## 18. Homotopies, simply-connected domains and Cauchy's theorem

A crucial point in our proof of Cauchy's theorem for a triangle was that the interior of the triangle was entirely contained in the open set on which our holomorphic function $f$ was defined. In general however, given a closed curve, it is not always easy to say what we mean by the "interior" of the curve. In fact there is a famous theorem, known as the Jordan Curve Theorem, which resolves this problem, but to prove it would take us too far afield. Instead we will take a slightly different strategy: in fact we will take two different approaches: the first using the notion of homotopy and the second using the winding number. For the homotopy approach, rather than focusing only on closed curves and their "interiors" we consider arbitrary curves and study what it means to deform one to another.

**Definition 18.1.** Suppose that $U$ is an open set in $\mathbb{C}$ and $a, b \in U$. If $\eta\colon [0,1] \to U$ and $\gamma\colon [0,1] \to U$ are paths in $U$ such that $\gamma(0) = \eta(0) = a$ and $\gamma(1) = \eta(1) = b$, then we say that $\gamma$ and $\eta$ are *homotopic* in $U$ if there is a continuous function $h\colon [0,1] \times [0,1] \to U$ such that

$$h(0,s) = a, \quad h(1,s) = b$$
$$h(t,0) = \gamma(t), \quad h(t,1) = \eta(t).$$

One should think of $h$ as a family of paths in $U$ indexed by the second variable $s$ which continuously deform $\gamma$ into $\eta$.

A special case of the above definition is when $a = b$ and $\gamma$ and $\eta$ are closed paths. In this case there is a constant path $c_a\colon [0,1] \to U$ going from $a$ to $b = a$ which is simply given by $c_a(t) = a$ for all $t \in [0,1]$. We say a closed path starting and ending at a point $a \in U$ is *null homotopic* if it is homotopic to the constant path $c_a$. One can show that the relation "$\gamma$ is homotopic to $\eta$" is an equivalence relation, so that any path $\gamma$ between $a$ and $b$ belongs to a unique equivalence class, known as its homotopy class.

**Definition 18.2.** Suppose that $U$ is a domain in $\mathbb{C}$. We say that $U$ is *simply connected* if for every $a, b \in U$, any two paths from $a$ to $b$ are homotopic in $U$.

**Lemma 18.3.** *Let $U$ be a convex open set in $\mathbb{C}$. Then $U$ is simply connected. Moreover if $U_1$ and $U_2$ are homeomorphic, then $U_1$ is simply connected if and only if $U_2$ is.*

*Proof.* Suppose that $\gamma\colon [0,1] \to U$ and $\eta\colon [0,1] \to U$ are paths starting and ending at $a$ and $b$ respectively for some $a, b \in U$. Then for $(s,t) \in [0,1] \times [0,1]$ let

$$h(t,s) = (1-s)\gamma(t) + s\eta(t)$$

It is clear that $h$ is continuous and one readily checks that $h$ gives the required homotopy. For the moreover part, if $f\colon U_1 \to U_2$ is a homeomorphism then it is clear that $f$ induces a bijection between continuous paths in $U_1$ to those in $U_2$ and also homotopies in $U_1$ to those in $U_2$, so the claim follows. $\square$

*Remark* 18.4. (*Non-examinable*) In fact, with a bit more work, one can show that any starlike domain $D$ is also simply-connected. The key is to show that a domain is simply-connected if all closed paths starting and ending at a given point $z_0 \in D$ are null-homotopic. If $D$ is star-like with respect to $z_0 \in D$, then if $\gamma\colon [0,1] \to D$ is a closed path with $\gamma(0) = \gamma(1) = z_0$, it follows $h(s,t) = z_0 + s(\gamma(t) - z_0)$ gives a homotopy between $\gamma$ and the constant path $c_{z_0}$.

Thus we see that we already know many examples of simply connected domains in the plane, such as disks, ellipsoids, half-planes. The second part of the above lemma also allows us to produce non-convex examples:

**Example 18.5.** Consider the domain

$$D_{\eta,\epsilon} = \{z \in \mathbb{C} : z = re^{i\theta} : \eta < r < 1, 0 < \theta < 2\pi(1-\epsilon)\},$$

where $0 < \eta, \epsilon < 1/10$ say, then $D_{\eta,\epsilon}$ is clearly not convex, but it is the image of the convex set $(0,1) \times (0, 1-\epsilon)$ under the map $(r,\theta) \mapsto re^{2\pi i\theta}$. Since this map has a continuous (and even differentiable) inverse, it follows $D_{\eta,\epsilon}$ is simply-connected. When $\eta$ and $\epsilon$ are small, the boundary of this set, oriented anti-clockwise, is a version of what is called a *key-hole contour*.

We are now ready to state our extension of Cauchy's theorem. The proof is given in the Appendices.

**Theorem 18.6.** *Let $U$ be a domain in $\mathbb{C}$ and $a, b \in U$. Suppose that $\gamma$ and $\eta$ are paths from $a$ to $b$ which are homotopic in $U$ and $f \colon U \to \mathbb{C}$ is a holomorphic function. Then*

$$\int_\gamma f(z)dz = \int_\eta f(z)dz.$$

*Remark* 18.7. Notice that this theorem is really more general than the previous versions of Cauchy's theorem we have seen – in the case where a holomorphic function $f \colon U \to \mathbb{C}$ has a primitive the conclusion of the previous theorem is of course obvious from the Fundamental theorem of Calculus[45], and our previous formulations of Cauchy's theorem were proved by producing a primitive for $f$ on $U$. One significance of the homotopy form of Cauchy's theorem is that it applies to domains $U$ even when there is no primitive for $f$ on $U$.

**Theorem 18.8.** *Suppose that $U$ is a simply-connected domain, let $a, b \in U$, and let $f \colon U \to \mathbb{C}$ be a holomorphic function on $U$. Then if $\gamma_1, \gamma_2$ are paths from $a$ to $b$ we have*

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

*In particular, if $\gamma$ is a closed oriented curve we have $\int_\gamma f(z)dz = 0$, and hence any holomorphic function on $U$ has a primitive.*

*Proof.* Since $U$ is simply-connected, any two paths from from $a$ to $b$ are homotopic, so we can apply Theorem 18.6. For the last part, in a simply-connected domain any closed path $\gamma \colon [0,1] \to U$, with $\gamma(0) = \gamma(1) = a$ say, is homotopic to the constant path $c_a(t) = a$, and hence $\int_\gamma f(z)dz = \int_{c_a} f(z)dz = 0$. The final assertion then follows from the Theorem 15.21. □

**Example 18.9.** If $U \subseteq \mathbb{C}\backslash\{0\}$ is simply-connected, the previous theorem shows that there is a holomorphic branch of $[\mathrm{Log}(z)]$ defined on all of $U$ (since any primitive for $f(z) = 1/z$ will be such a branch).

*Remark* 18.10. Recall that in Definition 16.8 we called a domain $D$ in the complex plane *primtive* if every holomorphic function $f \colon D \to \mathbb{C}$ on it had a primitive. Theorem 18.8 shows that any simply-connected domain is primitive. In fact the converse is also true – any primitive domain is necessarily simply-connected. Thus the term "primitive domain" is in fact another name for a simply-connected domain.

## 19. Winding numbers

Suppose that $\gamma \colon [0,1] \to \mathbb{C}$ is a closed path which does not pass through 0. We would like to give a rigorous definition of the number of times $\gamma$ "goes around the origin". Roughly speaking, this will be the change in argument $\arg(\gamma(t))$, and therein lies the difficulty, since $\arg(z)$ cannot be defined continuously on all of $\mathbb{C}\backslash\{0\}$. The next Proposition shows that we *can* however always define the argument as a continuous function *of the parameter* $t \in [0,1]$:

[45]Indeed the hypothesis that the paths $\gamma$ and $\eta$ are homotopic is irrelevant when $f$ has a primitive on $U$.

**Proposition 19.1.** *Let $\gamma\colon [0,1] \to \mathbb{C}\backslash\{0\}$ be a path. Then there is continuous function $a\colon [0,1] \to \mathbb{R}$ such that*

$$\gamma(t) = |\gamma(t)|e^{2\pi i a(t)}.$$

*Moreover, if $a$ and $b$ are two such functions, then there exists $n \in \mathbb{Z}$ such that $a(t) = b(t) + n$ for all $t \in [0,1]$.*

*Proof.* By replacing $\gamma(t)$ with $\gamma(t)/|\gamma(t)|$ we may assume that $|\gamma(t)| = 1$ for all $t$. Since $\gamma$ is continuous on a compact set, it is uniformly continuous, so that there is a $\delta > 0$ such that $|\gamma(s) - \gamma(t)| < \sqrt{3}$ for any $s, t$ with $|s - t| < \delta$. Choose an integer $n > 0$ such that $n > 1/\delta$ so that on each subinterval $[i/n, (i+1)/n]$ we have $|\gamma(s) - \gamma(t)| < \sqrt{3}$. Now on any half-plane in $\mathbb{C}$ we may certainly define a holomorphic branch of $[\mathrm{Log}(z)]$ (simply pick a branch cut along a ray in the opposite half-plane) and hence a continuous argument function, and if $|z_1| = |z_2| = 1$ and $|z_1 - z_2| < \sqrt{3}$, then the angle between $z_1$ and $z_2$ is at most $2\pi/3$. It follows there exists a continuous functions $a_i\colon [i/n, (i+1)/n] \to \mathbb{R}$ such that $\gamma(t) = e^{2\pi i a_i(t)}$ for $t \in [i/n, (i+1)/n]$. Now since $e^{2\pi i a_i(i/n)} = e^{2\pi i a_{i-1}(i/n)}$ $a_{i-1}(i/n)$ and $a_i(i/n)$ differ by an integer. Thus we can successively adjust the $a_i$ for $i > 1$ by an integer (as if $\gamma(t) = e^{2\pi i a_i(t)}$ then $\gamma(t) = e^{2\pi i (a(t)+n)}$ for any $n \in \mathbb{Z}$) to obtain a continuous function $a\colon [0,1] \to \mathbb{C}$ such that $\gamma(t) = e^{2\pi i a(t)}$ as required. Finally, the uniqueness statement follows because $e^{2\pi i (a(t) - b(t))} = 1$, hence $a(t) - b(t) \in \mathbb{Z}$, and since $[0,1]$ is connected it follows $a(t) - b(t)$ is constant as required. ☐

**Definition 19.2.** If $\gamma\colon [0,1] \to \mathbb{C}\backslash\{0\}$ is a closed path and $\gamma(t) = |\gamma(t)|e^{2\pi i a(t)}$ as in the previous lemma, then since $\gamma(0) = \gamma(1)$ we must have $a(1) - a(0) \in \mathbb{Z}$. This integer is called the *winding number* $I(\gamma, 0)$ of $\gamma$ around 0. It is uniquely determined by the path $\gamma$ because the function $a$ is unique up to an integer. By tranlation, if $\gamma$ is any closed path and $z_0$ is not in the image of $\gamma$, we may define the winding number $I(\gamma, z_0)$ of $\gamma$ about $z_0$ in the same fashion. Explicitly, if $\gamma$ is a closed path with $z_0 \notin \gamma^*$ then let $t\colon \mathbb{C} \to \mathbb{C}$ be given by $t(z) = z - z_0$ and define $I(\gamma, z_0) = I(t \circ \gamma, 0)$.

*Remark* 19.3. Note that if $\gamma\colon [0,1] \to U$ where $0 \notin U$ and there exists a holomorphic branch $L\colon U \to \mathbb{C}$ of $[\mathrm{Log}(z)]$ on $U$, then $I(\gamma, 0) = 0$. Indeed in this case we may define $a(t) = \Im(L(\gamma(t)))$, and since $\gamma(0) = \gamma(1)$ it follows $a(1) - a(0) = 0$ as claimed. Note also that the definition of the winding number only requires the closed path $\gamma$ to be continuous, not piecewise $C^1$. Of course as usual, we will mostly only be interested in piecewise $C^1$ paths, as these are the ones along which we can integrate functions.

We now see that the winding number has a natural interpretation in term of path integrals: Note that if $\gamma$ is piecewise $C^1$ then the function $a(t)$ is also piecewise $C^1$, since any branch of the logarithm function is in fact differentiable where it is defined, and $a(t)$ is locally given as $\Im(\log(\gamma(t))$ for a suitable branch.

**Lemma 19.4.** *Let $\gamma$ be a piecewise $C^1$ closed path and $z_0 \in \mathbb{C}$ a point not in the image of $\gamma$. Then the winding number $I(\gamma, z_0)$ of $\gamma$ around $z_0$ is equal to*

$$\frac{1}{2\pi i}\int_\gamma \frac{dz}{z - z_0}.$$

*In particular, if $\gamma_1, \gamma_2$ are two paths which are homotopic via a homotopy $h\colon [0,1] \times [0,1] \to \mathbb{C}\backslash\{z_0\}$ then $I(\gamma_1, z_0) = I(\gamma_2, z_0)$.*

*Proof.* If $\gamma\colon [0,1] \to \mathbb{C}$ we may write $\gamma(t) = z_0 + r(t)e^{2\pi i a(t)}$ (where $r(t) = |\gamma(t) - z_0| > 0$ is continuous and the existence of $a(t)$ is guaranteed by Proposition 19.1). Then we have

$$\begin{aligned}\int_\gamma \frac{dz}{z-z_0} &= \int_0^1 \frac{1}{r(t)e^{2\pi i a(t)}} . \left(r'(t) + 2\pi i r(t)a'(t)\right) e^{2\pi i a(t)} dt \\ &= \int_0^1 r'(t)/r(t) + 2\pi i a'(t) dt = [\log(r(t)) + 2\pi i a(t)]_0^1 \\ &= 2\pi i(a(1) - a(0)),\end{aligned}$$

since $r(1) = r(0) = |\gamma(0) - z_0|$. The last sentence now follows easily from Theorem 18.6. □

*Remark* 19.5. Note that in particular the integral formula for the winding number of course gives another proof that it only depends on the path $\gamma$. One can of course prove more directly that the winding number of two homotopic paths is constant – intuitively it is clear since it is a "continuously varying" function of the path, and thus as it is integer valued, it must be constant on homotopy classes of paths.

**Lemma 19.6.** *Let $U$ be an open set in $\mathbb{C}$ and let $\gamma\colon [0,1] \to U$ be a closed path. If $f(z)$ is a continuous function on $\gamma^*$ then the function*

$$I_f(\gamma, w) = \int_\gamma \frac{f(z)}{z-w} dz,$$

*is holomorphic*[46] *in $z$. In particular, if $f(z) = 1$ this shows that the function $z \mapsto I(\gamma, z)$ is a continuous function on $\mathbb{C}\backslash\gamma^*$, and hence, since it is integer-valued, it is constant on the connected components of $\mathbb{C}\backslash\gamma^*$.*

*Proof.* Fix $z_0 \in \mathbb{C}\backslash\gamma^*$. Since $\mathbb{C}\backslash\gamma^*$ is open, it suffices to show that $I_f(\gamma, z)$ is holomorphic in $B(z_0, r) \subseteq \mathbb{C}\backslash\gamma^*$ for some $r > 0$. Translating if necessary we may assume that $z_0 = 0$. Now since $0 \notin \gamma^*$ we have $2r = \min\{|\gamma(t)| : t \in [0,1]\} > 0$. We claim that $I_f(\gamma, z)$ is holomorphic in $B(0.r)$. Indeed if $w \in B(0,r)$ and $z \in \gamma^*$ it follows that $|w/z| < 1/2$. Moreover, since $\gamma^*$ is compact, $M = \sup\{|f(z)| : z \in \gamma^*\}$ is finite, and hence

$$|f(z).w^n/z^{n+1}| < \frac{M}{2r}(1/2)^n, \quad \forall z \in \gamma^*.$$

It follows from the Weierstrass $M$-test that the series $\sum_{n=0}^\infty \frac{f(z).w^n}{z^{n+1}}$ converges uniformly on $\gamma^*$ to $f(z)/(z-w)$. Thus for all $w \in B(0,r)$ we have

$$I_f(\gamma, w) = \int_\gamma \frac{f(z)dz}{z-w} = \sum_{n=0}^\infty \left(\int_\gamma \frac{f(z)}{z^{n+1}} dz\right) w^n,$$

hence $I_f(\gamma, w)$ is given by a power series in $B(0,r)$ and hence is holomorphic there as required.

