GEOMETRY OF LINEAR MATRIX INEQUALITIES

Lecture Notes
Universität Konstanz, Summer 2013

Daniel Plaumann

Contents

Daniel Plaumann
Universität Konstanz
Summer 2013

1. LINEAR MATRIX INEQUALITIES AND SPECTRAHEDRA

1.1. INTRODUCTION

Given real polynomials $g_1, \ldots, g_r \in \mathbb{R}[x]$ in n variables $x = (x_1, \ldots, x_n)$, let

$$S = \mathcal{S}(g_1, \ldots, g_r) = \left\{u \in \mathbb{R}^n \mid g_1(u) \geqslant 0, \ldots, g_r(u) \geqslant 0\right\}$$

be the **basic closed (semialgebraic) set** in $\mathbb{R}^n$ defined by $g_1, \ldots, g_r$. The convex hull

$$\operatorname{conv}(S) = \left\{\sum_{i=0}^{n} \lambda_i u_i \mid 0 \leqslant \lambda_i \leqslant 1, \sum_{i=0}^{n} \lambda_i = 1, u_i \in B\right\}$$

of S is again semialgebraic and can be thought of as a linearisation of S. How can we describe it in terms of inequalities? What is its boundary? Can we find a description that reflects the convexity and is well suited for computations? These are some of the questions that we will attempt to answer through the study of linear matrix inequalities.

Even though the motivation comes from computational problems and in particular from optimisation, our focus will be on geometry and theoretical foundations. The basic building blocks will be the affine-linear slices of the cone of positive semidefinite matrices. These convex sets are called spectrahedra, and our goal will be the realisation of other convex sets as spectrahedra or as projections of spectrahedra. To get started, we will first need to study the spectrahedra themselves in some detail.

GENERAL REFERENCES

[Barvinok] A. Barvinok. *A course in convexity*, vol. 54 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

[CAG-Book] G. Blekherman, P. Parrilo, and R. Thomas (editors). *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization. 2012. To appear. http://www.math.washington.edu/~thomas/frg/frgbook/SIAMBookFinalvNov12-2012.pdf

[Forst-Hoffmann] W. Forst and D. Hoffmann. *Optimization—theory and practice*. Springer Undergraduate Texts in Mathematics and Technology. Springer, New York, 2010.

[Marshall] M. Marshall. *Positive polynomials and sums of squares*, vol. 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008.

[Convexity-LN] D. Plaumann. *Vorlesung über Konvexität*. Lecture notes, Universität Konstanz, 2011. http://www.math.uni-konstanz.de/~plaumann/KonvexWS11/konvex.pdf

[Prestel-Delzell] A. Prestel and C. N. Delzell. *Positive polynomials*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2001.

[Rockafellar] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1970.

1.2. SPECTRAHEDRA

A **polyhedron** is a subset of $\mathbb{R}^n$ described by a finite number of linear inequalities. The compact polyhedra are exactly the **polytopes**, i.e. the convex hulls of finitely many points. These are the simplest (and the most extensively studied) convex sets.

Spectrahedra, our basic objects, are generalisations of polyhedra that comprise many more sets but still share some of the good properties of polyhedra. First, we need to recall a bit of linear algebra. We use the notations

$$\begin{aligned}\mathrm{Mat}_{k\times l}(R) &= \text{space of all matrices of size } k\times l \text{ with entries in } R,\\ \mathrm{Mat}_k(R) &= \text{space of all square matrices of size } k \text{ with entries in } R,\\ \mathrm{GL}_k(R) &= \text{group of invertible square matrices of size } k \text{ with entries in } R,\\ \mathrm{Sym}_k(R) &= \text{space of all symmetric matrices of size } k \text{ with entries in } R,\end{aligned}$$

where R is any field (or ring). The dimension of $\mathrm{Sym}_k(R)$ over R (or rank as a free R-module) is $\frac{1}{2}(k+1)k$. Throughout, matrices will always be assumed real, unless specified otherwise. So we write Mat_k, Sym_k, etc. to denote $\mathrm{Mat}_k(\mathbb{R})$, $\mathrm{Sym}_k(\mathbb{R})$, etc.

Recall that a real symmetric matrix $A \in \mathrm{Sym}_k$ is called **positive semidefinite** if

$$v^T A v \geqslant 0$$

for all $v \in \mathbb{R}^k$. It is called **positive definite** if the inequality is strict for all $v \neq 0$. We write

$$\begin{aligned}\mathrm{Sym}_k^+ &= \text{cone of positive semidefinite matrices of size } k\\ \mathrm{Sym}_k^{++} &= \text{cone of positive definite matrices of size } k.\end{aligned}$$

Note that Sym_k^+ is the closure of Sym_k^{++} and Sym_k^{++} is the interior of Sym_k^+.

A matrix $A \in \mathrm{Mat}_k$ defines both the linear map $\mathbb{R}^k \to \mathbb{R}^k$, $v \mapsto Av$, and the bilinear form $\mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$, $(v,w) \mapsto v^T A w$. But after a change of basis, given by $U \in \mathrm{GL}_k$, the linear map in the new coordinates is given by $U^{-1}AU$, while the bilinear form is given by $U^T A U$. The most basic fact about real symmetric matrices is that **all eigenvalues are real**. This is easy to prove: If $\lambda \in \mathbb{C}$ is a complex eigenvalue of a real symmetric matrix A and $v \in \mathbb{C}^k$ is a non-zero eigenvector, then

$$\lambda\|v\|^2 = \lambda v^T\overline{v} = (Av)^T\overline{v} = v^T A^T \overline{v} = v^T\overline{Av} = \overline{\lambda} v^T \overline{v} = \overline{\lambda}\|v\|^2$$

and since $\|v\| \neq 0$ for $v \neq 0$, we conclude $\lambda = \overline{\lambda}$.

Thus if a symmetric matrix A is viewed as a linear map and $U^{-1}AU$ is diagonal, the diagonal entries are the eigenvalues, all of which are real. On the other hand, if A is viewed as a bilinear form, we can make $U^T A U$ diagonal and normalise the entries to be ± 1 or 0. The difference between the number of positive and negative signs is the **signature**. Rank and signature are the only invariants, by Sylvester's theorem. In particular, we see that a symmetric matrix is positive semidefinite if and only if its signature equals its rank.

From this it is not a priori clear that the signature is in any way related to the eigenvalues. But the **principal axes theorem** says that any symmetric matrix possesses an orthonormal basis of eigenvectors. This means that, given $A \in \mathrm{Sym}_k$, there exists $U \in \mathrm{GL}_k$ with $U^T = U^{-1}$ such that $U^T A U$ is diagonal. In this case, the diagonal entries of $U^T A U = U^{-1}AU$ are the eigenvalues of A and the signs of the eigenvalues also determine the signature.

We sum up these well-known facts in

Theorem 1.1. *A real symmetric matrix has only real eigenvalues, and it is positive definite (resp. semidefinite) if and only if all its eigenvalues are positive (resp. positive or zero).*

Exercise 1.1. Let A be a real symmetric matrix of size k and rank r.

(a) Show that the following are equivalent:
   (1) The matrix A is positive semidefinite.
   (2) There exists a $k \times r$-matrix B with $A = BB^T$.
   (3) There exists a positive semidefinite $k \times k$-matrix P with $A = P^2$.

(b) If $BB^T = CC^T$ with B and C of the same size $k \times r$, there exists an orthogonal $r \times r$-matrix U such that $B = CU$.

(c) For $A \in \mathrm{Sym}_k^+$, the matrix P in (a) is uniquely determined (and commonly denoted $\sqrt{A}$).

Exercise 1.2. Let $A \in \mathrm{Sym}_k^+$ and $v \in \mathbb{R}^n$. Show that $v^T A v = 0$ implies $Av = 0$.

For a symmetric matrix A, we will write $A \succcurlyeq 0$ if it is positive semidefinite and $A \succ 0$ if it is positive definite. We extend this to a partial order on Sym_k by writing $A \succcurlyeq B$ if $A - B \succcurlyeq 0$.

Now we are ready to define spectrahedra.

Definition 1.2. A **spectrahedron** in $\mathbb{R}^n$ is the inverse image of Sym_k^+ under an affine-linear map $\mathbb{R}^n \to \mathrm{Sym}_k$, for some k.

Spectrahedra are clearly convex, since the inverse image of any convex set under an affine linear map is again convex. Occasionally, we may find it convenient not to fix coordinates and consider a spectrahedron in a finite dimensional real vector space V given as the inverse image of $\mathrm{Sym}(W)^+$ under an affine linear map $\Phi\colon V \to \mathrm{Sym}(W)$, for some finite-dimensional real vector space W.

When working with matrices, we can write things out explicitly: An affine-linear map $\Phi\colon \mathbb{R}^n \to \mathrm{Sym}_k$ is given by an $n+1$-tuple of symmetric matrices $A_0, A_1, \ldots, A_n \in \mathrm{Sym}_k$ via $\Phi(u) = A_0 + \sum_{i=1}^n u_i A_i$ and the corresponding spectrahedron is the set

$$\Phi^{-1}(\mathrm{Sym}_k^+) = \left\{u \in \mathbb{R}^n \mid (A_0 + u_1 A_1 + \cdots + u_n A_n) \succcurlyeq 0\right\} \subset \mathbb{R}^n.$$

We can view the expression $A_0 + u_1 A_1 + \cdots + u_n A_n$ as a polynomial of degree 1 with matrix coefficients, evaluated in the point u. Alternatively, we can think of it as a matrix with polynomial entries

$$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n = \begin{bmatrix} \ell_{11}(x) & \cdots & \ell_{1k}(x) \\ \vdots & \ddots & \vdots \\ \ell_{1k}(x) & \cdots & \ell_{kk}(x) \end{bmatrix}$$

with $\ell_{11}, \ldots, \ell_{kk} \in \mathbb{R}[x]$ of degree at most 1, and call this a **linear matrix polynomial**. (If not specified otherwise, linear matrix polynomials will always be real and symmetric.) Thus $A(x)$ is an element of $\mathrm{Sym}_k(\mathbb{R}[x])$ (which is isomorphic as an $\mathbb{R}$-Algebra to the polynomial ring $\mathrm{Sym}_k[x]$). If $A(x)$ is any linear matrix polynomial in n variables, we write

$$\mathcal{S}(A) = \left\{u \in \mathbb{R}^n \mid A(u) \succcurlyeq 0\right\}$$

for the spectrahedron defined by A, just as for ordinary polynomials. Spectrahedra are therefore the sets of solutions to **linear matrix inequalities**.

Examples 1.3.

(1) Any polyhedron is also a spectrahedron. For if $P = \mathcal{S}(\ell_1,\dots,\ell_k)$ is a polyhedron defined by polynomials $\ell_1,\dots,\ell_k \in \mathbb{R}[x]$ of degree 1, then $P = \mathcal{S}(A)$ for the diagonal matrix polynomial $A = \mathrm{Diag}(\ell_1,\dots,\ell_k)$. While any real symmetric matrix can be diagonalised, the same is not true for matrix polynomials, because finitely many matrices need not be *simultaneously* diagonalisable (see also Exercise 1.5). Therefore, not every spectrahedron is a polyhedron.

(2) For example, the linear matrix polynomial

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + x \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} + y \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1-x & y \\ y & 1+x \end{bmatrix}$$

in two variables x and y defines the closed unit disc in $\mathbb{R}^2$.

(3) More generally, the closed unit ball in $\mathbb{R}^n$ is a spectrahedron, defined by the linear matrix polynomial

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & x_1 \\ 0 & 1 & \cdots & 0 & x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_n \\ x_1 & x_2 & \cdots & x_n & 1 \end{bmatrix}.$$

(4) For a more interesting example, let $\mathbb{R}[x]_d$ be the vector space of polynomials of degree at most d in $x = (x_1,\dots,x_n)$, and let $m = (f_1,\dots,f_k)^T$ be a k-tuple of polynomials spanning $\mathbb{R}[x]_d$ (e.g. the monomial basis, with $k = \binom{n+d}{n}$). The map

$$\Psi_m \colon \begin{cases} \mathrm{Sym}_k & \to & \mathbb{R}[x]_{2d} \\ A & \mapsto & m^T A m \end{cases}.$$

is linear and surjective. Given $f \in \mathbb{R}[x]_{2d}$, any element of the fibre $\Psi_m^{-1}(f)$ is called a **Gram matrix** of f and $G_m(f) = \Psi_m^{-1}(f) \cap \mathrm{Sym}_k^+$ the **Gram spectrahedron** of f (with respect to m). The Gram spectrahedron is non-empty if and only if f is a sum of squares of elements in V_d. For given $A \in G_m(f)$, we can write $A = B^T B$ with B of size $\mathrm{rk}(A) \times k$ (Exercise 1.1). Then $f = m^T A m = (Bm)^T Bm$ is a sum of $\mathrm{rk}(A)$ squares. Conversely, if $f = \sum_{i=1}^r g_i^2$, we can write $(g_1,\dots,g_r)^T = Bm$ for some $B \in \mathrm{Mat}_{r\times k}$. Then $f = (Bm)^T Bm = m^T(B^T B)m$, so that $B^T B \in G_m(f)$ and $\mathrm{rk}(B^T B) = r$. In particular, the shortest sums-of-squares-representations of f correspond to the Gram matrices of minimal rank.

Moreover, $G_m(f)$ classifies the representations of f as a sum of squares up to orthogonal equivalence: For if $A \in G_m(f)$ is split as $A = B^T B = C^T C$ with B and C both of size $r \times k$, where $r = \mathrm{rk}(A)$, then $B = UC$ for some orthogonal matrix U of size $r \times r$ (Exercise 1.1). Conversely, given such U and a representation $f = \sum_{i=1}^r g_i^2$, then $(h_1,\dots,h_r)^T = U(g_1,\dots,g_r)^T$ gives another representation $f = \sum_{i=1}^r h_i^2$ belonging to the same Gram matrix.

Exercise 1.3. Verify examples (2) and (3) above.

Exercise 1.4. It clearly makes sense to generalise spectrahedra from the real symmetric to the complex hermitian case. Let $A(x)$ be a complex hermitian linear matrix polynomial of size k. Show that there exists a real symmetric linear matrix polynomial $B(x)$ of size $2k$ such that $\mathcal{S}(A) = \mathcal{S}(B)$.

Exercise 1.5.
(a) True or false? The spectrahedron $\mathcal{S}(A)$ is a polyhedron if and only if the matrices $A_0, \dots, A_n$ are pairwise commuting.
(b) In general, if $A = A_0 + x_1 A_1 + \cdots + x_n A_n$ is a linear matrix polynomial of size k in which $A_0, \dots, A_n$ are simultaneously equivalent to diagonal matrices (meaning there exists $U \in \mathrm{GL}_k$ such that all $U^T A_i U$ are diagonal), then $\mathcal{S}(A)$ is a polyhedron. To what extent does the converse hold?
(c) For more on this topic, see [**BRS11**].

Exercise 1.6. Let $A(x) = A_0 + \sum_{i=1}^n x_i A_i$ be a linear matrix polynomial and put $A'(x) = \sum_{i=1}^n x_i A_i$. Show that $\mathcal{S}(A)$ is a cone if and only if $\mathcal{S}(A) = \mathcal{S}(A')$.

While the definition of spectrahedron is simple enough, there are two different ways to think about it, corresponding to somewhat different geometric pictures.

- We may think of a spectrahedron as a subset of $\mathbb{R}^n$ defined by a linear matrix inequality. This we can rewrite as an infinite system of ordinary linear inequalities: If $A(x)$ is a linear matrix polynomial, then $\ell_v(x) = v^T A(x) v$ is a polynomial of degree at most 1 in the variables x, for any $v \in \mathbb{R}^k$, and
$$\mathcal{S}(A) = \left\{u \in \mathbb{R}^n \mid \ell_v(u) \geqslant 0 \text{ for all } v \in \mathbb{R}^k\right\}.$$
We may think of $\mathbb{R}^k$ as a parameter space for the linear inequalities describing the convex set $\mathcal{S}(A)$. Note that *any* closed convex set is described by an infinite family of linear inequalities: Given a closed convex subset C of $\mathbb{R}^n$, let $\mathcal{L} = \left\{\ell \in \mathbb{R}[x] \mid \deg(\ell) \leqslant 1, \ell|_C \geqslant 0\right\}$, then $C = \{u \in \mathbb{R}^n \mid \forall \ell \in \mathcal{L}\colon\ \ell(u) \geqslant 0\}$, by the separation theorem for closed convex sets. What makes spectrahedra special is the simple parametrisation of $\mathcal{L}$ in terms of a linear matrix polynomial.
- We may also think of a spectrahedron as a set of matrices. If $A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$ is a linear matrix polynomial and $\Phi_A\colon \mathbb{R}^n \to \mathcal{S}_k$ the affine-linear map $u \mapsto A_0 + u_1 A_1 + \cdots + u_n A_n$, consider the set of matrices $\mathrm{im}(\Phi_A) \cap \mathrm{Sym}_k^+$. If Φ_A is injective, we can identify this with $\mathcal{S}(A)$. The geometric picture here is that of a cone (namely Sym_k^+) sliced by a linear subspace. The subspace we are slicing with is the span of $A_1, \dots, A_n$ shifted by A_0. For instance, this is the way we would think about the Gram spectrahedra in Example 1.3(4).

  If Φ_A is not injective, then $A(x)$ is basically a cylinder over $\mathrm{im}(\Phi_A) \cap \mathrm{Sym}_k^+$, given in terms of the kernel of Φ_A (see Exercise 1.7 below), so we do not loose too much by passing to the image.

  It is instructive to make the analogy with polyhedra again: Let $P = \left\{u \in \mathbb{R}^n \mid \ell_1(u) \geqslant 0, \dots, \ell_k(u) \geqslant 0\right\}$ and consider the affine-linear map $\Phi\colon \mathbb{R}^n \to \mathbb{R}^k$, $u \mapsto (\ell_1(u), \dots, \ell_k(u))$. Then P is the inverse image of the **positive orthant** $(\mathbb{R}^k)_+$ under Φ. Thus, again up to injectivity of Φ, any polyhedron is a slice of a standard cone $(\mathbb{R}^k)_+$ for some k, just as in the case of spectrahedra.

Exercise 1.7. Let $A = A_0 + \sum_{i=1}^n x_i A_i$ be a linear matrix polynomial. Let $B_1, \dots, B_m$ be a basis of $\mathrm{span}(A_1, \dots, A_n)$ and put $B = A_0 + \sum_{i=1}^m x_i B_i$. Show that there exists a bijective linear transformation $\psi\colon \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^{n-m}$ such that $\psi(\mathcal{S}(A)) = \mathcal{S}(B) \times \mathbb{R}^{n-m}$ and $\Phi_A = (\Phi_B \times 0) \circ \Psi$.

1.3. FIRST PROPERTIES OF SPECTRAHEDRA

Proposition 1.4. *Spectrahedra are convex and basic closed.*

Proof. Let $S = \mathcal{S}(A)$ for some linear matrix polynomial A. Convexity is clear from the definition. We must show that there exist polynomials $g_1, \ldots, g_r \in \mathbb{R}[x]$ such that $S = \mathcal{S}(g_1, \ldots, g_r)$. To say that $A(u)$ is positive semidefinite for $u \in \mathbb{R}^n$ is saying that all its eigenvalues are positive or zero. The eigenvalues are the roots of the characteristic polynomial $\chi_{A(u)}(t) = \det(tI_k - A(u))$. Thus u is a point in S if and only if $\chi_{A(u)}(-t)$ has no positive roots. By the subsequent lemma, this is the case if and only if all coefficients of $(-1)^k \chi_{A(u)}(-t)$ are greater than or equal to zero. Thus we can take $g_1, \ldots, g_r$ to be the coefficients of $(-1)^k \chi_{A(x)}(-t)$ as a polynomial in t. □

Lemma 1.5. *Let $f \in \mathbb{R}[t]$ be a monic polynomial in one variable and assume that all roots of f are real. Then f has no positive roots if and only if all its coefficients are non-negative.*

Proof. Clearly, if all coefficients of f are greater than or equal to zero, it cannot have positive roots. Conversely, if the roots of f are $-\alpha_1, \ldots, -\alpha_k$ with $\alpha_i \geqslant 0$, then all coefficients of $f = (x + \alpha_1)\cdots(x + \alpha_k)$ are positive or zero. □

While the positive definite matrices are the interior of the cone of positive semidefinite matrices, it may still happen that a linear matrix polynomial A is nowhere positive definite even if $\mathcal{S}(A)$ has non-empty interior. For a trivial example, we could always artificially enlarge A by adding zeros, since

$$\mathcal{S}(A) = \mathcal{S}\begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}.$$

This degenerate case we would often like to exclude.

Definition 1.6. A linear matrix polynomial A of size k is called **monic** if $A(0) = I_k$, the identity matrix.

Lemma 1.7. *Let A be a linear matrix polynomial of size k.*

(1) If A is monic, the interior of $\mathcal{S}(A)$ is the set

$$\{u \in \mathbb{R}^n \mid A(u) > 0\}$$

(2) If 0 is an interior point of $\mathcal{S}(A)$, then there exists a monic linear matrix polynomial B of size $\operatorname{rk}(A(0))$ with $\mathcal{S}(A) = \mathcal{S}(B)$.

Proof. (1) Use that $\mathrm{Sym}_k^{++} = \mathrm{int}(\mathrm{Sym}_k^{+})$. (Giving an exact argument is Exercise 1.8).
(2) Let $A = A_0 + x_1 A_1 + \cdots + x_1 A_n$. Since $0 \in \mathcal{S}(A)$, we must have $A_0 \geqslant 0$. Hence there exists $U \in \mathrm{GL}_k$ such that

$$U^T A_0 U = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

where $r = \operatorname{rk}(A_0)$. Write

$$U^T A U = \begin{bmatrix} B & C \\ C^T & B' \end{bmatrix},$$

where B and B' are linear matrix polynomials of size r resp. $k-r$, and C is a non-symmetric linear matrix polynomial of size $r \times (k - r)$. We claim that $C = 0$ and $B' = 0$. By the choice of U, the constant term of B' is zero, say $B' = x_1 B_1' + \cdots + x_n B_n'$ with $B_i' \in \mathrm{Sym}_{k-r}$. Since 0 is an interior point of $\mathcal{S}(A) \subset \mathcal{S}(B')$, there is $\varepsilon > 0$ such that $\pm\varepsilon B_i' > 0$ for $i = 1, \ldots, n$. That is impossible, unless $B_i' = 0$ for $i = 1, \ldots, n$. Now if W is an open neighbourhood of 0 with $B(u) > 0$ for all $u \in W$, then $B'(u) = 0$ implies $C(u) = 0$ for all $u \in W$ (see exercise below). This implies $C = 0$, and the lemma is proved. □

Exercise 1.8. Prove part (1) of the lemma.

Exercise 1.9. Let

$$A = \begin{bmatrix} A_1 & A_2 \\ A_2^T & 0 \end{bmatrix}$$

be a real block-matrix with A_1 symmetric. Show that if A is positive semidefinite, then $A_2 = 0$.

Remark 1.8. This lemma essentially says that it is enough to consider spectrahedra defined by monic linear matrix polynomials. For if $S = \mathcal{S}(A)$ is any spectrahedron, let V be its affine hull. Then S has non-empty relative interior in V and after a translation we may assume that 0 is in the relative interior of S. We may then change coordinates, replace $\mathbb{R}^n$ by V, and assume that S is given by a monic linear matrix polynomial.

While this argument works fine as a first reduction step in a proof, actually computing a monic representation in large examples can be a difficult and computationally expensive task, which is usually avoided whenever possible.

Corollary 1.9. *A spectrahedron S has non-empty interior if and only if there exists a linear matrix polynomial A such that $S = \mathcal{S}(A)$ and A is positive definite in some point of S. In this case,* $\mathrm{int}(S) = \{u \in \mathbb{R}^n \mid A(u) > 0\}$.

Proof. If $A(u) > 0$ for some $u \in S$, we may translate and assume $u = 0$. Then $C = A(0)^{-1}$ is positive definite and $A'(x) = \sqrt{C}A(x)\sqrt{C}$ is a monic linear matrix polynomial with $\mathcal{S}(A) = \mathcal{S}(A')$. With this observation, the claim follows from the lemma. □

Remark 1.10. If A is a monic linear matrix polynomial of size k, the interior of $\mathcal{S}(A)$ is the basic open set defined by the principal minors of $A(x)$, which are the determinants of $A(x)$ with the last k rows and columns deleted, for $k = 0, \ldots, k-1$. These are k polynomials of ascending degree $1, \ldots, k$. But in general, it is not true that the closure of a basic open set $\mathcal{U}(g_1, \ldots, g_r) = \{u \in \mathbb{R}^n \,|\, g(u) > 0\}$ is the basic closed set $\mathcal{S}(g_1, \ldots, g_r)$, and it is indeed not true that a matrix is positive semidefinite if all its principal minors are non-negative. (Just take the diagonal matrix $\mathrm{Diag}(0, -1)$, all of whose principal minors are 0). What is true, however, is that a matrix is positive semidefinite if and only if all its diagonal minors are non-negative (which are the determinants of the submatrices of $A(x)$ obtained by deleting all rows and columns with indices in some subset of $\{1, \ldots, k\}$). This gives another proof of Prop. 1.4. However, this description uses 2^k inequalities rather than k.

So what makes spectrahedra special among convex semialgebraic sets? First, spectrahedra are basic closed, as we have seen. There are indeed closed semialgebraic sets that are not basic, so these cannot be spectrahedra.

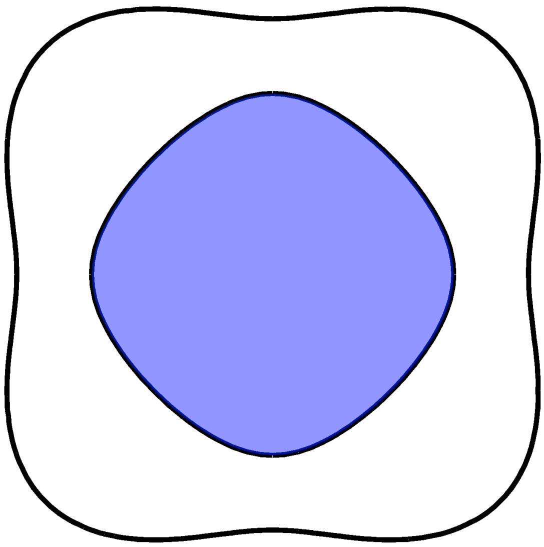
Exercise 1.10. Let $u_1 = (-1, 0)$ and $u_2 = (1, 0)$ in $\mathbb{R}^2$ and let $S = B_1(u_1) \cup B_1(u_2)$, i.e. the union of two discs of radius 1 about u_1 and u_2. Show that S is basic closed, but the convex hull of S is not. (Getting the idea is more important than a rigorous proof.) This set is called the **football stadium**.

Exercise 1.11. Let $C = \left\{(b,c) \in \mathbb{R}^2 \,\middle|\, \text{For all } x \in \mathbb{R}\colon x^4 - bx^2 + (1/4)c \geqslant 0\right\}$. Show that C is closed and convex, but not basic closed.

Another property of spectrahedra is that all their faces are exposed. We will discuss this in the proper context later.

Finally, there is a very restrictive necessary condition, called **hyperbolicity** (or **real-zero property**). This comes from the fact that symmetric matrices have real eigenvalues. If $A = I_k + \sum_{i=1}^n x_i A_i$ is a monic linear matrix polynomial of size k, the determinant $f = \det(A)$ is a polynomial (of degree at most k) in $\mathbb{R}[x]$ which vanishes on the boundary of the spectrahedron $\mathcal{S}(A)$. Since all the matrices $\sum u_i A_i$ for $u \in \mathbb{R}^n$ are real symmetric, their characteristic polynomials $\det(tI_k - \sum u_i A_i)$ have only real roots. Since $\det(tI_k - \sum u_i A_i) = t^k \det(A(t^{-1}u)) = t^k f(t^{-1}u)$, this means that the polynomial f has only real roots when restricted to any line $\operatorname{span}(u) = \{t^{-1}u \mid t \in \mathbb{R}^*\} \cup \{0\}$ through 0. This can be seen for a polynomial of degree 4 in two variables in the picture below.

An important paper in which such polynomials were studied in connection with spectrahedra is [**HV07**]. Some of the results there will be discussed in detail later on. For now, the basic point we wish to make is simply that spectrahedra are very special convex sets.



A hyperbolic plane curve of degree 4

Example 1.11. The set $\{(u,v) \in \mathbb{R}^2 \mid u^4 + v^4 \leqslant 1\}$ is known as the **TV screen**. It is basic closed and convex but not a spectrahedron. The reason is that $x^4 + y^4 - 1$ does not satisfy the hyperbolicity condition above.

Exercise 1.12. Prove that the TV screen is indeed not a spectrahedron.

REFERENCES

[BRS11] A. Bhardwaj, P. Rostalski, and R. Sanyal. Deciding polyhedrality of spectrahedra, 2011. Preprint. http://arxiv.org/abs/1102.4367

[HV07] J. W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. *Comm. Pure Appl. Math.*, **60** (5), 654–674, 2007. http://arxiv.org/abs/math/0306180

Daniel Plaumann
Universität Konstanz
Summer 2013

2. PROJECTED SPECTRAHEDRA AND DUALITY

2.1. OVERVIEW

A **projected spectrahedron** is the image of a spectrahedron under an affine-linear map. Such an image is again convex and semialgebraic (by quantifier elimination) but need not be a spectrahedron. One reason is that a projected spectrahedron is not necessarily closed, since linear maps are not closed. For example, the projection of the spectrahedron

$$\left\{(u,v)\in\mathbb{R}^2 \;\middle|\; \begin{bmatrix} u & 1\\ 1 & v\end{bmatrix} \succcurlyeq 0\right\}$$

onto the first factor is the open interval $(0,\infty)$. But there is much more to it.

Example 2.1.

(1) The TV-screen $C=\left\{(u,v)\in\mathbb{R}^2 \mid u^2+v^4\leqslant 1\right\}$ is not a spectrahedron (Example 1.11). But the spectrahedron

$$S=\left\{(u,v,a,b)\in\mathbb{R}^4 \;\middle|\; \begin{bmatrix} 1+a & b & & & & \\ b & 1-a & & & & \\ & & 1 & u & & \\ & & u & a & & \\ & & & & 1 & v\\ & & & & v & b\end{bmatrix}\succcurlyeq 0\right\}$$

in $\mathbb{R}^4$ maps onto the TV-screen under the projection $(u,v,a,b)\mapsto(u,v)$. For $(u,v,a,b)\in S$ satisfies $a^2+b^2\leqslant 1$ and $u^2\leqslant a$, $v^2\leqslant b$, hence $u^4+v^4\leqslant a^2+b^2\leqslant 1$. Conversely, any point $(u,v)\in C$ lifts to the point $(u,v,u^2,v^2)\in S$.

(2) Let $\mathbb{R}[x]_d$ be the vector space of polynomials in x of degree at most d and let

$$\Sigma_{2d}=\left\{f_1^2+\cdots+f_r^2 \mid f_1,\ldots,f_r\in\mathbb{R}[x]_d,\ r\in\mathbb{N}\right\}$$

be the cone of sums of squares (*sos-cone*). The sos-cone is not a spectrahedron. In fact, it is closed and convex, but it is not basic closed if $d>1$. (In Exercise 1.11 we identified an affine-linear slice that is not basic closed).

It is however a projected spectrahedron, namely it is the image of Sym_k^+ under the Gram map $\mathrm{Sym}_k\to\mathbb{R}[x]_{2d}$, $A\mapsto m^TAm$, where m is a vector of polynomials spanning $\mathbb{R}[x]_d$ (see Example 1.3(4)).

Exercise 2.1. Represent the football stadium as a projected spectrahedron.

Note first that we speak of *projected* spectrahedra, because every affine-linear map can be factored into an injective one followed by a projection. This simple fact translates into the following statement.

