

Complex Analysis II (Michaelmas Term)

Amit Einav¹

2024-2025

Contents

1	The Complex Plane and Riemann Sphere	1
1.1	Complex numbers	1
1.2	Exponential and trigonometric functions	5
1.3	Logarithms and complex powers	6
1.4	The Riemann Sphere and extended complex plane	11
2	Metric Spaces	14
2.1	Metric spaces	14
2.2	Open and closed sets	17
2.3	Convergence and continuity	22
2.4	Sequential Compactness and Compactness	29
3	Complex Differentiation	34
3.1	Complex differentiability	34
3.2	Cauchy-Riemann equations	35
3.3	Connected sets and zero derivatives	39
3.4	The angle-preserving properties of holomorphic functions	41
3.5	Biholomorphic maps	44
4	Möbius transformations	47
4.1	Definition and first properties of Möbius transformations	47
4.2	Fixed points, the cross-ratio, and the three points Theorem	50
4.3	Circles and lines	53
4.4	The Riemann sphere, revisited	56
4.5	Möbius transformations preserving the upper half plane or unit disc	57
4.6	Finding biholomorphic maps between domains	61
5	Notions of Convergence and Power Series	65
5.1	Pointwise and uniform convergence	65
5.2	Locally uniform convergence	70

5.3	Complex power series	72
6	Complex integration over contours	78
6.1	Definition of contour integrals	78
6.2	The Fundamental Theorem of Calculus	84
A	Additional lemmas from the Gappy notes	91
A.1	Complex function	91
A.2	Möbius transformations	93
A.3	Biholomorphic domains - revisited	97

¹These notes are based on the original notes of Jens Funke from many years ago. Chapters 1-5 evolved through Stephen, Michael Magee, Sabine Bögli, Katie Gittins, and Amit Einav. Chapter 6 onwards have gone through a similar evolution through Jens, Thanasis Bouganis, Michael & Stephen.

Chapter 1

The Complex Plane and Riemann Sphere

1.1 Complex numbers

A *complex number* z is a quantity of the form $z = x + iy$, where x, y are real numbers and i is the *imaginary unit*. We denote by $\mathbb{C}$ the set of all complex numbers.

Algebra

We can add, subtract and multiply complex numbers: If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ then

$$\begin{aligned} z_1 \pm z_2 &:= (x_1 \pm x_2) + i(y_1 \pm y_2), \\ z_1 z_2 &:= (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1). \end{aligned}$$

Notice that addition simply corresponds to adding the individual components. In general we denote by $\operatorname{Re}(z) = x$ the *real part of* z , and by $\operatorname{Im}(z) = y$ the *imaginary part of* z . By the definition of multiplication we have $i^2 = -1$, and using this we see that multiplication corresponds to ‘multiplying out the brackets’: $(x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 + i^2 y_1 y_2) + i(x_1 y_2 + x_2 y_1)$.

We can also divide complex numbers. For $z_2 \neq 0$ (here we use the shorthand $0 = 0 + 0i$) we have

$$\frac{z_1}{z_2} = \frac{x_1 + iy_1}{x_2 + iy_2} = \frac{(x_1 + iy_1)(x_2 - iy_2)}{(x_2 + iy_2)(x_2 - iy_2)} = \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + i \frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2} \in \mathbb{C}.$$

The quantity we used to make the denominator real is important. In general, for $z = x + iy$ we call $\bar{z} := x - iy$ the *complex conjugate of* z .

We immediately have a multiplicative inverse

$$z^{-1} := \frac{1}{z} = \frac{x}{x^2 + y^2} - i \frac{y}{x^2 + y^2}.$$

WARNING: While most of the nice properties of $\mathbb{R}$ hold in $\mathbb{C}$, we do not have notions of $\leq$, $<$, $\geq$ or $>$; the set $\mathbb{C}$ is not ‘ordered’ and expressions like $z_1 < z_2$ have no meaning.

How do we visualise the complex numbers?

There are various ‘models’ for the complex numbers. The most commonly used/most intuitive is to think of $\mathbb{C}$ as a copy of $\mathbb{R}^2$ equipped with a map

$$\mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2 : ((x_1, y_1), (x_2, y_2)) \mapsto (x_1x_2 - y_1y_2, x_1y_2 + x_2y_1).$$

So, a copy of $\mathbb{R}^2$ with a way of multiplying¹ (and dividing) vectors! Indeed there is an obvious bijection $f : \mathbb{R}^2 \rightarrow \mathbb{C}$ given by $f((x, y)) = x + iy$. As a result, we often draw complex numbers on the usual (x, y) -plane: such a picture is called an *Argand diagram* (see Figure 1.1).

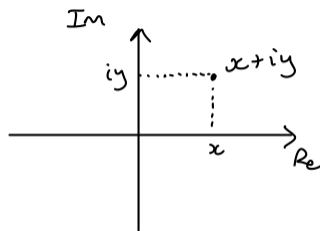


Figure 1.1: Argand diagram.

On $\mathbb{R}^2$ there is a natural notion of size, and we use it in $\mathbb{C}$: we call the quantity $|z| := \sqrt{x^2 + y^2}$ the *modulus* or *absolute value* of $z (= x + iy)$.

Lemma 1.1 (Important Properties of Complex numbers).

1. $z_1 z_2 = 0 \iff z_1 = 0 \text{ or } z_2 = 0$,
2. $|z| = \sqrt{z \bar{z}}$,
3. $\operatorname{Re}(z) = \frac{z + \bar{z}}{2}$ and $\operatorname{Im}(z) = \frac{z - \bar{z}}{2i}$,

¹The multiplication of complex numbers is commutative and associative. The fact that such a multiplication exists in 2 dimensions is truly remarkable: there is no such multiplication on $\mathbb{R}^3$, not even if we relax the condition that it is commutative! The search for an associative multiplication on $\mathbb{R}^3$ lead Hamilton to discover (a non-commutative) one on $\mathbb{R}^4$ instead which was his famous discovery of the quaternions.

$$4. z^{-1} = \frac{\bar{z}}{|z|^2}.$$

Remark 1.2. Property 1. is very important, and makes $\mathbb{C}$ an *integral domain* (see Algebra II).

Polar form/coordinates of a complex number

Now we have a notion of distance, as in $\mathbb{R}^2$ we can implement a change of variables $z(x, y) \rightarrow z(r, \theta)$. Let $r = |z|$ and let θ denote the anticlockwise angle measured from the real axis. (Angles measured clockwise will be considered negative.) We call θ the *argument* of z (for $z \neq 0$) and write $\arg(z) = \theta$. We then have the following *polar coordinates* for z :

$$z = r(\cos \theta + i \sin \theta),$$

which we write in shorthand as $z = re^{i\theta}$ (see Figure 1.2).

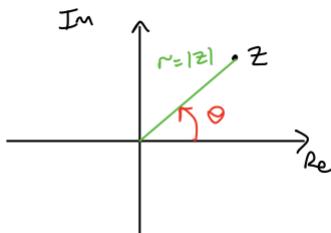


Figure 1.2: Polar coordinates.

For example, $i = e^{i\pi/2}$ and $1 + i = \sqrt{2}e^{i\pi/4}$. We have $|-1/\sqrt{2} - i\sqrt{3}/\sqrt{2}| = \sqrt{2}$, so

$$-1/\sqrt{2} - i\sqrt{3}/\sqrt{2} = \sqrt{2}(-1/2 - i\sqrt{3}/2) = \sqrt{2}e^{-i2\pi/3}.$$

Note that $\arg(z)$ is only defined up to multiples of 2π ; for example $i = e^{i\pi/2} = e^{i5\pi/2} = e^{-i3\pi/2}$. Strictly speaking $\arg(i) = \pi/2 + 2\pi k$, for any $k \in \mathbb{Z}$ (and so $\arg$ is a one-to-many function!). As a result, we need to be careful; we choose a fixed interval in which to express the argument: the *principal value* of $\arg(z)$ is the value in the interval $(-\pi, \pi]$ and will be denoted $\text{Arg}(z)$. So $\text{Arg}(i) = \pi/2$ and $\text{Arg}(-1) = \pi$ for example.

Lemma 1.3 (Properties of argument). *We have the following properties of the argument:*

1. $\arg(z_1 z_2) = \arg(z_1) + \arg(z_2) \bmod 2\pi$
2. $\arg(1/z) = -\arg(z) \bmod 2\pi$
3. $\arg(\bar{z}) = -\arg(z) \bmod 2\pi$.

When we say two real numbers are equal mod 2π we mean they differ by an integer multiple of 2π .

It is nice to have a geometric picture of what the algebraic operations on complex numbers mean.

Lemma 1.4. *Geometrically, multiplication in $\mathbb{C}$ is given by a dilated rotation; i.e., if $z_1 = r_1 e^{i\theta_1}$ and $z_2 = r_2 e^{i\theta_2}$ then*

$$z_1 z_2 = r_1 r_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)) = r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$

In particular, multiplying by z_2 constitutes an (anticlockwise) rotation of z_1 by θ_2 degrees, followed by a dilation with factor r_2 . Furthermore, addition represents a translation and conjugation represents a reflection in the real axis. Taking the real or imaginary part of a complex number z represents a projection of z onto the real or imaginary axis respectively.

Proof. By the standard double angle formula we have

$$\begin{aligned} z_1 z_2 &= r_1 r_2 (\cos \theta_1 + i \sin \theta_1) (\cos \theta_2 + i \sin \theta_2) \\ &= r_1 r_2 ((\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2) + i(\sin \theta_1 \cos \theta_2 + \sin \theta_2 \cos \theta_1)) \\ &= r_1 r_2 (\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)). \end{aligned}$$

The geometric interpretations of addition, conjugation, and real/imaginary parts are fairly obvious (see Figure 1.3). $\square$

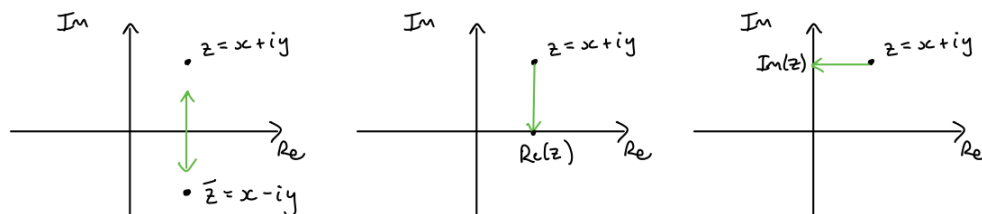


Figure 1.3: Geometric interpretation of conjugation and taking real and imaginary parts.

Corollary 1.5. 1. $|z_1 z_2| = |z_1| |z_2|$,

2. *De Moivre's formula:* $(\cos \theta + i \sin \theta)^n = \cos(n\theta) + i \sin(n\theta)$.

The modulus also has the following important properties.

1. (Triangle inequality) $|z_1 + z_2| \leq |z_1| + |z_2|$
2. $|z| \geq 0$ and $|z| = 0 \iff z = 0$
3. $\max(|\operatorname{Re}(z)|, |\operatorname{Im}(z)|) \leq |z| \leq |\operatorname{Re}(z)| + |\operatorname{Im}(z)|$

Remark 1.6. The first two properties above along with Corollary 1.5 make the modulus a norm on $\mathbb{C}$ (see later, Definition 2.2).

Simple complex functions and geometry (examples)

- We can also use functions to define regions in the complex plane. Consider the set of points z which satisfy the inequality $|z - i| < |z + i|$. This is precisely the points in $\mathbb{C}$ whose distance to i is strictly smaller than their distance to $-i$. Thus, the inequality represents the upper half plane $\mathbb{H} := \{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}$.
- Note that the equation $|z - i| = 1$ represents a circle centred at i of radius 1.

1.2 Exponential and trigonometric functions

Definition 1.7 (Complex exponential). We define the **complex exponential function** $\exp : \mathbb{C} \rightarrow \mathbb{C}$ by

$$\exp(z) := e^x(\cos y + i \sin y). \quad (z = x + iy)$$

As shorthand we write $\exp(z) = e^z$.

Remark 1.8. We will see later that $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ as in real analysis. We could have started with this as the definition instead.

Proposition 1.9. *We have the following properties of the complex exponential function:*

1. $e^z \neq 0$ for all $z \in \mathbb{C}$,
2. $e^{z_1+z_2} = e^{z_1}e^{z_2}$,
3. $e^z = 1 \iff z = 2\pi ik$ for some $k \in \mathbb{Z}$,
4. $e^{-z} = 1/e^z$,
5. $|e^z| = e^{\operatorname{Re}(z)}$.

Proof. Most are straightforward to check. 3. is very important: $\exp(z) = 1 \iff e^x \cos y = 1$ and $e^x \sin y = 0$. Since $e^x > 0$, the latter is equivalent to $\sin y = 0$ and so $y = n\pi$ for some $n \in \mathbb{Z}$. Thus we have $\exp(z) = 1 \iff e^x \cos(n\pi) = 1 \iff e^x(-1)^n = 1 \iff n$ is even and $e^x = 1 \iff x = 0$ and $y = 2k\pi$ ($k \in \mathbb{Z}$). $\square$

Corollary 1.10. *We have $\exp(2\pi i) = 1$ and $\exp(\pi i) = -1$. The latter is **Euler's formula**.*

Corollary 1.11. *The complex exponential function is $2\pi i$ -periodic; that is, $\exp(z + 2k\pi i) = \exp(z)$ for any $k \in \mathbb{Z}$.*

Remark 1.12. The above implies $\exp$ is determined entirely by the values it takes in any horizontal strip of width 2π in the complex plane. Note also that $\exp$ is ‘unbounded’, since by 5. the modulus $|\exp(z)|$ gets arbitrarily large as $\operatorname{Re}(z)$ increases.

Definition 1.13 (Trigonometric functions). All as functions from $\mathbb{C} \rightarrow \mathbb{C}$, we define

$$\begin{aligned}\sin(z) &:= \frac{1}{2i}(e^{iz} - e^{-iz}) & \cos(z) &:= \frac{1}{2}(e^{iz} + e^{-iz}) \\ \sinh(z) &:= \frac{1}{2}(e^z - e^{-z}) & \cosh(z) &:= \frac{1}{2}(e^z + e^{-z})\end{aligned}$$

(For $z = x$ real these coincide with the real functions. We will eventually get power series expressions for them from the one for the exponential function.)

Remark 1.14. All the usual double angle formulae and equations relating the functions hold: e.g., $\cos^2(z) + \sin^2(z) = 1$. Additionally, notice that we have $\cosh(iz) = \cos(z)$ and $\cos(iz) = \cosh(z)$, while $\sinh(iz) = i \sin(z)$ and $\sin(iz) = i \sinh(z)$. All four functions are unbounded.

1.3 Logarithms and complex powers

We use the notation $\mathbb{C}^* = \mathbb{C} - \{0\}$, i.e. the set of nonzero complex numbers.

Lemma 1.15. *[Inverting the exponential function] For every $w \in \mathbb{C}^*$, the equation*

$$e^z = w \tag{1.1}$$

has a solution z . Furthermore, if we write $w = |w|e^{i\phi}$ with $\phi = \operatorname{Arg}(w)$, then all solutions to (1.1) are given by

$$z = \log |w| + i(\phi + 2k\pi) \quad \text{for } k \in \mathbb{Z}. \tag{1.2}$$

Here, $\log |w|$ is the usual natural logarithm of the real number $|w|$. Note that there are infinitely many solutions.

Proof. If z is of the form as in (1.2) for some given $k \in \mathbb{Z}$, then

$$e^z = e^{\log |w| + i(\phi + 2k\pi)} = e^{\log |w|} e^{i(\phi + 2k\pi)} = e^{\log |w|} e^{i\phi} = |w| e^{i\phi} = w,$$

by Proposition 1.9, Part 2. and Corollary 1.10. Thus, z is a solution.

To see all solutions are of the given form, first write $z = x + iy$ and assume $e^z = w$. Since $e^x e^{iy} = e^z = w = |w|e^{i\phi}$, we have $|e^z| = e^x = |w|$. Thus $x = \log |w|$. Moreover, dividing both sides by $|w|$ we have $e^{iy} = e^{i\phi}$ and so $e^{i(y-\phi)} = 1$. It follows from Proposition 1.9, Part 3. that all solutions are given by $i(y - \phi) = 2k\pi i$ for some $k \in \mathbb{Z}$; in other words, $y = \phi + 2k\pi$. $\square$

We now come to an important topic called *branch cuts*.

First we give a high-level description of why branch cuts are necessary. We have just described for fixed $w_0 \neq 0$, exactly what the possible numbers are such that $e^z = w_0$. These values of z could all reasonably be called $\log(w_0)$. Can we make this choice of solution to $e^z = w_0$ vary nicely as we move w_0 a little bit? ². Of course, we will run into trouble at 0, since $e^z = 0$ has no solutions. Bearing this in mind, could we at least define a ‘log’ function that is ‘continuous’ on $\mathbb{C} \setminus \{0\}$? Since we don’t know the definition of continuous yet, let us just ask that the function should not jump abruptly when we move from a point to a nearby one.

The answer is **no**, and let’s see why by trying to come up with one.

Let’s suppose $w_0 = 1$ and we pick a solution to $e^z = w_0$. The obvious one is $z = 0$ so let’s pick that. Now let w be close to 1. If we’ve found a solution to $e^z = w$ then we know from Lemma 1.15 that it must be of the form

$$z = \log |w| + i(\text{Arg}(w) + 2\pi k)$$

for some $k \in \mathbb{Z}$. Since w is close to 1, we know $\log |w|$ is close to 0 and $\text{Arg}(w)$ is close to 0. Now, if k is not zero, then z will not be close to 0 (any integer that is not zero has absolute value at least one!). So k must be zero for z to be close to 0. *We just argued that if we move w a little, and if we want $\log$ to be ‘continuous’, we must choose the argument of w ‘continuously’.*

Now let’s take our idea a little further. If we move w on a path beginning at 1, following the unit circle anticlockwise, until we reach 1 again, what happens to our solutions to $e^z = w$ if we are choosing them continuously as above? In other words, what happens to the argument of w ? Since we are moving anticlockwise and we are choosing the argument continuously, it is increasing as we go around the unit circle. So just before we complete the circle, the argument of w is just below 2π . This is a big problem, since it means there are values of w just below 1 on the unit circle where we have been forced to set $\log(w)$ very close to $2\pi i$. On the other hand, we began by assuming $\log(1) = 0$. So the way we have tried to do things, our $\log$ function is going to have a jump discontinuity below 1.

No matter what way we try to define $\log$ continuously on $\mathbb{C} \setminus \{0\}$, we will run into a similar problem. On the other hand, if R_θ is any ray of the form

²To be precise, we would like the solution to $e^z = w$ to vary as a continuous function of w ; see later for the precise definition of continuous function.

$$R_\theta = \{re^{i\theta} : r \in \mathbb{R}, r \geq 0\} \subset \mathbb{C}$$

then it is possible to define a continuous function $\log(z)$ on $\mathbb{C} \setminus R_\theta$. (One can think of cutting out this ray as cutting out the points where $\log$ will have a jump discontinuity).

Definition 1.16 (Complex logarithm functions). For any two real numbers $\theta_1 < \theta_2$ with $\theta_2 - \theta_1 = 2\pi$, let $\arg$ be the choice of argument function with values in $(\theta_1, \theta_2]$. Then the function

$$\log(z) := \log|z| + i \arg(z)$$

is called a **branch of logarithm**. It has a jump discontinuity along the ray $R_{\theta_1} = R_{\theta_2}$. This ray is called a **branch cut**.

If we choose $\arg(z) = \text{Arg}(z) \in (-\pi, \pi]$, then we obtain a branch of logarithm called the **principal branch of log**. We write Log for this principal branch: it is given by the formula

$$\text{Log}(z) := \log|z| + i \text{Arg}(z).$$

The principal branch of logarithm has a ‘jump discontinuity’ along the ray given by the non-positive real axis $\mathbb{R}_{\leq 0}$.

Remark 1.17. • Any time one talks about a function called $\log$, one *has* to declare which branch of $\log$ we use. This is normally done simply by stating the interval $(\theta_1, \theta_2]$ where $\arg(z)$ lives.

- As soon as we define continuous functions, we will easily be able to see that the branch of $\log$ corresponding to $\arg(z) \in (\theta_1, \theta_2]$ is continuous on $\mathbb{C} \setminus R_{\theta_1}$.
- The principal branch, Log , agrees with the natural logarithm $\log$ on the real line; that is, for $x > 0$ we have $\text{Log}x = \log x$. For this reason *we will always use the principal branch unless otherwise stated*.

Lemma 1.18. [*Properties of logarithms*] We have the following properties when using any given branch of logarithm:

1. $e^{\log z} = z$ for any $z \in \mathbb{C} \setminus \{0\}$, but,
2. in general, $\log(zw) \neq \log z + \log w$, and
3. in general, $\log(e^z) \neq z$.

We would now like to define functions giving powers of complex numbers. We already know from the world of real numbers that to define a function giving for example, a square root of a positive real number, we have to make a choice of whether to take the positive or negative root. To take a root of a complex number, we have to make a similar choice, but we have more freedom.

Definition 1.19 (Complex powers). For $w \in \mathbb{C}$ fixed, by choosing any branch of $\log$ we can define a branch of the function $z \mapsto z^w$ by the expression

$$z^w := \exp(w \log z).$$

For example, if $w = 1/n$ and we use the principal branch we get

$$z^{1/n} = e^{(\log |z| + i \operatorname{Arg}(z))/n} = |z|^{1/n} e^{i \operatorname{Arg}(z)/n}.$$

Warning: different branches of $\log$ can give different power functions! So we must always specify which branch of $\log$ we are using.

Remark 1.20. Now that we have defined complex powers, we should check that our exponential function matches up with the concept of ‘raising e to the power z ’ for a suitable choice of $\log e$. The natural choice of $\log e$ is 1. Then, e raised to the power z should agree with computing $\exp(z \log e) = \exp(z)$ as we expected.

Finding values of $\log$ and complex powers (examples)

(a) Using the principal branch of $\log$, we find $\log(1 - i)$ and $(1 - i)^{1/2}$. We have $|1 - i| = \sqrt{2}$ and $\operatorname{Arg}(1 - i) = -\pi/4$. Thus, $1 - i = \sqrt{2}e^{-i\pi/4}$. Therefore, $\operatorname{Log}(1 - i) = \log |1 - i| + i \operatorname{Arg}(1 - i) = \log \sqrt{2} - i\pi/4$, and

$$(1 - i)^{1/2} = \exp\left(\frac{1}{2} \operatorname{Log}(1 - i)\right) = \exp\left(\frac{1}{2} \left(\log \sqrt{2} - i\frac{\pi}{4}\right)\right) = \exp\left(\log \sqrt[4]{2} - i\frac{\pi}{8}\right) = \sqrt[4]{2}e^{-i\frac{\pi}{8}}.$$

(b) Using the principal branch and the previous example

$$(1 - i)^i = \exp(i \operatorname{Log}(1 - i)) = \exp\left(\frac{\pi}{4} + i \log \sqrt{2}\right) = e^{\pi/4} e^{i \log \sqrt{2}}.$$

(c) Again, using the principal branch

$$2^{1/2} = \exp\left(\frac{1}{2} \log 2\right) = \exp(\log \sqrt{2}) = \sqrt{2}.$$

What about the other root? It comes from using a different branch: if we let $\log$ be the branch of logarithm corresponding to $\arg(z) \in (\pi, 3\pi]$ then $\log(z) = \log |z| + i(\operatorname{Arg}(z) + 2\pi)$, so we have $\log(2) = \log 2 + i2\pi$ and

$$2^{1/2} = \exp\left(\frac{1}{2}(\log 2 + i2\pi)\right) = \exp(\log \sqrt{2} + i\pi) = \sqrt{2}e^{i\pi} = -\sqrt{2}.$$

Remark 1.21. All n th roots can be found this way (see Assignments). In particular, for $z \in \mathbb{C}^*$ all n th roots of z are of the form

$$z^{1/n} = |z|^{1/n} \exp\left(i \frac{\operatorname{Arg}(z)}{n} + \frac{2k\pi i}{n}\right) \quad \text{for } k = 0, \dots, n-1.$$

Visualizing complex functions

The ‘graph’ of a complex-valued function $f : \mathbb{C} \rightarrow \mathbb{C}$ is 4-dimensional, so difficult to visualise - we certainly can’t draw it. We can employ other techniques to get a grasp on complex functions:

- We can graph the real-valued function $|f| : \mathbb{C} \rightarrow \mathbb{R}$. For example, consider the complex function $\cos z$. When $z = x$ is purely real, we have that $|f(z)| = |\cos(x)|$ is obviously just the modulus of the real cosine function. But for $z = iy$ purely complex we have $|f(z)| = |\cos(iy)| = |\cosh y| = \cosh y$. So in the imaginary direction f simply looks like cosh!
- It is often useful to visualise complex functions by considering how they map regions of the complex plane. Consider the image of the ‘right half-plane’ $\mathbb{H}_R := \{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$ under the map $f(z) = z^2$. Note that $\mathbb{H}_R = \{z \in \mathbb{C} : -\pi/2 < \operatorname{Arg}(z) < \pi/2\}$. If $z = re^{i\theta} \in \mathbb{H}_R$ then $z^2 = zz = r^2e^{i2\theta}$ has argument $2\theta \in (-\pi, \pi)$. Thus, f maps $\mathbb{H}_R$ to $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$, where $\mathbb{R}_{\leq 0}$ denotes the negative real axis (including 0), see Figure 1.4. The map is onto, since for every $w = se^{i\phi} \in \mathbb{C} \setminus \mathbb{R}_{\leq 0}$ (so $\phi \in (-\pi, \pi)$) we can find $z \in \mathbb{H}_R$ such that $f(z) = w$; namely we can choose $z = \sqrt{s}e^{i\phi/2} \in \mathbb{H}_R$.

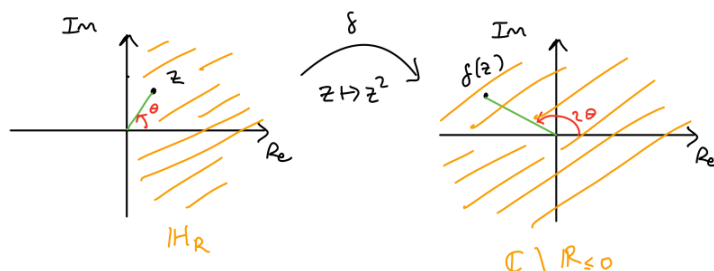


Figure 1.4: Geometric picture of how $f(z) = z^2$ transforms the right half-plane.

Similarly, the left half plane $\mathbb{H}_L = \{z \in \mathbb{C} : \operatorname{Re}(z) < 0\}$ is mapped to $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$. Moreover, f maps both the strictly positive imaginary axis $i\mathbb{R}_{>0}$ and the strictly negative imaginary axis $i\mathbb{R}_{<0}$ (given by $i\mathbb{R}_{>0} = \{iy \in \mathbb{C} : y > 0\}$ and $i\mathbb{R}_{<0} = \{iy \in \mathbb{C} : y < 0\}$ respectively) to the strictly negative real axis $\mathbb{R}_{<0}$.

Adding the observation that $f(0) = 0$, we have that $f(z) = z^2$ in essence maps $\mathbb{C}$ to two copies of itself (except for the origin, which is only attained once in the image - remember this different behaviour at the origin later!)

- **Branches of log.** Let $\log$ be the branch of logarithm corresponding to $\arg(z) \in (\theta_1, \theta_2]$. Then $\log$ maps $\mathbb{C} \setminus R_{\theta_1}$ to the infinite horizontal strip

$$\{z \in \mathbb{C} : \theta_1 < \operatorname{Im}(z) \leq \theta_2\}.$$

Infinite rays emanating from 0 map to horizontal lines, and circles centred at zero, minus their intersection with R_{θ_1} , map to vertical line segments (see Figure 1.5).

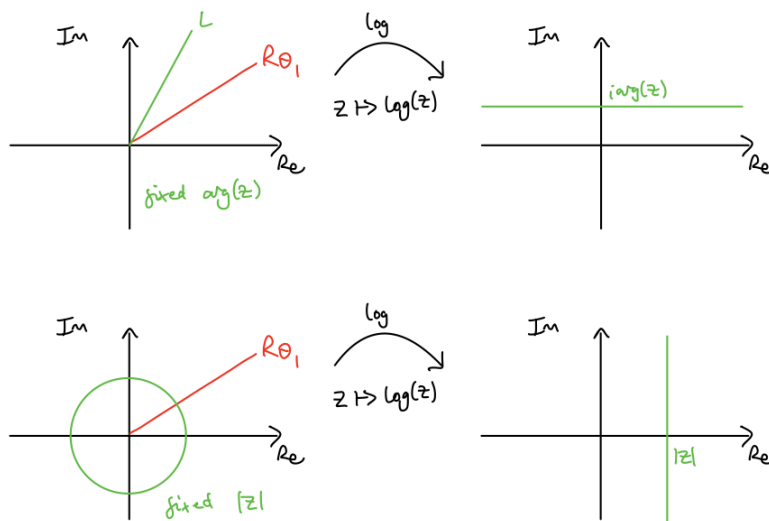


Figure 1.5: Geometric picture of how $f(z) = \log(z)$ transforms rays and circles.

1.4 The Riemann Sphere and extended complex plane

It is very useful at various points of the course to extend the complex plane by adding a point ‘at infinity’. To do this, we create a new object called ‘infinity’, denoted ∞ , and consider the set

$$\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}.$$

At the moment, we have accomplished nothing really. What will be useful later is that we can think of the point ∞ as being glued ‘nicely’ onto $\mathbb{C}$. The correct way to do this is by introducing the *Riemann sphere*.

The Riemann sphere

Consider the unit sphere $S^2 := \{(x, y, s) \in \mathbb{R}^3 : x^2 + y^2 + s^2 = 1\}$ in $\mathbb{R}^3$ and consider a copy of $\mathbb{C}$ embedded in $\mathbb{R}^3$ by identifying $\mathbb{C} = \mathbb{R}^2$ with the (x, y) -plane. Explicitly, a point $x + iy \in \mathbb{C}$ corresponds to the point $(x, y, 0) \in \mathbb{R}^3$.

