Chapter 4

Nonnegative Polynomials and Sums of Squares

Grigoriy Blekherman

A central question, for both practical and theoretical reasons, is how to efficiently test whether a polynomial p is nonnegative. We reformulate this problem in the following way: given a nonnegative polynomial p, how do we efficiently find a representation of p, so that nonnegativity of p is apparent from this representation? In other words, how do we efficiently represent p as an "obviously nonnegative" polynomial? Some polynomials are obviously nonnegative. If we can write p as a sum of squares of polynomials, then it is clear that p is nonnegative just from this presentation. Very importantly, if p is a sum of squares then its sums of squares representation can be efficiently computed via semidefinite programming. This connection was described in detail in Chapter 3. As we will see, the set of sums of squares is a *projected spectrahedron*, while the set of nonnegative polynomials is far more challenging computationally. The main question for this chapter is: what is the relationship between nonnegative polynomials and sums of squares?

4.1 Introduction

Our story begins in 1885, when twenty-three-year-old David Hilbert was one of the examiners in the Ph.D. defense of twenty-one-year-old Hermann Minkowski. During the examination Minkowski claimed that there exist nonnegative polynomials that are not sums of squares. Although he did not provide an example or a proof, his argument must have been convincing, as he defended successfully.

Three years later Hilbert published a paper in which he classified all of the (few) cases, in terms of degree and number of variables, in which nonnegative polynomials are the same as sums of squares. In all other cases Hilbert showed that there exist nonnegative polynomials that are not sums of squares. Interestingly, Hilbert did not provide an explicit example of such polynomials. The first explicit example was found only seventy years later and is due to Theodore Motzkin. In fact, Motzkin was not aware of what he constructed. Olga Taussky-Todd, who was present during the seminar in which Motzkin described his construction, later notified him that he found the first example of a nonnegative polynomial that is not a sum of squares [22].

We examine the relationship between nonnegativity and sums of squares in two different fundamental ways. We first consider the structures that prevent sums of squares from capturing all nonnegative polynomials, and show that equality occurs precisely when these structures are not present. We then examine in detail the smallest cases where there exist nonnegative polynomials that are not sums of squares and show that the inequalities separating nonnegative polynomials from sums of squares have a simple and elegant structure. Second, we look at the quantitative relationship between nonnegative polynomials and sums of squares. Here we show that when the degree is fixed and the number of variables grows, there are significantly more nonnegative polynomials than sums of squares. We also apply these ideas to studying the relationship between sums of squares and convex polynomials. While the techniques we develop for the two approaches are quite different in nature, the unifying theme is that we examine the sets of nonnegative polynomials and sums of squares geometrically. Algebraic geometry is at the forefront of our examination of fundamental differences between nonnegative polynomials and sums of squares, while convex geometry and analysis are used to examine the quantitative relationship.

The chapter is structured as follows: After discussing Hilbert's theorem and Motzkin's example in Section 4.2, we begin a detailed examination of the underlying causes of differences between nonnegative polynomials and sums of squares in Section 4.3. On the way we will see that nonnegative polynomials and sums of squares form fascinating convex sets. Section 4.4 is devoted to the examination of these objects from the point of view of convex algebraic geometry. We note that many basic questions remain open.

The fundamental reasons for the existence of nonnegative polynomials that are not sums of squares come from Cayley–Bacharach theory in classical algebraic geometry and, in fact, Hilbert's original proof of his theorem already used some of these ideas. We begin developing the necessary techniques in Section 4.5. Duality from convex geometry and its interplay with commutative algebra will play a central role in our investigation. Section 4.6 develops the duality ideas and presents a unified proof of the equality cases of Hilbert's theorem. Sections 4.7 and 4.8 investigate the smallest cases in which there exist nonnegative polynomials that are not sums of squares. We show that this situation fundamentally arises from the existence of Cayley–Bacharach relations and present some consequences.

We proceed by examining the quantitative relationship between nonnegative polynomials and sums of squares in Section 4.9. This is done by establishing bounds on the volume of sets of nonnegative polynomials and sums of squares, and analytic aspects of convex geometry come to the fore in this examination. We will explain that if the degree is fixed and the number of variables is allowed to grow, then there are significantly more nonnegative polynomials than sums of squares [5].

This happens despite the difficulty of constructing explicit examples of nonnegative polynomials that are not sums of squares, and numerical evidence that sums of squares approximate nonnegative polynomials well if the degree and number of variables is small [19]. The question of precisely when nonnegative polynomials begin to significantly overtake sums of squares is currently poorly understood.

Section 4.10 presents an application of the volume ideas to showing that there exist homogeneous polynomials that are convex functions but are not sums of squares. There is no known explicit example of such a polynomial, and this is the only known method of showing their existence.

4.2 A Deeper Look

We first reduce the study of nonnegative polynomials and sums of squares to the case of homogeneous polynomials, which are also called forms. A polynomial $p(x_1, \ldots, x_n)$ of degree d can be made homogeneous by introducing an extra variable x_{n+1} and multiplying every monomial in p by a power of x_{n+1}, so that all monomials have the same degree. More formally, let $\bar{p}$ be the homogenization of p:

$$\bar{p} = x_{n+1}^d p\left(\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}\right).$$

Exercise 4.1. Let p be a nonnegative polynomial. Show that $\bar{p}$ is a nonnegative form. Also show that if p is a sum of squares, then $\bar{p}$ is a sum of squares as well.

Given a form $\bar{p}$ we can dehomogenize it by setting $x_{n+1} = 1$. Dehomogenization clearly preserves nonnegativity and sums of squares. Therefore the study of nonnegative polynomials and sums of squares in n variables is equivalent to studying forms in $n+1$ variables. From now on we restrict ourselves to the case of forms.

Let $\mathbb{R}[x]_{\mathbf{d}}$ be the vector space of real forms in n variables of degree d. In order to be nonnegative a form must have even degree, and therefore our forms will have even degree $2d$. Inside $\mathbb{R}[x]_{\mathbf{2d}}$ sit two closed convex cones: the cone of nonnegative polynomials,

$$P_{\mathbf{n},\mathbf{2d}} = \{p \in \mathbb{R}[x]_{\mathbf{2d}} \mid p(x) \geq 0 \text{ for all } x \in \mathbb{R}^n\},$$

and the cone of sums of squares,

$$\Sigma_{\mathbf{n},\mathbf{2d}} = \left\{p \in \mathbb{R}[x]_{\mathbf{2d}} \mid p(x) = \sum q_i^2 \text{ for some } q_i \in \mathbb{R}[x]_{\mathbf{d}}\right\}.$$

Exercise 4.2. Show that $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ are closed, full-dimensional convex cones in $\mathbb{R}[x]_{\mathbf{2d}}$. (Hint: Consider Exercise 4.17.)

We now come to the first major theorem concerning nonnegative polynomials and sums of squares.

