A$. We compute

$$f(z)-[f(z_0)+f'(z_0)(z-z_0)] = \rho(z) \quad \text{and} \quad f(z)-[f(z_0)+A(z-z_0)] = (f'(z_0)-A)(z-z_0)+\rho(z).$$

Then we have (2.5.7) if and only if

$$|\rho(z)| < |(f'(z_0) - A)(z - z_0) + \rho(z)|. \tag{2.5.8}$$

By the reverse triangle inequality,

$$|(f'(z_0) - A)(z - z_0) + \rho(z)| > |f'(z_0) - A||z - z_0| - |\rho(z)|,$$

and so (2.5.8) will follow if we show

$$|\rho(z)| < |f'(z_0) - A||z - z_0| - |\rho(z)|.$$

This is equivalent to

$$|\rho(z)| < \frac{|f'(z_0) - A|}{2}. \tag{2.5.9}$$

By (2.5.6), we may choose $\delta > 0$ such that if $|z - z_0| < \delta$, then (2.5.9) holds. ■

2.5.3. Fundamental properties of derivatives.

The recent good news was that the definition of the derivative, at the formulaic level, does not change for functions of a complex variable. The new good news is that neither do the "differentiation rules," mostly. Here is a familiar result.

2.5.11 Theorem (Differentiability implies continuity). *Suppose that $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ is differentiable at $a \in \mathcal{D}$. Then f is also continuous at a.*

Proof. This proof does not use anything special about complex numbers, but it is a good opportunity to revisit some definitions and properties of sequences. The goal is to show that $\lim_{z\to a} f(z) = f(a)$, so we take any sequence (z_k) in $\mathcal{D} \setminus \{a\}$ with $z_k \to a$, and we want to show $f(z_k) \to f(a)$. This is equivalent to showing $(f(z_k) - f(a)) \to 0$.

Now, the difference $f(z_k) - f(a)$ looks like the numerator of the difference quotient in the definition of $f'(a)$, and, since

$$\lim_{z\to a} \frac{f(z) - f(a)}{z - a} = f'(a),$$

and since $z_k \to a$ with $z_k \neq a$ for all k, we have

$$\lim_{k\to\infty} \frac{f(z_k) - f(a)}{z_k - a} = f'(a).$$

We can make the difference that we care about $(f(z_k) - f(a))$ look like the difference quotient that we understand by introducing $z_k - a$. So, we multiply by 1 and find

$$f(z_k) - f(a) = [f(z_k) - f(a)] \cdot 1 = [f(z_k) - f(a)] \cdot \frac{z_k - a}{z_k - a} = \left(\frac{f(z_k) - f(a)}{z_k - a}\right)(z_k - a).$$

Since

$$\lim_{k\to\infty} \frac{f(z_k) - f(a)}{z_k - a} = f'(a) \quad \text{and} \quad \lim_{k\to\infty} (z_k - a) = 0,$$

we have

$$\lim_{k\to\infty} \big(f(z_k) - f(a)\big) = \lim_{k\to\infty} \left(\frac{f(z_k) - f(a)}{z_k - a}\right)(z_k - a) = \left(\lim_{k\to\infty} \frac{f(z_k) - f(a)}{z_k - a}\right)\left(\lim_{k\to\infty} (z_k - a)\right)$$
$$= f'(a) \cdot 0 = 0. \qquad \blacksquare$$

Conversely, if f is not continuous at a, then f cannot be differentiable at a.

Algebraic properties of derivatives likewise carry over from the real world to the complex. While we cannot import proofs from corresponding results for sequences as with our prior proofs of limits, since there are no differentiation results for sequences, nonetheless the proofs of the following theorems are more or less identical to the real-variable case.

2.5.12 Theorem (Algebraic properties of derivatives). *Let f, $g\colon \mathcal{D} \to \mathbb{C}$ be differentiable at $a \in \mathcal{D}$.*

(i) [Linearity of derivatives] *$f + g$ is differentiable at a, and $(f + g)'(a) = f'(a) + g'(a)$.*

(ii) [Linearity of derivatives] *αf is differentiable at a, and $(\alpha f)'(a) = \alpha f'(a)$.*

(iii) [Product rule] *fg is differentiable at a, and $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$.*

2.5.13 Theorem (Chain rule). *Let $\mathcal{D}_1$, $\mathcal{D}_2 \subseteq \mathbb{C}$ and let $f\colon \mathcal{D}_1 \to \mathbb{C}$ be differentiable at $a \in \mathcal{D}_1$. Suppose also that $f(z) \in \mathcal{D}_2$ for all $z \in \mathcal{D}_1$. If $g\colon \mathcal{D}_2 \to \mathbb{C}$ is differentiable at $f(a)$,* *then $g \circ f\colon \mathcal{D}_1 \to \mathbb{C}$ is differentiable at a, and $(g \circ f)'(a) = g'(f(a))f'(a)$.*

2.5.14 Example. Since $\exp' = \exp$, and since the rules for derivatives work as they should, we obtain the familiar derivatives for the trigonometric functions. We consider $\cos(z) = (\exp(iz) + \exp(-iz))/2$.

First, here is why cos is entire. Use the definition of the derivative to show that $z \mapsto z$ is differentiable. Next, use algebraic properties of the derivative to show that $z \mapsto \pm iz$ is differentiable. Third, use the chain rule and the differentiability of exp to show that $z \mapsto \exp(\pm iz)$ is differentiable. Last, use more algebraic properties of the derivative to show that $z \mapsto (\exp(iz) + \exp(-iz))/2$ is differentiable. Thus cos is differentiable.

Now we actually compute the derivative:

$$\cos'(z) = \frac{i\exp(iz) - i\exp(-iz)}{2} = i\left(\frac{\exp(iz) - \exp(-iz)}{2}\right) = i^2\left(\frac{\exp(iz) - \exp(-iz)}{2i}\right)$$
$$= -\left(\frac{\exp(iz) - \exp(-iz)}{2i}\right) = -\sin(z).$$

2.5.15 Problem (!). Check that $\sin'(z) = \cos(z)$ for all $z \in \mathbb{C}$.

2.5.4. The reverse chain rule.

How can we differentiate something for which we have less pleasant a formula than algebraic or exponential or trigonometric functions—something like a logarithm? We know that $\exp(\mathrm{Log}(z)) = z$ for all $z \in \mathbb{C} \setminus \{0\}$, so if Log is differentiable, then the chain rule leads us to expect that

$$1 = \exp'(\mathrm{Log}(z))\,\mathrm{Log}'(z) = \exp(\mathrm{Log}(z))\,\mathrm{Log}'(z) = z\,\mathrm{Log}'(z), \tag{2.5.10}$$

and therefore $\mathrm{Log}'(z) = 1/z$, as usual. But why should Log be differentiable in the first place? The answer lies in a deeper examination of the composition properties of the logarithm.

We begin with a lemma about difference quotients that will serve us well both here and in various future appearances.

2.5.16 Lemma (Difference quotient). *Let $f\colon \mathcal{D} \to \mathbb{C}$ be differentiable. Fix $a \in \mathcal{D}$ and define*

$$\phi\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto \begin{cases} \dfrac{f(z) - f(a)}{z - a}, & z \in \mathcal{D} \setminus \{a\} \\ f'(a), & z = a. \end{cases}$$

Then ϕ is differentiable on $\mathcal{D} \setminus \{a\}$ and continuous on $\mathcal{D}$.

Proof. Continuity on $\mathcal{D} \setminus \{a\}$ will follow from differentiability on $\mathcal{D} \setminus \{a\}$. Continuity at a

follows from the calculation

$$\lim_{z\to a}\phi(z)=\lim_{z\to a}\frac{f(z)-f(a)}{z-a}=f'(a)=\phi(a).$$

Now for the differentiability on $\mathcal{D}\setminus\{a\}$: this is essentially the quotient rule. The map $z\mapsto f(z)-f(a)$ is differentiable on $\mathcal{D}$ as f is differentiable on $\mathcal{D}$ and $f(a)$ is constant; the map $z\mapsto z-a$ is differentiable on $\mathbb{C}$, so the quotient $z\mapsto \big(f(z)-f(a)\big)/(z-a)$ is differentiable as long as the denominator is not zero, i.e., on $\mathcal{D}\setminus\{a\}$. ■

Now we prove a theorem that is a kind of "reverse" of the chain rule. Nothing in this theorem requires the independent variable to be complex or real, and this proof could have been done just as well in a real analysis class. But we think it is a good illustration of how the difference quotient behaves, and we will use difference quotients in several key places in the future.

2.5.17 Theorem (Reverse chain rule). *Let $\mathcal{D}_1,\mathcal{D}_2\subseteq\mathbb{C}$ with $\mathcal{D}_1\subseteq\mathcal{D}_2$. Let $f\colon\mathcal{D}_1\to\mathcal{D}_2$ be continuous and let $g\colon\mathcal{D}_2\to\mathbb{C}$ be differentiable. Suppose that $g(f(z))=z$ for all $z\in\mathcal{D}_1$ and $g'(f(z))\neq 0$ for all $z\in\mathcal{D}_1$. Then f is differentiable on $\mathcal{D}_1$ and*

$$f'(z)=\frac{1}{g'(f(z))}.$$

Proof. First, if we also know that f is differentiable, then the formula for f' follows from the chain rule as usual. Indeed, since $g(f(z))=z$, we differentiate both sides to find $g'(f(z))f'(z)=1$, and then we solve for $f'(z)$. But here we do not know that f is differentiable, so we have work to do.

Fix $a\in\mathcal{D}_1$. We need to show that

$$\lim_{z\to a}\frac{f(z)-f(a)}{z-a}=\frac{1}{g'(f(z))}.$$

The hypothesis $g(f(z))=z$ for all z lets us rewrite the difference quotient as

$$\frac{f(z)-f(a)}{z-a}=\frac{f(z)-f(a)}{g(f(z))-g(f(a))},\quad z\neq a. \tag{2.5.11}$$

Observe that $f(z)-f(a)\neq 0$ for all $z\in\mathcal{D}_1\setminus\{a\}$. Indeed, if $f(z)=f(a)$, then

$$z=g(f(z))=g(f(a))=a. \tag{2.5.12}$$

So, we can use (2.5.11) to say

$$\frac{f(z)-f(a)}{z-a}=\frac{1}{\dfrac{g(f(z))-g(f(a))}{f(z)-f(a)}}. \tag{2.5.13}$$

The denominator now has the form of the difference quotient function from Lemma 2.5.16. Specifically, if we put

$$\phi \colon \mathcal{D}_2 \to \mathbb{C} \colon w \mapsto \begin{cases} \dfrac{g(w)-g(f(a))}{w-f(a)}, \ w \in \mathcal{D}_2 \setminus \{f(a)\} \\ \\ g'(f(a)), \ w = f(a), \end{cases}$$

then (2.5.13) becomes

$$\frac{f(z)-f(a)}{z-a} = \frac{1}{\phi(f(z))}.$$

The reasoning in (2.5.12) ensures that $\phi(f(z)) \neq 0$ for all $z \neq a$. Continuity of the difference quotient function and ϕ then implies

$$\lim_{z \to a} \frac{f(z)-f(a)}{z-a} = \lim_{z \to a} \frac{1}{\phi(f(z))} = \lim_{w \to f(a)} \frac{1}{\phi(w)} = \frac{1}{\phi(f(a))} = \frac{1}{g'(f(a))}. \qquad \blacksquare$$

Now we can differentiate, among other things, the logarithm and in particular justify the formal calculations in (2.5.10).

2.5.18 Example. Define

$$f \colon \mathbb{C} \setminus \{0\} \to \mathbb{C} \colon z \mapsto \operatorname{Log}(z) \quad \text{and} \quad g \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto \exp(z).$$

Then $g(f(z)) = z$ for all $z \in \mathbb{C} \setminus \{0\}$, f is continuous, and g is differentiable. Moreover,

$$g'(f(z)) = \exp'(\operatorname{Log}(z)) = \exp(\operatorname{Log}(z)) = z \neq 0 \text{ for } z \in \mathbb{C} \setminus \{0\}.$$

The reverse chain rule therefore grants our desire:

$$\operatorname{Log}'(z) = f'(z) = \frac{1}{g'(f(z))} = \frac{1}{z}.$$

2.5.19 Problem (!). Where is $\log_\alpha$ differentiable and what is $\log_\alpha'$?

2.5.20 Problem ($\star$). Let $n \geq 1$ be a positive integer and let $\mathcal{D} \subseteq \mathbb{C} \setminus \{0\}$. A BRANCH OF THE nTH ROOT IN $\mathcal{D}$ is a function $f \colon \mathcal{D} \to \mathbb{C}$ such that $[f(z)]^n = z$ for all $z \in \mathcal{D}$. (For example, if we recall the definition $z^{1/2} = e^{(1/2)\log(z)}$, it follows that $f(z) = e^{\operatorname{Log}(z)/2}$ is a branch of the square root—"second root" just sounds wrong—in $\mathcal{D} = \mathbb{C} \setminus (-\infty, 0]$.) Use the "reverse chain rule" to show that if $f \colon \mathcal{D} \to \mathbb{C}$ is a continuous branch of the nth root in $\mathcal{D} \subseteq \mathbb{C} \setminus \{0\}$, then f is differentiable on $\mathcal{D}$ with

$$f'(z) = \frac{f(z)}{nz}$$

for all $z \in \mathcal{D}$. Is this what you expected from the power rule in real-variable calculus?

2.5.5. The derivative of a function of a real variable.

The derivative is fundamentally a limit, and we know that limits interact well with real and imaginary parts. What does this tell us in the context of derivatives? Not much, unfortunately.

Suppose that $f\colon \mathcal{D} \to \mathbb{C}$ is differentiable at $a \in \mathcal{D}$. Are **??**f] and $\text{Im}[f]$ differentiable at a? And can we recover their derivatives at a from the value $f'(a)$?

We think just about the real part here. Recall from (2.1.1) that $\text{Re}[f]$ is the function

$$\text{Re}[f]\colon \mathcal{D} \to \mathbb{R}\colon z \mapsto \text{Re}[f(z)].$$

If $\text{Re}[f]$ is differentiable at a, then the limit

$$\text{Re}[f]'(a) = \lim_{h\to 0} \frac{\text{Re}[f](a+h) - \text{Re}[f](a)}{h} \tag{2.5.14}$$

must exist, and we would probably like this limit to equal $\text{Re}[f'(a)]$. If so, then we would have $\text{Re}[f]'(a) = \text{Re}[f'(a)]$. Of course, by definition of $\text{Re}[f]$, this is equivalent to the existence of the limit

$$\text{Re}[f]'(a) = \lim_{h\to 0} \frac{\text{Re}[f(a+h) - f(a)]}{h} \tag{2.5.15}$$

So, does (2.5.14), equivalently, (2.5.15), exist? We know that the limit

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

exists, and we know that real parts and limits commute:

$$\text{Re}[f'(a)] = \text{Re}\left[\lim_{h\to 0} \frac{f(a+h) - f(a)}{h}\right] = \lim_{h\to 0} \text{Re}\left[\frac{f(a+h) - f(a)}{h}\right]. \tag{2.5.16}$$

If

$$\text{Re}\left[\frac{f(a+h) - f(a)}{h}\right] = \frac{\text{Re}[f(a+h) - f(a)]}{h} \tag{2.5.17}$$

then the limit in (2.5.15) exists, and we can equate (2.5.15) and (2.5.16) to conclude $\text{Re}[f]'(a) = \text{Re}[f'(a)]$. But there is no reason to expect that (2.5.17) is true when h is complex and nonreal, and if the domain $\mathcal{D}$ of f contains complex, nonreal numbers, then we can expect (Remark 2.5.4) that we must consider complex, nonreal h in (**??**).

In that case, the definition of complex division will certainly cause an interaction between h and the numerator that will alter the real parts of everything. We will soon see that there is a powerful connection between the differentiability of a function of a complex variable and the *partial* derivatives of its real and imaginary parts, when they are viewed as functions of two *real* variables. For now, here is a much tamer result when the independent variable is real.

2.5.21 Theorem. *Let $I \subseteq \mathbb{R}$ be an interval and let $a \in I$. Let $f\colon I \to \mathbb{C}$ be a function and put $u := \text{Re}[f]$ and $v := \text{Im}[f]$. Then f is differentiable at a if and only if both u and v are differentiable at a in the sense of Definition 2.5.1 (and consequently also in the sense* *of Theorem 2.5.2). That is, f is differentiable at a if and only if both the limits*

$$u'(a) = \lim_{t\to a} \frac{u(t)-u(a)}{t-a} \quad \text{and} \quad v'(a) = \lim_{t\to a} \frac{v(t)-v(a)}{t-a}$$

exist (with the limits taken as one-sided as in Definition 2.2.25 if a is an endpoint of I). In this case,

$$f'(a) = u'(a) + iv'(a).$$

2.5.22 Problem ($\star$). Here is the proof of this theorem. Let a, $b \in \mathbb{R}$ with $a \leq b$ and let $f\colon [a,b] \to \mathbb{C}$. Using only Definition 2.5.1 and results from Section 2.2.4, prove the following.

(i) f is differentiable at a if and only if the limit

$$L_+ := \lim_{\tau\to a^+} \frac{f(\tau)-f(a)}{\tau-a}$$

exists, in which case $L_+ = f'(a)$.

(ii) f is differentiable at b if and only if the limit

$$L_- := \lim_{\tau\to b^-} \frac{f(\tau)-f(b)}{\tau-b}$$

exists, in which case $L_- = f'(b)$.

(iii) f is differentiable at $t \in (a,b)$ if and only if the limits

$$L_+ := \lim_{\tau\to t^+} \frac{f(\tau)-f(t)}{\tau-t} \quad \text{and} \quad L_- := \lim_{\tau\to t^-} \frac{f(\tau)-f(t)}{\tau-t}$$

exist and are equal, in which case $f'(t) = L_+ = L_-$.

This is where we finished on Friday, February 9, 2024.

2.5.23 Example. The function $f\colon \mathbb{R} \to \mathbb{C}\colon t \mapsto e^{it}$ is differentiable, and by the chain rule $f'(t) = ie^{it}$. We also have $f(t) = u(t) + iv(t)$ with $u(t) = \cos(t)$ and $v(t) = \sin(t)$. Since $u'(t) = -\sin(t)$ and $v'(t) = \cos(t)$, we have

$$u'(t) + iv(t) = -\sin(t) + i\cos(t) = i^2\sin(t) + i\cos(t) = i\big(\cos(t) + i\sin(t)\big) = ie^{it} = f'(t),$$

as expected.

2.6. The Cauchy–Riemann equations.

Now we take answer the question left dangling at the start of Section 2.5.5: what is the relationship between a function's derivative and the (partial?) derivatives of its real and imaginary parts?

2.6.1. Some formal analysis.

The situation is the same as at the start of Section 2.5.5: we have a function $f \colon \mathcal{D} \to \mathbb{C}$ that is differentiable at $a \in \mathcal{D}$, and we want to learn about $\operatorname{Re}[f]$. The whole problem there was that, typically,

$$\operatorname{Re}\left[\frac{f(a+h)-f(a)}{h}\right] \neq \frac{\operatorname{Re}[f(a+h)-f(a)]}{h}$$

for h complex and nonreal (and nonzero, of course).

But what if we just took h to be real? More precisely, we know that

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

and so

$$f'(a) = \lim_{k \to \infty} \frac{f(a+h_k)-f(a)}{h_k} \tag{2.6.1}$$

for any sequence (h_k) such that $a + h_k \in \mathcal{D}$ and $h_k \neq 0$. Suppose that we take each h_k to be real. Then, repeating our work from the start of Section 2.5.5, we find

$$\operatorname{Re}[f'(a)] = \lim_{k \to \infty} \frac{\operatorname{Re}[f(a+h_k)-f(a)]}{h_k}.$$

What does this limit mean?

Rewrite

$$f(x+iy) = u(x,y) + iv(x,y), \qquad u(x,y) := \operatorname{Re}[f(x+iy)] \quad \text{and} \quad v(x,y) := \operatorname{Im}[f(x+iy)].$$

Here we are thinking of u and v as real-valued functions on the set $\widetilde{\mathcal{D}} := \{(x,y) \in \mathbb{R}^2 \mid x+iy \in \mathcal{D}\}$. Also, suppose $a = x_0 + iy_0$. Then for h_k real, we have

$$f(a+h_k) = f((x_0+h_k)+iy_0) = u(x_0+h_k, y_0) + iv(x_0+h_k, y_0) \quad \text{and} \quad f(a) = u(x_0,y_0) + iv(x_0,y_0),$$

so

$$\frac{\operatorname{Re}[f(a+h_k)-f(a)]}{h_k} = \frac{u(x_0+h_k, y_0) - u(x_0,y_0)}{h_k}.$$

Thus for any real sequence (h_k) such that $h_k \to 0$ with $h_k \neq 0$ and $(x_0+h_k)+iy_0 \in \mathcal{D}$, we have

$$\operatorname{Re}[f'(a)] = \lim_{k \to \infty} \frac{u(x_0+h_k, y_0) - u(x_0,y_0)}{h_k}.$$

That is, the limit

$$\lim_{h \to 0} \frac{u(x_0+h, y_0) - u(x_0,y_0)}{h}$$

exists, where here h is understood to be real. This limit is the definition of the partial derivative $u_x(x_0, y_0)$. We conclude

$$\text{Re}[f'(x_0 + iy_0)] = u_x(x_0, y_0).$$

This tells us about the differentiability of u in relation to the differentiability of f—exactly the information that we were seeking, if not the information that we were expecting. In particular, we have not said anything about whether the limit

$$\lim_{h \to 0} \frac{\text{Re}[f(a+h) - f(a)]}{h}$$

exists with h allowed to be complex and nonreal, and so we cannot say anything about whether $\text{Re}[f]$ is differentiable.

If we replace Re with Im in all of the calculations above, we can obtain

$$\text{Im}[f'(x_0 + iy_0)] = v_x(x_0, y_0).$$

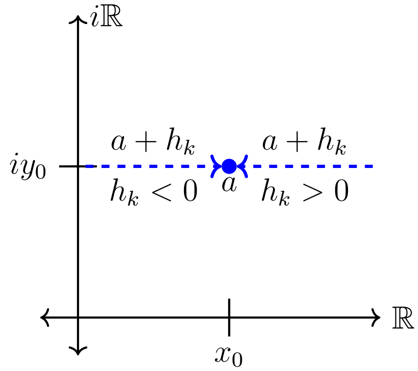
2.6.1 Problem (!). Do that.

We conclude

$$f'(x_0 + iy_0) = u_x(x_0, y_0) + iv_x(x_0, y_0). \tag{2.6.2}$$

This is a relationship between the differentiability of f and the differentiability of its real and imaginary parts—not $\text{Re}[f]$ and $\text{Im}[f]$ as functions of the complex variable z but u and v as functions of the real variables x and y.

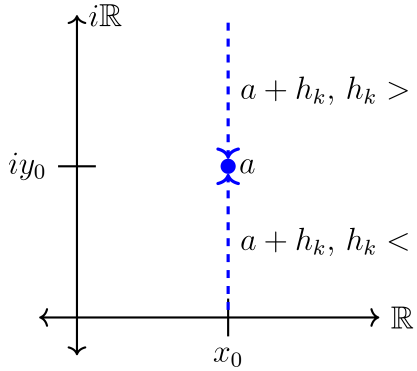
All of the work above was analytic in the sense that we worked with definitions, formulas, and sequences. Here is the geometric perspective: we approached $x_0 + iy_0$ "in the real direction" by adding small real numbers h_k.



In the process of approaching $x_0 + iy_0$ "in the real direction," we learned about the x-partials of u and v. Since the y-"slot" of u and v corresponds to the imaginary component of the input to f, we might also try approaching $x_0 + iy_0$ "in the imaginary direction." We still take (h_k) to be a sequence of real numbers with $h_k \neq 0$ and $h_k \to 0$, but now we assume that $a + ih_k \in \mathcal{D}$. Then

$$f'(a) = \lim_{k \to \infty} \frac{f(a + ih_k) - f(a)}{ih_k}. \tag{2.6.3}$$

Here is the picture.



Taking real parts, we have

$$\operatorname{Re}[f'(a)] = \lim_{k \to \infty} \operatorname{Re}\left[\frac{f(a + ih_k) - f(a)}{ih_k}\right]. \tag{2.6.4}$$

The simplification of the real part is a little more complicated now, but only a little, since, for $w \in \mathbb{C}$ and $h \in \mathbb{R} \setminus \{0\}$, we have

$$\operatorname{Re}\left(\frac{w}{ih}\right) = \frac{\operatorname{Im}(w)}{h} \quad \text{and} \quad \operatorname{Im}\left(\frac{w}{ih}\right) = -\frac{\operatorname{Re}(w)}{h}.$$

2.6.2 Problem (!). Check this.

Thus

$$\operatorname{Re}\left[\frac{f(a + ih_k) - f(a)}{ih_k}\right] = \frac{v(x_0, y_0 + h_k) - v(x_0, y_0)}{h_k},$$

so we conclude

$$\operatorname{Re}[f'(a)] = \lim_{k \to \infty} \frac{v(x_0, y_0 + h_k) - v(x_0, y_0)}{h_k}.$$

This limit holds for any real sequence (h_k) such that $h_k \to 0$ with $h_k \neq 0$ and $x_0 + i(y_0 + h_k) \in \mathcal{D}$, and so the limit

$$\lim_{h \to 0} \frac{v(x_0, y_0 + h) - v(x_0, y_0)}{h}$$

exists. This limit is the definition of the partial derivative $v_y(x_0, y_0)$. We conclude

$$\operatorname{Re}[f'(x_0 + iy_0)] = v_y(x_0, y_0).$$

Previously we calculated this and found $\operatorname{Re}[f'(x_0 + iy_0)] = v_y(x_0, y_0)$. This gives even more information than we had before: a relationship between u_x and v_y. Namely, if $f(x, y) = u(x, y) + iv(x, y)$ is differentiable at $x_0 + iy_0$, it appears that

$$u_x(x_0, y_0) = v_y(x_0, y_0). \tag{2.6.5}$$

2.6.3 Problem (!). Starting with the limit (2.6.4) in the work above, swap the roles of Re and Im to conclude

$$u_y(x_0, y_0) = -v_x(x_0, y_0) \tag{2.6.6}$$

and also

$$f'(x_0 + iy_0) = v_y(x_0, y_0) - iu_y(x_0, y_0). \tag{2.6.7}$$

Together, the partial differential equations (2.6.5) and (2.6.6)—which together we call the **Cauchy–Riemann equations**—and the identities (2.6.2) and (2.6.7) relate the derivative behavior of f to the derivative behavior of its real and imaginary parts. This is what we failed to achieve at the start of Section 2.5.5, where we had to restrict the function's domain to be real. (Note that here $\mathcal{D}$ cannot consist solely of real numbers if we want to approach $x_0 + iy_0$ "in the imaginary direction.")

Here is what we seem to have proved.

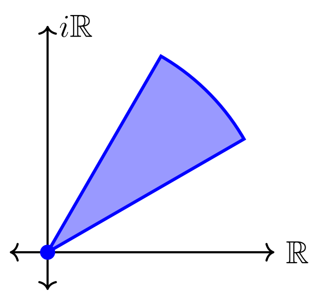
2.6.4 Theorem (Cauchy–Riemann, wrong version). *Let $f \colon \mathcal{D} \to \mathbb{C}$ be differentiable at $x_0 + iy_0 \in \mathcal{D}$. Write $u(x, y) = \mathrm{Re}[f(x + iy)]$ and $v(x, y) = \mathrm{Im}[f(x + iy)]$. Then the partial derivatives $u_x(x_0, y_0)$, $u_y(x_0, y_0)$, $v_x(x_0, y_0)$, and $v_y(x_0, y_0)$ all exist and satisfy*

$$\begin{cases} u_x(x_0, y_0) = v_y(x_0, y_0) \\ u_y(x_0, y_0) = -v_x(x_0, y_0). \end{cases}$$

Moreover,

$$f'(x_0 + iy_0) = u_x(x_0, y_0) + iv_x(x_0, y_0) = v_y(x_0, y_0) - iu_y(x_0, y_0).$$

But here is what is wrong about this: everything hinged on being able to approach $x_0 + iy_0$ "in the real and imaginary directions" while remaining in $\mathcal{D}$. That is, we assumed that we could write $f'(a)$ in the special forms (2.6.1) and (2.6.3), and this hinged on having real sequences (h_k) with $h_k \to 0$, $h_k \neq 0$, and $a + h_k$, $a + ih_k \in \mathcal{D}$. Why should such sequences exist? For example, if $\mathcal{D}$ is the sector below and $a = 0$, then no such sequences exist, because the real and imaginary axes do not intersect this sector except at 0.



The lesson is that for Theorem 2.6.4 to be true, we need some stronger geometry on $\mathcal{D}$. We now develop that.

2.6.2. Open sets.

The existence of a limit at a point hinges on the consistent behavior of a function at that point, regardless of the direction of approach. The existence of a limit also presumes that some direction of approach to that point is possible from elsewhere in the function's domain. Frequently we have used the possibility of multiple directions of approach to break a limit; conversely, restricting the directions of approach by restricting the function to a subset of its domain may artificially show that the *restricted* function has a limit on its *restricted* domain—see Problem 2.4.7 for the strange observation that $\lim_{t\to-1} \operatorname{Arg}\big|_{(-\infty,0)}(t)$ exists, although we know well that $\lim_{z\to-1} \operatorname{Arg}(z)$ does not.

Limits are at their strongest when we can approach the point in question not from one or two directions, not just from the left or the right, but from *every* possible direction. If the function has consistent behavior on every avenue of approach to the point, then the function's behavior near that point is very, very well-behaved indeed. This presumes that the function under consideration is defined on a ball centered at the point in question, and this suggests that we work with functions defined not on arbitrary subsets of $\mathbb{C}$ but on the following special kind.

2.6.5 Definition. *A set $\mathcal{D} \subseteq \mathbb{C}$ is* **OPEN** *if for each $z \in \mathcal{D}$, there is $r > 0$ such that $\mathcal{B}(z; r) \subseteq \mathcal{D}$.*

This is where we finished on Monday, February 12, 2024.

2.6.6 Example. **(i)** The whole complex plane is open. Indeed, given $z \in \mathbb{C}$, take $r > 0$ to be any positive number—say, $r = 1$. Then, certainly, $\mathcal{B}(z; 1) \subseteq \mathbb{C}$.

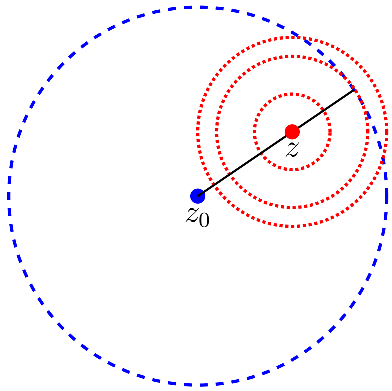
(ii) Open balls are open—it would be a horrible misnomer if they were not. To see this, fix $z_0 \in \mathbb{C}$ and $r > 0$. Take $z \in \mathcal{B}(z_0; r)$. We want to find $s > 0$ such that $\mathcal{B}(z; s) \subseteq \mathcal{B}(z_0; r)$. We want s to satisfy

$$w \in \mathcal{B}(z; s) \Longrightarrow w \in \mathcal{B}(z_0; r),$$

equivalently,

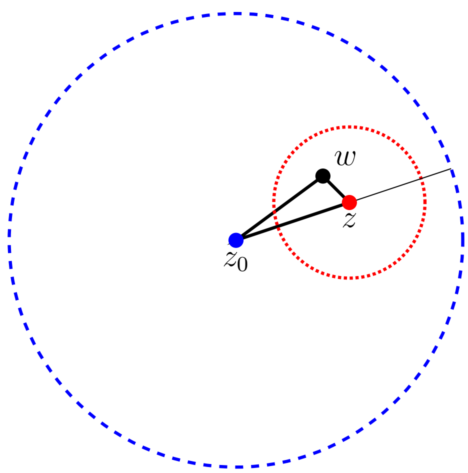
$$|w - z| < s \Longrightarrow |w - z_0| < r.$$

Here is a picture of several balls $\mathcal{B}(z; s)$ drawn for different values of s.



The picture suggests that if we take the radius of the red ball—which is s—to be no larger than the distance between z and the boundary of the blue ball—which is $r - |z - z_0|$, and which is positive since $|z - z_0| < r$—then everything will work. However, if we take the radius any larger, then the red ball would go outside the blue ball. So, we take $s = r - |z - z_0|$.

Suppose that $w \in \mathcal{B}(z; s)$. Here is another picture.



We know that $|w - z| < s = r - |z - z_0|$, and we want to show $|w - z_0| < r$. We know more about the distances $|w - z|$ and $|z - z_0|$ than we do about $|w - z_0|$, so we make those two distances show up by adding zero and using the triangle inequality:

$$|w - z_0| = |w - z + z - z_0| \le |w - z| + |z - z_0| < s + |z - z_0| = r - |z - z_0| + |z - z_0| = r.$$

2.6.7 Problem (!). Prove that punctured balls are open. [Hint: *drawing a picture may help you "avoid" the punctured center.*]

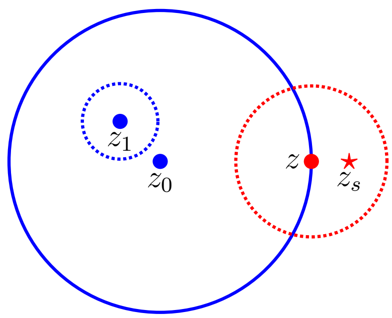
2.6.8 Problem (+). **(i)** Generalize Problem 2.6.7 to show that if $\mathcal{D} \subseteq \mathbb{C}$ is open and $z \in \mathcal{D}$, then $\mathcal{D} \setminus \{z\}$ is still open.

(ii) Show that if $\mathcal{D}_1, \mathcal{D}_2 \subseteq \mathbb{C}$ are open, then so is $\mathcal{D}_1 \cup \mathcal{D}_2$.

2.6.9 Problem ($\star$). Show that $\mathbb{C} \setminus i\mathbb{R}$ is open.

2.6.10 Example. Closed balls, as the name should suggest, are not open. Fix $z_0 \in \mathbb{C}$ and $r > 0$. The closed ball $\overline{\mathcal{B}}(z_0; r)$ is open if and only if for all $z \in \overline{\mathcal{B}}(z_0; r)$, there is $s > 0$ such that $\mathcal{B}(z; s) \subseteq \overline{\mathcal{B}}(z_0; r)$. So, the closed ball $\overline{\mathcal{B}}(z_0; r)$ is not open if and only if there exists $z \in \overline{\mathcal{B}}(z_0; r)$ such that for all $s > 0$ it is the case that $\mathcal{B}(z; s) \not\subseteq \overline{\mathcal{B}}(z_0; r)$. That is, we need to find some $z \in \overline{\mathcal{B}}(z_0; r)$ such that for all $s > 0$ there is $w \in \mathcal{B}(z; s)$ with $w \notin \overline{\mathcal{B}}(z_0; r)$. (Got all that?)

Drawing a picture helps: a point "inside" the ball, i.e., in $\mathcal{B}(z_0; r)$ is not problematic. Indeed, if $z_1 \in \mathcal{B}(z_0; r)$, then since open balls are open, there is $s > 0$ such that $\mathcal{B}(z_1; s) \subseteq \mathcal{B}(z_0; r) \subseteq \overline{\mathcal{B}}(z_0; r)$. So, we need to work with points $z \in \overline{\mathcal{B}}(z_0; r) \setminus \mathcal{B}(z_0; r) = \partial\mathcal{B}(z_0; r)$. For simplicity, try $z = z_0 + r$. Then the picture below leads us to expect that $\mathcal{B}(z_0 + r; s) \not\subseteq \overline{\mathcal{B}}(z_0; r)$ for any $s > 0$.



Figuring out a precise $w_s \in \mathcal{B}(z_0 + r; s) \setminus \overline{\mathcal{B}}(z_0; r)$ for each $s > 0$ is a good exercise.

2.6.11 Problem (!). Figure that out.

2.6.12 Problem (!). Let $I \subseteq \mathbb{R}$ be nonempty. Prove that I is not open. Conclude that open intervals in $\mathbb{R}$ are not open in the sense of Definition 2.6.5.

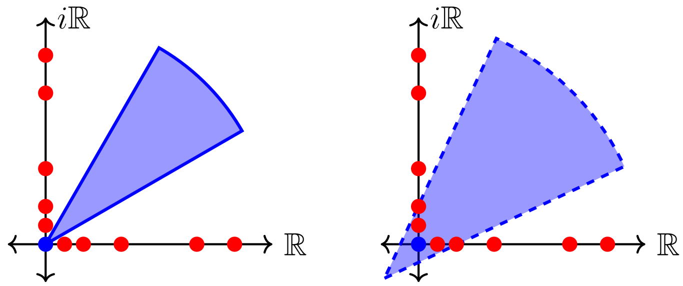
This is where we finished on Wednesday, February 14, 2024.

2.6.13 Example. Here is how open sets permit the "freedom of approach" that we so desperately need for our "proof" of the Cauchy–Riemann equations in Section 2.6.1 to work out. Let $\mathcal{D} \subseteq \mathbb{C}$ be open and $z_0 \in \mathcal{D}$. We want to show that if (h_k) is a sequence in $\mathbb{R}$ with $h_k \to 0$, then for k sufficiently large we have $z_0 + h_k \in \mathcal{D}$ and $z_0 + ih_k \in \mathcal{D}$.

Here is the key idea: since $h_k \to 0$, we have $z_0 + h_k \to z_0$ and $z_0 + ih_k \to z_0$. Theorem 1.3.7

tells us that for any $r > 0$, if k is sufficiently large, we will have $z_0 + h_k$, $z_0 + ih_k \in \mathcal{B}(z_0; r)$. So, we just take $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$. Then there is $N \in \mathbb{N}$ such that if $k \geq N$, we have $z_0 + h_k$, $z_0 + ih_k \in \mathcal{B}(z_0; r)$ and thus $z_0 + h_k$, $z_0 + ih_k \in \mathcal{D}$.

To see this in action, contrast how the real and imaginary sequences (drawn as dots) approach the origin in both pictures below; the set on the left is not open, and so those sequences never belong to it, unlike the set on the right.



2.6.3. The Cauchy–Riemann equations, done correctly.

We can correct Theorem 2.6.4 by adding the hypothesis that the domain of the function under consideration is open. Once we do that, all of our work in Section 2.6.1 becomes valid, and we actually get a little bit more.

2.6.14 Theorem (Cauchy–Riemann equations, correct version). *Suppose that $\mathcal{D} \subseteq \mathbb{C}$ is open and let $f: \mathcal{D} \to \mathbb{C}$ be a function. Write $f(x+iy) = u(x,y)+iv(x,y)$, where we think of u and v as being defined on the set $\widetilde{\mathcal{D}} := \{(x,y) \in \mathbb{R}^2 \mid x+iy \in \mathcal{D}\}$. In the following we write u_x, u_y, v_x, and v_y for the partial derivatives of u and v with respect to x and y.*

(i) *Suppose that f is differentiable at a point $z = x+iy \in \mathcal{D}$. Then the partial derivatives u_x, u_y, v_x, and v_y exist at (x,y) and satisfy the* CAUCHY–RIEMANN EQUATIONS

$$\begin{cases} u_x(x,y) = v_y(x,y) \\ u_y(x,y) = -v_x(x,y). \end{cases} \tag{2.6.8}$$

Moreover,

$$f'(x+iy) = u_x(x,y) + iv_x(x,y) = v_y(x,y) - iu_y(x,y). \tag{2.6.9}$$

(ii) *Let $x+iy \in \mathcal{D}$ and let $r > 0$ be such that $\mathcal{B}(x+iy; r) \subseteq \mathcal{D}$. Suppose that the four partial derivatives u_x, u_y, v_x, and v_y exist and are continuous on $\mathcal{B}(x+iy; r)$. Moreover, suppose that the partials satisfy the Cauchy–Riemann equations (2.6.8) at $x+iy$. Then f is differentiable at $x+iy$ and (2.6.9) holds.*

The proof of (ii) is cloistered in Appendix C.1, as it involves a technical estimate best justified with integrals and the fundamental theorem of calculus.

2.6.15 Problem (!). Use the formal work in Section 2.6.1 and Example 2.6.13 to prove part (i). [Hint: *there is very little new work to be done here. The goal is to show that the limits*

$$\lim_{h\to 0}\frac{u(x_0+h,y_0)-u(x_0,y_0)}{h},\qquad \lim_{h\to 0}\frac{v(x_0+h,y_0)-v(x_0,y_0)}{h},$$
$$\lim_{h\to 0}\frac{u(x_0,y_0+h)-u(x_0,y_0)}{h},\quad \textit{and}\quad \lim_{h\to 0}\frac{v(x_0,y_0+h)-v(x_0,y_0)}{h}$$

exist, where h is assumed to be real in each limit, and that the equalities of (2.6.8) *hold. To prove the existence of the limits, work with sequences. To show that $x_0+iy_0+h_k$ and $x_0+iy_0+ih_k$ belong to $\mathcal{D}$ for k large, when (h_k) is a sequence in $\mathbb{R}$ with $h_k\to 0$, use Example 2.6.13. Then copy and paste from Section 2.6.1. Once you know that the sequential limits work out for an arbitrary real sequence (h_k) with $h_k\to 0$, you know that the partial derivatives exist, right?*]

2.6.16 Remark. *A good mnemonic for remembering the Cauchy–Riemann equations is to look at the* **Jacobian matrix** *for $f(x+iy)=u(x,y)+iv(x,y)$, which is*

$$\begin{bmatrix} u_x & v_x \\ u_y & v_y \end{bmatrix}.$$

The diagonal entries are equal and the off-diagonal entries are negatives of each other.

2.6.17 Example. (i) Previously, in Example 2.5.8, we saw that the function $f\colon\mathbb{C}\to\mathbb{C}\colon z\mapsto\overline{z}$ was not differentiable at any point in $\mathbb{C}$. To show that, we had to use the definition of the derivative. Now we can use the Cauchy–Riemann equations. Write

$$f(x+iy)=\overline{x+iy}=x-iy,$$

so with

$$u(x,y):=x\quad\text{and}\quad v(x,y):=-y,$$

we have

$$f(x+iy)=u(x,y)+iv(x,y).$$

Now we differentiate:

$$u_x=1,\qquad u_y=0,\qquad v_x=0,\qquad\text{and}\qquad v_y=-1.$$

We clearly have $u_x\neq v_y$, and so the Cauchy–Riemann equations do not hold. It is the case, though, that $u_y=-v_x$.

(ii) Euler's formula gives

$$\exp(x+iy)=e^x\big(\cos(y)+i\sin(y)\big)=e^x\cos(y)+ie^x\sin(y),$$

and so with
$$u(x,y) := e^x \cos(y) \quad \text{and} \quad v(x,y) := e^x \sin(y),$$
we have
$$\exp(x+iy) = u(x,y) + iv(x,y).$$
We compute
$$u_x = e^x \cos(y), \qquad u_y = -e^x \sin(y), \qquad v_x = e^x \sin(y), \qquad \text{and} \qquad v_y = e^x \cos(y)$$
to see that $u_x = v_y$ and $u_y = -v_x$. So, the Cauchy–Riemann equations hold for exp on $\mathbb{C}$, and therefore exp is entire. Moreover,
$$\exp'(x+iy) = u_x(x,y) + iv_x(x,y) = e^x \cos(y) + ie^x \sin(y) = \exp(x+iy),$$
with the last equality being Euler's formula again.

This shows that if we only know calculus for real-valued exponentials, sines, and cosines on $\mathbb{R}$, then we get the expected results for the exponential on $\mathbb{C}$.

2.6.18 Problem (+). (Requires some linear algebra.) Let $\mathcal{D} \subseteq \mathbb{C}$ and put $\widetilde{\mathcal{D}} := \{(x,y) \in \mathbb{R}^2 \mid x+iy \in \mathcal{D}\}$. Let $f \colon \mathcal{D} \to \mathbb{C}$ and write $f(x+iy) = u(x,y) + iv(x,y)$, where $u(x,y) := \text{Re}[f(x+iy)]$ and $v(x,y) := \text{Im}[f(x+iy)]$. Then we obtain a function
$$\mathbf{f} \colon \widetilde{\mathcal{D}} \to \mathbb{R}^2 \colon (x,y) \mapsto \big(u(x,y), v(x,y)\big). \tag{2.6.10}$$
We can relate the differentiability of the complex-valued f (as a function of a complex variable) to the differentiability of the vector-valued $\mathbf{f}$ (as a function of two real variables) by defining an adequate notion of derivative for a function from (a subset of) $\mathbb{R}^2$ to $\mathbb{R}^2$.

Here is that notion. For $\mathbf{x} = (x,y) \in \mathbb{R}^2$, put $\|\mathbf{x}\| := \sqrt{x^2+y^2}$. For $\mathbf{g} \colon \widetilde{\mathcal{D}} \subseteq \mathbb{R}^2 \to \mathbb{R}$, say that $\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{g}(\mathbf{h}) = 0$ if for all $\epsilon > 0$, there exists $\delta > 0$ such that if $0 < \|\mathbf{h}\| < \delta$, then $\|\mathbf{g}(\mathbf{h})\| < \epsilon$. Then a function $\mathbf{f} \colon \widetilde{\mathcal{D}} \subseteq \mathbb{R}^2 \to \mathbb{R}^2$ is **DIFFERENTIABLE** at $\mathbf{x} \in \widetilde{\mathcal{D}}$ if there exists a matrix $A \in \mathbb{R}^{2\times 2}$ such that
$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) - A\mathbf{h}\|}{\|\mathbf{h}\|} = 0.$$
We call A the **DERIVATIVE** of $\mathbf{f}$ at $\mathbf{x}$, and we write $D\mathbf{f}(\mathbf{x}) := A$.

Show that with $\mathbf{f}$ defined in (2.6.10), f is differentiable at $z = x+iy \in \mathcal{D}$ (per Definition 2.5.1) if and only if $\mathbf{f}$ is differentiable at $\mathbf{x} = (x,y)$ and if A has the form
$$A = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$
If f is differentiable, what are α and β? (Optional part: how does this remind you of Problem B.2.13?)

The best results that we will find in this course are for differentiable functions defined on open sets. The Cauchy–Riemann equations are one such example. Because of the primacy of open sets for domains, we give differentiable functions defined on open sets a special name.

2.6.19 Definition. *A function $f\colon \mathcal{D} \to \mathbb{C}$ is* **HOLOMORPHIC** *if $\mathcal{D}$ is open and if f is differentiable on $\mathcal{D}$.*

2.6.4. The differential equation $f' = 0$.

The Cauchy–Riemann equations appear to reduce knowledge of a holomorphic function's derivative to knowledge of the partial derivatives of its real and imaginary parts—in other words, reducing a problem in complex analysis to a problem in real multivariable calculus. A function is particularly simple if either its real or imaginary part is identically zero, that is, if the function is strictly real-valued or strictly imaginary-valued.

Say that $f\colon \mathcal{D} \to \mathbb{C}$ is holomorphic and strictly real-valued, so $\text{Im}[f(z)] = 0$ for all $z \in \mathcal{D}$. Then if we write $f(x+iy) = u(x,y) + iv(x,y)$, we have $v(x,y) = 0$, and so immediately $v_x(x,y) = 0$. Moreover, the Cauchy–Riemann equations give

$$u_x(x,y) = v_y(x,y) = 0,$$

too, and so

$$f'(x+iy) = u_x(x,y) + iv_x(x,y) = 0.$$

2.6.20 Problem (!). Show that if $f\colon \mathcal{D} \to \mathbb{C}$ is holomorphic and $\text{Im}[f(z)] = 0$ for all $z \in \mathcal{D}$, then $f'(z) = 0$ for all $z \in \mathcal{D}$.

Our intuition with calculus probably leads us to believe that a function whose derivative is identically zero must be constant. That is, if $f'(z) = 0$ for all $z \in \mathcal{D}$, there must be $c \in \mathbb{C}$ such that $f(z) = c$ for all $z \in \mathcal{D}$. This intuition is wrong.

2.6.21 Example. Define

$$f\colon \mathbb{C} \setminus i\mathbb{R} \to \mathbb{C}\colon z \mapsto \begin{cases} -1, & \text{Re}(z) < 0 \\ 1, & \text{Re}(z) > 0. \end{cases}$$

Here we could use the definition of the derivative to see that f is differentiable on $\mathbb{C} \setminus i\mathbb{R}$ and $f'(z) = 0$ for all $z \in \mathbb{C} \setminus i\mathbb{R}$, and yet clearly f is not constant. Last, by Problem 2.6.9, the set $\mathbb{C} \setminus i\mathbb{R}$ is open, and so f is holomorphic.

The problem with the preceding example is that the domain of the function in question has a "gap" in it, namely, the imaginary axis. Our real-variable calculus intuition that if $f' = 0$, then f is constant hinges on having an interval for the domain—and intervals have no "gaps" in them. We can prove this easily if we accept the mean value theorem.

2.6.22 Theorem (Mean value). *Let a, $b \in \mathbb{R}$ with $a < b$ and let $f\colon [a,b] \to \mathbb{R}$ be* *continuous with f differentiable on (a, b). Then there is $\tau \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{b - a} = f'(\tau).$$

2.6.23 Corollary. *Let $I \subseteq \mathbb{R}$ be an interval. Suppose that $f: I \to \mathbb{R}$ is differentiable with $f'(t) = 0$ for all $t \in I$. Then f is constant on I.*

Proof. Fix $t_0 \in I$; we will show that $f(t) = f(t_0)$ for all $t \in I$. First assume $t \in I$ and $t < t_0$. Since I is an interval, $[t, t_0] \subseteq I$; then f is continuous on $[t, t_0]$ and differentiable on (t, t_0). The mean value theorem then implies

$$\frac{f(t_0) - f(t)}{t_0 - t} = f'(\tau) = 0$$

for some $\tau \in (t, t_0)$. Thus $f(t_0) = f(t)$. The proof when $t_0 < t$ is identical. ■

The key step in this proof was using the fact that I was an interval and thus $[t, t_0] \subseteq I$ when t, $t_0 \in I$ with $t \leq t_0$. Removing the assumption that I is an interval breaks the corollary; for example, if $I = \mathbb{R} \setminus \{0\}$, then the restriction $g := f\big|_I$ with f given by Example 2.6.21 is differentiable on I with $g' = 0$ and yet g is not constant.

The following problem indicates that the mean value theorem is not, in general, true for real-variable functions that are complex-and-non-real-valued—even when those functions are defined on real *intervals*.

2.6.24 Problem ($\star$). Define

$$f: [0, 2\pi] \subseteq \mathbb{R} \to \mathbb{C}: t \mapsto e^{it}.$$

Why are the results of the mean value theorem false for f?

Based on this problem, we might give up hope of generalizing Corollary 2.6.23 to the complex world. This would be premature.

2.6.25 Problem (!). Suppose that $I \subseteq \mathbb{R}$ is an interval and $f: I \to \mathbb{C}$ is differentiable with $f'(t) = 0$ for all $t \in I$. Prove that f is constant on I. [Hint: *since there is no mean value theorem for complex-valued functions, we cannot quite cite Corollary 2.6.23 directly—but we can apply it to* $\text{Re}[f]$ *and* $\text{Im}[f]$.]

Returning to complex-valued functions on (open) subsets of $\mathbb{C}$, it turns out that holomorphic functions with identically zero derivatives are not so terribly far from being constant.

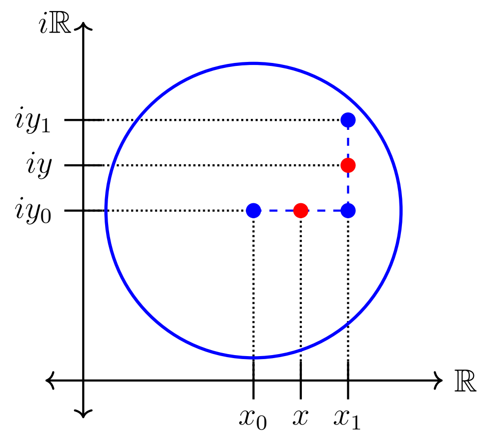
2.6.26 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ be open. A function $f: \mathcal{D} \to \mathbb{C}$ is* LOCALLY CONSTANT *if for each $z \in \mathcal{D}$ and $r > 0$ such that $\mathcal{B}(z; r) \subseteq \mathcal{D}$, the restriction $f\big|_{\mathcal{B}(z;r)}$ is constant: there is $c \in \mathbb{C}$ such that $f(w) = c$ for all $w \in \mathcal{B}(z; r)$.*

2.6.27 Theorem. *Suppose that $f\colon \mathcal{D} \to \mathbb{C}$ is holomorphic with $f'(z) = 0$ for all $z \in \mathcal{D}$. Then f is locally constant on $\mathcal{D}$.*

This is where we finished on Monday, February 19, 2024.

Proof. Fix $z_0 = x_0 + iy_0 \in \mathcal{D}$ and $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$. We want to show that $f(z_0) = f(z)$ for all $z \in \mathcal{B}(z_0; r)$.

Write $f(x + iy) = u(x, y) + iv(x, y)$. Since f is holomorphic, the Cauchy–Riemann equations tell us that the partial derivatives u_x, u_y, v_x, and v_y exist on $\mathcal{D}$. (Strictly speaking, those four partials are defined on $\widetilde{\mathcal{D}} = \{(x, y) \in \mathbb{R}^2 \mid x + iy \in \mathcal{D}\}$, but this is excessive notation.) Moreover, since $f'(x + iy) = 0$ for all $x + iy \in \mathcal{D}$, the identities (2.6.9) imply $u_x = u_y = v_x = v_y = 0$ on $\mathcal{D}$. It therefore suffices to show that u and v are constant on $\mathcal{B}(z_0; r)$, i.e., for $x_1 + iy_1 \in \mathcal{B}(z_0; r)$, we have $u(x_1, y_1) = u(x_0, y_0)$ and $v(x_1, y_1) = v(x_0, y_0)$. (Again, strictly speaking, u and v are not defined on $\mathcal{B}(z_0; r)$ but rather on the isomorphic set $\widetilde{\mathcal{B}}(z_0; r) := \{(x, y) \in \mathbb{R}^2 \mid x + iy \in \mathcal{B}(z_0; r)\}$. But this is also too excessive to bring up in polite company.) We will only do this for u, as the proof for v is analogous. Additionally, we will assume $x_0 < x_1$ and $y_0 < y_1$. There are many other arrangements of the inequalities, but their proofs are all similar.



Since $x_0 + iy_0$, $x_1 + iy_1 \in \mathcal{B}(z_0; r)$, Problem 2.3.6 tells us that $x + iy \in \mathcal{B}(z_0; r)$ for all $x \in [x_0, x_1]$ and all $y \in [y_0, y_1]$. The picture above also suggests a proof strategy. Since $u_x = 0$, the function u is constant in the "x-direction," and so the function $u(\cdot, y_0)$ is going to be constant on $[x_0, x_1]$. So $u(x_0, y_0) = u(x_1, y_1)$. And since $u_y = 0$, u is also constant in the "y-direction," and so $u(x_1, \cdot)$ is constant on $[y_0, y_1]$. So $u(x_1, y_0) = u(x_1, y_1)$.

Now we make this precise by defining

$$g\colon [x_0, x_1] \to \mathbb{R}\colon x \mapsto u(x, y_0) \quad \text{and} \quad h\colon [y_0, y_1] \to \mathbb{R}\colon y \mapsto u(x_1, y).$$

Then

$$g'(x) = u_x(x, y_0) = 0 \quad \text{and} \quad h'(y) = u_y(x_1, y) = 0$$

for all x and y, and so both g and h are constant. Consequently,

$$u(x_0, y_0) = g(x_0) = g(x_1) = u(x_1, y_0) = h(y_0) = h(y_1) = u(x_1, y_1).$$

This proves that u is constant on $\mathcal{B}(x_0 + iy_0; r)$, as desired. ■

2.6.28 Problem ($\star$). Suppose that $f: \mathbb{C} \to \mathbb{C}$ is entire with $f'(z) = 0$ for all $z \in \mathbb{C}$. Prove that f is constant on $\mathbb{C}$ (not just locally constant).

2.6.29 Problem (+). Previously we saw that the differentiability of $\exp(\cdot)$ was largely a consequence of the functional equation $\exp(z + w) = \exp(z)\exp(w)$. However, we could start with the derivative properties of the exponential and obtain the functional equation. Along the way, we rely on the fact that if the derivative of an entire function is identically zero, then that function is constant. Suppose that $f: \mathbb{C} \to \mathbb{C}$ is entire and satisfies the initial value problem

$$\begin{cases} f'(z) = f(z), \ z \in \mathbb{C} \\ f(0) = 1. \end{cases} \tag{2.6.11}$$

(i) Show that f satisfies the functional equation $f(z + w) = f(z)f(w)$ for all z, $w \in \mathbb{C}$. [Hint: *fix* z, $w \in \mathbb{C}$ *and define* $g(\xi) := f(z + w - \xi)f(\xi)$. *Show that* $g'(\xi) = 0$ *for all* ξ.]

(ii) Show that the only solution to the IVP (2.6.11) is $f(z) = \exp(z)$. [Hint: *certainly* exp *is a solution, but why is it the only solution? Obtain from* (2.6.11) *the equation* $f'(z)\exp(-z) + f(z)[-\exp(-z)] = 0$ *and recognize the product rule. This is the integrating factor method from differential equations.*]

Now we revisit the situation at the start of this section.

2.6.30 Example. Suppose that $f: \mathcal{D} \to \mathbb{C}$ is holomorphic and strictly real-valued, so $\text{Im}[f(z)] = 0$ for all $z \in \mathcal{D}$. Write $f(x+iy) = u(x, y)+iv(x, y)$, with $u(x, y) = \text{Re}[f(x+iy)]$ and $v(x, y) = \text{Im}[f(x + iy)]$. Then $v = 0$, and so, since f is holomorphic, the Cauchy–Riemann equations imply $u_x(x, y) = v_x(x, y) = 0$ for all $x + iy \in \mathcal{D}$. Then $f'(x + iy) = u_x(x, y) + iv_x(x, y) = 0$ for all $x + iy \in \mathcal{D}$. We conclude that f is locally constant on $\mathcal{D}$ (but maybe not constant).

2.6.31 Problem ($\star$). Generalize this example as follows. Suppose $f: \mathcal{D} \to \mathbb{C}$ is holomorphic with either $\text{Re}[f]$ or $\text{Im}[f]$ locally constant on $\mathcal{D}$. Then f itself is locally constant on $\mathcal{D}$.

2.6.32 Problem (+). Find all holomorphic functions $f: \mathcal{D} \to \mathbb{C}$ such that $\overline{f}$ is also holomorphic on $\mathcal{D}$.

This suggests that the values of a holomorphic function defined on an open subset of $\mathbb{C}$ must exhibit a certain "diversity." It is no problem for a function $f: [a, b] \subseteq \mathbb{R} \to \mathbb{R}$ to be differentiable, strictly real-valued, and not locally constant; this is daily life in real-valued calculus, of course. But as soon as we expand the domain of f to be an open subset of $\mathbb{C}$ (and no nonempty subset of $\mathbb{R}$ is open), then f cannot take just real (or just imaginary) values without being very "dull."

Can we improve these results? Are there any situations in which an identically zero derivative guarantees a genuinely constant function, not just a locally constant one? The answer is yes, but it requires more hypotheses on the domain $\mathcal{D}$ and some further topological machinery. We will now turn toward constructing that machinery, as it will play a key role in our development of the integral.

3. INTEGRAL CALCULUS

The integral is fundamentally a tool for *representing* functions and *extracting and measuring data* about functions. We have already met one integral representation of a function in real-variable calculus via the fundamental theorem:

$$f(t) = f(t_0) + \int_{t_0}^{t} f'(\tau)\, d\tau$$

when f is differentiable on an interval I, f' is continuous on I, and t_0, $t \in I$. We will redevelop the fundamental theorem (more or less) from scratch here and see other, possibly deeper (and possibly better) representations of functions via integrals. A typical calculus course emphasizes less the "data extraction" aspect of the integral, so that will be largely new here, but we probably saw an argument that the number

$$\frac{1}{b-a}\int_a^b f(t)\, dt$$

is a good measure of the "average value" of f on the interval $[a, b]$. This is one data point about a function that integrals extract and measure. A course in partial differential equations might develop integral norms like

$$\left(\int_a^b |f(t)|^p\, dt\right)^{1/p}, \ 1 \le p < \infty,$$

as good measures of the "size" of f on $[a, b]$ from different perspectives. There will be yet other points of data extracted and measured by integrals.

We will develop an integral for complex-valued functions of a complex variable that is sufficiently robust both to represent functions adequately and to extract and measure meaningful data about those functions. We will build this integral out of two tools—the familiar (. . .one hopes. . .) definite integral from real-variable calculus and the notion of a parametrized "path" or "curve" in the two-dimensional plane (which, in principle, we also met in calculus). Since our results on the "complex" integral will have many parallels with properties of the "real" integral, we postpone a review of the "real" integral and start by developing paths first. This will also draw directly on the differential calculus that we just completed and answer an unresolved question about the differential equation $f' = 0$ to boot.

3.1. Paths, curves, contours.

3.1.1. *Smooth paths.*

Here we study functions $\gamma\colon [a, b] \subseteq \mathbb{R} \to \mathbb{C}$, with a, $b \in \mathbb{R}$ and $a \le b$. We will need the notions of "one-sided" continuity and differentiability at a and b from Problems 2.4.8 and 2.5.22.

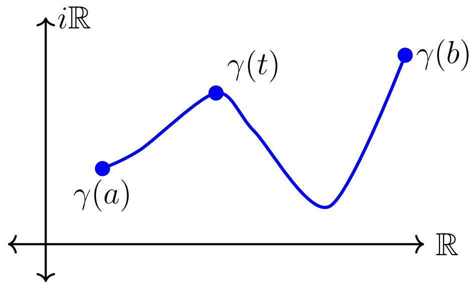
3.1.1 Definition. **(i)** *A function* $f\colon [a, b] \subseteq \mathbb{R} \to \mathbb{C}$ *is* CONTINUOUSLY DIFFERENTIABLE *if* γ *is differentiable and if* γ' *is continuous.*

(ii) *A* **SMOOTH PATH IN** $\mathcal{D} \subseteq \mathbb{C}$ *is a continuously differentiable function* $\gamma \colon [a,b] \subseteq \mathbb{R} \to \mathcal{D}$ *for some* $a, b \in \mathbb{R}$ *with* $a \le b$.

(iii) *The* **IMAGE** *of a smooth path* $\gamma \colon [a,b] \subseteq \mathbb{R} \to \mathbb{C}$ *is its range, i.e., the set*

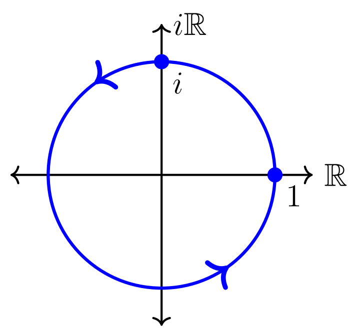
$$\gamma([a,b]) = \{\gamma(t) \mid t \in [a,b]\}.$$

Common synonyms for "path" are **CURVE** and **CONTOUR**. We plot the image of a path parametrically in the complex plane as we do a parametric curve in $\mathbb{R}^2$.



3.1.2 Example. The map $\gamma \colon [0, 2\pi] \to \mathbb{C} \colon t \mapsto e^{it}$ is a smooth path, and the image of this path is the unit circle. That is,

$$\left\{e^{it} \mid 0 \le t \le 2\pi\right\} = \{z \in \mathbb{C} \mid |z| = 1\}.$$

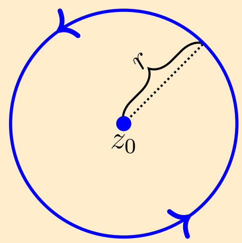


The image of this path has an inherent orientation or trajectory: it starts at 1 and "moves counterclockwise" to i, then to -1, then to $-i$ and, last, back to 1. That is, $\gamma(0) = \gamma(1)$, and so we might think that γ is "closed." We mark the orientation with euphemistic arrows on the parametric curve.

3.1.3 Problem (!). Let $z_0 \in \mathbb{C}$ and $r > 0$. Explain why the map

$$\gamma \colon [0, 2\pi] \to \mathbb{C} \colon t \mapsto z_0 + re^{it}$$

is a smooth path and show that its image is the circle $\partial\mathcal{B}(z_0; r)$. The "orientation" of this image is what we expect from years with the unit circle: a "counterclockwise" trajectory.

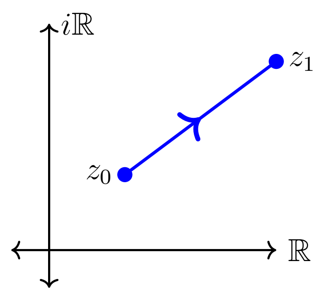


$$\gamma(t) = z_0 + re^{it},\ 0 \le t \le 2\pi$$

In addition to circles, line segments are also among the most important paths that we will study. One motivation for the following definition is the recollection from multivariable calculus that the line segment between vectors $\mathbf{x}_1$, $\mathbf{x}_2 \in \mathbb{R}^2$ is the set of all points of the form $(1-t)\mathbf{x}_1 + t\mathbf{x}_2$ for $0 \le t \le 1$, which can be experimentally verified by graphing.

3.1.4 Definition. *Let z_1, $z_2 \in \mathbb{C}$. The* **LINE SEGMENT FROM** z_1 **TO** z_2 *is the path*

$$\gamma\colon [0,1] \to \mathbb{C}\colon t \mapsto (1-t)z_1 + tz_2.$$



Sometimes we will abuse terminology and refer to the *image* of this line segment as the line segment itself. That is, we will also call the set

$$[z_1, z_2] := \{(1-t)z_1 + tz_2 \mid 0 \le t \le 1\}$$

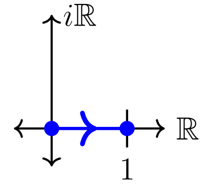
the line segment from z_1 to z_2. And sometimes we will denote the function γ in Definition 3.1.4 by $[z_1, z_2]$ as well.

3.1.5 Example. The line segment from 0 to 1 is the path

$$\gamma\colon [0,1] \to \mathbb{C}\colon t \mapsto [(1-t)\cdot 0] + (1 \cdot t) = t,$$

and so the image of γ is

$$\{\gamma(t) \mid 0 \le t \le 1\} = \{t \mid 0 \le t \le 1\} = [0,1].$$



3.1.6 Problem ($\star$). **(i)** Sometimes it is convenient to represent the same line segment in multiple ways. Let z_1, $z_2 \in \mathbb{C}$. Show that

$$\{(1-t)z_1 + tz_2 \mid 0 \le t \le 1\} = \{\tau z_1 + (1-\tau)z_2 \mid 0 \le \tau \le 1\}$$
$$= \{tz_1 + \tau z_2 \mid 0 \le t,\ \tau \le 1 \text{ and } t + \tau = 1\}$$

(ii) If a, $b \in \mathbb{R}$ with $a \le b$, then of course we want to think of the line segment from a to b as the interval $[a, b]$. Show that this is still the case per Definition 3.1.4. That is, show

$$\{x \in \mathbb{R} \mid a \le x \le b\} = \{(1-t)a + tb \mid 0 \le t \le 1\}.$$

3.1.7 Problem ($\star$). Let $a \in \mathbb{C}$ and $r > 0$. Show that if z_1, $z_2 \in \mathcal{B}(a; r)$, then $[z_1, z_2] \subseteq \mathcal{B}(a; r)$. [Hint: *as always, start by drawing a picture. Then think carefully about the definitions of the sets* $[z_1, z_2]$ *and* $\mathcal{B}(a; r)$.]

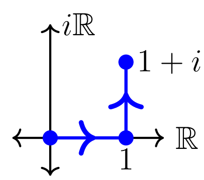
Many meaningful "paths" in $\mathbb{C}$ are not really smooth; they fail to be differentiable at a select number of points, but they fail in a tame way. We now show how to construct such paths out of genuinely smooth paths.

3.1.2. Composition of paths and nonsmooth paths.

Let γ_1 be the line segment from 0 to 1, and let γ_2 be the line segment from 1 to $1+i$. That is,

$$\gamma_1 \colon [0,1] \to \mathbb{C} \colon t \mapsto t \quad \text{and} \quad \gamma_2 \colon [0,1] \to \mathbb{C} \colon t \mapsto (1-t) + t(1+i). \tag{3.1.1}$$

The images of these paths are sketched below.



It looks like the line segment from 0 to 1 just continues into the line segment from 1 to $1+i$, and so we might ask if there is a "natural" way to combine these two paths into one function. That is, is there a map $\gamma \colon [a, b] \to \mathbb{C}$ whose image is $\gamma_1([0,1]) \cup \gamma_2([0,1])$, and does γ somehow restrict "naturally" into γ_1 and γ_2? There is, although this map will not quite be a *smooth* path.

Here is how we can construct such a γ. It takes from "time" 0 to 1 for γ_1 to get from 0 to 1, and then from "time" 0 to 1 for γ_2 to get from 1 to $1+i$. Maybe it should take from "time" 0 to 2 for γ to get from 0 to $1+i$; from "time" 0 to 1, γ should act like γ_1, and from "time" 1 to 2, γ should act like γ_2. So, we put

$$\gamma\colon [0,2] \to \mathbb{C}\colon t \mapsto \begin{cases} \gamma_1(t), & 0 \le t \le 1 \\ \gamma_2(t-1), & 1 \le t \le 2. \end{cases} \tag{3.1.2}$$

3.1.8 Problem (!). Suppose $\gamma_1\colon [a_1,b_1] \to \mathbb{C}$ and $\gamma_2\colon [a_2,b_2] \to \mathbb{C}$ are smooth paths with $\gamma_1(b_1) = \gamma_2(a_2)$. Abbreviate $\gamma := \gamma_1 \oplus \gamma_2$ and $I := [a_1, b_1 + (b_2 - a_2)]$.

(i) Show that $\gamma(0) = 0$ and $\gamma(2) = 1+i$.

(ii) Show that $\gamma([0,2]) = \gamma_1([0,1]) \cup \gamma_2([0,1])$.

(iii) Show that γ is continuous on $[0,2]$.

(iv) Show that γ is differentiable on $[0,1) \cup (1,2]$ and that γ' is continuous.

(v) Show that the limits $\lim_{t\to 1^-} \gamma'(t)$ and $\lim_{t \to 1^+} \gamma'(t)$ exist.

Here is the abstraction of this situation.

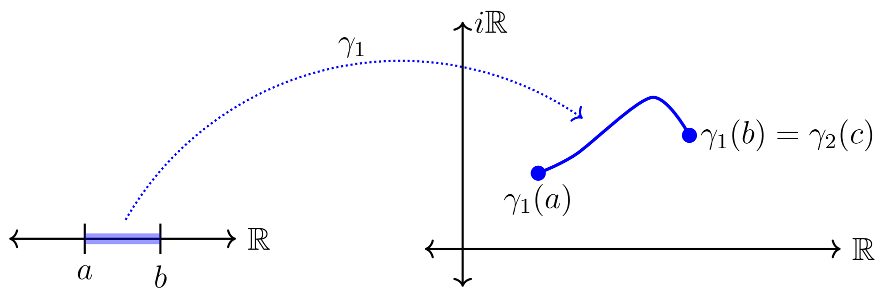
3.1.9 Definition. *Suppose $\gamma_1\colon [a_1,b_1] \to \mathbb{C}$ and $\gamma_2\colon [a_2,b_2] \to \mathbb{C}$ are smooth paths with $\gamma_1(b_1) = \gamma_2(a_2)$. Then the* COMPOSITION *of γ_1 and γ_2 is the map*

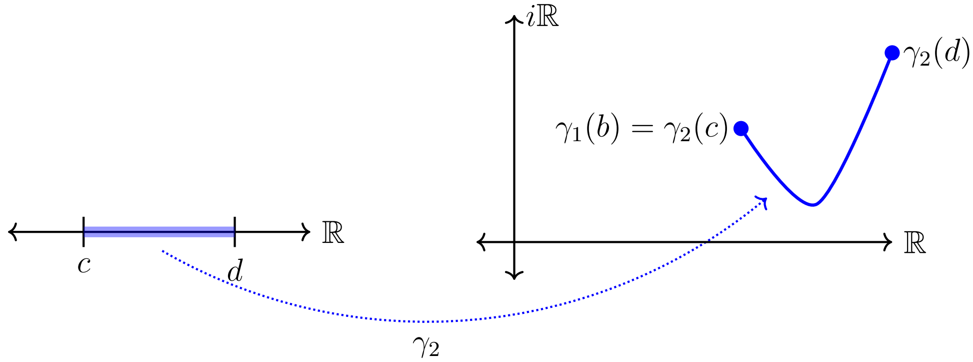
$$\gamma_1 \oplus \gamma_2\colon [a_1, b_1 + (b_2 - a_2)] \to \mathbb{C}\colon t \mapsto \begin{cases} \gamma_1(t), & a_1 \le t \le b_1 \\ \gamma_2(t + a_2 - b_1), & b_1 \le t \le b_1 + (b_2 - a_2). \end{cases}$$

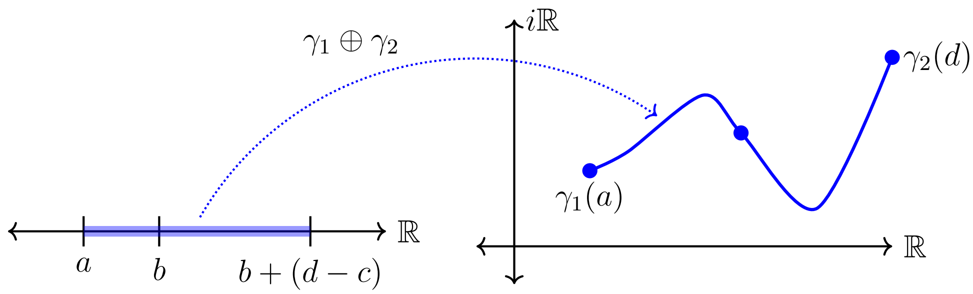
Sometimes this path is denoted by $\gamma_1 + \gamma_2$ or $[\gamma_1, \gamma_2]$ instead.