Let $N = (0, 0, 1) \in S^2$ denote the ‘north pole’. For any point $v \in S^2 \setminus \{N\}$, there is a unique straight line $L_{N,v}$ passing through N and v . Since $v \neq N$, this line is not parallel to the (x, y) -plane. Hence it intersects the (x, y) -plane in a unique point $(x, y, 0)$. This corresponds to the point $x + iy \in \mathbb{C}$. We have defined a map $P : S^2 \setminus \{N\} \rightarrow \mathbb{C}$ by $P(v) = x + iy$ in the notation of the preceding discussion. The map P is called the **stereographic projection** (from the north pole).

What is the formula for stereographic projection? Let $(x, y, s) \in S^2 \setminus \{N\}$. Note that $s \neq 1$. The equation of the line passing through the point $(x, y, s) \in S^2$ and the North Pole $N = (0, 0, 1) \in S^2$ is given by

$$\gamma(t) = N + \left(\begin{pmatrix} x \\ y \\ s \end{pmatrix} - N \right) t = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} x \\ y \\ s-1 \end{pmatrix} t, \quad (t \in \mathbb{R}).$$

This clearly intersects the plane when $t = \frac{1}{1-s}$. Thus

$$P : (x, y, s) \mapsto \frac{x}{1-s} + \frac{iy}{1-s}.$$

It is possible to find an inverse to P : given any point $z \in \mathbb{C}$, draw the straight line passing through N and z (thinking of $\mathbb{C}$ as the (x, y) -plane as before). This straight line passes through S^2 in exactly one point. Hence P is a bijection that identifies $S^2 \setminus \{N\}$ with $\mathbb{C}$.

Now we have identified $\mathbb{C}$ with $S^2 \setminus \{N\}$ via the map P , it gives us a natural way to view the added point ∞ of $\hat{\mathbb{C}}$. It should correspond to adding back in the north pole to $S^2 \setminus \{N\}$. In other words, we should think of $\hat{\mathbb{C}}$ simply as the entire sphere S^2 !

In fact, one can show that we have the following correspondences/mappings:

<u>In S^2</u>		<u>In $\hat{\mathbb{C}}$</u>	(1.3)
N	$\longleftrightarrow$	∞	
S	$\longleftrightarrow$	0	
Equator	$\longleftrightarrow$	Unit circle $\{z \in \mathbb{C} : z = 1\}$	
(open) Southern hemisphere	$\longleftrightarrow$	$\mathbb{D} := \{z \in \mathbb{C} : z < 1\}$	
(open) Northern hemisphere	$\longleftrightarrow$	$\hat{\mathbb{C}} \setminus \overline{B}_1(0) = \hat{\mathbb{C}} \setminus \{z \in \mathbb{C} : z \leq 1\}$	

Algebraically:

$$\begin{array}{lll} (x, y, s) & \longleftrightarrow & \frac{x + iy}{1 - s} \quad (\text{Stereo. Proj.}) \\ \frac{1}{|z|^2 + 1} (2\operatorname{Re}(z), 2\operatorname{Im}(z), |z|^2 - 1) & \longleftrightarrow & z \quad (\text{Inverse Stereo. Proj.}) \end{array}$$

Remark 1.22. Note that we could have used the south pole $S = (0, 0, -1)$, rather than N , to define the projection. In that case we would have the correspondence $(x, y, s) \mapsto \frac{x + iy}{1 + s}$ (and you can check that the map $f(z) = 1/\bar{z}$ takes $\frac{x + iy}{1 - s}$ to $\frac{x + iy}{1 + s}$).

Definition 1.23. The *Riemann sphere* is the unit sphere $S^2 \subset \mathbb{R}^3$ along with the stereographic projections from the north and south pole.

Remark 1.24. Later in your studies you might learn that the Riemann sphere is a special example of a *Riemann surface*. The purpose of considering the two stereographic projection maps as part of the definition is that any point in S^2 is in the domain of one of the projections, so informally speaking, the maps allow us to think of a region nearby to any point in S^2 as a region inside $\mathbb{C}$.

Chapter 2

Metric Spaces

2.1 Metric spaces

Now we have another way of thinking about $\hat{\mathbb{C}}$ - as a sphere in $\mathbb{R}^3$ - it looks like there might be (at least!) two natural ways to measure the distance between two points z and w in the extended complex plane (or indeed in $\mathbb{C}$).

- The modulus $|z - w|$ of their difference in $\mathbb{C}$.
- The distance in $\mathbb{R}^3$ between their stereographic representatives on the sphere S^2 .

In both cases, the *set* is the same, but there is a different notion of *distance*. A **metric space** is a set together with a ‘distance’ function that satisfies certain axioms.

Definition 2.1 (Metric spaces). A metric space is a set X together with a function $d : X \times X \rightarrow \mathbb{R}_{\geq 0}$ such that for all $x, y, z \in X$

- **(D1) Positivity.** $d(x, y) \geq 0$ and $d(x, y) = 0 \iff x = y$;
- **(D2) Symmetry.** $d(x, y) = d(y, x)$;
- **(D3) Triangle inequality.** $d(x, y) \leq d(x, z) + d(z, y)$.

The function d is called a **metric** and we will often denote a metric space by (X, d) .

Examples of Metrics

1. The metric induced by the modulus function $|\cdot|$ on $\mathbb{R}$ or $\mathbb{C}$. We can define a distance function d on $\mathbb{R} \times \mathbb{R}$ or $\mathbb{C} \times \mathbb{C}$ by the formula $d(x, y) = |x - y|$. This metric satisfies (D1)-(D3) by the properties 1. and 2. of the modulus we gave just after Corollary 1.5.

2. The Euclidean norm on $\mathbb{R}^n$ or $\mathbb{C}^n$ For vectors $\mathbf{x}$ in $\mathbb{R}^n$ (or $\mathbb{C}^n$), the function

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

is a metric. It is easy to check properties (D1)–(D3). Note that this norm comes from an *inner-product*. For example, for $n = 2$, the real Euclidean norm on $\mathbb{R}^2$ comes from the usual dot product $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ and the complex Euclidean norm on $\mathbb{C}^2$ comes from the inner product $\langle z, w \rangle = z_1 \bar{w}_1 + z_2 \bar{w}_2$; that is $\|z\|_2 = \sqrt{\langle z, z \rangle}$.

3. • (a) Metrics induced from inner products in vector spaces

More generally, given any finite dimensional real vector space V with a (positive definite) inner product $\langle \cdot, \cdot \rangle$, then

$$d(v, w) := \|v - w\| = \sqrt{\langle v - w, v - w \rangle} \quad (v, w \in V)$$

is a metric. Properties (D1) and (D2) are obvious, property (D3) follows from **Cauchy-Schwarz**: $|\langle v, w \rangle| \leq \|v\| \cdot \|w\|$ - see Assignments.

- (b) Metrics induced from norms in vector spaces

Even more generally, so long as a vector space has a ‘nice’ notion of the ‘size’ of each vector, we can define a metric in the obvious way. Such a notion is in generality referred to as a *norm*:

Definition 2.2. [Norms and normed vector spaces] Given any real or complex vector space V , a function $\|\cdot\| : V \rightarrow \mathbb{R}_{\geq 0}$ is a **norm** if it satisfies (for $v, w \in V$)

- (N1) $\|v\| \geq 0$ and $\|v\| = 0 \iff v = 0$;
- (N2) $\|\lambda v\| = |\lambda| \cdot \|v\|$ for $\lambda \in \mathbb{R}$ or $\mathbb{C}$;
- (N3) $\|v + w\| \leq \|v\| + \|w\|$ (the triangle inequality).

Note that (N3) implies $\|v - w\| \geq |\|v\| - \|w\||$ (the reverse triangle inequality). A vector space equipped with a norm is called a **normed vector space**. The metric given by $d(v, w) := \|v - w\|$ then always defines a metric (it is easy to check properties (D1)–(D3)). In particular, since the modulus function on $\mathbb{C}$ is a norm, the metric we get from the modulus function comes from a norm.

4. ℓ_p -norm on $\mathbb{R}^n$ or $\mathbb{C}^n$ ($p \geq 1$)

The above suggests that a vector space could be home to many different norms (so many different metrics). But, not all norms arise from an inner product as in

(3a); for example, for vectors $\mathbf{x}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$, the function

$$\|\mathbf{x}\|_p := \sqrt[p]{\sum_{i=1}^n |x_i|^p}$$

defines a norm for every $p \geq 1$, called the ℓ_p -**norm**. But, for $p \neq 2$ (the Euclidean norm) it does not arise from an inner product. When $p = 1$, the ℓ_1 -norm is simply given by the sum of the size of the components $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$ and is sometimes referred to as the **Taxicab norm**.

5. ℓ_∞ -norm on $\mathbb{R}^n$ (or $\mathbb{C}^n$)

The function

$$\|\mathbf{x}\|_\infty := \max_{i=1, \dots, n} |x_i|$$

also defines a norm, called the ℓ_∞ -**norm** (or the **sup-norm**), thus it also defines a metric. It is in some sense the ‘limiting notion’ of the ℓ_p norms.

6. Riemannian (chordal) metric on $\hat{\mathbb{C}}$

Let $f : \hat{\mathbb{C}} \rightarrow S^2$ be the (inverse of the) stereographic projection. Then the function

$$d(z, w) := \|f(z) - f(w)\|_2 \quad (z, w \in \hat{\mathbb{C}}),$$

where $\|\cdot\|_2$ is Euclidean norm in $\mathbb{R}^3$ (so, the ℓ_2 -norm), is a metric on $\hat{\mathbb{C}}$. It is called the **Riemannian metric** (or **chordal metric**). Note that with respect to this metric, the distance from 0 to i is the same as the distance from i to ∞ , for example!

7. Discrete metric

Let X be a non-empty finite set. Then (for $x, y \in X$) the function

$$d(x, y) := \begin{cases} 0 & \text{if } x = y, \\ 1 & \text{if } x \neq y, \end{cases}$$

defines a metric, called the **discrete metric**. It is easy to check (D1)-(D3). In this case, (X, d) is called a **discrete metric space**.

8. Function spaces

There are many of these, such as the space $X = C([a, b])$ of continuous functions on an interval $[a, b]$. The function

$$\|f\| := \max_{x \in [a, b]} |f(x)|$$

defines a norm, and thus a metric (see Analysis III for more examples).

9. Subspace metric

Any non-zero subset $Y \subset X$ of a metric space X is itself a metric space with respect to the same metric (this is easy to check). Unless mentioned otherwise, this will always be the metric we will use on Y .

2.2 Open and closed sets

Since we have a general notion of distance in any metric space X , we can define balls in the space. These will be the key to understanding the topology of the space.

Definition 2.3 (Balls in a metric space). Let (X, d) be a metric space, $x \in X$ and let $r > 0$ be a real number. Then:

- The **open ball** $B_r(x)$ of radius r centred at x is

$$B_r(x) := \{y \in X : d(x, y) < r\}.$$

- The **closed ball** $\bar{B}_r(x)$ of radius r centred at x is

$$\bar{B}_r(x) := \{y \in X : d(x, y) \leq r\}.$$

Visualizing balls (examples)

1. Let $X = \mathbb{C}$ and $d(z, w) = |z - w|$, then $B_1(0) = \mathbb{D} = \{z : |z| < 1\}$ as before. More generally $B_r(z_0)$ is the usual ball of radius r around z_0 , not including its boundary circle. $\bar{B}_r(z_0)$ is the ball of radius r around z_0 , including its boundary circle. This is the most important example from the point of view of Complex Analysis.
2. Let us consider the unit ball $B_1(\mathbf{0})$ in $\mathbb{R}^2$ with respect to the ℓ_p -norms, for $p = 1, 2$ and ∞ .

For $p = 2$ the unit ball $B_1(\mathbf{0})$ is the usual Euclidean ball - so the inside of the unit circle centred at the origin. For $p = \infty$, the equation $\max\{|x|, |y|\} < 1$ (for $(x, y) \in \mathbb{R}^2$) clearly defines the interior of a square with vertices $(1, 1), (-1, 1), (1, -1)$ and $(-1, -1)$.

For the ℓ_1 -norm a little care is needed. We are interested in the points $(x, y) \in \mathbb{R}^2$ with $|x| + |y| < 1$. In the 1st quadrant this means $y < 1 - x$, in the 2nd it means $y < 1 + x$, in the 3rd we have $y > -1 - x$ and in the 4th its $y > x - 1$. Thus, the unit ball is the interior of a diamond with vertices $(1, 0), (0, 1), (-1, 0)$ and $(0, -1)$.

Definition 2.4. [Open/closed sets in a metric space] Let (X, d) be a metric space. Then:

- A subset $U \subseteq X$ is **open (in X)** if for every point $x \in U$ there exists $\epsilon > 0$ such that $B_\epsilon(x) \subset U$.
- A subset $U \subseteq X$ is **closed (in X)** if its complement $U^c := X \setminus U$ is open.

Remark 2.5. Sets in a metric space can be open and closed at the same time! For example, the empty set $\emptyset$ and the whole metric space X are both open and closed. Such sets are referred to as ‘*clopen*’.

Lemma 2.6. [Open balls are open] In a metric space, the open ball $B_r(x)$ is open!

Proof. Let $y \in B_r(x)$ with $d(x, y) = s$ (and so $s < r$). We need to show there exists $\epsilon > 0$ such that $B_\epsilon(y) \subseteq B_r(x)$. Simply take $\epsilon = r - s > 0$. Then for every $z \in B_\epsilon(y)$ we have

$$d(x, z) \stackrel{(D3)}{\leq} d(x, y) + d(y, z) < s + \epsilon = r.$$

Thus, $z \in B_r(x)$ as required. $\square$

Remark 2.7. It can also be shown that in a metric space the closed ball $\overline{B}_r(x)$ is closed (see Assignments).

Open sets (examples/warnings)

1. All of the previously encountered subsets $\mathbb{H}, \mathbb{D}, \mathbb{C}^*$ and $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$ of the complex plane are open (see Assignments). The 1st quadrant $\Omega_1 := \{z \in \mathbb{C} : \operatorname{Re}(z) > 0, \operatorname{Im}(z) > 0\}$ is open. To see this, for $z \in \Omega_1$ consider the ball $B_r(z)$ where $r = \min(\operatorname{Re}(z), \operatorname{Im}(z))/2$.
2. Let X be a discrete metric space, so $d(x, y) := \begin{cases} 0 & \text{if } x = y. \\ 1 & \text{if } x \neq y. \end{cases}$ Then, for $x \in X$ and $r > 0$ we have

$$B_r(x) := \begin{cases} \{x\} & \text{if } r \leq 1. \\ X & \text{if } r > 1. \end{cases}$$

Therefore, (by Lemma 2.6) every singleton $\{x\}$ is an open set with respect to the discrete metric. Moreover, the complement $X \setminus \{x\}$ is also open, since for any $y \in X \setminus \{x\}$ (that is, any $y \neq x$ in X) and any $r < 1$ the open ball $B_r(y) = \{y\}$ is contained in $X \setminus \{x\}$. Thus, all balls are clopen with respect to the discrete metric! In fact, any subset $Y \subseteq X$ of a discrete metric space is clopen!

3. Sets don't have to be either open or closed. For example, $[0, 1)$ is neither open nor closed in $\mathbb{R}$ (with respect to the standard metric $|\cdot|$) - simply check the point $x = 0$ in $[0, 1)$ and the point $x = 1$ in the complement $(-\infty, 0) \cup [1, \infty)$. However, recall that any subset of a metric space is itself a metric space (w.r.t the same metric) - the *subspace metric*. Thus the pair $([0, 1), |\cdot|)$ is a metric space - but then (by the remark after Definition 2.4) the set $[0, 1)$ is open!

Key: Open and closed sets are really relative notions, depending on the ambient space (as well as the metric).

Notation: When we say a subset of $\mathbb{R}$ or $\mathbb{R}^n$ or $\mathbb{C}$ are open/closed, we will mean with respect to the standard norms $|\cdot|$ and $\|\cdot\|_2$ and $|\cdot|$ respectively. Most sets we encounter do not simply look like open/closed balls, so it will be useful to have rules for union and intersection:

Lemma 2.8. [*Unions and intersections of open sets*] Let (X, d) be a metric space. Then:

1. Arbitrary unions of open sets are open; that is

$$\bigcup_{i \in I} U_i \text{ is open, for any (possibly infinite) collection of open sets } U_i.$$

2. Finite intersections of open sets are open; that is

$$\bigcap_{i=1}^n U_i \text{ is open, for any finite collection of open sets } U_i.$$

Proof. 1. Let $x \in \bigcup_{i \in I} U_i$. Then, by definition, it must be contained in the set U_j for some $j \in I$. Since U_j is open there must exist a ball $B_\epsilon(x)$ centred at x lying in U_j . But then $B_\epsilon(x) \subseteq U_j \subseteq \bigcup_{i \in I} U_i$ as required.

2. Let $x \in \bigcap_{i=1}^n U_i$. By definition $x \in U_i$ for every $i = 1, \dots, n$. But, since they are all open, for every U_i there must exist $r_i > 0$ such that $B_{r_i}(x) \subset U_i$. Now simply take $\epsilon = \min_{i=1, \dots, n} (r_i)$. Then for every i we have $B_\epsilon(x) \subseteq B_{r_i}(x)$ and so

$$B_\epsilon(x) \subseteq \bigcap_{i=1}^n B_{r_i}(x) \subseteq \bigcap_{i=1}^n U_i.$$

□

Corollary 2.9 (Unions and intersections of closed sets). *Let (X, d) be a metric space. Then:*

1. *Finite unions of closed sets are closed.*
2. *Arbitrary intersections of closed sets are closed.*

Proof. De Morgan's laws together with Lemma 2.8. □

We have the following remarks about open and closed sets.

- An infinite intersection of open sets is not necessarily open (see Assignments). Similarly, an infinite union of closed sets is not necessarily closed: e.g., the union of closed intervals in $\mathbb{R}$,

$$\bigcup_{i=1}^{\infty} \left[\frac{1}{i}, 1 - \frac{1}{i} \right] = (0, 1) \quad \text{is open in } \mathbb{R}.$$

- The next generalisation of a metric space you will encounter (see Topology III) is called a *Topological space* $\mathcal{T}$. There, the only stipulations are the existence of open sets such that

$$(i) \emptyset \text{ and } \mathcal{T} \text{ are open ;} \quad (ii) \text{ Lemma 2.8 holds.}$$

We have a hierarchy:

Inner-product space $\implies$ Normed space $\implies$ Metric space $\implies$ Topological space.

- Why have we been looking at examples in $\mathbb{R}^2$ rather than $\mathbb{C}$? It turns out that both these spaces are ‘topologically equivalent’, that is, they have the same open sets - this is obvious since the complex modulus is essentially just the Euclidean norm on $\mathbb{R}^2$.

As we have seen, some sets are neither open nor closed. It will be useful to ask what the largest possible open set is inside a given set. Similarly, what is the smallest closed set containing a given set?

Definition 2.10 (Interior points, closure, boundary, exterior). Let A be a subset of a metric space (X, d) .

- The **interior** A^0 of A is defined by

$$A^0 := \{x \in A : \text{there exists an open set } U \subseteq A \text{ such that } x \in U\}.$$

- The **closure** $\bar{A}$ of A is the complement of the interior of the complement:

$$\bar{A} := \left((A^c)^0 \right)^c = \{x \in X : U \cap A \neq \emptyset \text{ for every open set } U \text{ with } x \in U\}.$$

- The **boundary** ∂A of A is the closure without the interior:

$$\partial A := \bar{A} \setminus A^0 \quad \left[= (A^0)^c \cap \left((A^c)^0 \right)^c = \left(A^0 \cup (A^c)^0 \right)^c \right]$$

- Don't confuse closure with conjugation! Closure concerns sets in any metric space, conjugation concerns points in $\mathbb{C}$.
- Clearly the interior and exterior are open and clearly the boundary is closed. The closure is also closed (see Assignments) - in fact, it is often defined more simply as $\bar{A} := X \setminus (X \setminus A)^0$, from which the closedness is obvious - it is the complement of an open set.
- The boundary matches our naive notion. Broadly speaking, the interior of a set consists of all the points that are not on its 'edge', and to form the closure of a set you simply add all the missing edge points. Indeed, we have the following additional properties of a subset $A \subset X$ (see Assignments):

$$\begin{aligned} (a) \ A \text{ is open} &\iff \partial A \cap A = \emptyset \iff A = A^0; & \text{In fact } A^0 &= \bigcup_{\substack{U \subseteq A \\ U \text{ open}}} U; \\ (b) \ A \text{ is closed} &\iff \partial A \subseteq A \iff A = \bar{A}; & \text{In fact } \bar{A} &= \bigcap_{\substack{A \subseteq F \\ F \text{ closed}}} F; \end{aligned}$$

That is, the interior A^0 is the largest open set contained in A and the closure $\bar{A}$ is the smallest closed set containing A . Convince yourself that all the definitions reflect your intuitive notions, say, for the plane!

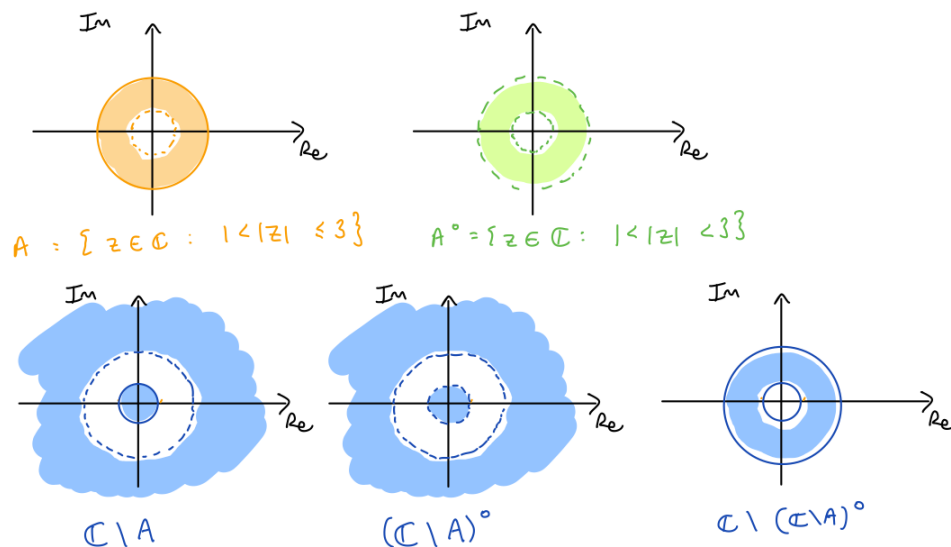
- There is a nice "geometric" description for the boundary of a set

$$\partial A = \{x \in X : \text{for all open sets } U \text{ containing } x, \text{ there exist } y, z \in U \text{ with } y \in A \text{ and } z \in A^c\}.$$

- In $\mathbb{R}^n$ or $\mathbb{C}^n$ for simple sets we only have to replace strict inequality with equality (or vice versa) to obtain the closure (or interior). For example, for $A = \{z \in \mathbb{C} : 1 < |z| \leq 3\}$, we have $A^0 = \{z \in \mathbb{C} : 1 < |z| < 3\}$, $\bar{A} = \{z \in \mathbb{C} : 1 \leq |z| \leq 3\}$, and $\partial A = \{z \in \mathbb{C} : |z| = 1\} \cup \{z \in \mathbb{C} : |z| = 3\}$ (see Figure 2.1).

Similarly

$$\overline{\{z \in \mathbb{C} : 1 < \operatorname{Re}(z) \leq 3, |\operatorname{Im}(z)| < 1\}} = \{z \in \mathbb{C} : 1 \leq \operatorname{Re}(z) \leq 3, |\operatorname{Im}(z)| \leq 1\}$$


 Figure 2.1: The set A with its interior and closure.

and

$$\{z \in \mathbb{C} : 1 < \text{Re}(z) \leq 3, |\text{Im}(z)| < 1\}^0 = \{z \in \mathbb{C} : 1 < \text{Re}(z) < 3, |\text{Im}(z)| < 1\}.$$

In fact, $\overline{B_r(x)} = \bar{B}_r(x)$ for any ball in $\mathbb{R}^n$ or $\mathbb{C}^n$.

However, this is not true in every metric space (see Assignments) - there are metric spaces for which $\overline{B_r(x)} \neq \bar{B}_r(x)$; that is, the smallest closed set containing the open ball $B_r(x)$ is not the closed ball $\bar{B}_r(x)$! [Hint: what if the open ball is already closed!?!]

2.3 Convergence and continuity

With our general notion of distance (a metric) comes a natural notion of convergence.

Definition 2.11. [Limits and convergence in a metric space] We say a sequence $\{x_n\}$ in a metric space (X, d) **converges to** $x \in X$ if we have

$$\lim_{n \rightarrow \infty} d(x_n, x) = 0.$$

That is,

for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for every $n > N$.

We write “ $x_n \rightarrow x$ as $n \rightarrow \infty$ ”, or “ $\lim_{n \rightarrow \infty} x_n = x$ ”.

Convergent sequences (example)

As mentioned, the chordal metric on $\hat{\mathbb{C}}$ is $d(z, w) = \|f(z) - f(w)\|_2$, where $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{R}^3$ and f is the inverse Stereographic projection given by

$$f(z) = \left(\frac{2\operatorname{Re}(z)}{|z|^2 + 1}, \frac{2\operatorname{Im}(z)}{|z|^2 + 1}, \frac{|z|^2 - 1}{|z|^2 + 1} \right).$$

Show that with respect to this metric the sequence $\{ki\}_{k \in \mathbb{N}}$ in $\hat{\mathbb{C}}$ converges to $\infty \in \hat{\mathbb{C}}$.

Since $|ki| = k$ and $f(\infty) = (0, 0, 1)$ we have

$$\begin{aligned} d(ki, \infty) &= \|f(ki) - f(\infty)\|_2 = \left\| \left(0, \frac{2k}{k^2 + 1}, \frac{k^2 - 1}{k^2 + 1} \right) - (0, 0, 1) \right\|_2 \\ &= \left\| \left(0, \frac{2k}{k^2 + 1}, -\frac{2}{k^2 + 1} \right) \right\|_2 \\ &= \sqrt{\left(\frac{2k}{k^2 + 1} \right)^2 + \left(\frac{-2}{k^2 + 1} \right)^2} \rightarrow 0 \quad \text{as } k \rightarrow \infty. \end{aligned}$$

Thus, the sequence indeed converges to ∞ . This is quite an odd notion as we are used to saying sequences ‘diverge’ if they tend to infinity. The key is that convergence depends on the metric being used.

Limits in $\mathbb{C}$ with the standard metric (Very important!).

The above definition of limit in a metric space also gives us an example of limits in $\mathbb{C}$ with the standard metric. This says that if z_n is a sequence of complex numbers, then $\lim_{n \rightarrow \infty} z_n = z$ if and only if

‘for all $\epsilon > 0$, there exists $N > 0$ such that for all $n \geq N$, $|z_n - z| < \epsilon$ ’.

Note this is the same definition as in Analysis I, but replacing the absolute value on the real line by the modulus on $\mathbb{C}$. Importantly, by the same proofs as in Analysis I, **limits in the complex plane follow the COLT rules.**

Furthermore there is a very important link between convergence in the complex plane and real convergence (see Analysis I for the proof). Let $\{z_n\}_{n \in \mathbb{N}}$ be a sequence of complex numbers $z_n = x_n + iy_n$. Then, for any fixed $z_0 = x_0 + iy_0 \in \mathbb{C}$ we have

$$\lim_{n \rightarrow \infty} z_n = z_0 \iff \lim_{n \rightarrow \infty} x_n = x_0 \quad \text{and} \quad \lim_{n \rightarrow \infty} y_n = y_0.$$

In other words, the sequence $\{z_n\}$ converges iff the real sequences $\{\operatorname{Re}(z_n)\}$ and $\{\operatorname{Im}(z_n)\}$ converge (see Assignments).

Let’s return to our general setting of metric spaces and prove some properties of limits.

Lemma 2.12. *[Limits and open sets] Let (X, d) be a metric space. Then:*

1. A sequence can have at most one limit.
2. We have

$$\lim_{n \rightarrow \infty} x_n = x \iff \forall \text{ open } U \text{ with } x \in U, \exists N \in \mathbb{N} \text{ such that } \forall n > N \quad x_n \in U.$$

Hence the notion of a limit in a metric space can be stated in terms only of its open sets.

Proof. 1. Assume $\lim_{n \rightarrow \infty} x_n = x$ and $\lim_{n \rightarrow \infty} x_n = y$. Then we have for each n , by the triangle inequality $d(x, y) \leq d(x, x_n) + d(x_n, y)$, so taking the limit as $n \rightarrow \infty$ gives

$$d(x, y) \leq \lim_{n \rightarrow \infty} d(x, x_n) + \lim_{n \rightarrow \infty} d(x_n, y) = 0 + 0 = 0,$$

so $d(x, y) = 0$, hence $x = y$ by property (D1) of metric spaces.

2. ($\Rightarrow$): Assume $\lim_{n \rightarrow \infty} x_n = x$ and that U is open with $x \in U$. By definition there exists $r > 0$ such that $B_r(x) \subseteq U$ and an $N \in \mathbb{N}$ such that $d(x_n, x) < r$ for every $n > N$. Thus, $x_n \in B_r(x) \subseteq U$ for every $n > N$.

($\Leftarrow$): Let $\epsilon > 0$. We wish to find $N \in \mathbb{N}$ such that $d(x_n, x) < \epsilon$ for $n > N$. Consider the ball $B_\epsilon(x)$. It is open and contains x , and so there exists $N \in \mathbb{N}$ such that $x_n \in B_\epsilon(x)$ for every $n > N$. This is precisely the statement that $d(x_n, x) < \epsilon$ for $n > N$. □

Remark 2.13. The key to completing proofs of this type is to write down the definitions in your assumptions and also write precisely what you need to prove. Usually, doing this leads very quickly to the proof.

