Under revision for publication in the American Mathematical Monthly.
All comments are very much welcome!

THE FUNDAMENTAL THEOREM OF ALGEBRA MADE EFFECTIVE: AN ELEMENTARY REAL-ALGEBRAIC PROOF VIA STURM CHAINS

MICHAEL EISERMANN

L'algèbre est généreuse ; elle donne souvent plus qu'on lui demande. (d'Alembert)

ABSTRACT. Sturm's famous theorem (1829/35) provides an elegant algorithm to count and locate the real roots of any given real polynomial. In his residue calculus of complex functions, Cauchy (1831/37) extended this to an algebraic method to count and locate the complex roots of any given complex polynomial. We give a real-algebraic proof of Cauchy's theorem starting from the axioms of a real closed field, without appeal to analysis. This allows us to algebraically formalize Gauss' geometric argument (1799) and thus to derive a real-algebraic proof of the Fundamental Theorem of Algebra, stating that every complex polynomial of degree n has n complex roots. The proof is elementary inasmuch as it uses only the intermediate value theorem and arithmetic of real polynomials. It can thus be formulated in the first-order language of real closed fields. Moreover, the proof is constructive and immediately translates to an algebraic root-finding algorithm. The latter is sufficiently efficient for moderately sized polynomials, but in its present form it still lags behind Schönhage's nearly optimal numerical algorithm (1982).


Carl Friedrich Gauß (1777–1855)


Augustin Louis Cauchy (1789–1857)


Charles-François Sturm (1803–1855)

1. INTRODUCTION AND STATEMENT OF RESULTS

1.1. **Historical origins.** Sturm's theorem [51, 52], announced in 1829 and published in 1835, provides an elegant and ingeniously simple algorithm to determine for each real polynomial $P \in \mathbb{R}[X]$ the number of real roots in any given interval $[a,b] \subset \mathbb{R}$. Sturm's result solved an outstanding problem of his time and earned him instant fame.

In his residue calculus of complex functions, outlined in 1831 and fully developed in 1837, Cauchy [8, 9] extended Sturm's method to determine for each complex polynomial $F \in \mathbb{C}[Z]$ the number of complex roots in any given rectangle $[a,b] \times [c,d] \subset \mathbb{R}^2 \cong \mathbb{C}$.

Date: first version March 2008; this version compiled May 13, 2009.
2000 *Mathematics Subject Classification.* 12D10; 26C10, 30C15, 65E05, 65G20.
Key words and phrases. constructive and algorithmic aspects of the fundamental theorem of algebra, real closed field, Sturm chains, Cauchy index, algebraic winding number, root-finding algorithm, computer algebra, numerical approximation.

Unifying the real and the complex case, we give a real-algebraic proof of Cauchy's theorem, starting from the axioms of a real closed field, without appeal to analysis. This allows us to algebraicize Gauss' geometric argument (1799) and thus to derive an elementary, real-algebraic proof of the Fundamental Theorem of Algebra, stating that every complex polynomial of degree n has n complex roots. This classical theorem is of theoretical and practical importance, and our proof attempts to satisfy both aspects. Put more ambitiously, we strive for an optimal proof, which is elementary, elegant, and effective.

The logical structure of such a proof was already outlined by Sturm in 1836, but his article [53] lacks the elegance and perfection of his famous 1835 mémoire. This may explain why his sketch found little resonance, was not further worked out, and became forgotten by the end of the 19th century. The contribution of the present article is to save the real-algebraic proof from oblivion and to develop Sturm's idea in due rigour. The presentation is intended for non-experts and thus contains much introductory and expository material.

1.2. **The theorem and its proofs.** In its simplest form, the Fundamental Theorem of Algebra says that every non-constant complex polynomial has at least one complex zero. Since zeros split off as linear factors, this is equivalent to the following formulation:

Theorem 1.1 (Fundamental Theorem of Algebra)**.** *For every polynomial*

$$F = Z^n + c_{n-1}Z^{n-1} + \cdots + c_1 Z + c_0$$

with complex coefficients $c_0, c_1, \ldots, c_{n-1} \in \mathbb{C}$ *there exist* $z_1, z_2, \ldots, z_n \in \mathbb{C}$ *such that*

$$F = (Z - z_1)(Z - z_2) \cdots (Z - z_n).$$

Numerous proofs of this theorem have been published over the last two centuries. According to the tools used, they can be grouped into three major families (§7):

(1) Analysis, using compactness, analytic functions, integration, etc.;
(2) Algebra, using symmetric functions and the intermediate value theorem;
(3) Algebraic topology, using some form of the winding number.

The real-algebraic proof presented here is situated between (2) and (3) and combines Gauss' winding number with Cauchy's index and Sturm's algorithm. It enjoys several remarkable features:

- It uses only the intermediate value theorem and arithmetic of real polynomials.
- It is elementary, in the colloquial as well as the formal sense of first-order logic.
- All arguments and constructions extend verbatim to all real closed fields.
- The proof is constructive and immediately translates to a root-finding algorithm.
- The algorithm is easy to implement and reasonably efficient in medium degree.
- It can be formalized to a computer-verifiable proof (theorem *and* algorithm).

Each of the existing proofs has its special merits. It should be emphasized, however, that a non-constructive existence proof only "announces the presence of a treasure, without divulging its location", as Hermann Weyl put it: "It is not the existence theorem that is valuable, but the construction carried out in its proof." [63, p. 54–55]

I do not claim the real-algebraic proof to be the shortest, nor the most beautiful, nor the most profound one, but overall it offers an excellent cost-benefit ratio. A reasonably short proof can be extracted by omitting all illustrative comments; in the following presentation, however, I choose to be comprehensive rather than terse.

1.3. **The algebraic winding number.** Our arguments work over every ordered field $\mathbf{R}$ that satisfies the intermediate value property for polynomials, i.e., a *real closed field* (§2). We choose this starting point as the axiomatic foundation of Sturm's theorem (§3). (Only for the root-finding algorithm in Theorem 1.11 and Section 6 must we additionally assume $\mathbf{R}$ to be an archimedian, which amounts to $\mathbf{R} \subset \mathbb{R}$.) We then deduce that the field $\mathbf{C} = \mathbf{R}[i]$ with $i^2 = -1$ is algebraically closed, and moreover establish an algorithm to locate the roots of any given polynomial $F \in \mathbf{C}[Z]$. The key ingredient is the construction of an algebraic

winding number (§4–§5), extending the ideas of Cauchy [8, 9] and Sturm [52, 53] in the setting of real algebra:

Theorem 1.2 (algebraic winding number)**.** *Consider an ordered field* $\mathbf{R}$ *and its extension* $\mathbf{C} = \mathbf{R}[i]$ *where* $i^2 = -1$*. Let* Ω *be the set of piecewise polynomial loops* $\gamma\colon [0,1] \to \mathbf{C}^*$*,* $\gamma(0) = \gamma(1)$*, where* $\mathbf{C}^* = \mathbf{C} \smallsetminus \{0\}$*. If* $\mathbf{R}$ *is real closed, then we can construct a map* $w\colon \Omega \to \mathbb{Z}$*, called* algebraic winding number*, satisfying the following properties:*

- (W0) *Computation:* $w(\gamma)$ *is defined as half the Cauchy index of* $\frac{\operatorname{re}\gamma}{\operatorname{im}\gamma}$*, recalled below, and can thus be calculated by Sturm's algorithm via iterated euclidean division.*
- (W1) *Normalization: if* γ *parametrizes the boundary* $\partial\Gamma \subset \mathbf{C}^*$ *of a rectangle* $\Gamma \subset \mathbf{C}$*, positively oriented as in Figure 1, then*

$$w(\gamma) = \begin{cases} 1 & \text{if } 0 \in \operatorname{Int}\Gamma, \\ 0 & \text{if } 0 \in \mathbf{C} \smallsetminus \Gamma. \end{cases}$$

- (W2) *Multiplicativity: for all* $\gamma_1, \gamma_2 \in \Omega$ *we have* $w(\gamma_1 \cdot \gamma_2) = w(\gamma_1) + w(\gamma_2)$*.*
- (W3) *Homotopy invariance: for all* $\gamma_0, \gamma_1 \in \Omega$ *we have* $w(\gamma_0) = w(\gamma_1)$ *if* $\gamma_0 \sim \gamma_1$*, that is, whenever* γ_0 *and* γ_1 *are (piecewise polynomially) homotopic in* $\mathbf{C}^*$*.*

The geometric idea is very intuitive: $w(\gamma)$ counts the number of turns that γ performs around 0 (see Figure 1). Theorem 1.2 turns the geometric idea into a rigorous algebraic construction and provides an effective calculation via Sturm chains.

Remark 1.3. The algebraic winding number is slightly more general than stated in Theorem 1.2. The algebraic definition (W0) of $w(\gamma)$ also applies to loops γ that pass through 0. Normalization (W1) extends to $w(\gamma) = \frac{1}{2}$ if 0 is in an edge of Γ, and $w(\gamma) = \frac{1}{4}$ is 0 is one of the vertices of Γ. Multiplicativity (W2) continues to hold provided that 0 is not a vertex of γ_1 or γ_2. Homotopy invariance (W3) applies only if γ does not pass through 0.

Remark 1.4. The existence of the algebraic winding number over $\mathbf{R}$ relies on the intermediate value theorem for polynomials. (Such an map does not exist over $\mathbb{Q}$, for example.) Conversely, its existence implies that $\mathbf{C} = \mathbf{R}[i]$ is algebraically closed and hence $\mathbf{R}$ is real closed (see Remark 2.6). More precisely, given any ordered field $\mathbf{K}$, Theorem 1.2 holds for the real closure $\mathbf{R} = \mathbf{K}^c$ (see Theorem 2.5): properties (W0), (W1), (W2) restrict to loops over $\mathbf{K}$, and it is the homotopy invariance (W3) that is equivalent to $\mathbf{K}$ being real closed.

Remark 1.5. Over the real numbers $\mathbb{R}$, several alternative constructions are possible:

(1) Covering theory, applied to $\exp\colon \mathbb{C} \twoheadrightarrow \mathbb{C}^*$ with covering group $\mathbb{Z}$.
(2) Fundamental group, $w\colon \pi_1(\mathbb{C}^*, 1) \xrightarrow{\sim} \mathbb{Z}$ via Seifert–van Kampen.
(3) Homology, $w\colon H_1(\mathbb{C}^*) \xrightarrow{\sim} \mathbb{Z}$ via Eilenberg–Steenrod axioms.
(4) Complex analysis, analytic winding number $w(\gamma) = \frac{1}{2i\pi}\int_\gamma \frac{\mathrm{d}z}{z}$ via integration.
(5) Real algebra, algebraic winding number $w\colon \Omega \to \mathbb{Z}$ via Sturm chains.

Each of the first four approaches uses some characteristic property of the real numbers (such as local compactness, metric completeness, or connectedness). As a consequence, these topological or analytical constructions do not extend to real closed fields.

Remark 1.6. Over $\mathbb{C}$ the *algebraic winding number* coincides with the *analytic winding number* given by Cauchy's integral formula

$$w(\gamma) = \frac{1}{2\pi i}\int_\gamma \frac{\mathrm{d}z}{z} = \frac{1}{2\pi i}\int_0^1 \frac{\gamma'(t)}{\gamma(t)}\,\mathrm{d}t. \tag{1.1}$$

This is called the *argument principle* and is intimately related to the covering map $\exp\colon \mathbb{C} \twoheadrightarrow \mathbb{C}^*$ and the fundamental group $\pi_1(\mathbb{C}^*, 1) \cong \mathbb{Z}$. Cauchy's integral (1.1) is the ubiquitous technique of complex analysis and one of the most popular tools for proving the Fundamental Theorem of Algebra.

In this article we develop an independent, purely algebraic proof avoiding integrals, transcendental functions, and covering spaces. Seen from an elevated viewpoint, our approach interweaves real-algebraic geometry and effective algebraic topology. In this general setting Theorem 1.2 and its real-algebraic proof seem to be new.

1.4. **The Fundamental Theorem of Algebra.** I have highlighted Theorem 1.2 in order to summarize the real-algebraic approach, combining geometry and algebra. The first step in the proof (§4) is to study the algebraic winding number $w(F|\partial\Gamma)$ of a polynomial $F \in \mathbf{C}[Z]$ along the boundary of a rectangle $\Gamma \subset \mathbf{C}$, positively oriented as in Figure 1.

Example 1.7. Figure 1 (right) displays $F(\partial\Gamma)$ for $F = Z^5 - 5Z^4 - 2Z^3 - 2Z^2 - 3Z - 12$ and $\Gamma = [-1,+1] \times [-1,+1]$. Here the winding number is seen to be $w(F|\partial\Gamma) = 2$.

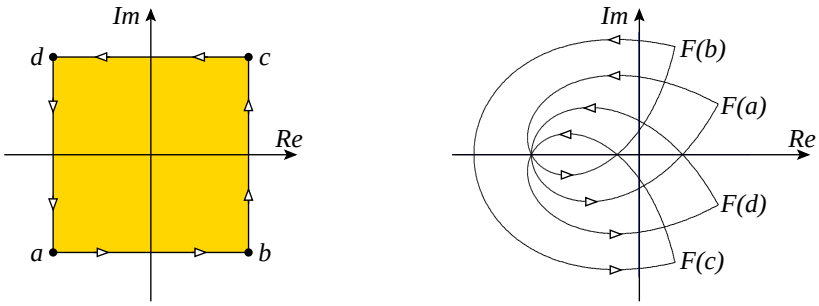


FIGURE 1. The winding number $w(F|\partial\Gamma)$ of a polynomial $F \in \mathbf{C}[Z]$ with respect to a rectangle $\Gamma \subset \mathbf{C}$

We then establish the algebraic generalization of Cauchy's theorem for $\mathbf{C} = \mathbf{R}[i]$ over a real closed field $\mathbf{R}$, extending Sturm's theorem from real to complex polynomials:

Theorem 1.8. *If $F \in \mathbf{C}[Z]$ does not vanish in any of the four vertices of the rectangle $\Gamma \subset \mathbf{C}$, then the algebraic winding number $w(F|\partial\Gamma)$ equals the number of roots of F in Γ:*

- *Each root of F in the interior of Γ counts with its multiplicity.*
- *Each root of F in an edge of Γ counts with half its multiplicity.*

Remark 1.9. The hypothesis that $F \neq 0$ on the vertices is very mild and easy enough to check in every concrete application. Unlike the integral formula (1.1), the algebraic winding number behaves well if zeros lie on (or close to) the boundary. This is yet another manifestation of the oft-quoted wisdom of d'Alembert that *algebra is generous; she often gives more than we ask of her*. Apart from its aesthetic appeal, the uniform treatment of all configurations simplifies theoretical arguments and practical implementations alike.

The second step in the proof (§5) formalizes the geometric idea of Gauss' dissertation (1799), which becomes perfectly rigorous and nicely quantifiable in the algebraic setting:

Theorem 1.10. *For each polynomial $F = c_0 + c_1 Z + \cdots + c_{n-1} Z^{n-1} + c_n Z^n$ in $\mathbf{C}[Z]$ of degree $n \geq 1$ we define its Cauchy radius to be $\rho_F := 1 + \max\{|c_0|, |c_1|, \ldots, |c_{n-1}|\}/|c_n|$. Then every rectangle Γ containing the disk $\{z \in \mathbf{C} \mid |z| < r\}$ satisfies $w(F|\partial\Gamma) = n$.*

Theorems 1.8 and 1.10 together imply that $\mathbf{C}$ is algebraically closed: each polynomial $F \in \mathbf{C}[Z]$ of degree n has n roots in $\mathbf{C}$, each counted with its multiplicity; more precisely, the square $\Gamma = [-\rho_F, \rho_F]^2 \subset \mathbf{C}$ contains n roots of F.

Applied to the field $\mathbb{C} = \mathbb{R}[i]$ of complex numbers, this result is traditionally called the *Fundamental Theorem of Algebra*, following Gauss, although nowadays it would be more appropriate to call it the "fundamental theorem of complex numbers".

We emphasize that the algebraic approach via Cauchy indices proves much more than mere existence of roots. It also establishes a root-finding algorithm (§6.2):

Theorem 1.11 (Fundamental Theorem of Algebra, effective version)**.** *For every polynomial $F \in \mathbb{C}[Z]$ of degree $n \geq 1$ there exist $c, z_1, \ldots, z_n \in \mathbb{C}$ such that*

$$F = c(Z - z_1) \cdots (Z - z_n).$$

The algebraic winding number provides an explicit algorithm to locate all roots $z_1, \ldots, z_n$ of F: starting from some rectangle containing all n roots, as in Theorem 1.10, we can subdivide and keep only those rectangles that actually contain roots, using Theorem 1.8. All computations can be carried out using Sturm chains according to Theorem 1.2. By iterated bisection we can thus approximate all roots to any desired precision.

Once sufficient approximations have been obtained, one can switch to Newton's method, which converges much faster but vitally depends on good starting values (§6.3).

Remark 1.12. In the real-algebraic setting of this article we consider the field operations $(a,b) \mapsto a+b$, $a \mapsto -a$, $(a,b) \mapsto a \cdot b$, $a \mapsto a^{-1}$ in $\mathbf{R}$ and the comparisons $a = b$, $a < b$ as primitive operations. In this sense our proof yields an algorithm over $\mathbf{R}$. Over the real numbers $\mathbb{R}$ this point of view was advanced by Blum–Cucker–Shub–Smale [6] by extending the notion of Turing machines to hypothetical "real number machines".

In order to carry out the required real-algebraic operations on a Turing machine, however, a more careful analysis is necessary (§6.1). At the very least, in order to implement the required operations for a given polynomial $F = c_0 + c_1 Z + \cdots + c_n Z^n$, we have to assume that for the ordered field $\mathbb{Q}(\mathrm{re}(c_0), \mathrm{im}(c_0), \ldots, \mathrm{re}(c_n), \mathrm{im}(c_n))$ the above primitive operations are computable in the Turing sense. See §6 for a more detailed discussion.

1.5. **Why yet another proof?** There are several lines of proof leading to the Fundamental Theorem of Algebra, and literally hundreds of variants have been published over the last 200 years (see §7). Why should we care for yet another proof?

The motivations for the present work are three-fold:

First, on a philosophical level, it is satisfying to minimize the hypotheses and the tools used in the proof, and simultaneously maximize the conclusion.

Second, when teaching mathematics, it is advantageous to have different proofs to choose from, adapted to the course's level and context.

Third, from a practical point of view, it is desirable to have a constructive proof, even more so if it directly translates to a practical algorithm.

In these respects the present approach offers several attractive features:

(1) The proof is elementary, and a thorough treatment of the complex case (§4–§5) is of comparable length and difficulty as Sturm's treatment of the real case (§2–§3).
(2) Since the proof uses only first-order properties (and not compactness, for example) all arguments hold verbatim over any real closed field (§2.3).
(3) The proof is constructive in the sense that it establishes not only existence but also provides a method to locate the roots of F (§6.2).
(4) The algorithm is fairly easy to implement on a computer and sufficiently efficient for medium-sized polynomials (§6.4).
(5) Its economic use of axioms and its algebraic character make this approach ideally suited for a formal, computer-verified proof (§6.6).
(6) Since the real-algebraic proof also provides an algorithm, the correctness of an implementation can likewise be formally proved and computer-verified.

1.6. **Sturm's forgotten proof.** Attracted by the above features, I have worked out the real-algebraic proof for a computer algebra course in 2008. The idea seems natural, or even obvious, and so I was quite surprised not to find any such proof in the modern literature. While retracing its history (§7), I was even more surprised when I finally unearthed very similar arguments in the works of Cauchy and Sturm (§7.4). Why have they been lost?

Our proof is, of course, based on very classical ideas. The geometric idea goes back to Gauss in 1799, and all algebraic ingredients are present in the works of Sturm and Cauchy in the 1830s. Since then, however, they have evolved in very different directions:

Sturm's theorem has become a cornerstone of real algebra. Cauchy's integral is the starting point of complex analysis. Their algebraic method for counting complex roots, however, has transited from algebra to applications, where its conceptual and algorithmic simplicity are much appreciated. Since the end of the 19th century it is no longer found in algebra text books, but is almost exclusively known as a computational tool, for example in the Routh–Hurwitz theorem on the stability of motion. After Sturm's outline of 1836, this algebraic tool seems not to have been employed to *prove* the existence of roots.

In retrospect, the proof presented here is thus a fortunate rediscovery of Sturm's algebraic vision (§7.5). This article gives a modern, rigorous, and complete presentation, which means to set up the right definitions and to provide elementary, real-algebraic proofs.

1.7. **How this article is organized.** Section 2 briefly recalls the notion of real closed fields, on which Sturm's theorem and the theory of Cauchy's index are built.

Section 3 presents Sturm's theorem [52] counting real roots of real polynomials. The only novelty is the extension to boundary points, which is needed in Section 4.

Section 4 proves Cauchy's theorem [9] counting complex roots of complex polynomials, by establishing the multiplicativity (W2) of the algebraic winding number.

Section 5 establishes the Fundamental Theorem of Algebra via homotopy invariance (W3), recasting the classical winding number approach in real algebra.

Section 6 discusses algorithmic aspects, such as Turing computability, the efficient computation of Sturm chains and the cross-over to Newton's local method.

Section 7, finally, provides historical comments in order to put the real-algebraic approach into a wider perspective.

The core of our real-algebraic proof is rather short (§4–§5). It seems necessary, however, to properly develop the underlying tools and to arrange the details of the real case (§2–§3). Algorithmic and historical aspects (§6–§7) complete the picture. I hope that the subject justifies the length of this article and its level of detail.

Annotation 1.1. (Organization) I have tried to keep the exposition as elementary as possible. This requires to strike a balance between terseness and verbosity – in cases of doubt I have opted for the latter: in this annotated student version, some complementary remarks are included that will most likely not appear in the published version. They are set in small font, as this one, and numbered separately in order to ensure consistent references.

CONTENTS

7. **Historical remarks.** 7.1. Solving polynomial equations. 7.2. Gauss' first proof. 7.3. Gauss' further proofs. 7.4. Sturm, Cauchy, Liouville. 7.5. Sturm's algebraic vision. 7.6. Further development in the 19th century. 7.7. 19th century textbooks. 7.8. Survey of proof strategies. 7.9. Constructive and algorithmic aspects.

A. **Application to the Routh–Hurwitz stability theorem.**

B. **Brouwer's fixed point theorem.**

2. REAL CLOSED FIELDS

There can be no purely algebraic proof of the Fundamental Theorem of Algebra in the sense that ordered fields and the intermediate value property of polynomials must enter the picture (see Remark 2.6 below). This is the natural setting of real algebra, and constitutes precisely the minimal hypotheses that we will be using.

We shall use only elementary properties of ordered fields, which are well-known from the real numbers (see for example Cohn [11, §8.6–§8.7]). In order to make the hypotheses precise, this section sets the scene by recalling the notion of a real closed field, on which Sturm's theorem is built, and sketches its analytic, algebraic, and logical context.

Annotation 2.1. (Fields) We assume that the reader is familiar with the algebraic notion of a *field*. In order to highlight the field axioms formulated in first-order logic, we recall that a field $(\mathbf{R},+,\cdot)$ is a set $\mathbf{R}$ equipped with two binary operations $+\colon \mathbf{R}\times\mathbf{R}\to\mathbf{R}$ and $\cdot\colon \mathbf{R}\times\mathbf{R}\to\mathbf{R}$ satisfying the following three groups of axioms:

First, addition enjoys the following four properties, saying that $(\mathbf{R},+)$ is an abelian group:

(A1) associativity: For all $a,b,c\in\mathbf{R}$ we have $(a+b)+c=a+(b+c)$.
(A2) commutativity: For all $a,b\in\mathbf{R}$ we have $a+b=b+a$.
(A3) neutral element: There exists $0\in\mathbf{R}$ such that for all $a\in\mathbf{R}$ we have $0+a=a$.
(A4) opposite elements: For each $a\in\mathbf{R}$ there exists $b\in\mathbf{R}$ such that $a+b=0$.

The neutral element $0\in\mathbf{R}$ whose existence is required by axiom (A3) is unique by (A2). This ensures that axiom (A4) is unambiguous. The opposite element of $a\in\mathbf{R}$ required by axiom (A4) is unique and denoted by $-a$.

Second, multiplication enjoys the following four properties, saying that $(\mathbf{R}\smallsetminus\{0\},\cdot)$ is an abelian group:

(M1) associativity: For all $a,b,c\in\mathbf{R}$ we have $(a\cdot b)\cdot c=a\cdot(b\cdot c)$.
(M2) commutativity: For all $a,b\in\mathbf{R}$ we have $a\cdot b=b\cdot a$.
(M3) neutral element: There exists $1\in\mathbf{R}$, $1\neq 0$, such that for all $a\in\mathbf{R}$ we have $1\cdot a=a$.
(M4) inverse elements: For each $a\in\mathbf{R}$, $a\neq 0$, there exists $b\in\mathbf{R}$ such that $ab=1$.

