Lietuvos matematikos rinkinys
Proc. of the Lithuanian Mathematical Society, Ser. A
Vol. 66, 2025, pages 11–15
https://doi.org/10.15388/LMR.2025.44493

Short proof of the fundamental theorem of algebra

Aivaras Novikas[a], **Juozas Juvencijus Mačys**[b]

[a] *Institute of Mathematics, Vilnius University, Lithuania*
[b] *Institute of Data Science and Digital Technologies, Vilnius University, Lithuania*
Corresponding author: aivaras.novikas@mif.vu.lt
E-mail: juozas.macys@mif.vu.lt

Received June 26, 2025; published December 21, 2025

Abstract. High school students may be introduced not only to important mathematical statements, such as the fundamental theorem of algebra, but also to their proofs. This famous theorem has various proofs, and its proofs have various modifications. The article discusses one such proof that is accessible to students.

Keywords: real polynomials; fundamental theorem of algebra; proof in school mathematics

AMS Subject Classification: 97H30; 12D05

1 Introduction

It is well known that a real polynomial

$$P(x) = \sum_{k=0}^{n} a_k x^k, \quad a_n \neq 0,\ n \geqslant 2. \tag{1}$$

is always divisible by a real polynomial of degree two. Therefore, $P(x)$ can always be factored as the product of real polynomials of degree one and two. This fact is known as *the fundamental theorem of algebra* (*FTA*) *for real polynomials*. Thus, to prove this theorem, it is sufficient to assume that polynomial (1) has no divisors $x - c$ with $c \in \mathbb{R}$ and prove that it is divisible by a real quadratic polynomial.

Although FTA, first rigorously proved in the 19th century, is not the most important statement of modern algebra, it is a significant fact about real polynomials, from which its no less significant version for complex polynomials can be easily derived.

Various proofs of FTA have been proposed [2]. However, they employ methods that are beyond the scope of school mathematics. In some cases, the proof can be simplified [1, 3, 4] and made accessible to inquisitive high school students while also providing them an opportunity for a deeper understanding of mathematics. Such a proof of FTA is presented here. After some preparatory work in Section 2, we prove FTA in Section 3. In Section 4, we provide some additional explanations that may be useful if this proof is presented to high school students.

2 Initial observations

We assume that the real polynomial (1) has no divisors of the form $x-c$, where $c \in \mathbb{R}$. We have to prove that it is divisible by a real quadratic polynomial. If FTA holds for $a_n^{-1}P(x)$, then it holds for $P(x)$ as well. Hence, we assume that $a_n = 1$.

We have $P(x) = x^n \cdot (1 + a_{n-1}x^{-1} + \cdots + a_0 x^{-n}) > 0$ for all real values of x sufficiently large. If $P(c) \leqslant 0$ for some $c \in \mathbb{R}$, then $P(c) = 0$ for some $c \in \mathbb{R}$ (the function $y = P(x)$ is continuous), and $P(x)$ is divisible by $x - c$ (Bezout's theorem). Thus, $P(c) > 0$ for each $c \in \mathbb{R}$.

Since $P(x)$ has no real roots, we will prove that it is divisible by a real quadratic polynomial with a negative discriminant. Such polynomial has two complex roots, which can be expressed in the polar form

$$c_{1,2} = r(\cos\varphi \pm i\sin\varphi), \quad r \in (0; +\infty),\ \varphi \in (0; \pi). \tag{2}$$

If $P(x)$ itself has roots (2), then it is divisible by the quadratic polynomial

$$p(x) = (x - c_1)(x - c_2) = x^2 - x(c_1 + c_2) + c_1c_2 = x^2 - 2xr\cos\varphi + r^2.$$

We have to prove that $P(c_1) = P(c_2) = 0$ for some numbers c_1 and c_2 as in (2). By de Moivre's formula, the values $P(c_1)$ and $P(c_2)$ can be expressed in a form $P(c_{1,2}) = U(r,\varphi) \pm iV(r,\varphi)$, where

$$U(r,\varphi) = \sum_{k=0}^{n} a_k r^k \cos k\varphi, \qquad V(r,\varphi) = \sum_{k=0}^{n} a_k r^k \sin k\varphi. \tag{3}$$

The equalities $P(c_1) = P(c_2) = 0$ follow from $U(r,\varphi) = V(r,\varphi) = 0$.

