Root-Finding Algorithms | Instructor: Padraic Bartlett

Finding *All* the Roots: Sturm's Theorem

Day 2 | *Mathcamp 2013*

In our last lecture, we studied two root-finding methods that each took in a polynomial $f(x)$ and an interval $[a, b]$, and returned a root of that function on that interval. This was great for the problem we asked at the start of the class — how to find a root of a quintic polynomial — but is not necessarily so great for many other problems we may want to study. For example, we may want to find a root of a degree-6 polynomial! In this situation, we can't obviously use either of the methods we've examined earlier: we have no obvious way to find an interval on which an even-order polynomial changes sign. Also, we may often want to find not just one root, but **all of the roots** of a given polynomial! Our earlier methods only give us one root, which is not necessarily very useful to us in practice.

In this lecture we're going to study Sturm's theorem, a tool that helps with both of these problems.

1 Sturm's Theorem

In order to state Sturm's theorem, we need to make some definitions.

1.1 Definitions and Notation

Definition. A **Sturm chain** is a finite sequence of polynomials $p_0(x), p_1(x), \dots p_m(x)$ of decreasing degree with the following properties:

1. $p_0(x)$ is square-free: i.e. it has no factors of the form $(q(x))^2$, for any polynomial $q(x)$.
2. If a is a root of $p(x)$, then the sign of $p_1(a)$ is the same as the sign of $p'(a)$, and in particular is nonzero.
3. If a is a root of $p_i(x)$ for some i with $0 < i < m$, then both $p_{i-1}(a), p_{i+1}(a)$ are nonzero. Moreover, the sign of $p_{i-1}(a)$ is the opposite of the sign of $p_{i+1}(a)$.
4. $p_m(x)$'s sign is constant and nonzero for all x.

Sturm chains look very strange initially. Before we get into why we'd want to study them, however, we offer an example of how to construct Sturm chains.

Proposition. Given a square-free polynomial $p(x)$, the following construction gives us a Sturm chain:

- $p(x) = p_0(x)$.
- $p_1(x) = p'(x)$.

- $p_2(x) = -\text{rem}(p_0(x), p_1(x)) = p_1(x)q_0(x) - p_0(x)$, where $\text{rem}(p_0(x), p_1(x))$ is the remainder when **polynomial long division**[1] is is used to divide $p_0(x)$ by $p_1(x)$, and $q_0(x)$ is the quotient of said long division.
- $p_3(x) = -\text{rem}(p_1(x), p_2(x)) = p_2(x)q_1(x) - p_1(x)$.
- In general, define p_k inductively as $-\text{rem}(p_{k-2}(x), p_{k-1}(x)) = p_{k-1}(x)q_{k-2}(x) - p_{k-2}(x)$.
- Repeat this process until we arrive at some m such that $\text{rem}(p_{m-1}, p_m) = 0$, where $p_m(x) \neq 0$. (Because our degrees are decreasing by at least one at each step, this will eventually occur.)

Proof. By definition, we know the first two properties in the definition of Sturm chains hold. So the only nontrivial conditions that we need to examine are the last two properties of Sturm chains.

To see the third property, we'll use a proof by induction. First notice that $p_0(x)$ and $p_1(x)$ do not share any common roots: this is because if a is a root of $p(x)$, we can write $p(x) = (x-a) \cdot q(x)$, for some other polynomial $q(x)$ that doesn't have a root at a (we know that $q(x)$ doesn't have a root at a because $p(x)$ is squarefree.) Then, if we differentiate, we get that $p'(x) = q(x) + (x-a)q'(x)$, which is equal to $q(a)$ at a, and in particular is nonzero.

For our inductive step, notice that $p_{i-1}(x) = p_i(x)q_{i-1}(x) - p_{i+1}(x)$. Therefore, if $p_{i+1}(x), p_i(x)$ had a common root for some i, this root would be shared with $p_{i-1}(x)$, and (by recursion) this root would be shared with all of our polynomials. We showed, however, that $p_0(x), p_1(x)$ do not share a common root, so this is impossible. Therefore, at any root a of $p_i(x)$, for $0 < i < m$, we've shown that both $p_{i-1}(a)$ and $p_{i+1}(a)$ must be nonzero.