3.1.10 Problem (!). Check that if $b_1 \le t \le b_1 + b_2 - a_2$, then $a_2 \le t - b_1 + a_2 \le b_2$, and so $\gamma_2(t - b_1 + a_2)$ is defined if γ_2 is defined on $[b_2, a_2]$.

Here is a visualization of composition.







Like γ_1 and γ_2, $\gamma_1 \oplus \gamma_2$ is a map from a closed, bounded subinterval of $\mathbb{R}$ into $\mathbb{C}$. It is continuous on its domain, but it may not be differentiable (let alone continuously differentiable) at the "connecting" point b_1.

3.1.11 Problem ($\star$). This problem generalizes Problem 3.1.8. Suppose $\gamma_1 \colon [a_1, b_1] \to \mathbb{C}$ and $\gamma_2 \colon [a_2, b_2] \to \mathbb{C}$ are smooth paths with $\gamma_1(b_1) = \gamma_2(a_2)$. Abbreviate $\gamma := \gamma_1 \oplus \gamma_2$ and $I := [a_1, b_1 + (b_2 - a_2)]$.

(i) Show that $\gamma(a_1) = \gamma_1(a_1)$ and $\gamma(b_1 + (b_2 - a_2)) = \gamma_2(b_2)$.

(ii) Show that $\gamma(I) = \gamma_1([a_1, b_1]) \cup \gamma_2([a_2, b_2])$.

(iii) Show that γ is continuous on I.

(iv) Show that γ is differentiable on $I \setminus \{b_1\}$ and that $\gamma' \colon I \setminus \{b_1\} \to \mathbb{C}$ is continuous.

(v) Show that the limits $\lim_{t \to b_1^-} \gamma'(t)$ and $\lim_{t \to b_1^+} \gamma'(t)$ exist.

(vi) What more must we assume to guarantee that $\gamma_1 \oplus \gamma_2$ is differentiable?

This is where we finished on Wednesday, February 21, 2024.

3.1.12 Problem (!). Here is a generalization of the preceding example. Let z_1, z_2, $z_3 \in \mathbb{C}$. Show that the line segment from z_1 to z_2 is parametrized by

$$\gamma_1 \colon [0,1] \to \mathbb{C} \colon t \mapsto (1-t)z_1 + tz_2$$

and the line segment from z_2 to z_3 is parametrized by

$$\gamma_2 \colon [0,1] \to \mathbb{C} \colon t \mapsto (1-t)z_2 + tz_3.$$

Since $\gamma_1(1) = z_1 = \gamma_2(0)$, we can compose γ_1 and γ_2. Show that

$$(\gamma_1 \oplus \gamma_2)(t) = \begin{cases} (1-t)z_1 + tz_2, \ 0 \le t \le 1 \\ tz_2 + (t-1)z_3, \ 1 \le t \le 2. \end{cases}$$

Use this to show that the map γ in (3.1.2) is in fact the composition $\gamma_1 \oplus \gamma_2$ of the paths from (3.1.1).

3.1.13 Remark. *We will compose various paths with line segments often enough that it is worth having a special notation for that. Let z_1, $z_2 \in \mathbb{C}$ and let γ be the line segment from z_1 to z_2. If μ is a path with terminal point z_1, then we will write*

$$\mu \oplus [z_1, z_2] \quad \textit{instead of} \quad \mu \oplus \gamma.$$

Likewise, if ν is a path with initial point z_2, then we will write

$$[z_1, z_2] \oplus \nu \quad \textit{instead of} \quad \gamma \oplus \nu.$$

3.1.14 Definition. *Suppose that $\gamma_k \colon [a_k, b_k] \to \mathbb{C}$, $1 \le k \le n$, are smooth paths with $\gamma_k(b_k) = \gamma_{k+1}(a_{k+1})$ for $k = 1, \ldots, n-1$. We define their composition $\oplus_{k=1}^n \gamma_k$ recursively via*

$$\oplus_{k=1}^n \gamma_k := \begin{cases} \gamma_1 \oplus \gamma_2, \ n = 2 \\ \\ \left(\oplus_{k=1}^{n-1} \gamma_k\right) \oplus \gamma_n, \ n \ge 2. \end{cases} \tag{3.1.3}$$

Sometimes this composition is denoted by $\gamma_1 \oplus \cdots \oplus \gamma_n$, $\gamma_1 + \cdots + \gamma_n$, or $[\gamma_1, \ldots, \gamma_n]$.

3.1.15 Problem ($\star$). Suppose that $\gamma_k \colon [a_k, b_k] \to \mathbb{C}$, $1 \le k \le n$, are paths with $\gamma_k(b_k) = \gamma_{k+1}(a_{k+1})$ for $k = 1, \ldots, n-1$.

(i) Define

$$\tau_k := \begin{cases} a_1, \ k = 0 \\ b_1, \ k = 1, \\ \tau_{k-1} + b_k - a_k, \ k \ge 2. \end{cases}$$

Show that the domain of $\oplus_{k=1}^{n}\gamma_k$ is $[a_1, b_1 + \sum_{k=2}^{n}(b_k - a_k)]$.

(ii) Define

$$\mathbb{1}_{[\tau_{k-1},\tau_k]}(t) := \begin{cases} 1, & \tau_{k-1} \le t \le \tau_k \\ 0, & t < \tau_{k-1} \text{ or } t > \tau_k. \end{cases}$$

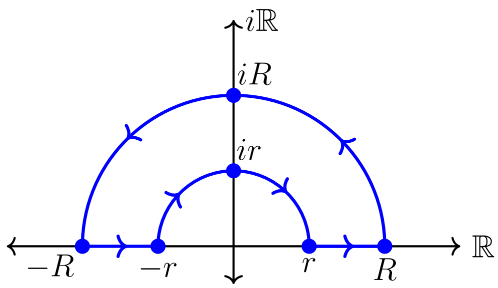
Show that

$$\left(\oplus_{k=1}^{n} \gamma_k\right)(t) = \sum_{k=1}^{n} \mathbb{1}_{[\tau_{k-1},\tau_k]}(t)\mu_k(t + a_k - \tau_{k-1}).$$

[Hint: *a full proof requires induction on n. To get a feel for how the τ_k work as they do and the structure of the sum above with the "indicator" functions $\mathbb{1}_{[\tau_{k-1},\tau_k]}$, just use Definition 3.1.14 recursively for $n = 4$, which will be large enough to be illustrative but not large enough to be overly annoying.*]

Nonetheless, when we consider a "large" composition like $\oplus_{k=1}^{n}\gamma_k$ in (3.1.3) above, we will rarely need to know what the domain of $\oplus_{k=1}^{n}\gamma_k$ actually is; it usually suffices to keep track of the individual domains of the components.

3.1.16 Example. Let $0 < r < R$. We will find four paths γ_1, γ_2, γ_3, γ_4 such that the image of $\oplus_{k=1}^{4}\gamma_k$ is the path below.



The line segment from $z = r$ to $z = R$ is parametrized by

$$\gamma_1(t) := (1 - t)r + tR = (R - r)t + r, \ 0 \le t \le 1.$$

The upper half of the circle of radius R with "counterclockwise" orientation is parametrized by

$$\gamma_2(t) := Re^{it}, \ 0 \le t \le \pi.$$

The line segment from $z = -R$ to $z = -r$ is parametrized by

$$\gamma_3(t) := (1 - t)(-R) + t(-r) = (t - 1)R - rt = (R - r)t - R, \ 0 \le t \le 1.$$

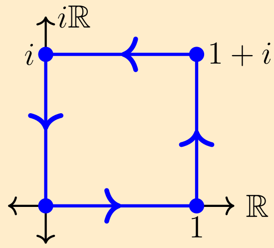
And (if we think hard about it) the upper half of the circle of radius r with "clockwise" orientation is parametrized by

$$\gamma_4(t) := re^{i(\pi - t)}, \ 0 \le t \le \pi.$$

The idea here is that γ_4 starts at $re^{i\pi} = -r$ at $t = 0$, passes through $re^{i(\pi-\pi/2)} = re^{i\pi/2} = ir$ at $t = \pi/2$, and ends at $re^{i(\pi-\pi)} = r$ at $t = \pi$. So, γ_4 needs to "reverse" the path $t \mapsto re^{it}$ on $[0, \pi]$, a notion that we will make precise shortly.

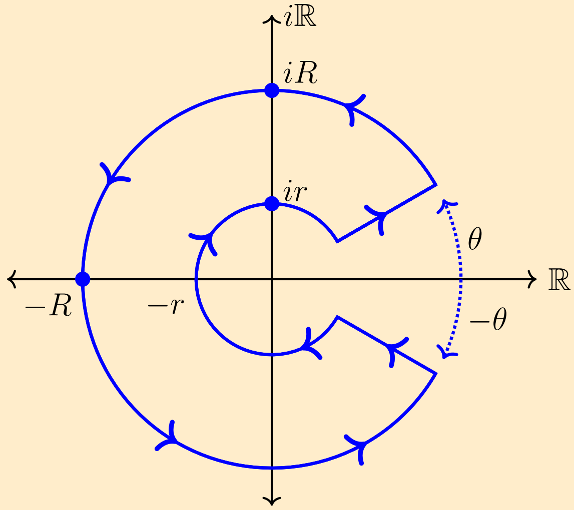
In the preceding example, we could write a piecewise formula for $\oplus_{k=1}^4 \gamma_k$ over some domain $[0, b]$ for some $b > 0$. However, we will actually never use such a formula when we work with compositions of paths later, and such a formula would only obscure the four individual domains above.

3.1.17 Problem (!). Find formulas for four paths γ_k, $k = 1, 2, 3, 4$, such that the image of $\oplus_{k=1}^4 \gamma_k$ is the unit square sketched below.



You do not need to find a "common domain" for the composition but instead can just give formulas for the four paths as in Example 3.1.16. Note that the curve as drawn is oriented roughly "counterclockwise" in the sense that as you traverse the curve in the direction indicated the "inside" square stays on your left. This is the same phenomenon that happens when we orient a circle counterclockwise.

3.1.18 Problem ($\star$). Let $0 < r < R$ and $0 < \theta < \pi/2$. Parametrize the "keyhole contour" below by finding formulas for four paths γ_1, γ_2, γ_3, and γ_4 such that the image of $\oplus_{k=1}^4 \gamma_k$ is the "keyhole contour" below. The angle of the "opening" is 2θ radians. Again, you do not need to find a "common domain" for the composition but instead can just give formulas for the four paths as in Example 3.1.16.



Do you see a "counterclockwise" orientation to this curve?

Now, finally, we relax the smoothness requirement of paths ever so slightly to obtain a more practically useful version.

3.1.19 Definition. **(i)** *A* **PATH IN** $\mathcal{D} \subseteq \mathbb{C}$ *is a map* $\gamma\colon [a,b] \to \mathcal{D}$ *of the form* $\gamma = \oplus_{k=1}^{n} \gamma_k$, *where each* γ_k *is a smooth path in* $\mathcal{D}$.

(ii) *The* **IMAGE** *of the path* $\gamma\colon [a,b] \to \mathbb{C}$ *is, again, its range:*

$$\gamma([a,b]) = \{\gamma(t) \mid t \in [a,b]\}.$$

(iii) *The* **INITIAL POINT** *of the path* $\gamma\colon [a,b] \to \mathbb{C}$ *is* $\gamma(a)$, *and its* **TERMINAL POINT** *is* $\gamma(b)$.

(iv) *A path* $\gamma\colon [a,b] \to \mathbb{C}$ *is* **CLOSED** *if* $\gamma(a) = \gamma(b)$, *i.e., if it has the same initial and terminal points.*

3.1.20 Example. The map

$$\gamma\colon [-1,1] \to \mathbb{C}\colon t \mapsto |t|$$

is a path in $\mathbb{C}$ because we can write it as $\gamma = \gamma_1 \oplus \gamma_2$, where

$$\gamma_1\colon [-1,0] \to \mathbb{C}\colon t \mapsto -t \quad \text{and} \quad \gamma_2\colon [0,1] \to \mathbb{C}\colon t \mapsto t,$$

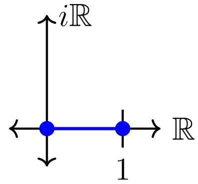
both of which are smooth paths. However, γ itself is not a smooth path, as it is not differentiable at 0:

$$\lim_{\tau \to 0^-} \frac{\gamma(\tau) - \gamma(0)}{\tau - 0} = -1 \neq 1 = \lim_{\tau \to 0^+} \frac{\gamma(\tau) - \gamma(0)}{\tau - 0}.$$

The image of γ is the set

$$\{\gamma(t) \mid t \in [-1,1]\} = \{|t| \mid t \in [-1,1]\} = [0,1].$$

Plotting this image parametrically in $\mathbb{C}$ yields what to our real-variable-calculus-trained eyes is a smooth graph.



We should be careful not to make assumptions about the smoothness of a path from the appearance of its image. Also, note that $\gamma(-1) = 1 = \gamma(1)$, so this path is closed—it "turns around" at $t = 0$ and "doubles back" over itself. For this reason, we did not put any euphemistic arrows on the plot above.

If γ_1 and γ_2 are paths such that the terminal point of γ_1 is the initial point of γ_2, then we can compose γ_1 and γ_2 exactly as in Definition 3.1.9 and obtain a new path (which need not be a smooth path, since neither γ_1 nor γ_2 is assumed to be smooth here). Likewise, we can compose an arbitrary number of paths and get a new path, just like Definition 3.1.14.

3.1.21 Problem (!). What is the image of the path

$$\gamma \colon [-1,1] \to \mathbb{C} \colon t \mapsto 0?$$

Is it the same as the image in the previous example?

3.1.22 Problem (!). When is a line segment a closed path?

3.1.23 Problem ($\star$). Show that the paths

$$\gamma_1 \colon [-1,1] \to \mathbb{C} \colon t \mapsto t + i|t|$$

and

$$\gamma_2 \colon [-1,1] \to \mathbb{C} \colon t \mapsto t^3 + i|t|^3$$

have the same images. Which path is smooth? *Here is the lesson: the image of a path may look like the graph of a function from (a subset of) $\mathbb{R}$ to $\mathbb{R}$, and yet the smoothness of that graph may have nothing to do with the smoothness of that path. (Compare this to Example 3.1.20.)*

3.1.24 Problem (+). Show that a map $\gamma \colon [a,b] \to \mathbb{C}$ is a path if and only if there exists a subset $\{t_k\}_{k=0}^n \subseteq [a,b]$ such that $t_0 = a$, $t_n = b$, and $t_{k-1} < t_k$ for $1 \le k \le n$, and $\gamma\big|_{[t_{k-1},t_k]}$ is continuously differentiable. For this reason, paths are sometimes called **PIECEWISE CONTINUOUSLY DIFFERENTIABLE** functions. (Again, the notions of "one-sided" continuity and differentiability at a and b from Problems 2.4.8 and 2.5.22 will be helpful.) Compare this to Problem 3.1.11. Then show that every path is in fact continuous.

3.1.25 Problem (+). Fill in the details of the following argument to prove that the image of a path is never open. Suppose that $\gamma\colon [a,b] \to \mathbb{C}$ is a path and let $\Gamma = \gamma([a,b])$.

(i) Explain why the extreme value theorem provides $t_* \in [a,b]$ such that $|\gamma(t_*)| = \max_{a \le t \le b} |\gamma(t)|$.

(ii) If Γ is open, there is $r > 0$ such that $\mathcal{B}(\gamma(t_*); r) \subseteq \Gamma$. Show that if $\text{Re}[\gamma(t_*)] \ge 0$, then $|\gamma(t_*) + r/2| > |\gamma(t_*)|$, and if $\text{Re}[\gamma(t_*)] \le 0$, then $|\gamma(t_*) - r/2| > |\gamma(t_*)|$.

(iii) Why is this a contradiction? As usual, drawing pictures will help.

3.1.26 Problem (+). **(i)** Let $\gamma\colon [a,b] \to \mathbb{C}$ be a curve and let $z_0 \in \mathbb{C} \setminus \gamma([a,b])$. Show that there exists a point on the image of γ that is "closest" to z_0 in the sense that

$$\min_{a \le t \le b} |\gamma(t) - z_0| = |\gamma(t_0) - z_0|$$

for some $t_0 \in [a,b]$. [Hint: *use the extreme value theorem on the function* $d(t) := |\gamma(t) - z_0|$.]

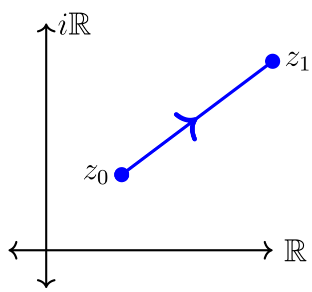
(ii) Draw a picture illustrating this phenomenon for $\gamma(t) = e^{it}$ on $[0, 2\pi]$ and $z_0 = 2$.

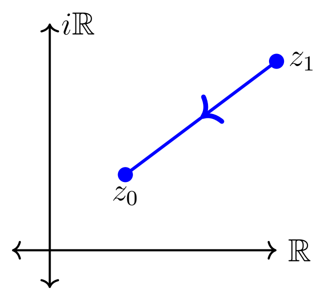
3.1.3. Reversing paths.

3.1.27 Example. Let z_1, $z_2 \in \mathbb{C}$ be distinct points. Consider the paths

$$\gamma_1(t)\colon [0,1] \to \mathbb{C}\colon t \mapsto (1-t)z_1 + tz_2 \quad \text{and} \quad \gamma_2(t)\colon [0,1] \to \mathbb{C}\colon t \mapsto (1-t)z_2 + tz_1.$$

The path γ_1 parametrizes $[z_1, z_2]$, while the path γ_2 parametrizes $[z_2, z_1]$. The images of these paths are the same, since as sets $[z_1, z_2] = [z_2, z_1]$. However, $\gamma_1(0) = z_1 \ne z_2 = \gamma_2(0)$, so γ_1 and γ_2 are distinct functions. (In fact, one can check that $\gamma_1(t) = \gamma_2(t)$ if and only if $t = 1/2$, so these functions certainly are not equal.) This calculation also tells us that the initial point of γ_1 is the terminal point of γ_2, and vice-versa. It appears, then, that γ_1 and γ_2 both "trace out" the same image but in the "reverse direction."





In fact, a little algebra shows

$$\gamma_1(t) = \gamma_2(1-t), \ 0 \le t \le 1. \tag{3.1.4}$$

We formally define this notion of "reverse."

3.1.28 Definition. *Suppose that $\gamma\colon [a,b] \to \mathbb{C}$ is a path. The* **REVERSE** *of γ is the path*

$$\gamma^-(t) := \gamma(a+b-t), \ a \le t \le b.$$

Some sources denote this path by $-\gamma$ or γ^* instead.

3.1.29 Example. We encountered a version of the path

$$\gamma\colon [0,\pi] \to \mathbb{C}\colon t \mapsto e^{it}$$

in Example 3.1.16). The image of this path is the upper half of the unit circle with the usual "counterclockwise" trajectory. The reverse of this path is

$$\gamma^-\colon [0,\pi] \to \mathbb{C}\colon t \mapsto \gamma(0+\pi-t) = e^{i(\pi-t)},$$

and this is exactly what we used in that example.

3.1.30 Problem (!). Let $\gamma\colon [a,b] \to \mathbb{C}$ be a path.

(i) Check that $\gamma^-(a) = \gamma(b)$ and $\gamma^-(b) = \gamma(a)$, so γ^- does indeed "reverse" the initial and terminal points of γ.

(ii) Check that if $a \le t \le b$, then $a \le a+b-t \le b$, and so $\gamma(a+b-t)$ is indeed defined if γ is defined on $[a,b]$.

(iii) Check that γ and γ^- have the same image.

3.1.31 Problem ($\star$). Let $\gamma\colon [a,b] \to \mathbb{C}$ be a path.

(i) Show that if $\phi := \gamma^-$, then $\phi^- = \gamma$. That is, show $(\gamma^-)^- = \gamma$.

(ii) Let $\mu\colon [c,d] \to \mathbb{C}$ also be a path such that the initial point of μ is the terminal point of γ. By considering the domains of $(\gamma \oplus \mu)^-$ and $\mu^- \oplus \gamma^-$, explain why we should not expect $(\gamma \oplus \mu)^- = \mu^- \oplus \gamma^-$ in general. However, if $a = c = 0$ and $b = d = 1$, show that the equality $(\gamma \oplus \mu)^- = \mu^- \oplus \gamma^-$ is true. (In practice, we will shortly see that could always reparametrize γ and μ so that both are defined on $[0,1]$, and any question that we have about $(\gamma \oplus \mu)^-$ and $\mu^- \oplus \gamma^-$ would likely be invariant under this parametrization.)

3.1.4. Reparametrizing paths.

Note carefully that a path is a function, while the image of a path is a set. A given set in $\mathbb{C}$ may be the image of many paths; for example, the unit circle is also the image of $\gamma_k\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto e^{ikt}$ for any $k \in \mathbb{Z}$.

3.1.32 Problem (!). Prove this. That is, show that if $k \in \mathbb{Z}$, then

$$\{z \in \mathbb{C} \mid |z| = 1\} = \left\{e^{ikt} \mid 0 \le t \le 2\pi\right\}.$$

We should distinguish precisely among "set," "path," and "image."

3.1.33 Definition. *A set $\Gamma \subseteq \mathbb{C}$ is* **PARAMETRIZED** *by the path $\gamma\colon [a, b] \to \mathbb{C}$ if the image of γ is Γ, i.e., if $\Gamma = \gamma([a, b])$. In this case we say that γ is a* **PARAMETRIZATION** *of Γ.*

3.1.34 Example. Here are four different parametrizations of the the unit circle, which is the set $\{z \in \mathbb{C} \mid |z| = 1\}$:

$$\begin{aligned}
\gamma_1&\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto e^{it} \\
\gamma_2&\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto e^{-it} \\
\gamma_3&\colon [0, \pi] \to \mathbb{C}\colon t \mapsto e^{2it} \\
\gamma_4&\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto e^{4it}.
\end{aligned}$$

The path γ_1 is probably what we think of as the "usual" parametrization, which "traces out" the unit circle "counterclockwise." (Hopefully the overabundance of quotation marks emphasizes that none of these words or phrases has been given a rigorous mathematical definition yet.) The path γ_2 traces out the unit circle clockwise, e.g., $\gamma_2(\pi/2) = -i$, whereas $\gamma_2(\pi/2) = i$. The path γ_3 traces out the unit circle in "half the time" as γ_1 and γ_2, e.g., $\gamma_3(\pi/4) = i$. And the path γ_4 traces out the unit circle a whopping four times, e.g., $\gamma_4(t) = 1$ for $k = 0$, $\pi/2$, π, $3\pi/2$, and 2π.

It turns out that the paths γ_1, γ_2, and γ_3 are closely related and can be "obtained" from each other in ways that we now make precise. First, $\gamma_2 = \gamma_1^-$, as we readily calculate $\gamma_1^-(t) = \gamma_1(0 + 2\pi - t) = e^{-it} = \gamma_2(t)$. However, γ_3 is something new.

Observe that $\gamma_3(t) = \gamma_1(2t)$, and the map $\varphi_{13}\colon [0, \pi] \to [0, 2\pi]\colon t \mapsto 2t$ is continuously differentiable with $\varphi_{13}'(t) = 2 > 0$ for all t. Observe also that $\gamma_1(t) = \gamma_3(t/2)$, and the map $\varphi_{31}\colon [0, 2\pi] \to [0, \pi]\colon t \mapsto t/2$ is continuously differentiable with $\varphi_{31}'(t) = 1/2 > 0$ for all t. This dual way of viewing γ_3 as the composition $\gamma_3 = \gamma_1 \circ \varphi_{13}$ and of viewing γ_1 as the composition $\gamma_1 = \gamma_3 \circ \varphi_{31}$ is an illustration of a more general phenomenon.

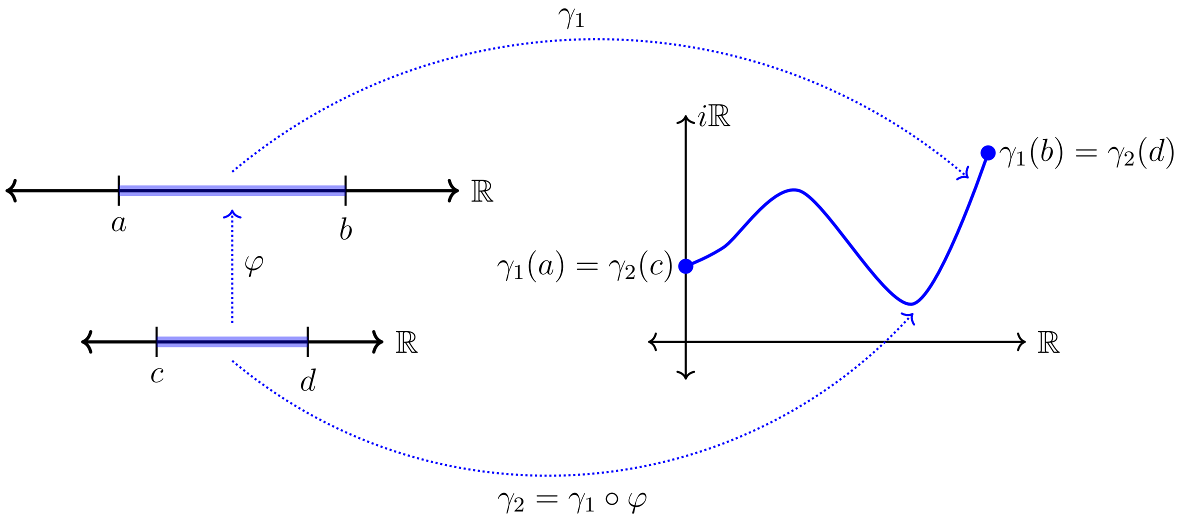
3.1.35 Definition. *Let $\gamma_1\colon [a, b] \to \mathbb{C}$ and $\gamma_2\colon [c, d] \to \mathbb{C}$ be paths. Then γ_1 and γ_2 are* **EQUIVALENT** *if there is a continuously differentiable map $\varphi\colon [c, d] \to [a, b]$ such that*

(i) *$\varphi'(t) > 0$ for all $t \in [c, d]$.*

(ii) *$\varphi(c) = a$ and $\varphi(d) = b$.*

(iii) *$\gamma_2(t) = \gamma_1(\varphi(t))$ for all $t \in [c, d]$.*

We say that γ_1 and γ_2 are **REPARAMETRIZATIONS** *of each other.*



The condition that $\varphi'(t) > 0$ in this definition ensures that φ is strictly increasing on $[a, b]$ and therefore one-to-one. (If φ' is ever negative, then φ would be decreasing, and so $\gamma_1 \circ \varphi$ could "double back" on itself and not have the same "trajectory" as γ_1.) This, morally, encodes the idea that $\gamma_2 = \gamma_1 \circ \varphi$ "traces out the same image" as γ_1 does in the "same orientation."

Additionally, the hypotheses on φ imply that φ is invertible and that its derivative is continuously differentiable. That is, if we assume parts (i) and (ii) of Definition 3.1.35, then there is a continuously differentiable map $\psi\colon [a, b] \to [c, d]$ such that $\varphi(\psi(\tau)) = \tau$ for all $\tau \in [c, d]$, $\psi(\varphi(t)) = t$ for all $t \in [c, d]$, and $\psi'(t) > 0$ for all t. Then we have $\gamma_1(\tau) = \gamma_1(\varphi(\psi(\tau))) = \gamma_2(\psi(\tau))$ for all $\tau \in [a, b]$. Moreover, $\psi(a) = \psi(\varphi(c)) = c$ and $\psi(b) = \psi(\varphi(d)) = d$. In other words, it is irrelevant in part (iii) whether γ_2 is written as a composition of γ_1 and some function, or whether γ_1 is written as a composition of γ_2 and some function. (More generally, the equivalence of paths is indeed an equivalence relation on the set of all paths.)

3.1.36 Problem (!). It is sometimes convenient to assume that the domain of a path is the interval $[0, 1]$. Show that it is always possible to reparametrize a path $\gamma\colon [a, b] \to \mathbb{C}$ by finding a continuously differentiable map $\psi\colon [a, b] \to [0, 1]$ that satisfies the conditions of Definition 3.1.35.

3.1.37 Problem ($\star$). Show that if the path γ_1 is a reparametrization of the path γ_2, then γ_1 and γ_2 have the same image.

3.1.38 Problem ($\star$). Is the reverse of a path ever a reparametrization of that path?

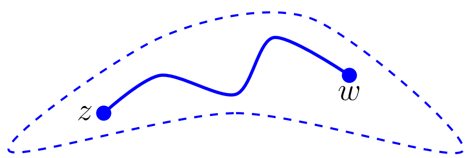
This is where we finished on Friday, February 23, 2024.

3.1.5. Connectedness.

We conclude with an application of paths that strengthens our prior results about functions with identically zero derivatives. First, we need to augment the geometry of our underlying domains.

3.1.39 Definition. *A set $\mathcal{D} \subseteq \mathbb{C}$ is* **CONNECTED** *if for any z, $w \in \mathcal{D}$, there is a path $\gamma \colon [a, b] \to \mathcal{D}$ such that $\gamma(a) = z$ and $\gamma(b) = w$. Sometimes such a set is called* **PATH-CONNECTED**, *not just connected.*

Informally, any points z, $w \in \mathcal{D}$ can be "connected" by a path that lies entirely in $\mathcal{D}$.



3.1.40 Example. **(i)** Any open ball $\mathcal{B}(z_0; r)$ is connected. Given z, $w \in \mathcal{B}(z_0; r)$, it is intuitively plausible that we could connect them by the line segment from z to z_0 and then from z_0 to w, or just by the line segment from z to w. This turns out to be true and is mostly a matter of working through definitions: if z, $w \in \mathcal{B}(z_0; r)$, why is any point on the line segment from z to w also in $\mathcal{B}(z_0; r)$?

(ii) The set $\mathbb{C} \setminus i\mathbb{R}$ (which we encountered in Example 2.6.21) is not connected. Intuitively, any curve connecting a point z with $\text{Re}(z) < 0$ to a point w with $\text{Re}(w) > 0$ must pass through the imaginary axis. Proving this requires some thought, possibly involving the intermediate value theorem to establish this "pass through" claim.

3.1.41 Problem ($\star$). Prove both of the claims in the previous example.

3.1.42 Theorem. *Suppose that $\mathcal{D} \subseteq \mathbb{C}$ is open and connected. If $f \colon \mathcal{D} \to \mathbb{C}$ is differentiable with $f'(z) = 0$ for all $z \in \mathcal{D}$, then f is constant on $\mathcal{D}$: there is $c \in \mathbb{C}$ such that $f(z) = c$ for all $z \in \mathcal{D}$.*

Proof. Fix z, $w \in \mathcal{D}$; we will show that $f(z) = f(w)$. Since $\mathcal{D}$ is path-connected, there is a path $\gamma \colon [a, b] \to \mathcal{D}$ such that $\gamma(a) = z$ and $\gamma(b) = w$.

First suppose that γ is smooth. Set $g(t) := f(\gamma(t))$, so g is also differentiable on $[a, b]$ by the chain rule, and $g'(t) = f'(\gamma(t))\gamma'(t) = 0$. By the mean value theorem, $\text{Re}[g]$ and $\text{Im}[g]$ are constant, so g is constant. Thus $f(z) = g(a) = g(b) = f(w)$.

Now suppose that $\gamma = \oplus_{k=1}^{n}\gamma_k$ for some smooth paths γ_k. For simplicity, take $n = 2$ and use Problem 3.1.24 to find $t_1 \in (a, b)$ such that $\gamma\big|_{[a,t_1]}$ and $\gamma\big|_{[t_1,b]}$ are continuously differentiable. Define

$$g_1 : [a, t_1] \to \mathcal{D} : t \mapsto f(\gamma\big|_{[a,t_1]}(t)) \quad \text{and} \quad g_2 : [t_1, b] \to \mathcal{D} : f(\gamma\big|_{[t_1,b]}(t)).$$

Of course, $g_1(t) = f(\gamma(t))$ for $t \in [a, t_1]$ and $g_2(t) = f(\gamma(t))$ for $t \in [t_1, b]$, but the domains of g_1 and g_2 are different intervals, so g_1 and g_2 are different functions. However, the utility of taking different domains is that g_1 and g_2 are now differentiable, with $g_1' = 0$ and $g_2' = 0$, so g_1 is constant on $[a, t_1]$ and g_2 is constant on $[t_1, b]$. Then

$$f(z) = g_1(a) = g_1(t_1) = f(\gamma(t_1)) = g_2(t_1) = g_2(b) = f(w).$$

We can generalize this argument to include an arbitrary $n \geq 2$ functions $g_k := f \circ \gamma\big|_{[t_{k-1},t_k]}$, and we obtain $g_k(t_{k-1}) = g_k(t_k) = f(\gamma(t_k)) = g_{k+1}(t_k) = g_{k+1}(t_{k+1}) = f(\gamma(t_{k+1}))$ for $k = 1, \ldots, n-1$. ■

3.1.43 Problem (!). Reread the proof of Theorem 2.6.27. Recall that we fixed $z_0 \in \mathcal{D}$ and took $r_0 > 0$ such that $\mathcal{B}(z_0; r_0) \subseteq \mathcal{D}$. By Problem 3.1.41, the ball $\mathcal{B}(z_0; r_0)$ is open and connected, and so f is constant on $\mathcal{B}(z_0; r_0)$. Where in the proof of Theorem 2.6.27 did we use the connectedness of $\mathcal{B}(z_0; r_0)$?

3.1.44 Problem ($\star$). Let $\alpha_1, \alpha_2 \in \mathbb{R}$ with $\alpha_1 \neq \alpha_2$. What do Problem 2.5.19 and Theorem 3.1.42 say about how $\log_{\alpha_1}$ and $\log_{\alpha_2}$ are related?

3.1.45 Problem (+). Let $\mathcal{D} \subseteq \mathbb{C}$ be nonempty and connected.

(i) Suppose that $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, where both $\mathcal{D}_1$ and $\mathcal{D}_2$ are open and $\mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing$. Argue by contradiction as follows that if $\mathcal{D}_2 \neq \varnothing$, then $\mathcal{D}_1 = \varnothing$.

Suppose instead that both $\mathcal{D}_1$ and $\mathcal{D}_2$ are nonempty. Explain why the function

$$f : \mathcal{D} \to \mathbb{C} : z \mapsto \begin{cases} 1, & z \in \mathcal{D}_1 \\ 2, & z \in \mathcal{D}_2 \end{cases}$$

is defined, holomorphic, locally constant, and not constant. Conclude that $\mathcal{D}$ cannot be connected.

(ii) Let $\mathcal{D} = \mathbb{C} \setminus \mathbb{R}$, $\mathcal{D}_1 = \{z \in \mathbb{C} \mid \text{Im}(z) < 0\}$, and $\mathcal{D}_2 = \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}$. Then $\mathcal{D}_1$ and $\mathcal{D}_2$ are open, $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, with $\mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing$ but $\mathcal{D}_1 \neq \varnothing$ and $\mathcal{D}_2 \neq \varnothing$. Draw a picture of this situation. Then draw a curve with initial point in $\mathcal{D}_1$ and terminal point in $\mathcal{D}_2$. Point out on your drawing how this curve shows that $\mathcal{D}$ is not connected.

3.2. Definite integrals.

We have said that the integral is a key tool for both representing functions and for extracting and measuring meaningful data about functions. Here we primarily take up the question of representing functions via integrals—specifically, representing antiderivatives via integrals.

3.2.1 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$. A holomorphic function $F \colon \mathcal{D} \to \mathbb{C}$ is an* ANTIDERIVATIVE *of $f \colon \mathcal{D} \to \mathbb{C}$ if $F'(z) = f(z)$ for all $z \in \mathcal{D}$.*

3.2.2 Example. **(i)** The function $F(z) = z$ is an antiderivative of the function $f(z) = 1$ on $\mathcal{D} = \mathbb{C}$.

(ii) By the chain rule, the function $F(t) = -ie^{it}$ satisfies $F'(t) = -i^2e^{it} = e^{it}$ on $\mathcal{D} = \mathbb{R}$, and so F is an antiderivative of $f(t) = e^{-it}$.

(iii) The function Log is differentiable on $\mathcal{D} = \mathbb{C} \setminus (-\infty, 0]$ and satisfies

$$\mathrm{Log}'(z) = \frac{1}{z}$$

there. The function $\log_{2\pi}$ is differentiable on $\mathcal{D} = \mathbb{C} \setminus [0, \infty)$ and satisfies

$$\log_{2\pi}'(z) = \frac{1}{z}$$

there. So, the function $f \colon \mathbb{C} \setminus \{0\} \to \mathbb{C} \colon t \mapsto 1/z$ has antiderivatives on two different subsets of $\mathbb{C} \setminus \{0\}$, but neither Log nor $\log_{2\pi}$ is an antiderivative on all of $\mathbb{C} \setminus \{0\}$. We will see that f cannot have an antiderivative on all of $\mathbb{C} \setminus \{0\}$.

When $\mathcal{D} = [a, b] \subseteq \mathbb{R}$ is a real interval, it turns out that there is not much new about antidifferentiation on $\mathcal{D}$; one simply antidifferentiates the real and imaginary parts of $f \colon [a, b] \to \mathbb{C}$. But when $\mathcal{D} \subseteq \mathbb{R}$ is open and therefore genuinely two-dimensional, the antiderivative problem becomes much more surprising, rather like the question of differentiating. (Recall that derivatives of a function defined on an open set $\mathcal{D} \subseteq \mathbb{C}$ are much more interesting than those of functions defined on an interval $[a, b] \subseteq \mathbb{R}$.) We need two tools to resolve the antiderivative problem. We have already mastered the first: paths will play a key role, as we will "integrate over" paths, not just intervals in $\mathbb{R}$. That is, we will study *line integrals*, first for their role as antiderivatives and subsequently for their tremendous value as *instruments* that reveal key features of functions.

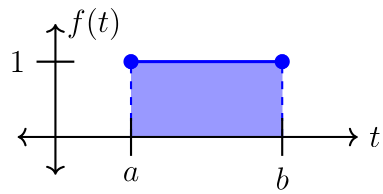
But to construct these line integrals, we need a second tool: a definite integral for functions defined on a closed bounded interval $[a, b] \subseteq \mathbb{R}$ but now taking values in $\mathbb{C}$. We will build this integral out of the ordinary (Riemann) integral.

3.2.1. Properties of "good" integrals.

What is an integral? We will separate the question of what an integral *is* from the question of what an integral *does*. The former can be quite technical to define precisely, but the latter is actually quite simple. Here are four fundamental "behaviors" that a "good" integral should exhibit.

($\int$1) First, the integral of a function $f\colon [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ should somehow measure the net area of the region between the graph of f and the interval $[a,b]$. Since the most fundamental area is the area of a rectangle, we should expect

$$\int_a^b 1\ dt = b - a.$$

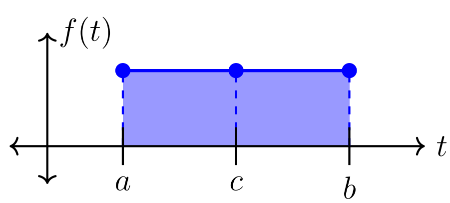


($\int$2) If f is nonnegative, the net area of the region between the graph of f and the interval $[a,b]$ should be the genuine area of the region between the graph of f and the interval $[a,b]$, and this should be a positive quantity. So, we expect that if $0 \le f(t)$ on $[a,b]$, then

$$0 \le \int_a^b f(t)\ dt.$$

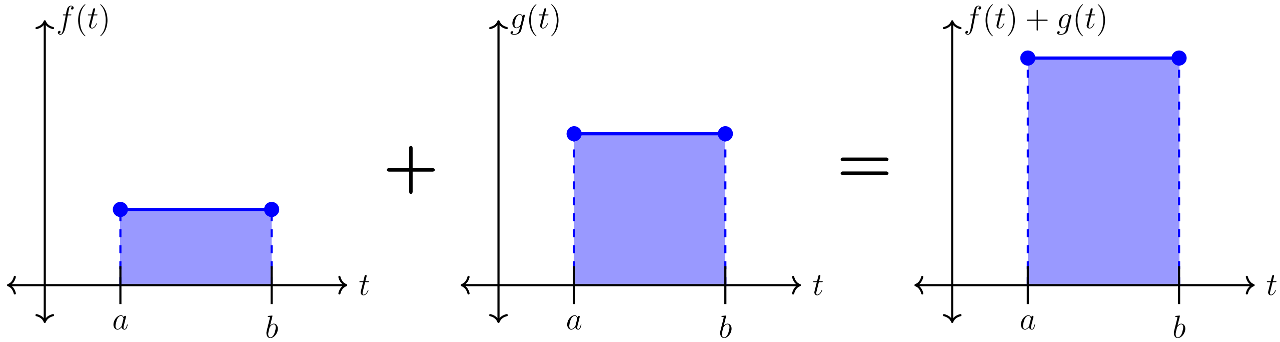
($\int$3) If we divide the region between the graph of f and the interval $[a,b]$ into multiple components, measure the net area of those components, and add those net areas together, we should get the total net area of the region between the graph of f and the interval $[a,b]$. There are many such ways to accomplish this division, but perhaps one of the most straightforward is to split $[a,b]$ up into two or more subintervals and consider the net areas of the regions between the graph of f and those subintervals. So, we expect that if $a \le c \le b$, then

$$\int_a^c f(t)\ dt + \int_c^b f(t)\ dt = \int_a^b f(t)\ dt.$$



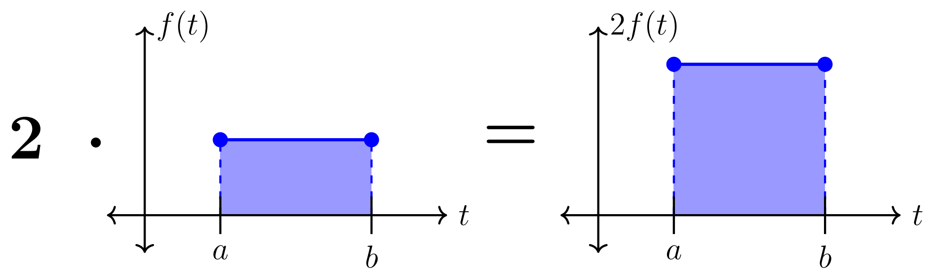
($\int$4) Adding two functions f, $g\colon [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ should "stack" the graphs of f and g on top of each other. Then the region between the graph of f and the interval $[a,b]$ gets "stacked" on top of region between the graph of g and the interval $[a,b]$. Consequently, the net area of the region between the graph of $f+g$ and the interval $[a,b]$ should just be the sum of these two areas:

$$\int_a^b f(t)\ dt + \int_a^b g(t)\ dt = \int_a^b \big[f(t) + g(t)\big]\ dt.$$



Next, multiplying a function $f\colon [a,b] \subseteq \mathbb{R} \to \mathbb{R}$ by a constant $\alpha \in \mathbb{R}$ should somehow "scale" the net area of the region between the graph of f and the interval $[a,b]$ by that factor α. For example, the area under the graph of $2f$ over $[a,b]$ should be double the area under the graph. Consequently, the net area of the region between the graph of αf and the interval $[a,b]$ should be the product

$$\int_a^b \alpha f(t)\ dt = \alpha \int_a^b f(t)\ dt.$$



These four properties are exactly the properties of a "good" integral that we will need—no more, no less. Below, we will assert that we can always integrate continuous functions in a manner consistent with the properties above. Before that, we give in to temptation and drop one aspect of integral notation.

3.2.3 Remark. *Contrary to everything that we are taught in calculus, we will typically not write a variable of integration unless we actually need one for clarification (say, to write out the formula for the integrand explicitly, or when changing variables). That is, we write*

$$\int_a^b f, \quad \text{not} \quad \int_a^b f(t)\ dt \quad \text{or} \quad \int_a^b f(\tau)\ d\tau.$$

However, when we do include the variable of integration, we follow the custom that any variable may be used, thus

$$\int_a^b f = \int_a^b f(t)\ dt = \int_a^b f(\tau)\ d\tau = \int_a^b f(s)\ ds = \cdots .$$

Also, we will use the words "integral" and "definite integral" more or less interchangeably. (Eventually we will meet a "line integral," and we will sometimes call that just an "integral"—we will add the adjectives "definite" or "line" as needed for emphasis.) But we will never use the words "indefinite integral."

Our view of the definite integral will be "dynamic": the integral is characterized by what it does. And integrals act on both integrands and limits of integration.

3.2.4 Theorem. *Let $I \subseteq \mathbb{R}$ be an interval and denote by $\mathcal{C}(I)$ the set of all continuous functions from I to $\mathbb{R}$. There exists a map*

$$\int : \{(f, a, b) \mid f \in \mathcal{C}(I),\ a, b \in I\} \to \mathbb{R} \colon (f, a, b) \mapsto \int_a^b f$$

with the following properties.

($\int$1) [Constants] *If a, $b \in I$, then*

$$\int_a^b 1 = b - a.$$

($\int$2) [Monotonicity] *If $f \in \mathcal{C}(I)$ and a, $b \in I$ with $a \leq b$ and $0 \leq f(t)$ for all $t \in [a, b]$, then*

$$0 \leq \int_a^b f.$$

If in particular $0 < f(t)$ for all $t \in [a, b]$, then

$$0 < \int_a^b f.$$

($\int$3) [Additivity of the domain] *If $f \in \mathcal{C}(I)$ and a, b, $c \in I$, then*

$$\int_a^c f + \int_c^b f = \int_a^b f.$$

($\int$4) [Linearity in the integrand] *If f, $g \in \mathcal{C}(I)$, a, $b \in I$, and $\alpha \in \mathbb{R}$, then*

$$\int_a^b (f + g) = \int_a^b f + \int_a^b g \quad \text{and} \quad \int_a^b \alpha f = \alpha \int_a^b f.$$

The number $\int_a^b f$ is the DEFINITE INTEGRAL OF f FROM a TO b. *Specifically,*

$$\int_a^b f = \lim_{n \to \infty} \frac{b - a}{n} \sum_{k=1}^{n} f\left(a + \frac{k(b - a)}{n}\right). \tag{3.2.1}$$

Properties ($\int 4$) encodes the linearity of the integral as an operator on the integrand with the limits of integration fixed, while property ($\int 3$) is its **ADDITIVITY** over subintervals with the integrand fixed. Property ($\int 2$) encodes the idea that a nonnegative function should have a nonnegative integral, while property ($\int 1$) defines the one value of the integral that it most certainly should have from the point of view of area.

The terms of the sequence on the right of (3.2.1) are the right-endpoint Riemann sums for f over $[a, b]$. Taking this limit as the definition of the integral—and tacitly assuming that the sequence of Riemann sums converges if f is continuous—we can prove properties ($\int 1$), ($\int 2$), and ($\int 4$) quite easily. Property ($\int 3$) is not so obvious from (3.2.1), and in fact this property hinges on expressing $\int_a^b f$ as a "limit" of several kinds of Riemann sums, not just the right-endpoint sum. And then there is still the challenge of ensuring that limits of all sorts of "well-behaved" Riemann sums for f (including, but not limited to, left and right endpoint and midpoint sums) all converge to the same number. Moreover, it is plausible that one might want to integrate functions that are not continuous.

These deeper questions of integration, while tremendously worthwhile, will have no bearing on our further study of complex analysis. We will only need to integrate continuous functions, and we will only need properties ($\int 1$), ($\int 2$), ($\int 3$), and ($\int 4$).

3.2.2. *The definite integral of a complex-valued function.*

So, equipped with the integral for real-valued functions, we turn to the complex-valued case.

3.2.5 Definition. *Let $f\colon I \subseteq \mathbb{R} \to \mathbb{C}$ be continuous and let a, $b \in I$. The* **INTEGRAL** *of f from a to b is*

$$\int_a^b f := \int_a^b \operatorname{Re}(f) + i \int_a^b \operatorname{Im}(f). \tag{3.2.2}$$

3.2.6 Remark. *The terms $(b-a)n^{-1}\sum_{k=1}^n f(a + k(b-a)/n)$ in the Riemann sum limit (3.2.1) are perfectly well-defined for a function $f\colon I \subseteq \mathbb{R} \to \mathbb{C}$, if a, $b \in I$. Thus one could in principle prove Theorem 3.2.4 not assuming that f is only real-valued. There are, however, certain advantages to assuming that f is indeed real-valued—namely, the ability to manipulate inequalities involving Riemann sums.*

The complex-valued integral inherits many properties from the real-valued version.

3.2.7 Problem ($\star$). Let $I \subseteq \mathbb{R}$ be an interval and f, $g\colon I \to \mathbb{C}$ be continuous. Let a, b, $c \in I$ and $\alpha \in \mathbb{C}$. Using only Definition 3.2.5 and Theorem 3.2.4, prove the following.

(i) $\operatorname{Re}\left(\int_a^b f\right) = \int_a^b \operatorname{Re}(f)$ and $\operatorname{Im}\left(\int_a^b f\right) = \int_a^b \operatorname{Im}(f)$

(ii) $\overline{\int_a^b f} = \int_a^b \overline{f}$, where $\overline{f}(t) := \overline{f(t)}$

(iii) [Generalization of ($\int 1$)] $\displaystyle\int_a^b \alpha = \alpha(b-a)$

(iv) [Generalization of ($\int 3$)] $\displaystyle\int_a^c f + \int_c^b f = \int_a^b f$

(v) [Generalization of ($\int 4$)] $\displaystyle\int_a^b (f+g) = \int_a^b f + \int_a^b g$ and $\displaystyle\int_a^b \alpha f = \alpha \int_a^b f$

(vi) $\displaystyle\int_a^a f = 0$

(vii) $\displaystyle\int_a^b f = -\int_b^a f$

3.2.8 Problem ($\star$). Use induction to generalize additivity as follows. Let $I \subseteq \mathbb{R}$ be an interval and $f \colon I \to \mathbb{C}$ be continuous. If $t_0, \ldots, t_n \in I$, then

$$\int_{t_0}^{t_n} f = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} f.$$

Note that Problem 3.2.7 does not discuss the monotonicity of the integral, as inequalities do not make sense for functions that are complex-and-not-real-valued. If we return to real-valued functions, then we can extend monotonicity in a useful way.

3.2.9 Problem ($\star$). Let $I \subseteq \mathbb{R}$ be an interval.

(i) Suppose that f, $g \colon I \to \mathbb{R}$ are continuous and a, $b \in \mathbb{R}$ with $a \le b$. If $f(t) \le g(t)$ for all $t \in [a, b]$, show that

$$\int_a^b f \le \int_a^b g. \tag{3.2.3}$$

(ii) Suppose that $f \colon I \to \mathbb{R}$ is continuous and there are m, $M \in \mathbb{R}$ such that $m \le f(t) \le M$ for all $t \in [a, b]$. Show that

$$m(b-a) \le \int_a^b f \le M(b-a). \tag{3.2.4}$$

A double application of (3.2.3) yields one of the most important estimates on integrals possible.

3.2.10 Theorem (Real triangle inequality). *Let $I \subseteq \mathbb{R}$ be an interval, let $f \colon I \to \mathbb{R}$ be* *continuous, and let a, $b \in I$ with $a \le b$. Then*

$$\left|\int_a^b f\right| \le \int_a^b |f|. \tag{3.2.5}$$

Proof. Use the inequalities $-|f(t)| \le f(t) \le |f(t)|$ and (3.2.3) to find

$$\int_a^b -|f| \le \int_a^b f \le \int_a^b |f|.$$

By linearity, this is

$$-\int_a^b |f| \le \int_a^b f \le \int_a^b |f|,$$

and by properties of absolute value, this is equivalent to (3.2.5). ■

The triangle inequality is also true in the complex-valued setting, but it needs a new proof, since the proof of Theorem 3.2.10 used monotonicity.

3.2.11 Theorem (Complex triangle inequality). *Let $I \subseteq \mathbb{R}$ be an interval, let $f \colon I \to \mathbb{C}$ be continuous, and let a, $b \in I$ with $a \le b$. Then*

$$\left|\int_a^b f\right| \le \int_a^b |f|.$$

3.2.12 Problem ($\star$). Prove this theorem as follows.

(i) First explain why the inequality is immediately true if $\int_a^b f = 0$.

(ii) Suppose, then, in the following that $\int_a^b f \ne 0$. Use Problem 1.6.4 to obtain the inequality

$$\left|\int_a^b f\right| \le \left|\operatorname{Re}\left(e^{-i\theta}\int_a^b f\right)\right|. \tag{3.2.6}$$

(iii) Justify each of the following with a reference to some result about integrals from this section:

$$e^{-i\theta}\int_a^b f = \int_a^b e^{-i\theta} f, \tag{3.2.7}$$

$$\operatorname{Re}\left(\int_a^b e^{-i\theta} f\right) = \int_a^b \operatorname{Re}(e^{-i\theta} f), \tag{3.2.8}$$

and

$$\left|\int_a^b \operatorname{Re}(e^{-i\theta} f)\right| \le \int_a^b |\operatorname{Re}(e^{-i\theta} f)|. \tag{3.2.9}$$

(iv) Explain why $|\operatorname{Re}(e^{-i\theta}f(t))| \leq |f(t)|$ and conclude

$$\int_a^b |\operatorname{Re}(e^{-i\theta}f)| \leq \int_a^b |f|. \tag{3.2.10}$$

(v) Put it all together to get the complex triangle inequality.

3.2.13 Problem (!). **(i)** Show that if we remove the hypothesis $a \leq b$, then the triangle inequality becomes

$$\left|\int_a^b f\right| \leq \left|\int_a^b |f|\right|.$$

Why is the extra modulus on the right necessary here?

(ii) Suppose that $f\colon I \to \mathbb{R}$ is continuous and $a,\, b \in I$. Show that

$$\left|\int_a^b f\right| \leq |b-a| \max_{0 \leq t \leq 1} |f((1-t)a+tb)|. \tag{3.2.11}$$

[Hint: *the annoying part about this problem is that we are not specifying $a \leq b$. Prove the two cases $a \leq b$ and $b \leq a$ separately. Use the fact that if $J := \{(1-t)a+tb \mid 0 \leq t \leq 1\}$, then $J = [a,b]$ if $a \leq b$, and $J = [b,a]$ if $b \leq a$.*]

3.2.3. The fundamental theorem of calculus.

We now have only a handful of results about the definite integral, and yet they are enough to prove the fundamental theorem of calculus. (Conversely, by themselves, they do not help us evaluate integrals more complicated than $\int_a^b \alpha$ for $\alpha \in \mathbb{C}$!) This is our first rigorous verification that an integral gives a meaningful representation of a function. Specifically, integrals represent antiderivatives.

3.2.14 Theorem (FTC1). *Let $f\colon I \to \mathbb{C}$ be continuous and fix $a \in I$. Define*

$$F\colon I \to \mathbb{C}\colon t \mapsto \int_a^t f$$

Then F is an antiderivative of f on I.

This is where we finished on Monday, February 26, 2024.

Proof. Fix $t \in I$. We need to show that F is differentiable at t with $F'(t) = f(t)$. That is, we want

$$\lim_{h \to 0} \frac{F(t+h)-F(t)}{h} = f(t),$$

equivalently,

$$\lim_{h\to 0} \frac{F(t+h) - F(t) - hf(t)}{h} = 0.$$

We first compute

$$\begin{aligned} F(t+h) - F(t) &= \int_a^{t+h} f(\tau)\ d\tau - \int_a^t f(\tau)\ d\tau \\ &= \int_a^{t+h} f(\tau)\ d\tau + \int_t^a f(\tau)\ d\tau \\ &= \int_t^{t+h} f(\tau)\ d\tau. \end{aligned}$$

Next,

$$hf(t) = f(t)[(t+h) - t] = f(t)\int_t^{t+h} 1\ d\tau = \int_t^{t+h} f(t)\ d\tau.$$

We then have

$$F(t+h) - F(t) - hf(t) = \int_t^{t+h} f(\tau)\ d\tau - \int_t^{t+h} f(t)\ d\tau = \int_t^{t+h} \big[f(\tau) - f(t)\big]\ d\tau.$$

Note that this is one instance in which using the variable of integration τ clarifies the fact that t is constant here. It therefore suffices to show that

$$\lim_{h\to 0} \frac{1}{h} \int_t^{t+h} \big[f(\tau) - f(t)\big]\ d\tau = 0, \tag{3.2.12}$$

and we do that in the following lemma. ■

Before stating and proving the lemma, we discuss why we should expect (3.2.13) to be true. We are sending $h \to 0$ and dividing by h, so the factor $1/h$ will be "large." Thus the integral $\int_t^{t+h} \big[f(\tau) - f(t)\big]\ d\tau$ needs to be "small" as $h \to 0$ to counterbalance this division by a "large" number; in fact, this integral needs to be "very small," since we want the limit to be 0, not just a finite number. However, this integral is indeed "very small" because it is "small" in *two* places. First, taking $h \to 0$ means that the limits of integration t and $t+h$ are "close," and so the interval of integration is "small." Second, because these limits of integration are close, each $\tau \in [t, t+h]$ is "close" to t, and thus, by continuity, $f(\tau)$ and $f(t)$ are "close," thus the integrand is "small."

3.2.15 Lemma. *Let $I \subseteq \mathbb{R}$ be an interval and let $f\colon I \to \mathbb{C}$ be continuous. Then*

$$\lim_{h\to 0} \frac{1}{h} \int_t^{t+h} \big[f(\tau) - f(t)\big]\ d\tau = 0$$

for any $t \in I$.

Proof. We use the squeeze theorem. Problem 3.2.13 gives us the estimate

$$\left|\frac{1}{h}\int_t^{t+h} \left[f(\tau) - f(t)\right]\, d\tau\right| \le \frac{1}{|h|}|t+h-h| \max_{0\le s\le 1} |f((1-s)t+s(t+h))-f(t)| = \max_{0\le s\le 1} |f(t+sh)-f(t)|.$$

We now need to show that

$$\lim_{h\to 0}\max_{0\le s\le 1} |f(t+sh)-f(t)| = 0.$$

This will involve the definition of continuity.

Let $\epsilon > 0$, so our goal is to find $\delta > 0$ such that if $0 < |h| < \delta$, then

$$\max_{0\le s\le 1} |f(t+sh)-f(t)| < \epsilon. \tag{3.2.13}$$

Since f is continuous at t, there is $\delta > 0$ such that if $|t-\tau| < \delta$, then $|f(\tau) - f(t)| < \epsilon$. Suppose $0 < |h| < \delta$. Then if $0 \le s \le 1$, we have

$$|(t+sh) - t| = |sh| \le |h| < \delta,$$

thus (3.2.13) holds. ■

3.2.16 Problem (!). Prove that the left limit in (3.2.13) holds. What specific changes are needed when $h < 0$?

With FTC1, we can prove a second version that facilitates the computation of integrals via antiderivatives.

3.2.17 Corollary (FTC2). *Let $I \subseteq \mathbb{R}$ be an interval and let $f\colon I \to \mathbb{C}$ be continuous. If F is any antiderivative of f on I, then*

$$\int_a^b f = F(b) - F(a)$$

for all a, $b \in I$.

Proof. Let $F_\star(t) = \int_a^t f$, so $F_\star$ is also an antiderivative of f. Put $h = F_\star - F$, so $h' = 0$ on I. Problem 2.6.25 implies that h is constant, say, $h(t) = h(a)$ for all t. Then $F_\star(t) = F(t) + h(a)$ for all t, so

$$\int_a^b f = F_\star(b) = h(b) + F(b) = h(a) + F(b) = F_\star(a) - F(a) + F(b) = F(b) - F(a)$$

since $F_\star(a) = 0$. ■

3.2.18 Example. Let $k \in \mathbb{Z} \setminus \{0\}$. Since $F(t) = e^{it}/ik$ is an antiderivative of $f(t) = e^{ikt}$, we have

$$\int_0^{2\pi} e^{ikt}\,dt = \frac{e^{ikt}}{ik}\bigg|_{t=0}^{t=2\pi} = \frac{e^{2\pi ik} - e^{0\cdot k}}{i} = \frac{1-1}{i} = 0.$$

3.2.19 Problem (+). We can use the definite integral to give an explicit definition of the natural logarithm. Recall that we originally defined $\ln\colon (0,\infty) \to \mathbb{R}$ as the unique map satisfying $\exp(\ln(t)) = t$ for all $t \in \mathbb{R}$. Such a map exists since exp is one-to-one on $\mathbb{R}$. However, we did not have an explicit formula for ln in the same way that we did for the exponential as a power series. Now we can develop a formula for the natural logarithm using our tried-and-true philosophy that mathematical objects are defined by *what they do.*

(i) We expect (from Example 2.5.18) that ln satisfies

$$\begin{cases} \ln'(t) = 1/t \\ \ln(1) = 0. \end{cases}$$

Use FTC2 to conclude that we should have

$$\ln(t) = \int_1^t \frac{d\tau}{\tau}.$$

(ii) We can start from this integral representation of the natural logarithm and obtain an inverse for the exponential. Put

$$L\colon (0,\infty) \to \mathbb{R}\colon t \mapsto \int_1^t \frac{d\tau}{\tau}.$$

Our goal is to show that $e^{L(t)} = t$ for all $t > 0$. One way to do this is to show that $e^{-L(t)}t = 1$ for all t, and one way to show *that* is to define $f(t) := e^{-L(t)}t$ and to show that f solves the initial value problem

$$\begin{cases} f'(t) = 0 \\ f(1) = 1 \end{cases}$$

and then use FTC2. Do just that.

We will eventually develop a similar integral representation for the principal logarithm and, in the process, obtain an explicit formula for the principal argument that does not rely on classical trigonometry. But first we will need a notion of integral for functions of a complex variable.

3.2.4. Consequences of the fundamental theorem of calculus.

The fundamental theorems of calculus are, of course, the keys to both substitution and integration by parts, two of the most general techniques for evaluating integrals in terms of simpler functions.

3.2.20 Theorem (Substitution). *Let I, $J \subseteq \mathbb{R}$ be intervals with a, $b \in J$. Let $\varphi\colon J \to I$ be continuously differentiable and let $f\colon I \to \mathbb{C}$ be continuous. Then*

$$\int_a^b (f \circ \varphi)\varphi' = \int_{\varphi(a)}^{\varphi(b)} f.$$

Proof. Put

$$F(t) := \int_{\varphi(a)}^{t} f,$$

so $F'(t) = f(t)$ for all $t \in I$ by FTC1. Next, put

$$G(t) := F(\varphi(t)),$$

so

$$G'(t) = F'(\varphi(t))\varphi'(t) = f(\varphi(t))\varphi'(t).$$

That is, G is an antiderivative of $(f \circ \varphi)\varphi'$, and so

$$\int_a^b (f \circ \varphi)\varphi' = G(b) - G(a)$$

by FTC2. But

$$G(b) - G(a) = F(\varphi(b)) - F(\varphi(a)) = \int_{\varphi(a)}^{\varphi(b)} f - \int_{\varphi(a)}^{\varphi(a)} f = \int_{\varphi(a)}^{\varphi(b)} f. \qquad \blacksquare$$

3.2.21 Example. We revisit the integral of Example 3.2.18. Let $k \in \mathbb{Z} \setminus \{0\}$ and put $\varphi(\tau) = k\tau$. Then

$$\begin{aligned}\int_0^{2\pi} e^{ik\tau}\, d\tau = \frac{1}{k}\int_0^{2\pi} e^{ik\tau}k\, d\tau = \frac{1}{k}\int_0^{2\pi} e^{i\varphi(\tau)}\varphi'(\tau)\, d\tau = \frac{1}{k}\int_{\varphi(0)}^{\varphi(2\pi)} e^{it}\, dt = \frac{1}{k}\int_0^{2k\pi} e^{it}\, dt \\ = \frac{1}{ik}e^{it}\Big|_{t=0}^{t=2\pi} = \frac{e^{2\pi i} - e^0}{ik} = \frac{1-1}{ik} = 0.\end{aligned}$$

Note that taking $\varphi(\tau) = ik\tau$ does not work, as then φ is not real-valued, which is essential in the substitution theorem.

3.2.22 Problem ($\star$). Suppose that $f\colon \mathbb{R} \to \mathbb{C}$ is continuous and p-**PERIODIC** for some $p \in \mathbb{R}$, in the sense that $f(t + p) = f(t)$ for all $t \in \mathbb{R}$. Then the integral of f over any interval of length p is the same:

$$\int_a^{a+p} f = \int_0^p f$$

for all $a \in \mathbb{R}$. Give two proofs of this identity as follows.

(i) Define

$$F(a) := \int_a^{a+p} f$$

and use FTC1 and the p-periodicity of f to show that $F'(a) = 0$ for all a. Then F is constant by Problem 2.6.25, so $F(a) = F(0)$ for all a.

(ii) First explain why

$$\int_a^{a+p} f = \int_0^p f + \left(\int_p^{a+p} f - \int_0^a f\right).$$

Then substitute $u = t - p$ to show

$$\int_p^{a+p} f = \int_0^a f(t-p)\, dt$$

and use the p-periodicity of f.

3.2.23 Problem (+). We can use integrals to prove the familiar limit

$$e = \lim_{t\to\infty} \left(1 + \frac{1}{t}\right)^t.$$

Here we interpret the power as the principal power.

(i) Use properties of the logarithm and exponential to conclude that the desired limit holds if

$$\lim_{t\to\infty} t \int_t^{t+1} \frac{d\tau}{\tau} = 1.$$

(ii) Change variables to show that

$$\int_t^{t+1} \frac{d\tau}{\tau} = \int_0^1 \frac{d\tau}{t+\tau}.$$

(iii) Rewrite

$$\left| t \int_t^{t+1} \frac{d\tau}{\tau} - 1 \right| = \left| \int_0^1 \frac{\tau}{t+\tau}\, d\tau \right|,$$

estimate the integral on the right, and use the squeeze theorem to obtain

$$\lim_{t\to\infty} \left| \int_0^1 \frac{\tau}{t+\tau}\, d\tau \right| = 0.$$

This is where we finished on Wednesday, February 28, 2024.

A recurring theme of our subsequent applications of integrals will be that we are trying to estimate or control some kind of difference (this is roughly 90% of analysis), and it turns out to be possible to rewrite that difference in a tractable way by introducing an integral. It may be possible to manipulate further terms under consideration by rewriting them as integrals, too. The fundamental identity that we will use in the future is (3.2.14) below.

3.2.24 Example. FTC2 allows us to rewrite a functional difference as an integral. When we incorporate substitution, we can get a very simple formula for that difference. Suppose that $I \subseteq \mathbb{R}$ is an interval, $f \colon I \to \mathbb{C}$ is continuously differentiable, and t, $t+h \in I$. Then

$$f(t+h) - f(t) = \int_t^{t+h} f'.$$

We will reverse engineer substitution and make the limits of integration simpler and the integrand more complicated. (This turns out to be a good idea.) Specifically, we are integrating over the interval $[t, t+h]$, and we recall its parametrization as a path: put

$$\varphi \colon [0,1] \to \mathbb{R} \colon \tau \mapsto= (1-\tau)t + \tau(t+h) = t + h\tau.$$

Then $\varphi'(\tau) = h$, $\varphi(0) = t$, and $\varphi(1) = t+h$, so substitution implies

$$\int_t^{t+h} f' = \int_0^1 f'(t+h\tau)h \; d\tau.$$

Thus if f is differentiable and f' continuous on an interval containing t and $t+h$, then

$$f(t+h) - f(t) = h \int_0^1 f'(t+h\tau) \; d\tau. \tag{3.2.14}$$

This equality allows us to control the distance between $f(t+h)$ and $f(t)$ using the explicit factor of h on the right above and the triangle inequality on the integral with the constant limits of 0 and 1. In particular, knowing the size of f' controls the difference. We could obtain a similar result from the mean value theorem (at least, if f is real-valued), but the explicit formula (3.2.14) eliminates a possibly vague "existential" result from the MVT.

This identity can be generalized to partial derivatives, e.g., if $f = f(t,s)$ is differentiable with respect to t and f_t is continuous, then

$$f(t+h,s) - f(t,s) = h \int_0^1 f_t(t+\tau h, s) \; d\tau.$$

3.2.25 Problem (!). Prove the following variant of Example 3.2.24: if $I \subseteq \mathbb{R}$ is an interval, $f \colon I \to \mathbb{C}$ is continuously differentiable, and a, $b \in I$, then

$$f(b) - f(a) = (b-a) \int_0^1 f'(a + t(b-a)) \; dt.$$

Dividing by $b-a$, this gives us a variant of the mean value theorem for complex-valued functions, as the "ordinary" mean value theorem need not hold for complex, nonreal-valued functions (recall Problem 2.6.24).

Integration by parts works nicely for complex-valued functions of a real variable, because the product rule, the FTC, and antiderivatives work as we think they should in this setting.

3.2.26 Theorem (Integration by parts). *Suppose that f, $g\colon [a,b] \to \mathbb{C}$ are differentiable with f', g' continuous. Then*

$$\int_a^b fg' = f(b)g(b) - f(a)g(a) - \int_a^b f'g. \tag{3.2.15}$$

Proof. Put $H(t) = f(t)g(t)$, so the product rule (for complex-valued functions of a real variable) gives

$$H'(t) = f'(t)g(t) + f(t)g'(t).$$

Then FTC2 gives

$$\int_a^b H' = H(b) - H(a) = f(b)g(b) - f(a)g(a). \tag{3.2.16}$$

But by linearity

$$\int_a^b H' = \int_a^b f'g - \int_a^b fg', \tag{3.2.17}$$

and so (3.2.15) follows by equating (3.2.16) and (3.2.17). ■

3.2.5. Arc length.

We have said that integrals are tools for representing functions and extracting data about functions. So far, we have seen two related representations of functions as integrals: representing an antiderivative of f as the integral $F(t) = \int_a^t f$, and representing the difference $f(t+h) - f(t)$ as the convenient expression $h\int_0^1 f'(t+h\tau)\, d\tau$. Now we perform an elementary data extraction with integrals.

In calculus we learned that if $f\colon [a,b] \to \mathbb{R}$ is continuously differentiable, then, by a limiting argument with Riemann sums, the integral

$$\int_a^b |f'(t)|\, dt$$

captures the natural notion of the "length" of the graph of f. We import this concept to paths.

3.2.27 Definition. *The* ARC LENGTH *of a smooth path $\gamma\colon [a,b] \subseteq \mathbb{R} \to \mathbb{C}$ is*

$$\ell(\gamma) := \int_a^b |\gamma'(t)|\, dt.$$

3.2.28 Example. The smooth path $\gamma\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto e^{it}$ parametrizes the unit circle, and of course we expect its arc length to be 2π. We check this:

$$\ell(\gamma) = \int_0^{2\pi} |\gamma'(t)|\, dt = \int_0^{2\pi} |ie^{it}|\, dt = \int_0^{2\pi} 1\, dt = 2\pi,$$

as expected.