The neutral element $1\in\mathbf{R}$ whose existence is required by axiom (M3) is unique by (M2). This ensures that axiom (M4) is unambiguous. The inverse element of $a\in\mathbf{R}$ required by axiom (M4) is unique and denoted by a^{-1}.

Third, multiplication is distributive over addition:

(D) distributivity: For all $a,b,c\in\mathbf{R}$ we have $a\cdot(b+c)=(a\cdot b)+(a\cdot c)$.

Annotation 2.2. (Ordered fields) An *ordered field* is a field $\mathbf{R}$ with a distinguished set of positive elements, denoted $x>0$, compatible with the field operations in the following sense:

(O1) trichotomy: For each $x\in\mathbf{R}$ we have either $x>0$ or $x=0$ or $-x>0$.
(O2) addition: For all $x,y\in\mathbf{R}$ the conditions $x>0$ and $y>0$ imply $x+y>0$.
(O3) multiplication: For all $x,y\in\mathbf{R}$ the conditions $x>0$ and $y>0$ imply $xy>0$.

From these axioms follow the usual properties, see Cohn [11, §8.6], Jacobson [25, §5.1] or Lang [28, §XI.1]. We define the ordering $x>y$ by $x-y>0$. The weak ordering $x\geq y$ means $x>y$ or $x=y$. The inverse ordering $x<y$ is defined by $y>x$, and likewise $x\leq y$ is defined by $y\geq x$. Intervals in $\mathbf{R}$ will be denoted, as usual, by

$$[a,b]=\{x\in\mathbf{R}\mid a\leq x\leq b\},\qquad]a,b]=\{x\in\mathbf{R}\mid a<x\leq b\},$$
$$]a,b[=\{x\in\mathbf{R}\mid a<x<b\},\qquad [a,b[=\{x\in\mathbf{R}\mid a\leq x<b\}.$$

Every ordered field $\mathbf{R}$ inherits a natural topology generated by open intervals: a subset $U\subset\mathbf{R}$ is open if for each $x\in U$ there exists $\delta>0$ such that $]x-\delta,x+\delta[\subset U$. We can thus apply the usual notions of topological spaces and continuous functions. Addition and multiplication are continuous, and so are polynomial functions.

For every $x\in\mathbf{R}$ we have $x^2\geq 0$ with equality if and only if $x=0$. The polynomial X^2-a can thus have a root $x\in\mathbf{R}$ only for $a\geq 0$; if it has a root, then among the two roots $\pm x$ we can choose $x\geq 0$, denoted $\sqrt{a}:=x$. For $x\in\mathbf{R}$ we define the absolute value to be $|x|:=x$ if $x\geq 0$ and $|x|:=-x$ if $x\leq 0$. We remark that $|x|=\sqrt{x^2}$.

We record the following properties, which hold for all $x,y\in\mathbf{R}$:

(1) $|x|\geq 0$, and $|x|=0$ if and only if $x=0$.
(2) $|x+y|\leq|x|+|y|$ for all $x,y\in\mathbf{R}$.

(3) $|x \cdot y| = |x| \cdot |y|$ for all $x, y \in \mathbf{R}$.

Annotation 2.3. (Rings) A *ring* $(\mathbf{R}, +, \cdot)$ is only required to satisfy axioms (A1-A4), (M1-M3), and (D) but not necessarily (M4). This is sometimes called a *commutative ring with unit*, for emphasis, but we will have no need for this distinction. For every ring $\mathbf{R}$ we denote by $\mathbf{R}^* = \mathbf{R} \setminus \{0\}$ the set of its non-zero elements. A ring $\mathbf{R}$ is called *integral* if for all $a, b \in \mathbf{R}^*$ we have $ab \in \mathbf{R}^*$. Every integral ring $\mathbf{R}$ can be embedded into a field; the smallest such field is unique and thus called the *field of fractions* of $\mathbf{R}$. Every ordered ring is integral, and the ordering uniquely extends to its field of fractions. For example, the ring of integers $\mathbb{Z}$ has as field of fractions the field of rational numbers $\mathbb{Q}$. In this article we will study the ring $\mathbf{R}[X]$ polynomials over some ordered field $\mathbf{R}$, as explained below, which has as field of fractions the field of rational functions $\mathbf{R}(X)$.

2.1. **Real numbers.**

As usual we denote by $\mathbb{R}$ the field of real numbers, that is, an ordered field $(\mathbb{R}, +, \cdot, <)$ such that every non-empty bounded subset $A \subset \mathbb{R}$ has a least upper bound in $\mathbb{R}$. This is a very strong property, and in fact it characterizes $\mathbb{R}$:

Theorem 2.1. *For every ordered field* $\mathbf{R}$ *the following conditions are equivalent:*

(1) *The ordered set* $(\mathbf{R}, <)$ *satisfies the least upper bound property.*
(2) *Each interval* $[a, b] \subset \mathbf{R}$ *is compact as a topological space.*
(3) *Each interval* $[a, b] \subset \mathbf{R}$ *is connected as a topological space.*
(4) *The intermediate value property holds for all continuous functions* $f : \mathbf{R} \to \mathbf{R}$.

Any two ordered fields satisfying these properties are isomorphic by a unique field isomorphism. The construction of the real numbers shows that one such field exists. □

Annotation 2.4. (Sketch of proof) Existence and uniqueness of the field $\mathbb{R}$ of real numbers form the foundation of any analysis course. Most analysis books prove (1) ⇒ (2) ⇒ (4), while (3) ⇔ (4) is essentially the definition of connectedness. Here we only show (4) ⇒ (1), in the form ¬(1) ⇒ ¬(4).

Let $A \subset \mathbf{R}$ be non-empty and bounded above. Define $f : \mathbf{R} \to \{\pm 1\}$ by $f(x) = 1$ if $a \leq x$ for all $a \in A$, and $f(x) = -1$ if $x < a$ for some $a \in A$. In other words, we have $f(x) = 1$ if and only if x is an upper bound. If f is discontinuous in x, then $f(x) = +1$ but $f(y) = -1$ for all $y < x$, whence $x = \sup A$. If A does not have a least upper bound in $\mathbf{R}$, then f is continuous but does not satisfy the intermediate value property.

2.2. **Real closed fields.**

The field $\mathbb{R}$ of real numbers provides the foundation of analysis. In the present article it appears as the most prominent example of the much wider class of real closed fields. The reader who wishes to concentrate on the classical case may skip the rest of this section and assume $\mathbf{R} = \mathbb{R}$ throughout.

Annotation 2.5. (Polynomials) In the sequel we shall assume that the reader is familiar with the polynomial ring $\mathbf{K}[X]$ of some ground ring $\mathbf{K}$, see Jacobson [25, §2.9–§2.12] or Lang [28, §II.2, §IV.1]. We briefly recall some notation. Let $\mathbf{K}$ be a ring, that is, satisfying axioms (A1-A4), (M1-M3), and (D) of Annotation 2.2, but not necessarily (M4). There exists a ring $\mathbf{K}[X]$ characterized by the following two properties: First, $\mathbf{K}[X]$ contains $\mathbf{K}$ as a subring and X as an element. Second, every non-zero element $P \in \mathbf{K}[X]$ can be uniquely written as

$$P = c_0 + c_1 X + \cdots + c_n X^n \quad \text{where} \quad n \in \mathbb{N} \text{ and } c_0, c_1, \ldots, c_n \in \mathbf{K}, c_n \neq 0.$$

In this situation $\mathbf{K}[X]$ is called the *ring of polynomials* over $\mathbf{K}$ in the variable X, and each element $P \in \mathbf{K}[X]$ is called a *polynomial* over $\mathbf{K}$ in X. In the above notation we call $\deg P := n$ the *degree* and $\operatorname{lc} P := c_n$ the *leading coefficient* of P. The zero polynomial is special: we set $\deg 0 := -\infty$ and $\operatorname{lc} 0 := 0$.

Annotation 2.6. (Polynomial functions) The ring $\mathbf{K}[X]$ has the following universal property: for every ring $\mathbf{K}'$ containing $\mathbf{K}$ as a subring and every element $x \in \mathbf{K}'$ there exists a unique ring homomorphism $\Phi : \mathbf{K}[X] \to \mathbf{K}'$ such that $\Phi|_{\mathbf{K}} = \mathrm{id}_{\mathbf{K}}$ and $\Phi(X) = x$. Explicitly, Φ sends $P = c_0 + c_1 X + \cdots + c_n X^n$ to $P(x) = c_0 + c_1 x + \cdots + c_n x^n$. In particular each polynomial $P \in \mathbf{K}[X]$ defines a polynomial function $f_P : \mathbf{K} \to \mathbf{K}$, $x \mapsto P(x)$. If $\mathbf{K}$ is an infinite integral ring, for example an ordered ring or field, then the map $P \mapsto f_P$ is injective, and we can thus identify each polynomial $P \in \mathbf{K}[X]$ with the associated polynomial function $f_P : \mathbf{K} \to \mathbf{K}$.

Annotation 2.7. (Roots) We shall mainly deal with polynomials over ordered – hence infinite – fields. In particular we can identify polynomials and their associated polynomial functions. Traditionally equations have *roots* and functions have *zeros*. In this article we use both words “roots” and “zeros” synonymously.

Definition 2.2. An ordered field $(\mathbf{R}, +, \cdot, <)$ is *real closed* if it satisfies the intermediate value property for polynomials: whenever a polynomial $P \in \mathbf{R}[X]$ satisfies $P(a)P(b) < 0$ for some $a < b$ in $\mathbf{R}$, then there exists $x \in]a, b[$ such that $P(x) = 0$.

Example 2.3. The field $\mathbb{R}$ of real numbers is real closed by Theorem 2.1 above. The field $\mathbb{Q}$ of rational numbers is not real closed, as shown by the example $P = X^2 - 2$ on $[1,2]$. The algebraic closure $\mathbb{Q}^c$ of $\mathbb{Q}$ in $\mathbb{R}$ is a real closed field. In fact, $\mathbb{Q}^c$ is the smallest real closed field, in the sense that $\mathbb{Q}^c$ is contained in any real closed field. Notice that $\mathbb{Q}^c$ is much smaller than $\mathbb{R}$, in fact $\mathbb{Q}^c$ is countable whereas $\mathbb{R}$ is uncountable.

Remark 2.4. The theory of real closed fields originated in the work of Artin and Schreier [3, 4]. Excellent textbook references include Jacobson [25, chapters I.5 and II.11], Cohn [11, chapter 8], and Bochnak–Coste–Roy [7, chapter 1]. For the present article, Definition 2.2 above is the natural starting point because it captures the essential geometric feature. It deviates, however, from Artin–Schreier's algebraic definition [3], which says that an ordered field is real closed if no proper algebraic extension can be ordered. For a proof of their equivalence see [11, Prop. 8.8.9] or [7, §1.2].

Every archimedian ordered field can be embedded into $\mathbb{R}$, see [11, §8.7]. The field $\mathbb{R}(X)$ of rational functions can be ordered (in many different ways, see [7, §1.1]) but does not embed into $\mathbb{R}$. Nevertheless it can be embedded into some real closure:

Theorem 2.5 (Artin–Schreier [3, Satz 8])**.** *Every ordered field* $\mathbf{K}$ *admits a real closure, i.e., a real closed field* $\mathbf{R} \supset \mathbf{K}$ *that extends the ordering and is algebraic over* $\mathbf{K}$*. Any two real closures of* $\mathbf{K}$ *are isomorphic via a unique field isomorphism fixing* $\mathbf{K}$*.* □

The real closure is thus much more rigid than the algebraic closure. In a real closed field $\mathbf{R}$ every positive element has a square root, and so the ordering on $\mathbf{R}$ can be characterized in algebraic terms: $x \geq 0$ if and only if there exists $r \in \mathbf{R}$ such that $r^2 = x$. In particular, if a field $\mathbf{R}$ is real closed, then it admits precisely one ordering.

Remark 2.6. Artin and Schreier [3, Satz 3] have shown that if a field $\mathbf{R}$ is real closed, then $\mathbf{C} = \mathbf{R}[i]$ is algebraically closed, recasting the classical algebraic proof of the Fundamental Theorem of Algebra (§7.8.2). Conversely [4], if $\mathbf{C}$ is algebraically closed and contains a subfield $\mathbf{R}$ such that $1 < \dim_{\mathbf{R}}(\mathbf{C}) < \infty$, then $\mathbf{R}$ is real closed and $\mathbf{C} = \mathbf{R}[i]$. We shall not use this striking result, but it underlines that we have chosen minimal hypotheses.

Annotation 2.8. (Finiteness conditions) In the sequel we will not appeal to the least upper bound property, nor compactness nor connectedness. In particular we will not use analytic methods such as integration, nor transcendental functions such as exp, sin, cos, ... The intermediate value property for polynomials is a sufficiently strong hypothesis. In order to avoid compactness, a sufficient finiteness condition will be the fact that a polynomial $P = c_n X^n + c_{n-1} X^{n-1} + \cdots + c_1 X + c_0$ of degree n over a field $\mathbf{K}$ can have at most n roots in $\mathbf{K}$.

In general P can have *less* than n roots, of course, as illustrated by the classical example $X^2 + 1$ over $\mathbb{R}$. The fact that P cannot have *more* than n roots relies on commutativity (M2) and invertibility (M4). For example $X^2 - 1$ has four roots in the non-integral ring $\mathbb{Z}/8\mathbb{Z}$ of integers modulo 8, namely ± 1 and ± 3. On the other hand, $X^2 + 1$ has infinitely many roots in the skew field $\mathbb{H} = \mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k$ of Hamilton's quaternions [14, chap. 7], namely every combination $ai + bj + ck$ with $a, b, c \in \mathbb{R}$ such that $a^2 + b^2 + c^2 = 1$. The limitation on the number of roots makes the theory of fields very special. We will repeatedly use it as a crucial finiteness condition.

2.3. **Elementary theory of ordered fields.** The axioms of an ordered field $(\mathbf{R}, +, \cdot, <)$ are formulated in first-order logic, which means that we quantify over elements of $\mathbf{R}$, but not over subsets, functions, etc. By way of contrast, the characterization of the field $\mathbb{R}$ of real numbers (Theorem 2.1) is of a different nature: here we have to quantify over subsets of $\mathbb{R}$, or functions $\mathbb{R} \to \mathbb{R}$, and such a formulation requires second-order logic.

The algebraic condition for an ordered field to be real closed is of first order. It is given by an axiom scheme where for each degree $n \in \mathbb{N}$ we have one axiom of the form

$$\begin{aligned} (2.1) \quad \forall a, b, c_0, c_1, \ldots, c_n \in \mathbf{R} \big[& (c_0 + c_1 a + \cdots + c_n a^n)(c_0 + c_1 b + \cdots + c_n b^n) < 0 \\ & \Rightarrow \exists x \in \mathbf{R} \big((x-a)(x-b) < 0 \ \wedge \ c_0 + c_1 x + \cdots + c_n x^n = 0 \big) \big]. \end{aligned}$$

First-order formulae are customarily called *elementary*. For a given ordered field $\mathbf{R}$, the collection of all first-order formulae that are true over $\mathbf{R}$ is called the *elementary theory* of $\mathbf{R}$. Tarski's theorem [25, 7] says that all real closed fields share the same elementary

theory: if an assertion in the first-order language of ordered fields is true over one real closed field, for example the real numbers, then it is true over any other real closed field. (This no longer holds for second-order logic, where $\mathbb{R}$ is singled out.) Tarski's theorem is a vast generalization of Sturm's technique, and so is its effective formulation, called *quantifier elimination*, which provides explicit decision procedures. We will not use Tarski's theorem; it only serves to situate our approach in its logical context.

From Tarski's meta-mathematical viewpoint it is not surprising that the *statement* of the Fundamental Theorem of Algebra generalizes to an arbitrary real closed field, because in each degree it is of first order. It is remarkable, however, to construct a first-order *proof* that is as direct and elegant as the second-order version. The real-algebraic proof presented here achieves this goal and, moreover, is geometrically appealing and algorithmically effective.

Annotation 2.9. (Geometry) Tarski's theorem implies that euclidean geometry, seen as cartesian geometry modeled on the vector space $\mathbb{R}^n$, remains unchanged if the field $\mathbb{R}$ of real numbers is replaced by any other real closed field $\mathbf{R}$. This is true as far as its first-order properties are concerned, and these comprise all of classical geometry.

Annotation 2.10. (Decidability) The elementary theory of real closed fields can be recursively axiomatized, as seen above. By Tarski's theorem it is complete in the sense that any two models of it share the same elementary theory. This implies decidability. This also shows that the first-order theory of euclidean geometry is decidable.

3. Sturm's theorem for real polynomials

This section recalls Sturm's theorem for polynomials over a real closed field – a gem of 19th century algebra and one of the greatest discoveries in the theory of polynomials.

Remark 3.1. It seems impossible to surpass the elegance of the original mémoires by Sturm [52] and Cauchy [9]. One technical improvement of our presentation, however, seems noteworthy: The inclusion of boundary points streamlines the arguments so that they will apply seamlessly to the complex setting in §4. The necessary amendments render the development hardly any longer nor more complicated. They pervade, however, all statements and proofs, so that it seems worthwhile to review the classical arguments in full detail.

3.1. **Counting sign changes.** For every ordered field $\mathbf{R}$ we define $\operatorname{sign}\colon \mathbf{R} \to \{-1, 0, +1\}$ by $\operatorname{sign}(x) = +1$ if $x > 0$, $\operatorname{sign}(x) = -1$ if $x < 0$, and $\operatorname{sign}(0) = 0$. Given a finite sequence $s = (s_0, \dots, s_n)$ in $\mathbf{R}$, we say that the pair (s_{k-1}, s_k) presents a *sign change* if $s_{k-1}s_k < 0$. The pair presents *half a sign change* if one element is zero while the other is non-zero. In the remaining cases there is no sign change. All cases can be subsumed by the formula

$$V(s_{k-1}, s_k) := \tfrac{1}{2}\big|\operatorname{sign}(s_{k-1}) - \operatorname{sign}(s_k)\big|. \tag{3.1}$$

Definition 3.2. For a finite sequence $s = (s_0, \dots, s_n)$ in $\mathbf{R}$ the *number of sign changes* is

$$V(s) := \sum_{k=1}^{n} V(s_{k-1}, s_k) = \sum_{k=1}^{n} \tfrac{1}{2}\big|\operatorname{sign}(s_{k-1}) - \operatorname{sign}(s_k)\big|. \tag{3.2}$$

For a finite sequence $(S_0, \dots, S_n)$ of polynomials in $\mathbf{R}[X]$ and $a \in \mathbf{R}$ we set

$$V_a\big(S_0, \dots, S_n\big) := V\big(S_0(a), \dots, S_n(a)\big). \tag{3.3}$$

For the difference at two points $a, b \in \mathbf{R}$ we use the notation $V_a^b := V_a - V_b$.

Annotation 3.1. The number $V(s_0, \dots, s_n)$ does not change if we multiply all $s_0, \dots, s_n$ by some constant $q \in \mathbf{R}^*$. Likewise, $V_a^b(S_0, \dots, S_n)$ remains unchanged if we multiply all $S_0, \dots, S_n$ by some polynomial $Q \in \mathbf{R}[X]^*$ that does not vanish in $\{a, b\}$. Such operations will be used repeatedly later on.

Remark 3.3. There is no universal agreement how to count sign changes because each application requires its specific conventions. While there is no ambiguity for $s_{k-1}s_k < 0$ and $s_{k-1}s_k > 0$, some arbitration is needed to take care of possible zeros. Our definition has been chosen to account for boundary points in Sturm's theorem, as explained below.

The traditional way of counting sign changes, following Descartes and Budan–Fourier, is to extract the subsequence $\hat{s}$ by discarding all zeros of s and to define $\hat{V}(s) := V(\hat{s})$. (This counting rule is non-local whereas in (3.2) only neighbours interact.) As an illustration we recall Descartes' rule of signs Budan–Fourier's generalization [40, chap. 10]:

Theorem 3.4 (Descartes' rule of signs)**.** *For every polynomial $P = c_0 + c_1X + \cdots + c_nX^n$ over an ordered field $\mathbf{R}$, the number of positive roots, each counted with its multiplicity, satisfies the inequality*

$$\#_{\text{mult}} \left\{ x \in \mathbf{R}_{>0} \;\middle|\; P(x) = 0 \right\} \quad \leq \quad \hat{V}(c_0, c_1, \ldots, c_n).$$

Theorem 3.5 (Budan–Fourier)**.** *Let $P \in \mathbf{R}[X]$ be a polynomial of degree n. The number of roots in $]a,b] \subset \mathbf{R}$, each counted with its multiplicity, satisfies the inequality*

$$\#_{\text{mult}} \left\{ x \in]a,b] \;\middle|\; P(x) = 0 \right\} \quad \leq \quad \hat{V}_a^b(P, P', \ldots, P^{(n)}).$$

If $\mathbf{R}$ is real closed, then the difference $(\text{r.h.s.} - \text{l.h.s.})$ is always an even integer. Equality holds for every interval $]a,b] \subset \mathbf{R}$ if and only if P has n roots in $\mathbf{R}$.

The upper bounds are very easy to compute but they often overestimate the number of roots. This was the state of knowledge before Sturm's ground-breaking discovery in 1829.

3.2. **The Cauchy index.** Index theory is based on judicious counting. Instead of counting zeros of $\frac{P}{Q}$ it is customary to count poles of $\frac{Q}{P}$, which is of course equivalent.

Definition 3.6. We denote by $\lim_a^+ f$ and $\lim_a^- f$ the right and left limit, respectively, of a rational function $f \in \mathbf{R}(X)^*$ in a point $a \in \mathbf{R}$. The *Cauchy index* of f in a is defined as

$$\text{ind}_a(f) := \text{ind}_a^+(f) - \text{ind}_a^-(f) \quad \text{where} \quad \text{ind}_a^\varepsilon(f) := \begin{cases} +\frac{1}{2} & \text{if } \lim_a^\varepsilon f = +\infty, \\ -\frac{1}{2} & \text{if } \lim_a^\varepsilon f = -\infty, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

Less formally, we have $\text{ind}_a(f) = +1$ if f jumps from $-\infty$ to $+\infty$, and $\text{ind}_a(f) = -1$ if f jumps from $+\infty$ to $-\infty$, and $\text{ind}_a(f) = 0$ in all other cases. For example, we have $\text{ind}_0(\frac{1}{x}) = +1$ and $\text{ind}_0(-\frac{1}{x}) = -1$ and $\text{ind}_0(\pm\frac{1}{x^2}) = 0$.

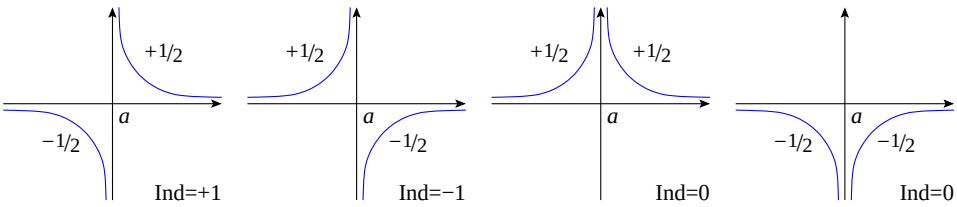


FIGURE 2. A pole a and its Cauchy index $\text{ind}_a(f) = \text{ind}_a^+(f) - \text{ind}_a^-(f)$

Remark 3.7. The limits $\lim_a^\pm f$ are just a convenient notation for purely algebraic quantities: we can factor $f = (X-a)^m g$ with $m \in \mathbb{Z}$ and $g \in \mathbf{R}(X)^*$ such that $g(a) \in \mathbf{R}^*$.

- If $m > 0$, then $\lim_a^\varepsilon f = 0$ for both $\varepsilon \in \{+,-\}$.
- If $m = 0$, then $\lim_a^\varepsilon f = g(a)$ for both $\varepsilon \in \{+,-\}$.
- If $m < 0$, then $\lim_a^\varepsilon f = \varepsilon^m \cdot \text{sign}\, g(a) \cdot (+\infty)$.

In the first case f has a zero of order m in a; for $m \geq 0$ we have $\lim_a^\varepsilon f \in \mathbf{R}$ and thus $\text{ind}_a^\varepsilon(f) = 0$. In the last case f has a pole of order $|m|$ in a, and $\text{ind}_a^\varepsilon(f) = \frac{1}{2}\varepsilon^m \cdot \text{sign}\, g(a)$.

Annotation 3.2. (Rational functions as maps) Here we wish to interpret rational functions $f \in \mathbf{R}(X)$ as maps. The right way to do this is to extend the affine line $\mathbf{R}$ to the projective line $\mathbf{PR} = \mathbf{R} \cup \{\infty\}$.

