universität
innsbruck
Institut für Mathematik

Real Algebra and Geometry

Tim Netzer

Contents

Introduction: What is real algebra and geometry?

Probably the most fundamental question in mathematics is about solvability of equations. *Systems of linear equations* over fields are examined in linear algebra, *systems of polynomial equations over the integers* lead to number theory, *systems of differential equations* are considered in analysis...

Classical *algebraic geometry* deals with systems of (nonlinear) polynomial equations over fields, the solutions are also taken from the field. The field is often assumed to be algebraically closed. Solving an equation is algebra, which accounts for the name *algebraic* geometry. But it is often useful to consider the whole set of solutions as a geometric object, called a *variety*. This is why the area is also called *geometry*. Fundamental results, such as *Hilbert's Nullstellensatz*, describe these geometric objects in terms of polynomial functions defined on them. This leads to algebraic certificates for solvability of such systems of equations, for example: a system has no solution if and only if the ideal generated by the equations contains 1. Such certificates can then also be examined algorithmically, for example via *Gröbner bases*.

Real algebraic geometry now deals with the special case that all polynomial equations are defined over the reals $\mathbb{R}$, and one is mostly interested in the set of *real solutions*. One can also consider more general so-called *real closed fields*. What might first look like a subfield of classical algebraic geometry, turns out to be an interesting field in itself, allowing for many completely new questions. For example, the field of real numbers admits an *ordering* $\geqslant$. So one can consider polynomial *inequalities* as well, which doesn't make sense over $\mathbb{C}$. The set of solutions to a system of polynomial inequalities is called a *semialgebraic set*, and it can clearly also be considered as a geometric object. Semialgebraic sets can have nonempty interior, and so the ring of all polynomial functions on them will not lead to a good classification. It makes much more sense to consider the *nonnegative polynomial*

functions on a semialgebraic set, and try to describe them algebraically. Such results are called *Positivstellensätze,* and we will see many of them in these lecture notes. Again, such results can be used to classify solvability of systems of polynomial inequalities, and they also allow for computational approaches, for example via semidefinite optimization.

The first important Positivstellensatz is formulated in *Hilbert's 17th Problem:* Every real polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$, which is nonnegative at each point of $\mathbb{R}^n$, is a sum of squares of rational functions. This result was proven in 1926 by Emil Artin, and it constitutes the beginning of modern real algebra. We will prove this nontrivial theorem during the course. Further Positivstellensätze classify nonnegative polynomials on certain subsets of $\mathbb{R}^n$. A particularly important such result is Schmüdgen's Positivstellensatz from 1991, being the first statement providing a sums-of-squares representation without rational functions, i.e. without denominators. Also this result we will prove in these lecture notes.

The mentioned results also allow for numerous applications. There are connection to functional analysis, to optimization, to convexity, and to theoretical quantum physics. One important example is Lasserre's optimization method, which uses Schmügen's Theorem to transform a hard polynomial optimization problem into a more tractable semidefinite program. We will explain some of these applications here as well.

These notes are lecture notes for a master's course on real algebra and geometry. Thus we do not always cite all references carefully. But to great parts, this text is based upon some standard references, in particular the books *Real Algebraic Geometry* by Bochnak, Coste & Roy [2], *Positive Polynomials and Sums of Squares* by Marshall [4], *Positive Polynomials* by Prestel & Delzell [5], and some unpublished lecture notes by Claus Scheiderer. But mistakes are clearly my own, and I am happy for any hint how the improve these lecture notes.

Chapter 1

Ordered Fields

This first chapter deals with ordered fields. We introduce the notion of a *real closed field,* a generalization of the field $\mathbb{R}$ of real numbers. We provide methods to count roots of polynomials over real closed fields. We show that every ordered field admits a unique real closure. We then prove the *projection theorem* for semi-algebraic sets over real closed fields, and deduce *quantifier elimination* and *Tarski's transfer principle*. These results will lead to a solution of Hilbert's 17th Problem in the following chapter.

1.1 Orderings on Fields

Definition 1.1.1. Let M be a nonempty set. A **linear ordering** on M is a binary relation $\leqslant$, such that for all $a, b, c \in M$ we have:

$a \leqslant a$	reflexivity
$a \leqslant b, b \leqslant c \Rightarrow a \leqslant c$	transitivity
$a \leqslant b, b \leqslant a \Rightarrow a = b$	anti-symmetry
$a \leqslant b$ or $b \leqslant a$	linearity/completeness.

We write $a < b$ if $a \leqslant b$ and $a \neq b$. △

Definition 1.1.2. Let K be a field. A **field ordering** is a linear ordering on K, which additionally fulfills the following conditions, for all $a, b, c \in K$:

$a \leqslant b \Rightarrow a + c \leqslant b + c$	compatibility with addition
$0 \leqslant a, 0 \leqslant b \Rightarrow 0 \leqslant ab$	compatibility with multiplication.

We then call $(K, \leqslant)$ an **ordered field**. △

Lemma 1.1.3. *In any ordered field $(K, \leqslant)$, the following holds for all $a, b, c \in K$:*

(i) $0 \leqslant a^2$.

(ii) $a \leqslant b, 0 \leqslant c \Rightarrow ac \leqslant bc$.

(iii) $0 < a < b \Rightarrow 0 < b^{-1} < a^{-1}$.

(iv) If $b \neq 0$, then $0 \leqslant ab \Leftrightarrow 0 \leqslant ab^{-1}$.

(v) $0 < \underbrace{1 + \cdots + 1}_{n}$ *for all $n \in \mathbb{N}$. In particular we have* $\mathrm{char(K)} = 0$, *i.e.* $\mathbb{Q} \subseteq K$.

Proof. (i) If $0 \leqslant a$, this follows from compatibility with multiplication. If $a \leqslant 0$, then by adding $-a$ we obtain $0 \leqslant -a$, and so $a^2 = (-a)^2 \geqslant 0$. (ii) From $a \leqslant b$ we obtain $b - a \geqslant 0$ by adding $-a$, and thus

$$c(b - a) = cb - ca \geqslant 0.$$

By adding ca the result follows. (iii) Assume $0 < a$. If $a^{-1} \leqslant 0$, then $1 = a^{-1}a < 0$ (again by compatibility with multiplication, after passing to $-a^{-1}$), which contradicts (i). So now let $0 < a < b$. Then

$$(a^{-1} - b^{-1})ab = b - a > 0.$$

From $0 < ab$ we obtain $0 < (ab)^{-1}$, and thus finally $b^{-1} < a^{-1}$. (iv) follows by multiplying with b^2 or $(b^{-1})^2$, respectively. (v) follows by iteratively adding 1 to the inequality $0 < 1$, using transitivity of $\leqslant$. □

Example 1.1.4. (i) $\mathbb{R}$ and $\mathbb{Q}$ admit the well-known field orderings. These are in fact the only field orderings here, as we will see later.
(ii) Let $\mathbb{R}(t)$ be the rational function field in one variable t over $\mathbb{R}$, i.e. the field of fractions of the polynomial ring $\mathbb{R}[t]$. Elements of $\mathbb{R}(t)$ are thus fractions of polynomials in t, under the usual equivalence relation. We can understand such elements also as functions on $\mathbb{R}$, defined up to finitely many points (called *poles*). If two fractions define the same element in $\mathbb{R}(t)$, they define the same function, whereever both are defined.
We now want to find all field orderings of $\mathbb{R}(t)$. For this we fix some $a \in \mathbb{R}$ and set for $f, g \in \mathbb{R}(t)$

$$f \leqslant_{a+} g \quad :\Leftrightarrow \quad \exists \epsilon > 0 \, \forall r \in (a, a + \epsilon) \quad f(r) \leqslant g(r).$$

Since f and g have only finitely many poles, this is well-defined, and its easy to convince oneself that it defines a field ordering, wich coincides with the known one when restricted to $\mathbb{R}$. Note that

$$a <_{a_+} t <_{a_+} b \text{ for all } b \in \mathbb{R} \text{ with } a < b.$$

The variable t is thus larger than a, but just infinitesimally so with respect to $\mathbb{R}$. In the same way we get an ordering $\leqslant_{a_-}$, for which t is infinitesimally smaller than a. There are two more orderings, $\leqslant_\infty$ und $\leqslant_{-\infty}$. One defines

$$f \leqslant_\infty g \quad :\Leftrightarrow \exists s \in \mathbb{R}\, \forall r \in (s, \infty) \quad f(r) \leqslant g(r),$$

and $\leqslant_{-\infty}$ analogously with $(-\infty, s)$. For these orderings we have

$$\mathbb{R} <_\infty t \text{ and } t <_{-\infty} \mathbb{R},$$

respectively. In total we have found the following field orderings on $\mathbb{R}(t)$:

$$\{\leqslant_{-\infty}, \leqslant_\infty, \leqslant_{a_-}, \leqslant_{a_+} | \ a \in \mathbb{R}\}.$$

We now check that these are indeed all field orderings. For this we first observe that a field ordering is already uniquely determined by the position of t with respect to $\mathbb{R}$ (by compatibility with addition and multiplication). Now let $\leqslant_?$ be an ordering on $\mathbb{R}(t)$. We consider

$$I := \{b \in \mathbb{R} \mid b \leqslant_? t\} \text{ and } J := \{b \in \mathbb{R} \mid t \leqslant_? b\}.$$

In case $I = \emptyset$ we clearly have $t <_? \mathbb{R}$, and obtain $\leqslant_?=\leqslant_{-\infty}$; for $J = \emptyset$ analogously $\leqslant_?=\leqslant_\infty$. If $I \neq \emptyset \neq J$, then there exists a unique $a \in \mathbb{R}$ with $I \leqslant a \leqslant J$, since $\mathbb{R}$ is Dedekind complete. Now either $t < a$ or $a < t$, and we thus have either $\leqslant_?=\leqslant_{a_-}$ or $\leqslant_?=\leqslant_{a_+}$.
(iii) Since $\mathbb{Q}(t)$ can be embedded to $\mathbb{R}(t)$, all orderings from $\mathbb{R}(t)$ can be restricted to $\mathbb{Q}(t)$. If a is transcendental, then $\leqslant_{a_-}$ and $\leqslant_{a_+}$ coincide on $\mathbb{Q}(t)$. If a is algebraic, they differ (Exercise 1). △

Definition 1.1.5. An ordering $\leqslant$ on K is called **Archimedean**, if for every $a \in K$ there exists some $n \in \mathbb{N}$ with $a \leqslant n$. △

Example 1.1.6. (i) The well-known orderings on $\mathbb{R}$ and $\mathbb{Q}$ are Archimedean.
(ii) None of the orderings on $\mathbb{R}(t)$ are Archimedean. There always exists some $f \in \mathbb{R}(t)$ with $\mathbb{R} < f$. For $\leqslant_{a_+}$ one can for example take $f = 1/(t - a)$.
(iii) If $a \in \mathbb{R}$ is transcendental over $\mathbb{Q}$, then $\leqslant_{a_-}$ and $\leqslant_{a_+}$ induce (the same) Archimedean ordering on $\mathbb{Q}(t)$. The other ordering are not Archimedean (Exercise 1). △

Lemma 1.1.7. *If $(K, \leqslant)$ is Archimedean, then $\mathbb{Q}$ is dense in K, i.e. for $a < b \in K$ there is some $q \in \mathbb{Q}$ with $a < q < b$.*

Proof. Choose $m \in \mathbb{N}$ with $(b-a)^{-1} < m$. We multiply with the positive element $b-a$ and obtain $1 < m(b-a)$, and thus $ma < mb-1$. Now let $n \in \mathbb{Z}$ be minimal with $mb \leqslant n+1$. Then

$$ma < mb - 1 \leqslant n < mb,$$

and thus $a < n/m < b$. □

Theorem 1.1.8 (Embedding Theorem)**.** *Every Archimedean ordered field can be order-embedded into $\mathbb{R}$.*

Proof. Let $(K, \leqslant)$ be Archimedean. Then $\mathbb{Q}$ is dense in K, by Lemma 1.1.7, and we define $\varphi\colon K \to \mathbb{R}$ as follows. For $a \in K$ consider

$$I_a = \{r \in \mathbb{Q} \mid r \leqslant a\} \text{ and } J_a = \{r \in \mathbb{Q} \mid a \leqslant r\}.$$

In $\mathbb{R}$ there is some x with $I_a \leqslant x \leqslant J_a$, since $\mathbb{R}$ is Dedekind complete. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, there is exactly one such x, and we define $\varphi(a) = x$. One checks that φ is additive und multiplicative, and thus an embedding. The fact that φ respects the ordering is also easy to see (Exercise 2). □

Example 1.1.9. For transcendental $a \in \mathbb{R}$ we have the Archimedean ordering $\leqslant_{a-}$ $(=\leqslant_{a+})$ on $\mathbb{Q}(t)$. Indeed, it comes from an embedding into $\mathbb{R}$, namely sending the variable t to a. Since a is transcendental, this mapping is well-defined and injective on $\mathbb{Q}(t)$. △

So far, an ordering is a binary relation on the field K. But since it is supposed to be compatible with $+$, we already know it completely if we know which elements are larger than 0. In this way we can reduce from a binary to a unary relation, i.e. to a subset of K. This is often much easier to handle.

Definition 1.1.10. Let K be a field. A subset $T \subseteq K$ is called a **preordering** , if

$$T + T \subseteq T, \quad T \cdot T \subseteq T, \quad K^2 \subseteq T, \quad -1 \notin T.$$

Here, K^2 denotes the set of squares in K. If additionally

$$T \cup -T = K$$

holds, we call T an **ordering** of K. Orderings a mostly denoted by P. △

We will justify the double use of *ordering* shortly. But let us start with some easy remarks:

Remark 1.1.11. (i) The set

$$\Sigma K^2 = \left\{ \sum_{i=1}^{n} a_i^2 \mid n \in \mathbb{N}, a_i \in K \right\}$$

of all sums of squares (*sos*) of elements of K is a preordering if and only if

$$-1 \notin \Sigma K^2.$$

In this case it is the smallest preordering, i.e. contained in all other preorderings of K. In particular, every preordering contains 0 and 1.
(ii) One has

$$a = \left(\frac{a+1}{2}\right)^2 - \left(\frac{a-1}{2}\right)^2$$

for all $a \in K$, i.e. every element is a difference of two squares. Using this, the condition $-1 \notin T$ for preorderings can be replaced by $T \neq K$.
(iii) For preorderings we always have $T \cap -T = \{0\}$. Indeed, $x, -x \in T$ for some $x \neq 0$ would imply

$$-1 = x \cdot (-x) \cdot \left(\frac{1}{x}\right)^2 \in T,$$

a contradiction.
(iv) If $P \subseteq P'$ are both orderings of K, then $P = P'$. Indeed, $0 \neq x \in P'$ implies that $-x \in P$ is impossible, by (3). So $x \in P$, since P is an ordering. △

Theorem 1.1.12. *For every field ordering $\leqslant$ on K, the set*

$$P_{\leqslant} := \{a \in K \mid 0 \leqslant a\}$$

is an ordering in the sense of Definition 1.1.10. Conversely, if $P \subseteq K$ is an ordering in the sense of Definition 1.1.10, then setting

$$a \leqslant_P b \quad :\Leftrightarrow \quad b - a \in P$$

gives rise to a field ordering $\leqslant_P$ on K. Both constructions are inverse to each other, and thus provide bijections between the sets of orderings of type $\leqslant$ a P.

Proof. Exercise 3. □

Example 1.1.13. On $\mathbb{R}(t)$ the following sets are orderings:

$$P_1 = \left\{ f/g \mid fg = \sum_{i=k}^{d} a_i t^i, a_d > 0 \right\} \cup \{0\}$$

$$P_2 = \left\{ f/g \mid fg = \sum_{i=k}^{d} a_i t^i, a_k > 0 \right\} \cup \{0\},$$

by Exercise 4. Which ordering from Example 1.1.4 do they correspond to? △

We now examine how preorderings can be extended to orderings.

Lemma 1.1.14. *Let T be a preordering of K, and assume $x \in K \setminus T$. Then*

$$T' := T - xT = \{s - xr \mid s, r \in T\}$$

is again a preordering, with $T \subseteq T'$ and $-x \in T$. In words: If something is not positive, we can make it negative.

Proof. $T \subseteq T'$ follows from $s = s - x \cdot 0$, and $-x \in T'$ follows from $-x = 0 - x \cdot 1$. We obviously have

$$T' + T' \subseteq T',\ T' \cdot T' \subseteq T',\ \text{ and } K^2 \subseteq T \subseteq T'.$$

It remains to show that $-1 \notin T'$. So assume $-1 = s - xr$ with $s, r \in T$. Then clearly $r \neq 0$. From the identitiy

$$r^{-1} = \left(\frac{1}{r}\right)^2 \cdot r \in T$$

we now deduce that

$$x = r^{-1} \cdot rx = r^{-1} \cdot (s + 1) \in T,$$

a contradiction. □

Theorem 1.1.15. *Every preordering T of a field is contained in an ordering P. We even have*

$$T = \bigcap_{T \subseteq P \text{ ordering}} P.$$

Proof. Let $T \subseteq K$ be a preordering. The set

$$\mathcal{T} = \{T' \subseteq K \mid T \subseteq T' \text{ preordering}\}$$

is nonempty ($T \in \mathcal{T}$), partially ordered by inclusion, and every chain admits an upper bound (the union over all chain elements). By Zorn's Lemma there exists a maximal element $P \in \mathcal{T}$. To see that P is actually an ordering, let $x \in K \setminus P$. With Lemma 1.1.14 we see that $P - xP$ again lies in $\mathcal{T}$. Maximality implies $P = P - xP$, and in particular $-x \in P$. So we have shown $P \cup -P = K$.
The inclusion $T \subseteq \bigcap_{T \subseteq P} P$ is obvious. So let $x \notin T$. Then by Lemma 1.1.14 there exists a preordering $T' \supseteq T$ with $-x \in T'$. Now let P be an ordering with $T' \subseteq P$. Then also $T \subseteq P$, and since $-x \in P$ we have $x \notin P$. Thus x does not belong to the intersection on the right-hand side. □

Theorem 1.1.16. *A field K admits an ordering if and only if $-1 \notin \Sigma K^2$. An element is nonnegative at each ordering of K if and only if it is a sum of squares in K.*

Proof. If K admits an ordering P, then

$$-1 \notin P \supseteq \Sigma K^2.$$

If conversely $-1 \notin \Sigma K^2$, then the sums of squares constitute a preordering, which by Theorem 1.1.15 can be extended to an ordering. Since ΣK^2 is contained in each ordering, we have

$$\Sigma K^2 = \bigcap_{P \text{ ordering}} P. \qquad \square$$

Example 1.1.17. (i) Let $f \in \mathbb{R}(t)$ be a rational function which is nonnegative at each point where it is defined. Then f is nonnegative at each ordering on $\mathbb{R}(t)$, which is clear from our complete classification of orderings. Thus f is a sum of squares. This is the trivial one-dimensional case of Hilbert's 17th Problem (which can in fact be improved greatly).
(ii) On $\mathbb{C}$ there is no field ordering, since $-1 = i^2$.
(iii) On $\mathbb{Q}$ there exists exactly one ordering, namely $\Sigma\mathbb{Q}^2$.
(iv) Also $\mathbb{R}$ admits only one ordering, namely $\Sigma\mathbb{R}^2$. We even have $\Sigma\mathbb{R}^2 = \mathbb{R}^2$ here. △

Definition 1.1.18. A field K is called **real**, if it admits at least one ordering. We have already seen that this equivalent to $-1 \notin \Sigma K^2$. It is also equivalent to the fact that $a_1^2 + \cdots + a_n^2 = 0$ implies $a_i = 0$ for all i. △

1.2 Order Extensions and Real Closed Fields

Let L/K be a field extension, and let $\leqslant$ be an ordering on K. We would like to extend the ordering to L, i.e. find an ordering $\leqslant'$ on L, that coincides with $\leqslant$ for elements from K. In view of orderings as subsets of the fields, it means that for the ordering $P \subseteq K$ we want to find an ordering $P' \subseteq L$ with

$$P' \cap K = P.$$

In this case, (L, P') (or $(L, \leqslant')$, respectively) is called an **order extension** of (K, P) (or $(K, \leqslant)$, respectively).

Lemma 1.2.1. *Let (K, P) be an ordered field, and L/K a field extension. Then P extends to L if and only if*

$$-1 \notin T_L(P) := \left\{ \sum_{i=1}^{n} p_i \ell_i^2 \mid n \in \mathbb{N}, \ell_i \in L, p_i \in P \right\},$$

i.e. if $T_L(P)$ is a preordering on L.

Proof. If $T_L(P)$ is a preordering on L, by Theorem 1.1.15 there exists an ordering P' of L with

$$T_L(P) \subseteq P'.$$

From $P \subseteq T_L(P)$ we obtain $P \subseteq P' \cap K$. But since $P' \cap K$ is clearly an ordering of K, we obtain $P = P' \cap K$ from Remark 1.1.11 (iv). So P' extends P.
Let conversely be P' an ordering of L, extending P. Then

$$-1 \notin P' \supseteq T_L(P)$$

proves the other implication. □

The following result states that orderings can be extended to quadratic extensions if and only if they are obtained by adjoining the square root of something positive.

Theorem 1.2.2. *Let (K, P) be an ordered field, $a \in K \setminus K^2$, and*

$$L := K(\sqrt{a}) = K[t]/(t^2 - a).$$

Then P extends to L if and only if $a \in P$.

Proof. Let P' be an extension of P to L. In L we have $a = (\sqrt{a})^2 \in P'$, so $a \in P' \cap K = P$. Conversely, let $a \in P$, and assume there exists an identity

$$-1 = \sum_i p_i(a_i + b_i\sqrt{a})^2$$

with $p_i \in P, a_i, b_i \in K$. Expansion yields

$$-1 = \sum_i p_i a_i^2 + p_i b_i^2 a + 2\sqrt{a}\sum_i a_i b_i,$$

and comparing coefficients shows

$$-1 = \sum_i p_i a_i^2 + p_i b_i^2 a \in P,$$

a contradiction. So we have $-1 \notin T_L(P)$, and Lemma 1.2.1 implies that P admits an extension to L. □

Theorem 1.2.3. *Let L/K be a finite field extension of odd degree. Then every ordering of K extends to L.*

Proof. Assume for contradiction that the statement if false. Then there exists a field extension L/K of odd degree, and an ordering P of K that does not extend to L. We choose such field extension of minimal degree.
In characteristic zero, every algebraic extension is separable, and thus by the Primitive Element Theorem the extension L/K is simple, i.e. of the form

$$L = K(\alpha) = K[t]/(f),$$

where f is the minimal polynomial of the generator α over K. Here, $\deg(f) = 2n + 1$ is the degree of the field extension.
Since P does not admit an extension to L by assumption, Lemma 1.2.1 implies the existence of an identity $-1 = \sum_i p_i \ell_i^2$ with all $\ell_i \in L$. This translates to the polynomial identity

$$1 + \sum_i p_i f_i^2 = h \cdot f, \tag{1.1}$$

with $f_i, h \in K[t]$. We can assume $\deg(f_i) \leqslant 2n$ for all i here. So the degree on the left hand side of (1.1) is at most $4n$, and it is even. Every term $p_i f_i^2$ indeed has

even degree, and a leading coefficient from P. These coefficients cannot sum to zero, since $P \cap -P = \{0\}$.
So $\deg(h) \leqslant 2n-1$ is odd. Now let $h_1 \in K[t]$ be an irreducible factor of h of odd degree, and let β be a zero of h_1 (from its splitting field). We set $L' := K(\beta)$ and obtain a field extension L'/K of odd degree $\leqslant 2n-1$, and by substituting β into the equation (1.1), we obtain an identity

$$-1 = \sum_i p_i \delta_i^2,$$

where $\delta_i = f_i(\beta) \in L'$. By Lemma 1.2.1, P does also not extend to L', contradicting the minimality of L. □

Theorem 1.2.4. *Every ordering of K extends to the rational function field $K(t)$.*

Proof. If some ordering P of K did not extend, we would obtain an identity

$$-1 = \sum_i p_i \left(\frac{f_i}{g}\right)^2$$

with $f_i, g \in K[t]$ and $p_i \in P \setminus \{0\}$, again by Lemma 1.2.1. We can assume that g does not have a common divisor with all f_i. We multiply with g^2 and evaulate at 0:

$$-g(0)^2 = \sum_i p_i f_i(0)^2 \in P.$$

If $g(0) \neq 0$, we obtain $-1 \in P$ by multiplying with the square $g(0)^{-2}$, a contradiction. If $g(0) = 0$, we conclude $f_i(0) = 0$ for all i, and all polynomials are divisible by t. This is also a contradiction. □

Definition 1.2.5. A field K is called **real closed**, if it is real, but none of its nontrivial algebraic extensions is real. △

Example 1.2.6. The real numbers $\mathbb{R}$ admit an ordering. The only nontrivial algebraic extension of $\mathbb{R}$ is $\mathbb{C}$, which does not admit an ordering. Therefore, $\mathbb{R}$ is a real closed field. △

Note that the notion of a real closed field does only imply that at least one ordering exists, but does not specify a particular ordering. The next lemma says that this is indeed not necessary.

Lemma 1.2.7. *If K is real closed, it admits exactly one ordering, namely $P = K^2$.*

Proof. Let P be an ordering of K. For $a \in P \setminus K^2$ we could extend P to the nontrivial extension $K(\sqrt{a})$, by Theorem 1.2.2. This is impossible, since K is real closed. So we have $P = K^2$. □

Before we can prove one of the most important theorems on real closed fields, we need the following auxiliary lemma:

Lemma 1.2.8. *Let K be a field for which K^2 is an ordering. Then every element from $K(\sqrt{-1})$ is a square.*

Proof. Let $z = a + b\sqrt{-1}$ be an element from $K(\sqrt{-1})$, i.e. $a, b \in K$. There exists $c \in K$ with $c^2 = a^2 + b^2$, and in fact we can choose $c \geqslant 0$ with respect the (only) ordering $P = K^2$ on K. From

$$(c + a)(c - a) = c^2 - a^2 = b^2 \geqslant 0$$

and the fact that either $c + a \geqslant 0$ or $c - a \geqslant 0$, both inequalities must in fact hold. So there are $d, e \geqslant 0$ with

$$d^2 = (c + a)/2 \text{ and } e^2 = (c - a)/2.$$

Then $(2de)^2 = (c + a)(c - a) = b^2$, and thus $\pm 2de = b$. We finally obtain

$$(d \pm e\sqrt{-1})^2 = d^2 - e^2 \pm 2de\sqrt{-1} = a + b\sqrt{-1} = z.$$ □

The following theorem is an important and very useful characterization of real closed fields:

Theorem 1.2.9 (Artin & Schreier, 1926)**.** *For a field K, the following properties are equivalent:*

(i) *K is real closed.*

(ii) *K^2 is an ordering of K, and every polynomial $p \in K[t]$ of odd degree has a zero in K.*

(iii) *$K \neq K(\sqrt{-1})$ and $K(\sqrt{-1})$ is algebraically closed.*

Proof. *(i)* $\Rightarrow$ *(ii)*: Lemma 1.2.7 says that K^2 is an ordering of K. Now let $p \in K[t]$ be of odd degree. By passing to an irreducible factor of odd degree, we can assume that p is irreducible. Then $L := K[t]/(p)$ is a field extension of K of odd

degree, onto wich the ordering extends by Theorem 1.2.3. Since K is real closed, this implies $L = K$, i.e. $\deg(p) = 1$. So p has a zero in K.
$(ii) \Rightarrow (iii)$: From $-1 \notin K^2$ we get $K \neq K(\sqrt{-1})$. Now we have to show that $K(\sqrt{-1})$ does not admit a nontrivial algebraic extension. To this end, let L be a finite algebraic extension of $K(\sqrt{-1})$. We can assume the extension L/K to be Galois, by passing to the normal hull if necessary. Let $G = \mathrm{Gal}(L/K)$, H a 2-Sylow-subgroup of G, and F the intermediate field corresponding to H:

$$\begin{array}{cc} L & \{\mathrm{id}\} \\ 2^e \,| & |\, 2^e \\ F & H \\ \text{odd} \,| & |\, \text{odd} \\ K & G \end{array}$$

Since by (ii) there do not exist nontrivial odd extensions of K, this implies $F = K$ and $|G| = 2^e$. For the subgroup $G_1 := \mathrm{Gal}(L/K(\sqrt{-1}))$ we thus have $|G_1| = 2^{e-1}$, and we will show $e - 1 = 0$, i.e. $L = K(\sqrt{-1})$. If this wasn't true, we could find a subgroup H_1 of G_1 with $|H_1| = 2^{e-2}$. For the corresponding intermediate field F_1 we would have the following:

$$\begin{array}{cc} L & \{\mathrm{id}\} \\ 2^{e-2} \,| & |\, 2^{e-2} \\ F_1 & H_1 \\ 2 \,| & |\, 2 \\ K(\sqrt{-1}) & G_1 \\ 2 \,| & |\, 2 \\ K & G \end{array}$$

By Lemma 1.2.8, $K(\sqrt{-1})$ does not have a quadratic extension, since each such extension arises by adjoining a square-root. So $e = 1$, and thus $L = K(\sqrt{-1})$.
$(iii) \Rightarrow (i)$: We first show $\Sigma K^2 = K^2$. So let $a, b \in K$. Since $K(\sqrt{-1})$ is algebraically closed, there are $c, d \in K$ with

$$(c + d\sqrt{-1})^2 = a + b\sqrt{-1},$$

so $c^2 - d^2 = a$ and $2cd = b$. We now compute

$$a^2 + b^2 = c^4 - 2c^2d^2 + d^4 + 4c^2d^2 = c^4 + 2c^2d^2 + d^4 = (c^2 + d^2)^2 \in K^2.$$

From $K \neq K(\sqrt{-1})$ we obtain $-1 \notin K^2 = \Sigma K^2$, and so K is real. Since every algebraic extension of K embeds into the algebraic closure $K(\sqrt{-1})$, $K(\sqrt{-1})$ is the only such extension. It is obviously not real. □

Corollary 1.2.10. *Let R be a real closed field, and $R' \subseteq R$ algebraically closed in R. Then R' is also real closed.*

Proof. Exercise 6. □

Theorem 1.2.11. *Let R be a real closed field. Then the following is true:*

(i) The only irreducible monic polynomials in $R[t]$ are of the form $t - a$ and $(t-a)^2 + b^2$, with $a, b \in R, b \neq 0$.

(ii) (Intermediate Value Theorem for polynomials) If for $p \in R[t]$ and $a, b \in R$ we have $p(a) < 0 < p(b)$, then there is some $c \in (a, b)$ with $p(c) = 0$.

Proof. (i): Since $R(\sqrt{-1})$ is algebraically closed, irreducible polynomials over R are of degree at most 2. If they are monic, the only possibilities are thus $t - a$ and

$$t^2 - 2at + c = (t - a)^2 + (c - a^2),$$

with $a, c \in R$. Polynomials of the second type are irreducible if and only if they do not have a root in R, i.e. if $a^2 - c \notin R^2$, and this is equivalent to $c - a^2 \in R^2 \setminus \{0\}$. (ii): Write p as a product of irreducible polynomials. A sign change between a and b can only come from a linear factor $t - c$ with $c \in (a, b)$, which gives rise to a root as desired. □

Theorem 1.2.12 (Rolle's Theorem for polynomials)**.** *Let R be a real closed field and $p \in R[t]$. If $p(a) = p(b)$ holds for some $a < b$ in R, then there exists $c \in R$ with $a < c < b$ and $p'(c) = 0$. Here, p' denotes the usual (formal) derivative of p.*

Proof. Exercise 7. □

1.3 Real Zeros of Polynomials

In this section we describe a method to count real zeros of polynomials, without actually having to compute them. The results are interesting for themselves, but also play an important role in the proof of the Theorem of Tarski and Seidenberg, which is most fundamental for the proof of Hilbert's 17th Problem below.

First let K be an arbitrary field, and

$$p = t^d + a_1 t^{d-1} + a_2 t^{d-2} + \cdots + a_{d-1} t + a_d \in K[t]$$

a monic polynomial over K. We denote the zeros of p (from the algebraic closure of K) by $\alpha_1, \ldots, \alpha_d$. For every $r \in \mathbb{N}$ we define the **r-th Newton sum** of p as

$$\nu_r(p) := \alpha_1^r + \cdots + \alpha_d^r.$$

In word, these are the sums of r-th powers of the zeros of p. It is well-known that it is hard (almost impossible) to compute the zeros from the coefficients of p exactly, in general. But surprisingly, the Newton sums *can* be computed easily from the coefficients of p alone.

Definition 1.3.1. The **companion matrix** of the polynomial $p \in K[t]$ is defined as

$$C(p) := \begin{pmatrix} 0 & & & -a_d \\ 1 & \ddots & & -a_{d-1} \\ & \ddots & 0 & \vdots \\ & & 1 & -a_1 \end{pmatrix} \in \mathrm{M}_d(K). \qquad \triangle$$

It is easy to check

$$\det(tI_d - C(p)) = p,$$

i.e. p is the characteristic polynomial of the matrix $C(p)$ (Exercise 13). So the eigenvalues of $C(p)$ are the zeros of p (all over the algebraic closure of K). The eigenvalues of $C(p)^k$ are thus the k-th powers of the zeros of p. So we have

$$\mathrm{tr}(C(p)^k) = \nu_k(p)$$

where tr denotes the trace. Note that the trace of a matrix over K is the sum of its eigenvalues on the one hand, and the sums of its diagonal elements, on the other hand. This is clear since both values are (the negative of) the second-highest coefficient of the characteristic polynomial of the matrix.

Now we can deduce from this argument that the Newton sums are integer polynomial expressions in the coefficients of p, which can easily be computed. For example, we obtain

$$\begin{aligned}\nu_0(p) &= d\\ \nu_1(p) &= -a_1\\ \nu_2(p) &= a_1^2 - 2a_2.\end{aligned}$$

In fact, ν_i is of total degree i.

Definition 1.3.2. Let K be a field and $p \in K[t]$ a monic polynomial of degree d. The matrix

$$\mathcal{H}(p) := (\nu_{i+j}(p))_{i,j=0,\dots,d-1} = \begin{pmatrix} \nu_0(p) & \nu_1(p) & \cdots & \nu_{d-1}(p)\\ \nu_1(p) & \nu_2(p) & \cdots & \nu_d(p)\\ \vdots & & & \vdots\\ \nu_{d-1}(p) & \nu_d(p) & \cdots & \nu_{2d-2}(p)\end{pmatrix}$$

is called the **Hermite matrix of** p. △

The (i,j)-entry of $\mathcal{H}(p)$ only depends on $i+j$, i.e. is constant along the counter-diagonal. A matrix of this form is also called a **Hankel matrix**. The (i,j)-entry of $\mathcal{H}(p)$ is a polynomial expression of degree $i+j$ in the coefficients of p. So $\mathcal{H}(p)$ can be computed explicitly from p.

Let $M \in \mathrm{Sym}_d(K)$ be a symmetric matrix over the field K, where we assume $\mathrm{char}(K) \neq 2$. Then there exists an invertible matrix $S \in \mathrm{Gl}_d(K)$, such that

$$S^t M S = \mathrm{diag}(a_1, \dots, a_d)$$

is of diagonal form. Now if (K, P) is an ordered field (which requires $\mathrm{char}(K) = 0$), we define the **signature** of M as follows:

$$\mathrm{sign}_P M := \sum_{i=1}^{d} \mathrm{sign}_P(a_i).$$

Here we set

$$\mathrm{sign}_P(a) := \begin{cases} 1: & a \in P \setminus \{0\}\\ -1: & -a \in P \setminus \{0\}\\ 0: & a = 0\end{cases}$$

In other words, the signature is the number of positive minus the number of negative diagonal elements in a diagonalization of M (as a symmetric bilinear form). *Sylvester's Law of Inertia* states that the signature is well-defined, i.e. does not depend on the choice of diagonalization. The proof is usually done over $\mathbb{R}$, but works over an arbitrary ordered field analogously.
The following theorem provides a method to count zeros of polynomials over real closed fields, without actually having to compute them. Since the Hermite matrix is easy to compute, this is an important and relevant result.

Theorem 1.3.3. *Let R be a real closed field and $p \in R[t]$ a monic polynomial of degree $\geqslant 1$. We then have:*

(i) $\operatorname{rank} \mathcal{H}(p) =$ *number of different roots of p in $R(\sqrt{-1})$.*

(ii) $\operatorname{sign} \mathcal{H}(p) =$ *number of different roots of p in R.*

Proof. Let $\alpha_1, \ldots, \alpha_d$ be all the roots of p in $R(\sqrt{-1})$. We set

$$\omega_i := (1, \alpha_i, \ldots, \alpha_i^{d-1})^t$$

and observe

$$\mathcal{H}(p) = \sum_{i=1}^{d} \omega_i \omega_i^t.$$

Now assume w.l.o.g. that $\alpha_1, \ldots, \alpha_s$ are the pairwisely different roots of p, and that α_i has multiplicity n_i. The vectors $\omega_1, \ldots, \omega_s$ are then linearly independent (over $R(\sqrt{-1})$), since the Vandermonde matrix to pairwisely different numbers has full rank. From

$$\mathcal{H}(p) = \sum_{i=1}^{s} n_i \cdot \omega_i \omega_i^t, \tag{1.2}$$

we can thus read off that $\mathcal{H}(p)$ as rank s (c.f. Exercise 8 and note that $n_i \omega_i \omega_i^t = (\sqrt{n_i}\omega_i)(\sqrt{n_i}\omega_i)^t$ holds). This proves (i).
Since p has coefficients from R, its zeros come in complex conjugate pairs, i.e. if $\alpha = a + b\sqrt{-1}$ is a zero, then so is $\overline{\alpha} = a - b\sqrt{-1}$. We can thus assume that

$$(\alpha_1, \ldots, \alpha_s) = (\alpha_1, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_q, \overline{\alpha}_{r+1}, \ldots, \overline{\alpha}_q)$$

with $\alpha_i \in R$ for $i \leqslant r$ and $\alpha_i \in R(\sqrt{-1}) \setminus R$ for $i > r$. Equation (1.2) then reads

$$\begin{aligned}\mathcal{H}(p) &= \sum_{i=1}^{r} n_i \cdot \omega_i \omega_i^t + \sum_{i=r+1}^{q} n_i \cdot (\omega_i \omega_i^t + \overline{\omega}_i \overline{\omega}_i^t) \\ &= \sum_{i=1}^{r} n_i \cdot \omega_i \omega_i^t + \sum_{i=r+1}^{q} 2n_i \cdot \mathrm{Re}(\omega_i)\mathrm{Re}(\omega_i)^t - \sum_{i=r+1}^{q} 2n_i \cdot \mathrm{Im}(\omega_i)\mathrm{Im}(\omega_i)^t.\end{aligned}$$

Furthermore, since

$$\omega_1, \ldots, \omega_r, \omega_{r+1}, \ldots, \omega_q, \overline{\omega}_{r+1}, \ldots, \overline{\omega}_q$$

are linearly independent, so are

$$\omega_1, \ldots, \omega_r, \mathrm{Re}(\omega_{r+1}), \ldots, \mathrm{Re}(\omega_q), \mathrm{Im}(\omega_{r+1}), \ldots, \mathrm{Im}(\omega_q) \in R^d.$$

This implies

$$\operatorname{sign} \mathcal{H}(p) = q - (q - r) = r,$$

by Exercise 8. Note again that the multiplicities n_i are positive and thus admit square-roots in R, which allows to replace $n_i \omega_i \omega_i^t$ by $(\sqrt{n_i}\omega_i)(\sqrt{n_i}\omega_i)^t$. □

Example 1.3.4. Consider $p = t^2 + 1 \in R[t]$. We compute

$$\mathcal{H}(p) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

The rank of this matrix is 2, its signature is 0. So p has two different roots in $R(\sqrt{-1})$, but none in R. This can of course be confirmed easily in this example, by computing the two zeros $\pm\sqrt{-1}$. △

Corollary 1.3.5. *The monic polynomial $p \in R[t]$ has zeros only in R if and only if $\mathcal{H}(p)$ is positive semidefinite, i.e. if its diagonalization has no negative entry.*

Proof. By Theorem 1.3.3, all zeros of p lie in R if and only if

$$\operatorname{rank} \mathcal{H}(p) = \operatorname{sign} \mathcal{H}(p)$$

holds. But this just means that the diagonalization has no negative entries. □

We will also need to count zeros under additional side constraints. For this purpose we now generalize the Hermite matrix. Again let $p \in K[t]$ be a monic polynomial over the field K, whose zeros in the algebraic closure are denoted by $\alpha_1, \ldots, \alpha_d$. Let $q \in K[t]$ be another polynomial, and define the **generalized Newton sums** as

$$\nu_r(p,q) := \alpha_1^r \cdot q(\alpha_1) + \cdots + \alpha_d^r \cdot q(\alpha_d)$$

as well as the **generalized Hermite matrix** as

$$\mathcal{H}(p,q) := (\nu_{i+j}(p,q))_{i,j=0,\ldots,d-1}.$$

For $q = 1$ we obtain the already established notions from above. It is easy to see that also the generalized Newton sums are integer polynomial expressions in the coefficients of p and q, which can easily be computed: for $q = \sum_{j=0}^{d'} b_j t^j$ we have

$$\nu_r(p,q) = \sum_i \alpha_i^r \cdot \sum_j b_j \alpha_i^j = \sum_j b_j \sum_i \alpha_i^{j+r} = \sum_j b_j \nu_{j+r}(p).$$

This gives rise to the following generalization of Theorem 1.3.3:

Theorem 1.3.6. *Let R be a real closed field and $p \in R[t]$ a monic polynomial of degree $\geqslant 1$. Let $q \in R[t]$ be another polynomial. We then have:*

(i) $\operatorname{rank} \mathcal{H}(p,q) =$ *number of different roots α of p in $R(\sqrt{-1})$ with $q(\alpha) \neq 0$.*

(ii) $\operatorname{sign} \mathcal{H}(p,q) = \sum_{\alpha \in R, p(\alpha)=0} \operatorname{sign} q(\alpha)$.