Therefore, it suffices to prove the following statement for each real polynomial (1) satisfying $a_n = 1$ and $P(c) > 0$, $c \in \mathbb{R}$: *there exist such* $r \in (0; +\infty)$ *and* $\varphi \in (0; \pi)$ *that both expressions* (3) *are equal to* 0. We prove this statement in the next section.

3 The proof of FTA

For each pair (r,φ), where $r \in [0; +\infty)$ and $\varphi \in [0; \pi]$, we define $U(r,\varphi)$ and $V(r,\varphi)$ by (3). For short, we denote $U = U(r,\varphi)$, $V = V(r,\varphi)$. Note that $U = V = 0$ is impossible if $\varphi = 0$, $\varphi = \pi$ or $r = 0$, as $U(r,\varphi) = P(\pm r) \neq 0$ in these cases. We will prove that $U = V = 0$ for at least one suitable pair (r,φ) by contradiction.

Hence, we assume that $U^2 + V^2 > 0$ for any $r \in [0; +\infty)$, $\varphi \in [0; \pi]$. We denote

$$X = \frac{U}{\sqrt{U^2 + V^2}}, \qquad Y = \frac{V}{\sqrt{U^2 + V^2}}. \tag{4}$$

Then $X^2+Y^2=1$, and in the Cartesian plane Oxy, each point $(X,Y)=(X_0,Y_0)$ is on the unit circle centered at O:

$$X=\cos\alpha, \qquad Y=\sin\alpha, \qquad \alpha=\alpha(r,\varphi)\in[0;2\pi). \tag{5}$$

Let $r\geqslant 0$ be fixed. As φ increases from 0 to π, the point (X_0,Y_0) moves on a circle. The pair of functions (X,Y) in variable φ determines the movement. We call this pair an r-*trajectory*. It is continuous on $[0;\pi]$, and so is α on any interval $I\subset[0;\pi]$, where $(X,Y)\neq(1,0)$. For any fixed $r\geqslant 0$, the following observations hold.

(i) If $\varphi=0$ or π, then $(U,V)=(P(r),0)$ or $(P(-r),0)$, respectively. In both cases, $U>0$, $(X,Y)=(1,0)$ and $\alpha=0$ (thus, each r-trajectory is closed).

(ii) If $r=0$, then $(U,V)=(P(0),0)$ and $(X,Y)=(1,0)$ (for each φ). Hence, in the case of the 0-trajectory, α is constant and equal to 0.

(iii) Let $r>0$. Then there is only a finite number of $\varphi\in[0;\pi]$, for which $V=Y=0$, i.e. for which $(X,Y)=(\pm1,0)$. Indeed, by the recurrence formula $\sin kx=2\cos x\cdot\sin(k-1)x-\sin(k-2)x$, it can be shown inductively that $\sin kx$ $(k=1,2,\dots)$ is expressible as a product of $\sin x$ and a real polynomial of degree $k-1$ in $\cos x$. By (3), it follows that $V=\sin\varphi\cdot Q(\cos\varphi)$, where Q is a real polynomial (depending on r) with $\deg Q=n-1$. Such a polynomial has at most $n-1$ roots $c\in\mathbb{R}$, and each of the equations $\sin\varphi=0$ and $\cos\varphi=c$, where $Q(c)=0$, has at most two solutions $\varphi\in[0;\pi]$. Therefore, the equation $V=0$ has only a finite number of solutions φ.

(iv) We consider the suspicious points φ of the interval $[0;\pi]$, at which $(X,Y)=(1,0)$ and the function α could be discontinuous (jumping from $\alpha\approx 0$ to $\alpha\approx 2\pi$ or vice versa). By (iii), the number of such φ is finite. We can remove the condition $\alpha\in[0;2\pi)$ in the definition (5) of α and modify the function α (in variable φ) as follows. We keep $\alpha(r,0)=0$ and add appropriate multiples of 2π at the suspicious points and on the intervals of the remaining values of φ so that α becomes continuous for all $\varphi\in[0;\pi]$.

