The Fundamental Theorem of Algebra

Let $f(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n$, where n is a positive integer, $\alpha_j \in \mathbb{C}$ for each j, and $\alpha_n \neq 0$. The Fundamental Theorem of Algebra says that there exists a $\gamma \in \mathbb{C}$ such that $f(\gamma) = 0$.

Note that if $f(\gamma_1) = 0$ then $z - \gamma_1$ is a factor of $f(z)$, and so we may write $f(z) = (z - \gamma_1)g(z)$, where g is a polynomial of degree $n - 1$. If $n - 1 > 1$ then the same reasoning yields a factor $z - \gamma_2$ of $g(z)$, and so it follows by repeated applications of the argument that $f(z)$ may be express as a product of n factors of degree 1: $f(z) = \alpha(z - \gamma_1)(z - \gamma_2)\cdots(z - \gamma_n)$.

It is clear that a polynomial of odd degree with real coefficients has at least one real root. This is a trivial application of the Intermediate Value Theorem: if p is such a polynomial then $p(x) \to \infty$ as $x \to \infty$ and $p(x) \to -\infty$ as $x \to -\infty$ (or the other way round if the leading coefficient of p is negative), and so there must be a point where $p(x)$ changes sign. It is natural to look for some generalization of this argument that can be applied to complex polynomial functions. Now investigation of complex functions of a complex variable is made more difficult by the fact that to graph $w = f(z)$ one would need to use four real dimensions, one each for the real and imaginary parts of z and of w. Nevertheless, it may be true that a four-dimensional being would consider the Fundamental Theorem of Algebra just as obvious as the above statement about real polynomials of odd degree.

The idea we shall apply is as follows. Imagine two complex planes, one for z and one for w. For each point in the z-plane the polynomial f gives us a corresponding point $w = f(z)$ in the w-plane. Now imagine z traversing a curve in its plane. The corresponding values of w yield a curve in the other plane. For example, if r is a fixed nonnegative real number, then the set $\{ z \mid |z| = r \}$ is a circle in the z-plane, and its image, the set $\{ f(z) \mid |z| = r \}$, will be some closed (possibly self-intersecting) curve in the w-plane. The circle in the z plane can be given parametrically as $z = r(\cos\theta + i\sin\theta)$, where $\theta \in [0, 2\pi]$. And $w = f(r(\cos\theta + i\sin\theta))$, for $\theta \in [0, 2\pi]$, gives a parametric form for the image of this circle in the w-plane.

The simplest case to consider is the polynomial $h(z) = z^n$. Since

$$(r(\cos\theta + i\sin\theta))^n = r^n(\cos(n\theta) + i\sin(n\theta)),$$

and since $n\theta$ goes from 0 to $2\pi n$ as θ goes from 0 to 2π, we see that in this case w moves n times round a circle of radius r^n as z moves once around a circle of radius r. If we modified this to $h(z) = \alpha z^n$, where $\alpha \neq 0$, the curve in the w-plane would have radius $|\alpha| r^n$, the initial and final point would be αr^n, and it would still be traversed n times.

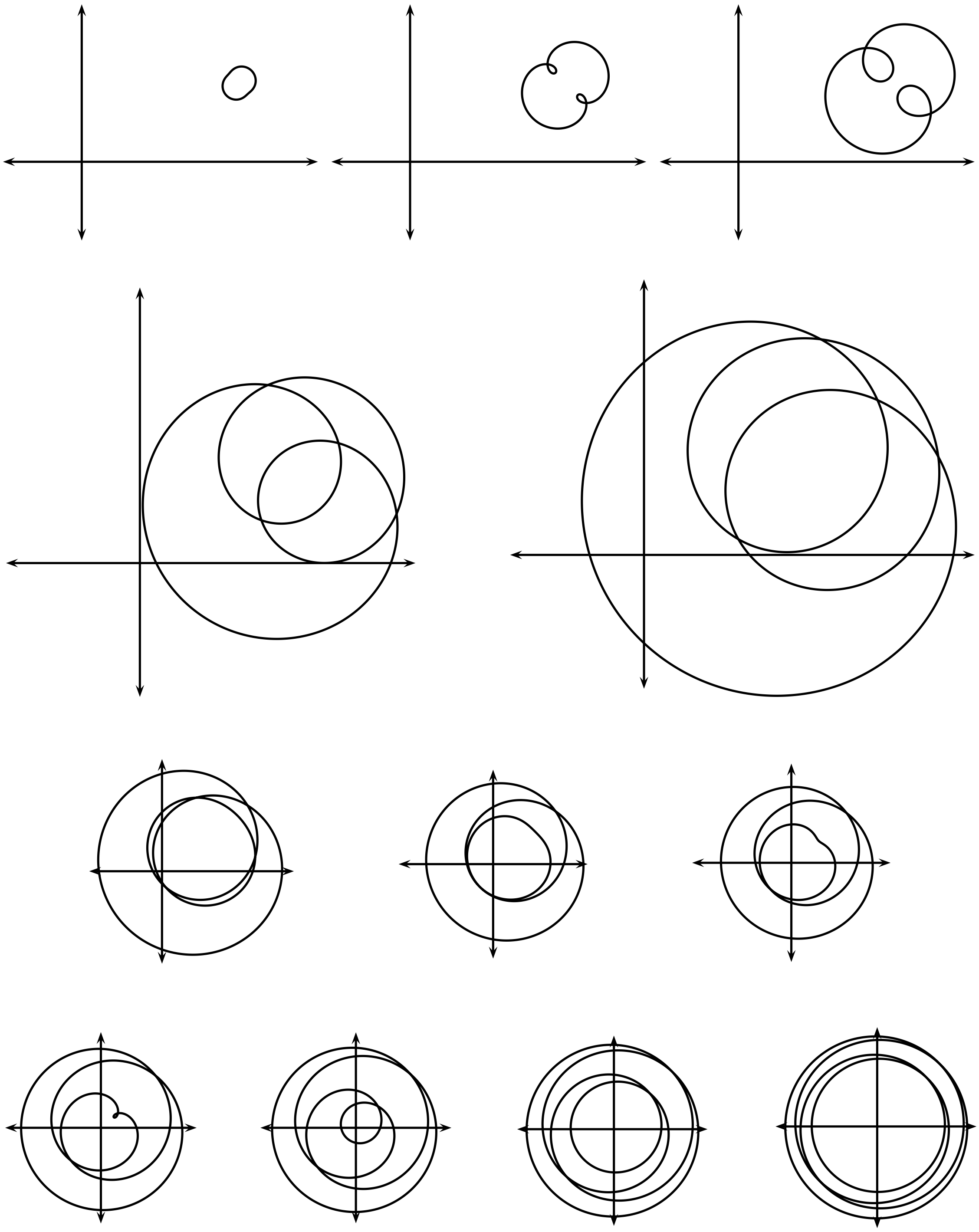
Now consider the images of the circles $\{ z \mid |z| = r \}$ under the polynomial function $f(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_n z^n$. Let us write $\mathcal{C}_r = \{ f(z) \mid |z| = r \}$, the image curve in the w-plane. If $r = 0$ then of course the circle reduces to a single point and its image $\mathcal{C}_0$ is also a single point, namely $f(0) = \alpha_0$. We can assume that $\alpha_0 \neq 0$, for otherwise $\gamma = 0$ is a root of f, and nothing more needs to be proved. The curve $\mathcal{C}_r$ clearly varies continuously with r, in some sense which we shall not attempt to make rigorous. So we can ensure that $\mathcal{C}_r$ is contained in any given small neighbourhood of α_0 by choosing r to be sufficiently small. In particular, if the small neighbourhood does not contain the origin the origin then the number of times that $\mathcal{C}_r$ winds around the origin, in the sense described in the next paragraph, is zero.

