Lecture 2: Algebraic functions and analytic continuation

September 9, 2014

In this lecture we'll define the notion of an algebraic function, and state a non-trivial necessary and sufficient condition for algebraicity, which we'll prove later in the course.

Let's start with another general principle from complex analysis: the *identity principle*.

Theorem 0.1. *Let D be an open disk in $\mathbb{C}$. Suppose given two holomorphic functions $f, g : D \to \mathbb{C}$ which agree on some smaller open disk $D' \subset D$. Then $f = g$ on all of D.*

Proof. Let r denote the radius of D'. Consider the points along the line segment joining the center of D' to that of D which are evenly spaced at distance $r/2$ apart, starting with the center of D'; label them $z_0, z_1, \dots z_n$. Since z_1 lies in D', it follows that the derivatives of f and g to all orders agree at z_1. Therefore the Taylor expansions of f and g agree at z_1. Now recall a theorem from last lecture: the Taylor series expansion of a function converges to that function wherever that function is defined and holomorphic. It follows in particular that f and g agree on the disc D'_1 of radius r centered at z_1. Continuing in this manner, we find inductively that f and g agree on all the analogous disks D'_i, and in particular agree on D'_n. But D'_n contains the center of D; thus f and g have the same Taylor expansions at the center of D, and hence by the same theorem agree on all of D. □

There is an extension of this theorem, where D is replaced by a more general open subset of $\mathbb{C}$. The relevant notion is *connectedness*.

Definition 0.2. *Let U be an open subset of $\mathbb{C}$. We say that U is (path)-connected if for every two points $z, w \in U$, there is a continuous function (a "path") $\gamma : [0,1] \to U$ such that $\gamma(0) = z$ and $\gamma(1) = w$.*

An example of a non-connected open subset is a subset U which is the union of two disjoint open disks. Giving a holomorphic function on such a U is the same thing as separately giving a holomorphic function on each of the disks. Thus the identity principle is certainly not valid on U. But on the other hand:

Theorem 0.3. *The identity principle holds for all connected open subsets of $\mathbb{C}$.*

Proof. Let U be the connected open subset, and let $f, g : U \to \mathbb{C}$ be the two holomorphic functions which agree on some open disk $D \subset U$. Let $z \in U$ be arbitrary; we want to show that $f(z) = g(z)$.

For this, choose a path $\gamma : [0,1] \to U$ with $\gamma(0) \in D$ and $\gamma(1) = z$. Since U is open and γ is continuous, we can choose, for every $t \in I$, an open disk $D_t \subset U$ centered at $\gamma(t)$ and an open interval $I_t \subset [0,1]$ centered at t with $\gamma(I_t) \subset D_i$.

Now, we recall a topological fact about the unit interval $[0,1]$: **Suppose given, for every $t \in [0,1]$, an open interval I_t centered at t. Then there exist real numbers $0 = t_0 < t_1 < \dots < t_N = 1$ such that for each $n = 0, \dots, N-1$, the intervals I_{t_n} and $I_{t_{n+1}}$ intersect nontrivially.** This is essentially the so-called *compactness* of the unit interval.

Applying this to our situation, we deduce that there are open disks D_i for $i = 0, \dots, N$ such that D_i and D_{i+1} intersect nontrivially, D_0 is centered at $\gamma(0)$, and D_N is centered at $\gamma(1) = z$. For any two open disks which intersect we can find an open disk inside the intersection, so by the identity principle for disks we deduce that if $f = g$ on D_i, then $f = g$ on D_{i+1}. But on the other hand, D and D_0 also intersect nontrivially, and we know $f = g$ on D. Thus inductively we deduce that $f = g$ on D_N, hence $f(z) = g(z)$, as desired. $\square$

Basically, the idea is that we shoot out a path from where we know $f = g$ to where we don't know that. Then we crawl along this path by finitely many intersecting open disks, to inductively deduce that $f = g$ along the whole path. In this manner, the values of a holomorphic function (on a connected open subset) near any point uniquely determine what it can be at every point.