If we represent a path γ as the composition of smooth paths $\gamma = \oplus_{k=1}^n \gamma_k$, then we should expect that the sum $\sum_{k=1}^n \ell(\gamma_k)$ gives a meaningful measurement of the arc length of γ. However, what if we have another representation of γ as $\gamma = \oplus_{j=1}^m \mu_j$, where each μ_j is also a smooth path? Do we have $\sum_{k=1}^n \ell(\gamma_k) = \sum_{j=1}^m \ell(\mu_j)$? This is a question of the "well-definedness" of arc length for (nonsmooth) paths, and it is the sort of problem that arises whenever we define a quantity in terms of a chosen "representation" of an object. Is the quantity independent of the chosen representation? Of course, yes.

We prove the following theorem in Appendix C.2.

3.2.29 Theorem. *Arc length is well-defined in the sense that if a path γ can be expressed as compositions of smooth paths via*

$$\gamma = \oplus_{k=1}^n \gamma_k \quad \text{and} \quad \gamma = \oplus_{j=1}^m \mu_j,$$

then

$$\sum_{k=1}^n \ell(\gamma_k) = \sum_{j=1}^m \ell(\mu_j).$$

3.2.30 Problem (!). **(i)** Let $k \in \mathbb{Z}$ and define $\gamma_k\colon [0, 2\pi] \subseteq \mathbb{R} \to \mathbb{C}\colon t \mapsto e^{ikt}$. What is $\ell(\gamma_k)$? Is this what you expected?

(ii) What is the arc length of a line segment? Is it what you expected?

(iii) Suppose that γ_1 and γ_2 are equivalent paths; for simplicity, assume that both are continuously differentiable. Show that $\ell(\gamma_1) = \ell(\gamma_2)$. Does the positivity of φ' from Definition 3.1.35 matter here? How does this compare to part (iv) of Problem 3.3.9?

3.3. Line integrals.

We extend the definite integral to functions of a complex variable as a line integral. While there is some reasonable motivation for the following definition as a "limit" of certain Riemann sums, we do not consider that. Instead, we take the position that the line integral is the most natural generalization of the definite integral that is also the best instrument for extracting critical information about functions, although its full utility will not be apparent for some time.

3.3.1. Definition and properties of line integrals.

3.3.1 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ and let $f \colon \mathcal{D} \to \mathbb{C}$ be continuous. Let $\gamma \colon [a,b] \subseteq \mathbb{R} \to \mathcal{D}$ be a smooth path. Then the* **LINE INTEGRAL OF f OVER γ** *is*

$$\int_\gamma f = \int_\gamma f(z)\ dz := \int_a^b f(\gamma(t))\gamma'(t)\ dt = \int_a^b (f \circ \gamma)\gamma'. \tag{3.3.1}$$

3.3.2 Remark. *The integrand in (3.3.1) is the product $(f \circ \gamma)\gamma'$. This is a continuous function since γ is continuously differentiable. Thus Definition 3.2.5 applies.*

As with definite integrals, we will often omit the variable of integration in line integrals and include it for clarity when necessary. When we do include it, we continue the custom that we can change the symbol at will:

$$\int_\gamma f = \int_\gamma f(z)\ dz = \int_\gamma f(w)\ dw = \int_\gamma f(\xi)\ d\xi = \cdots.$$

We will frequently integrate over lines and circles, and so the following two examples contain extremely important calculations.

3.3.3 Example. Parametrize the line segment $[0, i]$ by

$$\gamma \colon [0,1] \subseteq \mathbb{R} \to \mathbb{C} \colon t \mapsto (1-t)0 + ti = it.$$

Then $\gamma'(t) = i$ for all t. The function $f(z) := \overline{z}$ is continuous on $\mathbb{C}$, and so we may compute

$$\int_\gamma \overline{z}\ dz = \int_0^1 \overline{\gamma(t)}\gamma'(t)\ dt = \int_0^1 \overline{it}(i)\ dt = \int_0^1 -it(i)\ dt = \int_0^1 t\ dt = \frac{1}{2}.$$

3.3.4 Example. Let $z_0 \in \mathbb{C}$, $r > 0$, and $n \in \mathbb{Z}$ and parametrize the circle of radius r centered at z_0 by $\gamma(t) := z_0 + re^{it}$ for $t \in [0, 2\pi]$. Then $\gamma'(t) = ire^{it}$, and so

$$\int_\gamma (z-z_0)^n\ dz = \int_0^{2\pi} \big((z_0+re^{it})-z_0\big)^n (ire^{it})\ dt = ir\int_0^{2\pi} (re^{it})^n e^{it}\ dt = ir^{n+1}\int_0^{2\pi} e^{i(n+1)t}\ dt.$$

If $n = -1$, then

$$\int_\gamma \frac{dz}{z-z_0} = i\int_0^{2\pi} 1\ dt = 2\pi i.$$

If $n \neq -1$, then since $F(t) := e^{i(n+1)t}/(i(n+1))$ is an antiderivative of $f(t) := e^{i(n+1)t}$, we have

$$\int_\gamma (z-z_0)^n\ dz = ir^{n+1}\frac{e^{i(n+1)t}}{i(n+1)}\bigg|_{t=0}^{t=2\pi} = ir^{n+1}\left(\frac{1-1}{i(n+1)}\right) = 0.$$

Here it is essential that $n \in \mathbb{Z}$ for the expression $(z-z_0)^n$ to be unambiguously defined when $z \neq z_0$, and for the antiderivatives to work out correctly.

3.3.5 Remark. *Since we will integrate over line segments and circles so often, we will use a special, suggestive notation for their line integrals that will relieve us from writing out their parameterizations each time. Assume below that f is continuous at least on the path(s) over which the integration takes place.*

(i) *For z_1, $z_2 \in \mathbb{C}$, define*

$$\int_{[z_1,z_2]} f(z)\ dz := (z_2 - z_1) \int_0^1 f((1-t)z_1 + tz_2)\ dt.$$

This line integral is oriented "from z_1 to z_2."

(ii) *For $z_0 \in \mathbb{C}$ and $r > 0$, define*

$$\int_{|z-z_0|=r} f(z)\ dz := ir \int_0^{2\pi} f(z_0 + re^{it})e^{it}\ dt.$$

This line integral is oriented with the circle traversed "counterclockwise." (As needed, we may change the variable z in $|z - z_0| = r$ to w or some other symbol.)

In particular, the previous example shows

$$\int_{|z-z_0|=r} (z - z_0)^n\ dz = \begin{cases} 0, & n \in \mathbb{Z} \setminus \{-1\} \\ 2\pi i, & n = -1. \end{cases}$$

3.3.6 Problem (!). Let $I \subseteq \mathbb{R}$ be an interval and let $f \colon I \to \mathbb{R}$ be continuous. Show that for any a, $b \in I$, we have

$$\int_{[a,b]} f = \int_a^b f,$$

where the integral on the left is the line integral over the path $[a, b]$, and the integral on the right is the ordinary Riemann integral.

3.3.7 Problem ($\star$). **(i)** Let $f \colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ be continuous and let $z_0 \in \mathcal{D}$, $r > 0$ with $\overline{\mathcal{B}}(z_0; r) \subseteq \mathcal{D}$. Show that

$$\int_{|z-z_0|=r} f(z)\ dz = \int_{|z|=r} f(z + z_0)\ dz.$$

(ii) Find a continuous function $g \colon \overline{\mathcal{B}}(0; 1) \to \mathbb{C}$ such that

$$\int_{|z-z_0|=r} f(z)\ dz = \int_{|z|=1} g(z).$$

[Hint: *this function g will depend on both z_0 and r.*]

Now we extend our definition of the line integral to all paths, not just smooth ones. Recall that any path can be written as the composition of smooth paths, i.e., if γ is a path, then there are smooth paths $\gamma_1, \ldots, \gamma_n$ such that $\gamma = \oplus_{k=1}^n \gamma_k$. Naturally, then, we want to put

$$\int_\gamma f := \sum_{k=1}^n \int_{\gamma_k} f.$$

However, as with arc length, there is a question of well-definedness. That is, if we also have $\gamma = \oplus_{j=1}^m \mu_j$ for some smooth paths μ_j, we need to check that

$$\sum_{k=1}^n \int_{\gamma_k} f = \sum_{j=1}^m \int_{\mu_j} f.$$

3.3.8 Problem (+). Check that. [Hint: *adapt the proof of Theorem 3.2.29 given in Appendix C.2. The necessary changes mostly involve the integrands, not the limits of integration or the specific arrangement of the integrals in that proof.*]

The line integral enjoys mostly obvious properties that generalize those of the definite integral.

3.3.9 Problem ($\star$). Let $\mathcal{D} \subseteq \mathbb{C}$ and let $f: \mathcal{D} \to \mathbb{C}$ be continuous. The following results hold for all paths, but in your work you may assume that the paths are smooth. In the context of Problem 3.3.6, how do parts (i), (ii), and (iii) below generalize results from Problem 3.2.7?

(i) Let γ_1 and γ_2 be paths in $\mathcal{D}$ and suppose that the terminal point of γ_1 is the initial point of γ_2. Show that

$$\int_{\gamma_1 \oplus \gamma_2} f = \int_{\gamma_1} f \oplus \int_{\gamma_2} f.$$

(ii) Let γ be a path in $\mathcal{D}$. Show that

$$\int_{\gamma^-} f = -\int_\gamma f.$$

(iii) Let γ be a path in $\mathcal{D}$, let $g: \mathcal{D} \to \mathbb{C}$ also be continuous, and let $\alpha \in \mathbb{C}$. Show that

$$\int_\gamma (f+g) = \int_\gamma f + \int_\gamma g \quad \text{and} \quad \int_\gamma \alpha f = \alpha \int_\gamma f.$$

(iv) Show that if γ_1 and γ_2 are equivalent paths in $\mathcal{D}$, i.e., γ_1 is a reparametrization of γ_2, then

$$\int_{\gamma_1} f = \int_{\gamma_2} f.$$

Does the positivity of φ' as in Definition 3.1.35 really matter here?

3.3.10 Problem ($\star$). Let $\mathcal{D} \subseteq \mathbb{C}$ and let $f\colon \mathcal{D} \to \mathbb{C}$ be continuous.

(i) Let γ be a path in $\mathcal{D}$. What is the value of

$$\int_{\gamma \oplus \gamma^-} f?$$

Does this remind you of a result from Problem 3.2.7?

(ii) Fix $z_0 \in \mathcal{D}$ and let γ be the "constant" path $\gamma\colon [a,b] \subseteq \mathbb{R} \to \mathcal{D}\colon t \mapsto z_0$. What is the value of

$$\int_\gamma f?$$

Does this remind you of a result from Problem 3.2.7?

(iii) Explain why we should expect, in general, that

$$\overline{\int_\gamma f} \neq \int_\gamma \overline{f},$$

and give a specific example of f and γ for which the equality does not hold.

3.3.2. *The fundamental theorem of calculus for line integrals.*

The fundamental theorem of calculus nicely extends to line integrals and thereby generalizes the FTC for definite integrals.

3.3.11 Theorem (FTC for line integrals). *Let $\mathcal{D} \subseteq \mathbb{C}$ and $f\colon \mathcal{D} \to \mathbb{C}$ be continuous. Suppose that $F\colon \mathcal{D} \to \mathbb{C}$ is an antiderivative of f. Then if $\gamma\colon [a,b] \subseteq \mathbb{R} \to \mathcal{D}$ is a path,*

$$\int_\gamma f = F(\gamma(b)) - F(\gamma(a)).$$

Proof. We only give the proof in the special case that γ is smooth. Then

$$\int_\gamma f = \int_a^b (f \circ \gamma)\gamma' = \int_a^b (F \circ \gamma)' = F(\gamma(b)) - F(\gamma(a)).$$

Otherwise, express γ as the composition of smooth paths, apply the result just proved to each of those paths, add, and simplify using the **TELESCOPING** identity $\sum_{k=m}^n (z_{k+1} - z_k) = z_{n+1} - z_m$, which is valid for any set $\{z_k\}_{k=m}^{n+1} \subseteq \mathbb{C}$. ■

3.3.12 Problem (!). Redo as much as possible of Example 3.3.4 using the FTC for line integrals. What goes wrong at $n = -1$?

This is where we finished on Friday, March 1, 2024.

3.3.13 Example. Define $\gamma\colon [0,\pi] \to \mathbb{C}\colon t \mapsto e^{it}$, so γ is the upper half of the unit circle. We study

$$\int_\gamma \frac{dz}{z}.$$

(i) There is no need for any fancy machinery here, as we could just use the definition of the line integral:

$$\int_\gamma \frac{dz}{z} = \int_0^\pi \frac{\gamma'(t)}{\gamma(t)}\, dt = \int_0^\pi \frac{ie^{it}}{e^{it}}\, dt = \int_0^\pi i\, dt = i\pi.$$

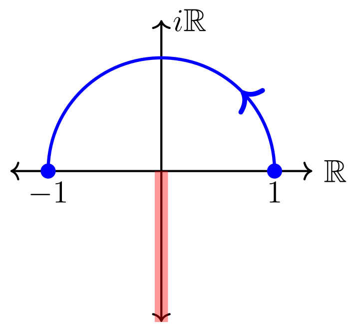
(ii) However, it may be instructive to see some fancy machinery. Any branch of the logarithm worth its salt should be an antiderivative of the function $f(z) := 1/z$. However, f can be defined on $\mathbb{C} \setminus \{0\}$, but we know from Problem 3.3.14 that f cannot have an antiderivative on $\mathbb{C} \setminus \{0\}$. Rather, for any $\alpha \in \mathbb{R}$, the branch $\log_\alpha$ is an antiderivative of f on $\mathbb{C} \setminus \{re^{i\alpha} \mid r \geq 0\}$. If we want to evaluate a line integral of f using the fundamental theorem of calculus, we will need to be careful with both how we view the domain of f and what antiderivative we choose.

Our instinct is probably to work with the principal logarithm, since we have $\operatorname{Log}'(z) = 1/z$, and so we expect

$$\int_\gamma \frac{dz}{z} = \operatorname{Log}(-1) - \operatorname{Log}(1) = i\pi. \tag{3.3.2}$$

This certainly agrees with the work above using the definition, but our reasoning is wrong. The principal logarithm is not differentiable on $(-\infty, 0]$, and so in particular Log is not differentiable at -1. We cannot invoke the fundamental theorem of calculus for line integrals. (Specifically, the problem is that we cannot choose a set $\mathcal{D}$ such that both $\gamma(t) \in \mathcal{D}$ for all $t \in [0,\pi]$ and Log is differentiable on $\mathcal{D}$.)

(iii) Instead of Log, we could choose a branch of the logarithm whose branch cut does not intersect the image of γ, for example, $\log_{-\pi/2}$, whose branch cut is the negative imaginary axis.



Then

$$\int_\gamma \frac{dz}{z} = \log_{-\pi/2}(-1) - \log_{-\pi/2}(1) = i \arg_{-\pi/2}(-1) - i \arg_{-\pi/2}(1).$$

Recall that the values $\arg_{-\pi/2}(z)$ need to satisfy

$$-\frac{\pi}{2} < \arg_{-\pi/2}(z) < \frac{3\pi}{2} \qquad \text{and} \qquad z = |z|e^{i \arg_{-\pi/2}(z)}.$$

Since $-1 = e^{i\pi}$ and $1 = e^{i\cdot 0}$ with $-1, 0 \in (-\pi/2, 3\pi/2)$, we have $\arg_{-\pi/2}(-1) = \pi$ and $\arg_{-\pi/2}(1) = 0$, thus

$$\int_\gamma \frac{dz}{z} = i\pi,$$

which is what we expected from the definition.

3.3.14 Problem (!). **(i)** Let $\mathcal{D} \subseteq \mathbb{C}$ and $f \colon \mathcal{D} \to \mathbb{C}$ be continuous. Show that if f has an antiderivative on $\mathcal{D}$ and γ is a closed path in $\mathcal{D}$, then

$$\int_\gamma f = 0.$$

(ii) Compute

$$\int_{|z|=1} \overline{z}\, dz.$$

Does $f(z) = \overline{z}$ have an antiderivative on $\mathcal{D} = \mathbb{C}$?

(iii) Compute

$$\int_{|z|=1} \frac{dz}{z}.$$

Does $f(z) = 1/z$ have an antiderivative on $\mathcal{D} = \mathbb{C} \setminus \{0\}$? Compare your conclusion to the previous example and discuss the validity of the slogan "The branch cut gets in the way."

3.3.3. The ML-inequality.

We have not yet stated a triangle inequality for line integrals; in fact, the natural (but, alas, naive) estimate

$$\left|\int_\gamma f\right| \le \int_\gamma |f|$$

does not even make sense.

3.3.15 Problem (!). Why not? Explain why we should not expect the quantity $\int_\gamma |f|$ to be real-valued, and therefore it has no place in an inequality.

Instead, the concept of arc length permits the correct adaptation of the triangle inequality for line integrals. The following estimate is sometimes called the "ML-inequality" or "ML-estimate" because the right side is the product of a *m*aximum and an arc *l*ength. In particular, it is an extension of part (ii) of Problem 3.2.9.

3.3.16 Theorem (ML-inequality). *Let $\mathcal{D} \subseteq \mathbb{C}$ and suppose that $f \colon \mathcal{D} \to \mathbb{C}$ is continuous. Let $\gamma \colon [a, b] \subseteq \mathbb{R} \to \mathbb{C}$ be a path in $\mathcal{D}$. Then*

$$\left|\int_\gamma f\right| \le \left(\max_{a \le t \le b} |f(\gamma(t))|\right) \ell(\gamma).$$

Proof. We prove this only for smooth paths γ and leave the proof in the general case as an exercise. The definition of the line integral and the triangle inequality for definite integrals yield the following estimate:

$$\left|\int_\gamma f\right| = \left|\int_a^b f(\gamma(t))\gamma'(t)\ dt\right| \le \int_a^b |f(\gamma(t))\gamma'(t)|\ dt = \int_a^b |f(\gamma(t))||\gamma'(t)|\ dt.$$

The function

$$g\colon [a,b] \subseteq \mathbb{R} \to \mathbb{R}\colon t \mapsto |f(\gamma(t))|$$

is continuous (because f, γ, and the modulus are all continuous), and so g has a maximum on the closed, bounded interval $[a,b]$ by the extreme value theorem. (Here it is important that g is real-valued, as otherwise the notion of maximum does not make sense.) Then for all $t \in [a,b]$, we have

$$|f(\gamma(t))||\gamma'(t)| \le M|\gamma'(t)|,$$

and so monotonicity for the definite integral of a real-valued function provides

$$\int_a^b |f(\gamma(t))\gamma'(t)|\ dt \le M\int_a^b |\gamma'(t)|\ dt = M\ell(\gamma).$$ ■

3.3.17 Problem (!). Prove the ML-inequality in the case that γ is a path that is not necessarily smooth. [Hint: *suppose* $\gamma = \oplus_{k=1}^n \gamma_k$, *with* γ_k *smooth. First show that* $\left|\int_\gamma f\right| \le \sum_{k=1}^n \left|\int_{\gamma_k} f\right|$. *How does this help?*]

3.3.18 Problem (!). Suppose that a, $b \in \mathbb{R}$ with $a \le b$ and $f\colon [a,b] \to \mathbb{C}$ is continuous. Use the fact that $\int_a^b f = \int_{[a,b]} f$ (Problem 3.3.6) and the ML-inequality to show

$$\left|\int_a^b f\right| \le \left(\max_{a\le t\le b} |f(t)|\right)(b-a).$$

Compare this to (3.2.11).

3.3.19 Problem ($\star$). Let $\mathcal{D} = \{z \in \mathbb{C} \mid |\operatorname{Im}(z)| \le 1\}$, i.e., $\mathcal{D}$ is an infinite horizontal strip of width 2 containing the real line. Suppose that $f\colon \mathcal{D} \to \mathbb{C}$ is continuous and satisfies $|f(z)| \le |\operatorname{Re}(z)|^{-1}$ when $z \in \mathcal{D}$ with $|z| \ge 1$. Use the ML-inequality and the squeeze theorem to show

$$\lim_{R\to\infty} \left|\int_{[R,R+i]} f\right| = 0.$$

3.4. Independence of path.

At last we tackle the problem of antiderivatives on a general subset of $\mathbb{C}$. FTC1 tells us that if $I \subseteq \mathbb{R}$ is an interval and $f\colon I \to \mathbb{C}$ is continuous, then f has an antiderivative, and specifically FTC1 *constructs* an antiderivative for f. For any fixed $a \in I$, an antiderivative is $F(t) := \int_a^t f$; from the point of view of the line integral (recall Problem 3.3.6), we have integrated f over the line segment $[a, t]$. This approach to antiderivatives will not quite succeed if we broaden the domain beyond real intervals.

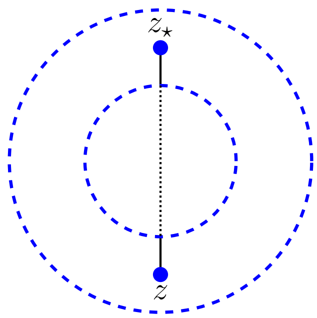
First, continuity alone does not guarantee antiderivatives; the functions in parts (ii) and (iii) of Problem 3.3.14 are continuous on their domains but do not have antiderivatives. Rather, part (i) gives a necessary condition for the existence of an antiderivative: the integral around a closed path must be zero.

Second, even if we knew that a continuous function under consideration integrated to zero around closed paths, how might we try to construct its antiderivative? Could we replicate the technique of FTC1? We could fix some $z_\star \in \mathcal{D}$ and try to "base" our antiderivative there. We might then try to define an antiderivative as

$$F(z) := \int_{[z_\star, z]} f,$$

where $[z_\star, z]$ is the line segment from $z_\star$ to z.

This presumes that $[z_\star, z] \subseteq \mathcal{D}$ for any $z \in \mathcal{D}$, as f needs to be defined over $[z_\star, z]$ for $\int_{[z_\star, z]} f$ to be defined. However, depending on the geometry of $\mathcal{D}$, we have no guarantee that $[z_\star, z] \subseteq \mathcal{D}$ for all $z \in \mathcal{D}$.



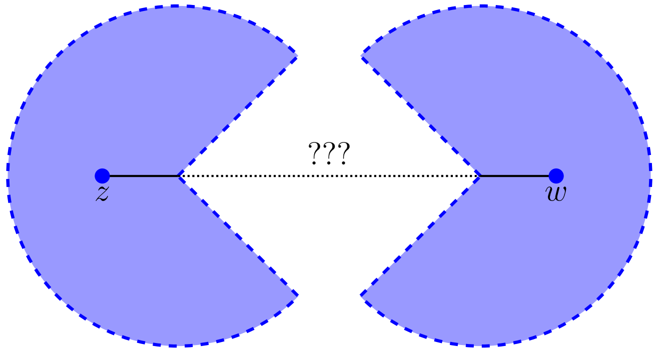
The next option would be not to restrict ourselves to line segments. Suppose we take an arbitrary path γ_z in $\mathcal{D}$ whose initial point is $z_\star$ and whose terminal point is z. Then we could define

$$F(z) := \int_{\gamma_z} f \tag{3.4.1}$$

and perhaps that would be an antiderivative of f.

There are, again, problems with this approach. First, we have no guarantee that there is a point $z_\star \in \mathcal{D}$ such that for any $z \in \mathcal{D}$, there is also a path in $\mathcal{D}$ connecting $z_\star$ and z;

perhaps $\mathcal{D}$ is not connected.



Of course, we could assume that $\mathcal{D}$ is connected, and, in general, we will. (Remember that by Theorem 3.1.42 connectedness is the key to having $f' = 0$ imply what we think it should imply.) We will also assume that $\mathcal{D}$ is open for technical reasons, whose utility will not yet be apparent (although the role of openness in the Cauchy–Riemann equations should be enough to convince us that it *might* be helpful).

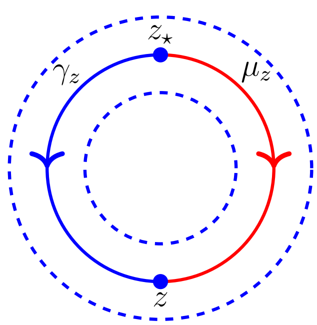
3.4.1 Definition. *A set $\mathcal{D} \subseteq \mathbb{C}$ that is both open and connected will be called, hereafter, a* **DOMAIN**. *Some books use the term* **REGION** *instead of domain.*

3.4.2 Problem (+). Prove that if $\mathcal{D}_1$, $\mathcal{D}_2 \subseteq \mathcal{D}$ are domains with $\mathcal{D}_1 \cap \mathcal{D}_2$ connected, then $\mathcal{D}_1 \cap \mathcal{D}_2$ is also a domain. [Hint: *if $\mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing$, then this intersection is (trivially) a domain, as it is vacuously true that $\varnothing$ is both open and connected. To keep things interesting, suppose $\mathcal{D}_1 \cap \mathcal{D}_2 \neq \varnothing$. To see why this intersection is open and connected, start, as always, by drawing pictures.*] What about $\mathcal{D}_1 \cup \mathcal{D}_2$?

Neither term in the preceding definition is ideal: every function has a domain, but not every function has a domain that is a domain! Every subset of $\mathbb{C}$ could reasonably be called a region, but not every region is a region!

Now, even if $\mathcal{D}$ is a domain, how do we know that the function F in (3.4.1) is *well-defined*? That is, perhaps there are paths γ_z and ϕ_z in $\mathcal{D}$ whose initial points are both $z_\star$ and whose terminal points are both z, but for which

$$\int_{\gamma_z} f \neq \int_{\phi_z} f.$$



In that case, would the antiderivative depend on which path we pick? How would we know which one to choose? Or could the integral of f over a path connecting $z_\star$ and z be "independent of path" in the sense that the integral is the same no matter what the path is (provided those endpoints $z_\star$ and z remain the same)?

This turns out to be a tremendously significant issue, so we first formalize it in a definition and then state and prove a theorem.

3.4.3 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$. A continuous function $f \colon \mathcal{D} \to \mathbb{C}$ is* **PATH INDEPENDENT ON** $\mathcal{D}$ *or* **INDEPENDENT OF PATH ON** $\mathcal{D}$ *if whenever γ_1 and γ_2 are paths in $\mathcal{D}$ with the same initial and terminal points, then*

$$\int_{\gamma_1} f = \int_{\gamma_2} f.$$

It is important to specify the set on which a function is path independent. There are functions $f \colon \mathcal{D} \to \mathbb{C}$ that are path independent on some smaller $\mathcal{D}_0 \subseteq \mathcal{D}$ (in the sense that $\int_\gamma f = \int_\mu f$ for all paths γ and μ in $\mathcal{D}_0$ that have the same initial and terminal points) but not on all of $\mathcal{D}$ (in the sense that there are paths $\widetilde{\gamma}$ and $\widetilde{\mu}$ in $\mathcal{D}$ with the same initial and terminal points but $\int_{\widetilde{\gamma}} f \neq \int_{\widetilde{\mu}} f$).

Combining the geometric hypothesis that $\mathcal{D}$ is a domain with the analytic hypothesis that f is path independent guarantees the existence of an antiderivative.

3.4.4 Theorem (Path independence). *Let $\mathcal{D} \subseteq \mathbb{C}$ be a domain and suppose that $f \colon \mathcal{D} \to \mathbb{C}$ is continuous and path independent. Then f has an antiderivative on $\mathcal{D}$. Specifically, fix $z_\star \in \mathcal{D}$ and, for $z \in \mathcal{D}$, let γ_z be any path with initial point $z_\star$ and terminal point z. Then the map*

$$F \colon \mathcal{D} \to \mathbb{C} \colon z \mapsto \int_{\gamma_z} f$$

is well-defined and holomorphic with $F' = f$.

This is where we finished on Monday, March 4, 2024.

Proof. Given $z \in \mathcal{D}$, such a path γ_z exists because $\mathcal{D}$ is a domain. The function F above is well-defined because f is independent of path: if μ_z is another path in $\mathcal{D}$ whose initial point is $z_\star$ and whose terminal point is z, then

$$\int_{\gamma_z} f = \int_{\mu_z} f.$$

We will show that F is differentiable on $\mathcal{D}$ with $F'(z) = f(z)$ for all $z \in \mathcal{D}$. The proof is very similar to that of FTC1 (Theorem 3.2.14). Fix $z \in \mathcal{D}$. We need to show that

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

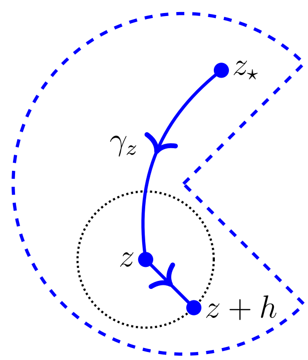
equivalently,

$$\lim_{h\to 0} \frac{F(z+h)-F(z)-hf(z)}{h} = 0. \tag{3.4.2}$$

Let $h \in \mathbb{C}\setminus\{0\}$ with $|h|$ small enough that $[z, z+h] \subseteq \mathcal{D}$. This is possible since $\mathcal{D}$ is open, and so there is $r > 0$ such that $\mathcal{B}(z;r) \subseteq \mathcal{D}$; then with $|h| < r$, we have $[z, z+h] \subseteq \mathcal{B}(z;r)$ by Problem 3.1.7. Let γ_z be any path in $\mathcal{D}$ with initial point $z_\star$ and terminal point z. Then $\gamma_z \oplus [z, z+h]$ is a path in $\mathcal{D}$ with initial point $z_\star$ and terminal point $z+h$, so

$$F(z+h) = \int_{\gamma_z \oplus [z,z+h]} f = \int_{\gamma_z} f + \int_{[z,z+h]} f.$$

Here is a sketch of this situation.



We therefore may calculate

$$F(z+h) - F(z) = \left(\int_{\gamma_z} f + \int_{[z,z+h]} f\right) - \int_{\gamma_z} f = \int_{[z,z+h]} f. \tag{3.4.3}$$

Parametrize the line segment $[z, z+h]$ by $t \mapsto (1-t)z + t(z+h) = z + th$, $0 \le t \le 1$, as usual, so $\gamma'(t) = h$ and

$$\int_{[z,z+h]} f = \int_0^1 f(z+th)h\ dt = h\int_0^1 f(z+th)\ dt.$$

We combine this with (3.4.3) to find

$$\begin{aligned} F(z+h) - F(z) - hf(z) = h\int_0^1 f(z+th)\ dt - hf(z) &= h\int_0^1 f(z+th)\ dt - h\int_0^1 f(z)\ dt \\ &= h\int_0^1 \big[f(z+th) - f(z)\big]\ dt. \end{aligned} \tag{3.4.4}$$

Then

$$\frac{F(z+h)-F(z)-hf(z)}{h} = \int_0^1 \big[f(z+th)-f(z)\big]\ dt. \tag{3.4.5}$$

To prove the desired limit (3.4.2), it therefore suffices to show

$$\lim_{h\to 0}\int_0^1 \big[f(z+th)-f(z)\big]\ dt = 0.$$

This is true by the continuity of f at z, as we show in the following lemma. ■

3.4.5 Lemma. *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f\colon \mathcal{D} \to \mathbb{C}$ be continuous. Then*

$$\lim_{h \to 0} \int_0^1 \left[f(z+th) - f(z)\right] \, dt = 0$$

for each $z \in \mathcal{D}$.

Proof. This proof, too, is very similar to that of FTC1 (Theorem 3.2.14). Fix $z \in \mathcal{D}$. We need to show that given $\epsilon > 0$, there is $\delta > 0$ such that if $h \in \mathbb{C} \setminus \{0\}$ with $|h| < \delta$, then

$$\left|\int_0^1 \left[f(z+th) - f(z)\right] \, dt\right| < \epsilon. \tag{3.4.6}$$

Since f is continuous at z, there is $\delta > 0$ such that if $w \in \mathcal{D}$ with $|w - z| < \delta$, then $|f(w) - f(z)| < \epsilon$. Now suppose $0 < |h| < \delta$. If $0 \le t \le 1$, then

$$|(z+th) - z| = |th| = |t||h| \le |h| < \delta,$$

and therefore

$$\max_{0 \le t \le 1} |f(z+th) - f(z)| < \epsilon.$$

Then the triangle inequality shows

$$\left|\int_0^1 \left[f(z+th) - f(z)\right] \, dt\right| \le \int_0^1 \left|f(z+th) - f(z)\right| \, dt < \int_0^1 \epsilon \, dt = \epsilon.$$

■

3.4.6 Problem (!). Compare the proofs for definite integrals of Theorem 3.2.14 (FTC1) and Lemma 3.2.15 to the proofs for line integrals of Theorem 3.4.4 and Lemma 3.4.5. Identify explicitly where the proofs are identical and how, if at all, they are different. [Hint: *in Lemma 3.2.15, h is strictly real; in Lemma 3.4.5, h can be nonreal.*]

We have now obtained a *sufficient* condition for a continuous function (whose domain is a...domain...) to have an antiderivative: the function should be path independent on that domain. We might ask if we could weaken or change this condition and still guarantee an antiderivative's existence.

We cannot.

In fact, earlier, in Problem 3.3.14 we saw a *necessary* condition for an antiderivative's existence: if a function has an antiderivative, then that function integrates to zero over closed paths. This condition turns out to be sufficient in that it implies path independence and thus the existence of an antiderivative. We collect these seemingly disparate results into one theorem.

3.4.7 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ be a domain and let $f\colon \mathcal{D} \to \mathbb{C}$ be continuous. The following are equivalent:*

(i) *f has an antiderivative on $\mathcal{D}$.*

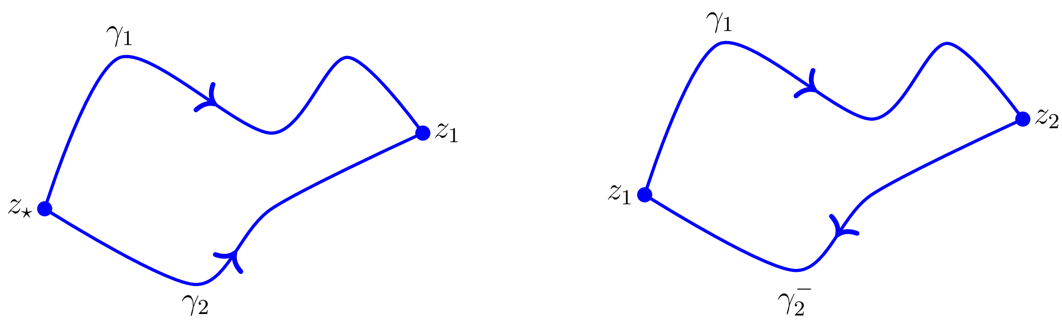
(ii) *If γ is a closed path in $\mathcal{D}$, then*

$$\int_\gamma f = 0.$$

(iii) *f is independent of path in $\mathcal{D}$.*

Proof. (i) $\Longrightarrow$ (ii) This is part (i) of Problem 3.3.14.

(ii) $\Longrightarrow$ (iii) Suppose that γ_1 and γ_2 are paths in $\mathcal{D}$ with the same initial point z_1 and the same terminal point z_2, as in the sketch below.



Then the path $\gamma_1 \oplus \gamma_2^-$ is closed, so part (ii) and properties of line integrals imply

$$0 = \int_{\gamma_1 \oplus \gamma_2^-} f = \int_{\gamma_1} f - \int_{\gamma_2} f$$

and so

$$\int_{\gamma_1} f = \int_{\gamma_2} f.$$

(iii) $\Longrightarrow$ (i) This is Theorem 3.4.4. ■

3.4.8 Problem ($\star$). Let

$$\mathcal{D} := \mathbb{C} \setminus \{0\} \quad \text{and} \quad \mathcal{D}_0 := \{z \in \mathbb{C} \mid \mathrm{Re}(z) > 0,\ \mathrm{Im}(z) > 0\},$$

so $\mathcal{D}$ and $\mathcal{D}_0$ are both domains. Define

$$f : \mathcal{D} \to \mathbb{C} : z \mapsto \frac{1}{z}.$$

(i) Explain why $f\big|_{\mathcal{D}_0}$ is path independent.

(ii) Find paths γ and μ in $\mathcal{D}$ with the same initial and terminal points such that $\int_\gamma f \neq \int_\mu f$. [Hint: *try different portions of the unit circle.*]

We now have new tools available in our quest for antiderivatives: we could check independence of path, or we could check that integrals over closed paths vanish. For a given function, both conditions are arguably somewhat difficult to check, as they require an *infinite* number of conditions to be met. The integral over *every* closed path must vanish—and any domain contains infinitely many closed paths (just consider all the circular ones). Or the integrals over *any* pair of paths with the same initial and terminal points must be the same—and we can probably find infinitely many paths connecting any two points in a domain.

These sufficient conditions for the existence of an antiderivative of a function on an open, connected subset of $\mathbb{C}$ are much more stringent (and annoying) than what guaranteed the existence of an antiderivative for a function on a subinterval of $\mathbb{R}$: just continuity. This suggests that the antiderivative problem in $\mathbb{C}$ is a much richer story than just being the next natural chapter in the evolution of calculus from limits to derivatives to integrals.

3.5. The Cauchy integral theorem.

We develop here a sufficient condition for a suitably nice function on a suitably nice domain of definition to have an antiderivative, with the goal that this condition is easier to check than the "for all"-dependent quantified statements of Theorem 3.4.7. We will once again have to impose more conditions on the geometric structure of the domain of definition and also on the analytic properties of the function. However, these conditions are completely natural in applications, and we will eventually see that the existence of an antiderivative effectively presupposes these conditions. That is, our additional hypotheses will not be that restrictive at all; they would be satisfied in any situation worth considering.

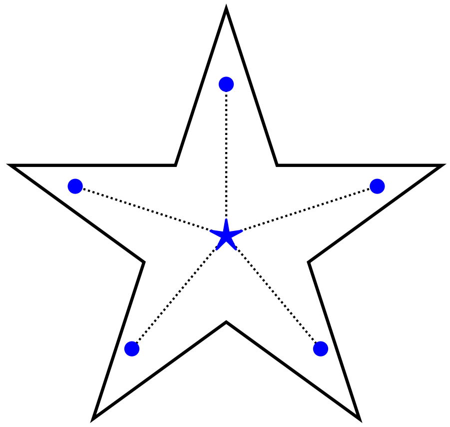
Quite quickly, however, we will move beyond the antiderivative problem. Its solution—more precisely, the *method* of its solution—will reveal incredibly important properties of suitably nice functions defined on suitably nice subsets of $\mathbb{C}$ that will propel the rest of our story.

3.5.1. Star-shaped domains.

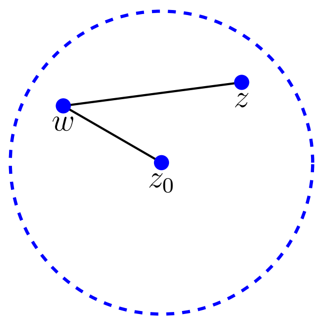
In Theorem 3.4.4, we worked on a domain $\mathcal{D}$ (i.e., an open and connected set), fixed a point $z_\star \in \mathcal{D}$, and integrated over paths connecting $z_\star$ to other $z \in \mathcal{D}$. We will consider those domains $\mathcal{D}$ for which the path connecting $z_\star$ to z is always the line segment $[z_\star, z]$.

3.5.1 Definition. *A set $\mathcal{D} \subseteq \mathbb{C}$ is* STAR-SHAPED *if there is a point $z_\star \in \mathcal{D}$ such that $[z_\star, z] \subseteq \mathcal{D}$ for all $z \in \mathcal{D}$. The point $z_\star$ is called a* STAR-CENTER *for $\mathcal{D}$. A* STAR-SHAPED DOMAIN *or a* STAR DOMAIN *is a domain that is also star-shaped.*

3.5.2 Example. **(i)** We should (unsurprisingly!) expect that the set below is star-shaped, and its star-center should be the point indicated by the symbol $\star$.

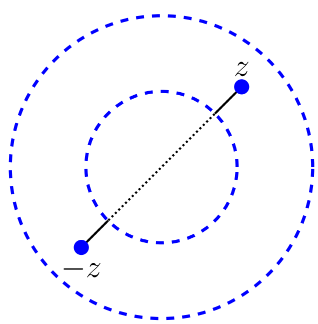


(ii) For any $z_0 \in \mathbb{C}$ and $r > 0$, the open ball $\mathcal{B}(z_0; r)$ is star-shaped, and any point in $\mathcal{B}(z_0; r)$ is a star-center. Below we see that the line segments from both the center of the ball z_0 and an arbitrary point z in the ball can reach any other point w in the ball.



This is where we finished on Wednesday, March 6, 2024.

(iii) Let $0 \leq r < R \leq \infty$. Any **ANNULUS** of the form $\mathcal{A} := \{z \in \mathbb{C} \mid r < |z| < R\}$ is not star-shaped: if $z \in \mathcal{A}$, then $-z \in \mathcal{A}$. However, $0 \in [z, -z]$ and $0 \notin \mathcal{A}$, so $[z, -z] \not\subseteq \mathcal{A}$. That is, no matter what point z we try to pick for the star-center, we cannot connect z to $-z$ by a line segment that is wholly contained in $\mathcal{A}$.



3.5.3 Problem ($\star$). Fill in the following technical details from Example 3.5.2.

(i) For any $z \in \mathbb{C}$, show that $0 \in [z, -z]$.

(ii) Fix $z_0 \in \mathbb{C}$ and $r > 0$. Show that if z, $w \in \mathcal{B}(z_0; r)$, then $[z, w] \subseteq \mathcal{B}(z_0; r)$. [Hint: *use the identity* $z_0 - ((1-t)z + tw) = (1-t)(z_0 - z) + t(z_0 - w)$.]

3.5.4 Problem (!). Let $z \in \mathbb{C} \setminus \{0\}$. Show that $0 \in [z, -z]$. Why does this tell you that no point in $\mathbb{C} \setminus \{0\}$ can be a star-center for $\mathbb{C} \setminus \{0\}$, and therefore that $\mathbb{C} \setminus \{0\}$ is not star-shaped?

3.5.5 Problem ($\star$). **(i)** Prove that any star-shaped set is connected.

(ii) A set $\mathcal{D} \subseteq \mathbb{C}$ is **CONVEX** if $[z, w] \subseteq \mathbb{C}$ for any z, $w \in \mathbb{C}$. Prove that every convex set is connected.

(iii) Is every connected set star-shaped? Is every convex set star-shaped?

3.5.6 Problem (+). Let $b > 0$. Prove that any strip

$$\mathcal{U}_b := \{z \in \mathbb{C} \mid |\operatorname{Im}(z)| < b\}$$

is a star domain. Find all of the star-centers of $\mathcal{U}_b$.

We can now show that with the additional geometric structure of the star domain, we can give a simple condition under which a function has an antiderivative. Namely, we show that if f is holomorphic on the star domain $\mathcal{D}$ and f' is continuous on $\mathcal{D}$, then $\int_\gamma f = 0$ for all closed paths γ in $\mathcal{D}$. By Theorem 3.4.7, this implies that f has an antiderivative on $\mathcal{D}$. Along the way, we will employ an important auxiliary technique called "differentiating under the integral."

3.5.2. Differentiating under the integral.

Suppose that $I \subseteq \mathbb{R}$ is an interval and a, $b \in \mathbb{R}$ with $a \le b$. Let $\mathcal{R}$ be the "rectangle"

$$\mathcal{R} = \left\{(t, s) \in \mathbb{R}^2 \;\middle|\; t \in I,\ a \le s \le b\right\}.$$

Let $f \colon \mathcal{R} \to \mathbb{C}$ be a function and denote by $f(t, \cdot)$ the map $f(t, \cdot) \colon [a, b] \to \mathbb{C} \colon s \mapsto f(t, s)$. If for some $t \in I$ this map $f(t, \cdot)$ is continuous on $[a, b]$, then the integral $\int_a^b f(t, s)\, ds$ is defined. If $f(t, \cdot)$ is continuous on $[a, b]$ for all $t \in I$, then we can define yet another function

$$\mathcal{I} \colon I \to \mathbb{C} \colon t \mapsto \int_a^b f(t, s)\, ds.$$

It is natural to ask if $\mathcal{I}$ is differentiable ("if it moves, differentiate it"), and since the integral has many properties in common with sums, and since *finite* sums and integrals can readily be interchanged, we might expect that

$$\mathcal{I}'(t) = \frac{d}{dt}\int_a^b f(t,s)\ ds = \int_a^b \frac{\partial}{\partial t}[f(t,s)]\ ds,$$

at least if f is differentiable with respect to t and if the partial derivative $f_t(t,\cdot)$ is continuous, so that the new integral on the right exists. Happily, this turns out to be the case, although the proof requires some nuance to make rigorous this interchange of derivative and integral.

3.5.7 Remark. *Note carefully in the derivative*

$$\frac{d}{dt}\int_a^b f(t,s)\ ds$$

that t is not a limit of integration of the integral, and so this is not a situation to which FTC1 applies. Note also that t is not the variable of integration.

3.5.8 Theorem (Leibniz's rule for differentiating under the integral). *Suppose that $I \subseteq \mathbb{R}$ is an interval and $a, b \in \mathbb{R}$ with $a \le b$. Put $\mathcal{R} = \{(t,s) \in \mathbb{R}^2 \mid t \in I,\ a \le t \le b\}$. Let $f \colon \mathcal{R} \to \mathbb{C} \colon (t,s) \mapsto f(t,s)$ be a continuous function such that f_t exists and is continuous on $\mathcal{R}$. Then the map*

$$\mathcal{I} \colon I \to \mathbb{C} \colon t \mapsto \int_a^b f(t,s)\ ds$$

defined and differentiable on I and

$$\mathcal{I}'(t) = \int_a^b f_t(t,s)\ ds.$$

We prove this in Appendix C.3. The goal is to show

$$\lim_{h\to 0}\frac{1}{h}\left(\int_a^b f(t+h,s)\ ds - \int_a^b f(t,s)\ ds\right) = \int_a^b f_t(t,s)\ ds,$$

and the challenge is "interchanging" the limit of the difference quotient and the integral. That is, we need to establish

$$\lim_{h\to 0}\int_a^b \frac{f(t+h,s)-f(t,s)}{h}\ ds = \int_a^b \lim_{h\to 0}\frac{f(t+h,s)-f(t,s)}{h}\ ds.$$

This requires some delicate estimates.

3.5.9 Problem (!). Let

$$\phi(t) := \left[\int_0^1 s\cos(s^2+t)\ ds\right].$$

Calculate ϕ' in two ways in two ways: first by evaluating the integral with FTC2 and differentiating the result and second by differentiating under the integral and then simplifying the result with FTC2. (The point is to convince you that differentiating under the integral gives the right answer.)

3.5.3. The Cauchy integral theorem.

Recall that a sufficient (and also necessary) condition for a function to have an antiderivative on a domain is that the integral of this function over any closed curve in the domain is 0. Cauchy's integral theorem, in turn, gives a sufficient condition for the integral of the function over any closed curve to be 0. Specifically, if the underlying domain is a star domain, and if the function under consideration is holomorphic on the domain, except possibly at a star-center, then the function integrates to 0 over any closed curve in the domain. Here is the precise statement of that result.

3.5.10 Theorem (Cauchy integral theorem). *Let $\mathcal{D} \subseteq \mathbb{C}$ be a star domain with star-center $z_\star$ and let $f\colon \mathcal{D} \to \mathbb{C}$ be continuous on $\mathcal{D}$ and holomorphic on $\mathcal{D} \setminus \{z_\star\}$. Then*

$$\int_\gamma f = 0$$

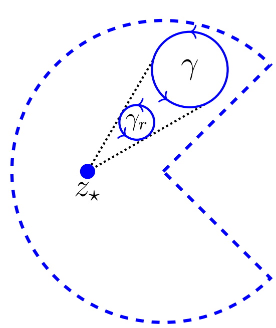
for any closed path γ in $\mathcal{D}$.

Proof. We give the proof under several simplifying assumptions; a fuller proof without these simplifications appears in Appendix C.4. First, suppose that f is holomorphic on all of $\mathcal{D}$ (including at the star-center $z_\star$) and that f' is continuous on $\mathcal{D}$. Second, suppose that γ is a smooth path. For simplicity, assume that γ has been (re)parametrized over $[0,1]$.

Since $z_\star$ is a star-center for $\mathcal{D}$ and $\gamma(t) \in \mathcal{D}$ for each $t \in [0,1]$, we have $[z_\star, \gamma(t)] \subseteq \mathcal{D}$ for each $t \in [0,1]$. Thus $(1-r)z_\star + r\gamma(t) \in \mathcal{D}$ for each $r \in [0,1]$ and $t \in [0,1]$. Define

$$\gamma_r\colon [0,1] \subseteq \mathbb{R} \to \mathcal{D}\colon t \mapsto (1-r)z_\star + r\gamma(t).$$

Then γ_r is a smooth path in $\mathcal{D}$ with $\gamma_1 = \gamma$ and $\gamma_0(t) = z_\star$. Here is a sketch.



We integrate f over γ_r and define

$$\mathcal{I}\colon [0,1] \subseteq \mathbb{R} \to \mathbb{C}\colon r \mapsto \int_{\gamma_r} f = \int_0^1 f((1-r)z_\star + r\gamma(t)) r\gamma'(t)\, dt.$$

Note that

$$\mathcal{I}(0) = 0 \quad \text{and} \quad \mathcal{I}(1) = \int_\gamma f.$$

We will show that $\mathcal{I}$ is constant on $[0,1]$ by computing $\mathcal{I}'$ via differentiation under the integral; we will obtain $\mathcal{I}'(r) = 0$ for each r, and thus $\mathcal{I}(1) = \mathcal{I}(0) = 0$.

The integrand here is

$$g(r,t) := f((1-r)z_\star + r\gamma(t))r\gamma'(t) = f(z_\star + (\gamma(t) - z_\star)r)r\gamma'(t).$$

Since γ is continuously differentiable on $[0,1]$ and since f is holomorphic on $\mathcal{D}$ with f' continuous, it follows that g is continuous on $\mathcal{R} := \{(r,t) \in \mathbb{R}^2 \mid 0 \le r,\ t \le 1\}$, that g is differentiable with respect to r on $\mathcal{R}$, and that g_r is continuous on $\mathcal{R}$. In particular, the product rule gives

$$g_r(r,t) = f'(z_\star + (\gamma(t) - z_\star)r)(\gamma(t) - z_\star)r\gamma' t(t) + f(z_\star + (\gamma(t) - z_\star)r)\gamma'(t).$$

Then

$$\mathcal{I}'(r) = \int_0^1 g_r(r,t)\,dt = \int_0^1 f'(z_\star + (\gamma(t) - z_\star)r)(\gamma(t) - z_\star)r\gamma' t(t)\,dt + \int_0^1 f(z_\star + (\gamma(t) - z_\star)r)\gamma'(t)\,dt. \tag{3.5.1}$$

We evaluate

$$\int_0^1 f'(z_\star + (\gamma(t) - z_\star)r)(\gamma(t) - z_\star)r\gamma' t(t)\,dt$$

using integration by parts. Take

$$\begin{aligned} u &= \gamma(t) - z_\star & dv &= f'(z_\star + (\gamma(t) - z_\star)r)r\gamma'(t)\,dt \\ du &= \gamma'(t)\,dt & v &= f'(z_\star + (\gamma(t) - z_\star)r). \end{aligned}$$

Then

$$\begin{aligned} \int_0^1 f'(z_\star + (\gamma(t) - z_\star)r)(\gamma(t) - z_\star)r\gamma' t(t)\,dt &= \big(\gamma(t) - z_\star\big)\big(f((1-r)z_\star + r\gamma(t))\big|_{t=0}^{t=1} \\ &\quad - \int_0^1 f'(z_\star + (\gamma(t) - z_\star)r)\,dt. \end{aligned} \tag{3.5.2}$$

Since γ is closed, we have $\gamma(0) = \gamma(1)$, and so it follows that

$$\big(\gamma(t) - z_\star\big)\big(f((1-r)z_\star + r\gamma(t))\big|_{t=0}^{t=1} = 0. \tag{3.5.3}$$

Combining (3.5.1), (3.5.2), and (3.5.3) yields $\mathcal{I}'(r) = 0$ for all $r \in [0,1]$. ■

3.5.11 Problem (!). Check that (3.5.3) is true. Remember that $\gamma(0) = \gamma(1)$.

3.5.12 Problem (+). Adapt the proof of the Cauchy integral theorem to the case where γ is only piecewise continuously differentiable. Proceed as follows. First, write $\gamma = \oplus_{k=1}^{n}\gamma_k$, where each γ_k is continuously differentiable on $[0,1]$ with $\gamma_{k-1}(1) = \gamma_k(0)$ for $k = 1,\ldots,n$. Then put $\gamma_{k,r}(t) := (1-r)z_\star + r\gamma_k(t)$ and $\gamma_r := \oplus_{k=1}^{n}\gamma_{k,r}$. Set $\mathcal{I}_k(r) := \int_{\gamma_{k,r}} f$, so $\mathcal{I}(r) = \sum_{k=1}^{n}\mathcal{I}_k(r)$. Differentiate under each integral and obtain

$$\mathcal{I}_k'(r) = \big(\gamma_k(t) - z_\star\big)\big(f((1-r)z_\star + r\gamma_k(t)\big)\big|_{t=0}^{t=1}.$$

Use this to recognize $\sum_{k=1}^{n}\mathcal{I}_k(r)$ as a telescoping sum, i.e., a sum of the form

$$\sum_{k=1}^{n}\mathcal{I}_k(r) = \sum_{k=1}^{n-1}(w_{k+1} - w_k) = w_n - w_0$$

for some $w_k \in \mathbb{C}$.

This is where we finished on Friday, March 8, 2024.

3.5.13 Problem (!). How does the Cauchy integral theorem help you do Example 3.3.4 for $n \neq -1$ very quickly? What goes wrong when $n = -1$?

3.5.14 Example. It is notoriously difficult (impossible) in calculus to find a formula in terms of "elementary functions" for an antiderivative of

$$f\colon \mathbb{R} \to \mathbb{C}\colon t \mapsto e^{t^2}.$$

However, we know that one exists on $\mathbb{R}$ because f is continuous, and so we can use the fundamental theorem of calculus: an antiderivative is

$$F\colon \mathbb{R} \to \mathbb{C}\colon t \mapsto \int_0^t f = \int_0^t e^{\tau^2}\, d\tau.$$

When we extend f to $\mathbb{C}$ as

$$\widetilde{f}\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto e^{z^2},$$

we observe that $f(z) = e^{z^2}$ is entire with $f'(z) = 2ze^{z^2}$, and $\mathbb{C}$ is of course a star domain. The Cauchy integral theorem implies that f has an antiderivative on $\mathbb{C}$.

In fact, we can construct this antiderivative (somewhat) explicitly. By Theorem 3.4.4, since f has an antiderivative on $\mathbb{C}$, f is also path independent on $\mathbb{C}$. Then since the line segment $[0,z]$ is contained in $\mathbb{C}$ for any $z \in \mathbb{C}$, that theorem says that we could take this antiderivative to be

$$\widetilde{F}\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto \int_{[0,z]} \widetilde{f} = \int_{[0,z]} e^{w^2}\, dw.$$

This leads to the wholly unsurprising result that

$$\widetilde{F}\Big|_{\mathbb{R}}(t) = \int_{[0,t]} \widetilde{f} = \int_0^t \widetilde{f}\Big|_{\mathbb{R}} = \int_0^t f = F(t).$$

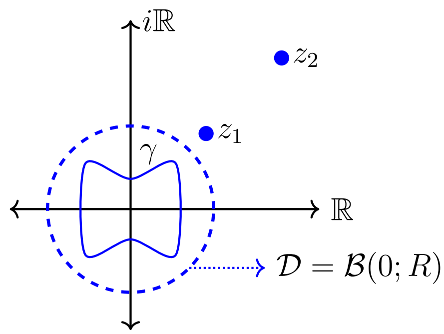
In other words, the Cauchy integral theorem permits us to extend our antiderivative F for f defined on $\mathbb{R}$ into an antiderivative $\widetilde{F}$ for $\widetilde{f}$ defined on $\mathbb{C}$. Everything works as it should, but it needs justification.

By the way, the logic leading to $\widetilde{F}$ was somewhat circular, no? We used the Cauchy integral theorem to prove that $\int_\gamma \widetilde{f} = 0$ for any closed curve γ in $\mathbb{C}$. This proves the independence of path of $\widetilde{f}$. And that proves that defining $\widetilde{F}$ as we did leads to an antiderivative of $\widetilde{f}$. The upshot with this reasoning is that we got the expected *explicit* antiderivative.

3.5.15 Example. Let $z_1, z_2 \in \mathbb{C}$ with $0 \le |z_1| < |z_2|$. Fix $0 < R < |z_1|$ and let γ be any closed curve in $\mathcal{B}(0; R)$. Then

$$\int_\gamma \frac{dz}{(z - z_1)(z - z_2)^2} = 0,$$

since the function $f(z) := 1/[(z - z_1)(z - z_2)^2]$ is holomorphic on the star domain $\mathcal{B}(0; R)$ with f' continuous there.



It is also possible, but more laborious, to obtain this result using a partial fractions decomposition and the fundamental theorem of calculus.

3.5.16 Problem (!). Do that, laboriously.

The Cauchy integral theorem is *the* tool unique to complex analysis that we have long awaited for extracting key properties of functions. We will use it constantly. In the next section we give an extended application of the Cauchy integral theorem to studying Fourier transforms. Appendix D revisits polar coordinates and constructs, more or less from scratch, the principal argument function using the Cauchy integral theorem.

3.5.4. Application to Fourier transforms.

We use the Cauchy integral theorem to prove an estimate on certain integrals that arise when taking Fourier transforms, which are integrals of the form

$$\int_{-\infty}^{\infty} f(t)e^{\pm ikt}\, dt,$$

where $f\colon \mathbb{R} \to \mathbb{C}$ is a function and $k \in \mathbb{R}$. The Fourier transform is a key tool in differential equations for converting ODE posed on $(-\infty, \infty)$ into algebraic equations and PDE into ODE, assuming that sufficient "decay" conditions at "spatial infinity" are assumed to guarantee the convergence of the Fourier integrals.

The goal here is less the estimate and more the proof, which involves recognizing an improper integral on $(-\infty, \infty)$ as part of a line integral, "shifting contours," and using several properties of line integrals. These techniques frequently appear in any application of complex analysis to studying integrals on (subintervals of) $\mathbb{R}$, which itself is a (surprisingly?) common application. We begin with a quick review of improper integrals.

3.5.17 Definition. *Suppose that $f\colon \mathbb{R} \to \mathbb{C}$ is continuous and that the limits*

$$\int_0^\infty f := \lim_{b\to\infty} \int_0^b f \quad \text{and} \quad \int_{-\infty}^0 f := \lim_{a\to-\infty} \int_a^0 f$$

exist. We then say that f is **IMPROPERLY INTEGRABLE** *over $(-\infty, \infty)$ and that the* **IMPROPER INTEGRAL OF f OVER** *$(-\infty, \infty)$ is*

$$\int_{-\infty}^\infty f := \int_{-\infty}^0 f + \int_0^\infty f.$$

We frequently test improper integrability via a kind of "absolute convergence" and "comparison" of improper integrals similar to Theorems 1.4.11 and 1.4.14 for series.

3.5.18 Theorem. *Suppose that $f\colon \mathbb{R} \to \mathbb{C}$ is continuous.*

(i) [Absolute integrability] If $|f|$ is improperly integrable on $(-\infty, \infty)$, then so is f, and the triangle inequality

$$\left|\int_{-\infty}^\infty f\right| \le \int_{-\infty}^\infty |g|$$

holds. In this case we say that f is **ABSOLUTELY INTEGRABLE** on $(-\infty, \infty)$.

(ii) [Principal value] If f is absolutely integrable, then

$$\int_{-\infty}^\infty f = \lim_{R\to\infty} \int_{-R}^R f.$$

(iii) [Comparison test] Suppose that $g\colon \mathbb{R} \to \mathbb{C}$ is continuous and absolutely integrable with $|f(t)| \le |g(t)|$ for all t. Then f is also absolutely integrable and

$$\int_{-\infty}^\infty |f| \le \int_{-\infty}^\infty |g|.$$

3.5.19 Problem (!). **(i)** Define

$$f\colon \mathbb{R} \to \mathbb{C}\colon t \mapsto t.$$

Show that

$$\lim_{R\to\infty} \int_{-R}^{R} t\ dt$$

exists but f is not improperly integrable on $(-\infty, \infty)$.

(ii) Let C, $q > 0$ and suppose that $f\colon \mathbb{R} \to \mathbb{C}$ is continuous with $|f(t)| \le Ce^{-q|t|}$ for all t. Show that f is improperly integrable on $(-\infty, \infty)$ and prove the estimate

$$\left|\int_{-\infty}^{\infty} f\right| \le \frac{2C}{q}.$$

We are going to prove an estimate that says that, under suitable hypotheses on f,

$$\left|\int_{-\infty}^{\infty} f(t)e^{\pm ikt}\ dt\right| \le Ce^{-b|k|}$$

for some constants C, $b > 0$. This shows that the Fourier integral on the left is "exponentially small" in k; informally, for "large" values of k, the integral is "really small." One might appreciate this result by doing the following problem.

3.5.20 Problem (+). **(i)** Suppose that $f\colon \mathbb{R} \to \mathbb{C}$ is n-times differentiable on $\mathbb{R}$ for some $n \ge 1$ and that $f^{(k)}$ is absolutely integrable on $(-\infty, \infty)$ for $1 \le k \le n$. Show that there exists $C_n > 0$ such that for any $k \in \mathbb{R} \setminus \{0\}$,

$$\left|\int_{-\infty}^{\infty} f(t)e^{\pm ikt}\ dt\right| \le \frac{C_n}{|k|^n}.$$

[Hint: *use induction on n to prove the identity*

$$\int_{-\infty}^{\infty} f(t)e^{\pm ikt}\ dt = \left(\mp\frac{1}{ik}\right)^n \int_{-\infty}^{\infty} f^{(n)}(t)e^{\pm ikt}\ dt.$$

It will be helpful to apply integration by parts to the identity

$$\int_{-\infty}^{\infty} g(t)e^{ikt}\ dt = \frac{1}{ik}\int_{-\infty}^{\infty} g(t)\frac{d}{dt}[e^{ikt}]\ dt$$

for various choices of g.]

(ii) Let $\epsilon_0 > 0$. A function $\mathcal{I}\colon (0, \epsilon_0) \to \mathbb{C}$ is **SMALL BEYOND ALL ALGEBRAIC ORDERS OF ϵ AS $\epsilon \to 0^+$** if for all $n \in \mathbb{N}$ there are $C_n > 0$, $\epsilon_n > 0$ such that $|\mathcal{I}(\epsilon)| \le C_n\epsilon^n$ for all $\epsilon \in (0, \epsilon_n)$. The estimate above shows that if f is infinitely differentiable on $\mathbb{R}$ and if each derivative $f^{(k)}$ is absolutely integrable on $(-\infty, \infty)$, then the Fourier integral $\int_{-\infty}^{\infty} f(t)e^{\pm ikt}\ dt$ is small beyond all algebraic orders of $|k|^{-1}$ as $|k|^{-1} \to 0$, equivalently, as $|k| \to \infty$.

Our overall goal here is to improve this estimate to the following. A function $\mathcal{I}\colon (0,\epsilon_0) \to \mathbb{C}$ is **EXPONENTIALLY SMALL IN ϵ AS $\epsilon \to 0^+$** if there are C, b, $\epsilon_\star > 0$ such that $|\mathcal{I}(\epsilon)| \le Ce^{-b/\epsilon}$ for all $\epsilon \in (0,\epsilon_\star)$. Since $b/\epsilon \to \infty$ as $\epsilon \to 0^+$, we have $e^{-b/\epsilon} \to 0$ as $\epsilon \to 0^+$, and our experience with exponentials in calculus should suggest that the convergence $e^{-b/\epsilon} \to 0$ is *really* fast.

Prove that if $\mathcal{I}$ is exponentially small in ϵ as $\epsilon \to 0^+$, then $\mathcal{I}$ is also small beyond all algebraic orders of ϵ as $\epsilon \to 0^+$. [Hint: *argue via L'Hospital's rule that* $\lim_{\epsilon\to 0^+} \epsilon^n e^{b/\epsilon} = \infty$ *and therefore there is* $\epsilon_n > 0$ *such that if* $0 < \epsilon < \epsilon_n$, *then* $1 \le \epsilon^n e^{b/\epsilon}$.] However, argue that the piecewise function

$$\mathcal{I}\colon (0,1) \to \mathbb{C}\colon \epsilon \mapsto \epsilon^n, \ 2^{-(n+1)} \le \epsilon < 2^{-n}$$

is small beyond all algebraic orders of ϵ as $\epsilon \to 0^+$ but not exponentially small in ϵ.

This problem shows that with enough hypotheses on f and *all* of its derivatives on the *real line*, we can eke out a small-beyond-all-algebraic orders estimate on its Fourier transform. Our result below shows that by replacing those derivative hypotheses with the assumption that f is holomorphic and suitably decaying on a *strip*, we can get the better estimate of exponential smallness. In other words, by changing our focus from the one-dimensional real line to (a subset of) the two-dimensional plane, we get stronger results.

Here is the precise result on Fourier transforms.

3.5.21 Theorem. *Let $b_0 > 0$ and let $\mathcal{U}_{b_0}$ be the strip*

$$\mathcal{U}_{b_0} := \{z \in \mathbb{C} \mid |\operatorname{Im}(z)| < b_0\}.$$

Suppose that $f\colon \mathcal{U}_{b_0} \to \mathbb{C}$ is holomorphic and there are M, $q > 0$ such that

$$|f(z)| \le Me^{-q|\operatorname{Re}(z)|}$$

for all $z \in \mathcal{U}_{b_0}$. Then

$$\left|\int_{-\infty}^{\infty} f(t)e^{\pm ikt}\,dt\right| \le \frac{2Me^{-b|k|}}{q}.$$

for all $k \in \mathbb{R}$ and all $b \in (0,b_0)$.

Proof. We give the proof only for the "$+ikt$" case with $k > 0$ and defer the other case, and various details in this proof, to Problem 3.5.22 below.

1. *Rewriting the Fourier integral as part of a line integral.* Problem 3.5.19 allows us to write

$$\int_{-\infty}^{\infty} f(t)e^{ikt}\,dt = \lim_{R\to\infty}\int_{-R}^{R} f(t)e^{ikt}\,dt, \tag{3.5.4}$$

and we manipulate this definite integral. The key idea is to recognize it as one component of a line integral to which we can apply the Cauchy integral theorem.

Fix $R > 0$ and let $0 < b < b_0$. Let γ_R be the rectangle

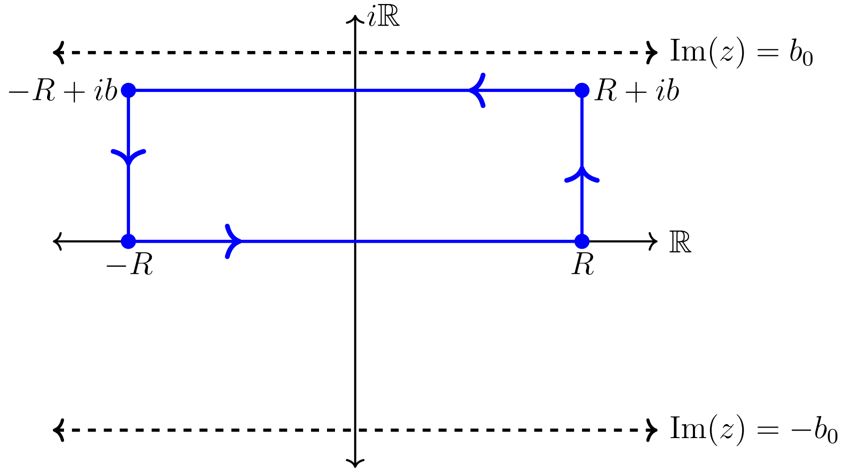
$$\gamma_R := [-R, R] \oplus [R, R+ib] \oplus [R+ib, -R+ib] \oplus [-R+ib, -R],$$

which we sketch below. Then γ_R is a closed curve in the star domain $\mathcal{U}_{b_0}$ (Problem 3.5.6), and the function

$$g \colon \mathcal{U}_{b_0} \to \mathbb{C} \colon z \mapsto f(z)e^{ikz}$$

is holomorphic on $\mathcal{U}_{b_0}$, so the Cauchy integral theorem implies

$$\int_{\gamma_R} g = 0.$$



We can also write this as

$$0 = \int_{\gamma_R} g = \underbrace{\int_{[-R,R]} g}_{\mathcal{I}_1(R)} + \underbrace{\int_{[R,R+ib]} g}_{\mathcal{I}_2(R)} + \underbrace{\int_{[R+ib,-R+ib]} g}_{\mathcal{I}_3(R)} + \underbrace{\int_{[-R+ib,-R]} g}_{\mathcal{I}_4(R)}.$$

Each of these four integrals plays a different role in controlling the Fourier integral.

2. *Shifting focus to a different line integral.* First, from (3.5.4) we have

$$\int_{-\infty}^{\infty} f(t)e^{ikt}\, dt = \lim_{R\to\infty} \int_{[-R,R]} g = \lim_{R\to\infty} \mathcal{I}_1(R).$$

Next, we claim that

$$\lim_{R\to\infty} \mathcal{I}_2(R) = 0 \quad \text{and} \quad \lim_{R\to\infty} \mathcal{I}_4(R) = 0. \tag{3.5.5}$$

If this is true, then

$$0 = \lim_{R\to\infty} \big(\mathcal{I}_1(R) + \mathcal{I}_3(R)\big), \tag{3.5.6}$$

and therefore

$$\int_{-\infty}^{\infty} f(t)e^{ikt}\, dt = -\lim_{R\to\infty} \mathcal{I}_3(R). \tag{3.5.7}$$

3. *Estimating $\mathcal{I}_3(R)$.* We parametrize the line segment $[R+ib, -R+ib]$ by

$$[0,1] \to \mathbb{C}\colon t \mapsto (1-t)(R+ib) + t(-R+ib) = R+ib-2tR,$$

thus

$$\mathcal{I}_3(R) = \int_{[R+ib,-R+ib]} g = -2R\int_0^1 g(R+ib-2tR)\, dt = -2R\int_0^1 f(R+ib-2tR)e^{ik(R+ib-2tR)}\, dt$$
$$= -2Re^{-bk}\int_0^1 f(R+ib-2tR)e^{ikR(1-2t)}\, dt.$$

The triangle inequality for definite integrals and the estimate on f therefore give

$$|\mathcal{I}_3(R)| \le 2Re^{-bk}\int_0^1 |f(R+ib-2tR)e^{ikR(1-2t)}\, dt| = 2Re^{-bk}\int_0^1 |f(R+ib-2tR)|\, dt$$
$$\le 2MRe^{-bk}\int_0^1 e^{-q|\operatorname{Re}(R+ib-2tR)|}\, dt = 2MRe^{-bk}\int_0^1 e^{-qR|1-2t|}\, dt.$$

This definite integral is not too hard to evaluate:

$$\int_0^1 e^{-qR|1-2t|}\, dt = \frac{1-e^{-qR}}{qR}. \tag{3.5.8}$$

All together, we conclude

$$\left|\int_{-\infty}^{\infty} f(t)e^{ikt}\, dt\right| = \lim_{R\to\infty} |\mathcal{I}_3(R)| \le \lim_{R\to\infty} 2MRe^{-bk}\left(\frac{1-e^{-qR}}{qR}\right) = \frac{2Me^{-bk}}{q}\lim_{R\to\infty}(1-e^{-qR})$$
$$= \frac{2Me^{-bk}}{q} = \frac{2Me^{-b|k|}}{q},$$

with the last equality holding since $k \ge 0$. This is the desired estimate on the Fourier integral.

This is where we finished on Monday, March 18, 2024.

