A Formalisation of Sturm's theorem in Isabelle/HOL

Manuel Eberl

September 17, 2013

1 Outline

This paper will give a detailed proof that closely follows the structure of the formal proof in Isabelle/HOL, but omits the proofs for basic lemmas about analysis and polynomials.

2 Notation

In this paper, $\mathbb{R}[X]$ will be used to denote the set of all polynomials with real coefficients. For a polynomial $P \in \mathbb{R}[X]$, we denote its derivative by P'. We say that some property holds in a neighbourhood of some $x_0 \in \mathbb{R}$ if there exists an $\varepsilon > 0$ such that the property holds for all $x \in (x_0 - \varepsilon; x_0 + \varepsilon) \setminus \{x_0\}$.

3 Proof of Sturm's theorem

3.1 Definitions

First, we define three notions that will be important in the proof of Sturm's theorem:
A *quasi-Sturm sequence* is a list of polynomials $P_0, P_1, \ldots, P_{n-1} \in \mathbb{R}[X]$ that fulfils the following properties:

- $n \geq 1$, i.e. the sequence is not empty
- P_n does not change its sign, i.e. $sgn(P_n(x)) = sgn(P_n(y))$ for all $x, y \in \mathbb{R}$
- for any $i \in \{0, \ldots, n-1\}$ and any root x_0 of P_{i+1}, $P_i(x_0)P_{i+2}(x_0) < 0$ holds

It can easily be seen that any nonempty sequence obtained by dropping a number of elements from the beginning of a quasi-Sturm sequence is again a quasi-Sturm sequence.

A *Sturm sequence* is a quasi-Sturm sequence that fulfils the following additional properties:

- $n \geq 2$, i.e. the sequence contains at least two polynomials
- for any root x_0 of P_0, $P_0(x)P_1(x)$ is negative in some sufficiently small interval $(x_0 - \varepsilon; x_0)$ and positive in some sufficiently small interval $(x_0; x_0 + \varepsilon)$[1]
- P_0 and P_1 have no common roots.

[1] In Isabelle/HOL, this is expressed as "eventually, the property $sgn(P_0P_1(x)) = \text{if } x > x_0 \text{ then } 1 \text{ else } -1$ holds at x_0".

In Isabelle, these three concepts are captured in *locales* of the same name.

Next, we define the notion of *sign changes*. Let $P_0, \ldots, P_{n-1} \in \mathbb{R}[X]$ be a sequence of polynomials and $x \in \mathbb{R}$. By evaluating the P_i at x, we obtain a sequence of real numbers $y_0, \ldots, y_{n-1}$. We now traverse this sequence from left to right and count how often the sign changes, i. e. how often we see a positive number and the last nonzero number in the sequence was negative (or vice versa). This is called the number of *sign changes* of the sequence $P_0, \ldots, P_{n-1}$ at the position x and is denoted as $\sigma(P_0, \ldots, P_{n-1}; x)$.
In Isabelle, this is realised by evaluating the P_i at x, deleting all zeros from the list, applying the *rem_adj_dups* operation on the list and taking the length of the remaining list minus one as the result. The *rem_adj_dups* function simply deletes equal adjacent elements in the list, e. g. it turns $[1, 2, 2, 1, 1, 1]$ into $[1, 2, 1]$.
A related notion are the sign changes "at infinity". For this, we compute the sequences

$$\left(sgn\left(\lim_{x\to-\infty} P_i(x)\right)\right)_{0\le i<n} \qquad \textit{and} \qquad \left(sgn\left(\lim_{x\to\infty} P_i(x)\right)\right)_{0\le i<n}$$

and count the sign changes in these sequences in the same way as before. The number of sign changes is called $\sigma_{-\infty}(P_0, \ldots, P_{n-1})$ and $\sigma_\infty(P_0, \ldots, P_{n-1})$, respectively. Note that the signs of these limits for a P_i can be determined by taking the sign of the leading coefficient of P_i.

