

STURM'S ROOT COUNTING THEOREM

KENNETH R. DRIESSEL

Remark: I mainly follow the books Basu, Pollack and Roy(2003) and van der Waerden(1970)

Date: November 20, 2006.

Let R be an ordered ring.

Definition: Let $a := (a_0, a_1, \dots, a_p)$ be a finite sequence of nonzero elements of R . Then the **number of sign changes** $V(a)$ in the sequence is defined by

- $V(a_0) := 0$
- $V(a_0, a_1, \dots, a_p) := V(a_1, \dots, a_p) + 1$ if $a_0 a_1 < 0$ and
- $V(a_0, a_1, \dots, a_p) := V(a_1, \dots, a_p)$ if $a_0 a_1 > 0$.

Let $a := (a_0, a_1, \dots, a_p)$ be a finite sequence of elements of R in which zeroes are allowed. Then the **number of sign changes** $V(a)$ in the sequence is defined by $V(a) := V(b)$ where b is the sequence obtained from a by dropping its zeroes.

Example: $V(1, -1, 2, 0, 3, -5, -2, 0, 3) = V(1, -1, 2, 3, -5, -2, 3) = 4$.

Proposition 1. *Let $a := (a_0, a_1, \dots, a_p)$ be a finite sequence of elements of R and let c be a nonzero element of R . Then $V(c \cdot a) = c \cdot V(a)$ where $c \cdot a = (ca_0, ca_1, \dots, ca_p)$.*

Definition: Let $\mathcal{P} := (P_0, P_1, \dots, P_d)$ be a finite sequence of univariate polynomials in $R[X]$ and let a and b be elements of $R \cup \{-\infty, \infty\}$. Then the **number of sign changes** of \mathcal{P} at a , denoted $V(\mathcal{P}; a)$ is defined by

$$V(\mathcal{P}; a) := V(P_0(a), P_1(a), \dots, P_d(a)).$$

And $V(\mathcal{P}; a, b) := V(\mathcal{P}, a) - V(\mathcal{P}, b)$.

Remark: Recall that for sufficiently large values x the sign of $P(x)$ is the same as the sign of the leading term of the polynomial P . We use this fact to interpret $P(\pm\infty)$.

Recall the following result. (See, for example, Rotman(2002).)

Proposition 2. *(Euclidean division algorithm) Let F be a field. Let $P, Q \in F[X]$ with $Q \neq 0$. Then there are unique polynomials A and R in $F[X]$ such that*

$$P = AQ + R$$

and $\text{degree}(R) < \text{degree}(Q)$.

Definition: The polynomial A is called the **quotient** and the polynomial R is called the **remainder** obtained by dividing P by Q .

Recall the following result. (See, for example, Rotman(2002).)

Proposition 3. (Euclidean algorithm) Let F be a field. Let $P, Q \in F[X]$. Let A_i and R_i be the sequence of quotients and remainders obtained by iterating the division algorithm starting with P and Q :

$$\begin{aligned} S_0 &:= P \\ S_1 &:= Q \\ S_0 &=: A_0 S_1 - S_2 \\ S_1 &=: A_1 S_2 - S_3 \\ &\dots \\ S_{n-2} &=: A_{n-2} S_{n-1} - S_n \\ S_{n-1} &=: A_{n-1} S_n. \end{aligned}$$

Then S_n is the greatest common divisor of the polynomials P and Q .

Definition: The sequence of polynomials determined by the Euclidean algorithm $S(P, Q) := (S_0, S_1, \dots, S_n)$ is called the **signed remainder sequence** determined by P and Q .

Proposition 4. Let F be a field. Let $P \in F[X]$ and let G be the greatest common divisor of P and P' . Let r be a root of P in F . Then r is a multiple root of P iff r is a root of G .

Proof. (\Rightarrow) Assume that r is a multiple root of P . Let

$$P(X) = (X - r)^m A(X)$$

where m is the multiplicity of r . Then

$$P'(X) = m(X - r)^{m-1} A(X) + (X - r)^m A'(X).$$

Since $m > 1$ we have that $X - r$ divides both P and P' . Hence $X - r$ divides G .