Finally, if $f = 1$, then since $I_1(\gamma, z) = I(\gamma, z)$ is integer-valued, it follows it must be constant on any connected component of $\mathbb{C}\backslash\gamma^*$ as required. □

*Remark* 19.7. If $\gamma$ is a closed path then $\gamma^*$ is compact and hence bounded. Thus there is an $R > 0$ such that the connected set $\mathbb{C}\backslash B(0,R) \cap \gamma^* = \emptyset$. It follows that $\mathbb{C}\backslash\gamma^*$ has exactly one unbounded connected component. Since

$$\left|\int_\gamma \frac{d\zeta}{\zeta - z}\right| \leq \ell(\gamma). \sup_{\zeta\in\gamma^*} |1/(\zeta - z)| \to 0$$

as $z \to \infty$ it follows that $I(\gamma, z) = 0$ on the unbounded component of $\mathbb{C}\backslash\gamma^*$.

[46] This Lemma is an easy generalization of Lemma 16.18 – essentially the same proof works.

**Definition 19.8.** Let $\gamma\colon [0,1] \to \mathbb{C}$ be a closed path. We say that a point $z$ is in the *inside*[47] of $\gamma$ if $z \notin \gamma^*$ and $I(\gamma, z) \neq 0$. The previous remark shows that the inside of $\gamma$ is a union of bounded connected components of $\mathbb{C}\backslash\gamma^*$. (We don't, however, know that the inside of $\gamma$ is necessarily non-empty.)

**Example 19.9.** Suppose that $\gamma_1\colon [-\pi,\pi] \to \mathbb{C}$ is given by $\gamma_1 = 1 + e^{it}$ and $\gamma_2\colon [0, 2\pi] \to \mathbb{C}$ is given by $\gamma_2(t) = -1 + e^{-it}$. Then if $\gamma = \gamma_1 \star \gamma_2$, $\gamma$ traverses a figure-of-eight and it is easy to check that the inside of $\gamma$ is $B(1,1) \cup B(-1,1)$ where $I(\gamma, z) = 1$ for $z \in B(1,1)$ while $I(\gamma, z) = -1$ for $z \in B(-1,1)$.

*Remark* 19.10. It is a theorem, known as the *Jordan Curve Theorem*, that if $\gamma\colon [0,1] \to \mathbb{C}$ is a simple closed curve, so that $\gamma(t) = \gamma(s)$ if and only if $s = t$ or $s,t \in \{0,1\}$, then $\mathbb{C}\backslash\gamma^*$ is the union of precisely one bounded and one unbounded component, and on the bounded component $I(\gamma, z)$ is either 1 or $-1$. If $I(\gamma, z) = 1$ for $z$ on the inside of $\gamma$ we say $\gamma$ is postively oriented and we say it is negatively oriented if $I(\gamma, z) = -1$ for $z$ on the inside.

The definition of winding number allows us to give another version of Cauchy's integral formula (sometimes called the *winding number* or *homology* form of Cauchy's theorem).

**Theorem 19.11.** *Let $f\colon U \to \mathbb{C}$ be a holomorphic function and let $\gamma\colon [0,1] \to U$ be a closed path whose inside lies entirely in $U$, that is $I(\gamma, z) = 0$ for all $z \notin U$. Then we have, for all $z \in U\backslash\gamma^*$,*

$$\int_\gamma f(\zeta)d\zeta = 0; \qquad \int_\gamma \frac{f(\zeta)}{\zeta - z}d\zeta = 2\pi i I(\gamma, z) f(z), \quad \forall z \in U\backslash\gamma^*.$$

*Moreover, if $U$ is simply-connected and $\gamma\colon [a,b] \to U$ is any closed path, then $I(\gamma, z) = 0$ for any $z \notin U$, so the above identities hold for all closed paths in such $U$.*

*Remark* 19.12. The "moreover" statement in fact just uses the fact that a simply-connected domain is primitive: if $D$ is a domain and $w \notin D$, then the function $1/(z - w)$ is holomorphic on all of $D$, and hence has a primitive on $D$. It follows $I(\gamma, w) = 0$ for any path $\gamma$ with $\gamma^* \subseteq D$.

*Remark* 19.13. This version of Cauchy's theorem has a natural extension: instead of integrating over a single closed path, one can integrate over formal sums of closed paths, which are known as *cycles*: if $a \in \mathbb{N}$ and $\gamma_1, \ldots, \gamma_k$ are closed paths and $a_1, \ldots, a_k$ are complex numbers (we will usually only consider the case where they are integers) then we define the integral around the formal sum $\Gamma = \sum_{i=1}^k a_i\gamma_i$ of a function $f$ to be

$$\int_\Gamma f(z)dz = \sum_{i=1}^k a_i \int_{\gamma_i} f(z)dz.$$

Since the winding number can be expressed as an integral, this also gives a natural defintion of the winding number for such $\Gamma$: explicitly $I(\Gamma, z) = \sum_{i=1}^k a_i I(\gamma_i, z)$. If we write $\Gamma^* = \gamma_1^* \cup \ldots \cup \gamma_k^*$ then $I(\Gamma, z)$ is defined for all $z \notin \Gamma^*$. The winding number version Cauchy's theorem then holds (with the same proof) for cycles in an open set $U$, where we define the inside of a cycle to be the set of $z \in \mathbb{C}$ for which $I(\Gamma, z) \neq 0$.

Note that if $z$ is inside $\Gamma$ then it must be the case that $z$ is inside some $\gamma_i$, but the converse is not necessarily the case: it may be that $z$ lies inside some of the $\gamma_i$ but does not lie inside $\Gamma$. One natural way in which cycles arise are as the boundaries of an open subsets of the plane: if $\Omega$ is an domain in the plane, then $\partial\Omega$, the boundary of $\Omega$ is often a *union* of curves rather than a single curve[48]. For example if $r < R$ then $\Omega = B(0,R)\backslash\bar{B}(0,r)$ has a boundary which is a union of

[47]The term *interior* of $\gamma$ might be more natural, but we have already used this in the first part of the course to mean something quite different.

[48]Of course in general the boundary of an open set need not be so nice as to be a union of curves at all.

two concentric circles. If these circles are oriented correctly, then the "inside" of the cycle $\Gamma$ which they form is precisely $\Omega$ (see the discussion of Laurent series below for more details). Thus the origin, although inside each of the circles $\gamma(0,r)$ and $\gamma(0,R)$, is not inside $\Gamma$. The cycles version of Cauchy's theorem is thus closest to Green's theorem in multivariable calculus.

As a first application of this new form of Cauchy's theorem, we establish the *Laurent expansion* of a function which is holomorphic in an annulus. This is a generalization of Taylor's theorem, and we already saw it in the special case of a function with a pole singularity.

**Definition 19.14.** Let $0 < r < R$ be real numbers and let $z_0 \in \mathbb{C}$. An open *annulus* is a set

$$A = A(r, R, z_0) = B(z_0, R) \backslash \bar{B}(z_0, r) = \{z \in \mathbb{C} : r < |z - z_0| < R\}.$$

If we write (for $s > 0$) $\gamma(z_0, s)$ for the closed path $t \mapsto z_0 + se^{2\pi it}$ then notice that the inside of the cycle $\Gamma_{r,R,z_0} = \gamma(z_0, R) - \gamma(z_0, r)$ is precisely $A$, since for any $s$, $I(\gamma(z_0, s), z)$ is 1 precisely if $z \in B(z_0, s)$ and 0 otherwise.

**Theorem 19.15.** *Suppose that $0 < r < R$ and $A = A(r, R, z_0)$ is an annulus centred at $z_0$. If $f : U \to \mathbb{C}$ is holomorphic on an open set $U$ which contains $\bar{A}$, then there exist $c_n \in \mathbb{C}$ such that*

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n, \quad \forall z \in A.$$

*Moreover, the $c_n$ are unique and are given by the following formulae:*

$$c_n = \frac{1}{2\pi i} \int_{\gamma_s} \frac{f(z)}{(z - z_0)^{n+1}} dz,$$

*where $s \in [r, R]$ and for any $s > 0$ we set $\gamma_s(t) = z_0 + se^{2\pi it}$.*

*Proof.* By translation we may assume that $z_0 = 0$. Since $A$ is the inside of the cycle $\Gamma_{r,R,z_0}$ it follows from the winding number form of Cauchy's integral formula that for $w \in A$ we have

$$2\pi i f(w) = \int_{\gamma_R} \frac{f(z)}{z - w} dz - \int_{\gamma_r} \frac{f(z)}{z - w} dz$$

But now the result follows in the same way as we showed holomorphic functions were analytic: if we fix $w$, then, for $|w| < |z|$ we have $\frac{1}{z-w} = \sum_{n=0}^{\infty} w^n / z^{n+1}$, converging uniformly in $z$ in $|z| > |w| + \epsilon$ for any $\epsilon > 0$. It follows that

$$\int_{\gamma_R} \frac{f(z)}{z - w} dz = \int_{\gamma_R} \sum_{n=0}^{\infty} \frac{f(z) w^n}{z^{n+1}} dz = \sum_{n \geq 0} \left( \int_{\gamma_R} \frac{f(z)}{z^{n+1}} dz \right) w^n.$$

for all $w \in A$. Similarly since for $|z| < |w|$ we have[49] $\frac{1}{w-z} = \sum_{n \geq 0} z^n / w^{n+1} = \sum_{n=-1}^{-\infty} w^n / z^{n+1}$, again converging uniformly on $|z|$ when $|z| < |w| - \epsilon$ for $\epsilon > 0$, we see that

$$\int_{\gamma_r} \frac{f(z)}{w - z} dz = \int_{\gamma_r} \sum_{n=-1}^{-\infty} f(z) w^n / z^{n+1} dz = \sum_{n=-1}^{-\infty} \left( \int_{\gamma_r} \frac{f(z)}{z^{n+1}} dz \right) w^n.$$

Thus taking $(c_n)_{n \in \mathbb{Z}}$ as in the statement of the theorem, we see that

$$f(w) = \frac{1}{2\pi i} \int_{\gamma_R} \frac{f(z)}{z - w} dz - \frac{1}{2\pi i} \int_{\gamma_r} \frac{f(z)}{z - w} dz = \sum_{n \in \mathbb{Z}} c_n z^n,$$

as required. To see that the $c_n$ are unique, one checks using uniform convergence that if $\sum_{n \in \mathbb{Z}} d_n z^n$ is any series expansion for $f(z)$ on $A$, then the $d_n$ must be given by the integral formulae above.

[49] Note the sign change.

Finally, to see that the $c_n$ can be computed using any circular contour $\gamma_s$, note that if $r \leq s_1 < s_2 \leq R$ then $f/(z-z_0)^{n+1}$ is holomorphic on the inside of $\Gamma = \gamma_{s_2} - \gamma_{s_1}$, hence by the homology form of Cauchy's theorem $0 = \int_\Gamma f(z)/(z-z_0)^{n+1}dz = \int_{\gamma_{s_2}} f(z)/(z-z_0)^{n+1}dz - \int_{\gamma_{s_1}} f(z)/(z-z_0)^{n+1}dz$. □

*Remark* 19.16. Note that the above proof shows that the integral $\int_{\gamma_R} \frac{f(z)}{z-w}dz$ defines a holomorphic function of $w$ in $B(z_0, R)$, while $\int_{\gamma_r} \frac{f(z)}{z-w}dz$ defines a holomorphic function of $w$ on $\mathbb{C}\backslash B(z_0, r)$. Thus we have actually expressed $f(w)$ on $A$ as the difference of two functions which are holomorphic on $B(z_0, R)$ and $\mathbb{C}\backslash\bar{B}(z_0, r)$ respectively.

**Definition 19.17.** Let $f: U\backslash S \to \mathbb{C}$ be a function which is holomorphic on a domain $U$ except at a discrete set $S \subseteq U$. Then for any $a \in S$ the previous theorem shows that for $r > 0$ sufficiently small, we have

$$f(z) = \sum_{n\in\mathbb{Z}} c_n(z-a)^n, \quad \forall z \in B(a, r)\backslash\{a\}.$$

We define

$$P_a(f) = \sum_{n=-1}^{-\infty} c_n(z-a)^n,$$

to be the *principal part* of $f$ at $a$. This generalizes the previous definition we gave for the principal part of a meromorphic function. Note that the proof of Theorem 19.15 shows that the series $P_a(f)$ is uniformly convergent on $\mathbb{C}\backslash B(a, r)$ for all $r > 0$, and hence defines a holomorphic function on $\mathbb{C}\backslash\{a\}$.

We can now prove one of the most useful theorems of the course – it is extremely powerful as a method for computing integrals, as you will see this course and many others.

**Theorem 19.18.** *(Residue theorem): Suppose that $U$ is an open set in $\mathbb{C}$ and $\gamma$ is a path whose inside is contained in $U$, so that for all $z \notin U$ we have $I(\gamma, z) = 0$. Then if $S \subset U$ is a finite set such that $S \cap \gamma^* = \emptyset$ and $f$ is a holomorphic function on $U\backslash S$ we have*

$$\frac{1}{2\pi i}\int_\gamma f(z)dz = \sum_{a\in S} I(\gamma, a)Res_a(f)$$

*Proof.* For each $a \in S$ let $P_a(f)(z) = \sum_{n=-1}^{-\infty} c_n(a)(z-a)^n$ be the principal part of $f$ at $a$, a holomorphic function on $\mathbb{C}\backslash\{a\}$. Then by definition of $P_a(f)$, the difference $f - P_a(f)$ is holomorphic at $a \in S$, and thus $g(z) = f(z) - \sum_{a\in S} P_a(f)$ is holomorphic on all of $U$. But then by Theorem 19.11 we see that $\int_\gamma g(z)dz = 0$, so that

$$\int_\gamma f(z)dz = \sum_{a\in S}\int_\gamma P_a(f)(z)dz$$

But by the proof of Theorem 19.15, the series $P_a(f)$ converges uniformly on $\gamma^*$ so that

$$\begin{aligned}\int_\gamma P_a(f)dz &= \int_\gamma \sum_{n=-1}^{-\infty} c_n(a)(z-a)^n = \sum_{n=1}^{\infty}\int_\gamma \frac{c_{-n}(a)dz}{(z-a)^n} \\ &= \int_\gamma \frac{c_{-1}(a)dz}{z-a} = I(\gamma, a)\mathrm{Res}_a(f),\end{aligned}$$

since for $n > 1$ the function $(z-a)^{-n}$ has a primitive on $\mathbb{C}\backslash\{a\}$. The result follows. □

*Remark* 19.19. In practice, in applications of the residue theorem, the winding numbers $I(\gamma, a)$ will be simple to compute in terms of the argument of $(z-a)$ – in fact most often they will be 0 or $\pm 1$ as we will usually apply the theorem to integrals around simple closed curves.

The Residue theorem gives us a very powerful technique for computing many kinds of integrals. In this section we give a number of examples of its application.

**Example 20.1.** Consider the integral $\int_0^{2\pi} \frac{dt}{1+3\cos^2(t)}$. If we let $\gamma$ be the path $t \mapsto e^{it}$ and let $z = e^{it}$ then $\cos(t) = \Re(z) = \frac{1}{2}(z+\bar{z}) = \frac{1}{2}(z+1/z)$. Thus we have

$$\frac{1}{1+3\cos^2(t)} = \frac{1}{1+3/4(z+1/z)^2} = \frac{1}{1+\frac{3}{4}z^2+\frac{3}{2}+\frac{3}{4}z^2} = \frac{4z^2}{3+10z^2+3z^4},$$

Finally, since $dz = izdt$ it follows

$$\int_0^{2\pi} \frac{dt}{1+3\cos^2(t)} = \int_\gamma \frac{-4iz}{3+10z^2+3z^4}dz.$$

Thus we have turned our real integral into a contour integral, and to evaluate the contour integral we just need to calculate the residues of the meromorphic function $g(z) = \frac{-4iz}{3+10z^2+3z^4}$ at the poles it has inside the unit circle. Now the poles of $g(z)$ are the zeros of the polynomial $p(z) = 3+10z^2+3z^4$, which are at $z^2 \in \{-3, -1/3\}$. Thus the poles inside the unit circle are at $\pm i/\sqrt{3}$. In particular, since $p$ has degree 4 and has four roots, they must all be simple zeros, and so $g$ has simple poles at these points. The residue at a simple pole $z_0$ can be calculated as the limit $\lim_{z \in z_0}(z-z_0)g(z)$, thus we see (compare with Remark 17.14) that

$$\begin{aligned}\mathrm{Res}_{z=\pm i/\sqrt{3}}(g(z)) &= \lim_{z\to\pm i/\sqrt{3}} \frac{-4iz(z-\pm i/\sqrt{3})}{3+10z^2+3z^4} = (\pm 4/\sqrt{3}).\frac{1}{p'(\pm i/\sqrt{3})} \\ &= (\pm 4/\sqrt{3}).\frac{1}{20(\pm i/\sqrt{3})+12(\pm i/\sqrt{3})^3} = 1/4i.\end{aligned}$$

It now follows from the Residue theorem that

$$\int_0^{2\pi} \frac{dt}{1+3\cos^2(t)} = 2\pi i\big(\mathrm{Res}_{z=i/\sqrt{3}}((g(z)) + \mathrm{Res}_{z=-i/\sqrt{3}}(g(z))\big) = \pi.$$

*Remark* 20.2. Often we are interested in integrating along a path which is not closed or even finite, for example, we might wish to understand the integral of a function on the positive real axis. The residue theorem can still be a power tool in calculating these integrals, provided we complete the path to a closed one in such a way that we can control the extra contribution to the integral along the part of the path we add.