4.2.1 Hilbert's Theorem

The first fundamental result about the relationship between $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$ was shown by Hilbert in 1888.

Theorem 4.3. *Nonnegative forms are the same as sums of squares, $P_{\mathbf{n,2d}} = \Sigma_{\mathbf{n,2d}}$, in the following three cases: $n = 2$ (univariate nonhomogeneous case), $2d = 2$ (quadratic forms), and $n = 3$, $2d = 4$ (ternary quartics). In all other cases there exist nonnegative forms that are not sums of squares.*

The proof of the three equality cases in Hilbert's theorem usually proceeds by treating each of the three cases separately. For example, it is a simple exercise to show that $P_{\mathbf{n,2}} = \Sigma_{\mathbf{n,2}}$.

Exercise 4.4. Deduce that $P_{\mathbf{n,2}} = \Sigma_{\mathbf{n,2}}$ from diagonalization of symmetric matrices.

We adopt a different approach: We begin by examining the structures that allow the existence of nonnegative forms that are not sums of squares. In Section 4.6.1 we show that the three cases of Hilbert's theorem are the only cases in which these structures do not exist. This provides a unified proof of the three equality cases of Hilbert's theorem, which are usually treated separately.

4.2.2 Motzkin's Example

The first explicit example of a nonnegative form that is not a sum of squares is due to Motzkin:

$$M(x, y, z) = x^4y^2 + x^2y^4 + z^6 - 3x^2y^2z^2.$$

The form M can be seen to be nonnegative by the application of the arithmetic mean-geometric mean inequality. Why is M not a sum of squares?

In the following exercises we develop a general method for showing that a form is not a sum of squares, based on the monomials that occur in the form. This method can also be applied to reduce the size of the semidefinite program that computes the sum of squares decomposition, as explained in Chapter 3. These ideas are originally due to Choi, Lam, and Reznick [22].

Exercise 4.5. For a polynomial p define its *Newton polytope* $\mathcal{N}(p)$ to be the convex hull of the vectors of exponents of monomials that occur in p. For example, if $p = x_1x_2^2 + x_2^2 + x_1x_2x_3$, then $\mathcal{N}(p) = \operatorname{conv}(\{(1,2,0),(0,2,0),(1,1,1)\})$, which is a triangle in $\mathbb{R}^3$.

Show that if $p = \sum q_i^2$, then

$$\mathcal{N}(q_i) \subseteq \frac{1}{2}\mathcal{N}(p).$$

Exercise 4.6. Calculate the Newton polytope of the Motzkin form and use Exercise 4.5 to show that the Motzkin form is not a sum of squares.

For much more on explicit examples of nonnegative polynomials that are not sums of squares see [22].

4.2.3 Quantitative Relationship

While Hilbert's theorem completely settles all cases of equality between $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ it does not shed light on whether these cones are close to each other, even if the cone of nonnegative polynomials is strictly larger. Due to the difficulty of constructing explicit examples and numerical evidence for a small number of variables and degrees, it is tempting to assume that $\Sigma_{\mathbf{n},\mathbf{2d}}$ approximates $P_{\mathbf{n},\mathbf{2d}}$ fairly well.

However, it was shown in [5] that if the degree $2d$ is fixed and at least 4, then as the number of variables n grows, there are significantly more nonnegative forms than sums of squares. We will make this statement precise and present a proof in Section 4.9. The main idea is that, although the cones themselves are unbounded, we can slice both cones with the same hyperplane, so that the section of each cone is compact. We then derive separate bounds on the volume of each section.

For now we would like to note that the bounds guarantee that the difference between $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ is large only for a very large number of variables n. Whether this is an artifact of the techniques used to derive the bounds is unclear. As we will see, for a small number of variables the distinction between $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ is quite delicate, and it is not known at what point $P_{\mathbf{n},\mathbf{2d}}$ becomes much larger than $\Sigma_{\mathbf{n},\mathbf{2d}}$.

We now begin a systematic examination of differences between nonnegative forms and sums of squares. It is actually possible to see that there exist nonnegative forms that are not sums of squares by considering values of forms on finitely many points. The following example will illustrate this idea and explain some of the major themes in our investigation.

4.3 The Hypercube Example

According to Hilbert's theorem the smallest cases where $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ differ are forms in 3 variables of degree 6, and forms in 4 variables of degree 4. We take a close look at an explicit example for the case of forms in 4 variables of degree 4. Let $S = \{s_1, \ldots, s_8\}$ be the following set of 8 points in $\mathbb{R}^4$:

$$S = \{\pm 1, \pm 1, \pm 1, 1\}.$$

We will see that there is a difference between nonnegative forms and sums of squares by simply looking at the values that nonnegative polynomials and sums of squares take on S. Accordingly, let us define a projection π from $\mathbb{R}[x]_{\mathbf{4},\mathbf{4}}$ to $\mathbb{R}^8$ given by evaluation on S:

$$\pi(f) = (f(s_1), \ldots, f(s_8)) \quad \text{for} \quad f \in \mathbb{R}[x]_{\mathbf{4},\mathbf{4}}.$$

We will explicitly describe the images of $P_{\mathbf{4},\mathbf{4}}$ and $\Sigma_{\mathbf{4},\mathbf{4}}$ under this projection. Let

$$P' = \pi(P_{\mathbf{4},\mathbf{4}}) \quad \text{and} \quad \Sigma' = \pi(\Sigma_{\mathbf{4},\mathbf{4}}).$$

As they are images of convex cones under a linear map, it is clear that both P' and Σ' are convex cones in $\mathbb{R}^8$. Although both P' and Σ' will turn out to be closed, projections of closed convex cones do not have to be closed in general.

Exercise 4.7. Construct a closed convex cone C in $\mathbb{R}^3$ and a linear map $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ such that $\pi(C)$ is not closed.

4.3.1 Values of Nonnegative Forms

We first look at values on S that are achievable by nonnegative forms. Let $\mathbb{R}^8_+$ be the nonnegative orthant of $\mathbb{R}^8$:

$$\mathbb{R}^8_+ = \{(x_1, \dots, x_8) \quad | \quad x_i \geq 0 \text{ for } i = 1, \dots, 8\}.$$

Since we are evaluating nonnegative polynomials, it is clear that $P' \subseteq \mathbb{R}^8_+$. We claim that, in fact, $P' = \mathbb{R}^8_+$. In other words, any 8-tuple of nonnegative numbers can be attained on S by a globally nonnegative form. By convexity of P' it suffices to show that all the standard basis vectors e_i are in P'. Moreover, substitutions $x_i \mapsto -x_i$ permute the set S, and therefore it is enough to show that $e_i \in P'$ for some i.