4. *Justifying the claim* (3.5.5). We just show that $\lim_{R\to\infty}\mathcal{I}_2(R) = 0$. We parametrize the line segment $[R, R+ib]$ by

$$[0,1] \to \mathbb{C}\colon t \mapsto (1-t)R + t(R+ib) = R+ibt,$$

and so

$$\mathcal{I}_2(R) = \int_0^1 g(R+ibt)ib\, dt = ib\int_0^1 f(R+ibt)e^{ik(R+ibt)}\, dt = ib\int_0^1 f(R+ib)e^{ikR}e^{-kbt}\, dt.$$

The triangle inequality for definite integrals, the estimate on f, and the estimate $e^{-kbt} \le 1$ for $0 \le t \le 1$ and k, $b > 0$ therefore give

$$|\mathcal{I}_2(R)| \le b\int_0^1 |f(R+ibt)e^{ikR}e^{-kbt}\, dt|\, dt \le Mb\int_0^1 e^{-q|\operatorname{Re}(R+ibt)|}\, dt = Mb\int_0^1 e^{-qR}\, dt$$
$$= Mbe^{-qR} \to 0 \text{ as } R \to \infty. \quad \blacksquare$$

3.5.22 Problem (+). Fill in the following omitted details from the proof of Theorem 3.5.21.

(i) The proof above shows

$$\left|\int_{-\infty}^{\infty} f(t)e^{ikt}\,dt\right| \le \frac{2Me^{-bk}}{q}$$

when $k \ge 0$. Now suppose $k < 0$. Use substitution to show

$$\int_{-\infty}^{\infty} f(t)e^{ikt}\,dt = \int_{-\infty}^{\infty} h(\tau)e^{i|k|\tau}\,d\tau, \qquad h(\tau) = f(-\tau),$$

and then use the result above for $k > 0$ to estimate $\int_{-\infty}^{\infty} h(\tau)e^{i|k|\tau}\,d\tau$. The final case on the integral follows by rewriting

$$\int_{-\infty}^{\infty} f(t)e^{-ikt}\,dt = \int_{-\infty}^{\infty} f(t)e^{i(-k)t}\,dt$$

and then using the result just proved for arbitrary $k \in \mathbb{R}$.

(ii) Explain precisely why Problem 3.5.19 permits (3.5.4).

(iii) Justify more precisely (3.5.6) and (3.5.7) using algebra of limits. [Hint: *if* $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$, *and* $\mathcal{I}_4$ *are functions such that* $\lim_{R\to\infty}(\mathcal{I}_1(R)+\mathcal{I}_2(R)+\mathcal{I}_3(R)+\mathcal{I}_4(R)) = 0$, *lim*$_{R\to\infty}\mathcal{I}_2(R) = 0$, $\lim_{R\to\infty}\mathcal{I}_4(R) = 0$, *and* $\lim_{R\to\infty}\mathcal{I}_1(R)$ *exists, use algebra to explain why* $\lim_{R\to\infty}\mathcal{I}_4(R)$ *exists with* $\lim_{R\to\infty}\mathcal{I}_4(R) = -\lim_{R\to\infty}\mathcal{I}_1(R)$.]

(iv) Use FTC2 and properties of absolute value to justify (3.5.8).

(v) Imitate the work in Step 4 to show that $\lim_{R\to\infty}\mathcal{I}_4(R) = 0$.

3.5.5. *Elementary domains.*

In the context of our quest for antiderivatives, the Cauchy integral theorem was a welcome result. In lieu of checking independence of path, it gave us a simple sufficient condition for the existence of an antiderivative: differentiability itself. That is, for a function defined on a star domain to have an antiderivative on that star domain, it suffices for the function to be differentiable. However, one might rightly quibble that the star domain is a very special geometric form. Are there more "relaxed" geometries that guarantee the existence of antiderivatives for suitably nice functions?

We answer our question with a (somewhat circular) definition.

3.5.23 Definition. *A domain (i.e., open and connected)* $\mathcal{D} \subseteq \mathbb{C}$ *is an* **ELEMENTARY DOMAIN** *if every holomorphic function on* $\mathcal{D}$ *has an antiderivative on* $\mathcal{D}$.

3.5.24 Problem (!). Is $\mathbb{C} \setminus \{0\}$ an elementary domain?

Certainly star domains are elementary domains, thanks to Cauchy's integral theorem, but are there others? It turns out that we can easily build new elementary domains out of given ones, and so in particular we can build elementary domains out of star domains. To do this, we need to be able to "glue" certain holomorphic functions together to produce a new holomorphic function that agrees, under certain restrictions, with the old ones.

3.5.25 Lemma (Merging). *Let $\mathcal{D}_1$, $\mathcal{D}_2 \subseteq \mathbb{C}$ be open. Let $f_1 \colon \mathcal{D}_1 \to \mathbb{C}$ and $f_2 \colon \mathcal{D}_2 \to \mathbb{C}$ be holomorphic, and suppose $f_1(z) = f_2(z)$ for all $z \in \mathcal{D}_1 \cap \mathcal{D}_2$. Then there is a unique holomorphic function $f \colon \mathcal{D}_1 \cup \mathcal{D}_2 \to \mathbb{C}$ such that $f\big|_{\mathcal{D}_1} = f_1$ and $f\big|_{\mathcal{D}_2} = \mathcal{D}_2$. Specifically,*

$$f(z) = \begin{cases} f_1(z), & z \in \mathcal{D}_1 \\ f_2(z), & z \in \mathcal{D}_2 \end{cases} \qquad \text{and} \qquad f'(z) = \begin{cases} f_1'(z), & z \in \mathcal{D}_1 \\ f_2'(z), & z \in \mathcal{D}_2. \end{cases}$$

Proof. First we prove uniqueness: if $g \colon \mathcal{D}_1 \cup \mathcal{D}_2 \to \mathbb{C}$ is holomorphic with $g\big|_{\mathcal{D}_1} = f_1$ and $g\big|_{\mathcal{D}_2} = f_2$, then necessarily $g = f$ as defined above. Now we prove existence. With f as defined above, first observe that f is well-defined; if $\mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing$, there is no question, and otherwise if $z \in \mathcal{D}_1 \cap \mathcal{D}_2$, then $f_1(z) = f_2(z)$, and so there is, again, no ambiguity in the definition of f. Next, we need to show that f is holomorphic. Fix $z \in \mathcal{D}_1 \cup \mathcal{D}_2$, let (z_k) be a sequence in $(\mathcal{D}_1 \cup \mathcal{D}_2) \setminus \{z\}$ with $z_k \to z$, and consider the following cases.

1. $z \in \mathcal{D}_1$. Since $\mathcal{D}_1$ is open, there is $r > 0$ such that $\mathcal{B}(z; r) \subseteq \mathcal{D}_1$. Then for k large, we have $z_k \in \mathcal{B}(z; r)$, and so for k large we have $z_k \in \mathcal{D}_1$ and thus $f(z_k) = f_1(z_k)$. Then (for k large)

$$\frac{f(z_k) - f(z)}{z_k - z} = \frac{f_1(z_k) - f_1(z)}{z_k - z} \to f_1'(z).$$

Since (z_k) was an arbitrary sequence in $(\mathcal{D}_1 \cup \mathcal{D}_2) \setminus \{z\}$ with $z_k \to z$, we see that f is differentiable at z and $f'(z) = f_1'(z)$.

2. $z \in \mathcal{D}_2$. The proof is identical, word for word, to the previous case, except for swapping $\mathcal{D}_1$ for $\mathcal{D}_2$ and f_1 for f_2. ■

3.5.26 Theorem. *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be elementary domains such that their intersection $\mathcal{D}_1 \cap \mathcal{D}_2$ is nonempty and connected. Then their union $\mathcal{D}_1 \cup \mathcal{D}_2$ is also an elementary domain.*

Proof. First, since $\mathcal{D}_1$ and $\mathcal{D}_2$ are domains with $\mathcal{D}_1 \cap \mathcal{D}_2$ nonempty and connected, $\mathcal{D}_1 \cup \mathcal{D}_2$ is also a domain. (Hopefully this was a conclusion from Problem 3.4.2.) So, for $\mathcal{D}_1 \cup \mathcal{D}_2$ to be an elementary domain, we want to show that an arbitrary holomorphic $f \colon \mathcal{D}_1 \cup \mathcal{D}_2 \to \mathbb{C}$ has an antiderivative on all of $\mathcal{D}_1 \cup \mathcal{D}_2$.

First, the restrictions $f\big|_{\mathcal{D}_1}$ and $f\big|_{\mathcal{D}_2}$ are also holomorphic; since $\mathcal{D}_1$ and $\mathcal{D}_2$ are elementary domains, there are holomorphic maps $F_1 \colon \mathcal{D}_1 \to \mathbb{C}$ and $F_2 \colon \mathcal{D}_2 \to \mathbb{C}$ such that $F_1'(z) = f(z)$ for $z \in \mathcal{D}_1$ and $F_2'(z) = f(z)$ for $z \in \mathcal{D}_2$. That is, $F_1' = f\big|_{\mathcal{D}_1}$ and $F_2' = f\big|_{\mathcal{D}_2}$.

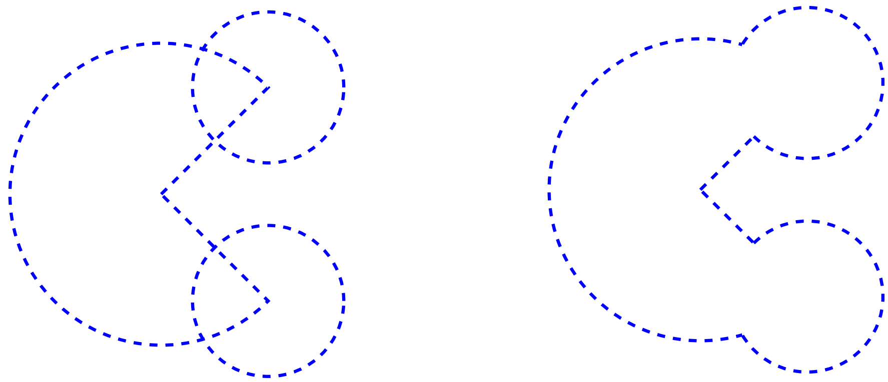
Now define

$$g\colon \mathcal{D}_1 \cap \mathcal{D}_2 \to \mathbb{C}\colon z \mapsto F_1(z) - F_2(z).$$

Then $g'(z) = 0$ for all $z \in \mathcal{D}_1 \cap \mathcal{D}_2$. Since $\mathcal{D}_1 \cap \mathcal{D}_2$ is a domain by Problem 3.4.2, Theorem 3.1.42 implies that g is constant on $\mathcal{D}_1 \cap \mathcal{D}_2$; take $C \in \mathbb{C}$ such that $g(z) = C$ for all $z \in \mathcal{D}_1 \cap \mathcal{D}_2$. Thus $F_1(z) = F_2(z) + C$ for all $z \in \mathcal{D}_1 \cap \mathcal{D}_2$.

The functions F_1 on $\mathcal{D}_1$ and $F_2 + C$ on $\mathcal{D}_2$ therefore satisfy the hypotheses of the merging lemma, and so there is a (unique) holomorphic function $F\colon \mathcal{D}_1 \cup \mathcal{D}_2 \to \mathbb{C}$ such that $F'(z) = f(z)$ for $z \in \mathcal{D}_1$ and $F'(z) = f(z)$ for $z \in \mathcal{D}_2$. Thus $F' = f$ on $\mathcal{D}_1 \cup \mathcal{D}_2$, so F is an antiderivative of f. ■

3.5.27 Example. Since open balls are star domains, we can "glue" overlapping balls onto an existing star domain and get an elementary domain that is (probably) not a star domain.



Not only are we assured that holomorphic functions on elementary domains have antiderivatives, there is also a transparent process for constructing them. Let $\mathcal{D} \subseteq \mathbb{C}$ be elementary and $f\colon \mathcal{D} \to \mathbb{C}$ be holomorphic, so f has an antiderivative on $\mathcal{D}$ and therefore is independent of path on $\mathcal{D}$ by Theorem 3.4.7. Then Theorem 3.4.4 guarantees that if we fix $z_\star \in \mathcal{D}$ and, for $z \in \mathcal{D}$, let γ_z be any path in $\mathcal{D}$ from $z_\star$ to z, the map

$$F\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto \int_{\gamma_z} f$$

is an antiderivative of f. This is the seemingly circular logic referenced in Example 3.5.14: first show the abstract existence of an antiderivative, then obtain path independence, then construct the explicit antiderivative.

3.6. The Cauchy integral formula and its consequences.

We have said that integrals *represent* functions and integrals *extract and measure data* about functions. So far, we have primarily seen integrals representing functions by using integrals to construct antiderivatives of functions. Now we will see integrals do both: a certain line integral will represent a holomorphic function in a very useful way, and a variant of this line integral will contain highly useful data about that function.

3.6.1. A deformation lemma.

A leitmotif of complex integration theory turns out to be deformation of curves. It may be possible to "deform" one curve onto another in a "continuous" way; if the underlying domain is suitably nice (possibly, but not necessarily, a star domain) and if the integrand is suitably nice (holomorphic), then a line integral of a function over one curve should equal a line integral of that function over the other curve. We saw this in our proof of the Cauchy integral theorem; the curve γ was deformed onto the "constant" curve $z_\star$, or, really, the line segment $[z_\star, z_\star]$, and the integral over this line segment was 0.

It is possible to make this notion of deformation very precise and to prove a version of the Cauchy integral theorem stating that the line integral of a holomorphic function is invariant under deformation of curves if the domain is geometrically suitable. We will not explore this and will instead be content with one very specific kind of deformation involving circles.

Recall from Example 3.3.4 that if $z_1 \in \mathbb{C}$ and $r_1 > 0$, then

$$\int_{|z-z_1|=r_1} \frac{dz}{z-z_1} = 2\pi i. \tag{3.6.1}$$

The key connection between the integrand and the path over which the integral is taken is the point z_1: this point z_1 is both the center of the circle over which the integral is taken and the one point at which the integrand fails to be holomorphic. In the following lemma, we relax the structure of the identity (3.6.1) to allow the center of the circle and the "bad" point of the denominator of the integrand to be different. Although it looks very specific, this particular integral will be quite important shortly, and the proof of the lemma below offers further practice with the Cauchy integral theorem, manipulating paths, and doing algebra with integrals. See Problem 3.6.6 for a nontrivial generalization of this result.

3.6.1 Lemma. *Let $z_0 \in \mathbb{C}$ and $r_0 > 0$, and let $z_1 \in \mathcal{B}(z_0; r_0)$. Then*

$$\int_{|z-z_0|=r_0} \frac{dz}{z-z_1} = 2\pi i.$$

Proof. First, trying to calculate this integral by definition, unlike the one in (3.6.1), is an exercise in futility and frustration. We have

$$\int_{|z-z_0|=r_0} \frac{dz}{z-z_1} = \int_0^{2\pi} \frac{r_0 i e^{it}}{(z_0 + r_0 e^{it}) - z_1}\, dt = ir_0 \int_0^{2\pi} \frac{e^{it}}{(z_0 - z_1) + r_0 e^{it}}\, dt.$$

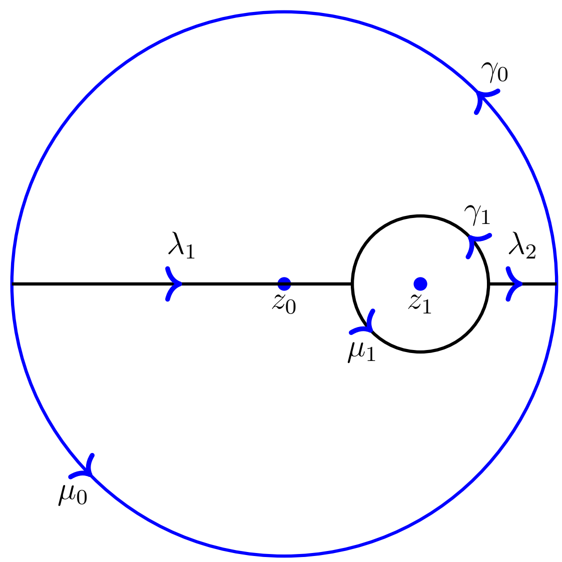
When $z_0 = z_1$, as in (3.6.1), the denominator collapses to $r_0 e^{it}$, which cancels nicely with much of the numerator. When $z_0 \neq z_1$, this definite integral is a piece of junk, and any attempt at antidifferentiating will fail. (Try it. In light of Theorem 3.2.20, why does the "substitution" $u = (z_0 - z_1) + r_0 e^{it}$ fail?)

Let $\rho > 0$ such that $\mathcal{B}(z_1; \rho) \subseteq \mathcal{B}(z_0; r_0)$ and take $r_1 = \rho/2$, so $\overline{\mathcal{B}}(z_1; r_1) \subseteq \mathcal{B}(z_0; r_0)$. We know, as stated above in (3.6.1), that

$$\int_{|z-z_1|=r_1} \frac{dz}{z-z_1} = 2\pi i.$$

We are going to "deform" the circle $|z - z_1| = r_1$ in a "continuous" manner onto the circle $|z - z_0| = r_0$, and the integral is sufficiently "robust" that its value of $2\pi i$ remains unchanged under this deformation. The three words in quotation marks in the previous sentence can be made mathematically rigorous with the notion of homotopy, which we will not pursue.

We split the circle of radius r_0 centered at z_0 and the circle of radius r_1 centered at z_1 into a number of auxiliary paths as sketched below.



The paths γ_0 and γ_1 are the upper halves of their respective circles, and μ_0 and μ_1 are the lower halves. The paths λ_1 and λ_2 are line segments. Then, abbreviating $f(z) := (z - z_1)^{-1}$,

$$\int_{|z-z_0|=r_0} f = \int_{\gamma_0 \oplus \mu_0} f \quad \text{and} \quad \int_{|z-z_1|=r_1} f = \int_{\gamma_1 \oplus \mu_1} f. \tag{3.6.2}$$

We already know that $\int_{\gamma_1 \oplus \mu_1} f = 2\pi i$; we will first use the Cauchy integral theorem to prove that

$$\int_{\lambda_1 \oplus \gamma_1^- \oplus \lambda_2 \oplus \gamma_0} f = 0 \quad \text{and} \quad \int_{\lambda_1 \oplus \mu_1 \oplus \lambda_2 \oplus \mu_0^-} f = 0,$$

and then we will use properties of line integrals and algebra to conclude

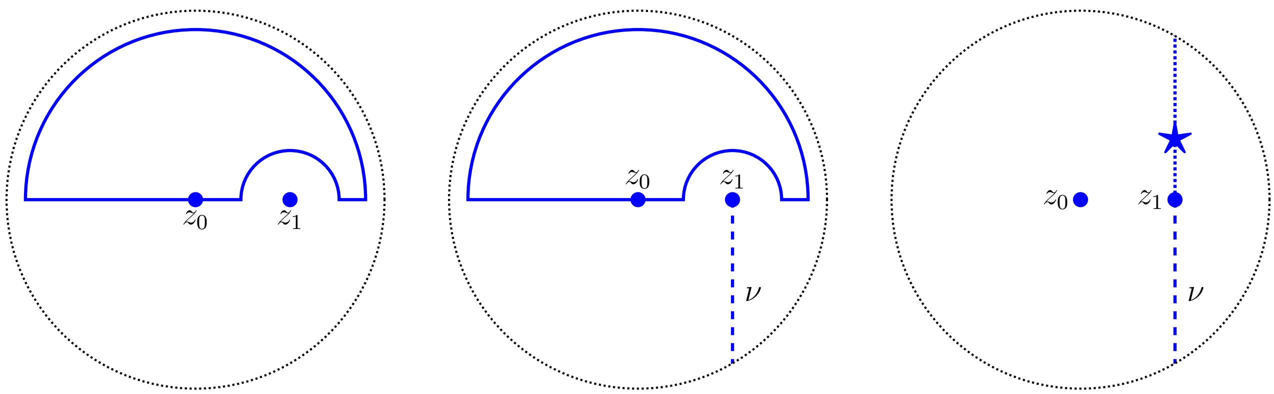
$$\int_{\gamma_0 \oplus \mu_0} f = \int_{\gamma_1 \oplus \mu_1} f.$$

That will give us the desired result.

This is where we finished on Wednesday, March 20, 2024.

Consider the path $\lambda_1 \oplus \gamma_1^- \oplus \lambda_2 \oplus \gamma_0$, which we draw in solid blue in the first circle below. This is a closed path contained in $\mathcal{B}(z_0; R)$ for any $R > r_0$; we draw a circle of radius R centered at z_0 in dotted black below. Delete from $\mathcal{B}(z_0; R)$ the line segment ν from z_1 to the circle of radius R centered at z_0 and call the resulting set $\mathcal{V}$; this is the second circle below. Then $\lambda_1 \oplus \gamma_1^- \oplus \lambda_2 \oplus \gamma_0$ is still a path in $\mathcal{V}$. Also, f is holomorphic on $\mathcal{V}$ since $z_1 \notin \mathcal{V}$.

Finally, $\mathcal{V}$ is a star domain; this is somewhat technical to prove precisely, but any point $\star$ on the dotted blue line in the third circle below will be a star center for $\mathcal{V}$.



The Cauchy integral theorem then implies that

$$\int_{\lambda_1 \oplus \gamma_1^- \oplus \lambda_2 \oplus \gamma_0} f = 0. \tag{3.6.3}$$

Exactly the same arguments show that

$$\int_{\lambda_1 \oplus \mu_1 \oplus \lambda_2 \oplus \mu_0^-} f = 0. \tag{3.6.4}$$

Equating (3.6.3) and (3.6.4) and using the algebra and arithmetic of line integrals shows

$$\int_{\gamma_0 \oplus \mu_0} f = \int_{\gamma_1 \oplus \mu_1} f, \tag{3.6.5}$$

and, by (3.6.2), this is the desired conclusion. ■

3.6.2 Problem (!). Carry out the algebra and arithmetic of line integrals to prove (3.6.5), assuming that (3.6.3) and (3.6.4) hold.

3.6.3 Problem (!). Find parametrizations for all the curves in the Death Star lemma.

3.6.4 Problem (+). Prove that the set $\mathcal{V}$ from the proof of the Death Star lemma is in fact a star domain.

3.6.5 Problem (⋆). Fix $r > 0$ and let $z \in \mathbb{C}$ with $|z| < r$ and $\operatorname{Im}(z) > 0$. Define

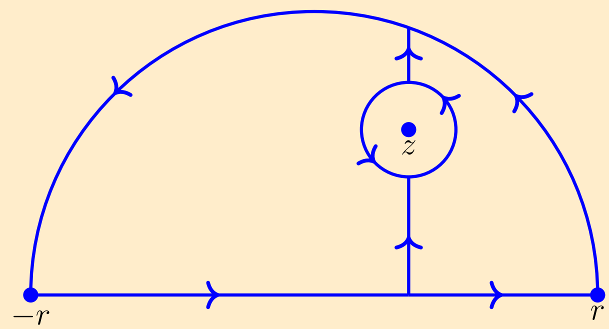
$$\gamma_r \colon [0, \pi] \to \mathbb{C} \colon t \mapsto re^{it}.$$

Show that

$$\int_{[-r,r] \oplus \gamma_r} \frac{dw}{w - z} = 2\pi i$$

using the following two different methods.

(i) Mimic the proof of the Death Star lemma by introducing some auxiliary curves as below.



(ii) Define

$$\mu_r \colon [\pi, 2\pi] \to \mathbb{C} \colon t \mapsto re^{it}.$$

Use the Death Star lemma to show

$$\int_{\gamma_r \oplus \mu_r} \frac{dw}{w-z} = 2\pi i$$

and then use the Cauchy integral theorem (what is the star domain?) to show

$$\int_{[-r,r]\oplus\mu_r^-} \frac{dw}{w-z} = 0.$$

Then add the two integrals above.

3.6.6 Problem ($\star$). Generalize the Death Star lemma as follows. Let $z_0 \in \mathbb{C}$ and $R > 0$. Let $0 < r_0 < R$ and suppose that $z_1 \in \mathcal{B}(z_0; r_0)$. Finally, suppose that $r_1 > 0$ is such that $\overline{\mathcal{B}}(z_1; r_1) \subseteq \mathcal{B}(z_0; r_0)$. Suppose that $f \colon \mathcal{B}(z_0; R) \setminus \{z_1\} \to \mathbb{C}$ is holomorphic. Show that

$$\int_{|z-z_0|=r_0} f = \int_{|z-z_1|=r_1} f.$$

[Hint: *start by drawing pictures of everything.*]

3.6.7 Remark. *Don't be too proud of this technological terror we've constructed in Lemma 3.6.1. The ability to deform one circle onto another and preserve the line integral is insignificant next to the power of the Cauchy theorems.*

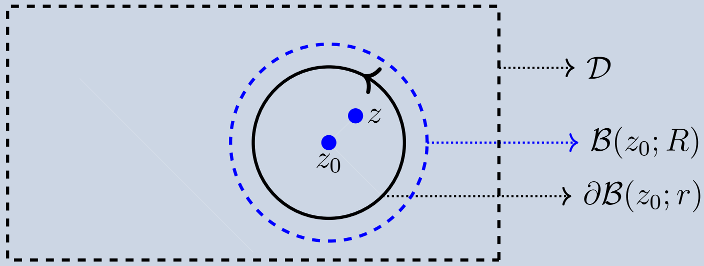
3.6.2. The Cauchy integral formula.

We will now prove one of the most important results in complex analysis, a formula that relates the values of a function in the interior of a ball to its values on the (circular) boundary of that ball. The full utility of this result will probably not be apparent right now, but it will serve us for the rest of the course.

3.6.8 Theorem (Cauchy integral formula). *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f\colon \mathcal{D} \to \mathbb{C}$ be holomorphic. Let $z_0 \in \mathcal{D}$ and $R > 0$ such that $\mathcal{B}(z_0; R) \subseteq \mathcal{D}$. Then*

$$f(z) = \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{w-z}\, dw \tag{3.6.6}$$

for $0 < r < R$ and all $z \in \mathcal{B}(z_0; r)$.



Proof. Fix $r \in (0, R)$ and $z \in \mathcal{B}(z_0; r)$. The Death Star lemma allows us to rewrite

$$f(z) = \frac{f(z)}{2\pi i}(2\pi i) = \frac{f(z)}{2\pi i} \int_{|w-z_0|=r} \frac{dw}{w-z} = \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(z)}{w-z}\, dw. \tag{3.6.7}$$

It therefore suffices to show

$$\int_{|w-z_0|=r} \frac{f(z)}{w-z}\, dw = \int_{|w-z_0|=r} \frac{f(w)}{w-z}\, dw,$$

and this is equivalent to

$$\int_{|w-z_0|=r} \frac{f(w)-f(z)}{w-z}\, dw = 0.$$

The form of the integrand above should call to mind the difference quotient lemma (Lemma 2.5.16), which tells us that the map

$$\phi\colon \mathcal{D} \to \mathbb{C}\colon w \mapsto \begin{cases} \dfrac{f(w)-f(z)}{w-z}, & w \in \mathcal{D} \setminus \{z\}, \\ \\ f'(z), & w = z \end{cases}$$

is holomorphic on $\mathcal{D} \setminus \{z\}$ and continuous on $\mathcal{D}$. In particular, ϕ is continuous on $\mathcal{B}(z_0; R)$ and holomorphic on $\mathcal{B}(z_0; R) \setminus \{z\}$. Recall that $\mathcal{B}(z_0; R)$ is a star domain and one of its (infinitely many) star centers is z. The Cauchy integral formula therefore implies that

$$0 = \int_{|w-z_0|=r} \phi(w)\, dw = \int_{|w-z_0|=r} \frac{f(w)-f(z)}{w-z}\, dw,$$

as desired. ■

This is far from the most general version of the Cauchy integral formula that exists. Eventually we will be able to replace the line integral over a circle with a line integral over an arbitrary closed path (like the path in the Cauchy integral theorem) and get essentially the same formula, up to an extra factor that encodes the "orientation" of the path. However, all the essential corollaries of the Cauchy integral formula for now just require circular paths.

3.6.9 Problem (!). How is the Death Star lemma a special case of the Cauchy integral formula?

3.6.10 Problem (+). Here is another proof of the Cauchy integral formula that relies primarily on the Death Star lemma and its generalization in Problem 3.6.6 rather than the Cauchy integral theorem. By Problem B.1.5, it suffices to show that

$$\left| f(z) - \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{w-z} \, dw \right| < \epsilon \tag{3.6.8}$$

for all $\epsilon > 0$. Take $s > 0$ small enough that $\overline{\mathcal{B}}(z; s) \subseteq \mathcal{D}$ and use the triangle inequality to estimate

$$\begin{aligned}\left| f(z) - \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{w-z} \, dw \right| &\le \left| f(z) - \frac{1}{2\pi i} \int_{|w-z|=s} \frac{f(w)}{w-z} \, dw \right| \\ &\quad + \frac{1}{2\pi} \left| \int_{|w-z|=s} \frac{f(w)}{w-z} \, dw - \int_{|w-z_0|=r} \frac{f(w)}{w-z} \, dw \right|.\end{aligned}$$

Use Problem 3.6.6 to explain why the second term on the right is 0; then adapt (3.6.7) to show

$$\left| f(z) - \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{w-z} \, dw \right| = \frac{1}{2\pi} \left| \int_{|w-z|=s} \frac{f(w)-f(z)}{w-z} \, dw \right|.$$

Use the triangle inequality to show

$$\frac{1}{2\pi} \left| \int_{|w-z|=s} \frac{f(w)-f(z)}{w-z} \, dw \right| \le \max_{0 \le t \le 2\pi} |f(z+se^{it}) - f(z)|,$$

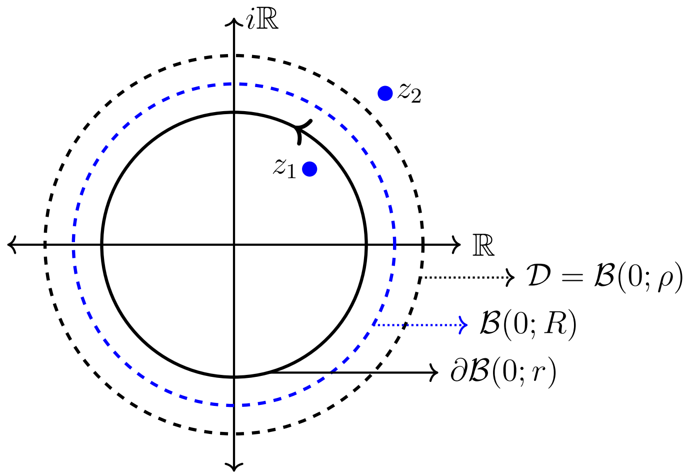
and then use the continuity of f at z to show that for s sufficiently small and all $t \in [0, 2\pi]$, we have

$$|f(z+se^{it}) - f(z)| < \epsilon.$$

Conclude that (3.6.8) is true.

3.6.11 Example. Let z_1, $z_2 \in \mathbb{C}$ with $0 \le |z_1| < |z_2|$. Choose $\rho > 0$ such that $|z_1| < \rho < |z_2|$ and consider the open set $\mathcal{D} = \mathcal{B}(0; \rho)$. Fix r, $R > 0$ with $|z_1| < r \le R \le \rho$. Then

$z_1 \in \mathcal{B}(0; r_1)$ and $\mathcal{B}(0; R) \subseteq \mathcal{B}(0; \rho)$. This is the content of the drawing below.



Then the Cauchy integral formula implies

$$\int_{|z|=r} \frac{dz}{(z-z_1)(z-z_2)^2} = \int_{|z|=r} \frac{1/(z-z_2)^2}{z-z_1}\, dz = \int_{|z-0|=r} \frac{f(z)}{z-z_1}\, dz = \frac{2\pi i}{(z_1-z_2)^2},$$

since the function $f(z) := 1/(z-z_2)^2$ is holomorphic on the open set $\mathcal{D} := \mathcal{B}(0; \rho)$, and since $z_1 \in \mathcal{B}(0; r)$ with $0 < r < R$ and $\mathcal{B}(0; R) \subseteq \mathcal{D}$.

3.6.12 Problem (!). Contrast the result (and the drawing) above with Example 3.5.15. Then redo Example 3.6.11 using partial fractions.

3.6.13 Problem ($\star$). Explain why the Cauchy integral formula does not (apparently) allow us to evaluate

$$\int_{|z|=2} \frac{dz}{z^2-1}.$$

Then rewrite the integrand using partial fractions and realize that the Cauchy integral formula (or maybe just the Death Star lemma!) does, in fact, apply.

This is where we finished on Friday, March 22, 2024.

The true value of the Cauchy integral formula (CIF) is not that it enables us to compute certain line integrals that would otherwise be difficult or impossible (although it does). Rather, the CIF provides an *integral representation* of a function, and integrals are *the* key instrument for extract information about functions.

Specifically, the CIF uses one-dimensional information about a function f—the values of $w \mapsto f(w)/(w-z)$ on the circle of radius r centered at z_0—to compute two-dimensional information about f—its values on the ball of radius r centered at z_0. This may feel similar to the fundamental theorem of calculus, which reads

$$\int_a^b f' = f(b) - f(a)$$

when f is differentiable on $[a, b]$ and f' is continuous on $[a, b]$. Both the CCIF and the FTC give information about a function from an integral whose integrand is related to that function.

The CIF, however, might have at least two advantages over our beloved FTC. First, the FTC requires information about the derivative on the whole interval $[a, b]$ to produce information about f at the endpoints; we need one-dimensional data (values on an interval) to get zero-dimensional data (the difference of values at the endpoints). Second, the FTC requires information about a function other than f (namely, the derivative of f), whereas the integrand in the CIF is really just f gussied up via division by a linear polynomial.

We only proved the Cauchy integral formula for line integrals over circles, whereas the Cauchy integral theorem holds for line integrals over arbitrary closed paths. We will eventually generalize the integral formula to permit more arbitrary closed paths, but that will also require us to account for a notion of "orientation" on the paths. As it stands, our version of the integral formula above is perfectly suited to give us a rich amount of information about functions.

3.6.3. The generalized Cauchy integral formula.

Here is the first of many deep consequences of the Cauchy integral formula. Suppose that the hypotheses of the Cauchy integral formula are met. That is, we have an open set $\mathcal{D}$ and a holomorphic function $f \colon \mathcal{D} \to \mathbb{C}$, and we have fixed $z_0 \in \mathcal{D}$ and $R > 0$ such that $\overline{\mathcal{B}}(z_0; R) \subseteq \mathcal{D}$. Then for any $r \in (0, R)$ and $z \in \mathcal{B}(z_0; r)$, we can write

$$f(z) = \frac{1}{2\pi i} \int_{|w-z_0|=r} g(w, z)\ dw, \qquad \text{where} \qquad g(z, w) := \frac{f(w)}{w - z}.$$

The map g is defined on the set

$$\mathcal{D}_0 := \left\{(z, w) \in \mathbb{C}^2 \ \middle|\ |z - z_0| < r,\ |w - z_0| = r\right\}.$$

In particular, for $(z, w) \in \mathcal{D}_0$, we have $z \neq w$. It should follow, then, that g is continuous on $\mathcal{D}_0$ (this needs some development, since we have not discussed continuity for functions defined on subsets of $\mathbb{C}^2$) and that g is differentiable with respect to z (this too needs development, since we have not discussed partial derivatives for functions of several complex variables), and that

$$g_z(z, w) = \frac{f(w)}{(w - z)^2}.$$

If this is all indeed true (it is), then we might expect that we could differentiate under the (line) integral as in Leibniz's rule (Theorem 3.5.8) and find

$$f'(z) = \frac{1}{2\pi i} \int_{|w-z_0|=r} g_z(z, w)\ dw = \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{(w - z)^2}\ dw.$$

Now look at this integrand. Exactly the same reasoning as above suggests that we can differentiate under the integral *again* to conclude that f' is differentiable and

$$f''(z) = 2\left(\frac{1}{2\pi i}\int_{|w-z_0|=r} \frac{f(w)}{(w-z)^3}\,dw\right).$$

Turn the crank and be convinced that f'' is differentiable...

If this reasoning holds, then we have discovered something remarkable. A holomorphic function is not just once differentiable but *infinitely* many times differentiable. This is definitely not true for functions defined on (subsets of) $\mathbb{R}$.

3.6.14 Problem ($\star$). Show that the function

$$f\colon \mathbb{R} \to \mathbb{C}\colon t \mapsto \begin{cases} t^2, & t \geq 0 \\ -t^2, & t < 0 \end{cases}$$

is differentiable on $\mathbb{R}$ and that f' is continuous on $\mathbb{R}$ but not differentiable at 0.

Moreover, it appears that we can represent all of a function's erivatives as a line integral of the quotient of the *original* function and a polynomial. This result is called the generalized Cauchy integral formula, and it has many proofs. The proof that we will give hinges on the venerable mathematical technique known as brute force.

3.6.15 Remark. *Brute force is the best force.*

Here is the brute force part of the proof; the proof of the following lemma is in Appendix C.5.

3.6.16 Lemma. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \partial\mathcal{B}(z_0; r) \to \mathbb{C}$ is continuous and let $k \geq 1$ be an integer. Define*

$$F_k\colon \mathbb{C} \setminus \mathcal{C}(z_0; r) \to \mathbb{C}\colon z \mapsto \int_{|w-z_0|=r} \frac{f(w)}{(w-z)^k}\,dw. \tag{3.6.9}$$

Then F_k is holomorphic with $F_k' = kF_{k+1}$.

In the following we denote the kth derivative of a function f by $f^{(k)}$, i.e.,

$$f^{(k)}(z) = \begin{cases} f(z), & k = 0 \\ (f^{(k-1)})'(z), & k \geq 1. \end{cases}$$

3.6.17 Theorem (Generalized Cauchy integral formula). *Suppose that $\mathcal{D} \subseteq \mathbb{C}$ is open and $f\colon \mathcal{D} \to \mathbb{C}$ is holomorphic. Then f is infinitely differentiable on $\mathcal{D}$. In particular, if $z_0 \in \mathcal{D}$ with $\mathcal{B}(z_0; R) \subseteq \mathcal{D}$, then for any $r \in (0, R)$, $z \in \mathcal{B}(z_0; R)$, and $k \geq 0$, the kth* *derivative of f is*

$$f^{(k)}(z) = \frac{k!}{2\pi i}\int_{|w-z_0|=r} \frac{f(w)}{(w-z)^{k+1}}\, dw. \tag{3.6.10}$$

Proof. We induct on k, starting with $k = 0$, i.e., $f^{(0)} = f$. Then (3.6.10) is just the Cauchy integral formula. Assume that (3.6.10) holds for some $k \ge 0$; then

$$f^{(k)} = \frac{k!}{2\pi i} F_{k+1},$$

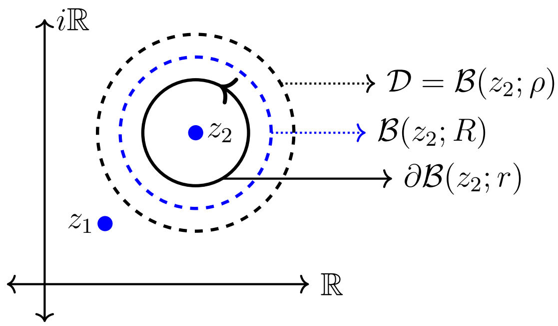
where F_{k+1} was defined in (C.5.1). Lemma 3.6.16 implies that F_{k+1} is holomorphic with $F_{k+1}' = (k+1)F_{k+2}$. Consequently, $f^{(k)}$ is differentiable with

$$f^{(k+1)}(z) = (f^{(k)})'(z) = (k+1)F_{k+2}(z) = \frac{(k+1)k!}{2\pi i}F_{k+2}(z) = \frac{(k+1)!}{2\pi i}\int_{|w-z_0|=r}\frac{f(w)}{(w-z)^{k+2}}\, dw.$$

This is the desired form of $f^{(k+1)}$ from (3.6.10). ■

Once again, we see an integral representing a function—specifically, the kth derivative of a function.

3.6.18 Example. Let $z_1, z_2 \in \mathbb{C}$ with $0 \le |z_1| < |z_2|$. Let $0 < \rho < |z_2| - |z_1|$, so $z_1 \notin \mathcal{B}(z_2;\rho)$. (Otherwise, we would have $|z_1 - z_2| < \rho$, and then the reverse triangle inequality would give $|z_2| - |z_1| < |z_1 - z_2| < \rho$.) Fix $0 < r < R \le \rho$, so $z_2 \in \mathcal{B}(z_2; r)$ and $\mathcal{B}(z_2; R) \subseteq \mathcal{B}(z_2;\rho)$. This is the content of the drawing below.



Then

$$\int_{|z-z_2|=r}\frac{dz}{(z-z_1)(z-z_2)^2} = \int_{|z-z_2|=r}\frac{1/(z-z_1)}{(z-z_2)^{1+1}}\, dz = 2\pi i \frac{d}{dz}\left[\frac{1}{z-z_1}\right]\Bigg|_{z=z_2} = -\frac{2\pi i}{(z_1-z_2)^2},$$

since the function $f(z) := 1/(z - z_1)$ is holomorphic on the open set $\mathcal{D} := \mathcal{B}(z_2;\rho)$ and the balls constructed above satisfy the hypotheses of the generalized Cauchy integral formula.

3.6.19 Problem (!). Can you redo this example with partial fractions?

3.6.20 Problem (!). **(i)** Redo Examples 3.5.15, 3.6.11, and 3.6.18 one after the other.

(ii) Why do these three examples all require different techniques? What are those techniques, and how are the results different?

(iii) Why do none of those techniques help us evaluate

$$\int_{|z|=r} \frac{dz}{(z-z_1)(z-z_2)^2},$$

where z_1, $z_2 \in \mathcal{B}(0;r)$?

At last, we can fully characterize when a function has an antiderivative. This effectively completes the third phase of our course—the integral calculus phase—and opens the way to a multiverse of complex analytic possibilities.

3.6.21 Problem (+). Let $\mathcal{D} \subseteq \mathbb{C}$ be a domain and suppose that $f \colon \mathcal{D} \to \mathbb{C}$ is continuous.

(i) Suppose here (and only here, i.e., not in the following parts) that $\mathcal{D}$ is also an elementary domain. Show that a function f is holomorphic if and only if f has an antiderivative on $\mathcal{D}$. [Hint: *one direction is the definition; for the other, if $F' = f$, what do you know about F''?*]

(ii) Show that f is holomorphic if and only if f is "locally antidifferentiable" in the sense that if $z_0 \in \mathcal{D}$ and $r > 0$ such that $\mathcal{B}(z_0;r) \subseteq \mathcal{D}$, then there is a holomorphic function $F \colon \mathcal{B}(z_0;r) \to \mathbb{C}$ such that $F'(z) = f(z)$ for all $z \in \mathcal{B}(z_0;r)$. [Hint: *an open ball $\mathcal{B}(z_0;r)$ is a star domain.*]

(iii) **[Morera's theorem]** Show that $f \colon \mathcal{D} \to \mathbb{C}$ is holomorphic if and only if $\int_\gamma f = 0$ for any closed curve γ whose image is contained in some ball $\mathcal{B}(z_0;r) \subseteq \mathcal{D}$. [Hint: *use the hint from the preceding part and the independence of path theorem.*]

(iv) Use Problem C.4.8 to show that in the preceding part, we can replace "all closed curves $\gamma \in \mathcal{D}$" with "all triangular paths $\partial\Delta(z_1,z_2,z_3)$ such that $\Delta(z_1,z_2,z_3) \subseteq \mathcal{D}$."

3.6.4. Liouville's theorem.

Here is a first result from that multiverse of possibilities. We have not used the following definition all that much, so now is a good time to bring it up.

3.6.22 Definition. *A holomorphic function $f \colon \mathbb{C} \to \mathbb{C}$ is called* **ENTIRE**. *That is, $f \colon \mathbb{C} \to \mathbb{C}$ is entire if f is differentiable at each $z \in \mathbb{C}$.*

If we replace $\mathbb{C}$ by $\mathbb{R}$ in the preceding definition, we are familiar with many functions that are infinitely differentiable on $\mathbb{R}$. And many of those functions are bounded; consider $f(t) = \sin(t)$, which satisfies $|\sin(t)| \leq 1$ for all $t \in \mathbb{R}$. It turns out that only the most trivial of bounded functions can be entire.

3.6.23 Theorem (Liouville). *Suppose that $f : \mathbb{C} \to \mathbb{C}$ is entire and bounded, i.e., there is $M > 0$ such that $|f(z)| \le M$ for all z. Then f is constant.*

Proof. We show that $f'(z) = 0$ for all z; since $\mathbb{C}$ is a domain, it follows that f is constant. Fix $z \in \mathbb{C}$ and $r > 0$. In the notation of the generalized Cauchy integral formula, we take $z_0 = z$, $R = 2r$, $k = 1$, and $\mathcal{D} = \mathbb{C}$. Then

$$f'(z) = \frac{1}{2\pi i} \int_{|w-z|=r} \frac{f(w)}{(w-z)^2}\, dw,$$

and if $|w - z| = r$, then we can estimate the integrand as

$$\left| \frac{f(w)}{(w-z)^2} \right| = \frac{|f(w)|}{|w-z|^2} = \frac{|f(w)|}{r^2} \le \frac{M}{r^2}.$$

Then the ML-inequality implies

$$|f'(z)| = \left| \frac{1}{2\pi i} \int_{|w-z|=r} \frac{f(w)}{(w-z)^2}\, dw \right| \le \frac{2\pi r M}{2\pi r^2} = \frac{M}{r^2}.$$

Since this is true for an arbitrary $r > 0$, we can use the squeeze theorem and send $r \to \infty$ to conclude $|f'(z)| = 0$, thus $f'(z) = 0$. ∎

3.6.24 Example. Previously we have seen that $\sin(\cdot)$ is unbounded on $\mathbb{C}$, e.g., by considering

$$\sin(iy) = \frac{e^{i(iy)} - e^{-i(iy)}}{2i} = \frac{e^{-y} - e^{y}}{2i},$$

thus

$$|\sin(iy)| = \frac{|e^{-y} - e^{y}|}{2} \to \infty \text{ as } y \to \pm\infty.$$

But even without this estimate, since we know that $\sin(\cdot)$ is entire and not constant (e.g., $\sin(0) = 0$ and $\sin(\pi/2) = 1$), we are guaranteed that $\sin(\cdot)$ is unbounded on $\mathbb{C}$. This is, of course, a marked contrast to the familiar estimate $|\sin(x)| \le 1$ for $x \in \mathbb{R}$.

As an application of Liouville's theorem, we derive a first (somewhat weak) version of the fundamental theorem of algebra, which states that every polynomial with complex coefficients has a root in $\mathbb{C}$. Note that not every polynomial with real coefficients has a root in $\mathbb{R}$ (think of the most famous quadratic in the world, which is, from one point of view, the reason this course exists).

3.6.25 Theorem. *Let $f(z) = \sum_{k=0}^{n} a_k z^k$ be a polynomial of degree $n \ge 1$, i.e., $a_0, \ldots, a_n \in \mathbb{C}$ and $a_n \ne 0$. Then f has a root in $\mathbb{C}$: there is $z_0 \in \mathbb{C}$ such that $f(z_0) = 0$.*

Proof. Suppose not. Then $f(z) \neq 0$ for all $z \in \mathbb{C}$, and so the function $1/f$ is defined on $\mathbb{C}$; moreover, g is holomorphic on $\mathbb{C}$. If we can show that $1/f$ is also bounded on $\mathbb{C}$, i.e., there is $M > 0$ such that $1/|f(z)| \leq M$ for all $z \in \mathbb{C}$, then Liouville's theorem will tell us that $1/f$ is constant. That is, there is $c \in \mathbb{C}$ such that $1/f(z) = c$ for all $z \in \mathbb{C}$ (and so in particular $c \neq 0$), and then $f(z) = 1/c$ for all $z \in \mathbb{C}$. But then f is not a polynomial of degree at least 1.

Here is the argument that $1/f$ is bounded. Multiple applications of the reverse triangle inequality show

$$|f(z)| \geq |a_n||z|^n - \sum_{k=0}^{n-1} |a_k||z|^k.$$

Define

$$h\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto |a_n|t^n - \sum_{k=0}^{n-1} |a_k|t^k,$$

so $|f(z)| \geq h(|z|)$, and h is a polynomial whose leading coefficient $|a_n|$ is positive. Thus $\lim_{t\to\infty} h(t) = \infty$, so there is $t_0 > 0$ such that if $t \geq t_0$, then $|h(t)| \geq 1$.

For $z \in \mathbb{C}$ with $|z| \geq t_0$, we then have

$$\frac{1}{|f(z)|} \leq \frac{1}{h(|z|)} \leq 1.$$

And since $1/f$ is continuous on the closed ball $\overline{\mathcal{B}}(0; t_0)$, there is $m > 0$ such that $1/|f(z)| \leq m$ for $z \in \mathcal{B}(0; t_0)$. All together, we have $1/|f(z)| \leq \max\{1, m\}$, and so $1/f$ is bounded, as desired. ■

This is where we finished on Monday, March 25, 2024.

4. THE MULTIVERSE OF ANALYTIC FUNCTIONS

4.1. Analyticity.

The fact that a once-differentiable function is really infinitely many times differentiable should be surprising, if not shocking. We will now develop a result that is nothing short of *staggering*.

4.1.1. Taylor series.

Suppose that $\mathcal{D} \subseteq \mathbb{C}$ is open and $z_0 \in \mathcal{D}$ with $R > 0$ such that $\mathcal{B}(z_0; R) \subseteq \mathcal{D}$. Fix $z \in \mathcal{B}(z_0; r)$ with $0 < r < R$. Then the Cauchy integral formula gives

$$f(z) = \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{w-z}\, dw. \tag{4.1.1}$$

We can manipulate the factor $1/(w-z)$ in the integrand in a powerful, critical way. We use three tricks.

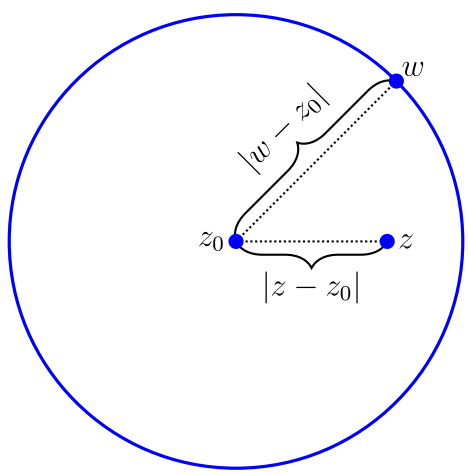
Trick 1. *Adding zero and factoring.* We rewrite the denominator as

$$w - z = w - z_0 + z_0 - z = w - z_0 - (z - z_0).$$

Since, in the line integral in (4.1.1), we presume $|w - z_0| = r > 0$, we have $w - z_0 \neq 0$, and so we may factor

$$w - z = w - z_0 - (z - z_0) = (w - z_0)\left(1 - \frac{z - z_0}{w - z_0}\right). \tag{4.1.2}$$

Trick 2. *Recognizing the geometric series.* Here is a drawing of how w, z, and z_0 are related to each other.



Hopefully the picture emphasizes that $|w - z| < |w - z_0|$, but this is easy to prove. We are assuming in (4.1.1) that $z \in \mathcal{B}(z_0; r)$ and $|w - z_0| = r$, thus $|z - z_0| < r = |w - z_0|$. So, the second factor in (4.1.2) has the form $1 - u$, where $u = (z - z_0)/(w - z_0)$. Thus $|u| < 1$, and so we can express that second factor as a geometric series.

Trick 3. *Deploying the geometric series.* Here we go: since $|(z-z_0)/(w-z_0)| < 1$, the geometric series gives

$$1 - \frac{z-z_0}{w-z_0} = \sum_{k=0}^{\infty} \left(\frac{z-z_0}{w-z_0}\right)^k = \sum_{k=0}^{\infty} \frac{(z-z_0)^k}{(w-z_0)^k},$$

and therefore

$$w - z = (w-z_0)\left(1 - \frac{z-z_0}{w-z_0}\right) = (w-z_0)\sum_{k=0}^{\infty} \frac{(z-z_0)^k}{(w-z_0)^k} = \sum_{k=0}^{\infty} \frac{(z-z_0)^k}{(w-z_0)^{k+1}}.$$

We combine all the tricks to rewrite

$$\int_{|w-z_0|=r} \frac{f(w)}{w-z}\, dw = \int_{|w-z_0|=r} f(w) \sum_{k=0}^{\infty} \frac{(z-z_0)^k}{(w-z_0)^{k+1}}\, dw$$

Suppose for the moment that we can "interchange" the line integral and the series, i.e.,

$$\int_{|w-z_0|=r} f(w) \sum_{k=0}^{\infty} \frac{(z-z_0)^k}{(w-z_0)^{k+1}}\, dw = \sum_{k=0}^{\infty} \int_{|w-z_0|=r} f(w) \frac{(z-z_0)^k}{(w-z_0)^{k+1}}\, dw. \tag{4.1.3}$$

This is certainly true if the series is just a finite sum ($\int_\gamma \sum_{k=0}^n f_k = \sum_{k=0}^n \int_\gamma f_k$) and morally it should smack of differentiating under the integral; both there and here we are swapping an integral and a limiting procedure. If we can perform this interchange, then

$$\sum_{k=0}^{\infty} \int_{|w-z_0|=r} f(w) \frac{(z-z_0)^k}{(w-z_0)^{k+1}}\, dw = \sum_{k=0}^{\infty} \left(\int_{|w-z_0|=r} \frac{f(w)}{(w-z_0)^{k+1}}\, dw\right)(z-z_0)^k.$$

After some rearranging, we conclude that if (4.1.3) is indeed permitted, then we have shown

$$f(z) = \sum_{k=0}^{\infty} \left(\frac{1}{2\pi i}\int_{|w-z_0|=r} \frac{f(w)}{(w-z_0)^{k+1}}\, dw\right)(z-z_0)^k. \tag{4.1.4}$$

If we put

$$a_k := \frac{1}{2\pi i}\int_{|w-z_0|=r} \frac{f(w)}{(w-z_0)^{k+1}}\, dw, \tag{4.1.5}$$

then this just compresses to

$$f(z) = \sum_{k=0}^{\infty} a_k (z-z_0)^k, \tag{4.1.6}$$

and we might remember from the generalized Cauchy integral formula that

$$a_k = \frac{f^{(k)}(z_0)}{k!}.$$

In other words, if (4.1.3) is true, then f is really a *power series*—at least locally, around a given point—and the coefficients in this power series expansion effectively arise from the generalized Cauchy integral formula.

We can be a little more precise about z. We have proved the identities (4.1.4) and (4.1.6), up to verifying (4.1.3), with the assumption that $z \in \mathcal{B}(z_0; r)$ for some $0 < r < R$, where $\mathcal{B}(z_0; R) \subseteq \mathcal{D}$. However, we can cut down on the annoying role of r slightly (but not entirely), and we do so in the following theorem.

4.1.1 Theorem (Taylor). *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f\colon \mathcal{D} \to \mathbb{C}$ be holomorphic. Let $z_0 \in \mathcal{D}$ and $R > 0$ such that $\mathcal{B}(z_0; R) \subseteq \mathcal{D}$. Then*

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k, \qquad a_k := \frac{f^{(k)}(z_0)}{k!} \tag{4.1.7}$$

for each $z \in \mathcal{B}(z_0; R)$. Equivalently,

$$a_k = \frac{1}{2\pi i} \int_{|w-z_0|=r} \frac{f(w)}{(w - z_0)^{k+1}}\, dw \tag{4.1.8}$$

for any $r \in (0, R)$.

Proof. We continue to defer verifying the interchange (4.1.3). Fix $z \in \mathcal{B}(z_0; R)$, so $|z - z_0| < R$. Take $r > 0$ such that $|z - z_0| < r < R$. Then $z \in \mathcal{B}(z_0; r)$, and all of the work above applies to give (4.1.7) and (4.1.8). The only reason that we needed to specify this "intermediate" radius r was to use the Cauchy integral formula (4.1.1). ■

This is where we finished on Wednesday, March 27, 2024.

The series (4.1.7) has a special name.

4.1.2 Definition. *Suppose that $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ is infinitely differentiable at $z_0 \in \mathcal{D}$. The* **TAYLOR SERIES OF f CENTERED AT z_0** *is the series*

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

Of course, this series is defined for any $z \in \mathbb{C}$ as a sequence of partial sums. Incidentally, calculus textbooks usually call the Taylor series at $z_0 = 0$ (if f is defined there and holomorphic on a ball centered at 0) the Maclaurin series. Outside of calculus classes, virtually no one uses this terminology.

Here are the major questions about Taylor series.

1. Does the series converge? That is, does

$$\lim_{n\to\infty} \sum_{k=0}^{n} \frac{f^{(k)}(z_0)}{k!}(z - z_0)$$

exist? Implicitly, for what z (if any) does this limit exist?

2. If this limit exists, is it $f(z)$? That is, do we have

$$f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(z_0)}{k!}(z - z_0)?$$

This second question is more nuanced than our previous questions of series convergence, because now we want to relate the value of the sum of the Taylor series of f specifically to f.

The good news is that Theorem 4.1.1 (mostly) answers these questions for holomorphic functions.

4.1.3 Corollary. *Let $f \colon \mathcal{D} \to \mathbb{C}$ be holomorphic and let $z_0 \in \mathcal{D}$. Suppose that $R > 0$ is such that $\mathcal{B}(z_0; R) \subseteq \mathcal{D}$. Then the Taylor series for f centered at z_0 converges to f on $\mathcal{B}(z_0; R)$. That is,*

$$f(z) = \sum_{k=0}^{\infty} \frac{f^{(k)}(z_0)}{k!}(z - z_0), \ \ z \in \mathcal{B}(z_0; R).$$

4.1.2. Interchange of series and integrals.

An interchange like (4.1.3) is, in the most abstract sense, a consequence of uniform convergence of a sequence/series of functions. However, we will not discuss the machinery of uniform convergence, as all of its applications in our course ultimately boil down to tractable arguments with geometric series. Here is the general structure of those arguments.

4.1.4 Theorem (Interchange). *Let $\mathcal{D} \subseteq \mathbb{C}$ and let f, $f_k \colon \mathcal{D} \to \mathbb{C}$ be continuous for $k \geq 0$. Suppose that for some $C > 0$ and $\rho \in (0, 1)$, the estimate*

$$|f_k(z)| \leq C\rho^k$$

holds for each $k \geq 0$ and all $z \in \mathcal{D}$. Then the following are true.

(i) *The series $\sum_{k=0}^{\infty} f_k(z)$ converges for each $z \in \mathcal{D}$.*

(ii) *The function*

$$f \colon \mathcal{D} \to \mathbb{C} \colon z \mapsto \sum_{k=0}^{\infty} f_k(z)$$

is continuous on $\mathcal{D}$.

(iii) *If γ is a path in $\mathcal{D}$, then*

$$\int_{\gamma} f = \sum_{k=0}^{\infty} \int_{\gamma} f_k. \tag{4.1.9}$$

That is, the series $\sum_{k=0}^{\infty} \int_{\gamma} f_k$ converges to $\int_{\gamma} f$.

4.1.5 Problem (+). Prove each part of the interchange theorem as outlined below.

(i) Use the comparison test and the geometric series to establish the convergence of the series $\sum_{k=0}^{\infty} f_k(z)$ for a given z.

(ii) For continuity, fix $z_0 \in \mathcal{D}$ and let $\epsilon > 0$. Choose an integer $n \geq 0$ such that $2C\rho^{n+1}(1-\rho) < \epsilon/2$. Then use the continuity of $f_0, \ldots, f_n$ to find $\delta > 0$ such that if $z \in \mathcal{D}$ with $|z - z_0| < \delta_k$, then $|f_k(z) - f_k(z_0)| < \epsilon/2(n+1)$. Last, estimate

$$|f(z) - f(z_0)| \leq \sum_{k=0}^{n} |f_k(z) - f_k(z_0)| + \sum_{k=n+1}^{\infty} \big(|f_k(z)| + |f_k(z_0)|\big).$$

Show that if $|z - z_0| < \min_{0 \leq k \leq n} \delta_k$, then both sums above are bounded by $\epsilon/2$. For the second sum, Problem 1.4.13 will be helpful.

(iii) Use the ML-inequality to show that for any $n \geq 0$,

$$\left|\sum_{k=0}^{n} \int_\gamma f_k - \int_\gamma\right| \leq \frac{C\ell(\gamma)\rho^{n+1}}{1-\rho}.$$

Then recall that $0 < \rho < 1$ and take the limit as $n \to \infty$.

4.1.6 Problem (+). Use the interchange theorem to prove (4.1.3) by estimating

$$\max_{|w-z_0|=r} \left| f(w) \frac{(z-z_0)^k}{(w-z_0)^{k+1}} \right| \leq M_r(f)\rho^k,$$

where

$$M_r(f) := \frac{1}{r} \max_{|w-z_0|=r} f(w) \quad \text{and} \quad \rho := \frac{|z-z_0|}{r}.$$

Explain why $\rho \in (0,1)$.

4.1.3. Power series.

We will now step away from (ostensibly) studying holomorphic functions to review some essential features of power series. We will return to discuss Taylor series extensively.

4.1.7 Definition. *Let (a_k) be a sequence in $\mathbb{C}$ and $z_0 \in \mathbb{C}$. The* **POWER SERIES CENTERED AT** z_0 **WITH COEFFICIENTS** (a_k) *is the series*

$$\sum_{k=0}^{\infty} a_k (z-z_0)^k.$$

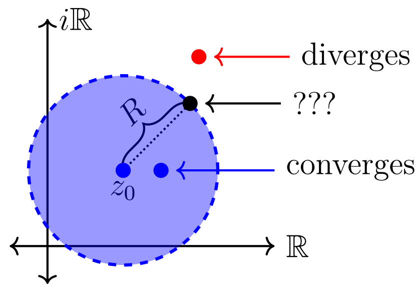
Recall that the symbol $\sum_{k=0}^{\infty} a_k(z-z_0)^k$ plays the dual role of denoting the sequence of partial sums $\left(\sum_{k=0}^{n} a_k(z-z_0)^k\right)$ and the limit of this sequence, if this limit exists. A power series carries z as an extra parameter, and so the convergence of a power series will depend on the value of z. In particular, a power series centered at z_0 always converges at $z = z_0$.

4.1.8 Problem (!). To what? Recall the convention of denoting $z^0 = 1$, even when $z = 0$.

We will now state a general convergence theorem for power series which we also likely saw for real power series in calculus. We will not prove it here, as the proof will not teach us anything new specifically about complex analysis.

4.1.9 Theorem. *Let (a_k) be a sequence in $\mathbb{C}$ and $z_0 \in \mathbb{C}$. There exists a unique (extended) real number $R \geq 0$ such that the power series $\sum_{k=0}^{\infty} a_k(z - z_0)^k$ converges for $|z - z_0| < R$ and diverges for $|z - z_0| > R$. This number R is the* **RADIUS OF CONVERGENCE** *of the power series.*

The radius of convergence R is an *extended* real number in the sense that we may have $R = \infty$, in which case the series converges for all $z \in \mathbb{C}$. The uniqueness of the radius of convergence means that if we have a number $r > 0$ such that the power series converges for $|z - z_0| < r$ and diverges for $|z - z_0| > r$, then r is the radius of convergence. Theorem 4.1.9 tells us nothing about what happens on the "boundary" of convergence, i.e., when $|z - z_0|$ equals the radius of convergence. Here is an illustration of what can happen for finite, positive R.



4.1.10 Problem (+). Let (a_k) be a sequence in $\mathbb{C}$ and let $z_0 \in \mathbb{C}$.

(i) Let $z_1 \in \mathbb{C} \setminus \{z_0\}$ such that the series $\sum_{k=0}^{\infty} a_k(z_1 - z_0)^k$ converges. Put $\rho = |z_1 - z_0| > 0$ and show that for some $C_\rho > 0$, the **FUNDAMENTAL ESTIMATE FOR POWER SERIES**

$$|a_k| \leq \frac{C_\rho}{\rho^k} \tag{4.1.10}$$

holds for all $k \geq 0$. [Hint: *use the test for divergence to show the existence of an integer $N \geq 0$ such that $|a_k|\rho^k \leq 1$ for all $k \geq N + 1$. Then put $C_\rho := \max_{0 \leq k \leq N} |a_k \rho^k|$.*]

(ii) Let $z_1 \in \mathbb{C} \setminus \{z_0\}$ such that the series $\sum_{k=0}^{\infty} a_k(z_1 - z_0)^k$ converges. Show that for all $z \in \mathbb{C}$ such that $|z - z_0| < |z_1 - z_0|$, the series $\sum_{k=0}^{\infty} a_k(z - z_0)^k$ converges absolutely. Conclude that the radius of convergence of the power series $\sum_{k=0}^{\infty} a_k(z - z_0)^k$ is at least $|z_1 - z_0|$. [Hint: *use the estimate* (4.1.10) *to compare $\sum_{k=0}^{\infty} |a_k(z - z_0)^k|$ to a geometric series.*]

(iii) Look ahead to Example 4.1.12 and explain why if the power series $\sum_{k=0}^{\infty} a_k(z-z_0)^k$ converges at $z = z_1$, then we cannot guarantee that it converges at all z with $|z - z_0| = |z_1 - z_0|$.

(iv) Suppose now that the radius of convergence of the power series $\sum_{k=0}^{\infty} a_k(z-z_0)^k$ is $R > 0$. Show that if $\rho \in (0, R)$, then there is $C_\rho > 0$ such that the estimate (4.1.10) holds for all $k \geq 0$. [Hint: *apply part (i) with* $z_1 = z_0 + \rho$.]

(v) With $R > 0$ as the radius of convergence of $\sum_{k=0}^{\infty} a_k(z-z_0)^k$, use the fundamental estimate for power series and the interchange theorem to show that the function $f(z) := \sum_{k=0}^{\infty} a_k(z-z_0)^k$ is continuous on $\mathcal{B}(z_0; R)$. [Hint: *fix* $0 < r < \rho < R$. *Show that* $|a_k(z-z_0)^k| \leq C_\rho(r/\rho)^k$ *when* $z \in \mathcal{B}(z_0; r)$. *Conclude by the interchange theorem that* f *is continuous on* $\mathcal{B}(z_0; r)$.]

While there is a formula for the radius of convergence R in terms of the coefficients (a_k), and while this formula *always* works, it is both complicated and unwieldy. Often it is best to use the ratio or root tests or to recognize the power series as the Taylor series for a holomorphic function. Indeed, we can paraphrase Theorem 4.1.1 and Corollary 4.1.3 in the following useful way.

4.1.11 Corollary. *Let* $f \colon \mathcal{D} \to \mathbb{C}$ *be holomorphic, let* $z_0 \in \mathcal{D}$, *and suppose* $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$ *for some* $r > 0$. *Then the radius of convergence of the Taylor series for* f *centered at* z_0 *is at least* r.

That is, for a holomorphic function, the radius of convergence of its Taylor series centered at some point in its domain is at least as large as the radius of any open ball centered at that point and contained in the domain. Unlike the Taylor series of a function of a real variable, we do not have to check any estimates on the remainder in the series; we just squeeze the largest open ball possible into the domain of our holomorphic function. Note the qualifying phrase "at least": it is possible for the Taylor series to converge on a larger ball than is contained in the domain. (Trivial example: take $\mathcal{D} = \mathcal{B}(0; 1)$ and $f \colon \mathcal{D} \to \mathbb{C} \colon z \mapsto 0$; then the Taylor series converges on $\mathbb{C}$.)

4.1.12 Example. As in real calculus, a power series may converge or diverge for $z \in \mathbb{C}$ with $|z - z_0| = R$. The behavior varies from series to series. It is even possible for a series to converge at some z with $|z - z_0| = R$ and diverge at others.

(i) The exponential power series

$$\sum_{k=0}^{\infty} \frac{z^k}{k!}$$

has center $z_0 = 0$ and coefficients $a_k = 1/k!$. Since this series converges for all $z \in \mathbb{C}$, as we saw in Example 1.4.21 via the ratio test (and as we have used ceaselessly since), the radius of convergence is $R = \infty$.

(ii) The familiar geometric series

$$\sum_{k=0}^{\infty} z^k$$

has center $z_0 = 0$ and coefficients $a_k = 1$. We saw in Theorem 1.4.12 that the geometric series converges for $|z| < 1$ and diverges for $|z| \geq 1$. We were even able to use some algebra and analysis to find a formula for the sum when $|z| < 1$. While the ratio test gave convergence for $|z| < 1$ and divergence for $|z| > 1$, we had to use other techniques to establish divergence at $|z| = 1$. In particular, the geometric series is an example of a power series that diverges at every point on its "boundary of convergence," i.e., at every point with $|z|$ equal to the radius of convergence.

(iii) Consider the series

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} z^k.$$

We use the ratio test and study

$$\left| \frac{(-1)^{k+1}}{(k+1)+1} z^{k+1} \cdot \frac{k+1}{(-1)^k z^k} \right| = |z| \frac{k+1}{k+2} \to |z| \text{ as } k \to \infty.$$

Strictly speaking, we should assume $z \neq 0$ here for the ratio test to work; at $z = 0$ we are guaranteed convergence, since that is the center of this power series. Thus (like our previous two examples) the series converges for $|z| < 1$ and diverges for $|z| > 1$.

When $|z| = 1$, we may have convergence or divergence: take $z = 1$ to see that the series is the alternating harmonic series

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} (1)^k = \sum_{k=0}^{\infty} (-1)^k \frac{1}{k+1} = -\sum_{j=1}^{\infty} \frac{(-1)^j}{j},$$

which converges. Take $z = -1$ to see that the series is the harmonic series

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k+1} (-1)^k = \sum_{k=0}^{\infty} \frac{1}{k+1} = \sum_{j=1}^{\infty} \frac{1}{j},$$

which diverges.

4.1.13 Problem (!). Use the ratio test to determine the radius of convergence R of the power series

$$\sum_{k=0}^{\infty} \frac{z^k}{k^2+1}.$$

Then use the comparison test to study the series when $|z| = R$.

4.1.14 Problem ($\star$). **Abel's test** for series convergence states that if (a_k) is a decreasing sequence of *real positive* numbers, i.e., $0 < a_{k+1} \leq a_k$ for all k, then the series $\sum_{k=0}^{\infty} a_k z^k$ converges for all $z \in \mathbb{C} \setminus \{1\}$ with $|z| = 1$. What does Abel's test say about the series in part (iii) of Example 4.1.12?

To determine the Taylor series for a function at a given point, we often have three options, which we list below from least to most preferred.

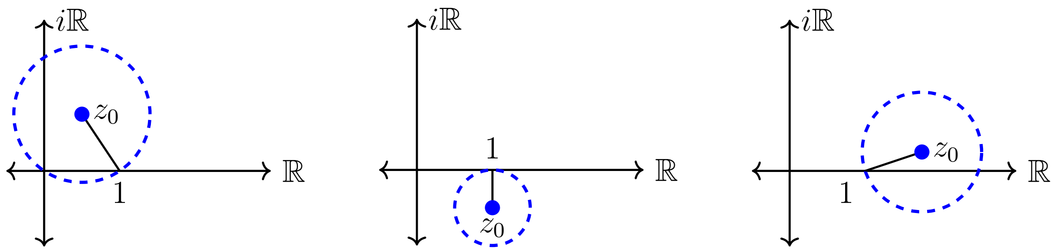
1. Calculate the coefficients using the generalized Cauchy integral formula, e.g., (4.1.5).

2. Calculate lots of derivatives of f and then use the fact that the kth coefficient is $f^{(k)}(z_0)/k!$.

3. Recognize f as some modification of a function whose Taylor series is known, and manipulate that known Taylor series.

4.1.15 Example. Let $f(z) = 1/(1-z)$. We know that f is holomorphic on $\mathbb{C} \setminus \{1\}$ and that $f(z) = \sum_{k=0}^{\infty} z^k$ for $|z| < 1$, but what is the Taylor series expansion for f centered at an arbitrary $z_0 \in \mathbb{C} \setminus \{1\}$, and what is the largest ball on which that series converges? There are several ways of proceeding here, which we broadly divide into geometric, analytic, and algebraic techniques.

(i) *Geometry.* We could draw pictures and just figure out what is the largest ball $\mathcal{B}(z_0; R)$ contained in $\mathbb{C} \setminus \{1\}$. Then we could use Theorem 4.1.1 or Corollary 4.1.11 to ensure convergence of the Taylor series on $\mathcal{B}(z_0; R)$. Note, though, that these results do not imply the *divergence* of the Taylor series outside $\mathcal{B}(z_0; R)$. Pretty quickly the pictures will convince us that $R = |1 - z_0|$. Then we would have to check convergence/divergence for $|z - z_0| > R$.



This is where we finished on Monday, April 1, 2024.