**Example 20.3.** If we have a function $f$ which we wish to integrate over the whole real line (so we have to treat it as an improper Riemann integral) then we may consider the contours $\Gamma_R$ given as the concatenation of the paths $\gamma_1\colon [-R,R] \to \mathbb{C}$ and $\gamma_2\colon [0,1] \to \mathbb{C}$ where

$$\gamma_1(t) = -R + t; \quad \gamma_2(t) = Re^{i\pi t}.$$

(so that $\Gamma_R = \gamma_2 \star \gamma_1$ traces out the boundary of a half-disk). In many cases one can show that $\int_{\gamma_2} f(z)dz$ tends to 0 as $R \to \infty$, and by calculating the residues inside the contours $\Gamma_R$ deduce the integral of $f$ on $(-\infty, \infty)$. To see this strategy in action, consider the integral

$$\int_0^\infty \frac{dx}{1+x^2+x^4}.$$

It is easy to check that this integral exists as an improper Riemann integral, and since the integrand is even, it is equal to

$$\frac{1}{2}\lim_{R\to\infty}\int_{-R}^{R} \frac{dx}{1+x^2+x^4}dx.$$

If $f(z) = 1/(1 + z^2 + z^4)$, then $\int_{\Gamma_R} f(z)dz$ is equal to $2\pi i$ times the sum of the residues inside the path $\Gamma_R$. The function $f(z) = 1/(1 + z^2 + z^4)$ has poles at $z^2 = \pm e^{2\pi i/3}$ and hence at $\{e^{\pi i/3}, e^{2\pi i/3}, e^{4\pi i/3}, e^{5\pi i/3}\}$. They are all simple poles and of these only $\{\omega, \omega^2\}$ are in the upper-half plane, where $\omega = e^{i\pi/3}$. Thus by the residue theorem, for all $R > 1$ we have

$$\int_{\Gamma_R} f(z)dz = 2\pi i\big(\mathrm{Res}_\omega(f(z)) + \mathrm{Res}_{\omega^2}(f(z))\big),$$

and we may calculate the residues using the limit formula as above (and the fact that it evaluates to the reciprocal of the derivative of $1 + z^2 + z^4$): Indeed since $\omega^3 = -1$ we have $\mathrm{Res}_\omega(f(z)) = \frac{1}{2\omega+4\omega^3} = \frac{1}{2\omega-4}$, while $\mathrm{Res}_{\omega^2}(f(z)) = \frac{1}{2\omega^2+4\omega^6} = \frac{1}{4+2\omega^2}$. Thus we obtain:

$$\begin{aligned}\int_{\Gamma_R} f(z)dz &= 2\pi i\big(\frac{1}{2\omega - 4} + \frac{1}{2\omega^2 + 4}\big)\\ &= \pi i\big(\frac{1}{\omega - 2} + \frac{1}{\omega^2 + 2}\big)\\ &= \pi i\big(\frac{\omega^2 + \omega}{2(\omega - \omega^2) - 5}\big) = -\sqrt{3}\pi/(-3) = \pi/\sqrt{3},\end{aligned}$$

(where we used the fact that $\omega^2 + \omega = i\sqrt{3}$ and $\omega - \omega^2 = 1$). Now clearly

$$\int_{\Gamma_R} f(z)dz = \int_{-R}^{R} \frac{dt}{1 + t^2 + t^4} + \int_{\gamma_2} f(z)dz,$$

and by the estimation lemma we have

$$\Big|\int_{\gamma_2} f(z)dz\Big| \leq \sup_{z\in\gamma_2^*} |f(z)|.\ell(\gamma_2) \leq \frac{\pi R}{R^4 - R^2 - 1} \to 0,$$

as $R \to \infty$, it follows that

$$\pi/\sqrt{3} = \lim_{R\to\infty} \int_{\Gamma_R} f(z)dz = \int_{-\infty}^{\infty} \frac{dt}{1 + t^2 + t^4}.$$

20.1. **Jordan's Lemma and applications.** The following lemma is a real-variable fact which is fundamental to something known as *convexity*. Note that if $x, y$ are vectors in any vector space then the set $\{tx + (1 - t)y : t \in [0, 1]\}$ describes the line segment between $x$ and $y$.

**Lemma 20.4.** *Let $g\colon \mathbb{R} \to \mathbb{R}$ be a twice differentiable function. Then if $[a, b]$ is an interval on which $g''(x) < 0$, the function $g$ is convex on $[a, b]$, that is, for $x < y \in [a, b]$ we have*

$$g(tx + (1 - t)y) \geq tg(x) + (1 - t)g(y), \quad t \in [0, 1].$$

*Thus informally speaking, chords between points on the graph of $g$ lie below the graph itself.*

*Proof.* Given $x, y \in [a, b]$ and $t \in [0, 1]$ let $\xi = tx + (1 - t)y$, a point in the interval between $x$ and $y$. Now the slope of the chord between $(x, g(x))$ and $(\xi, g(\xi))$ is, by the Mean Value Theorem, equal to $g'(s_1)$ where $s_1$ lies between $x$ and $\xi$, while the slope of the chord between $(\xi, g(\xi))$ and $(y, g(y))$ is equal to $g'(s_2)$ for $s_2$ between $\xi$ and $y$. If $g(\xi) < tg(x) + (1 - t)g(y)$ it follows that $g'(s_1) < 0$ and $g'(s_2) > 0$. Thus by the mean value theorem for $g'(x)$ applied to the points $s_1$ and $s_2$ it follows there is an $s \in (s_1, s_2)$ with $g''(s) = (g'(s_2) - g'(s_1))/(s_2 - s_1) > 0$, contradicting the assumption that $g''(x)$ is negative on $(a, b)$. $\square$

The following lemma is an easy application of this convexity result.

**Lemma 20.5.** *(Jordan's Lemma): Let $f\colon \mathbb{H} \to \mathbb{C}_\infty$ be a meromorphic function on the upper-half plane $\mathbb{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$. Suppose that $f(z) \to 0$ as $z \to \infty$ in $\mathbb{H}$. Then if $\gamma_R(t) = Re^{it}$ for $t \in [0, \pi]$ we have*

$$\int_{\gamma_R} f(z)e^{i\alpha z}dz \to 0$$

*as $R \to \infty$ for all $\alpha \in \mathbb{R}_{\geq 0}$.*

*Proof.* Suppose that $\epsilon > 0$ is given. Then by assumption we may find an $S$ such that for $|z| > S$ we have $|f(z)| < \epsilon$. Thus if $R > S$ and $z = \gamma_R(t)$, it follows that

$$|f(z)e^{i\alpha z}| =\leq \epsilon e^{-\alpha R\sin(t)}.$$

But now applying Lemma 20.4 to the function $g(t) = \sin(t)$ with $x = 0$ and $y = \pi/2$ we see that $\sin(t) \geq \frac{2}{\pi}t$ for $t \in [0, \pi/2]$. Similarly we have $\sin(\pi - t) \geq 2(\pi - t)/\pi$ for $t \in [\pi/2, \pi]$. Thus we have

$$|f(z)e^{i\alpha z}| \leq \begin{cases} \epsilon.e^{-2\alpha Rt/\pi}, & t \in [0, \pi/2] \\ \epsilon.e^{-2\alpha R(\pi-t)/\pi} & t \in [\pi/2, \pi] \end{cases}$$

But then it follows that

$$\Big|\int_{\gamma_R} f(z)e^{i\alpha z}dz\Big| \leq 2\int_0^{\pi/2} \epsilon R.e^{-2\alpha Rt/\pi}dt = \epsilon.\pi\frac{1 - e^{-\alpha R}}{\alpha} < \epsilon.\pi/\alpha,$$

Thus since $\pi/\alpha > 0$ is independent of $R$, it follows that $\int_{\gamma_R} f(z)e^{i\alpha z}dz \to 0$ as $R \to \infty$ as required. □

*Remark* 20.6. If $\eta_R$ is an arc of a semicircle in the upper half plane, say $\eta_R(t) = Re^{it}$ for $0 \leq t \leq 2\pi/3$, then the same proof shows that $\int_{\eta_R} f(z)e^{i\alpha z}dz$ tends to zero as $R$ tends to infinity. This is sometimes useful when integrating around the boudary of a sector of disk (that is a set of the form $\{re^{i\theta} : 0 \leq r \leq R, \theta \in [\theta_1, \theta_2]\}$).

It is also useful to note that if $\alpha < 0$ then the integral of $f(z)e^{i\alpha z}$ around a semicircle in the *lower* half plane tends to zero as the radius of the semicircle tends to infinity provided $|f(z)| \to 0$ as $|z| \to \infty$ in the lower half plane. This follows immediately from the above applied to $f(-z)$.

**Example 20.7.** Consider the integral $\int_{-\infty}^{\infty} \frac{\sin(x)}{x}dx$. This is an improper integral of an even function, thus it exists if and only if the limit of $\int_{-R}^{R} \frac{\sin(x)}{x}dx$ exists as $R \to \infty$. To compute this consider the integral along the closed curve $\eta_R$ given by the concatenation $\eta_R = \nu_R \star \gamma_R$, where $\nu_R\colon [-R, R] \to \mathbb{R}$ given by $\nu_R(t) = t$ and $\gamma_R(t) = Re^{it}$ (where $t \in [0, \pi]$). Now if we let $f(z) = \frac{e^{iz}-1}{z}$, then $f$ has a removable singularity at $z = 0$ (as is easily seen by considering the power series expansion of $e^{iz}$) and so is an entire function. Thus we have $\int_{\eta_R} f(z)dz = 0$ for all $R > 0$. Thus we have

$$0 = \int_{\eta_R} f(z)dz = \int_{-R}^{R} f(t)dt + \int_{\gamma_R} \frac{e^{iz}}{z}dz - \int_{\gamma_R} \frac{dz}{z}.$$

Now Jordan's lemma ensures that the second term on the right tends to zero as $R \to \infty$, while the third term integrates to $\int_0^{\pi} \frac{iRe^{it}}{Re^{it}}dt = i\pi$. It follows that $\int_{-R}^{R} f(t)dt$ tends to $i\pi$ as $R \to \infty$. and hence taking imaginary parts we conclude the improper integral $\int_{-\infty}^{\infty} \frac{\sin(x)}{x}dx$ is equal to $\pi$.

*Remark* 20.8. The function $f(z) = \frac{e^{iz}-1}{z}$ might not have been the first meromorphic function one could have thought of when presented with the previous improper integral. A more natural candidate might have been $g(z) = \frac{e^{iz}}{z}$. There is an obvious problem with this choice however, which is that it has a pole on the contour we wish to integrate around. In the case where the pole is simple (as it is for $e^{iz}/z$) there is standard procedure for modifying the contour: one indents it by a small circular arc around the pole. Explicitly, we replace the $\nu_R$ with $\nu_R^- \star \gamma_\epsilon \star \nu_R^+$ where $\nu_R^\pm(t) = t$

and $t \in [-R, -\epsilon]$ for $\nu_R^-$, and $t \in [\epsilon, R]$ for $\nu_R^+$ (and as above $\gamma_\epsilon(t) = \epsilon e^{i(\pi - t)}$ for $t \in [0, \pi]$). Since $\frac{\sin(x)}{x}$ is bounded at $x = 0$ the sum

$$\int_{-R}^{-\epsilon} \frac{\sin(x)}{x} dx + \int_{\epsilon}^{R} \frac{\sin(x)}{x} dx \to \int_{-R}^{R} \frac{\sin(x)}{x} dx,$$

as $\epsilon \to 0$, while the integral along $\gamma_\epsilon$ can be computed explicitly: by the Taylor expansion of $e^{iz}$ we see that $\mathrm{Res}_{z=0} \frac{e^{iz}}{z} = 1$, so that $e^{iz} - 1/z$ is bounded near 0. It follows that as $\epsilon \to 0$ we have $\int_{\gamma_\epsilon} (e^{iz}/z - 1/z) dz \to 0$. On the other hand $\int_{\gamma_\epsilon} dz/z = \int_{-\pi}^{0} (-\epsilon i e^{i(\pi-t)})/(e^{i(\pi-t)} dt = -i\pi$, so that we see

$$\int_{\gamma_\epsilon} \frac{e^{iz}}{z} dz \to -i\pi$$

as $\epsilon \to 0$.

Combining all of this we conclude that if $\Gamma_\epsilon = \nu_R^- \star \gamma_\epsilon \star \nu_R^+ \star \gamma_R$ then

$$\begin{aligned} 0 = \int_{\Gamma_\epsilon} f(z) dz &= \int_{-R}^{-\epsilon} \frac{e^{ix}}{x} dx + \int_{\gamma_\epsilon} \frac{e^{iz}}{z} dz + \int_{\epsilon}^{R} \frac{e^{ix}}{x} dx + \int_{\gamma_R} \frac{e^{iz}}{z} dz. \\ &= 2i \int_{\epsilon}^{R} \frac{\sin(x)}{x} + \int_{\gamma_\epsilon} \frac{e^{iz}}{z} + \int_{\gamma_R} \frac{e^{iz}}{z} dz \\ &\to 2i \int_{0}^{R} \frac{\sin(x)}{x} dx - i\pi + \int_{\gamma_R} \frac{e^{iz}}{z} dz. \end{aligned}$$

as $\epsilon \to 0$. Then letting $R \to \infty$, it follows from Jordans Lemma that the third term tends to zero so we see that

$$\int_{-\infty}^{\infty} \frac{\sin(x)}{x} dx = 2 \int_{0}^{\infty} \frac{\sin(x)}{x} dx = \pi$$

as required.

We record a general version of the calculation we made for the contribution of the indentation to a contour in the following Lemma.

**Lemma 20.9.** *Let $f \colon U \to \mathbb{C}$ be a meromorphic function with a simple pole at $a \in U$ and let $\gamma_\epsilon \colon [\alpha, \beta] \to \mathbb{C}$ be the path $\gamma_\epsilon(t) = a + \epsilon e^{it}$, then*

$$\lim_{\epsilon \to 0} \int_{\gamma_\epsilon} f(z) dz = Res_a(f).(\beta - \alpha)i.$$

*Proof.* Since $f$ has a simple pole at $a$, we may write

$$f(z) = \frac{c}{z - a} + g(z)$$

where $g(z)$ is holomorphic near $z$ and $c = \mathrm{Res}_a(f)$ (indeed $c/(z-a)$ is just the principal part of $f$ at $a$). But now as $g$ is holomorphic at $a$, it is continuous at $a$, and so bounded. Let $M, r > 0$ be such that $|g(z)| < M$ for all $z \in B(a, r)$. Then if $0 < \epsilon < r$ we have

$$\Big| \int_{\gamma_\epsilon} g(z) dz \Big| \leq \ell(\gamma_\epsilon) M = (\beta - \alpha)\epsilon.M,$$

which clearly tends to zero as $\epsilon \to 0$. On the other hand, we have

$$\int_{\gamma_\epsilon} \frac{c}{z - a} dz = \int_{\alpha}^{\beta} \frac{c}{\epsilon e^{it}} i\epsilon e^{it} dt = \int_{\alpha}^{\beta} (ic) dt = ic(\beta - \alpha).$$

Since $\int_{\gamma_\epsilon} f(z) dz = \int_{\gamma_\epsilon} c/(z-a) dz + \int_{\gamma_\epsilon} g(z) dz$ the result follows. ∎

20.2. **On the computation of residues and principal parts.** The previous examples will hopefully have convinced you of the power of the residue theorem. Of course for it to be useful one needs to be able to calculate the residues of functions with isolated singularities. In practice the integral formulas we have obtained for the residue are often not the best way to do this. In this section we discuss a more direct approach which is often useful when one wishes to calculate the residue of a function which is given as the ratio of two holomorphic functions.