Proof. With the same notation as in the proof of Theorem 1.3.3 we have

$$\mathcal{H}(p,q) = \sum_{i=1}^{s} n_i q(\alpha_i) \omega_i \omega_i^t,$$

with linearly independent vectors ω_i. This proves (i), after merging the coefficients $n_i q(\alpha_i)$ as square-roots in $R(\sqrt{-1})$ with the vectors. For (ii) we get terms of the form

$$n_i q(\alpha_i) \omega_i \omega_i^t$$

with $\alpha_i \in R$ and

$$n_i q(\alpha_i) \cdot \omega_i \omega_i^t + n_i \overline{q(\alpha_i)} \cdot \overline{\omega}_i \overline{\omega}_i^t$$

with $\alpha_i \notin R$. A term of the second type can be rewritten as

$$v_1 v_1^t - v_2 v_2^t$$

with new vectors $v_1, v_2 \in R^d$, spanning the same space as $\omega_i, \overline{\omega}_i$. We just have to replace ω_i by $\sqrt{n_i q(\alpha_i)}\omega_i$ and then continue as in the proof of Theorem 1.3.3. For the signature only the terms with $\alpha_i \in R$ are relevant. We can again replace a term $n_i q(\alpha_i)$ by either $1, -1$ or 0, depending on the sign. This proves the claim. □

We can now also increase the number of side-constraints, and take prescribed signs into account. So for $p, q_1, \ldots, q_m \in R[t]$ with $p \neq 0$ we want to compute the number

$$\#\{\alpha \in R \mid p(\alpha) = 0, q_1(\alpha) >, \ldots, q_m(\alpha) > 0\}.$$

Note that by replacing q_i by $-q_i$ we have also covered the condition $q_i(\alpha) < 0$, and by replacing p by $p^2 + q_i^2$ we have covered the condition $q_i(\alpha) = 0$. For $e \in \{1,2\}^m$ we set

$$q^e := q_1^{e_1} \cdots q_m^{e_m}.$$

Corollary 1.3.7. *Let R be a real closed field, and let $p, q_1, \ldots, q_m \in R[t]$ be polynomials with p monic. We then have*

$$\#\{\alpha \in R \mid p(\alpha) = 0, q_1(\alpha) > 0, \ldots, q_m(\alpha) > 0\} = \frac{1}{2^m} \sum_{e \in \{1,2\}^m} \operatorname{sign} \mathcal{H}(p, q^e).$$

Proof. From Theorem 1.3.6 we already know

$$\operatorname{sign} \mathcal{H}(p, q^e) = \sum_{\alpha \in R, p(\alpha)=0} \operatorname{sign} q^e(\alpha).$$

This implies

$$\begin{aligned} \sum_{e \in \{1,2\}^m} \operatorname{sign} \mathcal{H}(p, q^e) &= \sum_{\alpha \in R, p(\alpha)=0} \sum_{e \in \{1,2\}^m} \operatorname{sign} q_1^{e_1}(\alpha) \cdots q_m^{e_m}(\alpha) \\ &= \sum_{\alpha} \prod_{i=1}^{m} (\operatorname{sign} q_i(\alpha) + \operatorname{sign} q_i(\alpha)^2). \end{aligned}$$

Now $\operatorname{sign} q_i(\alpha) + \operatorname{sign} q_i(\alpha)^2$ equals 2 if and only if $q_i(\alpha) > 0$, and in all other cases it is 0. This proves the claim. □

1.4 The Real Closure

Definition 1.4.1. Let (K, P) be an ordered field. A field extension R of K is called **real closure of** (K, P), if R is real closed, R/K is algebraic, and the ordering on R extends P. △

Theorem 1.4.2. *Every ordered field admits a real closure.*

Proof. Fix an algebraic closure $\overline{K}$ of K, and consider the nonempty set

$$\mathcal{M} = \left\{(L, P') \mid K \subseteq L \subseteq \overline{K}, P' \cap K = P\right\}.$$

It is partially orderd by

$$(L_1, P_1) \preceq (L_2, P_2) \ :\Leftrightarrow\ L_1 \subseteq L_2 \text{ and } P_2 \cap L_1 = P_1,$$

and each chain $(L_i, P_i)_{i\in I}$ admits in $\mathcal{M}$ the upper bound $L = \bigcup_{i\in I} L_i$, equipped with the ordering $\bigcup_{i\in I} P_i$. Thus by Zorn's Lemma there exists a maximal element (R, Q) in $\mathcal{M}$, and we will show that R is real closed.
Every element $a \in Q$ must be a square in R, otherwise the ordering could be extended to the proper extension $R(\sqrt{a})$ (Theorem 1.2.2), contradicting maximality. So $Q = R^2$, and thus each field extension with an ordering is automatically an order extension of R (cf. Remark 1.1.11 (iv)). By maximality of R there can thus not exist any proper real field extension of R, and thus R is real closed. □

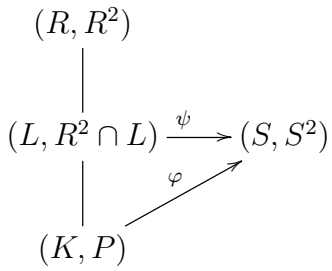
Our next goal is to show uniqueness of the real closure. We will need some auxiliary results first. The first lemma can already be seen as a weak version of the transfer principle of Tarski & Seidenberg, that we will later prove.

Lemma 1.4.3. *Let K be a field and R_1, R_2 two real closed extensions that induce the same ordering on K. Then for any $p \in K[t]$, the numbers of (different) roots of p in R_1 and R_2 coincide.*

Proof. By Theorem 1.3.3, the number of roots of p in R_1 and R_2 coincides with the signature of the Hermite matrix $\mathcal{H}(p)$ over the respective field. The entries of the Hermite matrix lie in K, however. Since R_1 and R_2 induce the same ordering on K, this signature is the same in both fields. □

We now need two results on extendability of order-embeddings.

Lemma 1.4.4. *Let R be a real closure of the ordered field (K, P), and $K \subseteq L \subseteq R$ an intermediate field with $[L : K] < \infty$. Let $\varphi\colon K \to S$ be an order-embedding into another real closed field S. Then φ admits an extension to L, i.e. an embedding $\psi\colon L \to S$ that respects the ordering $R^2 \cap L$ on L, with $\psi = \varphi$ on K.*



Proof. By the Primitive Element Theorem we have $L = K(\alpha)$ for some $\alpha \in L$. Let $p \in K[t]$ be the minimal polynomial of α over K. Then p as in R the zero α. By Lemma 1.4.3, p also has a zero in S, and thus there exists at least one extension $\psi\colon L \to S$ of φ. Let $\psi_1, \dots \psi_m$ be all of these extension, and assume none of them preserves the ordering $R^2 \cap L$). Then there are elements $b_1, \dots, b_m \in L$ with $b_i \in R^2$ and $\psi_i(b_i) < 0$ in S. We now consider the field $L' = L(\sqrt{b_1}, \dots, \sqrt{b_m}) \subseteq R$. None of the ψ_i extends to L', since squares must clearly be mapped to positive elements in S. Thus also φ cannot admit an extension to L' (because that would also be an extension of some ψ_i), contradicting the first part of the proof. □

Theorem 1.4.5. *Let (K, P) be an ordered field with a real closure R, and $\varphi\colon K \to S$ an order-respecting homomorphism into another real closed field S. Then there exists a unique extension $\psi\colon R \to S$ of φ, which automatically respects the ordering.*

Proof. First note that each homomorphism from L to S is automatically order-respecting, since positive elements are squares. Now consider the nonempty set

$$\mathcal{M} = \left\{(L, \psi) \mid K \subseteq L \subseteq R, \psi\colon L \to S, \psi(R^2 \cap L) \subseteq S^2, \psi = \varphi \text{ on } K\right\}.$$

As usually, it is partially ordered by the extension relation, and each chain admits an upper bound. Thus by Zorn's Lemma there exists a maximal element in $\mathcal{M}$, and by Lemma 1.4.4 it must fulfill $L = R$. So there exists an order-respecting extension $\psi\colon R \to S$.
For the uniqueness let $p \in K[t]$ be an irreducible polynomial. Then p has simple roots $\alpha_1 < \dots < \alpha_r$ in R, and by Lemma 1.4.3 it has roots $\beta_1 < \dots < \beta_r$ in S as well. Any extension ψ maps the α_i to the β_i, and since it is order-respecting, we

must have $\psi(\alpha_i) = \beta_i$ for all i. Since every element $\alpha \in R$ is the zero of some polynomial over K, ψ is uniquely determined. □

Corollary 1.4.6. *The real closure of an ordered field (K, P) is uniquely determined, up K-isomorphism. Even the isomorphism is uniquely determined.*

Proof. If R_1, R_2 are two real closures of (K, P), Theorem 1.4.5 guarantees the existence of two uniquely determined K-homomorphisms $\psi_1 \colon R_1 \to R_2$ and $\psi_2 \colon R_2 \to R_1$. Again by uniqueness we must have $\psi_2 \circ \psi_1 = \mathrm{id}_{R_1}$ (and vice versa), thus both homomorphisms are isomorphisms. □

Remark 1.4.7. So it makes sense to speak of *the real closure* of an ordered field, since it is uniquely determined up to isomorphism. We have seen that even the isomorphism is uniquely determined! It is known from the algebra course that also the algebraic closure of a field is uniquely determined up to isomorphism, but here several isomorphisms may exist. For example, both the identity and complex conjugation yield $\mathbb{R}$-automorphisms of $\mathbb{C}$. △

Example 1.4.8. Let

$$\mathbb{R}_0 := \{\alpha \in \mathbb{R} \mid \alpha \text{ algebraic over } \mathbb{Q}\}$$

be the relative algebraic closure of $\mathbb{Q}$ in $\mathbb{R}$. From Lemma 1.2.10 we know that $\mathbb{R}_0$ is real closed, and it is thus the real closure of $\mathbb{Q}$. So $\mathbb{R}_0$ is the smallest real closed field, i.e. it is contained in any other real closed field in a unique way. △

1.5 Semialgebraic Sets, the Projection Theorem, and the Transfer Principle of Tarski & Seidenberg

In the following let R always be a real closed field, and A an arbitrary subring of R. We set $\underline{x} = (x_1, \ldots, x_n)$. For polynomials $p_1, \ldots, p_m \in R[\underline{x}] = R[x_1, \ldots, x_n]$ we define

$$O_R(p_1, \ldots, p_m) := \{a \in R^n \mid p_1(a) > 0, \ldots, p_m(a) > 0\}$$

and

$$V_R(p_1, \ldots, p_m) := \{a \in R^n \mid p_1(a) = 0, \ldots, p_m(a) = 0\}\,.$$

In case it is clear which field R we have chosen, we often omit the index R.

Definition 1.5.1. (i) A subset of R^n is called A**-semialgebraic**, if it is a finite Boolean combination (unions, intersections, complements) of sets $O(p_1, \ldots, p_m)$ with $p_i \in A[\underline{x}]$. Instead of *R-semialgebraic* we also just say **semialgebraic**.
(ii) A set of the form $V(p_1, \ldots, p_m)$ with $p_i \in A[\underline{x}]$ is called A**-algebraic**. Instead of *R-algebraic* we also just say **algebraic**. △

Remark 1.5.2. (i) Every A-algebraic set is A-semialgebraic. The condition $p(a) = 0$ can indeed rewritten as

$$\neg\left(p(a) > 0\right) \wedge \neg\left((-p)(a) > 0\right).$$

In terms of sets, this corresponds to the following Boolean combination:

$$V(p) = O(p)^c \cap O(-p)^c.$$

(ii) In the definition of a semialgebraic set we can thus use conditions of the form

$$p(a) = 0,\ p(a) \geqslant 0,\ p(a) \leqslant 0,\ p(a) > 0,\ p(a) < 0$$

and Boolean combinations thereof. △

Lemma 1.5.3. *(i) Every A-algebraic set is of the form $V(p)$ for some $p \in A[\underline{x}]$.*
(ii) Every A-semialgebraic set $S \subseteq R^n$ has a description of the following form:

$$S = \bigcup_{i=1}^{r} \left(V(p_i) \cap O(q_{i1}, \ldots, q_{im_i})\right).$$

Proof. (i) follows directly from the observation

$$V(p_1, \ldots, p_m) = V(p_1^2 + \cdots + p_m^2),$$

cf. Definition 1.1.18. For (ii) one first checks that the class of all sets of this form is closed under finite Boolean combinations. Since it clearly contains all sets $O(p_1, \ldots, p_m)$, this proves the claim. □

Example 1.5.4. (i) The subset $\mathbb{Z} \subseteq \mathbb{R}$ is not semialgebraic. The graph of the sine function

$$\{(\alpha, \sin\alpha) \mid \alpha \in \mathbb{R}\} \subseteq \mathbb{R}^2$$

is not semialgebraic. The graph of the exponential function

$$\{(\alpha, e^\alpha) \mid \alpha \in \mathbb{R}\}$$

is not semialgebraic (Exercise 17).
(ii) Countable unions and intersections of semialgebraic sets are in general not semialgebraic (as can for example be seen with $\mathbb{Z}$). △

We can completely classify the semialgebraic subsets of the line:

Theorem 1.5.5. *The semialgebraic subsets of R are precisely the finite unions of intervals (closed, open, half-open, bounded and unbounded intervals with boundaries from R; in particular also single points).*

Proof. Obviously all intervals are semialgebraic. Conversely, the set of all finite unions of intervals is closed under finite Boolean combinations. So it is enough to show that each of the sets

$$O(p) = \{\alpha \in R \mid p(\alpha) > 0\}$$

for $p \in R[x]$ is such a finite union. By the Intermediate Value Theorem 1.2.11 (ii), a polynomial cannot change its sign in between two consecutive zeros. But a polynomial has only finitely many zeros, and thus $O(p)$ is indeed a finite union of (open) intervals. □

Theorem 1.5.6. *Let $p_1, \ldots, p_m \in A[x_1, \ldots, x_n]$ and*

$$\begin{aligned} p\colon R^n &\to R^m \\ a &\mapsto (p_1(a), \ldots, p_m(a)) \end{aligned}$$

the corresponding polynomial map. Then the inverse image $p^{-1}(T)$ of any A-semialgebraic (A-algebraic) set $T \subseteq R^m$ is again A-semialgebraic (A-algebraic).

Proof. For $q \in A[x_1, \ldots, x_m]$ we have

$$p^{-1}(O(q)) = \{a \in R^n \mid q(p_1(a), \ldots, p_m(a)) > 0\} = O(h)$$

where $h := q(p_1, \ldots, p_m) \in A[x_1, \ldots, x_n]$. The general case follows from the fact that taking inverse images is compatible with Boolean combinations. The case of an algebraic set is similar. □

Remark 1.5.7. (i) The polynomial *image* $p(V)$ of an algebraic set $V \subseteq R^n$ is in general not algebraic in R^m, and not even a Boolean combination of algebraic sets. For example, when projecting the set

$$\{(x, y) \in R^2 \mid y - x^2 = 0\}$$

onto the y-axis, one obtains $[0, \infty)$. But algebraic subsets of R are either finite or the full line R. Boolean combinations thereof are thus finite and cofinite sets.
(ii) The A-polynomial image of an A-semialgebraic set is again A-semialgebraic. This follows from the Projection Theorem, that we will now prove. △

We start with a technical lemma, saying that the signature of a matrix depends in a semialgebraic way on its coefficients:

Lemma 1.5.8. *For any $n \in \mathbb{N}$ and $k \in \mathbb{Z}$, the set*

$$\{M \in \mathrm{Sym}_n(R) \mid \operatorname{sign} M = k\} \subseteq M_n(R) \cong R^{n^2}$$

is $\mathbb{Z}$-semialgebraic. The semialgebraic description can be chosen independent of the real closed field R.

Proof. We prove the claim by induction on n, the case $n = 1$ is obvious. In the general case we proceed as in the diagonalization procedure for M as a bilinear form (a.k.a. the *symmetric Gauß Algorithm*). We write $M = (m_{ij})_{i,j}$ and can assume $M \neq 0$. After a suitable change of basis we can even assume $m_{11} \neq 0$. It is one of finitely many fixed base changes that will work, depending on which entry of M is nonzero. The matrix entries after such a base change can be computed $\mathbb{Z}$-polynomial from the initial entries. The different possible cases precisely translate to a union of semialgebraic sets.
After assuming $m_{11} \neq 0$, the vectors $e_i' := m_{11}e_i - m_{1i}e_1$ for $i = 2, \ldots, n$, together with e_1, form a basis R^n, with respect to which M has the form

$$\begin{pmatrix} m_{11} & 0 \\ 0 & M' \end{pmatrix}$$

for some symmetric matrix M' of size $n-1$. M' can be computed $\mathbb{Z}$-polynomial from M, by multiplication with the base change matrix from both sides. Now for M' the desired claim is true by induction hypothesis, and this also settles the general case. □

The following is one of the most important general results on semialgebraic sets.

Theorem 1.5.9 (Projection Theorem)**.** *Let $S \subseteq R^m \times R^n$ be an A-semialgebraic set, and $\pi\colon R^m \times R^n \to R^n; (x, y) \mapsto y$ the projection. Then $\pi(S) \subseteq R^n$ is again A-semialgebraic.*

Proof. It is enough to consider the case $m = 1$, the general case then follows by iteration. Since the image of a union is the union of the images, and by using Lemma 1.5.3 (ii), we can assume S to be of the form

$$\begin{aligned} S &= V(p) \cap O(q_1, \ldots, q_m) \\ &= \{(\alpha, a) \in R \times R^n \mid p(\alpha, a) = 0, q_1(\alpha, a) > 0, \ldots, q_m(\alpha, a) > 0\}, \end{aligned}$$

with $p, q_i \in A[t, \underline{x}]$. For $a \in R^n$ fixed, deciding whether $a \in \pi(S)$ holds means deciding whether there exists some $\alpha \in R$ with $(\alpha, a) \in S$. We can use the method of Hermite matrices for this. So we understand all polynomials as polynomials in the variable t, parametrized by the variables $\underline{x}$:

$$p = \sum_{i=0}^{d} p_i(\underline{x})t^i, \qquad q_j = \sum_{i=0}^{d_j} q_{ji}(\underline{x})t^i,$$

with $p_i, q_{ji} \in A[\underline{x}]$. For $a \in R^n$ and arbitrary $h \in A[t, \underline{x}]$ we set

$$h_a(t) := h(t, a) \in R[t].$$

Applying the method of Hermite matrices requires a case distinction, by the degree of p_a. So let

$$\Sigma_k := \{a \in R^n \mid \deg(p_a) = k\},$$

where we also allows for $k = -\infty$, in case $p_a \equiv 0$. Each of the sets Σ_k is A-semialgebraic in R^n, since it is defined by the vanishing or non-vanishing of the p_i. The full space R^n is a disjoint union of the Σ_k. So we are done, once we have shown that each set $\pi(S) \cap \Sigma_k$ is A-semialgebraic.
For $k = 0$ we have $\pi(S) \cap \Sigma_k = \emptyset$, since a polynomial of degree 0 has no zeros. The empty set is clearly A-semialgebraic.
So assume $k \geqslant 1$. For $a \in \Sigma_k$ we write $p_a = p_0(a) + p_1(a)t + \cdots + p_k(a)t^k$ with $p_k(a) \neq 0$. By Corollary 1.3.7 we have

$$a \in \pi(S) \Leftrightarrow \sum_{e \in \{1,2\}^m} \operatorname{sign} \mathcal{H}\left(\frac{1}{p_k(a)} p_a, q_a^e\right) > 0.$$

The entries of the Hermite matrices on the right are all integer polynomials in the $p_i(a)/p_k(a)$ and the $q_{ji}(a)$, see Section 1.3. The degree is bounded, depending on the degree of the appearing polynomials p, q_i. For computation of the signature we can clearly multiply all matrices with $p_k(a)^N$, for large enough (even) N. Then all entries are integer polynomials in the $p_i(a)$ and the $q_{ji}(a)$, and thus A-polynomial in a. Since all appearing signatures lie in between $-k$ and $+k$, there are only finitely many possibilities for these signatures to sum to a positive value. Each condition on the signature of a single matrix is A-semialgebraic in a, by Lemma 1.5.8. This proves the claim.
The remaining case is $k = -\infty$, i.e. $p_a = 0$. In this case we can add another equation to the description of S, without changing $\pi(S) \cap \Sigma_{-\infty}$. This reduces the problem to the previous case. A suitable equation is described in the following lemma. □

Lemma 1.5.10. *Let $q_1, \dots, q_m \in R[t]$ and set $q := q_1 \cdots q_m$. We consider*

$$p := (1 - q^2)q'.$$

If there is a point $\alpha \in R$ with $q_1(\alpha) > 0, \dots, q_m(\alpha) > 0$, then there is also such a point with $p(\alpha) = 0$, additionally.

Proof. The case that q (and thus all q_i) are constant is clear. We thus assume that q is not constant. Let $\alpha_1 < \alpha_2 < \cdots < \alpha_r$ be the zeros of q in R. In none of the intervals

$$(-\infty, \alpha_1), (\alpha_1, \alpha_2), \dots, (\alpha_{r-1}, \alpha_r), (\alpha_r, \infty)$$

any of the q_i changes its sign (Theorem 1.2.11 (ii)). So we are done, once we have shown that p has a zero in all of these intervals. In the bounded intervals, q' has a zero, by Rolle's Theorem1.2.12. Since $1 - q^2$ takes the value 1 at both α_1 and α_r, but is negative for large absolut values (q is not constant!), it must have a zero in both unbounded intervals as well. □

Corollary 1.5.11. *Every A-polynomial image of an A-semialgebraic set is again A-semialgebraic (cf. Remark 1.5.7 (ii)).*

Proof. Exercise 19. □

Remark 1.5.12. Note that the semialgebraic description of $\pi(S)$ in the projection theorem does only depend on the initial description of S, and *not* on the specific real closed field R. This is immediate from the proof, and admits a strong corollary known as **quantifier elimination**. △

Let us first introduce some notions from logic and model theory.

Definition 1.5.13. Again let A be a ring. A **prime formula over** A is a formula of the type

$$p(\underline{x}) > 0$$

where $p \in A[\underline{x}] = A[x_1, \dots, x_n]$ is a polynomial with coefficients from A. A general **formula over** A is then defined inductively. Every prime formula is a formula, and if φ, ψ are formulas, then so are

$$\varphi \wedge \psi, \quad \neg\varphi, \quad \exists x_i \, \varphi.$$

One can also use the well-known logical connectives:

$$\varphi \vee \psi, \quad \varphi \to \psi, \quad \forall x_i \, \varphi$$

as abbreviations for

$$\neg((\neg\varphi) \wedge (\neg\psi)), \quad \neg\varphi \vee \psi, \quad \neg(\exists x_i(\neg\varphi)).$$

We can also understand

$$p(\underline{x}) = 0 \text{ and } p(\underline{x}) \geqslant 0$$

as formulas, as abbreviations for

$$\neg(p(\underline{x}) > 0) \wedge \neg((-p)(\underline{x}) > 0) \text{ and } \neg(-p)(\underline{x}) > 0.$$

The expression $p(\underline{x}) = q(\underline{x})$ is short for $(p-q)(\underline{x}) = 0$.
The appearance of a variable x_i in a formula φ is called **free**, if it is not within the scope of a quantifier $\exists x_i$. Otherwise the appearance is called **bounded**. The set of all variables with a free appearance in φ is denoted $\mathrm{Fr}(\varphi)$. A formula φ with $\mathrm{Fr}(\varphi) = \emptyset$ is also called a **statement**. △

Now let R be a real closed extension field of the ring A. Then every A-formela φ with $\mathrm{Fr}(\varphi) \subseteq \{x_{i_1}, \ldots, x_{i_r}\}$ can be used to define a subset of R^r. In fact, just take all elements $a = (a_1, \ldots, a_r) \in R^r$, such that the formula φ is true in R, when every free appearance of each x_j is replaced by a_j. *Being true* of a statement is defined exactly as one would expect here (and we will not go into technical details thus). The set defined like this is denoted $\varphi(R)$:

$$\varphi(R) = \{a \in R^r \mid \varphi(a) \text{ holds in } R\}\,.$$

A statement is either true of false in the field R. The connection to the above is the following important observation:

Semialgebraic sets are precisely the sets $\varphi(R)$ for quantifier free formulas φ.

Example 1.5.14. (i) Let $A = \mathbb{Z}$ and

$$\varphi\colon \exists x_1\,(x_1x_2 = 1).$$

Then $\mathrm{Fr}(\varphi) = \{x_2\}$, and thus φ defines a subset $\varphi(R)$ of each real closed fiel R. On easily checks

$$\varphi(R) = R \setminus \{0\}.$$

(ii) Examples of statements are

$$\varphi_1\colon \forall x_1\ \left(x_1 > 0 \to \exists x_2\, x_2^2 = x_1\right)$$

and

$$\varphi_2\colon \exists x_1\,(x_1^2 = -1).$$

We already know that φ_1 holds in every real closed field, and φ_2 in none. △

We now obtain the following strong result.

Theorem 1.5.15 (Quantifier Elimination)**.** *Let A be a ring and φ a formula over A. Then there exists a quantifier free formula γ over A, with $\mathrm{Fr}(\varphi) = \mathrm{Fr}(\gamma)$ and*

$$\varphi(R) = \gamma(R)$$

for each real closed extension field R of A.

Proof. We can proceed inductively over the construction of φ. Prime formulas have no quantifiers, so there is nothing to show. Also the two constructions $\varphi \wedge \psi$ and $\neg\varphi$ do not add quantifiers. So finally assume $\varphi = \exists x_i\, \psi$ and that ψ is already quantifier free, by induction hypothesis. For each real closed extension field R of A, the set $\varphi(R)$ is the projection of the set $\psi(R)$ along the x_i-axis. The set $\psi(R)$ is semialgebraic, since ψ does not involve quantifiers. By the Projection Theorem 1.5.9 we obtain that also $\varphi(R)$ is semialgebraic, and can thus be described by a quantifier free formula γ. But we have already observed that γ can be chosen independent of R. This finishes the proof. □

This immediately implies the following strengthening of the above observation:

Semialgebraic sets are precisely the sets $\varphi(R)$ for arbitrary formulas φ.

Remark 1.5.16. In theory, quantifier elimination can be done algorithmically. As in the above proof, one proceeds inductively over the construction of the formula. Every existential quantifier is eliminated by the method described in the proof of the projection theorem. However, this method is not applicable in practice in general. △

Example 1.5.17. Consider the following formula over $\mathbb{Z}$:

$$\varphi\colon \exists t\; t^2 + xt + y = 0.$$

We have $\mathrm{Fr}(\varphi) = \{x, y\}$ and the set $\varphi(R) \subseteq R^2$ can be understood as the set of monic quadratic polynomials with a zero in R. It is well-known that a polynomial $t^2 + xt + y$ has a zero in R if and only if its *discriminant* $x^2 - 4y$ is a square in R, and this happens if and only if it is nonnegative. Thus the following quantifier free formula defines the same set as φ, over every real closed field:

$$x^2 - 4y \geqslant 0.$$ △

Quantifier elimination is a very strong statement. We can see this for example that the following corollary, known as the **Transfer Principle of Tarski & Seidenberg**.

Theorem 1.5.18 (Transfer Principle of Tarski & Seidenberg)**.** *Let K be a field with two real closed extension fields R_1, R_2, inducing the same ordering on K. Let φ be a statement over K. Then φ holds in R_1 if and only if it holds in R_2.*

Proof. Let γ be a quantifier free statement over K, which is equivalent to φ over each real closed field (Theorem 1.5.15). Now whether γ holds in R_i is determined already in K, since no quantifiers are involved. □

Example 1.5.19. (i) If a system of polynomial equations and inequalities (with polynomials over K) has a solution in some real closed extension field of K, then is has one in *each* such field (which induces the same ordering on K). The existence of a solution can indeed be formulate as a statement over K:

$$\exists x_1, \dots \exists x_n \colon \bigwedge_j p_j(\underline{x}) \geqslant 0 \wedge \bigwedge_j q_j(a) \neq 0 \wedge \bigwedge_j f_j(\underline{x}) = 0.$$

(ii) Statements such as the Intermediate Value Theorem 1.2.11 (ii) can be formulated as statements over $\mathbb{Z}$ (Exercise 18). Since every real closed field contains $\mathbb{Q}$ and induces the unique ordering there, we immediately obtain that the intermediate value theorem holds over any real closed field (however, we have used this fact in the above proofs leading to the transfer principle already, if one looks closely enough). But one can obtain many other important results for real closed fields easily in this fashion. △

Chapter 2

Globally Nonnegative Polynomials

In this chapter we will use our previous insights to study polynomials that are nonnegative at each point of $\mathbb{R}^n$. We will start with a positive solution to Hilbert's 17th Problem, and then examine some special cases more carefully.

2.1 Hilbert's 17th Problem

Let R be a real closed field. A polynomial $p \in R[\underline{x}] = R[x_1, \ldots, x_n]$ is called **nonnegative**, if it takes a nonnegative value at each point of R^n:

$$\forall a \in R^n : \quad p(a) \geqslant 0.$$

If one tries to write down a nonnegative polynomial explicitly, one usually comes up with something like $p = x_1^2$. A little more general, every **sum of squares**

$$p = q_1^2 + \cdots + q_m^2$$

with $q_i \in R[\underline{x}]$ is obviously a nonnegative polynomial. This rises the question whether these are all, or whether there exist nonnegative polynomials which are not sums of squares. Trying to find an explicit such example turns out to be very hard. However, already in 1888 Hilbert has shown that there *do* exist nonnegative polynomials that are not sums of squares. His proof was not constructive however, and it took until 1967 until an explicit such example was found by Motzkin. We will see this example later in these notes. Hilbert conjectured however that every nonnegative polynomial is a sum of squares of *rational functions*, i.e. fractions of polynomials. This conjecture, which became known as Hilbert's 17th Problem,

is indeed true, as was shown by Artin in 1926. Using the theory developed in the last chapter, we can now give a proof.

Theorem 2.1.1 (Hilbert's 17th Problem)**.** *A polynomial $p \in R[x_1, \dots, x_n]$ is nonnegative if and only if it is a sum of squares of rational functions, i.e. if there exist $q, q_1, \dots, q_m \in R[\underline{x}], q \neq 0$ with*

$$q^2 p = q_1^2 + \cdots + q_m^2.$$

Proof. "$\Leftarrow$": If $p(a) < 0$ for some $a \in R^n$, there would exists such a with $q(a) \neq 0$ additionally. But this would imply $(q^2 p)(a) < 0$, a contradiction to being a sum of squares.
"$\Rightarrow$": Let p be nonnegative. We will show that p, as an element of the field $R(\underline{x})$, is nonnegative for each field ordering. The statement then follows directly from Theorem 1.1.16.
Assume to the contrary that p is negative for a field ordering $\leqslant$ of $R(\underline{x})$, i.e. we have $p < 0$. We denote by $\widetilde{R}$ the real closure of $R(\underline{x})$ with respect to this ordering. Then in $\widetilde{R}$ the following formula over R is valid:

$$\exists x_1 \, \exists x_2 \, \cdots \, \exists x_n \quad p(x_1, \dots, x_n) < 0.$$

In fact, we can choose the variables x_i themselves, which are elements in $\widetilde{R}$. If the variables are plugged in for the variables, the polynomial remains unchanged, and it is negative as an element of $R(\underline{x})$ and thus of $\widetilde{R}$.
Since R is a subring both of the real closed fields R and $\widetilde{R}$, the Transfer Principle 1.5.18 implies that the statement also holds in R. So there exists $a \in R^n$ with $p(a) < 0$, contradicting nonnegativity of p. □

So for every nonnegative polynomial there exists an *algebraic certificate* that makes the nonnegativity obvious. However, the certificate involves *denominators*, which might be surprising, since p itself is a polynomial. We will examine to what extend denominators are necessary in the next section.

2.2 Sums of Squares of Polynomials

Let again R be a real closed field throughout this section. We start with the simple observation that *univariate* polynomials do not require denominators in the representation from Hilbert's 17th Problem.

Theorem 2.2.1. *Let $p \in R[t]$ be a univariate polynomial. Then p is nonnegative if and only if p is a sum of two squares of polynomials.*

Proof. Factor p into irreducible factors, which by Theorem 1.2.11 are all of the form $t - a$ or $(t - a)^2 + b^2$ with $a, b \in R, b \neq 0$. Polynomials of the second type are obviously sums of two squares. But each factor of the form $t-a$ must appear with an even power, since otherwise p would take negative values either left or right of a. Since the product of sums of two squares is again a sum of two squares (see the following remark), this proves the claim. □

Remark 2.2.2. In any commutative ring A we have for $a, b, c, d \in A$:

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2. \qquad \triangle$$

Another easy case is that of *quadratic polynomials*. We need the following lemma first:

Lemma 2.2.3. *Let $M \in \mathrm{Sym}_d(R)$ be a symmetric matrix. Then the following are equivalent:*

(i) $v^t M v \geqslant 0$ *for all* $v \in R^d$.

(ii) *There is* $S \in \mathrm{GL}_d(R)$ *with* $S^t M S = \mathrm{diag}(a_1, \ldots, a_d)$ *and* $a_i \geqslant 0$ *for all* i.

(iii) *All principal minors of M are nonnegative (in R).*

(iv) $M = \sum_{i=1}^{m} v_i v_i^t$ *for some* $m \in \mathbb{N}$ *and* $v_i \in R^d$.

(v) $M = \sum_{i=1}^{\mathrm{rank(M)}} v_i v_i^t$ *for some* $v_i \in R^d$.

Proof. For $R = \mathbb{R}$, the statement is well-known from linear algebra. The general case is proven in exactly the same way, and also follows directly from the transfer principle. □

Definition 2.2.4. A matrix $M \in \mathrm{Sym}_d(R)$ fulfilling the conditions from Lemma 2.2.3 is called **positive semidefinite**. △

Theorem 2.2.5. *Let $p \in R[x_1, \ldots, x_n]$ be a polynomial of degree 2. Then p is nonnegative if and only if it is a sum of squares of polynomials of degree 1.*

Proof. We can assume that p is *homogeneous* of degree 2. In fact we can homogenize p with the new variable x_0, i.e. we multiply each term with x_0 until is has degree 2. It is easy to see that the resulting polynomial is still nonnegative, if p was. If the homogenized polynomial is a sum of squares of linear polynomials, we obtain the desired representation for p by setting x_0 to zero.
Every homogeneous quadratic polynomial is of the form

$$p = \underline{x}^t M \underline{x} = \sum_{i,j=1}^{n} m_{ij} x_i x_j$$

for a symmetric matrix $M = (m_{ij})_{i,j} \in \mathrm{Sym}_n(R)$. Nonnegativity of p is now just condition (i) from Lemma 2.2.3, so M is positive semidefinite.
If we then write $M = \sum_i v_i v_i^t$ for certain $v_i \in R^n$ (condition (iv) from Lemma 2.2.3), we obtain

$$p = \underline{x}^t \left(\sum_i v_i v_i^t \right) \underline{x} = \sum_i (v_i^t \underline{x})^2.$$

This is the desired sums of square representation. □

Remark 2.2.6. Hilbert has shown that also each nonnegative polynomial in 2 variables of degree 4 is a sum of squares of polynomial. The proof is quite involved, so we do not cover it here. △

In all other cases there do exist nonnegative polynomials which are not sums of squares of polynomials. This was also already shown by Hilbert in 1888, but the first explicit such example is the **Motzkin polynomial** from 1967:

$$x^4y^2 + x^2y^4 - 3x^2y^2 + 1 \in \mathbb{Z}[x, y].$$

Theorem 2.2.7. *The Motzkin polynomial is nonnegative.*

Proof. 1st version: For positive numbers $a, b, c \geqslant 0$ we know that the geometric mean is smaller or equal than the arithmetic mean:

$$\sqrt[3]{abc} \leqslant \frac{1}{3}(a + b + c).$$

By setting $a = 1, b = x^4y^2, c = x^2y^4$ we obtain the result.

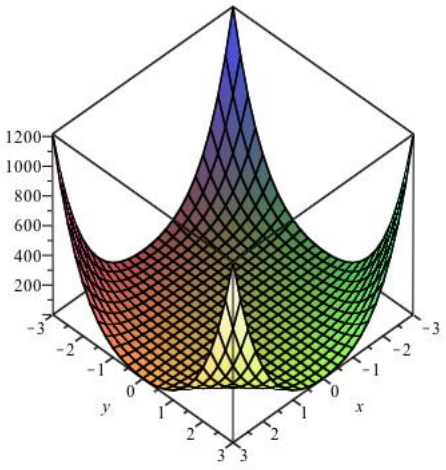


Figure 2.1: Graph of the Motzkin polynomial

2nd version: Let p be the Motzkin polynomial. The following identity is easily checked to be true:

$$(1+x^2)\cdot p = (1-x^2y^2)^2 + x^2(1-y^2)^2 + x^2y^2(1-x^2)^2.$$

This is a sums of squares representation with denomintor, which proves nonnegativity, as we have seen in Theorem 2.1.1.
3rd version: Again let p be the Motzkin polynomial. The following identity is easily checked to be true:

$$p(x^3,y^3) = q_1^2 + q_2^2 + q_3^2 + \tfrac{3}{4}q_4^2 + \tfrac{3}{4}q_5^2 + \tfrac{3}{4}q_6^2$$

with

$$q_1 = x^2y - \tfrac{1}{2}x^4y^5 - \tfrac{1}{2}x^6y^3, \quad q_2 = xy^2 - \tfrac{1}{2}x^3y^6 - \tfrac{1}{2}x^5y^4,$$
$$q_3 = 1 - \tfrac{1}{2}x^2y^4 - \tfrac{1}{2}x^4y^2, \quad q_4 = x^2y^4 - x^4y^2,$$
$$q_5 = x^3y^6 - x^5y^4, \quad q_6 = x^4y^5 - x^6y^3.$$

So $p(x^3,y^3)$ is a sum of squares of polynomials, and thus clearly nonnegative. Since every real number admits a third root, this implies nonnegativity also of p. □

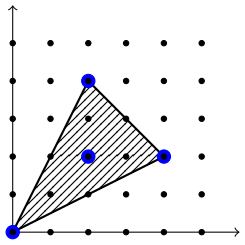
We now want to show that the Motzkin polynomial is not a sum of squares. This can be done very elementary, but for an elegant proof we need some preliminaries.

Definition 2.2.8. Let $p \in K[x_1, \ldots, x_n]$ be a polynomial over the field K. Write $p = \sum_{\alpha \in \mathbb{N}^n} p_\alpha \underline{x}^\alpha$, where $\underline{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and $p_\alpha \in K$. The **Newton polytope** of p is defined as

$$\mathcal{N}(p) := \operatorname{conv}\{\alpha \in \mathbb{N}^n \mid p_\alpha \neq 0\},$$

i.e. as the convex hull in $\mathbb{R}^n$ of the powers that appear in p. △

Example 2.2.9. The Newton polytope of the Motzkin polynomial is the convex hull of the points $(0,0), (2,2), (4,2), (2,4)$ in the plane:



△

The Newton polytope carries information about the behavior of the polynomial function at infinity. We will now make this precise. For $v \in \mathbb{R}^n$ and $d \in \mathbb{R}$ consider the halfspace

$$H_{v,d} := \{\alpha \in \mathbb{R}^n \mid \langle \alpha, v \rangle \geqslant d\}.$$

It is well-known that each polytope is the intersection of all its containing halfspaces.