Here is an intuitive account of what it means for a closed curve in the w-plane to wind around the origin a number of times. Assuming that the curve does not pass through the origin, the argument, $\arg(w)$, is defined for all points w on the curve. (Recall that the argument of $w \neq 0$ is the angle ψ such that $w = |w|(\cos\psi + i\sin\psi)$.) Now the argument is a many-valued function, but we may choose one particular value for the argument of the initial point and require that the argument varies continuously as we move along the curve. Since the final point is the same as the initial point, the value of the argument at the final point will differ from the initial value by $2\pi k$, for some integer k. This integer is the winding number. Observe that the winding number is not defined for curves that pass through the origin, but is defined for all other continuous closed curves. And continuously deforming a continuous closed curve cannot change the winding number, unless the deformation process at some stage gives a curve that passes through the origin. Intuitively, the winding number changes by 1 when a part of the curve is pushed across the origin.

Now we have seen that if r is small enough then $\mathcal{C}_r$ has a winding number of 0. But if r is large enough the absolute value of $\alpha_n z^n$ will exceed the absolute value of $\alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots + \alpha_{n-1} z^{n-1}$ for all z with $|z| = r$. In fact, as $r \to \infty$ the ratio of the absolute values of these two tends to infinity. The term $\alpha_n z^n$ dominates the other terms in $f(z)$. This means that the winding number for the curve $\mathcal{C}_r$ must be the same as it would have been if $f(z)$ were just $\alpha_n z^n$. In effect, what we are saying is that, when viewed from a distance, for large enough r the curve $\mathcal{C}_r$ is indistinguishable from a circle of very large radius traversed n times. But this winding number is n; so the continuous deformation of $\mathcal{C}_r$ from r close to zero to close to infinity must at some stage yield a curve with no winding number, for otherwise the winding number would have to have remained contstant at zero. So $\mathcal{C}_r$ passes through the origin for some value of r, and this means that $f(z) = 0$ for some z on the circle $\{\, z \mid |z| = r \,\}$.

To illustrate the above argument we have plotted the curves $\mathcal{C}_r$, for certain values of r, when $f(z) = (1+i)z^4 + 3z^3 - iz + (2+i)$. For the first five diagrams 1 cm represents 1 unit, and the values of r are (respectively) 0.2, 0.4, 0.5, 0.7 and 0.8. Clearly $\mathcal{C}_r$ passes through the origin for some value of r a little larger than 0.7.

The next seven diagrams are drawn to different scales chosen so as to keep the diagrams small. The values of r and the scales chosen are, respectively, as follows: $r = 1.0$ and 1 unit is 0.25 cm; $r = 1.2$ and 1 unit is 0.13 cm; $r = 1.4$ and 1 unit is 0.07 cm; $r = 1.8$ and 1 unit is 0.032 cm; $r = 3.1$ and 1 unit is 0.00485 cm; $r = 6.2$ and 1 unit is 0.0004 cm; $r = 12.4$ and 1 unit is 0.000025 cm.



You may also look at the animated display which is meant to show the the continous deformation of $\mathcal{C}_r$ as r changes.