(\Leftarrow) Assume that $G(r) = 0$. Then $X - r$ divides G . Hence $X - r$ divides both P and P' . Let $P = (X - r)Q$. Then $P' = Q + (X - r)Q'$. Since $X - r$ divides P' we see that $X - r$ divides Q and hence r is a multiple root of P . \square

Proposition 5. Root counting. (Sturm, 1835) Let R be a real closed field. Let P be a polynomial in $R[X]$. Let $a < b$ be elements of $R \cup \{-\infty, \infty\}$ which are not roots of P . Let $S(P, P')$ be the signed remainder sequence determined by P and its derivative P' . Then

$$\#\{r \in R : a < r < b \wedge P(r) = 0\} = V(S(P, P'); a) - V(S(P, P'); b).$$

Remark: Note that $V(S(P, P'); a)$ is the number of sign changes in the sequence $(S_0(a), S_1(a), \dots, S_n(a))$. The signed remainder sequence $S(P, P')$ is called the **Sturm chain** or **Sturm sequence** for P . The theorem states that the number of distinct roots between a and b is given by the number of sign changes in Sturm's chain which are lost in passing from a to b .

Proof. We begin by showing that we need only look at signed remainder sequences that start with polynomials with simple roots.

Let G be the greatest common divisor of P and P' . Note that the number of distinct roots of P between a and b is the same as the number of simple roots of P/G between a and b .

Note that $G(a) \neq 0$ and $G(b) \neq 0$ since a and b are not roots of P . Hence

- $V(S(P, P'); a) = V(S(P/G, P'/G); a)$ and
- $V(S(P, P'); b) = V(S(P/G, P'/G); b)$.

(Note, in particular, that G evenly divides every polynomial in the Sturm chain $S(P, P')$.)

So we consider the signed remainder sequence $\mathcal{S} := (S_0, S_1, \dots, S_k) := S(P/G, P'/G)$ defined by

$$\begin{aligned} S_0 &:= P/G \\ S_1 &:= P'/G \\ S_2 &:= A_0 S_1 - S_0 \\ &\dots \\ S_j &:= A_{j-2} S_{j-1} - S_{j-2} \\ &\dots \\ S_k &:= A_{k-2} S_{k-1} - S_{k-2} \\ 0 &:= A_{k-1} S_k - S_{k-1}. \end{aligned}$$

Claim: The last term S_k of this sequence is a nonzero element of R .

Note that the polynomials S_0 and S_1 are relatively prime.

Claim: No two successive terms of the sequence vanish at any element of R .

Suppose that $S_j(c) = S_{j+1}(c) = 0$ for some c in R . Then, from the definition of the chain, we get $S_{j+2}(c) = \dots = S_k(c) = 0$ which is a contradiction. This completes the proof of the claim.

We now consider how $V(\mathcal{S}; c)$ changes as c moves from a to b . In particular, we want to see that the following quantity remains constant:

$$\#\{r \in R : a < r < c \wedge P(r) = 0\} + V(\mathcal{S}; c).$$

Note that the roots of the polynomials S_j divide the interval (a, b) into subintervals. We consider various cases.

Case: The element c is inside one of the subintervals.

By the interpolation property, in the interior of any of the subintervals, all of the S_j retain their signs. Hence $V(\mathcal{S}, c)$ remains constant on such a subinterval.

Case: For some $j > 0$, $S_j(c) = 0$.

Consider the equation $S_{j+1} = A_{j-1} S_j - S_{j-1}$. From it we see that $S_{j+1}(c) = -S_{j-1}(c)$. Hence $S_{j+1}(c)$ and $S_{j-1}(c)$ have different signs. It follows that the polynomials S_{j+1} and S_{j-1} have different signs near c – namely, on the two subintervals adjacent to c . Hence the number $V(\mathcal{S}; c)$

does not change on the passage thru c . (To better see this situation, it helps to draw a picture.)

Case: The polynomial P vanishes at c ; in symbols, $P(c) = 0$.

Then there exists a positive integer l and a polynomial B in $R[X]$ such that

$$P(X) = (X - c)^l B(X) \text{ and } B(c) \neq 0.$$

(Note that l is the multiplicity of c as a root of P .) We have

$$P'(X) = l(X - c)^{l-1} B(X) + (X - c)^l B'(X).$$

Note, for x near c , we have

$$\text{Sign}(S_0(x)) = \text{Sign}((P/G)(x)) = \text{Sign}((x - c)B(x))$$

and

$$\text{Sign}(S_1(x)) = \text{Sign}((P'/G)(x)) = \text{Sign}(B(x)).$$

(In particular, note that $(X - c)^{l-1}$ is a factor of G since it divides $P(X)$ and $P'(X)$.) There are two sub-cases.

Sub-case: The sign B is positive at c ; in symbols, $B(c) > 0$.

