Complex Analysis (part 5): Local Behavior of Holomorphic Functions (by Evan Dummit, 2022, v. 1.00)

Contents

5 Local Behavior of Holomorphic Functions

In this chapter, our goal is to study the local behavior of holomorphic functions from various perspectives. We begin by examining ways to count the number of zeroes and poles of a meromorphic function in a given region and then describe ways to estimate the number of zeroes. Next, we establish the open mapping theorem: every nonconstant holomorphic function maps open sets to open sets. We then study holomorphic functions on the unit disc in some detail, providing some simple but powerful growth estimates and describing the action of fractional linear transformations on such functions. Finally, we discuss conformal mapping, which allows us to use holomorphic functions to transform complex regions into one another.

5.1 Locations of Zeroes and Poles

- In this section we will study the locations of zeroes and poles of holomorphic functions.

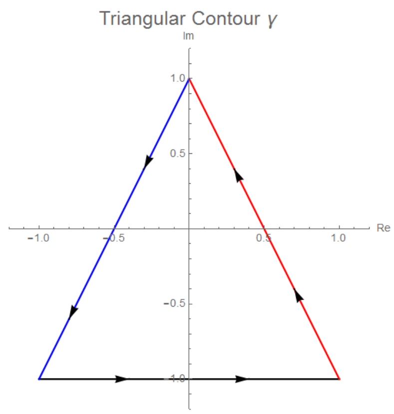
5.1.1 Counting Zeroes and Poles

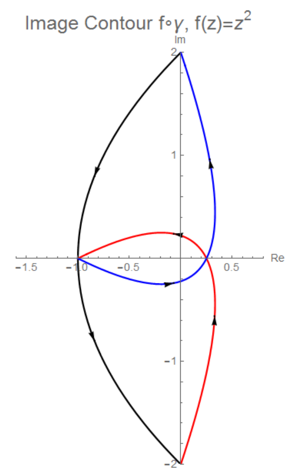
- Our first goal is to establish a procedure for counting the number of zeroes and poles a meromorphic function f possesses in a region R.
  - Suppose that f is meromorphic on R and not identically zero. Then by direct calculation we can see that f'/f has singularities only where f has a zero or a pole, since otherwise both f and f' are holomorphic and f is nonzero (so that the quotient f'/f is also holomorphic).
  - If f has a zero or pole at z_0 of order k, then by taking out a factor of $(z-z_0)^k$ from the Laurent expansion we see that $f(z) = (z-z_0)^k g(z)$ where $g(z_0) \neq 0$. Then $\dfrac{f'(z)}{f(z)} = \dfrac{k(z-z_0)^{k-1}g(z) + (z-z_0)^k g'(z)}{(z-z_0)^k g(z)} = \dfrac{k}{z-z_0} + \dfrac{g'(z)}{g(z)}$, and since $\dfrac{g'}{g}$ is holomorphic at z_0 because $g(z_0) \neq 0$, we see that f'/f has a simple pole at z_0 with residue k (the residue being positive when f has a zero and negative when f has a pole).
  - By applying the residue theorem to f'/f we can count zeroes and poles of f:

- Theorem (Counting Zeroes and Poles): Suppose that f is meromorphic on R, not identically zero, and γ is a simple closed counterclockwise contour in R such that f has no zeroes and no poles on γ. Then $\int_\gamma \frac{f'(z)}{f(z)}\,dz = 2\pi i \cdot [N_0(f) - N_\infty(f)]$ where $N_0(f)$ is the number of zeroes of f inside γ and $N_\infty(f)$ is the number of poles of f inside γ, both counted with multiplicity.
  - Proof: As noted above, the poles of $\frac{f'(z)}{f(z)}$ are all simple and the residue at z_0 is equal to the order of the zero of f at z_0 (if f has a zero) or the negative of the order of the pole at z_0 (if f has a pole).
  - Therefore, by the residue theorem, if we sum over all zeroes and poles, we see immediately that $\int_\gamma \frac{f'(z)}{f(z)}\,dz = 2\pi i \sum_{z_i \in \text{poles of } f'/f} \text{Res}_{f'/f}(z_i) = 2\pi i[N_0(f) - N_\infty(f)]$ as claimed.
- As an immediate obvious corollary, we can count the number of zeroes of a holomorphic function on a region:
- Corollary: Suppose that f is holomorphic on R, not identically zero, and γ is a simple closed counterclockwise contour in R such that f has no zeroes on γ. Then $\int_\gamma \frac{f'(z)}{f(z)}\,dz = 2\pi i \cdot N_0(f)$ where $N_0(f)$ is the number of zeroes of f inside γ counted with multiplicity.
  - Proof: Apply the theorem above and observe that f has no poles because it is holomorphic.
- As examples we can verify the calculation for some simple functions.
- Example: Verify the zero-counting result for $f(z) = z^3$ inside the circle $|z| = 1$.
  - Obviously f has a triple zero at $z = 0$. We have $f'/f = 3z^2/z^3 = 3/z$, and with γ the boundary of the unit circle we indeed get $\int_\gamma \frac{3}{z}\,dz = 3 \cdot 2\pi i$, as expected.
- Example: Verify the zero-and-pole-counting result for $f(z) = \frac{8z}{z-1}$ inside the circle $|z| = 2$.
  - Obviously f has a simple zero at $z = 0$ and a simple pole at $z = 1$. We have $f'(z) = -\frac{8}{(z-1)^2}$ so $f'/f = -\frac{1}{z(z-1)} = \frac{1}{z} - \frac{1}{z-1}$.
  - With γ the boundary of $|z| = 2$ we get $\int_\gamma \frac{f'(z)}{f(z)}\,dz = \int_\gamma [\frac{1}{z} - \frac{1}{z-1}]\,dz = 2\pi i - 2\pi i = 0$ as expected.
- It is not obvious why precisely one would think of considering the integral of $\frac{f'(z)}{f(z)}$ at all, but there are various ways to explain why this is in fact a natural idea.
  - The most direct is to observe that $\frac{f'(z)}{f(z)}$ is the logarithmic derivative $\frac{d}{dz}[\log[f(z)]]$, and so by the fundamental theorem of calculus, the integral $\int_\gamma \frac{f'(z)}{f(z)}\,dz$ should be "equal" to the difference in the values of $\log[f(z)]$ at the two endpoints of the contour.
  - Of course, the complex logarithm is multivalued, so this evaluation does not really make sense as written, but if we allow some abuse of notation, the integral should be the difference of two values of the complex logarithm at the endpoints of the contour, which (since the contour is closed) is an integer multiple of $2\pi i$.
  - We have encountered a similar phenomenon previously when we formalized the notion of winding numbers: in fact, we can rephrase this argument entirely in terms of winding numbers.
  - Explicitly, suppose we have a parametrization of γ as $\gamma(t)$ for $a \le t \le b$: then $\int_\gamma \frac{f'(z)}{f(z)}\,dz = \int_a^b \frac{f'(\gamma(t))}{f(\gamma(t))}\gamma'(t)\,dt$.

- But, motivated by the substitution $s = f(\gamma(t))$ with $ds = f'(\gamma(t))\gamma'(t)\,dt$, we can see that this integral also equals the line integral $\int_\Gamma \frac{1}{z}\,dz$ on the contour $\Gamma = f \circ \gamma$ defined with $\Gamma(t) = f(\gamma(t))$ for $a \le t \le b$.
- From our discussion of winding numbers, $\int_\Gamma \frac{1}{z}\,dz$ equals $2\pi i$ times the winding number of the contour $f \circ \gamma$ around $z = 0$.
- Thus, in other words, the zero-and-pole-counting calculation above can be equivalently formulated in terms of the winding behavior of the contour $\Gamma = f \circ \gamma$ around the origin.
- This interpretation is often called the argument principle since it is equivalent to tabulating the net change in the argument (i.e., the polar angle) of $f(\gamma(t))$ as t varies from a to b, which in turn is equivalent to counting the number of net crossings that the contour $f \circ \gamma$ makes across the positive real axis.

* Here is an illustration of the argument principle with the function $f(z) = z^2$ and the contour given by the counterclockwise boundary of the triangle with vertices $-1 - i$, $1 - i$, and i:





  - We can see that the triangle has winding number 1 around 0, whereas its image under the squaring map has winding number 2, reflecting the fact that $f(z) = z^2$ has 2 total zeroes (and no poles) counted with multiplicity inside the contour.

* Of course, as a practical matter, it is not generally so easy to compute the integral $\int_\gamma \frac{f'(z)}{f(z)}\,dz$ if f is sufficiently complicated in a way that makes it difficult to identify its zeroes and poles.
  - The zero-and-pole-counting theorem is of far more interest as a theoretical device.

5.1.2 Rouché's Theorem and Applications

* Suppose that f is holomorphic on a simply connected region R with boundary γ, and f is nonzero on γ.
  - Then, per the argument principle, the number of zeroes of f inside γ is equal to the winding number of $f \circ \gamma$ around the origin.
  - If we perturb f by a small amount to obtain a function g, then since γ is continuous the composition $f \circ \gamma$ will also be perturbed by a small amount to yield $g \circ \gamma$.
  - As long as the perturbation is small enough that it does not cause $f \circ \gamma$ to pass through the origin, then by our results on deformation of contours, $f \circ \gamma$ and $g \circ \gamma$ will have the same winding number around 0 and thus by the argument principle, f and g will have the same number of zeroes inside γ.
  - More precisely, we require that the homotopy function $h_t(s) = tf(s) + (1-t)g(s)$ not pass through the origin for any $s \in \gamma$ and any $0 \le t \le 1$.

