The Fundamental Theorem of Algebra

Rich Schwartz

October 1, 2014

1 Complex Polynomals

A *complex polynomial* is an expression of the form

$$P(z) = c_0 + c_1 z + \dots + c_n z^n,$$

where $c_0, \dots, c_n$ are complex numbers (called *coefficients*) and z is the variable. The number n is called *the degree* of P, at least when P is written so that $c_n \neq 0$. We'll always divide through by c_n so that $c_n = 1$.

A *root* of P is a value of z so that $P(z) = 0$. (Dividing through by c_n doesn't change the roots.) Some polynomials have no real roots, even if they have real coefficients. The polynomial $P(z) = z^2 + 1$ has this property.

The history of finding roots of polynomials goes back thousands of years. It wasn't until the 1800s, however, that we had a good picture of what is going on in general. The goal of these notes is to sketch a proof of the most famous theorem in this whole business.

Theorem 1.1 *Every complex polynomial has a root.*

This theorem is called the Fundamental Theorem of Algebra, and it is due to Gauss. It seems that Gauss proved the theorem in 1799, though his original proof had some gaps. The first complete proof is credited to Argand in 1806.

The proof I'm going to sketch has a "topological flavor". It only depends on general features of polynomials, and the notion of continuity. It seems more or less obvious, though some of these obvious steps are a little tricky to make precise.

2 Continuous Loops

A function f is *continuous* at a point x_0 if the following statement is true: For any $\epsilon > 0$ there is a $\delta > 0$ so that $|x - x_0| < \delta$ implies that $|f(x) - f(x_0)| < \epsilon$. Informally, f is continuous at x_0 if, when you change the value of x_0 a little, the value of f only changes a little. The function f is called *continuous* if it is continuous at all points where it is defined.

In a calculus class, this definition is usually made for real valued functions – i.e. for the case when x_0 and x and $f(x_0)$ and $f(x)$ are all real numbers. However, the definition also makes sense when x and x_0 are points in a circle and $f(x_0)$ and $f(x)$ are complex numbers. This leads to the main definition of this section: A *continuous loop* is a continuous map $f : C \to \boldsymbol{C}$, where C is a circle. So, the input to f is a point on the circle C, and the output is a complex number – i.e., a point in $\boldsymbol{C}$.

Given a complex polynomial P, and a positive number R, we can look at the map $P : C_R \to \boldsymbol{C}$. Here C_R is the set of complex numbers having norm R. That is, C_R is the circle of radius R centered at the origin. We are looking at what P does to the points on C_R. The map $P : C_R \to \boldsymbol{C}$ is a continuous loop. So, for each value of R, you get a different continuous loop.

3 Winding Number

Suppose that $f : C \to \boldsymbol{C}$ is a continuous loop with the property that the image $f(C)$ does not contain the origin. (The origin is the point 0.) We can assign an integer to the curve $f(C)$ like this: We trace C around counterclockwise. As we trace C around, we simultaneously trace $f(C)$ around, standing at the origin and looking at $f(C)$. When we have completed one full circuit, our head is looking in the same direction that we started but our neck has been twisted some number of times, either clockwise or counterclockwise. The *winding number* of $f(C)$ is defined to be k if our neck has been twisted k times counterclockwise and $-k$ if our neck has been twisted k times clockwise. The winding number measures how many times the curve $f(C)$ winds around the origin.

This definition can be made more formal mathematically, but it is a situation in which a picture says a thousand words. Figure 1 shows some continuous loops and their winding numbers.

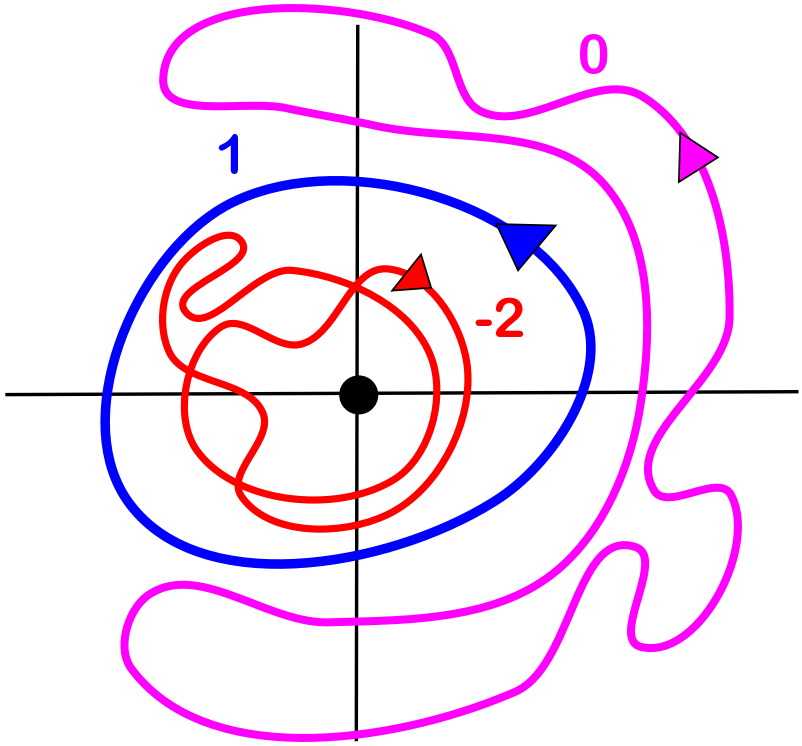

Figure 1: Some winding numbers.