Sequences in metric spaces also give a nice criterion to check if a set is closed. We will show soon that $F \subseteq X$ is closed if and only if every sequence in F which converges in X has its limit point in F . That is, if $\{x_n\}_{n \in \mathbb{N}}$ is in F and $\lim_{n \rightarrow \infty} x_n = x$ for some $x \in X$ then $x \in F$.

We can now define what it means for a function between two metric spaces to be continuous (this will incorporate many of the functions we have already encountered; e.g., $f : \mathbb{C} \rightarrow \mathbb{C}$; $\mathbb{R} \rightarrow \mathbb{C}$; or $\mathbb{C} \rightarrow \mathbb{R}$).

Definition 2.14 (Continuity). A map $f : (X_1, d_1) \rightarrow (X_2, d_2)$ between two metric spaces is called **continuous at** $x_0 \in X_1$ if

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad \text{such that} \quad \forall x \in X_1 \quad \text{we have} \quad d_1(x, x_0) < \delta \implies d_2(f(x), f(x_0)) < \epsilon.$$

We say a function f is **continuous on** X_1 if it is continuous at every point $x_0 \in X_1$. [Note that this is exactly the same as in Analysis I.]

Remark 2.15. Equivalently, one could write

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad \text{such that} \quad x \in B_\delta(x_0) \Rightarrow f(x) \in B_\epsilon(f(x_0)),$$

where the first ball is in X_1 and the second ball is in X_2 .

Lemma 2.16. *[Basic properties of continuous functions]*

1. *Products, sum, quotients of real/complex valued continuous functions on a metric space X are continuous. E.g., if $f : X \rightarrow \mathbb{C}$ and $g : X \rightarrow \mathbb{C}$ are continuous, then $f + g$ and fg and f/g are continuous (where defined).*
2. *Compositions of continuous functions are continuous. I.e., if $f : X_1 \rightarrow X_2$ and $g : X_2 \rightarrow X_3$ are continuous maps between metric spaces, then $g \circ f : X_1 \rightarrow X_3$ is continuous.*

Proof. Almost word-for-word from Analysis I. □

Examples of continuous functions on the complex plane (with the standard metric)

- The identity function is continuous.
- Constant functions are continuous.
- The functions $\text{Re}, \text{Im} : \mathbb{C} \rightarrow \mathbb{R}$ are continuous.
- The complex conjugation $z \mapsto \bar{z}$ is continuous as a map from $\mathbb{C} \rightarrow \mathbb{C}$.
- The modulus function $z \mapsto |z|$ is continuous as a map from $\mathbb{C} \rightarrow \mathbb{R}$.
- All of $\exp, \sin, \cos, \sinh$ and $\cosh$ are continuous on $\mathbb{C}$, as are all polynomials.
- If $\arg$ is the choice of argument function with values in $(\theta_1, \theta_2]$ then $\arg$ is continuous on $\mathbb{C} \setminus R_{\theta_1}$ (recall R_{θ_1} is the ray with angle θ_1)
- If $\log$ is a branch of $\log$ corresponding to an argument function as above, then $\log$ is continuous on $\mathbb{C} \setminus R_{\theta_1}$.

As with limits, it will be useful to restate continuity in terms of open sets. First, recall that for any function $f : X_1 \rightarrow X_2$ and any set $U \subseteq X_2$ we define the **preimage** $f^{-1}(U)$ **of U under f** by $f^{-1}(U) := \{x \in X_1 : f(x) \in U\}$.

Theorem 2.17. *[Continuity via open sets] Let X_1 and X_2 be metric spaces. Then:*

$$\begin{aligned} f : X_1 \rightarrow X_2 \text{ continuous} &\iff f^{-1}(U) \text{ is open in } X_1 \text{ for every open set } U \text{ in } X_2 \\ &\iff f^{-1}(F) \text{ is closed in } X_1 \text{ for every closed set } F \text{ in } X_2. \end{aligned}$$

Proof. For open sets:

- $(\Rightarrow)$: Let U be open in X_2 , and pick $x \in f^{-1}(U)$. Since U is open there exists $\epsilon > 0$ such that $B_\epsilon(f(x)) \subseteq U$ (since $f(x) \in U$). But f is continuous, so by definition there exists $\delta > 0$ such that if $y \in B_\delta(x)$ we have $f(y) \in B_\epsilon(f(x))$. Thus $f(y) \in U$, and so $y \in f^{-1}(U)$; since this is true for every $y \in B_\delta(x)$ we have shown $B_\delta(x) \subseteq f^{-1}(U)$ and so the set $f^{-1}(U)$ is open.
- $(\Leftarrow)$: Let $x \in X_1$ and $\epsilon > 0$. We need to find a $\delta > 0$ such that $y \in B_\delta(x) \Rightarrow f(y) \in B_\epsilon(f(x))$. First notice that (by Lemma 2.6) the ball $B_\epsilon(f(x))$ is open. By assumption, the preimage $f^{-1}(B_\epsilon(f(x)))$ of this ball is also open. The point x must be in this preimage (because the centre $f(x)$ is certainly in $B_\epsilon(f(x))$). Then, since the preimage is open there must exist an open ball $B_\delta(x)$ around x contained in $f^{-1}(B_\epsilon(f(x)))$. But, this is precisely the statement that $y \in B_\delta(x) \Rightarrow f(y) \in B_\epsilon(f(x))$ as required.

For closed sets see Assignments. □

Remark 2.18. • From the proof we see that we can be slightly more precise than the statement of the theorem. We have, for example

$$f : X_1 \rightarrow X_2 \text{ continuous} \iff f^{-1}(U) \text{ is open in } X_1 \text{ for every open set } U \text{ in } X_2$$

at $x \in X_1$ containing $f(x)$

- Note that for the $(\Leftarrow)$ direction we only actually needed the fact that $f^{-1}(B)$ was open for any open ball B in X_2 . It turns out that the open balls ‘generate’ all the open sets (via Lemma 2.8) - see Topology/Analysis III!
- Note that, by the second part of Theorem 2.17, if a function $f : X_1 \rightarrow X_2$ is continuous then $f^{-1}(\{x\})$ is closed for any $x \in X_2$.

As with limits, it turned out that continuity depends only upon the open sets in the respective metric spaces. This means we can use the continuity of known functions to prove the openness of very complicated sets.

Showing sets are open using continuity (examples)

- Show the following set is open:

$$U = \{(x, y) \in \mathbb{R}^2 : (x^2 + y^2) \sin^3(\sqrt{x^2 + 7}) > 2\}.$$

Well, the function

$$f : \mathbb{R}^2 \rightarrow \mathbb{R} : (x, y) \mapsto (x^2 + y^2) \sin^3(\sqrt{x^2 + 7})$$

is continuous by Lemma 2.16, because it is the product/composition of real valued continuous functions. Moreover,

$$U = \{(x, y) \in \mathbb{R}^2 : f((x, y)) > 2\} = f^{-1}((2, \infty)).$$

Since $(2, \infty)$ is open in $\mathbb{R}$ (see Assignments), the set U is the preimage of an open set under a continuous map and by Theorem 2.17 it is therefore open.

- We can actually do a little more using the following useful properties of the preimage (from Analysis I):

Useful properties of preimage

- $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.
- $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.
- $f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B)$.

Show the following set is open:

$$U = \{(x, y) \in \mathbb{R}^2 : xy > 1, x^2 + y^2 > 3\}.$$

Both

$$f(x, y) = xy \quad \text{and} \quad g(x, y) = x^2 + y^2$$

are continuous as functions $\mathbb{R}^2 \rightarrow \mathbb{R}$, and $U = f^{-1}((1, \infty)) \cap g^{-1}((3, \infty))$.

Since f and g are continuous and both $(1, \infty)$ and $(3, \infty)$ are open in $\mathbb{R}$, the preimages $f^{-1}((1, \infty))$ and $g^{-1}((3, \infty))$ are open (by Theorem 2.17). By Lemma 2.8 we have that U is open, since it is the intersection of two open sets.

Showing functions are not continuous using open sets (example)

- We can also use openness to prove a function is not continuous. Indeed, for $f : X_1 \rightarrow X_2$, if there exists an open set U in X_2 such that $f^{-1}(U)$ is not open in X_1 , then f is not continuous. For example, the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} \frac{xy}{x^2 + y^2}, & \text{if } (x, y) \neq 0, \\ 0 & \text{otherwise,} \end{cases}$$

isn't continuous at $(0, 0)$. Why? Consider the preimage $f^{-1}((-\epsilon, \epsilon))$. **Claim:** This preimage is not open for ϵ sufficiently small.

First note that the preimage in question contains $(0, 0)$ since $f((0, 0)) = 0$. To show the preimage is not open it is enough to show that any open ball in $\mathbb{R}^2$ centred at $(0, 0)$ is not contained in $f^{-1}((-\epsilon, \epsilon))$: Let $\epsilon < 1/4$, say, and for any

$\delta > 0$ consider the ball $B_\delta((0,0))$ centred at $(0,0)$. The point $(\delta/2, \delta/2)$ is in $B_\delta((0,0))$ since $\|(\frac{\delta}{2}, \frac{\delta}{2}) - (0,0)\|_2 = \|(\frac{\delta}{2}, \frac{\delta}{2})\|_2 = \sqrt{\frac{\delta^2}{4} + \frac{\delta^2}{4}} = \frac{\delta}{\sqrt{2}} < \delta$. But

$$f\left(\left(\frac{\delta}{2}, \frac{\delta}{2}\right)\right) = \frac{\frac{\delta}{2} \cdot \frac{\delta}{2}}{\frac{\delta^2}{4} + \frac{\delta^2}{4}} = \frac{1}{2} > \epsilon,$$

so $(\delta/2, \delta/2)$ is not in $f^{-1}((-\epsilon, \epsilon))$. Thus, for any $\delta > 0$ the ball $B_\delta((0,0))$ is not contained in $f^{-1}((-\epsilon, \epsilon))$ and so this preimage is not open.

Since $(-\epsilon, \epsilon)$ is open in $\mathbb{R}$ it follows from Theorem 2.17 that f is not continuous.

Why preimages?

Note that the use of preimages in Theorem 2.17, rather than images, is important. The same result is not true of images. E.g., the function $f(z) = |z|$ is continuous as a function $\mathbb{C} \rightarrow \mathbb{R}$, but it maps an open set in the complex plane $f : \mathbb{D} \rightarrow [0,1)$ to an interval that is neither open nor closed in $\mathbb{R}$.

Note that this f is actually a bijection from $\mathbb{R}_{\geq 0}$ to $\mathbb{R}_{\geq 0}$, and in $\mathbb{R}_{\geq 0}$ the interval $[0,1)$ is open! So, is the problem that we need the function to be bijective? No. For example, consider the metric spaces $X_1 = [0,1) \cup [2,3]$ and $X_2 = [0,2]$ with the usual (subspace) metric coming from the absolute value on $\mathbb{R}$. Define

$$f : X_1 \rightarrow X_2 : x \mapsto \begin{cases} x, & \text{if } x \in [0,1). \\ x-1, & \text{if } x \in [2,3]. \end{cases}$$

It is easy to check that f is a bijection and is continuous on its domain: (*Continuity is trivial for $x \neq 2$. For $x = 2$, pick $\epsilon > 0$, then for any $0 < \delta < 1$ we have $B_\delta(2) = [2, 2+\delta)$. Note that $f(2) = 1$ and so $B_\epsilon(f(2)) = (1-\epsilon, 1+\epsilon)$. To show f is continuous we must therefore find a δ so that $f(x) \in (1-\epsilon, 1+\epsilon)$ if $x \in [2, 2+\delta)$. Simply pick any $\delta < \epsilon$, for then:*

$$x \in [2, 2+\delta) \Rightarrow f(x) = x-1 \in [1, 1+\delta) \subset (1-\delta, 1+\delta) \subset (1-\epsilon, 1+\epsilon),$$

as required.) But, the set $[2,3]$ is open in X_1 (see Assignments) and its image $f([2,3]) = [1,2]$ is not open in X_2 .

Thus, we genuinely do need to use preimages. When can we use images of continuous functions to preserve properties of the sets in question? When can we find the maximum/minimum value taken by a function on a set? It turns out a key concept is that of sequential *compactness*.

2.4 Sequential Compactness and Compactness

Definition 2.19 (Sequential Compactness). A non-empty subset K of a metric space X is called **sequentially compact** if for any sequence $\{x_k\}_{k \in \mathbb{N}}$ in K there exists a convergent subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ with limit in K .

Note that the initial sequence in the definition does not have to converge. But what if it does? What is the link between a convergent sequence and its subsequences?

Lemma 2.20. *If $\{x_k\}_{k \in \mathbb{N}}$ is a convergent sequence in a metric space X , then any subsequence converges to the same limit.*

Proof. Assume $x_k \rightarrow x$. Let $\{x_{n_k}\}_{k \in \mathbb{N}}$ be a subsequence (here $n_1 < n_2 < \dots$). We know that for every $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $x_k \in B_\epsilon(x)$ for every $k > N$. So simply note that $n_k > k$, whence $x_{n_k} \in B_\epsilon(x)$ for $k > N$. Thus, $x_{n_k} \rightarrow x$. $\square$

It seems quite daunting to have to check every sequence in a set for convergent subsequences, so it will be useful to re-express sequential compactness in terms of our basic building blocks; that is, open and closed sets. However, exploring openness doesn't seem particularly fruitful; for example, consider the set $(0, 1)$ in $\mathbb{R}$; the sequence $\{1/n\}$ lies in $(0, 1)$, but its limit is $0 \notin (0, 1)$. Moreover, any subsequence must have the same limit 0 by Lemma 2.20, and so $(0, 1)$ is not sequentially compact. Let's try closedness.

Proposition 2.21. *[Closed sets and limits of sequences] We have*

$F \subset X$ is closed $\iff$ Every sequence in F which converges in X has its limit point in F .
(that is, if $x_n \in F$ and $x_n \rightarrow x$ for some $x \in X$, then $x \in F$.)

Proof.

- $(\Rightarrow)$: Assume F is closed and let $\{x_n\}_{n \in \mathbb{N}}$ be a sequence with $x_n \in F$ that converges to $x \in X$. We wish to show $x \in F$. For a contradiction, assume $x \notin F$; that is, assume $x \in X \setminus F$. We know $X \setminus F$ is open, so by definition there must exist an open ball $B_\epsilon(x)$ centred at x and contained in $X \setminus F$. But $x_n \rightarrow x$, so there exists $N \in \mathbb{N}$ such that $x_n \in B_\epsilon(x)$ for $n > N$. Thus $x_n \in X \setminus F$ for $n > N$ (that is, $x_n \notin F$ for $n > N$), which is a contradiction.
- $(\Leftarrow)$: We need to show $X \setminus F$ is open. Let $x \in X \setminus F$. We wish to show we can always find a ball $B_\epsilon(x)$ centred at x and contained in $X \setminus F$. If for some $n \in \mathbb{N}$ we have $B_{1/n}(x) \subseteq X \setminus F$ we are done. Otherwise, pick an element $x_n \in B_{1/n}(x) \cap F$ for each n . But then $x_n \rightarrow x$ and $x_n \in F$, but by assumption $x \notin F$. This contradicts the premise of the proof.

$\square$

Corollary 2.22. *[Relationship between sequential compactness and closedness]*

1. *Sequentially compact sets are closed.*
2. *Any closed subset of a sequentially compact subset is sequentially compact.*

Proof. 1. If F is sequentially compact and $\{x_k\}_{k \in \mathbb{N}}$ is a convergent sequence with $x_k \in F$ and limit $x \in X$, then by definition there must exist a convergent subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ with limit $x_0 \in F$. But, by Lemma 2.20, these limits must be the same; i.e., $x = x_0$. Thus $x \in F$ and by Proposition 2.21 the set F is closed.

2. Assume $F \subseteq K$ is closed and K is sequentially compact. Let $\{x_k\}_{k \in \mathbb{N}}$ be any sequence in F . Since each $x_k \in K$ there must exist a convergent subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ with limit $x \in K$. But, by Proposition 2.21 this limit must be in F . This shows F is sequentially compact. □

Are all closed sets sequentially compact? No: e.g., $[0, \infty)$ is closed in $\mathbb{R}$, but $x_n = n$ has no convergent subsequence. The problem here is that $[0, \infty)$ is unbounded.

Definition 2.23 (Bounded sets). A subset $A \subseteq X$ of a metric space X is **bounded** if

there exists $R > 0$ and $x \in X$ such that $A \subseteq B_R(x)$.

Lemma 2.24. *[Sequentially compact sets are bounded] Let $K \subseteq X$ be a sequentially compact subset of a metric space X . Then K is bounded.*

Proof. We employ a contrapositive argument. Assume K is not bounded and fix $x \in K$. For each $k \in \mathbb{N}$ we can therefore find a point $x_k \in K$ such that $d(x_k, x) \geq k$ (since $B_k(x)$ does not contain K). However, the sequence $\{x_k\}_{k \in \mathbb{N}}$ cannot have a convergent subsequence; given any potential limit point x_0 of x_{n_k} we have

$$d(x_{n_k}, x_0) \stackrel{(D3)}{\geq} d(x_{n_k}, x) - d(x, x_0) \geq n_k - d(x, x_0) \rightarrow \infty \text{ as } k \rightarrow \infty.$$

□

So, we know that sequentially compact sets are closed and bounded. Is that enough? Recall, the theorem of **Bolzano-Weierstrass** from Analysis I. It states precisely that the closed and bounded intervals $[a, b]$ in $\mathbb{R}$ are sequentially compact.

Theorem 2.25. *[Heine-Borel for $\mathbb{R}^n$]*

A subset K of $\mathbb{R}^n$ is sequentially compact $\iff K$ is closed and bounded.

Remark 2.26. Heine-Borel does not hold for arbitrary metric spaces.

The proof is simply induction on n , starting from the base ‘ $n = 1$ ’ case for subsets of $\mathbb{R}$. For this reason (and since this is a Complex Analysis course), we will prove it for $\mathbb{C}$ and leave the proof in higher dimensions for the enthusiastic reader - the statement for $\mathbb{C}$ corresponds to that of $\mathbb{R}^2$ in the statement of Theorem 2.25, since $\mathbb{C}$ with its standard metric can be identified with $\mathbb{R}^2$ with its standard metric.

Theorem 2.27 (Heine-Borel for $\mathbb{C}$).

A subset K of $\mathbb{C}$ is sequentially compact $\iff K$ is closed and bounded.

Proof. • $(\Rightarrow)$: This has been done via Corollary 2.22 and Lemma 2.24.

- $(\Leftarrow)$: Applying Bolzano-Weierstrass, we see that Heine-Borel holds for subsets of $\mathbb{R}$.

Next, let $K \subseteq \mathbb{C}$ be a closed and bounded subset of $\mathbb{C}$. It is very easy to show that if K is bounded then there exists $R > 0$ such that $K \subseteq B_R(0)$. Let $R > 0$ be such a number (so that $|z| < R$ for every $z \in K$) and let $\{z_k\}_{k \in \mathbb{N}}$ be a sequence in $\mathbb{C}$ with $z_n = x_n + iy_n$. Note that $|x_n| < R$ and $|y_n| < R$. We wish to show there is a convergent subsequence with limit in K .

The real interval $[-R, R]$ is closed and bounded, so by Heine-Borel for $\mathbb{R}$ the interval $[-R, R]$ is sequentially compact. Notice that $x_k \in (-R, R) \subset [-R, R]$. Since $\{x_k\}_{k \in \mathbb{N}}$ is a sequence in $[-R, R]$ there must be a convergent subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ with limit in $[-R, R]$, say $x_{n_k} \rightarrow x \in [-R, R]$.

Consider the corresponding complex subsequence $\{z_{n_k}\}_{k \in \mathbb{N}}$, where $z_{n_k} = x_{n_k} + iy_{n_k}$. Its imaginary part, the real sequence $\{y_{n_k}\}_{k \in \mathbb{N}}$, also lies in the sequentially compact set $[-R, R]$. So, it also must have a convergent subsequence $\{y_{n_{m_k}}\}_{k \in \mathbb{N}}$ (so a subsequence of a subsequence!!) which converges to some $y \in [-R, R]$.

Finally, take the (sub)subsequence $\{z_{n_{m_k}}\}_{k \in \mathbb{N}}$, where $z_{n_{m_k}} = x_{n_{m_k}} + iy_{n_{m_k}}$. The imaginary part converges to y and, by Lemma 2.20, the real part converges to x . Then, by the facts we discussed about convergence of complex sequences, the subsequence $\{z_{n_{m_k}}\}_{k \in \mathbb{N}}$ converges with limit $x + iy$. Since K is closed by assumption, it follows from Proposition 2.21 that this limit $x + iy$ is in K . This proves K is sequentially compact. $\square$

Remark 2.28. • The complex plane $\mathbb{C}$ is not sequentially compact with respect to the standard metric; e.g., the sequence $\{ik\}_{k \in \mathbb{N}}$ has no convergent subsequence. (Neither is $\mathbb{R}^n$.)

- The Riemann sphere S^2 in $\mathbb{R}^3$ is sequentially compact as a subset of $\mathbb{R}^3$ (with the usual Euclidean metric). Hence $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ is sequentially compact with respect to the chordal metric. (One way to prove this formally will be Theorem 2.30 below).

- The orthogonal group $O(n)$ and the unitary group $U(n)$ are sequentially compact; $SL_n(\mathbb{R})$ and $GL_n(\mathbb{R})$ are not (the same holds for matrices with entries in $\mathbb{C}$).

Finally, before stating the connection between continuous functions and sequentially compact sets, we restate continuity in terms of convergent sequences:

Lemma 2.29. *A function $f : X \rightarrow Y$ between two metric spaces is continuous at $x \in X$ if and only if*

$$\lim_{n \rightarrow \infty} f(x_n) = f(x) \quad \text{for every convergent sequence } \{x_n\}_{n \in \mathbb{N}} \text{ in } X \text{ with } x_n \rightarrow x.$$

Proof. See Assignments. □

Recall (from Analysis I) that non-empty sequentially compact sets in $\mathbb{R}$ have a minimal and maximal element. *[This is because $f(x) = |x|$ is continuous - the statement you have seen says continuous functions on sequentially compact sets attain their max/min.]*

Theorem 2.30 (Extreme Value Theorem). *Let $f : X \rightarrow Y$ be a map between two metric spaces. Then,*

$K \subset X$ is sequentially compact and f is continuous

$\implies$ the image $f(K)$ is sequentially compact in Y .

In particular, for $Y = \mathbb{R}$, any continuous real-valued function on a metric space X attains minima and maxima on sequentially compact sets.

Proof. Let $\{y_k\}_{k \in \mathbb{N}}$ be a sequence in $f(K)$, say with $f(x_k) = y_k$. We wish to show it has a convergent subsequence with limit in $f(K)$. Since K is sequentially compact, there must be a convergent subsequence $\{x_{n_k}\}_{k \in \mathbb{N}}$ (of the sequence $\{x_k\}_{k \in \mathbb{N}}$) with some limit x in K . Since f is continuous, we have by Lemma 2.29 that $x_{n_k} \rightarrow x$ implies $y_{n_k} = f(x_{n_k}) \rightarrow f(x)$. Thus $\{y_{n_k}\}_{k \in \mathbb{N}}$ has limit $f(x)$, and because $x \in K$ we must have $f(x) \in f(K)$. □

Remark 2.31. It follows that if K is a sequentially compact subset of $\mathbb{C}$, then $\operatorname{Re}(z)$, $\operatorname{Im}(z)$ and $|z|$ all attain maximum and minimum values on K . In term 2, we will find out a stronger statement under the additional assumptions that f is (complex) differentiable and K is ‘nice’; then the maximum modulus of a function occurs on the boundary of K . This is called the *maximum modulus theorem*.

As was mentioned before (just after the definition of sequential compactness) - it would be desirable to have a definition for sequential compactness that somehow revolves around open sets (as we saw how important they are). This is not straightforward but still possible. It goes through a more general notion of compactness:

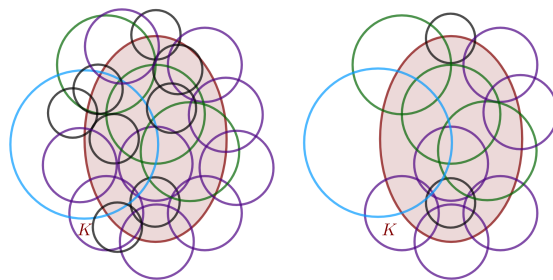


Figure 2.2: An “infinite” open cover of a set K on the left, and the finite open sub-cover of it on the right.

Definition 2.32. Let X be a metric space. We say that a subset K is **compact** if whenever $\{U_i : i \in I\}$ is a collection of open subsets $U_i \subseteq X$ with $K \subseteq \bigcup_{i \in I} U_i$, then there exists a *finite* subset $J \subseteq I$ with $K \subseteq \bigcup_{i \in J} U_i$.

The meaning of compactness is that any open cover admits a finite subcover. We have the following theorem (see Assignments)

Theorem 2.33. *Let X be a metric space and let K be a subset of X . Then K is sequentially compact if and only if K is compact.*

Chapter 3

Complex Differentiation

3.1 Complex differentiability

Differentiation of a complex function is defined in a similar way to that in $\mathbb{R}$, except using complex limits.

Definition 3.1 (Complex differentiability). A function $f : U \rightarrow \mathbb{C}$ defined on an open set U in $\mathbb{C}$ is **(complex) differentiable at $z_0 \in U$** if

$$\lim_{z \rightarrow z_0} \frac{f(z) - f(z_0)}{z - z_0} \text{ exists.}$$

We call this limit the **derivative of f at z_0** and write $f'(z_0)$ for the limit, i.e.

$$f'(z_0) = \lim_{h \rightarrow 0} \frac{f(z_0 + h) - f(z_0)}{h}.$$

Remark 3.2. • In the second formulation the quantity h is a complex number (not a real number), so the limit must exist from every direction. The second is often the more useful expression of the two.

- Note that if a function f is complex differentiable at z then it is continuous at z .

Differentiating complex functions from first principles (examples)

1. Rules for differentiating polynomials are the same. For example, consider $f(z) = z^2$ on $\mathbb{C}$. For any $z \in \mathbb{C}$ we have

$$\lim_{h \rightarrow 0} \frac{(z+h)^2 - z^2}{h} = \lim_{h \rightarrow 0} \frac{z^2 + 2hz + h^2 - z^2}{h} = \lim_{h \rightarrow 0} (2z + h) = 2z.$$

Thus, f is differentiable on $\mathbb{C}$ and $f'(z) = 2z$ as expected.

2. Consider $f(z) = \bar{z}$. For it to be differentiable we must obtain the same limit from every direction. But, considering limits from the purely real and purely imaginary directions, for every $z \in \mathbb{C}$ we have

$$\lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}}} \frac{\overline{z+h} - \bar{z}}{h} = \lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}}} \frac{h}{h} = 1,$$

yet

$$\lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}}} \frac{\overline{z+ih} - \bar{z}}{ih} = \lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}}} \frac{-ih}{ih} = -1.$$

Since z was arbitrary this shows f is not differentiable anywhere.

3. As in the real case, sums/products/quotients of complex differentiable functions are complex differentiable where defined (e.g., all polynomials/rational functions). In particular, **the product and quotient rules hold** for complex derivatives.
4. Composition of differentiable functions are complex differentiable where defined. In particular, **the chain rule holds** for complex derivatives. The proofs of 3. and 4. here are almost identical to those from Analysis I, so are excluded.
5. Generally, non-constant purely real/imaginary functions are not complex differentiable ; e.g., $\operatorname{Re}(z)$, $\operatorname{Im}(z)$, $|z|$ are nowhere differentiable as functions from $\mathbb{C} \rightarrow \mathbb{C}$.

3.2 Cauchy-Riemann equations

We have a basic way of determining whether a function is complex differentiable, but for more complicated functions we wish to find some nicer conditions. We consider a complex valued function f , defined for $z = x + iy$ in some set $U \subseteq \mathbb{C}$, as a function of the two real variables x and y . We write

$$f(z) = u(x, y) + iv(x, y)$$

and call $\operatorname{Re}(f) = u(x, y)$ the *real part of f* and $\operatorname{Im}(f) = v(x, y)$ the *imaginary part of f* .

For example, if $f(z) = z^2$ then $z^2 = (x + iy)^2 = x^2 - y^2 + i2xy$, so $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$.

When considered as real functions $\mathbb{R}^2 \rightarrow \mathbb{R}$ there are naturally two partial derivatives associated with each of u and v :

$$\begin{aligned} u_x(x, y) &= \frac{\partial}{\partial x} u(x, y) := \lim_{h \rightarrow 0} \frac{u(x+h, y) - u(x, y)}{h}, & u_y(x, y) &:= \lim_{h \rightarrow 0} \frac{u(x, y+h) - u(x, y)}{h}, \\ v_x(x, y) &:= \lim_{h \rightarrow 0} \frac{v(x+h, y) - v(x, y)}{h}, & v_y(x, y) &:= \lim_{h \rightarrow 0} \frac{v(x, y+h) - v(x, y)}{h}. \end{aligned}$$

Note that the limits above are standard real limits in $\mathbb{R}$ (so h is real).