- Since $h_t(s) = 0$ precisely when $tf(s) = (1-t)(-g(s))$, we must avoid the situation where one of $f(s)$ and $-g(s)$ is a nonnegative real multiple of the other. Conveniently, this is precisely the situation where the triangle inequality holds: $|f(s) - g(s)| = |f(s)| + |g(s)|$ if and only if one of $f(s)$ and $-g(s)$ is a nonnegative real multiple of the other.
- Therefore, if we insist on strict inequality in the triangle inequality, namely, that $|f(s) - g(s)| < |f(s)| + |f(s)|$, then the homotopy function $h_t(s)$ cannot take the value zero, and so we may deform the contour $f \circ \gamma$ into $g \circ \gamma$ without passing through the origin. By the argument principle, we can then conclude that f and g have the same number of zeroes inside R.
- We can formalize this as follows:

- Theorem (Rouché's Theorem): Suppose that f and g are holomorphic on a simply connected bounded region R with counterclockwise boundary γ. If $|f(z) - g(z)| < |f(z)| + |g(z)|$ for all $z \in \gamma$, then f and g have the same number of zeroes inside γ, counted with multiplicity.
  - The argument we gave above using deformation of contours is already essentially rigorous, but we will give another approach for illustration.
  - Proof: Suppose that $|f(z)| < |f(z) - g(z)| + |g(z)|$ requires both $f(z)$ and $g(z)$ not to vanish on γ, so (in particular) they also cannot be identically zero.
  - It therefore suffices to show that $\int_\gamma \dfrac{f'(z)}{f(z)}\,dz = \int_\gamma \dfrac{g'(z)}{g(z)}\,dz$ since these integrals calculate $2\pi i$ times the number of zeroes that f and g respectively have inside γ.
  - Additionally, we note that $\dfrac{f(z)}{g(z)}$ cannot be a nonpositive real number $-\alpha$ on γ: if $f(z) = -\alpha g(z)$ then $\left|\dfrac{f(z)}{g(z)} - 1\right| = |-\alpha - 1| = -\alpha - 1 = \left|\dfrac{f(z)}{g(z)}\right| - 1$ so multiplying by $g(z)$ would give $|f(z) - g(z)| = |f(z)| - |g(z)|$, contradicting the given hypothesis.
  - Now observe that $\int_\gamma \left[\dfrac{f'(z)}{f(z)} - \dfrac{g'(z)}{g(z)}\right] dz = \int_\gamma \dfrac{d}{dz}[\log \dfrac{f(z)}{g(z)}]\,dz$ where the logarithm's branch cut is taken along the negative real axis. Since $\dfrac{f(z)}{g(z)}$ does not take any nonpositive real values on γ, the integrand $\log \dfrac{f(z)}{g(z)}$ is holomorphic along γ.
  - Thus, by the fundamental theorem of calculus, the integral $\int_\gamma \dfrac{d}{dz}[\log \dfrac{f(z)}{g(z)}]\,dz$ is zero, and so we immediately conclude $\int_\gamma \dfrac{f'(z)}{f(z)}\,dz = \int_\gamma \dfrac{g'(z)}{g(z)}\,dz$ as desired.
  - Remark: The connection between this argument and the one above in terms of winding numbers is that for $F(z) = f(z)/g(z)$, the contour $F \circ \gamma$ never crosses the nonpositive real axis, and therefore it has winding number 0 around the origin.

- In practice, one often uses the weaker hypothesis $|f(z) - g(z)| < |f(z)|$ when applying Rouché's theorem.
  - Indeed, Rouché's original formulation was in terms of the perturbation $h(z) = f(z) - g(z)$, in which case the condition is $|h(z)| < |f(z)|$: the perturbation is smaller in absolute value than the original function.
  - By running the condition the other way one can also take the hypothesis $|h(z)| < |g(z)|$: the perturbation is smaller in absolute value than the new function.
  - The version using the weaker hypothesis $|f(z) - g(z)| < |f(z)| + |g(z)|$, which is symmetric in f and g, was given by Estermann in 1962. This version is also equivalent to the more obviously symmetric version of the theorem with hypothesis $|f(z) + g(z)| < |f(z)| + |g(z)|$ upon replacing g with $-g$ (since g and $-g$ obviously have the same zeroes with the same orders).

- Although the hypothesis in Rouché's theorem may seem rather unintuitive, there is a quite nice nice geometric interpretation.

- Imagine a person walking their leashed dog around a flagpole. The dog (being a dog) will run around its owner, sometimes staying close and other times moving far away, with the owner letting the leash out and reeling it back in as appropriate.
- The principle is that as long as the dog's leash is not long enough to reach all the way to the flagpole, from wherever the owner is currently located, then once the owner finishes circulating around the flagpole and returns to their starting position, the dog has circled the flagpole exactly the same number of times as the owner (or, in terms of a more likely situation in real life: the dog has not managed to get the leash tangled around the flagpole!).
- If the owner's position at time t is $f(z)$ while the dog's position is $g(z)$ for all $z \in \gamma$, then the leash represents the difference $f(z) - g(z)$, and so the condition simply says that $|f(z) - g(z)| < |f(z)|$: this is Rouché's original hypothesis. (Our formulation in terms of the owner and dog walking around the flagpole are rephrasing the zero-counting in terms of winding numbers using the argument principle.)
- In this formulation, the more general hypothesis $|f(z) - g(z)| < |f(z)| + |g(z)|$ states that the leash can never intersect the flagpole: although the owner and dog can be on opposite sides of the flagpole, the line joining them can never pass through the flagpole itself. (Intuitively, any "tangling" of the leash necessarily must start with the leash touching the flagpole.)

- <u>Example</u>: Show that $f(z) = z^4 + 2z + 5$ has no zeroes inside $|z| = 1$ but has four zeroes inside $|z| = 2$.
  - For $|z| = 1$ the largest contribution to f comes from the constant term 5, so we try comparing $f(z) = z^4 + 2z + 5$ to $g(z) = 5$ with associated perturbation $h(z) = f(z) - g(z) = z^4 + 2z$.
  - When $|z| = 1$ we have $|h(z)| = \left|z^4 + 2z\right| \le |z|^4 + 2|z| \le 3 < 5 = |g(z)|$, so the condition $|h(z)| < |g(z)|$ is satisfied.
  - Thus by Rouché's theorem, f and g have the same number of zeroes inside $|z| = 1$: namely, none, since g is nonzero everywhere.
  - For $|z| = 2$ the largest contribution to f comes from the term z^4, so we try comparing $f(z) = z^4 + 2z + 5$ to $g(z) = z^4$ with associated perturbation $h(z) = 2z + 5$.
  - When $|z| = 2$ we have $|h(z)| = |2z + 5| \le 2|z| + 5 \le 9 < 16 = |g(z)|$, so the condition $|h(z)| < |g(z)|$ is satisfied.
  - Thus by Rouché's theorem, f and g have the same number of zeroes inside $|z| = 1$: namely, 4, since g clearly has a single zero of multiplicity 4 at the origin.

- <u>Example</u>: Show that $f(z) = z^9 - 5iz^4 + 2$ has exactly five zeroes in the annulus $1 < |z| < 2$.
  - First observe that for $|z| = 1$ the middle term $-5iz^4$ is largest, while for $|z| = 2$ the leading term z^9 is largest.
  - For $|z| = 1$ we compare $f(z) = z^9 - 5iz^4 + 2$ to $g(z) = -5iz^4$ with associated perturbation $h(z) = z^9 + 2$.
  - When $|z| = 1$ we have $|h(z)| = \left|z^9 + 2\right| \le |z|^9 + 2 = 3 < 5 = \left|-5iz^4\right| = |g(z)|$ so by Rouché's theorem we see that f has the same number of zeroes as g inside $|z| = 1$, which is 4.
  - Likewise, for $|z| = 2$ we compare $f(z)$ to $g(z) = z^9$ with associated perturbation $h(z) = -5iz^4 + 2$.
  - When $|z| = 2$ we have $|h(z)| = \left|-5iz^4 + 2\right| \le 5|z|^4 + 2 \le 82 < 2^9 = |g(z)|$ so by Rouché's theorem we see that f has the same number of zeroes as g inside $|z| = 2$, which is 9.
  - So since there are 9 zeroes inside $|z| = 2$ but only 4 inside $|z| = 1$, and there are no zeroes on either circle, the remaining 5 are in the annulus $1 < |z| < 2$.

- <u>Example</u>: Show that $f(z) = 3 + z + e^z$ has exactly one zero in the left half-plane $\text{Re}(z) \le 0$, and that it is real.
  - Rouché's theorem only applies on bounded regions, so in order to count the zeroes of $f(z)$ on the unbounded left half-plane, we instead count the number of zeroes of f in the semicircular region $|z| \le R$ with $\text{Re}(z) \le 0$ and then take $R \to \infty$.
  - Since $|e^z| = e^{\text{Re}(z)} \le e$ is relatively small for z in the left half-plane, we can compare $f(z) = 3 + z + e^z$ to $g(z) = 3 + z$ with associated perturbation $h(z) = f(z) - g(z) = e^z$.

- Then for $\text{Re}(z) \le 0$ (hence in particular on the semicircle of interest) we have $|h(z)| \le e^{\text{Re}(z)} \le e$, while $|g(z)| \ge 3$ on the imaginary axis and $|f(z)| \ge |z| - 3 \ge R - 3$ on the semicircle $|z| = R$, so for $R > 3 + e$ we see that $|h(z)| < |g(z)|$.
- Thus for $R > 3 + e$ we see that f and g have the same number of zeroes inside the semicircle, which is clearly 1 since g has the single zero $z = -3$. Taking $R \to \infty$ shows that f has exactly one zero in the left half-plane, as desired.
- To see that the zero is real we simply invoke the intermediate value theorem: since $f(-4) = e^{-4} - 1 < 0$ and $f(-3) = e^{-3} > 0$ and $f(z) = 3 + x + e^x$ is continuous, by the intermediate value theorem f has a real root in the interval $(-4, -3)$, which by the above is the only root of f in the left half-plane.

- We can also use Rouché's theorem to give another quick proof of the fundamental theorem of algebra:
- Theorem (Fundamental Theorem of Algebra): If $p(z)$ is a polynomial of degree d, then $p(z)$ has d complex roots, counted with multiplicity.
  - Proof: Suppose $p(z) = a_d z^d + \sum_{k=0}^{d-1} a_k z^k$ with $a_d \neq 0$. For large $|z|$ the largest contribution to p comes from the leading term $a_d z^d$ so we apply Rouché's theorem to $f(z) = p(z)$ and $g(z) = a_d z^d$, with perturbation $h(z) = \sum_{k=0}^{d-1} a_k z^k$, on a sufficiently large circle $|z| = R$.
  - Explicitly, take $R > \max(1, \sum_{k=0}^{d-1} |a_k| / |a_d|)$, so that $1 < R$ and $\sum_{k=0}^{d-1} |a_k| < R |a_d|$.
  - Then for $|z| = R$ we have $|h(z)| = \left|\sum_{k=0}^{d-1} a_k z^k\right| \le \sum_{k=0}^{d-1} \left|a_k z^k\right| = \sum_{k=0}^{d-1} |a_k| R^k \le \sum_{k=0}^{d-1} |a_k| R^k \le \sum_{k=0}^{d-1} |a_k| R^{d-1} = R^{d-1} \sum_{k=0}^{d-1} |a_k| < R^{d-1} |a_d| R = \left|a_d z^d\right| = |f(z)|$.
  - Therefore, the condition $|h(z)| < |f(z)|$ for Rouché's theorem is satisfied, and so f and g have the same number of roots inside $|z| = R$, which is clearly d since g has a zero of multiplicity d at the origin.
  - Taking $R \to \infty$ yields that f has d roots in $\mathbb{C}$, as desired.