However, this does *not* mean that it's always possible to extend a holomorphic function from a small disk to a larger region. There are essentially two possible obstructions, one analytic and the other topological. We'll be concerned with the topological obstruction today. To illustrate it, we discuss *algebraic functions*.

First, we let $\mathbb{C}(z)$ denote the set of *rational functions of* z, i.e. the set of quotients $p(z)/q(z)$ where p and q are polynomials with $\mathbb{C}$-coefficients such that $q \neq 0$. Each $f \in \mathbb{C}(z)$ can be considered as a holomorphic function on the open subset U_f which is the complement in $\mathbb{C}$ of the finite set of zeros of the denominator of f.

Then we let $\mathbb{C}(z)[w]$ denote the set of polynomials in w with coefficients in $\mathbb{C}(z)$. (If there are no denominators in the coefficients, this is just a polynomial in two variables.) For a $P \in \mathbb{C}(z)[w]$, denote by U_P the intersection of all the U_f as f ranges over the coefficients of P. Thus $P(z, w)$ can be considered as a complex-valued function on $U_P \times \mathbb{C} \subset \mathbb{C}^2$.

A simple example to keep in mind is $P(z, w) = w^2 - z$. This is a polynomial of degree two in w. In this case there are no denominators, so P is a function on all of $\mathbb{C}^2$. A more complicated example would be

$$P(z, w) = \frac{z}{z^2 - i} \cdot w^3 + z \cdot w + \frac{1}{z}.$$

Here is our main definition.

Definition 0.4. *Let $f : U \to \mathbb{C}$ be a holomorphic function on a connected open subset of the complex plane. We say that f is* algebraic *if there exists a nonzero $P \in \mathbb{C}(z)[w]$ such that $U \subset U_P$ and*

$$P(z, f(z)) = 0$$

for all $z \in U$.

Thus, f is algebraic if there is a polynomial relationship between f and some rational function. One can think of this in the following way. If z_0 is a specific value in U_P, then $P(z_0, w) = 0$ is just a polynomial equation for w with complex coefficients. Then the equation above says in particular that $f(z_0)$ should be picking out one such solution w_0, out of the (generally) finitely many possibilities. But then the same should also remain true as z_0 varies. So an algebraic function is one which solves some family of algebraic equations, e.g. for the example $P(z, w) = w^2 - z$ we would be figuring out how to extract square roots, i.e. we are solving quadratic equations. The question is whether this can be done holomorphically in the parameter z.

The first main theorem is a *local existence and uniqueness theorem*:

Theorem 0.5. *Let $P \in \mathbb{C}(z)[w]$, and let $(z_0, w_0) \in \mathbb{C}^2$ be such that $z_0 \in U_P$, $P(z_0, w_0) = 0$, and $\partial_w P|_{(z_0,w_0)} \neq 0$. Then there exists an open disk D around z_0 such that there is a unique* *holomorphic function $f : D \to \mathbb{C}$ with $f(z_0) = w_0$ and satisfying*

$$P(z, f(z)) = 0$$

for all $z \in D$.

Thus, if you give an initial solution w_0 to the algebraic equation at the point z_0, then subject to a mild techinical hypothesis (the non-vanishing of the partial derivative) you can uniquely extend this to a holomorphic family of solutions in an neighborhood of z_0. For example, a choice of square root of 1 gives you a choice of square roots of all complex numbers close enough to 1. In fact, the usual power series expansion of the square root function converges in the disk $D(1,1)$ and verifies the above theorem in this specific case.

Proof. We will actually not use the specific nature of the function $P : U_P \times \mathbb{C} \to \mathbb{C}$: we will only need that it is a *holomorphic function of two variables*. The meaning of this is that P is continuously differentiable as a function of four real variables, and for each $(z, w) \in U_P \times \mathbb{C}$, the derivative

$$dP|_{(z,w)} : \mathbb{C}^2 \to \mathbb{C},$$

which is a priori only a real-linear map, is actually complex-linear.