More precisely, suppose that we have a function $F \colon U \to \mathbb{C}$ given to us as a ratio $f/g$ of two holomorphic functions $f, g$ on $U$ where $g$ is non-constant. The singularities of the function $F$ are therefore poles which are located precisely at the (isolated) zeros of the function $g$, so that $F$ is meromorphic. For convenience, we assume that we have translated the plane so as to ensure the pole of $F$ we are interested in is at $a = 0$. Let $g(z) = \sum_{n\geq 0} c_n z^n$ be the power series for $g$, which will converge to $g(z)$ on any $B(0,r)$ such that $\bar{B}(0,r) \subseteq U$. Since $g(0) = 0$, and this zero is isolated, there is a $k > 0$ minimal with $c_k \neq 0$, and hence

$$g(z) = c_k z^k (1 + \sum_{n\geq 1} a_n z^n),$$

where $a_n = c_{n+k}/c_k$. Now if we let $h(z) = \sum_{n=1}^{\infty} a_n z^{n-1}$ then $h(z)$ is holomorphic in $B(0,r)$ – since $h(z) = (g(z) - c_k z^k)/(c_k z^{k+1})$ – and moreover

$$\frac{1}{g(z)} = \frac{1}{c_k z^k}\big(1 + zh(z)\big)^{-1},$$

Now as $h$ is continuous, it is bounded on $\bar{B}(0,r)$, say $|h(z)| < M$ for all $z \in \bar{B}(0,r)$. But then we have, for $|z| \leq \delta = \min\{r, 1/(2M)\}$,

$$\frac{1}{g(z)} = \frac{1}{c_k z^k}\big(\sum_{n=0}^{\infty}(-1)^n z^n h(z)^n\big),$$

where by the Weierstrass $M$-test, the above series converges uniformly on $\bar{B}(0,\delta)$. Moreover, for any $n$, the series $\sum_{m\geq n}(-1)^m z^m h(z)^m$ is a holomorphic function which vanishes to order at least $n$ at $z = 0$, so that $\frac{1}{c_k z^k}\sum_{n\geq k}(-1)^n z^n h(z)^n$ is holmorphic. It follows that the principal part of the Laurent series of $1/g(z)$ is equal to the principal part of the function

$$\frac{1}{c_k z^k}\sum_{n=1}^{k}(-1)^{k-1} z^k h(z)^k.$$

Since we know the power series for $h(z)$, this allows us to compute the principal part of $\frac{1}{g(z)}$ as claimed. Finally, the principal part $P_0(F)$ of $F = f/g$ at $z = 0$ is just the $P_0(f.P_0(g))$, the principal part of the function $f(z).P_0(g)$, which again is straight-forward to compute if we know the power series expansion of $f(z)$ at 0 (indeed we only need the first $k$ terms of it). The best way to digest this analysis is by means of examples. We consider one next, and will examine another in the next section on summation of series.

**Example 20.10.** Consider $f(z) = 1/(z^2 \sinh(z)^3)$. Now $\sinh(z) = (e^z - e^{-z})/2$ vanishes on $\pi i\mathbb{Z}$, and these zeros are all simple since $\frac{d}{dz}(\sinh(z)) = \cosh(z)$ has $\cosh(n\pi i) = (-1)^n \neq 0$. Thus $f(z)$ has a pole or order 5 at zero, and poles of order 3 at $\pi i n$ for each $n \in \mathbb{Z}\backslash\{0\}$. Let us calculate the principal part of $f$ at $z = 0$ using the above technique. We will write $O(z^k)$ for the vector space of holomorphic functions which vanish to order $k$ at 0.

$$
\begin{aligned}
z^2\sinh(z)^3 &= z^2(z+\frac{z^3}{3!}+\frac{z^5}{5!}+O(z^7))^3 = z^5(1+\frac{z^2}{3!}+\frac{z^4}{5!}+O(z^6))^3\\
&= z^5(1+\frac{3z^2}{3!}+\frac{3z^4}{(3!)^2}+\frac{3z^4}{5!}+O(z^6))\\
&= z^5(1+\frac{z^2}{2}+\frac{13z^4}{120}+O(z^6))\\
&= z^5\left(1+z(\frac{z}{2}+\frac{13z^3}{120}+O(z^5))\right)
\end{aligned}
$$

Thus, in the notation of the above discussion, $h(z)=\frac{z}{2}+\frac{13z^3}{120}+O(z^5)$, and so, as $h$ vanishes to first order at $z=0$, in order to obtain the principal part we just need to consider the first two terms in the geometric series $(1+zh(z))^{-1}=\sum_{n=0}^{\infty}(-1)^n z^n h(z)^n$:

$$
\begin{aligned}
1/z^2\sinh(z)^3 &= z^{-5}\big(1+z(\frac{z}{2}+\frac{13z^3}{120}+O(z^5))\big)^{-1}\\
&= z^{-5}\big(1-z(\frac{z}{2}+\frac{13z^3}{120})+z^2\frac{z^2}{(2!)^2}+O(z^5)\big)\\
&= z^{-5}\big(1-\frac{z^2}{2}+(\frac{1}{4}-\frac{13}{120})z^4+O(z^5)\big)\\
&= \frac{1}{z^5}-\frac{1}{2z^3}+\frac{17}{120z}+O(z).
\end{aligned}
$$

Thus the principal part of $f(z)$ at 0 is $P_0(f)=\frac{1}{z^5}-\frac{1}{2z^3}+\frac{17}{120z}$, and $\mathrm{Res}_0(f)=17/120$.

There are other variants on the above method which we could have used: For example, by the binomial theorem for an arbitrary exponent we know that if $|z|<1$ then $(1+z)^{-3}=\sum_{n\geq 0}\binom{-3}{n}z^n = 1-3z+6z^2+\ldots$. Arguing as above, it follows that for small enough $z$ we have

$$
\begin{aligned}
\sinh(z)^{-3} &= z^{-3}.(1+\frac{z^2}{3!}+\frac{z^4}{5!}+O(z^6))^{-3}\\
&= z^{-3}\left(1+(-3)(\frac{z^2}{3!}+\frac{z^4}{5!})+6(\frac{z^2}{3!}+\frac{z^4}{5!})^2+O(z^6)\right)\\
&= z^{-3}\left(1-\frac{z^2}{2}+(\frac{-3}{5!}+\frac{6}{(3!)^2})z^4+O(z^6)\right)\\
&= z^{-3}\left(1-\frac{z^2}{2}+\frac{17z^4}{120}+O(z^6)\right)
\end{aligned}
$$

yielding the same result for the principal part of $1/z^2\sinh(z)^3$.

20.3. **Summation of infinite series.** Residue calculus can also be a useful tool in calculating infinite sums, as we now show. For this we use the function $f(z)=\cot(\pi z)$. Note that since $\sin(\pi z)$ vanishes precisely at the integers, $f(z)$ is meromorphic with poles at each integer $n\in\mathbb{Z}$. Moreover, since $f$ is periodic with period 1, in order to understand the poles of $f$ it suffices to calculate the principal part of $f$ at $z=0$. We can use the method of the previous section to do this:

We have $\sin(z)=z-\frac{z^3}{3!}+\frac{z^5}{5!}+O(z^7)$, so that $\sin(z)$ vanishes with multiplicity 1 at $z=0$ and we may write $\sin(z)=z(1-zh(z))$ where $h(z)=z/3!-z^3/5!+O(z^5)$ is holomorphic at $z=0$. Then

$$
\frac{1}{\sin(z)}=\frac{1}{z}(1-zh(z))^{-1}=\frac{1}{z}\big(1+\sum_{n\geq 1}z^n h(z)^n\big)=\frac{1}{z}+h(z)+O(z^2).
$$

Multiplying by $\cos(z)$ we see that the principal part of $\cot(z)$ is the same as that of $\frac{1}{z}\cos(z)$ which, using the Taylor expansion of $\cos(z)$, is clearly $\frac{1}{z}$ again. By periodicity, it follows that $\cot(\pi z)$ has a simple pole with residue $1/\pi$ at each integer $n \in \mathbb{Z}$.

We can also use this strategy[50] to find further terms of the Laurent series of $\cot(z)$: Since our $h(z)$ actually vanishes at $z = 0$, the terms $h(z)^n z^n$ vanish to order $2n$. It follows that we obtain all the terms of the Laurent series of $\cot(z)$ at 0 up to order 3, say, just by considering the first two terms of the series $1 + \sum_{n\geq 1} z^n h(z)^n$, that is, $1 + zh(z)$. Since $\cos(z) = 1 - z^2/2! + z^4/4!$, it follows that $\cot(z)$ has a Laurent series

$$\begin{aligned}\cot(z) &= (1 - \frac{z^2}{2!} + O(z^4)).(\frac{1}{z} + (\frac{z}{3!} - \frac{z^3}{5!} + O(z^5)))\\ &= \frac{1}{z} - \frac{z}{3} + O(z^3)\end{aligned}$$

The fact that $f(z)$ has simple poles at each integer will allow us to sum infinite series with the help of the following:

**Lemma 20.11.** *Let $f(z) = \cot(\pi z)$ and let $\Gamma_N$ denotes the square path with vertices $(N+1/2)(\pm 1 \pm i)$. There is a constant $C$ independent of $N$ such that $|f(z)| \leq C$ for all $z \in \Gamma_N^*$.*

*Proof.* We need to consider the horizontal and vertical sides of the square separately. Note that $\cot(\pi z) = (e^{i\pi z} + e^{-i\pi z})/(e^{i\pi z} - e^{-i\pi z})$. Thus on the horizontal sides of $\Gamma_N$ where $z = x \pm (N+1/2)i$ and $-(N+1/2) \leq x \leq (N+1/2)$ we have

$$\begin{aligned}|\cot(\pi z)| &= \left|\frac{e^{i\pi(x\pm(N+1/2)i)} + e^{-i\pi(x\pm(N+1/2)i)}}{e^{i\pi(x\pm(N+1/2)i} - e^{-i\pi(x\pm(N+1/2)i)}}\right|\\ &\leq \frac{e^{\pi(N+1/2)} + e^{-\pi(N+1/2)}}{e^{\pi(N+1/2)} - e^{-\pi(N+1/2)}}\\ &= \coth(\pi(N+1/2)).\end{aligned}$$

Now since $\coth(x)$ is a decreasing function for $x \geq 0$ it follows that on the horizontal sides of $\Gamma_N$ we have $|\cot(\pi z)| \leq \coth(3\pi/2)$.

On the vertical sides we have $z = \pm(N+1/2) + iy$, where $-N - 1/2 \leq y \leq N + 1/2$. Observing that $\cot(z + N\pi) = \cot(z)$ for any integer $N$ and that $\cot(z + \pi/2) = -\tan(z)$, we find that if $z = \pm(N+1/2) + iy$ for any $y \in \mathbb{R}$ then

$$|\cot(\pi z)| = |-\tan(iy)| = |-\tanh(y)| \leq 1.$$

Thus we may set $C = \max\{1, \coth(3\pi/2)\}$. $\square$

We now show how this can be used to sum an infinite series:

**Example 20.12.** Let $g(z) = \cot(\pi z)/z^2$. By our discussion of the poles of $\cot(\pi z)$ above it follows that $g(z)$ has simple poles with residues $\frac{1}{\pi n^2}$ at each non-zero integer $n$ and residue $-\pi/3$ at $z = 0$.

Consider now the integral of $g(z)$ around the paths $\Gamma_N$: By Lemma 20.11 we know $|g(z)| \leq C/|z|^2$ for $z \in \Gamma_N^*$, and for all $N \geq 1$. Thus by the estimation lemma we see that

$$\left(\int_{\Gamma_N} g(z)dz\right) \leq C.(4N+2)/(N+1/2)^2 \to 0,$$

as $N \to \infty$. But by the residue theorem we know that

$$\int_{\Gamma_N} g(z)dz = -\pi/3 + \sum_{\substack{n\neq 0,\\ -N\leq n\leq N}} \frac{1}{\pi n^2}.$$

[50]See Appendix II for more details on the generalities and justification of this method.

It therefore follows that
$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \pi^2/6$$

*Remark* 20.13. Notice that the contours $\Gamma_N$ and the function $\cot(\pi z)$ clearly allows us to sum other infinite series in a similar way – for example if we wished to calculate the sum of the infinite series $\sum_{n\geq 1} \frac{1}{n^2+1}$ then we would consider the integrals of $g(z) = \cot(\pi z)/(1+z^2)$ over the contours $\Gamma_N$.

*Remark* 20.14. (*Non-examinable – for interest only!*): Note that taking $g(z) = (1/z^{2k})\cot(\pi z)$ for any positive integer $k$, the above strategy gives a method for computing $\sum_{n=1}^{\infty} 1/n^{2k}$ (check that you see why we need to take even powers of $n$). The analysis for the case $k = 1$ goes through in general, we just need to compute more and more of the Laurent series of $\cot(\pi z)$ the larger we take $k$ to be.

One can show that $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ converges to a holomorphic function of $s$ for any $s \in \mathbb{C}$ with $\Re(s) > 1$ (as usual, we define $n^s = \exp(s.\log(n))$ where log is the ordinary real logarithm). As $s \to 1$ it can be checked that $\zeta(s) \to \infty$, however it can be shown that $\zeta(s)$ extends to a meromorphic function on all of $\mathbb{C}\backslash\{1\}$. The identity theorem shows that this extension is unique if it exists[51]. (This uniqueness is known as the principle of "analytic continuation".) The location of the zeros of the $\zeta$-function is the famous Riemann hypothesis: apart from the "trivial zeros" at negative even integers, they are conjectured to all lie on the line $\Re(z) = 1/2$. Its values at special points however are also of interest: Euler was the first to calculate $\zeta(2k)$ for positive integers $k$, but the values $\zeta(2k+1)$ (for $k$ a positive integer) remain mysterious – it was only shown in 1978 by Roger Apéry that $\zeta(3)$ is irrational for example. Our analysis above is sufficient to determine $\zeta(2k)$ once one succeeds in computing explicitly the Laurent series for $\cot(\pi z)$ or equivalently the Taylor series of $z\cot(\pi z) = iz + 2iz/(e^{2iz} - 1)$. See Appendix IV for more details.

20.4. **Keyhole contours.** There are many ingenious paths which can be used to calculate integrals via residue theory. One common contour is known (for obvious reasons) as a *keyhole contour*. It is constructed from two circular paths of radius $\epsilon$ and $R$, where we let $R$ become arbitrarily large, and $\epsilon$ arbitrarily small, and we join the two circles by line segments with a narrow neck in between. Explicitly, if $0 < \epsilon < R$ are given, pick a $\delta > 0$ small, and set $\eta_+(t) = t + i\delta$, $\eta_-(t) = (R-t) - i\delta$, where in each case $t$ runs over the closed intervals with endpoints such that the endpoints of $\eta_\pm$ lie on the circles of radius $\epsilon$ and $R$ about the origin. Let $\gamma_R$ be the positively oriented path on the circle of radius $R$ joining the endpoints of $\eta_+$ and $\eta_-$ on that circle (thus traversing the "long" arc of the circle between the two points) and similarly let $\gamma_\epsilon$ the path on the circle of radius $\epsilon$ which is negatively oriented and joins the endpoints of $\gamma_\pm$ on the circle of radius $\epsilon$. Then we set $\Gamma_{R,\epsilon} = \eta_+ \star \gamma_R \star \eta_- \star \gamma_\epsilon$ (see Figure 2). The keyhole contour can sometimes be useful to evaluate real integrals where the integrand is multi-valued as a function on the complex plane, as the next example shows:

**Example 20.15.** Consider the integral $\int_0^\infty \frac{x^{1/2}}{1+x^2}dx$. Let $f(z) = z^{1/2}/(1+z^2)$, where we use the branch of the square root function which is continuous on $\mathbb{C}\backslash\mathbb{R}_{>0}$, that is, if $z = re^{it}$ with $t \in [0, 2\pi)$ then $z^{1/2} = r^{1/2}e^{it/2}$.