3.2 Important auxiliary lemmas

Since Sturm's theorem is based on counting sign changes, a crucial part in proving it is to formalise the counting of sign changes in a way that is convenient for subsequent proofs. In this proof, this was done by proving that under certain conditions, the computation of the number of sign changes of a sequence can be "decomposed".

Lemma 1. *Let $P \in \mathbb{R}[X] \setminus \{0\}$. Then there exist $l, u \in \mathbb{R}$ such that $l < u$, all roots of P are in $(l; u)$ and*

$$\forall x \le l.\ sgn(P(x)) = sgn\left(\lim_{x\to-\infty} P(x)\right)$$
$$\forall x \ge u.\ sgn(P(x)) = sgn\left(\lim_{x\to\infty} P(x)\right)$$

Proof. Obviously, the limit of $P(x)$ at $-\infty$ (resp. ∞) is either ∞ or $-\infty$. Therefore, there exists an l (resp. a u) such that for all $x \le l$ (resp. $x \ge u$), $P(x)$ increases above every bound in the case of the limit ∞ or falls below every bound in case of the limit $-\infty$. In particular, if we choose 0 as the bound, we see that $P(x)$ will be positive for all values beyond some point in the case of ∞ or negative in the case of $-\infty$, which implies $sgn(P(x)) = sgn(\lim_{x\to-\infty} P(x))$ (resp. $sgn(P(x)) = sgn(\lim_{x\to\infty} P(x))$) for all x beyond that point. Therefore, bounds l and u that fulfil the last two requirements exist.
Also, since the $P(x)$ is positive or negative for all x below l (resp. above u), it has no roots outside the interval $(l; u)$. Lastly, should $l < u$ not hold already, we can simply increase u until it does. □

Lemma 2. *Let $P_0, \ldots, P_{n-1}$ be a sequence of polynomials and $i \in \{0, \ldots, n-1\}$. Furthermore, assume $P_i(x) \neq 0$. Then $\sigma(P_0, \ldots, P_{n-1}; x) = \sigma(P_0, \ldots, P_i; x) + \sigma(P_i, \ldots, P_{n-1}; x)$*

Proof. According to the informal definition of sign changes, this is obvious; for the Isabelle definition using list operation, it can be proven using simple properties of the *rem_adj_dups* and *filter* functions. □

Lemma 3. *Assume $P(x) \neq 0$ and $sgn(R(x)) = -sgn(P(x))$. Then $\sigma(P, R, Q; x) = 1$*

Proof. By case distinction □

Lemma 4. *Let $P_0, \ldots, P_{n-1} \in \mathbb{R}[X]$, $Q_0 \ldots, Q_{n-1} \in \mathbb{R}[X]$ be two sequences of polynomials. Assume $sgn(P_i(x_1)) = sgn(Q_i(x_2))$ for all i. Then $\sigma(P_0, \ldots, P_{n-1}; x_1) = \sigma(Q_0, \ldots, Q_{n-1}; x_2)$.*

Proof. Trivial. □

3.3 Sign changes of Sturm sequences

Lemma 5. *Let $P_0, \ldots, P_{n-1}$ be a quasi-Sturm sequence. Assume $P_0(x_0) \neq 0$. Then $\sigma(P_0, \ldots, P_n; x_0)$ is constant in the neighbourhood of x_0.*