Now, the hypothesis on ∂_w shows that $dP|_{(z_0,w_0)}$, when restricted to the second coordinate $0 \times \mathbb{C}$, is bijective. (Indeed, it is a complex-linear map $\mathbb{C} \to \mathbb{C}$ which sends 1 to $\lambda = \partial_w P|_{z_0,w_0}$; the only such map is multiplication by λ.) Therefore the implicit function theorem tells us that there exists a disk D containing z_0, a disk D' containing w_0, and a continuously real-differentiable function $f : D \to D'$ such that

$$P(z, f(z)) = 0.$$

To finish the proof, we need to show that f is holomorphic, and prove the uniqueness claim. To see that f is holomorphic, let $z \in D$. We recall the following description of $df|_z$ (valid potentially after shrinking D again), which also comes from the implicit function theorem: the kernel of $dP|_{(z,f(z))}$ projects bijectively to $\mathbb{C} \times 0$, and $df|_z$ is the composition of the inverse of this projection, followed by the projection to the other factor $0 \times \mathbb{C}$.

But now, since $dP|_{(z,f(z))}$ is complex-linear, the kernel is a complex subspace of $\mathbb{C}^2$, and both projections are complex-linear maps. But the inverse of a complex-linear map is also complex-linear, so it follows that $df|_z$ is complex linear, hence f is holomorphic on D.

Now for the uniqueness claim. The value of f at z_0 is constrained to be w_0. Then by iteratively differentiating the equation $P(z, f(z)) = 0$, we can successively calculate all the higher order derivatives of f at z_0 in terms of previous ones. Therefore we know the Taylor expansion of f a priori, so uniqueness follows from the identity principle. $\square$

Thus the local nature of algebraic functions is fairly clear: locally, they exist if you just specify a correct value at one point. But globally, it gets interesting. For example, one could ask the following. Fix a $P \in \mathbb{C}(z)[w]$, and let $S \subset \mathbb{C}$ denote the set of points $z \in \mathbb{C}$ where either z is the root of a denominator of a coefficient of P, or there exists a w such that $\partial_w P|_{(z,w)} = 0$. (Except in degenerate cases, S will be a finite set.) Then for every $z \in U := \mathbb{C} \setminus S$ we know that there is a locally defined holomorphic function f near z which satisfies

$$P(z, f(z)) = 0.$$

There is a natural question: is there a holomorphic function $f : U \to \mathbb{C}$ on all of U satisfying $P(z, f(z)) = 0$?

The answer is very much "no". The general principle behind this is the following: *locally non-uniqueness obstructs global existence*. To get going locally around z_0 we had to make an arbitrary choice of a companion value of w_0 out of a number of different possibilities; such an arbitrary choice will not be respected globally, in general. But let us see this concretely in an example.

Consider again $P(z,w) = w^2 - z$. Here $\partial_w P = 2w$; this is zero only when $w = 0$. Since $w^2 = z$, that's only possible when $z = 0$. So in this case the question is whether there is a holomorphic function $f : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ such that $f(z)^2 = z$.

Proposition 0.6. *There is no such function f.*

Proof. Replacing f by $-f$ if necessary, we can assume that $f(1) = 1$. Now, we can write down an explicit holomorphic function $g : U \to \mathbb{C}$ with $g(z)^2 = z$, where U denotes the complement of $(-\infty, 0]$ sitting inside the real axis. Namely, we can represent z in polar coordinates (r, θ) with θ constrained to lie in $(-\pi, \pi)$, and set

$$g(z) = (\sqrt{r}, \theta/2).$$

Clearly $g(z)^2 = z$; to see that g is holomorphic, one way to argue is as follows. From our description of the n^{th} power maps in the previous lecture it follows that $\varphi(z) = z^2$ gives a holomorphic bijection between the right-half plane $\Re(z) > 0$ in $\mathbb{C}$ and our region U; furthermore its derivative $2z$ is everwhere nonzero in this region. Thus φ is a biholomorphism, by the inverse function theorem. But the map g is nothing but an inverse to this biholomorphism; thus it is itself holomorphic.

