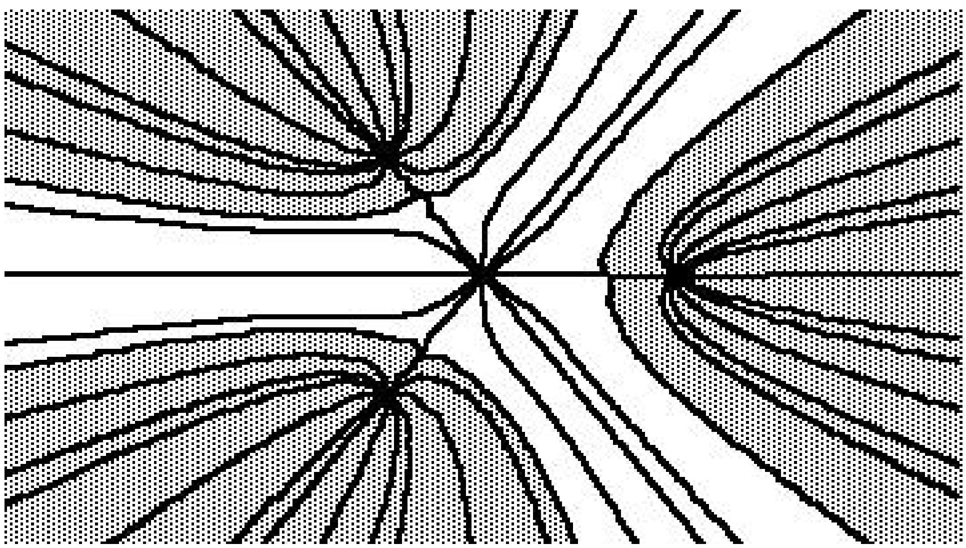
Theory of Equations

Lesson 7

by

Barry H. Dayton
Northeastern Illinois University
Chicago, IL 60625, USA

`www.neiu.edu/~bhdayton/theq/`



Chapter 3

Topology

The Fundamental Theorem of Algebra

3.1 History

Carl Friedrich Gauss, arguably the most brilliant mathematician in the history of mathematics, had a fascination with the Fundamental Theorem of Algebra and gave four proofs during his lifetime, the first one in his doctoral dissertation of 1799. In the first part of this chapter, like Gauss, we will give four proofs, unlike Gauss starting with Gauss' fourth and ending with a modern version of his first and fourth. Later in this chapter we will compare Gauss' first proof to Newton's method and explore where the topological ideas first enunciated by Gauss have taken the modern mathematician.

The first statement of the Fundamental Theorem of Algebra, henceforth to be known in this chapter as the FTA, was due to the Dutch mathematician Albert Girard (he was born in France, but worked in Holland) in 1629. Girard took complex numbers seriously and found explicit imaginary roots for several polynomials for which only real roots had been previously found. In addition he found new real roots for some previously "solved" equations, in each case bringing the total number of roots up to the degree. He summarizes his findings with the statement "All equations of algebra receive as many solutions as the denomination of the highest term shows." Girard does equivocate on "defective" equations with one or more coefficients equal to zero, but most historians of mathematics are willing to accept Girard's statement as the earliest statement of the FTA.

Descartes in his famous Third Book of La Geometrie explicity states that an equation of degree n can have at most n roots. However, Descartes' views on the FTA can merely be inferred from the following paragraph.

> Neither the true nor the false roots are always real; sometimes they are imaginary; that is, while we can always conceive of as many roots for each equation as I have already assigned, yet there is not always a definite quantity corresponding to each root so conceived of. Thus, while we may conceive of the equation $x^3 - 6xx + 13x - 10 = 0$ as having three roots, yet there is only one real root, 2, while the other two, however we may increase, diminish, or multiply them in accordance with the rules just laid down, remain always imaginary.