To see that at any root a of $p_i(x)$, the sign of $p_{i-1}(a)$ is the opposite of the sign of $p_{i+1}(a)$: just look at our construction. We know that $p_{i+1}(x) = p_i(x)q_{i-1}(x) - p_{i-1}(x)$, by definition. If we plug in a to this equation, we get $p_{i+1}(a) = -p_{i-1}(a)$, our claim. This finishes our proof that the third property holds.

We are left with the last property: i.e. showing that $p_m(x)$'s sign is a nonzero constant. To prove this, we refer to our construction again. We know that $\deg(p_i) > \deg(p_{i+1})$, for any i, by construction, because we're using polynomial long division. So our process eventually terminates, and yields some $p_m(x)$ such that $p_m(x)$ is both nonzero and satisfies $p_{m-1}(x) = q_{i-1}(x) \cdot p_m(x)$.

If $p_m(x)$ is not a constant, is this possible? Well, if $p_m(x)$ and $p_{m-1}(x)$ share $p_m(x)$ as a common factor, then by definition $p_{m-1}(x)$ also shares this factor, and (by recursion) so does **every** polynomial in our chain.

Can this happen? Well: again, look at $p_0(x)$ and $p_1(x)$. We know that if $p_m(x)$ is a factor of $p_0(x)$, then if we write $p_0(x) = p_m(x) \cdot q(x)$, we have $p_1(x) = p'_m(x) \cdot q(x) + p_m(x) \cdot q'(x)$. If $p_m(x)$ is a factor of $p_1(x)$ as well, we can use this equation to conclude that $p_m(x)$ is also a factor of $p'_m(x) \cdot q(x)$. Because $p'_m(x)$ has strictly lower degree than $p_m(x)$, we know that in order for this to be true, there must be some factor of $p_m(x)$ that is a factor of $q(x)$. However, this means that this factor occurs twice in $p_0(x) = p_m(x) \cdot q(x)$. In order for $p_0(x)$ to be squarefree, as claimed, this $p_m(x)$ must be a constant, as claimed.

[1] Given two polynomials p_0, p_1, there is a unique pair of polynomials q, r such that $p_0 = p_1 \cdot q + r$ and the degree of r is either 0 or lower than the degree of p_1. The process we use to find this q, r is polynomial long division; come talk to me if you haven't done this before, or have forgotten how it goes!

So this process generates a Sturm chain, as claimed.

1.2 Stating and Proving Sturm's Theorem

Sturm chains are pretty odd things; from their construction, it's not immediately obvious what they are, or why we care about them. As it turns out, however, they're incredibly useful things, that were very carefully and cleverly constructed so that the following theorem would be true:

Theorem. Take any squarefree polynomial $p(x)$, and any interval (a, b) such that $p_i(a), p_i(b) \neq 0$, for any i. Let $p_0(x), \ldots p_m(x)$ denote the Sturm chain corresponding to $p(x)$. For any constant c, let $\sigma(c)$ denote the number of changes in sign in the sequence $p_0(c), \ldots p_m(c)$. Then $p(x)$ has $\sigma(a) - \sigma(b)$ distinct roots in the interval (a, b).

Proof. First, notice that near any root a of $p(x)$, our function $p(x)$ is negative on one side of a and positive on the other side of a. This is because our function, when factored, looks like $(x - a) \cdot g(x)$, for some function $g(x)$ that does not have a root at a (because $p(x)$ is squarefree) and therefore has constant sign near a.

So. Imagine, for the moment, the function σ sweeping left-to-right across the real number line. We know that σ cannot change unless the sign of one of our polynomials $p_i(x)$ changes. Because polynomials are continuous, this can only happen if we pass over a zero of some $p_i(x)$. Therefore: at any such root of some $p_i(x)$, we want to show that σ decreases by 1 if we pass over a root of $p_0(x)$, and does not change if we pass over a zero of any other $p_i(x)$.

