Counting Real Roots of Polynomials

Caio De Naday Hornhardt

July 2022

1 Introduction

In the year 1829, Jacques Charles François Sturm presented an algorithm that precisely computes the number of roots (without accounting for multiplicities) of a real polynomial in a given interval. This is surprising, given how hard it is to actually solve polynomial equations.

Sturm is more famous for his work on differential equations, which gave rise to what is now known as Sturm-Liouville Theory. According to Hourya Benis-Sinaceur [BS88] , this was the first abstract theory on differential equations. "This mutation [on the approach to differential equations] stems, in Sturm, from the same disposition as that by which we will see him define Sturm sequences [see below]. It also expresses the conviction — at the time not very commonplace — that it is possible to obtain interesting mathematical statements about numerically indeterminate objects."

Let us wait no further and explain Sturm's algorithm. Consider a non-zero polynomial $f \in \mathbb{R}[x]$ and real numbers $a < b$, none of which are roots of f. Define $f_0 \coloneqq f$, $f_1 \coloneqq f'$ (the derivative of f) and then define recursively $f_{i+2} \coloneqq -r_{i+2}$, where r_{i+2} is the remainder of the division of f_i by f_{i+1}. In other words, f_{i+2} is the unique polynomial whose degree less than the degree of f_{i+1} for which there exists $q \in \mathbb{R}[x]$ such that

$$f_i(x) = q(x)\, f_{i+1}(x) - f_{i+2}(x).$$

We stop at the last non-zero polynomial f_n. We call $f_0, f_1, \ldots, f_n$ the *Sturm sequence* of f. We evaluate all polynomials of this sequence in both a and b, getting two sequences of real numbers. We then count the number of sign changes (ignoring the zeros) in the sequence $f_0(a), f_1(a), \ldots, f_n(a)$ and subtract from it the number of sign changes in the sequence $f_0(b), f_1(b), \ldots, f_n(b)$. The resulting number is the amount of roots of f in the interval (a, b), disregarding their multiplicities. This result is called *Sturm's Theorem*.

As we can see, Sturm's algorithm is based on a slightly modified version of Euclid's algorithm to find the greatest common divisor of f and f'. The signs introduced would be irrelevant if that was the purpose. But not in our case, since we are counting the sign changes.

The curious readers can jump to the proof that this algorithm works in the next section (put ref), or can even try to prove it by themselves. The proof is not hard but also does not give any clue on how would someone discover this ingenious procedure. This is what bugged the author and led him to write this text.

What follows is a compromise between mathematics and history of mathematics. More detailed historic approaches can be found in [BÎ1] and [BS88]. Here get only the main points from these two papers, and prove not only Sturm's Theorem but also two important results that preceded it: the Descarte's Rule of Signs and the Budan-Fourier Theorem. These are not logically necessary to prove our main result, but they contain the ideas that led to it.

But before we dwell on this mathematical genealogy, let us compute a simple example that illustrates the use of the algorithm above:

Example. We are going to deduce a well-known result on how many real solutions a quadratic equation has. Let $f(x) = f_0(x) = x^2 + Bx + C$. Then $f'(x) = f_1(x) = 2x + B$ and, by long division, we get that

$$x^2 + Bx + C = \left(\frac{1}{2}x + \frac{B}{4}\right)(2x + B) + \left(C - \frac{B^2}{4}\right),$$

so

$$f_2(x) = \frac{B^2}{4} - C.$$

Of course, the sign of $f_2(x)$ is the same one of the discriminant $\Delta \coloneqq B^2 - 4C$. Taking the a of our interval (a, b) to be a very big negative number, we get that the sign of $f_0(a)$ is positive while the sign of $f_1(a)$ is negative. Similarly, taking b to be a very big positive number, we get that the signs of $f_0(b)$ and $f_1(b)$ are both positive. Hence it all boils down to the sign of Δ:

- If $\Delta < 0$, the signs at a are $+\,-\,-$ and the signs at b are $+\,+\,-$. So there are $1 - 1 = 0$ real roots.
- If $\Delta > 0$, the signs at a are $+\,-\,+$ and the signs at b are $+\,+\,+$. So there are $2 - 0 = 2$ real roots.
- If $\Delta = 0$, since we ignore the zeros, the signs at a are $+-$ and the signs at b are $++$. So there is $1 - 0 = 1$ real root.