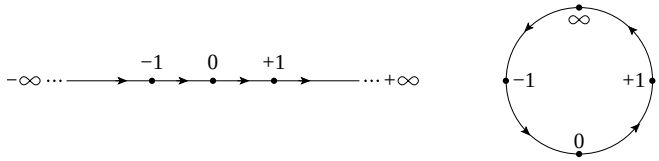
We construct $\mathbf{PR} = (\mathbf{R}^2 \setminus \{0\})/_\sim$ as the quotient of $\mathbf{R}^2 \setminus \{0\}$ by the quivalence $(p,q) \sim (s,t)$ defined by the condition that there exists $u \in \mathbf{R}^*$ such that $(p,q) = (ur,us)$. The equivalence class of (p,q) is denoted by $[p:q]$ and repesents the line passing through the origin $(0,0)$ and (p,q) in $\mathbf{R}^2$. The affine line $\mathbf{R}$ can be identified with $\{[p:1] \mid p \in \mathbf{R}\}$; this covers all points of $\mathbf{PR}$ except one: the point at infinity, $\infty = [1:0]$.

Likewise we construct $\mathbf{PR}(X) = (\mathbf{R}(X)^2 \setminus \{0\})/_\sim$ as the quotient of $\mathbf{R}(X)^2 \setminus \{0\}$ by the quivalence $(P,Q) \sim (R,S)$ defined by the condition that there exists $U \in \mathbf{R}(X)^*$ such that $(P,Q) = (UR,US)$. The equivalence class of (P,Q) is denoted by $[P:Q]$. Here $\mathbf{R}(X)$ can be identified with $\{[P:Q] \mid P,Q \in \mathbf{R}[X], Q \neq 0\}$ using only polynomials. Again this covers all points of $\mathbf{PR}(X)$ except one: the point at infinity, $\infty = [1:0]$.

Consider $f = [P:Q] \in \mathbf{PR}(X)$ with $P,Q \in \mathbf{R}[X]$. We can assume $\gcd(P,Q) = 1$ and set $m =: \max\{\deg P, \deg Q\}$. We then construct homogenous polynomials $\hat{P}, \hat{Q} \in \mathbf{R}[X,Y]$ by $X^k \mapsto X^k Y^{m-k}$. We have $(\hat{P}(x,y), \hat{Q}(x,y)) \neq (0,0)$ for all $(x,y) \neq (0,0)$ in $\mathbf{R}^2$, and the map $\hat{f}\colon \mathbf{PR} \to \mathbf{PR}$ given by $\hat{f}([x:y]) = [\hat{P}(x,y), \hat{Q}(x,y)]$ is well-defined.

This construction allows us to interpret every $f \in \mathbf{PR}(X)$ and in particular every rational fraction $f \in \mathbf{R}(X)$ as a map $\hat{f}\colon \mathbf{PR} \to \mathbf{PR}$. In the sequel most constructions for P/Q resp. $[P:Q]$ are slightly easier in the generic case where $P,Q \in \mathbf{R}[X]^*$, and are then extended to the exceptional cases where $P = 0$ or $Q = 0$.

Annotation 3.3. (Oriented line and circle) We can present the ordered field $\mathbf{R}$ as an oriented line, the two ends being denoted by $-\infty$ and $+\infty$. It is sometimes convenient to formally adjoin two further elements $\pm\infty$ and to extend the order of $\mathbf{R}$ to $\bar{\mathbf{R}} := \mathbf{R} \cup \{\pm\infty\}$ such that $-\infty < x < +\infty$ for all $x \in \mathbf{R}$. This turns $\bar{\mathbf{R}}$ into a closed interval.



We can think of the projective line $\mathbf{PR} = \mathbf{R} \cup \{\infty\}$ as an oriented circle. In the above picture this is obtained by identifying $+\infty$ and $-\infty$ in $\bar{\mathbf{R}}$. Even though we cannot extend the ordering of $\mathbf{R}$ to $\mathbf{PR}$, we can nevertheless define a sign function $\mathbf{PR} \to \{-1,0,+1\}$ by $\operatorname{sign}([p:q]) = \operatorname{sign}(pq)$, which simply means that $\operatorname{sign}(\infty) = 0$.

The intermediate value property now takes the following form: if $f \in \mathbf{R}(X)$ satisfies $f(a)f(b) < 0$ for some $a < b$ in $\mathbf{R}$, then there exists $x \in]a,b[$ such that $\operatorname{sign} f(x) = 0$, that is $f(x) = 0$ or $f(x) = \infty$.

Definition 3.8. For $a < b$ in $\mathbf{R}$ we define the Cauchy index of $f \in \mathbf{R}(X)^*$ on the interval $[a,b]$ by

$$\operatorname{ind}_a^b(f) := \operatorname{ind}_a^+(f) + \sum_{x \in]a,b[} \operatorname{ind}_x(f) - \operatorname{ind}_b^-(f). \tag{3.5}$$

The sum is well-defined because only finitely many $x \in]a,b[$ contribute.

For $b < a$ we define $\operatorname{ind}_a^b(f) := -\operatorname{ind}_b^a(f)$, and for $a = b$ we set $\operatorname{ind}_a^a(f) := 0$.

Finally, we set $\operatorname{ind}_a^b(\frac{R}{S}) := 0$ in the degenerate case where $R = 0$ or $S = 0$.

Remark 3.9. We opt for a more comprehensive definition (3.5) than usual, in order to take care of boundary points. We will frequently bisect intervals, and this technique works best with a uniform definition that avoids case distinctions. Moreover, we will have reason to consider piecewise rational functions in §4.

Proposition 3.10. *The Cauchy index enjoys the following properties (which formally resemble the properties of integration):*

(a) bisection: $\operatorname{ind}_a^b(f) + \operatorname{ind}_b^c(f) = \operatorname{ind}_a^c(f)$ *for all* $a,b,c \in \mathbf{R}$.

(b) invariance: $\operatorname{ind}_a^b(f \circ \tau) = \operatorname{ind}_{\tau(a)}^{\tau(b)}(f)$ *for every linear fractional transformation* $\tau\colon [a,b] \to \mathbf{R}$, $\tau(x) = \frac{px+q}{rx+s}$ *where* $p,q,r,s \in \mathbf{R}$, *without poles on* $[a,b]$.

(c) addition: $\operatorname{ind}_a^b(f+g) = \operatorname{ind}_a^b(f) + \operatorname{ind}_a^b(g)$ *if* f,g *have no common poles.*

(d) scaling: $\operatorname{ind}_a^b(gf) = \sigma \operatorname{ind}_a^b(f)$ *if* $g|_{[a,b]}$ *is of constant sign* $\sigma \in \{\pm 1\}$. □

Annotation 3.4. (Winding number) The Cauchy index $\mathbf{PR}(X) \to \frac{1}{2}\mathbb{Z}$, $f \mapsto \operatorname{ind}_a^b(f)$, counts the number of times that f crosses ∞ from $-$ to $+$ (clockwise in the figure of Annotation 3.3) minus the number of times that f crosses ∞ from $+$ to $-$ (counter-clockwise in the above figure). This geometric interpretation anticipates the winding number of loops in the plane constructed in §4.

Annotation 3.5. (Cauchy functions) Following Cauchy [9] we can define the index $\operatorname{ind}_a^b(f)$ not only for $f \in \mathbf{R}(X)$ but more generally for functions $f\colon [a,b] \to \mathbf{PR} = \mathbf{R} \cup \{\infty\}$ satisfying two natural conditions:

(1) f does not change sign without passing through 0 or ∞.

This allows us to define local indices for isolated poles: we set $\operatorname{ind}_a^+(f) = \frac{1}{2}\operatorname{sign} f(b)$ whenever $f(a) = \infty$ and there exists $b > a$ such that $f(]a,b]) \subset \mathbf{R}^*$: This means that the pole a is isolated on the right. We define $\operatorname{ind}_a^-(f)$ in the same way if the pole a is isolated on the left, and set $\operatorname{ind}_a^\pm(f) = 0$ in all other cases.

(2) f has only a finite number of (semi-)isolated poles in $[a,b]$.

This is needed to define $\operatorname{ind}_a^b(f)$ by a finite sum as in Equation (3.5) above. Examples include fractions $f = r/s$ where $r,s\colon [a,b] \to \mathbf{R}$ are continuous piecewise polynomial functions as in §4.

Example. Over the real numbers $\mathbb{R}$ we can consider functions $f\colon [a,b] \to \mathbb{R} \cup \{\infty\}$ such that for each point $x_0 \in [a,b]$ there exist one-sided neighbourhoods $U = [x_0, x_0+\varepsilon]$ resp. $U = [x_0-\varepsilon, x_0]$ with $\varepsilon > 0$, on which we have $f(x) = (x-x_0)^m g(x)$ with $m \in \mathbb{Z}$ and some continuous function $g\colon U \to \mathbb{R}^*$. Such a function f satisfies conditions (1) and (2), so that we can define its Cauchy index $\operatorname{ind}_a^b(f)$ as above. Examples include fractions $f = R/S$ where $R,S\colon [a,b] \to \mathbb{R}$ are piecewise real-analytic functions.

For emphasis we spell out the following definition:

Definition. We call $f\colon [a,b] \to \mathbf{PR}$ a *Cauchy function* if there exists a subdivision $a = t_0 < t_1 < \cdots < t_n = b$ such that on on each interval $[t_{k-1}, t_k]$ we have $f(x) = (x-t_{k-1})^m (x-t_k)^n g_k(x)$ with $m,n \in \mathbb{Z}$ and some continuous function $g_k\colon [t_{k-1},t_k] \to \mathbf{R}^*$ of constant sign. We can then define $\operatorname{ind}_a^b(f)$ as in Definition 3.8 above.

Annotation 3.6. (Nash functions) The notion of Cauchy function captures the requirements for counting poles as in Equation (3.5) above. If we also want to consider the derivative f', as in §3.3 below, then it suffices to assume each of the local functions g_k to be differentiable. The set of Cauchy functions is stable under taking products and inverses, but not sums. If we want a *ring*, then we should restrict attention to piecewise C^∞ Cauchy functions. This leads us to the classical analytic-algebraic setting:

Example (Nash functions). Let $\mathbf{R}$ be a real closed field. A *Nash function* is a map $f\colon [a,b] \to \mathbf{R}$ that is C^∞ and semi-algebraic [7, chap. 8]. Over the real numbers $\mathbb{R}$ this coincides with the class of real-analytic functions that are algebraic over $\mathbb{R}[X]$. Quotients of piecewise Nash functions are Cauchy functions, and thus seem to be a convenient and natural setting for defining and working with Cauchy indices over real closed fields.

3.3. **Counting real roots.** The ring $\mathbf{R}[X]$ is equipped with a derivation $P \mapsto P'$ sending each polynomial $P = \sum_{k=0}^n p_k X^k$ to its formal derivative $P' = \sum_{k=1}^n k p_k X^{k-1}$. This extends in a unique way to a derivation on the field $\mathbf{R}(X)$ sending $f = \frac{R}{S}$ to $f' = \frac{R'S - RS'}{S^2}$. This is an $\mathbf{R}$-linear map and satisfies Leibniz' rule $(fg)' = f'g + fg'$. For $f \in \mathbf{R}(X)^*$ the quotient f'/f is called the *logarithmic derivative* of f; it enjoys the following property:

Proposition 3.11. *For every $f \in \mathbf{R}(X)^*$ we have $\operatorname{ind}_a(f'/f) = +1$ if a is a zero of f, and $\operatorname{ind}_a(f'/f) = -1$ if a is a pole of f, and $\operatorname{ind}_a(f'/f) = 0$ in all other cases.*

Proof. We have $f = (X-a)^m g$ with $m \in \mathbb{Z}$ and $g \in \mathbf{R}(X)^*$ such that $g(a) \in \mathbf{R}^*$. By Leibniz' rule we obtain $\frac{f'}{f} = \frac{m}{X-a} + \frac{g'}{g}$. The fraction $\frac{g'}{g}$ does not contribute to the index because it does not have a pole in a. We conclude that $\operatorname{ind}_a(f'/f) = \operatorname{sign}(m)$. □

Corollary 3.12. *For every $f \in \mathbf{R}(X)^*$ and $a < b$ in $\mathbf{R}$ the index $\operatorname{ind}_a^b(f'/f)$ is the number of roots minus the number of poles of f in $[a,b]$, counted without multiplicity. Roots and poles on the boundary count for one half.* □

The corollary remains true for $f = \frac{R}{S}$ when $R = 0$ or $S = 0$, with the convention that we count only *isolated* roots and poles. Polynomials $P \in \mathbf{R}[X]$ have no poles, whence $\operatorname{ind}_a^b(P'/P)$ simply counts the number of (isolated) roots of P in $[a,b]$.

3.4. **The inversion formula.** While the Cauchy index can be defined over any ordered field $\mathbf{R}$, the following results require $\mathbf{R}$ to be real closed. The intermediate value property of polynomials $P \in \mathbf{R}[X]$ can then be reformulated quantitatively as $\operatorname{ind}_a^b(\frac{1}{P}) = V_a^b(1,P)$. More generally, we have the following result of Cauchy [9, §I, Thm. I]:

Theorem 3.13. *Let $\mathbf{R}$ be a real closed field, and consider $a < b$ in $\mathbf{R}$. If $P,Q \in \mathbf{R}[X]$ do not have common zeros in a nor b, then*

$$\operatorname{ind}_a^b\Big(\frac{Q}{P}\Big) + \operatorname{ind}_a^b\Big(\frac{P}{Q}\Big) = V_a^b(P,Q). \tag{3.6}$$

The inversion formula of Theorem 3.13 will follow as a special case from the product formula of Theorem 4.6. Its proof is short enough to be given separately here:

Proof. The statement is true if $P = 0$ or $Q = 0$, so we can assume $P, Q \in \mathbf{R}[X]^*$. Equation (3.6) remains valid if we divide P, Q by a common factor $U \in \mathbf{R}[X]$, because our hypothesis ensures that $U(a) \neq 0$ and $U(b) \neq 0$. We can thus assume $\gcd(P, Q) = 1$.

Suppose first that $[a, b]$ contains no pole. On the one hand, both indices $\operatorname{ind}_a^b\left(\frac{Q}{P}\right)$ and $\operatorname{ind}_a^b\left(\frac{P}{Q}\right)$ vanish in the absence of poles. On the other hand, the intermediate value property ensures that both P and Q are of constant sign on $[a, b]$, whence $V_a(P, Q) = V_b(P, Q)$.

Suppose next that $[a, b]$ contains at least one pole. Formula (3.6) is additive with respect to bisection of the interval $[a, b]$. It thus suffices to treat the case where $[a, b]$ contains exactly one pole. Bisecting once more, if necessary, we can assume that this pole is either a or b. Applying the symmetry $X \mapsto a + b - X$, if necessary, we can assume that the pole is a. Since Formula (3.6) is symmetric in P and Q, we can assume that $P(a) = 0$.

By hypothesis we have $Q(a) \neq 0$, whence Q has constant sign on $[a, b]$ and $\operatorname{ind}_a^b\left(\frac{P}{Q}\right) = 0$. Likewise, P has constant sign on $]a, b]$ and $\operatorname{ind}_a^b\left(\frac{Q}{P}\right) = \operatorname{ind}_a^+\left(\frac{Q}{P}\right)$. On the right hand side we find $V_a(P, Q) = \frac{1}{2}$, and for $V_b(P, Q)$ two cases occur:

- If $V_b(P, Q) = 0$, then $\frac{Q}{P} > 0$ on $]a, b]$, whence $\lim_a^+\left(\frac{Q}{P}\right) = +\infty$.
- If $V_b(P, Q) = 1$, then $\frac{Q}{P} < 0$ on $]a, b]$, whence $\lim_a^+\left(\frac{Q}{P}\right) = -\infty$.

In both cases we find $\operatorname{ind}_a^+\left(\frac{Q}{P}\right) = V_a^b(P, Q)$, whence Equation (3.6) holds. □

Annotation 3.7. (Local and global arguments) Reexamining the previous proof we can distinguish a local argument around a pole a, in the neighbourhoods $[a, a + \varepsilon]$ and $[a - \varepsilon, a]$ for some chosen $\varepsilon > 0$, and a global argument, for a given interval $[a, b]$, say without poles. The local argument only uses continuity and is valid for polynomials over any ordered field. It is in the global argument that we need the intermediate value property. This interplay of local and global arguments is a recurrent theme in the proofs of §4.5 and §5.1.

Annotation 3.8. (Reducing fractions) For arbitrary $P, Q \in \mathbf{R}[X]^*$ the inversion formula can be restated as

$$\operatorname{ind}_a^b\left(\tfrac{Q}{P}\right) + \operatorname{ind}_a^b\left(\tfrac{P}{Q}\right) = V_a^b\left(1, \tfrac{Q}{P}\right) = \tfrac{1}{2}\left[\operatorname{sign}\left(\tfrac{Q}{P} \mid b\right) - \operatorname{sign}\left(\tfrac{Q}{P} \mid a\right)\right]$$

with the convention $\operatorname{sign}(\infty) = 0$. This formulation has the advantage to depend only on the fraction $\frac{Q}{P}$ and not on the polynomials P, Q representing it. For reduced fractions we recover the formulation of Theorem 3.13.

Annotation 3.9. (Cauchy functions) The inversion formula holds more generally for all Cauchy functions, as defined in Annotation 3.5. Instead of dividing by $\gcd(P, Q)$, which is in general not defined, we simply divide by common roots or poles, so as to ensure that P, Q have no common roots nor poles on $[a, b]$.

3.5. **Sturm chains.** In the rest of this section we exploit the inversion formula of Theorem 3.13, and we will thus assume $\mathbf{R}$ to be real closed. We can then calculate the Cauchy index $\operatorname{ind}_a^b\left(\frac{R}{S}\right)$ by iterated euclidean division (§3.6). The crucial condition is the following:

Definition 3.14. A sequence of polynomials $(S_0, \ldots, S_n)$ in $\mathbf{R}[X]$ is a *Sturm chain* with respect to an interval $[a, b] \subset \mathbf{R}$ if it satisfies Sturm's condition:

$$\text{If } S_k(x) = 0 \text{ for } 0 < k < n \text{ and } x \in [a, b], \text{ then } S_{k-1}(x) S_{k+1}(x) < 0. \tag{3.7}$$

We will usually not explicitly mention the interval $[a, b]$ if it is understood from the context, or if $(S_0, \ldots, S_n)$ is a Sturm chain on all of $\mathbf{R}$. For $n = 1$ Condition (3.7) is void and should be replaced by the requirement that S_0 and S_1 have no common zeros.

Theorem 3.15. *If $(S_0, S_1, \ldots, S_{n-1}, S_n)$ is a Sturm chain in $\mathbf{R}[X]$, then*

$$\operatorname{ind}_a^b\left(\frac{S_1}{S_0}\right) + \operatorname{ind}_a^b\left(\frac{S_{n-1}}{S_n}\right) = V_a^b\left(S_0, S_1, \ldots, S_{n-1}, S_n\right). \tag{3.8}$$

Proof. The Sturm condition ensures that two consecutive functions S_{k-1} and S_k have no common zeros. For $n = 1$ Formula (3.8) reduces to the inversion formula of Theorem 3.13. For $n = 2$ the inversion formula implies that

$$\operatorname{ind}_a^b\left(\frac{S_1}{S_0}\right) + \operatorname{ind}_a^b\left(\frac{S_0}{S_1}\right) + \operatorname{ind}_a^b\left(\frac{S_2}{S_1}\right) + \operatorname{ind}_a^b\left(\frac{S_1}{S_2}\right) = V_a^b\left(S_0, S_1, S_2\right). \tag{3.9}$$

This is a telescopic sum: contributions to the middle indices arise at zeros of S_1, but at each zero of S_1 its neighbours S_0 and S_2 have opposite signs, which means that the middle terms cancel each other. Iterating this argument, we obtain (3.8) by induction on n. □

The following algebraic criterion will be used in §3.6 and §5.1:

Proposition 3.16. *Consider a sequence $(S_0,\dots,S_n)$ in $\mathbf{R}[X]$ such that*

$$A_kS_{k+1}+B_kS_k+C_kS_{k-1}=0 \qquad \textit{for} \qquad 0<k<n, \tag{3.10}$$

with $A_k,B_k,C_k\in\mathbf{R}[X]$ such that $A_k>0$ and $C_k\geq 0$ on $[a,b]$. Then $(S_0,\dots,S_n)$ is a Sturm chain on $[a,b]$ if and only if the terminal pair (S_{n-1},S_n) has no common zeros in $[a,b]$.

Proof. We assume that $n\geq 2$. If (S_{n-1},S_n) has a common zero, then the Sturm condition (3.7) is obviously violated. Suppose that (S_{n-1},S_n) has no common zeros in $[a,b]$. If $S_k(x)=0$ for $x\in[a,b]$ and $0<k<n$, then $S_{k+1}(x)\neq 0$. Otherwise Condition (3.10) would imply that $S_k,\dots,S_n$ vanish in x, which is excluded by our hypothesis. Now the equation $A_k(x)S_{k+1}(x)+C_k(x)S_{k-1}(x)=0$ with $A_k(x)S_{k+1}(x)\neq 0$ implies $C_k(x)S_{k-1}(x)\neq 0$. Using $A_k(x)>0$ and $C_k(x)>0$ we conclude that $S_{k-1}(x)S_{k+1}(x)<0$. □

Annotation 3.10. (Cauchy functions) Nothing so far is really special to polynomials: Definition 3.14, Theorem 3.15, and Proposition 3.16 extend verbatim to all Cauchy functions as defined in Annotation 3.5.

Annotation 3.11. (Mean value property) Assuming $A_k,C_k>0$ on $[a,b]$, the linear relation (3.10) resembles the mean value property of harmonic functions, here discretized to a graph in form of a chain. Is there a useful generalization of Conditions (3.7) or (3.10) to more general graphs?

Annotation 3.12. (A historic example) For many applications the case $A_k=C_k=1$ suffices, but the general setting is more flexible: A_k and C_k can absorb positive factors and thus purge the polynomials S_{k+1} and S_{k-1} of irrelevancy. The following example is taken from Kronecker (1872) citing Gauss (1849) in his course *Theorie der algebraischen Gleichungen*. [Notes written by Kurt Hensel, archived at the University of Strasbourg, available at `num-scd-ulp.u-strasbg.fr/429`, page 165.]

Example. We consider $P_0=X^7-28X^4+480$ and its derivative $P_1=P_0'=7X^2(X^4-16X)$. We set $S_0=P_0$ and $S_1=X^4-16X$, neglecting the positive factor $7X^2$. We wish to calculate $\operatorname{ind}_a^b(\frac{P_1}{P_0})=\operatorname{ind}_a^b(\frac{S_1}{S_0})$ by constructing a suitable Sturm chain. Euclidean division yields $P_2=(X^3-12)S_1-S_0=192X-480$, which we reduce to $S_2=2X-5$. Likewise $P_3=\frac{1}{16}(8X^3+20X^2+50X-3)S_2-S_1=\frac{15}{16}$ is reduced to $S_3=1$. We thus obtain a judiciously reduced Sturm chain (S_0,S_1,S_2,S_3) of the form $A_kS_{k+1}+B_kS_k+C_kS_{k-1}=0$ with $A_k,C_k>0$.

Annotation 3.13. (Orthogonal polynomials) Sturm sequences naturally occur for real *orthogonal polynomials* $P_0,P_1,P_2,\dots$, where $\deg P_k=k$ for all $k\in\mathbb{N}$. Here is a concrete and simple example:

Example. The sequence of *Legendre polynomials* $P_0,P_1,P_2,\dots$ starting with $P_0=1$ and $P_1=X$ satisfies the recursion $(k+1)P_{k+1}-(2k+1)XP_k+kP_{k-1}=0$ for all $k\geq 1$, and so $(P_0,\dots,P_n)$ is a Sturm chain.

Legendre polynomials are orthogonal with respect to the inner product $\langle f,g\rangle=\int_{-1}^{+1}f(x)g(x)\,\mathrm{d}x$. More generally, one can fix a measure μ on the real line $\mathbb{R}$, say with compact support, and consider the inner product $\langle f,g\rangle=\int f(x)g(x)\,\mathrm{d}\mu$. Orthogonality of $P_0,P_1,P_2,\dots$ means that $\langle P_k,P_\ell\rangle=0$ if $k\neq\ell$, and >0 if $k=\ell$. This entails a three-term recurrence relation $A_kP_{k+1}+B_kP_k+C_kP_{k-1}=0$ with constants $A_k,C_k>0$ and some polynomial B_k of degree 1, depending on k and μ. Orthogonal polynomials thus form a Sturm sequence. It follows that the real roots of each P_n are interlaced with those of its predecessor P_{n-1}, and that each P_n has n distinct real roots, strictly inside the smallest interval that contains the support of μ.