We first consider the case where $p_i(x)$ has a root, for some $i \geq 1$. Call it a. At a, we know that $p_{i-1}(a), p_{i+1}(a)$ are both nonzero and opposite signs. Furthermore, because these functions do not have a root at a, near a these signs do not change.

If $p_i(x)$ switches from being positive to negative, our sign pattern here goes from either $-++$ to $--+$ or $+--$ to $++-$, depending on the signs of $p_{i-1}(a), p_{i+1}(a)$. Similarly, if $p_i(x)$ switches from negative to positive, our sign pattern goes from either $--+$ to $-++$ or from $+--$ to $++-$. In any of these cases, the total number of sign changes is undisturbed. Therefore σ does not change.

Now, suppose that $p_0(x)$ has a root at a. If it switches from positive to negative, we know that in a small neighborhood of a our function must have negative derivative, as our polynomial is decreasing. Therefore our sign pattern has gone from $+-$ to $--$, and σ has decreased by 1. Similarly, if we switch from negative to positive, our function is increasing and has positive derivative near a; consequently, our sign pattern goes from $-+$ to $++$. Again, σ decreases by 1, which is what we claimed.

Therefore, $\sigma(a) - \sigma(b)$ is the total number of roots on the interval (a, b), as claimed.

We can even strengthen this to general polynomials as follows:

Corollary. Take any polynomial $p(x)$, and any interval (a, b) such that $p_i(a), p_i(b) \neq 0$, for any i. Then $p(x)$ has $\sigma(a) - \sigma(b)$ distinct roots in the interval (a, b).

Proof. Take any polynomial $p(x)$. Let $d(x)$ be the greatest common divisor of $p(x), p'(x)$. (In the case that $p(x)$ is squarefree, this is a constant, as proven earlier.)

Use the construction $p_0(x), p_1(x), \ldots p_m(x)$ that we created a Sturm chain with earlier. Now, divide every term in our chain by the quantity $d(x)$. Because $d(x)$ was a factor of both $p_0(x), p_1(x)$ it's a factor of every term of our chain by construction, so the resulting object is still a chain of polynomials. Furthermore, after doing so no two consecutive terms in our chain share a common nonconstant factor, again by construction. Because $p(x)$ has no roots at a, b, we know that $d(x)$ is nonzero at a, b, and therefore that this new chain has the same number of sign changes at a, b as the old chain did (because at a, b this $d(x)$ is just some nonzero constant.)

We claim that we can apply Sturm's theorem to this new chain, and it will give us the number of distinct roots of $p(x)/d(x)$ on the interval $[a, b]$. Note that $d(x)$ only contains repeated roots of $p(x)$; this is because it was formed by taking the gcd of $p(x)$ and another polynomial. Therefore, we know that if this new chain is Sturm, counting the number of distinct roots in $p(x)/d(x)$ is the same as counting the number of distinct roots of $p(x)$ itself.

This chain satisfies the first property of Sturm chains: we've forced the first term to be squarefree, by factoring out any repeated factors it has. (This is because any factor that shows up twice in $p(x)$ shows up at least once in $p'(x)$; prove this using the product rule and the derivative!)

It also satisfies the "no two consecutive zeroes" property. This is because if there was any repeated pair of zeroes, we would have two consecutive terms with a common factor, which (by using the recursive/inductive arguments we've made throughout this lecture) would force all of the terms of our sequence to share a common factor. In particular it would force $p(x)$ and $p'(x)$ to share a common factor, and we've made this impossible by factoring out such things.

It still satisfies the "alternating signs on either side of a root" property, because this property is unaltered by factoring out terms and is true in our original construction.

As well, its last term $p_m(x)$ is a nonzero constant, because (as shown earlier) we know that $p_m(x)$ is a factor of every other element in our chain, and we've factored out all common terms with degree greater than a constant.

