d'Alembert's Lemma

France Dacar, Jožef Stefan Institute

July 11, 2007

d'Alembert's Lemma for Polynomials and The Fundamental Theorem of Algebra

The proof of the 'fundamental theorem of algebra'[1] that we shall present is a descendant of the d'Alembert's proof of this theorem; we can say that it *is* the d'Alembert's proof as it is known today. The proof which d'Alembert gave in 1746 had serious weaknesses. It was not until the 1870s, when real numbers and continuity were properly defined, and the basic properties of continuous functions were rigorously proved, that the gaps in the d'Alembert's proof were finally filled in. By then, and also later on, the proof was modified time and again, but the leading idea of the proof withstood the trials of time; it is simplicity itself:

Let p be a nonconstant polynomial function of a complex variable. Then there exists a closed disc, centered at the origin, such that for each point z in the disc at which $p(z) \neq 0$ there exists another point z_1 in the disc with $|p(z_1)| < |p(z)|$. The function $|p|$ attains the minimum value at some point in the disc, and this minimum value must be 0.

The original proof of the the first statement was unduly complicated and not entirely correct, while the proof of the second statement rested on quite shaky grounds; both parts of the proof had to await clarification and consolidation of the notions of real numbers and continuity of functions before they could be made rigorous. And here is the entire proof, as we know it now,[2] laid out in three stages:

Let p be a nonconstant polynomial function of a complex variable.

(1) D'ALEMBERT'S LEMMA. If $p(a) \neq 0$, then every neighborhood of the point a contains a point z such that $|p(z)| < |p(a)|$.

Proof. The nonconstant polynomial function $p(a+w)$ of the complex variable w can be represented as

$$p(a+w) \;=\; p(a) + cw^m(1 + r(w))\,,$$

where the constant c is not 0, $m \geqslant 1$ is an integer, and r is a polynomial function with $r(0) = 0$. There exists $\varrho > 0$ such that $|cw^m| < |p(a)|$ and $|r(w)| < 1$ whenever $|w| < \varrho$. Let ε be any real number in the range $0 < \varepsilon < \varrho$. There exists w with $|w| = \varepsilon$ for which $cw^m = -\delta p(a)$, where $0 < \delta = |c|\,\varepsilon^m/\,|p(a)| < 1$; then $|p(a+w)| = |(1-\delta)p(a) - \delta p(a)\cdot r(w)| < (1-\delta)\,|p(a)| + \delta\,|p(a)| = |p(a)|$.

(2) Let $n \geqslant 1$ be the degree of the polynomial function p. Since $z^{-n}p(z)$ approaches the nonzero leading coefficient of p when $|z|$ increases to the infinity, there exists $R > 0$ such that $|p(z)| > |p(0)|$ for every point z on the circle $|z| = R$.

(3) Because the closed disc $|z| \leqslant R$ is a compact subset of the complex plane, the continuous function $|p|$, restricted to the disc, attains the minimum value at some point z_0 in the disc. Because of (2), z_0 is an interior point of the disc, and by d'Alembert's lemma, $|p(z_0)| = 0$.

[1] Some mathematical wit commented that it is dubious whether this famous theorem is really fundamental, that it is not always a theorem because it sometimes serves as a definition, and that in its classical form it is not of algebra but of analysis.

[2] A condensed version of the Cauchy's proof of the Fundamental Theorem of Algebra.

So this is the proof of the fundamental theorem of algebra in eighteen lines: short and sweet. However, it is short because it presumes that some things are already known. Below is a possible list of facts which, taken together, will put the d'Alembert's proof on solid grounds:

(1) If p is a polynomial function of complex variable and a is a complex number, then $p(a+w)$ is a polynomial function of the complex variable w. This seems obvious since manipulations with polynomials are so familiar to us that we barely notice them as something that had to be proved correct at some point. The rearrangement of $p(a+w)$ as a polynomial in w is in fact the main ingredient of algebra in the proof.

(2) Polynomial functions of a complex variable are continuous.

(3) Any complex number of absolute value 1 has a m-th root for any integer $m \geqslant 1$.

(4) Any comples numbers z and w satisfy the equality $|zw| = |z|\,|w|$ and the inequality $|z+w| \leqslant |z| + |w|$.

(5) If n is the degree of a polynomial function p of a complex variable, then $z^{-n}p(z)$ converges to the leading coefficient of p as $|z| \to \infty$.

(6) The absolute value of a complex number is continuous (follows from (4)).

(7) The inclusion map of a subspace into a topological space is continuous (by the definition of a subspace).