5.1.3 The Open Mapping Theorem, Local Invertibility and Local Preimages

- Using the argument principle and Rouché's theorem, we can establish another very fundamental local property of nonconstant holomorphic functions: namely, that they are open mappings, in the sense that they map open sets to open sets.
  - Before proving the main result, we make some preliminary remarks about the non-obviousness of this result: namely, that it fails completely for real-valued functions.
  - As a simple example, observe that the infinitely differentiable real-valued function $f(x) = \sin x$ maps the open interval $(0, 6\pi)$ to the non-open set $[-1, 1]$ (which is in fact closed).
  - Even non-holomorphic functions from $\mathbb{C}$ to $\mathbb{C}$ need not map open sets to open sets: for example, $f(z) = |z|^2$ maps the open disc $|z| < 1$ to the interval $[0, 1)$, which is not even open as a subset of $\mathbb{R}$ (much less $\mathbb{C}$), even though as a function on real and imaginary parts $f(x + iy) = x^2 + y^2$ has partial derivatives of all orders.
  - In general topology, the open mapping property is used as a definition of continuity: if X and Y are topological spaces, then $f : X \to Y$ is continuous precisely when the inverse-image relation f^{-1} maps open sets to open sets: more precisely, f is continuous precisely when for any open subset V of Y, the set $f^{-1}(V) = \{x \in X : f(x) \in V\}$ is an open subset of X.
  - Thus, the fact that nonconstant holomorphic functions map open sets to open sets is yet another reflection that their local behavior is far more pleasant than can be expected for more general functions.
  - More explicitly, nonconstant holomorphic functions are "locally onto": if $f(z_0) = w_0$, then near z_0, f takes all values sufficiently near w_0, in the sense that for any $\epsilon > 0$ there exists an $r > 0$ such that for any w with $|w - w_0| < r$, there exists some z with $|z - z_0| < \epsilon$ with $f(z) = w$.
  - Indeed, this property is equivalent to saying that f maps open sets, since the image under f of the disc $|z - z_0| < \epsilon$ necessarily contains an open disc $|w - w_0| < r$ around any arbitrary point w_0 in the image of f.

- We now establish the open mapping property of holomorphic functions:

- Theorem (Open Mapping Theorem): Suppose that U is a connected open set and $f : U \to \mathbb{C}$ is a nonconstant holomorphic function on U. Then $f(U)$ is open.
  - Proof: Let $z_0 \in U$: then since U is open it contains some open disc $|z - z_0| < R$. Let $f(z_0) = w_0$ and set $g(z) = f(z) - w_0$, so that g is zero at z_0.
  - Since f hence g is not constant on U, by our results on analytic functions since g is zero at z_0, it is nonzero on some punctured disc $0 < |z - z_0| < r$, for an appropriate positive $0 < r < R$. Take γ to be the circle $|z - z_0| = r/2$.
  - Then since g is continuous and γ is closed, by the extreme value theorem $|g(z)|$ attains its minimum M on γ, so $|g(z)| \geq M$, and since g is nonzero on γ we have $M > 0$.
  - Now for any $w \in \mathbb{C}$ with $|w - w_0| < M$, we have $|w - w_0| < M \leq |g(z)|$, so by Rouché's theorem, we see that $g(z) - (w - w_0)$ has the same number of zeroes inside γ as g does.
  - Since $g(z_0) = 0$ that means $g(z) - (w - w_0) = 0$ for some $|z - z_0| < r/2$, which is to say, for any w with $|w - w_0| < M$ there exists z with $|z - z_0| < r/2$ such that $f(z) = w$.
  - This shows that the image of f is open, and so f is an open mapping as claimed.

- As noted above, the open mapping theorem states that nonconstant holomorphic functions are locally onto. Another natural question is: are nonconstant holomorphic functions locally one-to-one?
  - Equivalently, we are asking whether nonconstant holomorphic functions are locally invertible, in the sense that for any z_0 there exists an open set U containing z_0 and an open set V containing $f(z_0)$ such that $f : U \to V$ has an inverse function $f^{-1} : V \to U$.
  - We have previously invoked the inverse function differentiation formula $(f^{-1})'(w) = \dfrac{1}{f'(f^{-1}(w))}$, which with $w = f(z_0)$ equivalently states that $(f^{-1})'(f(z_0)) = \dfrac{1}{f'(z_0)}$, and so under the reasonable suspicion that f^{-1} should be holomorphic, we must necessarily have $f'(z_0) \neq 0$ in order for $(f^{-1})'$ to be defined.
  - In fact, this condition is also sufficient, and in such a case the local inverse f^{-1} will also be holomorphic:

- Theorem (Local Invertibility): Suppose that f is holomorphic on an open region R. Then for any $z_0 \in R$ with $f'(z_0) \neq 0$, f is locally one-to-one near z_0, in the sense that there exists an open set U containing z_0 and an open set V containing $f(z_0)$ such that there exists a holomorphic function $g : V \to U$ such that $f(g(v)) = v$ for all $v \in V$ and $g(f(u)) = u$ for all $u \in U$.
  - The analogous result for real-differentiable functions is known as the inverse function theorem.
  - Proof: By translating in the domain and translating and rescaling f, we may assume that $z_0 = 0$, $f(z_0) = 0$ and $f'(z_0) = 1$, so that $f(z)$ has a series expansion $f(z) = z + z^2 q(z)$ where $q(z) = \sum_{n=2}^{\infty} a_n z^n$.
  - Then by our results on local series expansions, q is holomorphic near 0 hence is bounded on some disc $|z| < R$: say with $|q(z)| \leq M$ on this disc.
  - Now take $r < 1/(2M)$, select any ζ with $|\zeta| < r/2$, and consider $f_\alpha(z) = f(z) - \alpha$. We perturb f_α by $h(z) = z^2 q(z) - \alpha$ to compare f_α to $g(z) = z$.
  - For $|z| = r$, we have $|h(z)| \leq \left|z^2 q(z)\right| + |\alpha| \leq r^2 M + r/2 < r = |z| = |g(z)|$, so the hypothesis of Rouché's theorem is satisfied.
  - Therefore since g has a unique zero of order 1 inside $|z| < r$, so does f_α.
  - This means $f(z) = \alpha$ has a unique solution inside $|z| < r$ for any $|\alpha| < r/2$.
  - So now define $U = \{z : |z| < r \text{ and } |f(z)| < r/2\}$. This set is open because it is the intersection of the open disc $|z| < r$ with the open set $|f(z)| < r/2$ (which is open because f is continuous).
  - We have just shown that f maps U bijectively onto the open set $V = \{z : |z| < r/2\}$. Since f is an open mapping by the open mapping theorem, the inverse bijection $g : V \to U$ is continuous.
  - Next, since g is continuous, we may apply the inverse function differentiation formula to see that g is differentiable at z_0 and that $g'(0) = \dfrac{1}{f'(g(0))} = \dfrac{1}{f'(0)}$.

- Finally, since $f'(z_0) \neq 0$, since f' is continuous all of this discussion also applies to all points in some open disc containing z_0, so in fact g is differentiable on an open disc around z_0, meaning it is holomorphic.

- Let us now examine more closely the situation where $f(z)$ is not locally invertible at $z = z_0$.
  - For $f(z) = z^2$, we see that f is not locally invertible at $z = 0$, since $f'(0) = 0$. Indeed, we can see directly that f is not one-to-one on any disc $|z| < r$ centered at the origin: it is in fact two-to-one, since each point other than $z = 0$ in the image disc $|z| < r^2$ has exactly 2 preimages in the disc $|z| < r$.
  - In the same way we can see that $f(z) = z^d$ is d-to-one inside any disc $|z| < r$, since $f(z) = f(ze^{2\pi ik/d})$ for each $0 \leq k \leq d$, so each point other than $z = 0$ in the image disc $|z| < r^d$ has d preimages in the disc $|z| < r$.
  - It may seem that this local behavior is determined by the order of the zero of $f(z)$ at $z = z_0$, but the function $f(z) = z^2 + 1$ is also two-to-one on any disc $|z| < r$, for the same reason: shifting the image of the function does not affect the number of preimages.
  - Indeed, if we replace $f(z)$ by $f(z) - f(z_0)$, then the local behavior appears to be controlled by the order of the zero of $f(z) - f(z_0)$ at z_0.
  - This is in fact quite sensible if we look at the series expansion for f: if $f(z_0) = w_0$, and $f(z_0) - w_0$ has a zero of order d at z_0, then $f(z_0) = w_0 + a_d(z - z_0)^d + \sum_{n=d+1}^{\infty} a_n(z - z_0)^n$, and so for small $|z - z_0|$ we have $f(z_0) - w_0 = a_d(z - z_0)^d + O(|z - z_0|^{d+1})$, and the O-term will be a sufficiently small perturbation that $f(z)$ will still be locally d-to-one near z_0.
  - As with the local invertibility theorem, we can use Rouché's theorem to make the argument precise.

- Definition: If f is holomorphic and nonconstant on R and $z_0 \in R$, the multiplicity of z_0 is the order of the zero of $f(z) - f(z_0)$ at z_0. A point of multiplicity 1 is called simple.
  - We observe that points of multiplicity greater than 1 are the same as the zeroes of $f'(z)$, and are thus isolated under the assumption that f is nonconstant.

- Theorem (Local Preimages): Suppose that f is holomorphic and nonconstant on a connected open region R. For any $z_0 \in R$, if z_0 is a point of multiplicity d for f, then there exist positive ϵ and r such that each w with $0 < |\alpha - f(z_0)| < \epsilon$ has exactly d distinct preimages inside the disc $|z - z_0| < r$, and each preimage is a simple point.
  - The proof is essentially the same as the local invertibility theorem, aside from making a minor adjustment to avoid points of higher multiplicity.
  - Proof: If $d = 1$ the result follows immediately from the local invertibility theorem, so otherwise assume $d > 1$. As in the local invertibility theorem we may translate and rescale to assume $z_0 = 0$, $f(z_0) = 0$, and $f^{(d)}(z_0) = 1$. Then $f(z) = z^d + z^{d+1}q(z)$ for some holomorphic $q(z)$ with $|q(z)| \leq M$ for $|z| < R_1$.
  - As noted above, since the points of f of multiplicity > 1 are isolated, we may select R such that the only point of multiplicity > 1 for f inside $|z| < R_2$ is $z = 0$ itself.
  - Then take $r < \min(R_1, 1/(2M), R_2)$, select any ζ with $|\zeta| < r^d/2$, and consider $f_\alpha(z) = f(z) - \alpha$. We perturb f_α by $h(z) = z^{d+1}q(z) - \alpha$ to compare f_α to $g(z) = z^d$.
  - For $|z| = r$, we have $|h(z)| \leq \left|z^{d+1}q(z)\right| + |\alpha| \leq r^{d+1}M + r/2 < r^d = \left|z^d\right| = |g(z)|$, so the hypothesis of Rouché's theorem is satisfied.
  - Therefore since g has d total zeroes inside $|z| < r$, so does f_α.
  - This means $f(z) = \alpha$ has a total of d solutions (with multiplicity) $|z| < r$ for any $|\alpha| < r/2$. But since the only points with $0 < |z| < r < R_2$ are simple points, there are exactly d solutions to $f(z) = \alpha$ and they are all simple.

- We will remark that the "branch points" (or "ramified points"), which as we have seen are simply the zeroes of $f'(z)$, carry substantial geometric information about the behavior of the function f.

- We finish by noting the following useful property about images of boundaries under holomorphic maps.
  - Recall that if R is a region, then ∂R denotes the boundary of R.