Proposition 3.3. [*Cauchy-Riemann equations*] Let $f = u + iv$ be complex differentiable at z_0 . Then the real partial derivatives u_x, u_y, v_x, v_y exist at z_0 and satisfy the **Cauchy-Riemann equations**:

$$u_x(z_0) = v_y(z_0) \quad u_y(z_0) = -v_x(z_0).$$

Furthermore, the derivative of f at z_0 can be written as

$$\begin{aligned} f'(z_0) &= u_x(z_0) + iv_x(z_0) = v_y(z_0) - iu_y(z_0) \\ &= u_x(z_0) - iu_y(z_0) = v_y(z_0) + iv_x(z_0). \end{aligned}$$

Proof. Since f is complex differentiable at $z_0 = x_0 + iy_0$, the limit in the definition must exist and agree no matter which way we approach z_0 . Choosing the purely real and purely imaginary directions we have

$$\lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}}} \frac{f(z_0 + h) - f(z_0)}{h} = f'(z_0) = \lim_{\substack{h \rightarrow 0 \\ h \in \mathbb{R}}} \frac{f(z_0 + ih) - f(z_0)}{ih}.$$

We may split each limit into real and imaginary parts, by standard properties of complex limits. Since, when h is real,

$$\begin{aligned} f(z_0 + h) &= f((x_0 + h) + iy_0) = u(x_0 + h, y_0) + iv(x_0 + h, y_0), \quad \text{and} \\ f(z_0 + ih) &= f(x_0 + i(y_0 + h)) = u(x_0, y_0 + h) + iv(x_0, y_0 + h), \end{aligned}$$

the above reads

$$\begin{aligned} &\lim_{h \rightarrow 0} \frac{u(x_0 + h, y_0) - u(x_0, y_0)}{h} + i \lim_{h \rightarrow 0} \frac{v(x_0 + h, y_0) - v(x_0, y_0)}{h} = f'(z_0) \\ &= \frac{1}{i} \lim_{h \rightarrow 0} \frac{u(x_0, y_0 + h) - u(x_0, y_0)}{h} + \frac{i}{i} \lim_{h \rightarrow 0} \frac{v(x_0, y_0 + h) - v(x_0, y_0)}{h}. \end{aligned}$$

In other words, $u_x(z_0) + iv_x(z_0) = f'(z_0) = (1/i)u_y(z_0) + v_y(z_0)$. Since $1/i = -i$ the statement follows by comparing real and imaginary parts. $\square$

Returning to the function $f(z) = z^2$, note that the C-R equations indeed hold: We have $u(x, y) = x^2 - y^2$ and $v(x, y) = 2xy$ and so

$$u_x = 2x, \quad v_y = 2x, \quad u_y = -2y, \quad v_x = 2y.$$

Remark 3.4. [Warning] Note that Proposition 3.3 provides a way of showing a function is not differentiable at a given point. It cannot be used to prove a function is differentiable. I.e., it is only a necessary condition, not a sufficient one.

It turns out, by considering a complex function $f : \mathbb{C} \rightarrow \mathbb{C}$ as a real function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 : (x, y) \mapsto (u(x, y), v(x, y))$, that we have precisely

$$f \text{ is complex differentiable} \iff f \text{ is real differentiable} + \text{C-R equations hold};$$

so indeed we do have a converse:

Theorem 3.5. *Let $f = u + iv$ be defined on an open subset U of $\mathbb{C}$. Assume the partial derivatives u_x, u_y, v_x, v_y exist, are continuous, and satisfy the Cauchy-Riemann equations at $z_0 \in U$. Then f is complex differentiable at z_0 .*

Proof. Omitted. (If you're interested to see a proof of this theorem, then you can find one in Section 5.6, page 59, of the book "Introduction to Complex Analysis" by H. A. Priestley, Second Edition, Oxford University Press, 2009.) $\square$

Determining complex differentiability via Cauchy-Riemann (examples)

1. Let $f(z) = \exp(z) = e^x \cos y + ie^x \sin y$. Then

$$u_x = e^x \cos y, \quad v_y = e^x \cos y, \quad u_y = -e^x \sin y, \quad v_x = e^x \sin y.$$

All these functions are continuous as real functions (see Calculus I/AMV II) and the C-R equations hold. Thus, by Theorem 3.5, $\exp$ is differentiable everywhere in $\mathbb{C}$ and by Proposition 3.3

$$\exp'(z) = u_x + iv_x = e^x \cos y + ie^x \sin y = \exp(z).$$

2. By the chain rule, $f(z) = e^{iz}$ is differentiable and $f'(z) = ie^{iz}$. Since they are just sums of $\exp$, all of the functions $\sin, \cos, \sinh, \cosh$ are differentiable everywhere in $\mathbb{C}$ and you can verify

$$\sin'(z) = \cos z, \quad \cos'(z) = -\sin z, \quad \sinh'(z) = \cosh z, \quad \cosh'(z) = \sinh z.$$

Similarly, all polynomials/rational functions are differentiable with the usual formulae: e.g., for $a_0 \dots a_n$ complex;

$$(a_n z^n + \dots + a_2 z^2 + a_1 z + a_0)' = na_n z^{n-1} + \dots + 2a_2 z + a_1.$$

For the branch of $\log$ corresponding to arguments in $(\theta_1, \theta_2]$, the function $\log(z)$ is differentiable in $\mathbb{C} \setminus R_{\theta_1}$, i.e. at all points outside the branch cut. At these points, the derivative is given by $\log'(z) = 1/z$ (see Sheet 4 Q7 where this is proved for the principal branch).

3. $f(z) = \sin(z)/z^2$ is differentiable everywhere except $z = 0$, since it is the quotient of two complex differentiable functions (and is not defined when $z^2 = 0$). Furthermore,

$$[\sin(z)/z^2]' = \frac{(\cos z)(z^2) - (\sin z)(2z)}{(z^2)^2} = \frac{z \cos z - 2 \sin z}{z^3}.$$

4. Let $f(z) = f(x+iy) = (x^3 + 3x^2y - y^3 - x^2 - 2y^2) + i(-x^3 + 3xy^2 - y^3 + 4xy + 3y)$. Then,

$$\begin{aligned} u_x &= 3x^2 + 6xy - 2x, & v_y &= 6xy - 3y^2 + 4x + 3, \\ u_y &= 3x^2 - 3y^2 - 4y, & v_x &= -3x^2 + 3y^2 + 4y. \end{aligned}$$

Thus, the partial derivatives exist and are continuous everywhere. We have $u_y = -v_x$ everywhere, but $u_x = v_y$ if and only if $3x^2 - 2x = -3y^2 + 4x + 3$; that is, when $(x-1)^2 + y^2 = 2$. So, f is differentiable only on the circle of radius $\sqrt{2}$ centred at 1 in the complex plane!

Holomorphicity

In example 4. above, the function is only differentiable on a 1-dimensional subset of $\mathbb{C}$. This is a similar situation to a real function on $\mathbb{R}$ being differentiable only at a single point - not a very interesting function to work with from an analytical perspective. It will be useful for us to consider functions that are differentiable on genuine 2-dimensional sets in $\mathbb{C}$, for this will allow us to (later) express the functions using Taylor series. Such functions will turn out to have some quite remarkable properties.

Definition 3.6 (Holomorphic functions). A function $f : U \rightarrow \mathbb{C}$ defined on an open set $U \subset \mathbb{C}$ is **holomorphic on U** if it is complex differentiable at every point in U .

We say f is **holomorphic at $z_0 \in U$** if it is holomorphic on some open ball $B_\epsilon(z_0)$; (in other words, if there exists $\epsilon > 0$ such that f is complex differentiable at every point in $B_\epsilon(z_0)$).

Remark 3.7. • By the previous examples, exp, trig functions, and polynomials are holomorphic on all of $\mathbb{C}$. Logarithms and complex powers are holomorphic at points outside their branch cuts.

- The function in the previous example 4. is nowhere holomorphic despite being differentiable on a circle. To see this, simply note that any (open) ball in $\mathbb{C}$ centred at a point z on this circle must contain a point w not on the circle. The function is not complex differentiable at w , so is not holomorphic at z . Thus, being holomorphic is a strictly stronger property than being complex differentiable.

3.3 Connected sets and zero derivatives

We explore what it means for the derivative to be zero for a complex function. Consider, for example, the function $f : U \rightarrow \mathbb{C}$, where $U = \{z \in \mathbb{C} : |z| \neq 1\}$ and

$$f(z) = \begin{cases} 1 & \text{if } |z| < 1, \\ 2 & \text{if } |z| > 1. \end{cases}$$

The set U is open, the function f is clearly holomorphic on U and has zero derivative on U . However, it takes two different values - it is not constant. The problem is that f lives on two ‘unconnected’ subsets of the complex plane. We wish to find conditions on the set U under which we can conclude f is constant (as we are used to for real functions).

First, we need some terminology:

Definition 3.8. [Paths & path-connectedness]

1. A **path or curve** from $z \in \mathbb{C}$ to $w \in \mathbb{C}$ is a continuous function $\gamma : [0, 1] \rightarrow \mathbb{C}$ with $\gamma(0) = z$ and $\gamma(1) = w$. We say the path/curve is **closed** if $z = w$ (in this case, the endpoints of the path join up).
2. A path/curve is said to be **continuously differentiable, or C^1** , if its real part and imaginary parts are continuously differentiable on $[0, 1]$. At the end point 0 and 1 this means that real part and imaginary parts have right-sided derivatives at 0 and left-sided derivative at 1, and that the derivatives are continuous from the right at 0 and from the left at 1.

In that case we define

$$\gamma'(t) = (\operatorname{Re}(\gamma(t)))' + i(\operatorname{Im}(\gamma(t)))'.$$

3. We say a subset $U \subseteq \mathbb{C}$ is C^1 **path-connected** if for every pair of points $z, w \in U$ there exists a C^1 path from z to w such that $\gamma(t) \in U$ for all $t \in [0, 1]$. For simplicity, we will say path-connected instead of C^1 path connected in the remaining of this module.

In general paths can be defined from an interval $[a, b]$ instead of $[0, 1]$. It makes no difference since we can always reparametrise the paths. For instance, if $\gamma_1 : [a, b] \rightarrow X$ is a given function we just define $\gamma : [0, 1] \rightarrow X$ by

$$\gamma(t) = \gamma_1\left(\frac{t-a}{b-a}\right).$$

Conversely for $\gamma : [0, 1] \rightarrow X$ we define $\gamma_1 : [a, b] \rightarrow X$ by

$$\gamma_1(t) = \gamma((b-a)t + a).$$

Note that the continuity or differentiability of the maps is identical, as well as their image, so our definition of paths and C^1 paths remains intact.

Remark 3.9. • [Advanced/Warning] Usually people use ‘path-connected’ to mean you can connect two points simply by a continuous curve. We used a stronger definition to make our life easier.

Intuitively, a subset is path-connected if it has ‘one piece’; i.e., if you can draw a line between any two points in the set without the pencil leaving the subset (or the paper!).

Definition 3.10 (Domains). A **domain** D is an open, path-connected subset of $\mathbb{C}$. Some people call domains **regions**.

Showing sets are domains (examples)

1. We know both $\mathbb{C}$ and $B_r(z)$ (for any $z \in \mathbb{C}$ and $r > 0$) are open. They are also path-connected: Simply choose the line segment between a and b given by $\gamma(t) = a + (b - a)t$. This is clearly a smooth path, so $\mathbb{C}$ and $B_r(z)$ are domains.
2. The set $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$, on which the principal branches of $\log$ and complex powers are defined, is path-connected: Choose arbitrary $a, b \in \mathbb{C} \setminus \mathbb{R}_{\leq 0}$. If a lies on the positive real axis then the line segment again works (since it avoids the negative real axis). Otherwise, we could try letting

$$\gamma_1(t) = |a|e^{i\text{Arg}(a)(1-t)}, \quad \gamma_2(t) = |a| + (b - |a|)t.$$

The first path takes a to $|a|$, avoiding $\mathbb{R}_{\leq 0}$ by tracing around the circle of radius $|a|$ centred at the origin. The second joins $|a|$ to b via a straight line. See Figure 3.1. (You can reparametrize to make t run from 0 to 1 on the combined path if you like.) But the combined path running through γ_1 then γ_2 is not differentiable at the point where the paths γ_1 and γ_2 meet! One possibility for a differentiable path joining the points a and b would be the arc of the circle through a and b that does not cross $\mathbb{R}_{\leq 0}$.

We know $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$ is also open, so it is a domain.

3. The set $\{z \in \mathbb{C} : |z| \neq 1\}$ is not a domain. It is open, but it is not path-connected. Pick a point a with $|a| < 1$ and b with $|b| > 1$. We cannot draw a continuous path from a to b without crossing the circle $|z| = 1$; i.e., you cannot draw a continuous path between the points without leaving the set.

Previously we have seen that there is a chain rule for the composition of two complex differentiable functions. There is also the chain rule for the composition of a complex differentiable function and a smooth path:

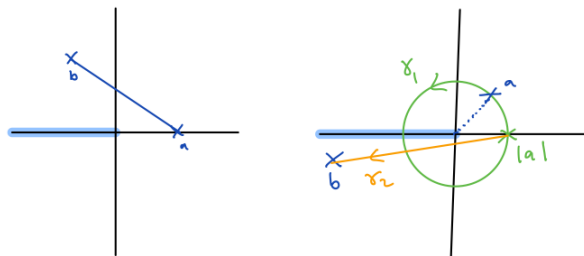


Figure 3.1: A depiction of the paths when a is on the real line, and when a is not on the real line.

Lemma 3.11. [Chain rule] Let $U \subset \mathbb{C}$ be an open set, $f : U \rightarrow \mathbb{C}$ be a holomorphic function on $\mathbb{C}$ and $\gamma : [0, 1] \rightarrow U$ be a smooth path. Then for $t_0 \in [0, 1]$ we have

$$(f \circ \gamma)'(t_0) = f'(\gamma(t_0))\gamma'(t_0).$$

Proof. Similar to the proof of the chain rule in Analysis I. $\square$

Theorem 3.12. Let $f : D \rightarrow \mathbb{C}$ be holomorphic on a domain $D \subseteq \mathbb{C}$. If $f'(z) = 0$ for every $z \in D$ then f is constant on D .

Proof. Since D is path-connected, it is enough to show f is constant on any smooth path γ lying in D . In other words we wish to show $f \circ \gamma$ is a constant function of t .

We have $(f \circ \gamma)'(t) = f'(\gamma(t))\gamma'(t) = 0$ by the chain rule. Since $f = u + iv$, we have

$$0 = (f \circ \gamma)'(t) = (u \circ \gamma)'(t) + i(v \circ \gamma)'(t),$$

which means both $(u \circ \gamma)'(t)$ and $(v \circ \gamma)'(t)$ are equal to 0 for all t . Hence u and v are constant along γ , so f is too. $\square$

Remark 3.13. The same statement holds if “ $f' = 0$ ” is replaced with “ f is purely real/imaginary”, or “ f has constant real/imaginary part”, or “ f has constant modulus” (see Sheet 4). So none of $\operatorname{Re}(z)$, $\operatorname{Im}(z)$, $|z|$ or $\operatorname{Arg}(z)$ are holomorphic anywhere (since they are non-constant real-valued functions).

3.4 The angle-preserving properties of holomorphic functions

Given a smooth path $\gamma : [0, 1] \rightarrow \mathbb{C}$, we can talk about the tangent vector to γ at a point $z_0 = \gamma(t_0)$ on the path.

Definition 3.14 (Conformal maps). We say a (real differentiable) map $f : D \rightarrow \mathbb{C}$ on a domain $D \subseteq \mathbb{C}$ is **conformal at z_0** if it preserves the angle and orientation between any two tangent vectors at z_0 . This is exactly the same as saying that it preserves the angle and orientation between any two smooth curves passing through z_0 . We say that f is **conformal** if it is conformal at all points in D .

It turns out that there is a close relationship between holomorphic and conformal maps.

Lemma 3.15. *[Holomorphic maps are conformal] A holomorphic map f with $f'(z_0) \neq 0$ is conformal at z_0 .*

Proof. Take a smooth path $\gamma : [0, 1] \rightarrow \mathbb{C}$ that passes through z_0 , and consider the tangent vector to γ at the point $z_0 = \gamma(t_0)$ on the path. What happens to this tangent vector under the application of a holomorphic function f ?

The path $f \circ \gamma$ is the path obtained by mapping γ under f . So to calculate what happens to the tangent vector $\gamma'(t_0)$ under f , we should calculate the tangent vector $(f \circ \gamma)'(t_0)$. This is done using the chain rule:

$$(f \circ \gamma)'(t_0) = f'(\gamma(t_0))\gamma'(t_0) = f'(z_0)\gamma'(t_0).$$

This describes what happens to tangent vectors geometrically: f transforms the tangent vector $\gamma'(t_0)$ by multiplying it by the complex number $f'(z_0)$.

But we understand what multiplication by complex numbers means geometrically: if $f'(z_0) \neq 0$ then multiplication by $f'(z_0)$ is dilation (by $|f'(z_0)|$) followed by a rotation (by $\text{Arg}(f'(z_0))$).

We now notice the following: both dilations and rotations preserve the angles and orientations between vectors! This completes the proof. $\square$

Determining where maps are conformal (examples)

- Let $f(z) = z^2$. Here, $f'(z) = 2z$ so $f'(z) = 0 \iff z = 0$ and so f is conformal on $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ (since it is holomorphic everywhere). We don't yet know whether f is conformal at the origin, but we can check: we know that $f'(z) = 2z$ so $f'(0) = 0$ and hence the action of f on tangent vectors at the origin is to multiply by 0, hence sending any tangent vector to zero.
- Let $f(z) = \frac{2}{3}z^3 + \frac{1}{2}(1 - 2i)z^2 - iz + 2 - 1$. This function is clearly holomorphic as it is just a polynomial. We have

$$f'(z) = 2z^2 + (1 - 2i)z - i = (2z + 1)(z - i),$$

which has zeros at $-\frac{1}{2}$ and i , so f is conformal on $\mathbb{C} \setminus \{-\frac{1}{2}, i\}$. *[Note, we can explicitly check what f does to tangent vectors at each point in this set: For*

example, at $z = i/2$ we have $f'(i/2) = (i+1)(-i/2) = (1-i)/2 = \frac{1}{\sqrt{2}}e^{-i\pi/4}$. Thus, at $z = i/2$ the function f rotates tangent vectors by $\text{Arg}(f'(i/2)) = -\pi/4$ degrees, then dilates by a factor of $|f'(z)| = 1/\sqrt{2}$.]

Rather than having to then check two paths through $z = 1/2$ and $z = i$ respectively, to determine whether f is conformal there, it will be convenient to have a complete description in terms of holomorphicity. The converse to Lemma 3.15 holds:

Proposition 3.16. [Conformal maps are holomorphic] Let D be a domain. If f is conformal at $z_0 \in D$ then f is complex differentiable at z_0 and $f'(z_0) \neq 0$. Therefore, if f is conformal on D , then f is holomorphic on D and $f'(z) \neq 0$ for all $z \in D$. Thus

$$f \text{ is conformal on } D \iff f \text{ is holomorphic with } f'(z) \neq 0 \text{ for all } z \in D.$$

Proof. (Non-examinable) Sketch: we know $f = u + iv$ is real differentiable when considered as a map $f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix}$. The total differential of this map at a point $z_0 = x_0 + iy_0$ (see AMV II) is the Jacobian matrix $D_{z_0} = \begin{pmatrix} u_x(x_0,y_0) & u_y(x_0,y_0) \\ v_x(x_0,y_0) & v_y(x_0,y_0) \end{pmatrix}$. To preserve angles it must be of the form $c \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ for some $c \in \mathbb{R}_{\neq 0}$. Thus D_{z_0} is of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ and the C-R equations are satisfied at z_0 . Since f is real differentiable by assumption, C-R equations are actually enough to show that f is complex differentiable at z_0 (we have not stated exactly this result before, but it holds). Since $\det D_z \neq 0$ we have $f'(z_0) = u_x(x_0, y_0) + iv_x(x_0, y_0) \neq 0$. $\square$

Remark 3.17. [Advanced] The idea of this proof, in reverse, can be adapted to give a more sophisticated proof of the CR equations!

Determining where maps are conformal (examples continued)

- Let $f(z) = xy + iy^2$. Then $u_x = y, v_y = 2y, u_y = x, v_x = 0$. So, the C-R equations only hold at $y = x = 0$. Hence f is not conformal at any point $z_0 \neq 0$, since it cannot be complex differentiable at z_0 (as C-R. equations fail to hold), and hence it is not conformal at z_0 by Proposition 3.16. What happens for $z_0 = 0$? There, $f'(0) = u_x(0, 0) + iv_x(0, 0) = 0 + i0 = 0$. So f cannot be conformal at 0 either by Proposition 3.16. Thus, f is nowhere conformal.

Visualising conformal maps

Here is a useful result, helpful in visualising the action of conformal maps:

Corollary 3.18. Any conformal map maps orthogonal grids in the (x, y) -plane to orthogonal grids.

Remark 3.19. [**Important warning**] We must allow our grids to be made up of arbitrary smooth curves, not necessarily straight lines (Corollary 3.18 is false if they are required to be made of straight lines).

How to visualise conformal maps (examples)

- Let $f(z) = z^2 = x^2 - y^2 + i2xy$ and consider the simple grid in the (x, y) -plane made up of lines parallel to the real/imaginary axes, separated by distance 1.

Given a line (not passing through the origin) in such a grid, say $x = a (\neq 0)$, the function f takes this line to the points $(a^2 - y^2, 2ay)$ in the (u, v) -plane. If $u = a^2 - y^2$ and $v = 2ay$, then $v^2 = 4a^2y^2$ and so $u = a^2 - (\frac{v}{2a})^2$ and the image defines a parabola.

Similarly, the line $y = b (\neq 0)$ is taken to the parabola $u = (\frac{v}{2b})^2 - b^2$ in the (u, v) -plane. By sketching these on the same axes we see that the parabolas indeed cross at right angles.

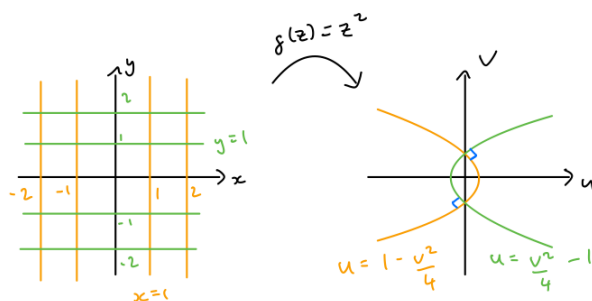


Figure 3.2: The image of the perpendicular lines $x = 1$, $y = 1$ is a pair of parabolas that cross at right angles.

- For the same example, consider the level curves $u(x, y) = a$ and $v(x, y) = b$ with $a, b \neq 0$. These trace out the curves $y^2 = x^2 - a$ and $y = \frac{b}{2x}$ respectively. By sketching these curves on a graph (for, say, $u = \pm 1, v = \pm 1$) we see that they are perpendicular. (They have to be, since $f(z)$ is conformal on $\mathbb{C} - \{0\}$ and they map to perpendicular straight lines.)

3.5 Biholomorphic maps

We wish to build a dictionary of conformal maps as a way of getting back and forth between domains of the complex plane. To do this, we need our maps to be invertible.

Definition 3.20 (Biholomorphic maps). Let D and D' be domains. We say that $f : D \rightarrow D'$ is **biholomorphic** if f is holomorphic, a bijection, and the inverse $f^{-1} : D' \rightarrow D$ is also holomorphic. A biholomorphic map f is called a **biholomorphism**. When f as above exists, we say that the domains D and D' are **biholomorphic** and write $f : D \xrightarrow{\sim} D'$.

Remark 3.21. [Advanced] It is possible to prove that if f as above is holomorphic and bijective, then it is automatic that its inverse is holomorphic.

Finding biholomorphisms (examples)

1. The function $\exp : \mathbb{C} \rightarrow \mathbb{C}^*$ is not biholomorphic since it is not injective (e.g., $e^z = e^{z+2n\pi i}$). How can we make it biholomorphic? We simply need to restrict it to a smaller domain where it will be injective. It is not too hard to check that

$$\exp(z_1) = \exp(z_2)$$

if and only if $\operatorname{Re}(z_1) = \operatorname{Re}(z_2)$ and $\operatorname{Im}(z_1) - \operatorname{Im}(z_2) \in 2\pi\mathbb{Z}$. So if we rule out different z_1 and z_2 whose imaginary parts differ by a non-zero integer multiple of 2π then we might be ok. We can do this by setting

$$D := \{z \in \mathbb{C} : \operatorname{Im}(z) \in (-\pi, \pi)\}.$$

If $z = x + iy \in D$ then $\exp(z) = e^x e^{iy}$ so the principal argument of $\exp(z)$ is never equal to π . On the other hand, it is not hard to check that $\exp$ maps D onto $\mathbb{C} - \mathbb{R}_{\leq 0}$. Moreover, by our choice of D , $\exp$ is injective on D . So $\exp$ is a bijection between D and $\mathbb{C} - \mathbb{R}_{\leq 0}$. Therefore it has an inverse function, and we already know what it is: the principal branch of Log . The function Log maps $\mathbb{C} - \mathbb{R}_{\leq 0}$ to D and is differentiable at every point of $\mathbb{C} - \mathbb{R}_{\leq 0}$, so it is holomorphic on $\mathbb{C} - \mathbb{R}_{\leq 0}$. In summary, when restricted to D , $\exp$ is biholomorphic and hence D and $\mathbb{C} - \mathbb{R}_{\leq 0}$ are biholomorphic.

2. Let $f(z) = z^2$. We have seen that this function is holomorphic on $\mathbb{C}$. However, as in the previous example, it is not injective since e.g. $f(1) = f(-1) = 1$. Note that $f(z_1) = f(z_2)$ with $z_1 \neq z_2$ if and only if $z_1 = \pm z_2$, so we may be able to create a biholomorphic function by restricting f to a domain that never simultaneously contains z and $-z$. Such a domain is the right half plane $\mathbb{H}_R := \{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$. We saw in a previous example that f maps $\mathbb{H}_R$ to $\mathbb{C} - \mathbb{R}_{\leq 0}$. In fact this is a bijection, with inverse given by

$$f^{-1} : \mathbb{C} - \mathbb{R}_{\leq 0} \rightarrow \mathbb{H}_R, \quad f^{-1}(z) := \exp\left(\frac{1}{2}\operatorname{Log}(z)\right).$$

Thus f^{-1} is holomorphic, since it is a composition of holomorphic functions, and hence f gives a biholomorphism $f : \mathbb{H}_R \xrightarrow{\sim} \mathbb{C} - \mathbb{R}_{\leq 0}$.

3. The ‘affine’ linear maps $z \mapsto az + b$ (for $a \in \mathbb{C}^*, b \in \mathbb{C}$) are biholomorphic $\mathbb{C} \xrightarrow{\sim} \mathbb{C}$.

As expected we can compose biholomorphic maps to construct new ones.

Lemma 3.22 (Automorphism groups). *Let $D \subset \mathbb{C}$ be a domain. The set of all biholomorphic maps $f : D \xrightarrow{\sim} D$ from D to itself forms a group under composition. We call this group the **automorphism group of D** and denote it by $\text{Aut}(D)$.*

Proof. We need only check the group properties:

- The identity map $\text{Id} : z \mapsto z$ is always in $\text{Aut}(D)$, and is clearly biholomorphic.
- We know composition of functions is associative, so $f \circ (g \circ h) = (f \circ g) \circ h$, for holomorphic maps.
- f^{-1} is the inverse of f since $f \circ f^{-1} = \text{Id} = f^{-1} \circ f$. Moreover, f^{-1} is in $\text{Aut}(D)$ since it is itself holomorphic by assumption and has holomorphic inverse $(f^{-1})^{-1} = f$.
- $\text{Aut}(D)$ is closed: $(f \circ g)$ is holomorphic (by the chain rule) and its inverse $(f \circ g)^{-1} := g^{-1} \circ f^{-1}$ is also the composition of two holomorphic functions, so holomorphic.

□

Remark 3.23. Note that example 1. shows conformal maps are not necessarily biholomorphic. On the other hand, it turns out (see Sheet 5) that all biholomorphic maps are conformal. [Whilst the converse is not true in general, conformal maps turn out to be ‘locally’ biholomorphic.]

Chapter 4

Möbius transformations

4.1 Definition and first properties of Möbius transformations

In this chapter we study a very special class of functions that play a truly fundamental role in mathematics, and have many beautiful properties. Recall that $GL_2(\mathbb{C})$ is the collection of matrices

$$\begin{aligned} GL_2(\mathbb{C}) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{C}, ad - bc \neq 0 \right\} \\ &= \{2 \times 2 \text{ complex matrices with nonzero determinant}\}. \end{aligned}$$

Also recall the extended complex plane $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ from Chapter 1.

Definition 4.1 (Möbius transformations). Given any matrix $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{C})$ we can define a function

$$M_T : \mathbb{C} \rightarrow \hat{\mathbb{C}}$$

by the formula

$$M_T(z) = \frac{az + b}{cz + d}$$

if $cz + d \neq 0$, and if $cz + d = 0$ we set $M_T\left(-\frac{d}{c}\right) = \infty$ when $c \neq 0$. The function M_T is called a **Möbius transformation**.

Remark 4.2. • [Advanced] The reason for excluding matrices for which $\det T = 0$ is because if $\det T = 0$ and at least one of c and d is non-zero, so that the definition makes sense, then T has rank 1, and this implies that $(a, b) = \lambda(c, d)$ for some $\lambda \in \mathbb{C}$. But this would give that for z with $cz + d \neq 0$, we would have $M_T(z) = \lambda$,

so the transformation would be constant. This is not a very interesting function to think about!

- For $T \in \text{GL}_2(\mathbb{C})$, by taking complex square roots, we can find a number such that $k^2 = \det T$. Then

$$M_T(z) = \frac{az + b}{cz + d} = \frac{\frac{az}{k} + \frac{b}{k}}{\frac{cz}{k} + \frac{d}{k}} = M_{\frac{1}{k}T}(z)$$

and

$$\det\left(\frac{1}{k}T\right) = \frac{1}{k^2} \det T = 1.$$

Thus, we may scale any $T \in \text{GL}_2(\mathbb{C})$ to obtain $T' = \frac{T}{k}$ such that $\det T' = 1$, and $M_T = M_{T'}$.