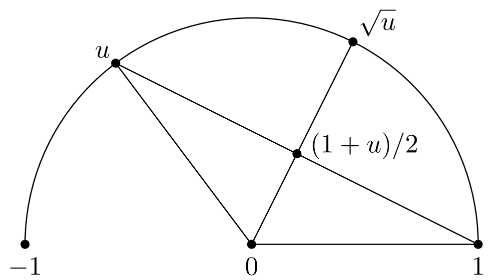
(8) The composition of continuous functions is continuous.

(9) A closed disc in the complex plane, with the topology induced from the topology of the complex plane, is a compact topological space.

(10) If X is a compact topological space and $f\colon X \to \mathbb{R}$ is continuous, then there exists in X a point x_0 such that $f(x_0) \leqslant f(x)$ for every point x of X.

The fact (3) is easy once we have at our disposal the surjective homomorphism $t \mapsto e^{it}$ of the additive group of real numbers onto the multiplicative group of complex numbers of absolute value 1. But such an assistance of the transcedental function exp somehow tarnishes the brilliance of the proof; can we get by without it? In fact we can, and here is how.

It suffices to prove the existence of a m-th root, for each integer $m \geqslant 1$, just for a complex number $u = c + si$ ($c, s \in \mathbb{R}$) on the upper half of the unit circle in the complex plane. The square root of u is easy to determine:



We have

$$\sqrt{u} \;=\; \frac{1+u}{|1+u|} \;=\; \frac{1+c+si}{\sqrt{2(1+c)}} \;=\; \sqrt{\frac{1+c}{2}}+i\sqrt{\frac{1-c}{2}}\,,$$

where the first two expressions are undefined at $u=-1$, while the third is defined even there and gives a correct result. Since our derivation relied on a geometric sketch, one may wonder whether the three expressions above really give the square root of u; to dispel any doubt, just square, say, the third expression.

This disposes of all the 2^k-th roots. Since the integer m can be written as $m = 2^k m'$ with m' odd, and a m'-th root of a 2^k-th root of u is a m-th root of u, it remains to deal with the case when m is odd.

For any point $x+yi$ ($x, y \in \mathbb{R}$) on the unit circle, the real component of $(x+yi)^m$ is determined by x:

$$\Re\big((x+yi)^m\big) \;=\; T_m(x)\,,$$

where T_m is the m-th Chebyshev polynomial (of the first kind),

$$T_m(x) \;=\; \sum_{0\leqslant 2k\leqslant m} \binom{m}{2k} x^{m-2k}(x^2-1)^k\,.$$

Since $(-1)^m = -1$ (because m is odd) and $1^m = 1$, we see that $T_m(-1) = -1$ and $T_m(1) = 1$. Suppose we are given a point $c+si$ on the upper half-circle. The polynomial T_m is continuous, so there exists c', $-1 \leqslant c' \leqslant 1$, at which $T_m(c') = c$; put $s' = \sqrt{1-c'^2}$. We have either $(c'+s'i)^m = c+si$, or $(c'+s'i)^m = c-si$ in which case $(c'-s'i)^m = c+si$. Done.

d'Alembert's Lemma for Holomorphic Functions and The Maximum Modulus Theorem

The proof of d'Alembert's Lemma easily adapts to arbitrary holomorphic functions. The Maximum Modulus Theorem is a straightforward corollary of the generalized Lemma.

D'ALEMBERT'S LEMMA FOR HOLOMORPHIC FUNCTIONS. Let Ω be a region (connected open subset of the complex plane). Let f be a nonconstant holomorphic function on Ω, let a be a point in Ω and V any neighborhood of a in Ω. Then V contains a point z_+ at which $|f(z_+)| > |f(a)|$, and if $f(a) \neq 0$, V contains a point z_- at which $|f(z_-)| < |f(a)|$.