- Theorem (Images of Boundaries): Suppose R is a bounded region and $f : R \to \mathbb{C}$ is a nonconstant holomorphic function. If U is an open subset of R, then $\partial f(U) \subseteq f(\partial U)$. Furthermore, if f is one-to-one (equivalently, if f' is nonzero on U), then in fact $\partial f(U) = f(\partial U)$.
  - We also remark here that if U is connected, then $f(U)$ is connected since f is continuous.
  - Proof: Since U is open, by the open mapping theorem, we see that $f(U)$ is open.
  - Also let $\overline{U} = U \cup \partial U$ be the closure of U: then $\overline{U}$ is closed and bounded, so since f is continuous, the closure $\overline{f(U)}$ is equal to the set $f(\overline{U})$. (Explicitly, the elements in $\overline{f(U)}$ are limits of the form $\lim_{n\to\infty} f(z_i)$ for $z_i \in U$, which by passing f through the limit are the same as limits of the form $f(\lim_{n\to\infty} z_i)$, and these are elements of $f(\overline{U})$.)
  - Then the boundary $\partial f(U) = \overline{f(U)} \backslash f(U) = f(\overline{U}) \backslash f(U) \subseteq f(\overline{U} \backslash U) = f(\partial U)$, as claimed.
  - If f is one-to-one then necessarily $f(U)$ and $f(\partial U)$ are disjoint, and so since $f(\partial U) \subseteq f(\overline{U}) = \overline{f(U)}$, we see $f(\partial U) \subseteq \overline{f(U)} \backslash f(U) = \partial f(U)$. This is the reverse containment so $\partial f(U) = f(\partial U)$.
- We remark that the condition that f be one-to-one is necessary to have $\partial f(U) = f(\partial U)$ in the theorem above.
  - Specifically, take U to be the open semidisc $|z| < 1$ with $\text{Im}(z) > 0$ and $f(z) = z^4$. Then $f(U)$ is the punctured disc $0 < |z| < 1$ while ∂U is the half-circle $|z| = 1$ with $\text{Im}(z) \geq 0$ along with the interval $[-1, 1]$ on the real line.
  - Then $f(\partial U)$ is the circle $|z| = 1$ along with the interval $[0, 1]$ on the real line, but $\partial f(U)$ is only the circle $|z| = 1$ along with the point 0.

5.2 Conformal Mapping

- Now that we have established various local properties of holomorphic functions, we broaden our discussion to encompass more global properties.
  - Our main goal is to study holomorphic functions $f : U \to V$ that are both one-to-one and onto.
  - Such functions are variously called analytic isomorphisms (reflecting the fact that they are structure-preserving analytic maps), biholomorphic functions (reflecting the fact that they are holomorphic functions having a holomorphic inverse), or conformal mappings (reflecting the fact that they have the conformal angle-preserving property we discussed long ago).
  - The main utility of having such a map from U to V is that they allow us to transfer holomorphic functions from U to V (and vice versa) via composition: thus, U and V behave essentially equivalently from the perspective of complex analysis, and so understanding behavior on U allows us to understand behavior on V.
  - We will begin by studying the class of conformal mappings on the entire complex plane: these are the fractional linear transformations. So as to present a more unified discussion, we first introduce the notion of the point at infinity and the extended complex plane in the context of projective geometry.
  - We then discuss various ways to construct conformal mappings among various sets, with a particular emphasis on conformal maps involving the unit disc, whose behavior is rather natural to study.

5.2.1 The Point at Infinity and the Extended Complex Plane

- For real-valued functions, we have the useful notion of a function diverging to $+\infty$ or $-\infty$ at a point, which we can formalize using variations on the ϵ-δ definition of limit.
  - Rather than bothering with the details, we will just observe that, for example, $1/x^2 \to +\infty$ and $-1/x^2 \to -\infty$ as $x \to 0$.
  - We also have the related notion of a function having a limit (or diverging) as $x \to +\infty$ or $x \to -\infty$: for example, we have $e^x \to +\infty$ as $x \to +\infty$ and $e^x \to 0$ as $x \to -\infty$.
- For complex-valued functions, there are many more ways for a function to diverge to ∞.

- For example, the function $f(z) = 1/|z|$ tends uniformly to $+\infty$ ("positive real infinity") as $z \to 0$, whereas $g(z) = i/|z|$ tends uniformly to $+i\infty$ ("imaginary infinity"), in the sense that $1/|z|$ grows large along the positive real axis and $i/|z|$ grows large along the positive imaginary axis.
- Like with real-valued functions we could make precise all of the different "directions" in which a function can diverge to ∞, but in fact for holomorphic functions, all of these directions can only occur together.
- As an illustration, consider $f(z) = 1/z$ as $z \to 0$. If we select different directions of approach (e.g., along the path $\gamma(t) = t \cdot e^{-i\theta}$ as $t \to 0$ for fixed θ) then $f(z)$ will approach ∞ along the ray corresponding to the polar angle θ.
- So in this sense, we can simply write $1/z \to \infty$ as $z \to 0$, with the understanding that $1/z$ approaches ∞ "in every possible direction" as $z \to 0$.
- Likewise, even though there are many possible limits we could evaluate for a meromorphic function $f(z)$ as "$z \to \infty$" (along the positive real axis, along the negative imaginary axis, along the polar ray $\theta = \pi/4$, along the curve $\gamma(t) = t + t^2 i$ as $t \to \infty$, etc.), all of these limits will in fact display the same behavior: either the function f is bounded and has a limit as $|z| \to \infty$, the function is unbounded and tends to ∞ in every direction, or the function is bounded along some paths and unbounded along others.
- This observation follows by considering the nature of the singularity of $f(1/z)$ at $z = 0$, since as $|z| \to \infty$ we have $1/z \to 0$, the behavior of $f(z)$ as $|z| \to \infty$ is the same as that of $f(1/z)$ as $z \to 0$.
- Thus, if $f(1/z)$ has a removable singularity at $z = 0$ then $f(z)$ will have a limit as $z \to \infty$, if $f(1/z)$ has a pole at $z = 0$ then $f(z)$ will tend to ∞ (uniformly) as $z \to 0$, and if $f(1/z)$ has an essential singularity at $z = 0$ then $f(z)$ will be bounded along some paths and unbounded along others.

- Let us record these observations:
- Definition: Suppose f is holomorphic on the region $|z| > R$ for some R. Then there are three types of behavior as $z \to \infty$ based on the singularity of $f(1/z)$ at $z = 0$.

1. If $f(1/z)$ has a removable singularity at $z = 0$, we say f is holomorphic at ∞. In this case $f(z)$ tends to a limit L as $|z| \to \infty$.
2. If $f(1/z)$ has a pole at $z = 0$, we say f has a pole at ∞. In this case $|f(z)| \to \infty$ uniformly as $|z| \to \infty$.
3. If $f(1/z)$ has an essential singularity at $z = 0$, we say f has an essential singularity at ∞. In this case $|f(z)|$ is bounded along some paths but unbounded along others as $|z| \to \infty$. In fact, by Picard's little theorem, for any $R > 0$, on the region $|z| > R$ the function $f(z)$ will take all values in $\mathbb{C}$ with at most one exception.

- Example: For $f(z) = z/(z-1)$, we have $f(1/z) = 1/(1-z)$ which is holomorphic at $z = 0$. Thus, $f(z)$ is holomorphic at ∞, and $f(z) \to 1$ (the value of $f(1/z)$ at $z = 0$) as $z \to \infty$.
- Example: For $f(z) = z^2$, we have $f(1/z) = 1/z^2$ which has a pole at $z = 0$. Thus, $f(z)$ has a pole at ∞, and $|f(z)| \to \infty$ as $z \to \infty$ (which here is obvious, since $|f(z)| = |z|^2$).
- Example: For $f(z) = e^z$, we have $f(1/z) = e^{1/z}$ which has an essential singularity at $z = 0$. Thus, $f(z)$ has an essential singularity at ∞, and $f(z)$ displays erratic behavior as $z \to \infty$ along different paths. Indeed, along the positive real axis we have $f(z) \to +\infty$ while along the negative real axis we have $f(z) \to 0$ and along the imaginary axis the values of f simply circulate around the unit circle (so f is bounded, indeed periodic, but does not approach a limit).
- As with other singularities, we can classify the nature of the singularity at ∞ using the Laurent expansion of f.
- Explicitly, if $f(z) = \sum_{n=-\infty}^{\infty} a_n z^n$ on $|z| > R$, then $f(1/z) = \sum_{n=-\infty}^{\infty} a_n z^{-n} = \sum_{n=-\infty}^{\infty} a_{-n} z^n$.
- Thus, if there are no terms a_{-n} with $n < 0$ (in other words, no terms a_n with $n > 0$) then f has a removable singularity at ∞, if there are only finitely many a_{-n} with $n < 0$ (i.e., finitely many a_n with $n > 0$) then f has a pole at ∞, and if there are infinitely many a_{-n} with $n < 0$ (i.e., infinitely many a_n with $n > 0$) then f has an essential singularity at ∞.
- As an immediate consequence, we see that if f is entire, then f is holomorphic at ∞ if and only if f is constant, f has a pole at ∞ if and only if f is a polynomial, and otherwise (if f is not a polynomial) then f has an essential singularity at ∞.

- We can make our discussion a bit more precise by including the point ∞ explicitly in our discussion. Intuitively, we "glue" the infinite extent of $\mathbb{C}$ into a single point at infinity, and extend the topology of $\mathbb{C}$ to include the point ∞ using the map $z \mapsto 1/z$:
- Definition: The extended complex plane is the set $\mathbb{C} \cup \{\infty\}$. A subset U of $\mathbb{C} \cup \{\infty\}$ is defined to be open if (i) $U \cap \mathbb{C}$ is open in $\mathbb{C}$ and (ii) if $\infty \in U$ then $U \cap \mathbb{C}$ must also contain the set $|z| > R$ for some R.
    - The idea is that we define the "open discs around ∞" to be ∞ along with the sets $|z| > R$, which are the images of the open discs $|z| < 1/R$ under the map $z \mapsto 1/z$. Then a subset U of $\mathbb{C} \cup \{\infty\}$ is open if and only if it contains an open disc around each of its points.
    - By construction, the map $z \mapsto 1/z$ on $\mathbb{C} \cup \{\infty\}$ is continuous, and so since this map is its own inverse, it is in fact a homeomorphism of the extended complex plane.
    - The extended complex plane is a special case of the topological construction known as one-point compactification (which takes a topological space X and makes it compact by adding one extra point, whose open neighborhoods are selected in a way that "closes up" all of the non-compact closed sets).
- Viewing ∞ as an actual point in the extended complex plane also clarifies why meromorphic functions are in fact quite natural to consider, while essential singularities are less natural.
    - Explicitly, if f is holomorphic on a region $R \backslash \{z_0\}$ and has a pole at z_0, we can now define $f(z_0) = \infty$.
    - Since $f(z) \to \infty$ uniformly as z approaches the pole z_0, the function $f : R \to \mathbb{C} \cup \{0\}$ is now actually continuous at z_0.
    - So we see that a pole is just a point where the (now everywhere continuous) function f takes the value ∞.
    - Hence, the meromorphic functions on R are simply the functions that are continuous everywhere (as functions from R to $\mathbb{C} \cup \{\infty\}$ and are holomorphic except at an isolated set of points).
    - We can also define what it means for a function to be meromorphic at ∞: we simply require that f be holomorphic or have a pole at ∞, and also that the singularity at ∞ be isolated (meaning that there is some R for which there is no pole of f in the open disc $|z| > R$ around ∞).
- There is an appealing geometric description of the extended complex plane using stereographic projection on the Riemann sphere.
    - Explicitly, consider the unit sphere $x^2 + y^2 + z^2 = 1$ in 3-dimensional real space. For any point $P = z = x + iy = (x, y, 0)$ in the xy-plane, let l be the line joining $(x, y, 0)$ to the north pole $(0, 0, 1)$ of the sphere, and consider the intersection of l with the sphere.
    - Explicitly, we can see that l is parametrized as $l(t) = t\langle x, y, 0\rangle + (1-t)\langle 0, 0, 1\rangle = \langle tx, ty, 1-t\rangle$, which intersects the sphere when $t^2x^2 + t^2y^2 + (1-t)^2 = 1$, yielding $t = \dfrac{2}{1+|z|^2}$. The point of intersection itself is then $\pi(z) = (\dfrac{2\mathrm{Re}(z)}{1+|z|^2}, \dfrac{2\mathrm{Im}(z)}{1+|z|^2}, \dfrac{1-|z|^2}{1+|z|^2})$.
    - We can see that this map is continuous, and as $|z| \to \infty$ approaches $(0, 0, 1)$ uniformly, so it is also one-to-one and onto. Additionally, the inverse function is also continuous (it is given explicitly by $\pi^{-1}(x_0, y_0, z_0) = \dfrac{x_0 + iy_0}{1 - z_0}$ for $z_0 \neq 1$ with $\pi^{-1}(0, 0, 1) = \infty$), so π gives a homeomorphism of the extended complex plane with the unit sphere.
    - The point at ∞ corresponds to the north pole, and all of the finite points in $\mathbb{C}$ form the rest of the sphere: the equator (the great circle lying in the xy-plane) corresponds to the unit circle, while the northern hemisphere corresponds to the exterior of the unit disc and the southern hemisphere corresponds to the interior of the unit disc.
- Another geometric description of the extended complex plane arises naturally from projective geometry.
    - Explicitly, we define the complex projective line $\mathbb{P}^1(\mathbb{C})$ to be the set of lines through the origin in $\mathbb{C}^2$. More explicitly, define $\mathbb{P}^1(\mathbb{C})$ to be the set of points $[z_0 : z_1]$ with $z_0, z_1 \in \mathbb{C}$ not both zero, where $P \sim Q$ if $P = \lambda Q$ for some nonzero $\lambda \in \mathbb{C}$.