2 Genealogy of Sturm's Theorem

In 1637, René Descartes' published one of his most influential works on Philosophy: *Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences*, or simply *Discourse on the Method.* It is there that his famous quote "I think, therefore I am." is written. As an appendix to *Discourse on the Method*, we can find Descartes' most celebrated work on Mathematics: *The Geometry.* It, for example, introduced the notation for the power of a number (e.g., x^3 for $x \cdot x \cdot x$) and popularized what is now known as the "Cartesian plane." In this same work, the following result is stated without proof:

Theorem 1 (Descartes' Rule of Signs, v. 1)**.** *Consider a non-zero polynomial with real coeficients $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$. The number of sign changes, ignoring the zeros, in the sequence $a_n, a_{n-1}, \ldots, a_0$ is an upper bound for the number of positive roots of $f(x)$, without accounting for multiplicities.*

We are going to give two proofs for this theorem. For the first one, it is better to consider the terms of f in increasing order of exponents. Roughly speaking, the idea is the following: to have a new root, we need to change the local monotonicity of the function f. For x close enough to 0, the monotonicity is controlled by a_1, and if a_0 has the same sign as a_1, we cannot have a root. As we increase x, the next terms of the polynomial become more relevant, and to have a change in monotonicity we need a change in sign.

Proof. By induction on n. If $n = 0$, f is constant: there are no sign changes and no roots. Now suppose $n \geq 1$ and that the result is true for polynomials with degree $n - 1$.

Write $f(x)$ as $a_0 + \cdots + a_n x^n$ and let $a_\ell x^\ell$ be the first non-zero term after a_0, so $f'(x) = \ell a_\ell x^{\ell-1} + \cdots + n a_n x^{n-1}$. Let $r_0, r_1, \ldots, r_k$ the roots of f'. The sign of f' is constant inside each of the intervals $(0, r_0]$, $(r_0, r_1]$, $\ldots$, $(r_{k-1}, r_k]$, and $(r_k, +\infty)$, hence there is at most one root of f inside each one of them. In particular, the number of roots of f greater than r_0 is at most the number of roots of f', which, by the induction hypothesis, is bounded by the number of sign changes in $a_\ell, \ldots, a_n$.

Let us now restrict ourselves to the interval $(0, r_0]$. If $a_0 = 0$, since we ignore the zeros, there isn't a sign change between a_0 and a_ℓ. Also, since $f(0) = 0$ and f is monotone, we cannot have a root in $(0, r_0]$. If $a_0 \neq 0$, let us factor $f'(x)$ as $x^{\ell-1}(\ell\, a_\ell + \cdots + n a_n x^{n-\ell})$. It follows that the sign of f' in $(0, r_0]$ is the same as the sign of a_ℓ and, hence, f is crescent if $a_\ell > 0$ and decrescent if $a_\ell < 0$. If a_0 has the same sign as a_ℓ, then f either starts positive and increases, or f starts negative and decreases. In both cases, there can be no root, concluding the proof. □

Our second proof will avoid derivatives and rely on more elementary methods. This approach actually allows us to give a more precise statement (Theorem 3).

Lemma 2. *Let $f(x)$ be a polynomial with real coefficients and let $r > 0$ be a real number. The amount of sign changes in the coefficients of $g(x) \coloneqq f(x)\,(x - r)$ is greater than the amount of sign changes in the coefficients of $f(x)$ by an odd number.*

Proof. Write $f(x) = a_n x^n + \cdots + a_0$ and $g(x) = b_{n+1} x^{n+1} + \cdots + b_0$. Clearly, we have that $b_{n+1} = a_n$, $b_0 = -a_0 r$ and, for $0 \leq i < n$, $b_{i+1} = a_i - a_{i+1}\,r$.