3.6. **Euclidean Sturm chains.** In the preceding paragraph we have defined Sturm chains and applied them to Cauchy indices. Everything so far is fairly general and not limited to polynomials. The crucial observation for polynomials is that the euclidean algorithm can be used to *construct* Sturm chains as follows:

Consider a rational function $f=\frac{R}{S}\in\mathbf{R}(X)^*$ represented by polynomials $R,S\in\mathbf{R}[X]^*$. Iterated euclidean division produces a sequence of polynomials starting with $P_0=S$ and $P_1=R$, such that $P_{k-1}=Q_kP_k-P_{k+1}$ and $\deg P_{k+1}<\deg P_k$ for all $k=1,2,3,\dots$. This process eventually stops when we reach $P_{n+1}=0$, in which case $P_n\sim\gcd(P_0,P_1)$.

Stated differently, this construction is the expansion of f into the continued fraction

$$f = \frac{P_1}{P_0} = \frac{P_1}{Q_1P_1 - P_2} = \frac{1}{Q_1 - \dfrac{P_2}{P_1}} = \frac{1}{Q_1 - \dfrac{1}{Q_2 - \dfrac{P_3}{P_2}}} = \cdots = \cfrac{1}{Q_1 - \cfrac{1}{Q_2 - \cfrac{\ldots}{Q_{n-1} - \cfrac{1}{Q_n}}}}.$$

Definition 3.17. Using the preceding notation, the *euclidean Sturm chain* $(S_0, S_1, \ldots, S_n)$ associated to the fraction $\frac{R}{S} \in \mathbf{R}(X)^*$ is defined by $S_k := P_k/P_n$ for $k = 0, \ldots, n$.

By construction, the chain $(S_0, S_1, \ldots, S_n)$ depends only on the fraction $\frac{R}{S}$ and not on the polynomials R, S chosen to represent it. Division by P_n ensures that $\gcd(S_0, S_1) = S_n = 1$ but preserves the equations $S_{k-1} + S_{k+1} = Q_k S_k$ for all $0 < k < n$. Proposition 3.16 then ensures that $(S_0, S_1, \ldots, S_n)$ is indeed a Sturm chain.

Annotation 3.14. (The euclidean cochain) The polynomials $(Q_1, \ldots, Q_n)$ suffice to reconstruct the fraction f. This presentation is quite economic because they usually have low degree; generically we expect $\deg(Q_k) = 1$.

We recover $(S_0, S_1, \ldots, S_n)$ working backwards from $S_{n+1} = 0$ and $S_n = 1$ by calculating $S_{k-1} = Q_k S_k - S_{k+1}$ for all $k = n-1, \ldots, 0$. This procedure also provides an economic way to evaluate $(S_0, S_1, \ldots, S_n)$ at $a \in \mathbf{R}$.

This indicates that, from an algorithmic point of view, the cochain $(Q_1, \ldots, Q_n)$ is of primary interest. From a mathematical point of view it is more convenient to use the chain $(S_0, S_1, \ldots, S_n)$.

Remark 3.18 (euclidean division). If $\mathbf{K}$ is a field, then for every $S \in \mathbf{K}[X]$ and $P \in \mathbf{K}[X]^*$ there exists a unique pair $Q, R \in \mathbf{K}[X]$ such that

$$S = PQ - R \qquad \text{and} \qquad \deg R < \deg P. \tag{3.11}$$

Here the negative sign has been chosen for the application to Sturm chains. Euclidean division works over every ring $\mathbf{K}$ provided that the leading coefficient c of P is invertible in $\mathbf{K}$. In general we can carry out pseudo-euclidean division: for all $S \in \mathbf{K}[X]$ and $P \in \mathbf{K}[X]^*$ over some integral ring $\mathbf{K}$ there exists a unique pair $Q, R \in \mathbf{K}[X]$ such that

$$c^d S = PQ - R \qquad \text{and} \qquad \deg R < \deg P, \tag{3.12}$$

where c is the leading coefficient of P and $d = \max\{0, 1 + \deg S - \deg P\}$. With a view to ordered fields it is advantageous to chose the exponent d to be even. (This is easy to achieve: if d is odd, then multiply Q and R by c and augment d by 1.) This will be applied in §5.1 to the polynomial ring $\mathbf{R}[Y, X] = \mathbf{K}[X]$ over $\mathbf{K} = \mathbf{R}[Y]$. Even for $\mathbb{Q}[X]$ it is often more efficient to work in $\mathbb{Z}[X]$ in order to avoid coefficient swell, see [18, §6.12].

Annotation 3.15. (Pseudo-euclidean division) For every ring $\mathbf{K}$, the degree $\deg\colon \mathbf{K}[X] \to \mathbb{N} \cup \{-\infty\}$ satisfies:

(1) $\deg(P+Q) \leq \sup\{\deg P, \deg Q\}$, with equality iff $\deg P \neq \deg Q$ or $\operatorname{lc}(P) + \operatorname{lc}(Q) \neq 0$.
(2) $\deg(PQ) \leq \deg P + \deg Q$, with equality iff $P = 0$ or $Q = 0$ or $\operatorname{lc}(P) \cdot \operatorname{lc}(Q) \neq 0$.

If $\mathbf{K}$ is integral, then $\deg(PQ) = \deg P + \deg Q$ and $\operatorname{lc}(PQ) = \operatorname{lc}(P) \cdot \operatorname{lc}(Q)$ for all $P, Q \in \mathbf{K}[X]^*$, and the polynomial ring $\mathbf{K}[X]$ is again integral. Moreover, for every $S \in \mathbf{K}[X]$ and $P \in \mathbf{K}[X]^*$ there exists a unique pair $Q, R \in \mathbf{K}[X]$ such that $c^d S = PQ - R$ and $\deg R < \deg P$, where $c = \operatorname{lc}(P)$ and $d = \max\{0, 1 + \deg S - \deg P\}$.

Existence: We proceed by induction on d. If $d = 0$, then $\deg S < \deg P$ and $Q = 0$ and $R = S$ suffice. If $d \geq 1$, then we set $M := \operatorname{lc}(S) \cdot X^{\deg S - \deg P}$ and $\tilde{S} := cS - PM$. We see that $\deg(S) = \deg(cS) = \deg(PM)$ and $\operatorname{lc}(cS) = \operatorname{lc}(PM)$, whence $\deg \tilde{S} < \deg S$. By hypothesis there exists $\tilde{Q}, R \in A[X]$ such that $c^{d-1}\tilde{S} = P\tilde{Q} + R$. We conclude that $c^d S = c^{d-1}\tilde{S} + c^{d-1}PM = PQ + R$ with $Q = \tilde{Q} + c^{d-1}M$.

Uniqueness: For $PQ + R = PQ' + R'$ with $\deg R < \deg P$ and $\deg R' < \deg P$, we find $P(Q - Q') = R' - R$, whence $\deg P + \deg(Q - Q') = \deg[P(Q - Q')] = \deg(R - R') < \deg P$. This is only possible for $\deg(Q - Q') < 0$, which means $Q - Q' = 0$. We conclude that $Q = Q'$ and $R = R'$.

Annotation 3.16. (Cauchy functions) The euclidean construction is tailor-made for polynomials, but perhaps it can be generalized to other classes of Cauchy functions. More explicitly, consider real-analytic functions $S_0, S_1\colon [a,b] \to \mathbb{R}$ or Nash functions $[a,b] \to \mathbf{R}$ over some real closed field $\mathbf{R}$. Even if a gcd is in general not defined, we can still eliminate common zeros. Is there some natural way to construct a sequence $(S_0, S_1, \ldots, S_n)$ satisfying $A_k S_{k+1} + B_k S_k + C_k S_{k-1} = 0$ as in Proposition 3.16 such that S_n has no zeros on $[a,b]$?

3.7. **Sturm's theorem.** Using the euclidean algorithm for constructing Sturm chains, we can now fix the following notation:

Definition 3.19. For $\frac{R}{S} \in \mathbf{R}(X)$ and $a,b \in \mathbf{R}$ we define the *Sturm index* to be

$$\operatorname{Sturm}_a^b\Big(\frac{R}{S}\Big) := V_a^b\big(S_0,S_1,\ldots,S_n\big),$$

where $(S_0,S_1,\ldots,S_n)$ is the euclidean Sturm chain associated to $\frac{R}{S}$. We include two exceptional cases: If $S=0$ and $R\neq 0$, the euclidean Sturm chain is $(0,1)$ of length $n=1$. If $R=0$, we take the chain (1) of length $n=0$. In both cases we obtain $\operatorname{Sturm}_a^b\big(\frac{R}{S}\big)=0$.

This definition is effective in the sense that the Sturm index $\operatorname{Sturm}_a^b\big(\frac{R}{S}\big)$ can immediately be calculated. Definition 3.8 of the Cauchy index $\operatorname{ind}_a^b\big(\frac{R}{S}\big)$, however, assumes knowledge of all roots of S in $[a,b]$. This difficulty is overcome by Sturm's celebrated theorem, equating the Cauchy index with the Sturm index over a real closed field:

Theorem 3.20 (Sturm 1829/35, Cauchy 1831/37)**.** *For every pair of polynomials $R,S \in \mathbf{R}[X]$ over a real closed field $\mathbf{R}$ we have*

$$\operatorname{ind}_a^b\Big(\frac{R}{S}\Big) = \operatorname{Sturm}_a^b\Big(\frac{R}{S}\Big). \tag{3.13}$$

Proof. Equation (3.13) is trivially true if $R=0$ or $S=0$, according to our definitions. We can thus assume $R,S\in\mathbf{R}[X]^*$. Let $(S_0,S_1,\ldots,S_n)$ be the euclidean Sturm chain associated to the fraction $\frac{R}{S}$. Since $\frac{R}{S}=\frac{S_1}{S_0}$ and $S_n=1$, Theorem 3.15 implies that

$$\operatorname{ind}_a^b\Big(\frac{R}{S}\Big) = \operatorname{ind}_a^b\Big(\frac{S_1}{S_0}\Big) + \operatorname{ind}_a^b\Big(\frac{S_{n-1}}{S_n}\Big) = V_a^b\big(S_0,S_1,\ldots,S_n\big) = \operatorname{Sturm}_a^b\Big(\frac{R}{S}\Big). \qquad \square$$

Remark 3.21. Sturm's theorem can be seen as an algebraic analogue of the fundamental theorem of calculus (or Stokes' theorem): it reduces a 1-dimensional counting problem on the interval $[a,b]$ to a 0-dimensional counting problem on the boundary $\{a,b\}$. We are most interested in the former, but the latter has the advantage of being easily calculable. Both become equal via the intermediate value property. In §4 we will generalize this to the complex realm, reducing a 2-dimensional counting problem on a rectangle Γ to a 1-dimensional counting problem on the boundary $\partial\Gamma$. This can be further generalized to arbitrary dimension, leading to an algebraic version of Kronecker's index [15].

Remark 3.22. Sturm's theorem is usually stated under two additional hypotheses, namely $\gcd(R,S)=1$ and $S(a)S(b)\neq 0$. Our formulation of Theorem 3.20 does not require any of these hypotheses, instead they are absorbed into our slightly refined definitions. The hypothesis $\gcd(R,S)=1$ is circumvented by formulating Definitions 3.8 and 3.19 such that both indices become well-defined on $\mathbf{R}(X)$. The case $S(a)S(b)=0$ is anticipated in Definitions 3.2 and 3.6 by counting boundary points correctly. Arranging these details is not only an aesthetic preoccupation: it clears the way for a uniform treatment of the complex case in §4 and ensures a simpler algorithmic formulation.

As an immediate consequence we obtain Sturm's classical theorem [52, §2]:

Corollary 3.23 (Sturm 1829/35)**.** *For every polynomial $P\in\mathbf{R}[X]^*$ we have*

$$\#\big\{x\in[a,b] \;\big|\; P(x)=0\big\} = \operatorname{ind}_a^b\Big(\frac{P'}{P}\Big) = \operatorname{Sturm}_a^b\Big(\frac{P'}{P}\Big), \tag{3.14}$$

where roots on the boundary count for one half. $\square$

Remark 3.24. The intermediate value property is essential. Over the field $\mathbb{Q}$ of rational numbers, for example, the function $f(x)=2x/(x^2-2)$ has no poles, whence $\operatorname{ind}_1^2(f)=0$. A Sturm chain is given by $S_0=X^2-2$ and $S_1=2X$ and $S_2=2$, whence $V_1^2(S_0,S_1,S_2)=1$. Thus the Sturm index does not count roots resp. poles in $\mathbb{Q}$ but in the real closure $\mathbb{Q}^c$.

Remark 3.25. By the usual bisection method, Formula (3.14) provides an algorithm to locate all real roots of any given real polynomial. Once the roots are well separated, one can switch to Newton's method (§6.3), which is simpler to apply and converges much faster – but vitally depends on good starting values.

Annotation 3.17. (Transformation invariance) If $f,g \in \mathbf{R}(X)$ and g has no poles in $[a,b]$, then $\operatorname{Sturm}_a^b(f \circ g) = \operatorname{Sturm}_{g(a)}^{g(b)}(f)$. If $\mathbf{R}$ is real closed, then $\operatorname{ind}_a^b(f \circ g) = \operatorname{ind}_{g(a)}^{g(b)}(f)$. To see this, assume $f = R/S$ and $g = P/Q$ with $P,Q,R,S \in \mathbf{R}[X]$ such that $\gcd(P,Q) = 1$ and $\gcd(R,S) = 1$. Since g has no poles, Q has no roots in $[a,b]$. If $(S_0,S_1,\dots,S_n)$ in $\mathbf{R}[X]$ is a Sturm chain on $[a,b]$, then so is $(P_0,P_1,\dots,P_n)$ defined by $P_k = Q^m S_k(P/Q)$ with $m = \max\{\deg S_0,\dots,\deg S_n\}$. Applied to the euclidean Sturm chain $(S_0,S_1,\dots,S_n)$ of $f = R/S$ this yields

$$\begin{aligned}\operatorname{Sturm}_{g(a)}^{g(b)}(f) = \operatorname{Sturm}_{g(a)}^{g(b)}\Big(\frac{S_1}{S_0}\Big) = V_{g(a)}^{g(b)}(S_0,S_1,\dots,S_n) &= V_a^b\big(S_0(P/Q),S_1(P/Q),\dots,S_n(P/Q)\big)\\ = V_a^b(P_0,P_1,\dots,P_n) = \operatorname{Sturm}_a^b\Big(\frac{P_1}{P_0}\Big) &= \operatorname{Sturm}_a^b\Big(\frac{S_1(P/Q)}{S_0(P/Q)}\Big) = \operatorname{Sturm}_a^b(f\circ g).\end{aligned}$$

We now conclude by Theorem 3.20. Again, the intermediate value property is essential. Consider for example $f(x) = \frac{1}{x-2}$ and $g(x) = x^2$ over $\mathbb{Q}$. Then $\operatorname{ind}_1^2(f \circ g) = 0$ differs from $\operatorname{ind}_{g(1)}^{g(2)}(f) = 1$.

4. Cauchy's theorem for complex polynomials

We continue to work over a real closed field $\mathbf{R}$ and consider its complex extension $\mathbf{C} = \mathbf{R}[i]$ where $i^2 = -1$. In this section we define the algebraic winding number $w(\gamma)$ for piecewise polynomial loops $\gamma\colon [0,1] \to \mathbf{C}$ and study in particular the winding number $w(F|\partial\Gamma)$ of a polynomial $F \in \mathbf{C}[Z]$ along the boundary of a rectangle $\Gamma \subset \mathbf{C}$. We then establish Cauchy's theorem (Corollary 4.10) stating that $w(F|\partial\Gamma)$ counts the number of roots of F in Γ.

Remark 4.1. Nowadays the winding number is most often defined via Cauchy's integral formula $w(F|\partial\Gamma) = \frac{1}{2\pi i}\int_{\partial\Gamma} \frac{F'(z)}{F(z)}\,\mathrm{d}z$. In his residue calculus of complex functions, Cauchy [8, 9] also described the algebraic calculation presented below. In the present article, we use exclusively the algebraic winding number and develop an independent, entirely algebraic proof. The real product formula, Theorem 4.6, seems to be new. The complex product formula, Corollaries 4.8, is well-known in the analytic setting using Cauchy's integral, but the algebraic approach reveals two noteworthy extensions:

- The algebraic construction is not restricted to the complex numbers $\mathbb{C} = \mathbb{R}[i]$ but works for $\mathbf{C} = \mathbf{R}[i]$ over an arbitrary real closed field $\mathbf{R}$.
- Unlike Cauchy's integral formula, the algebraic winding number can cope with roots of F on the boundary $\partial\Gamma$, as pointed out in the introduction.

4.1. **Real and complex fields.** Let $\mathbf{R}$ be an ordered field. For every $x \in \mathbf{R}$ we have $x^2 \geq 0$, whence $x^2 + 1 > 0$. The polynomial $X^2 + 1$ is thus irreducible in $\mathbf{R}[X]$, and the quotient $\mathbf{C} = \mathbf{R}[X]/(X^2+1)$ is a field. It is denoted by $\mathbf{C} = \mathbf{R}[i]$ with $i^2 = -1$. Each element $z \in \mathbf{C}$ can be uniquely written as $z = x + yi$ with $x,y \in \mathbf{R}$. We can thus identify $\mathbf{C}$ with $\mathbf{R}^2$ via the map $\mathbf{R}^2 \to \mathbf{C}$, $(x,y) \mapsto z = x + yi$, and define $\operatorname{re}(z) := x$ and $\operatorname{im}(z) := y$.

Using this notation, addition and multiplication in $\mathbf{C}$ are given by

$$\begin{aligned}(x+yi) + (x'+y'i) &= (x+x') + (y+y')i,\\ (x+yi)\cdot(x'+y'i) &= (xx'-yy') + (xy'+x'y)i.\end{aligned}$$

The ring automorphism $\mathbf{R}[X] \to \mathbf{R}[X]$, $X \mapsto -X$, fixes X^2+1 and thus descends to a field automorphism $\mathbf{C} \to \mathbf{C}$ that maps each $z = x + yi$ to its conjugate $\overline{z} = x - yi$. We have $\operatorname{re}(z) = \frac{1}{2}(z+\overline{z})$ and $\operatorname{im}(z) = \frac{1}{2i}(z-\overline{z})$. The product $z\overline{z} = x^2 + y^2 \geq 0$ vanishes if and only if $z = 0$. For $z \neq 0$ we thus find $z^{-1} = \frac{\overline{z}}{z\overline{z}} = \frac{x}{x^2+y^2} - \frac{y}{x^2+y^2}i$.

If $\mathbf{R}$ is real closed, then every $x \in \mathbf{R}_{\geq 0}$ has a square root $\sqrt{x} \geq 0$. For $z \in \mathbf{C}$ we can thus define $|z| := \sqrt{z\overline{z}}$, which extends the absolute value of $\mathbf{R}$. For all $u,v \in \mathbf{C}$ we have:

(0) $|\operatorname{re}(u)| \leq |u|$ and $|\operatorname{im}(u)| \leq |u|$.

(1) $|u| \geq 0$, and $|u| = 0$ if and only if $u = 0$.
(2) $|u \cdot v| = |u| \cdot |v|$ and $|\overline{u}| = |u|$.
(3) $|u+v| \leq |u| + |v|$.

All verifications are straightforward. The triangle inequality (3) can be derived from the preceding properties as follows. If $u+v=0$, then (3) follows from (1). If $u+v \neq 0$, then $1 = \frac{u}{u+v} + \frac{v}{u+v}$, and applying (0) and (2) we find

$$1 = \mathrm{re}\big(\tfrac{u}{u+v}\big) + \mathrm{re}\big(\tfrac{v}{u+v}\big) \leq \big|\tfrac{u}{u+v}\big| + \big|\tfrac{v}{u+v}\big| = \tfrac{|u|}{|u+v|} + \tfrac{|v|}{|u+v|}.$$

4.2. **Real and complex variables.** Just as we identify $(x,y) \in \mathbf{R}^2$ with $z = x+iy \in \mathbf{C}$, we consider $\mathbf{C}[Z]$ as a subring of $\mathbf{C}[X,Y]$ with $Z = X+iY$. The conjugation on $\mathbf{C}$ extends to a ring automorphism of $\mathbf{C}[X,Y]$ fixing X and Y, so that the conjugate of $Z = X+iY$ is $\overline{Z} = X - iY$. In this sense X and Y are real variables, whereas Z is a complex variable.

Every polynomial $F \in \mathbf{C}[X,Y]$ can be uniquely decomposed as $F = R + iS$ with $R,S \in \mathbf{R}[X,Y]$, namely $R = \mathrm{re}\,F := \frac{1}{2}(F+\overline{F})$ and $S = \mathrm{im}\,F := \frac{1}{2i}(F-\overline{F})$. In particular we thus recover the familiar formulae $X = \mathrm{re}\,Z$ and $Y = \mathrm{im}\,Z$.

For $F,G \in \mathbf{C}[X,Y]$ we set $F \circ G := F(\mathrm{re}\,G, \mathrm{im}\,G)$. The map $F \mapsto F \circ G$ is the unique ring endomorphism $\mathbf{C}[X,Y] \to \mathbf{C}[X,Y]$ that maps $Z \mapsto G$ and is equivariant with respect to conjugation, because $Z \mapsto G$ and $\overline{Z} \mapsto \overline{G}$ are equivalent to $X \mapsto \mathrm{re}\,G$ and $Y \mapsto \mathrm{im}\,G$.

4.3. **The algebraic winding number.** Given a polynomial $P \in \mathbf{C}[X]$ and two parameters $a < b$ in $\mathbf{R}$, the map $\gamma\colon [a,b] \to \mathbf{C}$ defined by $\gamma(x) = P(x)$ describes a polynomial path in $\mathbf{C}$. We define its winding number $w(\gamma)$ to be half the Cauchy index of $\frac{\mathrm{re}\,P}{\mathrm{im}\,P}$ on $[a,b]$:

$$w(P|[a,b]) := \tfrac{1}{2}\,\mathrm{ind}_a^b\big(\tfrac{\mathrm{re}\,P}{\mathrm{im}\,P}\big).$$

Remark 4.2. The definition is geometrically motivated as follows. Assuming that $\gamma(x) \neq 0$ for all $x \in [a,b]$, the winding number $w(\gamma)$ counts the number of turns that γ performs around 0: it changes by $+\frac{1}{2}$ each time γ crosses the real axis in counter-clockwise direction, and by $-\frac{1}{2}$ if the passage is clockwise. Our algebraic definition is slightly more comprehensive than the geometric one since it does not exclude zeros of γ.

More generally, we can consider a subdivision $a = x_0 < x_1 < \cdots < x_n = b$ in $\mathbf{R}$ and polynomials $P_1, \ldots, P_n \in \mathbf{C}[X]$ that satisfy $P_k(x_k) = P_{k+1}(x_k)$ for $k = 1, \ldots, n-1$. This defines a continuous, piecewise polynomial path $\gamma\colon [a,b] \to \mathbf{C}$ by $\gamma(x) := P_k(x)$ for $x \in [x_{k-1}, x_k]$. If $\gamma(a) = \gamma(b)$, then γ is a *loop*, i.e., a closed path. Its winding number is defined by

$$w(\gamma) := \sum_{k=1}^{n} w(P_k|[x_{k-1}, x_k]).$$

This is well-defined according to Proposition 3.10(a), because the winding number $w(\gamma)$ depends only on the path γ and not on the subdivision chosen to describe it.

4.4. **Rectangles.** Given $a,b \in \mathbf{C}$, the map $\gamma\colon [0,1] \to \mathbf{C}$ defined by $\gamma(x) = a + x(b-a)$ joins $\gamma(0) = a$ and $\gamma(1) = b$ by a straight line segment. Its image will be denoted by $[a,b]$. For $F \in \mathbf{C}[X,Y]$ we set $w(F|[a,b]) := w(F \circ \gamma)$ or, stated differently,

$$w(F|[a,b]) := w(F \circ G|[0,1]) \quad \text{where} \quad G = a + X(b-a).$$

This is the winding number of the path traced by $F(z)$ as z runs from a straight to b. For the reverse orientation we obtain $w(F|[b,a]) = -w(F|[a,b])$ according to Proposition 3.10(b).

A *rectangle* (with sides parallel to the axes) is a subset $\Gamma = [x_0,x_1] \times [y_0,y_1]$ in $\mathbf{C} = \mathbf{R}^2$ with $x_0 < x_1$ and $y_0 < y_1$ in $\mathbf{R}$. Its *interior* is $\mathrm{Int}\,\Gamma =]x_0,x_1[\times]y_0,y_1[$. Its *boundary* $\partial\Gamma$ consists of the four vertices $a = (x_0,y_0)$, $b = (x_1,y_0)$, $c = (x_1,y_1)$, $d = (x_0,y_1)$, and the four edges $[a,b]$, $[b,c]$, $[c,d]$, $[d,a]$ between them (see Figure 1).

Definition 4.3. Given a polynomial $F \in \mathbf{C}[X,Y]$ and a rectangle $\Gamma \subset \mathbf{C}$, we define the *algebraic winding number* as $w(F|\partial\Gamma) := w(F|[a,b]) + w(F|[b,c]) + w(F|[c,d]) + w(F|[d,a])$.