Lemma 2.2. *Let $P \subset \mathbb{R}^n$ be a projected spectrahedron. Then there exists a linear matrix polynomial $A(x,y)$ in variables $x = (x_1,\ldots,x_n)$ and $y = (y_1,\ldots,y_p)$, for some p, with*

$$P = \{u \in \mathbb{R}^n \mid \exists v \in \mathbb{R}^p \colon A(u,v) \succeq 0\}.$$

Proof. Let $P = \varphi(S)$ where $S = \mathcal{S}(B)$ is a spectrahedron in $\mathbb{R}^m$ and $\varphi\colon \mathbb{R}^m \to \mathbb{R}^n$ an affine linear map. Let $\Gamma\colon \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$, $v \mapsto (\varphi(v), v)$ be the Graph map. Since Γ is injective, $\Gamma(S)$ is a spectrahedron. Namely, $\Gamma(S) = \mathcal{S}(A)$ for

$$A(x,y) = \begin{bmatrix} B(y) & 0 & 0 \\ 0 & \varphi(y) - x & 0 \\ 0 & 0 & \varphi(y) + x \end{bmatrix},$$

where the entries $\varphi(y) - x$ and $\varphi(y) + x$ are diagonal blocks in the entries of x. Thus P has the desired representation. □

A representation of P as in the lemma is also called a **lifted linear matrix inequality representation**, an **extended formulation** or simply a **semidefinite representation**. If $A(x,y)$ is a linear matrix polynomial as above, we will often denote the projection $\mathbb{R}^{n+p} \to \mathbb{R}^n$ onto the first coordinates by π_x. Thus the lifted representation will be denoted by

$$P = \pi_x \mathcal{S}(A(x,y))$$

or just $P = \pi_x \mathcal{S}(A)$.

While spectrahedra are very special convex sets, it is not known whether projected spectrahedra have any distinguishing features at all (beyond the obvious).

Helton-Nie Conjecture. Every convex semialgebraic set is a projected spectrahedron.

The conjecture is open in general, but is well motivated by what is known. Helton and Nie proved a series of results which can be summed up as saying that any convex semialgebraic set with sufficiently regular boundary is a projectred spectrahedron. Recently, Scheiderer has given a proof of the full conjecture for subsets of the plane. Understanding and proving some of these results will be one of our main goals.

We will start in this lecture by establishing some basic properties of projected spectrahedra. In particular, we will show that all the usual operations preserving convexity, including convex duality, also preserve the property of being a projected spectrahedron.

2.2. CONES AND DUALITY

We will need some basics from convexity concerning cones and duality. Let V be a finite-dimensional real vector space. By a **cone** in V, we will always mean a convex cone, i.e. a non-empty subset $K \subset V$ such that $u + v \in K$ and $\alpha v \in K$ hold for all $u, v \in K$ and $\alpha \in \mathbb{R}$ with $\alpha \geq 0$. In particular, a cone always contains 0. A cone K is called **pointed** if

$K \cap (-K) = \{0\}$. For example, Sym_k^+ is a pointed cone in Sym_k, while any non-zero linear subspace of V is an example of a non-pointed cone. Given a subset S of V, we write

$$\mathrm{cone}(S) = \left\{ \sum_{i=1}^{k} \alpha_i u_i \mid u_i \in S, \alpha_i \geqslant 0, k \in \mathbb{N} \right\}$$

for the **conic hull** of S in V, the smallest cone in V containing S.

Proposition 2.3. *A cone $K \subset V$ is closed and pointed if and only if there exists a compact convex subset C of V with $0 \notin C$ and $K = \mathrm{cone}(C)$.*

Proof. See [**Barvinok**, II.8]. □

We write $V^* = \mathrm{Hom}(V, \mathbb{R}) = \{L \colon V \to \mathbb{R} \text{ linear}\}$ for the **dual space** of V, whose elements are the **linear functionals** on V. If $\varphi \colon V \to W$ is a linear map between vector spaces, the map $\varphi^* \colon W^* \to V^*$ given by $L \mapsto L \circ \varphi$ is again linear. By definition, it has the property

$$L(\varphi(v)) = \varphi^* L(v) \text{ for all } v \in V.$$

Exactness of the duality or direct computation also show $\mathrm{im}(\varphi^*) = \{L \in V^* \mid L|_{\ker(\varphi)} = 0\}$.

If $C \subset V$ is convex, we denote by

$$C^* = \left\{ L \in V^* \mid \text{For all } v \in C\colon\ L(v) \geqslant -1 \right\}$$

the **convex dual** of C. If $K \subset V$ is a cone, then

$$K^* = \left\{ L \in V^* \mid \text{For all } v \in C\colon\ L(v) \geqslant 0 \right\}.$$

For given $L \in K^*$ and $u \in K$, we must have $L(\alpha u) = \alpha L(u) \geqslant -1$ for all $\alpha > 0$, so $L(u) \geqslant 0$. This fact makes the duality theory for cones run somewhat more smoothly than for general convex sets. Note also that if $U \subset V$ is a linear subspace, then $U^* = U^\perp = \left\{ L \in V^* | L|_U = 0 \right\}$. (Since U^* also denotes the dual space of U, the notation $U^\perp$ is preferred.)

The fundamental fact is **biduality**, a consequence of the separation theorem for closed convex sets (see [**Barvinok**] or [**Convexity-LN**]).

Theorem 2.4 (Biduality). *For any convex subset C of V, we have*

$$(C^*)^* = \mathrm{clos}\big(\mathrm{conv}(C \cup \{0\})\big).$$

In particular,

(1) *if $C \subset V$ is a closed convex subset containing 0, then $(C^*)^* = C$.*
(2) *if $K \subset V$ is a closed cone, then $(K^*)^* = K$.* □

If V is a finite-dimensional euclidean space with scalar product $\langle -, - \rangle$, we can identify the dual space V^* with V using the map $V \to V^*$, $u \mapsto \langle u, - \rangle$. In this setting, the dual of a cone $K \subset V$ is $K^* = \{u \in V \mid \text{For all } v \in K\colon\ \langle u, v\rangle \geqslant 0\}$, and K is called **selfdual** if $K^* = K$. On the space V of matrices, the standard scalar product is given by the trace via

$$\langle A, B \rangle = \mathrm{tr}(AB^T)$$

for $A, B \in \mathrm{Mat}_k$.

The following will be essential.

Proposition 2.5. *The cone of real positive semidefinite matrices is selfdual.*

Proof. Let $A, B \in \mathrm{Sym}_k^+$ and write $A = PP^T$, $B = QQ^T$ (see Exercise 1.1). Then $\langle A, B\rangle = \langle PP^T, QQ^T\rangle = \mathrm{tr}(PP^TQQ^T) = \mathrm{tr}(Q^TPP^TQ) = \mathrm{tr}(Q^TP(Q^TP)^T) \geqslant 0$. (Here, we used that the trace is invariant under cyclic permutations.) Conversely, let $A \in \mathrm{Sym}_k$ with $\langle A, B\rangle \geqslant 0$ for all $B \in \mathrm{Sym}_k^+$. Then $v^TAv = \mathrm{tr}(v^TAv) = \mathrm{tr}(Avv^T) = \langle A, vv^T\rangle \geqslant 0$, for all $v \in \mathbb{R}^k$. □

Exercise 2.2. Let V_1 and V_2 be finite-dimensional euclidean spaces with scalar products $\langle -, -\rangle_1$ and $\langle -, -\rangle_2$. Given a linear map $\varphi\colon V_1 \to V_2$, show that there is a unique linear map $\varphi^*\colon V_2 \to V_1$ with

$$\langle \varphi(v), w\rangle_2 = \langle v, \varphi^*(w)\rangle_1$$

for all $v \in V_1$, $w \in V_2$. Verify that this corresponds to the dual map of φ under the identification $V_1 = V_1^*$, $V_2 = V_2^*$ via the scalar product.

Before returning to projected spectrahedra, we need a few more technical lemmas.

Lemma 2.6. *Let K_1 and K_2 be cones in V.*

(1) $(K_1 + K_2)^* = K_1^* \cap K_2^*$
(2) If K_1 and K_2 are closed, then $(K_1 \cap K_2)^* = \mathrm{clos}(K_1^* + K_2^*)$.

Proof. (1) is immediate. For (2), we use biduality and (1) to conclude $\mathrm{clos}(K_1^* + K_2^*) = (K_1^* + K_2^*)^{**} = (K_1^{**} \cap K_2^{**})^* = (K_1 \cap K_2)^*$. □

Exercise 2.3. Find an example of two closed cones K_1 and K_2 in $\mathbb{R}^3$ such that $K_1 + K_2$ is not closed.

Lemma 2.7. *Let $J \subset V$ and $K \subset W$ be cones and $\varphi\colon V \to W$ a linear map. Then*

(1) $\varphi(J)^* = (\varphi^*)^{-1}(J^*)$;
(2) $\varphi^{-1}(K)^* = \varphi^*((K \cap \mathrm{im}(\varphi))^*)$.

Proof. (1) This follows directly from the equality $\varphi^*(L(v)) = L(\varphi(v))$ for all $v \in V$.

(2) If $L \in (K \cap \mathrm{im}(\varphi))^* \subset W^*$, then $\varphi^*L(v) = L(\varphi(v)) \geqslant 0$ for all $v \in \varphi^{-1}(K)$, hence $\varphi^*L \in \varphi^{-1}(K)^*$. Conversely, since $\ker(\varphi) \subset \varphi^{-1}(K)$, any $L \in \varphi^{-1}(K)^*$ must vanish on $\ker(\varphi)$ and is therefore in the image of φ^*. Then if $L = \varphi^*L'$, we have $L'(\varphi(v)) = L(v) \geqslant 0$ whenever $\varphi(v) \in K$, hence $L' \in (K \cap \mathrm{im}(\varphi))^*$. □

Lemma 2.8. *Let K be a closed cone in V and U a linear subspace of V.*

(1) If $\mathrm{int}(K) \cap U \neq \varnothing$, then $K^* \cap U^\perp\{0\}$.
(2) If K is pointed and $K \cap U = \{0\}$, then $K + U$ is closed.

Proof. (1) Let $\pi\colon V \to V/U$ be the canonical projection. If $x \in \mathrm{int}(K) \cap U$, then $0 = \pi(x)$ is an interior point of $\pi(K)$ in V/U. Thus $\pi(K) = V/U$ and $K + U = \pi^{-1}(\pi(K)) = V$. Thus $K^* \cap U^\perp = (K + U)^* = V^* = \{0\}$.

(2) Since the cone K is closed and pointed, there is a compact convex subset S of V not containing 0 with $K = \mathrm{cone}(S)$, by Prop. 2.3. By hypothesis, $S \cap \ker(\pi) = \varnothing$, with π as before, so that $\pi(S)$ is again compact and convex with $0 \notin \pi(S)$. Hence $\pi(K) = \mathrm{cone}(\pi(S))$ is closed, again by Prop. 2.3, and so is $K + U = \pi^{-1}(\pi(K))$. □

2.3. OPERATIONS ON PROJECTED SPECTRAHEDRA

We are now ready for the main result of this chapter.

Theorem 2.9. *Let $P, Q \subset \mathbb{R}^m$, $R \subset \mathbb{R}^n$ be projected spectrahedra, and let $\varphi \colon \mathbb{R}^m \to \mathbb{R}^n$ be an affine-linear map. Then the following sets are again projected spectrahedra.*

(1) *Intersection:* $P \cap Q$
(2) *Cartesian product:* $P \times R$
(3) *Minkowski sum:* $P + Q$
(4) *Conic hull:* $\operatorname{cone}(P)$
(5) *Convex hull:* $\operatorname{conv}(P \cup Q)$
(6) *Linear image:* $\varphi(P)$
(7) *Inverse image:* $\varphi^{-1}(R)$
(8) *Convex dual:* $P^* \subset (\mathbb{R}^m)^*$
(9) *Closure:* $\operatorname{clos}(P)$
(10) *Relative interior:* $\operatorname{relint}(P)$

Proof. Let $P = \pi_x \mathcal{S}(A)$, $Q = \pi_x \mathcal{S}(B)$ for linear matrix polynomials $A(x,y), B(x,y')$ in variables $x = (x_1, \ldots, x_m)$, $y = (y_1, \ldots, y_p)$, $y' = (y'_1, \ldots, y'_{p'})$.

(1) We have $P \cap Q = \pi_x \mathcal{S}(C(x,y,y'))$ where

$$C(x,y,y') = \begin{bmatrix} A(x,y) & 0 \\ 0 & B(x,y') \end{bmatrix}.$$

(2) Clearly, $P \times \mathbb{R}^n = \{(u,u') \in \mathbb{R}^{m+n} \mid \exists v \in \mathbb{R}^p \colon A(u,v) \succcurlyeq 0\}$ is a projected spechtrahedron, and so is $\mathbb{R}^m \times R$. Therefore $P \times R = (P \times \mathbb{R}^n) \cap (\mathbb{R}^m \times R)$ is a projected spectrahedron.

(3) Let $\sigma \colon \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ be the linear map $(u,u') \mapsto u + u'$, then $P + Q = \sigma(P \times Q)$.

(4) Since P is already convex, the conic hull is simply given by

$$\operatorname{cone}(P) = \{u \in \mathbb{R}^m \mid \exists \lambda > 0 \mid \lambda^{-1} u \in P\} \cup \{0\}.$$

Write $A(x,y) = A_0 + A'(x,y)$ with $A'(0,0) = 0$. Now $\lambda^{-1}u \in P$ for $u \in \mathbb{R}^m$ and $\lambda > 0$ means $A(\lambda^{-1}u, \lambda^{-1}v) = A_0 + \lambda^{-1}A'(u,v) \succcurlyeq 0$ for some $v \in \mathbb{R}^p$ which is equivalent to $\lambda A_0 + A'(u,v) \succcurlyeq 0$. Thus the first guess is to look at $\{u \in \mathbb{R}^m \mid \exists \lambda \geqslant 0 \exists v \in \mathbb{R}^p \colon \lambda A_0 + A'(u,v) \succcurlyeq 0\}$. That almost works, but we run into trouble for $\lambda = 0$. To fix this, we define

$$C_i(x,y,s,t) = \begin{bmatrix} s & x_i \\ x_i & t \end{bmatrix},$$

so that $\mathcal{S}(C_i) = \{(u,v,\lambda,\mu) \mid \lambda, \mu \geqslant 0, \lambda\mu \geqslant u_i^2\}$. With this we can write

$$\operatorname{cone}(P) = \left\{u \in \mathbb{R}^m \mid \exists v, \lambda, \mu \mid \lambda A_0 + A'(u,v) \succcurlyeq 0 \text{ and } C_i(u,\lambda,\mu) \succcurlyeq 0 \text{ for } i = 1, \ldots, m\right\}.$$

To see this, let $u \in \operatorname{cone}(P)$. If $u \neq 0$, this means $\lambda^{-1}u \in P$ for some $\lambda > 0$. Then there exists μ such that $\lambda\mu \geqslant u_i^2$ for all i, and v such that $\lambda A_0 + A'(u,v) \succcurlyeq 0$. Conversely, if u is contained in the right hand side, we have $\lambda^{-1}u \in P$ for some $\lambda \geqslant 0$. If $\lambda > 0$, then $u \in \operatorname{cone}(P)$, as desired. If $\lambda = 0$, then $C_i(u,v,\lambda,\mu) \succcurlyeq 0$ implies $u_i = 0$ for all $i = 1, \ldots, n$, so that $u = 0$.

(5) Let $K = \operatorname{cone}(P \times \{1\}) + \operatorname{cone}(Q \times \{1\}) \subset \mathbb{R}^{m+1}$. Now K is a projected spectrahedron, hence so is $\operatorname{conv}(P \cup Q) = \{u \in \mathbb{R}^m \mid (u,1) \in K\}$. (The equality can be checked directly.)

(6) By definition.

(7) The set $\{(u,v) \in \mathbb{R}^m \times R \mid \varphi(u) = v\}$ is a projected spectrahedron, and $\varphi^{-1}(R)$ is the projection onto the first factor.

(8) Let $\Phi\colon \mathbb{R}^{m+p} \to \mathrm{Sym}_k$ be the map $(u,v) \mapsto A(u,v)$. Assume first that $A(0,0) = 0$, so that Φ is linear and $\mathcal{S}(A) = \Phi^{-1}(\mathrm{Sym}_k^+)$ is a cone. By Cor. 1.9, we may assume $\mathrm{im}(\Phi) \cap \mathrm{int}(\mathrm{Sym}_d^+) \neq \varnothing$. Using the preceding lemmas and the fact that Sym_k^+ is selfdual, we see that $\mathcal{S}(A)^* = \Phi^{-1}(\mathrm{Sym}_k^+)^* = \Phi^*((\mathrm{Sym}_k^+ \cap \mathrm{im}(\Phi))^*) = \Phi^*(\mathrm{Sym}_k^+ + \mathrm{im}(\Phi)^\perp)$ is a projected spectrahedron. Hence so is $P^* = (\pi_x \mathcal{S}(A))^* = (\pi_x^*)^{-1}\mathcal{S}(A)^*$. This proves the conic case.

In general, we know that the conic hull $P' = \mathrm{cone}(P \times \{1\}) \subset \mathbb{R}^{m+1}$ is a projected spectrahedron by (4). Hence so are $(P')^* = \{L' \in (\mathbb{R}^{m+1})^* \mid L'|_{P'} \geqslant 0\}$ and the intersection $Q = (P')^* \cap \{L' \in (\mathbb{R}^{m+1})^* \mid L'(0,1) = 1\}$. Let $\psi\colon (\mathbb{R}^{m+1})^* \to (\mathbb{R}^m)^*$ be the restriction map given by $(\psi L')(u) = L'(u,0)$ for $L' \in (\mathbb{R}^{m+1})^*$ and $u \in \mathbb{R}^m$. We claim that $P^* = \psi Q$. For given $L' \in Q$ and $u \in P$, we have $(\psi L')(u) = L'(u,0) = L'(u,1) - L'(0,1) \geqslant -1$, since $(u,1) \in P'$. Conversely, given $L \in (\mathbb{R}^m)^*$ with $L|_P \geqslant -1$, put $L'(v,\alpha) = L(v) + \alpha$ for $(v,\alpha) \in \mathbb{R}^{m+1}$. Then $L'(\lambda u, \lambda) = \lambda(L(u)+1) \geqslant 0$ for all $u \in P$ and $\lambda \geqslant 0$, so that $L' \in Q$ and $\psi L' = L$, proving the claim.

(9) Follows from (8), since we may assume $0 \in P$ and thus $\mathrm{clos}(P) = (P^*)^*$ by biduality.

(10) Since $\pi_x(\mathrm{relint}(\mathcal{S}(A))) = \mathrm{relint}(\pi_x(\mathcal{S}(A))) = \mathrm{relint}(P)$ (see Exercise 2.5 below), it suffices to show that $\mathrm{relint}(S)$ for $S = \mathcal{S}(A)$ is a projected spectrahedron. For any fixed point $u_0 \in \mathrm{relint}(S)$ and $u \in \mathbb{R}^{m+p}$, we have $u \in \mathrm{relint}(S)$ if and only if there exists $\varepsilon > 0$ such that $u + \varepsilon(u - u_o) \in S$, i.e. $A(u + \varepsilon(u - u_0)) \succcurlyeq 0$ (Exercise 2.4). Now we just write out $A(x) = A_0 + A'(x)$ with $A'(0) = 0$, put $\delta = \frac{1}{1+\varepsilon}$, compute

$$\delta A(u + \varepsilon(u - u_0)) = \delta A_0 + A'(u) - \delta\varepsilon A'(u_0)$$

and conclude that

$$\mathrm{relint}(S) = \left\{u \in \mathbb{R}^{m+p} \mid \exists \delta \in (0,1)\colon\ \delta A_0 + A'(u) + (\delta - 1)A'(u_0) \succcurlyeq 0\right\}.$$

Combining this with a representation of the open unit interval

$$(0,1) = \left\{\delta \in \mathbb{R} \;\middle|\; \exists \lambda\colon \begin{bmatrix} \lambda & 1 \\ 1 & \delta \end{bmatrix} \succcurlyeq 0 \text{ and } \begin{bmatrix} \lambda & 1 \\ 1 & 1-\delta \end{bmatrix} \succcurlyeq 0\right\},$$

we obtain a representation of $\mathrm{relint}(S)$ as a projected spectrahedron. □

Exercise 2.4. Let C be a convex subset of a real vector space V and let U be the affine hull of C. Recall that the **relative interior** of C is the set

$$\mathrm{relint}(C) = \left\{u \in C \mid \exists \varepsilon > 0\colon\ B_\varepsilon(u) \cap U \subset C\right\}.$$

(a) For any fixed point $u_0 \in \mathrm{relint}(C)$, show

$$\mathrm{relint}(C) = \left\{u \in V \mid \exists \varepsilon > 0\colon\ u + \varepsilon(u - u_0) \in C\right\}.$$

(b) Show that $\mathrm{relint}(C)$ is convex.

Exercise 2.5. Let $C \subset \mathbb{R}^m$ be a convex set and $\varphi\colon \mathbb{R}^m \to \mathbb{R}^n$ a linear map. Show that

$$\varphi(\mathrm{relint}(C)) = \mathrm{relint}(\varphi(C)).$$

Exercise 2.6. To what extent do the results in Thm. 2.9 also hold for spectrahedra? I.e. if P, Q, R are spectrahedra, which of (1)-(10) are again spectrahedra?

Note that the proof of Theorem 2.9 is completely constructive, so given concrete representations of P, Q, R as projected spectrahedra, the steps in the proof can be turned into concrete representations of the resulting convex sets.

We have already seen that the cone Σ_{2d} of sums of squares of polynomials of degree at most d in $\mathbb{R}[x]_{2d}$ is a projected spectrahedron. By Thm. 2.9(8), its dual Σ_{2d}^* is also a projected spectrahedron. But in fact, more is true.

Proposition 2.10. *The dual cone Σ_{2d}^* is a spectrahedron.*

Proof. By definition, we have

$$\Sigma_{2d}^* = \left\{ L \in \mathbb{R}[x]_{2d}^* \mid \text{For all } f \in \mathbb{R}[x]_d \colon L(f^2) \geqslant 0 \right\}.$$

Any linear functional $L \in \mathbb{R}[x]_{2d}$ defines a symmetric bilinear form

$$b_L \colon \begin{cases} \mathbb{R}[x]_d \times \mathbb{R}[x]_d & \to & \mathbb{R} \\ (g,h) & \mapsto & L(gh) \end{cases}$$

and from this we obtain a linear map

$$\Phi \colon \begin{cases} \mathbb{R}[x]_{2d}^* & \to & \mathrm{Sym}(\mathbb{R}[x]_d) \\ L & \mapsto & b_L \end{cases}.$$

Now Σ_{2d}^* is the spectrahedron $\Phi^{-1}(\mathrm{Sym}^+ \mathbb{R}[x]_d)$. $\square$

Using Thm. 2.9(8), this also shows again that Σ_{2d} is a projected spectrahedron.

2.4. SEMIDEFINITE PROGRAMMING

A **semidefinite programme** is a convex optimisation problem of a particular kind. In the optimisation literature, such a programme is usually written in the following form:

$$\text{(P)} \qquad \left. \begin{array}{ll} \text{Find} & \inf\langle B, X\rangle \\ \text{subject to} & \langle A_i, X\rangle = c_i \text{ for } i = 1,\dots,n \\ & X \geqslant 0 \end{array} \right\} \text{ in the variable } X \in \mathrm{Sym}_k,$$

where $A_1,\dots,A_n, B \in \mathrm{Sym}_k$ and $c \in \mathbb{R}^n$ are given. Thus the problem is to compute the minimum (or infimum) of the linear function $X \mapsto \langle B, X\rangle$ on the space of symmetric matrices under the constraint that X should be contained in the spectrahedron defined by the linear equations $\langle A_i, X\rangle = c_i$.

Starting in the 1990s, efficient algorithms for solving semidefinite programmes have been developed, based on so-called **interior-point methods**. This is the main reason for the current interest in spectrahedra, but is completely outside the scope of this course. (An overview is given in [**Convexity-LN**], more details for example in [**Forst-Hoffmann**]).

However, we want to make a few observations that also help to motivate later geometric results. First, duality plays an extremely important role. To the semidefinite programme (P) above (often called the **primal programme**), there is a corresponding **dual programme**

$$\text{(D)} \qquad \left. \begin{array}{ll} \text{Find} & \sup\langle c, y\rangle \\ \text{subject to} & \sum_{i=1}^n y_i A_i \geqslant B \end{array} \right\} \text{ in the variable } y \in \mathbb{R}^n.$$

The relation between (P) and (D) is a little mystifying at first. In particular, it is not clear how it translates into the cone duality we were looking at above (see Exercise 2.7).

Second, suppose we are given a general convex programming problem of the form

$$\left. \begin{array}{ll} \text{Find} & \inf L(u) \\ \text{subject to} & u \in C \end{array} \right\} \text{ in the variable } u \in \mathbb{R}^n.$$

where C is some convex subset of $\mathbb{R}^n$ and L a linear functional. If we wish to apply semidefinite programming methods, it makes little difference whether we can represent C as a spectrahedron $C = \mathcal{S}(A)$ or only as a projected spectrahedron $C = \pi_x \mathcal{S}(A(x,y))$. In either case, we just solve the programme for the spectrahedron $\mathcal{S}(A)$. What matters much more is how and whether we can actually find the representing linear matrix polynomial A and whether the size of the matrices and the number of extra variables y are not too large. So we should always keep the following in mind:

For optimisation, lifted representations are no worse than non-lifted representations, even though the geometry is very different.

Finally, we want to comment that the usefulness of duality in solving semidefinite programmes stems from optimality results like the following.

Theorem 2.11. *Consider the semidefinite programmes (P) and (D) above. Assume that the matrices $A_1, \dots, A_n$ are linearly independent and that both (P) and (D) possess strictly feasible points. Then X' is an optimal solution of (P) and y' an optimal solution of (D) if and only if*

$$\langle X', B - \sum_{i=1}^{n} y_i' A_i \rangle = 0$$

and there is no duality gap, which means that

$$\inf\langle B, X'\rangle = \sup\langle c, y'\rangle.$$

Proof. See [**Barvinok**, IV.7, Thm. 7.2 and Problem 2], or [**Convexity-LN**, §11,12]. □

Here, a strictly feasible point of (P) is a positive definite matrix X satisfying the constraints in (P) and similarly for (D). This theorem provides only one example of various assumptions one can make on the dual pair (P), (D) implying that there is no duality gap. The condition here is usually called the **interior point** or **Karush-Kuhn-Tucker** condition.

Exercise 2.7. Let $K_1 \subset V_1$ and $K_2 \subset V_2$ be cones in finite-dimensional euclidean spaces V_1 and V_2 with scalar products $\langle -,-\rangle_1$ and $\langle -,-\rangle_2$. A linear map $\varphi\colon V_1 \to V_2$ and elements $b \in V_1$ and $c \in V_2$ define the following dual pair of optimisation problems

$$\left.\begin{array}{ll} \text{Find} & \inf\langle b, x\rangle_1 \\ \text{subject to} & \varphi(x) - c \in K_2 \\ & x \in K_1 \end{array}\right\} \text{ in } x \in V \quad \text{and} \quad \left.\begin{array}{ll} \text{Find} & \sup\langle c, y\rangle_2 \\ \text{subject to} & \varphi^*(y) - b \in K_1^* \\ & y \in K_2^* \end{array}\right\} \text{ in } y \in V_2.$$

Verify that the duality of semidefinite programming is a special case of this general setup.

REFERENCES

[Ne07] A. Nemirovski. Advances in convex optimization: conic programming. In *International Congress of Mathematicians. Vol. I*, pp. 413–444. Eur. Math. Soc., Zürich, 2007. http://www2.isye.gatech.edu/~nemirovs/ICMNemirovski.pdf

[Ne10] T. Netzer. On semidefinite representations of non-closed sets. *Linear Algebra Appl.*, **432** (12), 3072–3078, 2010. http://dx.doi.org/10.1016/j.laa.2010.02.005

[NS09] T. Netzer and R. Sinn. A note on the convex hull of finitely many projections of spectrahedra, 2009. Online Note. http://arxiv.org/abs/0908.3386

Daniel Plaumann
Universität Konstanz
Summer 2013

3. POSITIVE POLYNOMIALS AND THE LASSERRE RELAXATION

The Lasserre relaxation in an approximation of the convex hull of a (compact) basic closed semialgebraic set through a sequence of projected spectrahedra. In this chapter, we review some background results on positive polynomials and describe the general setup.

3.1. POSITIVE POLYNOMIALS AND QUADRATIC MODULES

Let $S \subset \mathbb{R}^n$ be a semialgebraic set and let

$$\mathcal{P}(S) = \left\{ f \in \mathbb{R}[x] \mid \text{For all } u \in S\colon\ f(u) \geqslant 0 \right\}$$

be the set of non-negative polynomials on S. Understanding the structure of $\mathcal{P}(S)$ is a central goal of real algebraic geometry, for which we introduce some standard notions.

A **quadratic module** in $\mathbb{R}[x]$ (or indeed in any commutative ring with 1) is a subset $M \subset \mathbb{R}[x]$ such that

(1) $1 \in M$,
(2) $M + M \subset M$,
(3) $f^2 M \subset M$ for all $f \in \mathbb{R}[x]$.

A quadratic module is called a **preordering** if

(4) $M \cdot M \subset M$.

It is clear that $\mathcal{P}(S)$ above is a preordering. The sos-cone Σ is the smallest preordering, contained in all quadratic modules in $\mathbb{R}[x]$. More generally, for a finite set of polynomials $\underline{g} = \{g_1, \dots, g_r\} \subset \mathbb{R}[x]$, the set

$$M(\underline{g}) = \left\{ s_0 + s_1 g_1 + \cdots + s_r g_r \mid s_0, \dots, s_r \in \Sigma, \right\}$$

is clearly the smallest quadratic module containing $\underline{g}$, called the quadratic module generated by $\underline{g}$. The preordering generated by $\underline{g}$ can be written out as

$$P(\underline{g}) = \left\{ \sum_{i \in \{0,1\}^r} s_i g_1^{i_1} \cdots g_r^{i_r} \mid s_i \in \Sigma \right\}.$$

Notation 3.1. We will always implicitly define $g_0 = 1$, so that an element of $M(\underline{g})$ can be written in the form $\sum_{i=0}^r s_i g_i$ with $s_i \in \Sigma$.

Now if S is the basic closed set $\mathcal{S}(\underline{g})$, then it is clear from the definition that $M(\underline{g})$ and $P(\underline{g})$ are contained in $\mathcal{P}(S)$. Much research in real algebraic geometry revolves around the question of how close these inclusions are to equality.

Example 3.2.

(1) Every non-negative polynomial in one variable is a sum of (two) squares, so for $S = \mathbb{R}$, we have $\mathcal{P}(S) = \Sigma$. This becomes false in higher dimensions, i.e. $\Sigma \subsetneq \mathcal{P}(\mathbb{R}^n)$ for $n \geqslant 2$, by a classical result of Hilbert.

(2) Every $f \in \mathbb{R}[x]$ with $f|_{[0,1]} \geqslant 0$ is contained in $P(x(1-x))$, i.e. has a representation
$$f = s + t \cdot x(1-x)$$
for $s, t \in \Sigma$. (For general subsets of the line, see [**Marshall**, §2.7].)

Exercise 3.1. Show that $\mathcal{P}(\mathbb{R}) = \Sigma$ and $\mathcal{P}([0,1]) = M(x(1-x))$ as claimed above.

An important general result, which triggered a lot of further research on positive polynomials and sums of squares, is the following, often called *Schmüdgen's Positivstellensatz.*

Theorem 3.3 (Schmüdgen 1991). *If $\mathcal{S}(\underline{g})$ is compact, the preordering $P(\underline{g})$ contains all polynomials $f \in \mathbb{R}[x]$ with $f(u) > 0$ for all $u \in \mathcal{S}(\underline{g})$.*

Proof. See [**Marshall**, Cor. 6.1.2]. □

Definition 3.4. A quadratic module M in $\mathbb{R}[x]$ is called **archimedean**, if it contains a polynomial h such that $\mathcal{S}(h)$ is compact. If $M(\underline{g})$ is archimedean, we also say that $\underline{g}$ provides an **archimedean description** of the compact set $\mathcal{S}(\underline{g})$.