Proof. By induction over the length of the sequence

- if the sequence has length 1, the number of sign changes is, of course, 0 at any position.
- if the sequence has length 2, we know that P_1 does not change its sign anywhere by definition of a quasi-Sturm chain. Furthermore, since $P_0(x_0) \neq 0$, the sign of P_0 does not change in the neighbourhood of x_0. By Lemma 4, the number of sign changes of the sequence P_0, P_1 also remains constant in the neighbourhood of x_0.
- if the sequence has length ≥ 3 and $P_1(x_0) \neq 0$, we have $\sigma(P_0, \ldots, P_{n-1}; x_0) = \sigma(P_0, P_1; x_0) + \sigma(P_1, \ldots, P_{n-1}; x_0)$ from Lemma 2. Since $P_0(x_0) \neq 0$ and $P_1(x_0) \neq 0$, the signs of P_0 and P_1 do not change in the neighbourhood of x_0 and therefore $\sigma(P_1, \ldots, P_{n-1}; x_0)$ is constant in the neighbourhood of x_0. Also, the sequence $P_1, \ldots, P_{n-1}$ is again a quasi-Sturm sequence with $P_1(x_0) \neq 0$, i. e. the induction hypothesis can be applied and we can conclude that $\sigma(P_1, \ldots, P_{n-1}; x_0)$ is also constant in the neighbourhood of x_0.
- if the sequence has length ≥ 3 and $P_1(x_0) = 0$, we have $P_0(x_0)P_2(x_0) < 0$ by the definition of a quasi-Sturm chain and thus $P_0(x_0) \neq 0$, $P_2(x_0) \neq 0$, and $sgn(P_2(x_0)) = -sgn(P_0(x_0))$. With Lemma 2, we have $\sigma(P_0, \ldots, P_{n-1}; x_0) = \sigma(P_0, P_1, P_2; x_0) + \sigma(P_2, \ldots, P_{n-1}; x_0)$. Furthermore, since $P_0(x_0) \neq 0$ and $P_2(x_0) \neq 0$, $sgn(P_2(x)) = -sgn(P_0(x))$ holds in the entire neighbourhood of x_0 as the signs of P_0 and P_2 do not change in the neighbourhood of x_0, and with Lemma 3, we then have $\sigma(P_0, P_1, P_2; x) = 2$ in the entire neighbourhood of x_0. As for the second summand, we know that $P_2, \ldots, P_{n-1}$ is again a quasi-Sturm sequence and $P_2(x_0) \neq 0$, so we can apply the induction hypothesis and obtain that $\sigma(P_2, \ldots, P_{n-1}; x_0)$, too, is constant in the neighbourhood of x_0.

□

Lemma 6. *Let $P_0, \ldots, P_{n-1}$ be a Sturm sequence. Assume $P_0(x_0) = 0$. Then for x in the neighbourhood of x_0, we have $\sigma(P_0, \ldots, P_{n-1}; x) = \sigma(P_0, \ldots, P_{n-1}; x_0) + 1$ if $x < x_0$ and $\sigma(P_0, \ldots, P_{n-1}; x) = \sigma(P_0, \ldots, P_{n-1}; x_0)$ if $x \geq x_0$.*

Proof. From the definition of a Sturm chain, we know that $P_0(x_0) = 0$ implies $P_1(x_0) \neq 0$ and that in the neighbourhood of x_0, we have $P_0(x)P_1(x) < 0$ if $x < x_0$ and $P_0(x)P_1(x) > 0$ if $x > x_0$. Therefore, in the neighbourhood of x_0, we have $P_1(x) \neq 0$ and, with Lemma 5,

$\sigma(P_1, \ldots, P_{n-1}; x) = \sigma(P_1, \ldots, P_{n-1}; x_0)$.
Fix some x in that neighbourhood of x_0. By Lemma 2, we then have $\sigma(P_0, \ldots, P_{n-1}; x) = \sigma(P_0, P_1; x) + \sigma(P_1, \ldots, P_{n-1}; x) = \sigma(P_0, P_1; x) + \sigma(P_1, \ldots, P_{n-1}; x_0)$. What remains to be shown is that $\sigma(P_0, P_1; x) = 1$ for $x < x_0$ and $\sigma(P_0, P_1; x) = 0$ for $x > x_0$. This is a simple consequence of $P_0(x)P_1(x) < 0$ for $x < 0$ and $P_0(x)P_1(x) > 0$ for $x > 0$. □

Lemma 7. *Let $P_0, \ldots, P_{n-1}$ be a Sturm sequence with $P_0 \neq 0$. Fix $a, b \in \mathbb{R}$ with $a \leq b$. Then $\sigma(P_0, \ldots, P_{n-1}; a) - \sigma(P_0, \ldots, P_{n-1}; b) = |\{x \in (a; b] \mid P_0(x) = 0\}|$*

Proof. By induction over $|\{x \in (a; b].\ P_0(x) = 0\}|$ for arbitrary a, b.