(v) If $V\neq 0$ on some interval $I\subset[0;\pi]$, then the derivative of the ratio $X/Y=U/V$ (as a function in φ) exists, given by $(U'V-UV')/V^2$. The leading term (with respect to r) of the numerator $U'V-UV'$ equals

$$\bigl(r^n\cos n\varphi\bigr)'\cdot r^n\sin n\varphi-r^n\cos n\varphi\cdot\bigl(r^n\sin n\varphi\bigr)'=-nr^{2n},$$

while the sum S of the remaining terms satisfies $|S|<r^{2n-1}\cdot R$ for some constant $R>0$, not depending on r, φ or I. Hence, $(X/Y)'<0$ for $r=R$, $\varphi\in I$.

By (iv), we assume that, for each $r>0$, the number α continuously varies from 0 to $\alpha(r,\pi)$ as φ increases from 0 to π. Since $\sin(\alpha(r,\pi))=Y=0$ by (i), we conclude that $\alpha(r,\pi)$ is some multiple $2\pi m$ of 2π $(m\in\mathbb{Z})$. By (ii), we arrive at the same conclusions for $r=0$: in this case, α continuously varies from 0 to $2\pi m$, where $m=0$. We have two types of r-trajectories: for each $r\geqslant 0$, either $m=0$ or $m\neq 0$. Note that the 0-trajectory is of the first type.

By (iii) and (v), the derivative (with respect to φ) of $\cot\alpha=X/Y$ is negative for $r=R$, except at a finite number of points φ. Thus, the function $\alpha(R,\varphi)$ is increasing, except perhaps at these special points. Being continuous, it is increasing over the entire interval $[0;\pi]$. Hence, $\alpha(R,\pi)>0$, and the R-trajectory is of the second type.

We have r-trajectories of both types when $r \in [0; R]$. Thus, within this interval, the values of r for which the r-trajectories are of different types can be made arbitrarily close to each other, i.e. there exist such (infinite) sequences $q_1, q_2, \dots$ and $r_1, r_2, \dots$ with all terms q_k and r_k in $[0; R]$, that each q_k-trajectory is of the first type, each r_k-trajectory is of the second type, and $q_k - r_k \to 0$ as $k \to \infty$.

We recall that *each infinite sequence with all terms in an interval of a finite length has a convergent subsequence* (Bolzano–Weierstrass theorem). Hence, we assume that $r_k \to r_0 \in \mathbb{R}$ as $k \to \infty$ (otherwise, we could exclude some terms from the considered sequences). Then $q_k = (q_k - r_k) + r_k$ also converges to $0 + r_0 = r_0$.

For each $k = 1, 2, \dots$, we consider the difference $|\alpha(q_k, \varphi) - \alpha(r_k, \varphi)|$. This function in variable $\varphi \in [0; \pi]$ is equal to 0 at $\varphi = 0$ and to a positive multiple of 2π at $\varphi = \pi$. Being continuous, it takes all values between 0 and 2π, including the value π for some $\varphi = \varphi_k$. We obtain the sequence $\varphi_1, \varphi_2, \dots$. All its elements are in $[0; \pi]$, and we can assume again that this sequence converges to some $\varphi_0 \in \mathbb{R}$ (otherwise, we could exclude some terms from the considered sequences). The condition

$$\alpha(q_k, \varphi_k) - \alpha(r_k, \varphi_k) = \pm\pi$$

implies that the respective points $(X(q_k, \varphi_k), Y(q_k, \varphi_k))$ and $(X(r_k, \varphi_k), Y(r_k, \varphi_k))$ are the endpoints of a diameter of the circle $x^2 + y^2 = 1$. Thus,

$$X(q_k, \varphi_k) = -X(r_k, \varphi_k), \qquad Y(q_k, \varphi_k) = -Y(r_k, \varphi_k), \quad k = 1, 2, \dots.$$

By (4) and the continuity of trigonometric and power functions in (3), it follows that

$$X(q_k, \varphi_k) \to X(r_0, \varphi_0), \qquad -X(r_k, \varphi_k) \to -X(r_0, \varphi_0) \quad \text{as } k \to \infty.$$

Therefore, $X(r_0, \varphi_0) = -X(r_0, \varphi_0)$ and $X(r_0, \varphi_0) = 0$. Similarly, $Y(r_0, \varphi_0) = 0$. This contradicts our assumption $X^2 + Y^2 > 0$ and concludes the proof of FTA.