Theorem 3.5 (Putinar 1993). *If the quadratic module $M(\underline{g})$ is archimedean, then it contains all polynomials $f \in \mathbb{R}[x]$ with $f(u) > 0$ for all $u \in \mathcal{S}(\underline{g})$.*

Proof. See [**Marshall**, Thm. 5.6.1 and Thm. 7.1.1]. □

Corollary 3.6. *A finitely generated quadratic module $M(\underline{g})$ in $\mathbb{R}[x]$ is archimedean if and only if there exists a positive integer N such that $N - \sum_{i=1}^n x_i^2$ is contained in M.* □

Remark 3.7. Given Putinar's theorem, Schmüdgen's theorem can be rephrased as saying that $P(\underline{g})$ is archimedean whenever $S(\underline{g})$ is compact. One can find examples of such $\underline{g}$ for which $M(\underline{g})$ is not archimedean, which shows that Schmüdgen's theorem does not extend to quadratic modules without additional assumptions.

For practical purposes, if S is compact, the assumption that M should be archimedean is often considered quite mild, since one can just add the polynomial $N - \sum_{i=1}^r x_i^2$ to the description of S if $S \subset B_N(0)$. Since the representation of a positive polynomial in the quadratic module is simpler than in the preordering ($r+1$ summands instead of 2^r), the use of quadratic modules is often preferred.

On the other hand, we have the following negative result.

Theorem 3.8 (Scheiderer). *If S is a semialgebraic set of dimension at least 3, then $\mathcal{P}(S)$ is not a finitely generated quadratic module. In other words, if $S = \mathcal{S}(\underline{g})$ has dimension at least 3, there exists $f \in \mathcal{P}(S)$ with $f \notin M(\underline{g})$.*

Proof. See [**Marshall**, Prop. 2.6.2]. □

Of course, if S is compact, then any $f \in \mathcal{P}(S) \setminus P(\underline{g})$ must necessarily have a zero somewhere in S, by Schmügen's theorem.

Note that the cone $\mathcal{P}(S)$ lives in the infinite-dimensional vector space $\mathbb{R}[x]$. What if instead we consider the finite-dimensional slices $\mathcal{P}(S)_d = \mathcal{P}(S) \cap \mathbb{R}[x]_d$?

Proposition 3.9. *The cone $\mathcal{P}(S)_d$ is closed and semialgebraic in $\mathbb{R}[x]_d$.*

Proof. For each point $u \in \mathbb{R}^n$, let $L_u \in \mathbb{R}[x]_d^*$ be the linear functional $f \mapsto f(u)$. Then we can write

$$\mathcal{P}(S)_d = \bigcap_{u \in S} L_u^{-1}([0,\infty)).$$

Since the functionals L_u are continuous, this expresses $\mathcal{P}(S)_d$ as an intersection of closed sets, hence it is closed. Furthermore the projection onto the first factor of the set

$$\big\{(f,u,f(u)) \mid f \in \mathbb{R}[x]_d, u \in \mathbb{R}^n\big\} \cap (\mathbb{R}[x]_d \times S \times (-\infty,0)) \subset \mathbb{R}[x]_d \times \mathbb{R}^n \times \mathbb{R}$$

is semialgebraic and is the complement of $\mathcal{P}(S)_d$. Hence $\mathcal{P}(S)_d$ is semialgebraic. □

While $\mathcal{P}(S)_d$ is convex, closed and semialgebraic, it is usually not a spectrahedron, since (like Σ_{2d}) it is not basic closed. So how about finitely-generated quadratic modules instead? Let $M = M(g_1,\dots,g_r)$ and write

$$M_d = M \cap \mathbb{R}[x]_d.$$

Here, there is a crucial difference to the sos-cone. Namely, compare M_d with the cone

$$M_{\underline{g}}[d] = \left\{\sum_{i=0}^r s_i g_i \mid \deg(s_i g_i) \leqslant d \text{ for all } i = 0,\dots,r\right\},$$

in $\mathbb{R}[x]_d$, which we call the **truncation** of degree d of M (with respect to $\underline{g}$). In the case of the sos-cone, there is no difference between Σ_{2d} and $\Sigma[2d]$, because leading terms in a sum of squares cannot cancel. But for general quadratic modules, it is not true that $M_{\underline{g}}[d] = M_d$ or even that M_d is contained in $M_{\underline{g}}[e]$ for some $e \geqslant d$.

Exercise 3.2. Verify that $\Sigma_{2d} = \Sigma[2d]$ for all d.

Example 3.10. Let $g = (x(1-x))^3$ and $M = M(g) \subset \mathbb{R}[x]$ in the polynomial ring in one variable, describing the closed interval $\mathcal{S}(M) = [0,1]$. It is not hard to check that $x \notin M$. For suppose we had $x = s + tg$ with $s,t \in \Sigma$, then since $x^3|g$, we could conclude $x|s$. Since s is a sum of squares, this really implies $x^2|s$ (why?). So the right hand side would be divisible by x^2, a contradiction.

On the other hand, M is a preordering (since there is only one generator) and $\mathcal{S}(M) = [0,1]$ is compact, so M contains all strictly positive polynomials by Schmüdgen's theorem. In particular, $x + \varepsilon \in M$ holds for all $\varepsilon > 0$. Now if we had $x + \varepsilon \in M(e)$ for all $\varepsilon > 0$ and fixed $e \geqslant 0$, we could write

$$x + \varepsilon = s_\varepsilon + t_\varepsilon g$$

with $\deg(s_\varepsilon), \deg(t_\varepsilon) + 3 \leqslant e$. We could then (carefully!) take limits as $\varepsilon \to 0$ and conclude $x \in M$, a contradiction. That last argument is made precise in the following proposition.

Proposition 3.11. *If $M(\underline{g}) \cap (-M(\underline{g})) = \{0\}$, the cone $M_{\underline{g}}[d]$ is closed in $\mathbb{R}[x]_d$, for all $d \geqslant 0$.*

Proof. See [**Marshall**, Lemma 4.1.4]. □

Definition 3.12. Let $M = M(\underline{g})$ be a finitely generated quadratic module. For each integer d, we say that M is d-**stable** with respect to $\underline{g}$ if there exists an integer $e \geqslant 0$ such that $M_d \subset M_{\underline{g}}[e]$. We say that M is **stable** if it is d-stable for all degrees $d \geqslant 0$.

Note that if $M_d \subset M_{\underline{g}}[e]$, then $M_d = M_{\underline{g}}[e] \cap \mathbb{R}[x]_d$. It is also not hard to show that whether M is stable does not depend on the choice of generators $\underline{g}$. However, the notion of d-stability for fixed d *does* depend on the choice of generators.

Exercise 3.3. Show that stability is independent of the choice of generators.

By definition, a quadratic module $M(\underline{g})$ is stable if it admits **degree bounds** for representations of $f \in \mathcal{P}(\underline{g})$ in M that depend only on the degree of f. In compact situations, stability is only possible in exceptional situations of small dimension:

Exercise 3.4. Use Thm. 3.3, Thm. 3.8 and Prop. 3.11 to show: If $\mathcal{S}(\underline{g})$ is compact of dimension at least 3, then $P(\underline{g})$ is not stable. (Neither is $M(\underline{g})$, if the description is archimedean.) This is also true in dimension 2, by a (much deeper) result of Scheiderer.

3.2. THE LASSERRE RELAXATION

The basic building blocks for the Lasserre relaxation will be the dual cones $M_{\underline{g}}[d]^*$, which are spectrahedra by the following direct generalisation of Prop. 2.10.

Proposition 3.13. *Let $M = M(\underline{g})$ be a finitely generated quadratic module in $\mathbb{R}[x]$. For all $d \geqslant 0$, the dual cone $M_{\underline{g}}[d]^*$ is a spectrahedron. Hence $M_{\underline{g}}[d]$ is a projected spectrahedron.*

Proof. Let $d \geqslant 0$ and consider the finite-dimensional vector space

$$V = \big\{(p_0, \dots, p_r) \in \mathbb{R}[x]^{r+1} \mid 2\deg(p_i) + \deg(g_i) \leqslant d\big\}.$$

Just as in the proof of 2.10, we associate with $L \in \mathbb{R}[x]_d^*$ the bilinear form

$$b_L\colon \begin{cases} V \times V & \to & \mathbb{R} \\ \big((p_0, \dots, p_r)(p_0', \dots, p_r')\big) & \mapsto & \sum_{i=0}^r L(p_i p_i' g_i) \end{cases}$$

and obtain a linear map

$$\Phi\colon \begin{cases} \mathbb{R}[x]_d^* & \to & \mathrm{Sym}(V) \\ L & \mapsto & b_L \end{cases}$$

with $M_{\underline{g}}[d]^* = \Phi^{-1}(\mathrm{Sym}^+(V))$. To see this, let $L \in M_{\underline{g}}[d]^*$, then $b_L(\underline{p}, \underline{p}) = \sum_{i=0}^r p_i^2 g_i \geqslant 0$. Conversely, if b_L is positive semidefinite and $p \in \mathbb{R}[x]$ with $2\deg(p) + \deg(g_i) \leqslant d$, then $L(p^2 g_i) = b_L\big((0, \dots, p, \dots, 0), (0, \dots, p, \dots, 0)\big) \geqslant 0$, hence $L \in M_{\underline{g}}[d]^*$. □

Proposition 3.14. *Let $S = \mathcal{S}(\underline{g})$ be a basic closed subset of $\mathbb{R}^n$ and write*

$$M_{\underline{g}}[d]' = \big\{L \in M_{\underline{g}}[d]^* \mid L(1) = 1\big\}.$$

Consider the projection

$$\pi\colon \begin{cases} \mathbb{R}[x]_d^* & \to & \mathbb{R}^n \\ L & \mapsto & \big(L(x_1), \dots, L(x_n)\big). \end{cases}$$

Then $\mathrm{conv}(S) \subset \pi(M_{\underline{g}}[d]')$.

Proof. For $u \in S$, let

$$L_u\colon \begin{cases} \mathbb{R}[x]_d & \to & \mathbb{R} \\ f & \mapsto & f(u) \end{cases}.$$

Then $u = \pi(L_u)$, and since the polynomials in $M_{\underline{g}}[d]$ are non-negative on S, we have $L_u \in M_{\underline{g}}[d]^*$ and $L(1) = 1$. This implies $S \subset \pi(M_{\underline{g}}[d]')$, so also $\mathrm{conv}(S) \subset \pi(M_{\underline{g}}[d]')$. □

Definition 3.15. With notations as above, we define:

(1) For $d \geqslant 0$, the projected spectrahedron

$$\mathcal{L}_{\underline{g}}[d] = \pi(M_{\underline{g}}[d]')$$

in $\mathbb{R}^n$ is called the **Lasserre relaxation** of degree d of $\operatorname{conv}(S)$ with respect to $\underline{g}$.

(2) The Lasserre relaxation is **exact** in degree d if

$$\mathcal{L}_{\underline{g}}[d] = \operatorname{conv}(S),$$

or **exact up to closure** if these two sets have the same closure.

(3) The sequence of Lasserre relaxations $(\mathcal{L}_{\underline{g}}[d])_{d\in\mathbb{N}}$ **converges** to a subset $C \subset \mathbb{R}^n$ if

$$\operatorname{clos}(C) = \operatorname{clos}\Bigl(\bigcap_{d\geqslant 0} \mathcal{L}_{\underline{g}}[d]\Bigr).$$

Note that the chain $M_{\underline{g}}[1] \subset M_{\underline{g}}[2] \subset \cdots$ is ascending by definition, therefore the chains $M_{\underline{g}}[1]' \supset M_{\underline{g}}[2]' \supset \cdots$ and $\mathcal{L}_{\underline{g}}[1] \supset \mathcal{L}_{\underline{g}}[2] \supset \cdots$ are descending.

Our first goal is to characterise exactness in terms of stablity. We need the following version of the separation theorem for closed convex sets.

Proposition 3.16. *Let $C \subset \mathbb{R}^n$ be closed and convex. Given a point $u \in \mathbb{R}^n$, $u \notin C$, there exists a polynomial $\ell \in \mathbb{R}[x]$ with $\deg(\ell) = 1$ such that*

$$\ell|_C > 0 \qquad \textit{and} \qquad \ell(u) < 0.$$

Proof. See [**Barvinok**, Thm. III.1.3] or [**Convexity-LN**, Satz 5.4]. □

Exercise 3.5. Give a direct proof of Prop. 3.16.

Proposition 3.17. *Let $\underline{g} \subset \mathbb{R}[x]$ be finite and let $S = \mathcal{S}(\underline{g})$. For $d \geqslant 0$, consider the statements:*

(1) The Lasserre relaxation is exact up to closure in degree d, i.e.

$$\operatorname{conv}(S) \subset \mathcal{L}_{\underline{g}}[d] \subset \operatorname{clos}(\operatorname{conv}(S)).$$

(2) Every $\ell \in \mathbb{R}[x]$ with $\deg(\ell) = 1$ and $\ell|_S \geqslant 0$ is contained in $M_{\underline{g}}[d]$.

Then (2) implies (1). The converse also holds if S has non-empty interior.

Proof. (1) $\Longrightarrow$ (2). Let $\ell \in \mathbb{R}[x]$ be as in (2), say $\ell = \sum_{i=1}^n a_i x_i + b$. Since S has non-empty interior, we must have $M(\underline{g}) \cap (-M(\underline{g})) = \{0\}$. By Prop. 3.11, this implies that $M_{\underline{g}}[d]$ is closed and hence $M_{\underline{g}}[d] = (M_{\underline{g}}[d]^*)^*$ by biduality (Thm. 2.4). So if ℓ is not contained in $M_{\underline{g}}[d]$, there exists $L \in M_{\underline{g}}[d]^*$ with $L(\ell) < 0$. We claim that there also exists such L with $L(1) = 1$. If $L(1) \neq 0$, then $L(1) > 0$ since $1 \in M_{\underline{g}}[d]$, so we can just rescale. If $L(1) = 0$, take any $L_1 \in M_{\underline{g}}[d]^*$ with $L_1(1) = 1$ (e.g. a point evaluation) and let $L' = \alpha L + L_1$. Then L' has the desired property when α is sufficiently large.

It follows that $u = \bigl(L(x_1), \ldots, L(x_n)\bigr)$ is a point in $\mathcal{L}_{\underline{g}}[d] \subset \operatorname{clos}(S)$. On the other hand,

$$\begin{aligned}\ell(u) &= \ell\bigl(L(x_1), \ldots, L(x_n)\bigr) = \sum_{i=1}^n a_i L(x_i) + b = \sum_{i=1}^n a_i L(x_i) + L(b)\\ &= L(\ell) < 0,\end{aligned}$$

a contradiction. (Note that we needed $L(1) = 1$ to have $L(b) = b$).

(2) $\Longrightarrow$ (1). Let $C = \operatorname{clos}(\operatorname{conv}(S))$. Suppose that (1) does not hold, then, since $\operatorname{conv}(S) \subset \mathcal{L}_{\underline{g}}[d]$ by Prop. 3.14, there must exist $u \in \mathcal{L}_{\underline{g}}[d] \setminus C$. Hence, by Prop. 3.16, there is a polynomial $\ell \in \mathbb{R}[x]_1$ with $\ell|_S \geqslant 0$ and $\ell(u) < 0$. Since $u \in \mathcal{L}_{\underline{g}}[d]$, there exists $L \in M_{\underline{g}}[d]'$ such that $u = \big(L(x_1), \ldots, L(x_n)\big)$. Now since $\deg(\ell) = 1$ and $L(1) = 1$, we have

$$L(\ell) = \ell\big(L(x_1), \ldots, L(x_n)\big) = L(u) < 0,$$

which implies $\ell \notin M_{\underline{g}}[d]$. $\square$

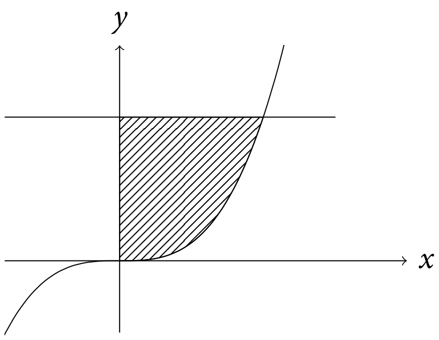
Corollary 3.18. *Let $S = \mathcal{S}(\underline{g})$ be an archimedean description of a compact set S with non-empty interior. The Lasserre relaxation of* $\operatorname{conv}(S)$ *with respect to $\underline{g}$ becomes exact if and only if the quadratic module $M(\underline{g})$ is* 1*-stable.*

Proof. Let $M = M(\underline{g})$. By Putinar's theorem, M_1 contains all $\ell \in \mathbb{R}[x]_1$ that are strictly positive on S. Since $M(\underline{g})$ is 1-stable, $M_1 = M_{\underline{g}}[d] \cap \mathbb{R}[x]_1$ for some d. Hence M_1 is closed by Prop. 3.11, since S has non-empty interior. So $M_{\underline{g}}[d]$ contains all $\ell \in \mathbb{R}[x]_1$ that are non-negative on S and $\operatorname{conv}(S) = \mathcal{L}_{\underline{g}}[d]$ by Prop. 3.17. (Note that $\operatorname{conv}(S)$ is closed.) $\square$

Theorem 3.19. *Let $S = \mathcal{S}(\underline{g})$ be an archimedean description of a compact set S. Then the Lasserre relaxations of S with respect to $\underline{g}$ converge to* $\operatorname{conv}(S)$.

Proof. We will show that $\operatorname{conv}(S) = \bigcap_{d\geqslant 0} \mathcal{L}_{\underline{g}}[d]$. The inclusion from left to right is clear by Prop. 3.14. Conversely, if $u \notin \operatorname{conv}(S)$, then there exists $\ell \in \mathbb{R}[x]_1$ with $\ell|_S > 0$ and $\ell(u) < 0$ by Prop. 3.16. By Putinar's theorem 3.5, we have $\ell \in M$ and hence $\ell \in M_{\underline{g}}[d]$ for some $d > 0$. This implies $u \notin \mathcal{L}_{\underline{g}}[d]$, by the same argument as before. $\square$

Example 3.20. To illustrate the Lasserre relaxation method, we discuss an example in detail. Let $g_1 = y - x^3$, $g_2 = x$, $g_3 = 1 - x$, $g_4 = y$, $g_5 = 1 - y$ in variables x, y. Put $M = M(\underline{g})$ and $S = \mathcal{S}(\underline{g})$. Clearly, S is already convex, and we claim that $S = \mathcal{L}_{\underline{g}}[3]$. We use Prop. 3.17 and show that $M_{\underline{g}}[3]$ contains all $\ell \in \mathbb{R}[x]_1$ with $\ell|_S \geqslant 0$.



Let $\ell \in \mathbb{R}[x]_1$ be such a polynomial. If $\ell(u) > 0$ for all $u \in S$, then ℓ will assume its mininum ε in some point of S, since S is compact. Since $\varepsilon \in M_{\underline{g}}[3]$, it suffices to show $m - \varepsilon \in M_{\underline{g}}[3]$. Also, if ℓ is already non-negative on the box $[0,1] \times [0,1]$, we can use Farkas's lemma (a standard convexity result; see for example [**Rockafellar**, Cor. 22.3.1] or [**Convexity-LN**, Exercise 10.3]) and conclude that ℓ is contained in $\operatorname{cone}(x, 1-x, y, 1-y) \subset M_{\underline{g}}[3]$.

Thus we are left with the case that ℓ describes a tangent to the cubic curve $y = x^3$ for $x \in (0,1)$. The tangent at a point (a, a^3) with $a \in (0,1)$ is given by the polynomial $\ell_a(x, y) = y - 3a^2x + 2a^3$. Direct computation now shows

$$\begin{aligned} \ell_a &= y - 3a^2x + 2a^3 = x^3 - 3a^2x + 2a^3 + (y - x^3) \\ &= (x-a)^2x + 2a\big(x-a\big)^2 + (y - x^3) \\ &= 2a\big(x-a\big)^2 + g_1 + (x-a)^2 g_2 \in M_{\underline{g}}[3]. \end{aligned}$$

Exercise 3.6. Compute the first and second Lasserre relaxation in the above example.

3.3. MODEL-THEORETIC CHARACTERISATION OF STABILTY

In this section, we will describe some abstract tools that can be used to prove the existence of degree bounds for sums of squares and quadratic modules. We will assume some familiarity with the theory of real-closed fields, in particular the Tarski principle, which says that a first-order formula in the language of ordered fields (or rings) holds in one real-closed field, say $\mathbb{R}$, if and only if it holds in *every* real-closed field. In fact, we are only interested in extension fields of $\mathbb{R}$ and the corresponding extensions of semialgebraic sets: If S is a semialgebraic subset of $\mathbb{R}^n$ and R is any real-closed extension field of $\mathbb{R}$, we write $S(R)$ for the **base extension** of S to R, which is just the subset of R^n described by the same formula[1] as S. Now an important consequence of the Tarski principle is that $S(R)$ is non-empty if and only if S is non-empty. So unlike the complex numbers or the algebraic field extensions of $\mathbb{Q}$ studied in number theory, the purpose of real closed extension fields of $\mathbb{R}$ is not to add solutions to polynomial systems. The point is rather that solvability remains the same, even though the underlying field may be radically different from $\mathbb{R}$ in other respects.

To understand this, recall that in a first-order formula we cannot quantify over the natural numbers or over subsets. This has two important consequences: (1) The archimedean axiom $\forall a \in \mathbb{R} \exists n \in \mathbb{N}\colon |a| < n$ is *not* a first order formula, and indeed the interesting real closed extension fields of $\mathbb{R}$ are **non-archimedean**, in other words, they contain infinitesimal elements. (2) A statement of the form "There exists a polynomial such that…" cannot be encoded in a first-order formula, but a statement of the form "There exists a polynomial of degree d such that…", for some fixed d, can be encoded. This provides the connection to the degree bounds in quadratic modules that we want to study.

Now, in precise technical terms, here is the statement we will need.

Theorem 3.21 ($\aleph_1$-Saturation). *There exists a real-closed extension field $\mathbb{R}^*$ of $\mathbb{R}$ with the following property: Every countable semialgebraic cover of a semialgebraic subset of $(\mathbb{R}^*)^n$ has a finite subcover. More precisely, any ultrapower $\mathbb{R}^* = \mathbb{R}^{\mathbb{N}}/\mathcal{F}$, where $\mathcal{F}$ is a non-principal ultrafilter on $\mathbb{N}$, has this property.*

Proof. See [**Prestel-Delzell**, Thm. 2.2.11]. □

Corollary 3.22. *There exists a real-closed extension field $\mathbb{R}^*$ of $\mathbb{R}$ such that the following holds. Any countable ascending chain*

$$S_1 \subset S_2 \subset S_3 \subset \cdots$$

of semialgebraic subsets of $\mathbb{R}^n$ either becomes stationary or else the union $\bigcup_{i\in\mathbb{N}} S_i(\mathbb{R}^*) \subset (\mathbb{R}^*)^n$ is not semialgebraic over $\mathbb{R}^*$. □

The ultrapower $\mathbb{R}^*$ can be written down more or less explicitly assuming that a non-principal ultrafilter $\mathcal{F}$ on $\mathbb{N}$ is given. (But it lies in the nature of non-principal ultrafilters that they exist only by virtue of the axiom of choice, so one cannot actually write one down.) To relieve the somewhat ethereal nature of the argument and to help understand what is going on here, we look at a more concrete example.

[1] A general semialgebraic set S in $\mathbb{R}^n$ is of the form $S = \bigcup_{i=1}^k \{u \in \mathbb{R}^n \mid g_{ij}(u) > 0, h_{ij}(u) = 0 \text{ for all } j = 1, \ldots, l\}$ for some finite family of polynomials $g_{ij}, h_{ij} \in \mathbb{R}[x]$. Then the base extension is simply the semialgebraic set $S(R) = \bigcup_{i=1}^k \{u \in R^n \mid g_{ij}(u) > 0, h_{ij}(u) = 0 \text{ for all } j = 1, \ldots, l\}$. Since R is real closed, this is independent of the description by the Tarski principle.

Example 3.23. Let $S_i = [-i, i] \subset \mathbb{R}$ and consider the ascending chain of closed intervals $S_1 \subset S_2 \subset \cdots$ in $\mathbb{R}$. Of course, this chain is not stationary, yet $\bigcup_{i\in\mathbb{N}} S_i = \mathbb{R}$ is semialgebraic. However, order the rational function field $\mathbb{R}(t)$ by making t infinitely large, i.e. larger than any constant in $\mathbb{R}$. (It is not hard to show that $\mathbb{R}(t)$ has a unique such order.) Then

$$\bigcup S_i(\mathbb{R}(t)) = \{f \in \mathbb{R}(t) \mid \exists n \in \mathbb{N}\colon |f| < n\}$$

is the convex hull of $\mathbb{Z}$ in $\mathbb{R}(t)$. This is *not* a semialgebraic subset of $\mathbb{R}(t)$.

The non-archimedean field $\mathbb{R}(t)$ is tiny compared to the ultrapower $\mathbb{R}^*$, but this example captures the nature of the compactness of $\mathbb{R}^*$ in the above theorem.

Exercise 3.7. Why is $\operatorname{conv}(\mathbb{Z}) \subset \mathbb{R}(t)$ not semialgebraic?

We now present an application to our problem of stability of quadratic modules. Given a finitely generated quadratic module $M = M(\underline{g})$ in $\mathbb{R}[x]$ and a real-closed extension field R of $\mathbb{R}$, consider the quadratic module $M_R(\underline{g})$ generated by $\underline{g}$ in $R[x]$. It turns out that M is stable if and only if $M_R(\underline{g})$ is the base extension of M to R for all $R/\mathbb{R}$.

Proposition 3.24. *Let $M = M(\underline{g})$ in $\mathbb{R}[x]$ be a finitely-generated quadratic module. The following are equivalent:*

1. *M is stable with respect to $\underline{g}$.*
2. *For all real-closed extension fields R of $\mathbb{R}$, $(M_R(\underline{g}))_d$ is semialgebraic for all $d \geqslant 0$.*
3. *The cone M_d is semialgebraic for all $d \geqslant 0$, and for all real-closed extension fields R of $\mathbb{R}$, $M_R(\underline{g})_d$ coincides with the base extension $M_d(R)$ of M_d to R.*

Proof. (1) $\Longrightarrow$ (3) Clearly, $M_{\underline{g}}[d]$ is semialgebraic for every $d \geqslant 0$ (it is even a projected spectrahedron). Then if M is stable, we have $M_d \subset M_{\underline{g}}[e]$ for some $e \geqslant 0$, so that $M_d = M_{\underline{g}}[e] \cap \mathbb{R}[x]_d$ is semialgebraic. Now consider the base-extension $M_d(R)$ and the set $M(R) = \bigcup_{d\in\mathbb{N}} M_d(R)$. Since addition (resp. multiplication) in $\mathbb{R}[x]$ is given by semialgebraic maps $\mathbb{R}[x]_d \times \mathbb{R}[x]_d \to \mathbb{R}[x]_d$ (resp. $\mathbb{R}[x]_d \times \mathbb{R}[x]_d \to \mathbb{R}[x]_{2d}$) and the corresponding maps in $R[x]$ are obtained by base extension, $M(R)$ is a quadratic module in $R[x]$ containing $\underline{g}$, hence $M_R(\underline{g}) \subset M(R)$. On the other hand, since M is stable we have $M(R)_d = M_d(R) = (M_{\underline{g}}[e] \cap \mathbb{R}[x]_d)(R) = M_R(\underline{g})[e] \cap R[x]_d \subset M_R(\underline{g})_d$. (Details are left as an exercise.)

(3) $\Longrightarrow$ (2) is clear.

(2) $\Longrightarrow$ (1) For $d, e \geqslant 0$, let $S_e = M_{\underline{g}}[e] \cap \mathbb{R}[x]_d$. For every real-closed extension field $R/\mathbb{R}$, the base extension $S_e(R)$ is equal to $M_R(\underline{g})[e] \cap R[x]_d$ and $\bigcup_{e\geqslant 0} S_e(R) = M_R(\underline{g})_d$, which is semialgebraic by hypothesis. Hence Thm. 3.21 implies that the ascending chain $S_1 \subset S_2 \subset \cdots$ in $\mathbb{R}[x]_d$ must become stationary, i.e. there is some e' with $S(e) = S(e')$ for all $e \geqslant e'$, hence $M_d = \bigcup_{e\geqslant 0} S_e = S_{e'} = M_{\underline{g}}[e'] \cap \mathbb{R}[x]_d$. So M is stable. $\square$

Exercise 3.8. Fill in the details in the proof of (1) $\Longrightarrow$ (3) above.

REFERENCES

[KL11] E. de Klerk and M. Laurent. On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems. SIAM J. Optimization 21, 824-832, 2011. http://homepages.cwi.nl/~monique/files/convex_pop_hierarchy_rev2.pdf

[La09] J. B. Lasserre. Convex sets with semidefinite representation. Math. Program. 120, 457–477, 2009. http://hal.archives-ouvertes.fr/docs/00/33/16/65/PDF/SDR-final.pdf

Daniel Plaumann
Universität Konstanz
Summer 2013

4. POSITIVE MATRIX POLYNOMIALS

4.1. OVERVIEW

In the previous chapter, we showed that proving the exactness of the Lasserre relaxation for the convex hull of a compact basic closed semialgebraic set is equivalent to establishing stability in degree 1 for the corresponding quadratic module. In other words, we need uniform degree bounds for representations of supporting hyperplanes obtained from Putinar's (or Schmüdgen's) theorem. On the other hand, we have already seen, in Example 3.10 and much more generally in Exercise 3.4, that such degree bounds cannot exist in great generality, i.e. for polynomials of arbitrary degree in arbitrary dimensions.

However, there do in fact exist non-uniform degree bounds, i.e. depending on other data than just the degree of the represented polynomial.

Theorem 4.1 (Putinar's theorem with degree bounds [**NS07**]).
Let $S = \mathcal{S}(\underline{g})$ be an archimedean description of a compact set. Given $\delta > 0$, every polynomial $f \in \mathbb{R}[x]$ satisfying $f(u) \geqslant \delta$ for all $u \in S$ admits a representation

$$f = s_0 + s_1 g_1 + \cdots + g_r s_r$$

in the quadratic module $M(\underline{g})$ where $s_0, \ldots, s_r \in \Sigma$ have degrees bounded by

$$\deg(s_i) \leqslant D(\underline{g}, \deg(f), \|f\|, \delta)$$

for $i = 0, \ldots, r$ (where $g_0 = 1$). □

Here, $\|f\| = \max\{f(u) \mid u \in S\}$ is the maximum norm of f on the compact set S, and the notation in the theorem means that the degree bound is provided by a fixed function D of the given arguments, i.e. a function $D\colon (X \times \mathbb{N} \times \mathbb{R}_+ \times \mathbb{R}_+) \longrightarrow \mathbb{N}$, where X is the set of finite subsets of $\mathbb{R}[x]$. Nie and Schweighofer have also determined the complexity of the function D (c.f. Remark 4.3 below).

This result by itself will not help us much: To prove exactness of the Lasserre relaxation, we would need to show that the degree bound in the theorem can be chosen independently of δ and $\|f\|$ if f has degree 1. The insight of Helton and Nie was that this could be proved under suitable regularity assumptions with the use of Lagrange functions. Their approach, to be discussed in the next chapter, requires a version of Putinar's theorem for (non-linear) matrix polynomials *and* with degree bounds.

By a **sum of squares** in the non-commutative ring $\mathrm{Sym}_k \mathbb{R}[x]$ we mean a matrix polynomial of the form $G^T G$ where G is a matrix polynomial of size $r \times k$ for some $r \geqslant 1$. Clearly, if a matrix polynomial is a sum of squares, it is positive semidefinite in every point of $\mathbb{R}^k$. Some exercises to warm up with the notion.

Exercise 4.1. Show that $F \in \mathrm{Sym}_k \mathbb{R}[x]$ is a sum of squares if and only if there exist $k \times k$-matrices $P_1, \ldots, P_s \in \mathrm{Mat}_k \mathbb{R}[x]$ such that $F = P_1^T P_1 + \cdots + P_s^T P_s$.