Exercise 4.8. Let $p \in \mathbb{R}[x]_{\mathbf{4},\mathbf{4}}$ be the following symmetric form:

$$p = \sum_{i=1}^{4} x_i^4 + 2 \sum_{i \neq j \neq k} x_i^2 x_j x_k + 4x_1x_2x_3x_4.$$

Show that p is nonnegative, and check that p vanishes on exactly 7 points in S. Conclude that $P' = \mathbb{R}^8_+$.

We have seen that all combinations of nonnegative values on S are realizable as values of a nonnegative form. We now look at why some values in $\mathbb{R}^8_+$ are not attainable by sums of squares. In the end we will completely describe the projection Σ'.

4.3.2 Values of Sums of Squares

In order to analyze the values of sums of squares, we need to take a look at the values of the forms that we are squaring. The values of quadratic forms on S are not linearly independent. Here is the unique (up to a constant multiple) linear relation between the values on the points s_i that all quadratic forms in 4 variables satisfy:

$$\sum_{s_i \text{ has even number of 1's}} f(s_i) = \sum_{s_i \text{ has odd number of 1's}} f(s_i). \tag{4.1}$$

Exercise 4.9. Verify that the relation (4.1) holds for all quadratic forms $f \in \mathbb{R}[x]_{\mathbf{4},\mathbf{2}}$ and that it is unique up to a constant multiple.

We are now ready to see how the relation (4.1) prevents sums of squares from attaining all values in $\mathbb{R}^8_+$.

Proposition 4.10 (Hilbert's original insight). *Let e_i be the ith standard basis vector in $\mathbb{R}^8$. Then $e_i \notin \Sigma'$ for all i.*

Proof. Since we did not attach a specific labeling to the points of S it will suffice to show that $e_1 \notin \Sigma' = \pi(\Sigma_{\mathbf{4},\mathbf{4}})$. Suppose that there exists $p \in \Sigma_{\mathbf{4},\mathbf{4}}$ such that $\pi(p) = e_1$. Write $p = \sum_j q_j^2$ for some $q_j \in \mathbb{R}[x]_{\mathbf{4},\mathbf{2}}$. The form p vanishes on $s_2, \ldots, s_8$, and it has value 1 on s_1. Since $p = \sum_j q_j^2$ it follows that each q_j vanishes on $s_2, \ldots, s_8$. Each q_j is a quadratic form in 4 variables, and therefore each q_j satisfies relation (4.1). From this relation it follows that $q_j(s_1) = 0$ for all j. Therefore $p(s_1) = 0$, which is a contradiction. ∎

Hilbert's original proof did not use an explicit example to show that the vectors e_i can be realized as values of a nonnegative form, which we did in Exercise 4.8. Instead he provided a recipe for constructing such a form, and proved that the construction works. We largely followed Hilbert's recipe to construct our counterexample. For more information on Hilbert's construction see [23].

4.3.3 Complete Description of Σ'

We can do better than just describing some points that are not in Σ'. Our next goal is to completely describe Σ' and, in particular, we will see how far the points e_i are from being the values of a sum of squares.

We use π to also denote the same evaluation projection on quadratic forms in 4 variables:

$$\pi(f) = (f(s_1), \ldots, f(s_8)) \quad \text{for} \quad f \in \mathbb{R}[x]_{\mathbf{4},\mathbf{2}}.$$

Let L be the projection of the entire vector space of quadratic forms:

$$L = \pi(\mathbb{R}[x]_{\mathbf{4},\mathbf{2}}).$$

Using relation (4.1) and Exercise 4.9 we see that L is a hyperplane in $\mathbb{R}^8$. Let C be the set of points that are coordinatewise squares of points in L:

$$C = \{(v_1^2, \ldots, v_8^2) \ \mid \ v = (v_1, \ldots, v_8) \in L\}.$$

We first show the following description of Σ'.

Lemma 4.11. *Σ' is equal to the convex hull of C:*

$$\Sigma' = \operatorname{conv}(C).$$

Proof. Let $v = (v_1, \ldots, v_8) \in L$. Then there exists a quadratic form $f \in \mathbb{R}[x]_{\mathbf{4},\mathbf{2}}$ such that $f(s_i) = v_i$ for $i = 1, \ldots, 8$. It follows that for the square of f we have $f^2(s_i) = v_i^2$. In other words,

$$\pi(f^2) = (v_1^2, \ldots, v_8^2), \quad \text{where} \quad v = (v_1, \ldots, v_8) = \pi(f).$$

Therefore we see that $C \subset \Sigma'$ and by convexity of Σ' it follows that $\operatorname{conv}(C) \subseteq \Sigma'$.

To prove the other inclusion, suppose that $p = \sum_i q_i^2 \in \Sigma_{\mathbf{4,4}}$. Then $\pi(q_i^2) \in C$ for all i and therefore $\Sigma' \subseteq \text{conv}(C)$. ∎

Let T_m be the subset of the nonnegative orthant $\mathbb{R}_+^m$ defined by the following m inequalities:

$$T_m = \left\{ (x_1, \dots, x_m) \in \mathbb{R}_+^m \quad \middle| \quad \sum_{i=1}^m \sqrt{x_i} \geq 2\sqrt{x_k} \quad \text{for all} \quad k \right\}.$$

We will show that $\Sigma' = T_8$. We begin with a lemma on the structure on T_m.

Lemma 4.12. *The set T_m is a closed convex cone. Moreover, T_m is the convex hull of the points $x = (x_1, \dots, x_m) \in \mathbb{R}_+^m$, where $\sum_{i=1}^m \sqrt{x_i} = 2\sqrt{x_k}$ for some k.*

Proof. The set T_m is defined as a subset of $\mathbb{R}^m$ by the following $2m$ inequalities: $x_k \geq 0$ and $\sqrt{x_1} + \cdots + \sqrt{x_m} \geq 2\sqrt{x_k}$ for all k. Therefore it is clear that T_m is a closed set.

For $x = (x_1, \dots, x_m) \in \mathbb{R}_+^m$ let $\|x\|_{1/2}$ denote the $L^{1/2}$-norm of x:

$$\|x\|_{1/2} = (\sqrt{x_1} + \cdots + \sqrt{x_m})^2.$$

We can restate the inequalities of T_m as $x_k \geq 0$ and $\|x\|_{1/2} \geq 4x_k$ for all k. Now suppose that $x, y \in T_m$ and let $z = \lambda x + (1-\lambda)y$ for some $0 \leq \lambda \leq 1$. It is clear that $z_k \geq 0$ for all k. It is known by the Minkowski inequality [11, p. 30] that $L^{1/2}$-norm is a concave function: $\|\lambda x + (1-\lambda)y\|_{1/2} \geq \lambda\|x\|_{1/2} + (1-\lambda)\|y\|_{1/2}$. Therefore

$$\|z\|_{1/2} \geq \lambda\|x\|_{1/2} + (1-\lambda)\|y\|_{1/2} \geq 4\lambda x_k + 4(1-\lambda)y_k = 4z_k \quad \text{for all} \quad k.$$

Thus T_m is a convex cone.

To show that T_m is the convex hull of the points where $\|x\|_{1/2} = 4x_k$ for some k we proceed by induction. The base case $m = 2$ is simple since T_2 is just a ray spanned by the point $(1, 1)$. For the induction step we observe that any convex set is the convex hull of its boundary. For any point on the boundary of T_m one of the defining $2m$ inequalities must be sharp. If a point x is on the boundary of T_m and $x_i \neq 0$ for all i, then the inequalities $x_i \geq 0$ are not sharp at x; therefore the inequality $\|x\|_{1/2} \geq 4x_k$ must be sharp for some k, and we are done.

If $x_i = 0$ for some i, then the point x lies in the set T_{m-1} in the subspace spanned by the $m-1$ standard basis vectors excluding e_i, and we are done by induction. ∎

Exercise 4.13. Show that the cone $T_4 \subset \mathbb{R}^4$ can be transformed via a nonsingular linear transformation into the dual cone of 3×3 positive semidefinite matrices with equal diagonal elements:

$$(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \quad \text{such that} \quad \begin{bmatrix} x_1 & x_2 & x_3 \\ x_2 & x_1 & x_4 \\ x_3 & x_4 & x_1 \end{bmatrix} \succeq 0.$$

If we restrict x_1 to being 1 then we obtain the *elliptope* $\mathcal{E}_3$, which we have already seen in Chapter 2.

We are now ready to completely describe Σ'.

Theorem 4.14. $\Sigma' = T_8$.

Proof. We rewrite the relation (4.1) in the form

$$\sum_{i=1}^{8} a_i f(s_i) = 0 \quad \text{for} \quad f \in \mathbb{R}[x]_{\mathbf{4,2}}, \tag{4.2}$$

and let $a = (a_1, \ldots, a_8)$ be the vector of coefficients, with $a_i = \pm 1$. It follows that $L = \pi(\mathbb{R}[x]_{\mathbf{4,2}})$ is the hyperplane in $\mathbb{R}^8$ perpendicular to a.

Since T_8 is a convex cone, to show the inclusion $\Sigma' \subseteq T_8$ it suffices by Lemma 4.11 to show that $C \subset T_8$. Let $v = (v_1, \ldots, v_8) \in L$ and $t = (v_1^2, \ldots, v_8^2) \in C$. By the relation (4.2) we have $a_1 v_1 + \cdots + a_8 v_8 = 0$ with $a_i = \pm 1$. Without loss of generality, we may assume that v_1 has the maximal absolute value among v_i. Multiplying the relation (4.2) by -1, if necessary, we can make $a_1 = -1$. Then we have $v_1 = a_2 v_2 + \cdots + a_8 v_8$. We can now write $\sqrt{t_1} = \pm\sqrt{t_2} \pm \sqrt{t_3} \pm \cdots \pm \sqrt{t_8}$ with the exact signs depending on a_i and signs of v_i. Therefore we see that $2\sqrt{t_1} \le \sqrt{t_1} + \cdots + \sqrt{t_8}$. Since v_1 has the largest absolute value among v_i, it follows that $2\sqrt{t_k} \le \sqrt{t_1} + \cdots + \sqrt{t_8}$ for all $1 \le k \le 8$. Hence we see that $\Sigma' \subseteq T_8$.

To show the reverse inclusion $T_8 \subseteq \Sigma'$ we use Lemma 4.12. It suffices to show that all points $x \in T_8$ with $2\sqrt{x_k} = \sqrt{x_1} + \cdots + \sqrt{x_8}$ for some k, are also in Σ'. Without loss of generality we may assume that $k = 1$ and we have $\sqrt{x_1} = \sqrt{x_2} + \cdots + \sqrt{x_8}$. Let $y = (y_1, \ldots, y_8)$ with $y_1 = -\sqrt{x_1}/a_1$ and $y_i = \sqrt{x_i}/a_i$ for $2 \le i \le 8$. It follows that $a_1 y_1 + \cdots + a_8 y_8 = 0$. Therefore $y \in \pi(\mathbb{R}[x]_{\mathbf{4,2}})$ and $y = \pi(q)$ for some quadratic form q. Then $\pi(q^2) = x$ and we are done. ∎

We can use Exercise 4.8 and Theorem 4.14 to visualize the discrepancy between P' and Σ'. Let's take a slice of both cones with the hyperplane H given by $x_1 + \cdots + x_8 = 1$. Recall that by Exercise 4.8 we have $P' = \mathbb{R}_8^+$. Therefore the slice of P' with H is the standard simplex. The slice of T_8 with H is the standard simplex with cut off corners. It was Hilbert's observation that the standard basis vectors e_i are not in Σ', and Theorem 4.14 tells us exactly how much is cut off around the corners.

We now take a short break from comparing $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$ to consider some convexity properties of these cones, such as boundary, facial structure, symmetries, and dual cones.

4.4 Symmetries, Dual Cones, and Facial Structure

4.4.1 Symmetries of $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$

The cones $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$ have a lot of built-in symmetries coming from linear changes of coordinates. Suppose that $A \in GL_n(\mathbb{R})$ is a nonsingular linear transformation of $\mathbb{R}^n$.

Exercise 4.15. Show that if $p(x) \in \mathbb{R}[x]_{\mathbf{2d}}$ is a nonnegative form, then $p(Ax)$ is also a nonnegative form in $\mathbb{R}[x]_{\mathbf{2d}}$. Similarly, if $p(x)$ is a sum of squares, then $p(Ax)$ is also a sum of squares.

In more formal terms, a nonsingular linear transformation A of $\mathbb{R}^n$ induces a nonsingular transformation ϕ_A of $\mathbb{R}[x]_{\mathbf{2d}}$, which maps $p(x) \in \mathbb{R}[x]_{\mathbf{2d}}$ to $p(A^{-1}(x))$. We say that the group $GL_n(\mathbb{R})$ acts on $\mathbb{R}[x]_{\mathbf{2d}}$. It follows from Exercise 4.15 that both cones $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ are invariant under this action. In other words, $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ are invariant under nonsingular linear changes of coordinates.

Exercise 4.16. Show that, up to a constant multiple, $r^{2d} = (x_1^2 + \cdots + x_n^2)^d$ is the only form in $\mathbb{R}[x]_{\mathbf{2d}}$ that is fixed under all orthogonal changes of coordinates; i.e., it is the only form in $\mathbb{R}[x]_{\mathbf{2d}}$ that satisfies

$$p(x) = p(Ax) \quad \text{for all} \quad A \in O_n,$$

where O_n is the group of orthogonal transformations of $\mathbb{R}^n$.

We note that even if a linear transformation A of $\mathbb{R}^n$ is singular, it still induces a linear transformation ϕ_A in the same way. However the linear map ϕ_A will also be singular. The map ϕ_A still sends $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ into themselves, but it will no longer preserve the cones. Closed convex cones in $\mathbb{R}[x]_{\mathbf{2d}}$ that are mapped into themselves under any linear change of coordinates are called *blenders* [24].

4.4.2 Dual Cone of $P_{\mathbf{n},\mathbf{2d}}$

Let K be a convex cone in a real vector space V. Let V^* be the dual vector space of linear functionals on V. The dual cone K^* is defined as the set of all linear functionals in V^* that are nonnegative on K:

$$K^* = \{\ell \in V^* \quad | \quad \ell(x) \geq 0 \quad \text{for all} \quad x \in K\}.$$

Many general aspects of duality will be discussed in Chapter 5. We examine the specific cases of cones of nonnegative polynomials and sums of squares.

Let's consider the dual space $\mathbb{R}[x]^*_{\mathbf{2d}}$ of linear functionals on $\mathbb{R}[x]_{\mathbf{2d}}$. We first observe that the dual cone of $P_{\mathbf{n},\mathbf{2d}}$ is conceptually simple. For $v \in \mathbb{R}^n$, let ℓ_v be the linear functional in $\mathbb{R}[x]^*_{\mathbf{2d}}$ given by evaluation at v:

$$\ell_v(f) = f(v) \quad \text{for} \quad f \in \mathbb{R}[x]_{\mathbf{2d}}.$$

By homogeneity of forms we know that nonnegativity on the unit sphere is equivalent to global nonnegativity. Therefore it is natural to think that the functionals ℓ_v with $v \in \mathbb{S}^{n-1}$ generate the dual cone $P^*_{\mathbf{n},\mathbf{2d}}$. Before we show that this is in fact the case we need a useful exercise from convexity.

Exercise 4.17. Let $K \subset \mathbb{R}^n$ be a compact convex set with the origin not in K. Show that the conical hull of K, cone(K), is closed. Construct an explicit example that shows that the condition $0 \notin K$ is necessary.

Lemma 4.18. *The dual cone $P^*_{\mathbf{n,2d}}$ of the cone of nonnegative forms is the conical hull of linear functionals ℓ_v with v on the unit sphere:*

$$P^*_{\mathbf{n,2d}} = \operatorname{cone}\left(\ell_v \mid v \in \mathbb{S}^{n-1}\right).$$

Proof. Let $L_{\mathbf{n,2d}} \subset \mathbb{R}[x]^*_{\mathbf{2d}}$ be the conical hull of functionals ℓ_v with $v \in \mathbb{S}^{n-1}$. The dual cone $L^*_{\mathbf{n,2d}}$ is the set of all forms $p \in \mathbb{R}[x]_{\mathbf{2d}}$ such that

$$\ell_v(p) = p(v) \geq 0 \quad \text{for all} \quad v \in \mathbb{S}^{n-1}.$$

Therefore we see that $L^*_{\mathbf{n,2d}} = P_{\mathbf{n,2d}}$. Using *biduality* we see that the dual cone $P^*_{\mathbf{n,2d}}$ is equal to the closure of $L_{\mathbf{n,2d}}$:

$$P^*_{\mathbf{n,2d}} = (L^*_{\mathbf{n,2d}})^* = \overline{L_{\mathbf{n,2d}}}.$$

We now just need to show that the cone $L_{\mathbf{n,2d}}$ is closed and then $\overline{L_{\mathbf{n,2d}}} = L_{\mathbf{n,2d}}$. Consider the set C of all linear functionals ℓ_v with $v \in \mathbb{S}^{n-1}$. The set C is given by a continuous embedding of the unit sphere $\mathbb{S}^{n-1}$ into $\mathbb{R}[x]^*_{\mathbf{2d}}$, and therefore C is compact. If we can show that the convex hull of C does not contain the origin, then we are done by applying Exercise 4.17.

Let $r^{2d} = (x_1^2 + \cdots + x_n^2)^d$ be the form in $\mathbb{R}[x]_{\mathbf{2d}}$ that is constantly 1 on the unit sphere. Suppose that $m = \sum c_v \ell_v \in \operatorname{conv}(C)$. Then it follows that $m(r^{2d}) = \sum c_v = 1$, and therefore m cannot be the zero functional in $\mathbb{R}[x]^*_{\mathbf{2d}}$. It follows that $\operatorname{conv}(C)$ is a compact convex set with $0 \notin C$ and we are done. □

Exercise 4.19. Use the *apolar inner product* from Chapter 3 to identify $\mathbb{R}[x]_{\mathbf{2d}}$ with the dual space $\mathbb{R}[x]^*_{\mathbf{2d}}$. Show that the dual cone $P^*_{\mathbf{n,2d}}$ is identified with the *cone of sums of* $2d$*th powers of linear forms*:

$$\left\{p \in \mathbb{R}[x]_{\mathbf{2d}} \;\middle|\; p = \sum q_i^{2d} \quad \text{with} \quad q_i \in \mathbb{R}[x]_{\mathbf{n,1}}\right\}.$$

Remark 4.20. *The map that sends a point $v \in \mathbb{R}^n$ to the form $(v_1x_1 + \cdots + v_nx_n)^{2d}$ is called the* $2d$*th Veronese embedding and its image is called the Veronese variety. It follows from Lemma* 4.18 *that the cone $P^*_{\mathbf{n,2d}}$ is the conical hull of the* $2d$*th Veronese variety. For more information and for connections to orbitopes we refer to* [25].

By applying spherical symmetries to functionals ℓ_v we obtain the following crucial corollary, which describes the extreme rays of $P^*_{\mathbf{n,2d}}$.

Corollary 4.21. *The functional ℓ_v spans an extreme ray of $P^*_{\mathbf{n,2d}}$ for all $v \in \mathbb{S}^{n-1}$, and the functionals ℓ_v form the complete set of extreme rays of $P^*_{\mathbf{n,2d}}$.*

The extreme rays of the cone $P^*_{\mathbf{n,2d}}$ have a very nice parametrization by points $v \in \mathbb{S}^{n-1}$. However, the cone $P^*_{\mathbf{n,2d}}$ is a very complex object from the computational and convex geometry point of view. For example, given a linear functional $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$, determining whether it belongs to the cone $P^*_{\mathbf{n,2d}}$ is known as the *truncated moment problem* in real analysis. Despite a long history, there are very few explicit and computationally feasible criteria for testing membership in $\mathbb{R}[x]^*_{\mathbf{2d}}$. For more on this approach see [15].

Decomposing a given linear functional in $\mathbb{R}[x]^*_{\mathbf{2d}}$ as a linear combination of the functionals ℓ_v, or equivalently by Exercise 4.19, decomposing a given form in $\mathbb{R}[x]_{\mathbf{2d}}$ as a linear combination of forms v^{2d} is known as the *symmetric tensor decomposition problem*. Again, despite a long history, many aspects of symmetric tensor decomposition remain unknown. For more information we refer to [14, 21].

4.4.3 Boundary of the Cone of Nonnegative Polynomials

The boundary and the interior of the cone of nonnegative forms $P_{\mathbf{n,2d}}$ are easy to describe given our knowledge of the dual cone $P^*_{\mathbf{n,2d}}$.

Exercise 4.22. Show that the interior of $P_{\mathbf{n,2d}}$ consists of forms that are strictly positive on $\mathbb{R}^n \setminus \{0\}$ and the boundary of $P_{\mathbf{n,2d}}$ consists of forms with a nontrivial zero.

We note that the situation is slightly different in the nonhomogeneous case. Let $f(x) = x^2 + 1$ be a univariate polynomial, and let P be the cone of nonnegative univariate polynomials of degree at most 4. Clearly $f \in P$ and f is strictly positive on $\mathbb{R}$. However, f lies on the boundary of P. Consider $g_\epsilon = f - \epsilon x^4$. For any $\epsilon > 0$ the polynomial g_ϵ will not be nonnegative. Therefore f is not in the interior of P, and it lies on the boundary of P.

The explanation for this phenomenon is that even though f is strictly positive on $\mathbb{R}$, when viewed as a polynomial of degree 4, f has a *zero at infinity*. The growth of $f(x)$ as x goes to infinity is only of order 2, and therefore we cannot subtract a nonnegative polynomial of degree 4 from f and have the difference remain nonnegative. The easiest way to see the zero at infinity is to homogenize f with an extra variable y: $\bar{f} = x^2y^2 + y^4$.

Note that if we set $y = 1$ in $\bar{f}$ we just recover f. However, $\bar{f}$ is not a strictly positive form on $\mathbb{R}^2 \setminus \{0\}$, since $\bar{f}$ has a nontrivial zero which comes from setting $y = 0$. In general, for a polynomial f in n variables of degree d, let f_d be the degree d component of f consisting of all terms of degree exactly d. *Zeroes at infinity* of f correspond to zeroes of f_d. This can be seen by homogenizing f with an extra variable. When we set this variable equal to 0 we obtain f_d.

4.4.4 Exposed Faces of $P_{\mathbf{n,2d}}$

Exposed faces of $P_{\mathbf{n,2d}}$ are conceptually easy to understand due to our knowledge of the extreme rays of the dual cone $P^*_{\mathbf{n,2d}}$ in Corollary 4.21. Maximal (by inclusion) faces of $P_{\mathbf{n,2d}}$ come from the vanishing of one extreme ray of the dual cone. Therefore it follows that maximal faces $F(v)$ of $P_{\mathbf{n,2d}}$ consist of all nonnegative forms that have a single common zero $v \in \mathbb{S}^{n-1}$:

$$F(v) = \{p \in P_{\mathbf{n,2d}} \quad | \quad \ell_v(p) = p(v) = 0\}.$$

We observe that a zero of a nonnegative form p is a local minimum. Therefore, if $p(v) = 0$, this implies that the gradient of p at v is zero as well, $\nabla p(v) = 0$. In other words, p must have a double zero at v.

Exercise 4.23 (Euler's relation). Show that for $p \in \mathbb{R}[x]_{\mathbf{d}}$ and all $v \in \mathbb{R}^n$ the following relation holds:

$$\langle \nabla p(v), v\rangle = d \cdot p(v).$$

From the above exercise it follows that for forms $p \in \mathbb{R}[x]_{\mathbf{2d}}$ the vanishing of the gradient at v, $\nabla p(v) = 0$, forces the form p to vanish at v as well, $p(v) = 0$. Therefore, for a nonnegative form $p \in P_{\mathbf{n,2d}}$ a single zero forces p to satisfy n linear conditions coming from $\nabla p(v) = 0$. It follows that the face $F(v)$ has codimension at least n.

Exercise 4.24. Show that the maximal faces $F(v)$ of $P_{\mathbf{n,2d}}$ have codimension exactly n in $\mathbb{R}[x]_{\mathbf{2d}}$.

All smaller exposed faces $F(v_1, \dots, v_k)$ come from the vanishing of several extreme rays $\ell_{v_1}, \dots, \ell_{v_k}$ of $P^*_{\mathbf{n,2d}}$. The face $F(v_1, \dots, v_k)$ has the form

$$F(v_1, \dots, v_k) = \{p \in P_{\mathbf{n,2d}} \quad | \quad p(v_1) = \dots = p(v_k) = 0,\ v_i \in \mathbb{S}^{n-1}\}.$$

Therefore $F(v_1, \dots, v_k)$ consists of all nonnegative forms with zeroes at prescribed points $v_1, \dots, v_k \in \mathbb{S}^{n-1}$. It is natural to expect that every additional zero increases the codimension of the exposed face by n so that $\operatorname{codim} F(v_1, \dots, v_k) = kn$. However, this intuition fails if the number of zeroes k is sufficiently large. In particular if we prescribe enough zeroes, it is not even clear when the face $F(v_1, \dots, v_k)$ is nonempty. The question of the dimension of $F(v_1, \dots, v_k)$ is quite complicated [6] and it is related to the celebrated Alexander–Hirschowitz theorem [17].

Exposed extreme rays of $P_{\mathbf{n,2d}}$ are also conceptually simple: a nonnegative form $p \in P_{\mathbf{n,2d}}$ is an exposed extreme ray of $P_{\mathbf{n,2d}}$ if and only if the variety defined by p is maximal among all varieties defined by nonnegative polynomials.

Exercise 4.25. Show that $p \in P_{\mathbf{n,2d}}$ is an exposed extreme ray of $P_{\mathbf{n,2d}}$ if and only if for all $q \in P_{\mathbf{n,2d}}$ with $V(p) \subseteq V(q)$ it follows that $q = \lambda p$ for some $\lambda \in \mathbb{R}$.

4.4.5 Nonexposed Faces of $P_{\mathbf{n,2d}}$

The cone $P_{\mathbf{n,2d}}$ has many nonexposed faces. If a form p has a zero at a point $v \in \mathbb{R}$, then it must have a double zero at v. Exposed faces of $P_{\mathbf{n,2d}}$ capture double zeroes on any set of points $v_1, \dots, v_k$, but exposed faces fail to capture zeroes of higher order.

Exercise 4.26. Show that x_1^{2d} is an extreme ray of $P_{\mathbf{n},\mathbf{2d}}$. Use Exercise 4.25 to conclude that x_1^{2d} is not exposed.

More generally, the following construction explains the origins of nonexposed faces of $P_{\mathbf{n},\mathbf{2d}}$. Consider a maximal face $F(v)$ of $P_{\mathbf{n},\mathbf{2d}}$. We can construct an exposed subface of $F(v)$ by considering nonnegative forms with zeroes at v and w for some $w \in \mathbb{S}^{n-1}$. We can also build nonexposed subfaces of $F(v)$ by considering nonnegative forms that are more singular at v.

Let $p \in F(v)$, so that p is a nonnegative form and $p(v) = 0$. Since 0 is the global minimum of p and $\nabla p(v) = 0$, it follows that the Hessian $\nabla^2 p(v)$ must be a positive semidefinite matrix. Let $F_w(v)$ be the set of all nonnegative forms p with zero at v whose Hessian at v is positive semidefinite and w lies in the kernel of $\nabla^2 p(v)$:

$$F_w(v) = \{p \in F(v) \quad | \quad \nabla^2 p(v) \cdot w = 0\}.$$

Exercise 4.27. Show that $F_w(v)$ is a face of $P_{\mathbf{n},\mathbf{2d}}$. Use the characterization of exposed faces of $P_{\mathbf{n},\mathbf{2d}}$ to show that $F_w(v)$ is not an exposed face of $P_{\mathbf{n},\mathbf{2d}}$.

4.4.6 Algebraic Boundaries

The boundaries of the cones $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$ are hypersurfaces in $\mathbb{R}[x]_{\mathbf{2d}}$. Suppose that we would like to describe these hypersurfaces by polynomial equations. This leads to the notion of *algebraic boundary* of the cones $P_{\mathbf{n},\mathbf{2d}}$ and $\Sigma_{\mathbf{n},\mathbf{2d}}$, which is obtained by taking the *Zariski closure* of the boundary hypersurfaces. As explained in Chapter 5, the algebraic boundary of $P_{\mathbf{n},\mathbf{2d}}$ is cut out by a single polynomial, the discriminant. The algebraic boundary of the cone of sums of squares is significantly more complicated.

Exercise 4.28. Show that the hypersurface cut out by the discriminant is a component of the algebraic boundary of $\Sigma_{\mathbf{n},\mathbf{2d}}$.

The above exercise shows that the algebraic boundary of $P_{\mathbf{n},\mathbf{2d}}$ is included in the algebraic boundary of $\Sigma_{\mathbf{n},\mathbf{2d}}$. This seems counterintuitive, but it occurs because we passed to the Zariski closures of the actual boundaries. We will see below that for $\Sigma_{\mathbf{3},\mathbf{6}}$ and $\Sigma_{\mathbf{4},\mathbf{4}}$ the algebraic boundary of the cone of sums of squares has one more component, which is described in Exercise 4.51.

4.5 Generalizing the Hypercube Example

We completely described the values of nonnegative forms and sums of squares on the specific set S of ± 1 vectors in $\mathbb{R}^4$ and we have seen, just from the evaluation on S, that there exist nonnegative forms in $\mathbb{R}[x]_{\mathbf{4},\mathbf{4}}$ that are not sums of squares.

However, these descriptions are limited to the specific set S. We now extend the arguments of Section 4.3 to work in far greater generality. We begin by explaining how the set S was chosen in the first place.

4.5.1 Hypercube Example Revisited

Let q_i be the three quadratic forms

$$q_1 = x_1^2 - x_2^2, \quad q_2 = x_1^2 - x_3^2, \quad q_3 = x_1^2 - x_4^2,$$

and let V be the set of common zeroes of q_i:

$$V = \{x \in \mathbb{R}^4 \quad | \quad q_i(x) = 0 \quad \text{for} \quad i = 1, 2, 3\}.$$

Viewed projectively V consists of eight points in the real projective space $\mathbb{RP}^3$. Viewed affinely V consists of eight lines, each line spanned by a point in S. We can extend much of what was proved about the values of nonnegative polynomials to zero-dimensional intersections in $\mathbb{RP}^{n-1}$.

4.5.2 Zero-Dimensional Intersections

Let V be a set of finitely many points in $\mathbb{RP}^{n-1}$:

$$V = \{\bar{s}_1, \ldots, \bar{s}_k\}.$$

Suppose that V is the complete set of real projective zeroes of some forms $q_1, \ldots, q_m$ of degree d:

$$V = \{x \in \mathbb{RP}^{n-1} \quad | \quad q_1(x) = \cdots = q_m(x) = 0\}.$$

For each $\bar{s_i} \in V$ let s_i be an affine representative of $\bar{s_i}$ lying on the line spanned by $\bar{s_i}$. Now let $S = \{s_1, \ldots, s_k\}$, be the set of affine representatives corresponding to the common zeroes of q_i.

Let's consider the values of nonnegative forms of degree $2d$ on S. Let $\pi_S\colon \mathbb{R}[x]_{\mathbf{2d}} \to \mathbb{R}^k$ be the evaluation projection:

$$\pi_S(f) = (f(s_1), \ldots, f(s_k)) \quad \text{for} \quad f \in \mathbb{R}[x]_{\mathbf{2d}}.$$

Let H be the image of $\mathbb{R}[x]_{\mathbf{2d}}$ and let P' be the image of $P_{\mathbf{n},\mathbf{2d}}$ under π_S:

$$H = \pi_S(\mathbb{R}[x]_{\mathbf{2d}}), \qquad P' = \pi_S(P_{\mathbf{n},\mathbf{2d}}).$$

We have an additional complication that H does not have to equal $\mathbb{R}^k$. We know, however, that P' must lie in H, and since we are evaluating nonnegative forms it follows that P' lies inside the nonnegative orthant of $\mathbb{R}^k$: $P' \subseteq \mathbb{R}^k_+$. Therefore it follows that P' lies inside the intersection of H and $\mathbb{R}^k_+$:

$$P' \subseteq H \cap \mathbb{R}^k_+.$$

The following theorem shows that this inclusion is almost an equality.

Theorem 4.29. *Let $\mathbb{R}^k_{++}$ be the positive orthant of $\mathbb{R}^k$. The intersection of H with the positive orthant is contained in P':*

$$H \cap \mathbb{R}^k_{++} \subset P'.$$

Before proving Theorem 4.29 we make some remarks. As we know from Exercise 4.7 we cannot simply conclude that $P' = H \cap \mathbb{R}^k_+$ using a closure argument, since a projection of a closed cone does not have to be closed. We now show that this occurs for evaluation projections as well.

Exercise 4.30. Let $S \subset \mathbb{R}^5$ be the set of 16 points $S = \{\pm 1, \pm 1, \pm 1, \pm 1, 1\}$. Show that S can be defined as a common zero set of four quadratic forms in $\mathbb{R}[x]_{\mathbf{5},\mathbf{2}}$, and use Theorem 4.29 to show that $\mathbb{R}^{16}_{++} \subseteq \pi_S(P_{\mathbf{5},\mathbf{4}})$. Show that the standard basis vectors $e_i \in \mathbb{R}^{16}$ are not in the image $\pi_S(P_{\mathbf{5},\mathbf{4}})$. In other words, the vectors e_i are not realized as values on S of a nonnegative form of degree 4 in 5 variables, but all strictly positive points in $\mathbb{R}^{16}_{++}$ are realized.