4 Remarks

In Section 2, we applied Bézout's theorem: if the polynomial (1) has a root c (thus $P(c) = 0$), then it is divisible by $x - c$. This theorem holds in both real and complex cases. It is implied by the equalities

$$P(x) = P(x) - P(c) = \sum_{k=0}^{n} a_k (x^k - c^k) = (x - c) P_1(x), \qquad \deg P_1 = n - 1.$$

Here, P has no other roots except for c and the roots of P_1. By induction on n, it follows that no polynomial (1) has more than n distinct (complex or real) roots. This was applied to the polynomial Q in part (iii) of Section 3.

The divisibility of the real polynomial (1), which has complex roots (2), by a quadratic polynomial $p(x)$ with the same two roots can be explained as follows. By repeatedly subtracting multiples of $p(x)$ from $P(x)$, we can repeatedly reduce the degree of the polynomial while c_1 and c_2 remain its roots until we obtain $\tilde{P}(x) = ax + b$. Since $\tilde{P}(c_1) = \tilde{P}(c_2) = 0$, it follows that $a = b = 0$. Thus, $P(x)$ is a multiple of $p(x)$.

In the proof of FTA, we applied easily comprehendible properties of continuous functions: if $f : I \to \mathbb{R}$ is continuous on $I = [a; b]$, then it takes all values between $f(a)$ and $f(b)$; the conditions $\{x_1, x_2, \dots\} \subset I$, $\lim_{n\to\infty} x_n = x_0$ imply that

$\lim_{n\to\infty} f(x_n) = f(x_0)$. Naturally, we assumed the continuity of trigonometric, power and polynomial functions. For a fixed $r > 0$, the continuity of α can be explained not only intuitively by the movement of the point (X_0, Y_0) but also by (5) and the continuity of X, Y and inverse trigonometric functions.

To obtain the sequences $q_1, q_2, \ldots$ and $r_1, r_2, \ldots$, we considered the partitioning of $[0; R]$ into two non-empty disjoint subsets (according to the types of r-trajectories) and concluded that their elements can be arbitrarily close to each other. This intuitive conclusion can be easily proved: if A is the subset containing 0 and our conclusion is wrong, then for any $x \in A$ and a fixed ε, we have $[x; x+\varepsilon] \cap [0; R] \subset A$. Taking $x = 0, \varepsilon, 2\varepsilon, \ldots$ implies that $A = [0; R]$, while the other subset is $\varnothing$ (a contradiction).

We also used the Bolzano–Weierstrass theorem, a fundamental statement of real analysis. It can be informally explained as follows. An interval of finite length that contains all terms of a given sequence is divided into two equal intervals, and the one containing infinitely many terms is selected. This action is then repeated with each newly selected interval. The increasing sequence of the left endpoints of the selected intervals converges to some a. Then a is in all these intervals and is the limit of any subsequence of the given sequence, where the k-th term is in the k-th selected interval.

See also [1] for approachable proofs or explanations of various mathematical concepts and theorems (like a polar form of the complex number and de Moivre's formula applied in our proof).

References

[1] R. Courant, H. Robbins. *What is Mathematics? An Elementary Approach to Ideas and Methods*. Oxford University Press, New York, 1996.

[2] B. Fine, G. Rosenberger. *The Fundamental Theorem of Algebra*. Springer-Verlag, New York, 1997.

[3] J.J. Mačys. Pagrindinė algebros teorema. *Liet. matem. rink. LMD darbai, Ser. B*, **63**:47–53, 2022. https://doi.org/10.15388/LMR.2022.29899.

[4] A. Novikas, J.J. Mačys. On an elementary proof of the fundamental theorem of algebra. *Lith. Math. J.*, **64**(4):507–513, 2024. https://doi.org/10.1007/s10986-024-09651-1.

REZIUMĖ

Trumpas pagrindinės algebros teoremos įrodymas

A. Novikas J.J. Mačys

Vidurinėje mokykloje mokinius galima supažindinti ne tik su tokiais reikšmingais matematikos teiginiais kaip pagrindinė algebros teorema, bet ir su jų įrodymais. Ši garsi teorema turi įvairių įrodymų, o jos įrodymai – įvairių modifikacijų. Straipsnyje aptariamas vienas mokiniams prieinamas įrodymas.

Raktiniai žodžiai: realieji polinomai; pagrindinė algebros teorema; matematinis įrodymas mokykloje