It is slightly annoying that at the moment, the domain of a Möbius transformation is not the same as its codomain. This would prevent us from composing two Möbius transformations. However, it is possible to extend the definition to describe where $\infty \in \hat{\mathbb{C}}$ gets mapped to: we declare that

$$M_T(\infty) = \begin{cases} \frac{a}{c} & \text{if } c \neq 0 \\ \infty & \text{if } c = 0 \end{cases}.$$

Now we have $M_T : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$.

Example 4.3. 1. $f(z) = z^{-1}$ is a Möbius transformation corresponding to the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Note that $|z| < 1$ and $z \neq 0$ if and only if $|f(z)| = |z^{-1}| = |z|^{-1} > 1$, so f maps the punctured unit ball $B_1(0) - \{0\}$ onto the outside of the closed unit ball. We have $f(0) = \infty$ and $f(\infty) = 0$, so f interchanges these two points.

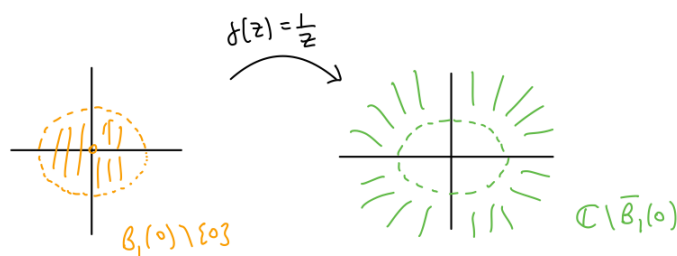


Figure 4.1: Image of $B_1(0) - \{0\}$ under f .

2. The **Cayley map** $f(z) = \frac{z-i}{z+i}$ corresponds to the matrix $\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$. We have

$$f(z) \in B_1(0) \iff |f(z)| < 1 \iff |z+i| > |z-i| \iff z \in \mathbb{H} = \{z \in \mathbb{C} : \operatorname{Im}(z) > 0\}.$$

Hence f maps the upper half plane to the open unit ball centered at 0. Moreover, $f(\infty) = 1$ and $f(-i) = \infty$.

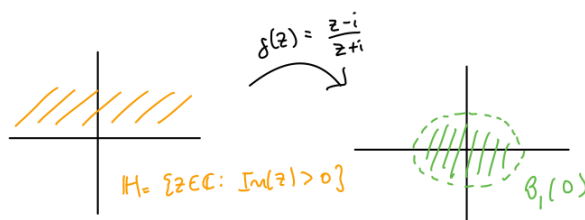


Figure 4.2: Image of upper half-plane under Cayley map.

Lemma 4.4. *The set of Möbius transformations form a group under composition. Furthermore,*

1. $M_{T_1} \circ M_{T_2} = M_{T_1 T_2}$.
2. $(M_T)^{-1} = M_{T^{-1}}$.
3. $M_T = \text{Id} \iff T = t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (t \in \mathbb{C}^*).$

Proof. See Assignments. □

Remark 4.5. Recall from Linear Algebra that $\text{GL}_2(\mathbb{C})$ forms a group under matrix multiplication. Lemma 4.4 says more than that the Möbius transformations form a group. It says that the mapping

$$T \mapsto M_T$$

is a group homomorphism between $\text{GL}_2(\mathbb{C})$ and the Möbius transformations! In other words, composing Möbius transformations is basically just multiplying matrices!

Lemma 4.6. *Let $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathbb{C})$. If $c = 0$, the Möbius transformation M_T gives a biholomorphic map*

$$M_T : \mathbb{C} \xrightarrow{\sim} \mathbb{C}.$$

If $c \neq 0$, then M_T gives a biholomorphic map

$$M_T : \mathbb{C} - \left\{ \frac{-d}{c} \right\} \xrightarrow{\sim} \mathbb{C} - \left\{ \frac{a}{c} \right\}.$$

Proof. If $c = 0$ then $a, d \neq 0$ since $\det T \neq 0$. Hence

$$M_T(z) = \frac{az + b}{d} = \frac{a}{d}z + \frac{b}{d}$$

is just an affine linear map which is holomorphic. It is a bijection with inverse given by $(M_T)^{-1} = \frac{d}{a}z - \frac{b}{a}$, which is holomorphic by the same token. So M_T gives a biholomorphic map $\mathbb{C} \xrightarrow{\sim} \mathbb{C}$.

If $c \neq 0$ let's check that M_T is complex differentiable at every $z \neq \frac{-d}{c}$. We can calculate

$$M'_T(z) = \frac{a(cz + d) - c(az + b)}{(cz + d)^2} = \frac{\det(T)}{(cz + d)^2}.$$

Therefore the derivative exists at all points of $\mathbb{C} - \{\frac{-d}{c}\}$ and hence M_T is holomorphic there. It is bijective with the inverse given by the inverse Möbius transformation, which is holomorphic by the same argument. Hence M_T is a biholomorphism from $\mathbb{C} - \{\frac{-d}{c}\} \xrightarrow{\sim} \mathbb{C} - \{\frac{a}{c}\}$. $\square$

Corollary 4.7. *A Möbius transformation M_T is conformal at all $z \in \mathbb{C}$ with $M_T(z) \neq \infty$.*

Proof. We know that biholomorphic maps are conformal, and M_T is a biholomorphism on $\mathbb{C}$ minus the points that map to ∞ . Alternatively, one can check that in both cases of the proof of Lemma 4.6, the derivative never vanishes. $\square$

Corollary 4.8. *Any Möbius transformation is a bijection from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$.*

Proof. By Lemma 4.4 M_T has an inverse $M_{T^{-1}}$, so that $M_T M_{T^{-1}} = M_{T^{-1}} M_T$ as maps from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$, so M_T is a bijection. $\square$

4.2 Fixed points, the cross-ratio, and the three points Theorem

Let $T \in \text{GL}_2(\mathbb{C})$ and M_T be a Möbius transformation. Then a point z is a fixed point of M_T if $M_T(z) = z$.

Lemma 4.9. *Let $T \in \text{GL}_2(\mathbb{C})$. If $M_T : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ is not the identity map, then M_T has at most 2 fixed points in $\hat{\mathbb{C}}$. In other words, if a Möbius transformation has three fixed points in $\hat{\mathbb{C}}$, then it is the identity.*

Proof. We split into two cases.

First suppose $M_T(\infty) = \infty$. From the definition, this can only happen if $c = 0$, so M_T preserves $\mathbb{C}$, and for $z \in \mathbb{C}$, $M_T(z) = \frac{a}{d}z + \frac{b}{d}$ with $a, d \neq 0$ (because $\det T \neq 0$). Such an affine linear map has at most one fixed point as follows:

- If $a = d$ then $b \neq 0$ since we assume M_T is not the identity. Then M_T is a translation by b/d which has no fixed points.
- If $a \neq d$ then M_T has a unique fixed point in $\mathbb{C}$ given by $z_0 = b/(d - a)$.

Now suppose $M_T(\infty) \neq \infty$. Then any fixed points of M_T are in $\mathbb{C}$. Suppose $z_0 \in \mathbb{C}$ is such that $M_T(z_0) = z_0$. We have

$$M \begin{pmatrix} a & b \\ c & d \end{pmatrix} (z_0) = z_0 \iff \frac{az_0 + b}{cz_0 + d} = z_0 \iff cz_0^2 + (d - a)z_0 - b = 0.$$

There are at most two roots of this quadratic equation, so at most two fixed points of M_T in $\hat{\mathbb{C}}$. $\square$

We introduce an important quantity called the *cross-ratio*.

Definition 4.10. Given four distinct points $z_0, z_1, z_2, z_3 \in \mathbb{C}$, the *cross-ratio* of these points is defined by

$$(z_0, z_1; z_2, z_3) := \frac{(z_0 - z_2)(z_1 - z_3)}{(z_0 - z_3)(z_1 - z_2)} = \frac{\frac{z_0 - z_2}{z_0 - z_3}}{\frac{z_1 - z_2}{z_1 - z_3}}.$$

We will denote the cross ratio of z_0, z_1, z_2, z_3 by $(z_0, z_1; z_2, z_3)$. We can extend the definition to the case that one of the points is ∞ by removing all differences involving that point, for example,

$$(\infty, z_1; z_2, z_3) := \frac{(z_1 - z_3)}{(z_1 - z_2)}.$$

We can use the cross ratio to prove that a Möbius transformation is uniquely determined by how it acts upon any three given points in $\hat{\mathbb{C}}$.

Theorem 4.11. [Three points Theorem] Let $\{z_1, z_2, z_3\}$ and $\{w_1, w_2, w_3\}$ be two sets of three ordered distinct points in $\hat{\mathbb{C}}$. Then there exists a unique Möbius Transformation f such that $f(z_i) = w_i$ for $i = 1, 2, 3$.

Proof.

Existence: Consider the functions

$$F(z) := (z, w_1; w_2, w_3), \quad G(z) := (z, z_1; z_2, z_3).$$

These are Möbius transformations with the properties that $F(w_1) = 1, F(w_2) = 0, F(w_3) = \infty$, and $G(z_1) = 1, G(z_2) = 0, G(z_3) = \infty$. Therefore $F^{-1} \circ G$ maps each z_i to w_i .

Uniqueness: Assume that there are two such maps, say f and $\tilde{f}$. Then the Möbius transformation $H := f^{-1} \circ \tilde{f}$ satisfies $H(z_i) = f^{-1}(\tilde{f}(z_i)) = f^{-1}(w_i) = z_i$. This shows H has three fixed points, z_1, z_2, z_3 , and by Lemma 4.9 it must therefore be trivial. Thus, $f = \tilde{f}$. $\square$

One has the following beautiful and fundamental fact about Möbius transformations:

Proposition 4.12. *[Möbius transformations preserve cross-ratio] Möbius transformations preserve the cross ratio: if z_0, z_1, z_2, z_3 are four distinct points in $\hat{\mathbb{C}}$, and f is a Möbius transformation, then*

$$(f(z_0), f(z_1); f(z_2), f(z_3)) = (z_0, z_1; z_2, z_3). \quad (*)$$

Proof. Let $w_i = f(z_i)$. Let F and G be the functions defined in the proof of Theorem 4.11. Recall that $F^{-1} \circ G$ mapped each z_i to w_i . So does f . Since there is a unique Möbius transformation with this property, we must have $f = F^{-1} \circ G$. Rearranging, $F \circ f = G$. But this is just the identity stated after applying both functions to z_0 :

$$(f(z_0), f(z_1); f(z_2), f(z_3)) = F \circ f(z_0) = G(z_0) = (z_0, z_1; z_2, z_3).$$

$\square$

Finding a Möbius transformation from three points (example)

Find the unique Möbius transformation $f : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ mapping the points $\{1, -1, i\}$ to the points $\{0, \infty, 1\}$. Our method is simply to rearrange (*): Notice that

$$\lim_{|w| \rightarrow \infty} \frac{f(z) - w}{w_1 - w} = \lim_{|w| \rightarrow \infty} \frac{\frac{f(z)}{w} - 1}{\frac{w_1}{w} - 1} = \frac{-1}{-1} = 1,$$

so (*) reduces to

$$\begin{aligned} 1 \cdot \frac{w_1 - w_3}{f(z) - w_3} &= \left(\frac{z - z_2}{z - z_3} \right) \left(\frac{z_1 - z_3}{z_1 - z_2} \right) \\ \iff \frac{0 - 1}{f(z) - 1} &= \left(\frac{z - (-1)}{z - i} \right) \left(\frac{1 - i}{1 - (-1)} \right) \\ \iff f(z) - 1 &= \frac{-2(z - i)}{(z + 1)(1 - i)} \\ \iff f(z) &= \frac{(-1 - i)z + (1 + i)}{(1 - i)z + (1 - i)} \\ \iff f(z) &= \left(\frac{-1 - i}{1 - i} \right) \left(\frac{z - 1}{z + 1} \right) \\ \iff f(z) &= \frac{-iz + i}{z + 1}. \end{aligned}$$

Möbius transformations methods

This leads us to a general strategy to find a Möbius transformation from how it acts on three points: Simply notice that since the cross ratio is preserved,

$$\left(\frac{f(z) - w_2}{f(z) - w_3} \right) \left(\frac{w_1 - w_3}{w_1 - w_2} \right) = \left(\frac{z - z_2}{z - z_3} \right) \left(\frac{z_1 - z_3}{z_1 - z_2} \right),$$

and to find f we need only rearrange this equation.

There is also a general strategy to find the image of a region D under a Möbius transformation M_T :

1. Find the image $M_T(\partial D)$ of the boundary ∂D .
2. Find the image $M_T(z_0)$ of a point $z_0 \in D$ in the interior.
3. The region D' bounded by $M_T(\partial D)$ and containing $M_T(z_0)$ is precisely the image of D under M_T , and

$$M_T : D \xrightarrow{\sim} D' = M_T(D).$$

4.3 Circles and lines

The following property of Möbius transformations is very useful when deciding how they map different regions of the complex plane.

Proposition 4.13. *Möbius transformations map circles and lines in $\hat{\mathbb{C}}$ to circles and lines in $\hat{\mathbb{C}}$.*

Remark 4.14. [Technical] We consider any line to pass through infinity, so that the above makes sense. By circles in $\hat{\mathbb{C}}$ we mean simply circles in $\mathbb{C}$.

In order to talk about what happens to circles and lines under Möbius transformations we need to know the equations of circles and lines. Of course, we already know the equation of a circle or line, but it turns out these can both be described by the same type of equation using complex numbers.

A circle of centre α and radius r is given by the equation

$$|z - \alpha|^2 = r^2$$

which can be rewritten as

$$r^2 = (z - \alpha)(\bar{z} - \bar{\alpha}) = z\bar{z} - \alpha\bar{z} - \bar{\alpha}z + \alpha\bar{\alpha}.$$

Introducing a new parameter $\beta = \alpha\bar{\alpha} - r^2 \in \mathbb{R}$ we have the equation

$$z\bar{z} - \alpha\bar{z} - \bar{\alpha}z + \beta = 0.$$

This is the general equation of a circle, given by parameters $\alpha \in \mathbb{C}$ and $\beta \in \mathbb{R}$ such that $|\alpha|^2 - \beta = r^2 > 0$.

What about lines? Well a line can always be written as a bisector:

$$\ell = \{z \in \mathbb{C} : |z - w_1| = |z - w_2|\}$$

with $w_1 \neq w_2 \in \mathbb{C}$. This means

$$(z - w_1)(\bar{z} - \bar{w}_1) = (z - w_2)(\bar{z} - \bar{w}_2)$$

or

$$z\bar{z} - w_1\bar{z} - \bar{w}_1z + w_1\bar{w}_1 = z\bar{z} - w_2\bar{z} - \bar{w}_2z + w_2\bar{w}_2.$$

Rearranging this, and writing $\alpha = w_1 - w_2 \neq 0$ and $\beta = w_1\bar{w}_1 - w_2\bar{w}_2 \in \mathbb{R}$, we have

$$-\alpha\bar{z} - \bar{\alpha}z + \beta = 0.$$

This is the general equation of a line in $\mathbb{C}$, given by parameters $\alpha \in \mathbb{C}$ and $\beta \in \mathbb{R}$ such that $\alpha \neq 0$.

Therefore we have proved

Lemma 4.15. *[Equation of circles and lines in $\mathbb{C}$] Given $\gamma, \beta \in \mathbb{R}$ and $\alpha \in \mathbb{C}$, the equation*

$$\gamma z\bar{z} - \alpha\bar{z} - \bar{\alpha}z + \beta = 0$$

describes a circle if $\gamma \neq 0$ and $|\alpha|^2 - \beta\gamma > 0$, and a line if $\gamma = 0$ and $\alpha \neq 0$. Conversely, any circle or line can be described by an equation of this form. $\square$

Proof of Proposition 4.13. Consider an arbitrary Möbius transformation M_T where $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We can assume by previous remarks that $\det T = 1$. If $c = 0$, then M_T is affine linear and is easily seen to preserve circles and lines (since rotations, dilations, and translations obviously do). So we can assume $c \neq 0$.

Note that

$$M_T(z) = \frac{az + b}{cz + d} = \frac{caz + cb}{c(cz + d)} = \frac{a(cz + d)}{c(cz + d)} + \frac{cb - ad}{c(cz + d)} = \frac{a}{c} - \frac{1}{c(cz + d)} = \frac{a}{c} - \left(\frac{1}{c^2}\right) \left(\frac{1}{z + \frac{d}{c}}\right).$$

Thus M_T is just some linear maps composed with the function $f(z) = 1/z$, called an **inversion**. Since linear maps preserve circles and lines, it is therefore enough to consider the action of $f(z) = 1/z$ on lines and circles. Moreover, we have $f^{-1}(z) = 1/z = f(z)$ so it is enough to consider one direction.

Let X be a circle or line. By Lemma 4.15, X is described by an equation $\gamma z\bar{z} - \alpha\bar{z} - \bar{\alpha}z + \beta = 0$ with $\gamma, \beta \in \mathbb{R}$. We claim that $f(X)$ is described by a similar equation. Indeed, if $z \in f(X)$ and $z \neq 0$, then $f(z) = f^{-1}(z) \in X$ and this happens if and only if

$$\gamma \left(\frac{1}{z} \right) \overline{\left(\frac{1}{z} \right)} - \alpha \overline{\left(\frac{1}{z} \right)} - \bar{\alpha} \left(\frac{1}{z} \right) + \beta = 0 \iff \beta z\bar{z} - \alpha z - \bar{\alpha}\bar{z} + \gamma = 0.$$

Finally, let's check this is the equation of a circle or a line.

First consider the case $\beta = 0$. If we started with a line we have $\alpha \neq 0$. If we started with a circle we know $|\alpha|^2 = |\alpha|^2 - \beta > 0$ so we have $\alpha \neq 0$ again. Therefore, the new equation describes a line.

Next, consider the case $\beta \neq 0$. Then we can divide by β to get $z\bar{z} - \frac{\alpha}{\beta}z - \frac{\bar{\alpha}}{\beta}\bar{z} + \frac{\gamma}{\beta} = 0$. This looks like the equation of a circle, but we have to check

$$\left\| \frac{\alpha}{\beta} \right\|^2 - \frac{\gamma}{\beta} > 0 \tag{4.1}$$

If we began with a line, then $\gamma = 0$ and $\alpha \neq 0$ so this is clearly true. If we began with a circle, then $\gamma = 1$. If $\beta < 0$ then (4.1) is obvious since both terms are nonnegative and $\gamma/\beta < 0$. If $\beta > 0$ then (4.1) follows from $|\alpha|^2 - \beta > 0$. $\square$

We use the term **circline** to refer to an object that is either a circle or line. Hence the previous Proposition could have been stated ‘Möbius transformations preserve circlines’. The fact that Möbius transformations preserve circles and lines is very powerful when combined with:

Remark 4.16. Any three distinct non-colinear points $z_1, z_2, z_3 \in \mathbb{C}$ uniquely determine a circle in $\mathbb{C}$ passing through those points. Any two distinct points uniquely determine a line passing through those points.

So to find out where a circle is mapped under a Möbius transformation, one simply needs to check where three points on the circle are mapped!

Finding the image of regions under Möbius transformations (example)

Find the image of the unit disc $\mathbb{D} := B_1(0) = \{w \in \mathbb{C} : |w| < 1\}$ under the Möbius transformation corresponding to the matrix $T = \begin{pmatrix} 2+2i & -2-6i \\ 1 & -1-2i \end{pmatrix}$. To determine the image we look at what happens to the boundary: we first find the image of the unit circle centred at the origin. Consider what happens to the four points on the unit circle,

1, i , -1 , $-i$ under the map $M_T(z) = \frac{(2+2i)z - (2+6i)}{z - (1+2i)}$. Under M_T :

$$\begin{aligned} 1 &\mapsto \frac{(2+2i) - (2+6i)}{1 - (1+2i)} = \frac{-4i}{-2i} = 2; \\ i &\mapsto \frac{(2+2i)i - (2+6i)}{i - (1+2i)} = \frac{-4 - 4i}{-1 - i} = 4; \\ -1 &\mapsto \frac{-(2+2i) - (2+6i)}{-1 - (1+2i)} = \frac{-4 - 8i}{-2 - 2i} \end{aligned}$$

By Proposition 4.13 the unit circle must be mapped to a circle, and the three calculations above tell us that the image of the unit circle must be the circle centred at $z = 3$ of radius 1. Note, the action of M_T is not something so simple as a rotation; for example we have

$$-i \mapsto \frac{-(2+2i)i - (2+6i)}{-i - (1+2i)} = \frac{-8i}{-1 - 3i} = \frac{12 + 4i}{5}, \quad \text{which is near the image of } -1 !!$$

What about the interior of this circle? Let's pick a point in the unit disc and see where it is taken. The point $z = 0$ is an obvious choice. We have

$$M_T(0) = \frac{(2+2i)0 - (2+6i)}{0 - (1+2i)} = \frac{2+6i}{1+2i} = \frac{(2+6i)(1-2i)}{5} = \frac{14+2i}{5},$$

which lies inside this circle in question. Thus, by continuity M_T maps $\mathbb{D}$ to the interior of the circle centred at $z = 3$ of radius 1 (it cannot map another point $z_0 \in \mathbb{D}$ to somewhere outside this circle as the image of the path from 0 to z_0 would have to cross the boundary, but Möbius transformations are conformal). Continuity (and the existence of an inverse) tell us the map must be onto and so

$$M_T : \mathbb{D} \xrightarrow{\sim} \{z \in \mathbb{C} : |z - 3| < 1\}.$$

4.4 The Riemann sphere, revisited

Since we saw earlier that we can identify $\hat{\mathbb{C}}$ with the Riemann sphere S^2 by stereographic projection from the north pole, we can now think of Möbius transformations as bijections from S^2 to S^2 ! In fact, our definitions were carefully chosen so that *Möbius transformations give continuous bijections from S^2 to S^2* .

This is a very fruitful perspective that can unify our picture of Möbius transformations. I encourage you to try to think about Möbius transformations in this way whenever you can!

Remark 4.17. It turns out that the definition of holomorphicity can be extended¹ to functions from S^2 to S^2 . In fact, *Möbius transformations give biholomorphic maps from*

¹The trick is to use the stereographic projection maps to transfer your function to $\mathbb{C}$, and use the definition of holomorphicity there.

S^2 to S^2 . Even better: *the Möbius transformations give all the biholomorphic maps from S^2 to S^2 .*

What do circles and lines in $\hat{\mathbb{C}}$ correspond to in the Riemann sphere? The answer is just circles! We can now add some more to our table of correspondences in (1.3):

<u>In S^2</u>		<u>In $\hat{\mathbb{C}}$</u>
N	$\longleftrightarrow$	∞
S	$\longleftrightarrow$	0
Geometrically:		
Circle not through N	$\longleftrightarrow$	Circle
Circle through N	$\longleftrightarrow$	Line

This also justifies our earlier convention that we think of all lines as passing through infinity.

Remark 4.18. [Advanced] One other nice fact is that stereographic projection itself is conformal! This means two tangent vectors to the sphere at the same point are mapped to two tangent vectors in $\mathbb{C}$ (at the same point) with the same angle.

Example 4.19. Consider the map $f : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$, $f(z) = i\frac{z-i}{z+i}$. Using stereographic projection, we can think of this as a transformation $\hat{f}$ of the sphere S^2 . What is it?

We have that $\hat{f}(N)$ should correspond to $f(\infty) = i$, $\hat{f}(0, 1, 0)$ should correspond to $f(i) = 0$ and $\hat{f}(0, 0, -1)$ should correspond to $f(0) = \frac{1}{i} = -i$. So $\hat{f}$ maps $N, (0, 1, 0), (0, 0, -1)$ to $(0, 1, 0), (0, 0, -1), (0, -1, 0)$. We might guess that $\hat{f}$ is simply a rotation of 90 degrees about the x -axis taking the ‘back’ hemisphere to the bottom/south hemisphere, and this turns out to be correct!

To see why this is true, let’s cheat slightly and assume that the rotation that we’ve guessed corresponds to a Möbius transformation. Then the Möbius transformation is determined by what it does to three points, and we’ve already checked that our guess agrees with the map f at three points ∞, i and 0 .

4.5 Möbius transformations preserving the upper half plane or unit disc

Notation: For a domain $D \subset \mathbb{C}$, denote by $\text{Mob}(D)$ the set of all Möbius transformations f such that $f(D) = D$, i.e. f maps D to D .

Proposition 4.20. [H2H] *Every Möbius transformation mapping $\mathbb{H}$ to $\mathbb{H}$ is of the form M_T with T in the group*

$$\text{SL}_2(\mathbb{R}) := \{T \in \text{Mat}_2(\mathbb{R}) : \det T = 1\}.$$

Conversely, every such Möbius transformation maps $\mathbb{H}$ to $\mathbb{H}$, and hence gives a biholomorphism from $\mathbb{H}$ to $\mathbb{H}$.

In more concise terms, we have

$$f \in \text{Mob}(\mathbb{H}) \iff f = M_T \text{ with } T \in \text{SL}_2(\mathbb{R}).$$

Remark 4.21. This gives us a group homomorphism $\text{SL}_2(\mathbb{R}) \rightarrow \text{Mob}(\mathbb{H})$ mapping $T \mapsto M_T$ and hence also a group homomorphism $\text{SL}_2(\mathbb{R}) \rightarrow \text{Aut}(\mathbb{H})$.

Proof. Any Möbius transformation $f : \mathbb{H} \rightarrow \mathbb{H}$ must map the boundary of $\mathbb{H}$ to the boundary of $\mathbb{H}$. Since this boundary is precisely the real line, $f : \mathbb{R} \cup \{\infty\} \rightarrow \mathbb{R} \cup \{\infty\}$. In particular, it must take the ordered set $\{1, 0, \infty\}$ to $\{x_1, x_2, x_3\}$ for some $x_i \in \mathbb{R} \cup \{\infty\}$. Consider the cross-ratio; assuming $x_i \neq \infty$ we have

$$\begin{aligned} \left(\frac{f(z) - x_2}{f(z) - x_3} \right) \left(\frac{x_1 - x_3}{x_1 - x_2} \right) &= \left(\frac{z - 0}{z - \infty} \right) \left(\frac{1 - \infty}{1 - 0} \right) = z \\ \iff (f(z) - x_2)(x_1 - x_3) &= z(f(z) - x_3)(x_1 - x_2) \\ \iff f(z) &= \frac{x_3(x_1 - x_2)z + x_2(x_3 - x_1)}{(x_1 - x_2)z + (x_3 - x_1)}. \end{aligned}$$

Thus, all coefficients are real, so $T \in \text{GL}_2(\mathbb{R})$.

Furthermore, if $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathbb{R})$ and $z = x + iy$, then

$$\begin{aligned} \text{Im}(M_T(z)) &= \text{Im} \left(\frac{az + b}{cz + d} \right) = \text{Im} \left(\frac{(az + b)(c\bar{z} + d)}{|cz + d|^2} \right) \\ &= \text{Im} \left(\frac{adz + bc\bar{z}}{|cz + d|^2} \right) \\ &= \frac{(ad - bc)y}{|cz + d|^2} = \frac{y \det T}{|cz + d|^2}. \end{aligned}$$

We have $z \in \mathbb{H} \iff y > 0$ so

$$M_T(z) \in \mathbb{H} \iff T \in \text{GL}_2(\mathbb{R}), \det T > 0$$

and so we can replace T by a real matrix of determinant 1 by scaling T by a real number. $\square$

Proposition 4.22. *[D2D] Every Möbius transformation from the unit disk $\mathbb{D}$ to itself is of the form M_T with T in the set*

$$\text{SU}(1, 1) := \left\{ T = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} : \alpha, \beta \in \mathbb{C}, \det T = |\alpha|^2 - |\beta|^2 = 1 \right\}.$$

Conversely, every such Möbius transformation maps $\mathbb{D}$ to $\mathbb{D}$ and hence gives a biholomorphic automorphism of $\mathbb{D}$.

In more concise terms, we have

$$f \in \text{Mob}(\mathbb{D}) \iff f = M_T \text{ with } T \in \text{SU}(1, 1).$$

Remark 4.23. • We obtain a group homomorphism

$$\text{SU}(1, 1) \rightarrow \text{Mob}(\mathbb{D}) \quad T \mapsto M_T$$

and hence also a group homomorphism $\text{SU}(1, 1) \rightarrow \text{Aut}(\mathbb{D})$.

- [Advanced] The group $\text{SU}(1, 1)$ is not the Special Unitary Group $\text{SU}(2)$. It is actually the set of matrices T which preserve the quadratic form $\langle z, w \rangle = \bar{z}_1 w_1 - \bar{z}_2 w_2$ over $\mathbb{C}$; so $\langle Tz, Tw \rangle = \langle z, w \rangle$ for $T \in \text{SU}(1, 1)$ and $z, w \in \mathbb{C}$.

Proof. ($\Rightarrow$): Let $M_T : \mathbb{D} \rightarrow \mathbb{D}$ be a Mob trans and consider the following picture:

$$\begin{array}{ccc} \mathbb{H} & \xrightarrow{f} & \mathbb{H} \\ M_C \downarrow & & \downarrow M_C \\ \mathbb{D} & \xrightarrow{M_T} & \mathbb{D} \end{array}$$

where M_C is the Cayley Map, so $C = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$.

We have that $f := M_{C^{-1}} \circ M_T \circ M_C$ is a Möbius transformation from $\mathbb{H}$ to $\mathbb{H}$. By Prop 4.20 (H2H) we have $f = M_S$ with $S \in \text{SL}_2(\mathbb{R})$, and by Lemma 4.4 we have $S = C^{-1}TC$.

Let $S = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc = 1$ and $a, b, c, d \in \mathbb{R}$. Then you can calculate that

$$T = CSC^{-1} = \frac{1}{2} \begin{pmatrix} (a+d) + i(b-c) & (a-d) - i(c+b) \\ (a-d) + i(c+b) & (a+d) - i(b-c) \end{pmatrix}.$$

Clearly this is of the required form. Finally, we have

$$\det T = \det(CSC^{-1}) = (\det C)^{-1} \cdot \det S \cdot \det C = \det S = 1$$

as required.