(ii) *Analysis.* We could differentiate f repeatedly and observe patterns:

$$f(z) = (1-z)^{-1}, \qquad f'(z) = -(1-z)^{-2}(-1) = (1-z)^{-2},$$
$$f''(z) = -2(1-z)^{-3}(-1) = 2(1-z)^{-3}, \qquad f^{(3)}(z) = -6(1-z)^{-4}(-1) = 6(1-z)^{-4}, \ldots$$

A formal induction argument establishes

$$f^{(k)}(z) = k!(1-z)^{-(k+1)},$$

and so the Taylor series for f centered at z_0 is

$$\sum_{k=0}^{\infty} \frac{k!(1-z_0)^{-(k+1)}}{k!}(z-z_0)^k = \sum_{k=0}^{\infty} \frac{1}{(1-z_0)^{k+1}}(z-z_0)^k. \tag{4.1.11}$$

Since f is holomorphic on the open set $\mathbb{C} \setminus \{1\}$, this Taylor series converges on any ball $\mathcal{B}(z_0; R)$ such that $\mathcal{B}(z_0; R) \subseteq \mathbb{C} \setminus \{1\}$. How can we find R just from the coefficients of this series? We could use the ratio test and calculate

$$\left|\frac{(z-z_0)^{k+1}}{(1-z_0)^{(k+1)+1}} \cdot \frac{(1-z_0)^{k+1}}{(z-z_0)^k}\right| = \frac{|z-z_0|}{|1-z_0|} \to \frac{|z-z_0|}{|1-z_0|} \text{ as } k \to \infty.$$

Thus the series converges for $|z-z_0| < |1-z_0|$ and diverges for $|z-z_0| > |1-z_0|$. Problem 4.1.16 discusses the divergence of this series when $|z-z_0| = |1-z_0|$.

(iii) *Algebra.* In lieu of the differentiation above, we could try to use a known Taylor series. Specifically, we would write, for z, $z_0 \in \mathbb{C} \setminus \{1\}$,

$$f(z) = \frac{1}{1-z} = \frac{1}{1-z_0+z_0-z} = \frac{1}{1-z_0-(z-z_0)} = \frac{1}{(1-z_0)\left[1-\left(\frac{z-z_0}{1-z_0}\right)\right]}$$

$$= \left(\frac{1}{1-z_0}\right)\left(\frac{1}{1-\left(\frac{z-z_0}{1-z_0}\right)}\right) = \frac{1}{1-z_0} f\left(\frac{z-z_0}{1-z_0}\right).$$

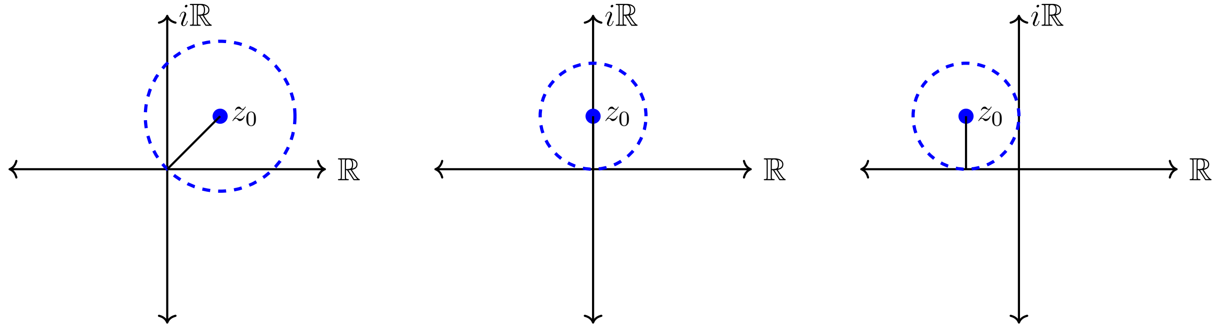
These calculations are just the tricks that we used with the geometric series earlier on the Cauchy integral formula. Since $f(w) = \sum_{k=0}^{\infty} w^k$ for $|w| < 1$, we therefore have

$$f(z) = \frac{1}{1-z_0}\sum_{k=0}^{\infty}\left(\frac{z-z_0}{1-z_0}\right)^k \qquad \text{for} \qquad \left|\frac{z-z_0}{1-z_0}\right| < 1,$$

and this gives the same Taylor series as above.

4.1.16 Problem (!). Use the test for divergence to show that the series (4.1.11) diverges when $|z-z_0| = |1-z_0|$.

4.1.17 Example. Since Log is holomorphic on $\mathbb{C} \setminus (-\infty, 0]$, its Taylor series centered at any $z_0 \in \mathbb{C} \setminus (-\infty, 0]$ converges on any ball $\mathcal{B}(z_0; r)$ such that $\mathcal{B}(z_0; r) \subseteq \mathbb{C} \setminus (-\infty, 0]$. Drawing some pictures, we might expect that the largest such ball has radius $|z_0|$ if $\text{Re}(z_0) \geq 0$ but radius $|\operatorname{Im}(z_0)|$ if $\operatorname{Re}(z_0) < 0$; this can be justified.



Now we find the actual Taylor series and compare its radius of convergence to what Taylor's theorem predicts. We compute some derivatives:

$$\operatorname{Log}'(z) = z^{-1}, \qquad \operatorname{Log}''(z) = -z^{-2}, \qquad \operatorname{Log}'''(z) = 2z^{-3}, \qquad \operatorname{Log}^{(4)}(z) = -6z^{-4}, \qquad \operatorname{Log}^{(5)}(z) = 24z^{-5}, \ldots,$$

and so, observing this pattern and/or inducting, we find

$$\operatorname{Log}^{(k)}(z) = (-1)^{k+1}(k-1)!z^{-k}.$$

Then the Taylor series for $\operatorname{Log}(\cdot)$ centered at any $z_0 \in \mathbb{C} \setminus (-\infty, 0]$ is

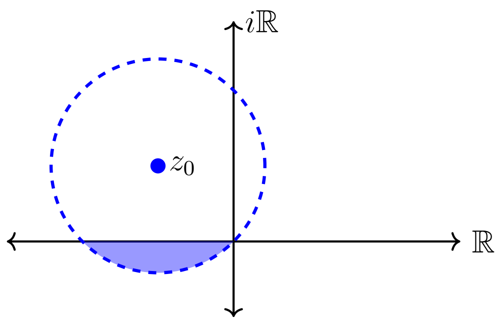
$$\sum_{k=0}^{\infty} \frac{f^{(k)}(z_0)}{k!}(z-z_0)^k = \operatorname{Log}(z_0) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}(k-1)!}{k!} z_0^{-k}(z-z_0)^k$$

$$= \operatorname{Log}(z_0) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{kz_0^k}(z-z_0)^k.$$

We can test the convergence of the series (starting with $k = 1$, since we can ignore finitely many terms in the series without affecting convergence) with the ratio test:

$$\left| \frac{(-1)^{(k+1)+1}}{(k+1)z_0^{k+1}}(z-z_0)^{k+1} \cdot \frac{kz_0^k}{(-1)^{k+1}(z-z_0)^k} \right| = \frac{k}{k+1}\left(\frac{|z-z_0|}{|z_0|}\right) \to \frac{|z-z_0|}{|z_0|} \text{ as } k \to \infty.$$

Consequently, the series converges if $|z - z_0| < |z_0|$, and so if $\operatorname{Re}(z_0) < 0$, then the series converges on a larger ball than can fit in the domain of Log. What happens outside this ball, i.e., on the shaded region below?



The ratio test tells us that the map

$$S\colon \mathcal{B}(z_0;|z_0|) \to \mathbb{C}\colon z \mapsto \operatorname{Log}(z_0) + \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{kz_0^k}(z-z_0)^k \tag{4.1.12}$$

is defined, and Taylor's theorem tells us that $S\big|_{\mathcal{B}(z_0;|\operatorname{Im}(z_0)|)} = \operatorname{Log}$. What, if anything, do the values of S at $z \in \mathcal{B}(z_0;|z_0|)$ with $\operatorname{Im}(z_0) \le 0$ have to do with Log?

4.1.18 Problem (!). Let $z_0 \in \mathbb{C}$ with $\operatorname{Re}(z_0) \ge 0$. Show that $\mathcal{B}(z_0;|z_0|) \cap (-\infty,0] = \varnothing$. [Hint: *show that if $x > 0$, then $|-x-z_0| > |z_0|$.*] Conclude that in this case, the largest ball on which the Taylor series for Log centered at z_0 converges is wholly contained in the domain of Log.

4.1.19 Problem ($\star$). This problem explores the behavior of the map S defined in (4.1.12) for $z_0 \in \mathbb{C}$ with $\operatorname{Re}(z_0) < 0$. For simplicity, we restrict to $\operatorname{Im}(z_0) > 0$, but similar results can be obtained for $\operatorname{Im}(z_0) < 0$. (Note that $z_0 \notin (-\infty,0]$, so Log is defined and holomorphic at z_0.)

(i) Show that $|\operatorname{Im}(z_0)| < |z_0|$. Explain why the Taylor series for Log centered at z_0 converges on a larger ball than is contained in the domain of Log.

(ii) Put $\theta_0 := \operatorname{Arg}(z_0)$; since $\operatorname{Re}(z_0) < 0$ and $\operatorname{Im}(z_0) > 0$, we have $\pi/2 < \theta_0 < \pi$. Show that

$$\theta_0 - \pi < 0 < \pi < (\theta_0 - \pi) + 2\pi,$$

so $[0,\pi] \subseteq (\theta_0 - \pi, (\theta_0 - \pi) + 2\pi]$.

(iii) Let $\mathcal{U} = \{z \in \mathbb{C} \mid \operatorname{Im}(z) > 0\}$, so $\mathcal{U}$ is the upper half-plane. With θ_0 as in the previous step, explain why $\operatorname{Arg}(z) = \arg_{\theta_0-\pi}(z)$ for all $z \in \mathcal{U}$ and therefore $\operatorname{Log}\big|_{\mathcal{U}} = \log_{\theta_0-\pi}\big|_{\mathcal{U}}$.

(iv) With z_0 and θ_0 as above, conclude that

$$S(z) = \begin{cases} \operatorname{Log}(z), & z \in \mathcal{B}(z_0;|z_0|) \cap \mathcal{U} \\ \operatorname{Log}(z) + 2\pi i, & z \in \mathcal{B}(z_0;|z_0|) \setminus \mathcal{U}. \end{cases}$$

[Hint: *argue first that the Taylor series for* Log *and* $\log_{\theta_0-\pi}$ *centered at* z_0 *have the same coefficients. Then, by considering the branch cut for* $\log_{\theta_0-\pi}$*, explain why* $\log_{\theta_0-\pi}$ *is analytic on* $\mathcal{B}(z_0;|z_0|)$*. Conclude that the Taylor series for* $\log_{\theta_0-\pi}$ *converges to* S *on* $\mathcal{B}(z_0;|z_0|)$*. What does this imply about* $S(z)$ *for* $z \in \mathcal{B}(z_0;|z_0|) \cap \mathcal{U}$*? To determine the value of* $S(z)$ *for* $z \in \mathcal{B}(z_0;|z_0|) \setminus \mathcal{U}$*, use the fact that* $\arg_{\theta_0-\pi}(z) = \operatorname{Arg}(z) + 2\pi$ *if* $\operatorname{Re}(z) < 0$ *and* $\operatorname{Im}(z) < 0$.]

4.1.20 Problem (!). **(i)** What is the Taylor series for Log centered at 1?

(ii) What is the Taylor series for

$$f\colon \mathbb{C}\setminus\{-1\}\to\mathbb{C}\colon z\mapsto \operatorname{Log}(z+1)?$$

(iii) How is all of this related to the series in part (iii) of Example 4.1.12 and its convergence on $\mathcal{B}(0;1)$?

This is where we finished on Wednesday, April 3, 2024.

Notwithstanding the oddities above, power series are some of the nicest functions in existence, because calculus-type computations with them are very easy—and so it is a wonder of nature that holomorphic functions are (locally) power series. Here is another theorem about power series that should be familiar from calculus; while it can be proved using methods of real-variable calculus (Problem 4.1.33), we will give some arguments illustrating the utility of the Cauchy integral theorem and formula and the fundamental estimate for power series. We might call this a "differentiation under the series" result in line with differentiating under the integral.

4.1.21 Theorem. *Suppose that the power series* $\sum_{k=0}^{\infty} a_k(z-z_0)^k$ *converges on* $\mathcal{B}(z_0;R)$*. Then the function* $f(z) := \sum_{k=0}^{\infty} a_k(z-z_0)^k$ *is holomorphic on* $\mathcal{B}(z_0;R)$ *with*

$$f^{(n)}(z)=\sum_{k=n}^{\infty}\left(\prod_{j=0}^{n-1}(k-j)\right)a_k(z-z_0)^{k-n}=\sum_{k=n}^{\infty}\frac{k!}{(k-n)!}a_k(z-z_0)^{k-n} \tag{4.1.13}$$

for each $z\in\mathcal{B}(z_0;R)$ *and each integer* $n\geq 0$*. In particular, the series in* (4.1.13) *converges on* $\mathcal{B}(z_0;R)$*, and*

$$a_k=\frac{f^{(n)}(z_0)}{k!}. \tag{4.1.14}$$

4.1.22 Problem (!). To motivate the equality (4.1.13), try differentiating $f(z)=z^k$ some n times, observe patterns, and try to rewrite the coefficients in the derivatives as quotients of factorials. For example, calculate $f',\ldots,f^{(6)}$ for $f(z)=z^5$.

Proof. First, the function f is continuous on $\mathcal{B}(z_0;R)$ by part (v) of Problem 4.1.10. Note that (4.1.14) follows directly from (4.1.13) by substituting $z=z_0$.

For simplicity, we start by assuming $z_0=0$, so the power series $f(z)=\sum_{k=0}^{\infty} a_k z^k$ converges on $\mathcal{B}(0;R)$.

1. *Sketch of the proof of holomorphy.* We will apply Morera's theorem (part (iii) of Problem 3.6.21) and show that $\int_\gamma f=0$ for any closed curve in $\mathcal{B}(0;R)$. Note that Morera's theorem can apply because f is continuous. Ideally we would do this with a simple interchange of summation and integration:

$$\int_\gamma f(z)\,dz = \int_\gamma \sum_{k=0}^{\infty} a_k z^k\,dz = \sum_{k=0}^{\infty} a_k \int_\gamma z^k\,dz = 0,$$

since each integral $\int_\gamma z^k\,dz$ vanishes by the fundamental theorem of calculus (or the Cauchy integral theorem). However, justifying the interchange is, as usual, a bit delicate.

2. *Sketch of the proof of the identity* (4.1.13). First we prove the case $n = 1$. We will use the Cauchy integral formula to obtain the series representation of the derivatives. Fix $z \in \mathcal{B}(0; R)$ and let $|z| < r < R$. Then

$$f'(z) = \frac{1}{2\pi i}\int_{|z|=r} \frac{f(w)}{(w-z)^2}\,dw = \frac{1}{2\pi i}\int_{|z|=r} \sum_{k=0}^{\infty} a_k \frac{w^k}{(w-z)^2}\,dw. \tag{4.1.15}$$

If we can interchange the series and the integral, then we have

$$f'(z) = \sum_{k=0}^{\infty} a_k \left(\frac{1}{2\pi i}\int_{|z|=r} \frac{w^k}{(w-z)^2}\,dw\right), \tag{4.1.16}$$

and the *generalized* Cauchy integral formula gives

$$\frac{1}{2\pi i}\int_{|z|=r} \frac{w^k}{(w-z)^2}\,dw = \begin{cases} 0, \ k=0 \\ kz^{k-1}, \ k \geq 1. \end{cases}$$

The case of a general $n \geq 1$ in (4.1.13) follows by induction.

3. *Rigorous proof of holomorphy.* Let $0 < r < R$. We will show that f is holomorphic on $\mathcal{B}(0; r)$ and that $f'(z) = \sum_{k=1}^{\infty} k a_k z^{k-1}$ for all $z \in \mathcal{B}(0; r)$; in particular, the series $\sum_{k=1}^{\infty} k a_k z^{k-1}$ converges on $\mathcal{B}(0; r)$. Then given $z \in \mathcal{B}(0; R)$, we just take r such that $|z| < r < R$ to conclude that f is differentiable at z with $f'(z)$ in the desired form. The identity (4.1.13) then follows by induction on n.

So, with $0 < r < R$, let $\rho > 0$ satisfy $r < \rho < R$. The fundamental estimate (4.1.10) for power series provides $C_\rho > 0$ such that $|a_k| < C_\rho \rho^{-k}$ for all k. Let γ be a closed curve in $\mathcal{B}(0; r)$; if $z \in \text{image}(\gamma)$, then $|z| < r$. Consequently,

$$|a_k z^k| < C_\rho \left(\frac{r}{\rho}\right)^k.$$

Since $r/\rho < 1$, the interchange theorem (Theorem 4.1.4) applies to show

$$\int_\gamma f = \int_\gamma \sum_{k=0}^{\infty} a_k z^k\,dz = \sum_{k=0}^{\infty} a_k \int_\gamma z^k\,dz,$$

and each integral $\int_\gamma z^k\,dz$ vanishes since each integrand is a polynomial and therefore holomorphic, and γ is closed.

4. *Rigorous proof of the identity* (4.1.13). The work in the sketch shows that we only need to justify the interchange of integration from (4.1.15) to (4.1.16). So, let $z \in \mathcal{B}(0; R)$ with $|z| < r < \rho < R$. By Problem 3.1.26, there is $d_0 > 0$ such that $d_0 \le |w - z|$ for all $w \in \mathbb{C}$ with $|w| = r$. Then the fundamental estimate for power series gives

$$\left|\frac{a_k w^k}{(w-z)^2}\right| \le \frac{C_\rho r^k}{d_0^2 \rho^k} = \frac{C_\rho}{d_0^2}\left(\frac{r}{\rho}\right)^k.$$

Since $r/\rho < 1$, the interchange lemma applies to show

$$\int_{|z|=r} \sum_{k=0}^{\infty} a_k \frac{w^k}{(w-z)^2}\,dw = \sum_{k=0}^{\infty} \int_{|z|=r} a_k \frac{w^k}{(w-z)^2}\,dw,$$

as desired. ■

4.1.23 Problem ($\star$). Complete the proof of Theorem 4.1.21 in the following steps.

(i) The proof above assumed $z_0 = 0$. Suppose now that the series $f(z) := \sum_{k=0}^{\infty} a_k (z - z_0)^k$ converges on $\mathcal{B}(z_0; R)$. Explain why $g(w) := \sum_{k=0}^{\infty} a_k w^k$ converges on $\mathcal{B}(0; R)$. Obtain $g'(w) = \sum_{k=1}^{\infty} k a_k w^{k-1}$ and then use $f(z) = g(z - z_0)$ to obtain the desired result for f.

(ii) Use induction on n to prove the general formula (4.1.13).

4.1.24 Example. Recognizing a given power series as the derivative of another is a useful skill. After staring at the series

$$\sum_{k=2}^{\infty} k(k-1) z^k$$

for a while, hopefully we agree that it looks like a second derivative, since the starting index is 2. Missing are the factorial quotient $k!/(k-2)!$ and the power z^{k-2}. For $k \ge 2$, we calculate

$$\frac{k!}{(k-2)!} = \frac{k(k-1)(k-2)!}{(k-2)!} = k(k-1),$$

and so all that is "wrong" in this series is the power of z. We therefore rewrite

$$\sum_{k=2}^{\infty} k(k-1) z^k = z^2 \sum_{k=2}^{\infty} \frac{k!}{(k-2)!} z^{k-2} = z^2 \left(\frac{d^2}{dz^2} \sum_{k=0}^{\infty} z^k\right) = z^2 \frac{d^2}{dz^2}\left[\frac{1}{1-z}\right]$$
$$= z^2 \left(\frac{2}{(1-z)^3}\right) = \frac{2z^2}{(1-z)^3}.$$

The work above illustrates our thought process in obtaining a "closed-form" expression for the series, but it is not the most rigorous way of proceeding. Instead, we might define

$$f \colon \mathbb{C} \setminus \{1\} \to \mathbb{C} \colon z \mapsto \frac{1}{1-z},$$

so

$$f(z) = \sum_{k=0}^{\infty} z^k \quad \text{and} \quad f''(z) = \sum_{k=2}^{\infty} \frac{k!}{(k-2)!} z^{k-2} = \sum_{k=2}^{\infty} k(k-1)z^{k-2},$$

with both equalities holding for $|z| < 1$. For any $z \neq 1$, we have $f''(z) = 2/(1-z)^3$, and so

$$\sum_{k=2}^{\infty} k(k-1)z^k = z^2 \sum_{k=2}^{\infty} k(k-1)z^{k-2} = z^2 f''(z) = \frac{2z^2}{(1-z)^3}.$$

4.1.25 Problem (!). Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f: \mathcal{D} \to \mathbb{C}$ be analytic. Suppose that f' is constant on $\mathcal{D}$. Find $a, b \in \mathbb{C}$ such that $f(z) = az + b$ for all $z \in \mathcal{D}$. [Hint: *what are the Taylor coefficients of* f*?*]

4.1.26 Problem (⋆). The coefficients of a power series are unique in the following sense. Let $z_0 \in \mathbb{C}$ and let (a_k) and (b_k) be sequences in $\mathbb{C}$ such that for some $R > 0$,

$$\sum_{k=0}^{\infty} a_k(z - z_0)^k = \sum_{k=0}^{\infty} b_k(z - z_0)^k$$

for all $z \in \mathcal{B}(z_0; R)$. In particular, both series converge on all of $\mathcal{B}(z_0; R)$. Show that $a_k = b_k$ for all k. [Hint: *let* f *be the difference of the series and use Theorem 4.1.21.*]

4.1.27 Problem (⋆). Here is another proof of Liouville's theorem (Theorem 3.6.23). Suppose that $f: \mathbb{C} \to \mathbb{C}$ is entire; explain why

$$f(z) = \sum_{k=0}^{\infty} a_k z^k, \qquad a_k := \frac{f^{(k)}(0)}{k!}$$

for all $z \in \mathbb{C}$. If $M > 0$ satisfies $|f(z)| \leq M$ for all $z \in \mathbb{C}$, use the generalized Cauchy integral formula to show

$$|a_k| \leq Mr^{-k}$$

for all $r > 0$ and $k \geq 0$. Send $r \to \infty$ and conclude $a_k = 0$ for $k \geq 1$.

4.1.4. Analytic functions.

We have shown that every holomorphic function is "locally" a power series, and, conversely, that every power series defines a holomorphic function. The latter result is what we expected from real variable calculus, which taught us that power series are infinitely differentiable; what is new in complex analysis is that differentiable functions are really power series, at least locally. (Recall that in analysis, something is "locally" true if it is true on a ball.) Here is a more compact and standard way of saying "locally a power series."

4.1.28 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$. A function $f \colon \mathcal{D} \to \mathbb{C}$ is* **ANALYTIC** *on $\mathcal{D}$ if for each $z_0 \in \mathcal{D}$, there is $r > 0$ and a sequence (a_k) in $\mathbb{C}$ such that*

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k \tag{4.1.17}$$

for each $z \in \mathcal{B}(z_0; r) \cap \mathcal{D}$.

Theorems 4.1.1 and 4.1.21 combine to tell us that analytic functions are precisely the holomorphic functions.

4.1.29 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ be open. A function $f \colon \mathcal{D} \to \mathbb{C}$ is analytic if and only if f is holomorphic, in which case the series expansion (4.1.17) of f about a point $z_0 \in \mathcal{D}$ is its Taylor series.*

4.1.30 Example. Many familiar functions are analytic on $\mathbb{C}$ because of how we chose to define them as power series. This includes the exponential, the sine, and the cosine:

$$e^z = \exp(z) = \sum_{k=0}^{\infty} \frac{1}{k!} z^k, \qquad \sin(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1}, \quad \text{and} \quad \cos(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k}.$$

The appearance of these functions is somewhat special because we have chosen to expand them as series centered at $z_0 = 0$ and because these series converge for all z. In general, the power series expansion (4.1.17) of an analytic function need not be valid on all of the function's domain $\mathcal{D}$, just on some open ball contained in $\mathcal{D}$. For the exponential, the sine, and the cosine, the expansions above are valid on all balls $\mathcal{B}(0; r)$ for any $r > 0$.

4.1.31 Remark. *It is not ideal to say that a function is analytic at a point; in this sense, using analytic as a synonym for holomorphic or differentiable is wrong. Analyticity is a property on a whole set; differentiability is a property localized at a point. Changing the values of a function somewhere on its domain can destroy analyticity on the domain but not differentiability at a point; changing the values of a function at a point can destroy differentiability.*

Although we do not usually employ the terminology "analytic" in real-variable calculus, it is entirely possible for a function defined on (a subinterval of) $\mathbb{R}$ to be analytic in the sense that for each point in that interval, the function equals its Taylor series around that point. Indeed, that is probably how we first rigorously met the exponential and trigonometric functions in calculus.

4.1.32 Definition. *Let $I \subseteq \mathbb{R}$ be an interval. A function $f \colon I \to \mathbb{R}$ is* **REAL ANALYTIC** *on I if for each $t_0 \in I$, there is a sequence of real numbers (a_k) and a real number $r > 0$*

such that for $t \in (t_0 - r, t_0 + r) \cap I$,

$$f(t) = \sum_{k=0}^{\infty} a_k (t - t_0)^k. \tag{4.1.18}$$

Theorem 4.1.21 holds for real analytic functions: a real analytic function is infinitely differentiable, and in the expansion (4.1.18), the coefficients satisfy $a_k = f^{(k)}(t_0)/k!$.

4.1.33 Problem (+). We can prove the $n = 1$ case of Theorem 4.1.21 without using the Cauchy theorems; indeed, this is how the proof works for real analytic functions. For simplicity, take $z_0 = 0$, so $f(z) = \sum_{k=0}^{\infty} a_k z^k$ converges on $\mathcal{B}(0; R)$.

Fix $0 < r < R$. We show that $\sum_{k=1}^{\infty} k a_k z^{k-1}$ converges on $\mathcal{B}(0; r)$ and that f is differentiable on $\mathcal{B}(0; r)$ with $f'(z) = \sum_{k=1}^{\infty} k a_k z^{k-1}$.

(i) First we show the convergence. Let $r < \rho < R$ and assume $|z| < r$. Use the fundamental estimate for power series (4.1.10) to show that for some $C_\rho > 0$ and all $k \geq 1$, the estimate

$$|k a_k z^{k-1}| \leq \frac{C_\rho}{r} k \left(\frac{r}{\rho}\right)^k$$

holds. Then show that the series

$$\sum_{k=1}^{\infty} k \left(\frac{r}{\rho}\right)^k$$

converges.

(ii) Continue to assume $|z| < r < \rho < R$. We show differentiability by establishing

$$\lim_{h \to 0} \frac{1}{h} \left[f(z+h) - f(z) - h \sum_{k=1}^{\infty} k a_k z^{k-1} \right] = 0.$$

First show

$$f(z+h) - f(z) - h \sum_{k=1}^{\infty} k a_k z^{k-1} = \sum_{k=2}^{\infty} a_k \big[(z+h)^k - z^k - k z^{k-1} \big].$$

Then, for $k \geq 2$, use the binomial theorem to obtain

$$(z+h)^k - z^k - k z^{k-1} = \sum_{j=0}^{k} \binom{k}{j} z^{k-j} h^j - z^k - k z^{k-1} = h^2 \sum_{\ell=0}^{k-2} \binom{k}{\ell+2} z^{k-(\ell+2)} h^\ell.$$

Next, choose $\delta > 0$ such that $r + \delta < \rho$, and assume $|h| < \delta$. Show that

$$\left| \sum_{\ell=0}^{k-2} \binom{k}{\ell+2} z^{k-(\ell+2)} h^\ell \right| \leq \frac{(r+\delta)^k}{\delta^2}$$

and conclude

$$\left|\frac{1}{h}\left[f(z+h)-f(z)-h\sum_{k=1}^{\infty}ka_kz^{k-1}\right]\right| \le \frac{C_\rho|h|}{\delta^2}\sum_{k=0}^{\infty}\left(\frac{r+\delta}{\rho}\right)^k.$$

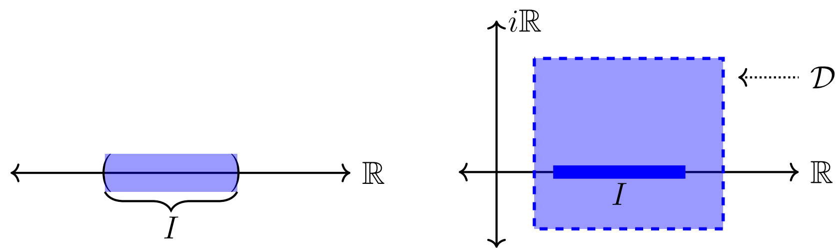
Why is this a good conclusion?

However, Theorem 4.1.29 does not remain true for real analytic functions. There are plenty of infinitely differentiable functions on $\mathbb{R}$ that are not real analytic; a classical counterexample is

$$f(t) := \begin{cases} e^{-1/t^2}, & t \neq 0 \\ 0, & t = 0. \end{cases}$$

One can show that f is infinitely differentiable on $\mathbb{R}$ and $f^{(k)}(0) = 0$ for all k. Thus the Taylor series for f centered at 0 converges to the zero function $\mathbb{R}$, and that is definitely not f. This is in line with our previous remarks that, as we know well from calculus, a function on $\mathbb{R}$ can be n-times differentiable but not $(n+1)$-times differentiable. Differentiability on $\mathbb{C}$ is much stronger: the existence of one derivative guarantees the existence of all derivatives *and* the convergence of the Taylor series back to the original function to boot.

But in the happy case that we do have a real analytic function $f\colon I \subseteq \mathbb{R} \to \mathbb{R}$, can we extend it to an analytic function on some open set $\mathcal{D} \subseteq \mathbb{C}$ with $I \subseteq \mathcal{D}$? After all, we did that quite successfully with the exponential and trigonometric functions.



Such an extension has a formal name.

4.1.34 Definition. *Let $\mathcal{D}_0 \subseteq \mathcal{D} \subseteq \mathbb{C}$. A function $f\colon \mathcal{D} \to \mathbb{C}$ is an* **ANALYTIC CONTINUATION** *of a function $f_0\colon \mathcal{D}_0 \to \mathbb{C}$ if f is analytic and if $f\big|_{\mathcal{D}_0} = f_0$, i.e., if $f(z) = f_0(z)$ for all $z \in \mathcal{D}_0$.*

So, when does a real analytic function have an analytic continuation from a real interval to an open subset of the plane? And if a function has an analytic continuation, is that continuation unique? That is, could a function f_0 have two analytic continuations, f_1 and f_2, with $f_1 \neq f_2$? Such a possibility should be frightening, as it might mean that there is more than one way to extend, say, the exponential to the plane—and so perhaps we have been working with the wrong exponential all along!

Of course, this is nonsense. Analytic continuations, if they exist, surely must be unique. The question is how to show it.

Forcing two functions f_1 and f_2 to be the same is really saying that $f_1 - f_2 = 0$. And so we will take up the study of the *zeros* of an analytic function: if f is analytic, what can we say about those z at which $f(z) = 0$? In particular, what is the minimum amount of data about a function that we need to conclude that it is *always* zero? (Not much.)

4.2. The zeros of an analytic function.

Power series are, euphemistically, "just" polynomials of "infinite" degree. A spot of work with the roots of polynomials, then, will motivate some of the broader results on the zeros of analytic functions that we will develop.

4.2.1. Roots of polynomials.

Let $f(z) = \sum_{k=0}^{n} a_k z^k$ be a polynomial of degree $n \geq 1$. Note that this formula for f is its Taylor expansion centered at 0, since $f^{(k)}(z) = 0$ for all integers $k \geq n+1$ and all $z \in \mathbb{C}$. By the fundamental theorem of algebra, f has a root $z_1 \in \mathbb{C}$. Since f is entire, we may expand f as a power series centered at z_1: $f(z) = \sum_{k=0}^{\infty} b_k (z - z_1)^k$. Here $b_0 = f(z_1) = 0$, and also $b_k = f^{(k)}(z_1)/k! = 0$ for $k \geq n+1$. Thus $f(z) = \sum_{k=1}^{n} b_k (z - z_1)^k$, and so we may factor

$$f(z) = (z - z_1)\sum_{k=1}^{n} b_k (z - z_1)^{k-1} = (z - z_1)p_1(z), \qquad p_1(z) = \sum_{j=0}^{n-1} b_{j+1}(z - z_1)^j.$$

We now recognize p_1 as a polynomial of degree $n-1$; if $n = 1$, then p_1 is constant, and in particular $p_1(z_1) \neq 0$. Otherwise, $f = 0$, and then f would not be a polynomial of degree at least 1. If $n \geq 2$, then either $p_1(z_1) \neq 0$, or $p_1(z_1) = 0$, in which case we can repeat the argument above and factor

$$p_1(z) = (z - z_1)p_2(z),$$

where p_2 is a polynomial of degree $n - 2$. In this case, we can rewrite

$$f(z) = (z - z_1)^2 p_2(z).$$

And then the process continues to allow us to conclude that for some integer $m_1 \geq 1$, there is a polynomial p_1 of degree $n - m_1$ such that $p_1(z_1) \neq 0$ and

$$f(z) = (z - z_1)^{m_1} p_1(z). \tag{4.2.1}$$

We want to call the integer m_1 the **MULTIPLICITY** or **ORDER** of z_1 as a root of f. As with most integer-dependent processes, a rigorous proof of the factorization (4.2.1) would use induction on n.

We could go further from (4.2.1) and say that, if $m < n$, then p_1 is a polynomial of degree at least 1, and therefore p_1 has a root z_2. Note that $z_2 \neq z_1$ since $p_1(z_1) \neq 0$. Then we could write $p_1(z) = (z - z_2)^{m_2} p_2(z)$, where $p_2(z_1) \neq 0$. And so on. Eventually we would factor

$$f(z) = a(z - z_1)^{m_1} \cdots (z - z_r)^{m_r}, \tag{4.2.2}$$

where $z_1, \ldots, z_r \in \mathbb{C}$ are distinct and $m_1, \ldots, m_r \geq 1$ are integers with $m_1 + \cdots + m_r = n$. The coefficient $a \in \mathbb{C} \setminus \{0\}$ is the constant polynomial that arises from the very last factorization of p_r, i.e., $p_r(z) = (z - z_r)^{m_r} a$. This factorization *is* the fundamental theorem of algebra, and a rigorous proof also needs induction.

This is where we finished on Friday, April 5, 2024.

Somewhat more important than the full factorization (4.2.2) for our work on the zeros of analytic is that if f is a polynomial with $f(z_0) = 0$, then $f(z) = (z - z_0)^m p(z)$ for some integer $m \geq 1$ and some polynomial p such that $p(z_0) \neq 0$.

4.2.1 Example. Let $f(z) = z^2 - 1$, so $f(z) = (z-1)(z+1)$. While this is the full factorization (4.2.2), in light of the previous remark consider just the root $z_0 = 1$. Here $p(z) = z + 1 = z - (-1)$, since $p(1) = 2 \neq 0$ and $f(z) = (z-1)p(z)$, so $m = 1$.

4.2.2. *Isolated zeros.*

Viewing analytic functions as "infinite degree polynomials," we will see just how much the behavior of zeros of analytic functions resembles the results above for polynomials.

4.2.2 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and suppose that $f \colon \mathcal{D} \to \mathbb{C}$ is analytic. Let $z_0 \in \mathcal{D}$ such that $f(z_0) = 0$ and take $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$. Then one, and only one, of the following holds:*

(i) *$f(z) = 0$ for all $z \in \mathcal{B}(z_0; r)$.*

(ii) *There is an analytic function $g \colon \mathcal{B}(z_0; r) \to \mathbb{C}$ and an integer $m \geq 1$ such that $f(z) = (z - z_0)^m g(z)$ for $z \in \mathcal{B}(z_0; r)$ and, additionally, $g(z_0) \neq 0$. The integer m is the smallest integer k such that $f^{(k)}(z_0) \neq 0$, and $g(z_0) = f^{(m)}(z_0)$. Moreover, there is $\rho \in (0, r]$ such that $f(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; \rho)$.*

Proof. Write $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ for $z \in \mathcal{B}(z_0; r)$, where $a_k = f^{(k)}(z_0)/k!$. We consider the following two cases on the coefficients.

(i) $a_k = 0$ for all k. Since $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ for all $z \in \mathcal{B}(z_0; r)$, we then have $f(z) = 0$ for all $z \in \mathcal{B}(z_0; r)$. This is (i).

(ii) There is $n \geq 1$ such that $a_n \neq 0$. Note that $a_0 = f(z_0) = 0$, so this is only possible for some $n \geq 1$. Now let $m \geq 1$ be the *smallest* integer satisfying $a_m \neq 0$. (That such a smallest integer exists is a consequence of the well-ordering property of the positive integers.) We may then write

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k = \sum_{k=m}^{\infty} a_k (z - z_0)^k = \sum_{j=0}^{\infty} a_{j+m} (z - z_0)^{j+m} = (z - z_0)^m \sum_{j=0}^{\infty} a_{j+m} (z - z_0)^j. \tag{4.2.3}$$

These equalities are valid for $z \in \mathcal{B}(z_0; r)$.

Then, for $z \in \mathcal{B}^*(z_0; r)$, we have

$$\sum_{j=0}^{\infty} a_{j+m}(z-z_0)^j = \frac{f(z)}{(z-z_0)^m}.$$

That is, the series on the left converges for $z \in \mathcal{B}^*(z_0; r)$, and certainly the series converges at $z = z_0$. Thus the map

$$g \colon \mathcal{B}(z_0; r) \to \mathbb{C} \colon z \mapsto \sum_{j=0}^{\infty} a_{j+m}(z-z_0)^j$$

is analytic. Moreover, we have the factorization $f(z) = (z-z_0)^m g(z)$ from (4.2.3), and by definition of g we compute $g(z_0) = a_m \neq 0$. This is (ii), except for the claim about ρ, which we leave as a problem below. ■

4.2.3 Problem (!). Prove the result above about ρ. [Hint: *use the continuity of g and the fact that $g(z_0) \neq 0$.*]

Case (ii) above has a special name.

4.2.4 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f \colon \mathcal{D} \to \mathbb{C}$ be analytic. Let $z_0 \in \mathcal{D}$ and let $m \geq 1$ be an integer. Then z_0 is an* **ZERO OF f OF ORDER (MULTIPLICITY) m** *if for some $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$, there is an analytic function $g \colon \mathcal{B}(z_0; r) \to \mathbb{C}$ such that $f(z) = (z-z_0)^m g(z)$ for $z \in \mathcal{B}(z_0; r)$ with $g(z_0) \neq 0$. In the case $m = 1$, the zero is sometimes called* **SIMPLE**.

4.2.5 Example. We find the zeros and their orders for several different functions.

(i) $f_1(z) = z^2$ on $\mathbb{C}$. Here $f_1(z) = 0$ if and only if $z = 0$, and we can basically read off from the definition of f_1 that 0 has order 2. Indeed, with $g(z) = 1$ for all z, we have $f_1(z) = (z-0)^2 g(z)$, and certainly $g(0) \neq 0$.

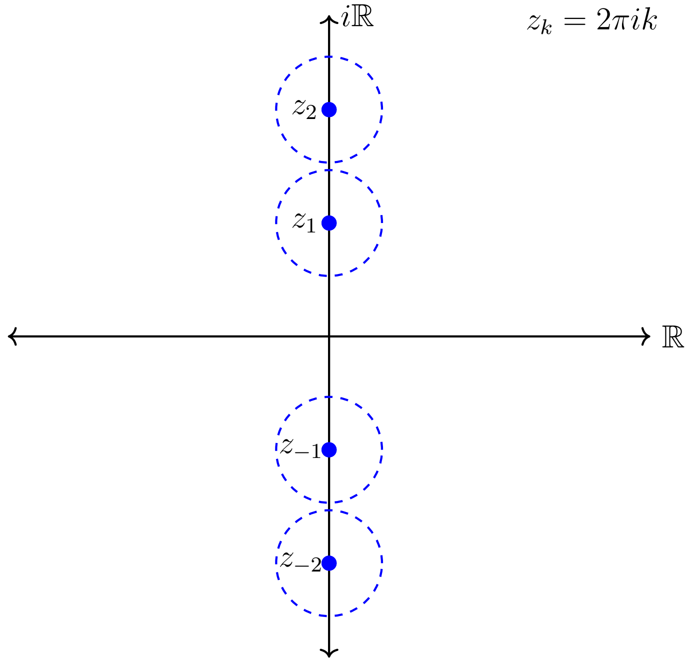
(ii) $f_2(z) = \operatorname{Log}(z)$ on $\mathbb{C} \setminus \{0\}$. Since $\operatorname{Log}(z) = \ln(|z|) + i \operatorname{Arg}(z)$, we have $\operatorname{Log}(z) = 0$ if and only if both $\ln(|z|) = 0$ and $\operatorname{Arg}(z) = 0$. First, $\ln(|z|) = 0$ if and only if $|z| = 1$, which happens if and only if $z = e^{it}$ for some $t \in (-\pi, \pi]$. So, at the very least, all zeros of Log lie on the unit circle. Next, since $-\pi < t \leq \pi$, we have $\operatorname{Arg}(e^{it}) = t$, and so $\operatorname{Arg}(e^{it}) = 0$ if and only if $t = 0$. Thus the only zero of Log is $e^{i \cdot 0} = 1$.

There is no transparent factorization of Log as $\operatorname{Log}(z) = (z-1)g(z)$ for some explicit function g, so we calculate derivatives to check the order of 1 as a root of Log. We do not have to go far: $\operatorname{Log}'(z) = 1/z$, so $\operatorname{Log}'(1) = 1 \neq 0$. Thus 1 is a zero of order 1 of Log, i.e., a simple zero.

(iii) $f_3(z) = e^{2z} - 2e^z + 1$ on $\mathbb{C}$. Here we use the factorization $w^2 - 2w + 1 = (w-1)^2$ to write $f_3(z) = (e^z - 1)^2$. Then $f_3(z) = 0$ if and only if $e^z = 1$, so the zeros of f are the numbers $2\pi ik$ for $k \in \mathbb{Z}$. We calculate $f_3'(z) = 2(e^z - 1)$, so $f_3'(2\pi ik) = 0$, and $f_3''(z) = 2e^z$, so $f_3''(2\pi ik) = 2 \neq 0$. Each zero therefore has order 2.

4.2.6 Problem (!). Show that each zero of the sine is simple.

The functions f_1 and f_2 in the preceding example each had only one zero, although they had different orders. The function f_3 had infinitely many zeros, all of the same order, but these zeros relate to each other "pairwise" in a special way: they are "isolated" from each other. We can see this just by plotting the zeros: around each zero we can draw a ball that does not intersect a neighboring ball.



We formalize this geometric observation.

4.2.7 Definition. *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f: \mathcal{D} \to \mathbb{C}$ be analytic. A point $z_0 \in \mathcal{D}$ is an* **ISOLATED ZERO OF** *f if there is $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$ and $f(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; r)$.*

4.2.8 Problem (!). Let $\mathcal{D} \subseteq \mathbb{C}$ be open, let $f: \mathcal{D} \to \mathbb{C}$ be analytic, and let $z_0 \in \mathcal{D}$ with $f(z_0) = 0$. Prove the following. [Hint: *the following are mostly consequences of Definition 4.2.7 and the mutual exclusivity of the two conclusions of Theorem 4.2.2.*]

(i) Prove that z_0 is an isolated zero of f if and only if z_0 is a zero of order m of f for some $m \geq 1$.

(ii) Prove that z_0 is not an isolated zero of f if and only if there is $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$ with $f(z) = 0$ for all $z \in \mathcal{B}(z_0; r)$.

So, if z_0 is an isolated zero of f, then z_0 is the only zero of f in this ball $\mathcal{B}(z_0; r)$; outside the ball, f certainly may have zeros.

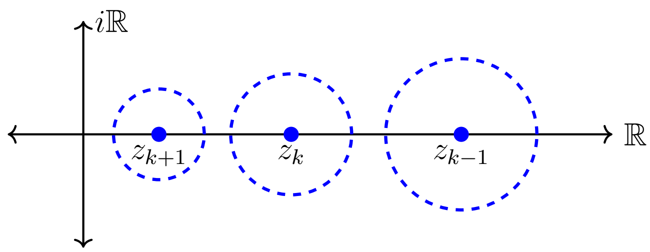
4.2.9 Problem (!). Show that there exists an entire function g such that $\sin(z) = zg(z)$. Explain why g must have some zeros.

Additionally, for different isolated zeros of the same function, there are no guarantees about the relative sizes of the balls surrounding them and excluding other zeros. In Example 4.2.5, the zeros of f_3 were a nice, uniform distance away from each other. The same is true for the roots of polynomials.

4.2.10 Problem (!). Explain why. Specifically, explain why if f is a polynomial, then there exists $d > 0$ such that if $f(z) = f(w) = 0$, then $|z - w| \geq d$.

This does not always happen.

4.2.11 Example. Let $\mathcal{D} = \mathbb{C} \setminus \{0\}$ and let $f(z) = \sin(\pi/z)$. Then f is analytic on $\mathcal{D}$ and $f(z) = 0$ if and only if $\pi/z = k\pi$ for some integer k. That is, the zeros of f are the numbers $z_k = 1/k$. These numbers are definitely isolated; after a bit of algebra, we can find $r_k > 0$ such that if $|z - 1/k| < r_k$, then $z \neq 1/j$ for any integer $j \neq k$. But note that $z_k \to 0$ as $k \to \infty$, and in particular the distance between successive zeros z_k and z_{k+1} shrinks as $k \to \infty$.



Although we cannot guarantee that the zeros of an analytic function are *all* a minimum distance apart, we can be assured that they are isolated, at least for a function that is not always zero. In other words, the only "interesting" zeros—those of a function that is not identically zero—must be isolated. We will actually prove a sort of converse to this statement and, in the process, demonstrate that only a small amount of data must be considered to guarantee that a function is always zero. From this, we will quickly extract a test for determining when two functions really are the same.

4.2.3. The identity principle.

4.2.12 Theorem. *Let $\mathcal{D} \subseteq \mathbb{C}$ be a domain and let $f \colon \mathcal{D} \to \mathbb{C}$ be analytic. The following are equivalent.*

(i) $f(z) = 0$ *for all* $z \in \mathcal{D}$.

(ii) *There is* $z_0 \in \mathcal{D}$ *such that* $f^{(k)}(z_0) = 0$ *for all* $k \geq 0$.

(iii) *There is a sequence* (z_k) *in* $\mathcal{D}$ *of distinct points (i.e.,* $z_k \neq z_j$ *for* $j \neq k$*) such that* $f(z_k) = 0$ *for all* k *and* $z_k \to z_0$ *for some* $z_0 \in \mathcal{D}$.

(iv) f *has a zero that is not isolated.*

This is where we finished on Monday, April 8, 2024.

Proof. (i) $\Longrightarrow$ (ii) This is essentially a direct calculation: if $f(z) = 0$ for all $z \in \mathcal{D}$, then, fixing z, we have

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

Thus $f'(z) = 0$ for all $z \in \mathcal{D}$. Proceeding inductively, we find $f^{(k)}(z) = 0$ for all $z \in \mathcal{D}$ and all integers $k \geq 0$. We can then take any point $z_0 \in \mathcal{D}$ to satisfy the condition in part (ii).

(ii) $\Longrightarrow$ (iii) Since $\mathcal{D}$ is open, we may fix $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$. Since $f^{(k)}(z_0) = 0$ for all k, we have $f(z) = \sum_{k=0}^{\infty} f^{(k)}(z_0)(z - z_0)^k/k! = 0$ for all $z \in \mathcal{B}(z_0; r)$. (In particular, if $\mathcal{D} = \mathcal{B}(z_0; r)$, this shows that $f(z) = 0$ for all $z \in \mathcal{D}$, i.e., this argument proves that (ii) implies (i) in this special case.) Now set $z_k := z_0 + r/(k+1)$. It is straightforward to check that $z_k \neq z_j$ for $j \neq k$, that $z_k \in \mathcal{B}(z_0; r) \subseteq \mathcal{D}$ for each k, and that $z_k \to z_0 \in \mathcal{D}$.

(iii) $\Longrightarrow$ (iv) We claim that z_0 is this zero that is not isolated, and we prove this by contradiction. If z_0 is isolated, then there is $r > 0$ such that $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$ and $f(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; r)$. Since $z_k \to z_0$, for k sufficiently large we have $z_k \in \mathcal{B}(z_0; r)$. And since the points z_k are all distinct, we have $z_k = z_0$ for at most one $k \geq 1$. Thus for k large, we really have $z_k \in \mathcal{B}^*(z_0; r)$. But $f(z_k) = 0$, which contradicts our prior conclusion that $f(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; r)$.

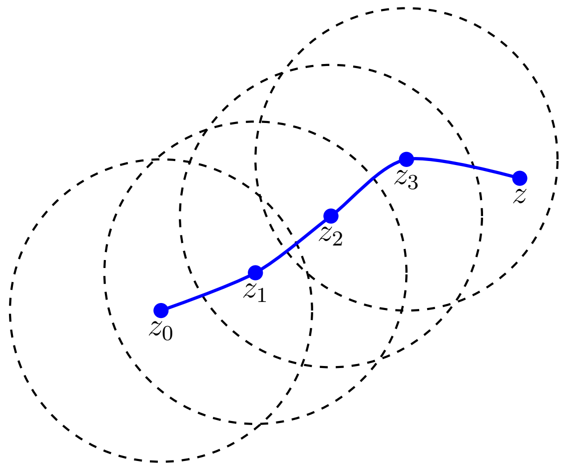
(iv) $\Longrightarrow$ (i) Let z_0 be the zero that is not isolated. Problem 4.2.8 gives $r > 0$ with $\mathcal{B}(z_0; r) \subseteq \mathcal{D}$, we have $f(z) = 0$ for all $z \in \mathcal{B}(z_0; r)$. If $\mathcal{D} = \mathcal{B}(z_0; r)$, then we are done. Otherwise, we need to do more work, and it is here that we will use for the first time in the proof the hypothesis that $\mathcal{D}$ is connected, not merely open.

We want to show that $f(z) = 0$ for all $z \in \mathcal{D}$. We give two arguments. The first is geometric and relies on an assertion about subsets of $\mathbb{C}$ that requires more technical tools from analysis than we care to develop here. The second is more rigorous but also possibly more opaque.

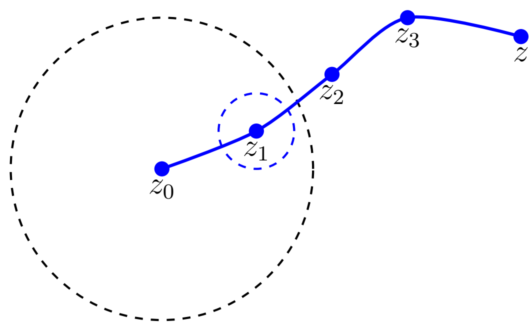
Argument #1. Let $z \in \mathcal{D}$ and let γ be a path in $\mathcal{D}$ from z_0 to z. It is possible to cover the image of γ by a finite sequence of overlapping balls of the same radius $\rho \leq r$ centered at points on the image of γ, starting with the point z_0, such that the center of the jth ball is contained in the $(j-1)$st ball, and such that each ball is contained in $\mathcal{D}$ with z belonging to the last ball. That is, if Γ is the image of γ, then for some $z_1, \ldots, z_n \in \Gamma$,

$$\Gamma \subseteq \bigcup_{k=0}^{n} \mathcal{B}(z_j; \rho) \subseteq \mathcal{D} \quad \text{and} \quad z_{j-1} \in \mathcal{B}(z_j; \rho), \ j = 1, \ldots, n \quad \text{with} \quad z \in \mathcal{B}(z_n; \rho).$$

For example, the situation could look like the following sketch, in which $n = 3$.

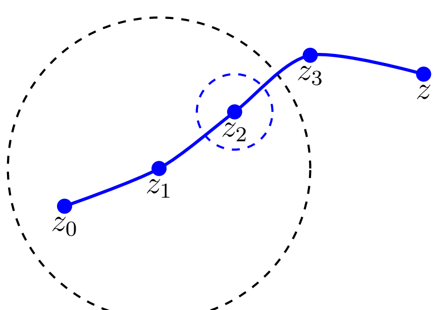


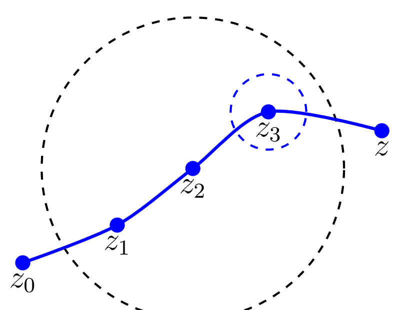
Here is how we exploit this "covering." We know that $f(w) = 0$ for all $w \in \mathcal{B}(z_0; \rho) \subseteq \mathcal{B}(z_0; r)$. We also know that $z_1 \in \mathcal{B}(z_0; \rho)$ and $\mathcal{B}(z_0; \rho)$ is open. So, take $s_1 > 0$ such that $\mathcal{B}(z_1; s_1) \subseteq \mathcal{B}(z_0; \rho)$.



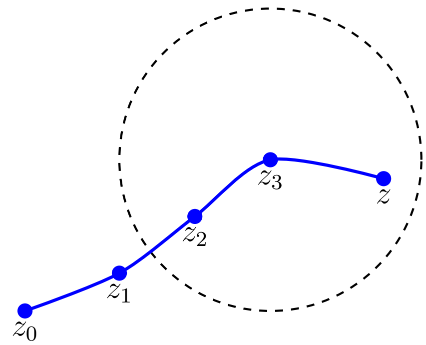
Then $f(w) = 0$ for all $w \in \mathcal{B}(z_1; s_1)$. Consequently, $f^{(k)}(z_1) = 0$ for all k; this is the implication of (ii) by (i), or common sense. Then $f(w) = 0$ for all $w \in \mathcal{B}(z_1; \rho)$ by the implication of (i) by (ii) in the special case that $\mathcal{D}$ is a ball, or more common sense.

Now iterate this reasoning: for each $j \geq 1$, we can show that $f(w) = 0$ for all $w \in \mathcal{B}(z_{j-1}; \rho)$ and that $\mathcal{B}(z_j; s_j) \subseteq \mathcal{B}(z_{j-1}; \rho)$ for some $s_j > 0$. Then $f(w) = 0$ for all $w \in \mathcal{B}(z_j; s_j)$, and so $f^{(k)}(z_j) = 0$ for all $k \geq 0$.





Hence $f(w) = 0$ for all $w \in \mathcal{B}(z_j;\rho)$. Repeat this reasoning until we stop at $j = n$ and conclude $f(w) = 0$ for all $w \in \mathcal{B}(z_n;\rho)$. Since $z \in \mathcal{B}(z_n;\rho)$, we are done.



The difficulty with this approach is the construction of this special "finite covering" of the image of γ, which needs, among other things, the tools of compactness and uniform continuity. Below we present a less geometrically obvious (but still geometrically motivated) proof that has the advantage of being logically self-contained to the tools that we already possess.

Argument #2. Put

$$\mathcal{D}_1 := \{z \in \mathcal{D} \mid z \text{ is an isolated zero of } f \text{ in } \mathcal{D} \text{ or } f(z) \neq 0\}$$

and

$$\mathcal{D}_2 := \{z \in \mathcal{D} \mid f(z) = 0 \text{ and } z \text{ is not an isolated zero of } f \text{ in } \mathcal{D}\}.$$

Note that $\mathcal{D}_2$ is nonempty, that $\mathcal{D}_1 \cup \mathcal{D}_2 = \mathcal{D}$, and that $\mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing$. We claim that $\mathcal{D}_1$ and $\mathcal{D}_2$ are both open; if this is true, then Problem 3.1.45 forces $\mathcal{D}_1 = \varnothing$ since $\mathcal{D}$ is a domain. Then $\mathcal{D} = \mathcal{D}_2$, in which case $f(z) = 0$ for all $z \in \mathcal{D}$.

We first show that $\mathcal{D}_1$ is open. If $z \in \mathcal{D}_1$ is an isolated zero of f in $\mathcal{D}$, let $r > 0$ be such that $\mathcal{B}(z;r) \subseteq \mathcal{D}$ with $f(w) \neq 0$ for $w \in \mathcal{B}^*(z_0;r)$. Thus $\mathcal{B}^*(z;r) \subseteq \mathcal{D}_1$, and since we know $z \in \mathcal{D}_1$ already, we conclude $\mathcal{B}(z;r) \subseteq \mathcal{D}$. If $z \in \mathcal{D}_1$ satisfies $f(z) \neq 0$, then by continuity there is $r > 0$ such that $\mathcal{B}(z;r) \subseteq \mathcal{D}$ and $f(w) \neq 0$ for $w \in \mathcal{B}(z;r)$. This implies that $w \in \mathcal{D}_1$ for all $w \in \mathcal{B}(z;r)$, and so $\mathcal{B}(z;r) \subseteq \mathcal{D}_1$. Either way, we have found $r > 0$ such that $\mathcal{B}(z;r) \subseteq \mathcal{D}_1$.

Now we show that $\mathcal{D}_2$ is open. If $z \in \mathcal{D}_2$, then $f(z) = 0$ and z is not an isolated zero of f in $\mathcal{D}$. So, for some $r > 0$ such that $\mathcal{B}(z;r) \subseteq \mathcal{D}$, we have $f(w) = 0$ for all $w \in \mathcal{B}(z;r)$. That is, each $w \in \mathcal{B}(z;r)$ is a zero of f; now we show that each w is a zero that is not isolated, which will imply $w \in \mathcal{D}_2$ and thus $\mathcal{B}(z;r) \subseteq \mathcal{D}_2$. Given $w \in \mathcal{B}(z;r)$, take $s > 0$ such that $\mathcal{B}(w;s) \subseteq \mathcal{B}(z;r)$. It is still the case that $f(\xi) = 0$ for all $\xi \in \mathcal{B}(w;s)$, so w is a zero of f in $\mathcal{D}$ that is not isolated, as desired. ■

4.2.13 Problem ($\star$). In the proof that part (iv) of Theorem 4.2.12 implies part (i), perhaps a more natural decomposition would be

$$\mathcal{D}_1 := \{z \in \mathcal{D} \mid z \text{ is an isolated zero of } f \text{ in } \mathcal{D}\}$$

and

$$\mathcal{D}_2 := \{z \in \mathcal{D} \mid z \text{ is not an isolated zero of } f \text{ in } \mathcal{D}\}.$$

Explain why $\mathcal{D}_1$ is not open, so this decomposition does not work.

4.2.14 Problem ($\star$). Give an example of an open set $\mathcal{D}$ and an analytic function $f : \mathcal{D} \to \mathbb{C}$ such that f is not identically zero on $\mathcal{D}$ but such that f has a zero in $\mathcal{D}$ that is not an isolated zero. That is, f and $\mathcal{D}$ should satisfy the following two conditions.

(i) There exists $z_1 \in \mathcal{D}$ such that $f(z_1) \neq 0$.

(ii) There exist $z_2 \in \mathcal{D}$ and $r > 0$ such that $\mathcal{B}(z_2; r) \subseteq \mathcal{D}$ and $f(z) = 0$ for all $z \in \mathcal{B}(z_2; r)$.

Such an open set $\mathcal{D}$ cannot be connected—why?

4.2.15 Problem (!). Does the situation of Example 4.2.11 contradict the equivalence of parts (i) and (iii) of Theorem 4.2.12?

While Theorem 4.2.12 is stated for the zeros of a function, this result carries over nicely to comparing two functions: just study where their difference is zero.

4.2.16 Corollary (Identity principle). *Let $\mathcal{D} \subseteq \mathbb{C}$ be a domain and let f_1, $f_2 : \mathcal{D} \to \mathbb{C}$ be analytic. Suppose that $f_1(z_k) = f_2(z_k)$ for a sequence (z_k) of distinct points in $\mathcal{D}$ such that $z_k \to z$ for some $z \in \mathcal{D}$. Then $f_1 = f_2$ on $\mathcal{D}$.*

Proof. Put $f = f_1 - f_2$ and use the equivalence of parts (i) and (iii) of Theorem 4.2.12. ■

Perhaps the most useful "test" to emerge from this theorem is part (iii): f need only be zero on a sequence of distinct points in $\mathcal{D}$ that converges to a point in $\mathcal{D}$ in order for us to conclude that f is always zero on $\mathcal{D}$! For example, if f is zero on a line segment in $\mathcal{D}$ (a one-dimensional subset of an open, and therefore two-dimensional, set), then f is zero on all of $\mathcal{D}$. This is only a very "little" amount of data!

4.2.17 Problem (!). Prove this ebullient claim. Specifically, let $\mathcal{D} \subseteq \mathbb{C}$ be a domain with z_1, $z_2 \in \mathcal{D}$ and $z_1 \neq z_2$. Suppose that f_1, $f_2 : \mathcal{D} \to \mathbb{C}$ are analytic with $f_1(z) = f_2(z)$ for all $z \in [z_1, z_2]$. Prove that $f_1 = f_2$ on $\mathcal{D}$.

This is where we finished on Wednesday, April 10, 2024.

4.2.18 Example. Many "functional identities" that are known on $\mathbb{R}$ remain true for functions extended analytically to $\mathbb{C}$. Often they can be proved brute-force (the best force) from the definitions of these analytic continuations, but we can also use the identity principle.

We know that $\ln(t_1 t_2) = \ln(t_1) + \ln(t_2)$ for t_1, $t_2 > 0$. We would like to say that

$\mathrm{Log}(z_1 z_2) = \mathrm{Log}(z_1) + \mathrm{Log}(z_2)$ for $z_1,\ z_2 \in \mathbb{C} \setminus \{0\}$, but this probably is not true for the entire plane. Nonetheless, we can say the following.

Fix $\tau > 0$ and define

$$f\colon (0,\infty) \to \mathbb{R}\colon t \mapsto \ln(t\tau) - \big[\ln(t) + \ln(\tau)\big].$$

Then, really, $f(t) = 0$ for all $t > 0$, and so certainly f is real analytic. (We are not saying anything about the real analyticity of ln, although since $\ln = \mathrm{Log}\big|_{(0,\infty)}$ and Log is holomorphic on $\mathbb{C} \setminus (-\infty, 0]$, we do obtain the real analyticity of ln from the analyticity of Log.)

Next, if $\tau > 0$ and $z \in \mathbb{C} \setminus (-\infty, 0]$, then $z\tau \in \mathbb{C} \setminus (-\infty, 0]$ as well. (Why? Consider the real and imaginary parts of $z\tau$.) Thus the function

$$\widetilde{f}\colon \mathbb{C} \setminus (-\infty, 0] \to \mathbb{C}\colon z \mapsto \mathrm{Log}(z\tau) - \big[\mathrm{Log}(z) + \mathrm{Log}(\tau)\big]$$

is analytic, since the principal logarithm is analytic except on the branch cut $(-\infty, 0]$. Furthermore, $\widetilde{f}(t) = f(t) = 0$ for all $t \in (0,\infty)$. By the identity principle, then, $\widetilde{f}(z) = 0$ for all $z \in \mathbb{C} \setminus (-\infty, 0]$. That is,

$$\mathrm{Log}(z\tau) = \mathrm{Log}(z) + \mathrm{Log}(\tau) \tag{4.2.4}$$

for all $z \in \mathbb{C} \setminus (-\infty, 0]$ and $\tau > 0$.

Now let $z,\ w \in \mathbb{C} \setminus \{0\}$, so $zw = |zw|e^{i[\mathrm{Arg}(z)+\mathrm{Arg}(w)]}$. In particular, $|zw| > 0$. If $e^{i[\mathrm{Arg}(z)+\mathrm{Arg}(w)]} \in \mathbb{C} \setminus (-\infty, 0]$, then we will have

$$\mathrm{Log}(zw) = \mathrm{Log}(|zw|) + \mathrm{Log}(e^{i[\mathrm{Arg}(z)+\mathrm{Arg}(w)]}) \tag{4.2.5}$$

So, when do we have $e^{i[\mathrm{Arg}(z)+\mathrm{Arg}(w)]} \in \mathbb{C} \setminus (-\infty, 0]$? Equivalently, when do we have $e^{i[\mathrm{Arg}(z)+\mathrm{Arg}(w)]} \notin (-\infty, 0]$? We know that $e^{i\theta} \in (-\infty, 0]$ if and only if $\theta = (2k+1)\pi$ for some $k \in \mathbb{Z}$. Since $-\pi < \mathrm{Arg}(z),\ \mathrm{Arg}(w) < \pi$, we have $-2\pi < \mathrm{Arg}(z) + \mathrm{Arg}(w) \le 2\pi$, and so the only way for $\mathrm{Arg}(z) + \mathrm{Arg}(w)$ to be an odd integer multiple of π is when $\mathrm{Arg}(z) + \mathrm{Arg}(w) = \pm\pi$.

We claim that if impose this restriction, then (hopefully familiar) properties of Log and Arg combine with (4.2.5) to imply $\mathrm{Log}(zw) = \mathrm{Log}(z) + \mathrm{Log}(w)$. In short, we will have shown

$$z,\ w \in \mathbb{C} \setminus \{0\} \text{ with } -\pi < \mathrm{Arg}(z) + \mathrm{Arg}(w) < \pi \Longrightarrow \mathrm{Log}(zw) = \mathrm{Log}(z) + \mathrm{Log}(w).$$

4.2.19 Problem (!). **(i)** Let $z,\ w \in \mathbb{C} \setminus \{0\}$ such that $-\pi < \mathrm{Arg}(z) + \mathrm{Arg}(w) < \pi$. Use (4.2.5) and various properties of Log and Arg to show that $\mathrm{Log}(zw) = \mathrm{Log}(z) + \mathrm{Log}(w)$.

(ii) Give an example of $z,\ w \in \mathbb{C}\setminus\{0\}$ such that $\mathrm{Log}(zw) \ne \mathrm{Log}(z)+\mathrm{Log}(w)$. [Hint: *it will be necessary that* $\mathrm{Arg}(z) + \mathrm{Arg}(w) = \pm\pi$, *and this can be achieved by taking* $z,\ w \in i\mathbb{R}$.]

4.2.4. Analytic continuation.

Now we can answer a major question that has been driving us since we first extended the exponential to the plane: is there only one way to extend a real analytic function into $\mathbb{C}$? Yes.

4.2.20 Theorem (Analytic continuation of real analytic functions). *Let $I \subseteq \mathbb{R}$ be an interval and let $f\colon I \to \mathbb{R}$ be real analytic. Then there exists a domain $\mathcal{D} \subseteq \mathbb{C}$ such that $I \subseteq \mathcal{D}$ and that f has a unique analytic continuation on $\mathcal{D}$.*

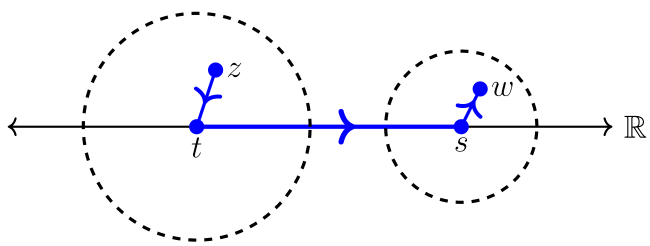
Proof. The uniqueness result is the identity theorem; see Problem 4.2.22.

Now we show existence. First we have to construct the domain $\mathcal{D}$. For each $t \in I$, there is $r_t > 0$ such that the Taylor series for f converges to f on $(t - r_t, t + r_t) \cap I$. We may as well make r_t so small that $(t - r_t, t + r_t) \subseteq I$. Then there is a sequence $(a_{k,t})$ of real numbers such that $f(\tau) = \sum_{k=0}^{\infty} a_{k,t}(\tau - t)^k$ for all $\tau \in (t - r_t, t + r_t)$. Specifically, $a_{k,t} = f^{(k)}(t)/k!$.

Now we set

$$\mathcal{D} := \bigcup_{t \in I} \mathcal{B}(t; r_t) = \{z \in \mathbb{C} \mid |z - t| < r_t \text{ for some } t \in I\}.$$

We claim that $\mathcal{D}$ is open and connected. For openness, fix $z \in \mathcal{D}$ and take $t \in I$ such that $z \in \mathcal{B}(t; r_t)$; since $\mathcal{B}(t; r_t)$ is open, there is $r > 0$ such that $\mathcal{B}(z; r) \subseteq \mathcal{B}(t; r_t)$. For connectedness, fix z, $w \in \mathcal{D}$. Take $z \in \mathcal{B}(t; r_t)$ and $w \in \mathcal{B}(s; r_s)$ for some t, $s \in I$. Let $\gamma = [z, t] \oplus [t, s] \oplus [s, w]$; then γ is a path in $\mathcal{D}$ with initial point z and terminal point w.



Next, we show that with the sequence $(a_{k,t})$ and the radius $r_t > 0$ defined above, the series $\sum_{k=0}^{\infty} a_{k,t}(z - t)^k$ converges for each $z \in \mathcal{B}(t; r_t)$. Specifically, we show that the series converges on $\mathcal{B}(t; s)$ for each $s < r_t$. So, fix some such s, so $t + s \in (t - r_t, t + r_t)$. Then the series $\sum_{k=0}^{\infty} a_{k,t}((t + s) - t)^k$ converges. We know that this series converges for any $z \in (t - r_t, t + r_t)$. Part (ii) of Problem 4.1.10 with $z_0 = t$ and $z_1 = t + s$ tells us that the series then converges for each $z \in \mathbb{C}$ with $|z - t| < |(t + s) - s| = s$, as desired.

Finally, we define the analytic continuation. First, for $t \in I$, define

$$f_t\colon \mathcal{B}(t; r_t) \to \mathbb{C}\colon z \mapsto \sum_{k=0}^{\infty} a_{k,t}(z - t)^k.$$

By the work above, f_t is analytic on $\mathcal{B}(t; r_t)$. Next, note that if $\mathcal{B}(t; r_t) \cap \mathcal{B}(s; r_s) \neq \varnothing$ for some t, $s \in I$, then by Problem 4.2.21 below, there is a sequence of distinct points (w_k) in $\mathcal{B}(t; r_t) \cap \mathcal{B}(s; r_s)$ such that $w_k \to w$ for some $w \in \mathcal{B}(t; r_t) \cap \mathcal{B}(s; r_s)$. Since $f_t(w_k) = f_s(w_k)$ for each k, the identity principle implies that $f_t(z) = f_s(z)$ for each $z \in \mathcal{B}(t; r_t) \cap \mathcal{B}(s; r_s)$.

Consequently, we may define

$$\widetilde{f}\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto f_t(z) \text{ if } z \in \mathcal{B}(t; r_t).$$

There is no ambiguity in this definition if $z \in \mathcal{B}(t; r_t) \cap \mathcal{B}(s; r_s)$ for two distinct t, $s \in I$, as the work above shows $f_t(z) = f_s(z)$. Finally, since each f_t is analytic on $\mathcal{B}(t; r_t)$, the function $\widetilde{f}$ is analytic on $\mathcal{D}$. And clearly $\widetilde{f}(t) = f(t)$ for each $t \in I$. ■

4.2.21 Problem ($\star$). Let z_1, $z_2 \in \mathbb{C}$ and r_1, $r_2 > 0$ such that $\mathcal{B}(z_1; r_1) \cap \mathcal{B}(z_2; r_2) \neq \varnothing$. Show that there exists a sequence of distinct points (w_k) in $\mathcal{B}(z_1; r_1) \cap \mathcal{B}(z_2; r_2)$ such that $w_k \to w$ for some $w \in \mathcal{B}(z_1; r_1) \cap \mathcal{B}(z_2; r_2)$. [Hint: *as usual, start by drawing a picture.*]

4.2.22 Problem (+). **(i)** Let $\mathcal{D} \subseteq \mathbb{C}$ be a domain and let $I \subseteq \mathbb{R}$ be a nonempty interval such that $I \subseteq \mathcal{D}$. Suppose that f_1, $f_2\colon \mathcal{D} \to \mathbb{C}$ are analytic with $f_1(t) = f_2(t)$ for all $t \in I$. Prove that $f_1 = f_2$ on $\mathcal{D}$. [Hint: *use Problem 4.2.17.*]

(ii) Prove that analytic continuations, whether of real analytic functions defined on a real interval or not, are unique. That is, suppose that $\mathcal{D}_0 \subseteq \mathbb{C}$ is a domain and $f\colon \mathcal{D}_0 \to \mathbb{C}$ is analytic. Let $\mathcal{D} \subseteq \mathbb{C}$ also be a domain with $\mathcal{D}_0 \subseteq \mathcal{D}$. Suppose that $\widetilde{f}_1$, $\widetilde{f}_2\colon \mathcal{D} \to \mathbb{C}$ are both analytic continuations of $\mathcal{D}_0$. Then $\widetilde{f}_1 = \widetilde{f}_2$.

