Math 3228 - Week 8

- Complex maps
- Fractional linear transformations
- Conformality and orientation preservation
- The Gamma function

* * * * *

Complex maps

- Let $f : \mathbf{C} \to \mathbf{C}$ be an analytic map, such as $f(z) = z^2$ or $f(z) = \exp(z)$ or $f(z) = i+iz$. Until now, we have mostly been thinking of these maps in algebraic terms, but now we will think of them more geometrically, as mapping the complex plane $\mathbf{C}$ to another copy of the complex plane $\mathbf{C}$. In particular we want to know, given various shapes $A \subset C$, what happens to the image $f(A) := \{f(z) : z \in A\}$ of this shape under the transformation f. We have already seen a little bit of this in the argument principle, when we began caring about the image $f(\gamma)$ of a closed curve γ.

- In real analysis, we can represent a function $f : \mathbf{R} \to \mathbf{R}$ visually as a graph, sketching the set of points $\{(x, f(x)) : x \in \mathbf{R}\}$ in $\mathbf{R}^2$. This is not so easy to accomplish for complex maps, because the analogue of a graph $\{(z, f(z)) : z \in \mathbf{C}\}$ would have to live in $\mathbf{C}^2$, which is four dimensional and thus not a particularly helpful visual aid. However, we can still get some visual understanding of a complex map by drawing the domain $\mathbf{C}$ and range $\mathbf{C}$ of the map separately (as opposed to being orthogonal axes of a four-dimensional space), and sketching some points in the domain and their corresponding image in the range.

- For example, take the function $f(z) = 2z$. This takes a complex number z in the domain as input, and returns another number $2z$ with the same argument but twice the magnitude; thus it dilates z by a factor of two about the origin. If A is any set in $\mathbf{C}$, then $f(A)$ will be twice as large as A (in both the horizontal and vertical directions) but will otherwise have the same shape and orientation. More generally, the map $f(z) = cz$ for any $c > 0$ is a dilation around the origin by c (though if $c < 1$ then this map is actually a contraction!). Note also the extreme case $c = 0$, which is a degenerate map that sends everything to a single point (the origin).

- Now consider the map $f(z) = -z$. This takes a number z to the number with the same magnitude but opposite phase; more generally, it takes any shape A in the complex plane $\mathbf{C}$ and reflects it through the origin (or what amounts to the same thing, it rotates A by 180° either clockwise or anticlockwise around the origin). One can compose this map with one of the previous dilation maps, e.g. $f(z) = 2z$, to form another map $f(z) = -2z$; this rotates by 180° and then dilates by 2.

- Now consider the map $f(z) = e^{i\theta}z$ for some angle θ. This takes a number z and increases its argument by θ, while keeping its magnitude fixed. In other words, f is a anti-clockwise rotation around the origin by θ (though if θ is negative, one can also think of this as a clockwise rotation around the origin by $-\theta$). Again, one can combine this with dilations, thus the map $f(z) = re^{i\theta}z$ rotates by θ and then dilates by r.

- Notice that all of the above maps preserve the location of the origin. Now consider a map $f(z) = z + z_0$ for some fixed complex number $z_0 = x_0 + iy_0$. This map takes a complex number z and shifts the real part to the right by x_0 and the imaginary part up by y_0; thus the map is a translation to the right by x_0 and up by y_0 (using the convention, of course, that a rightward translation by a negative number is the same as a leftward translation by a positive number, etc.) One can also compose this translation operation with dilations and rotations, but now one has to be careful about the order in which one does so, because the operations of dilation and translation don't commute. For instance, if one composes the translation map $f(z) := z + i$ with the dilation map $g(z) := 2z$, one obtains the map $g(f(z)) = 2(z + i) = 2z + 2i$, but if one does the dilation first and the translation second, one gets $f(g(z)) = 2z + i$ instead.

- Using translations, one can now describe dilations and rotations about any fixed reference point (not just the origin). Suppose, for instance, that one wanted to create a map $z \mapsto f(z)$ which rotated the complex plane by $90°$ anti-clockwise around $2i$; thus $2i$ stays fixed, while $2i + 1$ gets mapped to $3i$, etc. The way to do this is to start with the variable z, move it down by $2i$, $w := z - 2i$, and then rotate the new variable anti-clockwise around the *origin* by $90°$, $\zeta := iw$, and then move it back upwards by $2i$, $\omega := \zeta + 2i$. The final map is then

$$f(z) := \omega = \zeta + 2i = iw + 2i = i(z - 2i) + 2i = iz + 2i + 2.$$

- So far the mappings we have considered (dilations, rotations, translations, and combinations of the three) are *shape-preserving*; they may rotate, shift, enlarge or shrink an object but they don't affect the actual shape of the object - a square remains a square, a circle remains a circle, and so forth. In particular, they are also *angle-preserving* - two lines intersecting at an angle of θ remain intersecting at angle θ after any of the above transformations. They are also *orientation-preserving* - a curve traversed in the anti-clockwise direction will remain anti-clockwise (this is in contrast to *reflections* such as $f(z) := \overline{z}$, which are shape-preserving and angle-preserving but orientation-reversing). Now we will look at a different type of analytic map - one which is not shape preserving, although we shall see later that it is still angle-preserving and (mostly) orientation preserving. This map is the *inversion map* $f(z) := 1/z$.