We use the keyhole contour $\Gamma_{R,\epsilon}$. On the circle of radius $R$, we have $|f(z)| \leq R^{1/2}/(R^2-1)$, so by the estimation lemma, this contribution to the integral of $f$ over $\Gamma_{R,\epsilon}$ tends to zero as $R \to \infty$. Similarly, $|f(z)|$ is bounded by $\epsilon^{1/2}/(1-\epsilon^2)$ on the circle of radius $\epsilon$, thus again by the estimation lemma this contribution to the integral of $f$ over $\Gamma_{R,\epsilon}$ tends to zero as $\epsilon \to 0$. Finally, the

[51]It is this uniqueness and the fact that one can readily compute that $\zeta(-1) = -1/12$ that results in the rather outrageous formula $\sum_{n=1}^{\infty} n = -1/12$.

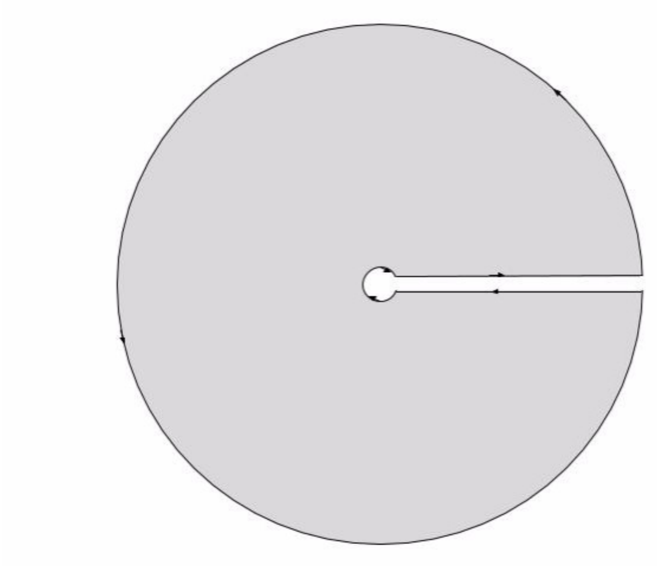

FIGURE 2. A keyhole contour.

discontinuity of our branch of $z^{1/2}$ on $\mathbb{R}_{>0}$ ensures that the contributions of the two line segments of the contour do not cancel but rather both tend to $\int_0^\infty \frac{x^{1/2}}{1+x^2}dx$ as $\delta$ and $\epsilon$ tend to zero.

To compute $\int_0^\infty \frac{x^{1/2}}{1+x^2}dx$ we evaluate the integral $\int_{\Gamma_{R,\epsilon}} f(z)dz$ using the residue theorem: The function $f(z)$ clearly has simple poles at $z = \pm i$, and their residues are $\frac{1}{2}e^{-\pi i/4}$ and $\frac{1}{2}e^{5\pi i/4}$ respectively. It follows that

$$\int_{\Gamma_{R,\epsilon}} f(z)dz = 2\pi i\left(\frac{1}{2}e^{-\pi i/4} + \frac{1}{2}e^{5\pi i/4}\right) = \pi\sqrt{2}.$$

Taking the limit as $R \to \infty$ and $\epsilon \to 0$ we see that $2\int_0^\infty \frac{x^{1/2}}{1+x^2}dx = \pi\sqrt{2}$, so that

$$\int_0^\infty \frac{x^{1/2}dx}{1+x^2} = \frac{\pi}{\sqrt{2}}.$$

## 21. The argument principle

**Lemma 21.1.** *Suppose that $f : U \to \mathbb{C}$ is a meromorphic and has a zero of order $k$ or a pole of order $k$ at $z_0 \in U$. Then $f'(z)/f(z)$ has a simple pole at $z_0$ with residue $k$ or $-k$ respectively.*

*Proof.* If $f(z)$ has a zero of order $k$ we have $f(z) = (z - z_0)^k g(z)$ where $g(z)$ is holomorphic near $z_0$ and $g(z_0) \neq 0$. It follows that

$$f'(z)/f(z) = \frac{k}{z - z_0} + g'(z)/g(z),$$

and since $g(z) \neq 0$ near $z_0$ it follows $g'(z)/g(z)$ is holomorphic near $z_0$, so that the result follows. The case where $f$ has a pole at $z_0$ is similar. □

*Remark* 21.2. Note that if $U$ is an open set on which one can define a holomorphic branch $L$ of $[\mathrm{Log}(z)]$ then $g(z) = L(f(z))$ has $g'(z) = f'(z)/f(z)$. Thus integrating $f'(z)/f(z)$ along a path $\gamma$ will measure the change in argument around the origin of the path $f(\gamma(t))$. The residue theorem allows us to relate this to the number of zeros and poles of $f$ inside $\gamma$, as the next theorem shows:

**Theorem 21.3.** *(Argument principle): Suppose that $U$ is an open set and $f: U \to \mathbb{C}$ is a meromorphic function on $U$. If $B(a,r) \subseteq U$ and $N$ is the number of zeros (counted with multiplicity) and $P$ is the number of poles (again counted with multiplicity) of $f$ inside $B(a,r)$ and $f$ has neither on $\partial B(a,r)$ then*

$$N - P = \int_\gamma \frac{f'(z)}{f(z)} dz,$$

*where $\gamma(t) = a + re^{2\pi i t}$ is a path with image $\partial B(a,r)$. Moreover this is the winding number of the path $\Gamma = f \circ \gamma$ about the origin.*

*Proof.* It is easy to check that $I(\gamma, z)$ is 1 if $|z-a| \leq 1$ and is 0 otherwise. Since Lemma 21.1 shows that $f'(z)/f(z)$ has simple poles at the zeros and poles of $f$ with residues the corresponding orders the result immediately from Theorem 19.18.

For the last part, note that the winding number of $\Gamma(t) = f(\gamma(t))$ about zero is just

$$\int_{f\circ\gamma} dw/w = \int_0^1 \frac{1}{f(\gamma(t))} f'(\gamma(t))\gamma'(t)dt = \int_\gamma \frac{f'(z)}{f(z)} dz$$

□

The argument principle is very useful – we use it here to establish some important results.

**Theorem 21.4.** *(Rouché's theorem): Suppose that $f$ and $g$ are holomorphic functions on an open set $U$ in $\mathbb{C}$ and $\bar{B}(a,r) \subset U$. If $|f(z)| > |g(z)|$ for all $z \in \partial B(a,r)$ then $f$ and $f+g$ have the same number of zeros in $B(a,r)$ (counted with multiplicities).*

*Proof.* Let $\gamma(t) = a + re^{2\pi i t}$ be a parametrization of the boundary circle of $B(a,r)$. We need to show that $(f+g)/f = 1 + g/f$ has the same number of zeros as poles (Note that $f(z) \neq 0$ on $\partial B(a,r)$ since $|f(z)| > |g(z)|$.) But by the argument principle, this number is the winding number of $h(\gamma(t))$ about zero, where $h(z) = 1 + f(z)/g(z)$. Since $|g(z)| < |f(z)|$ on $\gamma$ it follows that $|g(z)/f(z)| < 1$, so that the image of $\gamma^*$ under $1 + g/f$ lies entirely in the half-plane $\{z : \Re(z) > 0\}$, hence picking a branch of Log defined on this half-plane, we see that the integral

$$\int_\Gamma \frac{dz}{z} = \mathrm{Log}(f(\gamma(1)) - \mathrm{Log}(f(\gamma(0)) = 0$$

as required.

□

*Remark* 21.5. Rouche's theorem can be useful in counting the number of zeros of a function $f$ – one tries to find an approximation to $f$ whose zeros are easier to count and then by Rouche's theorem obtain information about the zeros of $f$.

**Example 21.6.** Suppose that $P(z) = z^4 + 5z + 2$. Then on the circle $|z| = 2$ we have $|z|^4 = 16 > 5.2 + 2 \geq |5z + 2|$ so that if $g(z) = 5z + 2$ we see that $P - g = z^4$ and $P$ have the same number of roots $B(0,2)$. It follows by Rouche's theorem that the four roots of $P(z)$ all have modulus less than 2. On the other hand, if we take $|z| = 1$, then $|5z + 2| \geq 5 - 2 = 3 > |z^4| = 1$, hence $P(z)$ and $5z + 2$ have the same number of roots in $B(0,1)$. It follows $P(z)$ has one root of modulus less than 1 and 3 of modulus between 1 and 2.

**Theorem 21.7.** *(Open mapping theorem): Suppose that $f: U \to \mathbb{C}$ is holomorphic and non-constant on a region $U$. Then for any open set $V \subset U$ the set $f(V)$ is also open.*

*Proof.* Suppose that $w_0 \in f(V)$, say $f(z_0) = w_0$. Then $g(z) = f(z) - w_0$ has a zero at $z_0$ which, since $f$ is nonconstant, is isolated. Thus we may find an $r > 0$ such that $g(z) \neq 0$ on $\bar{B}(z_0, r) \subset U$ and in particular since $\partial B(z_0, r)$ is compact, we have $|g(z)| \geq \delta > 0$ on $\partial B(z_0, r)$. But then if $|w - w_0| < \delta$ it follows $|w - w_0| < |g(z)|$ on $\partial B(z_0, r)$, hence by the argument principle $g(z)$ and

$h(z) = g(z) + (w_0 - w) = f(z) - w$ also has a zero in $B(z_0, r)$, that is, $f(z)$ takes the value $w$ in $B(z_0, r)$. Thus $B(w_0, \delta) \subseteq f(B(z_0, r))$ and hence $f(U)$ is open as required. □

*Remark* 21.8. Note that the proof actually establishes a bit more than the statement of the theorem: if $w_0 = f(z_0)$ then the multiplicity $d$ of the zero of the function $f(z) - w_0$ at $z_0$ is called the *degree* of $f$ at $z_0$. The proof shows that locally the function $f$ is $d$-to-1, counting multiplicities, that is, there are $r, \epsilon \in \mathbb{R}_{>0}$ such that for every $w \in B(w_0, \epsilon)$ the equation $f(z) = w$ has $d$ solutions counted with multiplicity in the disk $B(z_0, r)$.

**Theorem 21.9.** *(Inverse function theorem): Suppose that $f\colon U \to \mathbb{C}$ is injective and holomorphic and that $f'(z) \neq 0$ for all $z \in U$. If $g\colon f(U) \to U$ is the inverse of $f$, then $g$ is holomorphic with $g'(w) = 1/f'(g(w))$.*

*Proof.* By the open mapping theorem, the function $g$ is continuous, indeed if $V$ is open in $f(U)$ then $g^{-1}(V) = f(V)$ is open by that theorem. To see that $g$ is holomorphic, fix $w_0 \in f(U)$ and let $z_0 = g(w_0)$. Note that since $g$ and $f$ are continuous, if $w \to w_0$ then $f(w) \to z_0$. Writing $z = f(w)$ we have

$$\lim_{w \to w_0} \frac{g(w) - g(w_0)}{w - w_0} = \lim_{z \to z_0} \frac{z - z_0}{f(z) - f(z_0)} = 1/f'(z_0)$$

as required. □

*Remark* 21.10. Note that the non-trivial part of the proof of the above theorem is the fact that $g$ is continuous! In fact the condition that $f'(z) \neq 0$ follows from the fact that $f$ is bijective – this can be seen using the degree of $f$: if $f'(z_0) = 0$ and $f$ is nonconstant, we must have $f(z) - f(z_0) = (z - z_0)^k g(z)$ where $g(z_0) \neq 0$ and $k \geq 1$. Since we can chose a holomorphic branch of $g^{1/k}$ near $z_0$ it follows that $f(z)$ is locally $k$-to-1 near $z_0$, which contradicts the injectivity of $f$. For details see the Appendices. Notice that this is in contrast with the case of a single real variable, as the example $f(x) = x^3$ shows. Once again, complex analysis is "nicer" than real analysis!

## 22. The extended complex plane

When studying isolated singularities of a holomorphic function $f$, we observed that $f$ has a pole at a point $z_0$ if and only if $f(z) \to \infty$ as $z \to z_0$. This motivates the idea of extending the complex plane by adding a point $\infty$ "at infinity". In this section we want to develop this idea more fully and show that we can make sense of the notion of continuous and holomorphic functions on the extended plane $\mathbb{C} \cup \{\infty\} = \mathbb{C}_\infty$. We use two different approaches:

(1) Real geometry: The stereographic projection map will allow us to identify the plane $\mathbb{C} = \mathbb{R}^2$ with the complement of the point $(0, 0, 1)$ in the 2-sphere $\mathbb{S}^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$, so that the "north pole" $N = (0, 0, 1)$ becomes the point at infinity.
(2) Complex geometry: The set of lines $\mathbb{P}^1$ in $\mathbb{C}^2$, that is, one-dimensional subspaces of $\mathbb{C}^2$ contains a copy of $\mathbb{C}$ where $z \in \mathbb{C}$ is identified with the line through the vector $(z, 1)$. Every line but that through $(1, 0)$ is obtained in this way, so again we obtain $\mathbb{C}_\infty$ by identifying $\infty$ with the line $\mathbb{C}.(1, 0)$.

### 22.1. Stereographic projection.

22.1. **Stereographic projection.** Let $\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ be the unit sphere of radius 1 centred at the origin in $\mathbb{R}^3$, and view the complex plane as the copy of $\mathbb{R}^2$ inside $\mathbb{R}^3$ given by the plane $\{(x, y, 0) \in \mathbb{R}^3 : x, y \in \mathbb{R}\}$. Let $N$ be the "north pole" $N = (0, 0, 1)$ of the sphere $\mathbb{S}^2$. Given a point $z \in \mathbb{C}$, there is a unique line passing through $N$ and $z$, which intersects $\mathbb{S}\backslash\{N\}$ in a point $S(z)$. This map gives a bijection between $\mathbb{C}$ and $\mathbb{S}\backslash\{N\}$. Indeed, explicitly, if

$(X, Y, Z) \in \mathbb{S}\backslash\{N\}$ then it corresponds to[52] $z \in \mathbb{C}$ where $z = x + iy$ with $x = X/(1 - Z)$ and $y = Y/(1 - Z)$. Correspondingly, given $z = x + iy \in \mathbb{C}$ we have

$$S(z) = \big(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}\big) = \frac{1}{1 + |z|^2}\big(2\Re(z), 2\Im(z), |z|^2 - 1\big). \tag{22.1}$$

Thus if we set $S(\infty) = N$, then we get a bijection between $\mathbb{C}_\infty$ and $\mathbb{S}^2$, and we use this identification to make $\mathbb{C}_\infty$ into a metric space (and thus we obtain a notion of continuity for $\mathbb{C}_\infty$): As a subset of $\mathbb{R}^3$ equipped with the Euclidean metric $\mathbb{S}^2$ is naturally a metric space.

**Lemma 22.1.** *The metric induced on $\mathbb{C}_\infty$ by $S$ is given by*

$$d(z, w) = \frac{2|z - w|}{\sqrt{1 + |z|^2}\sqrt{1 + |w|^2}} \quad d(z, \infty) = \frac{2}{\sqrt{1 + |z|^2}}.$$

*for any $z, w \in \mathbb{C}$.*

*Proof.* First consider the case where $z, w \in \mathbb{C}$. Since $S(z), S(w) \in \mathbb{S}^2$ we see that $\|S(z) - S(w)\|^2 = 2 - 2S(z).S(w)$. But using (22.1) we see that

$$\begin{aligned} S(z).S(w) &= \frac{2(z\bar{w} + \bar{z}w) + (|z|^2 - 1)(|w|^2 - 1)}{(1 + |z|^2)(1 + |w|^2)} \\ &= \frac{2(z\bar{w} + \bar{z}w) + z\bar{z}w\bar{w} - z\bar{z} - w\bar{w} + 1}{(1 + |z|^2)(1 + |w|^2)} \\ &= 1 - \frac{2|z - w|^2}{(1 + |z|^2)(1 + |w|^2)} \end{aligned}$$

so that

$$d_2(S(z), S(w))^2 = \frac{4|z - w|^2}{(1 + |z|^2)(1 + |w|^2)}$$

as required. The case where one or both of $z, w$ is equal to $\infty$ is similar but easier. $\square$

*Remark* 22.2. Note that in particular, $S(z)$ tends to $N = (0, 0, 1)$ if and only if $|z| \to \infty$, thus our notation $z \to \infty$ now takes on a literal meaning, consistent with its previous definition. In particular, meromorphic functions on an open subset $U$ of $\mathbb{C}$ naturally extend to continuous functions from $U$ to $\mathbb{C}_\infty$.