- We use the notation $[z_0 : z_1]$ to evoke the idea of considering only the ratios between the coordinates, since for example we consider the points $[1 : 1]$ and $[2 : 2]$ to be the same since $[1 : 1] = \frac{1}{2}[2 : 2]$.
- When $z_1 \neq 0$ we see that $[z_0 : z_1] = [z_0/z_1 : 1]$ whereas when $z_1 = 0$ (in which case $z_0 \neq 0$ since both coordinates cannot be zero) we have $[z_0 : z_1] = [z_0 : 0] = [1 : 0]$.
- Therefore, if we identify the point $[\alpha : 1]$ with $\alpha \in \mathbb{C}$ and $[1 : 0]$ with ∞, we see that the extended complex plane corresponds with the points on the complex projective line $\mathbb{P}^1(\mathbb{C})$.
- Intuitively, the points on the projective line simply label the possible slopes of lines in $\mathbb{C}^2$: all elements from $\mathbb{C}$ along with ∞ (for vertical lines).
- Furthermore, $\mathbb{P}^1(\mathbb{C})$ inherits a natural quotient topology from $\mathbb{C}^2\backslash\{0\}$, and one may check directly that this topology agrees with the topology we defined above on the extended complex plane.
- The projective line (or more generally, projective n-space over an arbitrary field F, which is defined to be the set of lines through the origin in F^{n+1}) arises naturally in classical geometry.
- Geometrically speaking, parallel lines in $\mathbb{C}$, when extended to the projective line $\mathbb{P}^1(\mathbb{C})$, will intersect at the point at ∞.
- Because of the entirely uniform treatment of the point at ∞ as being entirely equivalent to any other point (rather than our approach above that requires special treatment of ∞, e.g., in classifying the behavior of the singularity there), many results have more natural statements when viewed projectively.

5.2.2 Fractional Linear Transformations

- We now study holomorphic bijections on various regions. To begin, we classify the holomorphic bijections from the extended complex plane to itself (one may reasonably view these as "meromorphic automorphisms" of the extended complex plane).
- We begin by showing that the only meromorphic functions on the extended complex plane are rational functions:
- Proposition (Meromorphic Functions on $\mathbb{C} \cup \{\infty\}$): A function $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is meromorphic everywhere if and only if it is a rational function of z.
  - Proof: Clearly each rational function is meromorphic on the extended complex plane, since any rational function has only finitely many poles (including potentially at ∞).
  - Conversely, suppose f is meromorphic on $\mathbb{C} \cup \{\infty\}$. Then f has no poles in $|z| > R$ for some R, and also f can only have finitely many poles in $|z| \leq R$ since otherwise the poles could not be isolated (as any infinite bounded sequence in $\mathbb{C}$ has a convergent subsequence by the Heine-Borel theorem).
  - Hence f has only finitely many poles in $\mathbb{C}$, so we may remove all of them by multiplying by an appropriate polynomial $q(z)$.
  - Then $q(z) \cdot f(z) = \sum_{n=0}^{\infty} a_n z^n$ can have only finitely many nonzero coefficients a_n since $q(z)f(z)$ cannot have an essential singularity at ∞. Therefore $q(z)f(z) = p(z)$ is a polynomial so that $f(z) = p(z)/q(z)$ is a rational function, as claimed.
- Now we can classify the meromorphic bijections of the extended complex plane using our characterization of one-to-one maps:
- Proposition (Meromorphic Bijections on $\mathbb{C} \cup \{\infty\}$): A function $f : \mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$ is meromorphic and one-to-one if and only if it is a nonconstant function of the form $f(z) = \dfrac{az + b}{cz + d}$ for some $a, b, c, d \in \mathbb{C}$.
  - Proof: Each nonconstant function $f(z) = \dfrac{az + b}{cz + d}$ is meromorphic everywhere and has an inverse function $f^{-1}(z) = \dfrac{dz - b}{-cz + a}$ by direct calculation, so each of these functions are meromorphic bijections.
  - Conversely, by the proposition above, if f is meromorphic everywhere then f must be a rational function of the form $\dfrac{p(z)}{q(z)}$ for some polynomials p and q.

- Now observe that this function can have at most one pole, since any pole z_0 is a location where $f(z_0) = \infty$ and f is a bijection. If the pole is at ∞ then replacing $f(z)$ with $f(1/z)$ yields a function with a pole at 0, so we may assume the pole is at a finite point z_0.
- Then we must have $f(z) = \dfrac{p(z)}{(z-z_0)^k}$ for some $p(z)$ with $p(z_0) \neq 0$. Now if $f(z)$ is one-to-one then $\dfrac{1}{f(z)} = \dfrac{(z-z_0)^k}{p(z)}$ is also one-to-one, but since this function is locally k-to-one near z_0 by our results, we must have $k = 1$.
- By applying the same argument to $\dfrac{1}{f(z)}$, we see that $p(z)$ must also have degree 1, and so $f(z)$ is a quotient of linear polynomials as claimed (and obviously it must be nonconstant in order to be one-to-one).

- These quotients of linear polynomials are quite fundamental and have many interesting properties:
- Definition: The fractional linear transformations are the nonconstant functions of the form $f(z) = \dfrac{az+b}{cz+d}$ for $a, b, c, d \in \mathbb{C}$. They are all bijections from the extended complex plane to itself.
  - We can see easily from the description that the unique zero of f is at $-b/a$ while the unique pole is at $-d/c$.
  - It is also easy to see that any fractional linear transformation can be obtained as a composition of scalings $z \mapsto az$ (which are in turn obtained as a composition of a dilation $z \mapsto rz$ and a rotation $z \mapsto e^{i\theta}z$), translations $z \mapsto z + b$, and inversions $z \mapsto 1/z$.
  - By basic linear algebra the function $f(z)$ is nonconstant whenever the coefficient vectors $\langle a, b\rangle$ and $\langle c, d\rangle$ are linearly independent in $\mathbb{C}^2$, which is equivalent to saying that the determinant $ad - bc$ is nonzero.
  - The fact that there is an obvious linear algebra condition arising here should raise some suspicion that linear algebra is more involved than may immediately be apparent.
- In fact, the fractional linear transformations arise very naturally, as actual linear transformations, when we work with the complex projective line $\mathbb{P}^1(\mathbb{C})$.
  - Explicitly, any linear transformation $T : \mathbb{C}^2 \to \mathbb{C}^2$ is of the form $T(z_0, z_1) = (az_0 + bz_1, cz_0 + dz_1)$ for some $a, b, c, d \in \mathbb{C}$.
  - These transformations naturally descend to linear transformations on $\mathbb{P}^1(\mathbb{C})$ via setting $T[z_0 : z_1] = [az_0 + bz_1 : cz_0 + dz_1]$, and they are well-defined on projective points since they are homogeneous: $T[\lambda z_0 : \lambda z_1] = [\lambda(az_0 + bz_1) : \lambda(cz_0 + dz_1)] = [az_0 + bz_1 : cz_0 + dz_1] = T[z_0 : z_1]$ for any $\lambda \neq 0$.
  - The only concern is that we may attempt to map a point $[z_0 : z_1]$ to an element not in $\mathbb{P}^1(\mathbb{C})$, which would require the coordinates $az_0 + bz_1$ and $cz_0 + dz_1$ to be zero simultaneously for some $(z_0, z_1) \neq 0$.
  - Thus we need to avoid the situation where the linear system $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ has a nonzero solution, which is equivalent to requiring the determinant $ad - bc$ of $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be nonzero.
  - Therefore, the linear transformations $T(z_0, z_1) = (az_0 + bz_1, cz_0 + dz_1)$ of $\mathbb{C}^2$ that yield well-defined maps on $\mathbb{P}^1(\mathbb{C})$ are precisely the ones with $ad - bc \neq 0$.
  - Finally, when we make the identification of the projective point $[z_0 : z_1] = [z_0/z_1 : 1]$ with the complex point $z = z_0/z_1$, since $T[z_0 : z_1] = [az_0 + bz_1 : cz_0 + dz_1] = [\dfrac{az_0 + bz_1}{cz_0 + dz_1} : 1] = [\dfrac{az+b}{cz+d} : 1]$, the corresponding complex map is simply $T(z) = \dfrac{az+b}{cz+d}$, our fractional linear transformation.
  - Indeed, the fact that $T(z) = \dfrac{az+b}{cz+d}$ arises as an actual linear transformation on the complex projective line is precisely the reason that T is called a fractional linear transformation.
  - Using the projective approach, however, gives us much more.