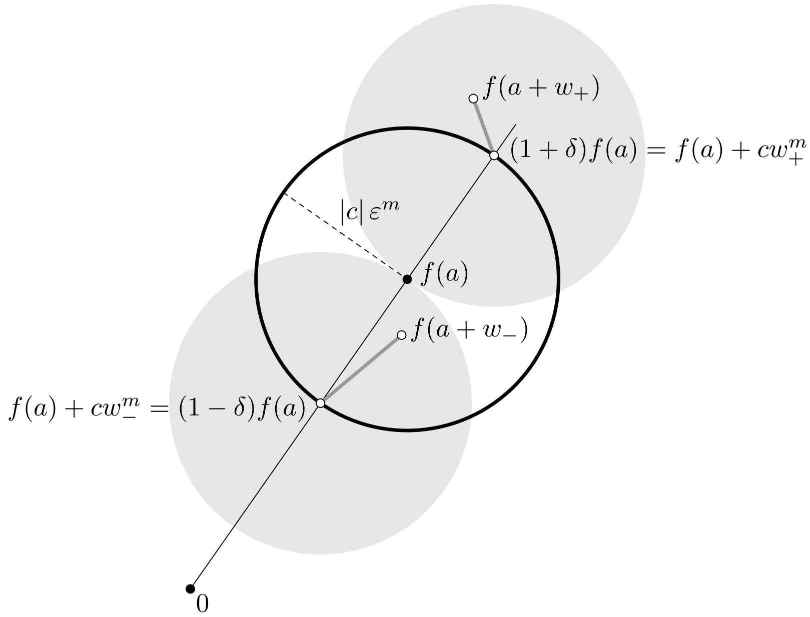
Proof. The nonconstant holomorphic function $f(a+w)$ of the complex variable w (varying in the region $\Omega - a$) can be represented as

$$f(a+w) \;=\; f(a) + cw^m(1+g(w))\,,$$

where c is a nonzero constant, $m \geqslant 1$ is an integer, and g is a holomorphic function on $\Omega - a$ with $g(0) = 0$. If $f(a) = 0$, then $f(a+w) \neq 0$ for all small enough $w \neq 0$, so there certainly exists a point $z_+ \in V$ at which $|f(z_+)| > 0 = |f(a)|$. From now on assume that $f(a) \neq 0$. There exists $\varrho > 0$ such that for every w with

$|w| < \varrho$ we have $a + w \in V$, $|cw^m| < |f(a)|$, and $|g(w)| < 1$. Let ε be any real number in the range $0 < \varepsilon < \varrho$. There exist w_+ and w_- with $|w_+| = |w_-| = \varepsilon$ for which $cw_+^m = \delta f(a)$ and $cw_-^m = -\delta f(a)$, where $0 < \delta = |c|\,\varepsilon^m / |f(a)| < 1$. Then $|f(a+w_+)| = |(1+\delta)f(a) + \delta f(a)\cdot g(w)| > (1+\delta)\,|f(a)| - \delta\,|f(a)| = |f(a)|$, and $|f(a+w_-)| = |(1-\delta)f(a) - \delta f(a)\cdot g(w)| < (1-\delta)\,|f(a)| + \delta\,|f(a)| = |f(a)|$.

The foregoing proof can be vividly retold as a story (a very short one) of a man walking his dog round a tree. The man, at $f(a) + cw^m$, is trodding a circular path of radius $|c|\,\varepsilon^m$ centered on the tree at $f(a)$. He has the dog, at $f(a+w)$, on an adjustable leash of the length $|cw^m g(w)|$ which is always shorter that $|c|\,\varepsilon^m$. The man walks round the tree m times, and during the walk he passes m times



through each of the points $(1+\delta)f(a) = f(a)+cw_+^m$ and $= (1-\delta)f(a) = f(a)+cw_-^m$. Whenever the man is at the point $(1+\delta)f(a)$, the dog is farther from the origin than the tree, and whenever he is at the point $(1-\delta)f(a)$, the dog is closer to the origin than the tree.

The Maximum Modulus Theorem. Let f be a nonconstant holomorphic function on a region Ω, and K a nonempty compact subset of Ω. Then

$$|f(a)| \;<\; \max_{z\in\partial K} |f(z)|$$

for every point a in the interior of K.

Proof. The continuous function $|f(z)|$ attains its maximum value at some point $z_0 \in K$, which is not an interior point of K, by d'Alembert's Lemma; thus $z_0 \in \partial K$, and $|f(z_0)| = \max_{z\in K} |f(z)| = \max_{z\in\partial K} |f(z)|$. If a is an interior point of K, there exists, by d'Alembert's Lemma, another interior point b of K with $|f(b)| > |f(a)|$, and hence $|f(a)| < |f(b)| \leqslant |f(z_0)| = \max_{z\in\partial K} |f(z)|$.

The following proposition about 'localization' of zeros of a holomorphic function is a straight rip-off of the final stage in the proof of the fundamental theorem of algebra.

Proposition. Let f be a holomorphic function on an open subset Ω of the complex plane, and let K be a compact subset of Ω. Suppose that K contains an interior point a such that

$$|f(a)| < |f(z)| \quad \text{for every } z \in \partial K\,.$$

If U is the connected component of the interior of K containing the point a, then f has a zero in U.