Suppose there are k sign changes in the sequence $a_n, \ldots, a_0$ and consider the indexes $i_1, i_2, \ldots, i_k$ where a_{i_1} is the first to have sign different from a_n, a_{i_2} is the first after a_{i_1} to change sign again, etc. Since $b_{i_1+1} = a_{i_1} - a_{i_1+1}\,r$ and a_{i_1+1} is either 0 or has the opposite sign of a_{i_0}, it follows that b_{i_1+1} has the same sign as a_{i_1}. Hence, b_{n+1} and b_{i_0+1} have opposite signs, so the number of sign changes between b_{n+1} and b_{i_0+1} is at least 1 (actually, it must be an odd number).

The same argument is valid replacing i_0 by i_ℓ, $0 < \ell \leq k$, and n by $i_\ell - 1$. It follows that we have at least k sign changes in the sequence $b_{n+1}, \ldots, b_{i_k+1}$. Since, by definition, there are no sign changes between a_{i_k} and a_0, we have that b_{i_k+1} has the opposite sign of $b_0 = -a_0\,r$. Hence, the number of sign changes between b_{i_k+1} and b_0 is at least 1 and, therefore, the number of sign changes in the sequence $b_{n+1}, b_n, \ldots, b_0$ is at least $k + 1$.

To show that the difference is an odd number, we simply replace all occurrences of "at least 1" above by "1 plus an even number." $\square$

Theorem 3 (Descartes' Rule of Signs, v. 2)**.** *Let $f(x)$ be a non-zero polynomial with real coefficients. The number of sign changes in the sequence of the coefficients of $f(x)$ is an upper bound for the number of positive roots,* ***accounting*** *for multiplicities. Moreover, the difference between these number of sign changes and the number of roots is even.*

Proof. We can write $f(x)$ as $h(x)\,(x - r_1) \cdots (x - r_k)$ where $h(x)$ has no positive roots and $r_1, \ldots, r_k$ are all the positive roots of f, with repetitions if needed.

We claim that the number of sign changes in the coefficients of $h(x)$ is even. Write $h(x) = x^\ell\,(c_m x^m + \cdots + c_0)$ where both c_m and c_0 are nonzero ($\ell = 0$ if 0 is not a root of h). If the number of sign changes in the sequence $c_m, \ldots, c_0$ was odd, c_m and c_0 would have different signs. But then $h(0)$ would have a different sign than $h(a)$ for a big enough. Since h is a continuous function, it would have a root between 0 and a, a contradiction.

The result follows by applying Lemma 2 iteratively. $\square$

Remark. The proof of Lemma 2 is strictly algebraic, but in the proof of Theorem 3, we used a fact about continuous functions from $\mathbb{R}$ to $\mathbb{R}$. It is unavoidable that we use somewhere that we are dealing with $\mathbb{R}$ and not an arbitrary ordered field. For instance, the "moreover" part is not true over $\mathbb{Q}$: in this case, the polynomial $x^2 - 2$ has no positive root, but the rule of signs gives us an odd upper bound.

Descartes's Rule of Signs can be easily adapted to compute the number of real roots greater than any real number a: we can rewrite $f(x)$ as a polynomial $g(y)$ where $y = x - a$, i.e., $g(y) = f(y + a)$. One possible way of finding the coefficients of g is using that g is the Taylor's polynomial of degree $n \coloneqq \deg f$ around the point a, so $g(y) = \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!} y^i$.

Corollary 4. *Let $f(x)$ be a non-zero polynomial with real coefficients and degree n, and let a be a real number. The number of sign changes in the sequence $f(a), f'(a), f^{(2)}(a), \ldots, f^{(n)}(a)$ is an upper bound for the number of roots of f that are greater than a, accounting for multiplicities.*

At this point, it is an educated guess that to find an upper bound on the number of roots of f in an interval $(a, b]$ one simply compute the number of sign changes in the sequence $f(a), f'(a), f^{(2)}(a), \ldots, f^{(n)}(a)$ and subtract from it the number of sign changes in the sequence $f(b), f'(b), f^{(2)}(b), \ldots, f^{(n)}(b)$. At least in the case the upper bound given by Corollary 4 is the actual number of roots in both intervals $(a, +\infty)$ and $(b, +\infty)$, this algorithm would give us the right amount of roots in $(a, b]$. Of course, the upper bound can only be considered an estimative, and it could be the case that the error in the estimative we get for the number of roots in $(b, +\infty)$ is bigger than the error of the estimative for $(a, +\infty)$. But both Budan (1807) and Fourier (1820) proved this was not the case.