4.3. Isolated singularities.

We now know a great deal about analytic functions, especially their power series expansions and their zeros. What happens if a function fails to be analytic, or holomorphic, or differentiable, on some proper subset of its domain? Depending on the geometry of that region of failure, we may still be able to say quite a lot about the function. Studying such failures is not just a natural evolution of our narrative—frequently applications demand consideration of functions that are not analytic in certain controlled ways.

We begin with the simplest failure of analyticity: the isolated singularity.

4.3.1 Definition. *Let $z_0 \in \mathbb{C}$ and $r > 0$. A function $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has an* **ISOLATED SINGULARITY** *at z_0 if f is analytic on $\mathcal{B}^*(z_0; r)$.*

4.3.2 Remark. *We will not study "non-isolated singularities." We might call a point $z_0 \in \mathbb{C}$ a non-isolated singularity of a function $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ if there is no $\rho \in (0, r)$ such that f is analytic on $\mathcal{B}^*(z_0; \rho)$. For example,* Log *is not analytic on any punctured ball centered at the origin because such a ball contains a "continuum" of singularities inherited from its intersection with the negative real axis. But there is very little more to say about the behavior of* Log *on the negative real axis than we have already exhaustively said; there is much more to say about isolated singularities.*

It may appear that there are lots of ways for a function to fail to be analytic at a single point in a ball, and lots of possible behaviors on that punctured ball, but the power of analyticity on the punctured ball is such that there are only really three situations to consider. The following three canonical examples, all of which are functions defined and analytic on $\mathbb{C} \setminus \{0\}$, will illustrate those three behaviors:

$$f(z) = \frac{e^z - 1}{z}, \qquad g(z) = \frac{e^z - 1}{z^2}, \qquad \text{and} \qquad h(z) = e^{1/z}.$$

The form of these functions illustrates a general truth: most isolated singularities arise in practice via some kind of division by 0.

4.3.1. Removable singularities.

If $f \colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has an isolated singularity at z_0, perhaps it is natural to ask about the limit behavior of f at z_0. Either the limit $\lim_{z \to z_0} f(z)$ exists, or it does not. If the limit does exist, our experience with removable discontinuities suggests that we can extend f to z_0 and retain continuity, perhaps analyticity.

We can.

4.3.3 Example. The function

$$f \colon \mathbb{C} \setminus \{0\} \to \mathbb{C} \colon z \mapsto \frac{e^z - 1}{z}$$

can be written, for $z \neq 0$, as

$$f(z) = \frac{1}{z}\left(\sum_{k=0}^{\infty} \frac{z^k}{k!} - 1\right) = \frac{1}{z}\sum_{k=1}^{\infty} \frac{z^k}{k!} = \sum_{k=1}^{\infty} \frac{z^{k-1}}{k!} = \sum_{j=0}^{\infty} \frac{z^j}{(j+1)}!.$$

Certainly this series converges when $z = 0$, and specifically it converges to 1. So, if we define

$$\widetilde{f} \colon \mathbb{C} \to \mathbb{C} \colon z \mapsto \sum_{j=0}^{\infty} \frac{z^j}{(j+1)}!,$$

then $\widetilde{f}$ is entire and $\widetilde{f}\big|_{\mathbb{C}\setminus\{0\}} = f$. In particular, note that $\lim_{z \to 0} f(z) = \widetilde{f}(0) = 1$. (We probably wanted to say this from the start thanks to L'Hospital's rule, except we never developed that for functions of a complex variable.) Thus we constructed an analytic continuation $\widetilde{f}$ of f by setting

$$\widetilde{f}(z) = \begin{cases} f(z), & z \neq 1 \\ \lim_{w \to 1} f(w), & z = 1. \end{cases}$$

This construction did not require all that much work, since f was essentially a power series centered at 0 in disguise.

This example generalizes in several ways. Here is the first generalization.

4.3.4 Theorem. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \mathcal{B}^*(z_0;r) \to \mathbb{C}$ is analytic and $L := \lim_{z\to z_0} f(z)$ exists. Then the function*

$$\widetilde{f}\colon \mathcal{B}(z_0;r) \to \mathbb{C}\colon z \mapsto \begin{cases} f(z), & z \neq z_0 \\ L, & z = z_0 \end{cases}$$

is analytic.

Proof. The ball $\mathcal{B}(z_0;r)$ is a star domain with star-center z_0, and the function $\widetilde{f}$ is continuous on $\mathcal{B}(z_0;r)$ and analytic on $\mathcal{B}^*(z_0;r)$. The Cauchy integral theorem implies that $\int_\gamma \widetilde{f} = 0$ for all closed curves γ in $\mathcal{B}(z_0;r)$, and so Morera's theorem (part (iii) of Problem 3.6.21) implies that $\widetilde{f}$ is analytic on $\mathcal{B}(z_0;r)$. ■

We now name this first kind of isolated singularity.

4.3.5 Definition. *Let $z_0 \in \mathbb{C}$ and $r > 0$. An analytic function $f\colon \mathcal{B}^*(z_0;r) \to \mathbb{C}$ has a* **REMOVABLE SINGULARITY** *at z_0 if the limit $\lim_{z\to z_0} f(z)$ exists.*

Theorem 4.3.4 says that any analytic function $f\colon \mathcal{B}^*(z_0;r) \to \mathbb{C}$ with a removable singularity at z_0 has an analytic continuation to that singularity. Conversely, the existence of an analytic continuation $\widetilde{f}\colon \mathcal{B}(z_0;r) \to \mathbb{C}$ of f implies that f has a removable singularity at z_0, since the limit $\lim_{z\to z_0} f(z) = \lim_{z\to z_0} \widetilde{f}(z)$ must exist by the continuity of $\widetilde{f}$ and the equality $f(z) = \widetilde{f}(z)$ on $\mathcal{B}^*(z_0;r)$.

4.3.6 Problem (!). Let $z_0 \in \mathbb{C}$ and $r > 0$. Show that an analytic function $f\colon \mathcal{B}^*(z_0;r) \to \mathbb{C}$ has a removable singularity at z_0 if and only if there is a sequence (a_k) such that

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k, \ z \in \mathcal{B}^*(z_0;r).$$

[Hint: *if f has this expansion, then argue that $f = \widetilde{f}\big|_{\mathcal{B}^*(z_0;r)}$, where $\widetilde{f}$ is this power series on all of $\mathcal{B}(z_0;r)$. Conversely, if f has a removable singularity, use Theorem 4.3.4 to get this analytic continuation of f.*]

4.3.7 Problem ($\star$). Let $\mathcal{D} \subseteq \mathbb{C}$ be open, let $f\colon \mathcal{D} \to \mathbb{C}$ be analytic, and let $z_0 \in \mathcal{D}$. Define

$$\phi\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto \begin{cases} \dfrac{f(z) - f(z_0)}{z - z_0}, & z \neq z_0 \\ \\ f'(z_0), & z = z_0. \end{cases}$$

(i) Show that ϕ is analytic on $\mathcal{D}$.

(ii) What is the Taylor series of ϕ centered at z_0?

(iii) Compare these results to the difference quotient lemma (Lemma 2.5.16).

(iv) How is this a generalization of Example 4.3.3?

4.3.8 Problem (+). Let $z_0 \in \mathbb{C}$ and $r > 0$. Use the following to prove the **Riemann removability criterion**: an analytic function $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has a removable singularity at z_0 if and only if f is bounded in the sense that for some $\rho \in (0, r]$ and $M > 0$, it is the case that $|f(z)| \le M$ for all $z \in \mathcal{B}^*(z_0; \rho)$.

(i) ($\Longleftarrow$) Use the fact that f has an analytic continuation to $\mathcal{B}(z_0; r)$ and the extreme value theorem on $\overline{\mathcal{B}}(z_0; r/2)$.

(ii) ($\Longrightarrow$) Since f is bounded, we can artificially force some limit behavior at z_0 by studying

$$g\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}\colon z \mapsto (z - z_0)f(z).$$

Use the squeeze theorem to show that $\lim_{z \to z_0} g(z) = 0$, so g has a removable singularity at z_0 and thus an analytic continuation $\widetilde{g}$ to $\mathcal{B}(z_0; r)$. Develop a power series expansion for f at z_0 based on this analytic continuation, its value $\widetilde{g}(0) = 0$, and the identity $f(z) = g(z)/(z - z_0)$ for $z \ne z_0$.

4.3.9 Problem ($\star$). **(i)** Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f\colon \mathcal{D} \to \mathbb{C}$ be analytic. Show that f has a removable singularity at every point of $\mathcal{D}$.

(ii) Let $\mathcal{D} \subseteq \mathbb{C}$ be open and $f\colon \mathcal{D} \to \mathbb{C}$ be continuous. Suppose that for some $z_0 \in \mathcal{D}$, f is analytic on $\mathcal{D} \setminus \{z_0\}$. Show that f is really analytic on $\mathcal{D}$.

4.3.2. Poles.

Suppose next that $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ is analytic but the limit $\lim_{z \to z_0} f(z)$ does not exist. As we know from calculus, there are different gradations of a limit not existing. An infinite limit (a vertical asymptote) technically does not exist as a real number, but knowing that a limit is infinite surely tells us more information than just saying that the limit does not exist.

4.3.10 Example. Consider the function

$$g\colon \mathbb{C} \setminus \{0\} \to \mathbb{C}\colon z \mapsto \frac{e^z - 1}{z^2}.$$

We study g through two approaches.

(i) Just as in Example 4.3.3, we can try to expand g as a power series:

$$g(z) = \frac{1}{z^2}\left(\sum_{k=0}^{\infty}\frac{z^k}{k!} - 1\right) = \frac{1}{z^2}\sum_{k=1}^{\infty}\frac{z^k}{k!} = \sum_{k=1}^{\infty}\frac{z^{k-2}}{k!} = z^{-1} + \sum_{k=2}^{\infty}\frac{z^{k-2}}{k!} = z^{-1} + \sum_{j=0}^{\infty}\frac{z^j}{(j+2)!}.$$

However, in contrast to the calculation in Example 4.3.3, we have not really expressed g as a power series but rather as the sum of some negative powers of z (well, one negative power) *and* a power series.

What is this saying about the limit behavior of g near 0? The series $\sum_{j=0}^{\infty} z^j/(j+2)!$ gives an entire function with $\lim_{z\to 0}\sum_{j=0}^{\infty} z^j/(j+2)! = 1/2$, and so perhaps we expect that

$$\lim_{z\to 0} g(z) = \lim_{z\to 0}\left(z^{-1} + \sum_{j=0}^{\infty}\frac{z^j}{(j+2)!}\right) = \lim_{z\to 0} z^{-1} + \frac{1}{2} = \infty.$$

However, this is nonsense, as we have never developed the notion of what an infinite limit for a function of a complex variable should mean.

(ii) We might recognize g as the quotient of f from Example 4.3.3 with z. That is,

$$g(z) = \frac{(e^z-1)/z}{z} = \frac{f(z)}{z}, \qquad \text{where} \qquad \lim_{z\to 0} f(z) = 1.$$

Does this mean that

$$\lim_{z\to 0} g(z) = \lim_{z\to 0}\frac{f(z)}{z} = \frac{1}{0} = \infty?$$

It strongly suggests that, but it more strongly suggests that we need a notion of infinite limit first.

This is where we finished on Friday, April 12, 2024.

4.3.11 Definition. *Let $z_0 \in \mathbb{C}$ and $r > 0$. For a function $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$, we write $\lim_{z\to z_0}|f(z)| = \infty$ if for all $M > 0$, there is $\delta \in (0, r]$ such that if $0 < |z - z_0| < \delta$, then $M < |f(z)|$.*

4.3.12 Example. We have

$$\lim_{z\to 0}\frac{1}{|z|} = \infty,$$

for given $M > 0$, we can take $\delta = 1/M$ to see that if $|z| < \delta$, then $M < 1/|z|$.

4.3.13 Problem (!). Show that

$$\lim_{z \to 0} \left| \frac{e^z - 1}{z^2} \right| = \infty.$$

[Hint: *the limit* $\lim_{z \to 0}(e^z - 1)/z = 1$ *will be helpful.*]

More generally, suppose that $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has an isolated singularity at z_0 with $\lim_{z \to z_0} |f(z)| = \infty$. Take $\delta > 0$ such that if $z \in \mathcal{B}^*(z_0; \delta)$, then $1 < |f(z)|$, so in particular $f(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; \delta)$. Then the function

$$g\colon \mathcal{B}^*(z_0; \delta) \to \mathbb{C}\colon z \mapsto \frac{1}{f(z)}$$

is defined and analytic. Moreover, it is not too much work to check that $\lim_{z \to z_0} g(z) = 0$.

4.3.14 Problem (!). Check this.

Then g has a removable singularity at z_0 and therefore an analytic continuation to $\mathcal{B}(z_0; \delta)$ of the form

$$\widetilde{g}\colon \mathcal{B}(z_0; \delta) \to \mathbb{C}\colon z \mapsto \begin{cases} 1/f(z), & z \neq z_0 \\ 0, & z = z_0. \end{cases}$$

Since $1/f(z) \neq 0$ for all $z \in \mathcal{B}^*(z_0; \delta)$, we see that $\widetilde{g}(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; \delta)$, too. Then $\widetilde{g}$ really has an isolated zero at z_0, and so there is an integer $m \geq 1$ and an analytic function $q\colon \mathcal{B}(z_0; \rho) \to \mathbb{C}$ for some $\rho \in (0, \delta]$ such that for $z \in \mathcal{B}(z_0; \rho)$,

$$\widetilde{g}(z) = (z - z_0)^m q(z) \quad \text{and} \quad q(z) \neq 0.$$

Thus for $z \in \mathcal{B}^*(z_0; \rho)$, we have

$$f(z) = \frac{1}{\widetilde{g}(z)} = \frac{1}{(z - z_0)^m q(z)} = \frac{1/q(z)}{(z - z_0)^m}.$$

Put $p(z) := 1/q(z)$ to conclude the following.

4.3.15 Theorem. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ is analytic and $\lim_{z \to z_0} |f(z)| = \infty$. Then there exist $\rho \in (0, r]$, an integer $m \geq 1$, and an analytic function $p\colon \mathcal{B}(z_0; \rho) \to \mathbb{C}$ such that $p(z_0) \neq 0$ and*

$$f(z) = \frac{p(z)}{(z - z_0)^m} \quad \text{for} \quad z \in \mathcal{B}^*(z_0; \rho).$$

This gives rise to another kind of named isolated singularity.

4.3.16 Definition. *Let $z_0 \in \mathbb{C}$ and $r > 0$. An analytic $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has a* **POLE OF ORDER** m *at z_0 if there exist $\rho \in (0, r]$, an integer $m \geq 1$, and an analytic function $p\colon \mathcal{B}(z_0; \rho) \to \mathbb{C}$ such that $p(z_0) \neq 0$ and*

$$f(z) = \frac{p(z)}{(z - z_0)^m} \quad \textit{for} \quad z \in \mathcal{B}^*(z_0; \rho). \tag{4.3.1}$$

4.3.17 Example. Consider again the function

$$f(z) = \frac{e^z - 1}{z^2}$$

on $\mathbb{C} \setminus \{0\}$. We might be tempted to say that f has a pole of order 2 at 0 because of the z^2 in the denominator, but this is wrong. The numerator satisfies $e^0 - 1 = 0$, and that is not what we should have in (4.3.1). Rather, we can rewrite

$$f(z) = \frac{(e^z - 1)/z}{z} = \frac{p(z)}{z},$$

where

$$p(z) = \begin{cases} (e^z - 1)/z, & z \neq 0 \\ 1, & z = 0 \end{cases}$$

is analytic by Example 4.3.3. Obviously $p(0) \neq 0$, so f has a pole of order 1 at 0.

4.3.18 Remark. *Informally, we might say that f has a pole of order m at z_0 if and only if $1/f$ has a zero of order m at z_0. This is not quite true, since $1/f(z) \neq 0$ for all z at which this quotient is defined (and it is not defined at z_0 anyway), but hopefully the euphemism is helpful.*

4.3.19 Problem (!). Let $z_0 \in \mathbb{C}$ and $r > 0$. Show that an analytic function $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has a pole of order m at z_0 if and only if there exist numbers $a_k \in \mathbb{C}$ for $1 \leq k \leq m$ and an analytic function $g\colon \mathcal{B}(z_0; r) \to \mathbb{C}$ such that

$$f(z) = \sum_{k=1}^{m} \frac{a_k}{(z - z_0)^k} + g(z), \ z \in \mathcal{B}^*(z_0; r) \quad \text{and} \quad a_m \neq 0.$$

[Hint: *when all else fails, give up and go back to the definition.*]

4.3.3. Essential singularities.

We have now seen two kinds of behaviors at isolated singularities: either $\lim_{z \to z_0} f(z)$ exists, or it does not but $\lim_{z \to z_0} |f(z)| = \infty$. The third possibility, simply, is that neither of these behaviors holds.

4.3.20 Example. Let
$$f: \mathbb{C} \setminus \{0\} \to \mathbb{C}: z \mapsto e^{1/z}.$$
Put $z_k = 1/2\pi ik$ to see that $z_k \to 0$ and $f(z_k) = e^{2\pi ik} = 1$. Thus $f(z_k) \to 1$ as well, and so it cannot be the case that $\lim_{z \to z_0} |f(z)| = \infty$.

Now put $w_k = 1/k$ to see that $w_k \to 0$ as well but $f(w_k) = e^k \to \infty$. Then the limit $\lim_{z \to z_0} f(z)$ cannot exist.

Finally, we examine the series behavior of f near 0:
$$f(z) = e^{1/z} = \sum_{k=0}^{\infty} \frac{(1/z)^k}{k!} = \sum_{k=0}^{\infty} \frac{z^{-k}}{k!}.$$
This representation of f contains *infinitely* many negative powers of z, unlike the pole in Example 4.3.10, which had only finitely many (specifically, one), and unlike the removable singularity in Example 4.3.3, which had none.

4.3.21 Definition. *Let $z_0 \in \mathbb{C}$ and $r > 0$. An analytic function $f: \mathcal{B}^*(z_0; r) \to \mathbb{C}$ has an* **ESSENTIAL SINGULARITY** *at z_0 if z_0 is neither a removable singularity nor a pole. That is, the limit $\lim_{z \to z_0} f(z)$ does not exist, but it is also not the case that $\lim_{z \to z_0} |f(z)| = \infty$.*

This is not the most helpful of definitions, as it requires us to check that two conditions do *not* hold. However, the situation of Example 4.3.20 in fact characterizes essential singularities. Along one "path of approach" to an essential singularity, a function blows up, but along a different, suitably chosen path, the function can become arbitrarily close to *any* $z \in \mathbb{C}$. In Example 4.3.20, we just saw that with the case of $z = 1$.

4.3.22 Problem (!). Fix $z \in \mathbb{C}$. Determine a sequence (z_k) such that $z_k \to 0$ and $e^{1/z_k} \to z$.

Given that a function with an essential singularity can become arbitrarily close to any complex number for inputs close to the singularity as well as become arbitrarily large, the words "nervous" and "erratic" are often used to describe behavior near essential singularities. Here is the precise statement of this behavior.

4.3.23 Theorem (Casorati–Weierstrass). *Let $z_0 \in \mathbb{C}$ and $r > 0$. Let $f: \mathcal{B}^*(z_0; r) \to \mathbb{C}$ be analytic. Then z_0 is an essential singularity of f if and only if both of the following hold.*

(i) *There is a sequence (w_k) in $\mathcal{B}^*(z_0; r)$ such that $w_k \to z_0$ and $|f(w_k)| \to \infty$.*

(ii) *For each $z \in \mathbb{C}$, there is a sequence (z_k) in $\mathcal{B}^*(z_0; r)$ such that $z_k \to z_0$ and $f(z_k) \to z$.*

Proof. ($\Longleftarrow$) The condition (i) means that $\lim_{z \to z_0} f(z)$ cannot exist (as a finite complex number), and therefore z_0 is not a removable singularity of f. The condition (ii) means that $\lim_{z \to z_0} |f(z)| \neq \infty$, and therefore z_0 is not a pole of f. The only other possibility is that z_0 is an essential singularity of f.

($\Longrightarrow$) This is Problem 4.3.24. ■

4.3.24 Problem (+). Prove the forward direction of the Casorati–Weierstrass theorem as follows.

(i) If z_0 is an essential singularity of f, then z_0 is not removable. By the Riemann removability criterion (Problem 4.3.8), f is not bounded near z_0. Manipulate quantifiers to produce the sequence (w_k).

(ii) Fix $z \in \mathbb{C}$ and suppose there is no sequence (z_k) with $z_k \to z_0$ and $f(z_k) \to z$. Argue that there exist M, $\rho > 0$ such that $M \leq |f(w) - z|$ for all $w \in \mathcal{B}^*(z_0; \rho)$. [Hint: *contradiction, manipulate quantifiers.*] Conclude that $g(w) := 1/[f(w) - z]$ is analytic and bounded on $\mathcal{B}^*(z_0; \rho)$; obtain an analytic continuation $\widetilde{g}$ of g to $\mathcal{B}(z_0; \rho)$ by Riemann removability. Solve for f as $f(w) = 1/\widetilde{g}(w) + z$ for $w \in \mathcal{B}^*(z_0; \rho)$. Depending on whether or not $\lim_{w \to z_0} \widetilde{g}(w) = 0$, conclude that either f has a pole at z_0 or a removable singularity there. This contradicts the assumption that f has an essential singularity at z_0.

We might also try to characterize essential singularities by the series behavior of functions near them.

4.3.25 Problem (!). Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ is analytic and that we can write

$$f(z) = \sum_{k=1}^{\infty} \frac{a_k}{(z - z_0)^k} + g(z), \ z \in \mathcal{B}^*(z_0; r),$$

where $g\colon \mathcal{B}(z_0; r) \to \mathbb{C}$ is analytic and infinitely many of the a_k are nonzero. Explain why f must have an essential singularity at z_0. [Hint: *Problems 4.3.6 and 4.3.19.*]

The challenge is the reverse: why, if f has an essential singularity at z_0, must f also have a series expansion of the peculiar form given in the problem above? This is true, but it is quite hard to show with only the tools that we have on hand. So, we need new tools.

4.4. Laurent series.

We introduced removable singularities, poles, and essential singularities via the limit behavior of the function at the singularity. Removable singularities lead to analytic continuations, poles lead to a nice fractional representation, and essential singularities lead to very nervous behaviors. It would, perhaps, be nice if there were one "unified" test that we could apply to singularities to determine their nature. We will develop such a test by examining the series behavior of functions near isolated singularities.

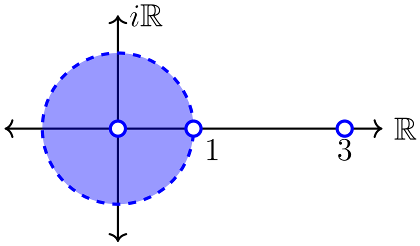
The pattern that emerged from our previous examples is that removable singularities at z_0 lead to ordinary power series at z_0; poles lead to series with negative powers of $z - z_0$, but only finitely many such negative powers (up to and including the order of the pole); and essential singularities have infinitely many negative powers of $z - z_0$. This pattern is indeed true, as we have verified for removable singularities in Problem 4.3.6 and for poles in Problem 4.3.19, but we have yet to verify it for essential singularities. We do this now. Moreover, we can do this in a more general context than the isolated singularity, which requires the function to be analytic on a punctured ball centered at z_0.

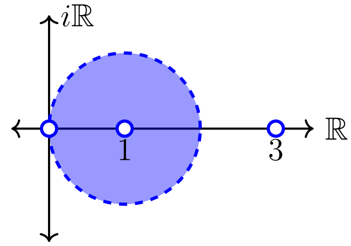
4.4.1 Example. The function

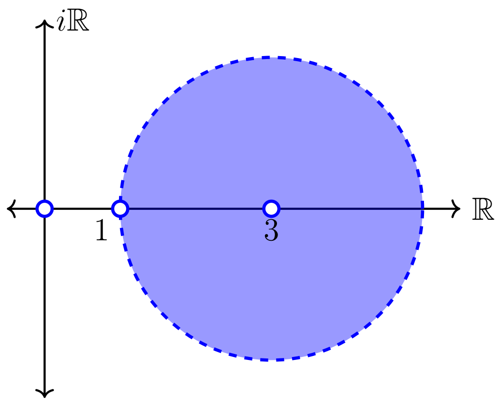
$$f(z) = \frac{1}{z(z-1)(z-3)}$$

is analytic on $\mathbb{C} \setminus \{0, 1, 3\}$ with simple poles at the points 0, 1, and 3. Much of our prior success hinged on working on open balls on which functions were analytic. Now we might try the next best thing: what are the largest ball-like subsets of $\mathbb{C}$ on which f is analytic? Such subsets would have to exclude the three poles.

We might start with the largest punctured balls on which f is analytic. These are the sets of points $z \in \mathbb{C}$ such that $0 < |z| < 1$, $0 < |z-1| < 1$, and $0 < |z-3| < 2$.

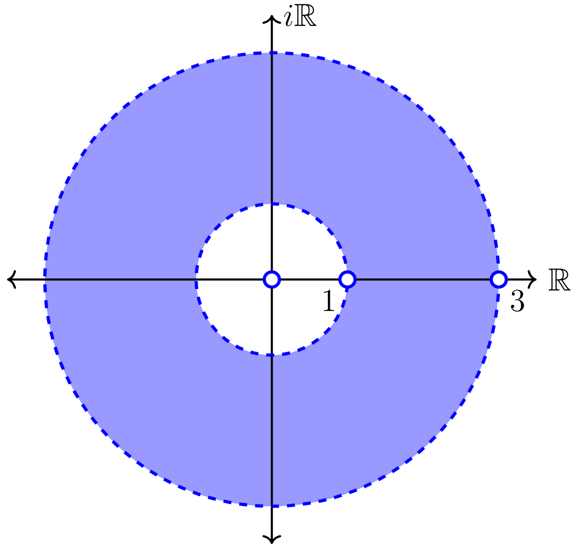




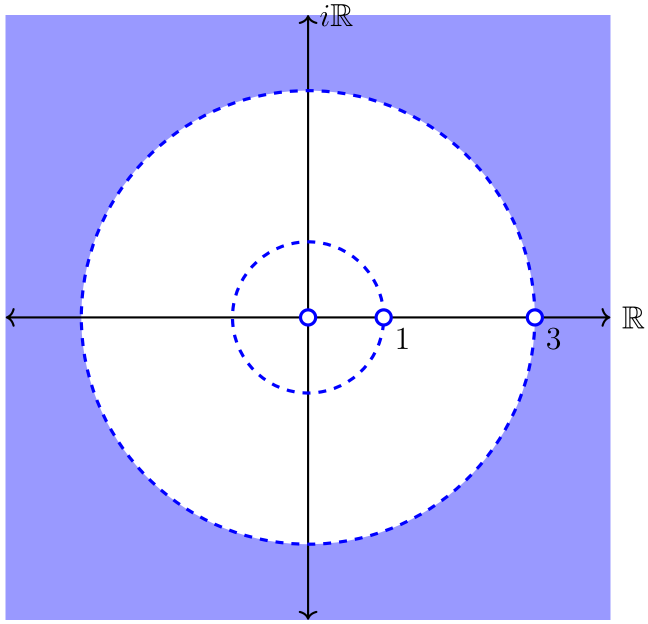


We might also consider regions "between" the singularities. One such region is the "ring" of points z such that $1 < |z| < 3$. This is really the open ball $\mathcal{B}(0; 3)$ with the closed ball

$\overline{\mathcal{B}}(0;1)$ removed from its center.



Another region is the set of z such that $3 < |z|$, which is the whole plane with the ball $\overline{\mathcal{B}}(0;3)$ removed.



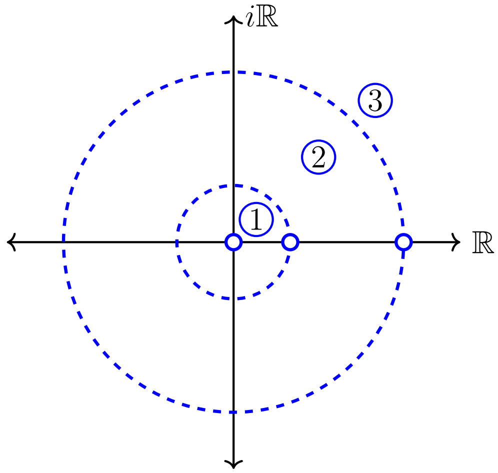
We place under one name the different subsets of $\mathbb{C}$ that appeared in the preceding example.

4.4.2 Definition. *Let $z_0 \in \mathbb{C}$ and $0 \le r < R \le \infty$. The* **annulus centered at** z_0 **of inner radius** r **and outer radius** R *is*

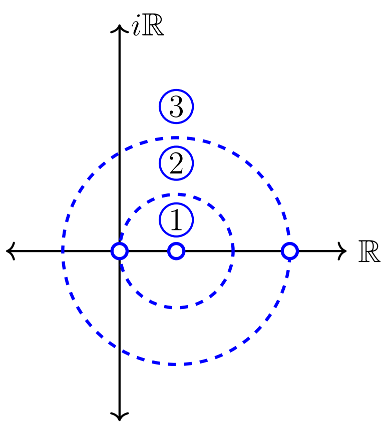
$$\mathcal{A}(z_0; r, R) := \{z \in \mathbb{C} \mid r < |z - z_0| < R\}.$$

4.4.3 Example. The function f from Example 4.4.1 is analytic on any annulus $\mathcal{A}(z_0; 0, R)$, where $z_0 \in \mathbb{C} \setminus \{0, 1, 3\}$ and $R > 0$ is chosen such that $\mathcal{B}(z_0; R) \subseteq \mathbb{C} \setminus \{0, 1, 3\}$. But at the points $z_0 = 0$, 1, 3, the choices of annuli $\mathcal{A}(z_0; r, R)$ on which f is analytic are more complicated. We need to choose r and R to exclude 0, 1, and 3; certainly taking $r > 0$ excludes z_0 from $\mathcal{A}(z_0; r, R)$, but what about the other two singularities? Here are the *largest* such annuli, and we remark on how we chose the radii to exclude the other singularities in addition to z_0.

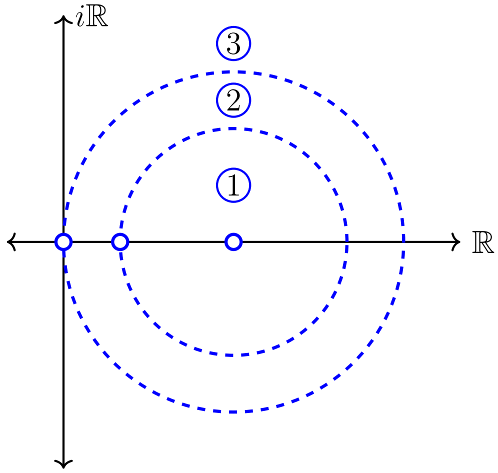
(i) $z_0 = 0$. They are (1) $\mathcal{A}(0; 0, 1)$ [radii chosen to exclude 1], (2) $\mathcal{A}(0; 1, 3)$ [radii chosen to exclude 1 and 3], and (3) $\mathcal{A}(0; 3, \infty)$.



(ii) $z_0 = 1$. They are (1) $\mathcal{A}(1; 0, 1)$ [radii chosen to exclude 0], (2) $\mathcal{A}(1; 1, 2)$ [radii chosen to exclude 0 and 1]; $\mathcal{A}(1; 2, \infty)$.



(iii) $z_0 = 3$. They are (1) $\mathcal{A}(3; 0, 2)$ [radii chosen to exclude 1]; (2) $\mathcal{A}(3; 2, 3)$ [radii chosen to exclude 1 and 0]; $\mathcal{A}(3; 2, \infty)$.



4.4.4 Problem (!). Let $z_0 \in \mathbb{C}$. Prove the following set equalities for annuli.

(i) If $0 < R < \infty$, then $\mathcal{A}(z_0; 0, R) = \mathcal{B}^*(z_0; R)$.

(ii) $\mathcal{A}(z_0; 0, \infty) = \mathbb{C} \setminus \{z_0\}$.

(iii) If $0 < r < \infty$, then $\mathcal{A}(z_0; r, \infty) = \mathbb{C} \setminus \overline{\mathcal{B}}(z_0; r)$.

4.4.5 Problem (!). Let $0 < z_1 < z_2$ and suppose $z_3 \in \mathbb{C}$ with $\text{Re}(z_3) > 0$ and $\text{Im}(z_3) > 0$. Draw the largest annuli centered at z_1, z_2, and z_3 such that none of z_1, z_2, or z_3 belong to any of these annuli. [Hint: *this is what we drew in Example 4.4.3.*]

This is where we finished on Monday, April 15, 2024.

We can now state the principal result about the series behavior of an analytic function on an annulus. Its proof is in Appendix C.6.

4.4.6 Theorem. *Let $z_0 \in \mathbb{C}$ and $0 \le r < R \le \infty$. Suppose that $f \colon \mathcal{A}(z_0; r, R) \to \mathbb{C}$ is analytic. Then there exist unique analytic functions*

$$f_r \colon \mathcal{B}(0; 1/r) \to \mathbb{C} \quad \text{and} \quad f_R \colon \mathcal{B}(0; R) \to \mathbb{C},$$

where we interpret $\mathcal{B}(0; 1/0) = \mathcal{B}(0; \infty) = \mathbb{C}$, such that $f_r(0) = 0$ and

$$f(z) = f_r\left(\frac{1}{z - z_0}\right) + f_R(z - z_0)$$

for each $z \in \mathcal{A}(z_0; r, R)$. We may expand f_r and f_R as power series centered at 0 to find

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \sum_{k=1}^{\infty} \frac{a_{-k}}{(z - z_0)^k} = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k, \tag{4.4.1}$$

where for each $k \in \mathbb{Z}$, the coefficient a_k satisfies

$$a_k = \frac{1}{2\pi i} \int_{|z-z_0|=s} \frac{f(z)}{(z - z_0)^{k+1}} \, dz \tag{4.4.2}$$

for any $s \in (r, R)$.

The ordered pair (f_r, f_R) is the **Laurent decomposition** *of f on $\mathcal{A}(z_0; r, R)$; the series (4.4.1) is the* **Laurent series** *of f on $\mathcal{A}(z_0; r, R)$; and the coefficients (4.4.2) are the* **Laurent coefficients** *of f on $\mathcal{A}(z_0; r, R)$. The function f_r is the* **principal part** *of the Laurent decomposition. The doubly infinite series on the right of (4.4.1) is defined to be the sum of the two series on the left.*

4.4.7 Remark. **(i)** *We often compress the series expansion in (4.4.1) to*

$$\sum_{k=-\infty}^{\infty} a_k (z - z_0)^k := \sum_{k=0}^{\infty} a_k (z - z_0)^k + \sum_{k=1}^{\infty} \frac{a_{-k}}{(z - z_0)^k}$$

More generally, we are taking the view that if (w_k) is a **doubly-infinite sequence***, i.e., a function from $\mathbb{Z}$ to $\mathbb{C}$, then*

$$\sum_{k=-\infty}^{\infty} w_k := \sum_{k=0}^{\infty} w_k + \sum_{k=1}^{\infty} w_{-k},$$

and the **doubly-infinite series** *on the left converges if and only if each series on the right converges. This view of doubly-infinite series is not universally applied (cf. the "Cauchy principal value" of Fourier series).*

(ii) *Above we called the function f_r the principal part of the Laurent series for f. The function*

$$\mathcal{A}(z_0; r, R) \to \mathbb{C} \colon z \mapsto f_r((z - z_0)^{-1}) = \sum_{k=1}^{\infty} \frac{a_{-k}}{(z - z_0)^k}$$

may also be called the principal part.

4.4.8 Problem (!). Explain why the principal part of the Laurent decomposition of a function at an isolated point is entire.

The formula (4.4.2) is useful for estimating the Laurent coefficients in terms of f, but it rarely provides an expedient way of actually calculating the coefficients. As with Taylor series, the strategy is to reduce a new Laurent expansion to an old one (or an old Taylor series).

Laurent decompositions and series meld analysis and geometry. The same function f may be defined on different annuli centered at a point z_0, and it is likely that f will have different Laurent decompositions and series on those different annuli. We saw this with Taylor series: changing the center of the series changes the coefficients of the series. But now the center of the annulus can stay the same, and if the radii change, so may the Laurent decomposition and series.

4.4.9 Example. We find a variety of Laurent decompositions and series for

$$f\colon \mathbb{C}\setminus\{0,1,3\} \to \mathbb{C}\colon z \mapsto \frac{6}{z(z-1)(z-3)} = \frac{2}{z} - \frac{3}{z-1} + \frac{1}{z-3},$$

which we previously considered in Examples 4.4.1 and 4.4.3, up to the convenient scaling factor of 6 (which makes the denominators in the partial fractions expansion above nicer).

These are mostly consequences of careful algebraic manipulations involving geometric series, which we spell out in full generality in Problem 4.4.11 below. More informally, the goal is to rewrite each of the rational functions in the partial fractions expansion of f above in the form

$$? \cdot \frac{1}{1-??}, \qquad |??| < 1.$$

Then we can expand the factor $1/(1-??)$ using the geometric series:

$$\frac{1}{1-??} = \sum_{k=0}^{\infty} (??)^k.$$

(i) *On* $\mathcal{A}(0;0,1)$. We want to write

$$f(z) = f_0\left(\frac{1}{z}\right) + f_1(z), 0 < |z| < 1,$$

where $f_0\colon \mathcal{B}(0;1/0) = \mathbb{C} \to \mathbb{C}$ and $f_1\colon \mathcal{B}(0;1) \to \mathbb{C}$ are analytic.

Since $|z| < 1$, the geometric series directly gives

$$-\frac{3}{z-1} = \frac{3}{1-z} = 3\sum_{k=0}^{\infty} z^k.$$

Next, we factor out 3 to find

$$\frac{1}{z-3} = \frac{1}{3\left(\frac{z}{3}-1\right)}.$$

Since $|z| < 1$, we also have $|z/3| < 1$, so this is the right set-up for the geometric series. (Had we factored out z to get $1/z(1-3/z)$, we would have $|3/z| < 1$ if and only if $3 < |z|$, and that is not the case in this annulus.) So,

$$\frac{1}{z-3} = -\frac{1}{3}\left(\frac{1}{1-\frac{z}{3}}\right) = -\frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{z}{3}\right)^k.$$

Note that the maps $z \mapsto -3/(z-1)$ and $z \mapsto 1/(z-3)$ are analytic at 0, so these expansions really just find their Taylor series at 0.

We therefore have

$$f(z) = \frac{2}{z} + 3\sum_{k=0}^{\infty} z^k - \frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{z}{3}\right)^k = \frac{2}{z} + \sum_{k=0}^{\infty}\left(3 - \frac{1}{3^{k+1}}\right) z^k.$$

This is the Laurent series for f. Its Laurent decomposition is given by

$$f_0(w) = 2w \quad \text{and} \quad f_1(w) = \sum_{k=0}^{\infty}\left(3 - \frac{1}{3^{k+1}}\right) w^k.$$

Changing variables to w is not all that necessary here, but hopefully it clears up the role of the reciprocal $1/z$ in $f_0(1/z)$, and later it will help us distinguish $z - z_0$ in the Laurent series from the behavior of f_0 and f_1. Note that f_0 is entire and f_1 contains a term that is a multiple of the geometric series, so f_1 is only defined on $\mathcal{B}(0;1)$. This is what we predicted at the start.

(ii) *On* $\mathcal{A}(0;1,3)$. We want to write

$$f(z) = f_1\left(\frac{1}{z}\right) + f_3(z), \ 1 < |z| < 3,$$

where $f_1 \colon \mathcal{B}(0;1/1) \to \mathbb{C}$ and $f_3 \colon \mathcal{B}(0;3) \to \mathbb{C}$ are analytic. We can no longer rewrite $-3/(z-1)$ with the geometric series exactly as before, since now $|z| > 1$. But this suggests dividing by z instead:

$$-\frac{3}{z-1} = -\frac{3}{z\left(1 - \frac{1}{z}\right)},$$

and here $|1/z| < 1$. Thus

$$-\frac{3}{z-1} = -\frac{3}{z}\left(\frac{1}{1 - \frac{1}{z}}\right) = -\frac{3}{z}\sum_{k=0}^{\infty}\frac{1}{z^k}.$$

Next, since $|z| < 3$, we have $|z/3| < 1$, and so the same calculations and reasoning as above give

$$\frac{1}{z-3} = -\frac{1}{3}\left(\frac{1}{1 - \frac{z}{3}}\right) = -\frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{z}{3}\right)^k.$$

Thus

$$f(z) = \frac{2}{z} - \frac{3}{z}\sum_{k=0}^{\infty}\frac{1}{z^k} - \frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{z}{3}\right)^k = \frac{2}{z} - \frac{3}{z} - \frac{3}{z}\sum_{k=1}^{\infty}\frac{1}{z^k} - \frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{z}{3}\right)^k$$

$$= -\frac{1}{z} + \sum_{k=1}^{\infty}(-3)\frac{1}{z^{k+1}} + \sum_{k=0}^{\infty}\left(-\frac{1}{3^{k+1}}\right)z^k = -\frac{1}{z} + \sum_{j=0}^{\infty}(-3)\frac{1}{z^j} + \sum_{k=0}^{\infty}\left(-\frac{1}{3^{k+1}}\right)z^k.$$

Here we have taken pains to write everything inside the sums, the better to see the coefficients on the powers of z. The Laurent decomposition is therefore

$$f_1(w) := -w + \sum_{j=0}^{\infty}(-3)w^j \quad \text{and} \quad f_3(w) := \sum_{k=0}^{\infty}\left(-\frac{1}{3^{k+1}}\right)w^k.$$

(iii) *On* $\mathcal{A}(0;3,\infty)$. We want to write

$$f(z) = f_3\left(\frac{1}{z}\right) + f_\infty(z),\ |3| < z(< \infty),$$

where $f_3\colon \mathcal{B}(0;1/3) \to \mathbb{C}$ and $f_\infty\colon \mathcal{B}(0;\infty) = \mathbb{C} \to \mathbb{C}$ are analytic. As before, since $3 < |z|$, we have $|1/z| < |3/z| < 1$, and so the expansion

$$-\frac{3}{z-1} = -\frac{3}{z}\sum_{k=0}^{\infty}\frac{1}{z^k}$$

is still valid. Next, however, we use the estimate $|3/z| < 1$ to obtain

$$\frac{1}{z-3} = \frac{1}{z\left(1-\frac{3}{z}\right)} = \frac{1}{z}\left(\frac{1}{1-\frac{3}{z}}\right) = \frac{1}{z}\sum_{k=0}^{\infty}\left(\frac{3}{z}\right)^k.$$

Thus

$$f(z) = \frac{2}{z} - \frac{3}{z}\sum_{k=0}^{\infty}\frac{1}{z^k} + \frac{1}{z}\sum_{k=0}^{\infty}\left(\frac{3}{z}\right)^k = \frac{2}{z} - \frac{3}{z} - \frac{3}{z}\sum_{k=1}^{\infty}\frac{1}{z^k} + \frac{1}{z} + \frac{1}{z}\sum_{k=1}^{\infty}\frac{3^k}{z^k}$$
$$= \sum_{k=1}^{\infty}(-3)\frac{1}{z^{k+1}} + \sum_{k=1}^{\infty}\frac{3^k}{z^{k+1}} = \sum_{k=1}^{\infty}(-3+3^k)\frac{1}{z^{k+1}} = \sum_{j=2}^{\infty}(-3+3^{j-1})\frac{1}{z^j}.$$

The Laurent decomposition is therefore

$$f_3(w) := \sum_{j=2}^{\infty}(-3+3^{j-1})w^j \quad \text{and} \quad f_\infty(w) = 0.$$

Here we are in the (surprising?) situation that the Laurent decomposition has no nonnegative powers of z!

This is where we finished on Wednesday, April 17, 2024.

(iv) *On* $\mathcal{A}(1;1,2)$. We want to write

$$f(z) = f_1\left(\frac{1}{z-1}\right) + f_2(z-1),\ 1 < |z-1| < 2,$$

where $f_1\colon \mathcal{B}(0;1/1) \to \mathbb{C}$ and $f_2\colon \mathcal{B}(0;2) \to \mathbb{C}$ are analytic. This requires artificially introducing $z-1$ into the quotients $2/z$ and $1/(z-3)$, and we do so by adding zero.

First,

$$\frac{2}{z} = \frac{2}{z-1+1}.$$

Since $|z-1| > 1$, we cannot expand this with a geometric series in powers of $z-1$. Instead, we divide and use the inequality $|1/(z-1)| < 1$ to obtain

$$\frac{2}{z} = \frac{2}{z-1+1} = \frac{2}{(z-1)\left(1+\dfrac{1}{z-1}\right)} = \frac{2}{z-1}\left(\frac{1}{1-\left(-\dfrac{1}{z-1}\right)}\right) = \frac{2}{z-1}\sum_{k=0}^{\infty}\left(-\frac{1}{z-1}\right)^k.$$

Next,

$$\frac{1}{z-3} = \frac{1}{z-1+1-3} = \frac{1}{(z-1)-2}.$$

We could either factor out $z-1$ or 2 to get a structure that looks like $1/(1-??)$ on the way to the geometric series. Since $|z-1| < 2$, we would have $2/|z-1| > 1$, so we factor out 2:

$$\frac{1}{z-3} = \frac{1}{(z-1)-2} = \frac{1}{2\left(\dfrac{z-1}{2}-1\right)}.$$

Since $|(z-1)/2| < 1$, we have

$$\frac{1}{z-3} = -\frac{1}{2}\left(\frac{1}{1-\dfrac{z-1}{2}}\right) = -\frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{z-1}{2}\right)^k.$$

Thus

$$\begin{aligned} f(z) &= \frac{2}{z-1}\sum_{k=0}^{\infty}(-1)^k\frac{1}{(z-1)^k} - \frac{3}{z-1} - \frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{z-1}{2}\right)^k \\ &= \frac{2}{z-1} + \frac{2}{z-1}\sum_{k=1}^{\infty}(-1)^k\frac{1}{(z-1)^k} - \frac{3}{z-1} + \sum_{k=0}^{\infty}\left(-\frac{1}{2^{k+1}}\right)(z-1)^k \\ &= -\frac{1}{z-1} + \sum_{k=1}^{\infty}2(-1)^k\frac{1}{(z-1)^{k+1}} + \sum_{k=0}^{\infty}\left(-\frac{1}{2^{k+1}}\right)(z-1)^k \end{aligned}$$

$$= -\frac{1}{z-1} + \sum_{j=2}^{\infty} 2(-1)^{j-1}\frac{1}{(z-1)^j} + \sum_{k=0}^{\infty}\left(-\frac{1}{2^{k+1}}\right)(z-1)^k.$$

The Laurent decomposition is therefore

$$f_1(w) = -w + \sum_{j=2}^{\infty} 2(-1)^{j-1}w^j \quad \text{and} \quad f_3(w) = \sum_{k=0}^{\infty}\left(-\frac{1}{2^{k+1}}\right)w^k.$$

Here, hopefully the placeholder variable w clarifies the different roles of the independent variable of f_1 and f_3 and the powers of $z-1$ in the series for f.

4.4.10 Problem (!). Find all of the other Laurent decompositions on the annuli drawn in Example 4.4.3.

4.4.11 Problem (!). The following identities are often useful when computing the Laurent series of a rational function with simple poles. Let z, $w \in \mathbb{C}$ with $|z| \neq |w|$. Show that

$$\frac{1}{z-w} = \begin{cases} \dfrac{1}{z\left(1-\dfrac{w}{z}\right)} = \displaystyle\sum_{k=0}^{\infty} \frac{w^k}{z^{k+1}}, & |w| < |z| \\[2em] -\dfrac{1}{w\left(1-\dfrac{z}{w}\right)} = -\displaystyle\sum_{k=0}^{\infty} \frac{z^k}{w^{k+1}}, & |z| < |w|. \end{cases}$$

4.4.12 Problem (!). Give an example of a function f with an essential singularity at some $z_0 \in \mathbb{C}$ such that the Laurent series $f(z) = \sum_{k=-\infty}^{\infty} a_k(z-z_0)^k$ has infinitely many coefficients a_k that *are* 0 with $k < 0$.

The Taylor series for a function analytic on a ball contains all the essential "data" for that function in its coefficients. If we know the countable sequence of coefficients in the Taylor series—a somewhat less than one-dimensional set of data—then (in principle) we know everything about that function in two dimensions on that ball, and in fact (in principle) on *any* ball centered at that point, regardless of the radius.

What data is contained in the Laurent coefficients of a function? Here we must remember that geometry, not just analysis, plays a role. In the preceding example, we saw that a function could have two very different Laurent series depending on the underlying annuli. If, in the case of an isolated singularity, we choose the annulus to be a punctured ball, we can glean a complete characterization of the singularity from the behavior of the Laurent coefficients. (Later we will see that one particular Laurent coefficient contains valuable data for our favorite activity of gathering other data: evaluating integrals.)

To ease our passage, we point out that if $f\colon \mathcal{A}(z_0; r, R) \to \mathbb{C}$ is analytic for some $z_0 \in \mathbb{C}$

and $0 \leq r < R \leq \infty$, and if we already know that

$$f(z) = \sum_{k=0}^{\infty} b_k (z - z_0)^k + \sum_{k=1}^{\infty} \frac{c_k}{(z - z_0)^k}$$

for $z \in \mathcal{A}(z_0; r, R)$ and some coefficients b_k, $c_k \in \mathbb{C}$, then by uniqueness, the coefficients b_k and c_k are the Laurent coefficients of f. Specifically, we could define

$$g_R \colon \mathcal{B}(0; R) \to \mathbb{C} \colon w \mapsto \sum_{k=0}^{\infty} b_k w^k \quad \text{and} \quad g_r \colon \mathcal{B}(0; 1/r) \to \mathbb{C} \colon w \mapsto \sum_{k=1}^{\infty} c_k w^k$$

to see that g_R and g_r are analytic and $g_r(0) = 0$. Since $f(z) = g_R(z - z_0) + g_r((z - z_0)^{-1})$ on $\mathcal{A}(z_0; r, R)$, the pair (g_r, g_R) is the Laurent decomposition of f on $\mathcal{A}(z_0; r, R)$.

Equipped with all of this information about the Laurent series, we are now able to characterize isolated singularities via the structure of Laurent coefficients. In particular, we confirm our prior expectation that essential singularities correspond to infinitely many negative powers in the series expansion.

4.4.13 Theorem. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f \colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ is analytic and let $f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k$ be the Laurent expansion of f on the annulus $\mathcal{A}(z_0; 0, r) = \mathcal{B}^*(z_0; r)$. Then*

(i) *f has a removable singularity at z_0 if and only if $a_k = 0$ for $k \leq -1$.*

(ii) *f has a pole of order $m \geq 1$ at z_0 if and only if $a_{-m} \neq 0$ and $a_k = 0$ for $k \leq -(m+1)$.*

(iii) *f has an essential singularity at z_0 if and only if $a_k \neq 0$ for infinitely many $k \leq -1$.*

Proof. **(i)** ($\Longrightarrow$) Suppose that f has a removable singularity at z_0. By Problem 4.3.6, we can write $f(z) = \sum_{k=0}^{\infty} b_k (z - z_0)^k$ for some $b_k \in \mathbb{C}$ and all $z \in \mathcal{B}^*(z_0; r)$. By the uniqueness of the Laurent series, we have $b_k = a_k$ for $k \geq 0$ and $a_k = 0$ for $k < 0$.

($\Longleftarrow$) If $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ for $z \in \mathcal{B}^*(z_0; r)$, then an analytic continuation of f to $\mathcal{B}(z_0; r)$ is just this series. Consequently, $\lim_{z \to z_0} f(z) = a_0$ exists, and so f has a removable singularity at z_0.

(ii) ($\Longrightarrow$) Suppose that f has a pole of order $m \geq 1$ at z_0. By Problem 4.3.19, there are $b_{-1}, \ldots, b_{-m} \in \mathbb{C}$ and an analytic function $g \colon \mathcal{B}(z_0; r) \to \mathbb{C}$ such that

$$f(z) = \sum_{k=1}^{m} \frac{b_{-k}}{(z - z_0)^k} + g(z), \ z \in \mathcal{B}^*(z_0; r), \quad \text{with} \quad b_{-m} \neq 0.$$

If we define

$$f_0 \colon \mathbb{C} \to \mathbb{C} \colon w \mapsto \sum_{k=1}^{m} b_{-k} w^k \quad \text{and} \quad f_r \colon \mathcal{B}(0; r) \to \mathbb{C} \colon w \mapsto g(w + z_0),$$

then (f_r, f_0) is a Laurent decomposition for f. In particular, $a_j = b_j$ for $j = -1, \ldots, -m$ and $a_j = 0$ for $j \leq -(m+1)$.

($\Longleftarrow$) Rewrite, for $z \in \mathcal{A}(z_0; r, R)$,

$$f(z) = \sum_{k=-m}^{\infty} a_k(z-z_0)^k = \sum_{k=-m}^{\infty} a_k(z-z_0)^{k+m}(z-z_0)^{-m} = \frac{1}{(z-z_0)^m}\sum_{k=-m}^{\infty} a_k(z-z_0)^{k+m}$$
$$= \frac{1}{(z-z_0)^m}\sum_{j=0}^{\infty} a_{j-m}(z-z_0)^j.$$

Put $p(z) := \sum_{j=0}^{\infty} a_{j-m}(z-z_0)^j$. Above we factored $f(z) = (z-z_0)^{-m}p(z)$, so the series $p(z)$ does converge for each $z \in \mathcal{A}(z_0; r, R)$. That is, the series converges for $r < |z-z_0| < R$, and so by Problem 4.1.10 it converges for all $z \in \mathcal{B}(z_0; R)$. Thus p is analytic on $\mathcal{B}(z_0; R)$. Moreover, $p(z_0) = a_{-m} \neq 0$. We conclude $f(z) = (z-z_0)^{-m}p(z)$ with p analytic on a ball centered at z_0 and $p(z_0) \neq 0$; hence f has a pole of order m at z_0.

(iii) ($\Longrightarrow$) Since z_0 is an essential singularity of f, z_0 is not a removable singularity, and so it cannot be the case that $a_k = 0$ for all $k \leq -1$. But z_0 is also not a pole, so it cannot be the case that $a_k = 0$ for all $k \leq -(m+1)$ for some integer $m \geq 1$. Thus, given any integer $m \geq 1$, there must be some integer $k < -m$ such that $a_k \neq 0$. We can therefore construct a sequence of infinitely many distinct points (a_{m_k}) such that $m_{k+1} < m_k < 0$ and $a_{m_k} \neq 0$ for all k.

($\Longleftarrow$) This is Problem 4.3.25. ■

4.4.14 Problem (+). Let $z_0 \in \mathbb{C}$ and $R > 0$. Suppose that $f\colon \mathcal{B}^*(z_0; R) \to \mathbb{C}$ is analytic. Here is another proof of the Riemann removability criterion (as stated in Problem 4.3.8) using Laurent series (and not using the "multiply by $z - z_0$" trick from that problem). The only hard thing to prove is the reverse direction (boundedness implies removable singularity), so let (a_k) be the Laurent coefficients of f; show that $a_k = 0$ for $k \leq -1$ by using the integral definition (4.4.2) for $s \in (0, \rho]$ and the ML-inequality. What happens in the limit of this integral as $s \to 0^+$?

4.4.15 Problem (⋆). Let $z_0 \in \mathbb{C}$ and $R > 0$. Suppose that $f\colon \mathcal{B}^*(z_0; R) \to \mathbb{C}$ is analytic. Prove that f has a removable singularity at z_0 if and only if $\lim_{z \to z_0}(z-z_0)f(z) = 0$. In the context of Problem 4.4.16, explain why we might euphemistically call a removable singularity a "pole of order 0."

4.4.16 Problem (+). Let $z_0 \in \mathbb{C}$ and $R > 0$. Suppose that $f\colon \mathcal{B}^*(z_0; R) \to \mathbb{C}$ is analytic. Prove that the following are equivalent.

(i) f has a pole of order $m \geq 1$ at z_0.

(ii) $\lim\limits_{z \to z_0}(z-z_0)^m f(z)$ exists and is nonzero.

(iii) $\lim\limits_{z \to z_0}(z-z_0)^{m+1} f(z) = 0$.

(iv) There exist $\rho \in (0, R]$ and $M > 0$ such that

$$|f(z)| \le \frac{M}{|z - z_0|^m} \text{ for } z \in \mathcal{B}^*(z_0; \rho).$$

4.5. Residue calculus.

So very many of our labors have involved line integrals. We built and characterized antiderivatives via line integrals, thereby completing one of the major stories of real-variable calculus in the complex setting. Moreover, we learned that the integral is *the* tool for extracting data about functions—specifically via the Cauchy integral formula and Taylor/Laurent coefficients. That story is more or less complete, and we will not typically succeed in finding antiderivatives for analytic functions on annuli.

4.5.1 Problem (!). Explain why. [Hint: *let* $z_0 \in \mathbb{C}$ *and* $0 < r < s < R \le \infty$ *and evaluate* $\int_{|z-z_0|=s} (z - z_0)^{-1} \, dz$. *Is the annulus* $\mathcal{A}(z_0; r, R)$ *an elementary domain?*]

4.5.1. Line integrals in annuli.

Nonetheless, we might ask what we can learn about line integrals of analytic functions over closed curves in annuli. Such integrals appeared so often in our former work that it is natural to pursue them further. So, let $z_0 \in \mathbb{C}$ and $0 \le r < R \le \infty$, and let $f \colon \mathcal{A}(z_0; r, R) \to \mathbb{C}$ be analytic. Let (f_R, f_r) be the Laurent decomposition of f in $\mathcal{A}(z_0; r, R)$, and let γ be a closed curve in $\mathcal{A}(z_0; r, R)$. Then

$$\int_\gamma f = \int_\gamma \big(f_R(z - z_0) + f_r((z - z_0)^{-1})\big) \, dz = \int_\gamma f_R(z - z_0) \, dz + \int_\gamma f_r((z - z_0)^{-1}) \, dz. \quad (4.5.1)$$

4.5.2 Problem (!). Recall that $f_R \colon \mathcal{B}(0; R) \to \mathbb{C}$ is analytic. Use this, the hypothesis that γ is a closed curve in $\mathcal{A}(z_0; r, R) \subseteq \mathcal{B}(z_0; R)$, and the Cauchy integral theorem to show that

$$\int_\gamma f_R(z - z_0) \, dz = 0.$$

Then (4.5.1) collapses to

$$\int_\gamma f = \int_\gamma f_r((z - z_0)^{-1}) \, dz. \quad (4.5.2)$$

In the notation of Theorem 4.4.6, write

$$f_r((z - z_0)^{-1}) = \sum_{k=1}^\infty \frac{a_{-k}}{(z - z_0)^k},$$

and remember that since γ is a curve in $\mathcal{A}(z_0; r, R)$, the point z_0 does not belong to the image of γ. Suppose that we can interchange the sum and integral to find

$$\int_\gamma f_r((z-z_0)^{-1})\ dz = \int_\gamma \sum_{k=1}^\infty a_{-k}\frac{a_k}{(z-z_0)^k}\ dz = \sum_{k=1}^\infty a_{-k}\int_\gamma \frac{dz}{(z-z_0)^k}. \tag{4.5.3}$$

This interchange can be justified by Theorem C.7.1 (which is really just an application of the original interchange theorem), and the integrals are all defined because z_0 does not lie on the image of γ. It then turns out that most of the series on the right of (4.5.3) will collapse to 0.

4.5.3 Problem (!). Use the fundamental theorem of calculus to show

$$\int_\gamma \frac{dz}{(z-z_0)^k} = 0,\ k \geq 2. \tag{4.5.4}$$

Also explain why (4.5.4) does not follow from the Cauchy integral theorem.

We combine (4.5.1), (4.5.2), (4.5.3), and (4.5.4) to conclude

$$\int_\gamma f = a_{-1}\int_\gamma \frac{dz}{z-z_0}. \tag{4.5.5}$$

For the purposes of calculating $\int_\gamma f$, all of the other data from the Laurent series was irrelevant; only the particular coefficient a_{-1} matters. Using the definition of a_{-1} from (4.4.2), the formula (4.5.5) reads

$$\int_\gamma f = \left(\frac{1}{2\pi i}\int_{|z-z_0|=s} f(z)\ dz\right)\left(\int_\gamma \frac{dz}{z-z_0}\right). \tag{4.5.6}$$

The line integral of f over γ is therefore the product of two integrals—one an integral of f over a (more or less) arbitrary circle, and one an integral of a "tame" rational function over the given curve γ. In other words, the data of the line integral—the curve γ and the integrand f—decouple into two integrals, one dependent on f (but not γ), and one dependent on γ (but not f), and both dependent on the center z_0 of the underlying annulus.

Both factors in (4.5.6) will reappear in our subsequent study of integrals in more general domains. We name and examine the second factor, adjusted slightly, first.

This is where we finished on Friday, April 19, 2024.

4.5.2. *The winding number.*

The following analytic concept remarkably encapsulates the geometric phenomenon of "orientation" for curves.

4.5.4 Definition. *Let γ be a closed curve in $\mathbb{C}$ and let $z \in \mathbb{C}$ be a point that is not in the* *image of γ. Then the* **WINDING NUMBER OF γ WITH RESPECT TO z** *is*

$$\chi(w;z) := \frac{1}{2\pi i}\int_\gamma \frac{dw}{w-z}.$$

Sometimes the winding number is called the **INDEX OF γ WITH RESPECT TO z.**

4.5.5 Problem (+). The winding number is indeed a "number" in the sense that it is an integer. In the following, let $\gamma\colon [0,1]\to\mathbb{C}$ be a continuously differentiable, closed curve and let $z\in\mathbb{C}$ not be in the image of γ.

(i) Show that $\chi(\gamma;z)\in\mathbb{C}$ if and only if

$$\exp\left(\int_\gamma \frac{dw}{w-z}\right) = 1.$$

(ii) Define

$$f\colon [0,1]\to\mathbb{C}\colon t\mapsto \int_0^1 \frac{\gamma'(\tau)}{\gamma(\tau)-z}\,d\tau.$$

Show that γ satisfies the ODE

$$\gamma'(t) - f'(t)\gamma(t) = -f'(t)z.$$

(iii) Multiply through by the integrating factor $e^{-f(t)}$ and conclude that

$$\gamma(t)e^{-f(t)} - \gamma(a)e^{-f(a)} = e^{-f(t)}z - e^{-f(a)}z.$$

(iv) Use this to show that

$$(\gamma(b)-z)(1-e^{f(b)}) = 0.$$

(v) Since z is not in the image of γ, conclude that $e^{f(b)}=1$, as desired.

(vi) How should you modify this argument for the case that γ is only piecewise continuously differentiable?

We can now rewrite (4.5.6) once again. Here is a summary of our work.

4.5.6 Theorem. *Let $z_0\in\mathbb{C}$ and $0\le r<R\le\infty$. Let $f\colon \mathcal{A}(z_0;r,R)\to\mathbb{C}$ be analytic and let γ be a closed curve in $\mathcal{A}(z_0;r,R)$. Then*

$$\int_\gamma f = \left(\int_{|z-z_0|=s} f(z)\,dz\right)\chi(\gamma;z_0),\ \ r<s<R.$$

We will develop and generalize this formula to the highly useful situation in which f has a finite number of isolated singularities within an elementary domain. First, however, we focus on the geometry of the winding number.

4.5.7 Example. Although it is not at all obvious at first glance, the winding number does what it promises. For $k \in \mathbb{Z} \setminus \{0\}$, $r > 0$, and $z_0 \in \mathbb{C}$, define

$$\gamma\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto z_0 + re^{ikt}.$$

Intuitively, we should view γ as "tracing out" the circle of radius r centered at z_0 a total number of $|k|$ times, with the circle oriented counterclockwise if $k > 0$ and clockwise if $k < 0$.

Now let $z \in \mathbb{C}$ with $|z - z_0| \neq r$. We can calculate

$$\int_\gamma \frac{dw}{w-z} = \begin{cases} 2\pi ik, & |z-z_0| < r \\ 0, & |z-z_0| > r, \end{cases} \tag{4.5.7}$$

and so

$$\chi(\gamma; z) = \begin{cases} k, & |z-z_0| < r \\ 0, & |z-z_0| > r. \end{cases}$$

In other words, $\chi(\gamma; z)$ "counts" the number of times that γ "winds around" z_0: either k times (with the sign of k indicating orientation) if z is "inside" the circle of radius r centered at z_0, or no times at all if z is "outside" this circle.

4.5.8 Problem ($\star$). Obtain the first identity in (4.5.7) by justifying each of the following equalities:

$$\int_0^{2\pi} \frac{rike^{ikt}}{z_0 + re^{ikt} - z}\, dt = \int_0^{2k\pi} \frac{rie^{i\tau}}{z_0 + re^{i\tau} - z}\, d\tau = \sum_{j=1}^{k} \int_{2(j-1)\pi}^{2j\pi} \frac{rie^{i\tau}}{z_0 + re^{i\tau} - z}\, d\tau$$
$$= k\int_0^{2\pi} \frac{rie^{i\tau}}{z_0 + re^{i\tau} - z}\, d\tau = k\int_{|w-z_0|=r} \frac{dw}{w-z} = 2\pi ik.$$

For the second, use the Cauchy integral theorem. What is the appropriate star domain?

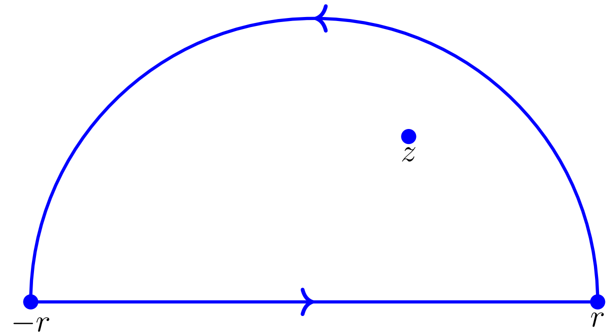
4.5.9 Example. In Problem 3.6.5, we calculated the following. Suppose that $r > 0$ and $z \in \mathbb{C}$ with $|z| < r$ and $\mathrm{Im}(z) > 0$. With

$$\gamma_r\colon [0, \pi] \to \mathbb{C}\colon t \mapsto re^{it},$$

we found

$$\chi(\gamma_r \oplus [-r, r]; z) = \frac{1}{2\pi i}\int_{[-r,r]\oplus\gamma_r} \frac{dw}{w-z} = 1.$$

That is, the semicircle drawn below "winds around" any point z in its "interior" exactly once.



Motivated by Examples 4.5.7 and 4.5.9, we can introduce some geometric notions for curves that we have heretofore avoided.

4.5.10 Definition. *Let γ be a closed curve in $\mathbb{C}$.*

(i) *The* **INTERIOR** *of γ is the set*

$$\operatorname{int}(\gamma) := \{z \in \mathbb{C} \mid \chi(\gamma; z) \neq 0\}.$$

(ii) *The* **EXTERIOR** *of γ is the set*

$$\operatorname{ext}(\gamma) := \{z \in \mathbb{C} \mid \chi(\gamma; z) = 0\}.$$

(iii) *The curve γ is* **POSITIVELY ORIENTED** *if $\chi(\gamma; z) > 0$ for all $z \in \operatorname{int}(\gamma)$ and* **NEGATIVELY ORIENTED** *if $\chi(\gamma; z) < 0$ for all $z \in \operatorname{int}(\gamma)$.*

4.5.11 Problem ($\star$). Let r, $b > 0$ and let

$$\gamma_{r,b} = [-r - ib, r - ib] \oplus [r - ib, r + ib] \oplus [r + ib, -r + ib] \oplus [-r + ib, r - ib].$$

(i) Draw a picture of $\gamma_{r,b}$.

(ii) Use the strategy of Problem 3.6.5 and the Cauchy integral theorem to compute $\chi(\gamma_{r,b}; z)$ for any $z \in \mathbb{C}$.

(iii) What is the interior of $\gamma_{r,b}$ and what is the exterior?

(iv) Is $\gamma_{r,b}$ positively or negatively oriented?

(v) Are these the results you expected from your picture?

4.5.3. The residue theorem.

The following situation often arises in practice. Let $\mathcal{D} \subseteq \mathbb{C}$ be an elementary domain—so $\mathcal{D}$ is open and connected, and if $h \colon \mathcal{D} \to \mathbb{C}$ is analytic and γ is a closed curve in $\mathcal{D}$, then $\int_\gamma h = 0$. Fix a finite number of distinct points $z_1, \ldots, z_n \in \mathcal{D}$, and let $f \colon \mathcal{D} \setminus \{z_k\}_{k=1}^n \to \mathbb{C}$

be analytic. Choose $r_k > 0$ such that $\mathcal{B}(z_k; r_k) \subseteq \mathcal{D}$ and if $j \neq k$, then $z_j \notin \mathcal{B}(z_k; r_k)$. Consequently, f is analytic on each $\mathcal{B}^*(z_k; R_k)$, and so each z_k is an isolated singularity of f. Let f_k be the principal part of the Laurent decomposition of $f\big|_{\mathcal{B}^*(z_k;r_k)}$, so each f_k is entire. (This is not exactly the notation that we used for principal parts before.)

It turns out that if we "remove" all the principal parts from f, then we are left with a rather nice function.

4.5.12 Lemma. *Under the hypotheses and notation above, the function*

$$g \colon \mathcal{D} \setminus \{z_k\}_{k=1}^n \to \mathbb{C} \colon z \mapsto f(z) - \sum_{k=1}^n f_k\left(\frac{1}{z - z_k}\right) \tag{4.5.8}$$

has removable singularities at z_k and consequently has an analytic continuation $\widetilde{g}$ on $\mathcal{D}$.

Proof. We do this only for $n = 2$, which will show all the steps necessary for the general case without some notational confusion. So, here

$$g(z) = f(z) - f_1\left(\frac{1}{z - z_1}\right) - f_2\left(\frac{1}{z - z_2}\right).$$

We want to show that

$$\lim_{z \to z_1} g(z) \quad \text{and} \quad \lim_{z \to z_2} g(z)$$

exist. We give the proof only for z_1, as the work for z_2 is almost identical. For z_1, in turn it suffices to show that

$$\lim_{z \to z_1} f_2\left(\frac{1}{z - z_2}\right) \qquad \text{and} \qquad \lim_{z \to z_2} f(z) - f_1\left(\frac{1}{z - z_1}\right)$$

exist.

The first limit is quite easy, since f_2 is entire and $z_1 \neq z_2$:

$$\lim_{z \to z_1} f_2\left(\frac{1}{z - z_2}\right) = f_2\left(\frac{1}{z_1 - z_2}\right).$$

For the second limit, let (f_1, h_1) be the full Laurent decomposition of $f\big|_{\mathcal{B}^*(z_1;r_1)}$ on the annulus $\mathcal{B}^*(z_1; r_1)$. To be fair, this is not the notation that we were using before, but what matters now is that $h_1 \colon \mathcal{B}(0; r_1) \to \mathbb{C}$ is analytic and

$$f(z) = f_1\left(\frac{1}{z - z_1}\right) + h_1(z - z_1), \ z \in \mathcal{B}^*(z_1; r_1).$$

Subtracting, we find

$$\lim_{z \to z_1} f(z) - f_1\left(\frac{1}{z - z_1}\right) = \lim_{z \to z_1} h_1(z - z_1) = h_1(0).$$

As we discussed above, the desired limit therefore exists, and in particular

$$\lim_{z\to z_1} g(z) = h_1(0) - f_2\left(\frac{1}{z_1 - z_2}\right).$$

The limit at z_2 can be calculated analogously, and the reasoning for $n \geq 3$ just involves more notation. ■

Let $\widetilde{g}$ be as in the preceding lemma and let γ be a closed curve in $\mathcal{D}$. Since $\widetilde{g}$ is analytic and $\mathcal{D}$ is an elementary domain, we have

$$\int_\gamma \widetilde{g} = 0.$$

Now we add the additional hypothesis that none of the points z_k belong to the image of γ. Then $\widetilde{g}(z) = g(z)$ for all z in the image of γ, and so

$$0 = \int_\gamma \widetilde{g} = \int_\gamma g.$$

Using the definition of g in (4.5.8), we have

$$0 = \int_\gamma f(z)\ dz - \sum_{k=1}^n \int_\gamma f_k\left(\frac{1}{z - z_k}\right)\ dz.$$

Since f_k is entire, it has the Taylor series expansion

$$f_k(w) = \sum_{j=1}^\infty a_{j,k} w^k,\ \ w \in \mathbb{C},$$

Theorem C.7.1 with $r = 0$ and $R = r_k$ shows

$$\int_\gamma f_k\left(\frac{1}{z - z_k}\right)\ dz = 2\pi i a_{1,k}\chi(\gamma; z_k) \tag{4.5.9}$$

and thus

$$0 = \int_\gamma f - 2\pi i \sum_{k=1}^n a_{1,k}\chi(\gamma; z_k). \tag{4.5.10}$$

This is where we finished on Monday, April 22, 2024.