- For example, we can see immediately that the composition of fractional linear transformations is precisely the same as composition of invertible linear transformations (namely, via matrix multiplication).
- Explicitly, if $T(z) = \dfrac{az+b}{cz+d}$ and $\tilde{T}(z) = \dfrac{\tilde{a}z+\tilde{b}}{\tilde{c}z+\tilde{d}}$, then $(T \circ \tilde{T})(z) = \dfrac{Az+B}{Cz+D}$ where $\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$.
- Likewise, for $T(z) = \dfrac{az+b}{cz+d}$ corresponding to $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the inverse transformation T^{-1} corresponds to the inverse matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \dfrac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ hence $T^{-1} = \dfrac{(dz-b)/(ad-bc)}{(-cz+a)/(ad-bc)} = \dfrac{dz-b}{-cz+a}$.
- Furthermore, we also see that the composition of fractional linear transformations has a group structure, since composition of linear transformations (equivalently, multiplication of invertible matrices) also has a group structure.
- Specifically, the group of fractional linear transformations is isomorphic to the group $GL_2(\mathbb{C})$ of invertible 2×2 matrices, modulo the subgroup that reduces to the identity transformation on $\mathbb{C} \cup \{\infty\}$, which one may check directly is simply the nonzero diagonal matrices.

- The fractional linear transformations also have an extremely useful geometric property: namely, they preserve the class of generalized circles: circles together with lines (where we view a line as being a circle with "radius ∞").
  - One may verify this directly for each of the simpler maps that can be composed to obtain a general fractional linear transformation: it is obvious that scalings and translation map circles to circles and lines to lines, and for the inversion map one may just compute directly the image of the circle $|z - \alpha| = r$ under the inversion map.
  - Using the Riemann sphere model, one may see that lines in $\mathbb{C}$ correspond to circles on the Riemann sphere passing through the north pole of the sphere while circles in $\mathbb{C}$ correspond to circles on the sphere not passing through the north pole (these calculations follow from some tedious uses of the explicit coordinate conversions between the Riemann sphere and the extended plane).
  - However, we can also obtain this result in a more conceptually clean way using the projective approach.
  - Explicitly, the circle $|z - \alpha| = r$ for $\alpha \in \mathbb{C}$ and $r > 0$ has equation $(z - \alpha)(\overline{z} - \overline{\alpha}) = r^2$ which yields $z\overline{z} - \overline{\alpha}z - \alpha\overline{z} + (|\alpha|^2 - r^2) = 0$, which in projective coordinates $z = z_0/z_1$ has the form $Az_0\overline{z_0} + Bz_0\overline{z_1} + \overline{B}\overline{z_0}z_1 + Cz_1\overline{z}_1 = 0$ for $A = 1$, $B = -\overline{\alpha}$, and $C = |\alpha|^2 - r^2$.
  - The line $ax + by = c$ for real a, b, c also has equation $(a + bi)z + (a - bi)\overline{z} = 2c$ which in projective coordinates also has the form $Az_0\overline{z_0} + Bz_0\overline{z_1} + \overline{B}\overline{z_0}z_1 + Cz_1\overline{z}_1 = 0$ with $A = 0$, $B = a + bi$, and $C = 2c$.
  - Therefore we see that the generalized circles are precisely the curves with projective equation $Az_0\overline{z_0} + Bz_0\overline{z_1} + \overline{B}\overline{z_0}z_1 + Cz_1\overline{z}_1 = 0$ for some real A and C and some complex B (not all zero): lines have $A = 0$ while circles have $A \neq 0$.
  - The equation $Az_0\overline{z_0} + Bz_0\overline{z_1} + \overline{B}\overline{z_0}z_1 + Cz_1\overline{z}_1 = 0$ can be written in matrix form as $\overline{\begin{bmatrix} z_0 & z_1 \end{bmatrix}} \begin{bmatrix} A & B \\ \overline{B} & C \end{bmatrix} \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} = 0$, or more compactly as $\mathbf{z}^\dagger M \mathbf{z} = \mathbf{0}$ where $\mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$, the dagger represents the adjoint (conjugate transpose), and the matrix M has $M^\dagger = M$ (i.e., M is Hermitian).
  - If we then apply a fractional linear transformation T to z, which corresponds to applying a left matrix multiplication to $\mathbf{z}$, the new equation is $(T\mathbf{z})^\dagger M(T\mathbf{z}) = 0$ which is equivalent to $\mathbf{z}^\dagger(T^\dagger MT)\mathbf{z} = 0$: this has the same form, except the new coefficient matrix is $T^\dagger MT$ (which is still Hermitian).
  - Therefore, we see in particular that applying a fractional linear transformation to a generalized circle yields another generalized circle.

- As an actual practical matter, to compute a fractional linear transformation taking one generalized circle to another, it is usually easiest to compose various simpler maps such as scalings, translations, and inversions.

- In particular, since the image of any generalized circle is another generalized circle, we only need to find the images of a few points to identify the image of any generalized circle under a fractional linear transformation.
- Of particular note is that a fractional linear transformation only maps one point to ∞ (namely, its pole), and so only the generalized circles passing through the pole will be mapped to lines; the others will be mapped to circles.

- <u>Example</u>: Find the images of $|z-1|=1$, $|z|=2$, and $|z|=1$ under the map $f(z)=\dfrac{z}{z-2}$.
  - The pole of $f(z)$ is at $z=2$, which lies on $|z-1|=1$. Thus this circle will be mapped to a line.
  - To identify which line we just find two points on it. For example, both $z=0$ and $z=1+i$ lie on $|z-1|=1$, so $f(0)=0$ and $f(1+i)=-i$ will lie on the image of $|z-1|=1$. So the desired line is simply the imaginary axis $\boxed{\mathrm{Re}(z)=0}$.
  - Likewise, the pole of $f(z)$ also lies on $|z|=2$ so it is also mapped to a line. Both $z=2i$ and $z=-2i$ lie on the circle, so $f(2i)=\dfrac{1-i}{2}$ and $f(-2i)=\dfrac{1+i}{2}$ both lie on the line. So the desired line is $\boxed{\mathrm{Re}(z)=\dfrac{1}{2}}$.
  - The circle $|z|=1$ does not contain the pole of f so it is mapped to another circle. This circle will contain the points $f(1)=-1$, $f(-1)=\dfrac{1}{3}$, and $f(i)=\dfrac{1-2i}{5}$. Using some basic geometry to construct the circumcenter of this triangle shows that the circumcenter is $-\dfrac{1}{3}$ and has radius $\dfrac{2}{3}$, so the circle is $\boxed{\left|z+\dfrac{1}{3}\right|=\dfrac{2}{3}}$.

- <u>Example</u>: Find a fractional linear transformation mapping $|z|=1$ to $|z-2i|=3$.
  - The original circle has radius 1 and the new circle has radius 3, so first we want to scale by a factor of 3. Then we just need to translate the original center 0 to the new center $2i$.
  - This yields the linear map $T(z)=\boxed{3z+2i}$.

- <u>Example</u>: Find a fractional linear transformation mapping $|z|=1$ to the real axis.
  - Here we want to map a circle to a line, so the pole of the fractional linear transformation needs to be on the circle. Since we may choose arbitrarily, let us take the pole at $z=1$, in which case we want $T(z)=\dfrac{az+b}{z-1}$ for some a,b.
  - Since the image is determined by the images of any two points on $|z|=1$, let us try sending $z=-1$ to 0, which requires $a=b$, and then try to map $z=i$ to a real number.
  - Since $T(i)=\dfrac{a(i-1)}{i-1}=a$, we can choose any real number a that makes T nonconstant (i.e., $a\neq 0$). So for example $T(z)=\boxed{\dfrac{z+1}{z-1}}$ will map $|z|=1$ to the real axis.

- In fact, since one-to-one holomorphic functions map boundaries to boundaries, we can even identify the images of regions under fractional linear transformations, as long as the region's boundary consists of generalized circles.
  - To identify which of the various possible regions having a particular generalized circle as its boundary is the actual image $f(R)$, we can simply calculate $f(z)$ for some z in the interior of R: then $f(z)$ must lie inside $f(R)$.
  - Furthermore, since holomorphic functions are continuous, they preserve connectedness of regions, so if the original region is connected, so is the new region.

- <u>Example</u>: Find the images of $|z-1|\leq 1$, $|z|>2$, and $1<|z|<2$ under the map $f(z)=\dfrac{z}{z-2}$.

- We calculated previously that $|z-1|=1$ maps to the line $\text{Re}(z)=0$, so the closed disc $|z-1|\leq 1$ must map either to the upper half-plane or the lower half-plane, since these are the only two connected regions having the real line as their boundary.
- To identify which one it is we can observe that the center of the circle $z=1$ is in the interior of the region, so $f(1)=-1$ must lie in the interior of the image. Thus the image of $|z-1|\leq 1$ is the lower half-plane $\boxed{\text{Re}(z)\leq 0}$.
- Likewise, $|z|=2$ maps to the line $\text{Re}(z)=1/2$ so the open region $|z|>2$ maps either to the region above or below the line. Since $z=3$ lies in the original region, $f(3)=3$ lies in the image, so the image is the upper half-plane $\boxed{\text{Re}(z)\geq 1/2}$.
- Finally, since f maps $|z|=1$ to the circle $\left|z+\dfrac{1}{3}\right|=\dfrac{2}{3}$ and f maps $|z|=2$ to the line $\text{Re}(z)=1/2$, it must map the region $1<|z|<2$ to a region bordering both the circle and the line, and there is only one possibility: the region outside the circle and to the left of the line: $\left|z+\dfrac{1}{3}\right|>\dfrac{2}{3}$ and $\text{Re}(z)<1/2$. Indeed, one can check that $z=3/2$ with $f(z)=-3$ lies in this region, as expected.

- There are many other interesting properties of fractional linear transformations acting on the extended complex plane. For example, their action is strictly 3-transitive, in the following sense:
- <u>Proposition</u> (3-Transitivity of FLTs): Suppose z_0, z_1, z_∞ are three distinct elements of the extended complex plane. Then there exists a unique fractional linear transformation T such that $T(z_0)=0$, $T(z_1)=1$, and $T(z_\infty)=\infty$. More generally, if $w_0, w_1, w_\infty$are any three distinct elements of the extended complex plane, then there exists a unique fractional linear transformation T such that $T(z_0)=w_0$, $T(z_1)=w_1$, and $T(z_\infty)=w_\infty$.
  - <u>Proof</u>: Suppose that T is a fractional linear transformation such that $T(z_0)=0$, $T(z_1)=1$, and $T(z_\infty)=\infty$.
  - Then z_0 is the unique zero of T and z_∞ is the unique pole of T, which determines T up to a scaling factor: explicitly, we must have $T(z)=a\dfrac{z-z_0}{z-z_\infty}$, where we interpret this equation as saying $T(z)=a(z-z_0)$ if $z_\infty=\infty$ and $T(z)=\dfrac{a}{z-z_\infty}$ if $z_0=\infty$.
  - Then the constant a is uniquely characterized by the condition $T(z_1)=1$, and so T exists and is unique, as claimed.
  - For the second part, suppose $S(z_0)=w_0$, $S(z_1)=w_1$, and $S(z_\infty)=w_\infty$. Then by the above there exists a unique T_w with $T_w(w_0)=0$, $T_w(w_1)=1$, $T_w(w_\infty)=\infty$. Then $T_w\circ S$ maps z_0 to 0, z_1 to 1, and z_∞ to ∞.
  - Since (again by the above) there is a unique T_z with $T_z(z_0)=0$, $T_z(z_1)=1$, $T_z(z_\infty)=\infty$, we have $T_w\circ S=T_z$ and so S must equal $T_w^{-1}\circ T_z$. But since $T_w^{-1}\circ T_z$ has the desired property, we see S exists and is unique.
  - <u>Remark</u>: One may check more explicitly that the unique fractional linear transformation with $T(z_0)=0$, $T(z_1)=1$, and $T(z_\infty)=\infty$ is given by the <u>cross ratio</u> $T(z)=\dfrac{z-z_0}{z-z_\infty}\dfrac{z_1-z_\infty}{z_1-z_0}$.