Proof. First assume that Ω is connected. Since the compact set K is not empty, its boundary is also not empty, thus $|f(a)| < |f(z)|$ for some z, so f is not constant. The continuous function $|f(z)|$ attains its minimum value at some point $z_0 \in K$. The point a demonstrates that $z_0 \notin \partial K$, so z_0 is interior to K; but then, by d'Alembert's Lemma, $f(z_0) = 0$.

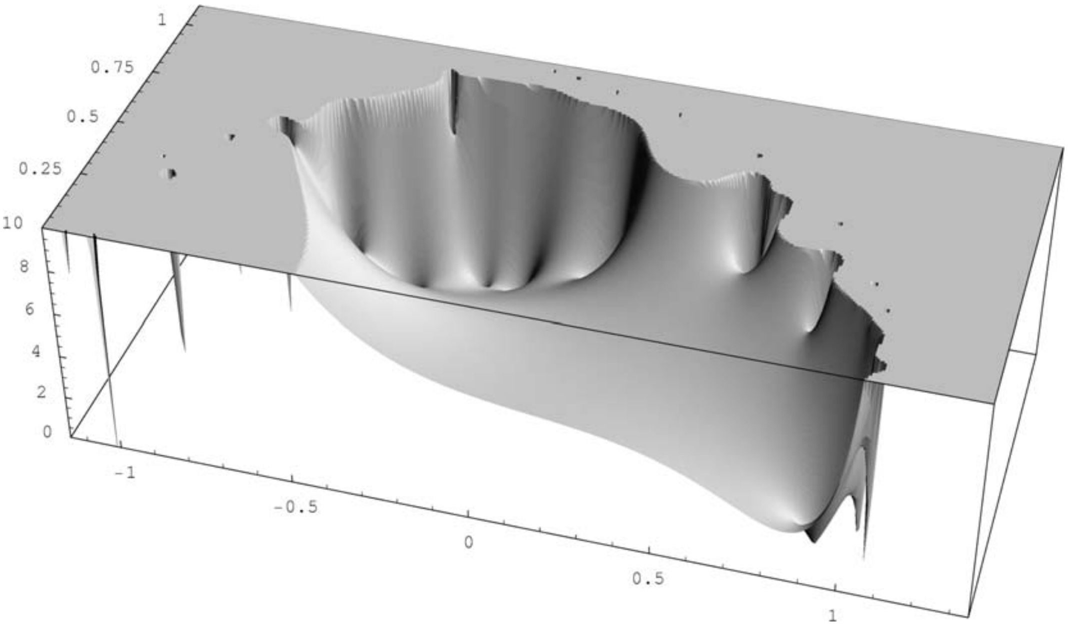
Now assume that Ω is just an open set in the complex plane. Let W be the connected component of Ω that contains the point a. Since the closure $\overline{U} \subseteq K \subseteq W$ is connected and contains point a, it is contained in W. The boundary $\partial\overline{U}$ is contained in ∂K because U is a connected component of the interior of K, hence U is the interior of $\overline{U}$. Now apply the first part of the proof to W and $\overline{U}$ in place of Ω and K.

We can reformulate the proposition in terms of the surface diagram of the function $|f|$ over Ω: every sinkhole in the surface goes all the way down to at least one zero of f. Thus by observing the surface diagram of a holomorphic function we can spot approximate locations of its zeros. This can be useful even with polynomials. Let us try this on some interesting polynomial, with many zeros. We choose the "look and say" polynomial:[3]

$$\begin{aligned} p(x) \;=\; & x^{71} - x^{69} - 2x^{68} - x^{67} + 2x^{66} + 2x^{65} + x^{64} - x^{63} - x^{62} - x^{61} - x^{60} \\ & - x^{59} + 2x^{58} + 5x^{57} + 3x^{56} - 2x^{55} - 10x^{54} - 3x^{53} - 2x^{52} + 6x^{51} \\ & + 6x^{50} + x^{49} + 9x^{48} - 3x^{47} - 7x^{46} - 8x^{45} - 8x^{44} + 10x^{43} + 6x^{42} \\ & + 8x^{41} - 5x^{40} - 12x^{39} + 7x^{38} - 7x^{37} + 7x^{36} + x^{35} - 3x^{34} + 10x^{33} \\ & + x^{32} - 6x^{31} - 2x^{30} - 10x^{29} - 3x^{28} + 2x^{27} + 9x^{26} - 3x^{25} + 14x^{24} \\ & - 8x^{23} - 7x^{21} + 9x^{20} + 3x^{19} - 4x^{18} - 10x^{17} - 7x^{16} + 12x^{15} + 7x^{14} \\ & + 2x^{13} - 12x^{12} - 4x^{11} - 2x^{10} + 5x^{9} + x^{7} - 7x^{6} + 7x^{5} - 4x^{4} + 12x^{3} \\ & - 6x^{2} + 3x - 6\,. \end{aligned}$$