Exercise 4.2. Let $F \in \mathrm{Sym}_k \mathbb{R}[x]$ be a matrix polynomial. Show that F is a sum of squares in $\mathrm{Sym}_k \mathbb{R}[x]$ if and only if the polynomial $y^T F y$ in variables $(x, y) = (x_1, \ldots, x_n, y_1, \ldots, y_k)$ is a sum of squares in $\mathbb{R}[x, y]$.

Exercise 4.3. Show that if $F \in \mathrm{Sym}_k \mathbb{R}[x]$ is a sum of squares, then $\det(F)$ and hence all diagonal minors of F are sums of squares in $\mathbb{R}[x]$. What about the converse?

Exercise 4.4. Show that every globally positive semidefinite matrix polynomial in one variable is a sum of squares. (When you get stuck, go and find [**CLR80**, Thm. 7.1].)

Putinar's theorem generalises to matrix polynomials as follows.

Theorem 4.2 (Putinar's theorem for matrix polynomials with degree bounds [**HN10**]). *Let $S = \mathcal{S}(\underline{g})$ be an archimedean description of a compact set. Given $\delta > 0$, every matrix polynomial $F \in \mathrm{Sym}_k \mathbb{R}[x]$ satisfying $F(u) \succcurlyeq \delta I_k$ for all $u \in S$ admits a representation*

$$F = S_0 + g_1 S_1 + \cdots + g_r S_r,$$

where $S_0, \ldots, S_r \in \mathrm{Sym}_k \mathbb{R}[x]$ are sums of squares with degrees bounded by

$$\deg(S_i) \leqslant D(\underline{g}, k, \deg(F), \|F\|, \delta)$$

for $i = 0, \ldots, r$.

Remark 4.3. More precisely, the degree bound can be chosen to be of the form

$$D(\underline{g}, k, \deg(F), \|F\|, \delta) = c\left(k^2 \deg(F)^2 \frac{\|F\|}{\delta}\right)^c,$$

where $c > 0$ depends only $\underline{g} = (g_1, \ldots, g_r)$.

The next question is how we should prove such a thing. There are at least two different approaches, but none is entirely simple or untechnical. Since the result is central to the approach of Helton and Nie, we want to give at least an idea. Below, we will first give a relatively quick proof of the matrix version of Putinar's theorem *without* degree bounds, followed by a rough sketch of a general technique for proving the existence of degree bounds.

The recent diploma theses of Randolf Ihrig [**Ih11**] and Roxana Heß [**He13**] give very good accounts of these results, including all the details we will have to omit.

4.2. POSITIVITY IN AFFINE ALGEBRAS

We briefly recall the correspondence between affine $\mathbb{R}$-varieties, (real) radical ideals and affine $\mathbb{R}$-algebras: For any ideal I in $\mathbb{R}[x]$, the set $V = \mathcal{V}_{\mathbb{C}}(I)$ of common complex zeros of elements in I is the **affine** $\mathbb{R}$**-variety** defined by I. We denote by $V(\mathbb{R}) = \mathcal{V}_{\mathbb{R}}(I)$ its real points. Conversely, for any subset $S \subset \mathbb{R}^n$, we write $\mathcal{I}(S)$ for the vanishing ideal of S in $\mathbb{R}[x]$. The **Nullstellensatz** says that $\mathcal{I}(\mathcal{V}_{\mathbb{C}}(I)) = \sqrt{I} = \{f \in \mathbb{R}[x] \mid \exists k \geqslant 0 \colon f^k \in I\}$ is the radical of I. The **coordinate ring** of the affine variety $V = \mathcal{V}_{\mathbb{C}}(I)$ is the residue class ring

$$\mathbb{R}[V] = \mathbb{R}[x]/\sqrt{I}.$$

The real analogue is the **real Nullstellensatz**, which says that $\mathcal{I}(\mathcal{V}_{\mathbb{R}}(I)) = \sqrt[\mathrm{re}]{I}$ is the **real radical**, defined by $\sqrt[\mathrm{re}]{I} = \{f \in \mathbb{R}[x] \mid \exists k \geqslant 0, s \in \Sigma : f^{2k} + s \in I\}$. It is easy to see that a vanishing ideal $\mathcal{I}(S)$ for $S \subset \mathbb{R}^n$ is real radical, i.e. $\sqrt[\mathrm{re}]{\mathcal{I}(S)} = \mathcal{I}(S)$. A standard result in real algebraic geometry says that the radical ideal $\sqrt{I}$ is real radical if and only if $V(\mathbb{R})$ is Zariski-dense in V.

The algebra $\mathbb{R}[V]$ is a reduced, finitely generated $\mathbb{R}$-algebra. Conversely, given any such algebra $\mathcal{A}$, fix finitely many generators $y_1, \ldots, y_n$ of $\mathcal{A}$ and consider the surjective ring homomorphism $\varphi : \mathbb{R}[x] \to \mathcal{A}$, $\varphi(x_i) = y_i$. Since $\mathcal{A}$ is reduced, $I = \ker(\varphi)$ is a radical ideal in $\mathbb{R}[x]$ and $\mathcal{A}$ is isomorphic to the coordinate ring of the affine $\mathbb{R}$-variety $\mathcal{V}_{\mathbb{C}}(I)$. Furthermore, this can be made independent of the choice of generators $y_1, \ldots, y_n$ by indentifying points in $\mathcal{V}_{\mathbb{C}}(I)$ with $\mathbb{R}$-algebra homomorphisms $\mathcal{A} \to \mathbb{C}$. In this way, the set $\mathrm{Hom}_{\mathbb{R}}(\mathcal{A}, \mathbb{C})$ can be regarded as the abstract variety corresponding to $\mathcal{A}$, with real points $\mathrm{Hom}_{\mathbb{R}}(\mathcal{A}, \mathbb{R})$.

Given an affine $\mathbb{R}$-variety V and elements $g_1, \ldots, g_r \in \mathbb{R}[V]$, we have a corresponding semialgebraic subset $\mathcal{S}_V(g_1, \ldots, g_r) = \{u \in V(\mathbb{R}) \mid g_1(u) \geqslant 0, \ldots, g_r(u) \geqslant 0\}$. Just as in the polynomial ring, we call a quadratic module $M \subset \mathbb{R}[V]$ **archimedean** if it contains an element g such that $\mathcal{S}_V(g)$ is compact. We need the following generalisation of Putinar's theorem (Thm. 3.5) to this setup. (Incidentally, the proof involves Schmüdgen's theorem).

Corollary 4.4 (Putinar's Theorem for affine algebras). *Let V be an affine $\mathbb{R}$-variety with coordinate ring $\mathbb{R}[V]$ and let $M \subset \mathbb{R}[V]$ be a finitely generated archimedean quadratic module. Then M contains all elements $f \in \mathbb{R}[V]$ such that $f(u) > 0$ for all $u \in \mathcal{S}_V(M)$.*

Proof. We can fix an embedding of V into affine space and just interpret Putinar's theorem modulo the vanishing ideal: Let $y_1, \ldots, y_n$ be generators of $\mathbb{R}[V]$ and let $I \subset \mathbb{R}[x]$ be the kernel of $\varphi : x_i \mapsto y_i$, so that $V(\mathbb{R})$ is identified with the algebraic subset $\mathcal{V}_{\mathbb{R}}(I)$ of $\mathbb{R}^n$. Let $g_1, \ldots, g_r \in \mathbb{R}[x]$ be such that $\varphi(g_1), \ldots, \varphi(g_r)$ generate M in $\mathbb{R}[V]$ and let $h_1, \ldots, h_s \in \mathbb{R}[x]$ be generators of the ideal I. Let $M_0 = \varphi^{-1}(M) = M + I$, then

$$M_0 = M(g_1, \ldots, g_r, h_1, \ldots, h_s, -h_1, \ldots, -h_s).$$

This follows from the observation that any element of the form $ph_i \in I$, for $p \in \mathbb{R}[x]$, can be rewritten as $ph_i = \left(\frac{p+1}{2}\right)^2 h_i + \left(\frac{p-1}{2}\right)^2 (-h_i)$.

We need to show that M_0 is archimedean. Since M is archimedean, there is $g \in \mathbb{R}[x]$ such that $\mathcal{S}_V(\varphi(g)) \subset V(\mathbb{R})$ is compact. By Schmüdgen's theorem 3.3, the preordering

$$P = P(g, h_1, \ldots, h_s, -h_1, \ldots, h_s)$$

contains all polynomials that are strictly positive on the compact set $\mathcal{S}(g) \cap \mathcal{V}_{\mathbb{R}}(I)$. Hence there is $g' \in P$ such that $\mathcal{S}(g') \subset \mathbb{R}^n$ is compact. On the other hand, we have $P(\varphi(g)) = M(\varphi(g)) \subset M$ and thus $P = \varphi^{-1}(P(\varphi(g))) \subset \varphi^{-1}(M) = M_0$, so M_0 is archimedean.

Now if $f \in \mathbb{R}[x]$ satisfies $\varphi(f)(u) > 0$ for all $u \in \mathcal{S}_V(M)$, this implies $f(u) \geqslant 0$ for all $u \in \mathcal{S}(M_0)$ and therefore $f \in M_0$, by Putinar's theorem 3.5. Hence $\varphi(f) \in \varphi(M_0) = M$. $\square$

4.3. PUTINAR'S THEOREM FOR MATRIX POLYNOMIALS

To prove Putinar's theorem for matrix polynomials, we will use an idea of Klep and Schweighofer, as presented in [**Ih11**] and [**He13**]. Let $F \in \mathrm{Sym}_k\, \mathbb{R}[x]$ be a matrix polynomial and let $\mathcal{A}_F$ be the commutative $\mathbb{R}[x]$-subalgebra of $\mathrm{Sym}_k\, \mathbb{R}[x]$ generated by F. Explicitly,

$\mathcal{A}_F$ consists of all expressions of the form $p(x,F)$ where $p \in \mathbb{R}[x,t]$ is a polynomial in x and one additional variable t.

Lemma 4.5. *Let $F \in \mathrm{Sym}_k \mathbb{R}[x]$ be a matrix polynomial and let*

$$\varphi\colon \begin{cases} \mathbb{R}[x,t] & \to & \mathcal{A}_F \\ p(x,t) & \mapsto & p(x,F) \end{cases}.$$

(1) The minimal polynomial μ_F of F in the polynomial ring $\mathbb{R}(x)[t]$ in one variable is contained in $\mathbb{R}[x,t]$ and generates the ideal $\ker(\varphi)$.
(2) The variety $\mathcal{V}_{\mathbb{R}}(\ker\varphi)$ corresponding to $\mathcal{A}_F$ is the hypersurface consisting of all points $(u,\lambda) \in \mathbb{R}^n \times \mathbb{R}$ such that λ is an eigenvalue of $F(u)$.

Proof. (1) Recall that μ_F is the unique monic polynomial in $\mathbb{R}(x)[t]$ of minimal degree in t such that $\mu_F(F) = 0$. By the Cayley-Hamilton theorem, it is a factor of the characteristic polynomial $\chi_F(t) = \det(tI_k - F(x,t))$, say $\chi_F = \mu_F \cdot r$ with $r \in \mathbb{R}(x)[t]$ monic. Let $c \in \mathbb{R}[x]$ be the least common multiple of the denominators of the (maximally reduced) coefficients of μ_F with respect to t, so that $c\mu_F \in \mathbb{R}[x][t]$ is primitive. By Gauss's lemma (in the form of 4.6 below), we have $\frac{1}{c}r \in \mathbb{R}[x][t]$. Since r is monic, this implies $\frac{1}{c} \in \mathbb{R}[x]$, hence $c \in \mathbb{R}^\times$ and $\mu_F \in \mathbb{R}[x][t]$. It now follows from Lemma 4.6 that μ_F divides any element of $\ker(\varphi)$ in $\mathbb{R}[x,t]$ and is therefore a generator. (2) follows from (1), since $\mathcal{V}_{\mathbb{R}}(\mu_F) = \mathcal{V}_{\mathbb{R}}(\chi_F)$. □

Lemma 4.6. *Let R be a factorial ring with field of fractions K and let $p, q \in R[t]$ be polynomials with q primitive. If $p = qr$ for some $r \in K[t]$, then $r \in R[t]$.*

Proof. This is a consequence of Gauss's lemma from algebra (see [**Lang**, Cor. IV.2.2]). □

Theorem 4.7 (Putinar's theorem for matrix polynomials)**.** *Let $S = \mathcal{S}(\underline{g})$ be an archimedean description of a compact set. Every matrix polynomial $F \in \mathrm{Sym}_k \mathbb{R}[x]$ such that $F(u)$ is positive definite for all $u \in S$ admits a representation*

$$F = S_0 + g_1 S_1 + \cdots + g_r S_r$$

where $S_0, \ldots, S_r \in \mathrm{Sym}_k$ are sums of squares of matrix polynomials in $\mathcal{A}_F \subset \mathrm{Sym}_k \mathbb{R}[x]$.

Proof. Let $\mathcal{A}_F$ be the subalgebra of $\mathrm{Sym}_k \mathbb{R}[x]$ generated by F and let $\varphi\colon \mathbb{R}[x,t] \to \mathcal{A}_F$ be as above with corresponding $\mathbb{R}$-variety $V = \mathcal{V}(\ker\varphi)$. To apply Putinar's theorem for affine algebras, we first need to check that the quadratic module generated by $\underline{g}$ in $\mathcal{A}_F$ is archimedean. Since $M(\underline{g})$ is archimedean, it contains $h \in \mathbb{R}[x]$ such that $\mathcal{S}(h) \subset \mathbb{R}^n$ is compact. Now $\mathcal{S}_V(\varphi(h)) = \{(u,\lambda) \in \mathbb{R}^n \times \mathbb{R} \mid h(u) \geqslant 0, \lambda \text{ an eigenvalue of } F(u)\} \subset V(\mathbb{R})$ is also compact, because the spectral radius of $F(u)$ (largest absolute value of an eigenvalue) is bounded on the compact set $\mathcal{S}(\underline{g})$. Thus $M_{\mathcal{A}_F}(\underline{g})$ is also archimedean. (See [**He13**, Satz 5.2.1] for a more careful version of this argument.)

Now $F \in \mathcal{A}_F$ regarded as a function on $\mathcal{V}_{\mathbb{R}}(\ker(\varphi)) \subset \mathbb{R}^{n+1}$ is just the polynomial t, i.e. it is the function $V \ni (u,\lambda) \mapsto \lambda$ where λ is an eigenvalue of $F(u)$ (Lemma 4.5(2)). Since F is positive definite, this is a strictly positive function on $\mathcal{S}_V(\underline{g})$ and we can apply Putinar's theorem in the form of Cor. 4.4 to conclude that F is contained in the quadratic module generated by $\underline{g}$ in $\mathcal{A}_F$. □

4.4. EXISTENCE OF DEGREE BOUNDS

To prove the existence of degree bounds, there are esssentially two approaches: First, via a more constructive proof of Putinar's theorem which allows for an analysis of the required degrees on the way. This was the original method in [**NS07**] and [**HN10**]. Alternatively, if one is only interested in the existence of bounds (and not in the precise asymptotic behaviour as in Remark 4.3), one can use model-theoretic ideas similar to those in §3.3.

Note first that if we could prove Putinar's theorem over any real closed field, this would give us uniform degree bounds, using a similar argument as in the proof of Prop. 3.24. However, we know that such bounds cannot exist and, consequently, Putinar's theorem does not hold over general real closed fields. So we need a more subtle idea.

Suppose we are given an archimedean description $\underline{g}$ of a compact set $S = \mathcal{S}(\underline{g})$ and a matrix polynomial F of size k that is positive definite on S. The key point is the use of the archimedean property of the reals: Since F is positive definite and S is compact, there exists $N \in \mathbb{N}$ such that $F(u) \geqslant (1/N)I_k$ for all $u \in S$. It turns out that this is the condition we should generalise to real closed fields. Given a real closed extension $R/\mathbb{R}$, the convex hull $\mathcal{O} = \operatorname{conv}(\mathbb{Z}) \subset R$ is a subring of R (called the canonical valuation ring). The proper generalisation of Putinar's theorem to arbitrary real closed fields is the following:

Theorem 4.8 (Putinar's theorem for matrix polynomials over real closed fields).
Let $S = \mathcal{S}(g_1, \dots, g_r)$ be an archimedean description of a compact set and let $R/\mathbb{R}$ be real closed with canonical valuation ring $\mathcal{O}$. Suppose that $F \in \operatorname{Sym}_k \mathcal{O}[x]$ is a matrix polynomial such that there exists $N \in \mathbb{N}$ with $F(u) \geqslant (1/N)I_k$ for all $u \in S$. Then F has a representation

$$F = S_0 + g_1 S_1 + \cdots + g_r S_r$$

where $S_0, \dots, S_r \in \operatorname{Sym}_k \mathcal{O}[x]$ are sums of squares of matrix polynomials. □

We do not give a proof here. First, one has to generalise Putinar's theorem for polynomials to the above setting. The necessary tools were developed by Jacobi and Prestel and can be found in [**Prestel-Delzell**, §8.3]. Thm. 4.8 can then be derived in a similar way as we have done above in Thm. 4.7 above (see [**He13**]). But we can now use this generalised version of Putinar's theorem to deduce the existence of the degree bounds we want.

Proof of Thm. 4.2. For fixed k, d and $N, D \in \mathbb{N}$, consider the set

$$P_{N,D} = \left\{ F = \sum_{|i| \leqslant d} A_i x^i \in \operatorname{Sym}_k \mathbb{R}[x] \;\middle|\; \begin{array}{l} |(A_i)_{r,s}| \leqslant N \text{ for all } 1 \leqslant r, s \leqslant k, |i| \leqslant d \\ \forall u \in S\colon F(u) \geqslant (1/N) I_k, \\ \exists S_0, \dots, S_r \text{ sums of squares in } \operatorname{Sym}_k \mathbb{R}[x] \\ \text{ with } \deg(S_i) \leqslant D \text{ and } F = \sum_{i=0}^r g_i S_i \end{array} \right\}$$

By Putinar's theorem for matrix polynomials (Thm. 4.7), the union $P_N = \bigcup P_{N,D}$ is semi-algebraic, namely it consists of all $F \in \operatorname{Sym}_k \mathbb{R}[x]$ with the absolute values of coefficients bounded by N and $F \geqslant (1/N)I_k$ on S. Now Thm. 4.8 exactly says that the same remains true for the base extensions $P_{N,D}(R)$ for any real closed extension $R/\mathbb{R}$. Thus we can apply Cor. 3.22 and conclude that the ascending chain $P_{N,1} \subset P_{N,2} \subset \cdots$ becomes stationary. This is what we wanted to show. (Note that, compared to the precise bound given in Remark 4.3, we now have a dependence on the size of the coefficients of F rather than the maximum norm, but since all norms on a finite-dimensional space are equivalent, this makes no difference for the existence of a bound.) □

REFERENCES

[CLR80] M.D. Choi, T.Y. Lam, B. Reznick. Real zeros of positive semidefinite forms I. Math. Z. **171**(1), p. 1–26, 1980.

[He13] R. Heß. Die Sätze von Putinar und Schmüdgen für Matrixpolynome mit Gradschranken. Diplomarbeit, Universität Konstanz, 2013.

[Ih11] R. Ihrig. Positivstellensätze für den Ring der Polynommatrizen. Diplomarbeit, Universität Konstanz, 2012.

[Lang] S. Lang. *Algebra*. Revised Third Edition, GTM **211**, Springer, New York, 2002.

[NS07] J. Nie and M. Schweighofer. One the complexity of Putinar's Positivstellensatz. *J. Complexity* **23**(1), p. 135–150, 2007. http://arxiv.org/abs/0812.2657

[HN10] J. W. Helton and J. Nie. Semidefinite representation of convex sets. *Math. Program.* **122**(1), p. 21–64, 2010. http://arxiv.org/abs/0705.4068

Daniel Plaumann
Universität Konstanz
Summer 2013

5. GENERAL EXACTNESS RESULTS

The goal of this section is to present the results of Helton and Nie in [**HN10**], which provide sufficient conditions for the exactness of the Lasserre relaxation. The theses of Rainer Sinn [**Si10**] and Tom Kriel [**Kr12**] both give very good accounts, with many improvements in proofs and exposition, and the presentation here borrows from both.

5.1. LAGRANGE MULTIPLIERS AND CONVEX OPTIMISATION

The basic idea of Helton and Nie is that in order to represent a linear polynomial defining a supporting hyperplane of a convex basic closed set in the corresponding quadratic module, it is helpful to study representations of its (non-linear) Lagrange function. To explain this, we need a bit of background and terminology from optimisation. The following is a special case of the *Karush-Kuhn-Tucker* theorem.

Theorem 5.1. *Let $S = \mathcal{S}(g_1, \ldots, g_r)$ be a basic closed set. Let $f \in \mathbb{R}[x]$ and assume that $u \in S$ is a point in which f attains its minimum on S. Assume further that there is $v \in \mathbb{R}^n$ with*

$$\begin{cases} \langle \nabla g_i(u), v \rangle > 0 & \text{if } \deg(g_i) \geqslant 2 \\ \langle \nabla g_i(u), v \rangle \geqslant 0 & \text{if } \deg(g_i) \leqslant 1 \end{cases}$$

whenever $g_i(u) = 0$. Then there exist $\lambda_1, \ldots, \lambda_r \geqslant 0$ such that

$$\begin{aligned} \nabla f(u) &= \sum_{i=1}^{r} \lambda_i \nabla g_i(u) \\ \lambda_i g_i(u) &= 0 \qquad\qquad \text{for all } i = 1, \ldots, r. \end{aligned}$$

(The same holds if $f, g_1, \ldots, g_r$ are just continuously differentiable functions.)

The constants $\lambda_1, \ldots, \lambda_r$ are called **Lagrange multipliers** for f at the **minimiser** u; the second statement, which says that the Lagrange multipliers for inactive inequalities are zero, is called *complementary slackness*. There are a number of conditions, called *constraint qualifications*, implying the existence of Lagrange multipliers in a minimiser. The one stated here is the *Mangasarian-Fromowitz* constraint qualification.

Proof. We prove only the special case where the objective function f is linear, following [**Si10**, Cor. A21]. This will be all we need. See [**Forst-Hoffmann**, Thm. 2.2.5 et seq.] for a full proof. So let $f = \ell \in \mathbb{R}[x]_1$ and let $u \in S$ be a minimiser of ℓ on S. It is not restrictive to assume that $u = 0$ and $\ell(u) = 0$. Furthermore, we may assume that $g_1(0) = \cdots = g_q(0) = 0$ and $g_{q+1}(0), \ldots, g_r(0) > 0$ for some $q \geqslant 1$, i.e. exactly the first q inequalities are active at 0.

We show that there are $\lambda_1, \ldots, \lambda_q \geqslant 0$ with $\nabla\ell(0) = \sum_{i=1}^q \lambda_i \nabla g_i(0)$ (and put $\lambda_{q+1} = \cdots = \lambda_r = 0$). Suppose not, then we may apply Farkas's lemma as spelt out below and conclude that there exists $w \in \mathbb{R}^n$ such that $\langle \nabla g_i(0), w\rangle \geqslant 0$ for $i = 1, \ldots, q$ but $\langle \nabla\ell(0), w\rangle < 0$. On the other hand, we may pick $v \in \mathbb{R}^n$ as in the hypothesis. Choose $\varepsilon > 0$ with $\langle \nabla\ell(0), w+\varepsilon v\rangle < 0$. Now if $1 \leqslant i \leqslant q$ and $\deg(g_i) \geqslant 2$, then $\langle \nabla g_i(0), w + \varepsilon v\rangle > 0$ for all $\varepsilon > 0$, which implies

$$g_i(\delta(w + \varepsilon v)) \geqslant 0$$

for all sufficiently small $\delta > 0$, since $g_i(0) = 0$. The same holds if $\deg(g_i) = 1$, since in this case we have $g_i(\delta(w + \varepsilon v)) = \delta\langle \nabla g_i(0), w + \varepsilon v\rangle \geqslant 0$. Finally, we may also assume that the same holds for the inactive inequalities ($i = q+1, \ldots, r$), by making δ smaller if necessary. But this implies that $\delta(w+\varepsilon v)$ is a point in S for which $\ell(\delta(w+\varepsilon v)) = \delta\langle\nabla\ell(0), w+\varepsilon v\rangle < 0$, contradicting the fact that $0 = \ell(0)$ is the minimum of ℓ on S. □

The following basic lemma was used in the proof (and in fact earlier in Example 3.20).

Lemma 5.2 (Farkas's lemma). *Let $c_1, \ldots, c_m \in \mathbb{R}^n$ and let*

$$P = \{u \in \mathbb{R}^n \mid \langle c_i, u\rangle \geqslant 0 \text{ for all } i = 1, \ldots, m\}.$$

Then $P^* = \operatorname{cone}(c_1, \ldots, c_m)$. In other words, for $c \in \mathbb{R}^n$ there either exists $u \in P$ such that $\langle c, u\rangle < 0$ or there exist $\lambda_1, \ldots, \lambda_m \geqslant 0$ with $c = \sum_{i=1}^m \lambda_i c_i$.

Proof. We have $\operatorname{cone}(c_1, \ldots, c_m)^* = P$, so that $P^* = \operatorname{cone}(c_1, \ldots, c_m)^{**} = \operatorname{cone}(c_1, \ldots, c_m)$ by biduality (Thm. 2.4). □

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is **concave** on a convex subset $C \subset \mathbb{R}^n$ if

$$f(\lambda u + (1-\lambda)v) \geqslant \lambda f(u) + (1-\lambda) f(v)$$

holds for all $u, v \in C$ and $\lambda \in [0,1]$. If the inequality is strict at all points, f is called *strictly concave*. It is called (strictly) *convex* if the opposite inequality holds, i.e. if $-f$ is (strictly) concave. Some essential facts are contained in the following exercises.

Exercise 5.1. Let $C \subset \mathbb{R}^n$ be a convex set and let $f : \mathbb{R}^n \to \mathbb{R}$.

(a) If f is continuously differentiable, then f is concave on $C \subset \mathbb{R}^n$ if and only if

$$f(v) \leqslant f(u) + \langle \nabla f(u), v - u\rangle$$

holds for all $u, v \in C$.

(b) If f is twice continuously differentiable, then f is concave on C if and only if its Hessian $(D^2 f)(u)$ is negative semidefinite for all $u \in C$. Furthermore, if $(D^2 f)(u)$ is negative definite on C, then f is strictly concave. Give an example showing that the converse is false.

Exercise 5.2. Let $C = \mathcal{S}(g_1, \ldots, g_r)$ be a basic closed set and suppose that $g_1, \ldots, g_r$ are concave on C. Show that C is convex.

Exercise 5.3. If $C \subset \mathbb{R}^n$ is compact and convex and $f : \mathbb{R}^n \to \mathbb{R}$ is concave on C, then there is an extreme point of C in which f attains its minimum on C.

Exercise 5.4.

(a) The set of extreme points of a compact convex subset of $\mathbb{R}^2$ is compact.

(b) Give an example of a compact convex subset of $\mathbb{R}^3$ whose set of extreme points is not closed.

Corollary 5.3. *Let $C = \mathcal{S}(g_1, \dots, g_r)$ be compact and convex with non-empty interior and suppose that $g_1, \dots, g_r$ are concave on C. Then Lagrange multipliers exist for any linear polynomial at any minimiser on C.*

Proof. Let $\ell \in \mathbb{R}[x]_1$ and let $u \in C$ be a minimiser of ℓ on C. Since C has non-empty interior, there exists a point $u_0 \in C$ with $g_i(u_0) > 0$ for $i = 1, \dots, r$. Now if $g_i(u) = 0$, then

$$0 < g_i(u_0) \leqslant g_i(u) + \langle \nabla g_i(u), u_0 - u\rangle = \langle \nabla g_i(u), u_0 - u\rangle,$$

since g_i is concave on C. Hence the hypotheses of Thm. 5.1 are satisfied for $v = u_0 - u$. □

Note that in the proof of the corollary we did not need that ℓ has degree one. But in this way, it relies only on the special case of Thm. 5.1 which we have proved.

5.2. THE HELTON-NIE THEOREMS

We now come to the main results of this chapter, starting with the simplest version.

Theorem 5.4. *Let $C = \mathcal{S}(g_1, \dots, g_r)$ be an archimedean description of a compact convex set with non-empty interior. Suppose that the following condition is satisfied:*

[SOS-Concavity] *The matrices $-D^2 g_i \in \operatorname{Sym}_n \mathbb{R}[x]$ are sums of squares for $i = 1, \dots, r$.*

Then C possesses an exact Lasserre relaxation with respect to $g_1, \dots, g_r$.

Here is the principal lemma needed for the proof.

Lemma 5.5. *Let $F \in \operatorname{Sym}_k \mathbb{R}[x]$ be a matrix polynomial which is a sum of squares, and fix $u \in \mathbb{R}^n$. Then the matrix polynomial*

$$G_u(x) = \int_0^1 \int_0^t F(u + s(x-u))\mathrm{d}s\mathrm{d}t,$$

(where the integration is carried out entry-wise) is again a sum of squares.

Proof. By Exercise 4.2, a matrix polynomial $G \in \operatorname{Sym}_k \mathbb{R}[x]$ is a sum of squares if and only if the polynomial $y^T G y \in \mathbb{R}[x, y]$, which is homogeneous of degree 2 in $y = (y_1, \dots, y_k)^T$, is a sum of squares. Thus, by hypothesis, the polynomial $f = y^T F(u+s(x-u))y \in \mathbb{R}[x, y, s]$ is a sum of squares and therefore possesses a positive semidefinite Gram matrix (c.f. Example 1.3(d)). This means $f = (Bm)^T Bm$ where m is a column vector of monomials in x, y, s and B a suitable real rectangular matrix. Now write $m = U(s) \cdot \overline{m}$ where $\overline{m}$ is a vector of monomials in x, y and U is a matrix polynomial in s of appropriate size[1]. So putting $A(s) = \big(BU(s)\big)^T BU(s)$, we find that $f = \overline{m}^T A(s) \overline{m}$ and hence

$$y^T G_u y = \int_0^1 \int_0^t f(x, y, s)\mathrm{d}s\mathrm{d}t = \overline{m}^T \left(\int_0^1 \int_0^t A(s)\mathrm{d}s\mathrm{d}t \right) \overline{m},$$

showing that G_u is a sum of squares, as claimed. (For a slicker proof, using integration of Banach space-valued functions, see [**Kr12**, Lemma 3.4.9].) □

[1]To see that this is possible, let $\overline{m}$ contain all monomials in x, y up to degree d. Let $U(s)$ be the rectangular matrix $[I_N, s \cdot I_N, \cdots, s^d \cdot I_N]^T \in \operatorname{Mat}_{(d+1)N \times N} \mathbb{R}[s]$, where $N = \operatorname{length}(\overline{m})$. Then $U(s) \cdot \overline{m}$ contains all monomials of degree d in x, y, s.

Proof of Thm. 5.4. Let $\ell \in \mathbb{R}[x]_1$ with $\ell|_C \geqslant 0$ and let $u \in C$ be a minimiser. Since the matrix polynomials $-D^2 g_i$ are sums of squares, the polynomials g_i are concave (Exercise 5.1). Then Cor. 5.3 guarantees the existence of Lagrange multipliers for ℓ in u, so that we have

$$\begin{aligned} \nabla \ell(u) &= \sum_{i=1}^{r} \lambda_i \nabla g_i(u) \\ \lambda_i g_i(u) &= 0 \qquad\qquad \text{for all } i = 1, \dots, r. \end{aligned}$$

for certain $\lambda_1, \dots, \lambda_r \geqslant 0$. It follows that the function $f_\ell = \ell - \ell(u) - \sum_{i=1}^{r} \lambda_i g_i$ and its gradient both vanish at u, so the fundamental theorem of calculus implies

$$f_\ell = \int_0^1 \frac{\partial}{\partial t} f_\ell(u + t(x-u))\mathrm{d}t = \int_0^1\int_0^t \frac{\partial^2}{\partial s^2} f_\ell(u + s(x-u))\mathrm{d}s\mathrm{d}t = \tag{5.6}$$

$$= \sum_{i=1}^{r} \lambda_i \cdot (x-u)^T \underbrace{\left(\int_0^1\int_0^t -D^2 g_i(u + s(x-u))\mathrm{d}s\mathrm{d}t \right)}_{=A_u^{(i)}(x)} (x-u).$$

By Lemma 5.5, the matrix polynomials $A_u^{(i)}$ are sums of squares, and therefore

$$\ell = \ell(u) + \sum_{i=1}^{r} \lambda_i g_i + \sum_{i=1}^{r} \lambda_i (x-u)^T A_u^{(i)} (x-u)$$

is a represention of ℓ in $\mathcal{M}(\underline{g})$ in which the degrees are bounded by $\max\{\deg(g_i)\}$. $\square$

A polynomial g whose Hessian is a sum of squares is called **sos-convex**. By Exercise 5.1, any such polynomial is convex. But as usual, the converse does not hold in general.