5.2.3 Conformal Maps and Analytic Isomorphisms

- Now that we have studied fractional linear transformations, which allow us to construct maps among regions bounded by lines and circles, we investigate how to use other holomorphic functions to solve conformal mapping problems.
  - If U is an open region, the classical terminology is to say that $f:U\to\mathbb{C}$ is a conformal map when f preserves angles.
  - As we showed during our original discussion of holomorphic functions, f preserves angles if and only if f is holomorphic on U and f' is everywhere nonzero on U.

- Per our results on local invertibility, the condition that f' is nonzero on U is equivalent to saying that f is locally one-to-one on U.
- If f is in fact globally one-to-one on U, then if we define $V = f(U)$, which is also an open set by the open mapping theorem, then the inverse function $f^{-1} : V \to U$ is also holomorphic, and f is a homeomorphism from U to V.
- In this situation we say that f is biholomorphic (holomorphic with holomorphic inverse), or equivalently that f is an analytic isomorphism between U and V.
- We remark that local invertibility (namely, that f' is nonvanishing) is not equivalent to global invertibility: for example, the exponential function $f(z) = e^z$ is locally invertible but also periodic hence not globally invertible on $\mathbb{C}$ (or indeed on any set with two points that differ by an integer multiple of $2\pi i$).
- Thus, analytic isomorphisms always preserve angles, but angle-preserving maps need not be analytic isomorphisms.

- As a practical matter, we are interested specifically in analytic isomorphisms (which we will refer to as "conformal maps", with the tacit insistence that these actually be bijections) since they have better properties.
  - For example, the composition $f \circ g$ of any two analytic isomorphisms $f : V \to W$ and $g : U \to V$ is also an analytic isomorphism (its inverse is $g^{-1} \circ f^{-1}$), as is the inverse $f^{-1} : V \to U$.
  - Thus in particular, for any open set U, the set of analytic isomorphisms $f : U \to U$ forms a group under composition (such maps are analytic automorphisms of U).
  - We say that two open sets U and V are conformally equivalent if there exists an analytic isomorphism $f : U \to V$. The observations above (along with the trivial observation that the identity function is an isomorphism of any open set with itself) show that conformal equivalence is indeed an equivalence relation.
  - The main utility of conformal equivalence is that it allows us to transfer questions involving holomorphic functions between U and V by composing appropriately with f or f^{-1}.
  - For example, we can see that if $f : U \to V$ and $g : U \to V$ are both analytic isomorphisms, then there exists a unique analytic automorphism $h : V \to V$ such that $g = h \circ f$: namely, $h = g \circ f^{-1}$.
  - Likewise, if $f : U \to V$ is an analytic isomorphism, then there is a bijection between analytic automorphisms of U and analytic automorphisms of V given by $\varphi \mapsto f \circ \varphi \circ f^{-1}$ (i.e., via conjugation by f).
  - Thus, if we want to understand the behavior of analytic isomorphisms involving a given set U, it suffices to construct an analytic isomorphism of U with a fixed set V, and then characterize the analytic automorphisms of V.

- A particular general case is the situation where U is simply connected. There turn out to be two distinct cases:

- Proposition (Analytic Isomorphisms of $\mathbb{C}$): Suppose $f : \mathbb{C} \to V$ is an analytic isomorphism. Then $V = \mathbb{C}$ and $f(z) = az + b$ is a nonconstant linear polynomial; conversely all such polynomials are analytic isomorphisms.
  - Proof: Suppose $f : \mathbb{C} \to V$ is a holomorphic bijection. Then f and f^{-1} map boundaries to boundaries, so since $\mathbb{C}$ has no boundary points, neither does V.
  - But the only nonempty region in $\mathbb{C}$ with no boundary points is $\mathbb{C}$ itself, so $V = \mathbb{C}$.
  - Since f is nonconstant, by Liouville's theorem f must be unbounded as $z \to \infty$ so f cannot be holomorphic at ∞, nor can it have an essential singularity by Picard's theorem (or Casorati-Weierstrass) since it would then fail to be one-to-one.
  - Thus f has a pole at ∞ so $f(\infty) = \infty$. This means f extends to be an analytic automorphism of the extended complex plane, hence by our previous characterization of such maps, f is a fractional linear transformation. Since the pole of f is at ∞ that means $f(z) = az + b$ for some nonzero a, as desired.
  - Conversely, it is obvious that $f(z) = az + b$ is an analytic automorphism of $\mathbb{C}$, since f has an inverse $f^{-1}(z) = (z - b)/a$.

- In the other situation, U is simply connected and a proper subset of $\mathbb{C}$. In this case, it turns out that there always exists an analytic isomorphism of U with the unit disc:

- Theorem (Riemann Mapping Theorem): Suppose U is a simply connected open region that is not the entire complex plane $\mathbb{C}$. Then there exists an analytic isomorphism of U with the open unit disc $D = \{z : |z| < 1\}$. More precisely, given $z_0 \in U$ there exists a unique analytic isomorphism $f : U \to D$ such that $f(z_0) = 0$ and $f'(z_0) > 0$.
  - As an immediate consequence, the Riemann mapping theorem implies that any two simply connected open regions U, V (excluding $\mathbb{C}$) are analytically isomorphic to one another. Furthermore, if we can find analytic isomorphisms $f : U \to D$ and $g : V \to D$, then $g^{-1} \circ f$ will give the desired analytic isomorphism.
  - Therefore, it suffices to construct analytic isomorphisms mapping regions onto the open unit disc.
  - We will not really use the result of this theorem except as motivation (namely, why we now just focus on writing down maps from a region to the open unit disc), so for brevity we will omit some of the technical details of the proof of this theorem.
  - Proof (Outline): For a given simply connected $U \neq \mathbb{C}$, consider the family $\mathcal{F}$ of all injective holomorphic functions $f : U \to D$ such that $f(z_0) = 0$.
  - This family is nonempty: because $U \neq \mathbb{C}$, U omits some point α, and then because U is simply connected, there exists a branch $g(z)$ of the complex logarithm $\log(z - \alpha)$ that is holomorphic and (thus) one-to-one on U.
  - Then for any $z_0 \in U$, $g(z)$ must be bounded away from $g(z_0) + 2\pi i$ (otherwise by taking an appropriate limit, there would exist points with $g(z) = g(\tilde{z}) + 2\pi i$), and so $[g(z) - g(z_0) - 2\pi i]^{-1}$ is bounded. By translation and rescaling one obtains a function $f(z) = a[g(z) - g(z_0) - 2\pi i]^{-1} + b$ such that $f(U) \subseteq D$ and $f(z_0) = 0$.
  - Now observe that for any $f \in \mathcal{F}$, the values $|f'(z_0)|$ are uniformly bounded: take γ to be the counterclockwise boundary of U and apply Cauchy's integral formula $f'(z_0) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{(z - z_0)^2}\,dz$. Since $|f(z)| \leq 1$ for all $z \in \gamma$ since the image of f lies in the unit disc, we see $|f'(z_0)|$ is uniformly bounded above (namely by $\frac{1}{2\pi}$ times the arclength of γ divided by the square of the distance from z_0 to γ).
  - The next claim is that there exists a function $f \in \mathcal{F}$ maximizing $|f'(z_0)|$. If M is the least upper bound of the values $|f'(z_0)|$ for $f \in \mathcal{F}$, then either there is a function with $|f'(z_0)| = M$ or there is a sequence of f_i with $|f_i'(z_0)| \to M$. In the latter case one can show that this sequence has a subsequence that converges uniformly on compact subsets of U, so by uniform convergence the limit function f has $|f'(z_0)| = M$. This function is also necessarily injective by a Rouché's theorem argument, and since $|f(z)| \leq 1$ for all $z \in U$, by the maximum modulus principle in fact $|f(z)| < 1$ for all $z \in U$, so this $f \in \mathcal{F}$.
  - Finally, we show that any injective holomorphic $f : U \to D$ maximizing $|f'(z_0)|$ is necessarily onto: if the image of f does not contain some point z_1, then by moving z_0 to 0 and z_1 to 0 one obtains an injective holomorphic $g : U \to D$ with $|g'(0)| = M$ and $0 \notin \mathrm{im}(g)$. Then since the image V of g is simply connected, there exists a branch of the complex logarithm holomorphic on $\mathrm{im}(g)$, and hence by extension there exists an $h : U \to D$ with $h(z)^2 = g(z)$ (namely, $h(z) = e^{[\log g(z)]/2}$). Then one immediately has $h \in \mathcal{F}$ and also $M = |g'(0)| = 2\,|h(0)|\,|h'(0)|$ so that $|h'(0)| > |g'(0)| = |f'(z_0)|$, which is a contradiction.

- Now we can give some simple examples of analytic isomorphisms of various different regions.

- Example: The function $f(z) = \frac{z - i}{z + i}$ is an analytic isomorphism from the upper half-plane $\mathrm{Re}(z) > 0$ to the unit disc D. Its inverse $f^{-1}(z) = -i\frac{z + 1}{z - 1}$ maps the unit disc to the upper half-plane.
  - Since f is a fractional linear transformation it suffices to observe that $f(z)$ maps the real axis to the unit circle, which in turn follows by noting that for t real we have $\left|\frac{t - i}{t + i}\right| = 1$, and that $f(i) = 0$ lies inside the image of f.
  - This map is often called the Cayley transform.

- Example: The function $f(z) = z^2$ is an analytic isomorphism from the first quadrant $\text{Re}(z) > 0$, $\text{Im}(z) > 0$ to the upper half-plane.
  - Thus, composing this map with the inverse of the Cayley transform yields an analytic isomorphism $z \mapsto -i\dfrac{z^2+1}{z^2-1}$ from the first quadrant to the unit disc.
  - Additionally, the inverse map $f^{-1}(z) = z^{1/2} = e^{[\text{Log} z]/2}$ (which is holomorphic on the upper half-plane since it is only badly behaved at $z = 0$) is an analytic isomorphism from the upper half-plane to the first quadrant.

- Example: The function $f(z) = \dfrac{z+1}{1-z}$ is an analytic isomorphism of the upper half-disc $|z| < 1$, $\text{Re}(z) > 0$ with the first quadrant.
  - To see this we simply calculate the action of f on the real axis (mapped to the real axis) and the circle $|z| = 1$ (mapped to the imaginary axis), and observe that $f(i/2) = \dfrac{3+4i}{5}$ lies in the first quadrant.
  - As a consequence, composing this map with the ones above yields an analytic isomorphism $z \mapsto \dfrac{z^2+1}{1-z^2}$ from the upper half-disc to the upper half-plane, and an analytic isomorphism $z \mapsto -\dfrac{i}{2}(z + \dfrac{1}{z})$ with the unit disc.