Stated differently, we have $w(F|\partial\Gamma) = w(F \circ \gamma)$ where the path $\gamma\colon [0,4] \to \mathbf{C}$ linearly interpolates between the vertices $\gamma(0) = a$, $\gamma(1) = b$, $\gamma(2) = c$, $\gamma(3) = d$, and $\gamma(4) = a$.

Proposition 4.4 (bisection property)**.** *Suppose that we bisect* $\Gamma = [x_0,x_2] \times [y_0,y_2]$

- *horizontally into* $\Gamma' = [x_0,x_1] \times [y_0,y_2]$ *and* $\Gamma'' = [x_1,x_2] \times [y_0,y_2]$,
- *or vertically into* $\Gamma' = [x_0,x_2] \times [y_0,y_1]$ *and* $\Gamma'' = [x_0,x_2] \times [y_1,y_2]$

where $x_0 < x_1 < x_2$ *and* $y_0 < y_1 < y_2$*. Then* $w(F|\partial\Gamma) = w(F|\partial\Gamma') + w(F|\partial\Gamma'')$.

Proof. This follows from Definition 4.3 by one-dimensional bisection and internal cancellation using Proposition 3.10. □

Proposition 4.5. *For a linear polynomial* $F = Z - z_0$ *with* $z_0 \in \mathbf{C}$ *we find*

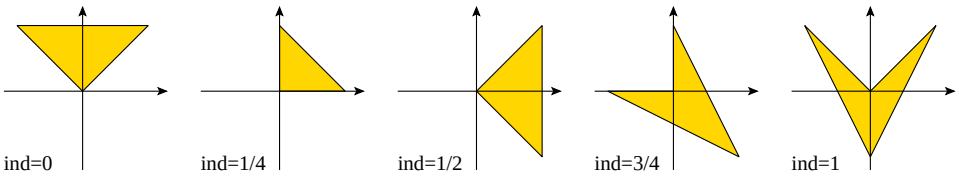
$$w(F|\partial\Gamma) = \begin{cases} 1 & \text{if } z_0 \text{ is in the interior of } \Gamma, \\ \frac{1}{2} & \text{if } z_0 \text{ is in one of the edges of } \Gamma, \\ \frac{1}{4} & \text{if } z_0 \text{ is in one of the vertices of } \Gamma, \\ 0 & \text{if } z_0 \text{ is in the exterior of } \Gamma. \end{cases}$$

Proof. By bisection, all configurations can be reduced to the case where z_0 is a vertex of Γ. By symmetry, translation, and homothety we can assume that $z_0 = a = 0$, $b = 1$, $c = 1+i$, $d = i$. Here an easy explicit calculation shows that $w(F|\partial\Gamma) = \frac{1}{4}$ by adding

$$\begin{aligned} w(F|[a,b]) &= w(X|[0,1]) = \tfrac{1}{2}\operatorname{ind}_0^1(\tfrac{X}{0}) = 0, \\ w(F|[b,c]) &= w(1+iX|[0,1]) = \tfrac{1}{2}\operatorname{ind}_0^1(\tfrac{1}{X}) = \tfrac{1}{4}, \\ w(F|[c,d]) &= w(1+i-X|[0,1]) = \tfrac{1}{2}\operatorname{ind}_0^1(\tfrac{1-X}{1}) = 0, \\ w(F|[d,a]) &= w(i-iX|[0,1]) = \tfrac{1}{2}\operatorname{ind}_0^1(\tfrac{0}{1-X}) = 0. \end{aligned}$$ □

Annotation 4.1. (Normalization) The factor $\frac{1}{2}$ in the definition of the winding number compared to the Cauchy index is chosen so as to achieve the normalization of Proposition 4.5. It also has a natural geometric interpretation. Compare the circle $\mathbf{S} = \{z \in \mathbf{C}\colon |z| = 1\}$ with the projective line $\mathbf{PR}$ of Annotation 3.3. The winding number $w(\gamma)$ of a path $\gamma\colon [0,1] \to \mathbf{C}^*$ is defined using the map $q\colon \mathbf{C}^* \to \mathbf{PR}$, $(x,y) \mapsto [x:y]$. The quotient map q is the composition of the deformation retraction $r\colon \mathbf{C}^* \to \mathbf{S}$, $z \mapsto z/|z|$, and the two-fold covering $p\colon \mathbf{S} \to \mathbf{PR}$, $(x,y) \mapsto [x:y]$. This means that *one* full circle in $\mathbf{C}^*$ maps to *two* full circles in $\mathbf{PR}$.

Annotation 4.2. (Angles) Proposition 4.5 generalizes from rectangles to convex polygons, and then to arbitrary polygons by suitable subdivision. The only subtlety occurs when z_0 is a vertex of the boundary $\partial\Gamma$: in general, we find $w(F|\partial\Gamma) \in \{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$, and one can easily construct examples showing that all possibilities are realized:



These examples illustrate how the result depends on the angle at 0 and its incidence with the real axis. The reference to the real axis breaks the rotational symmetry, and so $w(\gamma)$ may differ from $w(c\gamma)$ for some $c \in \mathbf{C}$, $|c| = 1$. Over $\mathbb{C}$ the average value $\overline{w}(\gamma) = \int_0^1 w(e^{2\pi it}\gamma)\,\mathrm{d}t \in [0,1]$ measures the angle at 0. For $\mathbf{C} = \mathbf{R}[i]$ over a real closed field $\mathbf{R}$ we can likewise define $\overline{w}(\gamma) := \lim_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N-1} w(e^{2\pi i/N}\gamma) \in \mathbb{R}$ for every piecewise polynomial loop $\gamma\colon [0,1] \to \mathbf{C}$. Measuring angles in this way does not follow the paradigm of effective calculation expounded here, but the definition of $\overline{w}(\gamma)$ might be useful in some other context. For the purpose of this article, however, it is only an amusing curiosity and will not be further developed.

4.5. **The product formula.** The product of two polynomials $F = P + iQ$ and $G = R + iS$ with $P,Q,R,S \in \mathbf{R}[X]$ is given by $FG = (PR - QS) + i(PS + QR)$. The following result relates the Cauchy indices of $\frac{P}{Q}$ and $\frac{R}{S}$ to that of $\frac{PR-QS}{PS+QR}$.

Theorem 4.6 (real product formula)**.** *Consider polynomials $P,Q,R,S \in \mathbf{R}[X]$ such that neither (P,Q) nor (R,S) have common roots in $a,b \in \mathbf{R}$. Then we have*

$$\operatorname{ind}_a^b\Big(\frac{PR-QS}{PS+QR}\Big) = \operatorname{ind}_a^b\Big(\frac{P}{Q}\Big) + \operatorname{ind}_a^b\Big(\frac{R}{S}\Big) - V_a^b\Big(1, \frac{P}{Q}+\frac{R}{S}\Big). \tag{4.1}$$

Remark 4.7. We have $\frac{P}{Q}+\frac{R}{S} = \frac{PS+QR}{QS} = \frac{\operatorname{im}(FG)}{\operatorname{im}(F)\operatorname{im}(G)}$. After simplification we find

$$V_a^b\big(1, \tfrac{PS+QR}{QS}\big) = \tfrac{1}{2}\big[\operatorname{sign}\big(\tfrac{PS+QR}{QS} \mid X \mapsto b\big) - \operatorname{sign}\big(\tfrac{PS+QR}{QS} \mid X \mapsto a\big)\big].$$

If a or b is a pole, this is evaluated using the convention $\operatorname{sign}(\infty) = 0$. For $(P=0, Q=1)$ or $(R=0, S=1)$ Theorem 4.6 reduces to the inversion formula of Theorem 3.13.

Proof. We can assume that $\gcd(P,Q) = \gcd(R,S) = 1$. If $Q=0$ or $S=0$ or $PS+QR=0$ then Formula (4.1) trivially holds, so we can assume $Q,S,PS+QR \in \mathbf{R}[X]^*$. Suppose first that $[a,b]$ does not contain any poles, that is, roots of the denominators Q, S, $PS+QR$. On the one hand, all three indices vanish in the absence of poles. On the other hand, the intermediate value property ensures that Q, S, and $PS+QR$ are of constant sign on $[a,b]$, whence $V_a^b\big(1, \frac{PS+QR}{QS}\big) = 0$.

Suppose next that $[a,b]$ contains at least one pole. Formula (4.1) is additive with respect to bisection of the interval $[a,b]$. We can thus assume that $[a,b]$ contains only one pole. Bisecting once more, if necessary, we can assume that this pole is either a or b. Applying the symmetry $X \mapsto a+b-X$, if necessary, we can assume that the pole is a. We thus have $V_a^b = \frac{1}{2}\operatorname{sign}(\frac{P}{Q}+\frac{R}{S} \mid X \mapsto b)$ and Q, S, $PS+QR$ are of constant sign on $]a,b]$. Applying the symmetry $(P,Q,R,S) \mapsto (P,-Q,R,-S)$, if necessary, we can assume that $V_a^b = +\frac{1}{2}$, which means that $\frac{P}{Q}+\frac{R}{S} > 0$ on $]a,b]$. We distinguish three cases:

First case. Suppose first that either $Q(a)=0$ or $S(a)=0$. Applying the symmetry $(P,Q,R,S) \mapsto (R,S,P,Q)$, if necessary, we can assume that $Q(a)=0$ and $S(a)\neq 0$. Then $PS+QR$ does not vanish in a, whence $\operatorname{ind}_a^b\big(\frac{PR-QS}{PS+QR}\big) = \operatorname{ind}_a^b\big(\frac{R}{S}\big) = 0$. We have $\lim_a^+ \frac{P}{Q} = \lim_a^+\big(\frac{P}{Q}+\frac{R}{S}\big) = +\infty$, whence $\operatorname{ind}_a^b\big(\frac{P}{Q}\big) = +\frac{1}{2}$ and Formula (4.1) holds.

Second case. Suppose that $PS+QR$ vanishes in a, but $Q(a)\neq 0$ and $S(a)\neq 0$. Then $\operatorname{ind}_a^b\big(\frac{P}{Q}\big) = \operatorname{ind}_a^b\big(\frac{R}{S}\big) = 0$, and we only have to study the pole of

$$\frac{PR-QS}{PS+QR} = \frac{\frac{P}{Q}\cdot\frac{R}{S}-1}{\frac{P}{Q}+\frac{R}{S}}. \tag{4.2}$$

In a the denominator vanishes and the numerator is negative:

$$\tfrac{P(a)}{Q(a)} + \tfrac{R(a)}{S(a)} = 0, \quad \text{whence} \quad \tfrac{P(a)}{Q(a)}\cdot\tfrac{R(a)}{S(a)} - 1 = -\tfrac{P^2(a)}{Q^2(a)} - 1 < 0.$$

This implies $\lim_a^+ \frac{PR-QS}{PS+QR} = -\infty$, whence $\operatorname{ind}_a^b\big(\frac{PR-QS}{PS+QR}\big) = -\frac{1}{2}$ and Formula (4.1) holds.

Third case. Suppose that a is a common pole of $\frac{P}{Q}$ and $\frac{R}{S}$, whence also of $\frac{PR-QS}{PS+QR}$. Since $\frac{P}{Q}+\frac{R}{S} > 0$ on $]a,b]$, we have $\lim_a^+ \frac{P}{Q} = +\infty$ or $\lim_a^+ \frac{R}{S} = +\infty$. Equation (4.2) implies that $\lim_a^+\big(\frac{PR-QS}{PS+QR}\big) = +\lim_a^+\big(\frac{P}{Q}\big)\cdot\lim_a^+\big(\frac{R}{S}\big)$. In each case Formula (4.1) holds. □

Corollary 4.8 (complex product formula)**.** *If $F,G \in \mathbf{C}[X,Y]$ do not vanish in any of the vertices of the rectangle $\Gamma \subset \mathbf{R}^2$, then $w(F\cdot G|\partial\Gamma) = w(F|\partial\Gamma) + w(G|\partial\Gamma)$.*

Proof. This follows from the real product formula of Theorem 4.6 and the fact that the boundary $\partial\Gamma$ forms a closed path. By excluding roots on the vertices we ensure that at each vertex both boundary contributions cancel each other. □

Remark 4.9. The same argument applies to the product of any two piecewise polynomial loops $\gamma_1,\gamma_2\colon [0,1] \to \mathbf{C}$, provided that vertices are not mapped to 0. This proves the multiplicativity (W2) stated in Theorem 1.2: $w(\gamma_1\cdot\gamma_2) = w(\gamma_1) + w(\gamma_2)$.

Corollary 4.10 (root counting)**.** *Consider a polynomial $F \in \mathbf{C}[Z]^*$ that splits into linear factors, such that $F = c(Z-z_1)\cdots(Z-z_n)$ for some $c, z_1, \ldots, z_n \in \mathbf{C}$. If none of the roots lies on a vertex of Γ, then $w(F|\partial\Gamma)$ counts the number of roots in Γ. Roots in the interior count with their multiplicity; roots on the boundary count with half their multiplicity.* □

Remark 4.11. In the preceding corollaries we explicitly exclude roots on the vertices in order to apply the real product formula (Theorem 4.6). One might wonder whether this is an artefact of our proof. While the degree 1 case of Proposition 4.5 is easy (and useful) there is no such simple rule in degree ≥ 2. As an illustration consider $\Gamma = [0,1] \times [0,1]$ and $F_t = Z \cdot (Z - 2 - it)$: here F_t has one root $z_1 = 0$ on a vertex and one root $z_2 = 2 + it$ outside of Γ. After a little calculation we find $w(F_1|\partial\Gamma) = 0$ and $w(F_0|\partial\Gamma) = \frac{1}{4}$ and $w(F_{-1}|\partial\Gamma) = \frac{1}{2}$. This shows that, in this degenerate case, the algebraic winding number depends on the configuration of all roots and not only on the roots in Γ. We will not further pursue this question, which is only of marginal interest, and simply exclude roots on the vertices. We emphasize once again that roots on the edges pose no problem.

Annotation 4.3. (Roots on vertices) Roots on vertices are special because our arbitrary reference to the real axis breaks the rotational symmetry, as illustrated in Annotation 4.2. The average winding number $\overline{w}(\gamma)$ of a piecewise polynomial path $\gamma\colon [0,1] \to \mathbf{C}$ repairs this defect by restoring rotational symmetry, such that $\overline{w}(\gamma_1\gamma_2) = \overline{w}(\gamma_1) + \overline{w}(\gamma_2)$ even if zeros happen to lie on vertices. For every polynomial $F \in \mathbf{C}[Z]^*$ and every polygonal domain $\Gamma \subset \mathbf{C}$, the average winding number $\overline{w}(F|\partial\Gamma)$ thus counts the number of roots of F in Γ, such that each root counts with α times its multiplicity, where $\alpha \in [0,1]$ measures the angle at the zero in Γ. For example, $\alpha \in \{1, \frac{1}{2}, \frac{1}{4}\}$ if Γ is a rectangle and the zero lies in $\operatorname{Int}\Gamma$, in an edge, or on a vertex, respectively.

Remark 4.12. If we assume that $\mathbf{C}$ is algebraically closed, then *every* polynomial $F \in \mathbf{C}[Z]$ factors as required in Corollary 4.10. So if you prefer some other existence proof for the roots, then you may skip the next section and still benefit from root location (Theorem 1.11). This seems to be the point of view adopted by Cauchy [8, 9] in 1831/37, which may explain why he did not attempt to use his index for a constructive proof of the Fundamental Theorem of Algebra. (In 1820 he had already given a non-constructive proof, see §7.8.1.) In 1836 Sturm and Liouville [55, 53] proposed to extend Cauchy's algebraic method for root counting so as to obtain an existence proof. This is our aim in the next section.

5. The Fundamental Theorem of Algebra

We continue to consider a real closed field $\mathbf{R}$ and its complex extension $\mathbf{C} = \mathbf{R}[i]$ where $i^2 = -1$. In the preceding sections we have constructed the algebraic winding number $w(F|\partial\Gamma)$ for $F \in \mathbf{C}[Z]^*$ and $\Gamma \subset \mathbf{C}$, and derived its multiplicativity. We can now establish our main result: an effective, real-algebraic proof of the Fundamental Theorem of Algebra.

Remark 5.1. The proof that we present here is inspired by classical arguments, based on the winding number of loops in the complex plane. The idea goes back to Gauss' dissertation (see §7.2) and has been much elaborated since. For $\mathbf{C} = \mathbf{R}[i]$ over a real closed field $\mathbf{R}$, the algebraic proof of Theorem 5.3 seems to be new.

5.1. **The winding number in the absence of zeros.** The crucial step is to show that $w(F|\partial\Gamma) \neq 0$ implies that F has a root in Γ. By contraposition, we will show that $w(F|\partial\Gamma) = 0$ whenever F has no zeros in Γ. The local version is easy:

Lemma 5.2 (local version)**.** *If $F \in \mathbf{C}[X,Y]$ satisfies $F(x,y) \neq 0$ for some point $(x,y) \in \mathbf{R}^2$, then there exists $\delta > 0$ such that $w(F|\partial\Gamma) = 0$ for every $\Gamma \subset [x-\delta, x+\delta] \times [y-\delta, y+\delta]$.*

Annotation 5.1. A proof can be improvised as follows. Suppose first that $\operatorname{im} F(x,y) > 0$. By continuity there exists $\delta > 0$ such that $\operatorname{im} F > 0$ on the rectangle $U = [x-\delta, x+\delta] \times [y-\delta, y+\delta]$. For every $\Gamma \subset U$ we then have $w(F|\partial\Gamma) = 0$. The case $\operatorname{im} F(x,y) < 0$ is analogous. If $\operatorname{im} F(x,y) = 0$ then our hypothesis ensures that $\operatorname{re} F(x,y) \neq 0$. Again there exists $\delta > 0$ such that $\operatorname{re} F \neq 0$ on the rectangle $U = [x-\delta, x+\delta] \times [y-\delta, y+\delta]$. Now Corollary 4.8 shows that $w(F|\partial\Gamma) = w(iF|\partial\Gamma) = 0$ as in the first case. The following detailed proof makes the choice of δ explicit and thus avoids case distinctions and the appeal to continuity.

Proof. Let us make the standard continuity argument explicit. For all $s,t \in \mathbf{R}$ we have $F(x+s,y+t) = a + \sum_{j+k\geq 1} a_{jk} s^j t^k$ with $a = F(x,y) \neq 0$ and certain coefficients $a_{jk} \in \mathbf{C}$. We set $M := \max \sqrt[j+k]{|a_{jk}/a|}$, so that $|a_{jk}| \leq |a| \cdot M^{j+k}$. For $\delta := \frac{1}{4M}$ and $|s|, |t| \leq \delta$ we find

$$\Big| \sum_{j+k\geq 1} a_{jk} s^j t^k \Big| \leq \sum_{n\geq 1} \sum_{j+k=n} |a| \cdot M^{j+k} \cdot |s|^j \cdot |t|^k \leq |a| \sum_{n\geq 1} (n+1)\big(\tfrac{1}{4}\big)^n = \tfrac{7}{9}|a|. \tag{5.1}$$

This shows that F does not vanish on $U := [x-\delta, x+\delta] \times [y-\delta, y+\delta]$. Corollary 4.8 ensures that $w(F|\partial\Gamma) = w(cF|\partial\Gamma)$ for every rectangle $\Gamma \subset U$ and every constant $c \in \mathbf{C}^*$. Choosing $c = i/a$ we can assume that $F(x,y) = i$. The Estimate (5.1) then shows that $\operatorname{im} F > 0$ on U, whence $w(F|\partial\Gamma) = 0$ for every rectangle $\Gamma \subset U$. □

While the preceding local lemma uses only continuity and holds over every ordered field, the following global version requires the field $\mathbf{R}$ to be real closed.

Theorem 5.3 (global version)**.** *Let $\Gamma = [x_0, x_1] \times [y_0, y_1]$ be a rectangle in $\mathbf{C}$. If $F \in \mathbf{C}[X,Y]$ satisfies $F(x,y) \neq 0$ for all $(x,y) \in \Gamma$, then $w(F|\partial\Gamma) = 0$.*

We remark that over the real numbers $\mathbb{R}$, a short proof can be given as follows:

Compactness proof. The rectangle $\Gamma = [x_0, x_1] \times [y_0, y_1]$ is covered by the family of open sets $U(x,y) =]x-\delta, x+\delta[\times]y-\delta, y+\delta[$ of Lemma 5.2, where δ depends on (x,y). Compactness of Γ ensures that there exists $\lambda > 0$, called a Lebesgue number of the cover, such that every rectangle $\Gamma' \subset \Gamma$ of diameter $< \lambda$ is contained in some $U(x,y)$. For all subdivisions $x_0 = s_0 < s_1 < \cdots < s_m = x_1$ and $y_0 = t_0 < t_1 < \cdots < t_n = y_1$, the bisection property ensures that $w(F|\partial\Gamma) = \sum_{j=1}^m \sum_{k=1}^n w(F|\partial\Gamma_{jk})$ where $\Gamma_{jk} = [s_{j-1}, s_j] \times [t_{k-1}, t_k]$. For $s_j = x_0 + j\frac{x_1 - x_0}{m}$ and $t_k = y_0 + k\frac{y_1 - y_0}{n}$ with m, n sufficiently large, each Γ_{jk} has diameter $< \lambda$, so Lemma 5.2 implies that $w(F|\partial\Gamma_{jk}) = 0$ for all j,k, whence $w(F|\partial\Gamma) = 0$. □

The preceding compactness argument applies only to the field $\mathbb{C} = \mathbb{R}[i]$ of complex numbers over $\mathbb{R}$ (§2.1) and not to an arbitrary real closed field (§2.2). In particular, it is no longer elementary in the sense that it uses a second-order property (§2.3). We therefore provide an elementary real-algebraic proof using Sturm chains:

Algebraic proof. Each $F \in \mathbf{C}[X,Y]$ can be written as $F = \sum_{k=0}^n f_k X^k$ with $f_k \in \mathbf{C}[Y]$. In this way we consider $\mathbf{R}[X,Y] = \mathbf{R}[Y][X]$ as a polynomial ring in one variable X over $\mathbf{R}[Y]$. Starting with $S_0, S_1 \in \mathbf{R}[X,Y]$ such that $\frac{S_1}{S_0} = \frac{\operatorname{re} F}{\operatorname{im} F}$, pseudo-euclidean division in $\mathbf{R}[Y][X]$, as explained in Remark 3.18, produces a chain $(S_0, \ldots, S_n)$ such that $c_k^2 S_{k-1} = Q_k S_k - S_{k+1}$ for some $Q_k \in \mathbf{R}[Y][X]$ and $c_k \in \mathbf{R}[Y]^*$ and $\deg_X S_{k+1} < \deg_X S_k$. We end up with $S_{n+1} = 0$ and $S_n \in \mathbf{R}[Y]^*$ for some n. (If $\deg_X S_n > 0$, then $\gcd(S_0, S_1)$ in $\mathbf{R}(Y)[X]$ is of positive degree and we can reduce the initial fraction $\frac{S_1}{S_0}$.)

Regular case. Assume first that S_n does not vanish in $[y_0, y_1]$. Proposition 3.16 ensures that specializing $(S_0, \ldots, S_n)$ in $Y \mapsto y \in [y_0, y_1]$ yields a Sturm chain in $\mathbf{R}[X]$, and likewise specializing $(S_0, \ldots, S_n)$ in $X \mapsto x \in [x_0, x_1]$ yields a Sturm chain in $\mathbf{R}[Y]$. In the sum over all four edges of Γ, all contributions cancel each other in pairs:

$$\begin{aligned} 2w(F|\partial\Gamma) &= +\operatorname{ind}_{x_0}^{x_1}\big(\tfrac{\operatorname{re} F}{\operatorname{im} F} \,\big|\, Y \mapsto y_0\big) + \operatorname{ind}_{y_0}^{y_1}\big(\tfrac{\operatorname{re} F}{\operatorname{im} F} \,\big|\, X \mapsto x_1\big) \\ &\quad + \operatorname{ind}_{x_1}^{x_0}\big(\tfrac{\operatorname{re} F}{\operatorname{im} F} \,\big|\, Y \mapsto y_1\big) + \operatorname{ind}_{y_1}^{y_0}\big(\tfrac{\operatorname{re} F}{\operatorname{im} F} \,\big|\, X \mapsto x_0\big) \\ &= +V_{x_0}^{x_1}\big(S_0, \ldots, S_n \,\big|\, Y \mapsto y_0\big) + V_{y_0}^{y_1}\big(S_0, \ldots, S_n \,\big|\, X \mapsto x_1\big) \\ &\quad + V_{x_1}^{x_0}\big(S_0, \ldots, S_n \,\big|\, Y \mapsto y_1\big) + V_{y_1}^{y_0}\big(S_0, \ldots, S_n \,\big|\, X \mapsto x_0\big) = 0. \end{aligned}$$

Singular case. In general we have to cope with a finite set $\mathscr{Y} \subset [y_0, y_1]$ of roots of S_n. We can change the rôles of X and Y and apply the euclidean algorithm in $\mathbf{R}[X][Y]$; this leads to a finite set of roots $\mathscr{X} \subset [x_0, x_1]$. We obtain a finite set $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ of singular points in Γ, where both chains fail. (These points are potential zeros of F.)