The first attempt to prove the FTA was made by Jean Le Rond D'Alembert in 1748. D'Alembert's statement of the FTA was that every polynomial can be factored as a product of linear and quadratic real polynomials, i.e. our Theorem 1.9.9. This point of view of the FTA dates back to Leibniz and Johann Bernoulli who needed this result to integrate rational functions. D'Alembert's proof was not very convincing at the time, however for his effort the FTA is still known to this day in France as D'Alembert's Theorem. D'Alembert viewed his theorem as a theorem of calculus and what he was lacking was a solid foundation of such concepts as limit and continuity. D'Alembert was well aware of the limitations of 18th century analysis and was the only important mathematician of that century to be concerned about the foundations of calculus, although his own efforts to rectify the problem were not successful.

At about the same time (1749) Leonhard Euler attempted a purely algebraic proof of the FTA in the form of D'Alembert's Theorem, starting however from the assumption that every polynomial of odd degree has at least one root. The key to his proof is an argument that every polynomial of degree 2^n can be factored into two polynomials of degree 2^{n-1}. D'Alembert's theorem then follows easily: for example to show a polynomial of degree 13 can be factored, multiply by x^3 to get a polynomial of degree 16 which can be factored into two polynomials of degree 8, which in turn can be factored into degree 4 polynomials etc. Although Euler was satisfied with his proof, other mathematicians, notably Gauss and Lagrange were not. Euler did however lay much of the groundwork of complex numbers (he gave the name to the number i) and was the first to prove (Leibniz had tried but failed) that if $a + bi$ was a root then so was $a - bi$.

Joseph Louis Lagrange attempted (in 1772) a proof of the FTA by trying to extend the explicit solution methods described by Cardano for degree 3 and 4 to equations of higher degree. While Lagrange also was not successful, his ideas on exact solutions led directly to the proofs of Abel and Galois on the impossibility of such a solution method. We will discuss the methods of Lagrange, Abel and Galois more completely in the next chapter.

It was then Gauss in 1799 (at age 22) who gave the first proof of the FTA (still in D'Alembert's formulation) that was generally accepted by the mathematical community. What Gauss had noticed was what D'Alembert had only hinted at and what Euler and Lagrange had missed was that the FTA was not a theorem on algebra but rather a theorem on the geometry (more precisely, the topology) of the complex plane. Ironically, to the modern mathematician Gauss' first and fourth proofs have a gap which is harder for us to fill in than the gap in D'Alembert's proof. Of course the gap now can be filled and Gauss's arguments paved the way for a correct view of the theorem.

In his later proofs Gauss takes the modern point of view of the FTA that it is only necessary to show the existence of one complex root for each complex polynomial. In this, Gauss's proofs are considered among the first examples of "non-constructive" proofs, i.e. proofs that assert the existence of some mathematical object without actually describing the object explicitly, i.e. Gauss does not actually give us a method for finding the roots. However we will see in this chapter that Gauss's proofs are not as non-constructive as sometimes supposed.

Before going on to look at Gauss's fourth proof, we mention that Gauss's second proof is along the lines of Euler's proof, he starts with a proof of the fact that every real polynomial of odd degree has a real root. Then given a polynomial of even degree, Gauss replaces the given polynomial by one of much higher odd degree, the roots of which, however allow one to find the roots of the original polynomial.

3.2 Gauss's Fourth Proof

Before proceeding with Gauss's proof we mention a key fact (essentially the complex form of Theorem 2.6.1) which is used somewhere, implicitly or explicitly, in every correct proof of the FTA.

Theorem 3.2.1 *Let $f(z)$ be a monic complex polynomial of degree n. Then*

$$\lim_{|z|\to\infty} \frac{f(z)}{z^n} = 1.$$

Proof: If $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots + z^n$ then $\frac{f(z)}{z^n} = \frac{a_0}{z^n} + \frac{a_1}{z^{n-1}} + \cdots + \frac{a_{n-1}}{z} + 1.$

This theorem says that for z of large modulus $f(z)$ is approximated by z^n closely in a relative sense, i.e. the relative error in approximating $f(z)$ by z^n, $|f(z) - z^n|/|f(z)|$, will be small. The absolute error may, and probably will, be quite large. For example consider $f(z) = 1 + z^7 + z^9, g(z) = z^9$. Given $z = 3 + 3i$ approximately $f(z) =$

$470119.5 + 420632.4i, g(z) = 445375.5 + 445375.5i$ so that $f(z) - g(z) = 24744.1 - 24743.1i$. This is a large number but the relative error is $34993/630828 = .055$ or about 5%. More importantly the argument of $f(z)$ will very closely approximate the argument of z^n for large z. In our example, for instance, the argument of $f(3 + 3i)$ is .7299 (radians) whereas the argument of $(3 + 3i)^9 = .7854$, i.e. a difference of about 3^o.

Figure 3.1: Gaussian Sea

We now turn to Gauss' fourth proof, for which we will adapt the discussion given in Uspensky. Assume $f(z)$ is a monic complex polynomial of degree 1 or greater. Consider the set of complex z such that the real part of $f(z)$ is positive. Uspensky calls this set the land, the complement the sea (see for example Figure 3.1 which is a computer drawn version of Uspensky's hand drawn example, here the land is white.) The boundary between the land and the sea is called the seashore. The seashore is defined by the condition that for all z on the seashore $Re\ f(z) = 0$. Thus it is enough to show that there is a point on the seashore where $Im\ f(z) = 0$, for if both the real and imaginary parts are 0 then $f(z) = 0$. But on the edge of the picture $f(z)$ is approximated by z^n (in this example z^5). Thus the points in which the seashore hits the boundary of the picture are approximately where z^n has real part 0. But $Re(z^n) = 0$ for all points with argument $\pi/2n + k\pi/n, k = 0, 1, \ldots, 2n - 1$ (there are always $2n$ of them) and the imaginary part of z^n alternates from positive to negative as we go around the boundary. If we start at a point where the seashore meets the boundary (in the general case we think of the boundary as some large square or circle with center at the origin) where going counterclockwise $Re\ f(z)$ is changing from positive to negative then the land (facing inwards to the center) is to our left. Further, at such points $Im\ f(z) > 0$ because $Im(z^n)$ is. Walking along the seashore keeping the land to the left we always reappear at the boundary an odd number of shorelines away. But here $Im(z^n) < 0$ so $Im\ f(z) < 0$. Since $Im\ f(z)$ is continuous on on the seashore (all over, in fact) going from positive to negative we must have crossed a point on the seashore where $Im\ f(z) = 0$.

Some comments on this proof are in order. The proof rests on the fact that the seashore is well behaved and really looks like the seashore as pictured in the example. Gauss does go to great lengths to show that the seashore is an algebraic curve. An algebraic curve is the solution set in the plane of an algebraic equation in the two variables x, y such as a line, a branch of a hyperbola or something of higher degree. At Gauss' time the theory of algebraic curves was not well enough developed to prove that going in along some branch one would necesssarily come out again. In fact, if the boundary was not an algebraic curve that might not happen. Gauss recognized this as a sticky point and promised (but never delivered) a proof, instead he stated "nobody, to my knowledge, has ever doubted it." The fact that $Im\ f(z)$ cannot change from positive to negative without crossing a point where it is zero follows from the fact that this piece of seashore is "connected." Here is a part of the proof where we get into "topology."

Maple Implementation

If you would like to draw Gaussian Seas you can use MAPLE. First define a polynomial, for example the sea above comes from $f := z^5 - 5z^4 + 9z^3 - 5z^2$. Then define a short procedure to evaluate the real part of $f(x + iy)$. For example

```
G := proc(x,y) local u;
  u := subs(z=x + y*I, f);
  RETURN(evalf(evalc(Re(u)));
end;
```

Now draw a contour plot using the command

```
plot3d(G(x,y), x=a..b, y=c..d, contours=[0]);
```

When you get the plot, rotate it so that you are looking from directly above and select the *patch and contour* option or just the *contour* option. Don't expect MAPLE to color the picture as Figure 3.1, that was produced by using MAPLE to find the contours (the shorelines) and using a graphics program to color it.

Exercise 3.2.1 (various point values) Draw the Gaussian Sea for various polynomials. Use MAPLE to find the contours and use a pencil or crayons to color in the seas.

3.3 Topological Proof

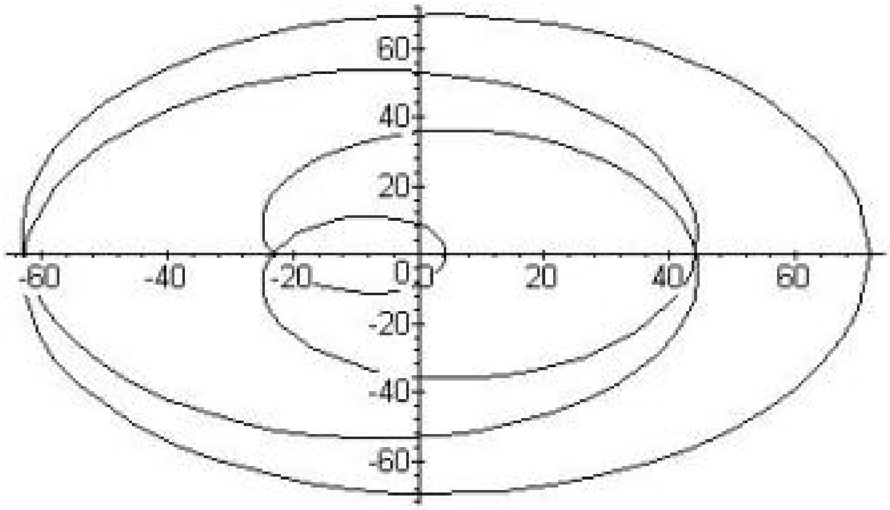
We now sketch another proof which uses topology. This proof uses a different property of curves. The origin of this proof is unclear, at any rate it appears in Birkhoff and MacLane's book, among other places. The advantage of this proof is that it most clearly shows the topological basis of the FTA.

Again we start with some preliminary ideas. Let $I = \{t | a \leq t \leq b\}$ be a closed interval of the real line. A *curve* is a continuous function from I to the set of complex numbers. A curve is *closed* if it starts and ends at the same place, i.e. if the curve is given by a function $\tau : I \to \boldsymbol{C}$ then τ is closed if $\tau(a) = \tau(b)$. For example, let $a = 0, b = 2\pi$ and R any non-zero real number. Then $\tau(t) = R(\cos t + i \sin t)$ is a closed curve, mainly the circle of radius R about the origin travelled counterclockwise.

The topological fact we will use is that every closed curve that does not pass through the origin has a *winding number* about the origin. Intuitively, if you travel all the way around the curve the winding number is the number of times you go around the origin counterclockwise. It is a standard theorem of topology that the winding number exists.

Now let $f(z)$ be a monic polynomial of degree $n \neq 0$. Let τ_R be the curve traversing the circle of radius R counterclockwise about the origin as above and $\Gamma_R(t) = f(\tau_R(t))$ be the closed curve travelled by $f(z)$ as z travels about the circle. Pick R_0 to be large so that $f(z)$ is approximated by z^n as in Theorem 3.2.1. We will denote the curve Γ_R for $R = R_0$ by Γ_0. Then $f(z)$ is not 0 on Γ_0 so the winding number of Γ_0 is defined. But it is easy to see that the winding number of $\tau_R(t)^n = R(\cos nt + i \sin nt)$ is n and since $f(z)$ is close to z^n on the circle of radius R_0 it follows that the winding number of Γ_0 is also n. We now argue by contradiction, suppose $f(z)$ has no roots. Then as we vary R from R_0 to 0 each of the curves Γ_R has a winding number. But for values of R close together the curves Γ_R must have the same winding number (here we are using topology again). Thus the winding number of Γ_R is independent of R. But when $R = 0$ the curve Γ_R is a (non-zero by assumption) constant and so has winding number 0. But this contradicts that the fact that the winding number was supposed to remain constant. Thus our assumption that $f(z)$ has no complex roots was wrong and we have proven the FTA.

Figure 3.2: Winding number = 4



It is interesting to note that one can prove using the methods of complex analysis that if τ is a simple closed counterclockwise curve in the complex plane (simple means that it does not cross itself, eg. a circle) and $f(z)$ is a polynomial (or other differentiable complex function) then the winding number of $f(\tau(t))$ is m where m is the sum of the multiplicities of all the zeros of $f(z)$ inside the curve τ.

Maple Implementation

This proof lends itself well to illustration by Maple. To plot the curve Γ_R first define a polynomial, for example the polynomial $f := z^4 + 2z^3 + z - 1$ is a good first choice. Then define a function $G(t, r)$ as follows (we define it in arrow notation):

```
G:= (t,r)-> evalf(subs(z=r*cos(t)+I*r*sin(t), f));
```

Now you can plot the curve, say Γ_2 as follows

```
plot([evalc(Re(G(t,2))), evalc(Im(G(t,2))), t=0..2*Pi]);
```

You should try this for several values of R, the Figure above uses this polynomial and $R = 2.5$.

A more interesting way to view this proof is to use the maple function `animate`. Before using `animate` don't forget the command `with(plots, animate);` to get the procedure from the plotting library. Then use

```
animate([evalc(Re(G(t,r))), evalc(Im(G(t,r))),
    t=0..2*Pi],r=a..b);
```

where a, b are the minimum and maximum values of r you wish to use. You will see the curve growing (or shinking) before your eyes. For the example indicated above you might try the ranges `0..1` , `1..2` and `3..10` . Note that the winding number changes when the curve passes through the orgin. The value of $z = \tau_R(t)$ at the point where the curve passes through the origin is a root of $f(z)$. Note that these graphics do not tell us what the roots are but we can estimate the modulus of the roots by figuring out which curves Γ_R pass through the origin.

3.4 Analytic Proof

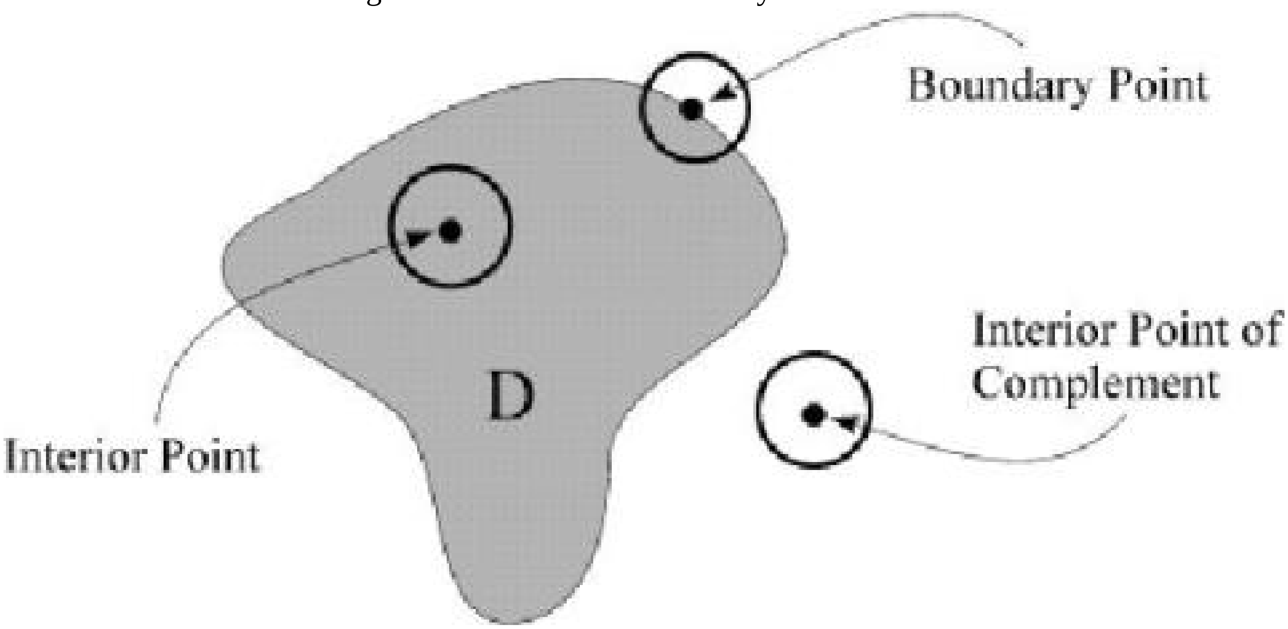
There are a number of proofs of the FTA that come from complex analysis, the first being Gauss's third proof. While most of these proofs are short they require the deep theorems (eg. Cauchy's Integral Theorem) derived in complex analysis. In this section we present one which can be proved quite simply in the special case of polynomials. Of our four proofs this one comes the closest to being complete, the missing steps being found in any advanced calculus textbook. Although the ideas behind this proof are attributed to D'Alembert, the proof itself is quite different from the real number based proof he wanted.

This time we start by recalling some definitions and theorems of advanced calculus or elementary metric topology.

Given a complex number α and positive real number r, the open disk of radius r about α is the set $\{z \in \boldsymbol{C} | \, |z - \alpha| < r\}$ of all complex numbers whose distance from α is less than r.

Let D be some set of complex numbers. A point α is in the *interior* of D if there is $r > 0$ so that the disk of radius r about α is completely contained in D. A point in the interior of D is also called an *interior point* of D.

Figure 3.3: Interior and Boundary



A point of $\boldsymbol{C}$, not necessarily in D, is called a *boundary point* of D if it is neither an interior point of D nor of the complement of D. Thus every disk about a boundary point contains both points in D and points not in D.

The set D is called *open* if every point of D is an interior point of D. D is called *closed* if every boundary point of D is actually a point of D. Alternatively, D is closed if its complement is open.

A set of complex numbers is called *bounded* if there is a positive real number M so that $|z| \leq M$ for all $z \in D$. A subset of $\boldsymbol{C}$ which is closed and bounded is called *compact*.

From a strictly analytic point of view the complex plane is no different from the real plane, thus the following theorem follows from a standard theorem which can be found in any advanced calculus textbook:

Theorem 3.4.1 *Let $f : D \to \boldsymbol{C}$ be a continuous function where D is a compact subset of the complex numbers $\boldsymbol{C}$. Then $|f(z)|$ attains a maximum and minimum value on D, i.e. there are points α, β in D so that $|f(\alpha)| \leq |f(z)| \leq |f(\beta)|$ for all $z \in D$.*

We will prove the following general theorem which gives a surprising property of polynomial functions in the complex plane, one which does not hold for real polynomials. We will later deduce the FTA from this theorem. The young reader is warned that the rest of this section contains $\epsilon's$, $\delta's$ and other Greek letters!

Theorem 3.4.2 (Minimum Modulus Theorem) *Let $f(z)$ be a complex polynomial of degree 1 or greater and D a compact subset of $\boldsymbol{C}$. Suppose that $|f(z)|$ attains its minimum value on D at α. Then either $f(\alpha) = 0$ or α is a boundary point of D.*

Proof: The conclusion of the theorem is logically equivalent to the statement that if c is an interior point of D, $f(c) \neq 0$ then there is a point $z \in D$ so that $|f(z)| < |f(c)|$. Thus we may prove this statement.

So assume c is an interior point of D with $f(c) \neq 0$. Let the Taylor's expansion of $f(z)$ about c be

$$f(z) = b_0 + b_1(z-c) + b_2(z-c)^2 + \cdots + b_n(z-c)^n.$$

Since $b_0 \neq 0$ we write b_0 = $f(c)$ in polar form as $b_0 = \epsilon(\cos\phi + i\sin\phi), \epsilon = |f(c)| > 0$. Now let k be the smallest positive integer such that $b_k \neq 0$, such a k exists because $f(z)$ is not a constant polynomial. Let $b_k = \rho(\cos\sigma + i\sin\sigma), \rho > 0$. For each positive real number t we let

$$z_t = c + t(\cos((\pi + \phi - \sigma)/k) + i\sin((\pi + \phi - \sigma)/k)).$$

The proof will be complete when we show that for small $t, z_t \in D$ and $|f(z_t)| < |f(c)|$.

Now $|z_t - c| = t$ and since c is by assumption an interior point of D there is a positive number δ_1 so that if $t < \delta_1$ then $z_t \in D$. Secondly, we can assume $t < 1$ so the larger j is the smaller $|(z_t - c)^j| = t^j$ is, and this difference becomes more pronounced the smaller t is. Thus we can find δ_2 so that if $t < \delta_2$ then

$$|b_{k+1}(z_t - c)^{k+1} + \cdots + b_n(z_t - c)^n| < \rho t^k/2$$

. Finally let δ_3 be such that if $t < \delta_3$ then $\rho t^k < \epsilon$. Now let $\delta = \min(\delta_1, \delta_2, \delta_3)$. Then for $t < \delta$, we have $z_t \in D$ and

$$\begin{aligned}
&b_0 - b_k(z_t - c)^k \\
&= \epsilon(\cos\phi + i\sin\phi) + \rho(\cos\sigma + i\sin\sigma) * t^k(\cos(\pi + \phi - \sigma) + i\sin(\pi + \phi - \sigma)) \\
&= \epsilon(\cos\phi + i\sin\phi) + \rho t^k(\cos(\pi + \phi) + i\sin(\pi + \phi)) \\
&= \epsilon(\cos\phi + i\sin\phi) - \rho t^k(\cos\phi + i\sin\phi) \\
&= (\epsilon - \rho t^k)(\cos\phi + i\sin\phi)
\end{aligned}$$

using DeMoivre's law and complex multiplication. Thus $|b_0 + b_k(z_t - c)^k| = \epsilon - \rho t^k$. By the triangle inequality

$$|f(z_t)| \leq |b_0 + b_k(z_t - c)^k| + |b_{k+1}(z_t - c)^{k+1} + \cdots + b_n(z_t - c)^n|$$
$$\leq \epsilon - \rho t^k + \rho t^k/2 = \epsilon - \rho t^k/2 < \epsilon = |f(c)|.$$

This completes the proof.

This proof is in the style of "hard analysis" but the crux of the proof is just that near c the function $f(z)$ is approximated closely by the function $b_0 + b_k(z - c)^k$. This latter funtion we can work with explicitly to make the term $b_k(z - c)^k$ small and of argument opposite that of b_0. Thus we see this function will take values with moduli smaller than b_0 near c.

Note that if $k > 1$ in the proof then c is a critical point of $f(z)$. This is the difference between the real and complex case. In the real case we are restricted to moving in two (opposite) directions around the critical point and this may not be enough to make $b_0 + b_k(z - c)^k$ smaller than b_0 (unlike when $k = 1$ where these two directions are enough.) Thus the function may have a minimum at the critical point. In the complex case we can move in all directions around the critical point so our function (or more precisely, its modulus) will not have a minimum at a critical point in the interior of the domain unless $f(c) = 0$. Thus the only alternative for the minimum is on the boundary of the domain.

We now deduce the FTA from the Minimum Modulus Theorem. Let $f(z)$ be a polynomial of degree $n > 0$. For z of large modulus $f(z)$ is approximated by z^n so that the relative error is less than 50% i.e. $|f(z)| > |z^n|/2$. But for $|z|$ large we have $|z^n| > 2|f(0)|$. Thus if R is a large positive number we have for $|z| = R$, $|f(z)| > |f(0)|$. Let $D = \{z||z| \leq R\}$ be the closed disk about the origin of radius D. For z on the boundary $|f(z)| > |f(0)|$ so $|f(z)|$ does not obtain its minimum on the boundary. By the Minimum Modulus Theorem $f(\alpha) = 0$ for some $\alpha \in D$.

The following Exercise/Example/Maple-implementation is recommended to understand this proof.

Exercise 3.4.1 In Maple consider the functions

```
f:= z^3-2*z+5;
A:= (x,y)->abs(subs(z=x+I*y,f));
```

The second function $A(x, y) = |f(x + i * y)|$ gives the modulus of the first polynomial as a function of the real and imaginary part of z respectively.

First do

```
plot3d(A(x,y), x=-1.5..1.5, y=-1..1);
```

and use the `style=patch, color=Z[HUE]` settings. Can you find the points of minimum modulus? Where are they?

Then do

```
plot3d(A(x,y), x=-2.5..2.5, y=-2..2);
```

with the same settings. Can you see the roots? Write up your findings.