Finally, we have the requirement that at any root a of the leading term in our chain, we have that the sign of the second term in our chain is the same as the sign of the derivative of the first term in our chain at a. Is this true? Well: the leading term in our chain is $\frac{p(x)}{d(x)}$, which has derivative $\frac{d}{dx}\left(\frac{p(x)}{d(x)}\right) = \frac{p'(x)d(x) - p(x)d'(x)}{d^2(x)}$. At a, this is $\frac{p'(a)d(a)}{d^2(a)} = \frac{p'(a)}{d(a)}$, if we use the observation that a is a root of $p(x)$. The second term in our chain is $\frac{p'(x)}{d(x)}$, which is precisely this at a; therefore this property is preserved.

So we have a chain that is Sturm: therefore, if we apply our theorem, we can count the number of roots in the leading polynomial! Because we've only factored out repeated factors, we know that this leading polynomial has the same number of distinct roots as our original polynomial $p(x)$: so counting the roots of this new polynomial is the same thing as counting roots of $p(x)$. As a consequence, because these factors (as noted) don't effect the total number of sign changes at a given point, we in fact have that $p(x)$ has $\sigma(a) - \sigma(b)$ distinct roots in the interval (a, b).

This is why we care about Sturm chains: they let us find **all** of the roots of a polynomial! In particular, consider the following algorithm for finding roots of a polynomial:

Root-Finding Algorithm 3: Sturm's Theorem

Input: A polynomial $p(x) = a_n x^n + \ldots a_1 x + a_0$, along with a desired error tolerance ϵ. As well, take as input an interval (a, b) that we want to find roots within. Require that $p_i(a), p_i(b) \neq 0$, for any $p_i(x)$ in the Sturm chain corresponding to $p(x)$.

1. Calculate the Sturm chain $p_0(x), \ldots p_m(x)$ corresponding to $p(x)$.
2. Take our interval (a, b). Using Sturm's theorem, find $\sigma(a) - \sigma(b)$, which tells us the total number of roots in (a, b).
3. If this total number of roots is zero, halt: there are no roots here. If the total number of roots is 1, use your favorite root-finding algorithm here.
4. Otherwise, the total number is k for some $k \geq 2$. If the length of this interval is smaller than 2ϵ, halt and report that there are k roots at $\frac{b+a}{2}$ within our error tolerances.
5. Otherwise, if $c = \frac{b+a}{2}$ is not a root of $p_i(x)$ for any i, divide this interval into two equal halves $(a, c), (c, b)$, and run step 2 on each interval. Else, if c was a root of one of these $p_i(m)$'s, add a tiny amount to c so that c still lies fairly close to the middle of (a, b) and is not a root. (The precise manner in which you do this doesn't matter too much, and probably depends on what you're coding your algorithm on. In any case, this should be relatively easy, as each p_i has at most m roots, so we're only trying to avoid finitely many points and can easily do so by adding very tiny bits to the midpoint of our interval.) Again, run step 2 on each of the intervals $(a, c), (c, b)$.

Output: Approximations to all of the roots of $p(x)$ within our interval, with an error of at most ϵ for any root.

We calculate an example here:

Example. Find all of the roots of $p(x) = x^6 - 4x^3 + x - 2$, with an error of $\pm\frac{1}{10}$.

Answer. We start by calculating Sturm chains, via our algorithm:

- $p_0(x) = x^6 - 4x^3 + x - 2$.
- $p_1(x) = 6x^5 - 12x^2 + 1$.
- $p_2(x) = -\text{rem}(x^6 - 4x^3 + x - 2, 6x^5 - 12x^2 + 1) = 2x^3 - \frac{5x}{6} + 2$.
- $p_3(x) = -\text{rem}(6x^5 - 12x^2 + 1, 2x^3 - \frac{5x}{6} + 2) = 18x^2 - \frac{25x}{24} + \frac{3}{2}$.
- $p_4(x) = -\text{rem}(2x^3 - \frac{5x}{6} + 2, 18x^2 - \frac{25x}{24} + \frac{3}{2}) = \frac{92687x}{93312} - \frac{5159}{2592}$.
- $p_5(x) = -\text{rem}(18x^2 - \frac{25x}{24} + \frac{3}{2}, \frac{92687x}{93312} - \frac{5159}{2592}) = -\frac{12568084416}{175324081}$.