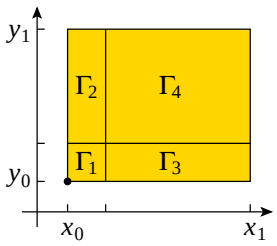


FIGURE 3. Isolating a singular point (x_0,y_0) within $\Gamma = [x_0,x_1] \times [y_0,y_1]$

By subdivision and symmetry we can assume that (x_0,y_0) is the only singular point in our rectangle $\Gamma = [x_0,x_1] \times [y_0,y_1]$. By hypothesis F does not vanish in (x_0,y_0), so we can apply Lemma 5.2 to $\Gamma_1 = [x_0,x_0+\delta] \times [y_0,y_0+\delta]$ with $\delta > 0$ sufficiently small such that $w(F|\partial\Gamma_1) = 0$. The remaining three rectangles $\Gamma_2 = [x_0,x_0+\delta] \times [y_0+\delta,y_1]$ and $\Gamma_3 = [x_0+\delta,x_1] \times [y_0,y_0+\delta]$ and $\Gamma_4 = [x_0+\delta,x_1] \times [y_0+\delta,y_1]$ do not contain any singular points, such that $w(F|\partial\Gamma_j) = 0$ by appealing to the regular case.

Summing over all sub-rectangles we conclude that $w(F|\partial\Gamma) = 0$. □

Annotation 5.2. The construction of the chain $(S_0,\dots,S_n)$ in $\mathbf{R}[Y][X]$ decreases the degree in X but usually increases the degree in Y. Here S_n is some crude form of the resultant of S_0 and S_1. We are rather careless about degrees here, and the usual approach via (sub)resultants would give better control. The crucial point in the proof, however, is that we can specialize $(S_0,\dots,S_n)$ in either X or Y and obtain a Sturm chain in the remaining variable, in the sense of Definition 3.14, by appealing to the algebraic criterion of Proposition 3.16. For subresultants a similar double specialization argument is less obvious and deserves further study.

5.2. **Counting complex roots.** The following result generalizes the real root count (§3.3) to complex roots.

Theorem 5.4. *Consider a polynomial $F \in \mathbf{C}[Z]^*$ and a rectangle $\Gamma \subset \mathbf{C}$ such that F does not vanish in the vertices of Γ. Then the winding number $w(F|\partial\Gamma)$ counts the number of roots of F in Γ. Roots on the boundary count for one half.*

Proof. We can factor $F = (Z-z_1)\cdots(Z-z_m)G$ such that $G \in \mathbf{C}[Z]^*$ has no roots in $\mathbf{C}$. The assertion follows from the product formula of Corollary 4.8. Each linear factor $(Z-z_k)$ contributes to the winding number as stated in Proposition 4.5. The factor G does not contribute to the winding number according to Theorem 5.3. (We will prove below that $m = \deg F$ and $G \in \mathbf{C}^*$.) □

Annotation 5.3. (Hypotheses) This corollary extends Sturm's theorem counting real roots, see Corollary 3.23. In both cases the intermediate value property of $\mathbf{R}$ is essential, see Remark 3.24. As a counterexample consider $\mathbf{R} = \mathbb{Q}$ and $\mathbf{C} = \mathbb{Q}[i]$. The winding number of $F = Z^2 - i$ in $\mathbf{C}[Z]$ with respect to $\Gamma = [0,1] \times [0,1] \subset \mathbf{C}$ is $w(F|\partial\Gamma) = 1$. This corresponds to the root $\frac{1}{2}\sqrt{2} + \frac{i}{2}\sqrt{2}$. Of course, this root does not lie in $\Gamma \subset \mathbb{Q}[i]$ but in $\mathbb{Q}^c[i]$.

Annotation 5.4. (Counting roots and poles of rational functions) We have focused on polynomials $F \in \mathbf{C}[Z]$, but Definition 4.3 of the winding number and the product formula of Corollary 4.8 immediately extend to rational functions $F \in \mathbf{C}(Z)$. It is then an easy matter to establish the following generalization:

Theorem. *Consider a rational function $F \in \mathbf{C}(Z)$ and a rectangle $\Gamma \subset \mathbf{C}$ such that the vertices of Γ are neither roots nor poles of F. Then $w(F|\partial\Gamma)$ counts the number of roots minus the number of poles of F in Γ. Boundary points count for one half.* □

5.3. **Homotopy invariance.** We wish to show that the winding number $w(F_t|\partial\Gamma)$ does not change if we deform F_0 to F_1. To make this precise we consider $F \in \mathbf{C}[Z,T]$ and denote by F_t the polynomial in $\mathbf{C}[Z]$ obtained by specializing $T \mapsto t \in [0,1]$.

Theorem 5.5. *Suppose that $F \in \mathbf{C}[Z,T]$ is such that for each $t \in [0,1]$ the polynomial $F_t \in \mathbf{C}[Z]$ has no roots on $\partial\Gamma$. Then $w(F_0|\partial\Gamma) = w(F_1|\partial\Gamma)$.*

Proof. Over the rectangle $\Gamma \subset \mathbf{C}$ with vertices $a,b,c,d \in \mathbf{C}$ we consider the cube $\Gamma \times [0,1]$ with vertices $a_0 = (a,0)$, $a_1 = (a,1)$, etc. The bottom rectangle $\Gamma_0 = \Gamma \times \{0\}$ has vertices a_0, b_0, c_0, d_0, whereas the top rectangle $\Gamma_1 = \Gamma \times \{1\}$ has vertices a_1, b_1, c_1, d_1.

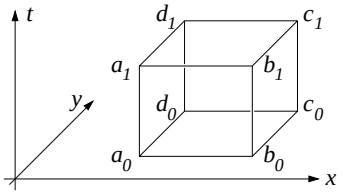


FIGURE 4. The cube $\Gamma \times [0,1]$ in $\mathbf{C} \times \mathbf{R}$

We can consider the polynomial $F \in \mathbf{C}[Z,T]$ as a map $\mathbf{C} \times \mathbf{R} \to \mathbf{C}$. By hypothesis F has no zero on $\partial\Gamma \times [0,1]$. Over each edge of Γ, say $[a,b]$, we have a rectangle $\tilde{\Gamma} = [a,b] \times [0,1]$. In the absence of zeros, Theorem 5.3 ensures that $w(F|\partial\tilde{\Gamma}) = 0$, that is,

$$w(F|[a_0,b_0]) - w(F|[a_1,b_1]) = w(F|[a_0,a_1]) - w(F|[b_0,b_1]).$$

In the sum over all four edges of Γ the terms on the right hand side cancel each other in pairs. We conclude that $w(F|\partial\Gamma_0) - w(F|\partial\Gamma_1) = 0$. □

Remark 5.6. The same argument holds for every piecewise polynomial homotopy $H\colon [0,1] \times [0,1] \to \mathbf{C}^*$ where $\gamma_t\colon [0,1] \to \mathbf{C}^*$, $\gamma_t(x) = H(x,t)$, is a closed path for each $t \in [0,1]$. This proves the homotopy invariance (W3) stated in Theorem 1.2: $w(\gamma_0) = w(\gamma_1)$.

5.4. **The global winding number of a polynomial.** Having all tools in hand, we can now prove Theorem 1.10, stating that $w(F|\partial\Gamma) = \deg F$ for every polynomial $F \in \mathbf{C}[Z]^*$ and every sufficiently large rectangle Γ. This can be quantified by Cauchy's bound:

Definition 5.7. For $F = c_0 + c_1 Z + \cdots + c_{n-1}Z^{n-1} + c_n Z^n$ in $\mathbf{C}[Z]$ with $c_n \neq 0$ we set $M = \max\{0, |c_0|, |c_1|, \ldots, |c_{n-1}|\}$ and define the *Cauchy radius* to be $\rho_F := 1 + M/|c_n|$.

Proposition 5.8. *If $z \in \mathbf{C}$ satisfies $|z| \geq \rho_F$, then $|F(z)| \geq |c_n| > 0$. Hence all roots of F in $\mathbf{C}$ are contained in the Cauchy disk $B(\rho_F) = \{z \in \mathbf{C} \mid |z| < \rho_F\}$.*

Proof. The assertion is true for $F = c_n Z^n$ where $M = 0$ and $\rho_F = 1$. In the sequel we can thus assume $M > 0$ and $\rho_F > 1$. For all $z \in \mathbf{C}$ satisfying $|z| \geq \rho_F$ we find

$$\begin{aligned} |F(z) - c_n z^n| &= |c_0 + c_1 z + \cdots + c_{n-1} z^{n-1}| \leq |c_0| + |c_1||z| + \cdots + |c_{n-1}||z^{n-1}| \\ &\leq M + M|z| + \cdots + M|z|^{n-1} = M\tfrac{|z|^n - 1}{|z| - 1} \leq |c_n|(|z|^n - 1). \end{aligned}$$

For the last inequality notice that $|z| \geq \rho_F$ implies $|z| - 1 \geq \rho_F - 1 = M/|c_n|$. We have

$$\begin{aligned} |c_n z^n| = |c_n z^n - F(z) + F(z)| \leq |c_n z^n - F(z)| + |F(z)|, \quad \text{whence} \\ |F(z)| \geq |c_n z^n| - |F(z) - c_n z^n| \geq |c_n||z|^n - |c_n|(|z|^n - 1) = |c_n| > 0. \end{aligned}$$

□

This proposition holds over any ordered field $\mathbf{R}$ because it uses only $|a+b| \leq |a| + |b|$ and $|a \cdot b| \leq |a| \cdot |b|$. It is not an existence result but only an a priori bound: if F has roots in $\mathbf{C}$, then they necessarily lie in $B(\rho_F)$. Now, over a real closed field $\mathbf{R}$, the winding number allows us to count all roots of F in $\mathbf{C}$ and to establish the desired conclusion:

Theorem 5.9. *For every polynomial $F \in \mathbf{C}[Z]^*$ and every rectangle $\Gamma \subset \mathbf{C}$ containing the Cauchy disk $B(\rho_F)$ we have $w(F|\partial\Gamma) = \deg F$.*

Proof. Given a polynomial $F = c_n Z^n + c_{n-1}Z^{n-1} + \cdots + c_0$ with $c_n \neq 0$ we deform $F_1 = F$ to $F_0 = c_n Z^n$ via $F_t = c_n Z^n + t(c_{n-1}Z^{n-1} + \cdots + c_0)$. For each $t \in [0,1]$ the Cauchy radius of F_t is $\rho_t = 1 + tM/|c_n|$, which shrinks from $\rho_1 = \rho_F$ to $\rho_0 = 1$. By the previous proposition, the polynomial $F_t \in \mathbf{C}[Z]$ has no roots on $\partial\Gamma$. We can thus apply Theorems 5.5 and 5.4 to conclude that $w(F_1|\partial\Gamma) = w(F_0|\partial\Gamma) = n$. □

This completes the proof of the Fundamental Theorem of Algebra: on the one hand Theorem 5.9 says that $w(F|\partial\Gamma) = \deg F$ provided that $\Gamma \supset B(\rho_F)$, and on the other hand Theorem 5.4 says that $w(F|\partial\Gamma)$ equals the number of roots of F in $\Gamma \subset \mathbf{C}$.

Annotation 5.5. (Degree bounds) The Fundamental Theorem of Algebra, in the form that we have just proven, states that if the field $\mathbf{R}$ is real closed, i.e., every polynomial $P \in \mathbf{R}[X]$ satisfies the intermediate value property over $\mathbf{R}$, then the field $\mathbf{C} = \mathbf{R}[i]$ is algebraically closed, i.e., every polynomial $F \in \mathbf{C}[Z]$ splits into linear factors over $\mathbf{C}$. Since we are working exclusively with polynomials, it is natural to study degree bounds.

We call an ordered field $\mathbf{R}$ *real d-closed* if every polynomial $P \in \mathbf{R}[X]$ of degree $\leq d$ satisfies the intermediate value property over $\mathbf{R}$. Likewise, we call a field $\mathbf{C}$ *algebraically d-closed* if every polynomial $F \in \mathbf{C}[Z]$ of degree $\leq d$ splits into linear factors over $\mathbf{C}$. It is easy to establish the following implication: if $\mathbf{R}$ is an ordered field such that $\mathbf{R}[i]$ is algebraically d-closed, then $\mathbf{R}$ is real d-closed. The converse seems to be open:

Question. If $\mathbf{R}$ is real d-closed, does this imply that $\mathbf{R}[i]$ is algebraically d-closed?

This is trivally true for $d = 1$. The answer is also affirmative for $d = 2,3,4$ because quadratic, cubic, and quartic equations can be solved by radicals of degree $n \leq d$, i.e., roots of $Z^n - c_0$ with $c_0 \in \mathbf{C}$, and these roots can be constructed in $\mathbf{R}[i]$ if $\mathbf{R}$ is real n-closed. Quartic equations can be reduced to auxialiary equations of degree ≤ 3, so if $\mathbf{R}$ is real 3-closed, then $\mathbf{R}[i]$ is algebraically 4-closed and $\mathbf{R}$ is in fact real 4-closed!

What happens in degree 5 and higher? An affirmative answer would be surprising… but a Galois-type obstruction seems unlikely, too. The arguments of this article immediately extend to refined versions with the desired degree bounds – the only exception is our algebraic proof of Theorem 5.3, where we construct a Sturm sequence in $\mathbf{R}[X,Y]$ with little control on the degrees. It seems to be an interesting research project to investigate this phenomenon in full depth and to prove optimal degree bounds.

6. Algorithmic aspects

The preceding development shows how to derive Cauchy's algebraic method for locating the roots of a complex polynomial, and this section discusses algorithmic questions.

Remark 6.1. The algorithm described here is often attributed to Wilf [66] in 1978, but it was already explicitly described by Sturm [53] and Cauchy [9] in the 1830s. It can also be found in Runge's *Encyklopädie* article [34, Band I, §I-B3a6] in 1898. Numerical variants are known as *Weyl's quadtree method* (1924) or *Lehmer's method* (1969), see §7.9. I propose to call it *Cauchy's method*, or *Cauchy's algebraic method* if emphasis is needed to differentiate it from Cauchy's analytic method using integration. For the theory of complex polynomials see Marden [33], Henrici [22], and Rahman–Schmeisser [40]; the latter contains extensive historical notes and an up-to-date guide to the literature.

6.1. **Turing computability.** The theory of ordered or orderable fields, nowadays called *real algebra*, was initiated by Artin and Schreier [3, 4] in the 1920s, culminating in Artin's solution [1] of Hilbert's 17th problem. Since the 1970s real-algebraic geometry is flourishing anew, see Bochnak–Coste–Roy [7], and with the advent of computers algorithmic and quantitative aspects have regained importance, see Basu–Pollak–Roy [5]. Sinaceur [49] presents a detailed history of Sturm's theorem and its multiple metamorphoses.

Definition 6.2. We say that an ordered field $(\mathbf{R}, +, \cdot, <)$ can be implemented on a Turing machine if each element $a \in \mathbf{R}$ can be coded as input/output for such a machine and each of the field operations $(a,b) \mapsto a+b$, $a \mapsto -a$, $(a,b) \mapsto a \cdot b$, $a \mapsto a^{-1}$ as well as the comparisons $a = b$, $a < b$ can be carried out by a uniform algorithm.

Example 6.3. The field $(\mathbb{R}, +, \cdot, <)$ of real numbers cannot be implemented on a Turing machine because the set $\mathbb{R}$ is uncountable: it is impossible to code all real numbers by finite strings over a finite alphabet, as required for input/output. This argument is independent of the chosen representation. If we insist on representing each and every real number, then this fundamental obstacle can only be circumvented by considering a hypothetical *real number machine* [6], which transcends the traditional setting of Turing machines.

Example 6.4. The subset $\mathbb{R}_{\text{comp}} \subset \mathbb{R}$ of computable real numbers, as defined by Turing [58] in his famous 1936 article, forms a countable, real closed subfield of $\mathbb{R}$. Each computable number a can be represented as input/output for a universal Turing machine by

an algorithm that approximates a to any desired precision. This overcomes the obstacle of the previous example by restriction to $\mathbb{R}_{\text{comp}}$. Unfortunately, not all operations of $(\mathbb{R}_{\text{comp}},+,\cdot,<)$ can be implemented: there exists no algorithm that for each computable real number a, given in form of an algorithm, determines whether $a=0$, or more generally determines the sign of a. (This is an instance of the notorious Entscheidungsproblem.)

Example 6.5. The algebraic closure $\mathbb{Q}^c$ of $\mathbb{Q}$ in $\mathbb{R}$ is, by definition, a real closed field; it is the smallest real closed field in the sense that it is contained in every real closed field. Unlike the field $\mathbb{R}_{\text{comp}}$ of computable real numbers, the much smaller field $(\mathbb{Q}^c,+,\cdot,<)$ can be implemented on a Turing machine [44, 43].

6.2. **A global root-finding algorithm.** We consider a complex polynomial

$$F = c_0 + c_1 Z + \cdots + c_n Z^n \quad \text{in} \quad \mathbb{C}[Z]$$

that we assume to be *implementable*, that is, we require the ordered field

$$\mathbb{Q}(\mathrm{re}(c_0),\mathrm{im}(c_0),\mathrm{re}(c_1),\mathrm{im}(c_1),\ldots,\mathrm{re}(c_n),\mathrm{im}(c_n)) \subset \mathbb{R}$$

to be implementable in the preceding sense. We begin with the following preparations:

- We divide F by $\gcd(F,F')$ to ensure that all roots of F are simple.
- We determine $r \in \mathbb{N}$ such that all roots of F are contained in the disk $B(r)$.

The following notation will be convenient: a 0-*cell* is a singleton $\{a\}$ with $a \in \mathbb{C}$; a 1-*cell* is an open line segment, either vertical $\{x_0\}\times]y_0,y_1[$ or horizontal $]x_0,x_1[\times\{y_0\}$ with $x_0<x_1$ and $y_0<y_1$ in $\mathbb{R}$; a 2-*cell* is an open rectangle $]x_0,x_1[\times]y_0,y_1[$ in $\mathbb{C}$.

It is immediate to check whether a 0-cell contains a root of F. Sturm's theorem (Corollary 3.23) allows us to count the roots of F in a 1-cell $]a,b[$: for $G=F(a+X(b-a))$ in $\mathbb{C}[X]$ calculate $P=\gcd(\mathrm{re}\,G,\mathrm{im}\,G)$ in $\mathbb{R}[X]$ and count roots of P in $]0,1[$. Cauchy's theorem (Corollary 4.10) allows us to count the roots in a 2-cell. In both cases the crucial subalgorithm is the computation of Sturm chains which we will discuss in §6.4 below.

Building on this, the root-finding algorithm successively refines a list $L_j=\{\Gamma_1,\ldots,\Gamma_{n_j}\}$ of disjoint cells such that:

- Each root of F is contained in exactly one cell $\Gamma\in L_j$.
- Each cell $\Gamma\in L_j$ contains at least one root of F.
- Each cell $\Gamma\in L_j$ has diameter $\le 3r\cdot 2^{-j}$.

More explicitly, the algorithm proceeds as follows:
We initialize $L_0=\{\Gamma\}$ with the square $\Gamma=]-r,+r[\times]-r,+r[$.
Given L_j we construct L_{j+1} by treating each cell $\Gamma\in L_j$ as follows:

(0) If Γ is a 0-cell, then retain Γ.
(1) If Γ is a 1-cell, then bisect Γ into two 1-cells of equal length.
Retain each new 1-cell that contains a root of F.
Retain the new 0-cell if it contains a root of F.
(2) If Γ is a 2-cell, then bisect Γ into four 2-cells of equal size.
Retain each new 2-cell that contains a root of F.
Retain each new 1-cell that contains a root of F.
Retain the new 0-cell if it contains a root of F.

Collecting all retained cells we obtain the new list L_{j+1}. After some initial iterations all roots will lie in disjoint cells $\Gamma_1,\ldots,\Gamma_n$, each containing precisely one root. Taking the midpoint $u_k\in\Gamma_k$, this can be seen as n approximate roots $u_1,\ldots,u_n$ each with an error bound $\delta_k\le\frac{3}{2}r\cdot 2^{-j}$ such that each u_k is δ_k-close to a root of F.

6.3. **Cross-over to Newton's local method.** For $F\in\mathbb{C}[Z]$ Newton's method consists in iterating the map $\Phi\colon \mathbb{C}\smallsetminus\mathscr{Z}(F')\to\mathbb{C}$ given by $\Phi(z)=z-F(z)/F'(z)$. Its strength resides in the following well-known property:

Theorem 6.6. *The fixed points of Newton's map Φ are the simple zeros of F, that is, $z_0 \in \mathbb{C}$ such that $F(z_0) = 0$ and $F'(z_0) \neq 0$. For each fixed point z_0 there exists $\delta > 0$ such that every initial value $u_0 \in B(z_0, \delta)$ satisfies $|\Phi^n(u_0) - z_0| \leq 2^{1-2^n} \cdot |u_0 - z_0|$ for all $n \in \mathbb{N}$.* □

The convergence is thus extremely fast, but the main obstacle is to find sufficiently good approximations $u_0 \approx z_0$ as starting values. Our global root-finding algorithm approximates all roots simultaneously, and the following simple criterion exploits this information:

Proposition 6.7. *Let $F \in \mathbb{C}[Z]$ be a separable polynomial of degree n. Suppose we have separated the roots in disjoint disks $B(u_k, \delta_k)$ for $k = 1, \ldots, n$ such that*

$$3n\delta_k \leq |u_k - u_j| \qquad \text{for all } j \neq k.$$

Then Newton's algorithm converges for each starting value u_k to the corresponding root $z_k \in B(u_k, \delta_k)$. More precisely, convergence is at least as fast as

$$|\Phi^n(u_k) - z_k| \leq 2^{-n} |u_k - z_k| \qquad \text{for all } n \in \mathbb{N}.$$

Remark 6.8. The hypothesis can be verified directly from the approximations $(u_k, \delta_k)_{k=1,\ldots,n}$ produced by the global root-finding algorithm of §6.2. Newton's method eventually converges much faster, and Proposition 6.7 only shows that right from the start Newton's method is at least as fast as bisection.

Proof. For $F = (Z - z_1) \cdots (Z - z_n)$ we have $F'/F = \sum_{j=1}^n (Z - z_j)^{-1}$. This entails $\Phi(z) = z - 1/\sum_{j=1}^n (z - z_j)^{-1}$, provided that $F(z) \neq 0$ and $F'(z) \neq 0$, whence

$$\frac{\Phi(z) - z_k}{z - z_k} = 1 - \frac{1}{\sum_{j=1}^n \frac{z - z_k}{z - z_j}} = \frac{\sum_{j \neq k} \frac{z - z_k}{z - z_j}}{1 + \sum_{j \neq k} \frac{z - z_k}{z - z_j}}.$$

By hypothesis we have approximate roots $u_1, \ldots, u_n$ such that $|u_k - z_k| \leq \delta_k$. Consider $z \in B(z_k, \delta_k)$, which entails $|z - u_k| \leq 2\delta_k$. The inequality $3n\delta_k \leq |u_k - u_j|$ for all $j \neq k$ implies $(3n-3)\delta_k + 2\delta_k + \delta_j \leq |u_k - u_j|$ and thus

$$|z - z_j| \geq |u_k - u_j| - 2\delta_k - \delta_j \geq (3n-3)\delta_k \qquad \text{for all } j \neq k.$$

This ensures that $\left|\frac{z - z_k}{z - z_j}\right| \leq \frac{\delta_k}{(3n-3)\delta_k} = \frac{1}{3(n-1)}$, whence $\left|\sum_{j \neq k} \frac{z - z_k}{z - z_j}\right| \leq \sum_{j \neq k} \left|\frac{z - z_k}{z - z_j}\right| \leq \frac{1}{3}$ and

$$\left|\frac{\Phi(z) - z_k}{z - z_k}\right| \leq \frac{\left|\sum_{j \neq k} \frac{z - z_k}{z - z_j}\right|}{1 - \left|\sum_{j \neq k} \frac{z - z_k}{z - z_j}\right|} \leq \frac{\frac{1}{3}}{1 - \frac{1}{3}} = \frac{1}{2}.$$

This shows that $|\Phi^n(z) - z_k| \leq 2^{-n} |z - z_k|$ for all $z \in B(z_k, \delta_k)$ and all $n \in \mathbb{N}$. In particular this holds for the starting value $z = u_k$ in $B(z_k, \delta_k)$. □

As an alternative to our tailor-made Proposition 6.7, the following theorem of Smale [6, chap. 8] provides a general convergence criterion in terms of local data. It applies in particular to polynomials, where it is most easily implemented.