- Strictly speaking, this map is not defined on all of $\mathbf{C}$; instead, one has to work in the extended complex plane $\mathbf{C} \cup \{\infty\}$ (also called the Riemann sphere), and adopt the conventions that $\frac{1}{0} = \infty$ and $\frac{1}{\infty} = 0$. (Note that the dilations, translations, and rotations can also be extended to the Riemann sphere, by adopting the convention that any dilate, translate, or rotation of ∞ still ends up at ∞.

- In polar co-ordinates, the inversion map $f(z) = 1/z$ takes the number $z = re^{i\theta}$ to the number $\frac{1}{z} = \frac{1}{r}e^{-i\theta}$. Thus, the inversion map replaces the magnitude of z with its reciprocal (so magnitudes greater than 1 become magnitudes less than 1, and vice versa) while the phase gets reflected across the x-axis (so points above the x-axis get reflected to points below the x-axis, and vice versa. Geometrically, one can think of inversion as reflection across the unit circle $\{z \in \mathbf{C} : |z| = 1\}$, followed by another reflection across the x-axis. (Inversion turns the complex plane inside-out and upside-down).

- Qualitatively: if you move z closer to the origin, then w moves further away, and vice versa; if you move z clockwise around the origin, then w moves anti-clockwise, and vice versa.

- Let us now see what the inversion map does to various geometric objects such as lines and circles. Let us first consider a line through the origin, e.g. $\{re^{i\theta} : r \in \mathbf{R}\}$ for some fixed angle θ (note that we allow r to be negative as well as positive, otherwise we would have a ray rather than a line). This is the line which makes an anti-clockwise angle of θ with the x-axis. Applying the inversion map would convert this line into $\{\frac{1}{r}e^{-i\theta} : r \in \mathbf{R}\}$, which is the line which makes a *clockwise* angle of θ with the x-axis (note that $1/r$ is just an arbitrary real number, since r was already arbitrary).

- Now let us consider what inversion does to a line that does not go through the origin. For instance, let us consider the line $\{z : \mathrm{Re}(z) = 1\}$. The image of this line under inversion is then $\{\frac{1}{z} : \mathrm{Re}(z) = 1\}$. Making the substitution $w := 1/z$ (so that $z = 1/w$), this is the same as $\{w : \mathrm{Re}(\frac{1}{w}) = 1\}$. To understand this set, we write w in Cartesian co-ordinates $w = x + iy$, so that

$$\frac{1}{w} = \frac{1}{x+iy} = \frac{x-iy}{(x+iy)(x-iy)} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}.$$

In particular, the real part of $\frac{1}{w}$ is $\frac{x}{x^2+y^2}$. Thus the image of the line $\{z : \mathrm{Re}(z) = 1\}$ under inversion is

$$\{x + iy : \frac{x}{x^2+y^2} = 1\}.$$

We can rewrite the equation $\frac{x}{x^2+y^2} = 1$ as $x^2 + y^2 - x = 0$, or by completing the square,

$$(x - \frac{1}{2})^2 + y^2 = \frac{1}{4}.$$

In other words, the image of the line $\{z : \mathrm{Re}(z) = 1\}$ under inversion is the circle centered at $\frac{1}{2} + 0i$ with radius $1/2$. Note that this circle goes through the origin; this is because the line goes to infinity, and inversion maps infinity to 0 and vice versa.

- More generally, one can show that the image of any line not going through the origin under inversion becomes a circle that does go through the origin. One could in principle write down a formula for this circle depending on the line, but this formula is somewhat messy to remember, and it is usually easier instead to do things by hand using algebraic considerations, or else geometric considerations. One fairly quick way to proceed is to use the fact that the closest point of the line to the origin will get mapped to the furthest point of the circle from the origin, since inversion replaces the magnitude of a number with its reciprocal. For instance, suppose we wish to invert the line $\{x + iy : x + y = 1\}$. The closest this line gets to the origin is at $\frac{1}{2} + \frac{1}{2}i$. Inversion maps this point to $1 - i$. Thus the image of the original line is a circle which goes through the origin and whose furthest point from the origin is $1 - i$; in other words, it is the circle with diameter connecting 0 with $1 - i$, and thus has center $\frac{1}{2} - \frac{1}{2}i$ and radius $\frac{\sqrt{2}}{2}$.

- Note that the operation of inversion is its own inverse: if $w = 1/z$, then $z = 1/w$. Thus if lines not going through the origin map to circles going through the origin, it is not surprising that conversely circles going through the origin map to lines not going through the origin. Again, to compute exactly which line corresponds to which circle it is easiest to remember that the farthest point of the circle from the origin maps to the closest point of the line to the origin. Thus for instance the circle $\{z \in \mathbf{C} : |z - 2i| = 2\}$ goes through the origin, and its furthest point from the origin is $4i$, so the line that the inversion map takes the circle to will have $\frac{1}{4i} = \frac{-1}{4}i$ as its closest approach to the origin. Since a line is orthogonal to the vector connecting the origin to its closest point, the line must be horizontal, i.e. it is the line $\{z : \mathrm{Re}(z) = \frac{-1}{4}\}$.