If we notice that for any $|x| \geq 2$, $|x|^6 > 4|x|^3 + |x| + |-2|$, we can conclude that all of $p(x)$'s roots occur on the interval $(-2, 2)$. So, we proceed via our algorithm. At $x = 2$, our Sturm chain has values (approximately)

$p_0(x)$	$p_1(x)$	$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$
32	145	16.33	71.42	$-3.8 \cdot 10^{-3}$	-71.68

At $x = -2$, our Sturm chain has values

$p_0(x)$	$p_1(x)$	$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$
92	-239	-12.33	75.58	-3.98	-71.68

Therefore, $\sigma(2) = 1, \sigma(-2) = 3$, and the total number of roots in $(-2, 2)$ is $3 - 1 = 2$.

We subdivide to try to improve these bounds. If we look at our chain at $x = 0$, we have

$p_0(x)$	$p_1(x)$	$p_2(x)$	$p_3(x)$	$p_4(x)$	$p_5(x)$
-2	1	2	1.5	-1.99	-71.68

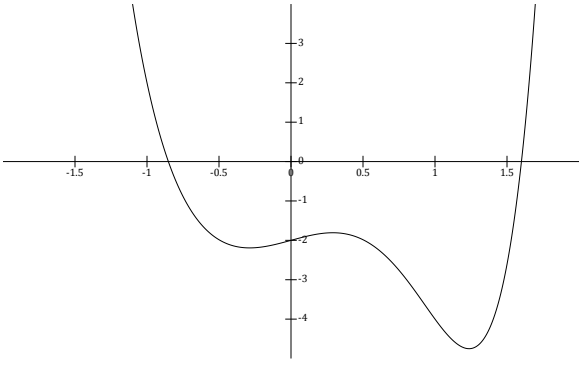
Therefore there is one root on $(-2, 0)$ and another root in $(0, 2)$. Because $p(-2) = 92, p(0) = -2, p(2) = 32$, we can apply the method of bisections on both of the intervals $[0, 2]$ and $[-2, 0]$:

interval $[a, b]$	midpoint of $[a, b]$	function at midpoint
$[-2, 0]$	-1	$p(-1) = 2.$
$[-1, 0]$	$-.5$	$p(-.5) \approx -1.98.$
$[-1, -.5]$	$-.75$	$p(-.75) \approx -.88.$
$[-1, -.75]$	$-.875$	$p(-.875) \approx .25.$
$[-.875, -.75]$	$-.8125$	

interval $[a, b]$	midpoint of $[a, b]$	function at midpoint
$[0, 2]$	1	$p(1) = -4.$
$[1, 2]$	1.5	$p(1.5) \approx -2.61.$
$[1.5, 2]$	1.75	$p(1.75) \approx 7.04.$
$[1.5, 1.75]$	1.625	$p(1.625) \approx .87.$
$[1.5, 1.625]$	1.5625	

1.5625 and $-.8125$ are thereby roots of our function to within our desired accuracy.

Graphing our function offers a nice verification that this process has indeed found both of our roots:



Root-Finding Algorithms | Instructor: Padraic Bartlett

Homework 2: Sturm's Theorem + More Root-Finding Problems

Day 2 | *Mathcamp 2013*

Starred problems are harder.

1. Can you find a function $f(x)$ and an interval $[a, b]$ such that both the method of false position and the bisection method converge to roots of $f(x)$, but they converge to different roots?

2. Use Sturm's theorem to find the roots of $x^4 - 3x^2 + x - 2$.

3. Similarly, use Sturm's theorem to find the roots of $x^6 + 2x^3 + x^2 + 1$.

4. Calculate the Sturm chains for an arbitrary quadratic polynomial $ax^2 + bx + c$. From looking at these chains, under what condition does your polynomial have two distinct roots? Exactly one root? No roots? Why are your results "expected?"