Then B is positive on an open interval containing c and S_0 is strictly increasing on this interval. It follows that $V(\mathcal{S}, x)$ decreases by one when moving thru c from left to right. (Again it helps to draw a figure.)

Sub-case: The sign of B is negative at c ; in symbols, $B(c) < 0$.

Then S_0 is strictly decreasing on the interior of an interval containing c . It follows again that $V(\mathcal{S}, x)$ decreases by one when moving thru c from left to right. (Again it helps to draw a figure.) \square

Anders Jensen provided the following example.

Example: Consider the polynomial $f = 8 - 4X + 6X^2 - 3X^3 - 2X^4 + X^5 = (X^2 + 1)(X - 2)^2(X + 2)$.

The Sturm sequence is as follows:

$$f_0 = 8 - 4X + 6X^2 - 3X^3 - 2X^4 + X^5$$

$$f_1 = -4 + 12X - 9X^2 - 8X^3 + 5X^4$$

$$f_2 = -\frac{192}{25} + \frac{56}{25}X - \frac{72}{25}X^2 + \frac{46}{25}X^3$$

$$f_3 = \frac{2500}{529} - \frac{17500}{529}X + \frac{8125}{529}X^2$$

$$f_4 = \frac{33856}{4225} - \frac{16928}{4225}X.$$

We evaluate the polynomials at $x = -3, -2, -1, 0, 1, 2, 3$ to get the following table:

x	$f_0(x)$	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$	$V(x)$
-3	-250	500	-90	128125/529	16928/845	3
-2	0	80	-192/5	70000/529	67712/4225	2
-1	18	-12	-366/25	28125/529	50784/4225	2
0	8	-4	-192/25	2500/529	33856/4225	2
1	6	-4	-162/25	-6875/529	16928/4225	2
2	0	0	0	0	0	0
3	50	140	114/5	23125/529	-16928/4225	1

Note, for example, that $f(-3) \neq 0$, $f(3) \neq 0$, the variation at -3 is 3 and the variation at 3 is 1. From Sturm's theorem we conclude that the polynomial f has two distinct real roots between -3 and 3.

Remark: Note that determining the number of distinct roots of a polynomial with real coefficients is an ill posed problem. Consider, for example, the polynomial $P(X) := X^2 + \epsilon$. If $\epsilon > 0$ then this polynomial has no real roots; if $\epsilon = 0$ then it has one repeated root; if $\epsilon < 0$ then it has two roots. Consequently, we should normally do the computations associated with Sturm's root counting theorem using exact arithmetic.

Remark: Let P be a polynomial with rational coefficients. Let $a < b$ be rational numbers which are not roots of P . Note that the Sturm computations of the variations $V(S(P, P'); a)$ and $V(S(P, P'); b)$ involve only arithmetic in the field of rational numbers. Hence the difference between these two numbers determines the number of distinct roots of P in any real closed field which contains the rational numbers.

More generally, suppose that F is an ordered field and $F \subseteq R$ where R is any real closed field. If P is a polynomial in $F[X]$ then we can do the variation calculations in F . We obtain a count of the number of the number of distinct roots of P in R .

The next result is a generalization of Sturm's root counting theorem.

Proposition 6. Root counting (*Sylvester, Tarski*). *Let R be a real closed field. Let P and Q be polynomials in $R[X]$. Let $a < b$ be elements of $R \cup \{-\infty, \infty\}$ which are not roots of P or Q . Let $S(P, P'Q)$ be the signed remainder sequence determined by polynomials P and $P'Q$. Then*

$$\begin{aligned} & \#\{r \in R : a < r < b \wedge P(r) = 0 \wedge Q(r) > 0\} - \\ & \#\{r \in R : a < r < b \wedge P(r) = 0 \wedge Q(r) < 0\} \\ & = V(S(P, P'Q); a) - V(S(P, P'Q); b). \end{aligned}$$

Proof. The proof is very similar to the proof of Sturm's root counting theorem. We omit some of the details.

We begin by eliminating the factors common to P and $P'Q$. Let G be the greatest common divisor of P and $P'Q$. Note that in the root counts we can ignore the roots of P which are also roots of $P'Q$. In particular, if $P(s) = 0$ and $Q(s) = 0$ then s does not enter into the root counts; if $P(s) = 0$ and $P'(s) = 0$ then s is a multiple root and gets counted correctly.