The geometry of the sphere nicely unites lines and circles in the plane as the following Lemma shows:

**Lemma 22.3.** *The map $S\colon \mathbb{C} \to \mathbb{S}$ induces a bijection between lines in $\mathbb{C}$ and circles in $\mathbb{S}$ which contain $N$, and a bijection between circles in $\mathbb{C}$ and circles in $\mathbb{S}$ not containing $N$.*

*Proof.* A circle in $\mathbb{S}$ is given by the intersection of $\mathbb{S}$ with a plane $H$. Any plane $H$ in $\mathbb{R}^3$ is given by an equation of the form $aX + bY + cZ = d$, and $H$ intersects $\mathbb{S}$ provided $a^2 + b^2 + c^2 > d^2$. Indeed to see this note that $H$ intersects the sphere in a circle if and only if its distance to the origin is less than 1. Since the closest vector to the origin on $H$ is perpendicular to the plane it is a scalar multiple of $(a, b, c)$, so it must be $\frac{d}{a^2+b^2+c^2}(a, b, c)$, hence $H$ is at distance $d^2/(a^2 + b^2 + c^2)$ from the origin and the result follows. Moreover, clearly $H$ contains $N$ if and only if $c = d$.

Now from the explicit formulas for $S$ we see that if $z = x + iy$ then $S(z)$ lies on this plane if and only if

$$\begin{aligned} &2ax + 2by + c(x^2 + y^2 - 1) = d(x^2 + y^2 + 1) \\ \iff\ &(c - d)(x^2 + y^2) + 2ax + 2by - (c + d) = 0 \end{aligned}$$

[52]Any point on the line between $N$ and $(X, Y, Z)$ can be written as $t(0, 0, 1) + (1 - t)(X, Y, Z)$ for some $t \in \mathbb{R}$. It is then easy to calculate where this line intersects the plane given by the equation $z = 0$.

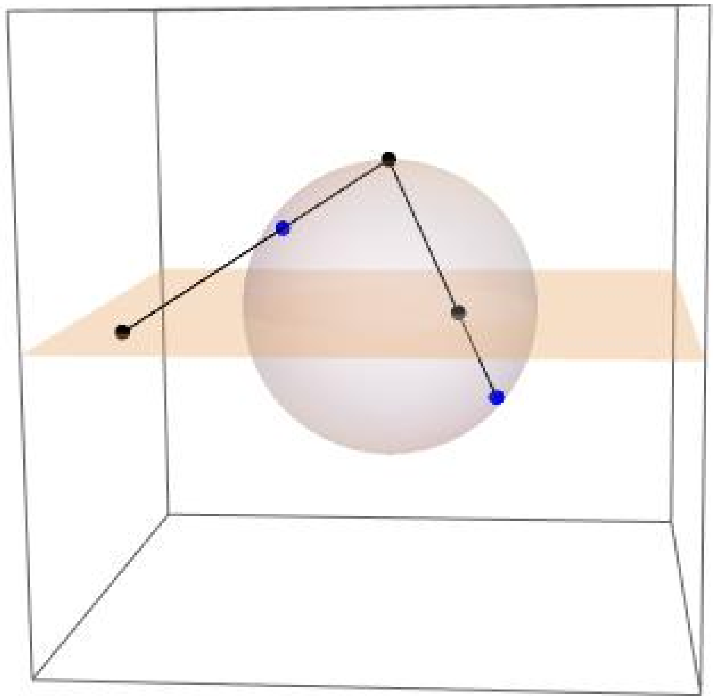

FIGURE 3. The stereographic projection map.

Clearly if $c = d$ this is the equation of a line, while conversely if $c \neq d$ it is the equation of a circle in the plane. Indeed if $c \neq d$, we can normalize and insist that $c - d = 1$, whence our equation becomes

$$(x + a)^2 + (y + b)^2 = (a^2 + b^2 + c + d) \tag{22.2}$$

that is, the circle with centre $(-a, -b)$ and radius $\sqrt{a^2 + b^2 + c + d}$. Note that the condition the plane intersected $\mathbb{S}$ becomes the condition that $a^2 + b^2 + c + d > 0$, that is, exactly the condition that Equation (22.2) has a non-empty solution set.

To complete the proof, we need to show that all circles and lines in $\mathbb{C}$ are given by the form of the above equation. When $c = d$ we get $2(ax + by - c) = 0$, and clearly the equation of every line can be put into this form. When $c \neq d$ as before assume $c - d = 1$, then letting $a, b, c + d$ vary freely we see that we can obtain circle in the plane as required. $\square$

22.2. **The projective line.** Our second approach to the extended complex plane is via the projective line $\mathbb{P}^1$: this is, as a set, simply the collection of one-dimensional subspaces of $\mathbb{C}^2$. If $e_1, e_2$ denote the standard basis of $\mathbb{C}^2$ then we have two natural subsets of $\mathbb{P}^1$, each naturally in bijection with $\mathbb{C}$. If we set $U_0 = \mathbb{P}^1 \backslash \mathbb{C}.e_1$ and $U_1 = \mathbb{P}^1 \backslash \mathbb{C}e_2$, then we have maps $i_0, i_\infty \colon \mathbb{C} \to \mathbb{P}^1$ given by $i_0(z) = \mathbb{C}.(ze_1 + e_2)$ and $i_\infty(z) = \mathbb{C}.(e_1 + ze_2)$ whose images are $U_0$ and $U_1$ respectively. Given a nonzero vector $(z, w) \in \mathbb{C}^2$ we will write $[z, w] \in \mathbb{P}^1$ for the line it spans. (The numbers $z, w$ are often called the *homogeneous coordinates* of $[z, w]$. They are only defined up to simultaneous rescaling.)

Thus $\mathbb{P}^1$ is covered by two pieces $U_0$ and $U_\infty$ whose union is all of $\mathbb{P}^1$. We can use this to make $\mathbb{P}^1$ a topological space: we say that $V$ is an open subset of $\mathbb{P}^1$ if and only if $V \cap U_0$ and $V \cap U_\infty$ are identified with open subsets of $\mathbb{C}$ via the bijections $i_0$ and $i_1$ respectively. It is a good exercise to check that this does indeed define a topology on $\mathbb{P}^1$ (in which both $U_0$ and $U_\infty$ are open, since $\mathbb{C}$ and $\mathbb{C}^\backslash\{0\}$ are open in $\mathbb{C}$. We however will take a more direct approach: Note that we can identify

$\mathbb{P}^1$ with $\mathbb{C}_\infty$ using the map $i_0\colon \mathbb{C} \to \mathbb{P}^1$ extending it to $\mathbb{C}_\infty$ by sending $\infty$ to $\mathbb{C}e_1$ and we can thus transport the metric on $\mathbb{C}_\infty$ (which of course we obtained in turn from our identification on $\mathbb{C}_\infty$ with $\mathbb{S}^2$) to that on $\mathbb{P}^1$. Perhaps surprisingly, this metric has a natural expression in terms of the Hermitian form $\langle\cdot,\cdot\rangle$ on $\mathbb{C}^2$ as the next Lemma shows:

**Lemma 22.4.** *The metric induced on $\mathbb{P}^1$ by its identification with $\mathbb{C}_\infty$ is given by*

$$d(L_1, L_2) = 2\sqrt{1 - \frac{|\langle v, w\rangle|^2}{\|v\|^2\|w\|^2}}$$

*where $v \in L_1\backslash\{0\}$ and $w \in L_2\backslash\{0\}$.*

*Proof.* Suppose $L_1 = [z, 1]$ and $L_2 = [w, 1]$. Then the formula in the statement of the Lemma gives

$$\begin{aligned} d(L_1, L_2) &= 2\sqrt{1 - \frac{|z\bar{w}+1|^2}{(1+|z|^2)(1+|w^2)}} \\ &= 2\sqrt{\frac{1+|z|^2+|w|^2+|z|^2|w|^2-|z|^2|w|^2-z\bar{w}-\bar{z}w-1}{(1+|z|^2)(1+|w|^2)}} \\ &= 2\sqrt{\frac{|z-w|^2}{(1+|z|^2)(1+|w|^2}} = \frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}} \end{aligned}$$

The case when $L_2 = \infty = \mathbb{C}e_1$ is similar but easier. $\square$

One advantage of thinking of $\mathbb{C}_\infty$ as the projective line is that we can use the charts $U_0$ and $U_\infty$ to define what it means for a function $f$ on $\mathbb{C}_\infty$ to be holomorphic:

**Definition 22.5.** Suppose that $f\colon W \to \mathbb{P}^1$ is a continuous function on an open subset $W$ of $\mathbb{P}^1$, and let $L \in V$. Suppose that $L \in U_p$ and $f(L) \in U_q$ where $p, q \in \{0, \infty\}$. Then $f^{-1}(U_l) \cap U_k$ is an open set in $\mathbb{P}^1$, which via $i_k$ (or rather its inverse) we can identify with an open subset $V$ of $\mathbb{C}$, and its image under $f$ lies in $U_q$ which we can identify with $\mathbb{C}$ via $i_q^{-1}$. Thus $f$ yields a continuous function $\tilde{f}\colon V \to \mathbb{C}$, where $\tilde{f} = i_q^{-1} \circ f \circ i_p$ and we say $f$ is holomorphic at $L$ if $\tilde{f}$ is holomorphic at $i_p(z) = L$.

Since most points in $\mathbb{P}^1$ lie in both $U_0$ *and* $U_\infty$ the above definition seems ambiguous. In fact, where there is a choice, it does not matter what which of $U_0$ or $U_\infty$ you pick. This is because $i_0^{-1} \circ i_\infty(z) = i_\infty^{-1} \circ i_0(z) = 1/z$ for all $z \in \mathbb{C}\backslash\{0\}$ and the function $1/z$ is holomorphic with holomorphic inverse (itself!) on $\mathbb{C}\backslash\{0\}$. This fact and the chain rule combine to show that the definition is independent of any choices. The essential point is that if $f(z)$ is holomorphic, so are $f(1/z), 1/f(z)$ and $1/f(1/z)$ wherever they are defined.

**Example 22.6.** Suppose that $U$ is an open subset of $\mathbb{C}$ and $f\colon U \to \mathbb{P}^1$ is holomorphic and suppose $z_0 \in U$ is such that $f(z_0) = \infty$. Then by continuity $f(z) \neq 0$ near $z_0$, so we can take $U_q = U_\infty$ and $U_p = U_0$. Then if we write $f([z:1]) = [1 : f_\infty(z)]$, it follows $i_\infty^{-1} \circ f \circ i_0(z) = f_\infty(z)$, and we simpy require $f_\infty(z)$ to be holomorphic at $z = z_0$ (with value 0 at $z = z_0$). This in particular means that, if $f$ is non-constant, $f_\infty(z_0) = 0$ is an isolated zero of $f_\infty$, so that close to $z_0$ we have $f_\infty(z) \neq 0$, and hence $f(z) \in U_0$. For such points we may write $f([z:1]) = [f_0(z):1]$. Since $[f_0(z):1] = f([z:1]) = [1 : f_\infty(z)]$ we see $f_0(z) = 1/f_\infty(z)$, hence the condition $f$ is holomorphic at $z_0$ is exactly our defintion that $f$ have a pole at $z_0$.

You can check using this definition that a holomorphic function $f\colon \mathbb{C} \to \mathbb{P}^1$ are precisely the meromorphic functions, and with a bit more work show that the holomorphic functions $f$ which are defined on all of $\mathbb{P}^1$ are exactly the set of rational functions.

We end this section by noting an illuminating connection between the extended complex plane and the notion of a simply connected domain in the plane.

**Theorem 22.7.** *A domain $D$ in $\mathbb{C}$ is simply-connected if and only if $\mathbb{C}_\infty \backslash D$ is connected.*

*Proof.* We can only sketch a proof of one direction of the theorem. Suppose that $\mathbb{C}_\infty \backslash D$ is connected, and let $\gamma$ be a closed path in $D$. Recall that $\mathbb{C}\backslash\gamma^*$ has exactly one unbounded component, $C$ say, and for $z \in C$ we have $I(\gamma, z) = 0$. In terms of the Riemann sphere, this is simply the component of $\mathbb{C}_\infty\backslash\gamma^*$ which contains $\infty$. Now $\mathbb{C}_\infty\backslash D \subset \mathbb{C}_\infty\backslash\gamma^*$ and since by assumption it is connected and contains $\infty$, we have $\mathbb{C}_\infty\backslash D \subset C$. Thus $I(\gamma, z) = 0$ for all $z \in \mathbb{C}\backslash D$, so that the inside of $\gamma$ lies entirely in $D$. But then Theorem 19.11 and Theorem 15.21 show that $D$ is a primitive domain, and hence, as discussed before, is simply-connected. ☐

## 23. Conformal transformations

Another important feature of the stereographic projection map is that it is *conformal*, meaning that it preserves angles. The following definition helps us to formalize what this means:

**Definition 23.1.** If $\gamma\colon [-1, 1] \to \mathbb{C}$ is a $C^1$ path which has $\gamma'(t) \neq 0$ for all $t$, then we say that the line $\{\gamma(t) + s\gamma'(t) : s \in \mathbb{R}\}$ is the *tangent line* to $\gamma$ at $\gamma(t)$, and the vector $\gamma'(t)$ is a tangent vector at $\gamma(t) \in \mathbb{C}$.

*Remark* 23.2. Note that this definition gives us a notion of tangent vectors at points on subsets of $\mathbb{R}^n$, since the notion of a $C^1$ path extends readily to paths in $\mathbb{R}^n$ (we just require all $n$ component functions are continuously differentiable). In particular, if $\mathbb{S}$ is the unit sphere in $\mathbb{R}^3$ as above, a $C^1$ path on $\mathbb{S}$ is simply a path $\gamma\colon [a, b] \to \mathbb{R}^3$ whose image lies in $\mathbb{S}$. It is easy to check that the tangent vectors at a point $p \in \mathbb{S}$ all lie in the plane perpendicular to $p$ – simply differentiate the identity $f(\gamma(t)) = 1$ where $f(x, y, z) = x^2 + y^2 + z^2$ using the chain rule.

We can now state what we mean by a conformal map:

**Definition 23.3.** Let $U$ be an open subset of $\mathbb{C}$ and suppose that $T\colon U \to \mathbb{C}$ (or $\mathbb{S}$) is continuously differentiable in the real sense (so all its partial derivatives exist and are continuous). If $\gamma_1, \gamma_2\colon [-1, 1] \to U$ are two paths with $z_0 = \gamma_1(0) = \gamma_2(0)$ then $\gamma_1'(0)$ and $\gamma_2'(0)$ are two tangent vectors at $z_0$, and we may consider the angle between them (formally speaking this is the difference of their arguments). By our assumption on $T$, the compositions $T \circ \gamma_1$ and $T \circ \gamma_2$ are $C^1$-paths through $T(z_0)$, thus we obtain a pair of tangent vectors at $T(z_0)$. We say that $T$ is *conformal* at $z_0$ if for every pair of $C^1$ paths $\gamma_1, \gamma_2$ through $z_0$, the angle between their tangent vectors at $z_0$ is equal to the angle between the tangent vectors at $T(z_0)$ given by the $C^1$ paths $T \circ \gamma_1$ and $T \circ \gamma_2$. We say that $T$ is conformal on $U$ if it is conformal at every $z \in U$.

One of the main reasons we focus on conformal maps here is because holomorphic functions give us a way of producing many examples of them, as the following result shows.

**Proposition 23.4.** *Let $f\colon U \to \mathbb{C}$ be a holomorphic map and let $z_0 \in U$ be such that $f'(z_0) \neq 0$. Then $f$ is conformal at $z_0$. In particular, if $f\colon U \to \mathbb{C}$ is has nonvanishing derivative on all of $U$, it is conformal on all of $U$ (and locally a biholomorphism).*

*Proof.* We need to show that $f$ preserves angles at $z_0$. Let $\gamma_1$ and $\gamma_2$ be $C^1$-paths with $\gamma_1(0) = \gamma_2(0) = z_0$. Then we obtain paths $\eta_1, \eta_2$ through $f(z_0)$ where $\eta_1(t) = f(\gamma_1(t))$ and $\eta_2(t) = f(\gamma_2(t))$. By the Chain Rule (see Lemma 25.7) we see that $\eta_1'(t) = Df_{z_0}(\gamma_1'(t))$ and $\eta_2'(t) = Df_{z_0}(\gamma_2'(t))$, and moreover if $f'(z_0) = \rho.e^{i\theta}$, then

$$Df_{z_0} = \rho.\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix},$$

(since the linear map given by multiplication by $f'(z_0)$ is precisely scaling by $\rho$ and rotating by $\theta$). It follows that if $\phi_1$ and $\phi_2$ are the arguments of $\gamma_1'(0)$ and $\gamma_2'(0)$, then the arguments of $\eta_1'(0)$ and $\eta_2'(0)$ are $\phi_1 + \theta$ and $\phi_2 + \theta$ respectively. It follows that the difference between the two pairs of arguments, that is, the angles between the curves at $z_0$ and $f(z_0)$, are the same.