- Finally, we consider what happens to circles that do not go through the origin. The easiest example of this are circles centered at the origin, such as $\{z \in \mathbf{C} : |z| = 2\}$. Since inversion replaces magnitudes with their reciprocal, it is clear that this circle inverts to the circle

$\{z \in \mathbf{C} : |z| = \frac{1}{2}\}$ (the reflection in phase is irrelevant since the circle encompasses all phases anyway).

- The last case we need to consider are circles that do not go through the origin, and are not centered at the origin. A typical such example is the circle $\{z : |z-2| = 1\}$. The image under inversion is $\{\frac{1}{z} : |z-2| = 1\}$, or under the change of variables $w := 1/z$, $\{w : |\frac{1}{w} - 2| = 1\}$. We could use Cartesian co-ordinates now, but it will make things easier later on if we simplify a bit first. Multiplying both sides by w, we can rewrite the equation $|\frac{1}{w} - 2| = 1$ as $|1-2w| = |w|$, which we can then square as $|1-2w|^2 = |w|^2$. If we write $w = x+iy$, then $1-2w = (1-2x) - 2iy$ and so $|1-2w|^2 = (1-2x)^2 + (-2y)^2 = 4x^2 + 4y^2 - 4x + 1$, while $|w|^2 = x^2 + y^2$. Thus the equation simplifies to $3x^2 + 3y^2 - 4x + 1$, which after completing the square becomes $3(x - \frac{2}{3})^2 + 3y^2 = \frac{1}{3}$, which (after dividing by 3) is the equation for the circle centered at $\frac{2}{3} + 0i$ with radius $\frac{1}{3}$. Thus, the inverse of a circle not going through the origin is another circle not going through the origin. This turns out to be true in general. Again, a relatively simple way to work out what the inverse of a circle is is to find the closest and furthest points from the origin, which will then map to the furthest and closest points from the origin. (Note that these two points always form a diameter of the circle in question - why?).

- Once one sees what the inversion map does to circles and lines, one can also see what it does to disks, half-planes, and exterior disks. Note that in some cases an inversion might map a disk to an exterior disk or vice versa; for instance the disk $\{z : |z| \leq 2\}$ clearly inverts to $\{z : |z| \geq \frac{1}{2}\}$. One easy way to work out which part of the complex plane a domain maps to is to first figure out where the boundary goes, and then take a test point inside the domain (e.g. the origin) to see which side of the boundary the domain ends up on. One can also work out how to invert line segments and arcs by a similar procedure.

- Viewed on the Riemann sphere, one can think of inversion as a $180°$ rotation around the axis connecting the antipodal points 1 and -1 on the sphere; thus ∞ is rotated down to 0, while the unit circle is flipped back onto itself.

- One can compose the inversion map with the dilation, rotation, and translation maps defined earlier, to generate a general family of transformations called *fractional linear transformations*, *projective linear transformations*, or *Möbius transformations* (named in honour of Augustus Möbius, of the Möbius strip fame). The general form of these transformations is $f(z) := \frac{az+b}{cz+d}$, where a, b, c, d are complex numbers and we impose the condition $ad - bc \neq 0$ to prevent the transformation from becoming degenerate (e.g. the map $f(z) = \frac{2z+3}{4z+6}$ is degenerate, as it collapses to the constant map $\frac{1}{2}$). Inversion, for instance, is the special case when $a = d = 1$ and $b = c = 0$, while translations and dilations come from the cases where $c = 0$ and $d = 1$. We observe that every fractional linear transformation can be "factorized" as a composition of the basic operations of translation, dilation, rotation, and inversion; for instance, the map $f(z) := \frac{z+i}{z-i}$ can be written as a vulgar fraction as $f(z) = 1 + \frac{2i}{z-i}$, so the map is a composition of a translation downward by $-i$ (that maps z to $z - i$), an inversion (which would map $z - i$ to $\frac{1}{z-i}$, a dilation by 2 and a rotation anti-clockwise by $90°$ (which would map $\frac{1}{z-i}$ to $\frac{2i}{z-i}$, followed by a translation to the right by 1. Thus, in principle, one can understand how this Möbius transform $f(z) = \frac{z+i}{z-i}$ transforms various geometric objects. For instance, to see what f does to the unit circle $\{z : |z| = 1\}$, we first shift downward by $-i$ (thus yielding the circle of radius 1 centered at $-i$), then invert (which gives the horizontal line $\{z : \mathrm{Im} z = \frac{1}{2}\}$ - why?), then multiply by $2i$ (which gives the line $\{z : \mathrm{Re}(z) = -1\}$ - why?) and then add 1 (to obtain the imaginary axis). This means that the unit disk $\{z : |z| < 1\}$ must map to either the left-half plane $\{z : \mathrm{Re} z < 0\}$ or the right half-plane $\{z : \mathrm{Re} z > 0\}$; to see which, we test the map at one point of the unit disk, e.g. the origin 0. Since $f(0) = -1$ is on the left half-plane, we see that the image of the unit disk must be the left-half-plane.

- Every Möbius transformation is invertible: if $w = \frac{az+b}{cz+d}$, then $wcz + wd = az + b$, thus $(cw - a)z = b - dw$, and hence $z = \frac{-dw+b}{cw-a}$. Note that the non-degeneracy condition $ad - bc = 0$ is unchanged. Also, the composition of any two Möbius transformations is still a Möbius transformation. (To see this, first observe that following a Möbius transformation $w = \frac{az+b}{cz+d}$ with an inversion $\zeta = \frac{1}{w}$ gives you another

Möbius transformation $\zeta = \frac{cz+d}{az+b}$. Similarly composing any Möbius transformation with a dilation, rotation or translation will also give another Möbius transformation. Since every Möbius transformation is itself a composition of dilations, rotations, translations, and inversions, the claim then follows). In other words, the set of all Möbius transformations form a group (with the identity element being of course the identity transformation $w = \frac{1z+0}{0z+1}$).

- Note that Möbius transforms always maps lines and circles to other lines and circles. Actually one can think of a line as like a circle of infinite radius. On the Riemann sphere, both lines and circles become spherical circles; the only difference is that lines go through the point at infinity whereas ordinary circles do not.

* * * * *

Constructing Möbius transforms

- Möbius transforms are useful for mapping one type of domain to another, for instance turning an unbounded domain into a bounded domain. Often, though one has to undergo a trial and error process in order to turn a given domain into a domain one wants, starting with the original domain and gradually getting closer to the desired goal.

- For instance, suppose we want to find a Möbius transform which maps the upper half-plane $\{z : \mathrm{Im}(z) > 0\}$ to the unit disk $\{z : |z| < 1\}$, and also maps the point i to the origin 0.

- At first glance none of the basic operations of translation, rotation, dilation, or inversion seem to make the half-plane look anything like the disk (note that if one inverts the upper half-plane one only gets the lower half-plane (why?)). The problem with inversion is that the boundary of the upper half-plane, i.e. the real axis, is going through the origin. So the solution is to move the boundary away from the origin first and *then* invert. For instance, if we start by translating the upper half-plane upward by i, to the half-plane $\{z : \mathrm{Im}(z) > 1\}$, and then invert, one obtains the disk centered at $-i/2$ with radius $1/2$, i.e. $\{z : |z + \frac{i}{2}| < \frac{1}{2}\}$ (why is this the case?). Meanwhile, the point i has shifted up to $2i$ and then invertedto $-i/2$, i.e. it is now in the center of the disk. If one then shifts this up by $\frac{i}{2}$, and then dilates by 2, one obtains the unit disk as desired, and i has mapped to the origin. The Möbius transformation that we have constructed this way can be written out explicitly by executing all of the above steps in the correct order:

$$z \mapsto z + i \mapsto \frac{1}{z+i} \mapsto \frac{1}{z+i} + \frac{i}{2} \mapsto 2(\frac{1}{z+i} + \frac{i}{2}),$$

which simplifies to the map $f(z) = \frac{iz+1}{z+i}$. (In this particular problem, it turns out there are multiple solutions; the map $f(z) = -\frac{iz+1}{z+i}$ will also work (why?)).

- Given any four complex numbers z_1, z_2, z_3, z_4, form the *cross ratio* $[z_1, z_2, z_3, z_4]$ by the formula

$$[z_1, z_2, z_3, z_4] = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_3)(z_2 - z_4)}.$$

The importance of this cross ratio is that it is preserved by Möbius transformations, i.e.

$$[z_1, z_2, z_3, z_4] = [f(z_1), f(z_2), f(z_3), f(z_4)]$$

for all Möbius transformations f. To see this, first observe that the claim is trivial if f is a translation $f(z) = z + z_0$ (since all the z_0 factors will cancel). Similarly if f is a dilation $f(z) = cz$ (since all the c factors will cancel) or a rotation $f(z) = e^{i\theta} z$ (same reason). For inversions $f(z) = 1/z$ the claim is a little trickier to see, but if one uses the identity

$$f(z_1) - f(z_2) = \frac{1}{z_1} - \frac{1}{z_2} = -\frac{z_1 - z_2}{z_1 z_2}$$

and then multiplies everything out one can also verify that inversion doesn't affect the cross ratio either. Since every Möbius transform is a combination of dilations, translations, rotations, and inversions, we thus see that none of the Möbius transforms affect the cross-ratio.

- One consequence of this formula is that Möbius transforms cannot map any four given points to any other four given points, unless their cross-ratios are equal; thus for instance there is no Möbius transform that maps $0,1,2,3$ to $0,1,2,4$. However, one can use this cross-ratio formula to find a Möbius transform f that maps any *three* given points z_1, z_2, z_3 to any other three given points w_1, w_2, w_3 (i.e. $f(z_1) = w_1$, $f(z_2) = w_2$, $f(z_3) = w_3$). Indeed, any such transformation must now obey the identity

$$[z_1, z_2, z_3, z] = [f(z_1), f(z_2), f(z_3), f(z_4)] = [w_1, w_2, w_3, f(z)]$$

and one can use this identity to solve for $f(z)$ in terms of z. For instance, if one wants to map the points $-1, 0, 1$ to $-1, i, 1$ respectively, we would solve the equation

$$[-1, 0, 1, z] = [-1, i, 1, f(z)]$$

i.e.

$$\frac{-(1-z)}{-2(-z)} = \frac{(-1-i)(1-f(z))}{-2(i-f(z))}$$

which after some algebra simplifies to $f(z) = \frac{z+i}{iz+1}$.

* * * * *

Conformality and orientation preservation

- Möbius transformations do not, in general, preserve the shape of objects. But a remarkable fact is that they are *conformal*, i.e. they preserve angles and orientations of objects. In fact this is a property not just of Möbius transformations, but of all analytic maps:

- **Theorem.** Let γ_1 and γ_2 be two differentiable paths starting at the same point z_0 (so $\gamma_1(0) = \gamma_2(0) = z_0$), which make an angle of θ at z_0 (i.e. the tangent vectors $\gamma_1'(0)$ and $\gamma_2'(0)$ make an angle of θ). Let f be any map which is analytic at z_0 with $f'(z_0)$ non-zero. Then $f(\gamma_1)$ and $f(\gamma_2)$ make an angle of θ at $f(z_0)$.

- **Proof.** By the chain rule, we see that

$$f(\gamma_1)'(0) = f'(\gamma_1(0))\gamma_1'(0) = f'(z_0)\gamma_1'(0)$$

and similarly

$$f(\gamma_2)'(0) = f'(\gamma_2(0))\gamma_2'(0) = f'(z_0)\gamma_2'(0).$$

Thus $f(\gamma_1)'(0)$ and $f(\gamma_2)'(0)$ are obtained from $\gamma_1'(0)$ and $\gamma_2'(0)$ by multiplication by a fixed non-zero complex number $f'(z_0)$. But this is just a dilation and rotation (by the magnitude and phase of $f'(z_0)$), and both of these operations clearly preserve the angle between $\gamma_1'(0)$ and $\gamma_2'(0)$. $\square$

- Thus for instance, two lines intersecting at right angles to each other will map under a Möbius transformation (or any other analytic map) to two curves that still intersect at right angles. The assumption that $f'(z_0)$ is non-zero is important; for instance, consider the map $f(z) = z^2$. The real and imaginary axes meet at right angles, but they map under f to the positive axis and negative axis respectively, which are $180°$ apart.

- Now we prove orientation preservation.

- **Theorem.** Let f be a function which is analytic at a point z_0 with $f'(z_0) \neq 0$. Then there exists an $\varepsilon > 0$ such that if γ is any simple closed curve in the ball $\{z \in \mathbf{C} : |z - z_0| < \varepsilon\}$ which goes around z_0 once anticlockwise, then $f(\gamma)$ also goes around $f(z_0)$ once anticlockwise.

- **Proof.** Consider the function $g(z) := f(z) - f(z_0)$. This function is analytic at z_0 and has a simple zero at z_0 (since $g(z_0) = 0$ and $g'(z_0) = f'(z_0) \neq 0$). Thus $g(z) = h(z)(z - z_0)$ for some function h which is analytic and non-zero near z_0. Since analytic functions are continuous, we know that $h(z)$ is in fact analytic and non-zero on some ball $\{z : |z - z_0| < \varepsilon\}$, and hence g is also analytic non-zero on this ball (except at z_0). In particular, if γ goes around z_0 in this ball once anticlockwise, then there is one zero of g inside γ and no poles. By the argument principle, this means that $g(\gamma)$ winds once anticlockwise around 0, which means that $f(\gamma)$ winds once anticlockwise around z_0 as claimed. $\square$

- Again, the assumption that $f'(z_0) = 0$ is important; if $f(z) = z^2$, then a small circle going once around the origin will map to a curve going twice around the origin. Also, the assumption that the curve is small is important; if one considers the inversion map $f(z) = 1/z$, then a small curve going anti-clockwise around 1 (say) will remain anticlockwise around 1 under inversion, but a very large curve going around anti-clockwise around both 1 and 0 will become clockwise and not go around 1 at all! (This can be seen from the argument principle).

- Note that if f is a Möbius transform, then $f'(z)$ is never zero (this is easiest to see by writing f as a vulgar fraction, e.g. $f(z) = \frac{az+b}{cz+d} = \frac{a}{c} + \frac{bc-ad}{c(cz+d)}$ and then differentiating to obtain $f'(z) = \frac{ad-bc}{(cz+d)^2}$, which cannot vanish since $ad - bc \neq 0$). Thus Möbius transforms always preserve angles, and also preserve the orientation of small curves (more precisely, they preserve the orientation of curves which don't go around the pole of the Möbius transform).

- Some Java applets illustrating this behaviour of analytic maps can be found at

$$\texttt{http://www.math.ucla.edu/}\sim\texttt{tao/3228}$$

- Conformal mappings are useful not only because they preserve angles and orientations, but they tend to preserve some other things as well (for instance, it turns out that they map harmonic functions to harmonic functions). One may then ask whether one can map any domain to any other. In the case of simply connected domains, this is true, and is a famous theorem (the proof of which is beyond the scope of this course, however):

- **Riemann mapping theorem.** Let D and D' be two simply connected domains. Then there is an invertible analytic map $f : D \to D'$ with f' non-vanishing (so f is always conformal).

- Thus, for instance, it is possible to map a solid square (or any other polygon) conformally to a disk. Such maps are known as *Schwarz-Christoffel transformations* and have many applications, for instance to fluid mechanics. The hypothesis of simply connectedness is important; one cannot map a disk to an annulus conformally, for instance.

* * * * *

The gamma function

- Up until now you have seen a number of useful analytic functions such as polynomials, Möbius transformations, exponential, trigonometric, and logarithm functions. But there are many other analytic functions out there, and one particularly useful one is the *Gamma function* $\Gamma(z)$. This function is defined for all complex numbers z (except for the negative integers $z = 0, -1, -2, \ldots$), but to begin with we will only define it for complex numbers with positive real part.

- **Definition.** If $\mathrm{Re}(z) > 0$, then we define $\Gamma(z) := \int_0^\infty t^{z-1}e^{-t}\, dt$.

- Let us first check that this integral is absolutely convergent. If $z = x + iy$, then $t^{z-1} = t^{x-1}t^{iy} = t^{x-1}e^{iy\ln t}$ has the magnitude of t^{x-1}, so it suffices to show that the integral $\int_0^\infty t^{x-1}e^{-t}$ is finite. We split this integral into $\int_0^1$ and $\int_1^\infty$. For $\int_0^1$ we observe that $t^{x-1}e^{-t} \leq t^{x-1}$, and since $x > 0$ we know that t^{x-1} is integrable on $[0, 1]$ (it has an antiderivative of t^x/x, which stays bounded on $[0, 1]$). For the integral $\int_1^\infty$ we use the fact that exponentials grow faster than polynomials, so for instance $t^{x+1}e^{-t}$ goes to zero as t goes to infinity. In particular the function $t^{x+1}e^{-t}$ must be bounded by some constant C, thus $t^{x-1}e^{-t}$ is bounded by C/t^2, which is integrable on $[1, \infty)$. So the Gamma function is well defined on the half-plane $\mathrm{Re}(z) > 0$.

- Next, we show that $\Gamma(z)$ is in fact analytic on this half-plane. By Morera's theorem (the converse to Cauchy's theorem; you should have seen it in an earlier week) it suffices to show that

$$\int_\gamma \Gamma(z)\, dz = 0$$

for all closed curves γ in the half-plane $\{z : \mathrm{Re}(z) > 0\}$. Observe that the real part of γ, i.e. $\{\mathrm{Re}(\gamma(t)) : t \in [0, 1]\}$ is a closed interval in the positive axis, say it is equal to $[\varepsilon, N]$. Then $\gamma \subset \{z : \varepsilon < \mathrm{Re}(z) < N\}$. Now we expand

$$\int_\gamma \Gamma(z)\, dz = \int_\gamma \int_0^\infty t^{z-1}e^{-t}\, dtdz.$$

Remember that t^{z-1} has magnitude t^{x-1}, which is bounded by $t^{\varepsilon-1} + t^{N-1}$. As observed above, $(t^{\varepsilon-1} + t^{N-1})e^{-t}$ is absolutely integrable on

$[0, \infty]$. Thus by Fubini's theorem (observing that γ has finite length and so does not affect the absolute integrability of the integrand) we can reverse the order of integration to obtain

$$\int_\gamma \Gamma(z)\ dz = \int_0^\infty \int_\gamma t^{z-1} e^{-t}\ dzdt.$$

But for $t > 0$, the function $t^{z-1} = e^{(z-1)\ln t)}$ is analytic in z, while e^{-t} does not depend on z. Thus by Cauchy's theorem the inner integral is zero, and thus the whole integral is zero. Thus $\Gamma(z)$ is indeed analytic. (One could also have proven this result by the Cauchy-Riemann equations and differentiation under the integral sign, but it is a little harder to justify differentiation under the integral sign than it is to swap integrals).