Theorem 6.9 (Smale 1986)**.** *Let $f \colon \mathbb{C} \supset U \to \mathbb{C}$ be an analytic function. Consider $u_0 \in U$ such that $f'(u_0) \neq 0$, and let $\eta = |f(u_0)/f'(u_0)|$ be the initial displacement in Newton's iteration. Suppose that $f(z) = \sum_{k=0}^\infty c_k (z - u_0)^k$ for all $z \in B(u_0, 2\eta)$. If*

$$|c_k| \leq (8\eta)^{1-k} |c_1| \qquad \text{for all } k \geq 2,$$

then f has a unique zero z_0 in $B(u_0, 2\eta)$, and Newton's iteration converges as

$$|\Phi^n(u_0) - z_0| \leq 2^{1-2^n} \cdot |u_0 - z_0| \qquad \text{for all } n \in \mathbb{N}.$$

6.4. **Cauchy index computation.** In this section we briefly consider bit-complexity. To simplify we shall work over the rational numbers $\mathbb{Q}$. For $R,S \in \mathbb{Q}[X]$, with $\gcd(R,S)=1$ say, we wish to calculate a Sturm chain $S_0 = S, S_1 = R, \ldots, S_n = 1, S_{n+1} = 0$ such that

$$a_k S_{k-1} + b_k S_{k+1} = Q_k S_k \quad \text{with} \quad Q_k \in \mathbb{Q}[X] \text{ and } a_k, b_k \in \mathbb{Q}_+. \tag{6.1}$$

Applying the usual euclidean algorithm to polynomials of degree $\leq n$, this takes $O(n^3)$ arithmetic operations in $\mathbb{Q}$. This over-simplification, however, neglects the notorious problem of coefficient swell, which plagues naïve implementations with exponential running time. This difficulty can be overcome replacing the euclidean remainder sequence by subresultants, which were introduced by Sylvester [56]. Habicht [21] systematically studied subresultants and used them to construct Sturm chains whose coefficients are polynomial functions in the input coefficients, and not rational functions as given by euclidean division. Subresultants have become a highly developed tool of computer algebra; we refer to Gathen–Gerhard [18, chapters 6 and 11] and the references cited therein. This should be taken into account when choosing or developing a library for polynomial arithmetic.

Annotation 6.1. (Data management) The *construction* of the Sturm chain is the most expensive step in the above root-finding algorithm. In the real case we have to construct this chain only once because we can reuse it in all subsequent iterations. In the complex case, each segment requires a separate computation: it is thus advantageous to store each segment with its corresponding Sturm chain, and each square with the four Sturm chains along the boundary, so as to reuse precious data as much as possible.

Theorem 6.10. *Let $F = c_n Z^n + c_{n-1} Z^{n-1} + \cdots + c_1 Z + c_0$ be a polynomial of degree n with Gaussian integer coefficients such that $|\mathrm{re}\, c_k| \leq 2^a$ and $|\mathrm{im}\, c_k| \leq 2^a$ for all $k = 0, \ldots, n$. Suppose that all roots of F lie in the disk $B(r)$. The above root-finding algorithm determines all roots of F to a precision $3r/2^b$ requiring $\tilde{O}(n^3 b(a+nb))$ bit-operations.*

Here the asymtotic complexity $\tilde{O}$ neglects logarithmic factors.

Proof. Suppose that $R,S \in \mathbb{Z}[X]$ are of degree $\leq n$ and all coefficients are bounded by $A = 2^a$. According to Lickteig–Roy [30] and Gathen–Gerhard [18, Cor. 11.17] the subresultant algorithm requires $\tilde{O}(n^2 a)$ bit-operations. This has to be iterated b times; coefficients are bounded by $A = 2^{a+nb}$. Since we assume all roots to be distinct, they ultimately become separated so that the algorithm has to follow n approximations in parallel. This multiplies the previous bound by a factor nb, so we arrive at $\tilde{O}(n^3 b(a+nb))$ bit-operations. □

Annotation 6.2. (Simplicity) The algebraic algorithm is straightforward to implement except for two standard subalgorithms, namely fast integer arithmetic and fast subresultant computation for integer polynomials. These subalgorithms are theoretically well-understood, and their complexity bounds are known and nearly optimal. Their implementation is laborious, but is available in general-purpose libraries for integer and polynomial arithmetic. The algebraic algorithm uses exact arithmetic and no approximations. This ensures that we do not have to worry about error propagation, which simplifies (formal) correctness proofs.

Annotation 6.3. (Parallelization) We can adapt the algorithm to find only *one* root of F, and according to the preceding proof its complexity is $\tilde{O}(n^2 b(a+nb))$, again neglecting terms of order $\log(n)$. This approach is parallelizable: whenever bisection separates the roots into non-empty clusters, these can then be processed by independent computers working in parallel. The parallel complexity thus drops to $\tilde{O}(n^2 b(a+nb))$.

6.5. **What remains to be improved?** Root-finding algorithms of bit-complexity $\tilde{O}(n^2(n+b))$ are the world record since the ground-breaking work of Schönhage [47] in the 1980s. Cauchy's algebraic method is of complexity $\tilde{O}(n^4 b^2)$ and thus comes close, but in its current form it remains one order of magnitude more expensive. Schönhage remarks:

> It is not clear whether methods based on Sturm sequences can possibly become superior. Lehmer [29] and Wilf [66] both do not solve the extra problems which arise, if there is a zero on the test contour (circle or rectangle) or very close to it. [47, p. 5]

Notice that we have applied the *divide-and-conquer* paradigm in the arithmetic subalgorithms, but not in the root-finding method itself. In Schönhage's method this is achieved

by approximately factoring F of degree n into two polynomials F_1, F_2 of degrees close to $\frac{n}{2}$. It is plausible but not obvious that a similar strategy can be put into practice in the algebraic setting. Some clever idea and a more detailed investigation are needed here.

Our development neatly solves the problem of roots on the boundary. Of course, approximating the roots of a polynomial $F \in \mathbb{C}[Z]$ can only be as good as the initial data, and we therefore assume that F is known exactly. This is important because root-finding is an ill-conditioned problem, see Wilkinson [67]. Even if exact arithmetic can avoid this problem during the computation, it comes back into focus when the initial data is itself only an approximation. In this more general situation the real-algebraic approach requires a detailed error analysis, ideally in the setting of interval arithmetic and recursive analysis.

6.6. **Formal proofs.** In recent years the theory and practice of *formal proofs* and *computer-verified theorems* has become a fully fledged enterprise. A prominent and much discussed example is the Four Colour Theorem, see Gonthier [20]. The computer-verified proof community envisages much more ambitious projects, such as the classification of finite simple groups. See the *Mathematical Components Manifesto* by Gonthier, Werner, and Bertot at `www.msr-inria.inria.fr/projects/math/manifesto.html`.

Such gigantic projects make the Fundamental Theorem of Algebra look like a toy example, but its formalization is by no means a trivial task. A constructive proof, along the lines of Hellmuth Kneser (1940) and Martin Kneser (1981), has been formalized by the FTA project at Nijmegen (`www.cs.ru.nl/~freek/fta`) using the COQ proof assistant (`pauillac.inria.fr/coq`). Work is in progress so as to extract the algorithm implicit in the proof (`c-corn.cs.ru.nl`).

The real-algebraic approach offers certain advantages, mainly its conceptual simplicity and its algorithmic character. The latter is an additional important aspect: the theorem is not only an existence statement but immediately translates to an algorithm. A formal proof of the *theorem* will also serve as a formal proof of the *implementation*. As a first step, Mahboubi [32] discusses a formal proof of the subresultant algorithm.

Annotation 6.4. (Ongoing debate) Computer-assisted proofs have been intensely debated, and their scope and mathematical reliability have been questioned. The approach is still in its infancy compared to traditional viewpoints, and its long-ranging impact on mathematics remains to be seen.

We should like to emphasize that the formalization of mathematical theorems and proofs and their computer verification may be motivated by several factors. Some theorems, of varying difficulty, have been formalized in order to show that this is possible in principle and to gain practical experience. While pedagogically important for proof formalization itself, the traditional mathematician will find no added value in such examples.

More complicated theorems, such as the examples above, warrant an intrinsic motivation for formalization and computer-verified proofs, because there is an enormous number of cases to be solved and verified. Whenever human fallibility becomes a serious practical problem, as in these cases, a trustworthy verification tool clearly has its merit. This is particularly true if the mathematical model is implemented on a computer, and a high level of security is required. It is in this realm that computer-assisted correctness proofs are most widely appreciated.

7. HISTORICAL REMARKS

The Fundamental Theorem of Algebra is a crowning achievement in the history of mathematics. In order to place our real-algebraic approach into perspective, this section sketches its historical context. For the history of the Fundamental Theorem of Algebra I refer to Remmert [41], Dieudonné [13, chap. II, §III], and van der Waerden [61, chap. 5]. The history of Sturm's theorem has been examined in great depth by Sinaceur [49].

7.1. **Solving polynomial equations.** The method to solve quadratic equations was known to the Babylonians. Not much progress was made until the 16th century, when del Ferro (around 1520) and Tartaglia (1535) discovered a solution for cubic equations by radicals. Cardano's student Ferrari extended this to a solution of quartic equations by radicals. Both formulae were published in Cardano's *Ars Magna* in 1545. Despite considerable efforts during the following centuries, no such formulae could be found for degree 5 and higher.

They were finally shown not to exist by Ruffini (1805), Abel (1825), and Galois (1831). This solved one of the outstanding problems of algebra, alas in the negative.

The lack of general formulae provoked the question whether solutions exist at all. The existence of n roots for each real polynomial of degree n was mentioned by Roth (1608) and explicitly conjectured by Girard (1629) and Descartes (1637). They postulated these roots in some extension of $\mathbb{R}$ but did not claim that all roots are contained in the field $\mathbb{C} = \mathbb{R}[i]$. Leibniz (1702) even speculated that this is in general not possible.

The first proofs of the Fundamental Theorem of Algebra were published by d'Alembert (1746), Euler (1749), Lagrange (1772), and Laplace (1795). In his doctoral thesis (1799) Gauss criticized the shortcomings of all previous tentatives and presented his first proof, which ranks among the monumental achievements of mathematics.

7.2. **Gauss' first proof.** Gauss considers $F = Z^n + c_{n-1}Z^{n-1} + \cdots + c_1 Z + c_0$; upon substitution of $Z = X + iY$ he obtains $F = R + iS$ with $R, S \in \mathbf{R}[X,Y]$. The roots of F are precisely the intersections of the two curves $R = 0$ and $S = 0$ in the plane. Near a circle $\partial\Gamma$ with sufficiently large radius around 0, these curves resemble those of Z^n. The latter are $2n$ straight lines passing through the origin. The circle $\partial\Gamma$ thus intersects each of the curves $R = 0$ and $S = 0$ in $2n$ points placed in an alternating fashion around the circle.

Prolongating these curves into the interior of Γ, Gauss concludes that the curves $R = 0$ and $S = 0$ must intersect somewhere inside the circle. This conclusion relies on certain (intuitively plausible) assumptions, which Gauss clearly states but does not prove.

> Satis bene certe demonstratum esse videtur, curvam algebraicam neque alicubi subito abrumpi posse (uti e.g. evenit in curva transscendente, cuius aequatio $y = 1/\log x$), neque post spiras infinitas in aliquo puncto se quasi perdere (ut spiralis logarithmica), quantumque scio nemo dubium contra hanc rem movit. Attamen si quis postulat, demonstrationem nullis dubiis obnoxiam alia occasione tradere suscipiam.[1] [19, Bd. 3, p. 27]

To modern standards Gauss' first proof is thus incomplete. The unproven assertions are indeed correct, and have later been rigorously worked out by Ostrowski [36, 37].

Notice that Gauss' argument shows $w(F|\partial\Gamma) = n$ by an implicit homotopy $F \sim Z^n$, and our development of the algebraic winding number exhibits a short and rigorous path to the desired conclusion. Our proof can thus be considered as an algebraic version of Gauss' first proof, suitably completed by the techniques of Sturm and Cauchy, and justified by the intermediate value theorem.

7.3. **Gauss' further proofs.** Gauss gave two further proofs in 1816, and a fourth proof in 1849 which is essentially an improved version of his first proof [61, chap. 5]. The second proof is algebraic (§7.8.2), the third proof uses integration (§7.8.3) and foreshadows Cauchy's integral formula for the winding number.

When Gauss published his fourth proof in 1849 for his doctorate jubilee, the works of Sturm (1835) and Cauchy (1837) had been known for several years, and in particular Sturm's theorem had immediately risen to international acclaim. In principle Gauss could have taken up his first proof and completed it by arguments similar to the ones presented here. This has not happened, however, so we can speculate that Gauss was perhaps unaware of the work of Sturm, Cauchy, and Sturm–Liouville on complex roots of polynomials. Completing Gauss' geometric argument, Ostrowski [37] mentions the relationship with the Cauchy index but builds his proof on topological arguments.

[1] It seems to have been proved with sufficient certainty that an algebraic curve can neither suddenly break off anywhere (as happens e.g. with the transcendental curve whose equation is $y = 1/\log x$) nor lose itself, so to say, in some point after infinitely many coils (like the logarithmic spiral). As far as I know, nobody has raised any doubts about this. Should someone demand it, however, then I will undertake to give a proof that is not subject to any doubt, on some other occasion. (Adapted from Prof. Ernest Fandreyer's translation, Fitchburg State College Library, Manuscript Collections, `www.fsc.edu/library/archives/manuscripts/gauss.cfm`)

7.4. **Sturm, Cauchy, Liouville.** In 1820 Cauchy proved the Fundamental Theorem of Algebra, using the existence of a global minimum z_0 of $|F|$ and a local argument showing that $F(z_0) = 0$, see §7.8.1. While the local analysis is rigorous, global existence requires some compactness argument, which was yet to be developed, see Remmert [41, §1.8].

Sturm's theorem for counting real roots was announced in 1829 [51] and published in 1835 [52]. It was immediately assimilated by Cauchy in his residue calculus [8], based on complex integration, which was published in 1831 during his exile in Turin. In 1837 he published a more detailed exposition [9] with analytic-geometric proofs, and explicitly recognizes the relation to Sturm's theorem [9, pp. 426–427,431].

In the intervening years, Sturm and Liouville [55, 53] had elaborated their own proofs of Cauchy's theorem, which they published in 1836. (Loria [31] and Sinaceur [49, I.VI] examine the interaction between Sturm, Liouville, and Cauchy in detail.) As opposed to Cauchy, their arguments are based on what they call the "first principles of algebra". In the terminology of their time this means the theory of complex numbers, including trigonometric coordinates $z = r(\cos\theta + i\sin\theta)$ and de Moivre's formula, but excluding integration. Furthermore they use sign variations and, of course, the intermediate value theorem of real functions, as well as tacit compactness arguments.

7.5. **Sturm's algebraic vision.** Sturm, in his article [53] continuing his work with Liouville [55], presents arguments which closely parallel our real-algebraic proof: the argument principle (Prop. 1, p. 294), multiplicativity (Prop. 2, p. 295), counting roots of a split polynomial within a given region (Prop. 3, p. 297), the winding number in the absence of zeros (Prop. 4, p. 297), and finally Cauchy's theorem (p. 299). One crucial step is to show that $w(F|\partial\Gamma) = 0$ when F does not vanish in Γ. This is solved by subdivision and a tacit compactness argument (pp. 298–299); our compactness proof of Theorem 5.3 completes his argument. Sturm then deduces the Fundamental Theorem of Algebra (pp. 300–302) and expounds on the practical computation of the Cauchy index $w(F|\partial\Gamma)$ using Sturm chains as in the real case (pp. 303–308).

Sturm's exposition strives for algebraic simplicity, but his arguments are ultimately based on geometric and analytic techniques. It is only on the final pages that Sturm employs his algebraic method for computing the Cauchy index. This mixed state of affairs has been passed on ever since, even though it is far less satisfactory than Sturm's purely algebraic treatment of the real case. Our proof shows that Sturm's algebraic vision of the complex case can be salvaged and his arguments can be put on firm real-algebraic ground.

We note that Sturm and Liouville explicitly exclude zeros on the boundary:

> Toutefois nous excluons formellement le cas particulier où, pour quelque point de la courbe *ABC*, on aurait à la fois $P = 0$, $Q = 0$: ce cas particulier ne jouit d'aucune propriété régulière et ne peut donner lieu à aucun théorème.[2] [55, p. 288]

This seems overly pessimistic in view of our Theorem 1.8 above. In his continuation [53], Sturm formulates the same problem much more cautiously:

> C'est en admettant cette hypothèse que nous avons démontré le théorème de M. Cauchy ; les modifications qu'il faudrait y apporter dans le cas où il aurait des racines sur le contour même *ABC*, exigeraient une discussion longue et minutieuse que nous avons voulu éviter en faisant abstraction de ce cas particulier.[3] [53, p. 306]

[2] We formally exclude, however, the case where for some point of the curve *ABC* we have simultaneously $P = 0$ and $Q = 0$: this special case does not enjoy any regular property and cannot give rise to any theorem.

[3] It is under this hypothesis that we have proven the theorem of Mr. Cauchy; the necessary modifications in the case where roots were on the contour *ABC* would require a long and meticulous discussion, which we have wanted to avoid by neglecting this special case.

It seems safe to say that our detailed discussion is just as "long and meticulous" as the usual development of Sturm's theorem. Modulo these details, the cited works of Gauss, Cauchy, and Sturm contain the essential ideas for the real-algebraic approach. It remained to work them out. To this end our presentation refines the techniques in several ways:

- We purge all arguments of transcendental functions and compactness assumptions. This simplifies the proof and generalizes it to real closed fields.
- The product formula (§4.5) and homotopy invariance (§5.3) streamline the proof and avoid tedious calculations.
- The uniform treatment of boundary points extends Sturm's theorem to piecewise polynomial functions and leads to straightforward algorithms.

7.6. **Further development in the 19th century.** Sturm's theorem was a decisive step in the development of algebra as an autonomous field, independent of analysis, in particular in the hands of Sylvester and Hermite. For a detailed discussion see Sinaceur [49].

In 1869 Kronecker [27] constructed his higher-dimensional index (also called Kronecker characteristic) using integration. His initial motivation was to generalize Sturm's theorem to higher dimensions, extending previous work of Sylvester and Hermite, but he then turned to analytic methods. Subsequent work was likewise built on analytic methods over $\mathbb{R}$: one gains in generality by extending the index to smooth or even continuous functions, but one loses algebraic generality, simplicity, and computability.

The problem of *stability of motion* led Routh [42] in 1878 and Hurwitz [23] in 1895 to count the number of complex roots having negative real part. With the celebrated Routh–Hurwitz theorem, the algebraic index has transited from algebra to application, where it survives to the present day. In the 1898 *Encyklopädie der mathematischen Wissenschaften* [34, Band I], Netto's survey on the Fundamental Theorem of Algebra (§I-B1a7) mentions Cauchy's algebraic approach only briefly (p. 236), while Runge's article on approximation of complex roots (§I-B3a6) discusses Cauchy's method in greater detail (pp. 418-422). In the 1907 *Encyclopédie des Sciences Mathématiques* [35], Netto and le Vavasseur give an overview of nearly 100 published proofs (tome I, vol. 2, §80–88), including Cauchy's argument principle (§87). The work of Sturm–Liouville [55, 53] is cited but the algebraic approach via Sturm chains is not mentioned.

7.7. **19th century textbooks.** While Sturm's theorem made its way from 19th century algebra to modern algebra textbooks and is still taught today, it seems that the algebraic approach to the complex case has been lost on the way. Let me illustrate this by two prominent and perhaps representative textbooks.

In his 1877 textbook *Cours d'algèbre supérieure*, Serret [48, pp. 118–132] presents the proof of the Fundamental Theorem of Algebra following Cauchy and Sturm–Liouville, with only minor modifications. Two decades later, Weber devoted over 100 pages to real-algebraic equations in his 1898 textbook *Lehrbuch der Algebra* [62], where he presents Sturm's theorem in great detail (§91-106). Calling upon Kronecker's geometric index theory (§100-102), he sketches how to count complex roots (§103-104). Quite surprisingly, he uses only $\operatorname{ind}\left(\frac{P'}{P}\right)$ and Corollary 3.23 where the general case $\operatorname{ind}\left(\frac{R}{S}\right)$ and Theorem 3.20 would have been optimal. Here Cauchy's algebraic method [9], apparently unknown to Weber, had gone much further concerning explicit formulae and concrete computations.

7.8. **Survey of proof strategies.** Since the time of Gauss numerous proofs of the Fundamental Theorem of Algebra have been developed. We refer to Remmert [41] for a concise overview and to Fine–Rosenberger [16] for a text-book presentation. As mentioned in §1.2, the proof strategies can be grouped into three families:

7.8.1. *Analysis.* Proofs in this family are based on the existence of a global minimum z_0 of $|F|$ and some local argument from complex analysis showing that $F(z_0) = 0$ (d'Alembert

1746, Argand 1814, Cauchy 1820). See Remmert [41, §2] for a presentation in its historical context, or Rudin [45, chap. 8] in the context of a modern analysis course. In its most succinct form, this is formulated by Liouville's theorem for entire functions. Such arguments are in general not constructive; for constructive refinements see [41, §2.5].

7.8.2. *Algebra*. Proofs in this family use the fundamental theorem of symmetric polynomials in order to reduce the problem from real polynomials of degree $2^k m$ with m odd to degree $2^{k-1}m'$ with m' odd (Euler 1749, Lagrange 1772, Laplace 1795, Gauss 1816, see [41, appendix]). The argument can be reformulated using Galois theory, see Cohn [11, Thm. 8.8.7], Jacobson [25, Thm. 5.2], or Lang [28, §VI.2, Ex. 5]. The induction is based, for $k = 0$, on real polynomials of odd degree, where the existence of at least one real root is guaranteed by the intermediate value theorem. This algebraic proof thus works over every real closed field. It is constructive but ill-suited to actual computations.

7.8.3. *Topology*. Proofs in this family use some form of the winding number $w(\gamma)$ of closed paths $\gamma\colon [0,1] \to \mathbb{C}^*$ (Gauss 1799/1816, Cauchy 1831/37, Sturm–Liouville 1836). The winding number appears in various guises, see Remark 1.5: in each case the difficulty is a rigorous construction and to establish its characteristic properties: normalization, multiplicativity and homotopy invariance, as stated in Theorem 1.2.

Our proof belongs to this last family. Unlike previous proofs, however, we do not base the winding number on analytical or topological arguments but on real algebra.

7.9. **Constructive and algorithmic aspects.** Sturm's method is eminently practical, by the standards of 19th century mathematics as for modern-day implementations. As early as 1840 Sylvester [56] wrote "Through the well-known ingenuity and proferred help of a distinguished friend, I trust to be able to get a machine made for working Sturm's theorem (...)". It seems, however, that such a machine was never built. Calculating machines had been devised by Pascal, Leibniz, and Babbage; the latter was Lucasian Professor of Mathematics at Cambridge when Sylvester studied there in the 1830s.

The idea of computing machinery seems to have been common among mid-19th century mathematicians. In a small note of 1846, Ullherr [60] remarks that the argument principle leads to a complex root-finding algorithm: "Die bei dem ersten Beweise gebrauchte Betrachtungsart giebt ein Mittel an die Hand, die Wurzeln der höheren Gleichungen mittels eines Apparates mechanisch zu finden." [4] No details are given.

For separating and approximating roots, the state of the art at the end of the 19th century has been surveyed in Runge's *Encyklopädie* article [34, Band I, §I-B3a].

In 1924 Weyl [64] reemphasized that the analytic winding number can be used to find and approximate the roots of F. In this vein Weyl formulated his constructive proof of the Fundamental Theorem of Algebra, which indeed translates to an algorithm: a careful numerical approximation can be used to calculate the integer $w(F|\partial\Gamma)$, see Henrici [22, §6.11]. While Weyl's motivation may have been philosophical, it is the practical aspect that has proven most successful. Variants of Weyl's algorithm are used in modern computer implementations for finding approximate roots, and are among the asymptotically fastest known algorithms. The question of algorithmic complexity was pursued by Schönhage [47] and others since the 1980s. See Pan [39] for an overview.

The fact that Sturm's and Cauchy's theorems together can be applied to count complex roots seems not to be as widely known as it should be. In the 1969 Proceedings [12] on constructive aspects of the Fundamental Theorem of Algebra, Cauchy's algebraic method is not mentioned. Lehmer [29] uses a weaker form, the Routh–Hurwitz theorem, although Cauchy's general result would have been better suited. Cauchy's method reappears in 1978 in a small note by Wilf [66], and is briefly mentioned in Schönhage's technical report [46,

[4] The viewpoint used in the first proof provides a method to find the roots of higher-degree equations by means of a mechanical apparatus.

p. 5]. Most often the computer algebra literature credits Weyl for the analytic-numeric algorithm, and Lehmer or Wilf for the algebraic-numeric method, but not Cauchy or Sturm. Even if Cauchy's index and Sturm's algorithm are widely used, their algebraic contributions to complex root location seem to be largely ignored.

