BULLETIN OF THE
GREEK MATHEMATICAL SOCIETY
Volume 53, 2007 (131–134)

The fundamental theorem of algebra: an elementary proof

Takis Sakkalis

Received 23/06/2005 Accepted 10/11/2005

Abstract

In this note an elementary proof of the fundamental theorem of algebra is presented. The proof is self-contained and is based on the Cauchy-Riemann equations for complex polynomials, Lemma 1.1, and on the second derivative test for real functions of two variables.

1. Introduction and the Result

Polynomials are perhaps the most simple and useful class of functions that appears in mathematics. Apart from their theoretical interest, they have enjoyed considerable attention over the last two decades in applied areas such as Computer Algebra, Computational Algebraic Geometry, etc. A polynomial $p(z)$ of degree n is a function of the form $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$, where $a_i \in \mathbf{C}$, $a_n \neq 0$. A root of $p(z)$ is a number $z_0 \in \mathbf{C}$ such that $p(z_0) = 0$. Observe that if z_0 is a root of $p(z)$, then $p(z)$ is divisible by $z - z_0$. In that regard, we say that z_0 has multiplicity k if $p(z) = (z - z_0)^k Q(z)$, where $Q(z)$ is a polynomial with $Q(z_0) \neq 0$.

Throughout this note we use the notation $z = x+iy = (x, y)$ for a complex number z. Similarly, for a polynomial $p(z) = p(x + iy) = f(x, y) + ig(x, y); f, g$ are the real and imaginary parts of p. The main idea of the proof is to analyze the (real) critical points of the function $F : \mathbf{R}^2 \to \mathbf{R}, F(x, y) = f^2(x, y) + g^2(x, y) = |p(z)|^2$. It turns out that if p and p' have no common roots, F has two kinds of critical points: (1) minima and (2) saddle points. The existence of at least one minimum, which in turn is necessarily a root of $p(z)$, completes the proof.

We begin with the following result whose first part gives an inductive proof of the remarkable Cauchy-Riemann equations for polynomials.

Lemma 1.1 *Let $p(z) \in C[z]$ be a polynomial of positive degree n. We write $p(z) = p(x + iy) = f(x, y) + ig(x, y)$ and $p'(z) = p'(x + iy) = q(x, y) + ir(x, y)$. Then,* *1* $\dfrac{\partial f}{\partial x} = \dfrac{\partial g}{\partial y}$ *and* $\dfrac{\partial f}{\partial y} = -\dfrac{\partial g}{\partial x}$ *[Cauchy-Riemann equations]*

2 $q = \dfrac{\partial f}{\partial x}$ *and* $r = \dfrac{\partial g}{\partial x}$

Proof. We first consider the special case where $p(z) = z^n$.

1. The proof will be inductive on n. For $n = 1$, then $p(z) = z = x + iy$ and thus the result holds. Assume that the result holds for $z^k = a(x, y) + ib(x, y)$ and consider $z^{k+1} = z^k \cdot z = [a(x, y) + ib(x, y)] \cdot (x + iy) = (ax - by) + i(ay + bx) = P(x, y) + iQ(x, y)$. In that case we have

$$\begin{aligned}\frac{\partial P}{\partial x} &= \frac{\partial a}{\partial x}x + a - \frac{\partial b}{\partial x}y \\ &= \frac{\partial b}{\partial y}x + a + \frac{\partial a}{\partial y}y \quad \text{[by the induction step]} \\ &= \frac{\partial Q}{\partial y}\end{aligned}$$

Similarly we can show that $\partial P/\partial y = -\partial Q/\partial x$.

2. Again, we will use induction on n. If $n = 1$, the result trivially holds. Suppose now that $p(z) = z^{k+1} = z^k \cdot z = [a(x, y) + ib(x, y)] \cdot (x + iy) = (ax - by) + i(ay + bx) = P(x, y) + iQ(x, y)$. Then $\partial P/\partial x = a + x\partial a/\partial x - y\partial b/\partial x = a + x\partial a/\partial x + y\partial a/\partial y = a + k\,a = (k + 1)a$, since $a(x, y)$ is a homogeneous polynomial of degree k. But $p'(z) = (k + 1)z^k = (k + 1) \cdot (a + ib)$ and thus $\partial P/\partial x = q$. Similarly we have $r = \partial g/\partial x$.

Finally, since partial derivatives behave well with respect to addition and scalar multiplication and a polynomial is nothing but a finite sum of constant multiples of powers of z^n, the proof is now complete. □

Theorem 1.2 [Fundamental Theorem of Algebra] *Every non constant monic polynomial* $p(z) = z^n + a_1 z^{n-1} + \cdots + a_n$*, where* $a_i \in \mathbf{C}$*, has a root in* $\mathbf{C}$*.*

Proof. The proof will be inductive on the degree n of $p(z)$. If $n = 1, p(z) = z + a_0$ and thus $p(-a_0) = 0$. By induction, we may assume that p and p' have no common roots. We again write $p(z) = p(x + iy) = f(x, y) + ig(x, y), p'(z) = p'(x + iy) = q(x, y) + ir(x, y)$, where f, g, q, r are real polynomials. Now consider the function $F = f^2 + g^2 = |p(z)|^2$. Observe that, $\lim_{|z|\to\infty} |p(z)| = \infty$ and thus $\lim_{|(x,y)|\to\infty} F(x, y) = \infty$. Also, in view of Lemma 1.1 we see that

$$2p(x + iy)\,\overline{p'}(x + iy) = 2f\frac{\partial f}{\partial x} + 2g\frac{\partial g}{\partial x} + i\left(2f\frac{\partial f}{\partial y} + 2g\frac{\partial g}{\partial y}\right) = \frac{\partial F}{\partial x} + i\frac{\partial F}{\partial y}$$

where $\overline{p'}$ is the complex conjugate of p'. Thus, the (real) critical points of F are precisely the roots of p and p', and therefore by induction are finite in number.