Lemma 2.2.10. *Let R be a real closed field and $p \in R[x_1, \ldots, x_n]$. For $v \in \mathbb{Q}^n$ and $d \in \mathbb{Q}$ the following are equivalent:*

(i) $\mathcal{N}(p) \subseteq H_{v,d}$.

(ii) *For each $a \in R^n$ we have*

$$\lim_{t \searrow 0} \left| t^{-d} \cdot p(a_1 t^{v_1}, \ldots, a_n t^{v_n}) \right| < \infty,$$

i.e. along the curves $t \mapsto (a_1 t^{v_1}, \ldots, a_n t^{v_n})$ for $t \searrow 0$, the polynomial p grows at most of degree d.

Proof. Write $p = \sum_\alpha p_\alpha \underline{x}^\alpha$ with $p_\alpha \in R$.
$(i) \Rightarrow (ii)$: By assumption we have $\langle \alpha, v \rangle \geqslant d$ for all $\alpha \in \mathbb{N}^n$ with $p_\alpha \neq 0$. Thus

$$t^{-d} \cdot p(a_1 t^{v_1}, \ldots, a_n t^{v_n}) = \sum_\alpha p_\alpha \cdot a^\alpha \cdot t^{\langle \alpha, v \rangle - d}.$$

All exponents here are nonnegative, and this implies (ii).
$(ii) \Rightarrow (i)$: Assume there exists an exponent $\alpha \in \mathbb{N}^n$ with $p_\alpha \neq 0$ and $\langle \alpha, v \rangle = e < d$. Choose e minimal and let $\{\alpha^{(1)}, \ldots, \alpha^{(m)}\}$ be the set of all such exponents α. There is a point $a \in R^n$ with $\gamma := \sum_{i=1}^m p_{\alpha^{(i)}} a^{\alpha^{(i)}} \neq 0$. Then

$$t^{-d} \cdot p(a_1 t^{v_1}, \ldots, a_n t^{v_n}) = \gamma \cdot t^{e-d} + h$$

where all terms in h have a degree larger than $e - d$ in t. Since $e - d$ is negative, the expression is unbounded for $t \searrow 0$. □

Now the crucial corollary for sums of squares representations is the following:

Corollary 2.2.11. *For all $p, q, q_1, \ldots, q_m \in R[\underline{x}]$ we have:*

(i) $\mathcal{N}(p^2) = 2\mathcal{N}(p) \quad (:= \{2a \mid a \in \mathcal{N}(p)\})$.

(ii) *If p, q are nonnegative, then $\mathcal{N}(p) \subseteq \mathcal{N}(p+q)$.*

(iii) *For $p = q_1^2 + \cdots + q_m^2$ we have $\mathcal{N}(q_i) \subseteq \frac{1}{2}\mathcal{N}(p)$ for all i.*

Proof. (i): For $v \in \mathbb{Q}^n$ and $d \in \mathbb{Q}$ we have $\mathcal{N}(p^2) \subseteq H_{v,d}$ if and only if for all a the function $t^{-d} p(a_1 t^{v_1}, \ldots, a_n t^{v_n})^2$ remains bounded for $t \searrow 0$, by Lemma 2.2.10. But this is clearly equivalent to

$$t^{-d/2} p(a_1 t^{v_1}, \ldots, a_n t^{v_n})$$

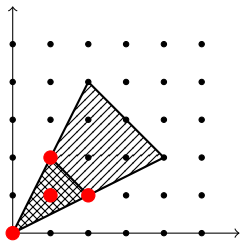
being bounded, and thus to $\mathcal{N}(p) \subseteq H_{v,d/2} = \frac{1}{2} H_{v,d}$, i.e. $2\mathcal{N}(p) \subseteq H_{v,d}$. But two polytopes that are contained in the same (rational) halfspaces coincide.
(ii): Let $\mathcal{N}(p+q) \subseteq H_{v,d}$, i.e.

$$t^{-d} \left(p(a_1 t^{v_1}, \ldots, a_n t^{v_n}) + q(a_1 t^{v_1}, \ldots, a_n t^{v_n}) \right)$$

remains bounded for $t \searrow 0$. Since p and q are nonegative, also the expression with p alone is bounded, and thus $\mathcal{N}(p) \subseteq H_{v,d}$. This again proves the claim.
(iii) is immediate from (i) and (ii). □

Theorem 2.2.12. *The Motzkin polynomial is not a sum of squares of polynomials.*

Proof. Assume $p = q_1^2 + \cdots + q_m^2$. Then $\mathcal{N}(q_i) \subseteq \frac{1}{2}\mathcal{N}(p)$ for all i, by Corollary 2.2.11. But this means that only the monomials $1, xy, x^2y, xy^2$ can appear in the q_i.



The monimial x^2y^2 thus arises in q_i^2 in a unique way, as the square of the monomial xy. In particular, its coefficient is nonnegative, and thus the coefficient of x^2y^2 in p would also be nonnegative. But this coefficient in p is actually -3, a contradiction. □

Remark 2.2.13. One can add an arbitrary positive constant to the Motzkin polynomial, and it will still not be a sum of squares of polynomials, by the same argument. So there exist *very positive* polynomials that are not sums of squares of polynomials. △

We will extend our examination of sums of squares a little further. Using Newton polytopes, we can give an elegant proof of the following result.

Lemma 2.2.14. *Let $q_1, \ldots, q_m \in R[x_1, \ldots, x_n]$ and $p = q_1^2 + \cdots + q_m^2$. Then*

$$\deg(p) = \max\{2 \cdot \deg(q_i) \mid i = 1, \ldots, m\}.$$

Proof. "$\leqslant$" is obvious. Now note that

$$\deg(q) = \max_{a \in \mathcal{N}(q)} \langle a, e \rangle$$

holds for $e = (1, \ldots, 1)$, for all polynomials q. Thus Corollary 2.2.11 (iii) implies

$$\deg(p) = \max_{a \in \mathcal{N}(p)} \langle a, e \rangle \geqslant \max_{a \in 2\mathcal{N}(q_i)} \langle a, e \rangle = 2 \cdot \max_{a \in \mathcal{N}(q_i)} \langle a, e \rangle = 2 \cdot \deg(q_i),$$

for all i. □

Now let us introduce **Gram matrices** of polynomials. By $R[\underline{x}]_d$ we denote the R-vectorspace of all polynomials in $\underline{x} = (x_1, \ldots, x_n)$ of degree $\leqslant d$. We again use the notation

$$\underline{x}^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n} \text{ and } |\alpha| := \alpha_1 + \cdots + \alpha_n.$$

The space $R[\underline{x}]_d$ has for example the monomial basis

$$\underline{X}_d := (\underline{x}^\alpha)_{\alpha \in \mathbb{N}^n, |\alpha| \leqslant d} = (1, x_1, x_2, \ldots, x_1^2, x_1x_2, \ldots),$$

and its dimension is $\Delta_d := \binom{n+d}{d}$. Now we consider the following linear map:

$$G \colon \mathrm{Sym}_{\Delta_d}(R) \to R[\underline{x}]_{2d}; \quad M \mapsto \underline{X}_d^t M \underline{X}_d.$$

For $M = (m_{\alpha\beta})_{|\alpha|,|\beta| \leqslant d}$ we thus have

$$G(M) = \sum_{|\alpha|,|\beta| \leqslant d} m_{\alpha\beta} \underline{x}^\alpha \underline{x}^\beta = \sum_{|\gamma| \leqslant 2d} \left(\sum_{\substack{\alpha+\beta=\gamma \\ |\alpha|,|\beta| \leqslant d}} m_{\alpha\beta} \right) \underline{x}^\gamma.$$

The map G is obviously surjective. For $p \in R[\underline{x}]_{2d}$ the set

$$G^{-1}(p) = \{M \mid \underline{X}_d^t M \underline{X}_d = p\}$$

is thus a nonempty affine subspace of $\mathrm{Sym}_{\Delta_d}(R)$.

Definition 2.2.15. The elements of $G^{-1}(p)$ are called the **Gram matrices of** p. $\triangle$

Example 2.2.16. $R[x_1, x_2]_1$ has the monomial basis $\underline{X}_1 = (1, x_1, x_2)$. The Gram map

$$G \colon \mathrm{Sym}_3(R) \to R[x_1, x_2]_2$$

thus has the following form:

$$\begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \mapsto (1, x_1, x_2) \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} \begin{pmatrix} 1 \\ x_1 \\ x_2 \end{pmatrix}$$
$$= a + 2bx_1 + 2cx_2 + dx_1^2 + 2ex_1x_2 + fx_2^2.$$

The polynomial $x_1^2 - 2x_1x_2 + x_2^2 + 2x_1 - 2x_2 + 1$ for example has the Gram matrix

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

In this example, every polynomial has a unique Gram matrix. But for $d \geqslant 2$ this fails. $\triangle$

Theorem 2.2.17. *A polynomial $p \in R[\underline{x}]_{2d}$ is a sums of squares of polynomials if and only if it admits a positive semidefinite Gram matrix M. In this case p is a sum of* $\mathrm{rank(M)}$ *many squares, in particular of at most $\binom{n+d}{d}$ many.*

Proof. Let $M \in \mathrm{Sym}_{\Delta_d}(R)$ be a positive semidefinite Gram matrix of p. By Lemma 2.2.3 we can find $\mathrm{rank}(M)$ many vectors $v_i \in R^{\Delta_d}$ with $M = \sum_i v_i v_i^t$. This implies

$$p = \underline{X}_d^t M \underline{X}_d = \sum_i \underline{X}_d^t v_i v_i^t \underline{X}_d = \sum_i (v_i^t \underline{X}_d)^2,$$

i.e. p is a sum of squares of polynomials, since $v_i^t \underline{X}_d \in R[\underline{x}]_d$.
Let conversely $p = \sum_i q_i^2$ be a sum of squares representation of p with $q_i \in R[\underline{x}]$. From Lemma 2.2.14 we obtain $q_i \in R[\underline{x}]_d$ for all i. We can thus write $q_i = v_i^t \underline{X}_d$ for some $v_i \in R^{\Delta_d}$, i.e. v_i is precisely the vector of coefficients of q_i in the monomial basis. We obtain

$$p = \sum_i (v_i^t \underline{X}_d)^2 = \sum_i \underline{X}_d^t v_i v_i^t \underline{X}_d = \underline{X}_d^t \left(\sum_i v_i v_i^t \right) \underline{X}_d,$$

and p thus possesses the positive semidefinite Gram matrix $\sum_i v_i v_i^t$. □

Example 2.2.18. (i) The (only) Gram matrix of $p = x_1^2 - 2x_1x_2 + x_2^2 + 2x_1 - 2x_2 + 1$ from Example 2.2.16 is positive semidefinite. We can even write it as vv^t with $v = (1, 1, -1)^t$. This implies $p = (x_1 - x_2 + 1)^2$.
(ii) Changing this slightly to $p = x_1^2 - 2x_1x_2 + x_2^2 + 2x_1 - 2x_2$, the (only) Gram matrix of p has a zero in the upper left corner. The upper left 2×2 minor is -1, and therefore p is not a sum of squares. △

Remark 2.2.19. We have seen that each sums of squares representation of p leads to a positive semidefinite Gram matrix. Conversely a positive semidefinite Gram matrix can lead to different sums of squares representations, since the decomposition $M = \sum_i v_i v_i^t$ is not unique in general. However, the positive semidefinite Gram matrices correspond to equivalence classes of similar sums of squares representations, when defined in the right sense. △

Sometimes the Newton polytope gives a better bound on the number of squares needed in a representation of a polynomial.

Theorem 2.2.20. *Let $p \in R[x_1, \ldots, x_n]$ be a sum of squares of polynomials, and let*

$$r := \left| \tfrac{1}{2}\mathcal{N}(p) \cap \mathbb{N}^n \right|$$

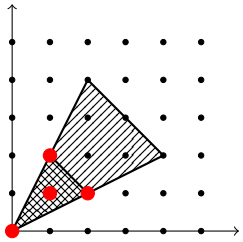
be the number of integer lattice points in $\frac{1}{2}\mathcal{N}(p)$. Then p is a sum of r squares of polynomials.

Proof. If $p = q_1^2 + \cdots + q_m^2$ for some $q_i \in R[\underline{x}]$, then in each q_i we can only have exponents from $\frac{1}{2}\mathcal{N}(p)$, by Theorem 2.2.11 (iii). So if we write

$$q_i = v_i^t \underline{X}_d,$$

in all of the v_i we have at most the same r entries nonzero. Thus the rank of $M = \sum_i v_i v_i^t$ is at most r, and since M is a positive semidefinite Gram matrix of p, the claim follows from Theorem 2.2.17. □

Example 2.2.21. If a sum of squares has the same Newton polytope as the Motzkin polynomial, it is a sum of at most 4 squares:



The general upper bound from Theorem 2.2.17 is only $\binom{2+3}{3} = 10$ here. △

In the following chapters we will examine polynomials that are nonnegative not globally, but only on certain semialgebraic subsets of $\mathbb{R}^n$. For this we will first develop the theory of ordered rings.

Chapter 3

Ordered Rings

Let A be a commutative ring with 1 throughout this chapter. Ring homomorphisms are assumed to map 1 to 1. If A is an integral domain, we denote its field of fractions by $\mathrm{Quot}(A)$. Again we will denote by R a real closed field.

3.1 Preorderings, Orderings and the Real Spectrum

Definition 3.1.1. A **preordering** on A is a subset $T \subseteq A$ with

$$T + T \subseteq T, \quad T \cdot T \subseteq T, \quad A^2 \subseteq T, \quad -1 \notin T.$$

The set $T \cap -T$ is called the **support** of T and denoted by $\mathrm{supp}(T)$. $\triangle$

The definition of a preordering is the same as for fields. However, there are differences regarding its properties.

Remark 3.1.2. (i) The set ΣA^2 of sums of squares in A is a preordering if and only if $-1 \notin \Sigma A^2$. In that case it is again the smallest preordering, i.e. contained in any other preordering.
(ii) If $\frac{1}{2} \in A$, then every element from A is a difference of two squares, as in Remark 1.1.11. The condition $-1 \notin T$ can thus be replaced by $T \neq A$ again.
(iii) Let $A = R[\underline{x}] = R[x_1, \ldots, x_n]$ be the polynomial ring over the real closed field R. For every nonempty subset $S \subseteq R^n$ we obtain the preordering

$$T_S := \{p \in R[\underline{x}] \mid p(a) \geqslant 0 \, \forall a \in S\}.$$

We have

$$\mathrm{supp}(T_S) = T_S \cap -T_S = \{p \in R[\underline{x}] \mid p(a) = 0 \, \forall a \in S\}.$$

For certain subsets S the support thus contains more than just the zero polynomial. This cannot happen in fields, as we have shown in Remark 1.1.11. Its proof in fact used division by elements, which is impossible in general rings. $\triangle$

Lemma 3.1.3. *If $T \subseteq A$ is a preordering with $T \cup -T = A$, then $\operatorname{supp}(T)$ is an ideal in A.*

Proof. Set $\mathfrak{p} := T \cap -T$. Then $0 \in \mathfrak{p}$ and $\mathfrak{p} + \mathfrak{p} \subseteq \mathfrak{p}$ is clear. We also clearly have $\pm T \cdot \mathfrak{p} \subseteq \mathfrak{p}$. From $A = T \cup -T$ thus follows $A \cdot \mathfrak{p} \subseteq \mathfrak{p}$. $\square$

Definition 3.1.4. An **ordering** on A is a preordering P, for which $P \cup -P = A$ holds, and for which $\operatorname{supp}(P)$ is a prime ideal of A. $\triangle$

Example 3.1.5. (i) The preordering T_S of $R[\underline{x}]$ from Remark 3.1.2 is an ordering if and only if $|S| = 1$.
(ii) If $A = K$ is a field, the new notion of an ordering coincides with the old one. By Remark 1.1.11 (iii) we in fact have $\operatorname{supp}(P) = \{0\}$, which is a prime ideal in K (the only one).
(iii) If $\varphi\colon A \to B$ is a ring homomorphism and $P \subseteq B$ an ordering, then $\varphi^{-1}(P)$ is an ordering of A. We have $\operatorname{supp}(\varphi^{-1}(P)) = \varphi^{-1}(\operatorname{supp}(P))$. In particular, for each integral domain A, we obtain via the embedding

$$A \subseteq \operatorname{Quot}(A)$$

an ordering with support $\{0\}$, for every field ordering of $\operatorname{Quot}(A)$
(iv) Let $P \subseteq A$ be an ordering with $\mathfrak{p} := \operatorname{supp}(P)$. Let $A/\mathfrak{p}$ be the factor ring modulo $\mathfrak{p}$ and $\pi_{\mathfrak{p}}\colon A \to A/\mathfrak{p}$ the canonical projection. Then $\overline{P} := \pi_{\mathfrak{p}}(P)$ is an ordering of $A/\mathfrak{p}$ with $\operatorname{supp}(\overline{P}) = \{0\}$ (Exercise 25).
(v) Consider the embedding $R[t] \subseteq R(t)$. On $R(t)$ we have the orderings

$$P_{a_+}, P_{a_-}, P_\infty, P_{-\infty},$$

see Example 1.1.4 (ii). These orderings induce orderings on $R[t]$ with support $\{0\}$. For example, we have

$$P_{a+} = \{p \in R[t] \mid \exists \varepsilon > 0 :\ p > 0 \text{ auf } (a, a+\varepsilon)\} \cup \{0\}.$$

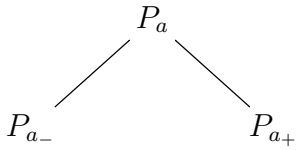
But on $R[t]$ we have more orderings! One example is

$$P_a = \{p \in R[t] \mid p(a) \geqslant 0\}$$

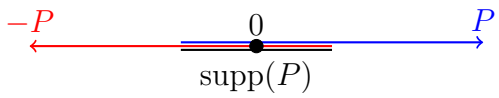
where we have

$$\operatorname{supp}(P_a) = \{p \in R[t] \mid p(a) = 0\} = (t - a).$$

We now observe $P_{a+} \subsetneq P_a$ and $P_{a-} \subsetneq P_a$. For example, we have $t - a \in P_a \setminus P_{a-}$ and $a - t \in P_a \setminus P_{a+}$.



Such phenomena cannot occur in fields, as we have seen in Remark 1.1.11 (iv). But the ordering P_a does indeed not come from $R(t)$.
(vi) As for fields, one can understand P as the set of elements larger than zero, with respect to a binary order relation $\leqslant$ on A. The axioms then translate to certain compatibility properties with the ring operations, similar to the case of a field. But now it can happen that both $a \leqslant 0$ snd $a \geqslant 0$ holds, without a being zero (exactly if $a \in \operatorname{supp}(P)$.) The relation is thus not necessarily antisymmetric. If we visualize the ordered ring by a line, it thus looks as follows:



△

Lemma 3.1.6. *Let P, P', P'' be orderings on A. We then have:*

(i) $P \subseteq P' \Rightarrow \operatorname{supp}(P) \subseteq \operatorname{supp}(P')$.

(ii) $P \subseteq P', \operatorname{supp}(P) = \operatorname{supp}(P') \Rightarrow P = P'$.

(iii) $P \subseteq P', P \subseteq P'' \Rightarrow P' \subseteq P''$ *or* $P'' \subseteq P'$.

Proof. (i) is obvious. For (ii) let $a \in P' \setminus P$. Then $-a \in P \subseteq P'$ and thus $a \in \operatorname{supp}(P') = \operatorname{supp}(P) \subseteq P$, a contradiction. For (iii) let $a \in P' \setminus P''$ and $b \in P'' \setminus P'$. Set $c = a - b$. Now $c \in P$ would imply $c \in P''$ and thus $a \in P''$, a contradiction. But on the other hand, $-c \in P$ would imply $-c \in P'$ and thus $b \in P'$, also a contradiction. This yields a total contradiction to $P \cup -P = A$. □

As for fields, we now examine how preorderings on rings can be enlarged.

Lemma 3.1.7. *Let $T \subseteq A$ be a preordering and $a, b \in A$ with $ab \in -T$. Then either $T + aT$ or $T + bT$ is again a preordering.*

Proof. We only have to show that -1 is not contained in one of the two potential preorderings. So assume that we have identities

$$-1 = t_1 + as_1 \text{ and } -1 = t_2 + bs_2,$$

for certain $t_1, t_2, s_1, s_2 \in T$. This implies

$$(1 + t_1)(1 + t_2) = abs_1s_2$$

and thus

$$-1 = t_1 + t_2 + t_1t_2 - abs_1s_2 \in T,$$

a contradiction. □

Theorem 3.1.8. *Every preordering of A is contained in an ordering of A.*

Proof. Let T be a preordering. Just as in the proof of Theorem 1.1.15 we use Zorn's Lemma to choose a maximal preordering that contains T. Now let $a \in A$. Since $a(-a) = -a^2 \in -P$, we obtain from Lemma 3.1.7 that either $P + aP$ or $P - aP$ is a preordering. Maximality of P thus implies either $a \in P$ or $-a \in P$. So we have shown $P \cup -P = A$. Lemma 3.1.3 now implies that $\mathfrak{p} := \mathrm{supp}(P)$ is an ideal, and what remains to show is the prime ideal property. So assume $a, b \in A$ fulfill $ab \in \mathfrak{p}$, but $a \notin \mathfrak{p}$. So we either have $a \notin P$ or $-a \notin P$. Let's assume w.l.o.g. that $a \notin P$ (just replace a by $-a$ in the other case). Then $P + aP$ is not a preordering anymore, by maximality of P. So Lemma 3.1.7 and maximality of P imply:

$$\begin{aligned} ab \in -P \quad &\Rightarrow \quad P + bP \text{ is a preordering } \Rightarrow \quad P = P + bP \ni b \\ a(-b) \in -P \quad &\Rightarrow \quad P - bP \text{ is a preordering } \Rightarrow \quad P = P - bP \ni -b \end{aligned}$$

But this just means $b \in \mathfrak{p}$. □

Corollary 3.1.9. *A ring possesses an ordering if and only if $-1 \notin \Sigma A^2$.*

Definition 3.1.10. (i) A ring is called **semireal** if $-1 \notin \Sigma A^2$. It is called **real** if $a_1^2 + \cdots + a_m^2 = 0$ always implies $0 = a_1 = \ldots = a_m$.
(ii) An ideal $I \subseteq A$ is called **semireal/real**, if A/I is semireal/real. △

Remark 3.1.11. (i) Real implies semireal.
(ii) For fields, the notions real and semireal coincide, as we have already seen in Definition 1.1.18.

(iii) For rings, the two notions are not equivalent. For example, consider

$$A = \mathbb{R}[x, y]/(x^2 + y^2).$$

Then A is not real, since $x^2+y^2 = 0$ holds in A, where $x, y \neq 0$. But A is semireal. To see this, note that elements from A can be evaluated at the origin, and sums of squares will be nonnegative there.
(iv) A ring is semireal if and only if it admits an ordering (Corollary 3.1.9). An integral domain is real if and only if it admits an ordering P with $\operatorname{supp}(P) = \{0\}$. To see this, let first be P such an ordering, and assume $a_1^2 + \cdots + a_m^2 = 0$. If we resolve for a_i^2 we get $a_i^2 \in \operatorname{supp}(P) = \{0\}$ and thus $a_i = 0$ for all i, using that A does not have zero divisors. Let conversely A be real. Then $\operatorname{Quot}(A)$ is a real field, as is easily checked. So $\operatorname{Quot}(A)$ admits an ordering (which has trivial support), and this induces an ordering on A with trivial support.
(v) An ideal is semireal if and only if it is contained in the support of an ordering. A prime ideal is real if and only if it is the support of an ordering, by (iv). △

Proposition 3.1.12. *Let A be an integral domain and $K = \operatorname{Quot}(A)$. Then the field orderings Q of K are in bijection with the ring orderings P of A with $\operatorname{supp}(P) = \{0\}$. The bijection is as follows:*

$$\begin{aligned} Q &\mapsto Q \cap A \\ P &\mapsto \operatorname{Quot}(P) := \left\{ \frac{a}{b} \mid ab \in P \right\}. \end{aligned}$$

Proof. Exercise 26. □

For a general prime ideal $\mathfrak{p}$ of A we consider its **residue field**

$$K_{\mathfrak{p}} := \operatorname{Quot}(A/\mathfrak{p}).$$

There is a natural ring homomorphism

$$\rho_{\mathfrak{p}} \colon A \xrightarrow{\pi_{\mathfrak{p}}} A/\mathfrak{p} \xrightarrow{\iota_{\mathfrak{p}}} K_{\mathfrak{p}}$$

where $\pi_{\mathfrak{p}}$ is the canonical projection (with kernel $\mathfrak{p}$), and $\iota_{\mathfrak{p}}$ is the inclusion into the die quotient field. The following results shows that the notion of a ring ordering can be reduced completely to the field case:

Theorem 3.1.13. *There is a bijection between the set of all ring orderings P of A, and the set of all pairs $(\mathfrak{p}, Q)$, where $\mathfrak{p}$ is a prime ideal of A and Q is a field ordering of its residue field $K_\mathfrak{p}$. The bijection is as follows (using the notation from Proposition 3.1.12):*

$$\begin{aligned} P &\mapsto (\mathfrak{p} := \operatorname{supp}(P), \operatorname{Quot}(\pi_\mathfrak{p}(P)) \\ (\mathfrak{p}, Q) &\mapsto \rho_\mathfrak{p}^{-1}(Q). \end{aligned}$$

Proof. For a ring ordering P of A we already know that $\operatorname{Quot}(\pi_\mathfrak{p}(P))$ is an ordering of the field $K_\mathfrak{p}$. This follows from Example 3.1.5 (iv) and Proposition 3.1.12. Conversely, for any ordering Q of the field $K_\mathfrak{p}$, the inverse image $\rho_\mathfrak{p}^{-1}(Q)$ is an ordering of the ring A, as seen in Example 3.1.5 (iii).
We now show that both constructions are mutually inverse. First take some P and set $\mathfrak{p} := \operatorname{supp}(P)$. Then

$$\rho_\mathfrak{p}^{-1}(\operatorname{Quot}(\pi_\mathfrak{p}(P)) = \pi_\mathfrak{p}^{-1}(\pi_\mathfrak{p}(P)) = P + \mathfrak{p} = P.$$

For the first equation we have used Proposition 3.1.12. If conversely $(\mathfrak{p}, Q)$ is given, we have

$$P := \rho_\mathfrak{p}^{-1}(Q) = \pi_\mathfrak{p}^{-1}(\iota_\mathfrak{p}^{-1}(Q)).$$

Since $\iota_\mathfrak{p}^{-1}(Q)$ is an ordering of $A/\mathfrak{p}$ with support $\{0\}$ by Proposition 3.1.12, the support of P is exactly $\mathfrak{p}$ (cf. Example 3.1.5 (iii)). If we apply $\pi_\mathfrak{p}$ to P, we obviously get $\iota_\mathfrak{p}^{-1}(Q)$, and by Proposition 3.1.12 we are done. □

Definition 3.1.14. Let A be a ring. The set of all orderings of A is called the **real spectrum of** A:

$$\begin{aligned} \operatorname{Sper}(A) :&= \{P \subseteq A \mid P \text{ ordering}\} \\ &= \{(\mathfrak{p}, Q) \mid \mathfrak{p} \text{ prime ideal}, Q \text{ ordering of } K_\mathfrak{p}\}. \end{aligned}$$ △

Example 3.1.15. (i) A real closed field R has exactly one ordering, so $\operatorname{Sper}(R)$ is a singleton. The same is true for $\operatorname{Sper}(\mathbb{Q})$.
(ii) Let $A = R[t]$. We have already constructed the orderings

$$P_{-\infty}, P_{a_-}, P_a, P_{a_+}, P_\infty$$

in Example 3.1.5 (v). Now let $P = (\mathfrak{p}, Q)$ be an arbitrary ordering of A. Since A is a principal ideal domain, every ideal is of the form (p) for some $p \in A$. Prime ideals are generated by irreducible polynomials (or zero). If $\mathfrak{p} = (0)$,

then by Proposition 3.1.12 the ordering P comes from $R(t)$, and is thus one of the $P_{-\infty}, P_{a-}, P_{a+}, P_{\infty}$. If $\mathfrak{p} = ((t-a)^2 + b^2)$ with $b \neq 0$, then

$$K_{\mathfrak{p}} = A/\mathfrak{p} = R[i]$$

which is easily checked. Thus $K_{\mathfrak{p}}$ is not real. For $\mathfrak{p} = (t-a)$ we have

$$K_{\mathfrak{p}} = A/\mathfrak{p} = R,$$

and Q is thus uniquely determined. Here, $\rho_{\mathfrak{p}} \colon A \to R$ is just the point evaluation in a, and thus $P = \rho_{\mathfrak{p}}^{-1}(Q) = P_a$. So we indeed already know all orderings of A:

$$\operatorname{Sper}(R[t]) = \left\{ P_{-\infty}, P_{a-}, P_a, P_{a+}, P_{\infty} \mid a \in R \right\}.$$

(iii) For $A = \mathbb{Z}$ one can check directly that $\Sigma\mathbb{Z}^2$ is the only ordering. Alternatively, we can use Theorem 3.1.13 as follows. The prime ideals of $\mathbb{Z}$ are (0) and (p) with $p \in \mathbb{Z}$ prime. We have $K_{(p)} = \mathbb{Z}/(p)$, and since $\operatorname{char}(K_{(p)}) \neq 0$, this field is not real. We further have $K_{(0)} = \mathbb{Q}$, and here we have the single ordering $\Sigma\mathbb{Q}^2$. Thus the only ordering of $\mathbb{Z}$ is $\Sigma\mathbb{Q}^2 \cap \mathbb{Z} = \Sigma\mathbb{Z}^2$. △

We will now introduce a point of view that might look very abstract at first, but has some nice conceptual advantages later on. Although the approach is strictly speaking not necessary to formulate the coming results, it makes them look much more illuminating.
We can understand elements of A as functions on $\operatorname{Sper}(A)$, even if A is an abstract ring. For $a \in A$ and $P = (\mathfrak{p}, Q) \in \operatorname{Sper}(A)$ we define

$$\hat{a}(P) = \hat{a}(\mathfrak{p}, Q) = \rho_{\mathfrak{p}}(a) \in K_{\mathfrak{p}}.$$

So when applying $\hat{a}$ to P, the element $a \in A$ itself is mapped to the residue field of $\mathfrak{p} = \operatorname{supp}(P)$, via the canonical homomorphism. Note that the image $\hat{a}(P)$ might lie in different fields, depending on P. To make it uniform, we must say

$$\hat{a} \colon \operatorname{Sper}(A) \to \bigcup_{\mathfrak{p} \text{ prime}} K_{\mathfrak{p}}.$$

Also note that the element $\hat{a}(P)$ only depends on the support $\mathfrak{p} = \operatorname{supp}(P)$ of P.

Example 3.1.16. (i) For a real field K we know that $\mathfrak{p} = \operatorname{supp}(P) = \{0\}$ holds for all $P \in \operatorname{Sper}(K)$. Thus always $K_{\mathfrak{p}} = K$ and $\rho_{\mathfrak{p}} = \mathrm{id}$. We obtain

$$\begin{aligned} \hat{a} \colon \operatorname{Sper}(K) &\to K \\ P &\mapsto a \end{aligned}$$

for all $a \in K$, i.e. each element induces a constant map.
(ii) For any real closed field R, we can understand R^n as a subset of the real spectrum $\operatorname{Sper} R[x_1, \ldots, x_n]$ of the polynomial ring, by identifying $a \in \mathbb{R}^n$ with

$$P_a = \{p \in R[\underline{x}] \mid p(a) \geqslant 0\}.$$

In this setup we have

$$\mathfrak{p} = \operatorname{supp}(P_a) = \{p \in R[\underline{x}] \mid p(a) = 0\} = (x_1 - a_1, \ldots, x_n - a_n)$$

and thus $\rho_{\mathfrak{p}} \colon R[\underline{x}] \to R$ is just point evaluation in a. So

$$\hat{p}(P_a) = p(a),$$

i.e. the function $\hat{p}$ coincides on R^n with the polynomial function p. But note that $\operatorname{Sper} R[\underline{x}]$ has more elements than just the P_a, on which $\hat{p}$ is also defined, and on which it might take values in other fields.
(iii) In case $A = R[t]$ we have determined $\operatorname{Sper}(A)$ completely. We have elements P_a with $a \in R$, and $\hat{p}(P_a) = p(a)$ holds, as shown in (ii). But for P_{a_+} we have $\mathfrak{p} = \operatorname{supp}(P_{a_+}) = (0)$, thus $K_{\mathfrak{p}} = R(t)$ and $\rho_{\mathfrak{p}} \colon R[t] \to R(t)$ is just the embedding. The same is true for the orderings P_{a_-}, $P_{-\infty}$ and P_∞. For $p \in R[t]$ we thus have

$$\begin{aligned} \hat{p} \colon \operatorname{Sper}(R[t]) &\to R \cup R(t) \\ P_a &\mapsto p(a) \in R \\ P_{a_+}, P_{a_-}, P_{-\infty}, P_\infty &\mapsto p \in R(t). \end{aligned}$$ △

For the just defined functions we can now define a notion of positivity. For $a \in A$ and an ordering $P = (\mathfrak{p}, Q)$, $\hat{a}(P)$ is an element of $K_{\mathfrak{p}}$, and Q is and ordering of this field. We now define

$$\begin{aligned} \hat{a}(P) > 0 \quad &:\Leftrightarrow \quad \hat{a}(P) >_Q 0 \text{ in } K_{\mathfrak{p}} \\ \hat{a}(P) \geqslant 0 \quad &:\Leftrightarrow \quad \hat{a}(P) \geqslant_Q 0 \text{ in } K_{\mathfrak{p}} \\ \hat{a}(P) = 0 \quad &:\Leftrightarrow \quad \hat{a}(P) = 0 \text{ in } K_{\mathfrak{p}}. \end{aligned}$$

Here we now really use Q, whereas for the definition of $\hat{a}(P)$ only $\mathfrak{p} = \operatorname{supp}(P)$ was relevant. For the reader who does not like these abstract function approach, the same can state without it. From the correspondence in Theorem 3.1.13 we get

$$\begin{aligned} \hat{a}(P) > 0 \quad &\Leftrightarrow \quad a \notin -P \\ \hat{a}(P) \geqslant 0 \quad &\Leftrightarrow \quad a \in P \\ \hat{a}(P) = 0 \quad &\Leftrightarrow \quad a \in \operatorname{supp}(P). \end{aligned}$$

Remark 3.1.17. Understanding elements from a ring as a function on $\mathrm{Sper}(A)$, although quite abstract and strictly speaking not necessary, has some great advantages. For example, we can often do computations in the ordered fields $(K_{\mathfrak{p}}, Q)$, with which we are already very familiar. But more importantly, it allows to understand some results as geometric Positivstellensätze, where the more elementary formulation wouldn't make this clear.
Theorem 1.1.15 and Theorem 1.1.16 from the first chapter are such examples. For the preordering of sums of squares in a field, the initial formulation was

$$\bigcap_{P \text{ ordering}} P \;=\; \Sigma K^2.$$

In words: if an element belongs to every ordering of K, then it is a sum of squares (and vice versa). This is algebraic and quite elementary, but does not admit a direct geometric interpretation. But now we can also look at it as follows. An element a is contained in every ordering if and only if the function $\hat{a}$ is nonnegative on the whole of $\mathrm{Sper}(K)$. So suddenly the result looks like an abstract Positivstellensatz:

$$\hat{a} \geqslant 0 \text{ on } \mathrm{Sper}(K) \;\Leftrightarrow\; a \in \Sigma K^2.$$

For Hilbert's 17th Problem we applied this to $K = R(\underline{x})$, and we could even relax the abstract positivity on $\mathrm{Sper}\,(R(\underline{x}))$ on the left to a much more concrete one, namely positivity on R^n. This turned the abstract Positivstellensatz into a concrete Positivstellensatz (Hilbert's 17th Problem):

$$p \geqslant 0 \text{ on } R^n \;\Leftrightarrow\; p \in \Sigma R(\underline{x})^2.$$

Reducing the abstract positivity to concrete positivity was the hardest part in the proof, we had to use the transfer principle and knowledge about real closures. However, this hard part also works for the polynomial ring, as we will now show, in a slightly more general version even. △

Theorem 3.1.18. *Let R be a real closed field and $p_1, \ldots, p_r, q_1, \ldots, q_s,\ f_1, \ldots, f_t \in R[x_1, \ldots, x_n]$. Then the following are equivalent:*

(i) There exists $a \in R^n$ with

$$\begin{aligned} p_1(a) \geqslant 0, \ldots, p_r(a) &\geqslant 0 \\ q_1(a) \neq 0, \ldots, q_s(a) &\neq 0 \\ f_1(a) = 0, \ldots, f_t(a) &= 0. \end{aligned}$$

(ii) There exists $P \in \operatorname{Sper}(R[\underline{x}])$ *with*

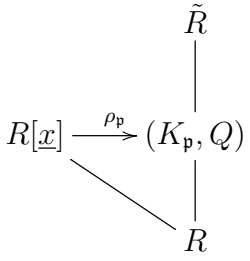
$$\begin{aligned}\hat{p}_1(P) \geqslant 0, \ldots, \hat{p}_r(P) &\geqslant 0\\ \hat{q}_1(P) \neq 0, \ldots, \hat{q}_s(P) &\neq 0\\ \hat{f}_1(P) = 0, \ldots, \hat{f}_t(P) &= 0.\end{aligned}$$

In particular, if some $p \in R[\underline{x}]$ *is nonnegative as a function on* R^n*, then* $\hat{p}$ *is nonnegative as a function on* $\operatorname{Sper}(R[\underline{x}])$*.*

Proof. $(i) \Rightarrow (ii)$ is clear from the fact that R^n embeds into $\operatorname{Sper}(R[\underline{x}])$, by identifying a with

$$P_a = \{p \in R[\underline{x}] \mid p(a) \geqslant 0\}.$$

For $(ii) \Rightarrow (i)$ let $P = (\mathfrak{p}, Q)$ be given as claimed. Let $\rho_\mathfrak{p} \colon R[\underline{x}] \to K_\mathfrak{p}$ be the homomorphism to the residue field and $\tilde{R}$ the real closure of $K_\mathfrak{p}$ with respect to Q:



In $K_\mathfrak{p}$ and thus also in $\tilde{R}$ we then have:

$$\begin{aligned}0 \leqslant \hat{p}_j(P) &= \rho_\mathfrak{p}(p_j) = p_j(\rho_\mathfrak{p}(x_1), \ldots, \rho_\mathfrak{p}(x_n))\\ 0 \neq \hat{q}_j(P) &= \rho_\mathfrak{p}(q_j) = q_j(\rho_\mathfrak{p}(x_1), \ldots, \rho_\mathfrak{p}(x_n))\\ 0 = \hat{f}_j(P) &= \rho_\mathfrak{p}(f_j) = f_j(\rho_\mathfrak{p}(x_1), \ldots, \rho_\mathfrak{p}(x_n))\end{aligned}$$

So in $\tilde{R}^n$ the point $(\rho_\mathfrak{p}(x_1), \ldots, \rho_\mathfrak{p}(x_n))$ fulfills all the desired equalities and inequalities, and by the Transfer Principle 1.5.18 there exists such a point in R^n as well. □

In the next section we will prove some abstract Positivstellensätze for arbitrary rings, similar to Theorem 1.1.15 for fields. Using Theorem 3.1.18 they can then be turned into concrete Positivstellensätze for the polynomial ring.

But let us first continue with the study of the real spectrum of rings. So let again be A an arbitrary ring. For $a_1, \ldots, a_m \in A$ we set

$$O(a_1, \ldots, a_m) := \{P \in \mathrm{Sper}(A) \mid \hat{a}_1(P) > 0, \ldots, \hat{a}_m(P) > 0\}$$
$$V(a_1, \ldots, a_m) := \{P \in \mathrm{Sper}(A) \mid \hat{a}_1(P) = 0, \ldots, \hat{a}_m(P) = 0\}\,.$$

Without the function notation we have

$$O(a_1, \ldots, a_m) = \{P \in \mathrm{Sper}(A) \mid a_1, \ldots, a_m \notin -P\}$$
$$V(a_1, \ldots, a_m) = \{P \in \mathrm{Sper}(A) \mid a_1, \ldots, a_m \in \mathrm{supp}(P)\}\,.$$

Definition 3.1.19. A **semialgebraic subset** of $\mathrm{Sper}(A)$ is a finite Boolean combination of sets of the form $O(a_1, \ldots, a_m)$, with $a_i \in A$. △

Remark 3.1.20. It is easily checked that Lemma 1.5.3 holds here as well, i.e. every semialgebraic set can be written as

$$\bigcup_i (V(a_i) \cap O(b_{i1}, \ldots, b_{im_i}))$$

for certain $a_i, b_{ij} \in A$. △

Example 3.1.21. If R^n is embedded into $\mathrm{Sper}(R[\underline{x}])$ as in Example 3.1.16 (ii), then the semialgebraic subsets of $\mathrm{Sper}(R[\underline{x}])$ induce exactly the already defined semialgebraic subsets of R^n. △

Definition 3.1.22. Let A be a ring.

(i) The **spectral topology** on $\mathrm{Sper}(A)$has the sets $O(a_1, \ldots, a_m)$ as a basis of open sets. The open sets are thus arbitrary unions of such sets.

(ii) The **constructible topology** on $\mathrm{Sper}(A)$ is the topology having all semialgebraic sets as a basis of open sets. For example, the sets $O(a_1, \ldots, a_m)$ and their complements form a subbasis. △

Obviously, the constructible topology is finer than the spectral topology, i.e. it has more open sets.

Theorem 3.1.23. *For any ring A, the constructible topology on $\mathrm{Sper}(A)$ is Hausdorff and quasi-compact (i.e. it has the finite open covering poperty). The spectral topology is also quasi-compact, but not Hausdorff in general.*

Proof. Take $P, Q \in \mathrm{Sper}(A)$ with $P \neq Q$. Then w.l.o.g. there exists some $a \in P \setminus Q$. So we have $Q \in O(-a)$ snd $P \in O(-a)^c$, and thus P and Q are separated by two disjoint open sets. So the constructible topology is Hausdorff.
The space $\mathrm{Sper}(R[t])$ for example, equipped with the spectral topology, is not Hausdorff. Some ordering P_a belongs to $O(p)$ if and only if $p(a) > 0$. But then p is strictly positive on some interval $(a-\epsilon, a+\epsilon)$, and thus p is also strictly positive at P_{a_-} and P_{a_+}. So none of these two orderings can be separated by disjoint open sets from P_a.
We will now show quasi-compactness of the constructible topology. This implies quasi-compactness of the spectral topology, since it has less open sets. We understand $\mathrm{Sper}(A)$ as a subset of

$$\{0,1\}^A = \{g \colon A \to \{0,1\}\}$$

by identifying a subset $P \subseteq A$ with its characteristic function. The finite set $\{0,1\}$ is clearly quasi-compact with respect to the finest topology, and by Tychonoff's Theorem, the product space $\{0,1\}^A$ is thus also quasi-compact. The product topology is the coarsest topology that makes all projections continuous, and the projections are just the evaluation maps in points of A, if elements from $\{0,1\}^A$ are understood as functions on A. On $\mathrm{Sper}(A)$, the induced topology is thus generated by sets $O(a)$ and their complements, i.e. it is the constructible topology. Since closed subsets of quasi-compact spaces are quasi-compact, it suffices to show that $\mathrm{Sper}(A)$ is a closed subset of $\{0,1\}^A$. So take

$$M \in \{0,1\}^A \setminus \mathrm{Sper}(A)$$

i.e. $M \subseteq A$ is not an ordering. We will construct an open set O that contains M, but no element from $\mathrm{Sper}(A)$. M can fail to be an ordering for different reasons. For example, there might exist $a, b \in M$ with $a + b \notin M$. In this case,

$$O := \{N \subseteq A \mid a, b \in N, a + b \notin N\}$$

is such a set. All other possibilities are proven similarly (Exercise 28). □

For the polynomial ring $R[\underline{x}]$, we have seen that R^n embeds into $\mathrm{Sper}(R[\underline{x}])$, and the latter space is quasi-compact with respect to both topologies defined by semi-algebraic sets. The space R^n is clearly not quasi-compact, the spectral topology for example induces the Euclidean topology on $\mathbb{R}^n$. As it turns out, the larger space $\mathrm{Sper}(R[\underline{x}])$ is not too large, and thus serves as a compactification of R^n.

Corollary 3.1.24. *R^n is dense in $\mathrm{Sper}(R[\underline{x}])$, with respect to the constructible (and thus also the spectral) topology.*

Proof. This is a direct consequence of Theorem 3.1.18, to be spelled out in Exercise 29. □

3.2 Positivstellensätze for Rings

For fields we have proven inTheorem 1.1.16 that

$$\bigcap_{P \text{ ordering}} P = \Sigma K^2$$

or respectively

$$\hat{a} \geqslant 0 \text{ on } \mathrm{Sper}(K) \Leftrightarrow a \in \Sigma K^2$$

holds. This statement is not true for all rings. Theorem 3.1.18 implies that the Motzkin polynomial belongs to each ordering of $A = R[x, y]$, but it is not a sum of squares in A, as we have seen. So we have to adapt Theorem 1.1.15 suitably to the ring case.

Proposition 3.2.1. *Let A be a ring, T a preordering, I an ideal, and G a multiplicatively closed subset of A with $1 \in G$. Then the following statements are equivalent:*

(i) There is no ordering $P \in \mathrm{Sper}(A)$ with

$$\begin{aligned} &\hat{t}(P) \geqslant 0 \text{ for all } t \in T \quad (\text{i.e. } T \subseteq P) \\ &\hat{i}(P) = 0 \text{ for all } i \in I \quad (\text{i.e. } I \subseteq \mathrm{supp}(P)) \\ &\hat{g}(P) \neq 0 \text{ for all } g \in G \quad (\text{i.e. } G \cap \mathrm{supp}(P) = \emptyset). \end{aligned}$$

(ii) There exist $i \in I, g \in G$ and $t \in T$ with $g^2 + t = i$.

Proof. $(ii) \Rightarrow (i)$ is easy: a possible counterexample $P = (\mathfrak{p}, Q) \in \mathrm{Sper}(A)$ would yield

$$0 = \hat{i}(P) = \widehat{g^2 + t}(P) = \hat{g}(P)^2 + \hat{t}(P) > 0,$$

a contradiction in the ordered field $(K_\mathfrak{p}, Q)$.

$(i) \Rightarrow (ii)$: Set $B := A/I$, consider the canonical projection $\pi\colon A \to B$ and set $\overline{G} := \pi(G)$, $\overline{T} = \pi(T)$. Since $\overline{G}$ is a multiplicatively closed subset of B, we can consider the localization of B by $\overline{G}$, i.e.

$$C := \overline{G}^{-1}B = \left\{ \frac{b}{g} \mid a \in B, b \in \overline{G} \right\},$$

where we have the usual equivalence relation:

$$\frac{b}{g} = \frac{c}{h} \quad :\Leftrightarrow \quad f(bh - cg) = 0 \text{ for some } f \in \overline{G}.$$

There is again a canonical homomorphism, namely

$$\iota\colon B \to C; \quad b \mapsto \frac{b}{1}.$$

So in total we have the following diagram

$$A \xrightarrow{\pi} B = A/I \xrightarrow{\iota} C = \overline{G}^{-1}B.$$

Now consider

$$T' := \left\{ \frac{t}{g^2} \mid t \in \overline{T}, g \in \overline{G} \right\} \subseteq C.$$

We distinguish two cases.
1st case: $-1 \in T'$, i.e. $f(g^2 + t) = 0$ for some $t \in \overline{T}, f, g \in \overline{G}$. This implies $(fg)^2 + f^2 t = 0$ in B, and by pulling everything back via π we obtain the desired identity in A.
2nd case: $-1 \notin T'$, so T' is a preordering of C. By Theorem 3.1.8 there exists an ordering P' of C with $T' \subseteq P'$. Then the inverse image

$$P := (\iota \circ \pi)^{-1}(P')$$

is an ordering of A. We obviously have $T \subseteq P$ and also $I \subseteq \operatorname{supp}(P)$, since already $\pi(i) = 0$ is true for all $i \in I$. Since for $g \in G$ the element $\iota(\pi(g))$ is invertible in C, it cannot belong to the ideal $\operatorname{supp}(P')$. Thus $g \notin \operatorname{supp}(P)$. So P fulfills all conditions from (i), and this means the second case cannot happen. □

For an arbitrary subset $T \subseteq A$ we write

$$W(T) := \left\{ P \in \operatorname{Sper}(A) \mid \hat{t}(P) \geqslant 0 \text{ for all } t \in T \right\} = \{ P \in \operatorname{Sper}(A) \mid T \subseteq P \}$$
$$V(T) := \left\{ P \in \operatorname{Sper}(A) \mid \hat{t}(P) = 0 \text{ for all } t \in T \right\} = \{ P \mid T \subseteq \operatorname{supp}(P) \}.$$

The following results always use the function notation for elements from the ring A. The reader might translate them to the more elementary algebraic formulation, if desired.

Theorem 3.2.2 (Abstract Positivstellensatz)**.** *Let $T \subseteq A$ be preordering. Then for each $a \in A$ we have*

$$\hat{a} > 0 \text{ on } W(T) \quad \Leftrightarrow \quad t_1 a = 1 + t_2 \text{ for certain } t_1, t_2 \in T.$$

Proof. "$\Leftarrow$": For $P \in W(T)$ we have

$$\hat{t}_1(P)\hat{a}(P) = \widehat{t_1 a}(P) = \hat{1}(P) + \hat{t}_2(P) = 1 + \hat{t}_2(P) > 0 \text{ in } K_{\mathfrak{p}}.$$

By dividing through the positive element $\hat{t}_1(P)$we get $\hat{a}(P) > 0$. For "$\Rightarrow$" use Proposition 3.2.1 with $I = (0), G = \{1\}$ and the preordering $T - aT$. □

Theorem 3.2.3 (Abstract Nichtnegativstellensatz)**.** *Let $T \subseteq A$ be a preordering. Then for each $a \in A$ we have*

$$\hat{a} \geqslant 0 \text{ on } W(T) \quad \Leftrightarrow \quad t_1 a = a^{2m} + t_2 \text{ for certain } t_1, t_2 \in T, m \in \mathbb{N}.$$

Proof. Exercise 30. □

Theorem 3.2.4 (Abstract real Nullstellensatz)**.** *Let $I \subseteq A$ be an ideal. Then for each $a \in A$ we have*

$$\hat{a} = 0 \text{ on } V(I) \quad \Leftrightarrow \quad a^{2m} + \sigma \in I \text{ for certain } m \in \mathbb{N}, \sigma \in \Sigma A^2.$$

Proof. Exercise 30. □

Definition 3.2.5. Let A be a ring and $I \subseteq A$ an ideal. The set

$$\operatorname{rrad}(I) := \left\{ a \in A \mid a^{2m} + \sigma \in I \text{ for some } m \in \mathbb{N}, \sigma \in \Sigma A^2 \right\}$$

is called the **real radical** of I. △

The usual *radical* of an ideal I is defined similarly, but without the sums of squares σ. The radical coincides with the intersection of all prime ideals above the ideal. For the real radical have instead have:

Theorem 3.2.6. *For every ideal $I \subseteq A$ we have*

$$\operatorname{rrad}(I) = \bigcap_{I \subseteq \mathfrak{p} \text{ real prime ideal}} \mathfrak{p}.$$

Proof. "$\subseteq$" is clear from the definition of a real ideal. For "$\supseteq$" let a be contained in every real prime ideal above I. This implies $\hat{a}(P) = 0$ for every ordering P with $I \subseteq \operatorname{supp}(P)$. From Theorem 3.2.4 we thus get $a \in \operatorname{rrad}(I)$. □

Note that an ideal I is real if and only if $I = \operatorname{rrad}(I)$. This follows from the fact that $a_1^2 + \cdots + a_m^2 \in I$ implies $a_i \in I$ for all i, if I is a real ideal.

3.3 Positivity on Semialgebraic Sets

We will now transform the abstract Positivstellensätze to concrete ones, exactly as for field. This can easily be done with Theorem 3.1.18. For finitely many polynomials $p_1, \ldots, p_r \in R[x_1, \ldots, x_n]$ we consider the smallest (potential) preordering that contains them:

$$T(p_1, \ldots, p_r) = \left\{ \sum_{e \in \{0,1\}^r} \sigma_e p_1^{e_1} \cdots p_r^{e_r} \mid \sigma_e \in \Sigma R[\underline{x}]^2 \right\}$$

as well as the so-called **basic closed semialgebraic set** in R^n:

$$W_R(p_1, \ldots, p_r) := \{a \in R^n \mid p_1(a) \geqslant 0, \ldots, p_r(a) \geqslant 0\}.$$

For $f_1, \ldots, f_t$ we recall the definition of the (real) variety

$$V_R(f_1, \ldots, f_t) = \{a \in R^n \mid f_1(a) = 0, \ldots, f_t(a) = 0\}$$

and the ideal

$$I(f_1, \ldots, f_t) = \left\{ \sum_i g_i f_i \mid g_i \in R[\underline{x}] \right\}$$

defined by the f_i.

Theorem 3.3.1 (Concrete Positivstellensatz)**.** *Let $p_1, \ldots, p_r \in R[\underline{x}]$. For $p \in R[\underline{x}]$ we then have*

$$p > 0 \text{ on } W_R(p_1, \ldots, p_r) \;\Leftrightarrow\; t_1 p = 1 + t_2 \text{ for certain } t_1, t_2 \in T(p_1, \ldots, p_r).$$

Proof. This is an immediate consequence of the abstract Positivstellensatz 3.2.2, since by Theorem 3.1.18 the condition $p > 0$ auf $W_R(p_1, \ldots, p_r)$ is equivalent to $\hat{p} > 0$ on $W(T(p_1, \ldots, p_r))$. □

In the same way we obtain:

Theorem 3.3.2 (Concrete Nichtnegativstellensatz). *Let $p_1, \ldots, p_r \in R[\underline{x}]$. Then for $p \in R[\underline{x}]$ we have*

$$p \geqslant 0 \text{ on } W_R(p_1, \ldots, p_r) \;\Leftrightarrow\; t_1 p = p^{2m} + t_2 \text{ with } m \in \mathbb{N}, t_1, t_2 \in T(p_1, \ldots, p_r).$$

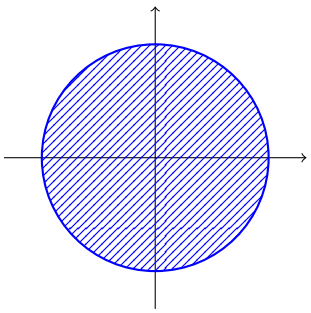
Theorem 3.3.3 (Concrete real Nullstellensatz). *Let $f_1, \ldots, f_t \in R[\underline{x}]$. Then for $p \in R[\underline{x}]$ we have*

$$p = 0 \text{ on } V_R(f_1, \ldots, f_t) \;\Leftrightarrow\; p \in \operatorname{rrad}(I(f_1, \ldots, f_t)).$$

Example 3.3.4. (*i*) For $q = 1 - x^2 - y^2 \in R[x, y]$, the set

$$W_R(q) = \left\{(a, b) \in R^2 \mid a^2 + b^2 \leqslant 1\right\}$$

is the unit disk:



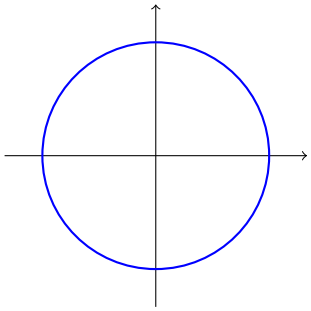
We have $T(q) = \{\sigma_1 + \sigma_2 q \mid \sigma_1, \sigma_2 \in \Sigma R[x]^2\}$. So if a polynomial $p \in R[x, y]$ is strictly positive on $W_R(q)$, there exists a representation

$$(\sigma_1 + \sigma_2 q)p = 1 + \tau_1 + \tau_2 q$$

with sums of squares $\sigma_1, \sigma_2, \tau_1, \tau_2$, and this makes the positivity of p obvious. If p is only nonnegative on $W_R(q)$, the representation looks like this:

$$(\sigma_1 + \sigma_2 q)p = p^{2m} + \tau_1 + \tau_2 q.$$

(*ii*) For the same $q = 1 - x^2 - y^2 \in R[x, y]$, the real variety $V_R(q)$ is the unit circle:



We have $I(f) = \{gf \mid g \in R[x,y]\}$, and a polynomial p vanishes on the unit circle if and only if $p \in \operatorname{rrad}(I(f))$, i.e. $p^{2m} + \sigma \in I(f)$. One can show that $I(f)$ is a real ideal (Exercise 31), and thus even $p \in I(f)$ in this case.
(iii) For $f = x^2 + y^2 \in R[x,y]$ we have $V_R(f) = \{(0,0)\}$. So if $p(0,0) = 0$, then $p^{2m} + \sigma = g \cdot (x^2 + y^2)$. Here the ideal is not real. For example, x vanishes at the origin, but does not belong to $I(f)$. However, a representation as from the real Nullstellensatz is $x^2 + y^2 \in I(f)$. △

Remark 3.3.5. Hilbert's Nullstellensatz classifies polynomials that vanish on the *complex variety* $V_C(I)$, where $C = R[i]$ is the algebraic closure of R. A polynomial vanishes on $V_C(I)$ if and only if some power lies in I. In Example 3.3.4 (iii) we can see that this fails for the smaller real variety $V_R(I)$. No power of x lies in (x^2+y^2). But in fact x does also not vanish on $V_C(I)$, since for example $(1,i) \in V_C(I)$. △

Theorem 3.3.6. *Let $I \subseteq R[\underline{x}]$ be an ideal. Then the following holds:*

(i) I is semireal if and only if $V_R(I) \neq \emptyset$.

(ii) If I is a radical ideal, then I is real if and only if $V_R(I)$ is R-Zariski-dense in $V_C(I)$.

Proof. (i) "⇒": By Remark 3.1.11 (v) there exists some $P \in \operatorname{Sper}(R[\underline{x}])$ with $I \subseteq \operatorname{supp}(P)$. Now if $I = (f_1, \ldots, f_t)$, then by Theorem 3.1.18 there also exists a point $a \in R^n$ with $f_i(a) = 0$ for all i, i.e. $V_R(I) \neq \emptyset$. "⇐" is clear from the fact that point evaluation in a point from $V_R(I)$ defines an algebra homomorphism from $R[\underline{x}]/I$ to R, which proves that $R[\underline{x}]/I$ is semireal.
(ii) "⇒": Let $p \in R[\underline{x}]$ be a polynomial with $p \equiv 0$ on $V_R(I)$. Then the concrete Nullstellensatz implies $p \in \operatorname{rrad}(I) = I$, and thus also $p \equiv 0$ on $V_C(I)$. "⇐": Let $p \in \operatorname{rrad}(I)$. Then $p \equiv 0$ on $V_R(I)$ and thus also $p \equiv 0$ on $V_C(I)$, by denseness. Hilbert's Nullstellensatz implies $p \in \operatorname{rad}(I) = I$, which proves that I is real. □

Remark 3.3.7. The concrete Positivstellensatz yields an algebraic criterion for a polynomial system of inequalities

$$p_1(\underline{x}) \geqslant 0, \ldots, p_r(\underline{x}) \geqslant 0$$

to be solvable/unsolvable. It is obviously unsolvable if and only if $-1 > 0$ holds on $W_R(p_1, \ldots, p_r)$. This is equivalent to $-t_1 = 1+t_2$ for certain $t_1, t_2 \in T(p_1, \ldots, p_r)$, and thus in fact to

$$-1 \in T(p_1, \ldots, p_r). \qquad \triangle$$

The Positivstellensätze that we haven proven so far all require *denominators*. In other words, the positive polynomial has to be multiplied with a certain polynomial, before it admits a good representation. The first Positivstellensatz *without denominators* is Schmüdgen's Theorem, that we will prove in the next chapter.

Chapter 4

Schmüdgen's Positivstellensatz

In this chapter we will prove Schmüdgen's Theorem about positive polynomials on compact sets. We will not give Schmüdgen's original proof from 1991, which is of functional analytic flavor, but a more algebraic proof, going back to Wörmann.

4.1 Archimedean Preorderings

Let again A be a commutative ring with 1.

Definition 4.1.1. A preordering $T \subseteq A$ is called **Archimedean**, if for all $a \in A$ there exists some $r \in \mathbb{N}$ with $r - a \in T$. $\triangle$

Note that we have encountered the same notion for orderings of fields already in the first chapter. Under the Archimedean assumption, we can strengthen the abstract Positivstellensatz 3.2.2 significantly, getting rid of the denominator almost completely. The following result is also true without the assumption $\mathbb{Q} \subseteq A$, but with a much more technical proof.

Theorem 4.1.2 (Abstract Archimedean Positivstellensatz)**.** *Let $\mathbb{Q} \subseteq A$ and $T \subseteq A$ an Archimedean preordering. Then for $a \in A$ we have*

$$\hat{a} > 0 \text{ on } W(T) \quad \Leftrightarrow \quad ka = 1 + t \text{ for certain } k \in \mathbb{N}, t \in T.$$

Proof. "$\Leftarrow$" is clear as usual. For "$\Rightarrow$" we first use the abstract Positivstellensatz 3.2.2 and obtain a representation

$$t_1 a = 1 + t_2$$

with $t_1, t_2 \in T$. We have $\ell - t_1 \in T$ for certain $\ell \in \mathbb{N}$, since T is Archimedean. Now consider the identity

$$\ell a + (r\ell - 1) = (\ell - t_1)(a + r) + (t_1 a - 1) + r t_1.$$

We can conclude that whenever $a + r \in T$ for some $r \geqslant 0$, then also

$$a + \left(r - \frac{1}{\ell}\right) \in T$$

holds (we have to divide by $\ell \in \mathbb{N}$ for this conclusion). But since T is Archimedean, we do have $a + r \in T$ for some $r \geqslant 0$. By iterated substraction of $\frac{1}{\ell}$ we finally even obtain some negative such r, and in particular

$$a - \tfrac{1}{k} \in T$$

for some $k \in \mathbb{N}$. This proves the claim. □

Surprisingly, a similar Nichtnegativstellensatz cannot be proven! We will later see some counterexamples.

4.2 Schmüdgen's Positivstellensatz

We can now make the abstract Archimedean Positivstellensatz concrete, using the tranfer principle as usual. But there is one more nice detail: if the semialgebraic set is bounded, the corresponding preodering is automatically Archimedean. However, this is only true over Archimedean real closed fields, i.e. subfields of $\mathbb{R}$. We will first prove this.

Proposition 4.2.1. *Let R be an Archimedean real closed field. Then a preordering $T \subseteq R[\underline{x}]$ is Archimedean if and only if*

$$r - \sum_{i=1}^{n} x_i^2 \in T$$

for some $r \in \mathbb{N}$.

Proof. "⇒" is clear. For "⇐" assume $r - \sum_i x_i^2 \in T$. This implies

$$\left(r + \frac{1}{4}\right) \pm x_j = \left(\frac{1}{2} \pm x_j\right)^2 + \left(r - \sum_i x_i^2\right) + \sum_{j \neq i} x_i^2 \in T.$$

All the variables can thus be exceeded with respect to T by some positive integer, and all coefficients from R as well, since R is Archimedean. For arbitrary polynomials we will now prove it by induction over their complexity. Therefore let $p_1, p_2 \in R[\underline{x}], r_1, r_2 \in \mathbb{N}$ and $r_1 \pm p_1 \in T, r_2 \pm p_2 \in T$. We then clearly have

$$(r_1 + r_2) \pm (p_1 + p_2) \in T$$

and

$$\begin{aligned} 3r_1r_2 - p_1p_2 &= (r_1 + p_1)(r_2 - p_2) + r_1(r_2 + p_2) + r_2(r_1 - p_1) \in T \\ 3r_1r_2 + p_1p_2 &= (r_1 + p_1)(r_2 + p_2) + r_1(r_2 - p_2) + r_2(r_1 - p_1) \in T. \end{aligned}$$

Since every polynomial arises from coefficients and variables by summation and multiplication, this finishes the proof. □

We call a subset $S \subseteq R^n$ **bounded**, if there exists some $r \in R$ with

$$\|a\|^2 := \sum_i a_i^2 \leqslant r$$

for all $a \in S$.

Theorem 4.2.2. *Let R be an Archimedean real closed field and $p_1, \ldots, p_r \in R[\underline{x}]$. Then the following are equivalent:*

(i) $W_R(p_1, \ldots, p_r) \subseteq R^n$ *is bounded.*

(ii) $T(p_1, \ldots, p_r)$ *is Archimedean.*

Proof. Set $T := T(p_1, \ldots, p_r)$ and $W := W_R(p_1, \ldots, p_r)$. $(ii) \Rightarrow (i)$ is easy: Since T is Archimedean, we have $r - \sum x_i^2 \in T$ for some $r \in \mathbb{N}$, and since elements from T are clearly nonnegative on W, W is bounded.
For $(i) \Rightarrow (ii)$ first choose $r \in \mathbb{N}$ with $p := r - \sum_i x_i^2 > 0$ on W. The concrete Positivstellensatz 3.3.1 implies $t_1 p = 1 + t$ for certain $t_1, t \in T$. We then have

$$(1 + t)p = t_1 p^2 \in T. \tag{4.1}$$

Now consider

$$T_0 := T(p) = \Sigma R[\underline{x}]^2 + \Sigma R[\underline{x}]^2 \cdot p.$$

This preordering is Archimedean by Proposition 4.2.1, and we have

$$(1 + t)T_0 \subseteq T, \tag{4.2}$$

by (4.1). Furthermore, (4.1) also implies

$$p + tr = p + tp + t\sum_i x_i^2 \in T.$$

We now choose some $s \in \mathbb{N}$ with $s - t \in T_0$. Then

$$(1+s)(s-t) = (1+t)(s-t) + (s-t)^2 \in T,$$

by (4.2). After dividing by the the positive number $1 + s$ we get $s - t \in T$. So finally

$$r(s+1) - \sum_i x_i^2 = rs + p = (p + tr) + r(s - t) \in T,$$

and thus T is Archimedean by Proposition 4.2.1. □

Theorem 4.2.3 (Schmüdgen's Positivstellensatz, a.k.a. Concrete Archimedean Positivstellensatz)**.** *Let R be an Archimedean real closed field, and let $p_1, \ldots, p_r \in R[\underline{x}]$ be such that $W_R(p_1, \ldots, p_r)$ is bounded. Then for all $p \in R[\underline{x}]$ we have*

$$p > 0 \text{ on } W_R(p_1, \ldots, p_r) \quad \Rightarrow \quad p \in T(p_1, \ldots, p_r).$$

Proof. Set $T := T(p_1, \ldots, p_r)$. As usual, from $p > 0$ on $W_R(p_1, \ldots, p_r)$ we get

$$\hat{p} > 0 \text{ on } W(T).$$

By Theorem 4.2.2 the predordering T is Archimedean, and we can apply the abstract Archimedean Positivstellensatz 4.1.2. We obtain $kp = 1 + t \in T$, and since we can divide in $R[\underline{x}]$ through the positive number k, this implies $p \in T$. □

Remark 4.2.4. (i) Again consider the unit disc in $\mathbb{R}^2$, defined by $1 - x^2 - y^2 \geqslant 0$. Every polynomial p that is strictly positive on the disc is of the form

$$p = \sigma_0 + \sigma_1(1 - x^2 - y^2)$$

with $\sigma_1, \sigma_2 \in \Sigma\mathbb{R}[x]^2$. This is a significant strengthening of Example 3.3.4 (i) for positive polynomials.
(ii) If the number r of defining polynomials p_i grows, the number of terms in a representation of p in T grows exponentially. The representation in $T(p_1, \ldots, p_r)$ uses all products of the p_i, of which there are 2^r many. In the next section we will comment on how to decrease this number. Is is in particular important for the applications of the Positivstellensätze that we will discuss in the next chapters.

(iii) The condition $p > 0$ in Schmüdgen's Positivstellensatz can in general not be replaced by $p \geqslant 0$. Consider

$$p_1 = (1 - t^2)^3 \in \mathbb{R}[t].$$

Then

$$W_{\mathbb{R}}(p_1) = [-1, 1]$$

is bounded, and the polynomial $p = 1 - t^2$ is nonnegative on this set. Assume there exists a representation

$$1 - t^2 = \sigma_0 + \sigma_1(1 - t^2)^3$$

with sums of squares σ_i. Then σ_0 would have to vanish at the points ± 1. Since σ_0 is a sum of squares, it vanished with an even multiplicity, and thus $(1 - t^2)^2$ would divide σ_0. After cancellation this would yield a representation

$$1 = \tilde{\sigma}_0(1 - t^2) + \sigma_1(1 - t^2)^2$$

with another sum of squares $\tilde{\sigma}_0$. Plugging in 1 for t then gives $1 = 0$, a contradiction.

(iv) Boundedness of $W_R(p_1, \ldots, p_m)$ in Schmüdgen's Positivstellensatz cannot be omitted. For example, consider $p_1 = t^3 \in \mathbb{R}[t]$, for which we have $W_{\mathbb{R}}(p_1) = [0, \infty)$. We have $t + 1 > 0$ on $W_{\mathbb{R}}(p_1)$. But if

$$t + 1 = \sigma_0 + \sigma_1 t^3$$

was a representation with sums of squares σ_i, the degree of σ_0 is even, whereas the degree of $\sigma_1 t^3$ is odd. Thus the degree on the right hand side is either even, which is a contradiction, or it is odd and $\geqslant 3$, which is also a contradiction.

(v) Schmüdgen's Positivstellensatz does not hold over non-Archimedean real closed fields. If R is a non-Archimedean extension of $\mathbb{R}$, it contains an infinitesimal positive element ε, i.e. we have $0 < \varepsilon < r$ for all $r \in \mathbb{R}, r > 0$. For p_1, p as in (iii) we have that $p + \varepsilon$ is strictly positive on $W_R(p_1)$. But one can show that $p + \varepsilon$ does not belong to $T(p_1)$ in $R[t]$, see Exercise 41.

(vi) We don't have any control over the degrees of the sums of squares σ_i in the representation of p in Schmüdgen's Positivstellensatz. They can in fact be much larger than the degree of p itself. Otherwise the result could be written as a formal statement, which would then hold over any real closed field, by the transfer principle. This contradicts (v). △

4.3 Some Remarks on Quadratic Modules

As we have already explained in Remark 4.2.4 (ii), the number of terms in a representation of $p \in T(p_1, \ldots, p_r)$ is 2^r in general, it thus grows exponentially with r. In applications this can soon lead to computational problems, which one would like to avoid. For this reason, the theory of *quadratic modules* is developed. We introduce the notion and give an overview about the most important results, but do not give any proofs here (see [5] for a through treatment). The reader who wants to skip this section just has to replace quadratic modules by preorderings in all of the following chapters.

Definition 4.3.1. Let A be a ring. A subset $M \subseteq A$ is called a **quadratic module** of A, if the following conditions are satisfied:

$$1 \in M,\ M + M \subseteq M,\ A^2 \cdot M \subseteq M,\ -1 \notin M. \qquad \triangle$$

Remark 4.3.2. (i) The smallest quadratic module in A is again the set of sums of squares ΣA^2 (at least if $-1 \notin \Sigma A^2$, otherwise no quadratic module exists). The (only) difference to preorderings is that quadratic modules need not be closed under multiplication, only under multiplication with (sums of) squares. So every preordering is a quadratic module, but in general not every quadratic module is a preordering. The name comes from the fact that quadratic modules resemble classical modules from algebra, but instead of a ring one uses the set of sums of squares as the underlying domain of scalars.
(ii) Given $a_1, \ldots, a_r \in A$, the smallest (potential) quadratic module containing the a_i arises from multiplying them with sums of squares and adding the terms, *but multiplication of the a_i among themselves is not necessary!* So we obtain the quadratic module generated by $a_1, \ldots, a_r$ as

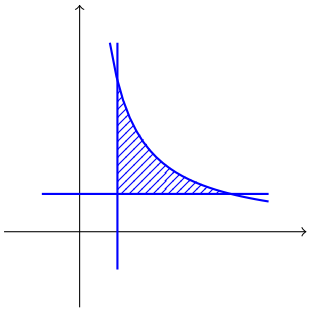
$$M(a_1, \ldots, a_r) := \left\{\sigma_0 + \sigma_1 a_1 + \cdots + \sigma_r a_r \mid \sigma_i \in \Sigma A^2\right\}$$

and the representation of its elements involves only sums of length $r + 1$. To see an example, in $A = \mathbb{R}[x, y]$ we have the quadratic module $M(x, y)$, which is not a preordering. One can easily check that $xy \notin M(x, y)$ holds (see Exercise 42).
(iii) Similar to orderings, one can develop a theory of *semiorderings*, which is built upon quadratic modules. Since we don't prove the coming results, we also do not develop the theory here. $\triangle$

Definition 4.3.3. A quadratic module M is called **Archimedean**, if for every $a \in A$ there exists some $r \in \mathbb{N}$ with $r - a \in M$. $\triangle$

We will now focus only on the polynomial ring $R[\underline{x}]$ over a real closed field. Given $p_1, \ldots, p_r \in R[\underline{x}]$, it is easily checked that if $M(p_1, \ldots, p_r)$ is Archimedean, then $W_R(p_1, \ldots, p_r) \subseteq R^n$ is bounded. This is proven exactly as in the easy direction of Theorem 4.2.2. For quadratic modules, the other implication is *false* in general.

Example 4.3.4. Choose $p_1 = x - \frac{1}{2}, p_2 = y - \frac{1}{2}, p_3 = 1 - xy \in \mathbb{R}[x, y]$. The set $W_{\mathbb{R}}(p_1, \ldots, p_3)$ is bounded:



Now assume

$$r - x = \sigma_0 + \sigma_1 \left(x - \frac{1}{2}\right) + \sigma_2 \left(y - \frac{1}{2}\right) + \sigma_3 \left(1 - xy\right)$$

for some $r \in \mathbb{N}$ and sums of squares σ_i. The homogeneous terms of highest degree on the right are

$$\tilde{\sigma}_0,\ \tilde{\sigma}_1 x,\ \tilde{\sigma}_2 y,\ -\tilde{\sigma}_3 xy,$$

where all $\tilde{\sigma}_i$, as terms of highest degrees of sums of squares, are again sums of squares. At least one of these terms has to have a degree $\geqslant 1$. We now consider the term of highest degree. If it is the first or the fourth, it has to cancel with the respective other one, since $n - x$ has odd degree. But from $\tilde{\sigma}_0 - xy\tilde{\sigma}_3 = 0$ we obtain $\tilde{\sigma}_0 = \tilde{\sigma}_3 = 0$, since all these terms are nonnegative on the second orthant. This contradicts maximality of the degree. Thus the maximal degree must be realized by the second or third term. But since $\tilde{\sigma}_1 x + \tilde{\sigma}_2 y = 0$ is also not possible (this time by nonnegativity on the first orthant) we must have

$$\tilde{\sigma}_1 x + \tilde{\sigma}_2 y = -x.$$

When we plug in $x = 1$ and $y = 1$, then the right hand side becomes negative, the left hand side is nonnegative, a contradiction. So $M(p_1, p_2, p_3)$ is not Archimedean. △

However, if R is Archimedean real closed, and $W_R(p_1, \ldots, p_r)$ is bounded, there exists some $r \in \mathbb{N}$ with

$$r - \sum_{i=1}^{n} x_i^2 \geqslant 0 \text{ on } W_R(p_1, \ldots, p_r).$$

If the extra generator

$$p_{r+1} := r - \sum_{i=1}^{n} x_i^2$$

is added, the semialgebraic set thus remains the same, but the quadratic module

$$M(p_1, \ldots, p_r, p_{r+1})$$

is now Archimedean. In fact by Proposition 4.2.1 already the quadratic module

$$M(p_{r+1}) \subseteq M(p_1, \ldots, p_r, p_{r+1})$$

is Archimedean, since for one generator we always have $M(p) = T(p)$. So making M Archimedean is not a big problem, and will increase the number of generators by at most one.
The following is the most important Positivstellensatz for quadratic modules.

Theorem 4.3.5 (Putinar's Positivstellensatz, a.k.a. Concrete Archimedean Positivstellensatz for Quadratic Modules)**.** *Let R be an Archimedean real closed field, and $p_1, \ldots, p_r \in R[\underline{x}]$ such that $M(p_1, \ldots p_r)$ is Archimedean. Then for all $p \in R[x]$ we have*

$$p > 0 \text{ on } W_R(p_1, \ldots, p_r) \Rightarrow p \in M(p_1, \ldots, p_r).$$

Example 4.3.6. Consider the unit cube

$$W = [-1, 1]^n \subseteq \mathbb{R}^n$$

defined by $1 \pm x_i \geqslant 0$ for $i = 1, \ldots, n$. The corresponding quadratic module M consist of all elements of the form

$$\sigma_0 + \sigma_1(1 - x_1) + \sigma_2(1 + x_1) + \cdots + \sigma_{2n-1}(1 - x_n) + \sigma_{2n}(1 + x_n).$$

We first check that M is Archimedean. We have

$$(1 - x_i)^2(1 + x_i) + (1 + x_i)^2(1 - x_i) = 2(1 - x_i^2),$$

thus $1 - x_i^2 \in M$, and thus also $n - \sum_i x_i^2 \in M$. So as explained before, M is Archimedean. So every polynomial p that is strictly positive on W is of the above form, whereas the representation from Schmüdgen's Positivstellensatz would involve 4^n terms. △

Chapter 5

Convexity and Optimization

The results that we have obtained so far can be used for polynomial optimization. To be able to explain this connection, we first explain some basics about *semidefinite optimization*.

5.1 Semidefinite Optimization

Let $\mathrm{Sym}_d(\mathbb{R})$ denote the real vectorspace of symmetric $d \times d$-matrices. For two symmetric matrices $A = (a_{ij})_{i,j}$ and $B = (b_{ij})_{i,j}$ we set

$$\langle A, B\rangle := \mathrm{tr}(AB) = \sum_{i,j} a_{ij}b_{ij}.$$

Here, tr denotes the trace of a matrix, i.e. the sum of its diagonal entries, and also the sum of its eigenvalues. $\langle \cdot, \cdot \rangle$ defines an inner product on $\mathrm{Sym}_d(\mathbb{R})$.
We denote by $\mathcal{P}_d$ the set of positive semidefinite symmetric matrices of size d (cf. Definition 2.2.4 and Lemma 2.2.3). The set $\mathcal{P}_d$ is a closed convex cone in $\mathrm{Sym}_d(\mathbb{R})$. By Lemma 2.2.3 we see that $\mathcal{P}_d$ is even basic closed semialgebraic, i.e. of the form

$$\mathcal{P}_d = W_{\mathbb{R}}(p_1, \ldots, p_r)$$

for certain polynomials $p_k \in \mathbb{R}[x_{ij} \mid i, j = 1, \ldots, d]$. For example, the principal minors of the matrix form such a set of defining polynomials. We write $A \succcurlyeq 0$ for $A \in \mathcal{P}_d$ and $B \succcurlyeq A$ for $B - A \succcurlyeq 0$. Similarly, $\succ$ denotes strict positive definiteness, i.e. $A \succ 0$ means that all eigenvalues of A are strictly positive, or that $v^tAv > 0$ holds for all $v \in \mathbb{R}^d \setminus \{0\}$.

Proposition 5.1.1. *(i) For $A \in \mathcal{P}_d$ and $P \in \mathrm{M}_d(\mathbb{R})$ we have $PAP^t \in \mathcal{P}_d$.*
(ii) $\langle \cdot, \cdot \rangle$ is invariant under conjugation with orthogonal matrices O, i.e.

$$\langle A, B \rangle = \langle OAO^t, OBO^t \rangle.$$

(iii) The convex cone $\mathcal{P}_d$ is self-dual with respect to $\langle \cdot, \cdot \rangle$, i.e.

$$A, B \in \mathcal{P}_d \Rightarrow \langle A, B \rangle \geqslant 0$$
$$\langle A, B \rangle \geqslant 0 \text{ for all } B \in \mathcal{P}_d \Rightarrow A \in \mathcal{P}_d.$$

(iv) If $A \succcurlyeq 0, B \succ 0$ and $\langle A, B \rangle = 0$, then $A = 0$.

Proof. (i) Clearly PAP^t is symmetric, and

$$v^t PAP^t v = (P^t v)^t A (P^t v) \geqslant 0.$$

(ii) We have

$$\begin{aligned} \langle OAO^t, OBO^t \rangle &= \mathrm{tr}(OAO^t OBO^t) = \mathrm{tr}(OABO^t) \\ &= \mathrm{tr}(ABO^t O) = \mathrm{tr}(AB) = \langle A, B \rangle. \end{aligned}$$

Here we have used cyclic invariance of the trace, and $O^t O = I_d$ since O is orthogonal.
(iii) Let $A, B \in \mathcal{P}_d$. By (i) and (ii) we can assume A to be diagonal with nonnegative diagonal entries. But the diagonal entries of the positive semidefinite matrix B are also nonnegative. This implies

$$\langle A, B \rangle = \sum_i a_{ii} b_{ii} \geqslant 0.$$

The same formula also shows that for a diagonal matrix A, all entries have to be nonnegative, for $\langle A, B \rangle \geqslant 0$ to hold for all $B \in \mathcal{P}_d$. Thus A really has to be positive semidefinite then. Now if A is not diagonal, then some $D := OAO^t$ is, and thus for $B \in \mathcal{P}_d$ we have

$$\langle D, B \rangle = \langle OAO^t, B \rangle = \langle A, O^t BO \rangle \geqslant 0.$$

Thus $D \succcurlyeq 0$, and so also $A \succcurlyeq 0$.
(iv) We can assume that B is diagonal with strictly positive diagonal entries. Then all diagonal entries of A must vanish, and this implies $A = 0$, since A is positive semidefinite (for example, look at the principal minors of size 2). □

Definition 5.1.2. Let $M, M_1, \ldots, M_m \in \mathrm{Sym}_d(\mathbb{R})$ and $\beta_1, \ldots, \beta_m \in \mathbb{R}$. The following optimization problem is called a **semidefinite optimization problem** (in **primal form**):

$$\begin{array}{ll} \text{find} & \inf \langle M, A\rangle \\ \text{s.t.} & \langle M_i, A\rangle = \beta_i \quad \text{for } i = 1, \ldots, m \\ & A \succcurlyeq 0. \end{array}$$

A **feasible point** is a matrix $A \succcurlyeq 0$ with $\langle M_i, A\rangle = \beta_i$ for all i. An **stricty feasible point** is a feasible point A that is positive definite, i.e. that also fulfills $A \succ 0$. The corresponding problem in **dual form** is:

$$\begin{array}{ll} \text{find} & \sup \sum_{i=1}^{m} \lambda_i \beta_i \\ \text{s.t.} & \sum_i \lambda_i M_i \preccurlyeq M. \end{array}$$

Here a feasible point is some $\lambda \in \mathbb{R}^m$ with $\sum \lambda_i M_i \preccurlyeq M$, and it is strictly feasible if $\sum_i \lambda_i M_i \prec M$. △

Remark 5.1.3. (i) For fixed $M \in \mathrm{Sym}_d(\mathbb{R})$, the map $A \mapsto \langle M, A\rangle$ is linear on $\mathrm{Sym}_d(\mathbb{R})$, and every linear map is of this form (this is true for any inner product on a finite-dimensional space). The conditions $\langle M_i, A\rangle = \beta_i$ define affine hyperplanes in $\mathrm{Sym}_d(\mathbb{R})$. So a semidefinite optimization problem in primal form formalizes the optimization of a linear function over an affine linear section of the convex cone $\mathcal{P}_d$ of positive semidefinite matrices.
(ii) In the dual problem, the linear function $\lambda \mapsto \beta^t\lambda$, defined on $\mathbb{R}^m$, is optimized. The domain of optimization is the set defined by the condition

$$\sum \lambda_i M_i \preccurlyeq M$$

which is clearly a closed convex set in $\mathbb{R}^m$.
(iii) The dual problem can be brought into the form of a primal problem, and vice versa. The dual problem can be seen as optimization inside the affine space

$$M + \mathrm{span}_{\mathbb{R}}(M_1, \ldots, M_m) \subseteq \mathrm{Sym}_d(\mathbb{R})$$

over the section with $\mathcal{P}_d$. But this is exactly what is done in a primal problem as well. Conversely, after choosing a basis of the affine space in the primal problem, the problem looks like a dual problem.
(iv) Semidefinite optimization problems are usually solved with numerical *interior point methods*, which can often determine the optimal value and even the optimal point efficiently.
(v) If the matrices M, M_i are all diagonal, the dual condition $\sum_i \lambda_i M_i \preceq M$ defines a *polyhedron*, i.e. a finite intersection of halfspaces. So a linear function is optimized over a polyhedron, which is known as **linear optimization**. Semidefinite optimization is thus a generalization of linear optimization. △

Theorem 5.1.4 (Duality Theorem of Semidefinite Optimization)**.** *Let p^* denote the optimal value of the primal problem and d^* the optimal value of the dual problem. Then*

$$d^* \leqslant p^*$$

holds. If both problems admit a feasible point, and one of them even a strictly feasible point, we even have

$$d^* = p^*.$$

Proof. If the dual problem does not have a feasible point, then $d^* = -\infty$ and the inequality is trivially fulfilled. The same is true if the primal problem does not have a feasible point.
So let $\lambda \in \mathbb{R}^m$ and $A \in \mathcal{P}_d$ be feasible for the dual and primal problem, respectively. From $M - \sum_i \lambda_i M_i \in \mathcal{P}_d$ and Proposition 5.1.1 (iii) we thus obtain

$$0 \leqslant \langle M - \sum_i \lambda_i M_i, A\rangle = \langle M, A\rangle - \sum_i \lambda_i \langle M_i, A\rangle,$$

i.e.

$$\sum_i \lambda_i \beta_i \leqslant \langle M, A\rangle.$$

Since d^* is the supremum over all expressions on the right, and p^* is the infimum over all expressions on the left, this proves $d^* \leqslant p^*$.
Now let $\lambda \in \mathbb{R}^m$ be strictly feasible for the dual problem, i.e. we have

$$M - \sum \lambda_i M_i \succ 0.$$

We first show that the following set is a closed convex cone:

$$K := \{(\langle A, M\rangle, \langle A, M_1\rangle, \ldots, \langle A, M_m\rangle) \mid A \in \mathcal{P}_d\} \subseteq \mathbb{R}^{m+1}.$$

It is obvious that K is a convex cone. Now let $(A_j)_{j\in\mathbb{N}}$ be a sequence in $\mathcal{P}_d$, such that the tuples

$$(\langle A_j, M\rangle, \langle A_j, M_1\rangle, \ldots, \langle A_j, M_m\rangle)$$

converge to some $r \in \mathbb{R}^{m+1}$, for $j \to \infty$. We can assume $A_j \neq 0$ for all A_j. Let $\|\cdot\|$ be any norm on $\mathrm{Sym}_d(\mathbb{R})$, for example the one induced by $\langle\cdot,\cdot\rangle$. Then w.l.o.g. there exists some $A \in \mathcal{P}_d \setminus \{0\}$ with

$$\frac{A_j}{\|A_j\|} \overset{j\to\infty}{\to} A,$$

by the Bolzano-Weierstraß Theorem and closedness of $\mathcal{P}_d$. By Proposition 5.1.1 (iv) we have

$$\begin{aligned} 0 < \langle A, M - \sum_i \lambda_i M_i\rangle &= \lim_j \frac{1}{\|A_j\|}\langle A_j, M - \sum_i \lambda_i M_i\rangle \\ &= \lim_j \frac{1}{\|A_j\|}\left(\langle A_j, M\rangle - \sum_i \lambda_i \langle A_j, M_i\rangle\right). \end{aligned}$$

The second factor converges to

$$r_0 - \lambda_1 r_1 - \cdots - \lambda_m r_m$$

and thus remains bounded. So the first factor cannot converge to zero, the norm of the A_j is thus also bounded. So we can assume w.l.o.g. that the sequence A_j itself converges, and this implies $r \in K$. This proves closedness of K.
Since both problems admit a feasible point, we have

$$-\infty < d^* \leqslant p^* < \infty.$$

Now let $p < p^*$. Then the tuple $(p, \beta_1, \ldots, \beta_m)$ does not belong to K, since otherwise there would exists a primal feasible point A with $\langle A, M\rangle = p < p^*$. So by the separation theorem for closed convex cones there exists a vector $\gamma \in \mathbb{R}^{m+1}$ with

$$0 \leqslant \gamma_0\langle A, M\rangle + \gamma_1\langle A, M_1\rangle + \cdots + \gamma_m\langle A, M_m\rangle \qquad \forall A \in \mathcal{P}_d \tag{5.1}$$

$$\gamma_0 p + \gamma_1\beta_1 + \cdots + \gamma_m\beta_m < 0. \tag{5.2}$$

By plugging in a primal feasible point A into (5.1),and comparing with (5.2), we get $\gamma_0 > 0$. We divide by γ_0 and see from (5.1) that

$$(-\gamma_1/\gamma_0, \ldots, -\gamma_m/\gamma_0)$$

is feasible for the dual problem (using Proposition 5.1.1 (iii) again). From (5.2) we then see that the corresponding dual value is $> p$. So $d^* > p$, and since $p < p^*$ was arbitrary, this proves $d^* = p^*$.
The case of a strictly feasible primal problem is proven similar, or reduced to the above case by formulating the dual problem as a primal problem, and vice versa. □

Remark 5.1.5. The duality theorem shows how semidefinite optimization problems can be solved numerically *with error bounds*. One just solves both primal and dual problem simultaneously. This usually means that d^* is approximated from below, and p^* is approximated from above. From $d^* \leqslant p^*$ we thus clearly know how far we are from the actual optimal values, at most. If $d^* = p^*$ holds, the two approximating sequences even converge to each other. △

In the following section we will see how Schmüdgen's and Putinar's Theorem can be used to compute the optimal value of a polynomial function on a semialgebraic set, using semidefinite optimization.

5.2 Lasserre's Optimization Method

Let polynomials $p_1, \ldots, p_r \in \mathbb{R}[x_1, \ldots, x_n]$ be given, and let

$$W := W_{\mathbb{R}}(p_1, \ldots, p_r) = \{a \in \mathbb{R}^n \mid p_1(a) \geqslant 0, \ldots, p_r(a) \geqslant 0\}$$

be the basic closed semialgebraic set defined by the p_i. For another polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ we are interested in its infimum on W:

$$p_* := \inf\{p(a) \mid a \in W\}.$$

Note that we do not assume W to be convex, nor p to be linear. So the problem is not in the form of a semidefinite optimization problem, nor in any other convex optimization form for which there exists efficient solution methods. Computing p_* will indeed be very hard in general.

By $\mathbb{R}[\underline{x}]_d$ we denote the finite-dimensional space of polynomials of degree $\leqslant d$, and by $M_d(p_1, \ldots, p_r)$ the set of those elements from the quadratic module $M(p_1, \ldots, p_r)$ that *obviously* belong to $\mathbb{R}[\underline{x}]_d$. To be precise, we define

$$M_d(p_1, \ldots, p_r) := \{\sigma_0 + \sigma_1 p_1 + \cdots + \sigma_r p_r \mid \deg(\sigma_0), \deg(\sigma_i p_i) \leqslant d\}.$$

So we take those elements that are of degree at most d because all summands in the representation are of degree at most d. Of course this can also be stated as

$$\deg(\sigma_i) \leqslant d - \deg(p_i).$$

However, note that obvious inclusion

$$M_d(p_1, \ldots, p_r) \subseteq M(p_1, \ldots, p_r) \cap \mathbb{R}[\underline{x}]_d$$

will be strict in general (cf. Example 4.2.4(vi))! We have

$$M(p_1, \ldots, p_r) = \bigcup_{d \in \mathbb{N}} M_d(p_1, \ldots, p_r)$$

and we call $M_d(p_1, \ldots, p_r)$ a **truncated quadratic module**. Now for each $d \in \mathbb{N}$ we set

$$p_{*,d} := \sup\{s \in \mathbb{R} \mid p - s \in M_d(p_1, \ldots, p_r)\}.$$

The following theorem describes Lasserre's optimization method.

Theorem 5.2.1. *Each $p_{*,d}$ is the optimal value of a semidefinite optimization problem that can be constructed explicitly from $p, p_1, \ldots, p_r$. We have*

$$p_{*,d} \leqslant p_*$$

for all d, and the sequence $(p_{*,d})_{d \in \mathbb{N}}$ is monotonically increasing. If $M(p_1, \ldots, p_r)$ is Archimedean, it converges to p_*.

Proof. Set $M := M(p_1, \ldots, p_r)$, $M_d := M_d(p_1, \ldots, p_r)$ and $W := W_{\mathbb{R}}(p_1, \ldots, p_r)$. If $p - s \in M_d$, then clearly $p - s \in M$ and thus $p - s \geqslant 0$ on W. So we get $s \leqslant p_*$ and thus clearly $p_{*,d} \leqslant p_*$. The sequence of the $p_{*,d}$ is obviously monotonically increasing, since $M_d \subseteq M_{d+1}$.
Now assume M is Archimedean and fix $\epsilon > 0$ arbitrary. Then $p - (p_* - \epsilon)$ is strictly positive on W, and by Theorem 4.3.5 we have

$$p - (p_* - \epsilon) \in M = \bigcup_{d \in \mathbb{N}} M_d.$$

So there exists some d with $p - (p_* - \epsilon) \in M_d$, i.e. $p_{*,d} \geqslant p_* - \epsilon$. This proves convergence.
What remains to show is how $p_{*,d}$ is the optimal value of a semidefinite optimization problem. For this we use the Gram matrices from Section 2.2. We consider the finite-dimensional vectorspace

$$V := \mathbb{R} \times \mathrm{Sym}_{\delta_0}(\mathbb{R}) \times \cdots \times \mathrm{Sym}_{\delta_r}(\mathbb{R}).$$

Here we choose δ_i so that $G(N_i)$ has degree at most $d - \deg(p_i)$, for every $N_i \in \mathrm{Sym}_{\delta_i}(\mathbb{R})$. In V we consider the affine linear subspace

$$H := \{(s, N_0, \ldots, N_r) \mid p - s = G(N_0) + G(N_1)p_1 + \cdots + G(N_r)p_r\}.$$

Then $p_{*,d}$ is the supremum of the linear function

$$(s, N_0, \ldots, N_r) \mapsto s$$

over the intersection of H with the set defined by the conditions $N_i \succcurlyeq 0$ for all i, by Theorem 2.2.17. This already shows that we have a semidefinite optimization problem. If we want an affine linear section of just *one* cone of positive semidefinite matrices, we can embed V into $\mathrm{Sym}_{1+\delta_0+\cdots+\delta_r}(\mathbb{R})$ by the rule

$$(s, N_0. \ldots, N_r) \mapsto \mathrm{diag}(s, N_0, \ldots, N_r). \qquad \square$$

Remark 5.2.2. (i) First recall that if $M(p_1, \ldots, p_r)$ is a preordering (for example, if instead of the p_i we use all of their products as generators), then the Archimedean assumption can be replaced by the condition that $W_{\mathbb{R}}(p_1, \ldots, p_r)$ is bounded (Theorem 4.2.2). Also recall that the Archimedean condition for general $M(p_1, \ldots, p_r)$ can be ensured by adding some polynomial $p_{r+1} = N - \sum_i x_i^2$ to the defining polynomials. If $W_{\mathbb{R}}(p_1, \ldots, p_r)$ is bounded, set set will not be changed, at least if N is large enough.
(ii) Lasserre's optimization method is implemented, for example in the free Matlab plugin *Yalmip*. As long as the number of variables and the degree of the involved polynomials is not too large, it solves such optimization problems quite efficiently. Often some refinements of the described methods are used to reduce complexity. For example, depending on the structure of the polynomials, certain more specific Positivstellensätze can be employed. $\triangle$

5.3 Spectrahedra

The set of feasible points of a semidefinite optimtization problem is called a *spectrahedron*. For a given set it might be hard to decide whether it is a spectrahedron or not, let alone find some explicit realization. We have already seen this phenomenon in the proof of Theorem 5.2.1, where we had to employ the Gram matrix method. But to be able to apply semidefinite programming successfully to many problems, it is important to understand its feasible sets better. We will thus go into this topic a little deeper now. We will restrict to spectrahedral *cones*, since this makes some things a little easier.

Definition 5.3.1. A **spectrahedral cone** is a set of the form

$$\mathcal{S}(M_1, \ldots, M_n) := \{a \in \mathbb{R}^n \mid a_1 M_1 + \cdots + a_n M_n \succcurlyeq 0\}$$

for certain symmetric matrices $M_1, \ldots, M_n \in \mathrm{Sym}_d(\mathbb{R})$. △

Remark 5.3.2. (i) A spectrahedral cone is the inverse image of the convex cone $\mathcal{P}_d$ of positive semidefinite matrices under a linear map $\mathbb{R}^n \to \mathrm{Sym}_d(\mathbb{R})$, where the map is given by

$$a \mapsto a_1 M_1 + \cdots + a_n M_n.$$

If the matrices M_i are linearly independent (what we will usually assume), we can also understand is as an intersection of $\mathcal{P}_d$ with a subspace.
(ii) Spectrahedral cones are basic closed semialgebraic (and thus closed) convex cones. This is immediately clear from the fact that $\mathcal{P}_d$ has these properties.
(iii) A *polyhedral cone* is of the form

$$\{a \in \mathbb{R}^n \mid v_1^t a \geqslant 0, \ldots, v_d^t a \geqslant 0\}$$

for certain $v_i \in \mathbb{R}^n$. Each such polyhedral cone is also a spectrahedral cone. In fact, if all M_i are diagonal, the condition $a_1 M_1 + \cdots + a_n M_n \succcurlyeq 0$ gives rise to precisely such finitely many linear inequalities.
(iv) Intersections of spectrahedral cones are spectrahedral. One just has to form block-diagonal sums of the defining matrices. △

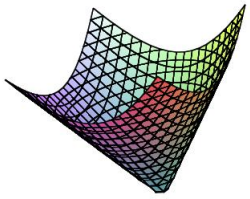
Example 5.3.3. (i) The spectrahedron $\mathcal{S}(M_1, M_2, M_3) \subseteq \mathbb{R}^3$ defined by

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ M_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

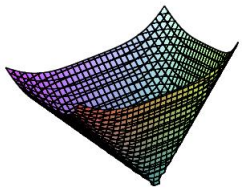
is the set defined by the following polynomial inequalities, as is easily checked:

$$x_1^2 - x_2^2 - x_3^2 \geqslant 0,\ x_1 \geqslant 0.$$

This is a circular cone, which is clearly not polyhedral:



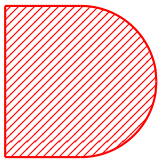
(*ii*) The convex cone defined by the conditions $x_1^4 - x_2^4 - x_3^4 \geqslant 0, x_1 \geqslant 0$ looks similar, but we will later see that it is not spectrahedral:



(*iii*) The convex cone whose cross-section is the set

$$[-1,0] \times [-1,1] \cup B \subseteq \mathbb{R}^2,$$

where B is the unit disk, is not spectrahedral. In fact it is not basic closed semi-algebraic (Exercise 34).



△

Example 5.3.4. Let $\mathbb{R}[\underline{x}]_d^*$ denote the *dual space* of the finite-dimensional space $\mathbb{R}[\underline{x}]_d = \mathbb{R}[x_1, \ldots, x_n]_d$. An element $\varphi \in \mathbb{R}[\underline{x}]_d^*$ is thus just a linear map

$$\varphi \colon \mathbb{R}[\underline{x}]_d \to \mathbb{R}.$$

Since linear maps are uniquely defined by their values on a basis, we can identify φ with the tuple of its values on the canonical monomial basis of $\mathbb{R}[\underline{x}]_d$, i.e.

$$\varphi = (\varphi(\underline{x}^\alpha))_{\alpha \in \mathbb{N}^n; |\alpha| \leqslant d}.$$

In this way $\mathbb{R}[\underline{x}]_d^*$ identifies with $\mathbb{R}^{\Delta_d}$.
Now let $p_1, \ldots, p_r \in \mathbb{R}[\underline{x}]$ be given. Again we consider the truncated quadratic module

$$M_d = M_d(p_1, \ldots, p_r) \subseteq \mathbb{R}[\underline{x}]_d$$

as in the last section. Now let

$$M_d^\vee = \{\varphi \in \mathbb{R}[\underline{x}]_d^* \mid \varphi \geqslant 0 \text{ on } M_d\}$$

be the *dual cone* of M_d. It is not hard to see that $M_d^\vee$ is in fact a spectrahedron in $\mathbb{R}[\underline{x}]_d^* = \mathbb{R}^{\Delta_d}$. Indeed, the condition $\varphi \geqslant 0$ on M_d consists of the single conditions

$$\varphi(q^2 p_i) \geqslant 0 \qquad \forall q \in \mathbb{R}[\underline{x}]_{k_i}$$

where k_i is chosen so that $q^2 p_i \in \mathbb{R}[\underline{x}]_d$. Writing $q = \sum_\alpha q_\alpha \underline{x}^\alpha$ we have

$$\varphi(q^2 p) = \varphi\left(\sum_{\alpha,\beta} q_\alpha q_\beta \underline{x}^{\alpha+\beta} p\right) = \sum_{\alpha,\beta} q_\alpha q_\beta \left(\varphi(\underline{x}^{\alpha+\beta} p)\right).$$

This being nonnegative for all choices of coefficients q_α just means that the matrix

$$\left(\varphi(\underline{x}^{\alpha+\beta} p)\right)_{\alpha,\beta}$$

is positive semidefinite. Its entries are linear combinations of the values $\varphi(\underline{x}^\alpha)$ (of course depending on p). This proves the claim. △

Every convex set has nonempty interior in its affine hull. So after replacing the ambient space by the affine hull of the convex set, one can always assume it to have nonempty interior.

Proposition 5.3.5. *Let $S \subseteq \mathbb{R}^n$ be a spectrahedral cone with interior point e. Then there exist symmetric matrices $M_1, \ldots, M_n$ with $S = \mathcal{S}(M_1, \ldots, M_n)$ and*

$$e_1 M_1 + \cdots + e_n M_n = I.$$

The interior of S is then defined by the condition $a_1 M_1 + \cdots + a_n M_a \succ 0$.

Proof. First write $S = \mathcal{S}(N_1, \ldots, N_n)$ with certain symmetric matrices N_i. We set

$$a \bullet N := a_1 N_1 + \cdots + a_n N_n.$$

Since e is an interior point of S, we have

$$0 \preccurlyeq (e \pm \varepsilon \delta_i) \bullet N = e \bullet N \pm \epsilon N_i$$

for some small $\varepsilon > 0$ and all i, where δ_i is the i-th standard basis vector. For every vector $v \in \ker(e \bullet N)$ we thus have

$$0 \leqslant v^t \left(e \bullet N \pm \varepsilon N_i\right) v = \pm \varepsilon v^t N_i v.$$

This implies $v^t N_i v = 0$ and thus also $0 = v^t \left(e \bullet N \pm \varepsilon N_i\right) v$. Since $e \bullet N \pm \varepsilon N_i$ is positive semidefinite, this implies

$$(e \bullet N \pm \varepsilon N_i) v = 0,$$

as can for example be seen from a decomposition into squares of rank 1 as in Lemma 2.2.3 (iv). But this further implies $N_i v = 0$, so we have shown

$$\ker(e \bullet N) \subseteq \ker(N_i).$$

After a change of basis all N_i simultaneously split off a block of zeros, that we can simply omit for the definition of S. The resulting new matrices M_i then fulfill $\ker(e \bullet M) = \{0\}$, i.e. $e \bullet M \succ 0$. After conjugation with an invertible matrix we can thus ensure $e \bullet M = I$.
It is clear that $a \bullet M \succ 0$ implies that a is an interior point of S. Conversely, if a is an interior point, then $a - \varepsilon e \in S$ for some $\varepsilon > 0$. This implies

$$0 \preccurlyeq (a - \varepsilon e) \bullet M = a \bullet M - \varepsilon e \bullet M = a \bullet M - \varepsilon I.$$

So $0 \prec \varepsilon I \preccurlyeq a \bullet M$. □

The following result reveals a connection between the geometry of spectrahedra and some property of real polynomials.

Theorem 5.3.6. *Let $M_1, \dots, M_n \in \mathrm{Sym}_d(\mathbb{R})$, and assume $e_1 M_1 + \cdots + e_n M_n = I$ holds for some $e \in \mathbb{R}^n$. Set*

$$h := \det\left(x_1 M_1 + \cdots + x_n M_n\right).$$

Then the polynomial $h \in \mathbb{R}[x_1, \dots, x_n]$ is homogeneous of degree d, one has $h(e) = 1$, and for each $a \in \mathbb{R}^n$ the polynomial

$$h_a(t) := h(a - te) \in \mathbb{R}[t]$$

has only real roots. One has

$$\mathcal{S}(M_1, \dots, M_n) = \{a \in \mathbb{R}^n \mid \textit{all roots of } h_a \textit{ are } \geqslant 0\}.$$

Proof. It is obvious that h is homogeneous. For $a \in \mathbb{R}^n$ we have

$$h_a(t) = \det((a - te) \bullet M) = \det(a \bullet M - tI),$$

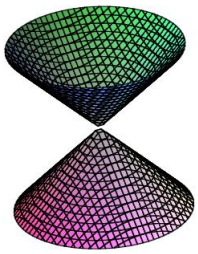
so h_a is the characteristic polynomial of $a \bullet M$. The zeros of h_a are thus the Eigenvalues of the symmetric matrix $a \bullet M$, which are all real. Furthermore, $a \bullet M$ is positive semidefinite if and only if all of these Eigenvalues/zeros are nonnegative. □

Definition 5.3.7. (i) A homogenous polynomial $h \in \mathbb{R}[x_1, \dots, x_n]$ is called **hyperbolic** in direction $e \in \mathbb{R}^n$, if $h(e) \neq 0$ and all zeros of $h_a(t) := h(a - te)$ are real, for all $a \in \mathbb{R}^n$.
(ii) If h is hyperbolic in direction e, the set

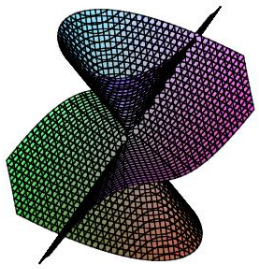
$$\Lambda_e(h) := \{a \in \mathbb{R}^n \mid \text{all zeros of } h_a \text{ are } \geqslant 0\}$$

is called the **hyperbolicity cone** of h (in direction e). △

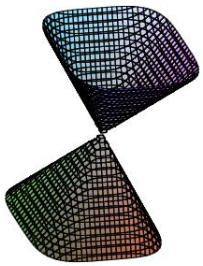
Example 5.3.8. (i) The polynomial $h = x_1^2 - x_2^2 - x_3^2$ is hyperbolic in direction $e = (1, 0, 0)$, as can be seen for example in the following picture:



On each vertical line there are 2 real intersection points, and since h is of degree 2, there can be no further complex zeros. The same is true for the following degree 3 polynomial $h = x_1^3 - x_1^2x_3 - x_1x_3^2 + x_1x_2^2 + x_3^3$:



The corresponding hyperbolicity cones are the upwards pointing and filled cones. (ii) The polynomial $h = x_1^4 - x_2^4 - x_3^4$ is not hyperbolic in direction $e = (1, 0, 0)$. On each vertical line there are only 2 real zeros. Since h is of degree 4, there must always exists two more non-real zeros, that cannot be seen in the picture. h is also not hyperbolic with respect to any other direction.



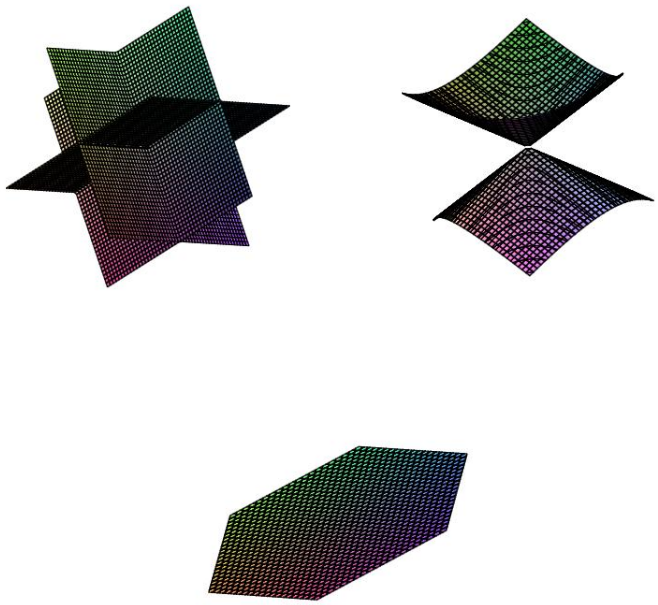
(iii) The **elementary symmetric polynomials**

$$s_{r,n} = \sum_{1 \leqslant i_1 < \cdots < i_r \leqslant n} x_{i_1} \cdots x_{i_r} \in \mathbb{R}[x_1, \ldots, x_n]$$

are all hyperbolic in direction $e = (1, \ldots, 1)$. This is best seen as follows. We have $s_{n,n} = x_1 \cdots x_n$, and thus

$$s_{n,n}(a - te) = s_{n,n}(a_1 - t, \ldots, a_n - t) = (a_1 - t) \cdots (a_n - t)$$

has the real zeros $a_1, \ldots, a_n$. Now $s_{r,n}(a - te)$ is precisely the r-th derivative of $s_{n,n}(a - te)$ with respect to t (up to constants). By Rolle's Theorem we always get a real zero of the derivative in between to consecutive zeros of the initial polynomial. Thus the derivative also just has real zeros. The following pictures show the real zero sets of $s_{3,3}$, $s_{2,3}$ und $s_{1,3}$:



△

Remark 5.3.9. (i) Hyperbolicity cones are in really convex cones. However, this is not obvious from the definition. It can be proven elementary but with a quite technical proof. We will instead deduce it from the Helton-Vinnikov Theorem (Theorem 5.3.12) below (see Exercise 50).
(ii) If h is hyperbolic in direction e, then for any e' in the interior of $\Lambda_e(h)$, h is also hyperbolic in direction e', with $\Lambda_e(h) = \Lambda_{e'}(h)$. Also this can be proven directly, or with the Helton-Vinnikov Theorem (Exercise 50). It is also quite plausible in the above pictures.
(iii) Every spectrahedral cone is a hyperbolicity cone. This follows from Theorem 5.3.6 (together with Proposition 5.3.5). △

Example 5.3.10. The cone K from Example 5.3.3 (ii), defined by

$$x_1^4 - x_2^4 - x_3^4 \geqslant 0, x_1 \geqslant 0$$

is not hyperbolic, and thus also not a spectrahedron. Indeed, if it was hyperbolic, there would exist a hyperbolic polynomial h with $h = 0$ on ∂K. By homogeneity we had $h = 0$ on the whole real zero set

$$V_{\mathbb{R}}(x_1^4 - x_2^4 - x_3^4) \subseteq \mathbb{R}^3.$$

The real Nullstellensatz then implies $h \in \operatorname{rrad}(I(x_1^4 - x_2^4 - x_3^4))$. But this ideal is real, which is checked similar to the case $1 - x_1^2 - x_2^2$ (settled in Exercise 31). So h would contain $x_1^4 - x_2^4 - x_3^4$ as a factor. But since this polynomial is not hyperbolic, h is also not hyperbolic. This is a contradiction. △

Every spectrahedral cone is hyperbolic. Hyperbolicity of a cone is often easier to check than the spectrahedral property. We have seen this in Example 5.3.10. This motivates the following conjecture:

Conjecture 5.3.11 (Generalized Lax Conjecture)**.** *Every hyperbolicity cone is spectrahedral.*

The conjecture is still open, there are only partial results. An important one is the Helton-Vinnikov Theorem, which we only cite. Its second part immediately follows from Theorem 5.3.6.

Theorem 5.3.12 (Helton & Vinnikov)**.** *Let $h \in \mathbb{R}[x_1, x_2, x_3]$ be hyperbolic in direction $e \in \mathbb{R}^3$, with $h(e) = 1$. Then there are matrices $M_1, M_2, M_3 \in \operatorname{Sym}_d(\mathbb{R})$ with $e_1M_1 + e_2M_2 + e_3M_3 = I$ and*

$$h = \det(x_1M_1 + x_2M_2 + x_3M_3).$$

In particular, every hyperbolicity cone in $\mathbb{R}^3$ is spectrahedral.

The full statement of the Helton-Vinnikov Theorem fails in higher dimensions:

Example 5.3.13. The polynomial $h = x_1^2 - x_2^2 - x_3^2 - x_4^2 \in \mathbb{R}[x_1, x_2, x_3, x_4]$ is hyperbolic in direction $e = (1, 0, 0, 0)$. We indeed have

$$h_a(t) = h(a_1 - t, a_2, a_3, a_4) = (a_1 - t)^2 - a_2^2 - a_3^2 - a_4^2,$$

and both zeros of this univariate polynomial are real. Now assume

$$h = \det(x_1M_1 + x_2M_2 + x_3M_3 + x_4M_4)$$

for certain symmetric matrices M_i, which must be of size 2, by homogeneity. Then the M_i must be linearly independent in the space $\mathrm{Sym}_2(\mathbb{R})$. If one was expressible as a linear combination of the others, we would have $h = q(A\underline{x})$ for some polynomial q in only three variables, and some matrix $A \in \mathrm{M}_{3\times 4}(\mathbb{R})$. For $0 \neq v \in \ker A$ and $\lambda \in \mathbb{R}$ we would obtain

$$1 = h(e) = h(e + \lambda v) = \lambda^2 h(v) + 2\lambda v_1 + 1,$$

implying $0 = v_1 = h(v) = -v_2^2 - v_3^2 - v_4^2$ and thus $v = 0$, a contradiction.
But in $\mathrm{Sym}_2(\mathbb{R})$ there exist at most three linearly independent matrices. So h does not admit a determinantal representation. △

Remark 5.3.14. Assume $h^r = \det(x_1 M_1 + \cdots + x_n M_n)$ with symmetric matrices and $e \bullet M = I$, for some $r \in \mathbb{N}$. Then

$$\Lambda_e(h) = \Lambda_e(h^r) = \mathcal{S}(M_1, \ldots, M_n)$$

is spectrahedral. In Example 5.3.13, and more general for all quadratic hyperbolic polynomials it is in fact always possible to find a determinantal representation of some power.
But there also exists a hyperbolic polynomial h of degree 4 in 4 variables, of which *no power* admits a determinantal representation as above. △

Remark 5.3.15. Brändén has shown that the hyperbolicity cones of all elementary symmetric polynomials $s_{r,n}$ are spectrahedral. He produced a determinantal representation of certain *multiples* $h \cdot s_{r,n}$, with a factor h that does not change the hyperbolicity cone. Much more is still not known about the generalized Lax conjecture. △

5.4 Spectrahedral Shadows

Definition 5.4.1. A **spectrahedral shadow** is the image of a spectrahedral cone under a linear map. △

Remark 5.4.2. (i) Any linear image of a polyhedron is again a polyhedron. This is not true for spectrahedra however, so spectrahedral shadows are strictly more general than spectrahedra. For example, consider the following convex cone:

$$K = \left\{(a, b, c, d, e) \in \mathbb{R}^5 \mid b^2 \leqslant da, c^2 \leqslant ea, d^2 + e^2 \leqslant a^2, 0 \leqslant a, d, e\right\}.$$

The conditions $b^2 \leqslant da, 0 \leqslant a, d$ for example translate to

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} \succeq 0,$$

and in this way one checks that K is indeed a spectrahedron. If we now project K to $\mathbb{R}^3$, using the map $(a, b, c, d, e) \mapsto (a, b, c)$, we obtain the convex cone

$$K' = \big\{(a, b, c) \mid a^4 \geqslant b^4 + c^4, a \geqslant 0\big\}$$

which is thus a spectrahedral shadow. In Example 5.3.10 we have seen that K' is not a spectrahedron.
(ii) In contrast to spectrahedra, the class of spectrahedral shadows is closed under most operation that apply to convex sets, for example duals, polars, Minkowski sums, closures and interiors. $\triangle$

There exists basically just one method to construct spectrahedral shadows directly. This method goes back to Lasserre and Parillo, and works as follow. Let again $p_1, \ldots, p_r \in \mathbb{R}[\underline{x}]$ be given. We consider the semialgebraic set

$$W_{\mathbb{R}}(p_1, \ldots, p_r) = \{a \in \mathbb{R}^n \mid p_1(a) \geqslant 0, \ldots, p_r(a) \geqslant 0\}$$

and the truncated quadratic module

$$M_d = M_d(p_1, \ldots, p_r) \subseteq \mathbb{R}[\underline{x}]_d.$$

We have seen in Example 5.3.4 that

$$M_d^\vee = \{\varphi \colon \mathbb{R}[\underline{x}]_d \to \mathbb{R} \text{ linear } \mid \varphi \geqslant 0 \text{ on } M_d\} \subseteq \mathbb{R}[\underline{x}]_d^* = \mathbb{R}^{\Delta_d}$$

is a spectrahedron. We now project this spectrahedron to $\mathbb{R}^n$, via the following map

$$\begin{aligned} \pi \colon \mathbb{R}[\underline{x}]_d^* &\to \mathbb{R}^n \\ \varphi &\mapsto (\varphi(x_1), \ldots, \varphi(x_n)). \end{aligned}$$

Written down explicitly in coordinates, we project the tuple $\varphi = (\varphi(\underline{x}^\alpha))_{|\alpha| \leqslant d}$ to its coordinates indexed by $x_1, \ldots, x_n$.
For a set $W \subseteq \mathbb{R}^n$ we denote by $\mathrm{cc}(W)$ its conic hull, i.e. the smallest convex cone in $\mathbb{R}^n$ that contains W. By $\overline{\mathrm{cc}}(W)$we denote its the closure, the smallest closed convex cone containing K.

Theorem 5.4.3. *Let $p_1, \dots, p_r \in \mathbb{R}[\underline{x}]$ and $W := W_{\mathbb{R}}(p_1, \dots, p_r)$. We then have:*
(i) The cone $\mathcal{L}_d := \pi(M_d^\vee)$ is a spectrahedral shadow with

$$\mathrm{cc}\,(W) \subseteq \mathcal{L}_{d+1} \subseteq \mathcal{L}_d$$

for all $d \in \mathbb{N}$
(ii) If for some $d \in \mathbb{N}$ every homogeneous linear polynomial $\ell \in \mathbb{R}[\underline{x}]_1$ that is nonnegative on W belongs to M_d, then

$$\mathrm{cc}\,(W) \subseteq \mathcal{L}_d \subseteq \overline{\mathrm{cc}}\,(W)\,.$$

Proof. (i): We already know that $\mathcal{L}_d$ is a spectrahedral shadow. The inclusion $\mathcal{L}_{d+1} \subseteq \mathcal{L}_d$ is also clear, since every $\varphi \in M_{d+1}^\vee$ gives rise to an element of $M_d^\vee$ by just restricting it to $\mathbb{R}[\underline{x}]_d$. Now every $a \in W$ defines a linear functional

$$\begin{aligned} \delta_a \colon \mathbb{R}[\underline{x}] &\to \mathbb{R} \\ p &\mapsto p(a) \end{aligned}$$

on the whole polynomial ring (even a ring homomorphism), which is nonnegative on the full quadratic module $M(p_1, \dots, p_r)$. By restricting it to $\mathbb{R}[\underline{x}]_d$ we obtain $\delta_a \in M_d^\vee$ for all $d \in \mathbb{N}$. Thus

$$a = (\delta_a(x_1), \dots, \delta_a(x_n)) \in \mathcal{L}_d \quad \text{for all } d.$$

Since $\mathcal{L}_d$ is a convex cone, this implies $\mathrm{cc}(W) \subseteq \mathcal{L}_d$ for all d.
For (ii) assume that $d \in \mathbb{N}$ is as required. By the classical separation theorem for closed convex cones, for any $a \notin \overline{\mathrm{cc}}(W)$ there is some homogeneous linear $\ell \in \mathbb{R}[\underline{x}]_1$ with $\ell \geqslant 0$ on $\overline{\mathrm{cc}}(W)$ and $\ell(a) < 0$. By assumption we have $\ell \in M_d$, and thus for each $\varphi \in M_d^\vee$

$$0 \leqslant \varphi(\ell) = \ell(\varphi(x_1), \dots, \varphi(x_n)).$$

This means $\ell \geqslant 0$ on $\mathcal{L}_d$, and thus $a \notin \mathcal{L}_d$. We have thus shown $\mathcal{L}_d \subseteq \overline{\mathrm{cc}}(W)$. □

Remark 5.4.4. (i) The last theorem says, that if all nonnegative linear polynomials on W have a representation in the quadratic module M with a *uniform degree bound*, then the conic hull of W is a spectrahedral shadow (up to closures).
(ii) It is easy to see that it suffices to represent the *strictly positive* polynomials with a uniform degree bound in M. The concrete Archimedean Positivstellensatz for quadratic modules 4.3.5 provides the existence of representations in the

Archimedean case. The problem of uniform degree bounds has to be settled independently.
(iii) There are cases in which such a uniform degree bound exists, and such where it does not, see Examples 5.4.5, 5.4.6 and 5.4.7 below.
(iv)There are further works of Helton & Nie, in which the method of Lasserre and Parillo is applied locally. They lead to large classes of sets which are indeed spectrahedral shadows.
(v) Scheiderer has recently shown that *every* convex semialgebraic cone in $\mathbb{R}^3$ is a spectrahedral shadow. He uses the above explained method and some deep results about sums of squares on curves.
(vi) Scheiderer has also shown that not every convex semialgebraic cone is a spectrahedral shadow. This refuts the so-called *Helton-Nie conjecture*. △

We will demonstrate the problem of representing nonnegative linear polynomials in some M_d in some examples. To keep the computations simple, we restrict to convex sets which are not convex cones. The results can be directly translated to the corresponding cones that have the given sets as compact cross-sections.

Example 5.4.5 (Farkas' Lemma)**.** Let $\ell_1, \ldots, \ell_r \in \mathbb{R}[\underline{x}]_1$ be of degree 1. Then

$$W = W_{\mathbb{R}}(\ell_1, \ldots, \ell_r) \subseteq \mathbb{R}^n$$

is a polyhedron. If another $\ell \in \mathbb{R}[\underline{x}]_1$ is nonnegative on $W \neq \emptyset$, there exist certain $\lambda_0, \ldots, \lambda_r \geqslant 0$ with

$$\ell = \lambda_0 + \lambda_1\ell_1 + \cdots + \lambda_r\ell_r.$$

In other words, we have $\ell \in M_1(\ell_1, \ldots, \ell_r)$. This follows for example from the Duality Theorem for linear/semidefinite optimization (Exercise 45). △

Example 5.4.6. This is another example in which uniform degree bounds exist for nonnegative linear polynomials. The set

$$W = W_{\mathbb{R}}(1 - x_1^4 - x_2^4) \subseteq \mathbb{R}^2$$

arises as the intersection of the cone from Example 5.3.3 (ii) with an affine plane. Let $\ell \in \mathbb{R}[\underline{x}]_1$ be nonnegative on W, and w.l.o.g. assume $\ell(a) = 0$ for some $a \in \partial W$. Up to scaling ℓ is then uniquely defined; for $a = (r, s)$ we obtain

$$\ell = 1 - r^3x_1 - s^3x_2.$$

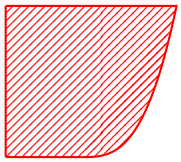
The polynomial

$$\ell - \lambda(1 - x_1^4 - x_2^4)$$

is then globally nonnegative, for some suitable $\lambda > 0$. This can be checked for example by checking its critical points. Since this polynomial is of degree 4 in 2 variables, it is a sum of squares of polynomials of degree 2 (see Remark 2.2.6). This shows $\ell \in M_4(1 - x_1^4 - x_2^4)$. △

Example 5.4.7. We finally provide an example in which uniform degree bounds for linear polynomials do not exist. Consider the set

$$W = W_{\mathbb{R}}(y - x^3, y, 1 - y, x + 1) \subseteq \mathbb{R}^2.$$



For $0 < r < 1$ we have the point $a = (r, r^3)$ in the boundary of W, and the unique (up to scaling) linear polynomial ℓ_a, which is nonnegative on W and vanishes at a is

$$\ell_a = 2r^3 - 3r^2x + y.$$

Now assume $\ell_a \in M_d$ for some *fixed* d and all $a = (r, r^3)$ with $r > 0$. That means we have representations

$$\ell_a = \sigma_0^{(a)} + \sigma_1^{(a)}(y - x^3) + \sigma_2^{(a)}y + \sigma_3^{(a)}(1 - y) + \sigma_4^{(a)}(x + 1) \tag{5.3}$$

with sums of squares $\sigma_i^{(a)}$ whose degrees are all uniformly bounded. By plugging in a we see that $\sigma_i^{(a)}(a) = 0$ must hold for all $i \neq 1$. We now pass to the limit for $a \to (0, 0)$. We can do that by scaling everything to keep all coefficients of the sums of squares bounded and then applying the Bolzano-Weierstraß Theorem. Alternatively, we could formulate the existence of representations (5.3) for all $r > 0$ as a first order statement, and apply it in some non-Archimedean real closed field for some infinitesimal $r > 0$. Similar to Exercise 41 we then obtain a representation over $\mathbb{R}$. We so obtain a representation

$$y = \sigma_0 + \sigma_1(y - x^3) + \sigma_2 y + \sigma_3(1 - y) + \sigma_4(x + 1) \tag{5.4}$$

with $\sigma_i(0, 0) = 0$ for all $i \neq 1$. By setting $y = 0$ we obtain

$$0 = \sigma_0(x, 0) + \sigma_1(x, 0)(-x^3) + \sigma_3(x, 0) + \sigma_4(x, 0)(x + 1).$$

Since $-x^3$ and $x+1$ are nonnegative on the whole interval $[-1,0]$, this implies $\sigma_1(x,0)=0$, i.e. y^2 divides σ_1. If we now set $x=0$ in (5.4), we get

$$y = \sigma_0(0,y) + \sigma_1(0,y)y + \sigma_2(0,y)y + \sigma_3(0,y)(1-y) + \sigma_4(0,y).$$

Since $\sigma_i(0,0)=0$ for all $i \neq 1$, y^2 divides all $\sigma_i(0,y)$ for $i \neq 1$. But y^2 divides σ_1 and thus also $\sigma_1(0,y)$. So y^2 divides the whole right-hand side, and thus also y, an obvious contradiction.

This contradiction really just arises from the assumption of a uniform degree bound, which allows us the pass the limit as $a \to (0,0)$. One can indeed show that all ℓ_a *do* belong to the quadratic module $M(y-x^3, y, 1-y, 1+x)$. But the degrees of the sums of squares go to infinity, as a approaches the origin.

This example can be extended to a much more general theorem. As soon as a convex and basic closed set W has a *non-exposed face*, uniform degree bounds for representations of nonnegative linear polynomials do not exist (in this example the origin is a non-exposed face). So for such sets, the method of Lasserre and Parillo cannot be used (directly) to show that they are spectrahedral shadows. $\triangle$

Chapter 6

The Moment Problem

The *moment problem* is a classical question from functional analysis. One is given a (multi-)sequence of real numbers, and wants to check whether it is the sequence of moments of some measure. In coordinate-free formulation, one is given a linear functional on the space of polynomials, and wants to check whether this functional is integration with respect to a measure. Haviland's Theorem provides a link to positive polynomials. If positive polynomials can be characterized by sums of squares/preorderings/quadratic modules, this leads to a classification of functionals with a measure representation which is quite easy and can often be checked with semidefinite optimization.

6.1 The Moment Problem and Haviland's Theorem

Let $\varphi\colon \mathbb{R}[\underline{x}] \to \mathbb{R}$ be a linear map (also called *functional*). We ask whether there exists a (Borel)-measure μ on $\mathbb{R}^n$, such that

$$\varphi(p) = \int_{\mathbb{R}^n} p \, d\mu$$

holds for all $p \in \mathbb{R}[\underline{x}]$. Note that this requires all the integrals on the right to exist, of course. In particular we must have

$$\mu(\mathbb{R}^n) = \int 1 d\mu = \varphi(1) < \infty.$$

Since φ is linear, it is enough to check whether

$$\varphi(\underline{x}^\alpha) = \int_{\mathbb{R}^n} \underline{x}^\alpha \mathrm{d}\mu$$

holds for all $\alpha \in \mathbb{N}^n$. The numbers on the right are called the **moments** of the measure μ. So we ask whether the prescribed multi-sequence $(\varphi(\underline{x}^\alpha))_{\alpha\in\mathbb{N}^n}$ of real numbers is the sequence of moments of a measure. One can also ask for the support of the measure to be contained in a prescribed set. These questions are known as the **moment problem**. Here are some classical results on the moment problem. We will later deduce them from Haviland's theorem and some of our Positivstellensätze.

Theorem 6.1.1 (Hamburger Moment Problem)**.** *A functional $\varphi\colon \mathbb{R}[t] \to \mathbb{R}$ has a representing measure μ on $\mathbb{R}$ if and only if $\varphi(p^2) \geqslant 0$ holds for all $p \in \mathbb{R}[t]$.*

Note that the condition $\varphi(p^2) \geqslant 0$ for all $p \in \mathbb{R}[t]$ is clearly necessary for a representing measure to exist. Squares are globally nonnegative, and integrating a nonnegative function gives a nonnegative number. Further note that the condition can also be written as

$$\varphi \in M_d(1)^\vee$$

for all d, using the notation from Example 5.3.4. So the condition states containment in a sequence of spectrahedra, which is accessible by semidefinite optimization.

Theorem 6.1.2 (Stieltjes Moment Problem)**.** *A functional $\varphi\colon \mathbb{R}[t] \to \mathbb{R}$ has a representing measure μ on $[0,\infty)$, i.e. $\mu((-\infty,0)) = 0$, if and only if*

$$\varphi(p^2) \geqslant 0 \text{ and } \varphi(p^2 \cdot t) \geqslant 0$$

for all $p \in \mathbb{R}[t]$, i.e. if $\varphi \in M_d(t)^\vee$ for all d.

Theorem 6.1.3 (Hausdorff Moment Problem)**.** *A functional $\varphi\colon \mathbb{R}[t] \to \mathbb{R}$ has a representing measure μ on $[0,1]$ if and only if*

$$\varphi(p^2) \geqslant 0 \text{ and } \varphi(p^2 \cdot t) \geqslant 0 \text{ and } \varphi(p^2(1-t)) \geqslant 0$$

for all $p \in \mathbb{R}[t]$, i.e. if $\varphi \in M_d(t, 1-t)^\vee$ for all d.

A very general classification of functionals with a representing measure is the following:

Theorem 6.1.4 (Haviland's Theorem)**.** *Let $W \subseteq \mathbb{R}^n$ be an arbitrary closed set and $\varphi\colon \mathbb{R}[\underline{x}] \to \mathbb{R}$ a functional. Then the following are equivalent:*

(i) There is a measure μ on W with $\varphi(p) = \int_W p \, d\mu$ for all $p \in \mathbb{R}[\underline{x}]$.

(ii) For any $p \in \mathbb{R}[\underline{x}]$ with $p \geqslant 0$ on W one has $\varphi(p) \geqslant 0$.

Remark 6.1.5. (i) The implication $(i) \Rightarrow (ii)$ in Haviland's Theorem is clear: the integral over a nonnegative function is nonnegative. So the other implication is the interesting one. One can deduce Haviland's Theorem from Riesz's Representation Theorem for functionals on the algebra of continuous functions on a locally compact Hausdorff space. This can for example be found in [4].
(ii) Unfortunately, the condition

$$\text{“}\varphi(p) \geqslant 0 \text{ for all nonnegative polynomials } p\text{”}$$

is a priori not easier to check than the question for a representing measure. But we can use results from real algebra here! In case that the set of nonnegative polynomials can be replaced by a finitely generated preordering/quadratic module, it becomes much simpler and accessible by semidefinite optimization. This can work even if not every nonnegative polynomial belongs to the preordering.
(iii) Every globally nonnegative polynomial $p \in \mathbb{R}[t]$ in *one variable* is a sum of squares (Theorem 2.2.1). So Hamburger's result immediately follows from Haviland's Theorem!
(iv) Every on $[0, \infty)$ nonnegative polynomial belongs to the quadratic module $M(t)$ (Exercise 32). So also Stieltjes' result follows from Haviland's Theorem.
(v) Every on $[0, 1]$ nonnegative polynomial belongs to the quadratic module $M(t, 1-t)$ (Exercise 32). So also Hausdorff's result follows from Haviland' Theorem. △

These considerations justify the following definitions. Let $p_1, \ldots, p_r \in \mathbb{R}[\underline{x}] = \mathbb{R}[x_1, \ldots, x_n]$ be given. We again consider the finitely generated quadratic module

$$M = M(p_1, \ldots, p_r) = \left\{\sigma_0 + \sigma_1 p_1 + \cdots + \sigma_r p_r \mid \sigma_i \in \Sigma\mathbb{R}[\underline{x}]^2\right\}.$$

The case of a finitely generated preordering is covered by this as well, since it is the quadratic module generated by the products of the p_i.
We consider its *dual cone* in the algebraic dual space $\mathbb{R}[\underline{x}]^*$:

$$M^\vee = M(p_1, \ldots, p_r)^\vee := \{\varphi \colon \mathbb{R}[\underline{x}] \to \mathbb{R} \text{ linear} \mid \varphi \geqslant 0 \text{ on } M\}$$

and the *double dual*

$$M^{\vee\vee} = M(p_1, \ldots, p_r)^{\vee\vee} := \{p \in \mathbb{R}[\underline{x}] \mid \varphi(p) \geqslant 0 \,\forall \varphi \in M^\vee\}.$$

Note that we define the double dual in $\mathbb{R}[\underline{x}]$ instead of the larger space $(\mathbb{R}[\underline{x}]^*)^*$. The basic closed set

$$W = W_{\mathbb{R}}(p_1, \ldots, p_r) = \{a \in \mathbb{R}^n \mid p_1(a) \geqslant 0, \ldots, p_r(a) \geqslant 0\}$$

is also familiar already, and we have already worked with the set of polynomials that are nonnegative on W. We call it the **saturation** of M from now on:

$$M^{\mathrm{sat}} = M(p_1, \ldots, p_r)^{\mathrm{sat}} := \{p \in \mathbb{R}[\underline{x}] \mid p \geqslant 0 \text{ on } W(p_1, \ldots, p_r)\}\,.$$

From now on we ignore the condition $-1 \notin M$ for quadratic modules and preorderings, to avoid some extra case distinctions. In case $-1 \in M$, i.e. $M = \mathbb{R}[\underline{x}]$, everything is trival anyway.

Theorem 6.1.6. *$M^{\vee\vee}$ is a quadratic module, and even a preordering, if M was. The saturation M^{sat} is always a preordering. We have the following inclusion:*

$$M \subseteq M^{\vee\vee} \subseteq M^{\mathrm{sat}}.$$

Proof. $M \subseteq M^{\vee\vee}$ is clear from the definition. We now show that $M^{\vee\vee}$ is indeed a quadratic module. For $p, q \in M^{\vee\vee}$ and $\varphi \in M^{\vee}$ we have

$$\varphi(p + q) = \varphi(p) + \varphi(q) \geqslant 0.$$

Thus $M^{\vee\vee}$ is closed under $+$. Now let $f \in \mathbb{R}[\underline{x}]$ be arbitrary. We have to show that $\varphi(f^2 p) \geqslant 0$ holds. For this we define a new functional

$$\psi\colon \mathbb{R}[\underline{x}] \to \mathbb{R}; \quad g \mapsto \varphi(f^2 g).$$

For any $m \in M$ we have $\psi(m) = \varphi(f^2 m) \geqslant 0$, since $f^2 m \in M$. So $\psi \in M^{\vee}$, and thus

$$0 \leqslant \psi(p) = \varphi(f^2 p).$$

This shows that $M^{\vee\vee}$ is a quadratic module. A similar argument shows that $M^{\vee\vee}$ is a preordering, if M was.
The set M^{sat} of all polynomials that are nonnegative on W is clearly a preordering, see Remark 3.1.2 (*iii*) (it was called T_W there). What remains to show is $M^{\vee\vee} \subseteq M^{\mathrm{sat}}$. For every point $a \in W$ we have the evaluation map

$$\delta_a\colon \mathbb{R}[\underline{x}] \to \mathbb{R}; \quad g \mapsto g(a)$$

which is clearly linear and belongs to $M^{\vee}$ (since polynomials from M are nonnegative on W!). Thus for $p \in M^{\vee\vee}$ we have

$$0 \leqslant \delta_a(p) = p(a),$$

i.e. $p \in M^{\mathrm{sat}}$. □

Remark 6.1.7. One can show that $M^{\vee\vee}$ is a closure of M with respect to a suitable topology, namely the *finest locally convex topology* on the vectorspace $\mathbb{R}[\underline{x}]$. In this topology a set is closed if and only if each intersection with a finite-dimensional subspace of $\mathbb{R}[\underline{x}]$ is closed (w.r.t. the euclidean topology on the finite-dimensional space). This is why we call $M^{\vee\vee}$ the **closure** of M. △

Definition 6.1.8. (i) M is called **closed** if $M = M^{\vee\vee}$ holds.
(ii) M has the **strong moment property** (SMP) if $M^{\vee\vee} = M^{\mathrm{sat}}$ holds.
(iii) M is called **saturated** if $M = M^{\mathrm{sat}}$ holds. △

The significance of (SMP) is summarized in the following result:

Theorem 6.1.9. *Let $M = M(p_1, \ldots, p_r)$ have (SMP) and set $W = W_{\mathbb{R}}(p_1, \ldots, p_r)$. Then a linear functional φ on $\mathbb{R}[\underline{x}]$ admits a representing measure on W if and only if $\varphi \in M^{\vee}$.*

Proof. Any $\varphi \in M^{\vee}$ is nonnegative on $M^{\vee\vee} = M^{\mathrm{sat}}$, and thus has a representing measure on W by Haviland's Theorem. □

Example 6.1.10. In Hamburger's, Stieltjes's and Hausdorff's moment problems, the quadratic modules $M(1)$, $M(t)$ and $M(t, 1-t)$ are even saturated (see Remark 6.1.5 (iii)-(v)), and thus have (SMP). △

Now Schmüdgen's and Putinar's Positivstellensätze give us the following strong result for bounded sets:

Theorem 6.1.11. *Let $p_1, \ldots, p_r \in \mathbb{R}[\underline{x}]$ and assume that $M = M(p_1, \ldots, p_r)$ is Archimedean. Then M has (SMP). This is automatic in case that $W = W_{\mathbb{R}}(p_1, \ldots, p_r)$ is bounded and M is a preordering.*

Proof. Take $p \in M^{\mathrm{sat}}$. Then $p + \epsilon > 0$ on W for all $\epsilon > 0$, and thus $p + \epsilon \in M$ by Theorem 4.3.5. For every $\varphi \in M^{\vee}$ we thus have

$$0 \leqslant \varphi(p + \epsilon) = \varphi(p) + \epsilon\varphi(1),$$

and this implies $\varphi(p) \geqslant 0$. So $p \in M^{\vee\vee}$. The last statement is Theorem 4.2.2. □

Example 6.1.12. For deciding whether a functional $\varphi\colon \mathbb{R}[\underline{x}] \to \mathbb{R}$ is integration with respect to a measure on the unit ball of $\mathbb{R}^n$, we have to check whether

$$\varphi(p^2) \geqslant 0, \qquad \varphi(p^2 \cdot (1 - x_1^2 - \cdots - x_n^2)) \geqslant 0$$

holds for all $p \in \mathbb{R}[\underline{x}]$. In fact the quadratic module $M(1 - x_1^2 - \cdots - x_n^2)$ is Archimedean (as argued in Section 4.3). △

In case of a non-compact set, there exists another result of Schmüdgen that allows the reduce the dimension when checking (SMP). We will just illustrate this in one example.

Example 6.1.13. Let $p_1 = 1-x^2 \in \mathbb{R}[x,y]$. Then $W = W_{\mathbb{R}}(p_1) \subseteq \mathbb{R}^2$ is the vertical strip over the interval $[-1,1]$. So far we don't know whether

$$M = M(p_1) = P(p_1)$$

has (SMP). The Nichtnegativstellensatz only provides sums-of-squares certificates with denominators here, and thus no direct relation between M^{sat} and $M^{\vee\vee}$. Now there exists a non-constant polynomial q that is bounded on W, for example $q = x$. *Schmüdgen's Fiber Theorem* says that we can restrict to all fibers of a bounded polynomial when checking (SMP). A fiber is the set of all points on which the polynomial takes a prescribed constant value. For example, $q = x$ takes the value $r \in [-1,1]$ on the vertical line $\{x = r\}$. Restricting to this line means to plug in r for x everywhere, giving rise to the quadratic module

$$M(1-r^2) = \Sigma\mathbb{R}[y]^2 \subseteq \mathbb{R}[y].$$

This module however has (SMP), by Hamburger's result. Since this is true for all r, and thus for all fibers of the bounded polynomial q, M itself has (SMP). So a functional $\varphi\colon \mathbb{R}[x,y] \to \mathbb{R}$ admits a representing measure on the vertical strip if and only if

$$\varphi(p^2) \geqslant 0 \text{ and } \varphi(p^2 \cdot (1-x^2)) \geqslant 0$$

holds for all $p \in \mathbb{R}[x,y]$. △

After having states several positive results on the moment problem, we will see some negative results in the next section.

6.2 Stability

Definition 6.2.1. Let $p_1, \ldots, p_r \in \mathbb{R}[\underline{x}]$. The quadratic module $M(p_1, \ldots, p_r)$ is called **stable**, if for every $d \in \mathbb{N}$ there exists some $e \in \mathbb{N}$ with

$$M(p_1, \ldots, p_r) \cap \mathbb{R}[\underline{x}]_d \subseteq M_e(p_1, \ldots, p_r).$$

Here we use again the truncated quadratic modules introduced in Section 5.2. △

Remark 6.2.2. (i) Stability just says that every polynomial $p \in M(p_1, \ldots, p_r)$ admits a representation with sums of squares of bounded degree, where the bound depends only on the degree of p.
(ii) A priori, it looks as if stability might depend on the generators of the quadratic module. However it can be shown that it does not, it is really a property of the quadratic module alone (Exercise 53).
(iii) We have seen in Example 4.2.4 (vi) that degree bounds cannot exist in general. So not every quadratic module is stable.
(iv) Stability seems to appear exactly in the cases where the set $W_{\mathbb{R}}(p_1, \ldots, p_r)$ is *very* non-compact. This is illustrated in the following result. △

Theorem 6.2.3. *If $W_{\mathbb{R}}(p_1, \ldots, p_r)$ contains a full-dimensional convex cone (with arbitrary vertex), then $M(p_1, \ldots, p_r)$ is stable.*

Proof. We can clearly assume that the vertex of the cone is in the origin. So let $B \subseteq \mathbb{R}^n$ be a full-dimensional ball such that

$$\mathrm{cc}(B) = \{\lambda a \mid a \in B, \lambda \geqslant 0\} \subseteq W_{\mathbb{R}}(p_1, \ldots, p_r) =: W.$$

If a polynomial q is nonnegative on W, we have $0 \leqslant q(\lambda b)$ for all $b \in B$ and $\lambda \geqslant 0$. So the homogeneous term of q of highest degree must be nonnegative on B. In fact, if $q = q_0 + \cdots + q_d$ is written as a sum of homogeneous terms, we have

$$0 \leqslant q(\lambda b) = q_0 + \lambda q_1(b) + \cdots + \lambda^d q_d(b),$$

and a polynomial with a negative leading coefficient would take negative values for some large enough λ.
If two polnomials are nonnegative on a set B with nonempty interior, they cannot sum to zero. There in fact exists a point at which both are positive.
So when adding two polynomials that are nonnegative on W, their leading forms cannot cancel. If we apply this to the single terms in the expression

$$\sigma_0 + \sigma_1 p_1 + \cdots + \sigma_r p_r,$$

we directly obtain stability of M, even with $e = d$. □

Remark 6.2.4. In Theorem 6.2.3 we not just get the strong statement $e = d$, but also that *every* representation of an element in $M(p_1, \ldots, p_r)$ reveals these degree bounds. This wouldn't be necessary for stability, in which just the existence of *at least one* such representation for every elements is required. But in fact most known results on stability prove this stronger statement. △

Example 6.2.5. (i) The quadratic module/preordering $M(1) = \Sigma\mathbb{R}[\underline{x}]^2$ is stable. But if we look closely at the definition of stability, we see that this is in fact trivial by definition. But since $W_{\mathbb{R}}(1) = \mathbb{R}^n$ contains a full-dimensional convex cone, the argument from the last proof gives something even stronger, namely the statement of Lemma 2.2.14.
(ii) The quadratic module/preordering $M(t) \subseteq \mathbb{R}[t]$ is stable, since $W_{\mathbb{R}}(t) = [0, \infty) \subseteq \mathbb{R}$ contains a full-dimensional convex cone. If we have

$$p = \sigma_0 + \sigma_1 t$$

with sums of squares σ_i, we even get $\deg(\sigma_0), \deg(\sigma_1 t) \leqslant \deg(p)$.
(iii) The quadratic module $M(x, y) \subseteq \mathbb{R}[x, y]$ is stable, since $W_{\mathbb{R}}(x, y)$ is the positive orthant in $\mathbb{R}^2$, which contains a full-dimensional convex cone. △

Stability allows us to show an equality in the chain $M \subseteq M^{\vee\vee} \subseteq M^{\mathrm{sat}}$:

Theorem 6.2.6. *If $M = M(p_1, \ldots, p_r)$ is stable and $W_{\mathbb{R}}(p_1, \ldots, p_r)$ has nonempty interior in $\mathbb{R}^n$, then M is closed, i.e. we have $M = M^{\vee\vee}$.*

Proof. We will show that $M \cap U$ is closed in U, for every finite-dimensional subspace U of $\mathbb{R}[\underline{x}]$. This means closedness in the finest locally convex topology, which is equivalent to $M = M^{\vee\vee}$.
Since every finite-dimensional subspace U is contained in some $\mathbb{R}[\underline{x}]_d$, it is sufficient to show that $M \cap \mathbb{R}[\underline{x}]_d$ is closed in $\mathbb{R}[\underline{x}]_d$, for all $d \geqslant 0$. By stability we have

$$M \cap \mathbb{R}[\underline{x}]_d = M_e \cap \mathbb{R}[\underline{x}]_d,$$

and it is thus enough to show closedness of each M_e.
But this is done as in Example 5.4.7. Since in any sequence from $M_e(p_1, \ldots, p_r)$ the degrees of all sums of squares remain bounded, one can pass to a limit (for example by using the Bolzano-Weierstraß Theorem coefficientwise, after a suitable normalization/scaling). It is needed that $W_{\mathbb{R}}(p_1, \ldots, p_r)$ has nonempty interior to see that scaling is in fact not even necessary (Exercise 54). □

Example 6.2.7. (i) The sums of squares $\Sigma\mathbb{R}[\underline{x}]^2$ are stable, and thus closed.
(ii) The quadratic module $M(x, y) \subseteq \mathbb{R}[x, y]$ is closed, i.e. we have

$$M(x, y) = M(x, y)^{\vee\vee} \subseteq M(x, y)^{\mathrm{sat}}.$$ △

In the last section we introduced the strong moment property (SMP). The following result provides a connection to stability.

Theorem 6.2.8. *Let $n \geqslant 2$ and assume $W_{\mathbb{R}}(p_1, \ldots, p_r)$ has nonempty interior in $\mathbb{R}^n$. Then stability and (SMP) for $M(p_1, \ldots, p_r)$ are mutually exclusive.*

Proof. Assume w.l.o.g. that $p_i \neq 0$ for all $i = 1, \ldots, r$. Then there exists a point in $W := W_{\mathbb{R}}(p_1, \ldots, p_r)$ at which all p_i are strictly positiv. Otherwise the product $p_1 \cdots p_r$ would vanish on W, implying $p_1 \cdots p_r = 0$ in $\mathbb{R}[\underline{x}]$, since W has nonempty interior. We can also assume that the origin is such a point, meaning that all p_i have a strictly positive constant term.
Now assume that $M = M(p_1, \ldots, p_r)$ is stable and has (SMP). By Theorem 6.2.6 we then even have $M = M^{\mathrm{sat}}$, i.e. M is saturated.
Now let $p \in \mathbb{R}[\underline{x}]$ be a globally nonnegative polynomial that is not a sum of squares in $\mathbb{R}[\underline{x}]$. This exists since $n \geqslant 2$, we can take the Motzkin polynomial for example. For $\lambda > 0$ we consider

$$p_\lambda := p(\lambda \underline{x}) = p(\lambda x_1, \ldots, \lambda x_n) \in \mathbb{R}[\underline{x}].$$

Then also each p_λ is globally nonnegative and thus belongs to $M^{\mathrm{sat}} = M$, in particular. The degree of all p_λ is the same, so by stability we get that all p_λ belong to sime fixed M_d. When writing down a corresponding representation for each p_λ and then replacing $\underline{x}$ by $\frac{1}{\lambda}\underline{x}$ again, we obtain representations

$$p = \sigma_0^{(\lambda)} + \sigma_1^{(\lambda)} p_1(\lambda^{-1}\underline{x}) + \cdots + \sigma_r^{(\lambda)} p_r(\lambda^{-1}\underline{x})$$

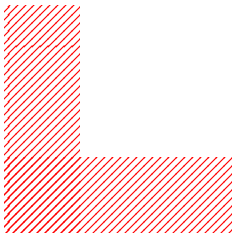
with sums of squares $\sigma_i^{(\lambda)}$ of bounded degree, independent of λ. Now as in Example 5.4.7 and Theorem 6.2.6 we pass to the limit as $\lambda \to \infty$, after scaling to keep all appearing coefficients bounded (use the Bolzano-Weierstraß Theorem again). The $p_i(\lambda^{-1}\underline{x})$ converge to $p_i(0) > 0$. So if a nontrivial scaling was really necessary, the left-hand side becomes 0, the right-hand side becomes a nontrivial sum of squares. This is a contradiction. If no scaling was necessary, we obtain a sum of squares on the right, again a contradiction. □

Example 6.2.9. (i) The sums of squares $\Sigma\mathbb{R}[\underline{x}]^2$ are stable and do thus not have (SMP) for $n \geqslant 2$. So for $n \geqslant 2$ there are positive functionals $\varphi\colon \mathbb{R}[\underline{x}] \to \mathbb{R}$ which do not admit a representing measure!
(ii) For $n = 1$ the sums of squares $\Sigma\mathbb{R}[t]^2$ are stable *and* have (SMP). In fact we have already seen that every globally nonnegative polynomial in one variable is a sum of squares, i.e. $\Sigma\mathbb{R}[t]^2$ is even saturated. So the condition $n \geqslant 2$ cannot be omitted in Theorem 6.2.8.
(iii) $M(x, y) \subseteq \mathbb{R}[x, y]$ is stable, thus closed and does not have (SMP).
(iv) We see again that an Archimedean quadratic modules in dimension $\geqslant 2$ is never stable (cp. Remark 6.2.2 (iii)). By Theorem 6.1.11 they have (SMP). △

To summarize, we have seen the following phenomenon (at least in dimension $\geqslant 2$): If the set W is relatively small (compact, for example), then M tends to have (SMP) and not to be stable. If W is very large instead (contains a full-dimensional cone, for example), then M tends to be stable and not have (SMP).

A nice exact definition of *large* is for example that no non-constant polynomial is bounded on W. With this definition, sets containing a full-dimensional cone are large, and compact sets are not large. However, whether stability holds precisely for such large sets is an unsolved problem. We will finally look at two more interesting examples.

Example 6.2.10. (i) Assume $W \subseteq \mathbb{R}^2$ contains a full-dimensional strip in both coordinate directions:

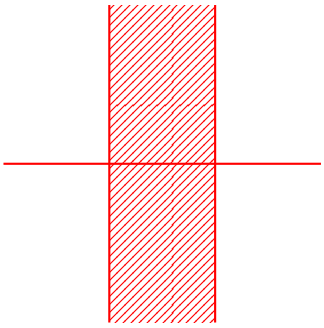


Then no non-constant polynomial is bounded on W (so W is large). From the vertical strip we deduce that a bounded polynomial cannot involve the variable y, from the horizontal strip we get the same for the variable x. Now one can see that when adding to polynomials that are nonnegative on W, the total degree can in fact decrease (other than in Theorem 6.2.3), but at most to one half of the original degrees (Exercise 55). If we apply this to the terms in a representation from $M(p_1, \ldots, p_r)$, we get

$$M \cap \mathbb{R}[\underline{x}]_d \subseteq M_{2d},$$

and thus M is indeed stable.

(ii) Consider the quadratic module $M((1-x^2)y^2) \subseteq \mathbb{R}[x,y]$, whose semialgebraic set $W_{\mathbb{R}}((1-x^2)y^2)$ is a vertical strip, together with the x-axis:



Again there are no non-trivial bounded polynomials. The vertical strip implies that a bounded polynomial contains no y, but no non-constant polynomial in x is bounded on the whole x-axis.
But here, $M((1-x^2)y^2)$ is not stable. If it was, then so would be $M(1-x^2)$ (Exercise 56). But this quadratic module has (SMP), as we have explained in Example 6.1.13, and it is thus not stable. But this example is quite artificial, since the set W contains a lower-dimensional component (the x-axis), which leads to a strange behavior. △

6.3 Saturation

In this section we will look at saturated quadratic modules in a little more detail. Recall that a quadratic module $M = M(p_1, \ldots, p_r) \subseteq \mathbb{R}[\underline{x}]$ is *saturated* if

$$M = M^{\mathrm{sat}}$$

holds, i.e. if M contains every polynomial that is nonnegative on the set $W = W_{\mathbb{R}}(p_1, \ldots, p_r)$. Alternatively, it means that M is closed and has (SMP) at the same time:

$$M = M^{\vee\vee} = M^{\mathrm{sat}}.$$

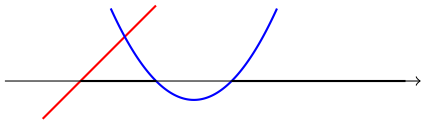
Apart from some examples in dimension 1 we haven't seen this property at all so far. Our Archimedean Positivstellensätze indeed always just applied to *strictly positive* polynomials.
We will first go back to the one-dimensional case, where everything works nicely (see Theorem 6.3.2). We next show that in dimension $\geqslant 3$ in fact *no* saturated finitely generated quadratic module does exist (Theorem 6.3.4). Dimension 2 is a special case in which some interesting things happen (which we will only cite).

But note that we have shown closedness $M = M^{\vee\vee}$ only via stability so far (Theorem 6.2.6). Already in dimension 2 this doesn't appear together with (SMP), and we thus cannot prove saturation in this way.
Now assume $n = 1$, i.e. we consider polynomial in one variable t, and semialgebraic subsets of $\mathbb{R}$. The basic closed semialgebraic sets are finite unions of closed intervals

$$(-\infty, a],\ [a, b],\ [b, \infty)$$

by Theorem 1.5.5 . For every such basic closed set there exist so-called *natural generators*. These are the polynomials that first come to mind when trying to define the set. Precisely, they are defined as follows. If W contains a smallest element a, then $t-a$ is one of the natural generators. If W contains a largest element b, then $b-t$ is one of the natural generators. If $a, b \in W$ with $a < b$ and $(a, b) \cap W = \emptyset$, then $(t-a)(t-b)$ is a natural generator. This altogether gives rise to polynomials $p_1, \ldots, p_r \in \mathbb{R}[t]$ with $W = W_{\mathbb{R}}(p_1, \ldots, p_r)$.

Example 6.3.1. The set $W = [-1, 0] \cup [1, \infty)$ has natural generators $t + 1$ and $t(t-1)$:



△

Theorem 6.3.2. *If $p_1, \ldots, p_r \in \mathbb{R}[t]$ are the natural generators of $W \subseteq \mathbb{R}$, then the preordering $T(p_1, \ldots, p_r)$ is saturated and stable.*

Proof. Take $p \geqslant 0$ on W. We factor p into irreducible polynomials, as in Theorem 1.2.11. We will show that the single factors can be grouped such that the products belong to $T = T(p_1, \ldots, p_r)$. Since the preordering T is multiplicatively closed, this proves the statement.
Factors of degree 2 are sums of squares and thus clearly belong to T. Without these factors, the polynomial is still nonnegative on W.
If a linear term $t - c$ for some $c \in \operatorname{int}(W)$ appears in the factorization, it has to appear with even multiplicity. Otherwise p would be negative in a neighborhood of c in W. We can also ignore such factors.
If W has a smallest element a, and a factor $t - a'$ for some $a' \leqslant a$ appears in p, it lies in T. This is because $t - a' = (t - a) + (a - a')$ and $t - a$ is one of the natural generators of T. The case of a largest element is similar.
Now what can finally still happen is that p has a zero in an interval $[a, b]$ with $a, b \in W, a < b$ and $W \cap (a, b) = \emptyset$. But then p must have another zero in the

same interval, since otherwise it wouldn't be nonnegative on W. Thus p contains a factor $(t-c)(t-d)$ with $a \leqslant c \leqslant d \leqslant b$. So we are done, once we have shown

$$(t-c)(t-d) \in T((t-a)(t-b)),$$

since $(t-a)(t-b) \in T$. Now one can show that there exists some $\lambda \in (0,1]$ such that

$$(t-c)(t-d) - \lambda(t-a)(t-b)$$

is globally nonnegative (Exercise 58). Since it is a sum of squares then, we are done.
Note that the proof already yields stability. In each step we have produced a representation with the best possible degree bound on the sums of squares. This remains true for products and thus even implies

$$T \cap \mathbb{R}[t]_d \subseteq T_d. \qquad \square$$

Remark 6.3.3. Without the natural generators, the preordering will in general not be saturated, as can be see from $t \notin T(t^3)$. $\triangle$

We now pass to dimension $\geqslant 3$. The following result says that $M(p_1, \ldots, p_r)$ is in fact *never* saturated, or equivalently that M^{sat} is never finitely generated as a preordering/quadratic module.

Theorem 6.3.4. *Let $n \geqslant 3$ and $p_1, \ldots, p_r \in \mathbb{R}[\underline{x}]$ be such that $W_{\mathbb{R}}(p_1, \ldots, p_r)$ has nonempty interior in $\mathbb{R}^n$. Then $M(p_1, \ldots, p_r)$ is not saturated.*

Proof. As argued before, we can assume that $W = W_{\mathbb{R}}(p_1, \ldots, p_r)$ contains a point at which all p_i are strictly positive, and this in fact the origin is this point. So all p_i have a strictly positive constant term.
Now let $h \in \mathbb{R}[\underline{x}]$ be *homogeneous* and globally nonnegative, but not a sum of squares. For $n \geqslant 3$ such a polynomial exists, take the homogenized Motzkin polynomial, for example. Then clearly $h \in M^{\mathrm{sat}}$. Now assume $h \in M(p_1, \ldots, p_r)$, i.e. there is a representation

$$h = \sigma_0 + \sigma_1 p_1 + \cdots + \sigma_r p_r \tag{6.1}$$

with sums of squares σ_i. In each of the terms $\sigma_i p_i$, the homogeneous term of lowest degree is a sum of squares (this uses that all p_i have a strictly positive constant term). The homogeneous term of lowest degree on the right in (6.1) is thus a nontrivial sum of squares. But since h is homogeneous, this term must coincide with h, a contradiction. $\square$

Remark 6.3.5. In the proof of Theorem 6.3.4 it might look as if it was always the same polynomial h that does not belong to $M(p_1, \ldots, p_r)$. But of course this is impossible, since it could just be added to the p_i, and would then clearly belong to the quadratic module. But in fact we have assumed h to be homogeneous with respect to the origin, and that all p_i are strictly positive there. In general, this requires a shift, and thus a shift of the polynomial h. But one thing is true: there is one fixed polynomial h (for example the Motzkin polynomial), of which always a suitable shift doesn't belong to M. △

So only in dimension 2 we can expect to find more interesting quadratic modules which are saturated. But usually it requires deep and/or technical methods to prove this. We will thus only state two examples:

Example 6.3.6. The quadratic modules $M(1 - x^2 - y^2)$ and $M(1 - x^2, 1 - y^2)$ are saturated. They define the unit disk and the unit square, respectively.
More generally, if $W = W_{\mathbb{R}}(p_1, \ldots, p_r) \in \mathbb{R}^2$ is bounded, the polynomials p_i are irreducible, and at each point of the boundary of W either just one of the p_i vanishes smoothly, or at most two of the p_i vanish smoothly and intersect transversally, then $M(p_1, \ldots, p_r)$ is saturated. This result was proven by Scheiderer and relies on a local-global principle and results for power series rings. △

Example 6.3.7. The quadratic module $M(1 - x^2) \subseteq \mathbb{R}[x, y]$ is saturated. This is very surprising, since the set $W_{\mathbb{R}}(1 - x^2) \subseteq \mathbb{R}^2$ is a vertical strip and thus not even compact. This rare case was found by Marshall, with a technical and highly nontrivial proof. △

Chapter 7

Non-Commutative Real Algebra and Geometry

In this chapter we discuss some aspects of a relatively new branch of real algebra and geometry, namely the *non-commutative* theory. For example, one tries to prove Positivstellensätze in non-commutative algebras. The corresponding geometry turns out to be much more complicated, but also much more insightful than in the classical setup. A good principle is to use matrix algebras or algebras of operators as guiding examples to develop the theory. This is why we start with an overview of these and some of their properties.

7.1 Matrix Algebras and Algebras of Operators

Let us consider two classes of examples that will guide us through the coming theory.

Matrix Algebras

For $d \in \mathbb{N}$ let $\mathrm{M}_d(\mathbb{C})$ be the set of $d \times d$-matrices with entries from $\mathbb{C}$. The set $\mathrm{M}_d(\mathbb{C})$ is endowed with the structure of a $\mathbb{C}$-algebra with multiplicative identity I_d. For $d \geqslant 2$ it is not commutative!
In context of positivity, we have usually used $\mathbb{R}$ as our ground field. But since we will always use an *involution* in the non-commutative setup anyway, whose set of fixed points play the role of real elements, we can pass to $\mathbb{C}$ here. This will make some of the later arguments simpler.

The involution that we will use on matrix algebras is the well-known $*$-operation, i.e. for $M = (m_{ij})_{i,j}$ we set

$$M^* = (\overline{m}_{ji})_{i,j}\,.$$

This defines a *conjugate-linear self-inverse antiautomorphism* on $\mathrm{M}_d(\mathbb{C})$, i.e. it fulfills

$$(\lambda M + \gamma N)^* = \overline{\lambda} M^* + \overline{\gamma} N^*,\ (M^*)^* = M,\ \text{ and }\ (MN)^* = N^* M^*$$

for all matrices $M, N \in \mathrm{M}_d(\mathbb{C})$ und $\lambda, \gamma \in \mathbb{C}$. These are precisely the axioms that we will later also use to define an involution abstractly.
A fixed point of the involution, i.e. a matrix that fulfills $M^* = M$, is called a **Hermitian matrix**, and the set of all these matrices forms an $\mathbb{R}$-subspace of $\mathrm{M}_d(\mathbb{C})$, but *not* a $\mathbb{C}$-subspace:

$$\mathrm{Her}_d(\mathbb{C}) := \mathrm{M}_d(\mathbb{C})_h := \{M \in \mathrm{M}_d(\mathbb{C}) \mid M^* = M\}\,.$$

Note that for $d \geqslant 2$, the space $\mathrm{Her}_d(\mathbb{C})$ is also *not* an algebra, i.e. not closed under multiplication.
In Definition 2.2.4 and Lemma 2.2.3 we have already encountered a notion of *positivity* that makes sense for matrices, also with complex entries. Recall that a matrix $M \in \mathrm{Her}_d(\mathbb{C})$ is called **positive semidefinite** if

$$v^* M v \geqslant 0 \quad \text{ for all } v \in \mathbb{C}^d.$$

Here, v^* denotes the conjugate-transposed vector of $v \in \mathbb{C}^d$. We denote this also by

$$M \succcurlyeq 0.$$

Note that we define positivity only for Hermitian matrices, just as we have restricted to real symmetric matrices before. This makes sense, for example since the numbers $v^* M v$ might not even be real otherwise:

Positivity takes place in the space of Hermitian elements only!

Now it can happen that the product of two positive matrices is not even Hermitian, and thus not positive again:

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$$

So the notion of a *preordering* cannot be defined in a general non-commutative setup, and we will have to restrict to quadratic modules.

Note that the statement of Lemma 2.2.3 remains true when adapted to the Hermitian setup. For example, M is positive semidefinite if and only if there exists a decomposition

$$M = S^*S$$

for some Matrix $S \in \mathrm{M}_d(\mathbb{C})$ (which can even be chosen Hermitian and positive semidefinite again). One just diagonalizes $U^*MU = D$ with a unitary matrix U and sets $S = U\sqrt{D}U^*$, where $\sqrt{D}$ is the diagonal matrix obtained from taking the real positive square-roots of the diagonal entries of D. This particular S is also denoted by $\sqrt{M}$. One also obtains this S as a polynomial expression $p(M)$ for some $p \in \mathbb{R}[t]$. In fact, any p that coincides with the square-root function on the spectrum of M will do. If we want some uniform polynomial expressions that work for *all* M, we can take a limit $p_n(M)$ for $n \to \infty$. We have to choose the p_n to approximate the square-root function ever closer, on ever larger intervals.
So this is already a first nice non-commutative Positivstellensatz:

$$M \succcurlyeq 0 \text{ if and only if } M \text{ is a Hermitian square in } \mathrm{M}_d(\mathbb{C}).$$

We see here: *In the non-commutative setup, squares are defined via the involution, i.e. we always consider Hermitian squares*. This is because usual squares of the form $Q \cdot Q$ will in general not even be positive, as can already be seen in the 1×1-case:

$$i \cdot i = -1.$$

We finish this class of examples with two interesting results on matrix algebras.

Theorem 7.1.1. $\mathrm{M}_d(\mathbb{C})$ *is a central simple algebra, i.e. it has only the two trivial two-sided ideals, and the center consist of only* $\mathbb{C} \cdot I_d$.

Proof. Let $0 \neq M \in \mathrm{M}_d(\mathbb{C})$. We show that the two-sided ideal generated by M is already the full ring:

$$\langle M \rangle = \mathrm{M}_d(\mathbb{C}).$$

Let E_{ij} be the matrix with a 1 at the (i,j)-position, and zeros elsewhere. The E_{ij} spa $\mathrm{M}_d(\mathbb{C})$ as a $\mathbb{C}$-vectorspace, and thus it is enough to show

$$E_{ij} \in \langle M \rangle$$

for all i, j. The identity

$$E_{ki}E_{ij}E_{jl} = E_{kl}$$

shows that it is enough to prove this for *one* of the matrices E_{ij}. Now if the (i,j)-entry of M is $m \neq 0$, then

$$E_{ij} = m^{-1} \cdot E_{ii} M E_{jj} \in \langle M \rangle.$$

Now assume that M belongs to the center of $\mathrm{M}_d(\mathbb{C})$, i.e. commutes with all other matrices. The identities

$$E_{ij} M = M E_{ij}$$

then show that M must be diagonal with identical diagonal entries. □

The next important statement is about subalgebras of matrix algebras. If $\mathcal{A} \subseteq \mathrm{M}_d(\mathbb{C})$ is a subalgebra, a subspace $V \subseteq \mathbb{C}^d$ is called **$\mathcal{A}$-invariant**, if $Mv \in V$ holds for all $v \in V$ and $M \in \mathcal{A}$. The two spaces $\{0\}$ and V are the two trivial invariant subspaces for any subalgebra $\mathcal{A}$.

Theorem 7.1.2 (Burnside's Theorem)**.** *If the subalgebra $\mathcal{A} \subseteq \mathrm{M}_d(\mathbb{C})$ does not admit a non-trivial invariant subspace, then $\mathcal{A} = \mathrm{M}_d(\mathbb{C})$.*

Proof. $\mathcal{A}$ operates transitively on $\mathbb{C}^d \setminus \{0\}$: for every $0 \neq v \in \mathbb{C}^d$ the set

$$\{0\} \subsetneq \{Mv \mid M \in \mathcal{A}\}$$

is clearly an $\mathcal{A}$-invariant subspace, and thus coincides with $\mathbb{C}^d$. We will now first show that $\mathcal{A}$ contains a matrix of rank 1. Choose $0 \neq P \in \mathcal{A}$. If $\mathrm{rank}(P) \geqslant 2$, choose $v_1, v_2 \in \mathbb{C}^d$ with Pv_1, Pv_2 linearly independent. Then choose $M \in \mathcal{A}$ with $MPv_1 = v_2$, using transitivity of the action. Now $PMPv_1$ and Pv_1 are linearly independent, and thus $PMP - \lambda P \neq 0$ holds for all $\lambda \in \mathbb{C}$. But there exists some λ_0 for which $PM - \lambda_0 I_d$ is not invertible on the space $P\mathbb{C}^d$, since $\mathbb{C}$ is algebraically closed, and each linear map thus has an eigenvalue (here we see why working over $\mathbb{C}$ is beneficial). Thus

$$(PM - \lambda_0 I_d)P$$

has a strictly smaller rank than P, but not zero. By iteration we obtain a matrix $Q \in \mathcal{A}$ of rank 1.
Every other matrix of rank 1 then also lies in $\mathcal{A}$. This uses transitivity of $\mathcal{A}$ on $\mathbb{C}^d$ again (Exercise 59). Since every matrix is a sum of matrices of rank 1, this proves $\mathcal{A} = \mathrm{M}_d(\mathbb{C})$. □

Now assume that $\mathcal{A} \subseteq \mathrm{M}_d(\mathbb{C})$ is even a $*$-subalgebra, i.e. for $M \in \mathcal{A}$ we also have $M^* \in \mathcal{A}$. If $\mathcal{A}$ admits a non-trivial invariant subspace V, then $V^\perp$ is also $\mathcal{A}$-invariant. This uses that $\mathcal{A}$ is closed under $*$:

$$\langle Mw, v\rangle = \langle w, \underbrace{M^* v}_{\in V}\rangle = 0$$

for $M \in \mathcal{A}, v \in V$ and $w \in V^\perp$. So after a unitary change of basis, all matrices from $\mathcal{A}$ have block-diagonal form, i.e. $\mathcal{A}$ is becomes a subalgebra of some algebra $\mathrm{M}_{d_1}(\mathbb{C}) \oplus \mathrm{M}_{d_2}(\mathbb{C})$ with $1 \leqslant d_1, d_2$ and $d_1 + d_2 = d$. This can clearly be iterated until the single blocks don't admit a nontrivial invariant subspace anymore, and then Burnside's Theorem says that they form a full matrix algebra. So in total we can assume

$$\mathcal{A} \subseteq \mathrm{M}_{d_1}(\mathbb{C}) \oplus \cdots \oplus \mathrm{M}_{d_r}(\mathbb{C})$$

with $d_1 + \cdots + d_r = d$, and the projection to each summand is surjective on $\mathcal{A}$.

Example 7.1.3. Let $\mathcal{A} \subseteq \mathrm{M}_d(\mathbb{C})$ be a *commutative* $*$-subalgebra. As just explained, we can assume

$$\mathcal{A} \subseteq \mathrm{M}_{d_1}(\mathbb{C}) \oplus \cdots \oplus \mathrm{M}_{d_r}(\mathbb{C})$$

and the projection to each block is surjective on $\mathcal{A}$. On the other hand, all elements of $\mathcal{A}$ commute, and this implies $d_i = 1$ for all i. So after a suitable unitary conjugation, $\mathcal{A}$ *consists of only diagonal matrices*. △

Operators on Hilbert Space

The second important class of examples are algebras of operators on a Hilbert space. Strictly speaking, this contains all the matrix algebras, which arise from finite-dimensional Hilbert spaces. But now we will mostly think about infinite-dimensional spaces.
Let $\mathcal{H}$ be a Hilbert space over $\mathbb{C}$. A linear map

$$T \colon \mathcal{H} \to \mathcal{H}$$

(also called **operator**) is continuous if and only if it is **bounded**, i.e. fulfills

$$\|Tv\| \leqslant C\|v\|$$

for some $C \geqslant 0$ and all $v \in \mathcal{H}$. The smallest such C (in infimum, to be precise) is called the **operator norm** of T, and is usually denoted by $\|T\|_{\mathrm{op}}$. The set of all continuous/bounded operators on $\mathcal{H}$ is denoted by $\mathcal{B}(\mathcal{H})$. Now the set $\mathcal{B}(\mathcal{H})$ turns

out to be a Banach space (with respect to the operator norm), and it is also an algebra with involution. Multiplication is composition of operators, the involution is given by the **adjoint** of an operator, which exists and is uniquely defined by the condition

$$\langle Tv, w\rangle = \langle v, T^*w\rangle$$

for all $v, w \in \mathcal{H}$. A fixed point of the involution is called a **self-adjoint operator**, and again the set

$$\mathcal{B}(\mathcal{H})_h := \{T \in \mathcal{B}(\mathcal{H}) \mid T^* = T\}$$

of all self-adjoint operators is an $\mathbb{R}$-subspace of $\mathcal{B}(\mathcal{H})$.

Example 7.1.4. Consider

$$\mathcal{H} = \ell^2(\mathbb{Z}) = \left\{ (a_i)_{i\in\mathbb{Z}} \mid a_i \in \mathbb{C}, \sum_i |a_i|^2 < \infty \right\},$$

the Hilbert space of all square-summable sequences, with inner product

$$\langle (a_i)_i, (b_i)_i\rangle = \sum_i a_i \overline{b}_i.$$

(i) Let $m = (m_i)_{i\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$ be a *bounded* sequence, i.e. $|m_i| \leqslant C$ holds for some $C \geqslant 0$ and all i. Then the following defines a bounded **multiplication operator** M_m on $\mathcal{H}$, with $\|M_m\|_{\mathrm{op}} \leqslant C$:

$$M_m \colon (a_i)_i \mapsto (m_i a_i)_i.$$

This is a straightforward generalization of a diagonal matrix. The adjoint operator M_m^* is multiplication with the complex conjugate sequence $\overline{m} = (\overline{m}_i)_i$, and thus M_m is self-adjoint if and only if all m_i are real.
(ii) The **left-shift operator** L is defined as

$$L \colon (a_i)_i \mapsto (a_{i-1})_i.$$

It is norm-preserving, i.e. fulfills $\|Lv\| = \|v\|$ for all $v \in \mathcal{H}$, in particular we have $\|L\|_{\mathrm{op}} = 1$ and $L \in \mathcal{B}(\mathcal{H})$. Its adjoint operator is the right-shift operator

$$L^* = R \colon (a_i)_i \mapsto (a_{i+1})_i.$$

So L is unitary but not self-adjoint. △

A self-adjoint operator $T \in \mathcal{B}(\mathcal{H})_h$ is called **positive semidefinite**, if

$$\langle Tv, v\rangle \geqslant 0 \quad \forall v \in \mathcal{H}.$$

We also denote this by $T \succcurlyeq 0$. The notion is thus a direct generalization of the known one for matrices. Also here we assume self-adjointness to ensure that all the numbers $\langle Tv, v\rangle$ are real:

$$\overline{\langle Tv, v\rangle} = \langle v, Tv\rangle = \langle T^*v, v\rangle = \langle Tv, v\rangle.$$

The spectral theory and functional calculus for self-adjoint operators in Hilbert spaces provide an exact analog to the Positivstellensatz for matrices: A self-adjoint operator $T \in \mathcal{B}(\mathcal{H})_h$ is positive semidefinite if and only if

$$T = S^*S$$

holds for some $S \in \mathcal{B}(\mathcal{H})$.

Example 7.1.5. The multiplication operator M_m from Example 7.1.4 (i), with a real sequence $m = (m_i)_i$, is self-adjoint, and positive semidefinite if and only if all $m_i \geqslant 0$. In this case, the multiplication operator $M_{\sqrt{m}}$ with $\sqrt{m} = (\sqrt{m_i})_i$ provides a square decomposition of M_m. △

Sometimes it will be necessary to consider *unbounded* operators, i.e. operators that are not continuous, and the underlying space need not even be a Hilbert space then. So let $\mathcal{D}$ be a $\mathbb{C}$-vectorspace with inner product, not necessarily complete. By

$$\mathcal{L}(\mathcal{D})$$

we denote the set of all linear maps $T\colon \mathcal{D} \to \mathcal{D}$, this time without any continuity/boundedness condition. They again form an algebra. If $T \in \mathcal{L}(\mathcal{D})$ is continuous on $\mathcal{D}$, we can extend T uniquely to a linear bounded map to the completion $\mathcal{H}$ of $\mathcal{D}$, i.e. we can understand T as an element of $\mathcal{B}(\mathcal{H})$ (see Exercise 64 for an even more general statement).
Now in general, the definition of adjoint operators in $\mathcal{L}(\mathcal{D})$ is problematic. We will just need the following: We call an operator $T \in \mathcal{L}(\mathcal{D})$ **self-adjoint**, if

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

holds for all $v, w \in \mathcal{D}$, and we call a self-adjoint operator **positive semidefinite**, if

$$\langle Tv, v\rangle \geqslant 0$$

holds for all $v \in \mathcal{D}$.

7.2 Algebras and Representations

We will now consider general $*$-algebras and their representations. Again we assume that every algebra has a unit element 1, and homomorphisms map 1 to 1.

Definition 7.2.1. (i) A $*$**-algebra** is a (not necessarily commutative) $\mathbb{C}$-algebra $\mathcal{A}$, together with an involution $*$, i.e. a conjugate-linear self-invers antiautomorphism:

$$(\lambda a + \gamma b)^* = \overline{\lambda} a^* + \overline{\gamma} b^*, \quad (a^*)^* = a, \quad (ab)^* = b^* a^*$$

for all $a, b \in \mathcal{A}$, $\lambda, \gamma \in \mathbb{C}$.
(ii) For a given $*$-algebra $\mathcal{A}$, the $\mathbb{R}$-subspace

$$\mathcal{A}_h := \{a \in \mathcal{A} \mid a^* = a\}$$

is called the space of **Hermitian/self-adjoint elements**.
(iii) By

$$\sum \mathcal{A}^2 := \left\{ \sum_{i=1}^{m} a_i^* a_i \mid m \in \mathbb{N}, a_i \in \mathcal{A} \right\}$$

we denote the set of **sums of Hermitian squares**. Clearly $\sum \mathcal{A}^2$ is a convex cone in $\mathcal{A}_h$, and

$$a^* \cdot \sum \mathcal{A}^2 \cdot a \subseteq \sum \mathcal{A}^2$$

holds for all $a \in \mathcal{A}$.
(iv) A $*$**-algebra homomorphism** is an algebra homomorphism

$$\pi \colon \mathcal{A} \to \mathcal{B}$$

between $*$-algebras, that also fulfills $\pi(a^*) = \pi(a)^*$ for all $a \in \mathcal{A}$.
(v) A **(bounded)** $*$**-representation** of $\mathcal{A}$ is a $*$-algebra homomorphism

$$\pi \colon \mathcal{A} \to \mathcal{B}(\mathcal{H})$$

for some Hilbert space $\mathcal{H}$. A $*$-representation is called **finite-dimensional**, if $\mathcal{H}$ is finite-dimensional. After fixing a basis it just means

$$\pi \colon \mathcal{A} \to \mathrm{M}_d(\mathbb{C}).$$

(vi) An **unbounded** $*$**-representation** is a $*$-algebra homomorphism

$$\pi \colon \mathcal{A} \to \mathcal{L}(\mathcal{D})$$

for some inner product space $\mathcal{D}$, fulfilling

$$\langle \pi(a)v, w\rangle = \langle v, \pi(a^*)w\rangle$$

for all $v, w \in \mathcal{D}$ and all $a \in \mathcal{A}$.
(vii) A **state** on $\mathcal{A}$ is a $\mathbb{C}$-linear functional $\varphi\colon \mathcal{A} \to \mathbb{C}$, that fulfills

$$\varphi(1) = 1 \text{ and } \varphi(a^*a) \geqslant 0$$

for all $a \in \mathcal{A}$. Often one additionally requires $\varphi(a^*) = \overline{\varphi(a)}$ for all a, but this in fact already follows from nonnegativity on squares (Exercise 60). In particular, $\varphi\colon \mathcal{A}_h \to \mathbb{R}$ is an $\mathbb{R}$-linear functional. △

Remark 7.2.2. The $*$-representations of a general $*$-algebra play the roles of geometric points corresponding to the algebra. One can *evaluate* an algebra element a at such a point π, by applying the representation to the point

$$a(\pi) := \pi(a)$$

and the result is a concrete operator/matrix. We will argue in the next example that in the commutative setup this corresponds to classical point evaluations.

Example 7.2.3. (i) Let $\mathcal{A}$ be a commutative $*$-algebra and $\pi\colon \mathcal{A} \to \mathrm{M}_d(\mathbb{C})$ a finite-dimensional $*$-representation. When we apply the result from Example 7.1.3 to the image of π, we see that we can assume

$$\pi\colon \mathcal{A} \to \mathrm{M}_1(\mathbb{C}) \oplus \cdots \oplus \mathrm{M}_1(\mathbb{C}),$$

i.e. π is just a d-tuple of $*$-algebra homomorphisms $\varphi_i\colon \mathcal{A} \to \mathbb{C}$. So increasing the dimensions of representations does not add anything new in the commutative case.
(ii) Consider $\mathcal{A} = \mathbb{C}[x_1, \ldots, x_n]$, with involution which is just complex conjugation of coefficients. We have

$$\mathbb{C}[x_1, \ldots, x_n]_h = \mathbb{R}[x_1, \ldots, x_n] \text{ and } \sum \mathbb{C}[x_1, \ldots, x_n]^2 = \sum \mathbb{R}[x_1, \ldots, x_n]^2,$$

so we are precisely in the well-known context of the previous chapters. By (i), the only relevant finite-dimensional $*$-representations of $\mathbb{C}[x_1, \ldots, x_n]$ are (after a possible change of basis) evaluations at points of $\mathbb{R}^n$. △

Remark 7.2.4. (i) Let $a \in \sum \mathcal{A}^2$ and let π be a (bounded or unbounded) $*$-representation of $\mathcal{A}$. Then $\pi(a)$ is positive semidefinite. In fact, for $a = \sum_i a_i^* a_i$ and any vector v we have

$$\langle \pi(a)v, v\rangle = \sum_i \langle \pi(a_i^*)\pi(a_i)v, v\rangle = \sum_i \langle \pi(a_i)v, \pi(a_i)v\rangle = \sum_i \|\pi(a_i)v\|^2 \geqslant 0.$$

So every sum of squares is nonnegative on the geometric set of $*$-representations. (ii) Every (bounded or unbounded) $*$-representation π of $\mathcal{A}$ gives rise to many states φ on $\mathcal{A}$. Indeed, for every $v \in \mathcal{H}$ ($v \in \mathcal{D}$, respectively) with $\|v\| = 1$ we get a state defined by

$$\varphi(a) := \langle \pi(a)v, v\rangle. \qquad \triangle$$

The important construction of *Gelfand, Neumark* and *Segal*, also called the **GNS-construction**, provides a converse to the last remark. Given a state $\varphi\colon \mathcal{A} \to \mathbb{C}$, one can construct a $*$-representation

$$\pi_\varphi\colon \mathcal{A} \to \mathcal{L}(\mathcal{D})$$

with a distinguished vector $v \in \mathcal{D}$, such that

$$\varphi(a) = \langle \pi_\varphi(a)v, v\rangle$$

holds for all $a \in \mathcal{A}$. Before explaining this in detail, we first provide two lemmas.

Lemma 7.2.5 (Cauchy-Schwarz Inequality)**.** *Let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be a state. Then for all $a, b \in \mathcal{A}$ we have*

$$|\varphi(b^*a)|^2 \leqslant \varphi(b^*b)\varphi(a^*a).$$

Proof. The Hermitian matrix

$$M := \begin{pmatrix} \varphi(a^*a) & \varphi(a^*b) \\ \varphi(b^*a) & \varphi(b^*b) \end{pmatrix} \in \mathrm{Her}_2(\mathbb{C})$$

is positive semidefinite. In fact, for $v = (v_1, v_2)^t \in \mathbb{C}^2$ we have

$$v^*Mv = \varphi\left((v_1a + v_2b)^*(v_1a + v_2b)\right) \geqslant 0.$$

Thus M has a nonnegative determinant, and we compute

$$\det(M) = \varphi(a^*a)\varphi(b^*b) - \varphi(a^*b)\varphi(b^*a) = \varphi(a^*a)\varphi(b^*b) - |\varphi(b^*a)|^2.$$

This proves the claim. $\square$

Lemma 7.2.6. *Let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be a state. Then*

$$N_\varphi := \{a \in \mathcal{A} \mid \varphi(a^*a) = 0\}$$

is a (nontrivial) left-ideal in $\mathcal{A}$.

Proof. Exercise 61. □

Now the GNS-construction works as follows. All statements that we do not prove are covered by Exercise 62. Let φ be a state on $\mathcal{A}$. We first equip the $\mathbb{C}$-vectorspace $\mathcal{A}$ with a sesquilinear form defined by

$$\langle a, b\rangle_\varphi := \varphi(b^*a),$$

which is clearly positive semidefinite, i.e. fulfills $\langle a, a\rangle_\varphi \geqslant 0$ for all a. To make $\langle\cdot,\cdot\rangle_\varphi$ strictly positive definite, i.e. an inner product, we have to factor out N_φ. On the $\mathbb{C}$-vectorspace

$$\mathcal{D} := \mathcal{A}/N_\varphi$$

we get a well-defined inner product $\langle\cdot,\cdot\rangle_\varphi$. Since N_φ is even a left-ideal, multiplication with elements from $\mathcal{A}$ from the the left is well-defined, and so every $a \in \mathcal{A}$ gives rise to a linear operator

$$m_a\colon \mathcal{D} \to \mathcal{D};\ d \mapsto ad.$$

In this way we obtain a $*$-representation

$$\begin{aligned}\pi_\varphi\colon \mathcal{A} &\to \mathcal{L}(\mathcal{D})\\ a &\mapsto m_a\end{aligned}$$

of $\mathcal{A}$. If $v \in \mathcal{D}$ is chosen as the residue class of 1, we get $\|v\|_\varphi = 1$ and

$$\langle\pi_\varphi(a)v, v\rangle_\varphi = \varphi(1^* \cdot a \cdot 1) = \varphi(a)$$

for all $a \in \mathcal{A}$. This is exactly what we wanted to show.

Since sets of $*$-representations can be seen as geometric spaces corresponding to the algebra, it is clear that *positivity* of elements should be defined by positivity on such representations.

Definition 7.2.7. Let $\mathcal{A}$ be a $*$-algebra and $\mathcal{F}$ a class of $*$-representations of $\mathcal{A}$. An element $a \in \mathcal{A}_h$ is called **nonnegative on** $\mathcal{F}$, if $\pi(a)$ is a positive semidefinite operator, for all $\pi \in \mathcal{F}$. △

Remark 7.2.8. By Remark 7.2.4 (i), sums of squares from $\mathcal{A}$ are nonnegative on every class $\mathcal{F}$ of $*$-representations of $\mathcal{A}$. △

Example 7.2.9. Let $\mathcal{A} = \mathbb{C}[x_1, \ldots, x_n]$ be as in Example 7.2.3 (ii), and let $\mathcal{F}$ be the set of all finite-dimensional $*$-representations of $\mathcal{A}$. We now already know that $p \in \mathbb{R}[x_1, \ldots, x_n]$ is nonnegative on $\mathcal{F}$ if and only if p is nonnegative at every point of $\mathbb{R}^n$. It would also be enough to consider the set of all one-dimensional $*$-representations only. We also know that not every such nonnegative polynomial is a sum of squares. △

Let us consider

$$\left(\sum \mathcal{A}^2\right)^{\vee\vee} := \{a \in \mathcal{A}_h \mid \varphi(a) \geqslant 0 \text{ for all states } \varphi \text{ on } \mathcal{A}\}$$

just as we have done for sums of squares in $\mathbb{R}[\underline{x}]$ (the extra condition $\varphi(1) = 1$ for states does not change anything, see Exercise 63). Again, this coincides with the closure of $\sum \mathcal{A}^2$ in the finest locally convex topology on the $\mathbb{R}$-vectorspace $\mathcal{A}_h$. The following result is a general Positivstellensatz for arbitrary $*$-algebras.

Theorem 7.2.10. *Let $\mathcal{A}$ be a $*$-algebra and $\mathcal{F}$ the class of all $*$-representations (bounded and unbounded) of $\mathcal{A}$. Then for all $a \in \mathcal{A}_h$ we have:*

$$a \text{ nonnegative on } \mathcal{F} \quad \Leftrightarrow \quad a \in \left(\sum \mathcal{A}^2\right)^{\vee\vee}$$

Proof. "$\Leftarrow$": Let $\pi\colon \mathcal{A} \to \mathcal{L}(\mathcal{D})$ be a $*$-representation and $v \in \mathcal{D}$ with $\|v\| = 1$. Then the rule

$$b \mapsto \langle \pi(b)v, v\rangle$$

defines a state on $\mathcal{A}$, and by assumption we thus have $\langle \pi(a)v, v\rangle \geqslant 0$. So a is nonnegative on π, and thus on the whole of $\mathcal{F}$.
For "$\Rightarrow$" let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be a state. From the GNS-construction we obtain a $*$-representation

$$\pi_\varphi\colon \mathcal{A} \to \mathcal{L}(\mathcal{D})$$

and some $v \in \mathcal{D}$ with $\varphi(a) = \langle \pi_\varphi(a)v, v\rangle_\varphi$. From $\pi_\varphi \in \mathcal{F}$ we thus get

$$\varphi(a) \geqslant 0,$$

the desired statement. □

Remark 7.2.11. From the proof of Theorem7.2.10 we see that we can choose $\mathcal{F}$ as the set of all $*$-representations on spaces $\mathcal{D}$ with

$$\dim(\mathcal{D}) \leqslant \dim(\mathcal{A}).$$

The space $\mathcal{D}$ that we construct in the GNS-construction is indeed always a quotient of $\mathcal{A}$. △

Example 7.2.12. For $\mathbb{C}[\underline{x}]$ we have seen in Example 6.2.7 (i) that

$$\Big(\sum \mathbb{C}[\underline{x}]^2\Big)^{\vee\vee} = \Big(\sum \mathbb{R}[\underline{x}]^2\Big)^{\vee\vee} = \sum \mathbb{R}[\underline{x}]^2$$

holds. So $p \in \mathbb{R}[\underline{x}]$ is a sum of squares if and only if p is nonnegative at each (bounded and unbounded) $*$-representation of $\mathbb{C}[\underline{x}]$. Nonnegativity at all finite-dimensional $*$-representation is not enough for this in general, since it is only equivalent to nonnegativity of p on $\mathbb{R}^n$ (see Example 7.2.9). △

As we have seen in the commutative setup already, Positivstellensätze become much stronger in case of Archimedean preorderings/quadratic modules. The same is true in the non-commutative case.

Definition 7.2.13. We call $\sum \mathcal{A}^2$ **Archimedean**, if for every $a \in \mathcal{A}_h$ there exists some $r > 0$ with

$$r - a \in \sum \mathcal{A}^2.$$

After dividing by r this is equivalent to

$$1 - \varepsilon a \in \sum \mathcal{A}^2$$

for some $\epsilon > 0$, i.e. to 1 being an **algebraic interior point** of $\sum \mathcal{A}^2$ in the $\mathbb{R}$-vectorspace $\mathcal{A}_h$ (meaning you can walk a little in each direction, starting from 1, without leaving the set). △

The following result is an Archimedean Positivstellensatz for $*$-algebras. We can significantly weaken the assumptions from Theorem 7.2.10, and also obtain a stronger result.

Theorem 7.2.14. *Let $\mathcal{A}$ be a $*$-algebra in which $\sum \mathcal{A}^2$ is Archimedean. Let $\mathcal{F}$ be the class of all bounded $*$-representations of $\mathcal{A}$. Then for all $a \in \mathcal{A}_h$ we have:*

$$a \text{ nonnegative on } \mathcal{F} \quad \Leftrightarrow \quad a + \varepsilon \in \sum \mathcal{A}^2 \quad \forall \varepsilon > 0.$$

Proof. "$\Leftarrow$" is proven as in Theorem 7.2.10. For "$\Rightarrow$" it is enough to show that $a \in (\sum \mathcal{A}^2)^{\vee\vee}$ holds. Since 1 is an interior point of $\sum \mathcal{A}^2$ this in fact implies $a + \varepsilon \in \sum \mathcal{A}^2$ for all $\varepsilon > 0$, using a suitable version of the Hahn-Banach Theorem.
We now proceed as in the proof of Theorem 7.2.10, we just have to ensure that all states on $\mathcal{A}$ give rise to *bounded* $*$-representations via the GNS-construction. So let φ be a state on $\mathcal{A}$ and π_φ the corresponding GNS-representation on $\mathcal{D} = \mathcal{A}/N_\varphi$. For $a \in \mathcal{A}$ we have $r - a^*a \in \sum \mathcal{A}^2$ for some $r > 0$, using the Archimedean property. Each vector $v \in \mathcal{D}$ is the residue class of some $b \in \mathcal{A}$, and we have

$$b^*(r - a^*a)b = rb^*b - b^*a^*ab \in \sum \mathcal{A}^2.$$

This implies

$$\|\pi_\varphi(a)v\|_\varphi^2 = \langle \pi_\varphi(a)v, \pi_\varphi(a)v\rangle_\varphi = \varphi(b^*a^*ab) \leqslant \varphi(rb^*b) = r\langle b, b\rangle_\varphi = r\|v\|_\varphi^2,$$

where the inequality follows from nonnegativity of φ on $\sum \mathcal{A}^2$. So $\pi_\varphi(a)$ is a bounded operator on $\mathcal{D}$ (with operator norm $\leqslant \sqrt{r}$, not even depending on φ), which extends uniquely to a bounded operator on the completion $\mathcal{H}$ of $\mathcal{D}$. In this way we can interpret

$$\pi_\varphi \colon \mathcal{A} \to \mathcal{B}(\mathcal{H})$$

as a bounded $*$-representation (see Exercise 64). □

Remark 7.2.15. In Theorem 7.2.14 we can restrict to the class of all bounded $*$-representations on Hilbert spaces $\mathcal{H}$ that admit a dense subspace $\mathcal{D} \subseteq \mathcal{H}$ with $\dim(\mathcal{D}) \leqslant \dim(\mathcal{A})$. If $\mathcal{A}$ has countable dimension, these are precisely called the **separable Hilbert spaces**. △

We have now proven two very general Positivstellensätze for $*$-algebras. In the following we will consider special classes of algebras, for which we can significantly improve upon the results, and also state the positivity more concretely.

7.3 Non-Commutative Polynomials

In this section we consider the algebra

$$\mathcal{A} = \mathbb{C}\langle \underline{z}\rangle = \mathbb{C}\langle z_1, \ldots, z_n\rangle$$

of polynomials in *non-commuting variables*. $\mathcal{A}$ is also called the **free algebra**. A **word** (or **monomial**) ω in the variables $z_1, \ldots, z_n$ is just a finite formal product

$$\omega = z_{i_1} z_{i_2} \cdots z_{i_k}$$

of some of the variables. The number k is called the **word length** (or **degree**) of ω. Since the variables do not commute, z_1z_2 and z_2z_1 are different words, for example. Words are multiplied by concatenating them formally, for example we have

$$z_1z_3 \cdot z_2z_1 = z_1z_3z_2z_1.$$

Now as a $\mathbb{C}$-vectorspace, the words form a basis of $\mathbb{C}\langle \underline{z} \rangle$:

$$\mathbb{C}\langle \underline{z} \rangle = \left\{ \sum_{\omega} c_\omega \cdot \omega \mid c_\omega \in \mathbb{C} \right\}$$

where all sums are finite. By $\mathbb{C}\langle \underline{z} \rangle_d$ we denote the finite-dimensional subspace spanned by words of length at most d.
Multiplication of words extends uniquely to an associative and distributive multiplication on $\mathbb{C}\langle \underline{z} \rangle$, making it an algebra which is non-commutative for $n \geqslant 2$. We finally consider the involution that is uniquely defined on $\mathbb{C}\langle \underline{z} \rangle$ by the conditions

$$z_i^* = z_i \text{ and } \lambda^* = \overline{\lambda}$$

for $\lambda \in \mathbb{C}$. Recall that involutions interchange the order of products, so it does not just act as conjugation on the coefficients here! We for example have

$$(iz_1z_2)^* = -iz_2z_1.$$

This also implies that $\mathbb{C}\langle \underline{z} \rangle_h$ is *not* equal to $\mathbb{R}\langle \underline{z} \rangle$, in contrast to the commutative polynomial algebra. For example, z_1z_2 is not Hermitian, whereas $iz_1z_2 - iz_2z_1$ in fact is. But note that we still have

$$\sum \mathbb{C}\langle \underline{z} \rangle^2 \cap \mathbb{R}\langle \underline{z} \rangle = \sum \mathbb{R}\langle \underline{z} \rangle^2,$$

which can easily be seen by splitting polynomials into real and imaginary parts with respect to its coefficients.

Proposition 7.3.1. *In the real vectorspace $\mathbb{C}\langle \underline{z} \rangle_h$ we have*

$$\sum \mathbb{C}\langle \underline{z} \rangle^2 = \left(\sum \mathbb{C}\langle \underline{z} \rangle^2 \right)^{\vee\vee}.$$

Proof. This is proven similar to the commutative case in Section 6.2, see Exercise 65. □

Remark 7.3.2. Since the variables z_i fulfill no relatios besides $z_i^* = z_i$, the $*$-representations of $\mathbb{C}\langle \underline{z} \rangle$ are easy to describe. For any n-tuple of self-adjoint operators $T_1, \ldots, T_n \in \mathcal{L}(\mathcal{D})$ we obtain a $*$-representation

$$\begin{aligned} \pi \colon \mathbb{C}\langle z_1, \ldots, z_n \rangle &\to \mathcal{L}(\mathcal{D}) \\ p &\mapsto p(T_1, \ldots, T_n), \end{aligned}$$

and each $*$-representation is of this form. So Theorem 7.2.10 can be stated as follows here: a Hermitian polynomial $p \in \mathbb{C}\langle \underline{z} \rangle_h$ is a sum of squares if and only if

$$p(T_1, \ldots, T_n) \succcurlyeq 0$$

holds for all choices of (bounded and unbounded) self-adjoint operators T_i. Although the sums of squares are *not* Archimedean here, there is a very surprising strengthening of this result, which is not true in the commutative setup (cp. Example 7.2.12). △

Theorem 7.3.3 (Helton's Theorem)**.** *If $p \in \mathbb{C}\langle \underline{z} \rangle_h$ fulfills*

$$p(M_1, \ldots, M_n) \succcurlyeq 0$$

for all tuples of $M_i \in \mathrm{Her}_d(\mathbb{C})$ and all $d \geqslant 1$, then

$$p \in \sum \mathbb{C}\langle \underline{z} \rangle^2.$$

In other words, p is a sum of Hermitian squares if and only if it is nonnegative on the set of all finite-dimensional $*$-representations.

Proof. By Theorem 7.2.10 and Proposition 7.3.1 it is enough to show that nonnegativity of p on all finite-dimensional $*$-representations already implies nonnegativity on *all* $*$-representations.
So let $\mathcal{D}$ be an inner product space and $T_1, \ldots, T_n \in \mathcal{L}(\mathcal{D})$ self-adjoint operators. We have to show

$$\langle p(T_1, \ldots, T_n)v, v \rangle \geqslant 0$$

for all $v \in \mathcal{D}$. So fix such $v \in \mathcal{D}$, choose $d \in \mathbb{N}$ with $p \in \mathbb{C}\langle \underline{z} \rangle_d$, i.e. only words of length at most d appear in p. We now consider

$$\mathcal{H} := \{ q(T_1, \ldots, T_n)v \mid q \in \mathbb{C}\langle \underline{z} \rangle_d \},$$

which is clearly a finite-dimensional subspace of $\mathcal{D}$ containing v. Let $P \in \mathcal{L}(\mathcal{D})$ be the orthogonal projection to $\mathcal{H}$, i.e. if $u_1, \dots, u_r$ is an orthonormal basis of $\mathcal{H}$, then

$$P(w) = \sum_{j=1}^{r} \langle w, u_j \rangle \cdot u_j$$

for all $w \in \mathcal{D}$. We now set

$$M_i := P \circ T_{i|_{\mathcal{H}}} \in \mathcal{L}(\mathcal{H}) = \mathcal{B}(\mathcal{H}) \quad \text{for } i = 1, \dots, n.$$

Now it is easily checked that the M_i are self-adjoint on $\mathcal{H}$ (using that P is self-adjoint). So we obtain a new and in fact *finite-dimensional* $*$-representation

$$\begin{aligned} \pi \colon \mathbb{C}\langle \underline{z} \rangle &\to \mathcal{B}(\mathcal{H}) \\ q &\mapsto q(M_1, \dots, M_n). \end{aligned}$$

By definition of $\mathcal{H}$, for every product $M_{i_1} \cdots M_{i_k}$ with $k \leqslant d$ we have

$$(M_{i_1} \circ \cdots \circ M_{i_k})v = (T_{i_1} \circ \cdots \circ T_{i_k})v.$$

In fact, as long as at most d of the T_i are applied to v consequtively, the result always lies in $\mathcal{H}$, and application of P in between does not change anything! Since all words in p are of length at most d, we thus have

$$p(M_1, \dots, M_n)v = p(T_1, \dots, T_n)v$$

and so also

$$\langle p(T_1, \dots, T_n)v, v \rangle = \langle p(M_1, \dots, M_n)v, v \rangle.$$

Now since p is nonnegative at all finite-dimensional $*$-representations, the expression on the right ist nonnegative. But this is what we wanted to show. □

Remark 7.3.4. (i) The dimension of the matrices M_i in Theorem 7.3.3 can even be bounded. The space $\mathcal{H}$ that we have constructed in the proof is of dimension at most

$$\dim \mathbb{C}\langle \underline{z} \rangle_d = \frac{n^{d+1} - 1}{n - 1}$$

if $p \in \mathbb{C}\langle \underline{z} \rangle_d$. So we have to check nonnegativity of p only on matrices of (at most) this size. This bound depends only on the degree of p and the number of variables.

(ii) Let $m \in \mathbb{C}\langle z_1, z_2\rangle_h$ be a non-commutative Hermitian version of the Motzkin polynomial (meaning we get the Motzkin polynomial when we let the variables commute). We could for example take

$$m = \frac{1}{2}z_1^4 z_2^2 + \frac{1}{2}z_2^2 z_1^4 + \frac{1}{2}z_1^2 z_2^4 + \frac{1}{2}z_2^4 z_1^2 - \frac{3}{2}z_1^2 z_2^2 - \frac{3}{2}z_2^2 z_1^2 + 1.$$

Then there exists some $2 \leqslant d \leqslant 127$ and $M_1, M_2 \in \mathrm{Her}_d(\mathbb{C})$ such that

$$m(M_1, M_2) \in \mathrm{Her}_d(\mathbb{C})$$

is *not* positive semidefinite! Otherwise m would be a sum of Hermitian squares by Theorem 7.3.3, and if we then let all variables commute, this would yield a sums of squares decomposition of the commutative Motzkin polynomial. △

7.4 Group Algebras

Let Γ be a group with identity element e (we will use multiplicative notation for groups). We take the elements of Γ as a basis for a complex vectorspace, denoted $\mathbb{C}\Gamma$:

$$\mathbb{C}\Gamma = \left\{ \sum_{g\in\Gamma} c_g \cdot g \mid c_g \in \mathbb{C}, \text{ only finitely many } c_g \neq 0 \right\}.$$

Now multiplication of Γ yields a multiplication of the basis vectors of $\mathbb{C}\Gamma$, and thus an associative and distributive multiplication on $\mathbb{C}\Gamma$:

$$\left(\sum_g c_g \cdot g\right) \cdot \left(\sum_g c'_g \cdot g\right) = \sum_g \left(\sum_{f\cdot h = g} c_f c'_h\right) \cdot g.$$

In this way $\mathbb{C}\Gamma$ becomes an algebra, called the (complex) **group algebra** of Γ. It is commutative if and only if Γ is. The identity element is $1 \cdot e$. We equip $\mathbb{C}\Gamma$ with the following involution:

$$\left(\sum_g c_g \cdot g\right)^* = \sum_g \overline{c_g} \cdot g^{-1} = \sum_g \overline{c_{g^{-1}}} \cdot g.$$

So an element $\sum_g c_g \cdot g$ is Hermitian if and only if

$$\overline{c_g} = c_{g^{-1}}$$

holds for all $g \in \Gamma$.

Example 7.4.1. (i) Consider $\Gamma = \mathbb{Z}$ with the usual addition. Then $\mathbb{C}\mathbb{Z}$ is commutative, and in fact isomorphic to the algebra

$$\mathbb{C}\mathbb{Z} \cong \mathbb{C}[t, t^{-1}]$$

of Laurent polynomials in one variable. Under this isomorphism, the basis vector $m \in \mathbb{Z}$ of $\mathbb{C}\mathbb{Z}$ corresponds to t^m. The corresponding involution on Laurent polynomials thus fulfills $(t^i)^* = t^{-i}$. So the Hermitian elements are *not* the real Laurent polynomials, which we would prefer however. Now note that we can also generate the algebra $\mathbb{C}[t, t^{-1}]$ by the Hermitian elements

$$x := \frac{t + t^{-1}}{2} \text{ and } y := \frac{t - t^{-1}}{2i}.$$

These fulfill the relation

$$x^2 + y^2 = 1,$$

and so there is a $*$-algebra isomorphism to the algebra

$$\mathbb{C}[x, y]/(x^2 + y^2 - 1) = \mathbb{C}[S^1]$$

of polynomial functions on the unit circle, now with involution $x^* = x, y^* = y$. So a Hermitian element is really just a polynomial with real coefficients. More general, $\mathbb{C}\mathbb{Z}^n$ is isomorphic to the algebra

$$\mathbb{C}[\underbrace{S^1 \times \cdots \times S^1}_{n}]$$

of polynomial functions on the n-dimensional torus, with canonical involution.
(ii) If $\Gamma = S_3$ is the permutation group of 3 elements, we obtain a 6-dimensional algebra $\mathbb{C}S_3$ which is not commutative.
(iii) By $\Gamma = F_n$ we denote the **free group** with n generators (which we usually call $z_1, \ldots, z_n$). An element in F_n is thus a word in the letters

$$z_1, \ldots, z_n \text{ and } z_1^{-1}, \ldots, z_n^{-1}.$$

The group operation is concatenation of words. The only valid relations are

$$z_i^{-1} z_i = z_i z_i^{-1} = e$$

for all i, where e is the empty word. The group algebra $\mathbb{C}F_n$ thus contains the free algebra $\mathbb{C}\langle z_1, \ldots, z_n\rangle$ as a subalgebra, but *not* as a $*$-subalgebra! In $\mathbb{C}\langle \underline{z}\rangle$ we have ja $z_i^* = z_i$, whereas in $\mathbb{C}F_n$ we have $z_i^* = z_i^{-1}$. △

Remark 7.4.2. Let us determine all $*$-representations of a group algebra $\mathbb{C}\Gamma$. For each $*$-representation

$$\pi\colon \mathbb{C}\Gamma \to \mathcal{L}(\mathcal{D})$$

we obtain a group homomorphism

$$\begin{aligned}\pi\colon \Gamma &\to \mathcal{U}(\mathcal{D})\\ g &\mapsto \pi(g).\end{aligned}$$

into the group of unitary operators on $\mathcal{D}$. Here we call $T \in \mathcal{L}(\mathcal{D})$ **unitary**, if there exist some $S \in \mathcal{L}(\mathcal{D})$ with

$$\langle Tv, w\rangle = \langle v, Sw\rangle \ \forall v, w \in \mathcal{D} \text{ and } TS = ST = \mathrm{id}_{\mathcal{D}}.$$

We use that $g^* = g^{-1}$ holds in $\mathbb{C}\Gamma$ for this. This already implies that each$*$-representation of $\mathbb{C}\Gamma$ is in fact a bounded representation, since unitary operators are bounded of norm 1. Conversely, every group homomorphism

$$\pi\colon \Gamma \to \mathcal{U}(\mathcal{D})$$

provides a $*$-representation of $\mathbb{C}\Gamma$, by the rule

$$\sum_g c_g \cdot g \mapsto \sum_g c_g \cdot \pi(g).$$

So the $*$-representations of $\mathbb{C}\Gamma$ are in one-to-one correspondence with group homomorphisms from Γ into groups of unitary operators. △

Example 7.4.3. The group homomorphisms of F_n into an arbitrary group G are obtained by prescribing an arbitrary image $g_i \in G$ for each of the generators z_i of F_n. A word like for example $z_1 z_2^{-1} z_1 z_3$ is then mapped to $g_1 g_2^{-1} g_1 g_3$.
So the $*$-representations of $\mathbb{C}F_n$ are given by n-tuples of unitary operators $U_i \in \mathcal{U}(\mathcal{H})$ on a Hilbert space. △

Proposition 7.4.4. *For every group Γ, the sums of squares $\sum \mathbb{C}\Gamma^2$ are Archimedean.*

Proof. For $a = \sum_g c_g \cdot g \in \mathbb{C}\Gamma$ we set

$$\|a\|_1 = \sum_g |c_g|.$$

Now one immediately checks the following identity:

$$\|a\|_1^2 - a^*a = \frac{1}{2}\sum_{g,h\in\Gamma} |c_g c_h|(1 - \frac{c_g\overline{c}_h}{|c_g c_h|}h^{-1}g)^*(1 - \frac{c_g\overline{c}_h}{|c_g c_h|}h^{-1}g).$$

This shows $\|a\|_1^2 - a^*a \in \sum \mathbb{C}\Gamma^2$. For an Hermitian element $a \in \mathbb{C}\Gamma_h$ and $r = \|a\|_1$ we thus have

$$r - a = \frac{1}{2r}\left((r-a)^*(r-a) + (r^2 - a^*a)\right) \in \sum \mathbb{C}\Gamma^2. \qquad \square$$

The next result in fact follows from Schmüdgen's Theorem 4.2.3:

Theorem 7.4.5. *Let Γ be a finitely generated Abelian group, and let $\mathcal{F}$ be the set of all one-dimensional $*$-representations of $\mathbb{C}\Gamma$. Then for $a \in \mathbb{C}\Gamma_h$ we have*

$$a \text{ nonnegative on } \mathcal{F} \;\Leftrightarrow\; a + \varepsilon \in \sum \mathbb{C}\Gamma^2 \quad \forall \varepsilon > 0.$$

Proof. Exercise 66, similar to Example 7.4.1 (i). $\square$

For the group algebra of the free group, a similar result is true, resembling Helton's Theorem 7.3.3:

Theorem 7.4.6. *Let $\Gamma = F_n$ be the free group and $\mathcal{F}$ the set of all finite-dimensional $*$-representations of $\mathbb{C}\Gamma$. Then for $a \in \mathbb{C}\Gamma_h$ we have*

$$a \text{ nonnegative on } \mathcal{F} \;\Leftrightarrow\; a + \varepsilon \in \sum \mathbb{C}\Gamma^2 \quad \forall \varepsilon > 0.$$

The nonnegativity on the left just means that any replacement of the z_i in a by unitary matrices results in a positive semidefinite matrix.

Proof. By Theorem 7.2.14 and Proposition 7.4.4 it suffices to show that nonnegativity on $\mathcal{F}$ implies nonnegativity at each (bounded) $*$-representation. So let

$$\pi : \mathbb{C}\Gamma \to \mathcal{B}(\mathcal{H})$$

be a $*$-representation and fix $v \in \mathcal{H}$. We have to show

$$\langle \pi(a)v, v\rangle \geqslant 0.$$

Let $d \in \mathbb{N}$ be such that $a \in \mathbb{C}\Gamma_d$, i.e. a is a linear combination of words in the z_i and z_i^{-1} of length at most d. We consider the finite-dimensional subspace

$$\mathcal{H}' := \{\pi(b)v \mid b \in \mathbb{C}\Gamma_d\}$$

of $\mathcal{H}$ and the orthogonal projection $P\colon \mathcal{H} \to \mathcal{H}$ onto $\mathcal{H}'$. With

$$T_i := \pi(z_i) \in \mathcal{U}(\mathcal{H})$$

we set

$$M_i := P \circ T_{i|_{\mathcal{H}'}} \in \mathcal{B}(\mathcal{H}').$$

We then have $M_i^* = P \circ T_i^* = P \circ T_i^{-1}$ on $\mathcal{H}'$. For a product $M_{i_1}^{\delta_1} \circ \cdots \circ M_{i_k}^{\delta_k}$ with $\delta_i \in \{1, *\}$ and $k \leqslant d$ we thus have, just as in the proof of Theorem 7.3.3:

$$(M_{i_1}^{\delta_1} \circ \cdots \circ M_{ik}^{\delta_k})v = (T_{i_1}^{\delta_1} \circ \ldots \circ T_{i_k}^{\delta_k})v.$$

But now the M_i will not be unitary in general, and thus not provide a new $*$-representation of $\mathbb{C}\Gamma$. But the M_i are *contractions*, i.e. both

$$\mathrm{id}_{\mathcal{H}'} - M_i^* M_i \text{ and } \mathrm{id}_{\mathcal{H}'} - M_i M_i^*$$

are positive semidefinite on $\mathcal{H}'$. This is a direct computation, using that the T_i are unitary and thus norm preserving. Now the following construction is known as *Choi's matrix trick*. We set

$$\widetilde{M_i} := \begin{pmatrix} M_i & \sqrt{\mathrm{id}_{\mathcal{H}'} - M_i M_i^*} \\ \sqrt{\mathrm{id}_{\mathcal{H}'} - M_i^* M_i} & -M_i^* \end{pmatrix} \in \mathcal{B}(\mathcal{H}' \oplus \mathcal{H}').$$

The $\widetilde{M_i}$ are now in fact unitary (Exercise 67), and for a product $\widetilde{M}_{i_1}^{\delta_1} \circ \cdots \circ \widetilde{M}_{i_k}^{\delta_k}$ as above, and $\tilde{v} := (v, 0) \in \mathcal{H}' \oplus \mathcal{H}'$ we have

$$\langle (\widetilde{M}_{i_1}^{\delta_1} \circ \cdots \circ \widetilde{M}_{i_k}^{\delta_k})\tilde{v}, \tilde{v}\rangle = \langle (M_{i_1}^{\delta_1} \circ \cdots \circ M_{i_k}^{\delta_k})v, v\rangle = \langle (T_{i_1}^{\delta_1} \circ \cdots \circ T_{i_k}^{\delta_k})v, v\rangle$$

(Exercise 67). Since the $\widetilde{M_i}$ as unitary matrices provide a finite-dimensional $*$-representation $\tilde{\pi}$ of $\mathbb{C}\Gamma$, and since $a \in \mathbb{C}\Gamma_d$, this implies

$$0 \leqslant \langle \tilde{\pi}(a)\tilde{v}, \tilde{v}\rangle = \langle \pi(a)v, v\rangle.$$

This finishes the proof. □

Remark 7.4.7. (i) In Theorem 7.4.6 one can even strengthen the statement

$$a + \varepsilon \in \sum \mathbb{C}\Gamma^2 \quad \forall \varepsilon > 0$$

to

$$a \in \sum \mathbb{C}\Gamma^2.$$

But this requires a technical refinement of the proof.
(*ii*) For $\Gamma = \mathbb{Z}^n$ with $n \geqslant 3$ we will *not* get the same strengthening $a \in \sum \mathbb{C}\Gamma^2$ in Theorem 7.4.5! In the commutative setup, there do not exist saturated preorderings in dimension $\geqslant 3$, by Theorem 6.3.4. We have proven this only for sets with nonempty interior in affine space $\mathbb{R}^n$, but it also holds for general real varieties. △

7.5 Matrix Polnomials

Let $\mathcal{A} = \mathrm{M}_d(\mathbb{C}[x_1, \ldots, x_n])$ be the algebra of $d \times d$-matrices with *polynomial* entries (but *commutative* polynomials!). Addition and multiplication are defined in the obvious way, the involution as follows:

$$(p_{ij})^* := \left(p_{ji}^*\right)$$

where $(\sum_\alpha c_\alpha \underline{x}^\alpha)^* = \sum_\alpha \overline{c}_\alpha \underline{x}^\alpha$ is the canonical involution on the polynomial ring $\mathbb{C}[\underline{x}] = \mathbb{C}[x_1, \ldots, x_n]$. Elements of $\mathcal{A}$ are called **matrix polynomials** or **polynomial matrices**. Every point $a \in \mathbb{R}^n$ yields a d-dimensional $*$-representation

$$\begin{aligned} \pi_a \colon \mathrm{M}_d(\mathbb{C}[\underline{x}]) &\to \mathrm{M}_d(\mathbb{C}) \\ M = (p_{ij}) &\mapsto M(a) = (p_{ij}(a)). \end{aligned}$$

Let $\mathcal{F} = \{\pi_a \mid a \in \mathbb{R}^n\}$. So a Hermitian matrix polynomial is nonnegative on $\mathcal{F}$, if any only if it is positive semidefinite pointwise on $\mathbb{R}^n$.
The following looks like a result of the non-commutative theory, but it in fact admits a completely commutative proof. The case of $d = 1$ is exactly Hilbert's 17th Problem.

Theorem 7.5.1 (Gondard & Ribenboim). *For $\mathcal{A} = \mathrm{M}_d(\mathbb{C}[x_1, \ldots, x_n])$ and $M \in \mathcal{A}_h$ we have*

$$M(a) \succcurlyeq 0 \text{ for all } a \in \mathbb{R}^n \;\Leftrightarrow\; p^*p \cdot M \in \sum \mathcal{A}^2 \text{ for some } 0 \neq p \in \mathbb{C}[x_1, \ldots, x_n].$$

In other words, M is nonnegative on $\mathcal{F} = \{\pi_a \mid a \in \mathbb{R}^n\}$ if and only if it is a sum of squares with scalar denominator in $\mathcal{A}$.

Proof. "⇐": For $a \in \mathbb{R}^n$ we have

$$0 \preccurlyeq \pi_a(p^*p \cdot M) = |p(a)|^2 \cdot M(a).$$

For $p(a) \neq 0$ this already implies $M(a) \succcurlyeq 0$, and a denseness argument implies the same for all $a \in \mathbb{R}^n$.
"⇒" We first assume $M \in \mathrm{M}_d(\mathbb{R}[\underline{x}])$, i.e. M is real symmetric. Over the field $\mathbb{R}(\underline{x})$ we can thus diagonalize M as follows:

$$M = P^t \cdot \mathrm{diag}(f_1, \ldots, f_d) \cdot P \tag{7.1}$$

for some $P \in \mathrm{GL}_d(\mathbb{R}(\underline{x}))$ and $f_i \in \mathbb{R}(\underline{x})$. From nonnegativity of M on $\mathbb{R}^n$ it follows that the $f_i \in \mathbb{R}(\underline{x})$ are nonnegative wherever they are defined. Theorem 2.1.1 implies $f_i \in \sum \mathbb{R}(\underline{x})^2$, as is easily checked. By multiplying (7.1) with some suitable p^2, we can clear all denominators. This proves the claim.
The general case $M \in \mathrm{M}_d(\mathbb{C}[\underline{x}])_h$ can be reduced to the real case. Write $M = M_1 + iM_2$ with $M_1, M_2 \in \mathrm{M}_d(\mathbb{R}[\underline{x}])$. Then

$$\widetilde{M} = \begin{pmatrix} M_1 & M_2 \\ -M_2 & M_1 \end{pmatrix} \in \mathrm{M}_{2d}(\mathbb{R}[\underline{x}])_h$$

is real symmetric, and again pointwise nonnegative. The statement for M now follows from the already proven result for $\widetilde{M}$ (Exercise 68). □

The End

(now solve Exercise 69)

Bibliography

[1] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry*, vol. 10 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, second edn., 2006.

[2] J. Bochnak, M. Coste, and M.-F. Roy. *Real algebraic geometry*, vol. 36 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)*. Springer-Verlag, Berlin, 1998.

[3] M. Knebusch and C. Scheiderer. *Einführung in die reelle Algebra*, vol. 63 of *Vieweg Studium: Aufbaukurs Mathematik [Vieweg Studies: Mathematics Course]*. Friedr. Vieweg & Sohn, Braunschweig, 1989.

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

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

[6] C. Scheiderer. Positivity and sums of squares: a guide to recent results. In *Emerging applications of algebraic geometry*, vol. 149 of *IMA Vol. Math. Appl.*, pp. 271–324. Springer, New York, 2009.

[7] K. Schmüdgen. Noncommutative real algebraic geometry—some basic concepts and first ideas. In *Emerging applications of algebraic geometry*, vol. 149 of *IMA Vol. Math. Appl.*, pp. 325–350. Springer, New York, 2009.

Exercises

Exercise 1. (i) Show that for $a \in \mathbb{R}$ the orderings $\leqslant_{a_-}$ und $\leqslant_{a_+}$ on $\mathbb{Q}(t)$ from Example 1.1.4 are different if and only if a is algebraic over $\mathbb{Q}$ ist.
(ii) Show that the orderings $\leqslant_{a_-}$ and $\leqslant_{a_+}$ respectively, are Archimedean on $\mathbb{Q}(t)$ if and only if a is transcendental over $\mathbb{Q}$.

Exercise 2. Show that the mapping used in the proof of Theorem 1.1.8 is an order-preserving ring homomorphism.

Exercise 3. Prove Theorem 1.1.12.

Exercise 4. Show that the following sets from Example 1.1.13 are orderings on $\mathbb{R}(t)$:

$$P_1 = \left\{ f/g \mid fg = \sum_{i=k}^{d} a_i t^i, a_d > 0 \right\} \cup \{0\}$$

$$P_2 = \left\{ f/g \mid fg = \sum_{i=k}^{d} a_i t^i, a_k > 0 \right\} \cup \{0\}$$

To which of the already constructed binary ordering relations do they correspond?

Exercise 5. Determine all orderings of $\mathbb{Q}(\sqrt{2})$.

Exercise 6. Prove Corollary 1.2.10.

Exercise 7. Prove Rolle's Theorem 1.2.12 for polynomials over arbitrary real closed fields.

Exercise 8. Let K be a field and $v_1, \ldots, v_d$ linearly independent (column) vectors from K^n. Show the following:
(i) The matrix $M = \sum_{i=1}^{d} v_i v_i^t$ has rank d.
(ii) The matrix $N = \sum_{i=1}^{s} v_i v_i^t - \sum_{i=s+1}^{d} v_i v_i^t$ has signature $s - (d - s)$, with respect to every ordering of K.

Exercise 9. We call an extension $(K, \leqslant) \subseteq (L, \leqslant)$ of ordered fields **Archimedean**, if for all $b \in L$ there exists some $a \in K$ with $b \leqslant a$.
Show that if L/K is algebraic, then the extension is automatically Archimedean.

Exercise 10. Let $(K, \leqslant)$ be an ordered field and $A \subseteq K$ a subring. Show the following: The set

$$\mathcal{O} := \{b \in K \mid \pm b \leqslant a \text{ for some } a \in A\}$$

is again a subring of K, with the following properties:

- $A \subseteq \mathcal{O}$
- $b_1 \leqslant c \leqslant b_2$ with $b_1, b_2 \in \mathcal{O} \Rightarrow c \in \mathcal{O}$ (**order convexity**)
- $c \in K \setminus \mathcal{O} \Rightarrow c^{-1} \in \mathcal{O}$ ($\mathcal{O}$ is a **valuation ring** of K).

Exercise 11. Let $\mathcal{O} \subseteq K$ be a valuation ring of the field K (i.e. $a \in K \setminus \mathcal{O} \Rightarrow a^{-1} \in \mathcal{O}$). Show that $\mathcal{O}$ has exactly one maximal ideal, namely

$$\mathfrak{m} = \left\{a \in \mathcal{O} \mid a^{-1} \notin \mathcal{O}\right\} = \mathcal{O} \setminus \mathcal{O}^{\times}.$$

Exercise 12. Let $(K, \leqslant) = (K, P)$ be an ordered field and A a subring of K. Let $\mathcal{O}$ be the convex valuation ring from Exercise 10, $\mathfrak{m}$ its maximal ideal (Exercise 11), and $\pi\colon \mathcal{O} \to \mathcal{O}/\mathfrak{m}$ the projection map. Show that

$$\pi(P \cap \mathcal{O})$$

is an ordering of the residue field $\mathcal{O}/\mathfrak{m}$. For $A = \mathbb{Z}$ it is archimedean.

Exercise 13. Show that

$$\det\left(tI_d - C(p)\right) = p$$

holds for all (monic) polynomials $p \in K[t]$.

Exercise 14. Compute the Hermite Matrix (see Definition 1.3.2) of the general polynomial $p = t^3 + at^2 + bt + c \in \mathbb{R}[t]$. Find conditions on the coefficients a, b, c, stating that p has at least two real zeros.

Exercise 15. Let $(K, \leqslant)$ be an ordered field, and $0 \neq p \in K[t]$ a polynomial whose roots all lie in K. Show the following:
The roots of p are all $\geqslant 0$ if and only if the coefficients of p have alternating signs, i.e. $p = \sum_{i=0}^{d} a_i t^i$ with $(-1)^i a_i \geqslant 0$.

Exercise 16. Find an explicit ordering $\leqslant$ of $\mathbb{R}(x, y)$, such that for polynomials $p \in \mathbb{R}[x, y]$ we have

$$0 < p(0, 0) \Rightarrow 0 \leqslant p.$$

Exercise 17. Show that the sets from Example 1.5.4 are not semialgebraic.

Exercise 18. (i) Formulate the statement of the Intermediate Value Theorem for polynomials of fixed degree d as a formal $\mathbb{Z}$-statement.
(ii) Formulate the statement, that every polynomial (of fixed degree d) takes a maximum on each interval $[a, b]$ as a formal $\mathbb{Z}$-statement.
(iii) Can you formulate the statement that $\leqslant$ is Archimedean on $\mathbb{R}$ as a formal $\mathbb{R}$-statement?

Exercise 19. (i) Prove Corollary 1.5.11.
(ii) Show that closure and interior of semialgebraic subsets of $\mathbb{R}^n$ are semialgebraic.

Exercise 20. Let R be real closed. A function $f\colon R \to R$ is called **definable**, if its graph

$$\Gamma(f) = \{(\alpha, f(\alpha)) \mid \alpha \in R\} \subseteq R^2$$

is semialgebraic. Let f be such a definable function.
(i) Show that there exists a polynomial $0 \neq p \in R[x, y]$ that vanishes on $\Gamma(f)$.
(ii) Show that there exists some $q \in R[t]$ with $|f(\alpha)| \leqslant q(\alpha)$ for all large enough $\alpha \in R$.
(iii) Does (ii) also hold on the whole of R?

Exercise 21. Let $S = \delta(R) = \delta'(R) \subseteq R^n$ be a semialgebraic set, defined by the two formulas δ and δ' in the free variables $x_1, \ldots, x_n$. Show the following:
(i) For every real closed extension field R_1 of R we have

$$\delta(R_1) = \delta'(R_1).$$

We denote this set by S_{R_1}.
(ii) If $S \subseteq R^2$ is the graph of a function, then so is $S_{R_1} \subseteq R_1^2$.

Exercise 22. Let $f\colon R \to R$ be a definable function. By Exercise 21 (ii), for each real closed extension field S of R we can canonically extend f to $f_S\colon S \to S$. Show that injectivity/surjectivity transfer from f to f_S.

Exercise 23. Show that for polynomials $p, q \in K[x_1, \ldots, x_n]$ (over an arbitrary field K) we always have

$$\mathcal{N}(pq) = \mathcal{N}(p) + \mathcal{N}(q).$$

Exercise 24. Let $A = \mathbb{R}[[t]]$ be the ring of formal power series in one variable.
(i) Show that A is an integral domain.
(ii) Show that $p \in A$ is invertible if and only if $p = \sum_{i=0}^{\infty} p_i t^i$ with $p_0 \neq 0$. Conclude that A has exactly one maximal ideal (i.e. A is a **local ring**).
(iii) Show that $p \in A$ is a square in A if and only if $p = \sum_{i=k}^{\infty} p_i t^i$ with k even and $p_k > 0$.
(iv) Find all orderings of the ring A.

Exercise 25. Prove the statement from Example 3.1.5 (iv).

Exercise 26. Prove Proposition 3.1.12.

Exercise 27. Let X be a compact Hausdorff space. Find all maximal orderings of the ring $C(X, \mathbb{R})$ of continuous real-valued functions on X.

Exercise 28. Complete the proof of Theorem 3.1.23.

Exercise 29. Prove Corollary 3.1.24.

Exercise 30. Prove the abstract Nichtnegativstellensatz (Theorem 3.2.3) and the abstract Nullstellensatz (Theorem 3.2.4).

Exercise 31. Show that the ideal $(1 - x^2 - y^2) \subseteq R[x, y]$ is real.

Exercise 32. Let $T(p_1, \ldots, p_m) \subseteq R[t]$ denote the preordering generated by the polynomials $p_1, \ldots, p_m$. Show the following:
(i) $t \notin T(t^3)$.
(ii) If $p \geqslant 0$ on $[0, \infty)$, then $p \in T(t)$.
(iii) If $p \geqslant 0$ on $[-1, 1]$, then $p \in T(1 - t, 1 + t)$.
(iv) $T(1 - t, 1 + t) = T(1 - t^2)$.

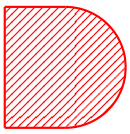
Exercise 33. Let $R[[x_1, \ldots, x_n]]$ be the formal power series ring in n variables. Show that for polynomials $p \in R[x_1, \ldots, x_n]$ we have

$$p \geqslant 0 \text{ in a neighborhood of zero} \Rightarrow \hat{p} \geqslant 0 \text{ on } \operatorname{Sper}(R[[x_1, \ldots, x_n]]).$$

Exercise 34. Let B be the unit disk in $\mathbb{R}^2$. Show that the set

$$([-1,0] \times [-1,1]) \cup B \subseteq \mathbb{R}^2$$

is not basic closed semialgebraic, i.e. not of the form $W_{\mathbb{R}}(p_1, \ldots, p_r)$ for certain $p_i \in \mathbb{R}[x,y]$.



Exercise 35. Let I be a nonempty set. A **filter** on I is a subset $\mathcal{F} \subseteq \mathcal{P}(I)$ of the power set of I, with the following properties:

$$\emptyset \notin \mathcal{F} \qquad A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F} \qquad A \in \mathcal{F}, A \subseteq B \Rightarrow B \in \mathcal{F}$$

An **ultrafilter** is filter $\mathcal{F}$ with the additional property

$$A \notin \mathcal{F} \Rightarrow I \setminus A \in \mathcal{F}.$$

Show the following:
(i) For each $i \in I$ the set $\mathcal{F}_i := \{A \subseteq I \mid i \in A\}$ is an ultrafilter (such ultrafilters are called **principal** ultrafilters).
(ii) If I is infinite, then $\{A \subseteq I \mid I \setminus A \text{ finite}\}$ is a filter (called the **filter of cofinite sets**).
(iii) Each filter is contained in some ultrafilter.
(iv) An ultrafilter is principal if and only if it does not contain the filter of cofinite sets.

Exercise 36. Let I be an infinite set and $\mathcal{F}$ an ultrafilter on I. Let $(K_i, \leqslant_i)$ be an ordered field, for each $i \in I$. Consider the commutative ring

$$R = \prod_{i \in I} K_i = \{(a_i)_{i \in I} \mid a_i \in K_i \text{ for all } i \in I\}.$$

Show the following:
(i) The set

$$\mathfrak{m} = \{(a_i)_i \mid \{i \mid a_i = 0\} \in \mathcal{F}\}$$

is a maximal ideal in R.

(ii) The relation

$$(a_i)_i \leqslant (b_i)_i :\Leftrightarrow \{i \mid a_i \leqslant_i b_i\} \in \mathcal{F}$$

induces a well-defined ordering on the field $R/\mathfrak{m}$.
(iii) If $\mathcal{F}$ is not principal and if I is countable, then the ordered field from (ii) is not Archimedean.

Exercise 37. Let I be a set and $\mathcal{F}$ an ultrafilter on I. Let $(K_i, \leqslant_i)$ be an ordered field, for each $i \in I$, and let $(K, \leqslant)$ be the ordered field constructed in Exercise 36 (ii). Let φ be formula over $\mathbb{Z}$ in the free variables $x_1, \ldots, x_n$, and let $a_1 = \overline{(a_{1i})}_{i\in I}, \ldots, a_n = \overline{(a_{ni})}_{i\in I}$ be elements from $K = \left(\prod_{i\in I} K_i\right)/\mathfrak{m}$.
Prove *Łos's Theorem*:

$$\begin{aligned}&\text{The statement } \varphi(a_1, \ldots, a_n) \text{ holds in } (K, \leqslant)\\ &\Leftrightarrow \{i \in I \mid \varphi(a_{1i}, \ldots, a_{ni}) \text{ holds in } (K_i, \leqslant_i)\} \in \mathcal{F}.\end{aligned}$$

Hint: Proceed by induction over the construction of the formula; considering $\exists$ instead of $\forall$ might be easier.

Exercise 38. Show that the property of being Archimedean of an ordered field cannot be formulated as a statement over $\mathbb{Z}$ (i.e. there is no statement over $\mathbb{Z}$ that holds in an ordered field if and only if it is Archimedean).

Hint: Use Exercises 36 and 37.

Exercise 39. Let $\mathcal{F}$ be a non-principal ultrafilter on $\mathbb{N}$, and for all $i \in \mathbb{N}$ let $R_i = R$ always be the same real closed field. Let

$$\widetilde{R} := \left(\prod_i R\right)/\mathfrak{m}$$

be the real closed extension field of R that we have constructed in Exercises 36 and 37 (why is $\widetilde{R}$ real closed? Why does it contain R?). Prove that $\widetilde{R}$ is $\aleph_1$**-saturated**, i.e.:

If a semialgebraic set in $\widetilde{R}^n$ is covered by countably many semialgebraic sets, then there exists a finite subcover.

Hint: If no finite subcover exists, then there is a sequence of elements a_i, such that each a_i belongs to the initial set but not to the i-th set from the covering. Then use a diagonal argument to construct an element from the initial set that does not belong to any of the covering sets.

Exercise 40. Prove *existence of degree bounds in Hilbert's 17th Problem*: For $n, d \in \mathbb{N}$ there exists some $d' \in \mathbb{N}$, such that for each real closed field R, and each globally nonnegative polynomial $p \in R[x_1, \ldots, x_n]$ of degree d, there exists a representation

$$q^2 p = p_1^2 + \cdots + p_m^2$$

with polynomials $q, p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$ of degree at most d'.

Hint: Apply Hilbert's 17th Problem in some $\aleph_1$-saturated field R; cp. Exercise 39.

Exercise 41. Let R be a non-Archimedean real closed extension field of $\mathbb{R}$, and let ε be an infinitesimal positive element from R. Show that

$$1 - t^2 + \varepsilon \notin T\left((1 - t^2)^3\right) \subseteq R[t].$$

Hint: Consider the order-convex hull $\mathcal{O}$ of $\mathbb{R}$ in R, and apply the factor map $\mathcal{O} \to \mathcal{O}/\mathfrak{m} \cong \mathbb{R}$.

Exercise 42. Show that in $\mathbb{R}[x, y]$ we have $xy \notin M(x, y)$.

Exercise 43. Find $p_1, \ldots, p_m \in \mathbb{R}[x_1, \ldots, x_n]$, such that $W_{\mathbb{R}}(p_1, \ldots, p_m) = \emptyset$ but

$$-1 \notin M(p_1, \ldots, p_m).$$

Exercise 44. Show that for diagonal matrices $M, M_1, \ldots, M_m$ we can omit the *strictly feasibly* condition in the Duality Theorem 5.1.4 (this is then known as the Duality Theorem for *linear optimization*).

Exercise 45. Prove *Farkas' Lemma* (cp. Example 5.4.5):

Let $\ell_1, \ldots, \ell_m \in \mathbb{R}[\underline{x}] = \mathbb{R}[x_1, \ldots, x_n]$ be polynomials of degree at most one, and assume the polyhedron

$$P = \{a \in \mathbb{R}^n \mid \ell_1(a) \geqslant 0, \ldots, \ell_m(a) \geqslant 0\}$$

that they define is not empty. Let $\ell \in \mathbb{R}[\underline{x}]$ be another polynomial of degree $\leqslant 1$. We then have

$$\ell \geqslant 0 \text{ on } P \;\Leftrightarrow\; \ell = r_0 + r_1\ell_1 + \cdots + r_m\ell_m \text{ for certain } r_i \in \mathbb{R}_{\geqslant 0}.$$

Hint: You can for example use the Duality Theorem with suitable diagonal matrices.

Exercise 46. Construct a semidefinite optimization problem for which strong duality $d^* = p^*$ fails.

Exercise 47. Construct a semidefinite optimization problem for which both primal and dual problems have feasible points, but in which the optimal values in both problems are not attained.

Exercise 48. Use a method of your choice (e.g. Lagrange's method) to compute minimum and maximum of the polynomial $x^2 + y + 1$ on the planar unit disk $W(1 - x^2 - y^2)$.

Exercise 49. Let $S \subseteq \mathbb{R}^n$ be a spectrahedral cone with nonempty interior. Show that there exist symmetric matrices $M_1, \ldots, M_n$ with $S = \mathcal{S}(M_1, \ldots, M_n)$, such that

$$\operatorname{int}(S) = \{a \in \mathbb{R}^n \mid a_1M_1 + \cdots + a_nM_n \succ 0\}\,.$$

Exercise 50. Let $h \in \mathbb{R}[x_1, \ldots, x_n]$ by hyperbolic in direction $e \in \mathbb{R}^n$. Show the following:
(i) The hyperbolicity cone $\Lambda_e(h)$ is a closed convex cone.
(ii) For $e' \in \operatorname{int}(\Lambda_e(h))$ the polynomial h is also hyperbolic in direction e'.
(iii) For e' as in (ii) we have $\Lambda_e(h) = \Lambda_{e'}(h)$.
Hint: You can use the Helton-Vinnikov Theorem 5.3.12, if you explain why.

Exercise 51. (i) The interior of a spectrahedral cone is a spectrahedral shadow.
(ii) The interior of a spectrahedral shadow is a spectrahedral shadow.

Exercise 52. Show the following:
(i) The dual cone of a spectrahedral cone S is a spectrahedral shadow.
Hint: W.l.o.g. assume that S has nonempty interior. You should now look closely at the proof of the Duality Theorem 5.1.4 again.
(ii) The dual cone of a spectrahedral shadow is a spectrahedral shadow.

Exercise 53. Show that stability of a quadratic module does not depend on the generators, i.e. if

$$M = M(p_1, \ldots, p_r) = M(q_1, \ldots, q_s)$$

and for each $d \in \mathbb{N}$ there exists some $d' \in \mathbb{N}$ with

$$M \cap \mathbb{R}[\underline{x}]_d \subseteq M_{d'}(p_1, \ldots, p_r),$$

then the same is true for $M(q_1, \ldots, q_s)$, maybe with a different d'.

Exercise 54. Do the limit process from Theorem 6.2.6 exactly.

Exercise 55. Prove the statement from Example 6.2.10 (i): If the set

$$W_{\mathbb{R}}(p_1, \ldots, p_r) \subseteq \mathbb{R}^2$$

contains a vertical and a horizontal strip, then $M(p_1, \ldots, p_r)$ is stable.

Exercise 56. Prove the statement from Example 6.2.10 (ii): $M((1 - x^2)y^2)$ is not stable.

Exercise 57. A finitely generated quadratic module $M = M(q_1, \ldots, q_r)$ in $\mathbb{R}[\underline{x}]$ is called **decidable**, if the set $M \cap \mathbb{R}[\underline{x}]_d$ is a semialgebraic subset of the finite-dimensional space $\mathbb{R}[\underline{x}]_d$, for all $d \geqslant 0$. Show the following:
(i) If M is stable, then M is decidable.
(ii) If M is saturated, then M is decidable.
(iii) Can you find an undecidable quadratic module?

Exercise 58. Let $a, b, c, d \in \mathbb{R}$ with $a \leqslant c \leqslant d \leqslant b$. Show that there exists some $\lambda \in [0, 1]$ for which the polynomial

$$(t - c)(t - d) - \lambda(t - a)(t - b)$$

is globally nonnegative.

Exercise 59. Complete the proof of Burnside's Theorem 7.1.2: If $\mathcal{A} \subseteq \mathrm{M}_d(\mathbb{C})$ is a subalgebra that acts transitively on $\mathbb{C}^d \setminus \{0\}$ and contains a matrix of rank 1, then $\mathcal{A} = \mathrm{M}_d(\mathbb{C})$.

Exercise 60. Let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be a state. Show that $\varphi(a^*) = \overline{\varphi(a)}$ holds for all $a \in \mathcal{A}$.

Exercise 61. Prove Lemma 7.2.6.

Exercise 62. Check all details from the GNS-construction in Section 7.2.

Exercise 63. Let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be a state with $\varphi(1) = 0$. Show that $\varphi \equiv 0$ holds.

Exercise 64. Let $\mathcal{D}$ be an inner product space, and assume $\pi\colon \mathcal{A} \to \mathcal{L}(\mathcal{D})$ is a $*$-representation for which $\pi(a)$ is a bounded operator on $\mathcal{D}$ for all $a \in \mathcal{A}$. Show that one can understand

$$\pi\colon \mathcal{A} \to \mathcal{B}(\mathcal{H})$$

as a bounded $*$-representation on the completion $\mathcal{H}$ of $\mathcal{D}$.

Exercise 65. Prove Proposition 7.3.1.

Exercise 66. Prove Theorem 7.4.5.

Exercise 67. Prove the statement from Choi's matrix trick in the proof of Theorem 7.4.6.

Exercise 68. Reduce the complex case to the real case in the proof of Theorem 7.5.1.

Exercise 69. Lean back and relax.