Now it is time to name these coefficients $a_{k,1}$.

4.5.13 Definition. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \mathcal{B}^*(z_0; r) \to \mathbb{C}$ is analytic, and let $f(z) = \sum_{k=-\infty}^\infty a_k(z - z_0)^k$ be the Laurent series for f on $\mathcal{B}^*(z_0; r)$. The* **residue of**

f **AT** *z_0 is the coefficient a_{-1}, and we write*

$$\operatorname{Res}(f; z_0) = a_{-1} = \frac{1}{2\pi i} \int_{|z-z_0|=s} f(z)\ dz,\ 0 < s < r.$$

There are, happily, many methods of computing residues that do not rely on the definition, either via the integral or the Laurent series, and we will eventually develop some of them. More immediately, all of our work up to and including (4.5.10) can be summarized in one theorem, the mightiest and proudest of the Cauchy theorems.

4.5.14 Theorem (Cauchy's residue theorem). *Let $\mathcal{D} \subseteq \mathbb{C}$ be an elementary domain and let $z_1, \ldots, z_n \in \mathbb{C}$ be distinct points. Let $f\colon \mathcal{D} \setminus \{z_k\}_{k=1}^n \to \mathbb{C}$ be analytic, and let γ be a closed curve in $\mathcal{D}$ such that no z_k belongs to the image of γ. Then*

$$\int_\gamma f = 2\pi i \sum_{k=1}^n \operatorname{Res}(f; z_k)\chi(\gamma; z_k).$$

As with Theorem 4.5.6, the residue theorem perfectly decouples the problem of computing a line integral into two distinct problems: the analytic problem of finding the residue (which involves the integrand and not the underlying curve) and the geometric problem of computing the winding number (which involves the curve and not the function)—the two problems are connected in that they both involve the isolated singularities of the integrand.

4.5.15 Problem ($\star$). Let $\mathcal{D} \subseteq \mathbb{C}$ be an elementary domain and let $f\colon \mathcal{D} \to \mathbb{C}$ be analytic.

(i) How does the residue theorem imply the Cauchy integral theorem? That is, why does the residue theorem imply $\int_\gamma f = 0$ in this case? (Note that we used the Cauchy integral theorem in the proof of the residue theorem, so logically the residue theorem is not independent of the Cauchy integral theorem.)

(ii) Show that the residue theorem implies the following more general version of the Cauchy integral formula: if γ is a closed curve in $\mathcal{D}$ and $z \in \mathcal{D}$ does not belong to the image of γ, then

$$\frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-z}\ dw = f(z)\chi(\gamma; z). \tag{4.5.11}$$

(iii) Show that (4.5.11) implies the version of the Cauchy integral formula in Theorem 3.6.8. [Hint: *use Example 4.5.7.*]

(iv) Show that, in fact, (4.5.11) implies the Cauchy integral theorem as stated in (i). [Hint: *by Problem 3.1.25, since $\mathcal{D}$ is open, there is some $z \in \mathcal{D}$ such that z is not in the image of γ. Set $g(w) = (w - z)f(w)$ and apply* (4.5.11) *to g in lieu of f.*]

4.5.16 Problem (!). Use Theorem 4.5.6 to prove the residue theorem in the very special case $n = 1$ and $\mathcal{D} = \mathcal{B}(z_1; r)$ for some $r > 0$.

4.5.17 Example. We revisit the integrand from Examples 3.5.15, 3.6.11, and 3.6.18 one last time. Let z_1, $z_2 \in \mathbb{C}$ with $0 \le |z_1| < z_2$ and let $r > 0$ such that z_1, $z_2 \in \mathcal{B}(0; r)$. We compute the value of

$$\int_{|z|=r} \frac{dz}{(z - z_1)(z - z_2)^2},$$

which (as we saw in Problem 3.6.19) cannot be done directly with the Cauchy integral theorem or the (generalized) Cauchy integral formula.

In the notation of the residue theorem, let $\mathcal{D} = \mathbb{C}$, so $\mathcal{D}$ certainly is an elementary domain (it is star-shaped). Define

$$f : \mathbb{C} \setminus \{z_1, z_2\} \to \mathbb{C} : z \mapsto \frac{1}{(z - z_1)(z - z_2)^2},$$

so f has isolated singularities at z_1 and z_2. Let

$$\gamma : [0, 2\pi] \to \mathbb{C} : t \mapsto re^{it},$$

so γ is a closed path in $\mathcal{D}$. Then

$$\int_{|z|=r} \frac{dz}{(z - z_1)(z - z_2)^2} = 2\pi i\big(\operatorname{Res}(f; z_1)\chi(\gamma; z_1) + \operatorname{Res}(f; z_2)\chi(\gamma; z_2)\big).$$

Example 4.5.7 shows that

$$\chi(\gamma; z_1) = \chi(\gamma; z_2) = 1,$$

so we just need to compute the residues. A partial fractions expansion for f will help:

$$f(z) = \frac{1}{(z_1 - z_2)^2(z - z_1)} - \frac{1}{(z_1 - z_2)^2(z - z_2)} - \frac{1}{(z_1 - z_2)(z - z_2)^2}. \tag{4.5.12}$$

In fact, this gives the Laurent decompositions (though not precisely the Laurent *series*) of f at z_1 and z_2:

$$f(z) = \underbrace{\frac{1}{(z_1 - z_2)^2(z - z_1)}}_{\text{pole at } z_1, \text{ analytic at } z_2} - \underbrace{\left(\frac{1}{(z_1 - z_2)^2(z - z_2)} + \frac{1}{(z_1 - z_2)(z - z_2)^2}\right)}_{\text{pole at } z_2, \text{ analytic at } z_1}.$$

The terms that are analytic at z_1 will not contribute any terms involving $(z - z_1)^{-1}$ to the Laurent series at z_1; the term that is analytic at z_2 will not contribute any terms involving $(z - z_2)^{-1}$ to the Laurent series at z_2. Thus

$$\operatorname{Res}(f; z_1) = \frac{1}{(z_1 - z_2)^2} \quad \text{and} \quad \operatorname{Res}(f; z_2) = -\frac{1}{(z_1 - z_2)^2}.$$

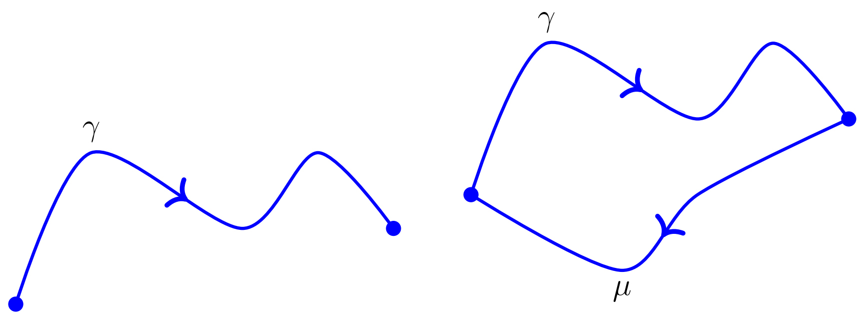
That is, $\operatorname{Res}(f; z_1) = -\operatorname{Res}(f; z_2)$, and so

$$\int_{|z|=r} \frac{dz}{(z - z_1)(z - z_2)^2} = 2\pi i\big(\operatorname{Res}(f; z_1) + \operatorname{Res}(f; z_2)\big) = 0.$$

4.5.18 Problem (!). Use the partial fractions expansion (4.5.12), the Death Star lemma, and the fundamental theorem of calculus to give a different proof that $\int_{|z|=r} f(z)\,dz = 0$. Such a calculation would have resolved the impasse in Problem 3.6.19 with the prior Cauchy theorems.

4.5.4. Evaluating real-valued integrals on (subintervals of) $\mathbb{R}$.

The residue theorem is a pathway to many abilities that some consider to be unnatural, at least from the point of view of real-variable calculus. One classical application is evaluating integrals, definite or improper, of functions defined on (subintervals of) $\mathbb{R}$. The broad idea is that one modifies the interval of integration into a path γ in $\mathbb{C}$, augments that path with another path μ so that $\gamma \oplus \mu$ is closed, and then integrates over this closed path and evaluates that integral with the residue theorem. Hopefully the auxiliary integral over μ is easy to control so that information is recovered about the desired integral over γ.



$$\int_\gamma f = \underbrace{\int_{\gamma\oplus\mu} f}_{\text{Residue theorem}} - \underbrace{\int_\mu f}_{\text{Something else?}}$$

There are many, many integrals that can be evaluated exactly using residue techniques. Here we do just one: we revisit the Fourier transform from Section 3.5.4 and instead of estimating the Fourier integral we calculate one precisely. Our goal is less to develop any systematic, broadly adaptable set of techniques and more to see just one specific application.

4.5.19 Theorem. *Define*

$$\operatorname{sech}\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto \frac{2}{e^t + e^{-t}}.$$

Then

$$\int_{-\infty}^{\infty} \operatorname{sech}(t) e^{\pm ikt}\,dt = \pi \operatorname{sech}\left(\frac{\pi k}{2}\right), \quad k \in \mathbb{R}. \tag{4.5.13}$$

The function sech (or more precisely a rescaling of its square) appears as a solution to the celebrated Korteweg–de Vries (KdV) partial differential equation, which models the behavior of nonlinear waves in shallow water; this function is a versatile tool in many areas of applied mathematics. First, the integral in (4.5.13) converges by the comparison test in Theorem

3.5.18 and, from that theorem and the evenness of sech, we have

$$\int_{-\infty}^{\infty} \operatorname{sech}(t)e^{\pm ikt}\,dt = \lim_{R\to\infty}\int_{-R}^{R} \operatorname{sech}(t)e^{ikt}\,dt.$$

4.5.20 Problem (!). Check that.

It therefore suffices to understand the definite integral $\int_{-R}^{R} \operatorname{sech}(t)e^{ikt}\,dt$ for large R. The integrand under consideration is really the restriction of

$$g(z) := \frac{2e^{ikz}}{e^z + e^{-z}}$$

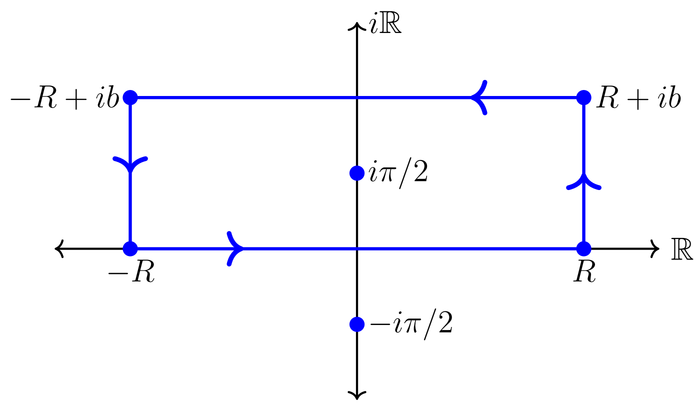
to $\mathbb{R}$. We are not indicating dependence on k (it is fixed in $\mathbb{R}$ throughout this discussion) and where g is defined for all z such that $e^z + e^{-z} \neq 0$.

4.5.21 Problem (!). Check that g is defined on $\mathbb{C} \setminus \{i\pi/2 + i\pi k \mid k \in \mathbb{Z}\}$. Also, check that g has poles of order 1 at each point $i\pi/2 + i\pi k$, $k \in \mathbb{Z}$.

We will view the interval $[-R, R]$ as the path γ discussed above and augment this interval with a path μ so that the composition $\gamma \oplus \mu$ is a rectangle, just like in Section 3.5.4. More precisely, fix b, $R > 0$ and consider the path

$$\gamma_{R,b} := [-R, R] \oplus [R, R + ib] \oplus [R + ib, -R + ib] \oplus [-R + ib, -R].$$

This is a closed path, and for $0 < b < \pi/2$, the integrand g is analytic on a star domain containing this path. (Take the star domain to be a slightly larger rectangle that does not contain $\pm i\pi/2$.) Then we have $\int_{\gamma_{R,b}} g = 0$, and we are back in the situation of Section 3.5.4. This will actually not be very helpful to us here, and so instead we will take $b \in (\pi/2, 3\pi/2)$, so that the interior (in the sense of Definition 4.5.10) of $\gamma_{R,b}$ contains only one isolated singularity of g: $i\pi/2$.



The residue theorem implies

$$\int_{\gamma_{R,b}} g = 2\pi i \operatorname{Res}\left(g; \frac{i\pi}{2}\right)\chi\left(\gamma_{R,b}; \frac{i\pi}{2}\right) = 2\pi i \operatorname{Res}\left(g; \frac{i\pi}{2}\right),$$

since $\chi(\gamma_{R,b}; i\pi/2) = 1$ by Problem 4.5.11. It therefore suffices to calculate this residue and to control

$$\mathcal{I}_{R,b} := \int_{[R,R+ib]\oplus[R+ib,-R+ib]\oplus[-R+ib,-R]} g \tag{4.5.14}$$

for large R; then we will be have

$$\int_{-R}^{R} g = \int_{[-R,R]} g = \int_{\gamma_{R,b}} g - \mathcal{I}_{R,b} = 2\pi i \operatorname{Res}\left(g; \frac{i\pi}{2}\right) - \mathcal{I}_{R,b}. \tag{4.5.15}$$

First we calculate the residue. We are really in the following situation.

4.5.22 Lemma. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that p, $q\colon \mathcal{B}(z_0; r) \to \mathbb{C}$ are analytic with $p(z_0) \neq 0$ and $q(z_0) = 0$. If q has a simple zero at z_0, then*

$$\operatorname{Res}\left(\frac{p}{q}; z_0\right) = \frac{p(z_0)}{q'(z_0)}.$$

4.5.23 Problem (!). Prove this lemma. [Hint: *write $q(z) = (z - z_0)r(z)$, where for some $\rho \in (0, r]$, the map $r\colon \mathcal{B}(z_0; \rho) \to \mathbb{C}$ is analytic with $r(z) \neq 0$ for all z. Explain why $r(z_0) = q'(z_0)$. Then factor*

$$\frac{p(z)}{q(z)} = \left(\frac{p(z)}{r(z)}\right)\frac{1}{z - z_0}.$$

Expand p/r in its Taylor series at z_0, identify the constant term, and conclude $\operatorname{Res}(p/q; z_0) = p(z_0)/r(z_0)$.]

Thus with $p(z) = 2e^{ikz}$ and $q(z) = e^{z} + e^{-z}$, we have

$$\operatorname{Res}\left(g; \frac{i\pi}{2}\right) = \operatorname{Res}\left(\frac{2e^{ikz}}{e^{z} + e^{-z}}; z = \frac{i\pi}{2}\right) = \frac{2e^{ik(i\pi/2)}}{e^{i\pi/2} - e^{-i\pi/2}} = \frac{2e^{-k\pi/2}}{i - (-i)} = -ie^{-k\pi/2}. \tag{4.5.16}$$

This is where we finished on Wednesday, April 24, 2024.

Now we return to the problem of controlling the integral $\mathcal{I}_{R,b}$. Our experience in Section 3.5.4 should suggest that the integrals over the vertical sides will vanish as $R \to \infty$.

4.5.24 Problem (⋆). Prove that. Specifically, show

$$\lim_{R\to\infty} \int_{[R,R+ib]} g = \lim_{R\to\infty} \int_{[-R+ib,-R]} g.$$

[Hint: *adapt the work from Step 4 in Section 3.5.4.*]

Now we manipulate the integral over the top side, and it is here that a precise choice of b will be helpful. The line segment $[R + ib, -R + ib]$ is parametrized by

$$\lambda_{R,b}\colon [0, 1] \to \mathbb{C}\colon t \mapsto (1 - t)(R + ib) + t(-R + ib) = R + ib - 2tR,$$

so

$$\int_{[R+ib,-R+ib]} g = -2R\int_0^1 g(R+ib-2tR)\ dt.$$

This is not a very transparent formula, but we might think that the integral over $[R+ib,-R+ib]$ should basically be the integral over $[-R,R]$, just reversed and shifted upwards. With this insight, we substitute $u = R-2tR$ to find

$$-2R\int_0^1 g(R+ib-2tR)\ dt = \int_R^{-R} g(u+ib)\ du = -\int_{-R}^{R} g(u+ib)\ du.$$

That looks much more like $\int_{-R}^{R} g$.

How much more? We have

$$\int_{-R}^{R} g(u+ib)\ du = \int_{-R}^{R} \frac{2e^{ik(u+ib)}}{e^{u+ib}+e^{-(u+ib)}}\ du = e^{-kb}\int_{-R}^{R}\frac{2e^{iku}}{e^{ib}e^{u}+e^{-ib}e^{-u}\ du}.$$

If we take $b=\pi$, then we will have $e^{\pm ib}=-1$, so the denominator will factor nicely, and we will still retain our restriction of keeping $b \in (\pi/2, 3\pi/2)$. Thus

$$\int_{[R+i\pi-R+i\pi]} g = -\int_{-R}^{R} g(u+i\pi)\ du = e^{-k\pi}\int_{-R}^{R}\frac{2e^{iku}}{e^{u}+e^{-u}}\ du = e^{-k\pi}\int_{-R}^{R} g$$

and so, by the definition of $\mathcal{I}_{R,b}$ in (4.5.14), we have

$$\mathcal{I}_{R,b} = \int_{[R,R+ib]\oplus[R+ib,-R+ib]\oplus[-R+ib,-R]} g = e^{-k\pi}\int_{-R}^{R} g + \int_{[R,R+ib]} g + \int_{[-R+ib,-R]} g. \quad (4.5.17)$$

We combine everything from (4.5.15), (4.5.16), and (4.5.17) to obtain

$$\int_{-R}^{R} g = 2\pi i \operatorname{Res}\left(g;\frac{i\pi}{2}\right) - \mathcal{I}_{R,b} = 2\pi i(-ie^{-k\pi/2}) - e^{-k\pi}\int_{-R}^{R} g - \int_{[R,R+ib]} g - \int_{[-R+ib,-R]} g.$$

A bit of algebra simplifies this to

$$(1+e^{-k\pi})\int_{-R}^{R} g = 2\pi e^{-k\pi/2} - \int_{[R,R+ib]} g - \int_{[-R+ib,-R]} g,$$

and then

$$\int_{-R}^{R} g = \frac{2\pi e^{-k\pi/2}}{1+e^{-k\pi}} - \frac{1}{1+e^{-k\pi}}\left(\int_{[R,R+ib]} g + \int_{[-R+ib,-R]} g\right).$$

Taking the limit as $R\to\infty$ and using Problem 4.5.24, we conclude

$$\int_{-\infty}^{\infty}\operatorname{sech}(t)e^{ikt}\ dt = \lim_{R\to\infty}\int_{-R}^{R} g = \frac{2\pi e^{-k\pi/2}}{1+e^{-k\pi}} = \pi\left(\frac{2}{e^{k\pi/2}+e^{-k\pi/2}}\right) = \pi\operatorname{sech}\left(\frac{k\pi}{2}\right),$$

as claimed in the theorem.

4.5.25 Problem (!). Explain why the exact result of Theorem 4.5.19 agrees with the estimate of Theorem 3.5.21.

4.5.5. The open mapping theorem.

We study here two consequences of the residue theorem on the way to proving our very last result, the open mapping theorem. The first two results count, up to multiplicities, the zeros and/or poles of an analytic function in certain regions; since many problems can be posed in the form "solve $f(z) = w$ for z given w," equivalently, "find the roots of $g(z) := f(z) - w$ given f and w," these are worthwhile tools to possess.

The open mapping theorem is one avenue toward finishing our story of complex analysis. We have said many times how special analytic functions are compared to infinitely differentiable or even real analytic functions on (subsets of) $\mathbb{R}$. In particular, in our study of the Cauchy–Riemann equations, we saw that analytic functions are genuinely two-dimensional objects in the sense that they cannot take strictly real or strictly imaginary values (i.e., their ranges cannot be contained in the coordinate axes) unless they are (locally) constant. Now we will see something stronger: the range of a nonconstant analytic function is open, and so around every point in the range we can fit a ball that is also contained in the range.

The result below relates a discrete quantity—the (difference between the) number of zeros and poles that a function has inside a closed path—to a continuous one—the line integral involving that path and that function.

4.5.26 Theorem (Counting). *Let $\mathcal{D} \subseteq \mathbb{C}$ be an elementary domain and let $p_1, \ldots, p_N \in \mathcal{D}$. Suppose that $f \colon \mathcal{D} \setminus \{p_k\}_{k=1}^N$ is analytic with poles of order n_k at p_k. Suppose also that $z_1, \ldots, z_M \in \mathcal{D} \setminus \{p_k\}_{k=1}^N$ are the zeros of f, and that m_k is the order of the zero at z_k. Finally, let γ be a closed path in $\mathcal{D} \setminus \{p_k\}_{k=1}^n$ such that $\chi(\gamma; p_k) = \chi(\gamma; z_k) = 1$ for all k. Then*

$$\sum_{k=1}^M m_k - \sum_{k=1}^N p_k = \frac{1}{2\pi i} \int_\gamma \frac{f'}{f}.$$

Proof. The map f'/f is analytic on $\mathcal{D} \setminus \left(\{p_k\}_{k=1}^N \cup \{z_k\}_{k=1}^M\right)$ with isolated singularities at z_k and p_k, so by the residue theorem

$$\int_\gamma \frac{f'}{f} = 2\pi i \left[\sum_{k=1}^N \operatorname{Res}\left(\frac{f'}{f}; p_k\right) \chi(\gamma; p_k) + \operatorname{Res}\left(\frac{f'}{f}; z_k\right) \chi(\gamma; z_k)\right].$$

Since $\chi(\gamma; p_k) = \chi(\gamma; z_k) = 1$, this reduces to

$$\int_\gamma \frac{f'}{f} = 2\pi i \left[\sum_{k=1}^N \operatorname{Res}\left(\frac{f'}{f}; p_k\right) + \operatorname{Res}\left(\frac{f'}{f}; z_k\right)\right].$$

We claim (and prove below) that

$$\operatorname{Res}\left(\frac{f'}{f}; p_k\right) = -n_k \quad \text{and} \quad \operatorname{Res}\left(\frac{f'}{f}; z_k\right) = m_k$$

and from this the theorem follows. ■

Continue to assume the notation and hypotheses of the counting theorem. We can view the integer

$$\mathsf{Z}(f;\gamma) := \sum_{k=1}^{M} m_k$$

as giving the number of zeros of f in the interior of γ, counting multiplicity, and the number

$$\mathsf{P}(f;\gamma) := \sum_{k=1}^{N} n_k$$

as the number of poles of f in the interior of γ, again counting multiplicity.

4.5.27 Problem ($\star$). Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \mathcal{B}^*(z_0;r) \to \mathbb{C}$ is analytic with $f(z) \neq 0$ on $\mathcal{B}^*(z_0;r)$. Prove that

$$\operatorname{Res}\left(\frac{f'}{f}; z_0\right) = \begin{cases} -m, & z_0 \text{ is a pole of order } m \text{ of } f \\ m, & z_0 \text{ is a zero of order } m \text{ of } f. \end{cases}$$

[Hint: *Write* $f(z) = (z - z_0)^{\pm m} p(z)$, *where* $p\colon \mathcal{B}(z_0;\rho) \to \mathbb{C}$ *is analytic for some* $\rho \in (0, r]$ *and* $p(z_0) \neq 0$. *Compute* f'/f *in terms of* z_0, m, *and* p *and look for the Laurent decomposition to expose the term containing* $(z - z_0)^{-1}$.]

The next result effectively says that if one function is a suitably small perturbation of another on the image of a closed path—"suitably small" to be made precise below—then both functions have, up to multiplicity, the same number of zeros inside the path.

4.5.28 Theorem (Rouché). *Let* $\mathcal{D} \subseteq \mathbb{C}$ *be open and suppose that* f, $g\colon \mathcal{D} \to \mathbb{C}$ *are analytic and* γ *is a closed path in* $\mathcal{D}$. *If*

$$|g(z) - f(z)| < |f(z)| \tag{4.5.18}$$

for all z *in the image of* γ, *then*

$$\mathsf{Z}(f;\gamma) = \mathsf{Z}(g;\gamma). \tag{4.5.19}$$

Proof. We first claim that neither f nor g has a zero in the image of γ. Otherwise, if $f(z_0) = 0$ for some z_0 in the image of γ, then (4.5.18) implies

$$0 = |f(z_0)| > |g(z_0) - f(z_0)| = |g(z_0)| \geq 0,$$

thus $0 < 0$, which is impossible. Similarly, if $g(z_0) = 0$ for some z_0 in the image of γ, then

$$|f(z_0)| > |g(z_0) - f(z_0)| = |f(z_0),$$

and this, too, is impossible.

Now, suppose that the domain of γ is $[a,b] \subseteq \mathbb{R}$. Then

$$\mu\colon [a,b] \to \mathbb{C}\colon t \mapsto \frac{g(\gamma(t))}{f(\gamma(t))}$$

is a path in $\mathbb{C}$, and more precisely μ is a path in $\mathcal{B}(1;1)$. We have

$$|g(\gamma(t)) - f(\gamma(t))| < |f(\gamma(t))|$$

for each t, so dividing by $|f(\gamma(t))|$ yields $|\mu(t) - 1| < 1$.

Finally, the map $\mathcal{B}(1;1) \to \mathbb{C}\colon z \mapsto 1/z$ is analytic on the star domain $\mathcal{B}(1;1)$, and so the Cauchy integral theorem implies

$$\int_\mu \frac{dz}{z} = 0. \tag{4.5.20}$$

(This is possibly the first time in our entire story that this kind of integral has vanished!) The definition of the line integral yields

$$\int_\mu \frac{dz}{z} = \int_\gamma \frac{(g/f)'}{g/g} \tag{4.5.21}$$

as well. But the counting theorem computes this as

$$\int_\gamma \frac{(g/f)'}{g/f} = \mathsf{Z}\left(\frac{g}{f};\gamma\right) - \mathsf{P}\left(\frac{g}{f};\gamma\right). \tag{4.5.22}$$

We combine (4.5.20), (4.5.21), and (4.5.22) to obtain

$$\mathsf{Z}\left(\frac{g}{f};\gamma\right) = \mathsf{P}\left(\frac{g}{f};\gamma\right). \tag{4.5.23}$$

We claim that

$$\mathsf{Z}\left(\frac{g}{f};\gamma\right) = \mathsf{Z}(g;\gamma) \quad \text{and} \quad \mathsf{P}\left(\frac{g}{f};\gamma\right) = \mathsf{Z}(f;\gamma). \tag{4.5.24}$$

Both identities should be intuitively plausible: the only zeros of g/f can come from g (a quotient is 0 if and only if the numerator is 0), and, since g is analytic, the only poles of g/f can come from f (poles effectively correspond to dividing by 0). Assuming (4.5.24) to be true, the equality in (4.5.24) then reduces to the desired $\mathsf{Z}(g;\gamma) = \mathsf{Z}(f;\gamma)$. ■

Here is how we can view g as a perturbation of f: by adding 0,

$$g = (g - f) + f.$$

Then g is a "suitably small" perturbation of f if $|g(z) - f(z)|$ is "suitably small," and the estimate (4.5.18) quantifies that precisely

4.5.29 Problem (!). Complete the proof of Rouché's theorem by filling in the following gaps.

(i) Prove (4.5.21). [Hint: *consider separately the cases that γ is a smooth path and only a path.*]

(ii) Prove (4.5.24). [Hint: *elaborate on the intuitive remarks in the proof.*]

We can view the estimate (4.5.18) in Rouché's theorem as saying that f and g are small perturbations of each other:

$$g(z) = \big(g(z) - f(z)\big) + f(z).$$

If the "perturbation" term $g(z) - f(z)$ is suitably small—namely, $|g(z) - f(z)| < |f(z)|$—then g and f have the same number of zeros in the interior of γ, counting multiplicities.

4.5.30 Problem (!). **(i)** Show that Rouché's theorem remains true if the estimate (4.5.18) is replaced by $|g(z) + f(z)| < |f(z)|$.

(ii) If instead of the estimate (4.5.18) we have $|g(z)| < |f(z)|$, how does the conclusion (4.5.19) change?

4.5.31 Theorem (Open mapping). *Let $\mathcal{D} \subseteq \mathbb{C}$ be open and let $f\colon \mathcal{D} \to \mathbb{C}$ be analytic and nonconstant (i.e., there exist z_1, $z_2 \in \mathcal{D}$ such that $f(z_1) \neq f(z_2)$). Then $f(\mathcal{D}) = \{f(z) \mid z \in \mathcal{D}\}$ is open.*

Proof. Fix $z_0 \in \mathcal{D}$. Our goal is to find $r > 0$ such that $\mathcal{B}(f(z_0); r) \subseteq f(\mathcal{D})$. That is, if $w \in \mathbb{C}$ with $|f(z_0) - w| < r$, then there is $z \in D$ such that $f(z) = w$.

The key idea in the proof is that $f(z) = w$ if and only if $f(z) - w = 0$. That is, this "surjectivity" problem is really a root-finding problem.

Define

$$g\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto f(z) - f(z_0).$$

Then $g(z_0) = 0$, but g is not identically zero, since f is not constant. Thus z_0 must be an isolated zero of g, so there is $s > 0$ such that $\mathcal{B}(z_0; s) \subseteq \mathcal{D}$ and $g(z) \neq 0$ for $z \in \mathcal{B}^*(z_0; s)$. By the extreme value theorem, the number

$$r := \min_{|z - z_0| \leq s/2} |g(z)|$$

is defined. We will show that $\mathcal{B}(f(z_0); r) \subseteq f(\mathcal{D})$.

To do this, fix $w \in \mathcal{B}(f(z_0); r)$ and consider the map

$$h\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto f(z) - w.$$

We want to be sure that h has a root, and specifically we will show that h has a root in $\mathcal{B}(z_0; s/2)$. Let $\gamma\colon [0, 2\pi] \to \mathbb{C}\colon t \mapsto z_0 + se^{it}/2$. Then, if z is in the image of γ,

$$|g(z) - h(z)| = |w - f(z_0)| < r \leq |g(z)|.$$

Thus Rouché's theorem implies $\mathsf{Z}(g;\gamma) = \mathsf{Z}(h;\gamma)$.

Since $g(z_0) = 0$, we have $1 \le \mathsf{Z}(g;\gamma)$, and so $1 \le \mathsf{Z}(h;\gamma)$. Consequently, h has at least one root in the ball $\mathcal{B}(z_0; s/2)$. ■

4.5.32 Problem (!). Let $\mathcal{D} \subseteq \mathbb{C}$ be open and $f : \mathcal{D} \to \mathbb{C}$ be analytic and nonconstant.

(i) Prove that the image of f cannot be the image of a path. [Hint: *Problem 3.1.25.*]

(ii) Prove that the image of f is in fact a domain. [Hint: *if γ is a path in $\mathcal{D}$, then $f \circ \gamma$ is a path in $f(\mathcal{D})$.*]

This is where we finished on Friday, April 26, 2024.

A. Very Elementary Set Theory

A.1. Sets and operations on sets.

This appendix corresponds to material covered on Monday, January 8, 2024.

We will frequently work with sets of real and complex numbers. To do so efficiently, we need a very small number of set-theoretic concepts.

A.1.1 Undefinition. *A* **SET** *is a collection of objects, called* **ELEMENTS**. *If x is an element of the set A, then we write $x \in A$, and if y is not an element of the set A, then we write $y \notin A$.*

This is an undefinition, not a definition, because we have not defined what “collections” or “objects” means. And we will not. If a set A consists of only finitely many elements, then we may denote A by listing those elements between curly braces. For example, the set consisting precisely of the numbers 1, 2, and 3 is $\{1, 2, 3\}$; the set consisting precisely of the number 1 is $\{1\}$, and $1 \in \{1\}$.

A.1.2 Example. Let $A = \{1, 2, 3\}$. Then $1 \in A$ but $4 \notin A$.

If U is a set, and if $P(x)$ is a statement that is either true or false for each $x \in U$, then we denote the set of all elements x of U for which $P(x)$ is true by

$$\{x \in U \mid P(x)\}.$$

A.1.3 Example. If $U = \{1, 2, 3\}$, then

$$\{x \in U \mid x \text{ is even}\} = \{2\}.$$

A.1.4 Definition. *A set A is a* **SUBSET** *of a set B if for each $x \in A$, it is the case that $x \in B$. That is, every element of A is an element of B. If A is a subset of B, we write $A \subseteq B$.*

In symbols,

$$A \subseteq B \iff (x \in A \Longrightarrow x \in B).$$

A.1.5 Example. $\{1, 2\} \subseteq \{1, 2, 3\}$ and $\{1, 2, 3\} \subseteq \{1, 2, 3\}$.

A.1.6 Definition. *Two sets A and B are* **EQUAL**, *written $A = B$, if $A \subseteq B$ and $B \subseteq A$. An element x and the set $\{x\}$ whose sole element is x cannot be equal: $x \neq \{x\}$.*

In symbols,

$$A = B \iff (x \in A \iff x \in B).$$

A.1.7 Hypothesis. **(i)** *There exists a set* $\varnothing$ *that contains no element. That is, if* x *is an element of any set* U*, then* $x \notin \varnothing$.

(ii) *An element* x *of a set* U *cannot be equal to the set* $\{x\}$ *whose only element is* x*. That is,* $x \neq \{x\}$.

A.1.8 Definition. *Let* A *and* B *be subsets of the set* U*. The* **UNION** *of* A *and* B *is the set*

$$A \cup B := \{x \in U \mid x \in A \text{ or } x \in B\},$$

the **INTERSECTION** *of* A *and* B *is the set*

$$A \cap B := \{x \in U \mid x \in A \text{ and } x \in B\},$$

and the **COMPLEMENT** *of* A *in* B *is the set*

$$B \setminus A := \{x \in B \mid x \notin A\}.$$

That is, $A \cup B$ *is the set of all elements in either* A *or* B *(or both),* $A \cap B$ *is the set of all elements in both* A *and* B*, and* $B \setminus A$ *is the set of all elements in* B *but not in* A.

A.1.9 Example. Let

$$A = \{1, 2, 3\} \quad \text{and} \quad B = \{2, 4, 6\}.$$

Then

$$A \cup B = \{1, 2, 3, 4, 6\},$$

$$A \cap B = \{2\},$$

and

$$B \setminus A = \{4, 6\}.$$

A.1.10 Problem (!). Let A and B be as in Example A.1.9. Determine the elements of each of the following sets.

(i) $A \setminus B$

(ii) $(A \setminus B) \cup B$

(iii) $(A \cap B) \setminus A$

(iv) $A \setminus \varnothing$

(v) $\varnothing \setminus B$

A.2. Composition of functions.

This appendix corresponds to material covered on Friday, January 12, 2024.

If the range of one function is contained in the domain of another function, we can compose them.

A.2.1 Definition. *Let A, B, C, and D be sets and let $f\colon A \to B$ and $g\colon C \to D$ be sets. Suppose that $f(A) \subseteq \mathbb{C}$. The* **COMPOSITION OF f WITH g** *is the function*

$$g \circ f\colon A \to D\colon x \mapsto g(f(x)).$$

A.2.2 Example. Define

$$f\colon \mathbb{C} \setminus \{\pm i\} \to \mathbb{C} \setminus \{0\}\colon z \mapsto z^2 + 1 \quad \text{and} \quad g\colon \mathbb{C} \setminus \{0\} \to \mathbb{C}\colon w \mapsto \frac{2}{w}.$$

Then

$$g(f(z)) = \frac{2}{f(z)} = \frac{2}{z^2+1}.$$

A.2.3 Problem. Let $f\colon A \to B$ and $g\colon C \to D$ be functions with $f(A) \subseteq C$. Prove that

$$g \circ f = \{(x, z) \mid (x, y) \in f \implies (y, z) \in g\}.$$

The tool of composition allows us to invert certain functions.

A.2.4 Theorem. *Let A and B be sets and suppose that $f\colon A \to B$ is* **INJECTIVE** *or* **ONE-TO-ONE** *in the sense that if $f(x_1) = f(x_2)$ for x_1, $x_2 \in A$, then $x_1 = x_2$. Let*

$$g := \{(f(x), x) \mid x \in A\} = \{(y, x) \mid (x, y) \in f\}.$$

Then $g\colon f(A) \to A$ is a function and

$$g(f(x)) = x \text{ for all } x \in A \quad \text{and} \quad f(g(y)) = y \text{ for all } y \in f(A). \tag{A.2.1}$$

Moreover, g is the only function from $f(A)$ to A to satisfy (A.2.1). *We call g the* **INVERSE** *of f, and we write $f^{-1} := g$.*

A.2.5 Problem ($\star$). Prove Theorem A.2.4. [Hint: *to show that $g\colon f(A) \to A$ is a function, think carefully about what aspects of Definition 1.2.2 need to be checked. The verification of* (A.2.1) *is a direct computation. Finally, suppose that $h\colon f(A) \to A$ satisfies* (A.2.1) *when g is replaced by h. We need to show $h = g$. If $w \in f(A)$, write $w = f(x)$ for some $x \in A$; what, then, are $g(w)$ and $h(w)$?*]

A.2.6 Example. Example 1.2.6 (or common sense) tells us that the function

$$h\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto z^2$$

is not one-to-one, since $h(1) = h(-1)$. Consider, however, the restriction $f := h\big|_{i\mathbb{R}_+}$, where $i\mathbb{R}_+ = \{iy \mid y \geq 0\}$. By (the work in) Example 1.2.6, the image of f is $f(\mathbb{R}_+) = (-\infty, 0]$. Moreover, we can check that f is one-to-one: if $f(iy_1) = f(iy_2)$ with y_1, $y_2 \geq 0$, then $-y_1^2 = -y_2^2$, thus $y_1^2 = y_2^2$. Then $\sqrt{y_1^2} = \sqrt{y_2^2}$, so by Remark 1.1.9 $y_1 = y_2$.

B. Rigorous Constructions of the Complex Numbers

B.1. The real numbers $\mathbb{R}$.

This appendix corresponds to material covered on Monday, January 8, 2024.

For almost all of our day-to-day experiences in this course, it will be perfectly adequate to say what we did at the start: that a complex number is an expression of the form $x + iy$, where x and y are real numbers and the symbol i satisfies $i^2 = -1$. But since this is a course in mathematics, we should strive for a deeper understanding of the symbol $x + iy$ at least once (and then promptly forget it). And to succeed at that, it might be helpful to think first about what *real* numbers are and defer studying that all-important symbol i a bit longer.

We will denote the set of real numbers by the symbol $\mathbb{R}$. We will typically use the letters s, t, and τ to represent real numbers (the third letter, τ, is the Greek letter "tau"). When considering real numbers in forming complex numbers, we will often use letters like x, y, u, and v.

The real numbers have many properties. Many, many properties. These properties are likely so familiar to us that we use them instinctively and without thinking. The purpose of this brief foray into real numbers is to make us think explicitly about those properties. It turns out that from a handful of axiomatic properties, we can derive all other useful properties of real numbers. This is an approach that we will often deploy throughout the course: start with a few reasonable axioms (which themselves may be proved in another context, like a real analysis class) and then develop everything from those axioms. And very often those axioms will arise from some "dynamic" property of the objects that we are studying.

More informally, the following will usually be our philosophy.

B.1.1 Hypothesis. *What things do defines what things are.*

Already in our remarks that the equation $t^2 + 1 = 0$ has no real solutions we have used several properties of real numbers. First, the left side of the equation presumes that additive and multiplicative operations are defined—given t, we can compute t^2 and then add that to 1. Second, the equivalence of this equation to $t^2 = -1$ presumes that the operation of addition has an inverse (the additive inverse of 1 is -1). Third, the fact that we cannot solve $t^2 = -1$ for a real number x because $t^2 \geq 0$ for all real x but $-1 < 0$ presumes that $\mathbb{R}$ has an ordering structure that interacts with multiplication and addition.

We now list those axiomatic aspects of $\mathbb{R}$ in order of the frequency with which we will explicitly use them in this course. Note that nowhere do we say conclusively what real numbers *are*, but we go on at length about what they *do* (and do not do).

($\mathbb{R}$1) We can do arithmetic with real numbers. The real numbers $\mathbb{R}$ satisfy the **FIELD AXIOMS**: there exist operations $+$ and $\cdot$ defined on pairs of real numbers such that for all s, $t \in \mathbb{R}$, the symbols $s + t$ and $s \cdot t$ are also real numbers, and the operations $+$ and $\cdot$ behave exactly as we expect. (Set-theoretically, by "operations" we mean that $+$ and $\cdot$ are functions from $\mathbb{R}^2$ to $\mathbb{R}$. But we will not discuss functions for some time yet.) By "exactly as we expect," we mean the following.

1. *Commutativity.* $s + t = t + s$ and $s \cdot t = t \cdot s$ for all s, $t \in \mathbb{R}$.

2. *Associativity.* $(s + t) + \tau = s + (t + \tau)$ and $(s \cdot t) \cdot \tau = s \cdot (t \cdot \tau)$ for all s, t, $\tau \in \mathbb{R}$.

3. *Distributivity.* $s \cdot (t + \tau) = (s \cdot t) + (s \cdot \tau)$ for all s, t, $\tau \in \mathbb{R}$.

4. *Identities.* There exist numbers 0, $1 \in \mathbb{R}$ such that $0 \neq 1$, $t + 0 = t$ and $s \cdot 1 = s$ for all s, $t \in \mathbb{R}$. Moreover, $0 \neq 1$.

5. *Inverses.* For each $t \in \mathbb{R}$, there is $-t \in \mathbb{R}$ such that $t + (-t) = 0$, and for each $s \in \mathbb{R} \setminus \{0\}$ there is $1/s \in \mathbb{R}$ such that $s \cdot (1/s) = 1$.

Of course, we call $+$ "addition" and $\cdot$ "multiplication," while "subtraction" is $s-t := s+(-t)$ and "division" is $s/t := s \cdot (1/t)$. After this introduction, we will usually denote multiplication by juxtaposition, e.g., $st = s \cdot t$ and $s(t + \tau) = s \cdot (t + \tau)$.

From these five classes of axioms, we can prove *everything* else about how arithmetic works. For example, we could show that the additive identity 0 is unique, and then that additive inverses are unique, and then that $0 \times t = 0$ for all $t \in \mathbb{R}$ (an interesting connection between the *additive* identity and the operation of *multiplication*). And then we could show that $-t = (-1) \cdot t$; in the axioms above, $-t$ is just a symbol for the additive inverse of t (just as $1/t$ is a symbol for the multiplicative inverse of t). The axioms do not specify any connection between the *additive* inverse and the identity for *multiplication*, but we can *prove* that such a connection is there.

Because of the arithmetic structure of the real numbers, they contain several other important kinds of numbers. First, the **NATURAL NUMBERS** are real numbers; intuitively, these are the positive whole numbers 1, 2, 3, and so on. More formally, we declare 1 to be a natural number and then recursively define $t \in \mathbb{R}$ to be a natural number if $t = k + 1$ for some natural number k. We denote the set of all natural numbers by $\mathbb{N}$.

Then we define the **INTEGERS** $\mathbb{Z}$ to be the natural numbers together with their additive inverses and the identity for addition. That is,

$$\mathbb{Z} := \mathbb{N} \cup \{0\} \cup \{-k \mid k \in \mathbb{N}\}.$$

Finally, we define the rational numbers $\mathbb{Q}$ to be quotients of integers (with nonzero denominators). That is,

$$\mathbb{Q} := \left\{ p \cdot \left(\frac{1}{q}\right) \,\middle|\, p,\ q \in \mathbb{Z},\ q \neq 0 \right\}.$$

Because $\mathbb{R}$ is closed under addition, multiplication, additive inverses, and multiplicative inverses, we have $\mathbb{N} \subseteq \mathbb{R}$, $\mathbb{Z} \subseteq \mathbb{R}$, and $\mathbb{Q} \subseteq \mathbb{R}$.

B.1.2 Example. From the five axioms for arithmetic we can prove many other familiar properties. Here are several.

(i) The number 0 is unique in the sense that if $\omega \in \mathbb{R}$ satisfies $t + \omega = t$ for all $t \in \mathbb{R}$, then

$\omega = 0$. Here is why: we can put $t = 0$ to find $0 = 0 + \omega = \omega$ by definition of 0.

(ii) Inverses are unique. For addition, if $t + s = 0$, then s should equal $-t$. We check this by assuming $t + s = 0$, adding $-t$ to both sides, and then using the definition of 0, the associativity of addition, and the definition of $-t$:

$$-t = -t + 0 = -t + (t + s) = (-t + t) + s = 0 + s = s.$$

(iii) We want to say that $0 \cdot t = 0$ for all $t \in \mathbb{R}$, but so far we only know how 0 interacts with addition. So, we introduce addition into the quantity $0 \cdot t$:

$$0 \cdot t = (0 + 0) \cdot t = (0 \cdot t) + (0 \cdot t).$$

Then we add $-(0 \cdot t)$ to both sides to get

$$\begin{aligned} 0 &= (0 \cdot t) + [-(0 \cdot t)] \text{ by definition of the inverse for addition} \\ &= [(0 \cdot t) + (0 \cdot t)] + [-(0 \cdot t)] \text{ since } 0 \cdot t = (0 \cdot t) + (0 \cdot t) \\ &= (0 \cdot t) + [(0 \cdot t) + [-(0 \cdot t)]] \text{ since addition is associative} \\ &= (0 \cdot t) + 0 \text{ by definition of the inverse for addition, again} \\ &= (0 \cdot t) \text{ by definition of 0.} \end{aligned}$$

(iv) When we see $-t$, our instinct is probably to think of it as the product $-t = (-1) \cdot t$. This is true, but it needs justification, since all we know about $-t$ is that it satisfies $t + (-t) = 0$. That is, from the axioms we only know how $-t$ interacts with addition, not multiplication (just like with 0 above).

What does it mean to have $-t = (-1) \cdot t$? We think about what $-t$ should do, and we conclude that we want $t + [(-1) \cdot t] = 0$, since inverses are unique. We check this by using the definition of 1 and the distributive property to compute

$$t + [(-1) \cdot t] = (1 \cdot t) + [(-1) \cdot t] = [1 + (-1)] \cdot t = 0 \cdot t = 0.$$

(v) As a consequence of the work above, we now have some particular results relating the identities for addition and multiplication in the context of both operations:

$$1 + (-1) = 0, \qquad 1 \cdot (-1) = -1, \qquad \text{and} \qquad 0 \cdot 1 = 0.$$

($\mathbb{R}$2) We can compare and order real numbers. The real numbers are ORDERED in the sense that we can "compare" two real numbers and obtain natural insights into their relative "sizes." Specifically, there exists a set $\mathcal{P} \subseteq \mathbb{R}$ with the following properties.

1. *1 is positive.* $1 \in \mathcal{P}$. (And so $\mathcal{P} \neq \varnothing$.)

2. *Closure.* The set $\mathcal{P}$ is closed under addition and multiplication: if $s, t \in \mathcal{P}$, then $s + t \in \mathcal{P}$ and $s \cdot t \in \mathcal{P}$.

3. *Trichotomy.* For each $t \in \mathbb{R}$, one, and only one, of the following holds: either $t = 0$ or

$t \in \mathcal{P}$ or $-t \in \mathcal{P}$.

Of course we call $\mathcal{P}$ the **POSITIVE NUMBERS**. We introduce an "ordering" on $\mathbb{R}$ by saying that for s, $t \in \mathbb{R}$, we have $s < t$ if $t - s \in \mathcal{P}$ and $s \le t$ if either $s < t$ or $s = t$. (And sometimes we write $t > s$ if $s < t$ and $t \ge s$ if $s \le t$.)

From the three axioms for $\mathcal{P}$ and the field axioms, we can obtain familiar results about the interaction of arithmetic and inequalities. For example, if $0 < s$ and $t < \tau$, then $s \cdot t < s \cdot \tau$.

B.1.3 Example. The ordering on $\mathbb{R}$ induced by $\mathcal{P}$ behaves exactly as we expect.

(i) For each $t \in \mathbb{R}$, we have $t \in \mathcal{P}$ if and only if $0 < t$. This is just the definition of $<$ with $s = 0$.

(ii) For all s, $t \in \mathbb{R}$, one, and only one, of the following holds: either $s = t$, or $s < t$, or $t < s$.

(iii) Multiplication by positive numbers preserves inequalities: we expect that if $0 < s$ and $t < \tau$, then $s \cdot t < s \cdot \tau$. Here is why. First, $s \cdot t < s \cdot \tau$ if and only if $0 < s \cdot \tau - s \cdot t$. Distribution lets us rewrite this second inequality as $0 < s \cdot (\tau - t)$. Since $0 < s$, we have $s \in \mathcal{P}$, and since $t < \tau$, we have $\tau - t \in \mathcal{P}$, and so $s \cdot (\tau - t) \in \mathcal{P}$.

($\mathbb{R}$3) We *cannot* find gaps among the real numbers. The natural numbers (and integers) definitely have gaps: there is no $k \in \mathbb{Z}$ such that $1 < k < 2$. The rational numbers do not have gaps in this sense (if p, $q \in \mathbb{Q}$ with $p < q$, then $r := (p+q)/2$ satisfies $p < r < q$), but the rationals do have gaps in the following less obvious sense. There is no rational number $r_\star$ such that $r_\star^2 = 2$, but given any $\epsilon > 0$, it is possible to find a rational number r such that $-\epsilon < r^2 - 2 < \epsilon$. More informally, we can approximate $\sqrt{2}$ as closely as we like with rationals, but $\sqrt{2}$ is irrational.

This is not the case with real numbers: they are **COMPLETE**. This is perhaps the most technical and least intuitive property of $\mathbb{R}$ to state; while it has been present throughout our lives, it is not as accessibly articulated as the field axioms or the ordering axioms. Here is one way to describe completeness.

1. *Any set of real numbers that is nonempty and bounded above has a least upper bound.* Suppose $A \subseteq \mathbb{R}$ is nonempty and **BOUNDED ABOVE** in the sense that there is $M \in \mathbb{R}$ such that $t \le M$ for all $t \in A$. (We call M an **UPPER BOUND** for A; an upper bound is not unique, for if M is an upper bound for A, then so is $M + 1$.) Then A has a **LEAST UPPER BOUND** in the sense that there exists $m \in \mathbb{R}$ such that (1) $t \le m$ for all $t \in A$ and (2) if $s < m$, then there is $t \in A$ such that $s < t \le m$. The first property of m ensures that m is an upper bound for A; the second property ensures that no s smaller than m can be an upper bound for A, and so m is the *least* upper bound for A.

A course in real analysis absolutely, completely, utterly hinges on completeness, but we will not discuss this property further in our course. Many results that we state but do not prove, such as convergence tests for series and the existence of definite integrals, do boil down to an invocation of completeness somewhere in their proofs.

We are proceeding under the assumption that there *exists* a set $\mathbb{R}$ with the properties ($\mathbb{R}$1), ($\mathbb{R}$2), and ($\mathbb{R}$3). It is possible to *prove* that there *is* such a set.

We will not do this. Instead, we present two useful consequences of all the axioms of $\mathbb{R}$. One is the fundamental motivation for the course, and the other is a remarkably versatile little inequality.

B.1.4 Theorem. *The additive inverse of the multiplicative identity has no square root in $\mathbb{R}$. That is, there is no $t \in \mathbb{R}$ such that $t \cdot t = -1$.*

Proof. We assume that there is $t \in \mathbb{R}$ such that $t \cdot t = -1$ and then we use trichotomy to derive a contradiction.

Case 1. $t = 0$. Then $t \cdot t = 0 \cdot 0 = 0$.

Case 2. $t > 0$. Then $t \cdot t > 0$, but $-1 < 0$.

Case 3. $t < 0$. Then $t \cdot t > 0$, but, again, $-1 < 0$.

We used four facts about real numbers in the cases above: (1) $0 \cdot 0 = 0$, (2) $t \cdot t > 0$ when $t > 0$, (3) $t \cdot t > 0$ when $t < 0$, and (4) $-1 < 0$. Each of those facts could be proved from the field axioms in ($\mathbb{R}$1) and the order axioms in ($\mathbb{R}$2), possibly with the help of some auxiliary results along the way. ■

B.1.5 Problem ($\star$). Let $x \geq 0$ be a real number with the property that if $\epsilon > 0$, then $x \leq \epsilon$. Prove that $x = 0$. [Hint: *if $x > 0$, what happens when we take $\epsilon = x/2$?*]

B.2. The complex numbers $\mathbb{C}$.

This appendix corresponds to material covered on Monday, January 8, 2024.

Our goal is now to give an airtight definition of the symbol $x + iy$ so that it enjoys all the arithmetic properties of a field, i.e., the five axioms in ($\mathbb{R}$1). The key insight comes from our geometric explorations that identified complex numbers $x + iy$ with ordered pairs (x, y) of real numbers x and y.

B.2.1 Definition. *Let A be a set and x, $y \in A$. The* **ORDERED PAIR** (x, y) *is the set*

$$(x, y) := \{\{x\}, \{x, y\}\}.$$

No one *ever* uses this definition of an ordered pair in practice, but its redeeming grace is the following theorem.

B.2.2 Theorem. *Let A be a set and x, y, u, $v \in A$. Then $(x, y) = (u, v)$ if and only if $x = u$ and $y = v$.*

B.2.3 Problem (!). Do *not* prove Theorem B.2.2—this is a course in complex analysis, not set theory. Instead, think about what needs to be done to prove it. [Hint: *what does "=" mean here?*]

We are going to define complex numbers as ordered pairs of real numbers. But wait! We want real numbers to be complex numbers. However, a real number cannot be an ordered pair of real numbers. . .right?

This suggests that we *really* need a new interpretation of *real* numbers, too.

B.2.4 Hypothesis. *There exists a set* $\widehat{\mathbb{R}}$*, which we call the* **TEMPORARY REAL NUMBERS***, on which there are defined two operations,* $+$ *and* $\cdot$*, which satisfy the field axioms (*$\mathbb{R}$*1). Moreover, there exists a set* $\widehat{\mathcal{P}} \subseteq \mathbb{R}$*, called the* **TEMPORARY POSITIVE NUMBERS***, that satisfies the order axioms (*$\mathbb{R}$*2) and such that* $\widehat{\mathbb{R}}$ *has the least upper bound property (*$\mathbb{R}$*3) with respect to the ordering on* $\widehat{\mathbb{R}}$ *defined by* $s < t$ *if* $t - s \in \widehat{\mathcal{P}}$*.*

B.2.5 Definition. (i) *A* **COMPLEX NUMBER** *is an ordered pair* (x, y)*, where* $x,\ y \in \widehat{\mathbb{R}}$*. We denote the set of all complex number numbers by*

$$\mathbb{C} = \left\{ (x, y) \;\middle|\; x,\ y \in \widehat{\mathbb{R}} \right\}.$$

(ii) *A* **REAL NUMBER** *is an ordered pair* $(x, 0)$*, where* $x \in \widehat{\mathbb{R}}$*. We denote the set of all real numbers by*

$$\mathbb{R} = \left\{ (x, 0) \;\middle|\; x \in \widehat{\mathbb{R}} \right\}.$$

The immediate upshot of this definition of $\mathbb{R}$ is that every real number is now a complex number. Momentarily we will create arithmetic in $\mathbb{C}$, and then we will show that $\mathbb{R}$ satisfies the field axioms ($\mathbb{R}$1), the order axioms ($\mathbb{R}$2), and the completeness axiom ($\mathbb{R}$3) with respect to those arithmetical and ordering operations. Invoking Hypothesis B.1.1, we will discard the temporary real numbers $\widehat{\mathbb{R}}$ and only work with $\mathbb{R}$ for the rest of the course, since $\mathbb{R}$ does what real numbers should do.

To define arithmetic in $\mathbb{C}$, it is helpful to remember how we formally added and multiplied in the past:

$$(x+iy)+(u+iv) = (x+u)+i(y+v) \quad \text{and} \quad (x+iy)(u+iv) = \cdots = (xu-yv)+i(xv+yu).$$

Based on these formulas, we define **COMPLEX ADDITION** by

$$(x, y) \oplus (u, v) := (x + u, y + v) \tag{B.2.1}$$

and **COMPLEX MULTIPLICATION** by

$$(x, y) \odot (u, v) := \big((x \cdot u) - (y \cdot v), (x \cdot v) + (y \cdot u)\big). \tag{B.2.2}$$

We are going to show that we can represent elements of $\mathbb{C}$ in the form $x + iy$ that we know and like. Specifically, direct calculations reveal

$$(0, 1) \odot (0, 1) = (-1, 0) \quad \text{and} \quad (x, y) = (x, 0) \oplus \big[(0, 1) \odot (y, 0)\big].$$

B.2.6 Problem (+). (Presumes knowledge of abstract algebra.) Prove that $\widehat{\mathbb{R}}$ and $\mathbb{R}$ are isomorphic as fields, where arithmetic in $\widehat{\mathbb{R}}$ uses $+$ and $\cdot$, while arithmetic in $\mathbb{R}$ uses $\oplus$ and $\odot$.

B.2.7 Example. The ordered pair $(1,0)$ is the identity for multiplication. We check

$$(x,y)\odot(1,0) = \big((x\cdot 1)-(y\cdot 0),(x\cdot 0)+(y\cdot 1)\big) = (x-0,0+y) = (x,y).$$

The operations $\oplus$ and $\odot$ satisfy the field axioms ($\mathbb{R}$1), and it is possible to define a subset of "positive" numbers in $\mathbb{R}$ (but not, as we shall see, in $\mathbb{C}$) that meet the ordering axioms ($\mathbb{R}$2). The set $\mathbb{R}$ also satisfies the least upper bound property with respect to the ordering induced by these positive numbers. And so $\mathbb{R}$ does what any good set of real numbers should do, and arithmetic works in the broader set $\mathbb{C}$ in the way that we fundamentally expect.

B.2.8 Problem (+). **(i)** Check that the rest of the field axioms from ($\mathbb{R}$1) hold for $\oplus$ and $\odot$, assuming that the field axioms hold for $+$ and $\cdot$ on $\widehat{\mathbb{R}}$. [Hint: *for the multiplicative inverses, think about what* (1.1.7) *says in the language of ordered pairs.*]

(ii) Suppose that $\widehat{\mathcal{P}} \subseteq \widehat{\mathbb{R}}$ satisfies the order axioms from ($\mathbb{R}$2). Put

$$\mathcal{P} := \left\{(x,0)\in\mathbb{R} \;\middle|\; x\in\widehat{\mathcal{P}}\right\}.$$

Show that $\mathcal{P}$ satisfies also the order axioms ($\mathbb{R}$2).

(iii) Write $(s,0)\prec(t,0)$ if $(t-s,0)\in\mathcal{P}$. Show that $\mathbb{R}$ has the least upper bound property ($\mathbb{R}$3) with respect to $\prec$, assuming that $\widehat{\mathbb{R}}$ has the least upper bound property with respect to $<$.

However, while we can compare any two elements of $\mathbb{R}$, we cannot compare any two elements of $\mathbb{C}$ and expect arithmetic to respect this comparison as it does on $\mathbb{R}$.

B.2.9 Problem ($\star$). Show that $\mathbb{C}$ cannot be ordered in the sense that there is no subset $\mathcal{P}_{\mathbb{C}}$ of $\mathbb{C}$ satisfying the order axioms of ($\mathbb{R}$2). Proceed by contradiction: assume there is $\mathcal{P}_{\mathbb{C}} \subseteq \mathbb{C}$ with the following three properties.

(i) $(1,0)\in\mathcal{P}_{\mathbb{C}}$.

(ii) If z, $w\in\mathcal{P}_{\mathbb{C}}$, then $z\oplus w\in\mathcal{P}_{\mathbb{C}}$ and $z\odot w\in\mathcal{P}_{\mathbb{C}}$.

(iii) If $(x,y)\in\mathbb{C}$, then one, and only one, of the following holds: either $(x,y)=(0,0)$, or $(x,y)\in\mathcal{P}_{\mathbb{C}}$, or $(-x,-y)\in\mathcal{P}_{\mathbb{C}}$.

Since $(0,1)\neq(0,0)$, it must be the case that either $(0,1)\in\mathcal{P}_{\mathbb{C}}$ or $(0,-1)\in\mathcal{P}_{\mathbb{C}}$. What contradictions result in either case? [Hint: *what do you know about* $(0,1)\odot(0,1)$ *and*

$(0,-1) \odot (0,-1)$?]

Although we cannot order $\mathbb{C}$ as we do $\mathbb{R}$, we do get what we really wanted in the first place: in contrast to Theorem B.1.4, the additive inverse of the multiplicative identity has a square root in $\mathbb{C}$.

B.2.10 Theorem. **(i)** $(0,1) \odot (0,1) = (-1,0)$.

(ii) $(0,y) = (0,1) \odot (y,0)$ *for all* $y \in \widehat{\mathbb{R}}$.

Proof. These are direct calculations.

(i) $(0,1) \odot (0,1) = \big((0 \cdot 1) - (1 \cdot 1), (0 \cdot 1) + (1 \cdot 0)\big) = (0-1, 0+0) = (-1,0)$.

(ii) $(0,1) \odot (y,0) = \big((0 \cdot y) - (1 \cdot 0), (0 \cdot 0) + (1 \cdot y)\big) = (0-1, 0+y) = (0,y)$. ■

Since the field axioms for $\oplus$ and $\odot$ fall out as they should, we have

$$(x,y) = (x,0) \oplus (0,y) = (x,0) \oplus \big[(0,1) \odot (y,0)\big]. \tag{B.2.3}$$

This is exactly the representation of a complex number that we expect: $x+iy$, if we identify x and $(x,0)$, y and $(y,0)$, and i and $(0,1)$. Indeed, working constantly in terms of ordered pairs and the baroque notation $\oplus$ and $\odot$ is, at best, wearying. Representing complex numbers as the symbols $x+iy$i much slicker, and the intuitive notions of arithmetic on $x+iy$ are all that we ever use in practice.

Going forward, we will use ordinary letters once again for real numbers; thus, a real number x is really an ordered pair $x = (\xi, 0)$, where ξ is a temporary real number. Likewise, we will use, as before, single letters such as z and w for complex numbers (by the way, a sentence like "$z > 0$" will always imply that z is a real number). We will write $+$ instead of $\oplus$ and $\cdot$ instead of $\odot$, but we can always refer to (B.2.1) and (B.2.2) if we need to comfort ourselves with arithmetic on ordered pairs. We put

$$i := (0,1) \tag{B.2.4}$$

to end where we started.

B.2.11 Theorem. $\mathbb{C} = \{x + iy \mid x,\ y \in \mathbb{R}\}$.

Proof. Use the definition of i and the identity (B.2.3). ■

B.2.12 Problem (!). Reread the two paragraphs preceding Theorem B.2.11 until you get a headache. Describe the intensity of your headache and what you did to get rid of it.

B.2.13 Problem (+). (Presumes knowledge of linear and abstract algebra.) The approach of this section is certainly not the only way to construct $\mathbb{C}$. We could also view $\mathbb{C}$ as a special set of 2×2 matrices. (We will not burden ourselves by giving a formal definition of a matrix.)

(i) Check that $\odot$ is really matrix-vector multiplication:

$$(x, y) \odot (u, v) = (xu - yv, xv + yu) = \begin{bmatrix} xu - yv \\ xv + yu \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{bmatrix} u & -v \\ v & u \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

(ii) Now put

$$\mathbb{R}_{\mathbb{C}}^{2\times 2} := \left\{ \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \,\middle|\, x,\ y \in \widehat{\mathbb{R}} \right\}$$

and check that $\mathbb{C}$ (as defined in Definition B.2.5) and $\mathbb{R}_{\mathbb{C}}^{2\times 2}$ are isomorphic as fields, where addition and multiplication in $\mathbb{C}$ are $\oplus$ and $\odot$, while addition and multiplication in $\mathbb{R}_{\mathbb{C}}^{2\times 2}$ are the usual addition and multiplication for 2×2 matrices.

(iii) To what subset of $\mathbb{R}_{\mathbb{C}}^{2\times 2}$ is $\mathbb{R}$ (as defined in Definition B.2.5) isomorphic as a field?

C. Assorted Proofs

C.1. The proof of part (ii) of Theorem 2.6.14.

We first need the tool of uniform continuity, as developed in real analysis or, more generally, metric space topology.

C.1.1 Lemma (Uniform continuity). *Suppose that $f\colon \mathcal{D} \subseteq \mathbb{C} \to \mathbb{C}$ is continuous, where $\mathcal{D}$ is a set of the form*

$$\mathcal{D} = \{x + iy \mid x_0 \le x \le x_1,\ y_0 \le y \le y_1\} \qquad \textit{or} \qquad \mathcal{D} = \overline{\mathcal{B}}(x_0 + iy_0; r_0). \tag{C.1.1}$$

Then for all $\epsilon > 0$, there exists $\delta > 0$ such that if w, $z \in \mathcal{D}$ with $|w - z| < \delta$, then

$$|f(w) - f(z)| < \epsilon.$$

We will not prove this lemma, but we contrast its "uniformity" with "ordinary continuity," which would say that for all $\epsilon > 0$ and $z \in \mathcal{D}$, there is $\delta > 0$ such that if $|w - z| < \delta$, then $|f(w) - f(z)| < \epsilon$. In "ordinary" continuity, the threshold δ can depend on both ϵ and z; in "uniform" continuity, the same δ works for the whole set $\mathcal{D}$. The key is that the two varieties of $\mathcal{D}$ in (C.1.1) are closed and bounded sets (and so Lemma C.1.1 turns out to hold for much more general $\mathcal{D}$ than these varieties, though we will not need them).

Now we restate and prove the theorem.

2.6.14 Theorem. *Suppose that $\mathcal{D} \subseteq \mathbb{C}$ is open and let $f\colon \mathcal{D} \to \mathbb{C}$ be a function. Write $f(x + iy) = u(x, y) + iv(x, y)$, where we think of u and v as being defined on the set $\widetilde{\mathcal{D}} := \big\{(x, y) \in \mathbb{R}^2 \bigm| x + iy \in \mathcal{D}\big\}$. In the following we write u_x, u_y, v_x, and v_y for the partial derivatives of u and v with respect to x and y.*

(ii) *Let $x + iy \in \mathcal{D}$ and let $r > 0$ be such that $\mathcal{B}(x + iy; r) \subseteq \mathcal{D}$. Suppose that the four partial derivatives u_x, u_y, v_x, and v_y exist and are continuous on $\mathcal{B}(x + iy; r)$. Moreover, suppose that the partials satisfy the Cauchy–Riemann equations*

$$\begin{cases} u_x(x, y) = v_y(x, y) \\ u_y(x, y) = -v_x(x, y). \end{cases}$$

at $x + iy$. Then f is differentiable at $x + iy$ and

$$f'(x + iy) = u_x(x, y) + iv_x(x, y) = v_y(x, y) - iu_y(x, y).$$

We first emphasize that the Cauchy–Riemann equations only need to hold at this particular point $x + iy$, not on all of $\mathcal{D}$. This proof uses four ideas. The first is to write out the limit that f' needs to satisfy. The second is to rewrite this limit using the Cauchy–Riemann equations so that the real part of the limit only involves u and the imaginary part only involves v. The third is to rewrite those real and imaginary parts by "adding zero" in a clever way to expose the fundamental theorem of calculus. And the fourth is to use the FTC to rewrite certain differences as integrals and then estimate those integrals using the triangle inequality and the continuity of the partials.

We want to show that

$$\lim_{h+ik\to 0} \frac{f((x+iy)+(h+ik)-f(x+iy)}{h+ik} = u_x(x,y)+iv_x(x,y),$$

equivalently,

$$\lim_{h+ik\to 0} \frac{f((x+iy)+(h+ik)-f(x+iy)-(h+ik)\big[u_x(x,y)+iv_x(x,y)\big]}{h+ik} = 0.$$

So, for all $\epsilon > 0$, we want to find $\delta > 0$ such that if $0 < |h+ik| < \delta$, then

$$\left|\frac{f((x+iy)+(h+ik)-f(x+iy)-(h+ik)\big[u_x(x,y)+iv_x(x,y)\big]}{h+ik}\right| < \epsilon. \qquad \text{(C.1.2)}$$

We have

$$\begin{aligned} f((x+iy)+(h+ik)-f(x+iy) &= \big[u(x+h,y+k)+iv(x+h,y+k)\big]-\big[u(x,y)+iv(x,y)\big] \\ &= \big[u(x+h,y+k)-u(x,y)\big]+i\big[v(x+h,y+k)-v(x,y)\big]. \end{aligned}$$

We compute

$$(h+ik)\big[u_x(x,y)+iv_x(x,y)\big] = \big[hu_x(x,y)-kv_x(x,y)\big]+i\big[hv_x(x,y)+ku_x(x,y)\big]$$

In the real part of this expression, use the Cauchy–Riemann equations to rewrite

$$-kv_x(x,y) = ku_y(x,y),$$

and in the imaginary part,

$$ku_x(x,y) = ku_y(x,y).$$

Then

$$(h+ik)\big[u_x(x,y)+iv_x(x,y)\big] = \big[hu_x(x,y)+ku_y(x,y)\big]+i\big[hv_x(x,y)+ku_y(x,y)\big].$$

This allows us to conclude

$$\begin{aligned} &f((x+iy)+(h+ik)-f(x+iy)-(h+ik)\big[u_x(x,y)+iv_x(x,y)\big] \\ &\quad = \Big(\big[u(x+h,y+k)-u(x,y)\big]-\big[hu_x(x,y)+ku_y(x,y)\big]\Big) \\ &\qquad + i\Big(\big[v(x+h,y+k)-v(x,y)\big]-\big[hv_x(x,y)+ku_y(x,y)\big]\Big). \qquad \text{(C.1.3)} \end{aligned}$$

We have now rewritten the real part of the limit so that it only involves u and the imaginary part so that it only involves v. Moreover, the real and imaginary parts are effectively the same: just replace u with v. So, we will only estimate the real part.