Exercise 5.5. Let $f \in \mathbb{R}[x]$ be a homogeneous polynomial.

(a) Show that if f is convex, then f is non-negative on $\mathbb{R}^n$.
(b) Show that if f is sos-convex, then f is a sum of squares.
(c) Give an example of a convex polynomial $f_0 \in \mathbb{R}[x]$ such that the homogenisation $x_0^{\deg f} f(x/x_0) \in \mathbb{R}[x_0, x_1, \dots, x_n]$ is not convex.

For an example of a convex sum of squares that is not sos-convex, see Ahmadi and Parillo [**AP09**]. There also exist convex homogeneous polynomials that are not sums of squares, even though not a single explicit example of such a polynomial is known (see Blekherman [**Bl09**]).

Example 5.7. Consider the TV-screen $C = \mathcal{S}(g)$, where $g = 1 - x^4 - y^4$. The Hessian of g is

$$D^2 g = \begin{bmatrix} -12x^2 & 0 \\ 0 & -12y^2 \end{bmatrix},$$

so g is sos-concave. Thus Thm. 5.4 implies that the Lasserre relaxation of C with respect to g becomes exact. (Note that $M(g)$ is a preordering and therefore archimedean.)

Exercise 5.6. Work out the construction in the proof of Thm. 5.4 for the TV-screen. In which degree does the Lasserre relaxation become exact?

Next, we present a more sophisticated version of Thm. 5.4 in which the defining polynomials are not required to be sos-concave. Instead, we will asssume that the concavity is strict, at least along the extreme part of the boundary, with a uniform lower bound on the curvature. The proof makes use of Putinar's theorem for matrix polynomials.

Theorem 5.8. *Let $C = \mathcal{S}(g_1, \dots, g_r)$ be an archimedean description of a compact convex set with non-empty interior. Suppose that the following condition is satisfied:*

[Concavity] *The function g_i is concave on C and if $u \in C$ is in the closure of the set of extreme points of C with $g_i(u) = 0$, then $D^2 g_i(u)$ is negative definite.*

Then C possesses an exact Lasserre relaxation with respect to $g_1, \dots, g_r$.

Proof. For any $1 \leqslant i \leqslant r$, let Z_i be the closure of the set of extreme points of C at which g_i vanishes. For $u \in \mathbb{R}^n$ write

$$A_u^{(i)}(x) = \int_0^1 \int_0^t -D^2 g_i(u + s(x-u)) \mathrm{d}s\mathrm{d}t \in \mathrm{Sym}_n \mathbb{R}[x].$$

Since $-D^2 g_i(u) > 0$ for all $u \in Z_i$ and $-D^2 g_i(v) \geqslant 0$ for $v \in C$ by hypotheses, it follows from linearity of integration that $A_u^{(i)}(v) > 0$ for all $v \in C$ and $u \in Z_i$. Thus by compactness of C and Z_i, there exists $\delta > 0$ with $A_u^{(i)}(v) \geqslant \delta I_n$ for all $u \in Z_i$, $v \in C$. We may therefore apply Putinar's theorem for matrix polynomials (Thm. 4.2) and obtain representations

$$A_u^{(i)}(x) = \sum_{j=0}^{r} g_j S_{j,u}^{(i)}$$

where each $S_{j,u}^{(i)} \in \mathrm{Sym}_n \mathbb{R}[x]$ is a sum of squares of degree bounded by

$$\deg(S_{j,u}^{(i)}) \leqslant D(\underline{g}, \deg(A_u^{(i)}), \|A_u^{(i)}\|, \delta).$$

Now we again use compactness of Z_i to make the bound independent of u. By taking $\|A^{(i)}\| = \max\{\|A_u^{(i)}\| \mid u \in Z_i\}$ and noting that $\deg(A_u^{(i)}) \leqslant \deg g_i$, we have

$$\deg(S_{j,u}^{(i)}) \leqslant D(\underline{g}, \max\{\deg(g_i)\}, \|A^{(i)}\|, \delta).$$

Now, as in the proof of Thm. 5.4, let $\ell \in \mathbb{R}[x]_1$ with $\ell|_C \geqslant 0$ and let $u \in C$ be a minimiser, which we may assume to be an extreme point of C (see Exercise 5.3). Again fix Lagrange multipliers $\lambda_1, \dots, \lambda_r \geqslant 0$ for ℓ at u (Cor. 5.3), so that

$$\begin{aligned} \nabla \ell(u) &= \sum_{i=1}^{r} \lambda_i \nabla g_i(u) \\ \lambda_i g_i(u) &= 0 \qquad\qquad \text{for all } i = 1, \dots, r. \end{aligned}$$

Then using identity (5.6), we obtain a representation

$$\ell = \ell(u) + \sum_{i=1}^{r} \lambda_i g_i + \sum_{i=1}^{r} \sum_{j=0}^{r} \big(\lambda_i (x-u)^T S_{j,u}^{(i)} (x-u)\big) g_j$$

in which the degrees are independent of u and hence of ℓ. □

Since not every strictly concave polynomial is sos-concave, it is clear that Thm. 5.8 can be applied to examples in which Thm. 5.4 fails. In fact, one can even show that *most* concave polynomials (in a suitable sense) are not sos-concave (see Blekherman [**Bl09**]). But explicit examples of such polynomials (as in [**AP09**]) are not easy to come by.

On the other hand, the assumption that the defining polynomials $\underline{g}$ be strictly concave on $\mathcal{S}(\underline{g})$ is still quite restrictive. We can further weaken the hypotheses if we make use of our freedom in choosing the defining polynomials $\underline{g}$ of the basic closed set $\mathcal{S}(\underline{g})$.

Definition 5.9. A twice continuously differentiable function $f:\mathbb{R}^n \to \mathbb{R}$ is called **strictly quasi-concave** at a point $u \in \mathbb{R}^n$ if the Hessian $D^2 f(u)$ is negative definite on the algebraic tangent space $\{v \in \mathbb{R}^n \mid \langle \nabla f(u), v\rangle = 0\}$, i.e. if

$$\text{For all } v \in \mathbb{R}^n \setminus \{0\}: \quad \langle \nabla f(u), v\rangle = 0 \implies v^T (D^2 f)(u) v < 0.$$

The definition is simple enough but not very intuitive. A more natural definition of (non-strict) quasi-concavity is contained in the following exercise.

Exercise 5.7. Let $C \subset \mathbb{R}^n$ be convex. A $\mathcal{C}^2$-function $f:\mathbb{R}^n \to \mathbb{R}$ is called *quasi-concave* on C if all its sublevel sets $C_a = \{u \in C \mid f(u) \geqslant a\}$, for $a \in \mathbb{R}$, are convex. Show that a strictly quasi-concave function, as defined above, is quasi-concave.

Example 5.10. The polynomial $f = xy$ is strictly quasi-concave on the open quadrant $(0,\infty) \times (0,\infty)$. Indeed, we compute $\nabla f = (y,x)^T$ and

$$D^2 f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Then given $(u,v) \in \mathbb{R}^2$, we have $\big((\nabla f)(u,v)\big)^\perp = \operatorname{span}(-u,v)$ and the restriction of $D^2 f$ to that line is $-2uv$, which is negative for $u,v > 0$. On the other hand, $D^2 f$ is constant and indefinite, so that f is not concave anywhere.

Lemma 5.11. *A twice continuously differentiable function $f:\mathbb{R}^n \to \mathbb{R}$ is strictly quasi-concave at a point $u \in \mathbb{R}^n$ if and only if there exists $M \geqslant 0$ such that*

$$D^2 f(u) - M \cdot \nabla f(u) \nabla f(u)^T < 0.$$

Proof. Put $v_0 = \nabla f(u)$ and $A = D^2 f(u)$ and suppose there exists M as above. Then given $v \in \mathbb{R}^n$ with $v_0^T v = 0$, we have $v^T A v = v^T A v - M v^T v_0 v_0^T v < 0$. Conversely, suppose that f is strictly quasi-concave at u. If $v_0 = 0$, quasi-concavity implies $A < 0$ and there is nothing to show, so assume $v_0 \neq 0$. Since A is negative definite and hence non-degenerate as a bilinear form on the subspace $V = \operatorname{span}(v_0)^\perp$, it admits an orthogonal complement, i.e. there exists $w \in \mathbb{R}^n$ such that $\mathbb{R}^n = V \oplus \operatorname{span}(w)$ and $v^T A w = w^T A v = 0$ for all $v \in V$. Then for any vector $v + \lambda w \in \mathbb{R}^n$, with $v \in V$ and $\lambda \in \mathbb{R}$, and for any $M \in \mathbb{R}$, we compute

$$(v + \lambda w)^T (A - M v_0 v_0^T)(v + \lambda w) = v^T A v + \lambda^2 (w^T A w - M \langle w, v_0\rangle^2).$$

So we choose $M \geqslant 0$ such that $M\langle w, v_0\rangle^2 > w^T A w$. (Note that $\langle w, v_0\rangle \neq 0$, since $w \notin V$.) □

Lemma 5.12 ([**Kr12**], Lemma 4.2.3). *For any $M > 0$, there exists a polynomial h in one variable that is a sum of squares and satisfies the following for all $t \in [-1,1]$:*

(1) $h(t) > 0$
(2) $h(t) + h'(t)t > 0$
(3) $\frac{2h'(t) + h''(t)t}{h(t) + h'(t)t} \leq -M$ □

Exercise 5.8. Give a proof of Lemma 5.12. *Suggestion:* Show first that

$$f(t) = \frac{1 - e^{-(M+1)t}}{(M+1)t}$$

is a $\mathcal{C}^2$-function with the desired properties.

Proposition 5.13. *Let $S = \mathcal{S}(g_1, \dots, g_r)$ be an archimedean description of a compact set and assume that the polynomials g_i are strictly quasi-concave on S. Then there exist $h_0, \dots, h_r \in M(g_1, \dots, g_r)$ which are strictly concave on S such that $S = \mathcal{S}(h_0, \dots, h_r)$.*

Proof. Choose $R > 0$ with $S \subset B_R(0)$. After rescaling, we may assume $g_i(B_R(0)) \subset [-1, 1]$ for $i = 1, \dots, r$. Now let $M \geqslant 0$, let $h \in \mathbb{R}[t]$ be as in Lemma 5.12 above and put

$$h_0 = R^2 - \sum_{i=1}^n x_i^2 \qquad \text{and} \qquad h_i = g_i \cdot h(g_i) \text{ for } i = 1, \dots, r.$$

Then $S \subset \mathcal{S}(\underline{h})$ is clear. Conversely, let $u \in \mathbb{R}^n \setminus S$. If $u \notin B_R(0)$, then $h_0(u) < 0$, hence $u \notin \mathcal{S}(\underline{h})$. If $u \in B_R(0) \setminus S$, then $-1 \leqslant g_i(u) < 0$ for some i and hence $(h(g_i))(u) = h(g_i(u)) > 0$, which implies $h_i(u) < 0$. Thus we have shown $S = \mathcal{S}(\underline{h})$. Also, since h is a sum of squares, so is $h(g_i)$, which implies $h_1, \dots, h_r \in M(\underline{g})$. Since $M(\underline{g})$ is archimedean, we also have $h_0 \in M(\underline{g})$.

Now we need to make sure that $h_0, \dots, h_r$ are strictly concave on S. The polynomial h_0 is everywhere strictly concave, since $D^2 h_0 = -I_n$. For the others, we compute

$$\begin{aligned} D^2 h_i &= \big(h(g_i) + h'(g_i) g_i\big) D^2 g_i + \big(2h'(g_i) + h''(g_i) g_i\big) \nabla g_i \cdot (\nabla g_i)^T \\ &= \big(h(g_i) + h'(g_i) g_i\big) \left(D^2 g_i + \frac{2h'(g_i) + h''(g_i) g_i}{h(g_i) + h'(g_i) g_i} \nabla g_i \cdot (\nabla g_i)^T \right) \\ &\leqslant \big(h(g_i) + h'(g_i) g_i\big)\big(D^2 g_i - M \cdot \nabla g_i \cdot (\nabla g_i)^T\big). \end{aligned}$$

Thus using property (2) of h and applying Lemma 5.11 to g_i, we can make $D^2 h_i$ negative definite at all points of S for $i = 1, \dots, r$, by choosing M sufficiently large. □

We sum up the results of this chapter in the following corollary, combining the various conditions in a single statement.

Corollary 5.14. *Let $C = \mathcal{S}(g_1, \dots, g_r)$ be an archimedean description of a compact convex set with non-empty interior and suppose that for each $i = 1, \dots, r$ one of the following two conditions is satisfied:*

[SOS-Concavity] *The matrices $-D^2 g_i \in \mathrm{Sym}_n \mathbb{R}[x]$ are sums of squares for $i = 1, \dots, r$.*
[Quasi-Concavity] *The function g_i is strictly quasi-concave on C.*

Then C possesses an exact Lasserre relaxation with respect to $g_1, \dots, g_r$.

Proof. Suppose that $g_1, \dots, g_q$ are sos-concave and $g_{q+1}, \dots, g_r$ are strictly quasi-concave, for some $1 \leqslant q \leqslant r$. Since $M(\underline{g})$ is archimedean, we have $R - \sum_{i=1}^n x_i^2 \in M(\underline{g})$ for some $R \geqslant 0$. By Prop. 5.13, we can replace $R - \sum_{i=1}^n x_i^2, g_{q+1}, \dots, g_r$ by strictly concave polynomials in $M(\underline{g})$ defining the same set and obtain a new description $C = \mathcal{S}(h_1, \dots, h_s)$ with $h_1, \dots, h_s \in M(\underline{g})$ such that each h_i for $i = 1, \dots, s$ is either sos-concave or strictly concave on C. Now the arguments in the proof of 5.4 and 5.8 can be combined by a simple case distinction between the sos-concave and strictly concave defining polynomials to show that C has an exact Lasserre relaxation with respect to $\underline{h}$ and therefore with respect to $\underline{g}$. □

Example 5.15. Let $g_1 = xy - 1$, $g_2 = 2 - x - y$ and $g_3 = x + y$. Since g_1 is strictly quasi-concave on $(0, \infty) \times (0, \infty)$ and g_2, g_3 are linear and therefore sos-concave, the basic closed set $\mathcal{S}(\underline{g})$ possesses an exact Lasserre relaxation with respect to $\underline{g}$. (It is in fact a spectrahedron).

Example 5.16. The statement in Cor. 5.14 is still not the best possible. Let $g_1 = y - x^3$, $g_2 = x$, $g_3 = 1 - x$, $g_4 = y$, $g_5 = 1 - y$ in variables x, y. We saw in Example 3.20 that the third Lasserre relaxation of $S = \mathcal{S}(\underline{g})$ with respect to $\underline{g}$ is exact. But the polynomial $g_1 = y - x^3$ has Hessian

$$D^2 g_1 = \begin{bmatrix} -6x & 0 \\ 0 & 0 \end{bmatrix}$$

and is therefore not strictly quasi-concave at the origin. Therefore, the exactness of the Lasserre relaxation does not follow from Cor. 5.14. We will revisit this example later.

REFERENCES

[AP09] A. A. Ahmadi and P. Parrilo. A convex polynomial that is not sos-convex. *Math. Program.*, **135**(1), p. 275–292, 2012. http://arxiv.org/abs/0903.1287

[Bl09] G. Blekherman. Convex forms that are not sums of squares. Preprint, 2009. http://arxiv.org/abs/0910.0656

[HN10] J. W. Helton and J. Nie. Semidefinite representation of convex sets. *Math. Program.* **122**(1), p. 21–64, 2010. http://arxiv.org/abs/0705.4068

[Kr12] T. Kriel. Semidefinite representability of closed semi-algebraic sets (in German). Bachelorarbeit, Universität Konstanz, 2012.

[Si10] R. Sinn. Spectrahedra and a relaxation of convex, semialgebraic sets. Diplomarbeit, Universität Konstanz, 2010. http://www.math.uni-konstanz.de/~sinn/thesis.pdf

Daniel Plaumann
Universität Konstanz
Summer 2013

6. NECESSARY CONDITIONS FOR EXACTNESS

So far we have only seen sufficient conditions for the exactness of the Lasserre relaxation but not a single example in which it demonstrably fails to become exact. In this chapter, we will fill this gap. The principal obstruction against exactness that we are going to use is contained in the following result by Gouveia and Netzer in [**GN11**, Prop. 4.1].

Proposition 6.1. *Let $S = \mathcal{S}(g_1, \dots, g_r) \subset \mathbb{R}^n$ and let $Z \subset \mathbb{R}^n$ be a line such that $S \cap Z$ has non-empty interior in Z. Assume that there exists a point $u_0 \in S$ in the relative boundary of $\overline{\mathrm{conv}(S)} \cap Z$ such that the gradients $\nabla g_i(u_0)$ are orthogonal to Z whenever $g_i(u_0) = 0$. Then all Lasserre relaxations $\mathcal{L}_{\underline{g}}[d]$ for $d \geqslant 1$ strictly contain $\overline{\mathrm{conv}(S)}$.*

Proof. Let Z and u_0 be as in the hypothesis. After a change of coordinates, we may assume $u_0 = 0$ and $Z = \{u \in \mathbb{R}^n \mid u_2 = \cdots = u_n = 0\}$. We may further assume that $u_1 \geqslant 0$ holds for all $u \in \overline{\mathrm{conv}(S)} \cap Z$. Let $h_i = g_i|_Z = g_i(x_1, 0, \dots, 0) \in \mathbb{R}[x_1]$ for $i = 1, \dots, r$ and consider the Lasserre relaxation $\mathcal{L}_{\underline{h}}[d] \subset Z$ for some $d \geqslant 1$. We have $\mathcal{L}_{\underline{h}}[d] \subset \mathcal{L}_{\underline{g}}[d] \cap Z$. (To see this, let $u \in \mathcal{L}_{\underline{h}}[d]$, i.e. $u = (L(x_1), 0, \dots, 0)$ for some $L \in M_{\underline{h}}[d]' \subset \mathbb{R}[x_1]^*$. Then L extends to $L_0 \in \mathbb{R}[x]^*$ via $L_0(f) = L(f(x_1, 0, \dots, 0))$ for $f \in \mathbb{R}[x]$. We have $L_0 \in M[\underline{g}]'$ since $f(x_1, 0, \dots, 0) \in M_{\underline{h}}[d]$ for any $f \in M_{\underline{g}}[d]$. Thus $u = (L_0(x_1), \dots, L_0(x_n)) \in \mathcal{L}_{\underline{g}}[d]$.)

Now let $c \in \mathbb{R}_+ \cup \{\infty\}$ with $\overline{\mathrm{conv}(S)} \cap Z = [0, c]$ and put $h_{r+1} = c - x_1$ if $c \in \mathbb{R}$ and $h_{r+1} = 1$ otherwise. Then $\mathcal{S}(h_1, \dots, h_{r+1}) = [0, c]$ and $\mathcal{L}_{\underline{h}, h_{r+1}}[d] \subset \mathcal{L}_{\underline{h}}[d]$. By Prop. 3.17, we have $\mathcal{L}_{\underline{h}, h_{r+1}} = [0, c]$ if and only if $M_{\underline{h}, h_{r+1}}[d]$ contains all $\ell \in \mathbb{R}[x_1]_1$ such that $\ell|_{[0,c]} \geqslant 0$. In particular, $\mathcal{L}_{\underline{h}, h_{r+1}} = [0, c]$ would imply $x_1 \in M_{\underline{h}, h_{r+1}}[d]$, so we would have a representation

$$x_1 = \sum_{i \in I} s_i h_i + \sum_{i \notin I} s_i h_i$$

where $h_0 = 1$ and we have split indices by putting $I = \{i \in \{0, \dots, r\} \mid h_i(0) > 0\}$, so that $h_i(0) = 0$ for $i \notin I$. Evaluating at 0, we see that $s_i(0) = 0$ for $i \in I$. So the s_i for $i \in I$ have no constant term and, since they are sums of squares, they have no linear term either. Now by hypothesis, the gradient $\nabla g_i(0)$ is orthogonal to Z for all $i \in \{1, \dots, r+1\} \setminus I$, which implies that h_i for $i \notin I$ also has no constant and no linear term. This is a contradiction, since x_1 is linear, so we conclude that no such representation of x_1 exists. It follows that $\mathcal{L}_{\underline{h}, h_{r+1}}[d]$ strictly contains $[0, c]$, hence $\mathcal{L}_{\underline{g}}[d]$ strictly contains $\overline{\mathrm{conv}(S)}$. □

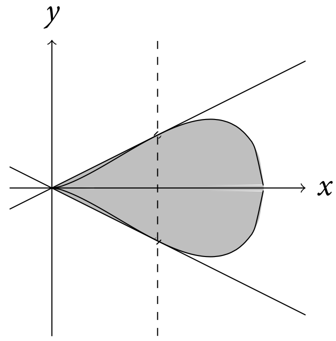
Corollary 6.2. *Let $S = \mathcal{S}(\underline{g})$ be basic closed with non-empty interior and suppose there exists $u_0 \in \left(\partial \overline{\mathrm{conv}(S)}\right) \cap S$ such that $\nabla g_i(u_0) = 0$ whenever $g_i(u_0) = 0$. Then all Lasserre relaxations $\mathcal{L}_{\underline{g}}[d]$ for $d \geqslant 1$ strictly contain $\overline{\mathrm{conv}(S)}$.*

Proof. Apply Prop. 6.1 to any line L through u_0 and $\mathrm{int}(S)$. □

Example 6.3.
Let $g = x^3(1-x) - y^2$ and put $S = \mathcal{S}(g)$. The origin is a singular point of $\mathcal{V}(g)$, i.e. $\nabla g(0) = 0$. So no Lasserre relaxation of $\operatorname{conv}(S)$ with respect to g is exact.
But a quick computation shows that

$$\operatorname{conv}(S) = \mathcal{S}(g, 2x-1) \cup \mathcal{S}(2y+x, x-2y, 1-2x)$$

which is not basic closed. The second set in the union is just a triangle while the other possesses an exact Lasserre relaxation by Cor. 5.14 (see Exercise 6.1 below).



Exercise 6.1. Show that $\operatorname{conv}(S)$ in the above example is a projected spectrahedron.

The necessary condition for exactness in Corollary 6.2 depends on the description of the basic closed set S. But there is also a more intrinsic geometric condition.

Let $C \subset \mathbb{R}^n$ be closed and convex. A **face** of C is a convex subset $F \subset C$ with the property that $\frac{1}{2}(u+v) \in F$ implies $u, v \in F$ whenever $u, v \in C$. A face F is called *proper* if $F \neq \emptyset, C$. It is called **exposed** if it is cut out by a supporting hyperplane, i.e. if there exists $\ell \in \mathbb{R}[x]_1$ with $\ell|_C \geqslant 0$ and $F = \{u \in \mathbb{C} \mid \ell(u) = 0\}$. Otherwise, F is called **non-exposed.**[1]

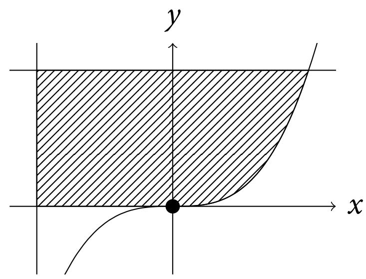
Exercise 6.2. Check that if $\ell \in \mathbb{R}[x]_1$ with $\ell|_C \geqslant 0$, then $\{u \in C \mid \ell(u) = 0\}$ is indeed a face of C.

Example 6.4. The faces of a polyhedron $\mathcal{S}(\ell_1, \dots, \ell_r)$ are exactly the vertices, edges, etc., defined by the vanishing of a subset of $\ell_1, \dots, \ell_r$ and are always exposed.

Example 6.5.
The convex set in the picture is the basic closed set

$$C = \mathcal{S}(y - x^3, 1 + x, y, 1 - y).$$

The origin is a non-exposed face of C, since the only linear polynomial $\ell \in \mathbb{R}[x]_1$ with $\ell|_C \geqslant 0$ and $\ell(0,0) = 0$ is $\ell = y$, which exposes the larger face $\{(u, 0) \in \mathbb{R}^2 \mid -1 \leqslant u \leqslant 0\}$ of C rather than just the origin.



Exercise 6.3. Show that no proper face of a closed convex set C contains a point of $\operatorname{relint}(C)$.

Exercise 6.4. Let C be closed and convex. Show that a convex subset F of C is a face if and only if $C \setminus F$ is convex and any convex subset of C containing F has strictly greater dimension than F.

Proposition 6.6. *Let $C \subset \mathbb{R}^n$ be closed and convex and let F be a face of C.*

(1) *Every face of F is also a face of C.*
(2) *F is closed.*
(3) *If $u \in \operatorname{relint}(F)$ and $\ell \in \mathbb{R}[x]_1$ with $\ell_C \geqslant 0$ and $\ell(u) = 0$, then $\ell|_F = 0$.*
(4) *Any point in the relative boundary of C is contained in a proper exposed face of C.*
(5) *For any point $u \in C$ there exists a unique face of C containing u in its relative interior. This is precisely the smallest face of C containing u.*
(6) *F is exposed if and only if for every face F' strictly containing F there is $\ell \in \mathbb{R}[x]_1$ with $\ell|_C \geqslant 0$ such that $\ell|_F = 0$ but $\ell|_{F'} \neq 0$.*

[1]Unfortunately, the terminology here is not uniform in the literature. It is equally common to call *face* what we call *exposed face* and use another term (e.g. *facelet* or *extremal convex subset*) for what we call *face*.

Proof. (1)–(3) Exercise.

(4) Let $u \in C \smallsetminus \mathrm{relint}(C)$. By the general separation theorem, there exists $\ell \in \mathbb{R}[x]_1$ with $\ell(u) = 0$ and $\ell > 0$ on $\mathrm{relint}(C)$. Thus $\{u \in C \mid \ell(u) = 0\}$ is the face we want. (See also [**Barvinok**, Thm. 2.7] or [**Convexity-LN**, Satz 5.1].)

(5) Let F be the intersection of all faces containing u. Then F is a face of C, and therefore obviously the smallest face containing u. If u were not contained in the relative interior of F, it would be contained in a proper face of F, which would also be a face of C, a contradiction. If F' is another face containing u, then $F \subset F'$ by definition of F. So if $F \subsetneq F'$, then u cannot be contained in the relative interior of F'.

(6) If F is exposed, there exists such ℓ by definition. Conversely, let F be a face of C with this property and consider $\mathcal{N} = \{\ell \in \mathbb{R}[x]_1 \mid \ell|_C \geqslant 0 \text{ and } \ell|_F = 0\}$. We choose a sequence $(U_k)_{k\geqslant 1}$ of open subsets in $\mathbb{R}^n$ such that $F = \bigcap_{k\geqslant 1} U_k$ and $\mathbb{R}^n \smallsetminus U_k$ is compact for every $k \geqslant 1$. (Explicitly, we may take $U_k = \bigcup_{u\in F} B_{\frac{1}{k}}(u) \cup (\mathbb{R}^n \smallsetminus \overline{B}_k(0))$.) Given a point $u \in C \smallsetminus F$, we can pick $\ell_u \in \mathcal{N}$ with $\ell_u(u) > 0$. For if we take just any $\ell \in \mathcal{N}$, then either $\ell(u) > 0$ or u is contained in the relative interior of some face containing F and there exists $\ell_u \in \mathcal{N}$ with $\ell_u(u) > 0$ by hypothesis. Now since $C \smallsetminus U_k$ is compact, we can find $u_1, \ldots, u_m \in C \smallsetminus U_k$ such that $\ell_k = \sum_{i=1}^m \ell_{u_i}$ is strictly positive on $C \smallsetminus F$. Then the convergent series

$$\ell = \sum_{k=1}^{\infty} \frac{\ell_k}{2^k \|\ell_k\|}$$

(where $\|\cdot\|$ is some norm on $\mathbb{R}[x]_1$) determines an element of $\mathcal{N}$ that exposes F. □

Exercise 6.5. Show that any maximal proper face of a closed convex set is exposed.

Theorem 6.7. *All faces of a spectrahedron are exposed.*

Proof. Consider first the psd cone Sym_k^+ itself. If $A \in \mathrm{Sym}_k^+$ is any psd matrix, it is not hard to show that the unique face of Sym_k^+ containing A in its relative interior is

$$F = \left\{B \in \mathrm{Sym}_k^+ \mid \ker A \subset \ker B\right\}.$$

(See [**Barvinok**, §II.12] or [**Convexity-LN**, Satz 6.4].) To see that such F is an exposed face, let C be a psd matrix with $\mathrm{im}\, C = \ker A$. (To prove the existence of such C, let $r = \mathrm{rk}\, A$ and choose an orthogonal matrix $U \in \mathrm{GL}_k$ such that $U^T A U$ is the diagonal matrix with the first r diagonal entries the non-zero eigenvalues of A and the remaining equal to 0. Let J be the diagonal matrix with the first r entries equal to 0 and the remaining equal to 1 and put $C = UJU^T$. Then $AC = 0$ and $\ker A \cap \ker C = \{0\}$, hence $\mathrm{im}\, C = \ker A$, as desired.) Now given $B \in \mathrm{Sym}_k^+$ with $\langle B, C\rangle = 0$, we have $BC = 0$ (since C is also psd) and hence $\ker A = \mathrm{im}\, C \subset \ker B$. This shows $F = \{B \in \mathrm{Sym}_k^+ \mid \langle B, C\rangle = 0\}$, so that F is exposed. (In the above notation, $\ell = \langle X, C\rangle$ with $X = (X_{ij})_{i\geqslant j}$ is the polynomial in $\mathbb{R}[X]_1$ exposing F.)

Now if $U \subset \mathrm{Sym}_k$ is an affine-linear subspace and F a face of the spectrahedron $U \cap \mathrm{Sym}_k^+$, let $A \in \mathrm{relint}(F)$ and let F' be the unique face of Sym_k^+ with $A \in \mathrm{relint}(F')$. Then $F' \cap U$ contains A in its relative interior, hence $F' \cap U = F$. By what we have just seen, $F' = \{B \in \mathrm{Sym}_k^+ \mid \langle B, C\rangle = 0\}$ for some $C \in \mathrm{Sym}_k^+$, so that $F = \{B \in U \cap \mathrm{Sym}_k^+ \mid \langle B, C\rangle = 0\}$.

In general, if $\varphi\colon \mathbb{R}^n \to \mathrm{Sym}_k$ is an affine linear map, the faces of $\varphi^{-1}(\mathrm{Sym}_k^+)$ are in bijection with the faces of $S = \mathrm{im}\,\varphi \cap \mathrm{Sym}_k^+$, and if H is a hyperplane in Sym_k exposing a face $H \cap S$, the corresponding face $\varphi^{-1}(H \cap S) = \varphi^{-1}(H) \cap \varphi^{-1}(S)$ of $\varphi^{-1}(S)$ is also exposed. □

Exercise 6.6. Fill in the details in the above proof.

This statement does not extend to projected spectrahedra since projections of exposed faces need not be exposed (see Example 6.9 below). However, an exact Lasserre relaxation of a convex basic closed set with respect to its describing inequalities cannot have any non-exposed faces. This was first shown in [**NPS10**]. It can also be easily deduced from Prop. 6.1.

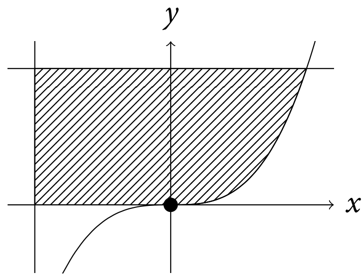
Exercise 6.7. Find an example of a speactrahedron in $\mathbb{R}^3$ whose projection onto the first two coordinates has a non-exposed face.