- Now we work out some values of $\Gamma(z)$. The easiest one to work out is $\Gamma(1)$:

  $$\Gamma(1) = \int_0^\infty e^{-t}\ dt = -e^{-t}|_0^\infty = -0 + 1 = 1.$$

  For other values of z, we use integration by parts. Observe that for any $T > \varepsilon > 0$ and $\mathrm{Re}(z) > 0$,

  $$\int_\varepsilon^T t^{z-1} e^{-t}\ dt = \frac{t^z}{z} e^{-t}|_\varepsilon^T - \int_\varepsilon^T \frac{t^z}{z}(-e^{-t})\ dt$$

  so taking limits as $\varepsilon \to 0$ and $T \to \infty$ we obtain

  $$\Gamma(z) = \int_0^\infty t^z e^{-t}\ dt = 0 + \int_0^\infty \frac{t^z}{z} e^{-t}\ dt = \frac{1}{z}\Gamma(z+1).$$

  In other words, we have

  $$\Gamma(z+1) = z\Gamma(z) \text{ whenever } \mathrm{Re}(z) > 0;$$

  thus $\Gamma(2) = 1$, $\Gamma(3) = 2$, $\Gamma(4) = 6$, and more generally $\Gamma(n+1) = n!$. Thus one can think of $\Gamma(z+1)$ as a generalization of the factorial function.

- Now we prove another identity for the Gamma function.

- **Theorem.** If $0 < \mathrm{Re}(z) < 1$, then $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$.
- **Proof.** We expand

$$\Gamma(z)\Gamma(1-z) = \int_0^\infty \int_0^\infty t^{z-1} e^{-t} s^{-z} e^{-s} \; dsdt$$

(this integral is absolutely convergent by the previous discussion). Making the change of variables $s = u - t$, this becomes

$$\int_0^\infty \int_t^\infty t^{z-1} e^{-u} (u-t)^{-z} \; dudt$$

or by swapping the t and u integration (and watching the limits of integration carefully)

$$\int_0^\infty \int_0^u t^{z-1} (u-t)^{-z} \; dt e^{-u} du.$$

Making another change of variables $t = uy$, this becomes

$$\int_0^\infty \int_0^1 u^{z-1} y^{z-1} u^{-z} (1-y)^{-z} \; udy e^{-u} du.$$

The powers of u all cancel, and we can separate the integrals as

$$(\int_0^1 y^{z-1}(1-y)^{-z} \; dy)(\int_0^\infty e^{-u} \; du).$$

The second integral is $\Gamma(1) = 1$ which can be discarded. To evaluate the first integral, we use yet another change of variables, $y = \frac{x}{x+1}$, where x goes from 0 to ∞. Then $dy = \frac{1}{(x+1)^2} \; dx$, and the above expression becomes

$$\int_0^\infty (\frac{x}{x+1})^{z-1} (\frac{1}{x+1})^{-z} \frac{1}{(x+1)^2} \; dx$$

which simplifies to

$$\int_0^\infty \frac{x^{z-1}}{x+1} \; dx = \int_0^\infty \frac{e^{(z-1)\ln x}}{x+1} \; dx.$$

Call this quantity X; our task is to show that $X = \frac{\pi}{\sin(\pi z)}$. We are now in a position to use contour integration. Let $f(\zeta)$ denote the function

$$f(\zeta) := \frac{e^{(z-1)\mathrm{Log}_0(\zeta)}}{\zeta + 1}$$

where Log_0 is the branch of the logarithm with imaginary part in $[0, 2\pi)$. This function is analytic everywhere except at -1 and on the positive real axis. At -1 it has a simple pole with residue

$$e^{(z-1)\mathrm{Log}_0(-1)} = e^{\pi i(z-1)} = -e^{\pi i z}.$$

Let $0 < r < 1 < R$. We now introduce the C-shaped contour γ which consists of the following four arcs: the straight line segment from $r + \varepsilon i$ to $R + \varepsilon i$ for some extremely small ε; the circular arc anti-clockwise from $R + \varepsilon i$ to $R - \varepsilon i$; the straight line segment from $R - \varepsilon i$ to $r - \varepsilon i$; and the circular arc clockwise from $r - \varepsilon i$ to $r + \varepsilon i$. This is a simple closed anticlockwise contour enclosing a simple pole at -1, so by the residue theorem

$$\int_\gamma f(\zeta)\, d\zeta = -2\pi i e^{\pi i z}.$$

On the other hand, on the line from $r+\varepsilon i$ to $R+\varepsilon i$, $\mathrm{Log}_0(\zeta)$ is basically the same as $\ln(\mathrm{Re}\zeta)$, and this integral becomes (as $\varepsilon \to 0$)

$$\int_r^R \frac{e^{(z-1)\ln x}}{x+1}\, dx.$$

Conversely, on the line $R - \varepsilon i$ to $r - \varepsilon i$, $\mathrm{Log}_0(\zeta)$ is basically the same as $\ln(\mathrm{Re}\zeta) + 2\pi i$, and this integral becomes (as $\varepsilon \to 0$)

$$\int_R^r \frac{e^{(z-1)(\ln x + 2\pi i)}}{x+1}\, dx = -e^{2\pi i z} \int_r^R \frac{e^{(z-1)\ln x}}{x+1}\, dx.$$

Now look at the big circle with radius R. This circle has length $2\pi R$ (as $\varepsilon \to 0$). On this circle, the real part of $\mathrm{Log}_0 \zeta$ is $\ln R$, and the imaginary part at most 2π, so we have

$$|e^{(z-1)\mathrm{Log}_0\zeta}| = e^{\mathrm{Re}((z-1)(\ln R + \mathrm{Arg}_0\zeta)} \leq e^{\ln R\mathrm{Re}(z-1)} e^{2\pi \mathrm{Im}(z-1)} = R^{\mathrm{Re}(z-1)} e^{2\pi \mathrm{Im}(z)}.$$