ACKNOWLEDGMENTS

Many colleagues had the kindness to comment on successive versions of this article and to share their expertise on diverse aspects of this fascinating topic. It is my heartfelt pleasure to thank Roland Bacher, Theo de Jong, Christoph Lamm, Bernard Parisse, Cody Roux, Marie-Françoise Roy, Francis Sergeraert, and Duco van Straten. The thoughtful suggestions of the referees greatly helped to improve the exposition.

REFERENCES

1. E. Artin, *Über die Zerlegung definiter Funktionen in Quadrate*, Abh. Math. Sem. Univ. Hamburg **5** (1926), 100–115, Collected Papers [2], pp. 273–288.
2. ———, *Collected Papers*, Edited by S. Lang and J. T. Tate, Springer-Verlag, New York, 1982, Reprint of the 1965 original.
3. E. Artin and O. Schreier, *Algebraische Konstruktion reeller Körper*, Abh. Math. Sem. Univ. Hamburg **5** (1926), 85–99, Collected Papers [2], pp. 258–272.
4. ———, *Eine Kennzeichnung der reell abgeschlossenen Körper*, Abh. Math. Sem. Univ. Hamburg **5** (1927), 225–231, Collected Papers [2], pp. 289–295.
5. S. Basu, R. Pollack, and M.-F. Roy, *Algorithms in real algebraic geometry*, second ed., Springer-Verlag, Berlin, 2006, Available at `perso.univ-rennes1.fr/marie-francoise.roy`.
6. L. Blum, F. Cucker, M. Shub, and S. Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998.
7. J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Springer-Verlag, Berlin, 1998.
8. A. L. Cauchy, *Sur les rapports qui existent entre le calcul des résidus et le calcul des limites*, Bulletin des Sciences de Férussac **16** (1831), 116–128, Œuvres [10], Série 2, tome 2, pp. 169–183.
9. ———, *Calcul des indices des fonctions*, Journal de l'École Polytechnique **15** (1837), 176–229, Œuvres [10], Série 2, tome 1, pp. 416–466.
10. ———, *Œuvres complètes*, Gauthier-Villars, Paris, 1882–1974, Available at `mathdoc.emath.fr/OEUVRES/`.
11. P. M. Cohn, *Basic algebra*, Springer-Verlag London Ltd., London, 2003.
12. B. Dejon and P. Henrici (eds.), *Constructive aspects of the fundamental theorem of algebra*, John Wiley & Sons Inc., London, 1969.
13. J. Dieudonné, *Abrégé d'histoire des mathématiques. 1700–1900.*, Hermann, Paris, 1978.
14. H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel, and R. Remmert, *Numbers*, Graduate Texts in Mathematics, vol. 123, Springer-Verlag, New York, 1991.
15. M. Eisermann, *Kronecker's index and Brouwer's fixed point theorem over real closed fields*, In preparation.
16. B. Fine and G. Rosenberger, *The fundamental theorem of algebra*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1997.
17. A. T. Fuller (ed.), *Stability of motion*, Taylor & Francis, Ltd., London, 1975, A collection of early scientific publications by E. J. Routh, W. K. Clifford, C. Sturm and M. Bôcher.
18. J. von zur Gathen and J. Gerhard, *Modern computer algebra*, second ed., Cambridge University Press, Cambridge, 2003.
19. C. F. Gauß, *Werke. Band I–XII*, Georg Olms Verlag, Hildesheim, 1973, Reprint of the 1863–1929 original, available at `resolver.sub.uni-goettingen.de/purl?PPN235957348`.
20. G. Gonthier, *A computer-checked proof of the four colour theorem*, Tech. report, Microsoft Research, Cambridge, 2004, 57 pages, available at `research.microsoft.com/~gonthier/4colproof.pdf`.
21. W. Habicht, *Eine Verallgemeinerung des Sturmschen Wurzelzählverfahrens*, Comment. Math. Helv. **21** (1948), 99–116.
22. P. Henrici, *Applied and computational complex analysis*, John Wiley & Sons Inc., New York, 1974.
23. A. Hurwitz, *Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt*, Math. Ann. **46** (1895), no. 2, 273–284, Math. Werke [24], Band 2, pp. 533–545. Reprinted in [17].
24. ———, *Mathematische Werke*, Birkhäuser Verlag, Basel, 1962–1963.
25. N. Jacobson, *Basic algebra I-II*, second ed., W. H. Freeman and Company, New York, 1985, 1989.
26. L. Kronecker, *Werke*, Chelsea Publishing Co., New York, 1968, Reprint of the 1895–1930 original.
27. ———, *Ueber Systeme von Functionen mehrer Variabeln*, Monatsberichte Akademie Berlin (1969), 159–193, 688–698, Werke [26], Band I, pp. 175–226.
28. S. Lang, *Algebra*, third ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002.

29. D. H. Lehmer, *Search procedures for polynomial equation solving*, [12], John Wiley & Sons Inc., 1969, pp. 193–208.
30. Th. Lickteig and M.-F. Roy, *Sylvester-Habicht sequences and fast Cauchy index computation*, J. Symbolic Comput. **31** (2001), no. 3, 315–341.
31. G. Loria, *Charles Sturm et son œuvre mathématique*, Enseign. Math. **37** (1938), 249–274.
32. A. Mahboubi, *Proving formally the implementation of an efficient gcd algorithm for polynomials*, Tech. report, INRIA, Nice, France, 2006, 15 pages, available at `hal.inria.fr/inria-00001270/en/`.
33. M. Marden, *Geometry of polynomials*, Second edition. Mathematical Surveys, No. 3, Amer. Math. Soc., Providence, R.I., 1966.
34. W. F. Meyer (ed.), *Encyklopädie der mathematischen Wissenschaften*, B. G. Teubner, Leipzig, 1898.
35. J. Molk (ed.), *Encyclopédie des Sciences Mathématiques*, Gauthier-Villars, Paris, 1907.
36. A. Ostrowski, *Über den ersten und vierten Gausschen Beweis des Fundamentalsatzes der Algebra*, vol. X.2, ch. 3 in [19], 1920, Collected Papers [38], vol. 1, pp. 538–553.
37. ———, *Über Nullstellen stetiger Funktionen zweier Variablen*, J. Reine Angew. Math. **170** (1933), 83–94, Collected Papers [38], vol. 3, pp. 269–280.
38. ———, *Collected Mathematical Papers*, Birkhäuser Verlag, Basel, 1983.
39. V. Y. Pan, *Solving a polynomial equation: some history and recent progress*, SIAM Rev. **39** (1997), no. 2, 187–220.
40. Q. I. Rahman and G. Schmeisser, *Analytic theory of polynomials*, London Mathematical Society Monographs. New Series, vol. 26, Oxford University Press, Oxford, 2002.
41. R. Remmert, *The fundamental theorem of algebra*, ch. 4 in [14], Springer-Verlag, New York, 1991.
42. E. J. Routh, *A treatise on the stability of a given state of motion*, Macmillan, London, 1878, Reprinted in [17], pp. 19–138.
43. M.-F. Roy, *Basic algorithms in real algebraic geometry and their complexity: from Sturm's theorem to the existential theory of reals*, Lectures in real geometry (Madrid, 1994), de Gruyter Exp. Math., vol. 23, de Gruyter, Berlin, 1996, pp. 1–67.
44. M.-F. Roy and A. Szpirglas, *Complexity of computation on real algebraic numbers*, J. Symbolic Comput. **10** (1990), no. 1, 39–51.
45. W. Rudin, *Principles of mathematical analysis*, third ed., McGraw-Hill Book Co., New York, 1976.
46. A. Schönhage, *The fundamental theorem of algebra in terms of computational complexity*, Tech. report, Math. Inst. Univ. Tübingen, Tübingen, Germany, 1982, 49 pages, available at `www.informatik.uni-bonn.de/~schoe/fdthmrep.ps.gz`.
47. ———, *Equation solving in terms of computational complexity*, Proc. Int. Congress of Math., Berkeley, 1986 (Providence, RI), Amer. Math. Soc., 1987, pp. 131–153.
48. J. A. Serret, *Cours d'algèbre supérieure*, Gauthier-Villars, Paris, 1877, Available at `gallica.bnf.fr/ark:/12148/bpt6k291135`.
49. H. Sinaceur, *Corps et Modèles*, Librairie Philosophique J. Vrin, Paris, 1991, Translated as [50].
50. ———, *Fields and Models*, Birkhäuser, Basel, 2008.
51. C.-F. Sturm, *Mémoire sur la résolution des équations numériques*, Bulletin des Sciences de Férussac **11** (1829), 419–422, Collected Works [54], pp. 323–326.
52. ———, *Mémoire sur la résolution des équations numériques*, Académie Royale des Sciences de l'Institut de France **6** (1835), 271–318, Collected Works [54], pp. 345–390.
53. ———, *Autres démonstrations du même théorème*, J. Math. Pures Appl. **1** (1836), 290–308, Collected Works [54], pp. 486–504, English translation in [17], pp. 189–207.
54. ———, *Collected Works*, Edited by J.-C. Pont, Birkhäuser, Basel, 2009, Some of the articles are also available at `www-mathdoc.ujf-grenoble.fr/pole-bnf/Sturm.html`.
55. C.-F. Sturm and J. Liouville, *Démonstration d'un théorème de M. Cauchy, relatif aux racines imaginaires des équations*, J. Math. Pures Appl. **1** (1836), 279–289, Collected Works [54], pp. 474–485.
56. J. J. Sylvester, *A method of determining by mere inspection the derivatives from two equations of any degree*, Philosophical Magazine **16** (1840), 132–135, Collected Papers [57], vol. I, pp. 54–57.
57. ———, *Collected Mathematical Papers*, Cambridge University Press, Cambridge, 1904–1912.
58. A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, Proc. Lond. Math. Soc., II. Ser. **42** (1936), 230–265, Collected Works [59], vol. IV, pp. 18–56.
59. ———, *Collected Works*, North-Holland Publishing Co., Amsterdam, 1992.
60. J. C. Ullherr, *Zwei Beweise für die Existenz der Wurzeln der höhern algebraischen Gleichungen*, J. Reine Angew. Math. **31** (1846), 231–234.
61. B. L. van der Waerden, *A history of algebra*, Springer-Verlag, Berlin, 1985.
62. H. Weber, *Lehrbuch der Algebra*, second ed., F. Vieweg & Sohn, Braunschweig, 1898, Reprint: Chelsea Pub Co, New York, 3rd edition, January 2000.
63. H. Weyl, *über die neue Grundlagenkrise der Mathematik. (vorträge, gehalten im mathematischen Kolloquium Zürich.)*, Math. Z. **10** (1921), 39–79, Ges. Abh. [65], Band II, pp. 143–180.
64. ———, *Randbemerkungen zu Hauptproblemen der Mathematik, II. Fundamentalsatz der Algebra und Grundlagen der Mathematik*, Math. Z. **20** (1924), no. 1, 131–150, Ges. Abh. [65], Band II, pp. 433–453.

65. ———, *Gesammelte Abhandlungen*, Springer-Verlag, Berlin, 1968.
66. H. S. Wilf, *A global bisection algorithm for computing the zeros of polynomials in the complex plane*, J. Assoc. Comput. Mach. **25** (1978), no. 3, 415–420.
67. J. H. Wilkinson, *The evaluation of the zeros of ill-conditioned polynomials*, Numer. Math. **1** (1959), 150–180.

APPENDIX A. APPLICATION TO THE ROUTH–HURWITZ STABILITY THEOREM

The algebraic winding number is a versatile tool beyond the Fundamental Theorem of Algebra. In certain applications it is important to determine whether a given polynomial $F \in \mathbb{C}[Z]$ has all of its roots in the left half plane $\mathbb{C}_{\mathrm{re}<0} = \{z \in \mathbb{C} \mid \mathrm{re}(z) < 0\}$. This question originated from the theory of dynamical systems and the problem of *stability of motion*:

Example A.1. Let $A \in \mathbb{R}^{n\times n}$ be a square matrix with real coefficients. The differential equation $y' = Ay$ with initial condition $y(0) = y_0$ has a unique solution $f\colon \mathbb{R} \to \mathbb{R}^n$ given by $f(t) = \exp(tA)y_0$. In terms of dynamical systems, the origin $a = 0$ is a fixed point; it is *stable* if all eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ of A satisfy $\mathrm{re}\,\lambda_k < 0$: in this case $\exp(tA)$ has eigenvalues $\exp(t\lambda_k)$ of absolute value < 1. The matrix $\exp(tA)$ is thus a contraction for all $t > 0$, and every initial value is attrated to $a = 0$, i.e., $f(t) \to 0$ for $t \to +\infty$.

Example A.2. The previous argument holds locally around fixed points of any dynamical system given by a differential equation $y' = \Phi(y)$ where $\Phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. Suppose that a is a fixed point, i.e., $\Phi(a) = 0$. It is *stable* if all eigenvalues of the matrix $A = \Phi'(a) \in \mathbb{R}^{n\times n}$ have negative real part: in this case there exists a neighbourhood V of a that is attracted to a: every trajectory $f\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, satisfying $f'(t) = \Phi(f(t))$ for all $t \geq 0$, starting at $f(0) \in V$ satisfies $f(t) \to a$ for $t \to +\infty$.

Given $F \in \mathbb{C}[Z]$ we can determine the number of roots with positive real part simply by considering the rectangle $\Gamma = [0, r] \times [-r, r]$ and calculating $w(F|\partial\Gamma)$ for r sufficiently large. (One could use the Cauchy radius ρ_F defined in §5.4.) Routh's theorem, however, offers a simpler solution by calculating the Cauchy index along the imaginary axis. This is usually proven using complex integration, but here we will give a real-algebraic proof. As usual we consider a real closed field $\mathbf{R}$ and its extension $\mathbf{C} = \mathbf{R}[i]$ with $i^2 = -1$.

Definition A.3. For every polynomial $F \in \mathbf{C}[Z]^*$ we define its *Routh index* as

$$\mathrm{Routh}(F) := \mathrm{ind}_{+r}^{-r}\left(\tfrac{\mathrm{re}\,F(iY)}{\mathrm{im}\,F(iY)}\right) + \mathrm{ind}_{-1/r}^{+1/r}\left(\tfrac{\mathrm{re}\,F(i/Y)}{\mathrm{im}\,F(i/Y)}\right) \tag{A.1}$$

for some arbitrary parameter $r \in \mathbf{R}_{>0}$; the result is independent of r by Proposition 3.10(b).

Remark A.4. We can decompose $F(iY) = R + iS$ with $R, S \in \mathbf{R}[Y]$ and compare the degrees $m = \deg S$ and $n = \deg R$. If $m \geq n$, then the fraction $\frac{R(1/Y)}{S(1/Y)} = \frac{Y^m R(1/Y)}{Y^m S(1/Y)}$ has no pole in 0, so the second index vanishes for r sufficiently large, and Equation (A.1) simplifies to

$$\mathrm{Routh}(F) = -\mathrm{ind}_{-\infty}^{+\infty}\left(\tfrac{\mathrm{re}\,F(iY)}{\mathrm{im}\,F(iY)}\right). \tag{A.2}$$

Example A.5. In general the second index in Equation (A.1) cannot be neglected, as illustrated by $F = (Z-1)(Z-2)$: here $F(iY) = -Y^2 - 3iY - 2$, whence

$$\tfrac{\mathrm{re}\,F(iY)}{\mathrm{im}\,F(iY)} = \tfrac{Y^2-2}{3Y} \quad \text{and} \quad \tfrac{\mathrm{re}\,F(i/Y)}{\mathrm{im}\,F(i/Y)} = \tfrac{1-2Y^2}{3Y}.$$

Both indices in Equation (A.1) contribute $+1$ such that $\mathrm{Routh}(F) = +2$.

Lemma A.6. *We have* $\mathrm{Routh}(Z - z_0) = \mathrm{sign}(\mathrm{re}\,z_0)$ *for all* $z_0 \in \mathbf{C}$.

Proof. For $F = Z - z_0$ we find $F(iY) = R + iS$ with $R = -\mathrm{re}\,z_0$ and $S = Y - \mathrm{im}\,z_0$. Thus $\mathrm{Routh}(F) = -\mathrm{ind}_{-\infty}^{+\infty}\left(\frac{R}{S}\right) = \mathrm{ind}_{-\infty}^{+\infty}\left(\frac{\mathrm{re}\,z_0}{Y - \mathrm{im}\,z_0}\right) = \mathrm{sign}(\mathrm{re}\,z_0)$. □

Lemma A.7. *We have* $\mathrm{Routh}(FG) = \mathrm{Routh}(F) + \mathrm{Routh}(G)$ *for all* $F, G \in \mathbf{C}[Z]^*$.

Proof. This follows from the real product formula stated in Theorem 4.6. □

Remark A.8. For every $c \in \mathbf{C}^*$ we have $\operatorname{Routh}(c) = 0$, whence $\operatorname{Routh}(cF) = \operatorname{Routh}(F)$. This can be used to ensure the favourable situation of Remark A.4, where $S = \operatorname{im} F(iY)$ has at least the same degree as $R = \operatorname{re} F(iY)$. If $\deg S < \deg R$, then it is advantageous to pass to iF, that is, to make the replacement $(R,S) \leftarrow (-S,R)$.

We can now deduce the following formulation of the famous Routh–Hurwitz theorem:

Theorem A.9. *The Routh index of every polynomial $F \in \mathbf{C}[Z]^*$ satisfies $\operatorname{Routh}(F) = p - q$ where p resp. q is the number of roots of F in $\mathbf{C}$ having positive resp. negative real part.*

Proof. The Fundamental Theorem of Algebra ensures that $F = c(Z - z_1) \cdots (Z - z_n)$ in $\mathbf{C}[X]$, so the Routh index formula follows from the preceding lemmas. □

Remark A.10. By a linear transformation $z \mapsto az + b$, with $a \in \mathbf{C}^*$ and $b \in \mathbf{C}$, we can map the imaginary line onto any other straight line, so we can apply the theorem to count roots in any half-space in $\mathbf{C}$. The transformation $z \mapsto \frac{z-1}{z+1}$ maps $\mathbf{R}i \cup \{\infty\}$ onto the unit circle, and the right half plane to the unit disk. Again by linear transformation we can thus apply the theorem to count roots in any given disk in $\mathbf{C}$.

Routh's criterion is often applied to real polynomials $P \in \mathbf{R}[X]$, as in the motivating examples above, which warrants the following more detailed formulation:

Corollary A.11. *Let $P = c_0 + c_1 X + \cdots + c_n X^n$ be a polynomial of degree n over $\mathbf{R}$, and let p resp. q be the number of roots of P in $\mathbf{C}$ having positive resp. negative real part. Then*

$$p - q = \operatorname{Routh}(P) = \begin{cases} -\operatorname{ind}_{-\infty}^{+\infty}\left(\frac{\operatorname{re} P(iY)}{\operatorname{im} P(iY)}\right) & \text{if } n \text{ is odd,} \\ +\operatorname{ind}_{-\infty}^{+\infty}\left(\frac{\operatorname{im} P(iY)}{\operatorname{re} P(iY)}\right) & \text{if } n \text{ is even.} \end{cases} \tag{A.3}$$

Both cases can be subsumed into the unique formula

$$q - p = \operatorname{ind}_{-\infty}^{+\infty}\left(\frac{c_{n-1}X^{n-1} - c_{n-3}X^{n-3} + \dots}{c_n X^n - c_{n-2}X^{n-2} + \dots}\right). \tag{A.4}$$

This implies Routh's criterion: All roots of P have negative real part if and only if $q = n$ and $p = 0$, which is equivalent to saying that the Cauchy index in (A.4) *evaluates to n.*

Routh's formulation via Cauchy indices is unrivaled in its simplicity, and can immediately be calculated using Sturm's theorem (§3.7). Hurwitz' formulation uses determinants, which has the advantage to produce explicit polynomial formulae in the given coefficients. See Henrici [22, §6.7], Marden [33, chap. IX], or Rahman–Schmeisser [40, chap. 11].

Appendix B. Brouwer's fixed point theorem

Brouwer's theorem states that every continuous map $f\colon [0,1]^n \to [0,1]^n$ of a cube in $\mathbb{R}^n$ to itself has a fixed point. While in dimension $n = 1$ this follows directly from the intermediate value theorem, the statement in dimension $n \geq 2$ is much more difficult to prove: one employs either sophisticated machinery (differential topology, Stokes' theorem, co/homology) or subtle combinatorial techniques (Sperner's lemma, Nash's game of Hex). All proofs use Brouwer's mapping degree, in a more or less explicit way, and the compactness of $[0,1]^n$ plays a crucial rôle. Such proofs are often non-constructive and do not address the question of locating fixed points.

Using the algebraic winding number we can prove Brouwer's theorem in a constructive way over real closed fields, restricting the statement from continuous to rational functions:

Theorem B.1. *Let $\mathbf{R}$ be a real closed field and let $P, Q \in \mathbf{R}(X,Y)$ be rational functions. Assume that P, Q have no poles in $\Gamma = [x_0, x_1] \times [y_0, y_1]$, so that they define a map $f\colon \Gamma \to \mathbf{R}^2$ by $f(x,y) = (P(x,y), Q(x,y))$. If $f(\Gamma) \subset \Gamma$, then there exists $z \in \Gamma$ such that $f(z) = z$.* □

Proof. The essential properties of the algebraic winding number stated in Theorem 1.2 extend to rational functions without poles. By translation and homothety we can assume that $\Gamma = [-1,+1] \times [-1,+1]$. We consider the homotopy $g_t = \mathrm{id} - t f$ from $g_0 = \mathrm{id}$ to $g_1 = \mathrm{id} - f$. For $z \in \partial\Gamma$ we have $g_t(z) = 0$ if and only if $t = 1$ and $f(z) = z$; in this case the assertion holds. Otherwise, we have $g_t(z) \neq 0$ for all $z \in \partial\Gamma$ and $t \in [0,1]$. We can then apply homotopy invariance to conclude that $w(g_1|\partial\Gamma) = w(g_0|\partial\Gamma) = 1$. Theorem 5.3 implies that there exists $z \in \operatorname{Int}\Gamma$ such that $g_1(z) = 0$, whence $f(z) = z$. □

Remark B.2. As for the Fundamental Theorem of Algebra, the algebraic proof of Theorem B.1 also provides an algorithm to approximate a fixed point to any desired precision. Here we have to assume the ordered field $\mathbf{R}$ to be archimedean, or equivalently $\mathbf{R} \subset \mathbb{R}$. Beginning with $\Gamma_0 = [-1,+1] \times [-1,+1]$ and bisecting successively, we can construct a sequence of subsquares $\Gamma = \Gamma_0 \supset \Gamma_1 \supset \cdots \supset \Gamma_k$ such that f has a fixed point on $\partial\Gamma_k$ or $w(\mathrm{id} - f|\partial\Gamma_k) \neq 0$. In the first case, a fixed point on the boundary $\partial\Gamma_k$ is signalled during the calculation of $w(\mathrm{id} - f|\partial\Gamma_k)$ and leads to a one-dimensional search problem. In the second case, we continue the two-dimensional approximation.

Remark B.3. Tarski's theorem says that all real closed fields share the same elementary theory. This implies that the statement of Brouwer's fixed point theorem generalizes from the real numbers $\mathbb{R}$ to every real closed field $\mathbf{R}$: as formulated above it is a first-order assertion in each degree. It is remarkable that there exists a first-order proof over $\mathbf{R}$ that is as direct as the usual second-order proof over $\mathbb{R}$. In this article we concentrate on dimension $n = 2$, but the algebraic approach generalizes to any finite dimension [15].

Remark B.4. Over the field $\mathbb{R}$ of real numbers the algebraic version implies the continuous version as follows. Since $\Gamma = [-1,+1] \times [-1,+1]$ is compact, every continuous function $f\colon \Gamma \to \Gamma$ can be approximated by polynomials $g_n\colon \Gamma \to \mathbb{R}^2$ such that $|g_n - f| \leq \frac{1}{n}$. The polynomials $f_n = \frac{n}{n+1} g_n$ satisfy $f_n(\Gamma) \subset \Gamma$ and $|f_n - f| \leq \frac{2}{n}$. For each n there exists $z_n \in \Gamma$ such that $f_n(z_n) = z_n$ according to Theorem B.1. Again by compactness of Γ we can extract a convergent subsequence. Assuming $z_n \to z$, we find

$$|f(z) - z| \leq |f(z) - f(z_n)| + |f(z_n) - f_n(z_n)| + |z_n - z| \to 0,$$

which proves $f(z) = z$.

INSTITUT FOURIER, UNIVERSITÉ GRENOBLE I, FRANCE
E-mail address: Michael.Eisermann@ujf-grenoble.fr
URL: www-fourier.ujf-grenoble.fr/~eiserm