The curve $f(C)$ is called *linked* if the winding number of $f(C)$ is nonzero. In Figure 1, the red and blue curves are linked and the magenta curve is not linked.

Here is the case of the winding number we are interested in. Imagine that P is a polynomial of degree n, and R is an enormous number. What is the winding number of $P : C_R \to \boldsymbol{C}$? Well, we can write P as

$$P = z^n + (\ldots)$$

When R is very large, the term z^n is much larger than the other terms, and the curve $P(C_R)$ looks almost like a circle of radius R^n which winds n times around the origin. The point here is that the z^n term accounts for most of the shape, and the other terms are just "noise". When R is small, the picture, of course, is much more complicated.

In short, when R is large enough, the image $P(C_R)$ is linked.

4 The End of the Proof

Suppose now that P is a degree n polynomial. We'll assume that P has no roots and derive a contradiction.

Consider the continuous loops $P(C_R)$, where R varies. When R is very small, $P(C_R)$ is just a tiny loop concentrated around the point $P(0) = c_0$. Here is what the picture looks like, more or less:

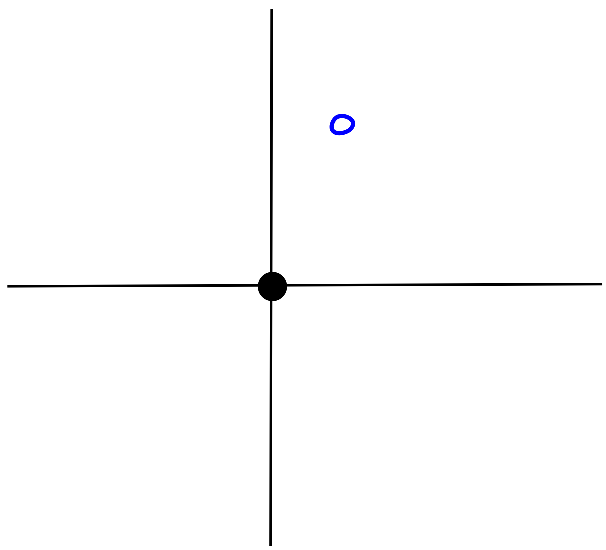

Figure 2: $P(C_R)$ for R small.

Note that $P(0) \neq 0$ because P has no roots. So, when R is very small, $P(C_R)$ is unlinked. On the other hand, when R is very large $P(C_R)$ is linked. How can this happen? If we just wiggle $P(C_R)$ around a bit, the winding number does not change. The point is that, when we trace out $P(C_R)$ and the nearby version of $P(C_R)$, our head is essentially looking in the same direction for both loops. So, at the end, our neck has been twisted the same number of times for both.

The only way for the winding number to change, as we vary R, is that $P(C_R)$ crosses over the origin for some value of R. But if $P(C_R)$ contains the origin, there is some $z \in C_R$ so that $P(z) = 0$. This is a contradiction.

In short, if P has no roots, the winding properties of $P(C_R)$ are the same for values of R. However, as we saw, the winding properties of $P(C_R)$ change as R varies. That's the end of the proof.

5 Making it Rigorous

The argument I sketched seems a bit too informal to count as a rigorous mathematical proof. So, in this section, I'll explain how this is typically made more rigorous. (Feel free to ignore this section.) The main step is making the notion of winding number precise. There are many ways to do this. One way to do it is through *line integrals*. This is something you learn about in a multivariable calculus class.

It turns out that the winding number of $f(C)$ can be defined as the line integral

$$\frac{1}{2\pi}\oint_{f(C)} \alpha.$$

Here

$$\alpha = \frac{-y}{x^2+y^2}dx + \frac{x}{x^2+y^2}dy.$$

The integral only makes sense when $x^2 + y^2 \neq 0$. So, we can only define the winding number for curves which do not contain 0.

In general, a 1-form $Pdx + Qdy$ is called *closed* if

$$\frac{\partial P}{\partial y} - \frac{\partial Q}{\partial x} = 0.$$

The form α has this property. For closed 1-forms, it follows from Green's theorem that the value of the line integral does not change if the path is perturbed – as long as the perturbation never leaves the domain where the 1-form is defined. In our case, the domain is the whole plane, minus the origin.

This is the justification for the claim that the winding number of $P(C_R)$ cannot change as R values unless $P(C_R)$ contains the origin for some R