In the real part, add and subtract $u(x,y)$, $u(x+h,y)$, $u(x,y+k)$ so that

$$\begin{aligned}\big[u(x+h,y+k)-u(x,y)\big]-\big[hu_x(x,y)+ku_y(x,y)\big] &= u(x+h,y+k)-u(x,y+k)\\ &\quad+u(x,y)-u(x+h,y)\\ &\quad+u(x+h,y)-u(x,y)-hu_x(x,y)\\ &\quad+u(x,y+k)-u(x,y)-ku_y(x,y).\end{aligned}$$

Then use the fundamental theorem of calculus (or Example 3.2.24) to rewrite the third line on the right above as

$$\begin{aligned}u(x+h,y)-u(x,y)-hu_x(x,y) &= h\int_0^1 u_x(x+h\tau,y)\;d\tau - hu_x(x,y)\int_0^1 1\tau\\ &= h\int_0^1 \big[u_x(x+h\tau,y)-u_x(x,y)\big]\;d\tau\end{aligned}$$

and likewise the fourth line as

$$u(x,y+k)-u(x,y)-ku_y(x,y) = k\int_0^1 \big[u_y(x,y+k\tau)-u_y(x,y)\big]\;d\tau.$$

Then write the first line as

$$u(x+h,y+k)-u(x,y+k) = h\int_0^1 u_x(x+h\tau,y+k)\;d\tau$$

and the second line as

$$u(x,y)-u(x+h,y) = -h\int_0^1 u_x(x+h\tau,y)\;d\tau,$$

so that together the first and second lines equal

$$u(x+h,y+k)-u(x,y+k)+u(x,y)-u(x+h,y) = h\int_0^1 \big[u_x(x+h\tau,y+k)-u_x(x+h\tau,y)\big]\;d\tau.$$

All together, we obtain

$$\frac{\big[u(x+h,y+k)-u(x,y)\big]-\big[hu_x(x,y)+ku_y(x,y)\big]}{h+ik} = I + II + III,$$

where

$$I = \frac{h}{h+ik}\int_0^1 \big[u_x(x+h\tau,y+k)-u_x(x+h\tau,y)\big]\;d\tau,$$

$$II = \frac{h}{h+ik}\int_0^1 \big[u_x(x+h\tau,y)-u_x(x,y)\big]\;d\tau,$$

and

$$III = \frac{k}{h+ik}\int_0^1 \big[u_y(x,y+k\tau)-u_y(x,y)\big]\;d\tau.$$

Estimating the difference quotient in u above therefore amounts to using the (real-valued) triangle inequality on these three integrals and exploiting the *uniform* continuity of the partial derivatives. Recall that the four partial derivatives are continuous on the ball $\mathcal{B}(x+iy;r) \subseteq \mathcal{D}$, and so they are continuous on the smaller *closed* ball $\overline{\mathcal{B}}(x+iy;r/2) \subseteq \mathcal{D}$, too. Uniform continuity then tells us that given $\epsilon > 0$, there is $\delta > 0$ such that if $\xi+i\eta$, $\grave{\xi}+i\grave{\eta} \in \overline{\mathcal{B}}(x+iy;r/2)$ and $\sqrt{(\xi-\grave{\xi})^2+(\eta-\grave{\eta})^2} < \delta$, then

$$|u_x(\xi,\eta) - u_x(\grave{\xi},\grave{\eta})| < \frac{\epsilon}{6} \quad \text{and} \quad |u_y(\xi,\eta) - u_y(\grave{\xi},\grave{\eta})| < \frac{\epsilon}{6}.$$

Suppose that $0 < |h+ik| < \delta$; this forces $|h| \le |h+ik| < \delta$ and $|k| \le |h+ik| < \delta$. Then

$$\sqrt{(x-x)^2+(y+k\tau-y)^2} = \sqrt{k^2\tau^2} = |k|\tau \le |k| < \delta,$$
$$\sqrt{(x+h\tau-(x+h\tau))^2-(y+k-y)^2} = \sqrt{k^2} = |k| < \delta,$$

and

$$\sqrt{(x-x)^2+(y+k\tau-y)^2} = \sqrt{k^2\tau^2} = |k|\tau < \delta.$$

Thus if $0 \le \tau \le 1$, we have

$$\big|u_x(x+h\tau,y+k) - u_x(x+h\tau,y)\big| < \frac{\epsilon}{6}, \qquad \big|u_x(x+h\tau,y) - u_x(x,y)\big| < \frac{\epsilon}{6},$$
$$\text{and} \quad \big|u_y(x,y+k\tau) - u_y(x,y)\big| < \frac{\epsilon}{6}.$$

Furthermore, since $|h| \le |h+ik|$ and $|k| \le |h+ik|$, we have

$$\frac{|h|}{|h+ik|} \le 1 \quad \text{and} \quad \frac{|k|}{|h+ik|} \le 1.$$

The triangle inequality and the estimates above therefore provide

$$|I| + |II| + |III| < \frac{\epsilon}{6} + \frac{\epsilon}{6} + \frac{\epsilon}{6},$$

and so we see that given $\epsilon > 0$, there is $\delta > 0$ such that if $0 < |h+ik| < \delta$, then

$$IV := \left|\frac{\big[u(x+h,y+k) - u(x,y)\big] - \big[hu_x(x,y) + ku_y(x,y)\big]}{h+ik}\right| < \frac{\epsilon}{2}.$$

Exactly the same arguments, using the same δ, show

$$V := \left|\frac{\big[v(x+h,y+k) - v(x,y)\big] - \big[hu_x(x,y) + ku_y(x,y)\big]}{h+ik}\right| < \frac{\epsilon}{2}.$$

By (C.1.3) and the triangle inequality, we have

$$\left|\frac{f((x+iy)+(h+ik) - f(x+iy) - (h+ik)\big[u_x(x,y) + iv_x(x,y)\big]}{h+ik}\right| \le IV + V < \epsilon,$$

and this is the desired estimate (C.1.2).

C.2. The proof of Theorem 3.2.29.

3.2.29 Theorem. *Arc length is well-defined in the sense that if a path γ can be expressed as compositions of smooth paths via*

$$\gamma = \oplus_{k=1}^{n}\gamma_k \quad \text{and} \quad \gamma = \oplus_{j=1}^{m}\mu_j,$$

then

$$\sum_{k=1}^{n}\ell(\gamma_k) = \sum_{j=1}^{m}\ell(\mu_j). \tag{C.2.1}$$

Proof. Since γ is a path, by Problem 3.1.24 there are $t_0, \ldots, t_N \in [a,b]$ such that $t_0 = a$, $t_N = b$, $t_{s-1} < t_s$, and $\nu_s := \gamma\big|_{[t_{s-1},t_s]}$ is continuously differentiable.

For $j = 1, \ldots, m$, let $\mu_j\colon [a_j, b_j] \subseteq \mathbb{R} \to \mathcal{D}$ be smooth paths such that $\gamma = \oplus_{j=1}^{m}\mu_j$. Since $\gamma = \oplus_{j=1}^{m}\mu_j$, Problem 3.1.15 tells us that $a = a_1$ and $b = b_1 + \sum_{j=2}^{m}(b_j - a_j)$. We show that

$$\sum_{j=1}^{m}\ell(\mu_j) = \sum_{s=1}^{N}\ell(\nu_s). \tag{C.2.2}$$

Replacing the μ_j with γ_k, we have the desired equality (C.2.1).

To prove (C.2.2), we use Problem 3.1.15 and put

$$\tau_j := \begin{cases} a = a_1, \ j = 0 \\ b_1, \ j = 1, \\ \tau_{j-1} + b_j - a_j, \ j \geq 2 \end{cases} \qquad \text{and} \qquad \mathbb{1}_{[\tau_{j-1},\tau_j]}(t) := \begin{cases} 1, \ \tau_{j-1} \leq t \leq \tau_j \\ 0, \ t < \tau_{j-1} \text{ or } t > \tau_j \end{cases}$$

to have

$$\gamma(t) = \big(\oplus_{j=1}^{m}\mu_j\big)(t) = \sum_{j=1}^{m}\mathbb{1}_{[\tau_{j-1},\tau_j]}(t)\mu_j(t + a_j - \tau_{j-1}),$$

and so

$$\gamma\big|_{[\tau_{j-1},\tau_j]}(t) = \mu_j(t + a_j - \tau_{j-1}).$$

In particular, γ is continuously differentiable on $[a,b] \setminus \{\tau_j\}_{j=0}^{m}$ and

$$\gamma\big|_{[\tau_{j-1},\tau_j]}{}'(t) = \mu_j'(t + a_j - \tau_{j-1}).$$

with the derivatives at $t = \tau_{j-1}$ and $t = \tau_j$ interpreted as in Problem 2.5.22. Now we consider cases on N.

Case 1: $N = 1$. Then γ is in fact continuously differentiable on all of $[a,b]$, and so

$$\sum_{s=1}^{N}\ell(\nu_s) = \sum_{s=1}^{N}\int_{t_{s-1}}^{t_s}|\nu_s'(t)|\,dt = \sum_{s=1}^{N}\int_{t_{s-1}}^{t_s}|\gamma'(t)|\,dt = \int_a^b|\gamma'(t)|\,dt = \sum_{j=1}^{m}\int_{\tau_{j-1}}^{\tau_j}|\gamma'(t)|\,dt. \tag{C.2.3}$$

For each $j \geq 1$, we have

$$\int_{\tau_{j-1}}^{\tau_j} |\gamma'(t)|\ dt = \int_{\tau_{j-1}}^{\tau_j} |\mu_j'(t + a_j - \tau_{j-1})|\ dt = \int_{a_j}^{b_j} |\mu_j'(u)|\ du = \ell(\mu_j) \tag{C.2.4}$$

via the substitution $u = t + a_j - \tau_{j-1}$ and the identities

$$\tau_{j-1} + a_j - \tau_{j-1} = a_j \quad \text{and} \quad \tau_j + a_j - \tau_{j-1} = \begin{cases} b_1 + a_1 - a_1,\ j = 1 \\ (\tau_{j-1} + b_j - a_j) + a_j - \tau_{j-1},\ j \geq 2 \end{cases} = b_j.$$

Combining (C.2.3) and (C.2.4) gives

$$\sum_{s=1}^{N} \ell(\nu_s) = \sum_{j=1}^{m} \ell(\mu_j),$$

as desired.

Case 2: $N \geq 2$. We may assume that γ is *not* continuously differentiable at any of the t_s with $1 \leq s < N$.

Here is why. First, if γ is continuously differentiable at every t_s, then γ is continuously differentiable on $[a, b]$, and we are back in the previous case. Second, if γ is continuously differentiable at some t_s, then γ is continuously differentiable on $[t_{s-1}, t_{s+1}]$, and so by relabeling and renumbering the t_s at the start of the proof, we may ignore this particular t_s.

Then we have $\{t_s\}_{s=0}^{N} \subseteq \{\tau_j\}_{j=0}^{m}$. Here is why. Certainly $t_0 = a = \tau_0$ and $t_N = b = \tau_m$. For $1 \leq s \leq N-1$, we are assuming that γ is not continuously differentiable at t_s, so if $t_s \neq \tau_j$ for some s and all j, then $t_s \in (\tau_{j_*-1}, \tau_{j_*})$ for some j_*. But then γ would be continuously differentiable at t_s. So, we can write $t_s = \tau_{j_s}$ for some j_s with $0 \leq j_s \leq m$; note that $j_0 = 0$, $j_N = m$, and $j_{s-1} < j_s$ for $1 \leq N$.

Then

$$\sum_{s=1}^{N} \ell(\nu_s) = \sum_{s=1}^{N} \int_{t_{s-1}}^{t_s} |\nu_s'(t)|\ dt = \sum_{s=1}^{N} \int_{\tau_{j_{s-1}}}^{\tau_{j_s}} |\nu_s'(t)|\ dt.$$

Fixing s, we have

$$\int_{\tau_{j_{s-1}}}^{\tau_{j_s}} |\nu_s'(t)|\ dt = \sum_{r=j_{s-1}+1}^{j_s} \int_{\tau_{r-1}}^{\tau_r} |\nu_s'(t)|\ dt.$$

If t satisfies

$$t_{s-1} = \tau_{j_{s-1}} \leq \tau_{r-1} \leq t \leq \tau_r \leq \tau_{j_s} = t_s,$$

then

$$\nu_s(t) = \nu\big|_{[t_{s-1}, t_s]}(t) = \nu\big|_{[\tau_{r-1}, \tau_r]}(t) = \mu_r(t + a_r - \tau_{r-1}),$$

and so

$$\int_{\tau_{r-1}}^{\tau_r} |\nu_s'(t)|\ dt = \int_{\tau_{r-1}}^{\tau_r} |\mu_r'(t + a_r - \tau_{r-1})|\ dt = \ell(\mu_r)$$

by the same substitution as in (C.2.4). Thus

$$\sum_{r=j_{s-1}+1}^{j_s} \int_{\tau_{r-1}}^{\tau_r} |\nu_s'(t)|\ dt = \sum_{r=j_{s-1}+1}^{j_s} \ell(\mu_r)$$

and so

$$\sum_{s=1}^{N} \ell(\nu_s) = \sum_{s=1}^{N} \sum_{r=j_{k-1}+1}^{j_k} \ell(\mu_r) = \sum_{j=1}^{m} \ell(\mu_j).$$

■

C.3. The proof of Theorem 3.5.8.

3.5.8 Theorem. *Suppose that $I \subseteq \mathbb{R}$ is an interval and $a,\ b \in \mathbb{R}$ with $a \le b$. Put $\mathcal{R} = \big\{(t,s) \in \mathbb{R}^2 \ \big|\ t \in I,\ a \le t \le b\big\}$. Let $f\colon \mathcal{R} \to \mathbb{C}\colon (t,s) \mapsto f(t,s)$ be a continuous function such that f_t exists and is continuous on $\mathcal{R}$. Then the map*

$$\mathcal{I}\colon I \to \mathbb{C}\colon t \mapsto \int_a^b f(t,s)\ ds$$

defined and differentiable on I and

$$\mathcal{I}'(t) = \int_a^b f_t(t,s)\ ds.$$

Proof. Fix $t \in I$. We want to show that

$$\lim_{h\to 0} \frac{\mathcal{I}(t+h) - \mathcal{I}(t)}{h} = \int_a^b f_t(t,s)\ ds,$$

equivalently,

$$\lim_{h\to 0} \frac{1}{h}\left(\mathcal{I}(t+h) - \mathcal{I}(t) - h\int_a^b f_t(t,s)\ ds\right) = 0.$$

That is, we want to show that for all $\epsilon > 0$, there is $\delta > 0$ such that if $|h| < \delta$, then

$$\left|\frac{1}{h}\left(\mathcal{I}(t+h) - \mathcal{I}(t) - h\int_a^b f_t(t,s)\ ds\right)\right| < \epsilon. \tag{C.3.1}$$

We compute

$$\mathcal{I}(t+h) - \mathcal{I}(t) - h\int_a^b f_t(t,s)\ ds = \int_a^b f(t+h,s)\ ds - \int_a^b f(t,s)\ ds - h\int_a^b f_t(t,s)\ ds$$
$$= \int_a^b \big[f(t+h,s) - f(t,s) - hf_t(t,s)\big]\ ds. \tag{C.3.2}$$

It therefore suffices to show

$$\lim_{h\to 0} \int_a^b \frac{f(t+h,s) - f(t,s) - hf_t(t,s)}{h}\ ds = 0. \tag{C.3.3}$$

By definition of the partial derivative, we know that

$$\lim_{h \to 0} \frac{f(t+h,s) - f(t,s) - hf_t(t,s)}{h} = 0$$

for any fixed t and s. Our challenge is now to make this limit hold "uniformly" over all $s \in [0,1]$ so that we can "pass the limit through the integral" in (C.3.3).

Example 3.2.24 allows us to rewrite

$$f(t+h,s) - f(t,s) = h \int_0^1 f_t(t+h\tau,s)\, d\tau,$$

and so

$$\int_a^b \big[f(t+h,s) - f(t,s) - hf_t(t,s)\big]\, ds = \int_a^b \left[h \int_0^1 f_t(t+h\tau,s)\, d\tau - hf_t(t,s)\right] ds. \quad \text{(C.3.4)}$$

Now rewrite

$$f_t(t,s) = f_t(t,s) \int_0^1 1\, d\tau = \int_0^1 f_t(t,s)\, d\tau,$$

so that

$$\int_a^b \left[h \int_0^1 f_t(t+h\tau,s)\, d\tau - hf_t(t,s)\right] ds = h \int_a^b \int_0^1 \big[f_t(t+h\tau,s) - f_t(t,s)\big]\, d\tau\, ds. \quad \text{(C.3.5)}$$

We combine (C.3.2), (C.3.4), and (C.3.5) to conclude that

$$\frac{1}{h}\left(\mathcal{I}(t+h) - \mathcal{I}(t) - h\int_a^b f_t(t,s)\, ds\right) = \int_a^b \int_0^1 \big[f_t(t+h\tau,s) - f_t(t,s)\big]\, d\tau\, ds,$$

and so we estimate with *two* applications of the triangle inequality that

$$\begin{aligned}\left|\frac{1}{h}\left(\mathcal{I}(t+h) - \mathcal{I}(t) - h\int_a^b f_t(t,s)\, ds\right)\right| &\le (b-a) \max_{a \le s \le b} \left|\int_0^1 \big[f_t(t+h\tau,s) - f_t(t,s)\big]\, d\tau\right| \\ &\le (b-a) \max_{a \le s \le b} \left(\max_{0 \le \tau \le 1} \big|f_t(t+h\tau,s) - f_t(t,s)\big|\right).\end{aligned}$$

Now we will use uniform continuity. Since I is an interval and $t \in I$, there are t_0, $t_1 \in I$ such that $t_0 < t < t_1$ and $[t_0, t_1] \subseteq I$. Then f_t is continuous on a set $\mathcal{D}$ of the first form in (C.1.1), and so given $\epsilon > 0$, there is $\delta > 0$ such that both $[t-\delta, t+\delta] \subseteq I$ and, if $|\xi - t| < \delta$, then

$$\big|f_t(\xi,s) - f_t(t,s)\big| < \frac{\epsilon}{b-a}$$

for all $s \in [a,b]$. What is critical here is that we can make the difference above uniformly small over all $s \in [a,b]$.

Take $0 < |h| < \delta$, so that $|(t+h\tau) - t| = |h||\tau < |h| < \delta$, since $0 \le \tau \le 1$. This guarantees

$$\big|f_t(t+h\tau,s) - f_t(t,s)\big| < \frac{\epsilon}{b-a},$$

and thus

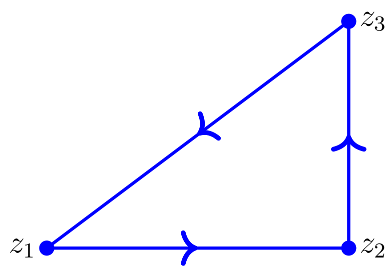
$$(b-a) \max_{a \le s \le b} \left(\max_{0 \le \tau \le 1} \big|f_t(t+h\tau,s) - f_t(t,s)\big|\right) < \epsilon$$

when $0 < |h| < \delta$. This proves the desired estimate (C.3.1). ■

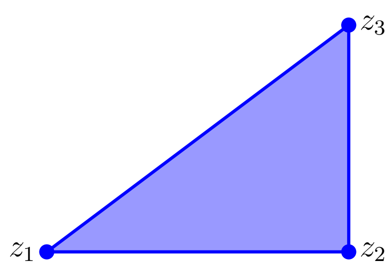
C.4. A fuller proof of Theorem 3.5.10.

We begin by calling upon the fearsome power of the triangle. This completes our return to kindergarten geometry begun with lines and circles.

What is a triangle? Let z_1, z_2, $z_3 \in \mathbb{C}$. Surely the path below is a triangle.

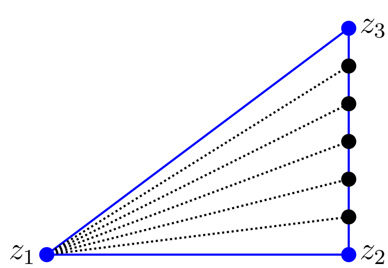


We recognize this path as the composition of three line segments in a particular order, namely $[z_1, z_2] \oplus [z_2, z_3] \oplus [z_3, z_1]$. However, we might also argue that the two-dimensional region below is a triangle as well.



Both "triangular paths" and "triangular regions" will be very useful to us, and so we should give precise definitions of them both, and use notation that distinguishes them. While we recognized the triangular path above as a composition of line segments, how might we tractably describe the triangular *region* above in terms of z_1, z_2, z_2?

One useful approach is to recognize the region as a *union* of line segments—specifically, all line segments whose initial point is z_1 and whose terminal point lies on the line segment $[z_2, z_3]$.



Based on this reasoning, we make the following definition.

C.4.1 Definition. *Let z_1, z_2, $z_3 \in \mathbb{C}$.*

(i) *The* **TRIANGLE** *spanned by z_1, z_2, and z_3 is the set*

$$\Delta(z_1, z_2, z_3) := \bigcup_{0 \le s \le 1} [z_1, (1-s)z_2 + sz_3] = \left\{(1-t)z_1 + t\big((1-s)z_2 + sz_3\big) \;\middle|\; 0 \le s,\ t \le 1\right\}. \tag{C.4.1}$$

(ii) *The* **TRIANGULAR PATH** *spanned by z_1, z_2, and z_3 is the closed path*

$$\partial\Delta(z_1, z_2, z_3) := [z_1, z_2] \oplus [z_2, z_3] \oplus [z_3, z_1]. \tag{C.4.2}$$

C.4.2 Problem ($\star$). Let z_1, z_2, $z_3 \in \mathbb{C}$.

(i) Prove that the order in which we specify the endpoints of a triangle is irrelevant in the sense that

$$\Delta(z_1, z_2, z_3) = \Delta\big(z_{\sigma(1)}, z_{\sigma(2)}, z_{\sigma(3)}\big)$$

for any function $\sigma \colon \{1, 2, 3\} \to \{1, 2, 3\}$ that is one-to-one and onto (i.e., any permutation). Explain why the order of the points matters very much when we are working with a triangular *path*.

(ii) Suppose that two or more of the points z_1, z_2, z_3 are equal, or that all three points belong to some line segment $[z, w]$. Prove that $\Delta(z_1, z_2, z_3)$ is really a line segment. (Remarkably, this "degenerate" case will not require any special treatment in our subsequent use of triangles!)

The key to a version of the Cauchy integral theorem that drops the hypothesis of continuity on f' is that f should integrate to 0 over triangles. This turns out to be true.

C.4.3 Theorem (Cauchy–Goursat theorem). *Suppose that f is holomorphic on an open set $\mathcal{D}$ (which need not be star-shaped or even a domain). Let z_1, z_2, $z_3 \in \mathcal{D}$ such that $\Delta(z_1, z_2, z_3) \subseteq \mathcal{D}$. Then*

$$\int_{\partial\Delta(z_1, z_2, z_3)} f = 0.$$

We will not prove this theorem here; its proof is a wonderful union of analysis (careful estimates using the definition of the derivative and the triangle inequality for integrals) and geometry (breaking a given triangle into an infinite sequence of nested triangles) and more analysis (estimating integrals over those nested triangles and finding a subsequence of triangles whose intersection is nonempty).

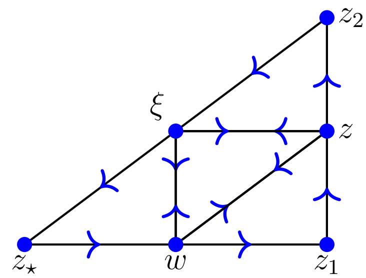
C.4.4 Corollary ("Relaxed" Cauchy–Goursat theorem). *Let $\mathcal{D} \subseteq \mathbb{C}$ be a star domain with star center $z_\star$. Suppose that $f \colon \mathcal{D} \to \mathbb{C}$ is continuous on $\mathcal{D}$ and holomorphic on*

$\mathcal{D} \setminus \{z_\star\}$. *Then*

$$\int_{\partial\Delta(z_\star, z_1, z_2)} f = 0$$

for any $z_1, z_2 \in \mathcal{D}$ *such that* $\Delta(z_\star, z_1, z_2) \subseteq \mathcal{D}$.

Proof. Let $w \in [z_\star, z_1]$, $z \in [z_1, z_2]$, and $\xi \in [z_2, z_\star]$ as drawn below.



Write $\int_{\partial\Delta(z_\star,z_1,z_2)} f$ as the sum of the integrals of f over the six line segments $[z_\star, w]$, $[w, z_1]$, $[z_1, z]$, $[z, z_2]$, $[z_2, \xi]$, and $[\xi, z_\star]$. Then add and subtract the integrals of f over the interior line segments $[w, \xi]$, $[z, w]$, and $[\xi, z]$. Conclude that $\int_{\partial\Delta(z_\star,z_1,z_2)} f$ is the sum of the integrals of f over the four triangles $\partial\Delta(z_\star, w, \xi)$, $\partial\Delta(w, z_1, z)$, $\partial\Delta(z, z_2, \xi)$, and $\partial\Delta(w, z, \xi)$. Furthermore, the integrals over the last three triangles are all 0, because those triangles are contained in $\mathcal{D} \setminus \{z_\star\}$; this set is open and f is holomorphic on $\mathcal{D} \setminus \{z_\star\}$, so the Cauchy–Goursat theorem applies there. We are left with

$$\int_{\partial\Delta(z_\star,z_1,z_2)} f = \int_{\partial\Delta(z_\star,w,\xi)} f.$$

We estimate

$$\max_{\eta \in \partial\Delta(z_\star,w,\xi)} |f(\eta)| \le \max_{\eta \in \partial\Delta(z_\star,z_1,z_2)} |f(\eta)| =: M,$$

and so

$$\left| \int_{\partial\Delta(z_\star,w,\xi)} f \right| \le M\big(|z_\star - w| + |w - \xi| + |\xi - z_\star|\big).$$

Since

$$\lim_{(z,w,\xi)\to(z_\star,z_\star,z_\star)} |z_\star - w| + |w - \xi| + |\xi - z_\star| = 0,$$

we conclude

$$\int_{\partial\Delta(z_\star,z_1,z_2)} f = \int_{\partial\Delta(z_\star,w,\xi)} f = 0.$$

■

C.4.5 Problem (!). Chase through the algebra of triangles, line segments, and integrals in the preceding proof. Specifically, carry out the direction to "add and subtract the integrals of f over the interior line segments $[w, \xi]$, $[z, w]$, and $[\xi, z]$ and conclude that $\int_{\partial\Delta(z_\star,z_1,z_2)} f$ is the sum of the integrals of f over the four triangles $\partial\Delta(z_\star, w, \xi)$, $\partial\Delta(w, z_1, z)$, $\partial\Delta(z, z_2, \xi)$, and $\partial\Delta(w, z, \xi)$."

At last we are ready to prove that a holomorphic function integrates to 0 around closed paths *without* assuming that the derivative is continuous and *without* assuming that the path is a triangle.

C.4.6 Theorem (Cauchy integral theorem). *Let $\mathcal{D} \subseteq \mathbb{C}$ be a star domain with star-center $z_\star$, and let $f \colon \mathcal{D} \to \mathbb{C}$ be continuous. If f is also holomorphic on $\mathcal{D} \setminus \{z_\star\}$, then*

$$\int_\gamma f = 0$$

for any closed path γ in $\mathcal{D}$.

Proof. We will show that

$$F(z) := \int_{[z_\star, z]} f$$

is an antiderivative of f on $\mathcal{D}$. The proof is very similar to that of Theorem 3.4.4, except we have replaced the general path connecting $z_\star$ and z with the line segment $[z_\star, z]$.

Fix $z \in \mathcal{D}$. As always, we want to show that

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

equivalently,

$$\lim_{h \to 0} \frac{F(z+h) - F(z) - hf(z)}{h} = 0.$$

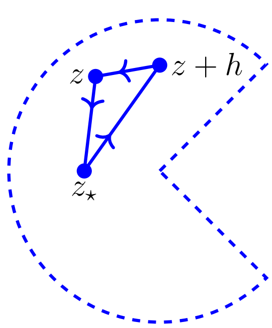
By Problem C.4.7 below, there is $r > 0$ such that if $h \in \mathbb{C}$ with $|h| < r$, then $\Delta(z_\star, z, z+h) \subseteq \mathcal{D}$. Assume that $h \in \mathbb{C}$ satisfies $|h| < r$ from now on.

We calculate

$$F(z+h) - F(z) = \int_{[z_\star, z+h]} f - \int_{[z_\star, z]} f = \int_{[z_\star, z+h]} f + \int_{[z, z_\star]} f.$$

If we add and subtract the integral of f over $[z+h, z]$, then we will have integrated f over the triangle $\partial\Delta(z_\star, z+h, z)$, and this integral is 0 by the relaxed Cauchy–Goursat theorem, since f is continuous on $\mathcal{D}$ and holomorphic on $\mathcal{D} \setminus \{z_\star\}$. That is,

$$\int_{\partial\Delta(z_\star, z+h, z)} f = 0.$$



So, we do just that:

$$\begin{aligned}F(z+h)-F(z) &= \int_{[z_\star,z+h]} f + \int_{[z+h,z]} f + \int_{[z,z_\star]} f - \int_{[z+h,z]} f \\ &= \int_{[z_\star,z+h]\oplus[z+h,z]\oplus[z,z_\star]} f + \int_{[z,z+h]} f \\ &= \int_{\partial\Delta(z_\star,z+h,z)} f + \int_{[z,z+h]} f \\ &= \int_{[z,z+h]} f.\end{aligned}$$

Then

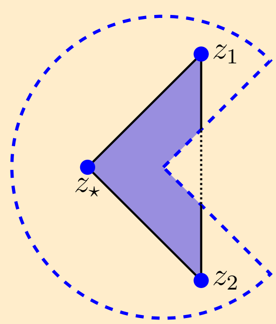
$$F(z+h)-F(z)-hf(z) = \int_{[z,z+h]} f - hf(z) = h\int_0^1 \big[f(z+th)-f(z)\big]\,dt,$$

as we previously calculated in (3.4.4). Since f is continuous on $\mathcal{D}$, Lemma 3.4.5 then implies

$$\lim_{h\to 0}\frac{F(z+h)-F(z)-hf(z)}{h} = \lim_{h\to 0}\int_0^1 \big[f(z+th)-f(z)\big]\,dt = 0,$$

as desired. ■

C.4.7 Problem ($\star$). Let $\mathcal{D}\subseteq\mathbb{C}$ be a star domain with star center $z_\star$ and let $z\in\mathcal{D}$. Since $\mathcal{D}$ is open, there is $r>0$ such that $\mathcal{B}(z;r)\subseteq\mathcal{D}$. Prove that if $h\in\mathbb{C}$ with $|h|<r$, then $\Delta(z_\star,z,z+h)\subseteq\mathcal{D}$. [Hint: *use the definition of a triangle as a union of line segments, the definition of an open ball, and a lot of estimates.*] Note that for *arbitrary* z_1, $z_2\in\mathcal{D}$, the triangle $\Delta(z_\star,z_1,z_2)$ need not be wholly contained in $\mathcal{D}$.



C.4.8 Problem ($\star$). Let $\mathcal{D}\subseteq\mathbb{C}$ be a star-domain with star-center $z_\star$ and suppose that $f\colon\mathcal{D}\to\mathbb{C}$ is continuous with the following property: for all $z\in\mathbb{C}$, there is $r>0$ such that if $h\in\mathbb{C}$ with $0<|h|<r$, then $\Delta(z_\star,z,z+h)\subseteq\mathcal{D}$ and $\int_{\Delta(z_\star,z,z+h)} f=0$. Reread the preceding problem and the proof of the Cauchy integral theorem and convince yourself that $\int_\gamma f=0$ for all closed paths γ in $\mathcal{D}$.

C.5. The proof of Lemma 3.6.16.

3.6.16 Lemma. *Let $z_0 \in \mathbb{C}$ and $r > 0$. Suppose that $f\colon \partial\mathcal{B}(z_0; r) \to \mathbb{C}$ is continuous and let $k \geq 1$ be an integer. Define*

$$F_k\colon \mathbb{C} \setminus \partial\mathcal{B}(z_0; r) \to \mathbb{C}\colon z \mapsto \int_{|w-z_0|=r} \frac{f(w)}{(w-z)^k}\, dw. \tag{C.5.1}$$

Then F_k is holomorphic with $F_k' = kF_{k+1}$.

Proof. This is essentially a differentiation under the integral argument for a very specific integrand. We need to show that for any $z \in \mathbb{C} \setminus \partial\mathcal{B}(z_0; r)$, we have

$$\lim_{h \to 0} \frac{F_m(z+h) - F_m(z)}{h} = mF_{m+1}(z),$$

equivalently,

$$\lim_{h \to 0} \frac{F_m(z+h) - F_m(z) - hmF_{m+1}(z)}{h}. \tag{C.5.2}$$

We compute

$$\begin{aligned} &F_m(z+h) - F_m(z) - hmF_{m+1}(z) \\ &\quad= \int_{|w-z_0|=r} f(w) \left[\frac{(w-z)^{m+1} - (w-z)((w-z)-h)^m - hm((w-z)-h)^m}{(w-z)^{m+1}((w-z)-h)}\right] dw. \end{aligned} \tag{C.5.3}$$

This calculation just requires finding a common denominator inside the integral.

We claim that for all integers $m \geq 1$, there is a function $P_m\colon \mathbb{C}^2 \to \mathbb{C}$ and a constant $C_m > 0$ such that

$$\xi^{m+1} - \xi(\xi+h)^m + mh(\xi+h)^m = h^2 P_m(\xi, h) \quad \text{and} \quad |P_m(\xi, h)| \leq C_m(|\xi| + |h|)^{m-1} \tag{C.5.4}$$

for all ξ, $h \in \mathbb{C}$. The proof of this claim is Problem C.5.1. With this claim in hand, we can estimate the integral on the right in (C.5.3) via the ML-inequality.

Problem 3.1.26 gives $d_0 > 0$ such that $d_0 < |w - z|$ for all w satisfying $|w - z_0| = r$. Since we are taking $h \to 0$, we may as well assume that

$$|h| \leq \min\left\{1, \frac{d_0}{2}\right\}. \tag{C.5.5}$$

The triangle inequality implies

$$|w - z| = |(w - z_0) + (z_0 - z)| \leq |w - z_0| + |z - z_0| = r + |z - z_0| =: \rho.$$

Then taking $\xi = w - z$ in (C.5.4) gives

$$\left|(w-z)^{m+1} - (w-z)((w-z)-h)^m - hm((w-z)-h)^m\right| = |P_m(w-z, -h)|$$

$$\le C_m|h|^2(|w-z|+|-h|)^{m-1} \le C_m|h|^2(\rho+1)^{m-1},$$

while the reverse triangle inequality implies

$$|(w-z)-h| \ge |w-z|-|h| \ge d_0-|h| \ge \frac{d_0}{2}.$$

We put

$$M := \max_{|w-z_0|=r} |f(w)|,$$

and use the ML-inequality to estimate

$$\big|F_m(z+h)-F_m(z)-hmF_{m+1}(z)\big| \le \frac{2\pi r M C_m|h|^2(\rho+1)^{m-1}}{d_0^{m+1}\left(\dfrac{d_0}{2}\right)}.$$

If we divide both sides by $|h|$, we conclude

$$\left|\frac{F_m(z+h)-F_m(z)-hmF_{m+1}(z)}{h}\right| \le C|h|, \qquad C := \frac{4MC_m\pi r(\rho+1)^{m-1}}{d_0^{m+2}}.$$

The squeeze theorem then yields the limit (C.5.2). ■

C.5.1 Problem (+). Prove the claim (C.5.4) using one of the following options.

(i) Add and subtract $(\xi+h)^{m+1}$ to find

$$\xi^{m+1}-\xi(\xi+h)^m+mh(\xi+h)^m = -\big((\xi+h)^{m+1}-\xi^{m+1}\big)+(m+1)h(\xi+h)^m.$$

Rewrite

$$(\xi+h)^{m+1}-\xi^{m+1} = (m+1)h\int_0^1 (\xi+th)^m \, dt$$

using the fundamental theorem of calculus and obtain

$$\xi^{m+1}-\xi(\xi+h)^m+mh(\xi+h)^m = (m+1)h\left(\int_0^1 \big[(\xi+h)^m-(\xi+th)^m\big]\, dt\right).$$

Use the fundamental theorem of calculus again to rewrite

$$\int_0^1 \big[(\xi+h)^m-(\xi+th)^m\big]\, dt = mh\int_0^1\int_0^1 (1-t)\big(\xi+th+\tau h(1-t)\big)^{m-1}\, d\tau\, dt.$$

Define

$$P_m(\xi,h) := m(m+1)\int_0^1\int_0^1 (1-t)\big(\xi+th+\tau h(1-t)\big)^{m-1}\, d\tau\, dt.$$

Prove the estimate on P_m using multiple applications of the triangle inequality.

(ii) Expand $(\xi+h)^m$ using the binomial theorem:

$$(\xi+h)^m = \sum_{k=0}^{m}\binom{m}{k}\xi^k h^{m-k} = \xi^m + m\xi^{m-1}h + \sum_{k=0}^{m-2}\binom{m}{k}\xi^k h^{m-k}.$$

Then do arithmetic.

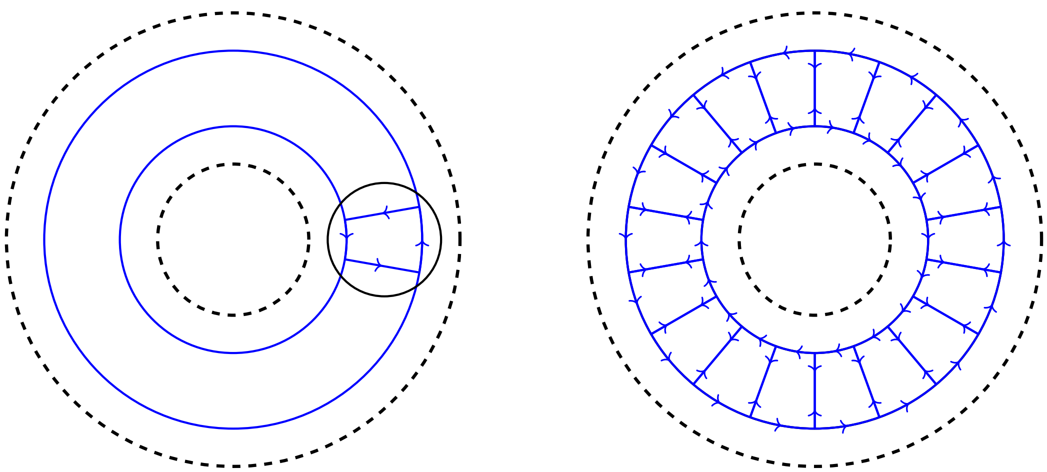
C.6. The proof of Theorem 4.4.6.

We first state and prove another "deformation" lemma. This resembles the Death Star lemma in that we show that the integral of a holomorphic function over one circle equals the integral of that function over another circle. However, now the circles remain centered at the same points and only the radii change; moreover, the function is not assumed to be holomorphic on a certain circle "interior" to both circles over which the integrals run.

C.6.1 Lemma. *Suppose that f is analytic on the annulus $\mathcal{A}(z_0; r, R)$ and $r < \rho < \mathrm{P} < R$. Then*

$$\int_{|z-z_0|=\rho} f = \int_{|z-z_0|=\mathrm{P}} f. \tag{C.6.1}$$

Proof. We "partition" the annulus $\mathcal{A}(z_0; \rho, \mathrm{P})$ into a family of "rectangles" $\gamma_0, \ldots, \gamma_n$ as in the sketch below.



More precisely, each "rectangle" is of the form

$$\gamma_k := [z_0 + \rho e^{i(k-1)\theta_n}, z_0 + \mathrm{P}e^{i(k+1)\theta_n}] \oplus \lambda_k \oplus [z_0 + \mathrm{P}e^{i(k+1)\theta_n}, z_0 + \rho e^{i(k+1)\theta_n}] \oplus \mu_k^-,$$

where

$$\theta_n = \frac{2\pi}{n}$$

for some positive integer n,

$$\lambda_k(t) = z_0 + \mathrm{P}e^{it}, \ \theta_k \le t \le \vartheta_k,$$

and

$$\mu_k(t) = z_0 + \rho e^{it},\ \theta_k \le t \le \vartheta_k.$$

The integer n is chosen to be large enough that each "rectangle" γ_k is contained in the ball $\mathcal{B}(z_k; s)$, where

$$z_k = z_0 + \left(\frac{\rho + \mathrm{P}}{2}\right) e^{ik\theta_n} \quad \text{and} \quad s := \frac{\mathrm{P} - \rho}{2} + \min\left\{\frac{R - \mathrm{P}}{2}, \frac{\rho - r}{2}\right\}.$$

This choice of center and radius for $\mathcal{B}(z_k; s)$ ensures $\mathcal{B}(z_k; s) \subseteq \mathcal{A}(z_0; r, R)$, so f is analytic on $\mathcal{B}(z_k; s)$. Since the ball $\mathcal{B}(z_k; s)$ is a star-domain, the Cauchy integral theorem implies $\int_{\gamma_k} f = 0$ for all k.

We then have

$$0 = \sum_{k=1}^{n} \int_{\gamma_k} f = \sum_{k=1}^{n} \int_{\lambda_k} f - \sum_{k=1}^{n} \int_{\mu_k} f = \int_{|z-z_0|=\mathrm{P}} f - \int_{|z-z_0|=\rho} f,$$

from which the equality (C.6.1) follows. ■

Now we restate and prove the Laurent decomposition.

4.4.6 Theorem. *Let $z_0 \in \mathbb{C}$ and $0 \le r < R \le \infty$. Suppose that $f\colon \mathcal{A}(z_0; r, R) \to \mathbb{C}$ is analytic. Then there exist unique analytic functions*

$$f_R\colon \mathcal{B}(0; R) \to \mathbb{C} \quad \text{and} \quad f_r\colon \mathcal{B}(0; 1/r) \to \mathbb{C},$$

where we interpret $\mathcal{B}(0; 1/0) = \mathcal{B}(0; \infty) = \mathbb{C}$, such that $f_r(0) = 0$ and

$$f(z) = f_R(z - z_0) + f_r\left(\frac{1}{z - z_0}\right)$$

for each $z \in \mathcal{A}(z_0; r, R)$. We may expand f_R and f_r as power series centered at 0 to find

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \sum_{k=1}^{\infty} \frac{a_{-k}}{(z - z_0)^k} = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k, \tag{C.6.2}$$

where for each $k \in \mathbb{Z}$, the coefficient a_k satisfies

$$a_k = \frac{1}{2\pi i} \int_{|z-z_0|=s} \frac{f(z)}{(z - z_0)^{k+1}}\, dz \tag{C.6.3}$$

for any $s \in (r, R)$.

Proof. We give the proof in the following steps.

1. *Reduction to the case $z_0 = 0$.* Suppose that the theorem is true for $z_0 = 0$ and define

$$g\colon \mathcal{A}(0; r, R) \to \mathbb{C}\colon z \mapsto f(z + z_0).$$

Then g is analytic, and so there is a Laurent decomposition (g_R, g_r) for g on $\mathcal{A}(0; r, R)$. That is, $g_R\colon \mathcal{B}(0; R)$ and $g_r\colon \mathcal{B}(0; 1/r)$ are analytic with $g_r(0) = 0$ and

$$g(z) = g_R(z) + g_r\left(\frac{1}{z}\right), \ z \in \mathcal{A}(0; r, R),$$

and so

$$f(z) = g(z - z_0) = g_R(z - z_0) + g_r\left(\frac{1}{z - z_0}\right), \ z \in \mathcal{A}(z_0; r, R).$$

Thus (g_R, g_r) is a Laurent decomposition for f on $\mathcal{A}(z_0; r, R)$ as well.

Suppose that (f_R, f_r) is another Laurent decomposition for f on $\mathcal{A}(z_0; r, R)$. Put $\widetilde{g}_R(z) := f_R(z+z_0)$ for $z \in \mathcal{B}(0; R)$ and $\widetilde{g}_r(\xi) := f_r(\xi+z_0)$ for $\xi \in \mathcal{B}(0; 1/r)$. Then $(\widetilde{g}_R, \widetilde{g}_r)$ is a Laurent decomposition for g on $\mathcal{A}(0; r, R)$, and so $\widetilde{g}_R = g_R$ and $\widetilde{g}_r = g_r$. Thus $f_R = g_R$ and $f_r = g_r$. This proves the uniqueness of the decomposition on $\mathcal{A}(z_0; r, R)$.

Finally, we discuss the coefficients. We have

$$g_R(z) = \sum_{k=0}^{\infty} a_k w^k \quad \text{and} \quad g_r(\xi) = \sum_{k=1}^{\infty} a_{-k}\xi^k, \quad a_k = \frac{1}{2\pi i}\int_{|w|=s} \frac{g(w)}{w^{k+1}}\,dw, \ r < s < R.$$

Part (i) of Problem 3.3.7 and the formula $g(z) = f(z + z_0)$ then give the formula (C.6.3) for a_k in terms of f.

2. *Uniqueness on* $\mathcal{A}(0; r, R)$. Suppose that an analytic function $g\colon \mathcal{A}(0; r, R) \to \mathbb{C}$ can be written as

$$g(z) = g_R(z) + g_r\left(\frac{1}{z}\right) \quad \text{and} \quad g(z) = \breve{g}_R(z) + \breve{g}_r\left(\frac{1}{z}\right)$$

for all $z \in \mathcal{A}(0; r)$ and some analytic functions g_R, $\breve{g}_R\colon \mathcal{B}(0; R) \to \mathbb{C}$ and g_r, $\breve{g}_r\colon \mathcal{B}(0; 1/r) \to \mathbb{C}$ with $g_r(0) = \breve{g}_r(0) = 0$. Put

$$h_R\colon \mathcal{B}(0; R) \to \mathbb{C}\colon z \mapsto g_R(z) - \breve{g}_R(z) \quad \text{and} \quad h_r\colon \mathcal{B}(0; 1/r) \to \mathbb{C}\colon \xi \mapsto g_r(\xi) - \breve{g}_r(\xi).$$

Then h_R and h_r are analytic and $h_r(0) = 0$. Additionally,

$$h_R(z) + h_r\left(\frac{1}{z}\right) = g(z) - g(z) = 0, \ z \in \mathcal{A}(0; r, R).$$

Now consider the analytic function

$$H_r\colon \mathcal{A}(0; r, \infty) \to \mathbb{C}\colon z \mapsto h_r\left(\frac{1}{z}\right),$$

which satisfies

$$h_R(z) = -H_r(z), \ z \in \mathcal{A}(0; r, R).$$

The merging lemma (Lemma 3.5.25) implies that the function

$$H\colon \mathbb{C} \to \mathbb{C}\colon z \mapsto \begin{cases} h_R(z), & |z| < R \\ -H_r(z), & |z| > r \end{cases}$$

is well-defined and analytic, i.e., entire.

We now show that H is bounded. First suppose $R < |z|$. Since $r < R$, if $R < |z|$, then $r < |z|$ and $1/|z| < 1/R < 1/r$, thus

$$|H(z)| = |H_r(z)| = \left|h_r\left(\frac{1}{z}\right)\right| \leq \max_{|w|\leq 1/R} |h_r(W)| =: M_R.$$

Since H is entire, the maximum

$$m_R := \max_{|z|\leq R} |H(z)|$$

certainly exists. And so $|H(z)| \leq \max\{M_R, m_r\}$ for any $z \in \mathbb{C}$. Thus H is indeed bounded; since H is also entire, by Liouville's theorem H is constant, say, $H(z) = c$ for all $z \in \mathbb{C}$.

Now let $n \geq r$ be an integer. Then

$$c = |H(n)| = |H_r(n)| = \left|h_r\left(\frac{1}{n}\right)\right| \to 0 \text{ as } n \to \infty.$$

The limit holds because h_r is analytic on $\mathcal{B}(0; 1/r)$ and $h_r(0) = 0$. Thus $c = 0$, and so $H(z) = 0$ for all z. We conclude, therefore, that $h_R(z) = 0$ for all $|z| < R$ and

$$0 = H_r(z) = h_r\left(\frac{1}{z}\right)$$

for all $|z| > r$, thus $h_r(w) = 0$ for all $|w| < r$. This proves that $g_R = \breve{g}_R$ and $g_r = \breve{g}_r$.

3. *Existence on subannuli* $\mathcal{A}(0; \rho, \mathrm{P})$. Assume that $g\colon \mathcal{A}(0; r, R) \to \mathbb{C}$ is analytic and let $r < \rho < \mathrm{P} < R$. Fix $z \in \mathcal{A}(0; \rho, \mathrm{P})$. By Problem 4.3.7, the function

$$\phi\colon \mathcal{A}(0; r, R) \to \mathbb{C}\colon w \mapsto \begin{cases} \dfrac{g(w) - g(z)}{w - z}, & w \neq z \\ \\ g'(z), & w = z \end{cases}$$

is analytic on $\mathcal{A}(0; \rho, \mathrm{P})$. Lemma C.6.1 therefore implies that

$$\int_{|w|=\rho} \phi(w)\ dw = \int_{|w|=\mathrm{P}} \phi(w)\ dw.$$

That is,

$$\int_{|w|=\rho} \frac{g(w) - g(z)}{w - z}\ dw = \int_{|w|=\mathrm{P}} \frac{g(w) - g(z)}{w - z},$$

which rearranges to

$$\int_{|w|=\rho} \frac{g(w)}{w - z}\ dw - g(z) \int_{|w|=\rho} \frac{dw}{w - z} = \int_{|w|=\mathrm{P}} \frac{g(w)}{w - z}\ dw - g(z) \int_{|w|=\mathrm{P}} \frac{dw}{w - z}.$$

Since $|z| > \rho$, the Cauchy integral formula implies

$$\int_{|w|=\rho} \frac{dw}{w - z} = 0,$$

while since $|z| < \mathrm{P}$, the Cauchy integral theorem implies

$$\int_{|w|=\mathrm{P}} \frac{dw}{w-z} = 2\pi i.$$

We therefore obtain

$$\int_{|w|=\rho} \frac{g(w)}{w-z}\, dw = \int_{|w|=\mathrm{P}} \frac{g(w)}{w-z}\, dw - 2\pi i g(z),$$

and, in turn,

$$g(z) = \frac{1}{2\pi i}\int_{|w|=\mathrm{P}} \frac{g(w)}{w-z}\, dw - \frac{1}{2\pi i}\int_{|w|=\rho} \frac{g(w)}{w-z}\, dw.$$

Lemma 3.6.16 implies that

$$g_{\mathrm{P}}(z) := \frac{1}{2\pi i}\int_{|w|=\mathrm{P}} \frac{g(w)}{w-z}\, dw \tag{C.6.4}$$

is analytic on $\mathbb{C} \setminus \mathcal{C}(0;\mathrm{P})$ and that

$$\widetilde{g}_\rho(\xi) := -\frac{1}{2\pi i}\int_{|w|=\rho} \frac{g(w)}{w-\xi}\, dw$$

is analytic on $\mathbb{C} \setminus \mathcal{C}(0;\rho)$. The work above shows

$$g(z) = g_{\mathrm{P}}(z) + \widetilde{g}_\rho(z),\ \ z \in \mathcal{A}(0;\rho,\mathrm{P}).$$

We really want to write g in the form

$$g(z) = g_{\mathrm{P}}(z) + g_\rho\left(\frac{1}{z}\right)$$

for some analytic function $g_\rho \colon \mathcal{B}(0;1/\rho) \to \mathbb{C}$, and so this suggests defining

$$g_\rho(\xi) := \begin{cases} \widetilde{g}_\rho\left(\dfrac{1}{\xi}\right),\ 0 < |\xi| < \dfrac{1}{\rho} \\ \\ 0,\ \xi = 0. \end{cases}$$

We now need to check that g_ρ is analytic on $\mathcal{B}(0;1/\rho)$. By definition, g_ρ is analytic on $\mathcal{B}^*(0;1/\rho)$, and for $\xi \in \mathcal{B}^*(0;1/\rho)$, we have

$$g_\rho(\xi) = \widetilde{g}_\rho\left(\frac{1}{\xi}\right) = -\frac{1}{2\pi i}\int_{|w|=\rho} \frac{g(w)}{w-\dfrac{1}{\xi}}\, dw = \frac{1}{2\pi i}\int_{|w|=\rho} \frac{g(w)}{\dfrac{1-\xi w}{\xi}}\, dw = \frac{1}{2\pi i}\int_{|w|=\rho} \frac{\xi g(w)}{1-\xi w}\, dw. \tag{C.6.5}$$

If we can show that $\lim_{\xi\to 0} g_\rho(\xi) = 0$, then g_ρ will be analytic on $\mathcal{B}(0; 1/\rho)$ by part (ii) of Problem 4.3.9. To do this, we may assume that $|\xi| \le 1/2\rho$ and use the reverse triangle inequality to bound

$$|1 - \xi w| \ge 1 - |\xi w| = 1 - \rho|\xi| \ge 1 - \rho\left(\frac{1}{2\rho}\right) = \frac{1}{2}.$$

Then

$$\left|\int_{|w|=\rho} \frac{g(w)}{1-\xi w}\, dw\right| \le \pi\rho M_\rho(g), \qquad M_\rho(g) := \max_{|w|=\rho} |g(w)|,$$

and so by the squeeze theorem

$$\lim_{\xi\to 0} g_\rho(\xi) = \frac{1}{2\pi i} \lim_{\xi\to 0} \xi \left(\int_{|w|=\rho} \frac{g(w)}{1-\xi w}\, dw\right) = 0.$$

4. *Existence on* $\mathcal{A}(0; r, R)$. Let $g\colon \mathcal{A}(0; r, R) \to \mathbb{C}$ be analytic. Step 3 above proves the existence of a Laurent decomposition (g_{P}, g_ρ) on any annulus $\mathcal{A}(0; \rho, \mathrm{P})$ with $r < \rho < \mathrm{P} < R$, and Step 2 shows that this decomposition is unique. (In Step 2, just replace r with ρ and R with P.) Now let $r < \rho_1 < \rho_2 < \mathrm{P}_2 < \mathrm{P}_1 < R$, so $\mathcal{A}(0; \rho_2, \mathrm{P}_2) \subseteq \mathcal{A}(0; \rho_1, \mathrm{P}_1) \subseteq \mathcal{A}(0; r, R)$. Let $(g_{\mathrm{P}_1}, g_{\rho_1})$ be the Laurent decomposition of g on $\mathcal{A}(0; \rho_1, \mathrm{P}_1)$ and let $(g_{\mathrm{P}_2}, g_{\rho_2})$ be the Laurent decomposition of g on $\mathcal{A}(0; \rho_2, \mathrm{P}_2)$. Then the restriction $\big(g_{\mathrm{P}_1}\big|_{\mathcal{A}(0;\rho_2,\mathrm{P}_2)}, g_{\rho_1}\big|_{\mathcal{A}(0;\rho_2,\mathrm{P}_2)}\big)$ is a Laurent decomposition for g on $\mathcal{A}(0; \rho_2, \mathrm{P}_2)$, and so $g_{\mathrm{P}_1}(z) = g_{\mathrm{P}_2}(z)$ for all $z \in \mathcal{B}(0; \mathrm{P}_2)$ and $g_{\rho_1}(\xi) = g_{\rho_2}(\xi)$ for all $\xi \in \mathcal{B}(0; 1/\rho_2)$.

Now define

$$g_R\colon \mathcal{B}(0; R) \to \mathbb{C}\colon z \mapsto g_{\mathrm{P}}(z),\ |z| < \mathrm{P} \quad \text{and} \quad g_r\colon \mathcal{B}(0; 1/\rho) \to \mathbb{C}\colon \xi \mapsto g_\rho(\xi),\ |\xi| < 1/\rho.$$

By the work above, these are well-defined, analytic functions with $g_r(0) = 0$. Moreover, if $z \in \mathcal{A}(0; r, R)$, and if $r < \rho < |z| < \mathrm{P} < R$, then

$$g(z) = g_{\mathrm{P}}(z) + g_\rho\left(\frac{1}{z}\right) = g_R(z) + g_r\left(\frac{1}{z}\right).$$

This proves the existence of the Laurent decomposition on $\mathcal{A}(0; r, R)$.

5. *Coefficients on* $\mathcal{A}(0; r, R)$. Let $g\colon \mathcal{A}(0; r, R) \to \mathbb{C}$ be analytic, and let (g_R, g_r) be its Laurent decomposition. Since g_R and g_r are analytic and $g_r(0) = 0$, they have power series expansions of the form

$$g_R(z) = \sum_{k=0}^{\infty} \alpha_k z^k \quad \text{and} \quad g_r(z) = \sum_{k=1}^{\infty} \beta_k z^k.$$

Now we need to calculate their coefficients α_k and β_k more transparently in terms of g.

We begin with an observation that may seem unmotivated but is in fact quite important. For any s_1, $s_2 \in (r, R)$ and $n \in \mathbb{Z}$, Lemma C.6.1 implies that

$$\int_{|w|=s_1} \frac{g(w)}{w^n}\, dw = \int_{|w|=s_2} \frac{g(w)}{w^n}\, dw.$$

For this reason, the numbers

$$a_k := \frac{1}{2\pi i} \int_{|w|=s} \frac{g(w)}{w^{k+1}}\, dw, \;\; s \in (r, R), \;\; k \in \mathbb{Z},$$

are defined independently of s.

Now fix $z \in \mathcal{B}(0; R)$ and let $\mathrm{P} \in (r, R)$ such that $|z| < \mathrm{P} < R$. By (C.6.4),

$$g_R(z) = g_{\mathrm{P}}(z) = \frac{1}{2\pi i} \int_{|w|=\mathrm{P}} \frac{g(w)}{w-z}\, dw = \frac{1}{2\pi i} \int_{|w|=\mathrm{P}} \frac{g(w)}{w} \left(\frac{1}{1 - \frac{z}{w}} \right) dw$$

$$= \frac{1}{2\pi i} \int_{|w|=\mathrm{P}} \frac{g(w)}{w} \sum_{k=0}^{\infty} \left(\frac{z^k}{w^k} \right) dw.$$

Here we have used the estimate $|z| < \mathrm{P} = |w|$ to invoke the geometric series. Then the interchange theorem (Theorem 4.1.4) allows us to conclude

$$g_R(z) = \frac{1}{2\pi i} \sum_{k=0}^{\infty} \int_{|w|=\mathrm{P}} \frac{g(w)z^k}{w^{k+1}}\, dw = \sum_{k=0}^{\infty} \left(\frac{1}{2\pi i} \int_{|w|=\mathrm{P}} \frac{g(w)}{w^{k+1}}\, dw \right) z^k, \; z \in \mathcal{A}(0; \rho, \mathrm{P}) = \sum_{k=0}^{\infty} a_k z^k.$$

Similarly, with $\xi \in \mathcal{B}(0; 1/r)$ and $\rho \in (r, R)$ such that $|\xi| < 1/\rho$, by (C.6.5) we have

$$g_r(\xi) = g_\rho(\xi) = \frac{1}{2\pi i} \int_{|w|=\rho} \frac{\xi g(w)}{1 - \xi w}\, dw = \frac{\xi}{2\pi i} \int_{|w|=\rho} g(w) \sum_{k=0}^{\infty} (\xi w)^k\, dw.$$

Here we have used the estimate $|\xi w| = |\xi|\rho < (1/\rho)\rho = 1$ to introduce the geometric series. Then the interchange theorem implies

$$g_r(\xi) = \frac{\xi}{2\pi i} \sum_{k=0}^{\infty} \int_{|w|=\rho} g(w)\xi^k w^k\, dw = \sum_{k=0}^{\infty} \left(\frac{1}{2\pi i} \int_{|w|=\rho} g(w) w^k\, dw \right) \xi^{k+1}$$

$$= \sum_{j=1}^{\infty} \left(\frac{1}{2\pi i} \int_{|w|=\rho} g(w) w^{j-1}\, dw \right) \xi^j = \sum_{j=1}^{\infty} \left(\frac{1}{2\pi i} \int_{|w|=\rho} \frac{g(w)}{w^{-j+1}}\, dw \right) \xi^j = \sum_{j=1}^{\infty} a_{-j} \xi^j. \quad \blacksquare$$

C.7. The proofs of equations (4.5.3) and (4.5.9).

We prove a theorem that encapsulates the situations of both equalities.

C.7.1 Theorem. *Let $0 \le r < R \le \infty$ and $z_0 \in \mathbb{C}$, and let $f \colon \mathcal{A}(z_0; r, R) \to \mathbb{C}$ be analytic. Let (f_R, f_r) be the Laurent decomposition of f on $\mathcal{A}(z_0; r, R)$. Let γ be a closed curve in $\mathcal{B}(0; 1/r)$ with $\mathcal{B}(0; 1/0) = \infty$. Then*

$$\int_\gamma f_r \left(\frac{1}{z - z_0} \right) dz = a_{-1} \int_\gamma \frac{dz}{z - z_0} = 2\pi i \operatorname{Res}(f; z_0) \chi(\gamma; z_0).$$

Proof. Recall that $f_r\colon \mathcal{B}(0;1/r) \to \mathbb{C}$ is analytic, with $\mathcal{B}(0;1/0) = \mathbb{C}$, and

$$f_r(w) = \sum_{k=1}^{\infty} a_{-k} w^k, \qquad a_{-k} := \frac{1}{2\pi i} \int_{|z-z_0|=s} f(z)(z-z_0)^{k-1}\, dz,\ r < s < R. \tag{C.7.1}$$

We will apply the interchange theorem (Theorem 4.1.4) to interchange the order of summation and integration and show

$$\int_\gamma \sum_{k=1}^{\infty} \frac{a_{-k}}{(z-z_0)^k}\, dz = \sum_{k=1}^{\infty} \int_\gamma \frac{a_{-k}}{(z-z_0)^k}\, dz. \tag{C.7.2}$$

For $k \geq 2$, the fundamental theorem of calculus (see Problem 4.5.3) provides

$$\int_\gamma \frac{dz}{(z-z_0)^k} = 0.$$

Then

$$\sum_{k=1}^{\infty} \int_\gamma \frac{a_{-k}}{(z-z_0)^k}\, dz = a_{-1} \int_\gamma \frac{dz}{z-z_0} = 2\pi i \operatorname{Res}(f;z_0)\chi(\gamma;z_0).$$

To justify the use of the interchange theorem, we first call upon Problem 3.1.26 to summon up $t_0 \in [a,b]$ such that

$$|\gamma(t) - z_0| \geq |\gamma(t_0) - z_0| =: d_0$$

for all $t \in [a,b]$. And since $\gamma(t) \in \mathcal{A}(z_0;r,R)$ for all t, we have $d_0 = |\gamma(t_0) - z_0| > r$. Then we estimate

$$\left| \frac{a_{-k}}{(z-z_0)^k} \right| \leq \frac{|a_{-k}|}{d_0^k} \tag{C.7.3}$$

for any $z \in \text{image}(\gamma)$. Next, from (C.7.1) and the ML-inequality, we estimate

$$|a_{-k}| \leq \frac{1}{2\pi}(2\pi s)M_s(f)s^{k-1} = M_s(f)s^k, \qquad M_s(f) := \max_{|z-z_0|=s} |f(z)|. \tag{C.7.4}$$

This is valid for any $s \in (r,R)$.

Combining (C.7.3) and (C.7.4), we have

$$\left| \frac{a_{-k}}{(z-z_0)^k} \right| \leq M_s(f) \left(\frac{s}{d_0} \right)^k.$$

This is valid for any $z \in \text{image}(\gamma)$, any $s \in (r,R)$, and any integer $k \geq 1$. Since $r < d_0$, we may choose $s \in (r, d_0)$ to ensure $s/d_0 \in (0,1)$. Then the interchange theorem applies to validate (C.7.2). ■

D. Polar Coordinates Revisited

We present here a construction of polar coordinates, specifically of the principal argument, that is independent of the annoying "existential" result in part ($\pi 4$) of Theorem 1.5.10. Specifically, we will use the Cauchy integral theorem to redevelop the principal logarithm and principal argument with explicit integral formulas.

We do assume the following.

D.1 Hypothesis. **(i)** *Theorems 1.5.2 and 1.5.4 about the exponential are true. (Most of these results are easy to prove anyway, except for the functional equation.)*

(ii) *There exists $\pi > 0$ such that $e^{i\tau} \notin (-\infty, 0)$ for $\tau \in (-\pi, \pi)$ and $e^{i\pi} = -1$. These results are consequences of Theorem 1.5.10, but we do not assume any part of that theorem here.*

(iii) *The function*

$$\ln\colon (0,\infty) \to \mathbb{R}\colon t \mapsto \int_1^t \frac{d\tau}{\tau}$$

satisfies $e^{\ln(t)} = t$ for all $t > 0$. This was developed in Problem 3.2.19. Moreover, we can write $\ln$ *(trivially) as the line integral*

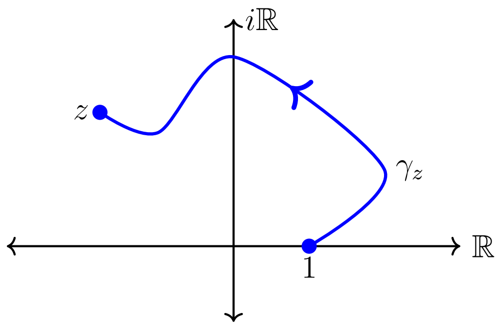
$$\ln(t) = \int_{[1,t]} \frac{d\tau}{\tau}$$

by Problem 3.3.6.

Assuming these, and only these, results about exponentials, natural logarithms, and π, we put $\mathcal{D} := \mathbb{C} \setminus (-\infty, 0]$, so $\mathcal{D}$ is a star domain.

D.2 Problem (!). Draw a picture to convince yourself that $\mathbb{C} \setminus (-\infty, 0]$ is a star domain. What should a star-center be? Prove it.

For $z \in \mathcal{D}$, let $\gamma_z\colon [0,1] \subseteq \mathbb{R} \to \mathcal{D}$ be a path with $\gamma_z(0) = 1$ and $\gamma_z(1) = z$.



Now put

$$L\colon \mathcal{D} \to \mathbb{C}\colon z \mapsto \int_{\gamma_z} \frac{dw}{w}.$$

D.3 Problem (+). Explain why L is well-defined and holomorphic on $\mathcal{D}$ with $L'(z) = 1/z$. Exactly the same strategy as incan then be used to show that $e^{L(z)} = z$ for all $z \in \mathcal{D}$; this is really just a Show that L solves the IVP

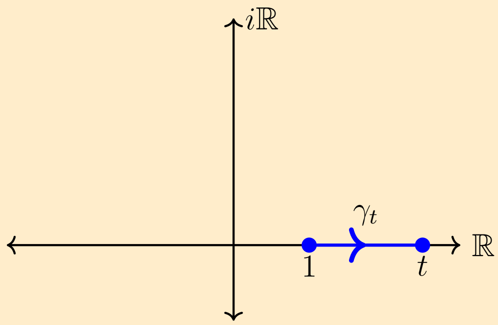
$$\begin{cases} L'(z) = 1/z \\ L(1) = 0 \end{cases}$$

and, using the same strategy as in part (ii) of Problem 3.2.19, deduce that $e^{L(z)} = z$ for all $z \in \mathcal{D}$.

We compute $L(z)$ for several increasingly more complicated forms of $z \in \mathcal{D}$.

D.4 Problem (+). **(i)** $z = t > 0$. Show that

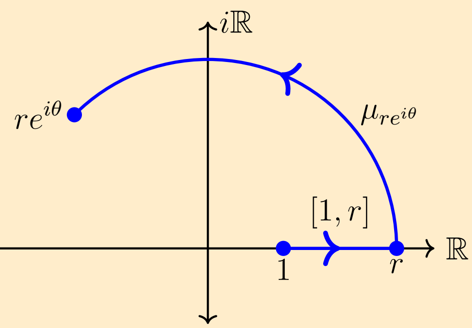
$$L(t) = \ln(t).$$



(ii) $z = re^{i\theta}$ for some $r > 0$ and $\theta \in (-\pi, \pi)$. Put

$$\mu_{re^{i\theta}} \colon [0,1] \to \mathbb{C} \colon t \mapsto re^{i\theta t},$$

so $[1, r] \oplus \mu_{re^{i\theta}}$ is a path in $\mathcal{D}$ from 1 to $re^{i\theta}$.



Show that

$$L(re^{i\theta}) = \ln(r) + i\theta.$$

[Hint: *use the definition of the line integral to evaluate* $\int_{\mu_{re^{i\theta}}} dw/w$.]

(iii) $z \in \mathcal{D}$ is arbitrary. Take $\gamma_z = [1, |z|] \oplus \mu_z$, where μ_z is any path in $\mathcal{D}$ from $|z|$ to z. Show that

$$L(z) = \ln(|z|) + iA(z), \qquad A(z) := -i \int_{\mu_z} \frac{dw}{w}.$$

We will show that A can be thought of as the principal argument Arg restricted to $\mathbb{C} \setminus (-\infty, 0]$. Namely, we will show the polar coordinates identity $z = |z|e^{iA(z)}$, the reality of A, and the bounds $-\pi < A(z) < \pi$. (Recall that A is defined on $\mathbb{C} \setminus (-\infty, 0]$, so we want to exclude π from the range of A.)

D.5 Problem (+). Let $z \in \mathcal{D}$.

(i) Use the identities $z = e^{L(z)}$ and $L(z) = \ln(|z|) + iA(z)$ to compute

$$z = |z|e^{iA(z)}.$$

(ii) Show that $|e^{iA(z)}| = 1$ and thus $e^{-\operatorname{Im}[A(z)]} = 1$. Conclude that $\operatorname{Im}[A(z)] = 0$ and therefore $A(z) \in \mathbb{R}$.

(iii) Conclude further that

$$\operatorname{Re}[L(z)] = \ln(|z|) \quad \text{and} \quad \operatorname{Im}[L(z)] = A(z)$$

and therefore that A is continuous on $\mathcal{D}$.

(iv) To obtain the desired bounds $-\pi < A(z) < \pi$, suppose instead that $|A(z)| \geq \pi$. Let $\nu\colon [0, 1] \to \mathcal{D}$ be a path from 1 to z, and put $f(t) := |A(\nu(t))|$. Use the intermediate value theorem to find $t_0 \in (0, 1)$ such that $f(t_0) = \pi$. Then use the polar coordinates identity to show

$$\nu(t_0) = |\nu(t_0)|e^{iA(\nu(t_0))} < 0.$$

Why is this a contradiction?

Here is what we have shown.

D.6 Theorem. *For $z \in \mathbb{C} \setminus (-\infty, 0]$, let γ_z be any path in $\mathbb{C} \setminus (-\infty, 0]$ with initial point 1 and terminal point z. Put*

$$L(z) := \int_{\gamma_z} \frac{dw}{w} \quad \text{and} \quad A(z) := \operatorname{Im}[L(z)].$$

Then the maps

$$\mathcal{A}\colon \mathbb{C} \setminus \{0\} \to (-\pi, \pi]\colon z \mapsto \begin{cases} A(z), & z \in \mathbb{C} \setminus (-\infty, 0] \\ \pi, & z \in (-\infty, 0) \end{cases}$$

and

$$\mathcal{L}\colon \mathbb{C}\setminus\{0\} \to \mathbb{C}\colon z \mapsto \begin{cases} L(z), & z \in \mathbb{C}\setminus(-\infty,0] \\ L(|z|)+i\pi, & z \in (-\infty,0) \end{cases}$$

satisfy the following.

(i) $e^{\mathcal{L}(z)} = z$ *for all* $z \in \mathbb{C}\setminus\{0\}$ *and* $\mathcal{A}(re^{i\theta}) = \theta$ *for* $\theta \in (-\pi,\pi]$.

(ii) $z = |z|e^{i\mathcal{A}(z)}$ *for all* $z \in \mathbb{C}\setminus\{0\}$.

(iii) $\operatorname{Re}[\mathcal{L}(z)] = \ln(|z|)$ *and* $\operatorname{Im}[\mathcal{L}(z)] = \mathcal{A}(z)$.

(iv) *The map* $\mathcal{A}$ *is continuous on* $\mathbb{C}\setminus(-\infty,0]$ *and discontinuous on* $(-\infty,0]$.

(v) *The map* $\mathcal{L}$ *is holomorphic on* $\mathbb{C}\setminus(-\infty,0]$ *with* $\mathcal{L}'(z) = 1/z$, *but* $\mathcal{L}$ *is not continuous on* $(-\infty,0]$.

Of course, we write $\operatorname{Log} = \mathcal{L}$ and $\operatorname{Arg} = \mathcal{A}$.

D.7 Problem (+). Using the definition of $\mathcal{A}$ given above, check that the method of Example 2.2.19 still works to show that $\mathcal{A}$ is discontinuous on $(-\infty,0)$. [Hint: $\mathcal{A}(re^{it}) = t$ *for* $r > 0$ *and* $t \in (-\pi,\pi)$.] Then show that $\mathcal{A}$ is also discontinuous at 0. Along the way, be sure to demonstrate why none of these discontinuities are removable.

D.8 Problem (+). Let $\mathcal{D} \subseteq \mathbb{C}$ be an elementary domain. A **HOLOMORPHIC LOGARITHM** of a function $f\colon \mathcal{D} \to \mathbb{C}$ is a holomorphic function $L\colon \mathcal{D} \to \mathbb{C}$ such that $f(z) = e^{L(z)}$ for all $z \in \mathcal{D}$. This problem shows that $f\colon \mathcal{D} \to \mathbb{C}$ has a holomorphic logarithm on $\mathcal{D}$ if and only if f is both holomorphic and never 0 on $\mathcal{D}$.

(i) Show that if $f\colon \mathcal{D} \to \mathbb{C}$ has a holomorphic logarithm L on $\mathcal{D}$, then f is analytic, $f(z) \neq 0$ for all $z \in \mathcal{D}$, and

$$L'(z) = \frac{f'(z)}{f(z)}, \quad z \in \mathcal{D}. \tag{D.0.1}$$

Thus (D.0.1) determines L, up to a constant of integration.

(ii) Conversely, suppose that $f\colon \mathcal{D} \to \mathbb{C}$ is holomorphic and $f(z) \neq 0$ for all $z \in \mathcal{D}$. The identity (D.0.1) suggests that we might define a holomorphic logarithm of f on $\mathcal{D}$ as an antiderivative of f'/f; specifically, fix $z_\star \in \mathcal{D}$ and put

$$L(z) := C + \int_{\gamma_z} \frac{f'}{f},$$

where γ_z is a path in $\mathcal{D}$ with initial point $z_\star$ and terminal point z, and C is a constant of integration that we will determine later.

First explain why L is well-defined and satisfies (D.0.1). Then, following the method of part (ii) of Problem 3.2.19, put $g(z) = e^{-L(z)}f(z)$, show that $g'(z) = 0$, and determine the value of C that yields $g(z_\star) = 1$. Conclude, with this value of C, that $g(z) = 1$ and thus $f(z) = e^{L(z)}$.