Now, $f(1) = g(1)$ and both solve $P(z,w) = 0$. Moreover U is connected. So by the local uniqueness and the identity principle, we must have $f = g$.

But this is a problem, since g does not even admit a continuous extension to $\mathbb{C}$, let alone a holomorphic one. Indeed, if we approach the point -1 from above, then we get $g(-1) = i$; but if we approach it from below, we get $g(-1) = -i$. $\square$

How can we understand this phenomenon? One way is through the idea of *analytic continuation*. Recall how we proved the identity principle for connected open subsets: we had two functions which agreed somewhere, then we proved they had to agree at some farther point by moving along a path. But what if we only have one function, only defined near one point? We can still try to figure out what its values should be as we move along the path. That's the idea behind the following definition:

Definition 0.7. *Let $f : U \to \mathbb{C}$ be a holomoprhic function on some open subset of $\mathbb{C}$, and let $\gamma : [0,1] \to \mathbb{C}$ be an arbitrary path with $\gamma(0) \in U$. We say that f* can be analytically continued along γ *if and only if there exists a sequence*

$$0 = t_0 < t_1 < \ldots < t_N = 1$$

in $[0,1]$, open disks D_i around $\gamma(t_i)$ with $D_i \cap D_{i+1} \neq \emptyset$, and holomorphic functions $f_i : D_i \to \mathbb{C}$ such that f_i and f_{i+1} agree on $D_i \cap D_{i+1}$ and f_0 agrees with f on $D_0 \cap U$.

In such a case we say that $f_N(\gamma(1))$ is the value of f at $\gamma(1)$, as determined by analytic continuation along γ.

This is a reasonable definition, because the same argument as we used in the identity principle shows that the value $f_N(\gamma(1))$ is independent of everything except the path γ and the original function f. Furthermore, if the path γ were to travel in a region where f is defined and holomorphic, then $f_n(\gamma(1))$ would necessarily agree with $f(\gamma(1))$. We may abusively write $f(\gamma(1))$ for the result of analytic continuation of f along γ, but this is technically incorrect since it can depend on γ, not just $\gamma(1)$.

Now we can say in a different way what goes wrong with the global square root function: if we start with a square root function in a neighborhood of 1 and we try to analytically continue it, at first it looks like we're in business: the function can be analytically continued along any path which avoids the origin. But, there's a catch: the result of analytic continuation depends on more than the endpoint: it depends on the path chosen. For example, in trying to analytically continue to -1, if we go around 0 from above, then we get i; but if we go around 0 from below, then we get $-i$. These don't agree, so there can be no global square root function.

Now we can state a theorem we'll prove later, which gives a characterization of algebraic functions in terms of analytic continuation.

Theorem 0.8. *Let $f : U \to \mathbb{C}$ be a holomorphic function on a connected open set. Then f is algebraic if and only if there exists a finite subset $S \subset \mathbb{C}$ such that the following conditions are satisfied:*

1. *For any path $\gamma : [0,1] \to \mathbb{C}$ with $\gamma(1) \notin S$, there exists an analytic continuation of f along γ. Moreover, if we fix the endpoint $\gamma(1)$, the set of possible values $f(\gamma(1))$ for this analytic continuation is finite.*

2. *Whenever you analytically continue to z with z approaching a point $s \in S$, there is the estimate*
$$|f(z)| \leq C \cdot |z-s|^{-d}$$
for some $d \in \mathbb{N}$ and positive constant C. (Thus: f blows up no worse than a rational function at the missing points)

3. *Whenever you analytically continue to z with $|z|$ approaching ∞, there is the estimate*
$$|f(z)| \leq C \cdot |z|^d$$
for some $d \in \mathbb{N}$ and positive constant C. (Thus: f grows no faster than polynomially at ∞.)

The two conditions 2 and 3 should really be joined into one; this will become clear once we talk about the Riemann sphere.