- if $|\{x \in (a; b].\ P_0(x) = 0\}| = 0$, we have $P_0(x) \neq 0$ for all $x \in (a; b]$. We now distinguish two cases:
  - if $P_0(a) \neq 0$, we have $P_0(x) \neq 0$ for all $x \in [a; b]$. Lemma 5 then implies that $\sigma(P_0, \ldots, P_{n-1})$ is constant in the neighbourhood of all $x \in [a; b]$ and must therefore be constant in the entire interval $[a; b]$.
  - if $P_0(a) = 0$, we know from Lemma 3.3 that $\sigma(P_0, \ldots, P_{n-1}; x) = \sigma(P_0, \ldots, P_{n-1}; a)$ for all $x \in (a; a + \varepsilon)$ for some $\varepsilon > 0$, i. e. $\sigma(P_0, \ldots, P_{n-1})$ is constant in the right neighbourhood of a. Furthermore, using $P_0(x) \neq 0$ for all $x \in (a; b]$ and Lemma 5, that $\sigma(P_0, \ldots, P_{n-1})$ is constant in the neighbourhood of every $x \in (a; b]$. Therefore, $\sigma(P_0, \ldots, P_{n-1})$ is constant on the entire interval $[a; b]$.

  Since $\sigma(P_0, \ldots, P_{n-1})$ is constant on the entire interval $[a; b]$, we have $\sigma(P_0, \ldots, P_{n-1}; a) - \sigma(P_0, \ldots, P_{n-1}; b) = 0$, which is exactly the number of roots of P_0 in $(a; b]$.
- if $|\{x \in (a; b].\ P_0(x) = 0\}| = n > 0$, we take the smallest root of P_0 in the interval $(a; b]$ and call it x_2. With Lemma 3.3, we know that $\sigma(P_0, \ldots, P_{n-1}; x) = \sigma(P_0, \ldots, P_{n-1}; x_2) + 1$ for all x in a sufficiently small left neighbourhood of x_2. Let x_1 then be some value in $(a; x_2)$ that is in this neighbourhood. We know that P_0 has no roots in the interval $(a; x_2]$ and thus, by induction hypothesis, $\sigma(P_0, \ldots, P_{n-1}; a) = \sigma(P_0, \ldots, P_{n-1}; x_1) = \sigma(P_0, \ldots, P_{n-1}; x_2) + 1$. Furthermore, the number of roots of P_0 in the interval $(x_2; b]$ is exactly $n-1$, so by induction hypothesis, $\sigma(P_0, \ldots, P_{n-1}; x_2) - \sigma(P_0, \ldots, P_{n-1}; b) = n - 1$. Adding these two equations yields $\sigma(P_0, \ldots, P_{n-1}; a) - \sigma(P_0, \ldots, P_{n-1}; b) = n$.

□

Lemma 8. *Let $P_0, \ldots, P_{n-1}$ be a Sturm sequence with $P_0 \neq 0$. Then $\sigma_{-\infty}(P_0, \ldots, P_{n-1}) - \sigma_{\infty}(P_0, \ldots, P_{n-1}) = |\{x \mid P_0(x) = 0\}|$*

Proof. Using Lemma 1 for every P_i, we obtain $l, u \in \mathbb{R}$ with $l < u$ such that P_0 has no roots outside the interval $(l; u)$ and for all P_i, $sgn(P_i(l)) = sgn(\lim_{x \to -\infty} P_i(x))$ and $sgn(P_i(u)) = sgn(\lim_{x \to \infty} P_i(x))$. It is then easy to see that the number of sign changes of the sequence $P_0, \ldots, P_{n-1}$ can be determined by considering the signs of the limits of the P_i at $\pm\infty$ instead of $P_i(u)$ and $P_i(l)$. Since all roots of P_0 lie in the interval $(l; u)$, we have thus shown that the total number of roots of P_0 can be determined in the way described above. □

Note: similar statements for the usage of $\sigma_{-\infty}$ and σ_{∞} to compute the number of roots $> a$ or $\leq a$ for some fixed $a \in \mathbb{R}$ can be proven analogously.