($\Leftarrow$): Conversely, if $T \in \text{SU}(1, 1)$ then the same calculation in reverse shows that the matrix $S := C^{-1}TC$ is in $\text{SL}_2(\mathbb{R})$. Thus $M_S : \mathbb{H} \rightarrow \mathbb{H}$ is a Möbius transformation by Prop 4.20 (H2H), and by Lemma 4.4 the map $M_T := M_C \circ M_S \circ M_C^{-1}$ is a Möbius transformation $\mathbb{H} \rightarrow \mathbb{H}$. $\square$

We can actually say much more about what they look like:

Corollary 4.24 (D2D*).

1. Every Möbius transformation f from the unit disk $\mathbb{D}$ to itself can be written

$$f(z) = e^{i\theta} \frac{z - z_0}{\bar{z}_0 z - 1},$$

for some angle θ and $z_0 \in \mathbb{D}$ the unique point such that $f(z_0) = 0$.

2. All Möbius transformations of the unit disk to itself for which $f(0) = 0$ are rotations about 0.

Remark 4.25. The map $g(z) := \frac{z - z_0}{\bar{z}_0 z - 1}$ in the above swaps 0 and z_0 and is an “involution” (see Assignments; that is, $g \circ g = \text{Id}$). Furthermore, the map $z \mapsto e^{i\theta} z$ is a rotation. So all Möbius transformations of the unit disk are given by an involution followed by a rotation.

Proof. 1. By Proposition D2D we have

$$f(z) = \frac{az + b}{\bar{b}z + \bar{a}} = \left(-\frac{a}{\bar{a}}\right) \left(\frac{z - (-\frac{b}{\bar{a}})}{\left(-\frac{\bar{b}}{\bar{a}}\right)z - 1}\right),$$

so $z_0 = -b/a$. Moreover, since $|-a/\bar{a}| = 1$, we must have $-a/\bar{a} = e^{i\theta}$ for some $\theta \in (-\pi, \pi]$. All that remains is to check that $z_0 \in \mathbb{D}$. Since $|a|^2 - |b|^2 = 1$, we have

$$|z_0|^2 - 1 = \left|-\frac{b}{a}\right|^2 - 1 = \frac{|b|^2}{|a|^2} - 1 = -\frac{1}{|a|^2} < 0,$$

and so $|z_0| < 1$ as required.

2. We have

$$f(0) = 0 \iff e^{i\theta} \left(\frac{-z_0}{-1}\right) = 0 \iff z_0 = 0 \iff f(z) = -e^{i\theta} z,$$

and so f is a rotation. □

Finding automorphisms (example)

Find a Möbius transformation f from the closed unit disc onto the closed unit disc taking $\frac{i}{2}$ to 0 and $-i$ to 1. Since $\frac{i}{2} \mapsto 0$, by Corollary D2D* with $z_0 = \frac{i}{2}$ we have

$$f(z) = e^{i\theta} \left(\frac{z - \frac{i}{2}}{-\frac{i}{2}z - 1}\right) = e^{i\theta} \left(\frac{2z - i}{-iz - 2}\right)$$

for some θ . Since $-i \mapsto 1$, we have

$$1 = e^{i\theta} \left(\frac{-2i - i}{-1 - 2} \right) \iff 1 = e^{i\theta} i \iff e^{i\theta} = -i.$$

Thus

$$f(z) = -i \left(\frac{2z - i}{-iz - 2} \right) = i \left(\frac{2z - i}{iz + 2} \right) = \frac{2z - i}{z - 2i} \quad (\text{multiplying top and bottom by } -i).$$

Finding the image of geometric shapes under Möbius transformation (example)

Let F be the geometric figure made up of a line segment from 0 to -1 , a clockwise circular arc (tracing out the unit circle) from -1 to i , then a line segment from i to 0. Find the image of F under the Cayley Map $M_C(z) = \frac{z-i}{z+i}$.

Each section is a segment of a circle or a line, so must be taken to a segment of a circle or a line by Proposition 4.13. First, let us check where the three ‘vertices’ go. We have

$$0 \mapsto \frac{-i}{i} = -1; \quad -1 \mapsto \frac{-1-i}{-1+i} = \frac{(1+i)^2}{2} = i; \quad i \mapsto \frac{i-i}{i+i} = 0.$$

Let’s look at where the line segment from 0 to i goes:

$$\frac{i}{2} \mapsto \frac{\frac{i}{2} - i}{\frac{i}{2} + i} = \frac{-3i}{i} = -3,$$

so the line segment from 0 to i must map to the circle/line from -1 to 0, passing through $-\frac{1}{3}$; that is, it is the line segment from -1 to 0 on the real axis.

Now, if we wanted we could do the same for the other two sections; pick a point on each remaining line/circular arc and see where it maps to, thus determining whether each image is a line segment or a circular arc. However, there is a much quicker method: We know Mob trans are conformal and so preserve angles/orientation - so tracing round the shape we must have the same angles in the image of the figure. So, travelling from $M_C(0) = -1$, when we reach $M_C(i) = 0$ we must turn anticlockwise $\pi/2$ degrees and head towards $M_C(-1) = i$. Thus, the next section must simply be the straight line from 0 to i . Similarly, the final section must be a circular arc from i back to -1 . Thus, the image is just F . Actually, letting $D = \{z \in \mathbb{D} : \pi/2 < \text{Arg}(z) < \pi\}$ be the interior of F , one can check that D maps to D and so $M_{C^{-1}}$ (and therefore M_C) is actually in $\text{Aut}(D)$.

4.6 Finding biholomorphic maps between domains

1. Find a biholomorphism from $D := \{z \in \mathbb{D} : \text{Im}(z) < 0\}$ to $\mathbb{H}$. We build the map in various stages from simpler known maps.

- Step 1: We know the Cayley transform M_C is a map from $\mathbb{H}$ to $\mathbb{D}$, so its inverse $M_{C^{-1}}$ maps $\mathbb{D}$ to $\mathbb{H}$. This is a good thing to start with as it has the right image. To find the image of D under $M_{C^{-1}}$ first consider how it acts on two segments of the boundary:

- The line segment from -1 to 1 (through 0). We have $M_{C^{-1}} = \frac{iz+i}{-z+1}$ so

$$M_{C^{-1}}(-1) = \frac{-i+i}{1+1} = 0, \quad M_{C^{-1}}(1) = \frac{i+i}{-1+1} = \infty, \quad M_{C^{-1}}(0) = \frac{0+i}{0+1} = i.$$

Thus, the line segment from -1 to 1 (through 0) is taken to the line segment from 0 to ∞ (through i); so, the nonnegative imaginary axis. *[You could also have just checked that “ $-d/c$ ” = -1 is on the line, so it must be taken to another line.]*

- The circular arc from -1 to 1 (through $-i$). We have

$$M_{C^{-1}}(-i) = \frac{-i^2+i}{i+1} = 1.$$

Thus, the circular arc from -1 to 1 (through $-i$) is taken to the line segment from 0 to ∞ (through 1); so, the real axis. *[Instead, we could just have used conformality to deduce that this was the image - the angle and its orientation at $z = -1$ must be preserved, so the positive real axis had to be the image.]*

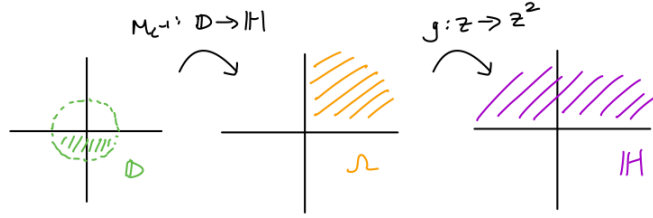
Combined, this tells us that the image of D under $M_{C^{-1}}$ is the first quadrant $\Omega = \{w \in \mathbb{C} : 0 < \text{Arg}(w) < \pi/2\}$ - by conformality the interior must stay on the ‘same side’ of each line segment. *[You could instead explicitly check what happens to an element in D , say $z = -i/2$.]*

- Step 2: We now need a biholomorphic map from Ω to $\mathbb{H}$. We already know of one from earlier: the map $g : z \mapsto z^2 : \Omega \xrightarrow{\sim} \mathbb{H}$. Since compositions of biholomorphic maps are biholomorphic, this gives us the map we want; let $f := g \circ M_{C^{-1}}$ then

$$f : D \xrightarrow{\sim} \mathbb{H}.$$

We can write down the map f explicitly by composing the formulae of the functions:

$$f(z) = g \circ M_{C^{-1}}(z) = g\left(\frac{iz+i}{-z+1}\right) = \left(\frac{iz+i}{-z+1}\right)^2.$$

Figure 4.3: The sequence of transformations to obtain f .

2. Find the image of $\mathbb{D} \setminus \mathbb{R}_{\leq 0} := \{z \in \mathbb{D} : -\pi < \text{Arg} z < \pi\}$ under the map $h(z) = z^{1/2}$. Thus, find a biholomorphic map from $\mathbb{D} \setminus \mathbb{R}_{\leq 0}$ to the unit disc $\mathbb{D}$.

We have (using the principal branch)

$$z = re^{i\theta} (\theta \in (-\pi, \pi), 0 < r < 1) \iff z^{1/2} = \sqrt{r}e^{i\theta/2} \left(\frac{\theta}{2} \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), 0 < \sqrt{r} < 1 \right).$$

We know the complex power function is biholomorphic on $\mathbb{C} - \mathbb{R}_{\leq 0}$, so

$$h : \mathbb{D} \setminus \mathbb{R}_{\leq 0} \xrightarrow{\sim} \{z \in \mathbb{D} : -\pi/2 < \text{Arg} z < \pi/2\}$$

maps $\mathbb{D} \setminus \mathbb{R}_{\leq 0}$ biholomorphically to the (open) right half of the unit disc.

We also know rotations are biholomorphic on $\mathbb{C}$ and so $r(z) = e^{-\pi/2} z = -iz$ maps the right half of the unit disc biholomorphically to the lower half of the unit disc.

Using the previous example f from 1. we can now write down a biholomorphic map from $D := \{z \in \mathbb{D} : \text{Im}(z) < 0\}$ to $\mathbb{H}$, and then map $\mathbb{H}$ (via the Cayley Map) to $\mathbb{D}$. Namely, since $M_C(z) = \frac{z-i}{z+i} : \mathbb{H} \xrightarrow{\sim} \mathbb{D}$ we have $\tilde{f} : \mathbb{D} \setminus \mathbb{R}_{\leq 0} \xrightarrow{\sim} \mathbb{D}$, where

$$\tilde{f}(z) := (M_C \circ f \circ r \circ h)(z) = \frac{\left(\frac{i(-iz^{1/2})+i}{-(-iz^{1/2})+1}\right)^2 - i}{\left(\frac{i(-iz^{1/2})+i}{-(-iz^{1/2})+1}\right)^2 + i} = \frac{(z^{1/2} + i)^2 - i(iz^{1/2} + 1)^2}{(z^{1/2} + i)^2 + i(iz^{1/2} + 1)^2},$$

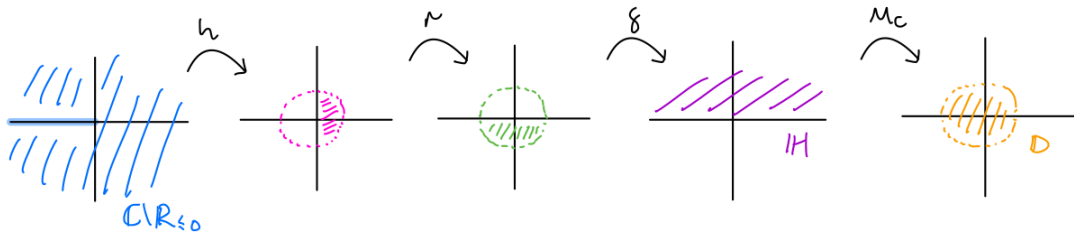
which you can simplify further if you wish.

3. Translations are biholomorphic on $\mathbb{C}$ so for example

$$f(z) = z - i : \{w \in \mathbb{C} : \text{Im}(w) - 1 > \text{Re}(w)\} \xrightarrow{\sim} \{w \in \mathbb{C} : \text{Im}(w) > \text{Re}(w)\}.$$

Since

$$g(z) = e^{-i\frac{\pi}{4}} z = \frac{1-i}{\sqrt{2}} z : \{w \in \mathbb{C} : \text{Im}(w) > \text{Re}(w)\} \xrightarrow{\sim} \mathbb{H}$$


 Figure 4.4: The sequence of transformations to obtain f .

we have that $\tilde{f} := g \circ f$ satisfies

$$\tilde{f}(z) = \{w \in \mathbb{C} : \operatorname{Im}(w) - 1 > \operatorname{Re}(w)\} \xrightarrow{\sim} \mathbb{H}.$$

Explicitly, $\tilde{f}(z) = \left(\frac{1-i}{\sqrt{2}}\right)(z - i)$.

Chapter 5

Notions of Convergence in complex analysis and power series

5.1 Pointwise and uniform convergence

Now we have good knowledge of basic continuous/holomorphic functions and their properties, we would like to be able to construct new continuous functions from them by taking limits.

Given a sequence $\{f_n\}_{n \in \mathbb{N}}$ of functions f_n on a subset of a metric space (say, on a region in $\mathbb{C}$), we want the limit function $f(x) := \lim_{n \rightarrow \infty} f_n(x)$ (if it exists) to behave “nicely”; i.e., we want it to carry over useful properties of the functions f_n . One important case is that of an infinite series of functions

$$\sum_{n=0}^{\infty} f_n := \lim_{N \rightarrow \infty} \sum_{n=0}^N f_n;$$

for example, a power series $\sum_{n=0}^{\infty} a_n(z - z_0)^n$ in $\mathbb{C}$. First we need a concrete notion of what it means for a sequence of functions between two metric spaces to converge. Our initial naive idea is the following version:

Definition 5.1 (Pointwise convergence). Let (X, d_X) and (Y, d_Y) be two metric spaces. A sequence of functions $\{f_n\}_{n \in \mathbb{N}} : X \rightarrow Y$ **converges pointwise (on X) to f** if every $x \in X$ the **limit function** $f(x) := \lim_{n \rightarrow \infty} f_n(x)$ exists in Y . In other words, we have

$\forall x \in X$ and $\forall \epsilon > 0$, there exists $N \in \mathbb{N}$ such that $\forall n > N$, $d_Y(f_n(x), f(x)) < \epsilon$.

Note that N depends on $x \in X$.

Remark 5.2. Note that for any given $x \in X$ we have that $f_n(x)$ is just a sequence of points in Y . This definition is precisely what we get by applying Definition 2.11 to the sequence $f_n(x)$ at every point $x \in X$ independently.

The problem with pointwise convergence (examples)

1. Consider the sequence of functions $f_n = x^n$ on $[0, 1]$. It is easy to see that sequence is pointwise convergent on $[0, 1]$ with limit

$$f(x) = \begin{cases} 0, & \text{if } x < 1. \\ 1, & \text{if } x = 1. \end{cases}$$

This is a sequence of continuous, differentiable functions on a compact set, but the limit function is not continuous!

2. We encounter this issue again in the complex plane. Let $f_n : \mathbb{C} \rightarrow \mathbb{C} : z \mapsto z^n$. We split into three cases.

$|z| < 1$: Pick $z \in \mathbb{D}$ and $\epsilon > 0$. Then, we can certainly find $N \in \mathbb{N}$ such that $|z|^N < \epsilon$ (for example, take any $N > \log \epsilon / \log |z|$). Thus, for every $n > N$ we have

$$|f_n(z) - 0| = |z|^n < |z|^N < \epsilon,$$

and so $\lim_{n \rightarrow \infty} f_n(z) = 0$ in $\mathbb{D}$.

$|z| = 1$: When $|z| = 1$ the point z rotates around the unit circle $\partial\mathbb{D}$ by $\text{Arg}(z)$ every iteration. For any $z \neq 1$ this sequence clearly doesn't converge, but for $z = 1$ we have $\lim_{n \rightarrow \infty} f_n(z) = \lim_{n \rightarrow \infty} 1 = 1$.

$|z| > 1$: Here, the value $|z|^n$ is unbounded, so the limit does not exist.

To conclude, the sequence $f_n(z)$ is not pointwise convergent on $\mathbb{C}$, however, notice that it is pointwise convergent on $\mathbb{D} \cup \{1\}$ with limit function

$$f(z) = \begin{cases} 0, & \text{if } z \in \mathbb{D}. \\ 1, & \text{if } z = 1. \end{cases}$$

Again, this is not continuous.

3. Was the problem that the region where the function converges wasn't open? From the previous examples it seems like this could be the case (it was fine on the interior of balls). No! For example, the sequence of continuous functions $f_n(x) = \arctan(nx)$ converges pointwise on all of $\mathbb{R}$ to the not-continuous function

$$f(x) = \begin{cases} \pi/2, & \text{if } x > 0. \\ 0, & \text{if } x = 0. \\ -\pi/2, & \text{if } x < 0. \end{cases}$$

Key: Pointwise convergence does not preserve continuity. We need a better notion.

Definition 5.3 (Uniform convergence). We say a sequence of functions $\{f_n\}_{n \in \mathbb{N}} : X \rightarrow Y$ **converges uniformly (on X) to (the limit function) f** if we have

$$\forall \epsilon > 0, \quad \exists N \in \mathbb{N} \quad \text{such that} \quad \forall n > N \quad d(f_n(x), f(x)) < \epsilon, \quad \forall x \in X.$$

Note that N here does not depend on the specific choice of $x \in X$ - the same N works for all of them!

Remark 5.4. Uniform convergence trivially implies pointwise convergence (to the same limit).

The following is a generalization of a theorem from Analysis I.

Theorem 5.5. *[Uniform limits of continuous functions are continuous] Let (X, d_X) and (Y, d_Y) be two metric spaces and let $\{f_n\}_{n \in \mathbb{N}} : X \rightarrow Y$ be a sequence of continuous functions that converges uniformly to f on X . Then f is continuous on X .*

Proof. We will show f is continuous at any given $x_0 \in X$. Let $\epsilon > 0$. Since $f_n \rightarrow f$ uniformly we have that

$$\exists N \in \mathbb{N} \quad \text{such that} \quad \forall n > N \quad d_Y(f(x), f_n(x)) < \frac{\epsilon}{3} \quad \forall x \in X. \quad (5.1)$$

We know each f_n is continuous at x_0 , therefore there exists $\delta > 0$ such that

$$\forall x \text{ with } d_X(x, x_0) < \delta, \quad \text{we have} \quad d_Y(f_n(x), f_n(x_0)) < \frac{\epsilon}{3}. \quad (5.2)$$

Pick some fixed $n > N$. Then, for any x with $d_X(x, x_0) < \delta$ we have

$$\begin{aligned} d_Y(f(x), f(x_0)) &\stackrel{(D3)}{\leq} d_Y(f(x), f_n(x)) + d_Y(f_n(x), f(x_0)) \\ &\stackrel{(D3)}{\leq} d_Y(f(x), f_n(x)) + d_Y(f_n(x), f_n(x_0)) + d_Y(f_n(x_0), f(x_0)) \\ &\stackrel{(5.1)}{<} \frac{\epsilon}{3} + d_Y(f_n(x), f_n(x_0)) + \frac{\epsilon}{3} \\ &\stackrel{(5.2)}{<} \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Thus f is continuous at x_0 . □

We would like to develop criteria to determine when sequences of complex valued functions converge uniformly. Here is the first, if we already know the sequence converges pointwise:

Lemma 5.6. *[Test for uniform convergence] Let $f_n : X \rightarrow \mathbb{C}$ be a sequence of functions converging pointwise to a limit function f .*

1. *If $|f_n(x) - f(x)| \leq s_n$ for every $x \in X$, where $\{s_n\}_{n \in \mathbb{N}}$ is some sequence in $\mathbb{R}_{\geq 0}$ (independent of x) with $\lim_{n \rightarrow \infty} s_n = 0$, then f_n converge uniformly to f on X .*
2. *If there exists a sequence $x_n \in X$ such that $|f_n(x_n) - f(x_n)| \geq c$ for some positive constant c , then f_n does not converge uniformly to f on X .*

Proof. Very easy.

1. Let $\epsilon > 0$ and $x \in X$. Since $s_n \rightarrow 0$ there exists $N \in \mathbb{N}$ with $s_n < \epsilon$ for $n > N$. Thus, for $n > N$ we have $|f_n(x) - f(x)| \leq s_n < \epsilon$.
2. The statement obviously implies the negation of uniform continuity; it states that there exists $\epsilon (= c)$ such that for all sufficiently large $n \in \mathbb{N}$ there is a point $x (= x_n) \in X$ for which $|f_n(x) - f(x)| \geq \epsilon$.

□

Checking for uniform convergence (example)

Consider the sequence of functions $f_n(z) = e^z + \frac{1}{n}$ and $g_n(z) = e^z + \frac{z}{n}$. It is easy to see that both functions converge pointwise to the exponential function $f(z) = e^z$: for every fixed $z \in \mathbb{C}$ we have $\lim_{n \rightarrow \infty} (e^z + \frac{1}{n}) = e^z = \lim_{n \rightarrow \infty} (e^z + \frac{z}{n})$.

Notice that $f_n \rightarrow f$ uniformly because for every $z \in \mathbb{C}$ we have

$$|f_n(z) - f(z)| = \left| \left(e^z + \frac{1}{n} \right) - e^z \right| = \frac{1}{n};$$

so we may take $s_n = 1/n$ in Lemma 5.6 part 1..

However $g_n \not\rightarrow f$ uniformly - the giveaway is that in writing

$$|g_n(z) - f(z)| = \left| \left(e^z + \frac{z}{n} \right) - e^z \right| = \frac{|z|}{n};$$

the difference depends on $|z|$, and $|z|$ is unbounded in $\mathbb{C}$. So, we simply notice that upon taking $z_n = n$ that $|g_n(z_n) - f(z_n)| = 1$ and by Lemma 5.6 part 2. there is no uniform convergence.

However, for $\rho > 0$ the sequence g_n does converge uniformly in any ball $B_\rho(0)$ in the complex plane. For then $|g_n(z) - f(z)| = |z|/n < s_n := \rho/n$.

As mentioned, we are mainly interested in the convergence of infinite series.

Theorem 5.7. [Weierstrass M-test] Let $f_n : X \rightarrow \mathbb{C}$ be a sequence of functions such that $|f_n(x)| \leq M_n$ for all $x \in X$ and some sequence of non-negative numbers $\{M_n\}_{n \in \mathbb{N}}$ such that

$$\sum_{n=1}^{\infty} M_n < \infty.$$

Then $S_N(x) = \sum_{n=1}^N f_n(x)$ converges uniformly on X to some limit function $S : X \rightarrow \mathbb{C}$ which we denote by

$$S(x) = \sum_{n=1}^{\infty} f_n(x).$$

In particular, if all the functions $f_n(x)$ are continuous on X then $S(x) = \sum_{n=1}^{\infty} f_n(x)$ is also continuous on X .

Proof. Excluded - see similar proof from Analysis I. Instead, we will prove a stronger result later. $\square$

Determining the convergence of series via the M-test (example)

Show that

$$\sum_{n=1}^{\infty} \frac{|2z|^{3n}}{3^{2n} n^2}$$

converges uniformly on $\overline{\mathbb{D}}$ and let $f(z)$ be its limit function. Is $f(z)$ continuous on $\overline{\mathbb{D}}$?

Note that when $|z| \leq 1$ we have

$$\frac{|2z|^{3n}}{3^{2n} n^2} \leq \frac{2^{3n}}{3^{2n} n^2} = \left(\frac{8}{9}\right)^n \left(\frac{1}{n^2}\right) < \frac{1}{n^2}.$$

We know $\sum_{n \in \mathbb{N}} 1/n^2$ converges, so taking $M_n = 1/n^2$ the Weierstrass M-test implies the function converges uniformly to some limit function f . Furthermore, since every f_n was continuous, then so is f (by Theorem 5.5).

Have we found our ideal definition of convergence? Indeed, we record one other crucial property of uniform convergence from Analysis I:

Theorem 5.8. Assume a sequence of functions $f_n : [a, b] \rightarrow \mathbb{R}$ converge uniformly on an interval $[a, b]$ to some function f , and that $\{f_n\}$ are all continuous. Then

$$\lim_{n \rightarrow \infty} \int_a^c f_n(x) dx = \int_a^c f(x) dx, \quad \text{for all } c \in [a, b].$$

In particular, if $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly on an interval $[a, b]$ and if $\{f_n\}_{n \in \mathbb{N}}$ are continuous for all $n \in \mathbb{N}$ then for any $c \in [a, b]$ we have that:

$$\int_a^c \left(\sum_{n=1}^{\infty} f_n(x) \right) dx = \sum_{n=1}^{\infty} \int_a^c f_n(x) dx.$$

Remark 5.9. If the convergence is not uniform all bets are off! We will heavily use this in term 2 for contour integrals.

Everything looks good, but there is a slight issue to think about.

The problem with uniform convergence (an important example)

Let us return to the sequence of functions $f_n(z) = z^n$. We saw that it converges pointwise on the (open) unit disc $\mathbb{D}$ to the function $f(z) = 0$. Is this convergence uniform?

Let's see if we can find a sequence such that $|f_n(z_n) - f(z_n)| = c$ for some $c > 0$. For simplicity, let's just try to find a sequence of positive real numbers: We want

$$c = |f_n(z_n) - f(z_n)| = |(z_n)^n - 0| = (z_n)^n,$$

so simply take $z_n = c^{1/n}$. We need this sequence to be in $\mathbb{D}$ so let's set $c = 1/2$. Then with $z_n = (1/2)^{1/n} \in \mathbb{D}$, by construction we have

$$|f_n(z_n) - f(z_n)| = \left| \left(\left(\frac{1}{2} \right)^{1/n} \right)^n - 0 \right| = \left(\frac{1}{2} \right)^{n/n} = \frac{1}{2}.$$

It follows from Lemma 5.6 part 2. that the convergence is not uniform. BUT, notice that the limit function $f(z) = 0$ is trivially continuous on all of $\mathbb{D}$ - our notion can't even conclude that the constant function is continuous for a very basic example - we have in some sense been too restrictive!

Key: Uniform convergence is actually too restrictive in terms of preservation of continuity.

5.2 Locally uniform convergence

Definition 5.10 (Locally uniform convergence). Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of functions in a metric space X . We say $\{f_n\}$ **converges locally uniformly (on X) to (the limit function) f** , if for every $x \in X$ there exists an open set $U \subset X$ containing x on which $\{f_n\}_{n \in \mathbb{N}}$ converges uniformly to f .

Theorem 5.11. *Let $\{f_n\}_{n \in \mathbb{N}}$ be a sequence of continuous functions, which converges locally uniformly on X to a limit function f . Then f is continuous on X .*

Proof. Locally uniform convergence gives for any $x \in X$ that the sequence converges uniformly on some open U containing x . Hence f is continuous on U by Theorem 5.5. Hence f is continuous at x . Since x was arbitrary, f is continuous. $\square$

Remark 5.12. This is in many ways the "right" notion for convergence of a sequence of functions. It was championed by Weierstrass. In term 2, we will see that the limit of a locally uniform convergent sequence of holomorphic functions is again holomorphic!!

Locally uniform but not uniform convergence (example)

Again, consider $f_n(z) = z^n$. We know this sequence converges pointwise to $f = 0$ on the (open) unit disc $\mathbb{D}$, but not uniformly. However, the convergence is locally uniform.

For $w \in \mathbb{D}$, we can find $r < 1$ such that $w \in B_r(0)$ (this will be our open set U). Then for all $z \in B_r(0)$, we have $|z^n| < r^n$ and $\lim_{n \rightarrow \infty} r^n = 0$. Hence by Lemma 5.6 part 1. (with $s_n := r^n$), we have uniform convergence in $B_r(0)$.

Note the ‘counter example’ $(1/2)^{1/n}$ from above fails since $\lim_{n \rightarrow \infty} (1/2)^{1/n} = 1$, so at some point the sequence $(1/2)^{1/n}$ leaves the ball $B_r(0)$.

Remark 5.13. Proving locally uniform convergence on every ball $B_r(0)$ in $\mathbb{D}$ (for $0 < r < 1$) is enough to prove that the limit function f is continuous on all of $\mathbb{D}$. In fact, to show a limit function f is continuous on a set X , it turns out to be enough to prove uniform convergence on all compact sets in X :

Theorem 5.14. [Local M-test] *Let X be a metric space and let $f_n : X \rightarrow \mathbb{C}$ be a sequence of continuous functions such that for any $y \in X$, there is an open $U \subset X$ containing y and constants $M_n > 0$ with $\sum_{n=1}^{\infty} M_n < \infty$ and $|f_n(x)| \leq M_n$ for all $x \in U$. Then $\sum_{n=1}^{\infty} f_n$ converges locally uniformly to a continuous function on X .*

Proof. [Hard but worthwhile!] If we prove $\sum_{n=1}^{\infty} f_n$ converges locally uniformly on X , then it converges to a continuous function on X by Theorem 5.11. So we need to show that if $F_k = \sum_{n=1}^k f_n$, then the sequence F_k converges locally uniformly on X , i.e. we want to prove:

There is a function $F : X \rightarrow \mathbb{C}$ such that for any $y \in X$ there is an open U with $y \in U \subset X$ and with F_n converging uniformly to F on U .

Step 1. The sequence is pointwise convergent. Given $y \in X$, the hypothesis implies that there are constants $M_n > 0$ such that $|f_n(y)| \leq M_n$ and $\sum_{n=1}^{\infty} M_n$ converges. Note that

$$|F_k(y)| = \left| \sum_{n=1}^k f_n(y) \right| \leq \sum_{n=1}^k |f_n(y)| \leq \sum_{n=1}^k M_n$$

Since as $k \rightarrow \infty$, the RHS converges, it must be bounded, say by L , so for all k , $|F_k(y)| \leq L$ and hence the sequence $F_k(y)$ is bounded. Hence it is contained in some large closed ball, which is compact by Heine-Borel, and hence there is a subsequence $F_{k_j}(y)$ that converges to some value $F(y)$. We want to prove the whole sequence also converges to $F(y)$.