Theorem 6.8. *Let $C = \mathcal{S}(\underline{g})$ be convex with non-empty interior. If C has a non-exposed face, then no Lasserre relaxation of C with respect to $\underline{g}$ is exact.*

Proof. Let $F \subset C$ be a non-exposed face. By Prop. 6.6(6), there exists some face F' of C containing F such that $\ell|_{F'} = 0$ for all $\ell \in \mathbb{R}[x]_1$ with $\ell|_F = 0$ and $\ell|_C \geqslant 0$. Let $u_0 \in \operatorname{relint}(F)$ and $u_1 \in \operatorname{relint}(F;)$, and let Z be the line through u_0 and u_1. Then $C \cap Z$ has non-empty interior in Z and u_0 is a point in the relative boundary of $C \cap Z$. So we can apply Prop. 6.1 once we have verified the condition on the gradients. Suppose that $g_i(u_0) = 0$ for some i. Since C is convex and $g_i|_C \geqslant 0$, we must have $\langle \nabla g_i(u_0), u - u_0 \rangle \geqslant 0$ for all $u \in C$. In other words, the polynomial $\ell = \langle \nabla g_i(u_0), x - u_0 \rangle \in \mathbb{R}[x]_1$ is non-negative on C. Since $u_0 \in \operatorname{relint}(F)$ and $\ell(u_0) = 0$, we must have $\ell|_F = 0$ and hence also $\ell|_{F'} = 0$. So ℓ vanishes in u_0 and u_1 and thus on all of Z, which means that $\nabla g_i(u_0)$ is orthogonal to Z. □

Example 6.9.
Returning to the convex set

$$C = \mathcal{S}(y - x^3, 1 + x, y, 1 - y).$$



from Example 6.5, we conclude from Thm. 6.8 that C cannot have an exact Lasserre relaxation with respect to the defining polynomials $y - x^3, 1 + x, y, 1 - y$, since the origin is a non-exposed face. However, C is the convex hull of the face $\{-1\} \times [0, 1]$ and $C \cap \mathcal{S}(x)$, both of which are projected spectrahedra. This is clear for the first, while for the second we gave an explicit proof in Example 3.20.

Remarks 6.10.

(1) Note that Thm. 6.8 only applies when $\mathcal{S}(\underline{g})$ is already convex. It may in fact happen that $\operatorname{conv}(\mathcal{S}(\underline{g}))$ has an exact Lasserre relaxation with respect to $\underline{g}$ even when $\operatorname{conv}(S)$ has a non-exposed face. For example, Gouveia and Netzer show for the football stadium $\operatorname{conv}(\mathcal{S}(g))$ defined by $g = -((x+1)^2 + y^2 - 1)((x-1)^2 + y^2 - 1)$ (cf. Exercise 1.10) that $\operatorname{conv}(\mathcal{S}(g)) = \mathcal{L}_g[6]$ (see [**GN11**, Prop. 4.12]).

(2) Furthermore, Thm. 6.8 only applies when the Lasserre relaxation is indeed exact. An inexact Lasserre relaxation of a convex basic closed set may well have non-exposed faces. For example, the set in Example 6.9 has a Lasserre relaxation with a non-exposed face [**GN11**, Cor. 4.11].

REFERENCES

[GN11] J. Gouveia and T. Netzer. Positive polynomials and projections of spectrahedra. *SIAM J. Optimization*, **21**(3), p. 960–976, 2011. http://arxiv.org/abs/0911.2750

[NPS10] T. Netzer, D. Plaumann, M. Schweighofer. Exposed faces of semidefinitely representable sets. *SIAM J. Optimization*, **20**(4), p. 1944–1955, 2010. http://arxiv.org/abs/0902.3345

7. HYPERBOLIC POLYNOMIALS

In this chapter, we return to the study of spectrahedra, rather than projected spectrahedra, and their connection with hyperbolic and determinantal polynomials. The main goal is the celebrated Helton-Vinnikov theorem, which provides a complete characterisation of the two-dimensional convex semialgebraic sets that can be represented by linear matrix inequalities, i.e. a characterisation of the plane spectrahedra.

It turns out to be technically more convenient to deal with cones rather than general convex sets. However, everything we show about convex cones and spectrahedral cones has an analogue for convex sets and spectrahedra that can be obtained by taking slices.

7.1. HYPERBOLICITY

Definition 7.1. A homogeneous polynomial $f \in \mathbb{R}[x]$ is called **hyperbolic** with respect to a point $e \in \mathbb{R}^n$ if $f(e) \neq 0$ and if the univariate polynomial $f(u + te) \in \mathbb{R}[t]$ has only real roots, for every $u \in \mathbb{R}^n$. It is called **strictly hyperbolic** if it is hyperbolic and the roots of $f(u + te)$ are all distinct, for every $u \in \mathbb{R}^n$, $u \notin \mathbb{R} \cdot e$.

Examples 7.2.

(1) The polynomial $f = x^2 + y^2 - z^2 \in \mathbb{R}[x, y, z]$ is (strictly) hyperbolic with respect to $e = (0, 0, 1)$, since $f(u + te) = u_1^2 + u_2^2 - (u_3 + t)^2$ has discriminant $4(u_1^2 + u_2^2) > 0$ and therefore two distinct real roots in t, for all $u = (u_1, u_2, u_3) \in \mathbb{R}^3 \setminus \mathbb{R} \cdot e$.

(2) The polynomial $f = x^4 + y^4 - z^4 \in \mathbb{R}[x, y, z]$ is not hyperbolic with respect to any point in $\mathbb{R}^3$. (This is the homogenized version of the polynomial defining the TV screen; see Example 1.11). In particular, for $e = (0, 0, 1)$, one can check that $f(u + te)$ has two real but also a pair of non-real complex-conjugate roots.

(3) The determinant $\det(X)$ of a general symmetric $d \times d$-matrix $X = (X_{ij})_{1 \leqslant i \leqslant j \leqslant d}$, regarded as a polynomial on Sym_d and thus an element of $\mathbb{R}[X_{ij} \mid 1 \leqslant i \leqslant j \leqslant d]$, is hyperbolic with respect to the point $e = I_d$. This is because for any $A \in \mathrm{Sym}_d$, the roots of $\det(A - tI_d)$ are exactly the eigenvalues of A. In particular,

$$\mathrm{Sym}_d^+ = \left\{A \in \mathrm{Sym}_d \mid \det(A - tI_d) \text{ has only non-negative roots}\right\}.$$

(4) Let $A = \sum_{i=1}^n x_i A_i$ be a homogeneous linear matrix polynomial of size $d \times d$ and suppose there exists $e \in \mathbb{R}^n$ with $A(e) = I_d$. Then $f = \det(A) \in \mathbb{R}[x]$ is homogeneous of degree d and hyperbolic with respect to e. Again, this is because $f = (u - te) = \det(A(u) - tI_d)$ is the characteristic polynomial of the symmetric matrix $A(u)$ and therefore has only real roots. More generally, the same remains true if

$A(e)$ is just any positive definite matrix, by considering $\sqrt{A(e)}^{-1} \cdot A(x)\sqrt{A(e)}^{-1}$. Note that the spectrahedral cone $\mathcal{S}(A)$ can also be expressed in terms of f as

$$\begin{aligned}\mathcal{S}(A) &= \{u \in \mathbb{R}^n \mid A(u) \succeq 0\} \\ &= \{u \in \mathbb{R}^n \mid f(u - te) \text{ has only non-negative roots}\}.\end{aligned}$$

In view of these examples, we make the following definition.

Definition 7.3. Let $f \in \mathbb{R}[x]$ be hyperbolic with respect to e. The set

$$\overline{C}_e(f) = \{u \in \mathbb{R}^n \mid f(u - te) \text{ has only non-negative roots}\}$$

is called the (closed) **hyperbolicity cone** of f with respect to e.

In this sense, one can think of hyperbolic polynomials as generalised characteristic polynomials and the hyperbolicity cones as generalised psd cones. But in spite of the name, it is not apparent from the definition that $\overline{C}_e(f)$ is indeed a convex cone. Also, the name suggests that f should be hyperbolic with respect to any point in the (interior of the) hyperbolicity cone. These statements can be proved directly (see for example [**Convexity-LN**, §13]). Instead, we will deduce them later from the Helton-Vinnikov theorem.

In example 7.2(4), we have already seen the following.

Proposition 7.4. *If A is a homogeneous linear matrix polynomial with $A(e) \succ 0$ for some $e \in \mathbb{R}^n$, the spectrahedral cone $\mathcal{S}(A)$ coincides with the hyperbolicity cone $\overline{C}_e(\det(A))$.* □

The hyperbolicity condition is therefore necessary for a cone to be spectrahedral. Thus the following statement is not surprising, though a little additional work is needed for the proof, which we will omit here (see [**HV07**, §2]).

Theorem 7.5. *Let $C \subset \mathbb{R}^n$ be a closed semialgebraic convex cone with non-empty interior and let $f \in \mathbb{R}[x] \setminus \{0\}$ be the unique polynomial of minimal degree vanishing on the boundary of C. Then C is a hyperbolicity cone if and only if f is hyperbolic with respect to some point $e \in \operatorname{int}(C)$ and, in this case, $C = \overline{C}_e(f)$.* □

For example, this shows that the cone $\{u \in \mathbb{R}^3 \mid u_1^4 + u_2^4 \leqslant u_3^4\}$ (the cone over the TV screen) is not spectrahedral, since the polynomial $x^4 + y^4 - z^4$ is not hyperbolic (c.f. 7.2(2)).

Given a homogeneous polynomial $f \in \mathbb{R}[x]$ and a symmetric linear matrix polynomial A such that $f = \det(A)$, we say that A is a **symmetric (linear) determinantal representation** of f. If, in addition, f is hyperbolic with respect to e and $A(e) \succ 0$, we say that the determinantal representation is **definite**.

Exercise 7.1. Show that every hyperbolic polynomial in two variables possesses a definite symmetric determinantal representation.

But it is in fact not hard to see that not every hyperbolic polynomial possesses a symmetric determinantal representation.

Proposition 7.6. *If $n \geqslant 4$ and $d \geqslant 7$, there exist hyperbolic polynomials in n variables of degree d that do not possess a symmetric determinantal representation.*

Proof. The set of hyperbolic polynomials of degree d has non-empty interior in $\mathbb{R}[x]_{(d)}$, the space of homogeneous polynomials of degree d. Indeed, every strictly hyperbolic polynomial is an interior point of that set, since the roots of a univariate polynomial depend continuously on the coefficients (see Exercise 7.2). The dimension of the vector space $\mathbb{R}[x]_{(d)}$ is $\binom{n+d-1}{d}$. On the other hand, if $f \in \mathbb{R}[x]_{(d)}$ has a symmetric determinantal representation $f = \det(A)$, then A must be of size $d \times d$ and the space of homogeneous linear matrix polynomials of size d in n variables has dimension $n\binom{d+1}{2}$. The map taking A to $\det(A)$ is polynomial, so the dimension of the image cannot increase. It follows that if every hyperbolic polynomial is to possess a symmetric determinantal representation, we must have

$$n\binom{d+1}{2} = \frac{n(d+1)d}{2} \geqslant \frac{(n+d-1)!}{d!(n-1)!} = \binom{n+d-1}{d}.$$

This is equivalent to $n!(d+1)!d \geqslant 2(n+d-1)!$. Now one can check directly that this inequality fails to hold for $n \geqslant 4$ and $d \geqslant 7$. $\square$

Exercise 7.2. Fix $e \in \mathbb{R}^n$ and denote by $\mathcal{H}_e \subset \mathbb{R}[x]_{(d)}$ the set of hyperbolic polynomials of degree d with respect to e.

(a) Show that every strictly hyperbolic polynomial of degree d is an interior point of $\mathcal{H}_e$.
(b) Show that the strictly hyperbolic polynomials are dense in $\mathcal{H}_e$.
Hint: If $f \in \mathbb{R}[t]$ has only real roots, examine the roots of $f + \alpha f'$ for $\alpha \in \mathbb{R}$.

More is true: $\mathcal{H}_e$ is connected, simply connected and conincides with the closure of its interior, which is the set of strictly hyperbolic polynomials. (See Nuij [**Nu68**]).

Exercise 7.3. Examine the inequality in the proof of Prop. 7.6 for $d = 2, 3, 4, 5, 6$.

However, if we do the count of parameters in the above proof for $n = 3$, we find that the resulting inequality holds for all d. In 1957, it was conjectured by Peter Lax, in connection with the study of hyperbolic PDEs in [**La57**], that every hyperbolic polynomial in three variables possesses a definite determinantal representation. This became known as the **Lax conjecture**. It was proved in [**HV07**] through the work of Vinnikov and Helton-Vinnikov.

Theorem 7.7 (Helton-Vinnikov). *Every hyperbolic polynomial in three variables possesses a definite symmetric determinantal representation.*

Corollary 7.8. *Every three-dimensional hyperbolicity cone is spectrahedral.*

We will not give a full proof of the Helton-Vinnikov theorem. However, we will prove a weaker version in section 7.3 below that will still imply the above corollary.

Since it is clear from the count in Prop. 7.6 that the Helton-Vinnikov theorem cannot extend to the case $n \geqslant 4$, the search began for a suitable higher dimensional analogue. Various weaker versions have been proposed in recent years some of which have been disproved. Perhaps the most natural generalisation is simply the statement of Cor. 7.8.

Generalised Lax Conjecture. Every hyperbolicity cone is spectrahedral.

A few special cases of the conjecture are known, but in general it remains elusive.

7.2. DEFINITE DETERMINANTAL REPRESENTATIONS AND INTERLACING

To prove the Helton-Vinnikov theorem, it is helpful to understand it as a statement in two parts. The first part amounts to the construction of determinantal representations over $\mathbb{R}$ or $\mathbb{C}$, the second to a characterisation of those determinantal representations that are definite and therefore reflect the hyperbolicity. The following notion will be used to address the second part.

Definition 7.9. Let $f, g \in \mathbb{R}[t]$ be univariate polynomials with $\deg(f) = d$ and $\deg(g) = d-1$ and suppose that f and g have only real roots. Denote the roots of f by $\alpha_1 \leqslant \cdots \leqslant \alpha_d$ and the roots of g by $\beta_1 \leqslant \cdots \leqslant \beta_{d-1}$. We say that g **interlaces** f if $\alpha_i \leqslant \beta_i \leqslant \alpha_{i+1}$ for all $i = 1, \ldots, d-1$. We say that g **strictly interlaces** f, if all these inequalities are strict.

If $f \in \mathbb{R}[x]_{(d)}$ is hyperbolic with respect to $e \in \mathbb{R}^n$, we say that $g \in \mathbb{R}[x]_{(d-1)}$ (strictly) **interlaces** f **with respect to** e if $g(u+te)$ (strictly) interlaces $f(u+te)$ in $\mathbb{R}[t]$ for every $u \in \mathbb{R}^n$, $u \notin \mathbb{R} \cdot e$. Note that this implies that g is hyperbolic with respect to e, as well.

Example 7.10. The simplest and most important example is the following: If $f \in \mathbb{R}[t]$ is a univariate polynomial with only real roots, then its derivative f' interlaces f. More generally, if $f \in \mathbb{R}[x]$ is hyperbolic with respect to $e \in \mathbb{R}^n$, the directional derivative

$$D_e(f) = \frac{\partial}{\partial t} f(x+te)\bigg|_{t=0} = \sum_{i=1}^{n} e_i \frac{\partial}{\partial x_i} f$$

interlaces f, since $f'(u+te) \in \mathbb{R}[t]$ interlaces $f(u+te) \in \mathbb{R}[t]$ for all $u \in \mathbb{R}^n$. If f is strictly hyperbolic, then $D_e f$ strictly interlaces f.

Lemma 7.11. *Suppose that $f \in \mathbb{R}[x]_{(d)}$ is irreducible and hyperbolic with respect to e. Fix g, h in $\mathbb{R}[x]_{(d-1)}$ where g interlaces f with respect to e. Then h interlaces f with respect to e if and only if $g \cdot h$ is non-negative or non-positive on $\mathcal{V}_{\mathbb{R}}(f)$.*

Proof. It suffices to prove this statement for the restriction of gh to a line $\{u + te \mid t \in \mathbb{R}\}$ for generic $u \in \mathbb{R}^n$. Thus we may assume that the roots of $f(u+te)$ are distinct from each other and from the roots of $g(u+te) \cdot h(u+te)$.

Suppose that $g \cdot h$ is non-negative on $\mathcal{V}_{\mathbb{R}}(f)$. By the genericity assumption, the product $g(u+te)h(u+te)$ is positive on all the roots of $f(u+te)$. Between consecutive roots of $f(u+te)$, the polynomial $g(u+te)$ has a single root and thus changes sign. For the product gh to be positive on these roots, $h(u+te)$ must also change sign and have a root between each pair of consecutive roots of $f(u+te)$. Hence h interlaces f with respect to e.

Conversely, suppose that g and h both interlace f. Between any two consecutive roots of $f(u+te)$, both $g(u+te)$ and $h(u+te)$ each have exactly one root, and their product has exactly two. It follows that $g(u+te)h(u+te)$ has the same sign on all the roots of $f(u+te)$. Taking $t \to \infty$ shows this sign to be the sign of $g(e)h(e)$, independent of the choice of u. Hence gh has the same sign at every point of $\mathcal{V}_{\mathbb{R}}(f)$. □

The version of the Helton-Vinnikov theorem that we are going to prove in the next section yields *Hermitian* determinantal representations, rather than symmetric ones. One can make a theory of spectrahedra defined by Hermitian matrices, which may be the natural point of view for certain questions. However, in terms of the class of sets that one obtains, this does not add anything new, essentially due to the following simple observation.

Lemma 7.12. *Let M be a homogeneous Hermitian linear matrix polynomial of size $d \times d$. Then there exists a real symmetric linear matrix polynomial N of size $2d \times 2d$ such that*

$$\{u \in \mathbb{R}^n \mid M(u) \succcurlyeq 0\} = \{u \in \mathbb{R}^n \mid N(u) \succcurlyeq 0\}$$

and $\det(N) = \det(M)^2$.

Proof. Write $M = A + iB$, where A is real symmetric and B is real skew-symmetric, and put $N = \begin{bmatrix} A & B \\ -B & A \end{bmatrix}$. To see that N has the desired property, apply the change of coordinates

$$U^T N \overline{U} = \begin{bmatrix} A - iB & 0 \\ 0 & A + iB \end{bmatrix} \quad \text{where} \quad U = \begin{bmatrix} \frac{1}{\sqrt{2}} \cdot I_d & \frac{i}{\sqrt{2}} \cdot I_d \\ \frac{i}{\sqrt{2}} \cdot I_d & \frac{1}{\sqrt{2}} \cdot I_d \end{bmatrix}.$$

In particular, $\det(N) = \det(\overline{M}) \det(M) = \det(M)^2$. □

Recall that the **adjugate matrix** of a $d \times d$-matrix A (also called *adjoint matrix* or *Cramer matrix*) is the $d \times d$-matrix A^{adj} whose (j,k)-entry is $(-1)^{j+k}$ times the $(d-1) \times (d-1)$-minor of A obtained by deleting the jth row and kth column and taking the determinant. The fundamental fact is *Cramer's rule*, which says that $A \cdot A^{\mathrm{adj}} = \det(A) \cdot I_d$. This is a general indentity holding for matrices with entries in any commutative ring.

If M is a (symmetric or hermitian) homogeneous linear matrix polynomial $d \times d$, its adjugate M^{adj} is, by definition, a homogeneous matrix polynomial of size $d \times d$ and degree $d-1$. The relation between M and M^{adj} will play a crucial role in the next section.

Definition 7.13. Let M be a Hermitian linear matrix polynomial of size $d \times d$. Then

$$\mathcal{C}(M) \;=\; \{\lambda^T M^{\mathrm{adj}} \overline{\lambda} \;\mid\; \lambda \in \mathbb{C}^d \smallsetminus \{0\}\},$$

is a subset of $\mathbb{R}[x]_{(d-1)}$, which we call the **system of hypersurfaces associated with** M.

Here is a useful identity that goes back to the work of Hesse in 1855 [**He1855b**].

Proposition 7.14. *Let M be a Hermitian matrix of linear forms. Then the polynomial*

$$(\lambda^T M^{\mathrm{adj}} \overline{\lambda})(\mu^T M^{\mathrm{adj}} \overline{\mu}) \;-\; (\lambda^T M^{\mathrm{adj}} \overline{\mu})(\mu^T M^{\mathrm{adj}} \overline{\lambda}) \tag{7.15}$$

is contained in the ideal generated by $\det(M)$, *for any* $\lambda, \mu \in \mathbb{C}^d$. *In particular, the polynomial* $(\lambda^T M^{\mathrm{adj}} \overline{\lambda})(\mu^T M^{\mathrm{adj}} \overline{\mu})$ *is non-negative on* $\mathcal{V}_{\mathbb{R}}(\det(M))$.

Proof. Consider a general $d \times d$-matrix of variables $X = (X_{ij})$. At a generic point in $\mathcal{V}_{\mathbb{C}}(\det(X))$, the matrix X has rank $d-1$. The identity $X \cdot X^{\mathrm{adj}} = \det(X) \cdot I_d$ implies that X^{adj} then has rank one at such a point. It follows that the 2×2-matrix

$$\begin{bmatrix} \lambda^T X^{\mathrm{adj}} \overline{\lambda} & \lambda^T X^{\mathrm{adj}} \overline{\mu} \\ \mu^T X^{\mathrm{adj}} \overline{\lambda} & \mu^T X^{\mathrm{adj}} \overline{\mu} \end{bmatrix}$$

also has rank at most one on $\mathcal{V}_{\mathbb{C}}(\det(X))$. Since the polynomial $\det(X)$ is irreducible, the determinant of this 2×2 matrix thus lies in the ideal generated by $\det(X)$. Restricting to $X = M$ gives the desired identity.

For the claim of non-negativity, note that $(\mu^T M^{\mathrm{adj}} \overline{\lambda}) = \overline{(\lambda^T M^{\mathrm{adj}} \overline{\mu})}$. So the polynomial $(\lambda^T M^{\mathrm{adj}} \overline{\lambda})(\mu^T M^{\mathrm{adj}} \overline{\mu})$ is equal to a polynomial times its conjugate modulo $\det(M)$ and is therefore non-negative on $\mathcal{V}_{\mathbb{R}}(\det(M))$. □

We can use this identity to determine whether a determinantal representation is definite.

Theorem 7.16. *Let $f \in \mathbb{R}[x]_{(d)}$ be irreducible and hyperbolic with respect to e, and let $f = \det(M)$ be a Hermitian determinantal representation of f. If some polynomial in $\mathcal{C}(M)$ interlaces f with respect to e, every polynomial in $\mathcal{C}(M)$ does and the matrix $M(e)$ is (positive or negative) definite.*

Proof. First, suppose that $g = \lambda^T M^{\mathrm{adj}} \overline{\lambda}$ interlaces f and let h be another element of $\mathcal{C}(M)$, say $h = \mu^T M^{\mathrm{adj}} \overline{\mu}$ where $\mu \in \mathbb{C}^d$. From Proposition 7.14, we see that the product $g \cdot h$ is non-negative on $\mathcal{V}_{\mathbb{R}}(f)$. Then, by Lemma 7.11, h interlaces f.

To show that $M(e)$ is definite, we first show that any two elements g, h of $\mathcal{C}(M)$ have the same sign at the point e. Since f is irreducible, the polynomial $g \cdot h$ cannot vanish on $\mathcal{V}_{\mathbb{R}}(f)$. By Proposition 7.14, the product $g \cdot h$ is non-negative on $\mathcal{V}_{\mathbb{R}}(f)$ and thus strictly positive on a dense subset of $\mathcal{V}_{\mathbb{R}}(f)$. Furthermore, since both g and h interlace f, they cannot have any zeros between e and $\mathcal{V}_{\mathbb{R}}(f)$ when restricted to any line $\{u + te \mid t \in \mathbb{R}\}$, for $u \in \mathbb{R}^n \setminus \{0\}$. So the product $g \cdot h$ must be positive at e. Now since $\lambda^T M^{\mathrm{adj}}(e) \overline{\lambda} \in \mathbb{R}$ has the same sign for all $\lambda \in \mathbb{C}^d$, the Hermitian matrix $M^{\mathrm{adj}}(e)$ is definite. Hence so is the matrix $M(e) = f(e)(M^{\mathrm{adj}}(e))^{-1}$. □

Remark 7.17. The converse of Thm. 7.16 also holds. (See [**PV13**, Thm. 3.3]).

We conclude this section with a useful lemma showing that the map taking a matrix with linear entries to the determinant is closed when restricted to definite representations, which it need not be in general.

Lemma 7.18. *Let $e \in \mathbb{R}^n$. The set of all homogeneous polynomials $f \in \mathbb{R}[x]_d$ with $f(e) = 1$ that possess a Hermitian determinantal representation $f = \det(M)$ such that $M(e)$ is positive definite is a closed subset of $\mathbb{R}[x]_d$.*

Proof. First we observe that if $f(e) = 1$ and $f = \det(M)$ with $M(e) > 0$, then f has such a representation M' for which $M'(e)$ is the identity matrix. To find it, we can decompose the matrix $M(e)^{-1}$ as $\overline{U} U^T$ for some complex $d \times d$-matrix U. Then $M' = U^T M \overline{U}$ is a definite determinantal representation of f with $M'(e) = I_d$.

Now let $f_k \in \mathbb{R}[x]_d$ be a sequence of polynomials converging to f such that $f_k = \det(M^{(k)})$ with $M^{(k)}(x) = x_0 M_0^{(k)} + \cdots + x_n M_n^{(k)}$ and $M^{(k)}(e) = I_d$. For each j, let e_j denote the jth unit vector. Since $f_k(te - e_j)$ is the characteristic polynomial of $M_j^{(k)}$, the eigenvalues of each $M_j^{(k)}$ converge to the zeros of $f(te - e_j)$. It follows that each sequence $(M_j^{(k)})_k$ is bounded. We may therefore assume that the sequence $M^{(k)}$ is convergent (after successively passing to a convergent subsequence of $M_j^{(k)}$ for each $j = 0, \ldots, n$) and conclude that $f = \det(\lim_{k \to \infty} M^{(k)})$. □

7.3. HYPERBOLIC CURVES AND THE HELTON-VINNIKOV THEOREM

The goal of this section is to show that every hyperbolic plane curve possesses a Hermitian determinantal representation. This is a weaker statement than the Helton-Vinnikov theorem (Thm. 7.7), which says that there even exists a symmetric determinantal representation, but it still suffices to characterise the plane spectrahedra. The proof follows [**PV13**].

First of all, we now speak of curves and convex subsets of the plane rather than of three-dimensional cones. This is because we consider projective varieties instead of affine cones, which gives a better geometrie picture. Thus a homogeneous polynomial $f \in \mathbb{R}[x,y,z]$ of degree d defines the **projective plane curve**

$$\mathcal{Z}_{\mathbb{C}}(f) = \{p \in \mathbb{P}^2(\mathbb{C}) \mid f(p) = 0\}$$
$$\mathcal{Z}_{\mathbb{R}}(f) = \{p \in \mathbb{P}^2(\mathbb{R}) \mid f(p) = 0\}$$

where $\mathbb{P}^2(\mathbb{C})$ is the complex projective plane. We use the letter $\mathcal{Z}$ to distinguish the plane projective curve from the affine cone $\mathcal{V}_{\mathbb{C}}(f) \subset \mathbb{C}^3$.

Recall how points in the projective plane are denoted in homogeneous coordinates: A point in $\mathbb{P}^2(\mathbb{C})$ is an equivalence class of points in $\mathbb{C}^3 \setminus \{0\}$ defining the same line through the origin. The point in $\mathbb{P}^2(\mathbb{C})$ corresponding to $(a,b,c) \in \mathbb{C}^3 \setminus \{0\}$ is denoted by $(a:b:c)$ and we have $(a:b:c) = (\lambda a : \lambda b : \lambda c)$ for all $\lambda \in \mathbb{C}^*$. In particular, $(a:b:c)$ is real if and only if $\lambda a, \lambda b, \lambda c$ are all real for some $\lambda \in \mathbb{C}^*$. (Thus $(i:i:i)$ is a real point, while $(1:1:i)$ is not.) Complex conjugation acts on $\mathbb{P}^2(\mathbb{C})$ via the rule $\overline{(a:b:c)} = (\overline{a} : \overline{b} : \overline{c})$, so that $\mathbb{P}^2(\mathbb{R})$ is exactly the set of fixed points of this action.

Let $f \in \mathbb{R}[x,y,z]$ be homogeneous and irreducible of degree d. We wish to find a determinantal representation $f = \det(M)$ where M is a Hermitian matrix of linear forms. We will describe a general method for constructing such a representation. The idea goes back to the work of Hesse in 1855 [**He1855a**] and was extended by Dixon in 1902 [**Di1902**].

Construction 7.19. Let $f, g \in \mathbb{R}[x,y,z]$ be homogeneous with f irreducible, $\deg(f) = d$ and $\deg(g) = d-1$. Assume that $\mathcal{Z}_{\mathbb{R}}(f) \cap \mathcal{Z}_{\mathbb{R}}(g) = \varnothing$.

(1) Put $S = \mathcal{Z}_{\mathbb{C}}(f) \cap \mathcal{Z}_{\mathbb{C}}(g)$ and let $T \subset S$ be such that $S = T \cup \overline{T}$ and $T \cap \overline{T} = \varnothing$.
(2) Consider the complex vector space

$$V = \mathcal{I}_{\mathbb{C}}(T) \cap \mathbb{C}[x,y,z]_{(d-1)} = \{h \in \mathbb{C}[x,y,z]_{(d-1)} \mid h|_T = 0\},$$

which is of dimension at least $\frac{(d+1)d}{2} - \frac{d(d-1)}{2} = d$ (the dimension of $\mathbb{C}[x,y,z]_{(d-1)}$ minus the number of conditions imposed by the vanishing at the points in T). Note that S may contain fewer than $d(d-1)$ points, in which case we have to take multiplicities into account in the definition of V to make things work, but we will ignore this point here. Put $a_{11} = g$ and extend to a linearly independent family

$$a_{11}, \ldots, a_{1d} \in V$$

(3) Fix j,k with $2 \leqslant j \leqslant k \leqslant d$. The polynomial $\overline{a_{1k}}a_{1k}$ vanishes on the intersection points S of $\mathcal{Z}_{\mathbb{C}}(f)$ and $\mathcal{Z}_C(g)$. If S consists of $d(d-1)$ distinct points, the homogeneous vanishing ideal of S is generated by f and $g = a_{11}$. Thus we obtain polynomials $p, q \in \mathbb{C}[x,y,z]$ such that

$$\overline{a_{1j}}a_{1k} = pf + qa_{11}.$$

Since $\overline{a_{1j}}a_{1k}$, f and a_{11} are all homogeneous, we can assume that p and q are also homogeneous, and we find $\deg(q) = d-1$.

Again, if S contains fewer than $d(d-1)$ points, we have to take multiplicities into account, but the statement remains true (using Max Noether's theorem).

Put $a_{jk} = q$. If $j = k$, then a_{11} and $\overline{a_{1j}}a_{1j}$ are both real and we take a_{jj} to be real as well. Finally put $a_{kj} = \overline{a_{jk}}$ for $j > k$.

(4) We denote by A_g the $d \times d$-matrix with entries a_{jk}. By construction, we have
$$a_{11}a_{jk} - a_{1k}a_{j1} \in (f).$$
for all $j < k$. Note that A_g depends not only on f and g, but also on the choice of splitting $S = T \cup \overline{T}$ and the choice of basis of V. We will ignore this and denote by A_g any matrix arising in this way.

Now we are ready for the main result of this section.

Theorem 7.20. *Let $f, g \in \mathbb{R}[x, y, z]$ be homogeneous with f irreducible,* $\deg(f) = d$ *and* $\deg(g) = d - 1$. *Assume that $\mathcal{Z}_{\mathbb{R}}(f) \cap \mathcal{Z}_{\mathbb{R}}(g) = \varnothing$ and let A_g be as in Construction 7.19.*

(1) Every entry of the adjugate matrix $(A_g)^{\mathrm{adj}}$ is divisible by f^{d-2} and the matrix
$$M_g = (1/f^{d-2})A_g^{\mathrm{adj}}$$
has linear entries. Furthermore, there exists $\gamma \in \mathbb{R}$ such that
$$\det(M_g) = \gamma f.$$
(2) Assume that f is strictly hyperbolic with respect to a point e and that g strictly interlaces f. Then $\gamma \neq 0$ and the matrix $M_g(e)$ is (positive or negative) definite.

The proof will make use of the following simple lemma.

Lemma 7.21. *Let A be a $d \times d$-matrix with entries in a factorial ring R. If $f \in R$ is irreducible and divides all 2×2-minors of A, then for every $1 \leqslant k \leqslant d$, the element f^{k-1} divides all $k \times k$-minors of A.*

Proof. By hypothesis, the claim holds for $k \leqslant 2$. So assume $k > 2$ and suppose that f^{k-2} divides all $(k-1) \times (k-1)$-minors of A. Let B be a submatrix of size $k \times k$ of A. From $B^{\mathrm{adj}}B = \det(B) \cdot I_k$ we conclude $\det(B^{\mathrm{adj}}) = \det(B)^{k-1}$.

Suppose $\det(B) = f^m g$ where f does not divide g. Then $\det(B)^{k-1} = f^{m(k-1)}g^{k-1}$. By assumption f^{k-2} divides all entries of B^{adj}, hence $f^{k(k-2)}$ divides its determinant $\det(B)^{k-1}$. Since f is irreducible, f does not divide g^{k-1}, so $f^{k(k-2)}$ must divide $f^{m(k-1)}$. Then $k(k-2) \leq m(k-1)$ which implies that $k - 1 \leq m$, as claimed. □

Proof of Thm. 7.20. (1) By construction, the 2×2 minors of A_g of the form $a_{11}a_{jk} - a_{1k}a_{j1}$, are divisible by f. Therefore, if $u \in \mathcal{Z}_{\mathbb{C}}(f)$ is a point with $a_{11}(u) \neq 0$, we can conclude that every row of $A_g(u)$ is a multiple of the first, so that $A_g(u)$ has rank 1. Since a_{11} is not divisible by f, it follows that $a_{11}(u) \neq 0$ holds on a Zariski-dense subset of $\mathcal{Z}_{\mathbb{C}}(f)$. So all the 2×2 minors of A_g are divisible by f. Since f is irreducible in $\mathbb{C}[x, y, z]$, this implies that all $(d-1) \times (d-1)$-minors of A_g are divisible by f^{d-2}, by Lemma 7.21.

The entries of A_g^{adj} have degree $(d-1)^2$ and f has degree d, so that $M_g = (1/f^{d-2}) \cdot A_g^{\mathrm{adj}}$ has entries of degree $(d-1)^2 - d(d-2) = 1$. Furthermore, by Lemma 7.21, $\det(A_g)$ is divisible by f^{d-1}. So $\det(A_g) = cf^{d-1}$ for some $c \in \mathbb{R}[x, y, z]$ and we obtain
$$\begin{aligned} \det(M_g) &= \det(f^{2-d}A_g^{\mathrm{adj}}) = f^{d(2-d)}\det(A_g^{\mathrm{adj}}) = f^{d(2-d)}\det(A_g)^{d-1} \\ &= f^{d(2-d)}c^{d-1}f^{(d-1)^2} = c^{d-1}f. \end{aligned}$$

Since $\det(M_g)$ has degree d, we see that c is a constant and we take $\gamma = c^{d-1}$.

(2) To show that $\det(M_g)$ is not the zero-polynomial, we begin by showing that

$$\lambda^T A_g \overline{\lambda}$$

is not the zero-polynomial, for any $\lambda \in \mathbb{C}^d$. As argued in the proof of (1), the matrix A_g has rank one at all points of $\mathcal{V}_{\mathbb{C}}(f)$ and for every $\lambda \in \mathbb{C}^d$, we have

$$a_{11} \cdot (\lambda^T A_g \overline{\lambda}) - (\lambda^T A_g e_1)(\overline{\lambda^T A_g e_1}) \in (f). \tag{7.22}$$

If $\lambda^T A_g \overline{\lambda}$ is identically zero, we conclude that $\lambda^T A_g e_1$ vanishes on $\mathcal{V}_{\mathbb{C}}(f)$. Since f has degree d and $\lambda^T A_g e_1$ has degree $d-1$, $\lambda^T A_g e_1$ must then vanish identically as well. This contradicts the linear independence of the polynomials $\overline{a_{11}}, \ldots, \overline{a_{1d}}$.

Now suppose that the claim is false and that $\det(M_g)$ is identically zero. From the proof of (1), it is clear that $\det(A_g)$ is then zero, as well. In particular, the determinant of $A_g(e)$ is zero, so there is some nonzero vector $\lambda \in \mathbb{C}^n$ in its kernel, and $\lambda^T A_g(e)\overline{\lambda}$ is also zero. But we have just shown that the polynomial $(\lambda^T A_g \overline{\lambda})$ is nonzero, and Eq. (7.22) shows that the product $a_{11} \cdot (\lambda^T A_g \overline{\lambda})$ is non-negative on $\mathcal{V}_{\mathbb{R}}(f)$. By Lemma 7.11, $(\lambda^T A_g \overline{\lambda})$ therefore interlaces f. Thus this polynomial cannot vanish at the point e and the determinant $\det(M_g)$ is not identically zero.

Now all we need to show is that $M_g(e)$ is definite. To do this, we show that A_g is the adjugate matrix of M_g. By construction, $M_g = f^{2-d} \cdot A_g^{\text{adj}}$. Taking adjugates, we see that

$$M_g^{\text{adj}} = \frac{1}{f^{(d-2)(d-1)}} \cdot (A_g^{\text{adj}})^{\text{adj}} = \frac{1}{f^{(d-2)(d-1)}} \cdot \det(A_g)^{d-2} \cdot A_g = c^{d-2} A_g,$$

where $\det(A_g) = cf^{d-1}$ as in the proof of (1) above. Thus a_{11} is a constant multiple of $e_1^T M_g^{\text{adj}} e_1$ and belongs to $\mathcal{C}(M_g)$. Since a_{11} interlaces f with respect to e, Theorem 7.16 implies that the matrix $M_g(e)$ is definite. □

Corollary 7.23. *Every hyperbolic polynomial in three variables possesses a definite Hermitian determinantal representation.*

Proof. Suppose $f \in \mathbb{R}[x, y, z]$ is irreducible and strictly hyperbolic with respect to e. Then the directional derivative $D_e f$ strictly interlaces f and can be used an input in Construction 7.19. By Thm. 7.20(2), this will result in a definite Hermitian determinantal representation of f. If f is strictly hyperbolic but not irreducible, then each irreducible factor of f is strictly hyperbolic and we can build a Hermitian determinantal representation of f as a block matrix form the representations of all factors.

In general, if $f \in \mathbb{R}[x, y, z]_{(d)}$ is (not necessarily strictly) hyperbolic with respect to e with $f(e) = 1$, there exists a sequence of strictly hyperbolic polynomials $(f_k) \subset \mathbb{R}[x, y, z]_{(d)}$ (with respect to e) and $f_k(e) = 1$ converging to f. Now each f_k has a Hermitian determinantal representation, hence so does f by Lemma 7.18. □

Corollary 7.24. *Every three-dimensional hyperbolicity cone is spectrahedral.*

Proof. This follows at once in view of Lemma. 7.12. □

Corollary 7.25. *Every hyperbolicity cone is a convex cone.*

Proof. Let $f \in \mathbb{R}[x]_{(d)}$ be hyperbolic with respect to e. For $u \in \overline{\mathcal{C}}_e(f)$ and $\alpha > 0$, the roots of $f(\alpha u + te)$ are those of $f(u + \alpha^{-1} te)$, so it is clear that $\alpha u \in \mathcal{C}_e(f)$. Given two points $u, v \in \overline{\mathcal{C}}_e(f)$, let V be the three-dimensional subspace spanned by e, u, v. Then $\overline{\mathcal{C}}_e(f) \cap V$

is the hyperbolicity cone of $f|_V$ and is therefore spectrahedral by the preceding corollary. In particular, it is convex and therefore contains $u + v$. $\square$

Corollary 7.26. *If f is hyperbolic with respect to e, then it is hyperbolic with respect to any interior point of $\overline{C}_e(f)$.*

Proof. This is easy to see if f has a Hermitian determinantal representation and therefore holds for three-dimensional hyperbolicity cones. The general case can be reduced to the three-dimensional one as in the proof of the corollary above. $\square$

Finally, let us see an example. Carrying out Construction 7.19 in practise is not an easy matter. The following example is taken directly from [**PV13**, Example 4.11].

Example 7.27. We apply Construction 7.19 to the quartic

$$f(x, y, z) = x^4 - 4x^2y^2 + y^4 - 4x^2z^2 - 2y^2z^2 + z^4, \tag{7.28}$$

which is hyperbolic with respect to the point $e = (1 : 0 : 0)$. This curve has two nodes, $(0 : 1 : 1)$ and $(0 : -1 : 1)$, so that f is not strictly hyperbolic. But the construction will still work, and this happens to simplify the explicit computations considerably. Figure 1 shows the real curve in the plane $\{x = 1\}$.

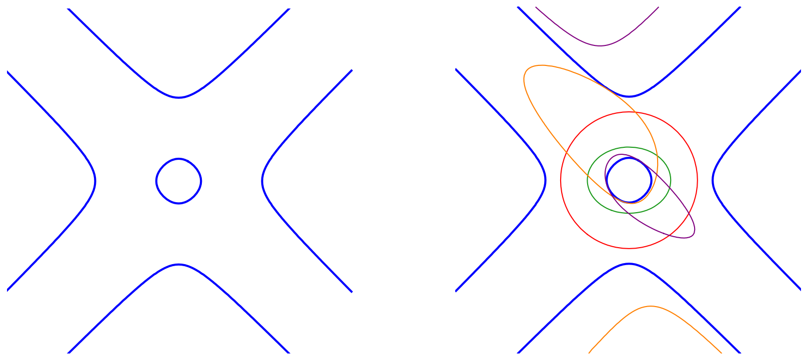


Figure 1. The hyperbolic quartic (7.28) and interlacing cubics.

We define a_{11} to be the directional derivative $\frac{1}{4}D_e f = x^3 - 2xy^2 - 2xz^2$. The curves $\mathcal{V}_{\mathbb{C}}(f)$ and $\mathcal{V}_{\mathbb{C}}(a_{11})$ intersect in the eight points $(2 : \pm\sqrt{3} : \pm i)$, $(2 : \pm i : \pm\sqrt{3})$ and the two nodes, $(0 : \pm 1 : 1)$, each with multiplicity 2, for a total of $4 \cdot 3 = 12$ intersection points, counted with multiplicities. We divide these points into two conjugate sets (making an arbitrary choice) and decompose $S = \mathcal{V}_{\mathbb{C}}(f) \cap \mathcal{V}_{\mathbb{C}}(a_{11})$ into $S = T \cup \overline{T}$ where

$$T = \left\{(0 : 1 : 1), (0 : -1 : 1), (2 : \sqrt{3} : i), (2 : -\sqrt{3} : i), (2 : i : \sqrt{3}), (2 : i : -\sqrt{3})\right\}.$$

The vector space of cubics in $\mathbb{C}[x, y, z]$ vanishing on these six points is four dimensional and we extend a_{11} to a basis $a_{11}, a_{12}, a_{13}, a_{14}$ for this space, where

$$\begin{aligned} a_{12} &= ix^3 + 4ixy^2 - 4x^2z - 4y^2z + 4z^3, \\ a_{13} &= -3ix^3 + 4x^2y + 4ixy^2 - 4y^3 + 4yz^2, \\ a_{14} &= -x^3 - 2ix^2y - 2ix^2z + 4xyz. \end{aligned}$$

Then, to find a_{22} for example, we write $a_{12} \cdot \overline{a_{12}}$ as an element of the ideal (f, a_{11}),

$$a_{12} \cdot \overline{a_{12}} = (13x^3 - 14xy^2 - 22xz^2) \cdot a_{11} + (16z^2 - 12x^2) \cdot f,$$

and set $a_{22} = 13x^3 - 14xy^2 - 22xz^2$. We proceed similarly for the remaining entries and eventually obtain the output of Construction 7.19, the Hermitian matrix of cubics

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ \overline{a_{12}} & a_{22} & a_{23} & a_{24} \\ \overline{a_{13}} & \overline{a_{23}} & a_{33} & a_{34} \\ \overline{a_{14}} & \overline{a_{24}} & \overline{a_{34}} & a_{44} \end{bmatrix}.$$

By taking the adjugate of A and dividing by f^2, we find the desired Hermitian determinantal representation,

$$M \;=\; \frac{1}{f^2}\cdot A^{\mathrm{adj}} \;=\; 2^5 \begin{bmatrix} 14x & 2z & 2ix-2y & 2i(y-z) \\ 2z & x & 0 & -ix+2y \\ -2ix-2y & 0 & x & ix-2z \\ -2i(y-z) & ix+2y & -ix-2z & 4x \end{bmatrix}.$$

The determinant of M is $2^{24}\cdot f$. As promised by Theorems 7.16 and 7.20, the cubics in $\mathcal{C}(M)$ interlace f (see Figure 1) and the matrix M is positive definite at the point $(1,0,0)$.

7.4. HYPERBOLICITY CONES AS PROJECTED SPECTRAHEDRA

Following Netzer and Sanyal in [**NS12**], we can also use the exactness results of Helton and Nie to study representations of hyperbolicity cones as projected spectrahedra. For a hyperbolic polynomial f with smooth hyperbolicity cone, this boils down to verifying quasi-concavity of f on the boundary. But this requires some care, since the concavity cannot be strict along lines through the origin, on which f is constant. Thus we have to take suitable affine slices. First, we will need the following basic lemma.

Lemma 7.29. *Let $f \in \mathbb{R}[x]$ be hyperbolic with respect to e. If $u \in \mathbb{R}^n$ is a point with $f(u) = 0$ and $\nabla f(u) \neq 0$, then $\frac{\partial}{\partial t} f(u+te)|_{t=0} \neq 0$.*

Proof. If $\nabla f(u) \neq 0$, then $\frac{\partial}{\partial s} f(u+sv)|_{s=0} \neq 0$ for generic $v \in \mathbb{R}^n$. Fix such $v \in \mathcal{C}_e(f) = \mathrm{int}(\overline{\mathcal{C}}_e(f))$ and consider the hyperbolic polynomial $f(ru+sv+te)$ in three variables r, s, t. By Thm. 7.23, this polynomial has a Hermitian determinantal representation

$$f(ru+sv+te) = \det(rA+sB+tC)$$

where B and C are positive definite, hence factor as $B = U\overline{U}^T$ and $C = V\overline{V}^T$. Now $s = 0$ is a simple root of $f(u+sv) = \det(A+sB)$, which means that $U^{-1}A(\overline{U^T})^{-1}$ has one-dimensional kernel. But then so does $V^{-1}A(\overline{V^T})^{-1}$, hence $t = 0$ is also a simple root of $f(u+te) = \det(A+tC)$. (For a direct proof of this lemma, see also [**PV13**, Lemma 2.4]). □

Lemma 7.30. *Let $f \in \mathbb{R}[x]$ be hyperbolic with respect to e with $f(e) > 0$ and suppose that $\overline{\mathcal{C}}_e(f)$ is pointed. Let H be any affine hyperplane with $e \in H$, $0 \notin H$. Then $f|_H$ is strictly quasi-concave at any point $u \in \overline{\mathcal{C}}_e(f) \cap H$ with $f(u) \neq 0$ or with $f(u) = 0$ and $\nabla f(u) \neq 0$.*

Proof. Let $H' = H - e$, a linear hyperplane in $\mathbb{R}^n$. Note that showing strict quasi-concavity of $f|_H$ at u amounts to the following: If $v \in H' \setminus \{0\}$ is such that $\frac{\partial}{\partial t} f(u+tv)|_{t=0} = 0$, then $\frac{\partial^2}{\partial t^2} f(u+tv)|_{t=0} < 0$ (see Exercise 7.4 below).

(1) Let $u \in \overline{\mathcal{C}}_e(f) \cap H$ with $f(u) \neq 0$, then $f(u) > 0$, so we may rescale and assume $f(u) = 1$. By Cor. 7.26, f is then hyperbolic with respect to u. Thus for any $v \in H' \setminus \{0\}$,

the univariate polynomial $f(v+tu)$ has only real roots. Since f is homogeneous, the same is true for $f(u+tv) = t^d f(t^{-1}u+v)$, where $d = \deg(f)$. Since $f(u) \neq 0$, all these roots are different from 0, hence we may write

$$f(u+tv) = (1+\lambda_1 t)\cdots(1+\lambda_k t)$$

where $k = \deg_t(f(u+tv)) \leqslant d$ and $-(1/\lambda_1),\ldots,-(1/\lambda_k)$ are the roots of $f(u+tv)$. Since $\overline{C}_e(f)$ is pointed, it does not contain any lines, so $f(u+tv)$ cannot be constant for any $v \in H' \setminus \{0\}$. So in this case, we must have $k > 0$. In particular, the coefficient of t in $f(u+tv)$ is $a_1 = \sum_{i=1}^k \lambda_i$ and the coefficient of t^2 is $a_2 = \frac{1}{2}(a_1^2 - \sum_{i=1}^k \lambda_i^2)$. So if $a_1 = 0$, then $a_2 < 0$, showing that $f|_H$ is strictly quasi-concave at u.

(2) Let $u \in \overline{C}_e(f)$ with $f(u) = 0$ and $\nabla f(u) \neq 0$. For any $v \in H' \setminus \{0\}$, the polynomial $f(re+s(u-e)+tv) \in \mathbb{R}[r,s,t]$ is the restriction of f to the subspace spanned by e,u,v, a hyperbolic polynomial of degree $d = \deg(f)$ in three variables r,s,t. Applying Thm. 7.23 and assuming $f(e) = 1$, we can find Hermitian matrices A,B of size $d \times d$ such that

$$f(re+s(u-e)+tv) = \det(rI_d + sA + tB).$$

Evaluating at $r = s = 1$, $t = 0$ gives $f(u) = \det(I_d + A)$, and since $\nabla f(u) \neq 0$, we must have $\mathrm{rk}(I_d + A) = d-1$ by Lemma 7.29. After a change of coordinates, we may assume $I_d + A = \mathrm{Diag}(1,\ldots,1,0)$. Write $B = (b_{jk})_{j,k \leqslant d}$ and expand $f(u+tv) = a_1 t + \cdots + a_k t^k$. Comparing coefficients on both sides of $f(u+tv) = \det(I_d + A + tB)$, we find

$$a_1 = b_{dd} \qquad \text{and} \qquad a_2 = \sum_{j=1}^{d-1} b_{jj}b_{dd} - b_{jd}\overline{b_{jd}}.$$

So if $a_1 = 0$, we cannot have $a_2 \geqslant 0$ unless $b_{jd} = 0$ for $j = 1,\ldots,d-1$. But that would imply $f(u+tv) = 0$ for all t, which is impossible under the assumption that $\overline{C}_e(f)$ is pointed. □

Exercise 7.4. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Show that f is strictly quasi-concave at a point $u \in \mathbb{R}^n$ if and only if $\frac{\partial}{\partial t} f(u+tv)|_{t=0} = 0$ implies $\frac{\partial^2}{\partial t^2} f(u+tv)|_{t=0} < 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$.

To be able to apply Lemma 7.30, we need a preliminary reduction step to make sure that the hypotheses are satisfied. To do this, we need to get rid of any affine-linear subspace contained in the hyperbolicity cone. In general, if $C \subset \mathbb{R}^n$ is non-empty, closed and convex, there is a unique linear subspace $V \subset \mathbb{R}^n$, called the **lineality space** of C, with the property $C = (C \cap V^\perp) + V$ and such that $C \cap V^\perp$ does not contain any affine linear subspace. The proof is contained in the following exercise.

Exercise 7.5. Let $C \subset \mathbb{R}^n$ be non-emtpy, closed and convex. Prove the following:

(a) If $V \subset \mathbb{R}^n$ is a linear subspace with $u + V \subset C$ for some $u \in C$, then $u + V \subset C$ for *all* $u \in C$.
(b) If V is a subspace with the property in (a), then $C = (C \cap V^\perp) + V$.
(c) There is a unique maximal subspace with the property in (a), called the *lineality space* of C.
(d) If V is the lineality space, then $C \cap V^\perp$ does not contain any affine-linear subspace.

Theorem 7.31. *Every hyperbolicity cone with smooth boundary is a projected spectrahedron.*

Proof. Let $f \in \mathbb{R}[x]$ be hyperbolic with respect to e and assume $f(e) > 0$. That the hyperbolicity cone $\overline{C}_e(f)$ has smooth boundary means $\nabla f(u) \neq 0$ for all $u \in \overline{C}_e(f) \setminus \{0\}$ with $f(u) = 0$. Let $V \subset \mathbb{R}^n$ be the lineality space of $\overline{C}_e(f)$. Then $\overline{C}_e(f) = (\overline{C}_e(f) \cap V^\perp) + V$ by

Exercise 7.5(b). If $\overline{\mathcal{C}}_e(f) \cap V^{\perp}$ is a projected spectrahedron, then so is $\overline{\mathcal{C}}_e(f)$ by Thm. 2.9(3). Thus, with a suitable choice of coordinates and using Exercise 7.5(d), we may assume that $\overline{\mathcal{C}}_e(f)$ contains no affine-linear subspace. In particular, the hyperbolicity cone $\overline{\mathcal{C}}_e(f)$ is then pointed, which implies that there exists an affine hyperplane $H \subset \mathbb{R}^n$ with $e \in H, 0 \notin H$ such that $\overline{\mathcal{C}}_e(f) = \operatorname{cone}(\overline{\mathcal{C}}_e(f) \cap H)$ and such that $\overline{\mathcal{C}}_e(f) \cap H$ is compact (c.f. Prop. 2.3).

Now Lemma 7.30 says that $f|_H$ is strictly quasi-concave at all points of $\overline{\mathcal{C}}_e(f) \cap H$. Furthermore, given $u \in \overline{\mathcal{C}}_e(f) \smallsetminus \{0\}$ with $f(u) = 0$, there exists a radius ε such that $\overline{B}_\varepsilon(u) \cap \mathcal{S}(f) = \overline{B}_\varepsilon \cap \overline{\mathcal{C}}_e(f)$, because $t = 0$ is a simple root of $f(u + te)$ by Lemma 7.29. In other words, f locally describes the hyperbolicity cone. Since the defining polynomial $\varepsilon^2 - \sum_{i=1}^n (x_i - u_i)^2$ of the closed ball $\overline{B}_\varepsilon(u)$ is everywhere strictly quasi-concave, we may apply Cor. 5.14 and conclude that $\overline{B}_\varepsilon \cap \overline{\mathcal{C}}_e(f) \cap H$ possesses an exact Lasserre relaxation. Using the compactness of the boundary of $\overline{\mathcal{C}}_e(f) \cap H$ in H, we can write $\overline{\mathcal{C}}_e(f) \cap H$ as the convex hull of finitely many projected spectrahedra. Applying Thm. 2.9(5), we see that $\overline{\mathcal{C}}_e(f) \cap H$ is a projected spectrahedron. Hence so is $\overline{\mathcal{C}}_e(f) = \operatorname{cone}(\overline{\mathcal{C}}_e(f) \cap H)$, by Thm. 2.9(4). □

REFERENCES

[Di1902] A. C. Dixon. Note on the reduction of a ternary quantic to a symmetrical determinant. *Cambr. Proc.* **11**, 350–351, 1902.

[HV07] J. W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. *Comm. Pure Appl. Math.*, **60** (5), 654–674, 2007. http://arxiv.org/abs/math/0306180

[He1855a] O. Hesse. Über die Doppeltangenten der Curven vierter Ordnung. *J. Reine Angew. Math.* **49** (1855) 279–332.
http://deutsche-digitale-bibliothek.de/item/CNDDPEUUXFQ4MFPKTHW6ZFM3DHUR3UJO

[He1855b] ———. Über Determinanten und ihre Anwendung in der Geometrie, insbesondere auf Curven vierter Ordnung. *J. Reine Angew. Math.*, **49**, 243 – 264, 1855
http://www.deutsche-digitale-bibliothek.de/item/KGWA3VX7UDHU625NLX2VBJ6ZPMZHWCUZ

[La57] P. D. Lax. Differential equations, difference equations and matrix theory. Comm. Pure Appl. Math., **11**, 175–194, 1958.

[NS12] T. Netzer, R. Sanyal. Smooth hyperbolicity cones are spectrahedral shadows. Preprint, 2012.
http://arxiv.org/abs/1208.0441.

[Nu68] W. Nuij. A note on hyperbolic polynomials. *Math. Scand.*, **23**, 69–72, 1968.

[PV13] D. Plaumann and C. Vinzant. Determinantal representations of hyperbolic plane curves: An elementary approach. To appear in *J. Symbolic Computation*, 2013. http://arxiv.org/abs/1207.7047

Daniel Plaumann
Universität Konstanz
Summer 2013

8. TWO-DIMENSIONAL CONVEX SETS

In this chapter, we discuss Scheiderer's recent solution in [**Sc12**] of the Helton-Nie conjecture in dimension two: Every convex semialgebraic subset of the plane is a projected spectrahedron. This is obtained as a consequence of a stronger result, namely the stability of quadratic modules defining 1-dimensional compact sets. Here, we will first deduce the two-dimensional Helton-Nie conjecture from this result and then sketch a proof of the stability theorem for curves.

8.1. CONVEX HULLS OF CURVES AND SCHEIDERER'S THEOREM

Some basic notation and results on affine varieties have already been discussed in Section 4.2. We recall the most important points: To an affine variety $V = \mathcal{V}_{\mathbb{C}}(I) \subset \mathbb{C}^n$ defined over $\mathbb{R}$ by an ideal $I \subset \mathbb{R}[x]$ in the real polynomial ring, corresponds the coordinate ring $\mathbb{R}[V] = \mathbb{R}[x]/\mathcal{I}(V)$ (where $\mathcal{I}(V) = \sqrt{I}$ is the vanishing ideal), a finitely-generated reduced $\mathbb{R}$-algebra. The variety V is treated as an abstract object encoded in $\mathbb{R}[V]$, independent of the choice of coordinates, i.e. the choice of the surjection $\mathbb{R}[x] \to \mathbb{R}[V]$. Recall also that a **morphism** $\varphi\colon V \to W$ between two affine varieties V and W over $\mathbb{R}$ is simply a real polynomial map from V to W (with respect to some embedding of V and W into affine space). Such a morphism induces a homomorphism $\varphi^*\colon \mathbb{R}[W] \to \mathbb{R}[V]$ of $\mathbb{R}$-algebras, given by $f \mapsto f \circ \varphi$.

If M is a finitely generated quadratic module in $\mathbb{R}[V]$, we want to make sense of the notion of stability for M, as defined in Section 3.1. The problem is that we have no well-defined notion of degree for elements in $\mathbb{R}[V]$. There are two solutions: Fixing coordinates and choosing $g_1, \dots, g_r \in \mathbb{R}[x]$ with $M = M(\overline{g_1}, \dots, \overline{g_r})$ and generators $\mathcal{I}(V) = (h_1, \dots, h_s)$ of the vanishing ideal, we can consider the quadratic module $M_0 = M(g_1, \dots, g_r, \pm h_1, \dots, \pm h_s)$ in $\mathbb{R}[x]$, which is just the preimage of M in $\mathbb{R}[x]$ under the residue map $\mathbb{R}[x] \to \mathbb{R}[x]/\mathcal{I}(V)$ (c.f. proof of Cor. 4.4). Then we say that M is stable if and only if M_0 is. We would have to show that this does not depend on the choice of coordinates.

More elegantly, we can just eliminate the notion of degree from the definition of stability: Let A be any $\mathbb{R}$-algebra and $M = M(\underline{g})$ a finitely generated quadratic module in A. Given a linear subspace W of A, we write

$$M_{\underline{g}}[W] = \left\{ \sum_{i=0}^{r} s_i g_i \mid s_i = \sum_j t_{ij}^2 \text{ with } t_{ij} \in W \text{ for } i = 1, \dots, r \right\}$$

and say that M is **stable** if for every finite-dimensional subspace U of A there exists a finite-dimensional subspace W of A such that $M_{\underline{g}} \cap U \subset M_{\underline{g}}[W]$. This definition agrees with the previous one for the polynomial ring and is independent of the choice of generators.

Exercise 8.1. Show that the notion of stability of a quadratic module $M(\underline{g})$ in an $\mathbb{R}$-algebra A, as defined above, does not depend on the choice of generators $\underline{g} \subset M(\underline{g})$.

Now let V be an affine $\mathbb{R}$-variety. We write $V(\mathbb{R})$ for the set of real points of V. Given a subset S of $V(\mathbb{R})$, the preordering $\mathcal{P}(S) = \{f \in \mathbb{R}[V] \mid f|_S \geqslant 0\}$ is called the **saturated preordering of** S. A finitely generated quadratic module $M = M(\underline{g})$, with $\underline{g} = \{g_1, \ldots, g_r\} \subset \mathbb{R}[V]$, is called **saturated** if $M = \mathcal{P}(\mathcal{S}(\underline{g}))$, where $\mathcal{S}(\underline{g}) = \{u \in V(\mathbb{R}) | g_1(u) \geqslant 0, \ldots, g_r(u) \geqslant 0\}$ is the basic-closed set defined by $\underline{g}$. As explained in Section 3.1, the saturated preordering $\mathcal{P}(S)$ of a semialgebraic set S is never finitely generated if $\dim(S) \geqslant 3$. On the other hand, it is finitely generated for any subset of the line (c.f. Example 3.2(2)). It turns out that this is also true for compact subsets of smooth algebraic curves. Recall that a variety is called **smooth** if it does not possess any singular points, neither real nor complex. For an affine hypersurface $\mathcal{V}_{\mathbb{C}}(f)$, $f \in \mathbb{R}[x]$ irreducible, this just means $\nabla f(u) \neq 0$ for all $u \in \mathcal{V}_{\mathbb{C}}(f)$. The affine variety V is called an affine **curve** if all of its irreducible components are one-dimensional.

Theorem 8.1 (Scheiderer [**Sc03**]). *Let Z be a smooth affine curve over $\mathbb{R}$ and $S \subset Z(\mathbb{R})$ a compact semialgebraic subset. Then the preordering $\mathcal{P}(S) \subset \mathbb{R}[Z]$ is finitely generated.* □

Since the saturated preordering $\mathcal{P}(S)$ is finitely generated, it makes sense to ask whether it is stable. To show that it is, we will consider its behaviour under real closed extensions of $\mathbb{R}$, just as we did for positive matrix polynomials in Section 4.4. Let $R/\mathbb{R}$ be a real closed field extension and let $K = R(\sqrt{-1})$ be the algebraic closure. If $V = \mathcal{V}_{\mathbb{C}}(h_1, \ldots, h_s) \subset \mathbb{C}^n$ is an affine variety defined over $\mathbb{R}$ by $h_1, \ldots, h_s \in \mathbb{R}[x]$ with coordinate ring $\mathbb{R}[V] = \mathbb{R}[x]/\sqrt{(h_1, \ldots, h_s)}$, then we can regard $\mathcal{V}_K(h_1, \ldots, h_s) \subset K^n$ as an affine variety defined over R with coordinate ring $R[V] = R[x]/\sqrt{h_1, \ldots, h_s}$. One can show that there is a canonical isomorphism $R[V] \cong \mathbb{R}[V] \otimes_{\mathbb{R}} R$, which is a more intrinsic description of $R[V]$. The main technical result of Scheiderer in [**Sc12**] is the following.

Theorem 8.2. *Let Z be a smooth affine curve over $\mathbb{R}$, let $S \subset Z(\mathbb{R})$ be a compact semialgebraic subset and let $P = \mathcal{P}(S) \subset \mathbb{R}[Z]$ be the saturated preordering of S. For any real closed field extension $R/\mathbb{R}$, the preordering P_R generated by P in $R[Z]$ is again saturated.*

Here, the semialgebraic subset $\mathcal{S}(P_R)$ of $Z(R)$ is the base extension $S(R)$, hence the theorem says exactly that $P_R = \mathcal{P}(S(R))$. We will discuss this result and its proof in detail in the next section. For now, we will just apply it to show stability of $\mathcal{P}(S)$.

Corollary 8.3. *For Z and S as above, the saturated preordering $\mathcal{P}(S)$ in $\mathbb{R}[Z]$ is finitely generated and stable.*

Proof. We only need to show stability of P. This comes as an application of Prop. 3.24, which also holds for stability of quadratic modules in general $\mathbb{R}$-algebras, with the same proof. That P_R is saturated implies that the intersection of P_R with any finite-dimensional subspace of $R[Z] = \mathbb{R}[Z] \otimes_{\mathbb{R}} R$ is semialgebraic over R. (The proof is completely analogues to that of Prop. 3.9). This holds for any real closed $R/\mathbb{R}$, so P is stable by Prop. 3.24. □

We now apply this result in the context of the Lasserre relaxation. Simply speaking, we would like to show that the convex hull of a compact 1-dimensional set in $\mathbb{R}^n$ possesses an exact Lasserre relaxation. This is easy to deduce from Cor. 8.3, but only for subsets of smooth curves. To avoid this assumption, we need a few more preparations. Since stability means that we have degree bounds for all non-negative polynomials, not only linear ones, we have more flexibility and coordinate-independence, which can be exploited as follows.

Proposition 8.4. *Let V be an affine $\mathbb{R}$-variety, let $S \subset V(\mathbb{R})$ be a semialgebraic set and $\varphi\colon V(\mathbb{R}) \to \mathbb{R}^n$ a morphism of varieties. If the saturated preordering $\mathcal{P}(S)$ in $\mathbb{R}[V]$ is finitely generated and stable, then $\overline{\mathrm{conv}(\varphi(S))}$ is a projected spectrahedron.*

Proof. Let $\varphi^*\colon \mathbb{R}[x] \to \mathbb{R}[V]$ be the homomorphism of $\mathbb{R}$-algebras induced by φ and fix generators $\mathcal{P}(S) = M(g_1, \dots, g_r)$, $g_1, \dots, g_r \in \mathbb{R}[V]$. Since $\mathcal{P}(S)$ is finitely generated and stable, there exists a finite-dimensional subspace W of $\mathbb{R}[V]$ such that $(\varphi^*\mathbb{R}[x]_1) \cap \mathcal{P}(S) \subset M_{\underline{g}}[W]$. Now, just as in Prop. 3.13, the convex set

$$M_{\underline{g}}[W]' = \{L \in \mathbb{R}[V]^* \mid L|_{M_{\underline{g}}[W]} \geqslant 0 \text{ and } L(1) = 1\}$$

is a spectrahedron. We can define the generalised Lasserre-relaxation $\mathcal{L}_W \subset \mathbb{R}^n$ as the image of $M_{\underline{g}}[W]'$ under the map $\pi\colon L \mapsto \big(L(\varphi^* x_1), \dots, L(\varphi^* x_n)\big)$. For $u \in S$, the functional $L_u \in \mathbb{R}[V]^*$ given by evaluation in u is contained in $M_{\underline{g}}[W]'$, which implies $\mathrm{conv}(\varphi(S)) \subset \mathcal{L}_W$. As in the proof of 3.14(2), suppose we had $u \in \mathcal{L}_W$, say $u = \pi(L)$ for $L \in M_{\underline{g}}[W]'$, but $u \notin \mathrm{clos}\big(\mathrm{conv}(\varphi(S))\big)$. Then we can pick $\ell \in \mathbb{R}[x]_1$ with $\ell|_{\varphi(S)} \geqslant 0$ and $\ell(u) < 0$ by the separation theorem (Prop. 3.16) and conclude $L(\varphi^*\ell) = \ell(u) < 0$, hence $\varphi^*\ell \notin M_{\underline{g}}[W]$, a contradiction. Thus $\overline{\mathrm{conv}(\varphi(S))} = \overline{\mathcal{L}_W}$ is a projected spectrahedron. □

Remark 8.5. In spite of its apparent generality, the hypotheses of this proposition can only be satisfied if V has dimension at most 1, by the main result of [**Sc05**].

Now if Z is any affine curve over $\mathbb{R}$, possibly singular, we can get rid of the singularities by passing to the **normalisation** of Z. If Z is irreducible, this is the curve corresponding to the integral closure of the domain $\mathbb{R}[Z]$ in its quotient field. This integral closure is again a finitely-generated $\mathbb{R}$-algebra and therefore corresponds to an affine curve $\widetilde{Z}$ over $\mathbb{R}$. The inclusion $\mathbb{R}[Z] \subset \mathbb{R}[\widetilde{Z}]$ corresponds to a morphism $\widetilde{Z} \to Z$ of curves, which is an isomorphism everywhere except over the singular points of Z. Since $\mathbb{R}[\widetilde{Z}]$ is integrally closed, the curve $\widetilde{Z}$ is smooth[1]. The map $\widetilde{Z} \to Z$ is surjective, but its restriction $\widetilde{Z}(\mathbb{R}) \to Z(\mathbb{R})$ to real points may be non-surjective. In fact, a point $u \in Z(\mathbb{R})$ is outside the image of $\widetilde{Z}(\mathbb{R})$ if and only if it is an isolated singularity of Z, in which case it is the image of (potentially several) pairs of complex-conjugate points in $\widetilde{Z}$. The map $\widetilde{Z}(\mathbb{R}) \to Z(\mathbb{R})$ is however proper in the euclidean topology, i.e.it is closed with compact fibres.

If Z has several irreducible components, say $Z = Z_1 \cup \dots \cup Z_k$, we can apply the normalisation separately to each component and obtain the normalisation $\widetilde{Z} = \widetilde{Z}_1 \sqcup \dots \sqcup \widetilde{Z}_k$ of Z (where $\sqcup$ denotes the disjoint union) with coordinate ring $\mathbb{R}[\widetilde{Z}] = \mathbb{R}[Z_1] \times \dots \times \mathbb{R}[Z_k]$.

With all this, we are ready for the first main result.

Theorem 8.6. *Every compact convex semialgebraic set $C \subset \mathbb{R}^n$ whose set of extreme points has dimension at most 1 is a projected spectrahedron.*

[1] In general, the singular locus of an irreducible affine variety V with integrally closed coordinate ring has codimension at least 2 in V. Thus if V is a curve, it must be smooth.

Proof. Let S be the closure of the set of extreme points of C, a compact semialgebraic set of dimension at most 1 by hypothesis. Let Z be the Zariski-closure of S, an affine variety of dimension at most 1. Let $Z_1, \dots, Z_k$ be the irreducible components of Z, let $\widetilde{Z}_i \to Z_i$ be the normalisation of each, and let $\widetilde{Z} \to Z$ be the normalisation of Z, given by $\widetilde{Z} = \widetilde{Z}_1 \sqcup \cdots \sqcup \widetilde{Z}_k$.

Now let $u_1, \dots, u_l \in S$ be the isolated singularities of Z contained in S. As explained above, $u_1, \dots, u_l$ do not lie in the image of $\widetilde{Z}(\mathbb{R}) \to Z(\mathbb{R})$. Consider the abstract variety

$$Y = \widetilde{Z} \sqcup p_1 \sqcup \cdots \sqcup p_l$$

where each p_i is a real point. Its real coordinate ring is the $\mathbb{R}$-algebra

$$\mathbb{R}[Y] = \mathbb{R}[\widetilde{Z}_1] \times \cdots \times \mathbb{R}[\widetilde{Z}_k] \times \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{l \text{ times}}.$$

The variety Y comes with a natural morphism $\varphi\colon Y \to Z$ which is the normalisation on $\widetilde{Z}$ and sends each p_i to u_i. Let $\widetilde{S}$ be the preimage of S in $Y(\mathbb{R})$ under φ. Then $\widetilde{S}$ is again compact and by Cor. 8.3, applied to each irreducible factor of Y, the saturated preordering $\mathcal{P}(\widetilde{S}) \subset \mathbb{R}[Y]$ is finitely generated and stable. By Prop. 8.4, this implies that $C = \overline{\operatorname{conv}(\varphi(\widetilde{S}))}$ is a projected spectrahedron. □

Corollary 8.7. *Every compact convex semialgebraic subset of $\mathbb{R}^2$ is a projected spectrahedron.*

Proof. By the Krein-Milman theorem, a compact convex set is the convex hull of its set of extreme points. For a convex semialgebraic set of dimension 2, these form a semialgebraic set of dimension at most 1, so Thm. 8.6 applies. □

Example 8.8. For example, we see again that the compact convex regions in Examples 6.1 and 6.9 are projected spectrahedra. Moreover, using the construction in the proof of Thm. 8.6, we can also produce an exact (generalised) Lasserre relaxation of these sets, while it was not previously clear whether that is possible.

Finally, we show how to deduce the full two-dimensional Helton-Nie conjecture from the above result. This amounts to dealing with closed, not necessarily compact, semialgebraic subsets of the plane, and then performing some surgery on the boundary in the general case. The latter uses the following result due to Netzer, which we do not prove here. But one can think of it as a (very much) refined version of the argument showing that the interior of a projected spectrahedron is a projected spectrahedron (Thm. 2.9(10)).

Theorem 8.9 (Netzer [**Ne10**]). *Let C, T be projected spectrahedra in $\mathbb{R}^n$ with $T \subset C$. Let $\mathcal{F}_T$ be the set of all faces F of C with $F \cap T \neq \varnothing$ and let*

$$T \looparrowleft C = \bigcup_{F \in \mathcal{F}_T} \operatorname{relint}(F).$$

Then $T \looparrowleft C$ is a projected spectrahedron. □

Exercise 8.2. Show the following (directly or using Netzer's theorem): If C is a projected spectrahedron and u an extreme point of C, then $C \smallsetminus \{u\}$ is a projected spectrahedron. *Hint:* Consider first the case in which u is an exposed face of C.

Now we have all we need.

Theorem 8.10. *Every convex semialgebraic subset of $\mathbb{R}^2$ is a projected spectrahedron.*

Proof. Let C be such a set. Suppose first that C is closed and let $\widehat{C} = \mathrm{clos}(\mathrm{cone}(C \times \{1\})) \subset \mathbb{R}^3$. Since $C \times \{1\} = \widehat{C} \cap (\mathbb{R}^2 \times \{1\})$, it suffices to show that $\widehat{C}$ is a projected spectrahedron. Let U be the lineality space of $\widehat{C}$ (c.f. 7.5). Then $\widehat{C} = (\widehat{C} \cap U^\perp) + U$ and $\widehat{C} \cap U^\perp$ is pointed, so we may assume that $\widehat{C}$ is pointed. In that case, by Prop. 2.3, there exists an affine hyperplane $H \subset \mathbb{R}^3$ such that $\widehat{C} = \mathrm{cone}(\widehat{C} \cap H)$ and $\widehat{C} \cap H$ is compact. Now $\widehat{C} \cap H$ is a projected spectrahedron by Thm. 8.6, hence so is $\widehat{C}$ by Thm. 2.9(4).

For the general case, we may assume as usual that $\mathrm{int}(C) \neq \varnothing$. We know that $\overline{C} = \mathrm{clos}(C)$ is a projected spectrahedron and $S = \overline{C} \smallsetminus C$ is a certain one-dimensional semialgebraic subset of the boundary of $\overline{C}$. Let $\mathcal{F}$ be the finite set of one-dimensional faces of $\overline{C}$. We decompose S as follows:

$$S_0 = \text{the relative interior of } S \cap \mathrm{Ex}(\overline{C}) \text{ inside } S$$
$$S_F = F \cap S \text{ for } F \in \mathcal{F}$$

Then S is the union of S_0, $\bigcup_{F \in \mathcal{F}} S_F$, and finitely-many extreme points $u_1, \ldots, u_k$ of $\overline{C}$. For $F \in \mathcal{F}$, let H_F be the open halfplane with $\mathrm{int}(C) \subset H_F$, $F \cap H_F = \varnothing$, so that $F \subset \partial H_F$. Put

$$\begin{aligned} C_0 &= \overline{C} \smallsetminus S_0 \\ C_F &= H_F \cup (\mathrm{clos}(H_F) \cap C) \\ C_u &= \overline{C} \smallsetminus \{u\} \text{ for } u \in \mathrm{Ex}(\overline{C}). \end{aligned}$$

By construction, we now have

$$C = C_0 \cap \bigcap_{F \in \mathcal{F}} C_F \cap C_{u_1} \cap \cdots \cap C_{u_k}.$$

It therefore suffices to show that each of the finitely many sets appearing in this intersection is a projected spectrahedron. The sets C_F are the union of an open halfplane and an interval and are therefore projected spectrahedra (either by a direct argument or using Thm. 2.9). For C_{u_i}, see Exercise 8.2 below. To deal with C_0, let

$$T = \mathrm{clos}\big(\mathrm{conv}(\partial\overline{C} \smallsetminus S_0)\big)$$

Since T is closed, convex and semialgebraic, we know that T is a projected spectrahedron. Now let $\mathcal{F}_T$ be the set of all faces F of $\overline{C}$ such that $F \cap T \neq \varnothing$ and let $T \looparrowleft \overline{C} = \bigcup_{F \in \mathcal{F}_T} \mathrm{relint}(F)$. We claim that $T \looparrowleft \overline{C} = C_0$, showing that C_0 is a projected spectrahedron, by Netzer's theorem 8.9. To see this let $u \in C_0$. If u is an interior point of $\overline{C}$, then clearly $u \in T \looparrowleft \overline{C}$ (unless $T = \varnothing$, which is a trivial case). If $u_0 \in \partial\overline{C}$, then it is not in S_0, hence it lies in T and therefore in $T \looparrowleft \overline{C}$. Conversely, given u in $\overline{C}$ but not in C_0, then it is a point of S_0. By the definition of S_0, this means that $\{u\}$ is the unique face of $\overline{C}$ containing u. Hence $u \notin T \looparrowleft \overline{C}$. This completes the proof. □

Remark 8.11. Using some more convex geometry, one can refine the argument in the first part of the proof of Thm. 8.10 to show that the closure of the convex hull of any semialgebraic set of dimension at most 1 is a projected spectrahedron. (see [**Sc12**, Thm. 6.1]).

8.2. SUMS OF SQUARES ON COMPACT CURVES AND BASE EXTENSION

The goal of this section is to sketch the proof of Thm. 8.2, which says that the saturated preordering of a compact subset of a smooth curve over $\mathbb{R}$ remains saturated when going up to a real closed extension field.

Example 8.12. Consider the situation on the real line, where everything is completely elementary. For the unit interval $S = [-1,1]$, the saturated preordering $P = \mathcal{P}(S)$ in $\mathbb{R}[x]$ is generated by the polynomial $1-x^2$. This is not hard to show directly. If $R/\mathbb{R}$ is a real closed field extension, the preordering $P_R = P(1-x^2)$ in $R[x]$ is still saturated. But that is because we can just do the same direct proof over R, not because of what we know over $\mathbb{R}$.

If Z is an affine curve over $\mathbb{R}$, the situation is far more complicated. One of the main results in [**Sc03**] says that if Z is smooth and $P = P(g_1,\dots,g_r)$ is a finitely generated preordering in $\mathbb{R}[Z]$ defining a compact set $S = \mathcal{S}(g_1,\dots,g_r) \subset Z(\mathbb{R})$, then P is saturated if and only if in each boundary point u of S in $Z(\mathbb{R})$ one of the generators g_i vanishes to order 1 (or two generators with opposing sign changes if u is an isolated point; see [**Sc03**, Thm. 5.17]). We can see this in the above example, too: The polynomial $1-x^2$ has simple roots 1 and -1, whereas $(1-x^2)^3$ has triple roots and therefore does not generate $\mathcal{P}([-1,1])$, as we showed in Example 3.10. However, the same statement does not hold for curves over a non-archimedean real closed field R. The proof of Thm. 8.2 requires completely new techniques. But it turns out that finding certain elements in P_R with simple zeros on the boundary of S still plays a role in the proof.

Let A be a commutative ring with $\frac{1}{2} \in A$ and let P be a preordering of A. In this generality, P is called **archimedean** if for every $f \in A$, there exists $n \in \mathbb{Z}$ such that $n \pm f \in P$. (It is a consequence of Schmüdgen's theorem that this definition agrees with the one given earlier for preorderings in the polynomial ring.) If I is a prime ideal in A, we let P_I denote the preordering generated by P in the localisation A_I. Explicitly, we have

$$P_I = \left\{ \frac{a}{s^2} \mid a \in P, s \in A \setminus I \right\}.$$

We will use the following local-global principle, developed earlier in [**Sc06**].

Theorem 8.13 (Archimedean local-global principle). *Let A be a ring containing $\frac{1}{2}$ and let P be an archimedean preordering of A. Then an element $f \in A$ is contained in P if and only if it is contained in P_M for every maximal ideal M of A.*

Proof. See [**Sc06**, Thm. 2.8] or [**Marshall**, Thm. 9.6.2]. □

To study the local preorderings P_M, we will need to work in the real spectrum. We quickly recall the basics and fix notations: The points of the **real spectrum** $\operatorname{Sper} A$ of A are the pairs $(I,\leqslant)$ where I is a prime ideal of A and $\leqslant$ an ordering of the residue field $\operatorname{Quot}(A/I)$. If $\alpha = (I,\leqslant)$ is a point in $\operatorname{Sper} A$, the prime ideal I is called the **support** of α, denoted $\operatorname{supp}(\alpha)$. An element $f \in A$ is regarded as a function on A, and $f(\alpha) \geqslant 0$ means that the class of f in $\operatorname{Quot}(A/\operatorname{supp}(\alpha))$ is non-negative with respect to the ordering given by α. If A_I is the localisation of A in a prime ideal I, the ideals of A_I are in canonical bijection with the ideals of A contained in I. Accordingly, the real spectrum $\operatorname{Sper} A_I$ can be considered as an open subset of $\operatorname{Sper} A$, consisting of those points $\alpha \in \operatorname{Sper} A$ with $\operatorname{supp}(\alpha) \subset I$.

Given a preordering P of A, we write

$$X_P = \{\alpha \in \operatorname{Sper}(A) \mid f(\alpha) \geqslant 0 \text{ for all } f \in P\}$$

for the subset of Sper A defined by P.

If V is an affine $\mathbb{R}$ variety, the real points $V(\mathbb{R})$ can be identified with a subset of $\operatorname{Sper}(\mathbb{R}[V])$. Namely, each point $v \in V(\mathbb{R})$ corresponds to a maximal ideal M_v in $\mathbb{R}[V]$ and the residue field $\mathbb{R}[V]/M_v \cong \mathbb{R}$ has a unique order, so that there is a unique point $\alpha_v \in \operatorname{Sper}(\mathbb{R}[V])$ with $\operatorname{supp}(\alpha_v) = M_v$. If P is a finitely generated preordering in $\mathbb{R}[V]$, then clearly $\mathcal{S}(P) \subset X_P$ under this identification.

The following proposition will be very important to reduce the number of maximal ideals we need to consider when applying the archimedean local global principle.

Proposition 8.14. *Let A be a local ring with $\frac{1}{2} \in A$ and let P be a preordering in A. Then every $f \in A$ with $f > 0$ on $X_P \subset$ Sper A is contained in P.*

Proof. See [**Sc10**, Prop. 2.1.]. ☐

Let $R/\mathbb{R}$ be a real closed field extension. We denote by $\mathcal{O}$ the convex hull of $\mathbb{Z}$ in R. It is a subring of R which clearly has the property that $a \in \mathcal{O}$ or $a^{-1} \in \mathcal{O}$ holds for every $a \in R$. It is therefore a valuation ring, i.e. there is a valuation $v\colon R \to \Gamma \cup \{\infty\}$ into some ordered group Γ such that $\mathcal{O} = \{a \in R \mid v(a) \geqslant 0\}$ is the valuation ring of v. The residue field of $\mathcal{O}$ modulo its maximal ideal $\mathfrak{m} = \{a \in \mathcal{O} \mid v(a) > 0\}$ is just $\mathbb{R}$ and we denote the residue class map $\mathcal{O} \to \mathcal{O}/\mathfrak{m} \cong \mathbb{R}$ by $a \mapsto \overline{a}$.

Now let V be an irreducible affine variety over $\mathbb{R}$ with coordinate ring $\mathbb{R}[V]$ and consider the base extension of V to R with coordinate ring $R[V] = \mathbb{R}[V] \otimes_{\mathbb{R}} R$. The valuation v extends to a map $w\colon R[V] \to \Gamma \cup \{\infty\}$ as follows: Given an element $f = \sum_{i=1}^r f_i \otimes a_i \in R[V] = \mathbb{R}[V] \otimes_{\mathbb{R}} R$ with $f_i \in \mathbb{R}[V]$, $a_i \in R$, we define

$$w(f) = \min\{v(a_i) \mid i = 1, \dots, r\}.$$

One can check that w has the same formal properties as v, i.e.

$$w(f+g) \geqslant \min\{w(f), w(g)\} \text{ and}$$
$$w(fg) = w(f) + w(g)$$

for all $f, g \in R[V]$. (The first is clear from the definition. The second is not difficult to show but uses the fact that $\mathbb{R}[V]$ is integral.)

We will mostly work in the coordinate ring of V with coefficients in the valuation ring, which is the ring $\mathcal{O}[V] = \mathbb{R}[V] \otimes_{\mathbb{R}} \mathcal{O}$. We have $\mathcal{O}[V] = \{f \in R[V] \mid w(f) \geqslant 0\}$. The residue map of $\mathcal{O}$ extends to a homomorphism $\mathcal{O}[V] \to \mathbb{R}[V]$, denoted $f \mapsto \overline{f}$, defined coefficient wise, i.e. $\overline{\sum_{i=1}^r f_i \otimes a_i} = \sum_{i=1}^r f_i \otimes \overline{a_i} = \sum_{i=1}^r \overline{a_i} f_i \in \mathbb{R}[V]$. Clearly, $\overline{f} = 0$ if and only if $w(f) > 0$. Likewise, the residue map also induces a map on points $V(\mathcal{O}) \to V(\mathbb{R})$, $v \mapsto \overline{v}$. (If $v \in V(\mathcal{O})$ is regarded as an $\mathbb{R}$-algebra homomorphism $v\colon \mathbb{R}[V] \to \mathcal{O}$, then $\overline{v}$ is just the composition of v with the residue map $\mathcal{O} \to \mathbb{R}$. If coordinates are fixed and $v = (v_1, \dots, v_n)$ is regarded as a tuple in $\mathcal{O}^n$, then $\overline{v} = (\overline{v}_1, \dots, \overline{v}_n) \in \mathbb{R}^n$.)

Lemma 8.15. *Let A be an $\mathbb{R}$-algebra and P an archimedean preordering in A. Then the preordering $P_{\mathcal{O}}$ generated by P in $A \otimes_{\mathbb{R}} \mathcal{O}$ is again archimedean.*

Proof. Note first that we have $P - P = A$. For given $f \in A$, choose $c \in \mathbb{R}$ with $c \pm f \in P$, then $f = \frac{1}{2}(c+f) - \frac{1}{2}(c-f) \in P - P$. It follows that any $f \in A \otimes_{\mathbb{R}} \mathcal{O}$ can be written in the form $f = \sum_{i=1}^r f_i \otimes a_i$ with $f_i \in P$ and $a_i \in \mathcal{O}$ for $i = 1, \ldots, r$. Now choose $0 < c_1 \in \mathbb{R}$ with $c_1 - f_i \in P$ for all i. Also, by definition of $\mathcal{O}$, there exists $c_2 \in \mathbb{R}$ with $0 < a_i < c_2$ for all i, hence $c_2 - a_i$ is a square in $\mathcal{O}$. Now we can write

$$\begin{aligned} rc_1c_2 - f &= \sum_{i=1}^r c_1c_2 \otimes 1 - \sum_{i=1}^r f_i \otimes a_i + \underbrace{\sum_{i=1}^r c_2 f_i \otimes 1 - \sum_{i=1}^r f_i \otimes c_2}_{=0} \\ &= c_2 \sum_{i=1}^r (c_1 - f_i) \otimes 1 + \sum_{i=1}^r f_i \otimes (c_2 - a_i) \end{aligned}$$

which is an element of $P_{\mathcal{O}}$, showing that $P_{\mathcal{O}}$ is archimedean. □

We are now ready for the proof of Thm. 8.2, which we restate.

Theorem. *Let Z be a smooth affine curve over $\mathbb{R}$, let $S \subset Z(\mathbb{R})$ be a compact semialgebraic subset and let $P = \mathcal{P}(S) \subset \mathbb{R}[Z]$ be the saturated preordering of S. For any real closed field extension $R/\mathbb{R}$, the preordering P_R generated by P in $R[Z]$ is again saturated.*

Proof. Let $f \in R[Z]$ with $f \geqslant 0$ on $S(R) = \mathcal{S}(P_R)$. First, we can find $c \in R$ with $w(f) = v(c^2)$, which implies $w(c^{-2}f) = 0$ and $c^{-2}f \in \mathcal{O}[Z]$. Since $c^{-2}f \in P_{\mathcal{O}}$ clearly implies $f \in P_R$, we may assume $f \in \mathcal{O}[Z]$ with $w(f) = 0$. We show $f \in P_{\mathcal{O}}$ in several steps.

(1) We wish to apply the archimedean local global principle Thm. 8.13 for $P_{\mathcal{O}} \subset \mathcal{O}[Z]$. Since S is compact, $P = \mathcal{P}(S)$ is archimedean and so $P_{\mathcal{O}}$ is archimedean by Lemma 8.15. So we only need to show that f is contained in P_M for every maximal ideal M of $\mathcal{O}[Z]$.

(2) The next step is to show that since f is non-negative on $S(R) \subset Z(R)$, it is indeed non-negative on the corresponding subset $X_{P_{\mathcal{O}}}$ of $\operatorname{Sper} \mathcal{O}[Z]$. We will omit the proof of this fact (see [**Sc12**, Lemma 3.5]).

(3) Now localising in a maximal ideal M, we know that f is non-negative on the subset

$$X_{P,M} = X_{P_{\mathcal{O}}} \cap \operatorname{Sper} \mathcal{O}[Z]_M$$

of $\operatorname{Sper} \mathcal{O}[Z]$. If $f > 0$ on $X_{P,M}$, then $f \in P_M$ by Prop. 8.14. So we need only consider those maximal ideals M of $\mathcal{O}[Z]_M$ that contain f and with $X_{P,M} \neq \varnothing$.

We claim that such a maximal ideal M is necessarily of the form

$$M_z = \left\{ \sum_{i=1}^r f_i \otimes a_i \mid \sum_{i=1}^r \overline{a_i} f_i(z) = 0 \text{ in } \mathbb{R} \right\},$$

the maximal ideal of $\mathcal{O}[Z]$ corresponding to a point $z \in S$. To see this, let $\alpha \in X_{P,M}$ with $\operatorname{supp}(\alpha) \subset M$ and let $\mathfrak{p} = \operatorname{supp}(\alpha) \cap \mathbb{R}[Z]$, a prime ideal in $\mathbb{R}[Z]$. Since $w(f) \neq 0$, we cannot have $\operatorname{supp}(\alpha) \subset \mathbb{R}[Z] \otimes \mathfrak{m}$ and so $\mathfrak{p} \neq (0)$. Since Z is a curve, this implies that $\mathfrak{p}$ is a maximal ideal of $\mathbb{R}[Z]$ and therefore corresponds to a point $z \in Z$. Since $\alpha \in X_{P,M}$, the point z must be real and contained in S. Now $I = \mathfrak{p} \otimes \mathcal{O}$ is an ideal of $\mathcal{O}[Z]$ with $I \subset \operatorname{supp}(\alpha) \subset M$ and $\mathcal{O}[Z]/I = \mathcal{O}$. So $\mathcal{O}[Z]/M$ is a field containing $\mathbb{R}$ and contained in $\mathcal{O}$, hence it is archimedean and thus coincides with $\mathbb{R}$. Thus M corresponds to a real point, which must be the point z.

(4) Our goal is to find an element $h \in P_{\mathcal{O}}$ with the property that $g = f/h$ is a unit in $\mathcal{O}[Z]_M$ and such that g is positive on $X_{P,M}$. When that is done, we can apply Lemma 8.14 and conclude $g \in P_M$, hence $f = gh \in P_M$, completing the proof.

(5) Showing the existence of h as in (4) is the most technical part of the proof and we will only give an outline. First, let $\mathcal{O}^c = \mathcal{O}[\sqrt{-1}]$ and consider the set of $\mathcal{O}^c$-points

$$U(z) = \{\zeta \in Z(\mathcal{O}^c) \mid \overline{\zeta} = z\}$$

that map to z under the residue map. We split the real zeros of f in $U(z)$ into two groups

$$\{\zeta \in U(z) \cap Z(R) \mid f(\zeta) = 0\} = \{\eta_1, \dots, \eta_r\} \cup \{\zeta_1, \dots, \zeta_s\}$$

in such a way that $\eta_1, \dots, \eta_r \in \operatorname{int}(S(R))$ and $\zeta_1, \dots, \zeta_s \notin \operatorname{int}(S(R))$. Furthermore, let $\{\omega_1, \dots, \omega_t\} \subset U(z)$ be a subset containing exactly one representative from each pair of non-real complex-conjugate zeros of f in $U(z)$.

Next, one computes the order of vanishing of $\overline{f}$ at z and finds ([**Sc06**, Prop. 3.13])

$$\operatorname{ord}_z(\overline{f}) = \sum_{j=1}^{r} \operatorname{ord}_{\eta_j}(f) + \sum_{k=1}^{s} \operatorname{ord}_{\zeta_k}(f) + 2\sum_{l=1}^{t} \operatorname{ord}_{\omega_l}(f).$$

The crucial point is now to show that for every point $\zeta \in U(z)$ with $f(\zeta) = 0$, there exists an element $h_\zeta \in P_{\mathcal{O}}$ with $w(h_\zeta) = 0$, $h_\zeta(\zeta) = 0$ and

$$\operatorname{ord}_z(\overline{h_\zeta}) = \begin{cases} 1 & \text{if } \zeta \in Z(R), \zeta \notin \operatorname{int}(S(R)) \\ 2 & \text{if } \zeta \in \operatorname{int}(S(R)) \text{ or } \zeta \notin Z(R). \end{cases}$$

This is proved in [**Sc12**, Lemma 4.10]. Given this, we can define

$$h = \prod_{j=1}^{r} (h_{\eta_j})^{\frac{1}{2}\operatorname{ord}_{\eta_j}(f)} \cdot \prod_{k=1}^{s} (h_{\zeta_k})^{\operatorname{ord}_{\zeta_k}(f)} \cdot \prod_{l=1}^{t} (h_{\omega_l})^{\operatorname{ord}_{\omega_l}(f)} \in P_{\mathcal{O}}.$$

which, by the above computation of $\operatorname{ord}_z(\overline{f})$, satisfies $\operatorname{ord}_z(\overline{h}) = \operatorname{ord}_z(\overline{f})$. This implies that $g = \frac{f}{h}$ is a unit in $\mathcal{O}[C]_M$. Since $h \in \mathcal{P}_{\mathcal{O}}$, it is clear that g is non-negative in all points of $X_{P,M}$ where h does not vanish. In the zeros of h, one can argue with continuity in $U(z) \cap Z(R)$, except when there are isolated points, which require an additional adjustment of h. □

REFERENCES

[Ne10] T. Netzer. On semidefinite representations of non-closed sets. *Linear Algebra Appl.*, **432** (**12**), 3072–3078, 2010. http://dx.doi.org/10.1016/j.laa.2010.02.005

[Sc03] C. Scheiderer. Sums of squares on real algebraic curves. *Math. Z.* **245**, 725–760, 2003.

[Sc05] ———. Non-existence of degree bounds for weighted sums of squares representations. *J. Complexity*, **21**(6), 823–844, 2005.

[Sc06] ———. Sums of squares on real algebraic surfaces. *Manuscripta Math.* **119**(4), 395–410, 2006.

[Sc10] ———. Weighted sums of squares in local rings and their completions, I. *Math. Z.* **266**, 1–19, 2010.

[Sc12] ———. Semidefinite representation for convex hulls of real algebraic curves. Preprint, 2012. http://arxiv.org/abs/1208.3865