- Example: The logarithm $f(z) = \text{Log}(z)$ is an analytic isomorphism of the upper half-plane with the strip $0 < \text{Im}(z) < \pi$.
  - This follows immediately from our description of $\text{Log}(z) = r + i\theta$ for $z = re^{i\theta}$ with $r > 0$ and $0 < \theta < \pi$.
  - The logarithm also gives an analytic isomorphism of the full plane with $[0, \infty)$ removed with the strip $0 < \text{Im}(z) < 2\pi$.

5.2.4 Functions on the Unit Disc, Harmonic Functions

- We can also classify the analytic automorphisms of the unit disc. To do this we will require a few related preliminary results about injective maps on the unit disc, all of which are generally referred to as Schwarz's lemma:

- Proposition (Schwarz's Lemma): Let D be the open unit disc $|z| < 1$ and let $f : D \to D$ be one-to-one and such that $f(0) = 0$.
  1. We have $|f(z)| \le |z|$ for all $z \in D$.
     - Proof: Since $f(0) = 0$, the power series expansion for f is of the form $f(z) = a_1 z + \sum_{n=2}^{\infty} a_n z^n$.
     - Then $g(z) = f(z)/z = a_1 + \sum_{n=1}^{\infty} a_{n+1} z^n$ is holomorphic on D after removing the removable singularity at 0, and also since $|f(z)| < 1$ by hypothesis we see that $|g(z)| < 1/|z|$ for all $|z| < 1$.
     - In particular, for $|z| < r$ we have $|g(z)| < 1/r$. Hence by the maximum modulus principle, since the maximum modulus of g on $|z| \le r$ must occur on the boundary and g is continuous, we see that $|g(z)| \le 1/r$ for $|z| \le r$.
     - Now taking $r \to 1$ from below yields $|g(z)| \le 1$ for $|z| < 1$, whence $|f(z)| \le |z|$ as desired.
  2. If $|f(z_0)| = |z_0|$ for any nonzero $z_0 \in D$, then f is a rotation: $f(z) = e^{i\theta} z$ for some θ.
     - Proof: If we have $|f(z_0)/z_0| = 1$ then we have the equality case for the maximum modulus principle, which only occurs when the function is constant. Thus $f(z)/z$ is constant, hence must equal $f(z_0)/z_0 = e^{i\theta}$ for some θ, so $f(z) = e^{i\theta} z$ as claimed.
  3. We have $|f'(0)| \le 1$ with equality if and only if f is a rotation.
     - Proof: As in (1) observe that $g(z) = f(z)/z$ is holomorphic on D after removing the removable singularity, and that $f'(0) = a_1 = g(0)$.
     - By (1) we have $|g(z)| \le 1$ for all $z \ne 0$, so taking $z \to 0$ yields $|g(0)| \le 1$.

- Now consider the case where $|f'(0)| = 1$. If g is nonconstant then g is an open mapping, so in particular the image of g near 0 would contain an open disc around $g(0)$. But since $|g(0)| = |f'(0)| = 1$ the image of g would include values of absolute value exceeding 1, which contradicts (1).
- Hence g must be constant, in which case f is a rotation just as in (2).

- Now we can classify the analytic automorphisms of the unit disc. As in the situation of the extended complex plane, they all turn out to be fractional linear transformations:

- Proposition (Analytic Automorphisms of the Disc): Let D be the unit disc $|z| < 1$ and suppose that $f : D \to D$ is an analytic automorphism (i.e., that f is holomorphic and a bijection). If $f(\alpha) = 0$ then there exists an angle θ such that $f(z) = e^{i\theta}\dfrac{\alpha - z}{1 - \overline{\alpha}z}$; conversely all such maps with $|\alpha| < 1$ and θ real are analytic automorphisms of the unit disc.
  - The analytic automorphisms of the disc are often called Möbius transformations
  - Proof: First we observe that all maps of the given form are analytic automorphisms of the unit disc. Since these maps are fractional linear transformations it suffices to observe that they map the unit circle to itself, and that they map an interior point to another interior point.
  - If $|z| = 1$ then $|f(z)| = \left|e^{i\theta}\right| \dfrac{|z-\alpha|}{|1-\overline{\alpha}z|} = \dfrac{1}{|\overline{z}|}\dfrac{|1-\alpha\overline{z}|}{|1-\overline{\alpha}z|} = 1$ since $|\overline{z}| = 1$ and $|1 - \alpha\overline{z}| = |1 - \overline{\alpha}z|$ since they are complex conjugates, so f maps the unit circle to itself. Furthermore, since $f(\alpha) = 0$, f maps the interior point α to another interior point 0, as required.
  - Now suppose that f is an analytic automorphism of the disc and that $f(\alpha) = 0$. Take $g(z) = \dfrac{\alpha - z}{1 - \overline{\alpha}z}$. Then g is an analytic automorphism of the unit disc so $h = f \circ g^{-1}$ is as well, and h maps 0 to 0.
  - Since h is an injective map from the disc to itself, by the Schwarz lemma we have $|h(z)| \le |z|$ for all z in the disc. But h^{-1} is also an injective map from the disc to itself, so $\left|h^{-1}(z)\right| \le |z|$ for all z in the disc, which upon applying h implies $|z| \le |h(z)|$ for all z in the disc.
  - Hence we have equality: $|h(z)| = |z|$ for all z, so h must be a rotation: $h(z) = e^{i\theta}z$ for some θ. Then $f(z) = (h \circ g)(z) = e^{i\theta}\dfrac{\alpha - z}{1 - \overline{\alpha}z}$, as claimed.

- One of the main utilities of being able to calculate all of these analytic isomorphisms is that they allow us to transfer solutions to certain types of equations from one region to another. The connection comes via harmonic functions:

- Definition: If $u(x, y)$ is a real-valued function of two variables (equivalently, a real-valued function of $z = x+iy$), we say u is harmonic when $u_{xx} + u_{yy} = 0$. Equivalently, in terms of the Laplacian operator $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = \dfrac{1}{4}\frac{\partial^2}{\partial z \partial \overline{z}}$, u is harmonic when $\Delta u = 0$.
  - From the Cauchy-Riemann equations, we can see immediately that if $f = u(x, y)+iv(x, y)$ is holomorphic, then u and v are both harmonic.
  - Inversely, if $u(x, y)$ is harmonic on a simply connected region U, then in fact u is the real part of a holomorphic function. Explicitly, if we define $h = u_x - iu_y$, then h is holomorphic since it satisfies the Cauchy-Riemann equations, and so since U is simply connected it has an antiderivative $f = p+iq$. Then $f' = p_x - ip_y$ so $p_x = u_x$ and $p_y = u_y$, meaning that p and u have the same partial derivatives hence are equal up to a constant. By subtracting the constant we obtain a holomorphic function f whose real part is u, as desired.
  - For a given $u = \mathrm{Re}(f)$, the imaginary part $v = \mathrm{Im}(f)$ is determined up to an additive constant, and is called a harmonic conjugate of u.
  - If u is a harmonic function on V and we have an analytic isomorphism $f : U \to V$ for some other region V, then the composition $g = u \circ f$ is also a harmonic function on U: explicitly, one may compute using the chain rule that $\Delta g = (\Delta u)\,|f|^2 = 0$.
  - Therefore, composition with an analytic isomorphism $f : U \to V$ allows us to transfer between harmonic functions on U and harmonic functions on V.

- As a consequence, if we wish to construct a particular harmonic function (e.g., one satisfying a particular boundary-value problem) on U, we may equivalently solve that problem on V and then transfer it to U.
  - One wide class of such functions arises from studying thermal flow: if $u(x, y)$ represents temperature, then in a steady state temperature distribution in the plane (i.e., constant over time), the heat equation $\Delta u = u_t$ is equivalent to requiring that u be harmonic.
  - Boundary-value problems naturally arise in this context in the guise of arranging heating elements around the boundary of a region (whose temperatures are assumed known and constant): the goal is then to determine the temperature distribution inside the region.
  - Another class of such functions arises from studying standing 2-dimensional standing waves: if $u(x, y)$ represents wave intensity, then in a standing wave pattern (constant over time), the wave equation $\Delta u = u_{tt}$ requires that u be harmonic. (Indeed, the connection with standing wave patterns is the reason for the name "harmonic function".)
  - In either situation, if the region is simply connected and bounded, then by the Riemann mapping theorem, it is analytically isomorphic to the open unit disc. Therefore, if we can solve the corresponding boundary-value problem for the unit disc, we can transfer back to obtain the solution on the original region.
  - But since harmonic functions are simply the real parts of holomorphic functions, we are equivalently seeking to reconstruct a holomorphic function f given its values on the boundary of the disc, and this is precisely done by Cauchy's integral formula. We need only rewrite it in a way that only involves the real part of f.

- Theorem (Poisson's Formula): Suppose u is harmonic on the unit disc. Then for any $r < 1$ we have $u(re^{i\theta}) = \dfrac{1-r^2}{2\pi} \displaystyle\int_0^{2\pi} \dfrac{u(e^{it})}{1 + r^2 - 2r\cos(t-\theta)}\, dt$.
  - Proof: Suppose $u = \mathrm{Re}(f)$ where f is holomorphic. If γ is the boundary of the unit disc, the Cauchy integral formula says that $f(z_0) = \dfrac{1}{2\pi i}\displaystyle\int_\gamma \dfrac{f(z)}{z - z_0}\, dz$.
  - By Cauchy's formula again, we also have $\dfrac{1}{2\pi i}\displaystyle\int_\gamma \dfrac{f(z)}{z - 1/\overline{z_0}}\, dz = 0$ since $1/\overline{z_0}$ is not inside the disc.
  - Subtracting yields $f(z_0) = \dfrac{1}{2\pi i}\displaystyle\int_\gamma f(z)\left[\dfrac{1}{z - z_0} - \dfrac{1}{z - 1/\overline{z_0}}\right] dz = \dfrac{1}{2\pi i}\displaystyle\int_\gamma \dfrac{1 - z_0\overline{z_0}}{(z - z_0)(1 - z\overline{z_0})} f(z)\, dz$.
  - Parametrizing via $\gamma(t) = e^{it}$ for $0 \le t \le 2\pi$ yields $f(z_0) = \dfrac{1}{2\pi i}\displaystyle\int_\gamma \dfrac{1 - z_0\overline{z_0}}{(e^{it} - z_0)(1 - e^{it}\overline{z_0})} f(e^{it}) i e^{it}\, dt = \dfrac{1}{2\pi}\displaystyle\int_\gamma \dfrac{1 - |z_0|^2}{|e^{it} - z_0|^2} f(e^{it})\, dt$ and now since the integrand is $f(e^{it})$ times a real number, extracting real parts yields $u(z_0) = \dfrac{1}{2\pi}\displaystyle\int_\gamma \dfrac{1 - |z_0|^2}{|e^{it} - z_0|^2} u(e^{it})\, dt$. Rearranging yields the desired formula.

Well, you're at the end of my handout. Hope it was helpful.