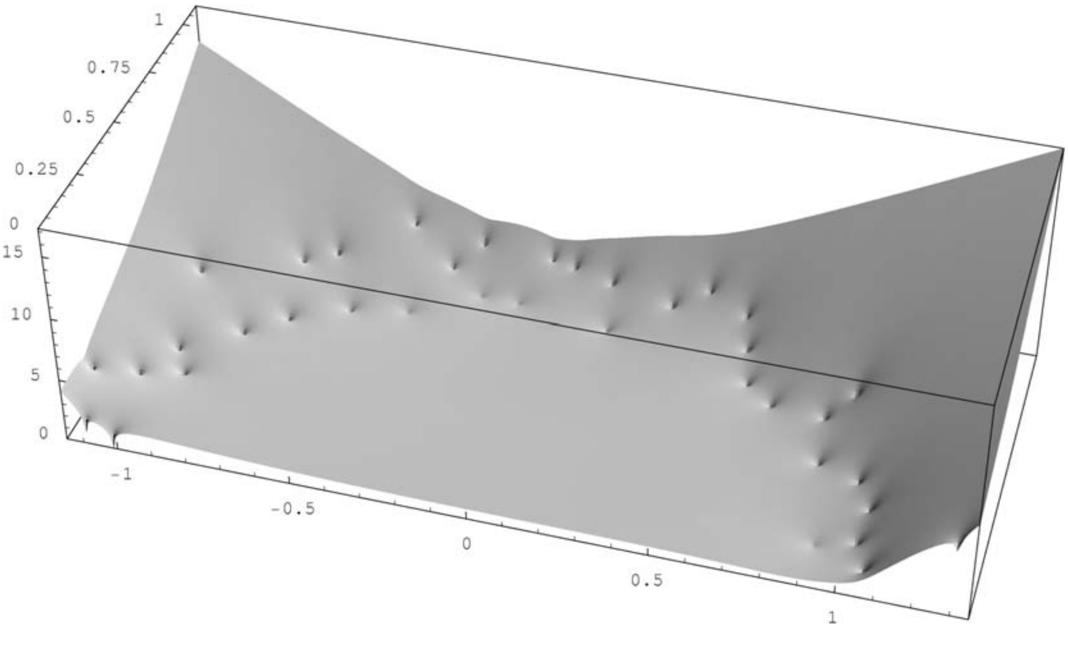
Its unique positive real zero $\lambda = 1.303577269\ldots$ is known as Conway's constant. Here is the surface diagram of $|p(x + yi)|$ over the rectangle $-1.15 \leqslant x \leqslant 1.35$, $0 \leqslant y \leqslant 1.1$, cut off at the elevation 10:

[3] See http://en.wikipedia.org/wiki/Look-and-say_sequence.



We can clearly see the sinkholes in the low part of the surface, but outside that we can detect some sinkholes only as small punctures in the cutoff plateau. Of certain zeros there is no trace. One such zero is λ, and there is good reason for it. The derivative of p at λ is $1.38 \cdot 10^8$, thus the sinkhole leading to the zero λ has at the elevation 10 a width $1.45 \cdot 10^{-7}$; since the surface is plotted on the grid of 361×159 points, it is no wonder that no grid point has managed to hit such a tiny target.

We can apply to $|p|$ some strictly increasing function $h\colon \mathbb{R}^{\geqslant 0} \to \mathbb{R}^{\geqslant 0}$ that preserves 0; the surface of $h(|p|)$ will have sinkholes and zeros precisely where the surface of $|p|$ has them. A good choice is $h(t) = \log(1+t)$. For small t, $h(t)$ is approximately t, thus at altitudes less than 0.5, say, the transformed surface will look almost the same as the original surface. On the other hand, very large values will be vigorously squashed down by h. Let us look at the surface of $\log_{10}(1+|p|)$, over the same region as above, this time with the full range of values shown:



Ha! We got a dimpled surface. Now we can see all the zeros: each dimple is the opening into a funnel leading down to a simple zero. Look at the zero λ. We see its funnel in cross-section, down to the altitude 5 or thereabouts, but lower than that there is nothing. This is quite realistic; at the scale the diagram is drawn, and at the resolution 300 dots per inch, the diameter of the funnel's stem below the altitude 5 is less than one dot.