Also note that for $k_j \geq k$

$$|F_{k_j}(y) - F_k(y)| = \left| \sum_{n=k+1}^{k_j} f_n(y) \right| \leq \sum_{n=k+1}^{k_j} |f_n(y)| \leq \sum_{n=k+1}^{k_j} M_n.$$

Taking the limit as $j \rightarrow \infty$, both the LHS and RHS converge, and we get

$$|F(y) - F_k(y)| \leq \sum_{n=k+1}^{\infty} M_n.$$

Now taking the limit as $k \rightarrow \infty$ of both sides, the RHS goes to zero (any sequence that converges must have tails that go to 0) and we obtain

$$\lim_{k \rightarrow \infty} |F(y) - F_k(y)| = 0.$$

So $F_k(y) \rightarrow F(y)$ as $k \rightarrow \infty$. Repeating this for all y , we get that $F_k \rightarrow F$ pointwise on X . (I.e. we didn't need the subsequences.)

Step 2. Locally uniform convergence. Given y as in the statement we want to prove, let U be the open set provided by hypothesis of the Theorem, and M_n be the provided constants for this U .

By the same estimate as we used before, we have for all x in U (this is the key point!) and for all $\ell > k$

$$|F_\ell(x) - F_k(x)| \leq \sum_{n=k+1}^{\ell} M_n.$$

Taking the limit as $\ell \rightarrow \infty$ we obtain $|F(x) - F_k(x)| \leq \sum_{n=k+1}^{\infty} M_n$ for all $x \in U$. Now since $\lim_{k \rightarrow \infty} \sum_{n=k+1}^{\infty} M_n = 0$, the test for uniform convergence (Lemma 5.6 part 1.) tells us that $F_k \rightarrow F$ uniformly on U , as required. $\square$

5.3 Complex power series

A (complex) power series is an expression of the form

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

with a_n and c complex numbers. Recall that from Analysis I we have:

Theorem 5.15. *There are three cases:*

1. $\sum_{n=0}^{\infty} a_n(z - c)^n$ converges only for $z = c$. ($R = 0$)
2. There exists $R > 0$ (**radius of convergence**) such that

- $\sum_{n=0}^{\infty} a_n(z-c)^n$ converges absolutely for $|z-c| < R$.

(We call $B_R(c)$ the **disc of convergence**.)

- $\sum_{n=0}^{\infty} a_n(z-c)^n$ diverges for $|z-c| > R$.

(Anything can happen on the circle $|z-c| = R$).

3. $\sum_{n=0}^{\infty} a_n(z-c)^n$ converges absolutely for all $z \in \mathbb{C}$. ($R = \infty$)

Proof. See Analysis I. □

Remark 5.16. The radius of convergence can be found using the formula

$$R = \frac{1}{\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}}.$$

We can replace $\limsup$ by $\lim$ when the limit exists.

Moreover, we have the following formula when the limit exist:

$$R = \lim_{n \rightarrow \infty} \frac{|a_n|}{|a_{n+1}|}.$$

We have the following key result:

Theorem 5.17. A power series $\sum_{n=0}^{\infty} a_n(z-c)^n$ with radius of convergence $0 < R \leq \infty$ converges uniformly on any ball $B_r(c)$ with $0 < r < R$. This implies the power series is locally uniformly convergent on its disc of convergence.

Proof. The second statement follows from the first since every $w \in B_R(c)$ is contained in some $B_r(c)$ with $r < R$.

So we will prove the convergence is uniform on $B_r(c)$ given $r < R$. Consider the point $z_0 = c + r$. We have $|z_0 - c| = r < R$, so $z_0 \in B_R(c)$ and the series $\sum_{n=1}^{\infty} a_n(z_0 - c)^n = \sum_{n=1}^{\infty} a_n r^n$ converges absolutely by Theorem 5.15; in other words

$\sum_{n=1}^{\infty} |a_n| r^n < \infty$. So, with $M_n = |a_n| r^n$ it follows that $|a_n(z-c)^n| \leq M_n$ for all $z \in B_r(c)$

and $\sum_{n=0}^{\infty} M_n < \infty$. Thus, the conditions of the M-test (Theorem 5.7) are satisfied and

the series $\sum_{n=1}^{\infty} a_n(z-c)^n$ converges uniformly on $B_r(c)$. □

Remark 5.18. Power series do not converge uniformly in the entire disc of convergence $B_R(c)$. The good news is that this is not necessary because we only need locally uniform convergence to conclude continuity!

Showing a power series converges (example)

For example, consider the series $\sum_{n=0}^{\infty} \frac{z^n}{n!}$. Applying the ratio test, we have

$$L = \lim_{n \rightarrow \infty} \frac{|z^{n+1}|/(n+1)!}{|z^n|/n!} = \lim_{n \rightarrow \infty} \frac{|z|}{n+1} = 0 < 1,$$

for every $z \in \mathbb{C}$; so $R = \infty$ and the series converges absolutely in $\mathbb{C}$. In fact, by Theorem 5.17 it converges locally uniformly in the whole complex plane to a continuous function. This series looks familiar, and actually agrees with the exponential function e^z we've defined. We'll be able to prove this later in the course.

Differentiation/integration of power series

Now we have an idea of what it means for a power series to converge, and where it converges, we would like to know when a power series represents a holomorphic function. In other words, we want to know when we can differentiate (and integrate) power series. We have the following generalization of a result from Analysis I for real power series.

Proposition 5.19. *[Term by term differentiation or integration preserves the radius of convergence] Let $\sum_{n=0}^{\infty} a_n(z-c)^n$ be a power series with radius of convergence $0 < R \leq \infty$. Then the formal derivatives and anti-derivatives*

$$\sum_{n=1}^{\infty} n a_n (z-c)^{n-1} \quad \text{and} \quad \sum_{n=0}^{\infty} \frac{a_n}{n+1} (z-c)^{n+1}$$

have the same radius of convergence R .

Remark 5.20. By 'formal' we mean that the above series (our guess for the derivative and anti-derivative) both define convergent series in the disc $B_R(c)$.

So, we know the series with the expected formulae for derivative and anti-derivative both converge, but do they genuinely represent the derivative and anti-derivative of the original series?

Theorem 5.21. *[Power series can be differentiated term by term in their disc of convergence] Let $\sum_{n=0}^{\infty} a_n(z-c)^n$ be a power series in $\mathbb{C}$, with radius of convergence $0 < R \leq \infty$,*

and let $f : B_R(c) \rightarrow \mathbb{C}$ be the resulting limit function. Then f is holomorphic on $B_R(c)$ with

$$f'(z) = \sum_{n=1}^{\infty} n a_n (z - c)^{n-1} \quad (5.3)$$

for $z \in B_R(c)$.

Proof. For simplicity, assume $c = 0$ (the case of general c is essentially the same).

We wish to show for each $w \in B_R(0)$ that $\lim_{z \rightarrow w} \frac{f(z) - f(w)}{z - w}$ exists and agrees with the expression in (5.3). Since convergence of $f(z)$ is absolute in the disc of convergence (meaning we can reorder sums), we have

$$f(z) - f(w) = \sum_{n=1}^{\infty} a_n (z^n - w^n) = \sum_{n=1}^{\infty} a_n (z - w) q_n(z),$$

where $q_n(z) = \sum_{k=0}^{n-1} w^k z^{n-1-k}$. So, for $z \neq w$

$$\frac{f(z) - f(w)}{z - w} = \sum_{n=1}^{\infty} a_n q_n(z) =: h(z).$$

Note the series makes sense even at w , so we view h as being defined there too.

We claim that the series defining $h(z)$ converges to a continuous function on $B_R(0)$. We'll prove this using the local M-test. Given $z_0 \in B_R(0)$, choose $r < R$ such that $w, z_0 \in B_r(0)$. We need constants M_n for this set $B_r(0)$ that control the terms $a_n q_n(z)$ defining h . We have for $z \in B_r(0)$

$$\begin{aligned} |a_n q_n(z)| &= \left| a_n \sum_{k=0}^{n-1} w^k z^{n-1-k} \right| \leq |a_n| \sum_{k=0}^{n-1} |w|^k |z|^{n-1-k} \\ &< |a_n| \sum_{k=0}^{n-1} r^k r^{n-1-k} = n |a_n| r^{n-1} =: M_n. \end{aligned}$$

We have $\sum_{n=1}^{\infty} M_n = \sum_{n=0}^{\infty} n |a_n| r^{n-1}$ which converges, since by Proposition 5.19 the series $\sum_{n=1}^{\infty} n a_n z^{n-1}$ has radius of convergence R , and so converges absolutely on $B_R(0)$, in particular at the point r . It follows from the Local M-test that the series defining h converges locally uniformly to a continuous function on $B_R(0)$. Hence

$$\lim_{z \rightarrow w} \frac{f(z) - f(w)}{z - w} = \lim_{z \rightarrow w} h(z) = h(w) = \sum_{n=1}^{\infty} a_n q_n(w) = \sum_{n=1}^{\infty} a_n \sum_{k=0}^{n-1} w^k w^{n-1-k} = \sum_{n=1}^{\infty} n a_n w^{n-1}$$

as required. $\square$

Corollary 5.22. *A power series f as in Theorem 5.21 with positive radius of convergence R can be differentiated infinitely many times and $f^{(k)}(z) := \sum_{n=k}^{\infty} k! \binom{n}{k} a_n (z - c)^{n-k}$ for $z \in B_R(c)$. This implies $f^{(k)}(c) = k! a_k$.*

Corollary 5.23 (Power series can be integrated term by term in their disc of convergence). *A power series f as in Theorem 5.21 with positive radius of convergence has a holomorphic antiderivative $F : B_R(c) \rightarrow \mathbb{C}$, that is, $F'(z) = f(z)$, and F is given by $F(z) := \sum_{n=0}^{\infty} \frac{a_n}{n+1} (z - c)^{n+1}$ for $z \in B_R(c)$.*

Convergence of familiar power series (examples)

1. The expected power series for $\sin(z)$, $\cos(z)$, $\sinh(z)$ and $\cosh(z)$ all converge locally uniformly to continuous functions on $\mathbb{C}$, and the derivatives/anti-derivatives match those expected. We will see later that the series genuinely do represent the functions in question.

More generally, when can a holomorphic function be represented via a power series? What about $\log(z)$? It is only defined (and is holomorphic) on $\mathbb{C} \setminus \mathbb{R}_{\leq 0}$ (and is not continuous if extended to the negative real line), so certainly cannot be defined by a power series on $\mathbb{D}$, say. Maybe it works if we change the point at which the series is expanded about (i.e., change the constant c from zero).

Indeed, one can show that $\sum_{n=1}^{\infty} (-1)^{n+1} \frac{z^n}{n}$ converges on $\mathbb{D}$. We expect this to

be the power series converging to $\log(1 + z)$, but what was so special about the taking a power series about the point $c = 1$? It turns out, nothing. We will find out that the holomorphic functions are precisely the functions that can be locally represented by power series at every point in their domain. As a result, we will see by Theorem 5.21 that every holomorphic function is infinitely many times complex differentiable!

2. Consider the *geometric* series $\sum_{n=1}^{\infty} z^n$. By the ratio test this sum converges when $|z| < 1$, so the radius of convergence is $R = 1$ and the series converges to a continuous function on its disc of convergence $\mathbb{D}$. To find the continuous function

the series converges to we need to consider the pointwise limit. Notice that

$$\sum_{n=1}^{\infty} z^n = \lim_{N \rightarrow \infty} \left(\sum_{n=1}^N z^n \right) = \lim_{N \rightarrow \infty} \left(\frac{1 - z^{N+1}}{1 - z} \right) = \frac{1}{1 - z}, \quad (\text{for } |z| < 1).$$

This limit function is defined and is continuous on all of $\mathbb{C} \setminus \{1\}$. So, in some sense the convergence of the series in the complex plane is limited to the unit disc $\mathbb{D}$ because it can't pass the 'pole' at $z = 1$.

3. We may determine the convergence of new power series by substitution. For example, notice that $|z| < 1 \iff |z^2| < 1 \iff |-z^2| < 1$, so by the substitution $z \mapsto z^2$ we have $\sum_{n=1}^{\infty} z^{2n}$ converges locally uniformly on $\mathbb{D}$ to $\frac{1}{1-z^2}$,

and similarly, by the substitution $z \mapsto -z^2$ we have $\sum_{n=1}^{\infty} (-1)^n z^{2n}$ converges locally uniformly to $\frac{1}{1+z^2}$.

These examples give us real insight into the reasons for convergence of the corresponding real power series. Consider the graph of the real function $y = 1/(1-x^2)$. It appears obvious why its real *interval of convergence* is the unit interval - there are asymptotes at $x = \pm 1$ that we can't 'get past' continuously. But on the other hand $y = \frac{1}{1+x^2}$ is a nice smooth looking graph everywhere on the real line - so why on earth is its interval of convergence also restricted to the unit interval? We can now see the answer - the interval of convergence of the real power series is restricted by the disc of convergence of the corresponding complex power series!!! The issue being the poles in the complex plane at $z = \pm i$ that we couldn't see when considering only the real version of the function. So hidden inside the real power series of nice continuous real function is actually some meaningful and significant complex analysis.

Key: Complex analysis can give us new information about real functions!

Chapter 6

Complex integration over contours

6.1 Definition of contour integrals

Our first aim is to give a meaning to the following

$$\int_{\gamma} f(z) dz,$$

where $f : U \rightarrow \mathbb{C}$ is a complex function on an open set U , and γ is a *curve* in U . This will be called the integral of f along the curve γ .

Our first step towards the above is to start by considering complex valued functions of a **real** variable. That is,

$$f : [a, b] \rightarrow \mathbb{C}$$

where $[a, b] \subset \mathbb{R}$. We note that such a function can be written as

$$f = u + iv, \quad \forall t \in [a, b],$$

where $u, v : [a, b] \rightarrow \mathbb{R}$ and given by $u = \operatorname{Re}(f)$ and $v = \operatorname{Im}(f)$. Moreover we note that f is continuous if and only if both u and v are continuous. We then define

$$\int_a^b f(t) dt := \int_a^b u(t) dt + i \int_a^b v(t) dt \in \mathbb{C}.$$

We note that $\operatorname{Re} \left(\int_a^b f(t) dt \right) = \int_a^b u(t) dt$ and $\operatorname{Im} \left(\int_a^b f(t) dt \right) = \int_a^b v(t) dt$.

Example 6.1. We consider the function $f(t) = t + it$. That is $u(t) = v(t) = t$. We compute

$$\begin{aligned}\int_0^1 f(t) dt &= \int_0^1 (t + it) dt = \int_0^1 t dt + i \int_0^1 t dt = \\ &= \left[\frac{t^2}{2} \right]_0^1 + i \left[\frac{t^2}{2} \right]_0^1 = \frac{1}{2} + i \frac{1}{2}.\end{aligned}$$

Before going on we check our definition of integrating complex valued functions of a real parameter is complex linear (as it should be!).

Lemma 6.2.

1. Let f_1 and f_2 be continuous functions from $[a, b]$ to $\mathbb{C}$. Then $\int_a^b (f_1(t) + f_2(t)) dt = \int_a^b f_1(t) dt + \int_a^b f_2(t) dt$.
2. For any complex number $c \in \mathbb{C}$, and continuous function $f : [a, b] \rightarrow \mathbb{C}$, $\int_a^b cf(t) dt = c \int_a^b f(t) dt$.

Proof. (Simple, included for completeness). For part 1, write $f_j = u_j + iv_j$ where u_j and v_j are the real and imaginary parts of f_j . Then using the definition, and the known linearity of real integrals,

$$\begin{aligned}\int_a^b (f_1(t) + f_2(t)) dt &= \int_a^b (u_1(t) + iv_1(t) + u_2(t) + iv_2(t)) dt \\ &= \int_a^b (u_1(t) + u_2(t) + i(v_1(t) + v_2(t))) dt \\ &= \int_a^b (u_1(t) + u_2(t)) dt + i \int_a^b (v_1(t) + v_2(t)) dt \\ &= \int_a^b u_1(t) dt + \int_a^b u_2(t) dt + i \left(\int_a^b v_1(t) dt + \int_a^b v_2(t) dt \right) \\ &= \int_a^b u_1(t) dt + i \int_a^b v_1(t) dt + \int_a^b u_2(t) dt + i \int_a^b v_2(t) dt \\ &= \int_a^b f_1(t) dt + \int_a^b f_2(t) dt.\end{aligned}$$

For part 2 we write $f = u + iv$ where u and v are real, and write $c = x + iy$. Then using linearity of real integrals as before,

$$\begin{aligned}
 \int_a^b cf(t)dt &= \int_a^b (x + iy)(u(t) + iv(t))dt \\
 &= \int_a^b (xu(t) - yv(t) + i(yu(t) + xv(t)))dt \\
 &= \int_a^b (xu(t) - yv(t))dt + i \int_a^b (yu(t) + xv(t))dt \\
 &= x \int_a^b u(t)dt - y \int_a^b v(t)dt + i \left(y \int_a^b u(t)dt + x \int_a^b v(t)dt \right) \\
 &= (x + iy) \left(\int_a^b u(t)dt + i \int_a^b v(t)dt \right) \\
 &= c \int_a^b f(t)dt.
 \end{aligned}$$

□

Recall from Definition 3.8 that a smooth curve in $\mathbb{C}$ is a continuously differentiable function $\gamma : [0, 1] \rightarrow \mathbb{C}$. In fact we can be more general and consider curves $\gamma : [a, b] \rightarrow \mathbb{C}$. For short, we will say that such curves are C^1 .

Remark 6.3. Let us write $\gamma(t) = u(t) + iv(t)$ with $u, v : [a, b] \rightarrow \mathbb{R}$. Then the derivative γ' is defined as

$$\gamma'(t) := u'(t) + iv'(t).$$

One needs to be careful for the meaning of the condition on smoothness at the endpoints. That is, there we demand that the one-sided derivative exists and is continuous from the one side. That is, $\gamma'(b) := \lim_{h \rightarrow 0^-} \frac{u(b+h) - u(b)}{h} + i \lim_{h \rightarrow 0^-} \frac{v(b+h) - v(b)}{h}$ exists and $\lim_{t \rightarrow b^-} \gamma'(t) = \gamma'(b)$.

Examples.

1. An example of a C^1 curve is

$$\gamma : [0, 2\pi] \rightarrow \mathbb{C}, \quad \gamma(\theta) = re^{i\theta},$$

where $r > 0$. (Here we use the letter θ to denote the parameter since it is more often used to denote angles.) We note that this is nothing else than the circle with center the origin and radius r . Moreover we note that as the parameter θ runs from 0 to 2π we run the curve on a anti-clockwise direction. It is easy to see that this is C^1 since

$$\gamma'(\theta) = (re^{i\theta})' = (r \cos'(\theta)) + ir \sin'(\theta) = rie^{i\theta}.$$

2. Another example is given by

$$\gamma : [0, 2\pi] \rightarrow \mathbb{C}, \quad \gamma(\theta) = re^{-i\theta},$$

where $r > 0$. We note the graph of this curve is exactly the same as above, namely the circle centred at the origin and radius r . However here we run the curve on the clockwise direction!

3. We now define the curve

$$\gamma : [0, 2] \rightarrow \mathbb{C},$$

$$\gamma(t) := \begin{cases} t + it, & 0 \leq t \leq 1 \\ t + i & 1 \leq t \leq 2. \end{cases}$$

It is easy to see that γ consists of a straight line connecting the origin and the point $1 + i$ followed by another line connecting the point $1 + i$ and the point $2 + i$. Note that this curve is not C^1 since the derivative at the point $t = 1$ does not exist (from the left it is $1 + i$ and from the right it is 1).

We are now ready to give the following definition.

Definition 6.4. Let $U \subset \mathbb{C}$ be an open set, and let $f : U \rightarrow \mathbb{C}$ be a continuous function. Let $\gamma : [a, b] \rightarrow U \subset \mathbb{C}$ be a C^1 -curve. Then we define the integral of f along the curve γ by

$$\int_{\gamma} f(z) dz := \int_a^b f(\gamma(t))\gamma'(t) dt.$$

Remark 6.5. Note the function f is of a complex variable z . Moreover $f(\gamma(t))$ makes sense since $\gamma([a, b]) \subset U$. Furthermore the condition that γ is C^1 guarantees the existence of $\gamma'(t)$. Finally the function $g(t) := f(\gamma(t))\gamma'(t)$ is a function of the real variable t , so we have already defined $\int_a^b g(t) dt$ above.

Basic Properties. We now note the following basic properties. Assume f is continuous.

1. We have

$$\int_{\gamma} (f_1(z) + f_2(z)) dz = \int_{\gamma} f_1(z) dz + \int_{\gamma} f_2(z) dz$$

Indeed we have by using Lemma 6.2,

$$\begin{aligned} \int_a^b (f_1(\gamma(t)) + f_2(\gamma(t)))\gamma'(t) dt &= \int_a^b f_1(\gamma(t))\gamma'(t) dt + \int_a^b f_2(\gamma(t))\gamma'(t) dt \\ &= \int_{\gamma} f_1(z) dz + \int_{\gamma} f_2(z) dz. \end{aligned}$$

2. For a $c \in \mathbb{C}$ we have

$$\int_{\gamma} cf(z) dz = c \int_{\gamma} f(z) dz.$$

Indeed,

$$\int_{\gamma} cf(z) dz := \int_a^b cf(\gamma(t))\gamma'(t) dt = c \int_a^b f(\gamma(t))\gamma'(t) dt = c \int_{\gamma} f(z) dz$$

where we used Lemma 6.2 for the middle equality.

3. Given $\gamma : [a, b] \rightarrow \mathbb{C}$ we define the curve $(-\gamma) : [-b, -a] \rightarrow \mathbb{C}$, by $(-\gamma)(t) := \gamma(-t)$. Then we have

$$\int_{-\gamma} f(z) dz = - \int_{\gamma} f(z) dz.$$

See Assignments.

Example 6.6.

1. We compute the integral $\int_{\gamma} dz$, where $\gamma(\theta) = re^{i\theta}$ with $0 \leq \theta \leq 2\pi$, with $r > 0$.

$$\begin{aligned} \int_{\gamma} dz &= \int_0^{2\pi} (re^{i\theta})' d\theta = \int_0^{2\pi} rie^{i\theta} d\theta = ri \int_0^{2\pi} e^{i\theta} d\theta = \\ &= ri \left(\int_0^{2\pi} \cos(\theta) d\theta + i \int_0^{2\pi} \sin(\theta) d\theta \right) = ri(0 + 0) = 0. \end{aligned}$$

2. We consider the curve

$$\gamma : [-\pi/2, \pi/2] \rightarrow \mathbb{C}, \quad \gamma(\theta) = 2e^{i\theta}$$

and we compute the integral $\int_{\gamma} \bar{z} dz$. We have

$$\int_{\gamma} \bar{z} dz = \int_{-\pi/2}^{\pi/2} \overline{(2e^{i\theta})} (2e^{i\theta})' d\theta = \int_{-\pi/2}^{\pi/2} 2e^{-i\theta} 2ie^{i\theta} d\theta = 4i \int_{-\pi/2}^{\pi/2} d\theta = 4\pi i.$$

In the last example we may parametrise the same curve differently. For example we may consider $\delta : [0, 1] \rightarrow \mathbb{C}$ where $\delta(t) = 2e^{\pi i(t-1/2)}$. It is easily seen that both γ and δ parametrise the very same curve, namely the semi-circle joining $-2i$ and $2i$ and lying entirely at the positive part of the x -axis. We may then ask whether $\int_{\gamma} f(z) dz = \int_{\delta} f(z) dz$. It turns out that this is indeed the case, as the following lemma shows.

Lemma 6.7 (Reparametrisation of curves). *Let $U \subset \mathbb{C}$ be an open set, $f : U \rightarrow \mathbb{C}$ be continuous, and let $\gamma : [a, b] \rightarrow U$ be a C^1 curve. If $\phi : [a', b'] \rightarrow [a, b]$ with $\phi(a') = a$ and $\phi(b') = b$ is continuously differentiable and we define*

$$\delta : [a', b'] \rightarrow \mathbb{C}, \quad \delta := \gamma \circ \phi \text{ (composition),}$$

then we have

$$\int_{\gamma} f(z) dz = \int_{\delta} f(z) dz.$$

Proof. We have

$$\begin{aligned} \int_{\delta} f(z) dz &= \int_{a'}^{b'} f(\delta(t)) \delta'(t) dt \\ &= \int_{a'}^{b'} f(\gamma(\phi(t))) (\gamma(\phi(t)))' dt \\ &= \int_{a'}^{b'} f(\gamma(\phi(t))) \gamma'(\phi(t)) \phi'(t) dt \end{aligned}$$

We change the variable $s := \phi(t)$ and get that $ds = \phi'(t)dt$, and we use the fact that $\phi(a') = a$ and $\phi(b') = b$. That is, the integral above is equal to

$$\int_a^b f(\gamma(s)) \gamma'(s) ds = \int_{\gamma} f(z) dz.$$

□

In the last example above we have, with $\phi(t) = \pi(t - \frac{1}{2})$ that $\delta = \gamma \circ \phi$. The next step is to consider more general curves, and not only C^1 curves.

Definition 6.8 (Contours). Let $\gamma : [a, b] \rightarrow \mathbb{C}$ be a curve, and suppose that there exist $a = a_0 < a_1 < a_2 < \dots < a_{n-1} < a_n = b$ such that the curves $\gamma_i : [a_{i-1}, a_i] \rightarrow \mathbb{C}$, $i = 1, 2, \dots, n$ defined by $\gamma_i(t) := \gamma(t)$ for $t \in [a_{i-1}, a_i]$ are C^1 curves. Then we say that γ is a piecewise C^1 -curve, or **contour**.

For a contour γ as above, we then define

$$\int_{\gamma} f(z) dz = \sum_{i=1}^n \int_{\gamma_i} f(z) dz.$$

This type of integral is called a **contour integral**. It is the main object of study of this chapter and plays a central role in Complex Analysis.

Remark 6.9. [Technical] Here one should check that the above is well-defined. That is, we may find a different $a = a'_0 < a'_1 < a'_2 < \dots < a'_{m-1} < a'_m = b$, and γ'_j to establish that γ is a contour. But then it is easy to show that

$$\sum_{i=1}^n \int_{\gamma_i} f(z) dz = \sum_{j=1}^m \int_{\gamma'_j} f(z) dz.$$

We will need one more definition regarding curves. We will need to add them. That is, if $\gamma : [a, b] \rightarrow \mathbb{C}$ and $\delta : [c, d] \rightarrow \mathbb{C}$ are two contours with $\gamma(b) = \delta(c)$ then we define the contour $\gamma \cup \delta : [a, b + d - c] \rightarrow \mathbb{C}$

$$(\gamma \cup \delta)(t) := \begin{cases} \gamma(t), & a \leq t \leq b \\ \delta(t + c - b), & b \leq t \leq b + d - c. \end{cases}$$

Then it is easy to see that

$$\int_{\gamma \cup \delta} f(z) dz = \int_{\gamma} f(z) dz + \int_{\delta} f(z) dz.$$

6.2 The Fundamental Theorem of Calculus

The Fundamental Theorem of Calculus in Analysis I tells us that we can calculate real integrals by finding an antiderivative of the quantity being integrated. There is a version of this theorem for contour integrals and, as you might expect, it is very useful in practice.

Theorem 6.10 (Complex Fundamental Theorem of Calculus (FTC)). *Let $U \subset \mathbb{C}$ be an open set and let $F : U \rightarrow \mathbb{C}$ be holomorphic with continuous derivative f . Then for any contour $\gamma : [a, b] \rightarrow U$ we have*

$$\int_{\gamma} f(z) dz = F(\gamma(b)) - F(\gamma(a)).$$

In particular if γ is closed, that is $\gamma(a) = \gamma(b)$, then we have that

$$\int_{\gamma} f(z) dz = 0.$$

Proof. First we consider the case of γ being a C^1 curve. Write $F = u + iv$. We have

$$\int_{\gamma} F'(z) dz = \int_a^b F'(\gamma(t)) \gamma'(t) dt$$

$$\begin{aligned} &= \int_a^b (F(\gamma(t)))' dt = \int_a^b (u(\gamma(t)))' dt + i \int_a^b (v(\gamma(t)))' dt \\ &= [u(\gamma(t))]_a^b + i [v(\gamma(t))]_a^b = F(\gamma(b)) - F(\gamma(a)). \end{aligned}$$

where here we note that $(F(\gamma(t)))' = \frac{dF(\gamma(t))}{dt}$ and for the second to last equality we used the Fundamental Theorem of Calculus from real analysis.

We can now extend the proof to any contour. Indeed all we need to observe is that (with notation as above)

$$\int_{\gamma} F'(z) dz = \sum_{i=1}^n \int_{\gamma_i} F'(z) dz = \sum_{i=1}^n [F(\gamma(t))]_{a_{i-1}}^{a_i} = F(\gamma(a_n)) - F(\gamma(a_0)) = F(\gamma(b)) - F(\gamma(a)).$$

□

Remark 6.11.

- [Looking ahead] Later we will prove that if a function F is holomorphic, then its derivative f is automatically continuous. So that assumption can be dropped from the FTC.
- Note that Theorem 6.10 states that if two curves $\gamma_1 : [a, b] \rightarrow \mathbb{C}$ and $\gamma_2 : [a', b'] \rightarrow \mathbb{C}$ have the same endpoints, that is $\gamma_1(a) = \gamma_2(a')$ and $\gamma_1(b) = \gamma_2(b')$ then

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

under the same hypothesis on F as in Theorem 6.10.

Important notation: Given a domain D such that there exists a bijective contour $\gamma : [a, b] \rightarrow \partial D$ with a continuous inverse $\gamma^{-1} : \partial D \rightarrow [a, b]$ and such that $\gamma'(t) \neq 0$, we define

$$\int_{\partial D} f(z) dz = \int_{\gamma} f(z) dz.$$

This notion is well defined and doesn't depend on γ due to our Reparametrisation of curves lemma, Lemma 6.7. When the boundary has no “end points”, such as the circle, we can extend to above definition to the case where $\gamma : [a, b] \rightarrow \partial D$ has the previously mentioned properties and $\gamma(a) = \gamma(b)$.