For the final part, note that if $f'(z_0) \neq 0$ then by the definition of the degree of vanishing, the function $f(z)$ is locally biholomorphic (see the proof of the inverse function theorem). □

**Example 23.5.** The function $f(z) = z^2$ has $f'(z)$ nonzero everywhere except the origin. It follows $f$ is a conformal map from $\mathbb{C}^\times$ to itself. Note that the condition that $f'(z)$ is non-zero is necessary – if we consider the function $f(z) = z^2$ at $z = 0$, $f'(z) = 2z$ which vanishes precisely at $z = 0$, and it is easy to check that at the origin $f$ in fact doubles the angles between tangent vectors.

**Lemma 23.6.** *The sterographic projection map $S\colon \mathbb{C} \to \mathbb{S}$ is conformal.*

*Proof.* Let $z_0$ be a point in $\mathbb{C}$, and suppose that $\gamma_1(t) = z_0 + tv_1$ and $\gamma_2(t) = z_0 + tv_2$ are two paths[53] having tangents $v_1$ and $v_2$ at $z_0 = \gamma_1(0) = \gamma_2(0)$. Then the lines $L_1$ and $L_2$ they describe, together with the point $N$, determine planes $H_1$ and $H_2$ in $\mathbb{R}^3$, and moreover the image of the lines under stereographic projection is the intersection of these planes with $\mathbb{S}$. Since the intersection of $\mathbb{S}$ with any plane is either empty or a circle, it follows that the paths $\gamma_1$ and $\gamma_2$ get sent to two circles $C_1$ and $C_2$ passing through $P = S(z_0)$ and $N$. Now by symmetry, these circles meet at the same angle at $N$ as they do at $P$. Now the tangent lines of $C_1$ and $C_2$ at $N$ are just the intersections of $H_1$ and $H_2$ with the plane tangent to $\mathbb{S}$ at $N$. But this means the angle between them will be the same as that between the intersection of $H_1$ and $H_2$ with the complex plane, since it is parallel to the tangent plane of $\mathbb{S}$ at $N$. Thus the angles between $C_1$ and $C_2$ at $P$ and $L_1$ and $L_2$ at $z_0$ coincide as required. □

23.1. **Mobius transformations.** Recall that we have identified $\mathbb{C}_\infty$ with the projective line $\mathbb{P}^1$. The general linear group $\mathrm{GL}_2(\mathbb{C})$ acts on $\mathbb{C}^2$ in the natural way, and this induces an action on the set of lines in $\mathbb{C}$. We thus get an action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathbb{P}^1$, and so on the extended complex plane. Explicitly, if $v = (z_1, z_2)^t$ spans a line $L = \mathbb{C}.v$ then if $g \in \mathrm{GL}_2(\mathbb{C})$ is given by a matrix

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we see that

$$g(L) = \mathbb{C}.g(v) = \mathbb{C}\begin{pmatrix} az_1 + bz_2 \\ cz_1 + dz_2 \end{pmatrix}.$$

In particular, using our embedding $i_0\colon \mathbb{C} \to \mathbb{P}^1$ we see that

$$g(i_0(z)) = \mathbb{C}.g\begin{pmatrix} z \\ 1 \end{pmatrix} = \mathbb{C}.\begin{pmatrix} az+b \\ cz+d \end{pmatrix} = \mathbb{C}.\begin{pmatrix} \frac{az+b}{cz+d} \\ 1 \end{pmatrix} = i_0(\frac{az+b}{cz+d}).$$

Note that $f(-d/c) = \infty$ and $f(\infty) = a/c$, as is easily checked using the fact that $\infty = [1:0] \in \mathbb{P}^1$.

**Definition 23.7.** The induced maps $z \mapsto \frac{az+b}{cz+d}$ from the extended complex plane to itself are known as *Mobius maps* or *Mobius transformations*. Since they come from the action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathbb{P}^1$ they automatically form a group. Note this means that every Mobius transformation is a bijection of the extended complex plane to itself, and moreover its inverse is also a Mobius transformation. In particular, since rational functions on $\mathbb{C}$ yield holomorphic functions on $\mathbb{C}_\infty$, every Mobius transformation gives an invertible holomorphic function on $\mathbb{C}_\infty$.

$$\mathrm{Mob} = \{f(z) = \frac{az+b}{cz+d} : ad - bc \neq 0\}.$$

[53]with domain $[-1, 1]$ say – or even the whole real line, except that it is non-compact.

Note that if we rescale $a, b, c, d$ by the same (nonzero) scalar, then we get the same transformation. In group theoretic terms, the map from $\mathrm{GL}_2(\mathbb{C})$ to Mob has a kernel, the scalar matrices, thus Mob is a *quotient group* of $\mathrm{GL}_2(\mathbb{C})$. As a quotient group it is usually denoted $\mathrm{PGL}_2(\mathbb{C})$ the *projective general linear group.*

Any Mobius transformation can be understood as a composition of a small collection of simpler transformations, as we will now show. This can be useful because it allows us to prove certain results about all Mobius transformations by checking them for the simple transformations.

**Definition 23.8.** A transformation of the form $z \mapsto az$ where $a \neq 0$ is called a *dilation.* A transformation of the form $z \mapsto z + b$ is called a *translation.* The transformation $z \mapsto 1/z$ is called *inversion.* Note that these are all Mobius transformations, and the inverse of a dilation is a dilation, the inverse of a translation is a translation, while inversion is an involution and so is its own inverse.

**Lemma 23.9.** *Any Mobius transformation can be written as a composition of dilations, translations and an inversion.*

*Proof.* Let $G$ denote the set of all Mobius transformations which can be obtained as compositions of dilations, translations and inversions. The set $G$ is a subgroup of Mob. We wish to show it is the full group of Mobius transformations.

First note that any transformation of the form $z \mapsto az+b$ is a composition of the dilation $z \mapsto az$ and the translation $z \mapsto z+b$. Moreover, if $f(z) = \frac{az+b}{cz+d}$ is a Mobius transformation and $c = 0$ then $f(z) = (a/d)z + (b/d)$ (note if $c = 0$ then $ad - bc \neq 0$ implies $d \neq 0$) and so is a composition of a dilation and a translation. If $c \neq 0$ then we have

$$\frac{az+b}{cz+d} = \frac{(a/c)(cz+d) + (b - da/c)}{cz+d} = \frac{a}{c} + (b - d/a)\frac{1}{cz+d}. \tag{23.1}$$

Now $z \mapsto \frac{1}{cz+d}$ is the composition of an inversion with the map $z \mapsto cz + d$, and so lies in $G$. But then by equation (23.1) we have $f(z)$ is a composition of this map with a dilation and a translation, and so $f$ lies in $G$. Since $f$ was an arbitrary transformation with $c \neq 0$ it follows $G = \text{Mob}$ as required. ☐

*Remark* 23.10. The subgroup of Mob generated by translations and dilations is the group of $\mathbb{C}$-linear affine transformations $\mathrm{Aff}(\mathbb{C}) = \{f(z) = az + b : a \neq 0\}$ of the complex plane. It is the stablizer of $\infty$ in Mob.

*Remark* 23.11. One should compare the statement of the previous Lemma with the theory or reduced row echelon form in Linear Algebra: any invertible $2 \times 2$ matrix will have the identity matrix as its reduced row echelon form, and the elementary row operations correspond essentially to the simple transformations which generate the Mobius group. This can be used to give an alternative proof of the Lemma.

As an example of how we can use this result to study Mobius transformations, we prove the following:

**Lemma 23.12.** *Let $f\colon \mathbb{C}_\infty \to \mathbb{C}_\infty$ be a Mobius transformation. Then $f$ takes circles to circles. (Here we view $\mathbb{C}_\infty$ as $\mathbb{S}^2$ so that by Lemma 22.3 a circle in $\mathbb{C}_\infty$ is a line or a circle in $\mathbb{C}$).*

*Proof.* Since a line in $\mathbb{C}$ is given by the equation $\Im(az) = s$ where $s \in \mathbb{R}$ and $|a| = 1$, while a circle is given by the equation $|z - a| = r$ for $a \in \mathbb{C}$, $r \in \mathbb{R}_{>0}$, it is easy to check that any dilation or translation takes a line to a line and a circle to a circle. On the other hand, we have seen that any circle can be described as the locus $C = \{z : |z - a| = k|z - b|\}$ where $a, b \in \mathbb{C}$ and $k \in (0, 1)$ and moreover we can assume $a, b \neq 0$ (see the remark after Lemma 12.3). But if $z \in C$ and $w = 1/z$

we have

$$|1/w - a| = k|1/w - b| \iff |w - 1/a||a| = k|b||w - 1/b| \iff |w - 1/a| = \frac{k|b|}{|a|}|w - 1/a|,$$

thus we see that the image of $C$ under inversion is the locus of points $w$ which satisfy the equation $|w - 1/a| = \frac{k|b|}{|a|}|w - 1/b|$, which is therefore a line or a circle as required. $\square$

Although it follows easily from what we have already done, it is worth high-lighting the following:

**Lemma 23.13.** *Mobius transformations are conformal.*

*Proof.* As we have already shown, any holomorphic map is conformal wherever its derivative is nonzero. Since a Mobius transformation $f$ is invertible everywhere with holomorphic inverse, its derivative must be nonzero everywhere and we are done.

One can also give a more explicit proof: If $f(z) = \frac{az+b}{cz+d}$ then it is easy to check that

$$f'(z) = \frac{ad - bc}{(cz + d)^2} \neq 0,$$

for all $z \neq -d/c$, thus $f$ is conformal at each $z \in \mathbb{C}\backslash\{-d/c\}$. Checking at $z = \infty, -d/c$ is similar: indeed at $\infty = [1 : 0]$ we use the map $i_\infty \colon \mathbb{C} \to \mathbb{P}^1$ given by $w \mapsto [1 : w]$ to obtain $f_\infty(w) = \frac{a+bw}{c+dw}$ and $f'_\infty(w) = \frac{bc-ad}{(c+dw)^2}$, which is certainly nonzero at $w = 0$ (and $i_\infty(0) = \infty$). $\square$

Since a Mobius map is given by the four entries of a $2 \times 2$ matrix, up to simultaneus rescaling, the following result is perhaps not too surprising.

**Proposition 23.14.** *If $z_1, z_2, z_3$ and $w_1, w_2, w_3$ are triples of pairwise distinct complex numbers, then there is a unique Mobius transformation $f$ such that $f(z_i) = w_i$ for each $i = 1, 2, 3$.*

*Proof.* It is enough to show that, given any triple $(z_1, z_2, z_3)$ of complex numbers, we can find a Mobius transformations which takes $z_1, z_2, z_3$ to $0, 1, \infty$ respectively. Indeed if $f_1$ is such a transformation, and $f_2$ takes $0, 1, \infty$ to $w_1, w_2, w_3$ respectively, then clearly $f_2 \circ f_1^{-1}$ is a Mobius transformation which takes $z_i$ to $w_i$ for each $i$.

Now consider

$$f(z) = \frac{(z - z_1)(z_2 - z_3)}{(z - z_3)(z_2 - z_1)}$$

It is easy to check that $f(z_1) = 0, f(z_2) = 1, f(z_3) = \infty$, and clearly $f$ is a Mobius transformation as required. If any of $z_1, z_2$ or $z_3$ is $\infty$, then one can find a similar transformation (for example by letting $z_i \to \infty$ in the above formula). Indeed if $z_1 = \infty$ then we set $f(z) = \frac{z_2-z_3}{z-z_3}$; if $z_2 = \infty$, we take $f(z) = \frac{z-z_1}{z-z_3}$; and finally if $z_3 = \infty$ take $f(z) = \frac{z-z_1}{z_2-z_1}$.

To see the $f$ is unique, suppose $f_1$ and $f_2$ both took $z_1, z_2, z_3$ to $w_1, w_2, w_3$. Then taking Mobius transformations $g, h$ sending $z_1, z_2, z_3$ and $w_1, w_2, w_3$ to $0, 1, \infty$ the transformations $hf_1g^{-1}$ and $hf_2g^{-1}$ both take $(0, 1, \infty)$ to $(0, 1, \infty)$. But suppose $T(z) = \frac{az+b}{cz+d}$ is any Mobius transformation with $T(0) = 0$, $T(1) = 1$ and $T(\infty) = \infty$. Since $T$ fixes $\infty$ it follows $c = 0$. Since $T(0) = 0$ it follows that $b/d = 0$ hence $b = 0$, thus $T(z) = a/d.z$, and since $T(1) = 1$ it follows $a/d = 1$ and hence $T(z) = z$. Thus we see that $hf_1g^{-1} = hf_2g^{-1} = \mathrm{id}$ are all the identity, and so $f_1 = f_2$ as required. $\square$

**Example 23.15.** The above lemma shows that we can use Mobius transformations as a source of conformal maps. For example, suppose we wish to find a conformal transformation which takes the upper half plane $\mathbb{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$ to the unit disk $B(0, 1)$. The boundary of $\mathbb{H}$ is the real line, and we know Mobius transformations take lines to lines or circles, and in the latter case this means the point $\infty \in \mathbb{C}_\infty$ is sent to a finite complex number. Now any circle is uniquely determined by three points lying on it, and we know Mobius transformations allow us to take any three points to any other three points. Thus if we take $f$ the Mobius map which sends $0 \mapsto -i$, and $1 \mapsto 1$, $\infty \mapsto i$ the real axis will be sent to the unit circle. Now we have

$$f(z) = \frac{iz+1}{z+i}$$

(one can find $f$ in a similar fashion to the proof of Proposition 23.14).

So far, we have found a Mobius transformation which takes the real line to the unit circle. Since $\mathbb{C}\backslash\mathbb{R}$ has two connected components, the upper and lower half planes, $\mathbb{H}$ and $i\mathbb{H}$, and similarly $\mathbb{C}\backslash\mathbb{S}^1$ has two connected components, $B(0,1)$ and $\mathbb{C}\backslash\bar{B}(0,1)$. Since a Mobius transformation is continuous, it maps connected sets to connected sets, thus to check whether $f(\mathbb{H}) = B(0,1)$ it is enough to know which component of $\mathbb{C}\backslash\mathbb{S}^1$ a single point in $\mathbb{H}$ is sent to. But $f(i) = 0 \in B(0,1)$, so we must have $f(\mathbb{H}) = B(0,1)$ as required.

Note that if we had taken $g(z) = (z+i)/(iz+1)$ for example, then $g$ also maps $\mathbb{R}$ to the unit circle $\mathbb{S}^1$, but $g(-i) = 0$, so[54] $g$ maps the lower half plane to $B(0,1)$. If we had used this transformation, then it would be easy to "correct" it to get what we wanted: In fact there are (at least) two simple things one could do: First, one could note that the map $R(z) = -z$ (a rotation by $\pi$) sends the upper half plane to the lower half place, so that the composition $g \circ R$ is a Mobius transformation taking $\mathbb{H}$ to $B(0,1)$. Alternatively, the inversion $j(z) = 1/z$ sends $\mathbb{C}\backslash\bar{B}(0,1)$ to $B(0,1)$, so that $j \circ g$ also sends $\mathbb{H}$ to $B(0,1)$. Explicitly, we have

$$g \circ R(z) = \frac{z-i}{iz-1} = \frac{-i(iz+1)}{i(z+i)} = -f(z), \quad j \circ g(z) = \frac{iz+1}{z+i} = f(z).$$

Note in particular that $f$ is far from unique – indeed if $f$ is any Mobius transformation which takes $\mathbb{H}$ to $B(0,1)$ then composing it with any Mobius transformation which preserves $B(0,1)$ will give another such map. Thus for example $e^{i\theta}.f$ will be another such transformation.

**Exercise 23.16.** Every Mobius transformation gives a biholomorphic map from $\mathbb{C}_\infty$ to itself, but they may not preserve the distance function $d_S$ on $\mathbb{P}^1$. What is the subgroup of Mob which are isometries of $\mathbb{P}^1$ with respect to the distance function $d_S$?