4 Constructing Sturm chains

4.1 The canonical Sturm chain

The canonical Sturm chain $P_0, \ldots, P_{n-1}$ of a polynomial $P \in \mathbb{R}[X]$ is constructed as follows:

$$\textit{where} \quad P_i = \begin{cases} P & \textit{for } i = 0 \\ P' & \textit{for } i = 1 \\ -P_{i-2} \textit{ mod } P_{i-1} & \textit{otherwise} \end{cases}$$

n is chosen such that the $n \geq 2$ and P_{n-1} is constant. The details of this are not important, since choosing a higher value for n only results in a number of zero polynomials at the end of the sequence, which do not change the result in any way. Note that this construction always terminates as the degree of the polynomials involved strictly decreases with every step, so that one reaches a constant polynomial in finitely many steps.

Lemma 9. *For all $i \in \{0, n-1\}$, we have* $\gcd(P_i, P_{i+1}) = \gcd(P, P')$

Proof. By induction on i using the fact that $\gcd(R, S) = \gcd(S, R \mod S) = \gcd(S, -R \mod S)$. □

Lemma 10. *Let $P \in \mathbb{R}[X]$ with no multiple roots. Then the canonical Sturm chain construction of P yields an actual Sturm chain.*

Proof. We simply prove the five conditions a Sturm chain has to satisfy:

- the sequence has at least length 2: obvious by definition
- P_{n-1} does not change its sign: obvious, since P_{n-1} is constant
- for any $i \in \{0, \ldots, n-1\}$ and any root x_0 of P_{i+1}, $P_i(x_0)P_{i+2}(x_0) < 0$ holds:
  By construction, we have $P_{i+2} = -P_i \textit{ mod } P_{i+1}$. With $P_{i+1}(x_0) = 0$, this implies that $P_{i+2}(x) = -P_i(x)$, and because of Lemma 9 and the fact that P has no multiple roots, we also have $P_i(x) \neq 0$ and thus $P_i(x)P_{i+2}(x) < 0$.
- for any root x_0 of P_0, $P_0(x)P_1(x)$ is negative in some sufficiently small interval $(x_0 - \varepsilon; x_0)$ and positive in some sufficiently small interval $(x_0; x_0 + \varepsilon)$:
  since $P_0 = P$ and $P_1 = P'$ are polynomials, there exists some neighbourhood of x_0 that does not contain any roots of either P_0 or P_1. For any $x < x_0$ in that neighbourhood, we can then apply the mean value theorem to obtain some $\xi \in [x; x_0]$ with $P(x) - P(x_0) = P'(\xi)(x - x_0)$, and since $P(x_0) = 0$ and $x - x_0 < 0$, $P(x) = P'(\xi)$. Furthermore, $P(x) \neq 0$ and $P'(\xi)$ has the same sign as $P'(x)$, since we chose a neighbourhood without roots; therefore we have $P(x)P'(x) < 0$. For any $x > x_0$, we can use the same argument to find that $P(x)P'(x) > 0$.
- P_0 and P_1 have no common roots:
  $P_0 = P$ and $P_1 = P'$, so if P_0 and P_1 had a common root, it would be a multiple root of P, which contradicts our assumption.

□

4.2 The case of multiple roots

Of course, one way to handle a polynomial with multiple roots is to "divide away" the "excess roots" by dividing the polynomial P by $\gcd(P, P')$. However, surprisingly, it turns out that as long as the interval bounds a and b are not multiple roots of P themselves, i.e. at least one of $P(a) \neq 0$ and $P'(a) \neq 0$ and at least one of $P(b) \neq 0$ and $P'(b) \neq 0$ holds, one can simply use the canonical Sturm chain without any modification. To prove this, we will first prove the following auxiliary result: