1. Sturm Sequences

(1.1) Definition. *Let K be an ordered field and let $c := (c_0, ..., c_n) \in K^{n+1}$. We define the* **variance** *of the tuple c to be*

$$\mathrm{var}(c) = \mathrm{card}\{i \in \{0, ..., n-1\} \mid \exists j > i : c_i \cdot c_j < 0 \text{ and } c_k = 0 \ (i < k < j)\}.$$

Hence $\mathrm{var}(c)$ *is the number of sign changes in $(c_0, ..., c_n)$ after crossing out all c_i which are zero.*

Observe that

$$(+) \qquad \mathrm{var}\, c = \mathrm{var}(c_0, ..., c_k) + \mathrm{var}(c_k, ..., c_n) \text{ whenever } k \in \{1, ..., n-1\} \text{ and } c_k \neq 0.$$

(1.2) Definition. *Let K be a field and let $f(X) \in K[X]$ be a polynomial. The Sturm sequence $\mathfrak{f}$ of f is the following tuple $\mathfrak{f} := (f_0, f_1, ..., f_d)$ of polynomials $f_i \in K[X]$:*
$f_0 := f$, $f_1 := f'$ and for each $i > 1$ let f_{i+1} be the negative of the remainder if we divide f_{i-1} by f_i. Hence

$$\begin{aligned}
f_0 &= f \\
f_1 &= f' \\
f_0 &= q_1 \cdot f_1 - f_2 \text{ with } q_1 \in K[X], \ \deg f_2 < \deg f_1 \\
&\ \ . \\
&\ \ . \\
&\ \ . \\
f_{i-1} &= q_i \cdot f_i - f_{i+1} \text{ with } q_i \in K[X], \ \deg f_{i+1} < \deg f_i \\
&\ \ . \\
&\ \ . \\
&\ \ . \\
f_{d-1} &= q_d \cdot f_d \text{ with } q_i \in K[X]
\end{aligned}$$

By induction we see that the natural number d as well as the polynomials $f_0, ..., f_d$ are well defined. The construction of $\mathfrak{f}$ differs from the euclidean algorithm applied to f and f' only in the choice of the sign of the remainders. The proof that the euclidean algorithm applied for f and f' computes the greatest common divisor of f and f' can be literally copied in order to see

$$f_d = \gcd(f, f').$$

□

(1.3) Theorem. *(Sturm, 1829)*
Let R be real closed, let $f(X) \in R[X]$ with $f \neq 0$ and let $(f_0, ..., f_d)$ be the Sturm sequence of f. If $a < b$ are elements from R such that $f(a), f(b) \neq 0$ then the number of different roots (so we don't count multiplicities) of f in (a, b) is

$$\mathrm{var}(f_0(a), ..., f_d(a)) - \mathrm{var}(f_0(b), ..., f_d(b))$$

Proof. For $i \in \{0, ..., d\}$ let $h_i := \frac{f_i}{f_d} \in K[X]$. Observe that by the definition of the

Sturm sequence $f_0, ..., f_d$ we have

$$f_{i-1} = q_i \cdot f_i - f_{i+1} \text{ with } q_i \in K[X],\ \deg f_{i+1} < \deg f_i$$

and therefore

$$(*) \qquad h_{i-1} = q_i \cdot h_i - h_{i+1},\ \deg h_{i+1} < \deg h_i\ (1 \le i < d).$$

Moreover for each $i \in \{1, ..., d\}$,

$$(\dagger) \qquad h_{i-1} \text{ and } h_i \text{ do not have common zeroes in } R,$$

otherwise $(*)$ implies that h_d has a zero; but $h_d = 1$. For $x \in R$ let

$$W(x) := \operatorname{var}(h_0(x), ..., h_d(x)).$$

<u>Claim 1.</u> If $c \in R$ with $f(c) \neq 0$, then $W(c) = \operatorname{var}(f_0(c), ..., f_d(c))$.
This is so, since $f(c) \neq 0$ implies $f_d(c) \neq 0$ and therefore

$$\operatorname{var}(f_0(c), ..., f_d(c)) = \operatorname{var}(f_d(c)h_0(c), ..., f_d(c)h_d(c)) = W(x).$$

<u>Claim 2.</u> h_0 and f have the same zero set in R.
To see claim 2 it is enough to prove $h_0(c) = 0$ for each zero c of f. Let $k > 0$ and $g(X) \in R[X]$, $g(X) \neq 0$ with $f(X) = (X - c)^k \cdot g(X)$. Since $k > 0$ we have

$$f'(X) = (X - c)^{k-1} \cdot (kg(X) + (X - c)g'(X)).$$

As $g(c) \neq 0$ the multiplicity of $X - c$ is $k - 1$ in f'. Since $f_d = \gcd(f, f')$ this shows that $X - c$ divides $h_0 = f/f_d$, in other words $h_0(c) = 0$.

Since $f(a), f(b) \neq 0$, claim 1 and claim 2 reduce the problem to show that the number of different zeroes of h_0 in (a, b) is equal to $W(a) - W(b)$. Let

$$h := h_0 \cdot ... \cdot h_d.$$

<u>Claim 3.</u> If $c < d$ are elements from R and h does not vanish in the interval $[c, d]$, then $W(X)$ is constant on $[c, d]$.
Claim 3 holds by the intermediate value property for polynomials.

<u>Claim 4.</u> If $i \in \{1, ..., d - 1\}$ and $c \in R$ is a zero of h_i, then there is some $\varepsilon > 0$ such that $\operatorname{var}(h_{i-1}(x), h_i(x), h_{i+1}(x)) = 1$ for $x \in (c - \varepsilon, c + \varepsilon)$.
As $h_i(c) = 0$, we get $h_{i-1}(c) = -h_{i+1}(c)$ from $(*)$. Since not both, h_i and h_{i+1} are zero in c, it follows $h_{i-1}(c) = -h_{i+1}(c) \neq 0$ and we may choose ε so that $\operatorname{sign} h_{i-1}(x) = -\operatorname{sign} h_i(x) \neq 0$ for all $x \in (c - \varepsilon, c + \varepsilon)$. Then, no matter what $h_i(x)$ is in $(c - \varepsilon, c + \varepsilon)$, we always have $\operatorname{var}(h_{i-1}(x), h_i(x), h_{i+1}(x)) = 1$ for $x \in (c - \varepsilon, c + \varepsilon)$.

For $i \in \{0, ..., d - 1\}$ let

$$W_i(x) := \operatorname{var}(h_i(x), ..., h_d(x)).$$

<u>Claim 5.</u> If $c \in R$ and $i \in \{0, ..., d - 1\}$ with $h_i(c) \neq 0$, then there is some $\varepsilon > 0$ such that $W_i(X)$ is constant on $(c - \varepsilon, c + \varepsilon)$.
Let $j_1 < ... < j_l$ be an enumeration of those indices $j \in \{i, ..., d\}$ such that $h_j(c) \neq 0$. Take ε so that
(a) $\operatorname{sign} h_{j_\alpha}(x) = \operatorname{sign} h_{j_\alpha}(c) \neq 0$ for all $x \in (c - \varepsilon, c + \varepsilon)$ and all $\alpha \in \{1, ..., l\}$.

By claim 4 we may shrink ε such that
(b) for each $j \in \{i+1, ..., d-1\}$ with $h_j(c) = 0$, $\operatorname{var}(h_{j-1}(x), h_j(x), h_{j+1}(x)) = 1$ for $x \in (c-\varepsilon, c+\varepsilon)$.

As $h_i(c) \neq 0$ by assumption and $h_d(c) \neq 0$ we have $j_1 = i$ and $j_l = d$. Thus by (+) and (a),

$$W_i(X) = w_1(x) + ... + w_{l-1}(x), \text{ with } w_\alpha(x) := \operatorname{var}(h_{j_\alpha}(x), ..., h_{j_{\alpha+1}}(x)) \ (x \in (c-\varepsilon, c+\varepsilon)).$$

By (†), $j_{\alpha+1} \leq j_\alpha + 2$. Hence, either $w_\alpha(x) = \operatorname{var}(h_{j_\alpha}(x), h_{j_\alpha+1}(x))$ $(x \in (c-\varepsilon, c+\varepsilon), \alpha \in \{1, ..., l-1\})$, which is constant on $(c-\varepsilon, c+\varepsilon)$ by (a), or,

$$w_\alpha(x) = \operatorname{var}(h_{j_\alpha}(x), h_{j_\alpha+1}(x), h_{j_\alpha+2}(x)) \ (x \in (c-\varepsilon, c+\varepsilon), \alpha \in \{1, ..., l-1\}),$$

which is constant on $(c-\varepsilon, c+\varepsilon)$ by (b).

Since $W_i(X) = w_1(x) + ... + w_{l-1}(x)$, with $w_\alpha(x)$ for $x \in (c-\varepsilon, c+\varepsilon)$, this shows claim 5.

<u>Claim 6.</u> If $c \in R$ is a zero of h_0, then there is some $\varepsilon > 0$ such that

$$W(x) = W(y) + 1 \text{ for all } x, y \text{ with } c - \varepsilon < x < c < y < c + \varepsilon.$$

Since $h_0(c) = 0$ we have $h_1(c) \neq 0$ by (†). Choose $\varepsilon > 0$ such that
(i) $W_1(X)$ is constant on $(c-\varepsilon, c+\varepsilon)$ (this is possible by claim 5),
(ii) $\operatorname{sign} h_1(x) = \operatorname{sign} h_1(c) \neq 0$ $(x \in (c-\varepsilon, c+\varepsilon))$ (this is possible, since $h_1(c) \neq 0$).
(iii) c is the unique zero of f in $(c-\varepsilon, c+\varepsilon)$. DAS BRAUCHT MAN VIELLEICHT NICHT

Let $k > 0$ and $g(X) \in R[X]$, $g(X) \neq 0$ with $f(X) = (X-c)^k \cdot g(X)$. Since $k > 0$ we have

$$f'(X) = (X-c)^{k-1} \cdot (kg(X) + (X-c)g'(X)).$$

For $x \in (c, c+\varepsilon)$ we have $\operatorname{sign} f(x) = \operatorname{sign} g(x)$ and $\operatorname{sign} f'(x) = \operatorname{sign}(kg(x) + (X-c)g'(x))$. By shrinking ε if necessary and since $g(c) \neq 0$ we see that $\operatorname{sign} f'(x) = \operatorname{sign} g(x)$ $(x \in (c, c+\varepsilon))$. It follows that $\operatorname{sign} h_0(x) = \operatorname{sign} h_1(x) \neq 0$ for all $x \in (c, c+\varepsilon)$, in other words

$$(**) \qquad \operatorname{var}(h_0(x), h_1(x)) = 0 \text{ for all } x \in (c, c+\varepsilon).$$

As $g(c) \neq 0$ the multiplicity of $X - c$ is $k-1$ in f'. Since $f_d = \gcd(f, f')$ and $h_0 = \frac{f}{f_d}$ the multiplicity of $X - c$ in h_0 is 1. Hence h_0 changes sign in c. By (∗∗) and (ii) we get

$$(***) \qquad \operatorname{var}(h_0(x), h_1(x)) = 1 \text{ for all } x \in (c-\varepsilon, c).$$

Now for $y \in (c, c+\varepsilon)$ we have $W(y) \overset{(+),(ii)}{=} \operatorname{var}(h_0(y), h_1(y)) + W_1(y) \overset{(**)}{=} W_1(y)$. Whereas for $x \in (c-\varepsilon, c)$ we have $W(x) \overset{(+),(ii)}{=} \operatorname{var}(h_0(x), h_1(x)) + W_1(x) \overset{(**)}{=} W_1(x) + 1$. Since $W_1(x)$ is constant on $(c-\varepsilon, c+\varepsilon)$ by (i), this shows claim 6.

Now we prove the Theorem. Let $c_1 < ... < c_m$ be the enumeration of the zeroes of h in $[a, b]$. Choose $\varepsilon > 0$ such that for each $j \in \{1, ..., m\}$ the following conditions are satisfied:
(1) If $h_0(c_j) = 0$, then $W(x) = W(y) + 1$ for all x, y with $c - \varepsilon < x < c < y < c + \varepsilon$. This is possible by claim 6.
(2) If $h_0(c_j) \neq 0$, then $W(X)$ is constant on $(c_j - \varepsilon, c_j + \varepsilon)$. This is possible by claim 5 applied to $i = 0$.
(3) $c_j + \varepsilon < c_{j+1} - \varepsilon$ $(j \in \{1, ..., m-1\})$.

Choose $d_j \in (c_j - \varepsilon, c_j)$ and $e_j \in (c_j, c_j + \varepsilon)$ $(1 \leq j \leq m)$, in particular

$$d_1 < c_1 < e_1 < d_2 < c_2 < e_2 < \ldots < d_m < c_m < e_m.$$

By enlarging d_1 and shrinking e_m if necessary, we may assume that all zeros of h in $[d_1, e_m]$ are among the $c_1, \ldots, c_m$ (note that a or b might be zeros of h).

By the choice of ε in (1), $W(x)$ decreases in the interval $[d_j, e_j]$ by 1 if and only c_j is a zero of h_0, whereas in in all other intervals $[d_j, e_j]$, $W(x)$ is constant by the choice of ε in (2). Finally $W(x)$ is constant in every interval $[e_i, d_{i+1}]$ $(1 \leq i < m)$ by claim 3.

Thus $W(d_1) - W(e_m)$ is the number of zeroes of h_0 in (d_1, e_m).

Since $f(a) \neq 0$, also $h_0(a) \neq 0$ and by our choice of d_1, h_0 does not have zeroes in the closed interval between a and d_1. Thus $W(d_1) = W(a)$. Similarly $W(e_m) = W(b)$. Hence $W(a) - W(b)$ is the number of zeroes of h_0 in (d_1, e_m), which is the number of zeroes of h_0 in (a, b). $\square$