Given two domains $D_1, D_2$ in the complex plane, one can ask if there is a conformal transformation $f\colon D_1 \to D_2$. Since a conformal transformation is in particular a homeomorphism, this is clearly not possible for completely arbitrary domains. However if we restrict to simply-connected domains (that is, domains in which any path can be continuously deformed to any other path with the same end-points), the following remarkable theorem shows that the answer to this question is yes! Since it will play a distinguished role later, we will write $\mathbb{D}$ for the unit disc $B(0,1)$.

**Theorem 23.17.** *(Riemann's mapping theorem): Let $U$ be an open connected and simply-connected proper subset of $\mathbb{C}$. Then there if $z_0 \in U$ there is a unique bijective conformal transformation $f\colon U \to \mathbb{D}$ such that $f(z_0) = 0$, $f'(z_0) > 0$.*

*Remark* 23.18. The proof of this theorem is beyond the scope of this course, but it is a beautiful and fundamental result. The proof in fact only uses the fact that on a simply-connected domain any holomorphic function has a primitive, and hence it in fact shows that such domains are simply-connected in the topological sense (since a conformal transformation is in particular a homeomorphism, and the disc in simply-connected). It relies crucially on *Montel's theorem* on families of holomorphic functions, see for example the text of Shakarchi and Stein for an exposition of the argument.

[54] A Mobius map is a continuous function on $\mathbb{C}_\infty$, and if we remove a circle from $\mathbb{C}_\infty$ the complement is a disjoint union of two connected components, just the same as when we remove a line or a circle from the plane, thus the connectedness argument works just as well when we include the point at infinity.

Note that it follows immediately from Liouville's theorem that there can be no bijective conformal transformation taking $\mathbb{C}$ to $B(0,1)$, so the whole complex plane is indeed an exception. The uniqueness statement of the theorem reduces to the question of understanding the conformal transformations of the disk $\mathbb{D}$ to itself.

Of course knowing that a conformal transformation between two domains $D_1$ and $D_2$ exists still leaves the challenge of constructing one. As we will see in the next section on harmonic maps, this is an important question. In simple cases one can often do so by hand, as we now show.

In addition to Mobius transformations, it is often useful to use the exponential function and branches of the multifunction $[z^\alpha]$ (away from the origin) when constructing conformal maps. We give an example of the kind of constructions one can do:

**Example 23.19.** Let $D_1 = B(0,1)$ and $D_2 = \{z \in \mathbb{C} : |z| < 1, \Im(z) > 0\}$. Since these domains are both convex, they are simply-connected, so Riemann's mapping theorem ensure that there is a conformal map sending $D_2$ to $D_1$. To construct such a map, note that the domain is defined by the two curves $\gamma(0,1)$ and the real axis. It can be convenient to map the two points of intersection of these curves, $\pm 1$ to 0 and $\infty$. We can readily do this with a Mobius transformation:

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

Now since $f$ is a Mobius transformation, it follows that $f_1(\mathbb{R})$ and $f_1(\gamma(0,1))$ are lines (since they contain $\infty$) passing through the origin. Indeed $f(\mathbb{R}) = \mathbb{R}$, and since $f$ had inverse $f^{-1} = \frac{z+1}{z-1}$ it follows that the image of $\gamma(0,1)$ is $\{w \in \mathbb{C} : |w-1| = |w+1|\}$, that is, the imaginary axis. Since $f(i/2) = (-3+4i)/5$ it follows by connectedness that $f(D_1)$ is the second quadrant $Q = \{w \in \mathbb{C} : \Re(z) < 0, \Im(z) > 0\}$.

Now the squaring map $s\colon \mathbb{C} \to \mathbb{C}$ given by $z \mapsto z^2$ maps $Q$ bijectively to the half-plane $H = \{w \in \mathbb{C} : \Im(w) < 0\}$, and is conformal except at $z = 0$ (which is on the boundary, not in the interior, of $Q$). We may then use a Mobius map to take this half-plane to the unit disc: indeed in Example 23.15 we have already seen that the Mobius transformation $g(z) = \frac{z+i}{iz+1}$ takes the lower-half plane to the upper-half plane.

Putting everything together, we see that $F = g \circ s \circ f$ is a conformal transformation taking $D_1$ to $D_2$ as required. Calculating explicitly we find that

$$F(z) = i\left(\frac{z^2 + 2iz + 1}{z^2 - 2iz + 1}\right)$$

*Remark* 23.20. Note that there are couple of general principles one should keep in mind when constructing conformal transformations between two domains $D_1$ and $D_2$. Often if the boundary of $D_1$ has distinguished points (such as $\pm 1$ in the above example) it is convenient to move these to "standard" points such as 0 and $\infty$, which one can do with a Mobius transformation. The fact that Mobius transformations are three-transitive and takes lines and circles to lines and circles and moreover act transitively on such means that we can always use Mobius transformations to match up those parts of the boundary of $D_1$ and $D_2$ given by line segments or arcs of circles. However these will not be sufficient in general: indeed in the above example, the fact that the boundary of $D_1$ is a union of a semicircle and a line segment, while that of $D_2$ is just a circle implies there is no Mobius transformation taking $D_1$ to $D_2$, as it would have to take $\partial D_1$ to $\partial D_2$, which would mean that its inverse would not take the unit circle to either a line or a circle. Branches of fractional power maps $[z^\alpha]$ are often useful as they allow us to change the angle at the points of intersection of arcs of the boundary (being conformal on the interior of the domain but not on its boundary).

23.2. **Conformal transformations and the Laplace equation.** In this section we will use the term *conformal map* or *conformal transformation* somewhat abusively to mean a holomorphic function whose derivative is nowhere vanishing on its domain of definition. (We have seen already that this implies the function is conformal in the sense of the previous section.) If there is a bijective conformal transformation between two domains $U$ and $V$ we say they are *conformally equivalent*.

Recall that a function $v\colon \mathbb{R}^2 \to \mathbb{R}$ is said to be *harmonic* if it is twice differentiable and $\partial_x^2 v + \partial_y^2 v = 0$. Often one seeks to find solutions to this equation on a domain $U \subset \mathbb{R}^2$ where we specify the values of $v$ on the boundary $\partial U$ of $U$. This problem is known as the *Dirichlet problem*, and makes sense in any dimension (using the appropriate Laplacian). In dimension 2, complex analysis and in particular conformal maps are a powerful tool by which one can study this problem, as the following lemma show.

**Lemma 23.21.** *Suppose that $U \subset \mathbb{C}$ is a simply-connected open subset of $\mathbb{C}$ and $v\colon U \to \mathbb{R}$ is twice continuously differentiable and harmonic. Then there is a holomorphic function $f\colon U \to \mathbb{C}$ such that $\Re(f) = v$. In particular, if $v$ is harmonic and twice continuously differentiable then it is analytic.*

*Proof.* (*Sketch*): Consider the function $g(z) = \partial_x v - i\partial_y v$. Then since $v$ is twice continuously differentiable, the partial derivatives of $g$ are continuous and

$$\partial_x^2 v = -\partial_y^2 v; \quad \partial_y \partial_x v = \partial_x \partial_y v,$$

so that $g$ satisfies the Cauchy-Riemann equations. It follows from Theorem 13.11 that $g$ is holomorphic. Now since $U$ is simply-connected, it follows that $g$ has a primitive $G\colon U \to \mathbb{C}$. But then it follows that if $G = a(z) + ib(z)$ we have $\partial_z G = \partial_x a - i\partial_y a = g(z) = \partial_x v - i\partial_y v$, hence the partial derivatives of $a$ and $v$ agree on all of $U$. But then if $z_0, z \in U$ and $\gamma$ is a path between then, the chain rule[55] shows that

$$\begin{aligned}\int_\gamma (\partial_x v + i\partial_y v)dz &= \int_0^1 (\partial_x(v(\gamma(t)) + i\partial_y v(\gamma(t)))\gamma'(t)dt \\ &= \int_0^1 \frac{d}{dt}(v(\gamma(t)))dt = v(z) - v(z_0),\end{aligned}$$

Similarly, we see that the same path integral is also equal to $a(z) - a(z_0)$. It follows that $a(z) = v(z) + (a(z_0) - v(z_0))$, thus if we set $f(z) = G(z) - (G(z_0) - v(z_0))$ we obtain a holomorphic function on $U$ whose real part is equal to $v$ as required.

Since we know that any holomorphic function is analytic, it follows that $v$ is analytic (and in particular, infinitely differentiable). ☐

The previous Lemma shows that, at least locally (in a disk say) harmonic functions and holomorphic functions are in correspondence – given a holomorphic function $f$ we obtain a harmonic function by taking its real part, while if $u$ is harmonic the previous lemma shows we can associate to it a holomorphic function $f$ whose real part equals $u$ (and in fact examining the proof, we see that $f$ is actually unique up to a purely imaginary constant). Thus if we are seeking a harmonic function on an open set $U$ whose values are a given function $g$ on $\partial U$, then it suffices to find a holomorphic function $f$ on $U$ such that $\Re(f) = g$ on the boundary $\partial U$.

Now if $H\colon U \to V$ was a bijective conformal transformation which extends to a homeomorphism $\bar{H}\colon \bar{U} \to \bar{V}$ which thus takes $\partial U$ homeomorphically to $\partial V$, then if $f\colon V \to \mathbb{C}$ is holomorphic, so is $f \circ H$. Thus in particular $\Re(f \circ H)$ is a harmonic function on $U$. It follows that we can use conformal transformations to transport solutions of Laplace's equation from one domain to another: if we can use a conformal transformation $H$ to take a domain $U$ to a domain $V$ where

[55] This uses the chain rule for a composition $g \circ f$ of real-differentiable functions $f\colon \mathbb{R} \to \mathbb{R}^2$ and $g\colon \mathbb{R}^2 \to \mathbb{R}$, applied to the real and imaginary parts of the integrand. This follows in exactly the same way as the proof of Lemma 25.7. See the remark after the proof of that lemma.

we already have a supply of holomorphic functions satisfying various boundary conditions, the conformal transformation $H$ gives us a corresponding set of holomorphic (and hence harmonic) functions on $U$. We state this a bit more formally as follow:

**Lemma 23.22.** *If $U$ and $V$ are domains and $G\colon U \to V$ is a conformal transformation, then if $u\colon V \to \mathbb{R}$ is a harmonic function on $V$, the composition $u \circ G$ is harmonic on $U$.*

*Proof.* To see that $u \circ G$ is harmonic we need only check this in a disk $B(z_0, r) \subseteq U$ about any point $z_0 \in U$. If $w_0 = G(z_0)$, the continuity of $G$ ensures we can find $\delta, \epsilon > 0$ such that $G(B(z_0, \delta)) \subseteq B(w_0, \epsilon) \subseteq V$. But now since $B(w_0, \epsilon)$ is simply-connected we know by Lemma 23.21 we can find a holomorphic function $f(z)$ with $u = \Re(f)$. But then on $B(z_0, \delta)$ we have $u \circ G = \Re(f \circ G)$, and by the chain rule $f \circ G$ is holomorphic, so that its real part is harmonic as required. $\square$

*Remark* 23.23. You can also give a more direct computational proof of the above Lemma. Note also that we only need $G$ to be holomorphic – the fact that it is a conformal equivalence is not necessary. On the other hand if we are trying to produce harmonic functions with prescribed boundary values, then we will need to use carefully chosen conformal transformations.

This strategy for studying harmonic functions might appear over-optimistic, in that the domains one can obtain from a simple open set like $B(0, 1)$ or the upper-half plane $\mathbb{H}$ might consist of only a small subset of the open sets one might be interested in. However, the Riemann mapping theorem (Theorem 23.17) show that *every* domain which is simply connected, other than the whole complex plane itself, is in fact conformally equivalent to $B(0, 1)$. Thus a solution to the Dirichlet problem for the disk at least in principal comes close[56] to solving the same problem for any simply-connected domain! For convenience, we will write $\mathbb{D}$ for the open disk $B(0, 1)$ of radius 1 centred at 0.

In the course so far, the main examples of conformal transformations we have are the following:

(1) The exponential function is conformal everywhere, since it is its own derivative and it is everywhere nonzero.
(2) Mobius transformations understood as maps on the extended complex plane are everywhere conformal.
(3) Fractional exponents: In cut planes the functions $z \mapsto z^\alpha$ for $\alpha \in \mathbb{C}$ are conformal (the cut removes the origin, where the derivative may vanish).

Let us see how to use these transformations to obtain solutions of the Laplace equation. First notice that Cauchy's integral formula suggests a way to produce solutions to Laplace's equation in the disk: Suppose that $u$ is a harmonic function defined on $B(0, r)$ for some $r > 1$. Then by Lemma 23.21 we know there is a holomorphic function $f\colon B(0, r) \to \mathbb{C}$ such that $u = \Re(f)$. By Cauchy's integral formula, if $\gamma$ is a parametrization of the positively oriented unit circle, then for all $w \in B(0, 1)$ we have $f(w) = \frac{1}{2\pi i}\int_\gamma f(z)/(z - w)dz$, and so

$$u(z) = \Re\big(\frac{1}{2\pi i}\int_\gamma \frac{f(z)dz}{z - w}\big).$$

Since the integrand uses only the values of $f$ on the boundary circle, we have almost recovered the function $u$ from its values on the boundary. (Almost, because we appear to need the values of it harmonic conjugate). The next lemma resolves this:

**Lemma 23.24.** *If $u$ is harmonic on $B(0, r)$ for $r > 1$ then for all $w \in B(0, 1)$ we have*

$$u(w) = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\frac{1 - |w|^2}{|e^{i\theta} - w|^2}d\theta = \frac{1}{2\pi}\int_0^{2\pi} u(e^{i\theta})\Re\big(\frac{e^{i\theta} + w}{e^{i\theta} - w}\big)d\theta.$$

[56]The issue is whether the conformal equivalence behaves well enough at the boundaries.

*Proof.* (*Sketch.*) Take, as before, $f(z)$ holomorphic with $\Re(f) = u$ on $B(0, r)$. Then letting $\gamma$ be a parametrization of the positively oriented unit circle we have

$$f(w) = \frac{1}{2\pi i}\int_\gamma \frac{f(z)dz}{z-w} - \frac{1}{2\pi i}\int_\gamma \frac{f(z)dz}{z-\bar{w}^{-1}}$$

where the first term is $f(w)$ by the integral formula and the second term is zero because $f(z)/(z - \bar{w}^{-1})$ is holomorphic inside all of $B(0, 1)$. Gathering the terms, this becomes

$$f(w) = \frac{1}{2\pi}\int_\gamma f(z)\frac{1-|w|^2}{|z-w|^2}\frac{dz}{iz} = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\frac{1-|w|^2}{|e^{i\theta}-w|^2}d\theta.$$

The advantage of this last form is that the real and imaginary parts are now easy to extract, and we see that

$$u(z) = \int_0^{2\pi} u(e^{i\theta})\frac{1-|w|^2}{|e^{i\theta}-w|^2}d\theta.$$

Finally for the second integral expression note that if $|z| = 1$ then

$$\frac{z+w}{z-w} = \frac{(z+w)(\bar{z}-\bar{w})}{|z-w|^2} = \frac{1-|w|^2+(\bar{z}w-z\bar{w})}{|z-w|^2}.$$

from which one readily sees the real part agrees with the corresponding factor in our first expression. $\square$

Now the idea to solve the Dirichlet problem for the disk $B(0, 1)$ is to turn this previous result on its head: Notice that it tells us the values of $u$ inside the disk $B(0, 1)$ in terms of the values of $u$ on the boundary. Thus if we are given the boundary values, say a (periodic) function $G(e^{i\theta})$ we might reasonably hope that the integral

$$g(w) = \frac{1}{2\pi}\int_0^{2\pi} G(e^{i\theta})\frac{1-|w|^2}{|e^{i\theta}-w|^2}d\theta,$$

would define a harmonic function with the required boundary values. Indeed it follows from the proof of the lemma that the integral is the real part of the integral

$$\frac{1}{2\pi i}\int_\gamma G(z)\frac{1}{z-w}dz,$$

which we know from Lemma 16.18 is holomorphic in $w$, thus $g(w)$ is certainly harmonic. It turns out that if $w \to w_0 \in \partial B(0, 1)$ then provided $G$ is continuous at $w_0$ then $g(w) \to G(w_0)$, hence $g$ is in fact a harmonic function with the required boundary value.