Let $C > 0$ be so that C is greater than any critical value of F. Consider the set $A = \{(x, y) \mid F(x, y) \leq C\}$. Since $\lim_{|(x,y)|\to\infty} F(x, y) = \infty$, A is a non empty compact subset of $\mathbf{R}^2$, and thus F has a minimum $m_0 \in \mathbf{R}^2$; we will show that m_0 is a root of $p(z)$. To achieve that, we will prove that every root of p' is not a *minimum* of F. First, note that for any $z = (x, y)$ the Hessian determinant

$$H(F)(z) = \begin{vmatrix} F_{xx}(z) & F_{xy}(z) \\ F_{yx}(z) & F_{yy}(z) \end{vmatrix} = 4\left(|p'(z)|^2 - |p(z)|^2\,|p''(z)|^2\right) \tag{1}$$

Suppose, now, that z_0 is a root of p', and let $m = k - 1 \geq 1$ be its multiplicity. By replacing z with $z - z_0$, we may assume that $z_0 = 0$. If $m = 1$, formula (1) shows that $H(F)(z_0) = -4\left(|p(z_0)|^2\,|p''(z_0)|^2\right) < 0$ and thus from the second derivative test for a function of two variables we get that z_0 is a saddle point of $F(x, y)$. If $m > 1$, then $p(z)$ must have the form $p(z) = z^k Q(z) + c$, where $Q(0) \cdot c \neq 0$. In that case, $F(x, y) = |p(z)|^2 = p(z)\,\overline{p(z)} = (z\bar{z})^k Q\overline{Q} + z^k Q\bar{c} + c\bar{z}^k\overline{Q} + c\bar{c}$. Then, the Taylor series of F around $(0, 0)$ takes the form

$$F(x, y) = c\bar{c} + z^k a\bar{c} + \bar{z}^k\bar{a}c + R(x, y)$$

where $a = Q(0)$ and $R(x, y)$ is a polynomial whose lowest degree term is of degree greater than k. We are going to show that if U is any neighborhood of $(0, 0)$, $F(x, y) - c\bar{c}$ takes on positive, as well as negative values in U. First, we shall need the following. Let $\delta > 0$ and let R_m^δ be the maximum value of $|R(x, y)|$ on the circle $K = \{z \mid z = \delta e^{it}\}$. Then, for each $z \in K$, $R(z) = \delta^{k+1} R^*(\delta, t)$. Thus, if $M^*(\delta)$ is the maximum value of $|R^*(\delta, t)|$ on K, we have $R_m^\delta = \delta^{k+1} \cdot M^*(\delta)$.

Let $w = a\bar{c} + \bar{a}c$ and $S(z) = z^k a\bar{c} + \bar{z}^k\bar{a}c$. Note that $w, S(z) \in \mathbf{R}$. For $b > 0$, denote by $b_k = \sqrt[k]{b}$. We now consider the following cases:

- $w \neq 0$. Pick $\varepsilon > 0$ so that $|w| > \varepsilon_k M^*(\varepsilon_k)$. Then, if u_1, u_2 are k-th roots of $\varepsilon, -\varepsilon$, respectively, we have $S(u_1) = \varepsilon w$ and $S(u_2) = -\varepsilon w$, and $|S(u_i)| = \varepsilon|w| > \varepsilon\varepsilon_k M^*(\varepsilon_k) = (\varepsilon_k)^{k+1} \cdot M^*(\varepsilon_k) = R_m^{\varepsilon_k} > |R(u_i)|$, $i = 1, 2$. Thus, $-|S(u_1)| < R(u_1) < |S(u_1)|$ and $-|S(u_2)| < R(u_2) < |S(u_2)|$. Therefore,

$$[S(u_1) + R(u_1)] \cdot [S(u_2) + R(u_2)] < 0 \tag{2}$$

- $w = 0$. Let $w_1 = a\bar{c} - \bar{a}c$. Then w_1 is a non zero purely imaginary number. Pick $\varepsilon > 0$ so that $|w_1| > \varepsilon_k M^*(\varepsilon_k)$. Then, if b_1, b_2 are k-th roots of $i\varepsilon, -i\varepsilon$, respectively, we have $S(b_1) = i\varepsilon w_1$ and $S(b_2) = -i\varepsilon w_1$, and $|S(b_i)| = \varepsilon|w_1| > |R(b_i)|$, $i = 1, 2$. Thus, we get

$$[S(b_1) + R(b_1)] \cdot [S(b_2) + R(b_2)] < 0 \tag{3}$$

Finally, let $D = \{z \mid |z| < r\}$ be an open disk centered at 0 of radius $r > 0$. Let, also, $\varepsilon > 0$ so that $\varepsilon_k < r$ and $|w| + |w_1| > 2\,\varepsilon_k M^*(\varepsilon_k)$. Then, using inequalities (2) or (3) we can find $z_1, z_2 \in D$ so that $S(z_1) + R(z_1) < 0$ and $S(z_2) + R(z_2) > 0$; that is

$$F(z_1) - c\bar{c} < 0 \quad \text{and} \quad F(z_2) - c\bar{c} > 0 \tag{4}$$

Thus, from (4) we get

$$F(z_1) < c\bar{c} = F(0,0) < F(z_2)$$

This proves that $z_0 = (0,0)$ cannot be a *minimum* of F. Therefore, m_0 has to be a root of p. □

◇ Takis Sakkalis
Mathematics Laboratory
Agricultural Univeristy of Atehns
Athens, 118 55 GREECE
stp@aua.gr