Note that $G(a) \neq 0$ and $G(b) \neq 0$ since a and b are not roots of P . Hence

- $V(S(P, P'Q); a) = V(S(P/G, P'Q/G); a)$ and
- $V(S(P, P'Q); b) = V(S(P/G, P'Q/G); b)$.

(Note, in particular, that G evenly divides every polynomial in the sequence $S(P, P'Q)$.)

So we consider the signed remainder sequence $\mathcal{S} := (S_0, S_1, \dots, S_k) := S(P/G, P'Q/G)$ defined by

$$\begin{aligned} S_0 &:= P/G \\ S_1 &:= P'Q/G \\ S_2 &:= A_0S_1 - S_0 \\ &\dots \\ S_j &:= A_{j-2}S_{j-1} - S_{j-2} \\ &\dots \\ S_k &:= A_{k-2}S_{k-1} - S_{k-2} \\ 0 &= A_{k-1}S_k - S_{k-1}. \end{aligned}$$

Claim: The last term S_k of this sequence is a nonzero element of R .

Note that the polynomials S_0 and S_1 are relatively prime.

Claim: No two successive terms of the sequence vanish at any element of R .

Suppose that $S_j(c) = S_{j+1}(c) = 0$ for some c in R . Then, from the definition of the chain, we get $S_{j+2}(c) = \dots = S_k(c) = 0$ which is a contradiction. This completes the proof of the claim.

We now consider how the quantities of interest change as c moves from a to b . In particular, we want to see that the following quantity remains constant:

$$N(c) := n^+(c) - n^-(c) + v(c)$$

where

$$\begin{aligned} n^+(c) &:= \#\{r \in R : a < r < c \wedge P(r) = 0 \wedge Q(r) > 0\}, \\ n^-(c) &:= \#\{r \in R : a < r < c \wedge P(r) = 0 \wedge Q(r) < 0\}, \\ v(c) &:= V(\mathcal{S}; c). \end{aligned}$$

Note that the roots of the polynomials S_j divide the interval (a, b) into subintervals. We consider various cases.

Case: The element c is inside one of the subintervals.

By the interpolation property, in the interior of any of the subintervals, all of the S_j retain their signs. Hence $N(c)$ remains constant on such a subinterval.

Case: For some $j > 0$, $S_j(c) = 0$.

Consider the equation $S_{j+1} = A_{j-1}S_j - S_{j-1}$. From it we see that $S_{j+1}(c) = -S_{j-1}(c)$. Hence $S_{j+1}(c)$ and $S_{j-1}(c)$ have different signs. It follows that the polynomials S_{j+1} and S_{j-1} have different signs near c -

namely, on the two subintervals adjacent to c . Hence the number $N(c)$ does not change on the passage thru c .

Case: The polynomial P vanishes at c ; in symbols, $P(c) = 0$.

Then there exists a positive integer l and a polynomial B in $R[X]$ such that

$$P(X) = (X - c)^l B(X) \text{ and } B(c) \neq 0.$$

(Note that l is the multiplicity of c as a root of P .) We have

$$P'(X)Q(X) = l(X - c)^{l-1}B(X)Q(X) + (X - c)^l B'(X)Q(X).$$

Note, for x near c , we have

$$\text{Sign}(S_0(x)) = \text{Sign}((x - c)B(x))$$

and

$$\text{Sign}(S_1(x)) = \text{Sign}((P'Q/G)(x)) = \text{Sign}(B(x)Q(x)).$$

(In particular, note that $(X - c)^{l-1}$ is a factor of G since it divides $P(X)$ and $P'(X)Q(X)$.) There are several sub-cases.

Sub-case: The sign B is positive at c and the sign of Q is positive at c : $B(c) > 0 \wedge Q(c) > 0$.

Then B is positive near c , S_0 is strictly increasing near c and $B(x)Q(x)$ is positive near c . It follows that n^+ increases by one and v decreases by one when moving thru c from left to right.

Sub-case: The sign B is positive at c and the sign of Q is negative at c : $B(c) > 0 \wedge Q(c) < 0$.

It follows that n^- increases by one and v increases by one when moving thru c from left to right.

Sub-case: The sign B is negative at c and the sign of Q is positive at c : $B(c) < 0 \wedge Q(c) > 0$.

It follows that n^+ increases by one and v decreases by one when moving thru c from left to right.

Sub-case: The sign B is negative at c and the sign of Q is negative at c : $B(c) < 0 \wedge Q(c) < 0$.

It follows that n^- increases by one and v increases by one when moving thru c from left to right.

□