Meanwhile, $\frac{1}{\zeta+1}$ has magnitude at most $\frac{1}{R-1}$. Thus the integral on this circle is bounded in magnitude by

$$2\pi R \frac{R^{\mathrm{Re}(z-1)} e^{2\pi \mathrm{Im}(z)}}{R-1}$$

which goes to zero as $R \to +\infty$ since $\mathrm{Re}(z-1) < 0$. Now consider the small circle with radius r. This circle has radius $2\pi r$, and arguing similarly to before

$$|e^{(z-1)\mathrm{Log}_0 \zeta}| = r^{\mathrm{Re}(z-1)} e^{2\pi \mathrm{Im}(z)}.$$

Meanwhile, $\frac{1}{\zeta+1}$ has magnitude at most $\frac{1}{1-r}$, thus the integral on this circle is bounded by

$$2\pi r \frac{r^{\mathrm{Re}(z-1)} e^{2\pi \mathrm{Im}(z)}}{1-r}$$

which goes to zero as $r \to 0$ (since $\mathrm{Re}(z-1) > -1$). Putting this all together we see that in the limit $r \to 0$ and $R \to +\infty$

$$(1 - e^{2\pi i z}) \int_0^\infty \frac{e^{(z-1)\ln x}}{x+1}\, dx = -2\pi i e^{\pi i z}$$

or in other words

$$X = \frac{-2\pi i e^{\pi i z}}{1 - e^{2\pi i z}} = \frac{\pi}{(e^{\pi i z} - e^{-\pi i z})/2i} = \frac{\pi}{\sin \pi z}$$

as desired. $\square$

- Thus for instance

$$\Gamma(1/2)\Gamma(1/2) = \frac{\pi}{\sin(\pi/2)} = \pi.$$

Since $\Gamma(1/2)$ is defined as the integral of a positive function, it is positive, and hence $\Gamma(1/2) = \sqrt{\pi}$, and thus $\Gamma(3/2) = \sqrt{\pi}/2$, etc. Thus for instance we can say that the factorial of $\frac{1}{2}$ is $\sqrt{\pi}/2$!

- Using these identities we can now extend Γ to the left half-plane also. We first need a technical lemma concerning the uniqueness of analytic extensions (also called analytic continuations).

- **Lemma.** Let D be a domain, and let $f(z)$, $g(z)$ be two analytic functions on D. If f and g are equal on a non-empty open subset of D, then they are in fact equal on all of D.

- **Proof.** If f and g are not equal on all of D, then $f - g$ is not the zero function. But then all the zeroes of $f - g$ are isolated, which contradicts the assumption that they agree on a non-empty open subset. $\square$

- We can now define the Γ function on the half-plane $\{z : \mathrm{Re}(z) > -1\}$, except at 0, by the formula

  $$\Gamma(z) := \Gamma(z + 1)/z.$$

  This new definition of Γ is clearly analytic in $\{z : \mathrm{Re}(z) > -1\} - \{0\}$, and agrees with the old definition of Γ back in $\{z : \mathrm{Re}(z) > 0\}$, so by the above Lemma it is the unique analytic extension of Γ to this domain. By construction, the identity $\Gamma(z + 1) = z\Gamma(z)$ now holds for the entire region $\{z : \mathrm{Re}(z) > -1\} - \{0\}$, and so we can repeat this process, extending Γ to the region $\{z : \mathrm{Re}(z) > -2\} - \{0, -1\}$ again by the formula $\Gamma(z) = \Gamma(z+1)/z$, and so forth. Eventually we can obtain a consistent definition of the Gamma function which is defined for all complex numbers except $0, -1, \ldots$. Observe that this procedure tells us that $\Gamma(z)$ has a simple pole at 0 (because $\Gamma(z + 1)$ is analytic and non-zero at 0, so when divided by z produces a simple pole), and then inductively we also have simple poles at -1, -2, etc. (Quiz: what are the residues at these poles? Start at 0 and work backwards).

- With this new definition of the Γ function, which extends the old one, the function $\Gamma(z)\Gamma(1 - z)$ is now analytic and defined on the entire complex plane except the integers. So is the function $\frac{\pi}{\sin(\pi z)}$. Since these functions agree on the strip $\{z : 0 < \mathrm{Re}(z) < 1\}$, they must then agree everywhere by the above uniqueness lemma. Thus we have

  $$\Gamma(z)\Gamma(1 - z) = \frac{\pi}{\sin(\pi z)}$$

  for all complex numbers z which are not integers. (This actually gives an alternate way to extend the Gamma function to the left-half plane). Note that this formula also shows that Γ has no zeroes, because the right-hand side has no zeroes.

- The Gamma function arises in many contexts, for instance it turns up in high-dimensional geometry when computing the volumes of balls or the surface area of spheres. Next week we will see how it interacts with another important analytic function, the *Riemann zeta function*.