For example, for $D = \{z \in \mathbb{C} : |z - c| < r, r > 0, c \in \mathbb{C}\}$ we have that

$$\partial D = \{z \in \mathbb{C} : |z - c| = r, r > 0, c \in \mathbb{C}\},$$

which is the bijective image of the C^1 curve $\gamma : [0, 2\pi) \rightarrow \mathbb{C}$

$$\gamma_{c,r}(\theta) = c + re^{i\theta}.$$

Consequently

$$\int_{|z-c|=r} f(z)dz = \int_{\gamma_{c,r}} f(z)dz = \int_0^{2\pi} f(c + re^{i\theta}) rie^{i\theta} d\theta.$$

Example 6.12. (Important) We now compute the integral

$$\int_{|z|=r} z^n dz := \int_{\gamma} z^n dz, \quad n \in \mathbb{Z}$$

where $\gamma(\theta) = re^{i\theta}$, $0 \leq \theta \leq 2\pi$. We first consider the case $n \neq -1$. Then we have

$$\int_{|z|=r} z^n dz = \int_{\gamma} z^n dz = \int_{\gamma} \left(\frac{z^{n+1}}{n+1} \right)' dz = \left[\frac{z^{n+1}}{n+1} \right]_r = 0,$$

that is we apply the theorem with $F(z) = \frac{z^{n+1}}{n+1}$ and take $U = \mathbb{C} \setminus \{0\}$, which clearly contains the curve γ .

However we cannot do the same with $n = -1$. Indeed, a natural candidate to consider as $F(z)$ would be the logarithm. But we have to make sure that the open set U where this is defined does include the curve γ . For example the principal branch of the logarithm will not do, since we remove the negative real axis, and the contour γ goes through it. Actually if we turn to the actual definition of the integral we see that

$$\int_{\gamma} \frac{1}{z} dz = \int_0^{2\pi} r^{-1} e^{-i\theta} rie^{i\theta} d\theta = i \int_0^{2\pi} d\theta = 2\pi i \neq 0.$$

Note that the integral is not zero. That is, we can conclude that there is no open set U that contains γ such that $\log(z)$ is well-defined, since if it were so, by the theorem above we would get that the integral had to be zero, which is not!

Our next step is to address the following question:

Problem 6.13. Given a function $f : U \rightarrow \mathbb{C}$, provide sufficient conditions such that there exists a holomorphic function $F : U \rightarrow \mathbb{C}$, with derivative $F' = f$. Such an F is usually called the **antiderivative** or **primitive** of f . In other words, is there a converse to the second part of the Fundamental Theorem of Calculus?

For this we need to be able to estimate the modulus of a contour integral, that is, bound $\left| \int_{\gamma} f(z) dz \right|$ from above. We start with a definition.

Definition 6.14 (Length of a contour). Let $\gamma : [a, b] \rightarrow \mathbb{C}$ be a contour. We define the *length* of γ by

$$L(\gamma) := \int_a^b |\gamma'(t)| dt.$$

We now prove

Lemma 6.15 (The Estimation Lemma). *Let $f : U \rightarrow \mathbb{C}$ be continuous and $\gamma : [a, b] \rightarrow U$ be a contour. Then*

$$\left| \int_{\gamma} f(z) dz \right| \leq L(\gamma) \sup_{\gamma} |f|,$$

where

$$\sup_{\gamma} |f| := \sup\{|f(z)| : z \in \gamma\}.$$

Proof. We start by proving the following: Given a continuous function $g : [a, b] \rightarrow \mathbb{C}$, we have that

$$\left| \int_a^b g(t) dt \right| \leq \int_a^b |g(t)| dt. \quad (6.1)$$

Indeed if we write $\int_a^b g(t) dt = re^{i\theta}$, with $r \geq 0$ then we have that

$$\begin{aligned} \left| \int_a^b g(t) dt \right| &= r = \operatorname{Re}(r) = \operatorname{Re} \left(e^{-i\theta} \int_a^b g(t) dt \right) \\ &= \operatorname{Re} \left(\int_a^b e^{-i\theta} g(t) dt \right) = \int_a^b \operatorname{Re} \left(e^{-i\theta} g(t) \right) dt \\ &\leq \int_a^b \left| e^{-i\theta} g(t) \right| dt = \int_a^b |g(t)| dt \end{aligned}$$

Using (6.1) we can now conclude the proof of the lemma. Indeed,

$$\left| \int_{\gamma} f(z) dz \right| = \left| \int_a^b f(\gamma(t)) \gamma'(t) dt \right| \leq \int_a^b |f(\gamma(t)) \gamma'(t)| dt,$$

where we used the above observation for $g(t) = f(\gamma(t)) \gamma'(t)$. But then we have that

$$\int_a^b |f(\gamma(t)) \gamma'(t)| dt \leq \sup_{\gamma} |f| \int_a^b |\gamma'(t)| dt = \sup_{\gamma} |f| L(\gamma).$$

□

The definition of a length of a curve is motivated from the analysis of functions of many variables and the study of integration over curves. There one shows that the infinitesimal length of a curve is exactly $|\gamma'(t)|dt$. This means that if we have a function $g : U \rightarrow \mathbb{R}$ and a curve $\gamma : [a, b] \rightarrow U$ we can define the integral of the *real valued function* g along the curve γ as

$$\int_a^b g(\gamma(t)) |\gamma'(t)| dt.$$

We have a special notation for this:

Notation: Let $U \subseteq \mathbb{C}$ be a open and let $g : U \rightarrow \mathbb{R}$ be continuous. For any $\gamma : [a, b] \rightarrow U$ we define

$$\int_{\gamma} g(z) d|z| := \int_a^b g(\gamma(t)) |\gamma'(t)| dt.$$

In the proof of our estimation lemma we have shown the important inequality:

$$\left| \int_{\gamma} f(z) dz \right| \leq \int_{\gamma} |f(z)| d|z|.$$

Example 6.16. We consider $\gamma : [0, \pi/2] \rightarrow \mathbb{C}$ given by $\gamma(\theta) = 2e^{i\theta}$. Then we get an upper bound for

$$\left| \int_{\gamma} \frac{z+4}{z^3-1} dz \right|$$

We have $L(\gamma) = \int_0^{\pi/2} |(2e^{i\theta})'| d\theta = \pi$, and for $z \in \gamma$, that is $|z| = 2$ we have

$$\left| \frac{z+4}{z^3-1} \right| = \frac{|z+4|}{|z^3-1|} \leq \frac{|z|+4}{||z|^3-1|} = \frac{6}{7}.$$

Here we used the triangle inequality for the numerator and the reverse triangle inequality for the denominator. Then, by Proposition 6.15,

$$\left| \int_{\gamma} \frac{z+4}{z^3-1} dz \right| \leq \frac{6\pi}{7}.$$

We are now ready to answer the problem stated above. The next theorem gives a converse to the second part of the FTC.

Theorem 6.17 (Converse to FTC). *Let $f : D \rightarrow \mathbb{C}$ be continuous on a domain D . If $\int_{\gamma} f(z) dz = 0$ for all closed contours γ in D , then there exists a holomorphic $F : D \rightarrow \mathbb{C}$ such that*

$$F'(z) = f(z).$$

Proof. We fix a point $a_0 \in D$. Then for any other point $w \in D$, by definition of D being a domain there is always a smooth path connecting a_0 and w and we write γ_w for such a path. Then we define the following function

$$F(w) := \int_{\gamma_w} f(z) dz.$$

Step 1. $F(w)$ doesn't depend on the contour we use. We first check that $F(w)$ does not depend on the choice of contour connecting a_0 and w . To see this, if $\tilde{\gamma}(w)$ is another

contour from a_0 to w , then we can consider the closed contour $C = \gamma(w) \cup (-\tilde{\gamma}(w))$ obtained by following $\gamma(w)$ then $\tilde{\gamma}(w)$ in the reverse direction. Then

$$0 = \int_C f(z) dz = \int_{\gamma(w) \cup (-\tilde{\gamma}(w))} f(z) dz = \int_{\gamma(w)} f(z) dz - \int_{\tilde{\gamma}(w)} f(z) dz$$

where the first equality is from the hypothesis that integrals of f over closed contours are zero. This gives

$$\int_{\gamma(w)} f(z) dz = \int_{\tilde{\gamma}(w)} f(z) dz$$

showing that $F(w)$ does not depend on the choice of $\gamma(w)$.

Step 2. F is holomorphic with derivative f . We now claim that F is holomorphic and moreover its derivative is f , that is, for every $w \in D$ we have that $F'(w)$ exists and $F'(w) = f(w)$, i.e.,

$$\lim_{h \rightarrow 0} \frac{F(w+h) - F(w)}{h} = f(w) \quad w \in D.$$

For any given $w \in D$ we fix an $r > 0$ such that $B_r(w) \subset D$. This ball exists since D is open. Then for any $h \in \mathbb{C}$ with $|h| < r$ we consider the straight line δ_h that connects the point w to $w+h$. A parametrisation of such a line is given by

$$\delta_h : [0, 1] \rightarrow D, \quad t \mapsto w + th.$$

We now consider the contour $\gamma_w \cup \delta_h$. This is a contour from a_0 to $w+h$, contained in D , so

$$F(w+h) = \int_{\gamma_w \cup \delta_h} f(z) dz = \int_{\gamma_w} f(z) dz + \int_{\delta_h} f(z) dz = F(w) + \int_{\delta_h} f(z) dz.$$

It is easily seen that $\int_{\delta_h} f(w) dz = hf(w)$ and hence we may rewrite the above equation as

$$F(w+h) = F(w) + hf(w) + \int_{\delta_h} (f(z) - f(w)) dz,$$

or for $h \neq 0$

$$\frac{F(w+h) - F(w)}{h} - f(w) = \frac{1}{h} \int_{\delta_h} (f(z) - f(w)) dz.$$

That is,

$$\left| \frac{F(w+h) - F(w)}{h} - f(w) \right| = \left| \frac{1}{h} \int_{\delta_h} (f(z) - f(w)) dz \right| = \frac{1}{|h|} \left| \int_{\delta_h} (f(z) - f(w)) dz \right|$$

and by Lemma 6.15 (note here it is essential that f is continuous) we have that the above quantity is bounded by

$$\leq \frac{1}{|h|} L(\delta_h) \sup_{z \in \delta_h} |f(z) - f(w)|.$$

But a simple calculation shows that $L(\delta_h) = |h|$ and hence we have that

$$\left| \frac{F(w+h) - F(w)}{h} - f(w) \right| \leq \sup_{z \in \delta_h} |f(z) - f(w)|.$$

Now we take the limit $h \rightarrow 0$, and we observe that the curve δ_h becomes just the point w . (More rigorously, every $z \in \delta_h$ has $|z - w| \leq h$.) Since f is continuous, we get that $\sup_{z \in \delta_h} |f(z) - f(w)| \rightarrow 0$ as $h \rightarrow 0$. Hence we have that $\left| \frac{F(w+h) - F(w)}{h} - f(w) \right| \rightarrow 0$ as $h \rightarrow 0$. That is,

$$\lim_{h \rightarrow 0} \frac{F(w+h) - F(w)}{h} = f(w).$$

Since w was any point in D we have established the theorem. $\square$

Remark 6.18. The above theorem established the existence of a holomorphic anti-derivative of f . However there may be more than one such F , after all we make quite a few choices in the construction of F above. For example we could have picked a different a_0 . However if there is another $\tilde{F}$ such that $\tilde{F}'(w) = f(w)$ for all $w \in D$, then we would have that $F'(w) = \tilde{F}'(w)$ or equivalently $(F - \tilde{F})'(w) = 0$. But we know from Theorem 3.12 that if a function defined over a domain has a zero derivative we have that the function is just a constant. That is, there is some $c \in \mathbb{C}$ such that $\tilde{F} = F + c$.

Appendix A

Additional lemmas from the Gappy notes

In this short appendix we write the lemmas which were added to the Gappy note to help explain complex maps and Möbius transformation (proofs that were given in the lecture are provided).

A.1 Complex function

Lemma A.1.

- The map z^n injectively takes an angular segment of length $\frac{2\pi}{n}$ which is open at one end and closed at the other from a circle of radius r to the entire circle of radius r^n . If the above segment is closed, or its size is larger than $\frac{2\pi}{n}$ the image is no longer injective.
- The map z^n injectively takes a ray of angle θ to a ray of angle $n\theta \bmod 2\pi$. Consequently, the map z^n injectively takes the wedge bounded by rays of angles θ_1 and θ_2 to the wedge bounded by rays of angles $n\theta_1 \bmod 2\pi$ and $n\theta_2 \bmod 2\pi$ if $n|\theta_1 - \theta_2| < 2\pi$. When $n|\theta_1 - \theta_2| \geq 2\pi$ the image is the entire complex plane (not injectively).
- We can define n different n -th “roots” which are inverses to the map z^n . We can write them all in the form

$$z^{\frac{1}{n}} = |z|^{\frac{1}{n}} e^{i\left(\frac{\text{Arg}(z)}{n} + \frac{2\pi k}{n}\right)}$$

with $k = 0, \dots, n-1$ (we remove the k index from the map, but one need to specify which k we are talking about). Note that for a fixed k the n -th root of z takes $\mathbb{C}$ into a anti clockwise rotation by $\frac{\pi}{n}$ of $R_{n,k-1}$.

Lemma A.2.

- The map e^z injectively takes a segment of length 2π which is open at one end and closed at the other from the line $x = c$ to the entire circle of radius e^c . If the above segment is closed, or its size is larger than 2π the image is no longer injective.
- The map e^z injectively takes the line $y = c$ to the ray of angle $c \bmod 2\pi$ without the origin.
- The map e^z injectively takes the set $\{z \in \mathbb{C} : \theta < \operatorname{Im}(z) \leq \theta + 2\pi\}$ to $\mathbb{C}^*$ where $\mathbb{C}^* := \mathbb{C} \setminus \{0\}$.

Lemma A.3. We have that for $z = x + iy$:

$$\sin(z) = \sin(x) \cosh(y) + i \cos(x) \sinh(y).$$

$$\cos(z) = \cos(x) \cosh(y) - i \sin(x) \sinh(y).$$

$$\sinh(z) = -i \sin(iz) = \sinh(x) \cos(y) + i \cosh(x) \sin(y)$$

$$\cosh(z) = \cos(iz) = \cosh(x) \cos(y) + i \sinh(x) \sin(y).$$

In addition we have that for all $z \in \mathbb{C}$

$$\sin(z)^2 + \cos(z)^2 = 1, \quad \cosh(z)^2 - \sinh(z)^2 = 1.$$

Lemma A.4.

- The map $\sin(z)$ injectively takes a segment of length 2π which is open at one end and closed at the other from the line $y = c$ to an ellipse when $c \neq 0$. When $c = 0$ we get the “squished” ellipse $[-1, 1] \times \{0\}$ and the map is injective for a segment of length π on which the real map $\sin(x)$ is injective.
- The map $\sin(z)$ injectively takes the line $x = c$ to a one sided hyperbola when $c \neq \pi k$ and $c \neq \frac{\pi}{2} + \pi k$ for all $k \in \mathbb{Z}$. When $c = \pi k$ or $c = \frac{\pi}{2} + \pi k$ for some $k \in \mathbb{Z}$ we get the “squashed” hyperbolas:
 - $\{0\} \times i\mathbb{R}$ if $c = \pi k$. The map is injective in this case.
 - $[1, \infty) \times \{0\}$ if $c = \frac{\pi}{2} + 2\pi k$. The map is injective on $\{z \in \mathbb{C} : z = \frac{\pi}{2} + 2\pi k + iy, y \geq 0\}$ and $\{z \in \mathbb{C} : z = \frac{\pi}{2} + 2\pi k + iy, y \leq 0\}$.
 - $(-\infty, 1] \times \{0\}$ if $c = -\frac{\pi}{2} + 2\pi k$. The map is injective on $\{z \in \mathbb{C} : z = -\frac{\pi}{2} + 2\pi k + iy, y \geq 0\}$ and $\{z \in \mathbb{C} : z = -\frac{\pi}{2} + 2\pi k + iy, y \leq 0\}$.

Similar statements can be done for $\cos(z)$ using the identity $\cos(z) = \sin(z + \frac{\pi}{2})$ (a shift by $\frac{\pi}{2}$ on the x -axis), and $\sinh(z)$ and $\cosh(z)$ using the identities

$$\sinh(z) = -i \sin(iz), \quad \cosh(z) = \cos(iz)$$

(rotations by $\pm \frac{\pi}{2}$ of the variable and image in the complex plane).

Lemma A.5.

- A branch of the map $\log(z)$ injectively takes a ray without the origin of angle $\theta \neq 0$, measured with respect to the branch cut, to the line $y = \theta$.
- A branch of the map $\log(z)$ takes a concentric circle of radius r , minus the branch cut, to a segment of length 2π on the line $x = \log r$ with its lowest point given by the angle that define the branch cut.

A.2 Möbius transformations**Lemma A.6.**

- The map $1/z$ injectively takes the set $\mathbb{D} \setminus \{0\} = \{z \in \mathbb{C} : 0 < |z| < 1\}$ to $\mathbb{D}^c = \{z \in \mathbb{C} : |z| > 1\}$. When considering $1/z$ over $\hat{\mathbb{C}}$ we find that it injectively takes $\mathbb{D}$ to $\mathbb{D}^c \cup \{\infty\}$. Similarly, $1/z$ injectively takes the set $\mathbb{D}^c$ to $\mathbb{D} \setminus \{0\}$ and on $\hat{\mathbb{C}}$ it takes $\mathbb{D}^c \cup \{\infty\}$ to $\mathbb{D}$.
- The map $1/z$ injectively takes the ray at angle θ without the origin to the ray at angle $-\theta$ without the origin. On $\hat{\mathbb{C}}$ the statement remains the same by adding the origin and ∞ to the rays.

Lemma A.7 (Building blocks of Möbius transformations). *Any Möbius transformation is composition of the following four type of maps:*

1. *Shift:* $z \rightarrow z + b$ for some $b \in \mathbb{C}$.
2. *Stretch/Compression:* $z \rightarrow \lambda z$ for some $\lambda \in \mathbb{R}_{>0}$.
3. *Rotation:* $z \rightarrow e^{i\theta} z$ for some $\theta \in (-\pi, \pi]$.
4. *Inversion:* $z \rightarrow \frac{1}{z}$.

Remark A.8. Note that in terms of matrices we have that:

1. A shift corresponds to the matrix $T = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$.
2. Stretch/Compression corresponds to the matrix $T = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}$.
3. Rotation corresponds to the matrix $T = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & 1 \end{pmatrix}$.

4. Inversion corresponds to the matrix $T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Proof. We start by noticing that for any $a, b \in \mathbb{C}$

$$az + b = |a| e^{i\text{Arg}(a)} z + b$$

which is a composition of rotation, stretch/compression and shift. In matrix form:

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} |a| & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\text{Arg}(a)} & 0 \\ 0 & 1 \end{pmatrix}.$$

This shows that the maps $az + b$ and $\frac{1}{az+b}$ can be achieved as composition of the given four maps.

We notice that when $a = 0$ we have that, as $b \neq 0$,

$$\frac{az + b}{cz + d} = \frac{1}{\frac{c}{b}z + \frac{d}{b}}$$

which we have just considered. We can assume moving forward that $a \neq 0$. We notice that

$$\frac{1}{\alpha \left(\frac{1}{az+b} \right) + \beta} = \frac{az + b}{a\beta z + (\alpha + b\beta)}$$

when we don't have that $a = b = 0$ and $\alpha = \beta = 0$. We want to have

$$c = a\beta, \quad d = \alpha + b\beta.$$

Since $a \neq 0$ we can choose $\beta = c/a$ and $\alpha = d - \frac{bc}{a} = \frac{ad-bc}{a}$. In other words when $a \neq 0$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{ad-bc}{a} & \frac{c}{a} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$$

□

Lemma A.9. *The map $f(z) = \frac{1}{z}$ takes a circle that doesn't pass through the origin to a circle. Moreover, if B is an open ball whose closure doesn't contain the origin then $f(B)$ is once again an open ball. Consequently, if $D \subseteq \mathbb{C}$ is a set then for any $z_0 \in \mathbb{D} \setminus \{0\}$ for which there exists $\epsilon > 0$ with $B_\epsilon(z_0) \subseteq D$ we have that there exists $\delta > 0$ such that $B_\delta(f(z_0)) \subseteq f(D)$ (i.e. f takes any non-zero interior point of D to an interior point of $f(D)$).*

Proof. The equation of a general circle in the complex plane is given by

$$|z - a| = r$$

or

$$(z - a)(\bar{z} - \bar{a}) = r^2$$

which is

$$z\bar{z} - a\bar{z} - \bar{a}z + (|a|^2 - r^2) = 0.$$

Assuming $z \neq 0$ and motivated by trying to recover $1/z$ we see that by dividing the above by $\frac{1}{z\bar{z}} = \frac{1}{|z|^2} > 0$ we get

$$1 - \bar{a}\frac{1}{z} - a\frac{1}{\bar{z}} + (|a|^2 - r^2)\frac{1}{z\bar{z}} = 0.$$

If we call $w = 1/z$ we find that the above is:

$$(|a|^2 - r^2)|w|^2 - \bar{a}\bar{w} - aw + 1 = 0.$$

Which looks like an equation for a circle! We just want to divide by $|a|^2 - r^2$. This is allowed since if $|a| = r$ the original circle would have passed through $z = 0$ which we didn't allow. We conclude that

$$|w|^2 - \frac{\bar{a}}{(|a|^2 - r^2)}\bar{w} - \frac{a}{(|a|^2 - r^2)}w + \frac{1}{(|a|^2 - r^2)} = 0.$$

We conclude that a circle $|z - a| = r$ which doesn't pass through the origin is mapped under $\frac{1}{z}$ to

$$\left| \frac{1}{z} - \frac{\bar{a}}{(|a|^2 - r^2)} \right| = \sqrt{\left| \frac{\bar{a}}{(|a|^2 - r^2)} \right|^2 - \frac{1}{(|a|^2 - r^2)}} = \frac{r}{|a|^2 - r^2},$$

which is again a circle. What happens to the ball $|z - a| < r$? If we assume that $\bar{D}$ does not contain the problematic point, $z = 0$, we must have that $r < |a|$ or $|a|^2 - r^2 > 0$. Following the steps above we see that the open ball is mapped to

$$(|a|^2 - r^2)|w|^2 - \bar{a}\bar{w} - aw + 1 < 0$$

which, since $|a|^2 - r^2 > 0$, can be written as

$$|w|^2 - \frac{\bar{a}}{(|a|^2 - r^2)}\bar{w} - \frac{a}{(|a|^2 - r^2)}w + \frac{1}{(|a|^2 - r^2)} < 0$$

or

$$\left| \frac{1}{z} - \frac{\bar{a}}{(|a|^2 - r^2)} \right| < \frac{r}{|a|^2 - r^2},$$

i.e. we get another ball!

To show the second part of the lemma we assume that $z_0 \neq 0$ is an interior point of D , i.e. there exists $\epsilon > 0$ such that $B_\epsilon(z_0) \subseteq D$. to show that $f(z_0)$ is an interior point of $f(D)$ we need to find $\delta > 0$ such that $B_\delta(1/z_0) \subseteq f(D)$. It is enough to show if $|z - z_0| < \epsilon$ for ϵ small enough which satisfies $\epsilon < |z_0|$ (the ball doesn't encircle the origin) we can find $\delta > 0$ such that

$$B_\delta\left(\frac{1}{z_0}\right) \subseteq B_{\frac{\epsilon}{|z_0|^2 - \epsilon^2}}\left(\frac{\overline{z_0}}{|z_0|^2 - \epsilon^2}\right).$$

Since

$$\left|\frac{1}{z_0} - \frac{\overline{z_0}}{|z_0|^2 - \epsilon^2}\right| = \frac{\epsilon^2}{|z_0|(|z_0|^2 - \epsilon^2)}$$

we see that if $\left|\frac{1}{z} - \frac{1}{z_0}\right| < \delta$

$$\begin{aligned} \left|\frac{1}{z} - \frac{\overline{z_0}}{|z_0|^2 - \epsilon^2}\right| &< \delta + \frac{\epsilon^2}{|z_0|(|z_0|^2 - \epsilon^2)} \\ &= \left(\frac{\delta(|z_0|^2 - \epsilon^2)}{\epsilon} + \frac{\epsilon}{|z_0|}\right) \frac{\epsilon}{|z_0|^2 - \epsilon^2}. \end{aligned}$$

A choice of ϵ and δ such that

$$\frac{\delta(|z_0|^2 - \epsilon^2)}{\epsilon} + \frac{\epsilon}{|z_0|} < 1$$

will be enough to conclude the proof (for instance $\epsilon < |z_0|/2$ and $\delta = \frac{\epsilon}{2(|z_0|^2 - \epsilon^2)}$). $\square$

Lemma A.10. *Let $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathbb{C})$ and let $D \subseteq \mathbb{C}$ be a open set. Then $M_T(D \setminus \{-\frac{d}{c}\})$ is open. Together with the fact that Möbius maps are bijections on $\hat{\mathbb{C}}$ we have that*

$$M_T\left(\partial D \setminus \left\{-\frac{d}{c}\right\}\right) = \partial M_T(D) \setminus \{\infty\}.$$

When $c = 0$ we consider $-\frac{d}{c}$ as ∞ which is outside of $\mathbb{C}$ and consequently $A \setminus \{\infty\} = A$ for any $A \subseteq \mathbb{C}$.

Proof. Using the fact that every Möbius transformation is a composition of shifts, stretch/compression, rotation, and inversion, we see that we only need to show that the claim is true for each such map (while being careful with the point $-d/c$). The claim is straightforward for shifts, stretch/compression, and rotation - there is also no point of infinity. We really only need to check inversions which is covered by our previous lemma.

The claim about the boundary follows from the above but required a bit more care. $\square$

A.3 Biholomorphic domains - revisited

Function	Maps the domain	To the domain	Biholomorphic?
$z^n, \quad n \in \mathbb{N}$	$\{0 < \text{Arg}(z) < \theta\} \cup \{0\}$	$\mathbb{C},$ $\{0 < \text{Arg}(z) < n\theta\} \cup \{0\},$ $\{0 < \text{Arg}(z) \leq \pi\} \cup \{-\pi < \text{Arg}(z) < n\theta - 2\pi\} \cup \{0\},$ if $\theta > \frac{2\pi}{n},$ if $\theta \leq \frac{2\pi}{n},$	When $\theta \leq \frac{2\pi}{n}.$
$z^n, \quad n \in \mathbb{N}$	$\{\theta_1 < \text{Arg}(z) < \theta_2\}$ with $-\pi < \theta_1 < \theta_2 \leq \pi$	a rotation by $n\theta_1$ of the image of $\{0 < \text{Arg}(z) < \theta_2 - \theta_1\}$ by z^n	When $\theta_2 - \theta_1 \leq \frac{2\pi}{n}.$
Principle branch of $z^{\frac{1}{n}}, n \in \mathbb{N}$	$\{\theta_1 < \text{Arg}(z) < \theta_2\} \cup \{0\}$ with $-\pi < \theta_1 < \theta_2 \leq \pi$	$\left\{\frac{\theta_1}{n} < \text{Arg}(z) < \frac{\theta_2}{n}\right\} \cup \{0\}$	Yes
e^z	$\{\theta_1 < \text{Im}(z) < \theta_2\}$	$\mathbb{C}^*,$ $\{\theta_1 \bmod 2\pi < \text{Arg}(z) < \theta_2 \bmod 2\pi\} \setminus \{0\},$ if $\theta_2 - \theta_1 > 2\pi,$ if $\theta_2 - \theta_1 \leq 2\pi.$	When $\theta_2 - \theta_1 \leq 2\pi$
$\text{Log}(z)$	$\{\theta_1 < \text{Arg}(z) < \theta_2\} \setminus \{0\}$ with $-\pi < \theta_1 < \theta_2 \leq \pi$	$\{\theta_1 < \text{Im}(z) < \theta_2\} \setminus \{0\}$	Yes
$\frac{1}{z}$	$\{r_1 < z < r_2\}$	$\left\{\frac{1}{r_2} < z < \frac{1}{r_1}\right\}$	Yes
Cayley map $f(z) = \frac{z-i}{z+i}$	$\mathbb{D}$	$\mathbb{H}_L$	Yes
Cayley map $f(z) = \frac{z-i}{z+i}$	$\mathbb{H}$	$\mathbb{D}$	Yes
Cayley map $f(z) = \frac{z-i}{z+i}$	$\mathbb{H} \cap \mathbb{D} = \{ z < 1, \text{Im}(z) > 0\}$	$\mathbb{H}_L \cap \mathbb{D} = \{ z < 1, \text{Re}(z) < 0\}$	Yes
Cayley map $f(z) = \frac{z-i}{z+i}$	$\mathbb{H} \cap \mathbb{H}_R = \{\text{Re} z > 0, \text{Im}(z) > 0\}$	$\mathbb{H}_L \cap \mathbb{D} = \{ z < 1, \text{Im}(z) < 0\}$	Yes
Möbius map $M_T(z)$ with real matrix T such that $\det T = 1$	$\mathbb{H}$	$\mathbb{H}$	Yes
$f(z) = e^{i\theta} \frac{z-z_0}{\bar{z}-\bar{z}_0}$ with $\theta \in (-\pi, \pi]$ and $z_0 \in \mathbb{D}$	$\mathbb{D}$	$\mathbb{D}$	Yes
General Möbius transformation	domains bounded by circlines	domains bounded by circline	Yes

where we have used the notation of $\mathbb{H}_- := \{z \in \mathbb{C} : \text{Im} z < 0\}$ for the lower half plane.