Theorem 5 (Budan-Fourier)**.** *Bla, bla*

But here we do not care about who deserves or not the accolade of having its naming attached to the theorem. Maybe Fourier was not the first to discover the theorem, but it was his proof and his analysis of how the mistakes in the number of roots appear that were the base for Sturm, who said "I based myself on Fourier."

Lemma 6. *Derivatives and multiple roots.*

Proof. bla □

Proof of Theorem 5. bla □

It is interesting to note that only the last case got the error $= 0$.

Sturm and Fourier together quote. Sturm-Liouville is an abstraction of Fourier series. In the same way Hourya quote.

But guy argues that the abstraction was not without another example to sustain. He read blah, where there is a work of sturm in a differential equation problem, heat and waves, see the paper of the guy for details. There Sturm gets a sequence, and voilà.

If instead of an upper bound we had the precise number of roots, we could then find the number of roots in any interval $(a, b]$ by computing blah minus blah. But it is only an upper bound, so we can't conclude this, not even that blah minus blah is an upper bound on the number of roots. Not as a corollary, at least. Actually, the latter turns out to be true. This was first published by Budan in yyyy (who stated his theorem in terms of the coefficients of the polynomials in $x - a$ and $x - b$ rather than in terms of derivatives) and by Fourier in yyyy. According to [11], "Fourier wanted to prove it knew it before... ." But here we do not care about who deserves or not the accolade of having its naming attached to the theorem. Maybe Fourier was not the first to discover the theorem, but it was his proof and his analysis of how the mistakes in the number of roots appear that were the base for Sturm, who said "I based myself on Fourier."

Proposition 7. *Multiple roots and derivatives.*

Theorem 8 (Budan-Fourier)**.** *Blah*

Proof. Blah. Four cases in the end. □

The last case is very interesting. All others introduce errors. Cite Hourya.

Definition 9. Generalized Sturm sequence.

Proposition 10. *If gcd of f and f' is* 1*, then the sturm sequence (def. 2) is a generalized sturm sequence.*

Theorem 11 (Sturm's theorem, abstract version)**.** *Blah*

Faster algorithms to compute a generalized sturm sequence, different from the one in Def 2 exist (cite the paper in Pocket).

Example 12. The example I have in my dissertation that is not the classical Sturm sequence

According to [11], Sturm didn't abstract the axioms of Definition 9 only from Fourier's work. Sturm also said that "it came from his works on difference equations." The guy goes beyond that. By reading blah, he finds a differential equation problem (whose Sturm's approach to it is detailed in [11]) that ends up in a sequence of polynomials that satisfies those axioms. Sturm is interested in the number of roots of y_0, and must have realized that the proof of Fourier worked for the sequence he had, and that only the exceptional case that doesn't introduce mistakes appeared. That sequence was produced by a difference equation of the form:

equation:blah

Then the guy conjectures that, putting the second polynomial as the derivative, and trying to figure out a equation like (1, above), the modified euclidian division is a fit.

(Put a reference to this paragraph with the hyperlink package) It only remains now on how deduce the Sturm theorem (cite) from Theorem 11. This is easy: instead of considering intervals $(a_i, b_i]$ with only one root of an element of the sequence, we can consider intervals of the form (a_i, b_i), where none of a_i and b_i are roots, blah blah.

References

[BÎ1] Maxime Bôcher, *The published and unpublished work of Charles Sturm on algebraic and differential equations*, Bull. Amer. Math. Soc. **18** (1911), no. 1, 1–18. MR 1559126

[BS88] Hourya Benis-Sinaceur, *Deux moments dans l'histoire du théorème d'algèbre de Ch. F. Sturm*, Rev. Histoire Sci. **41** (1988), no. 2, 99–132. MR 954562