***Proof of Theorem* 4.29.** Let $v = (v_1, \ldots, v_k) \in H \cap \mathbb{R}^k_{++}$. Since $v \in H$ there exists a form $f \in \mathbb{R}[x]_{\mathbf{2d}}$ such that $f(s_i) = v_i$. Let $g = q_1^2 + \cdots + q_m^2$, where q_i are the forms defining V. We claim that for large enough $\lambda \in \mathbb{R}$ the form $\bar{f} = f + \lambda g$ will be nonnegative, and since each q_i is zero on S we will also have $\pi_S\left(\bar{f}\right) = v$.

By homogeneity of $\bar{f}$ it suffices to show that it is nonnegative on the unit sphere $\mathbb{S}^{n-1}$. Furthermore, we may assume that the evaluation points s_i lie on the unit sphere. Since we are dealing with forms, evaluation on the points outside of the unit sphere amounts to rescaling of the values on $\mathbb{S}^{n-1}$.

Let $B_\epsilon(S)$ be the open epsilon neighborhood of S in the unit sphere $\mathbb{S}^{n-1}$. Since $f(s_i) > 0$ for all i, it follows that for sufficiently small ϵ the form f is strictly positive on $B_\epsilon(S)$:

$$f(x) > 0 \quad \text{for all} \quad x \in B_\epsilon(S).$$

The complement of $B_\epsilon(S)$ in $\mathbb{S}^{n-1}$ is compact, and therefore we can let m_1 be the minimum of g and m_2 be the minimum of f on $\mathbb{S}^{n-1} \setminus B_\epsilon(S)$. If $m_2 \geq 0$, then f itself is nonnegative and we are done. Therefore, we may assume $m_2 < 0$. We also note that since g vanishes on S only, it follows that m_1 is strictly positive.

Now let $\lambda \geq -\frac{m_2}{m_1}$. The form $\bar{f} = f + \lambda g$ is positive on $B_\epsilon(S)$. By construction of $B_\epsilon(S)$ we also see that the minimum of $\bar{f}$ on the complement of $B_\epsilon(S)$ is at least 0. Therefore $\bar{f}$ is nonnegative on the unit sphere $\mathbb{S}^{n-1}$, and we are done. □

We proved in Theorem 4.29 that any set of strictly positive values on the finite set S, coming from real zeroes of forms of degree d, can be achieved by a globally nonnegative form of degree $2d$. We now look at the values that sums of squares can take on such sets S.

4.5.3 Values of Sums of Squares

We recall from Section 4.3 that the reason that sums of squares could not achieve all the possible nonnegative values on the hypercube was that the values of quadratic forms on the hypercube satisfied a linear relation. The points of the hypercube come from common zeroes of the quadratic forms, as we have seen in Section 4.5.1.

There is a general theory in algebraic geometry on the number of relations that values of forms of certain degree have to satisfy on finite sets of points. These relations are known as Cayley–Bacharach relations. For more details we refer the reader to [10].

At first glance it is surprising that there should be any linear relation at all. If the points were chosen generically then the values of forms of degree d on these points would be linearly independent, at least until we have as many points as the dimension of the vector space of forms of degree d. However, our choice of points is not generic; point sets that come from common zeroes are special.

For the cases $\mathbb{R}[x]_{\mathbf{4,4}}$ and $\mathbb{R}[x]_{\mathbf{3,6}}$ it is easy to establish the existence of the linear relation by simple dimension counting. We explain the case of $\mathbb{R}[x]_{\mathbf{4,4}}$.

Since common zeroes of real forms do not have to be real, for this section we will work with complex forms. Suppose that $q_1, q_2, q_3 \in \mathbb{C}[x]_{\mathbf{4,2}}$ are complex quadratic forms in 4 variables. As before let V be the complete set of projective zeroes of some forms q_1, q_2, q_3:

$$V = \{\bar{x} \in \mathbb{CP}^3 \quad | \quad q_1(\bar{x}) = q_2(\bar{x}) = q_3(\bar{x}) = 0\}.$$

Three quadratic forms in $\mathbb{C}[x]_{\mathbf{4,2}}$ are expected to generically have $2^3 = 8$ common zeroes. Suppose that this is the case and let $V = \{\bar{s}_1, \dots, \bar{s}_8\}$.

For each $\bar{s}_i \in V$ let s_i be an affine representative of $\bar{s}_i$ lying on the line corresponding to $\bar{s}_i$. Let $S = \{s_1, \dots, s_8\}$, be the set of affine representatives corresponding to the common zeroes of q_i. Define $\pi_S : \mathbb{C}[x]_{\mathbf{4,2}} \to \mathbb{C}^8$ to be the evaluation projection.

Lemma 4.31. *The values of quadratic forms in* $\mathbb{C}[x]_{\mathbf{4,2}}$ *satisfy a linear relation on the points of* S*. In other words there exist* $\mu_1, \dots \mu_8 \in \mathbb{C}$ *such that*

$$\mu_1 f(s_1) + \cdots + \mu_8 f(s_8) = 0 \quad \textit{for all} \quad f \in \mathbb{C}[x]_{\mathbf{4,2}}. \tag{4.3}$$

Proof. The dimension of $\mathbb{C}[x]_{\mathbf{4,2}}$ is 10. Note that the kernel of π_S contains the three forms q_i, since each q_i evaluates to 0 on S. Therefore the dimension of the kernel of π_S is at least 3. It follows that the image of π_S has dimension at most $10 - 3 = 7$. Since the image of π_S lies inside $\mathbb{C}^8$, it follows that there exists a linear functional that vanishes on the image of π_S. This linear functional gives us the desired linear relation. □

Remark 4.32. *It is possible to show in the above proof that the dimension of the kernel of* π_S *is exactly* 3 *and therefore the linear relation* (4.3) *is unique. Furthermore, it can be shown that each* $\mu_i \neq 0$*, or, in other words, the unique linear relation has to involve all of the points of* S*.*

Exercise 4.33. Suppose that $q_1, q_2 \in \mathbb{C}[x]_{3,3}$ are two cubic forms intersecting in $3^2 = 9$ points in $\mathbb{CP}^2$. Let S be the set of affine representatives of the common zeroes of q_1 and q_2. Use the argument of Lemma 4.31 to show that the values of cubic forms on S satisfy a linear relation.

Exercise 4.34. The Robinson form

$$R(x,y,z) = x^6 + y^6 + z^6 - (x^4y^2 + x^2y^4 + x^4z^2 + x^2z^4 + y^4z^2 + y^2z^4) + 3x^2y^2z^2$$

is an explicit example of a nonnegative polynomial that is not a sums of squares. Let $q_1 = x(x+z)(x-z)$ and $q_2 = y(y+z)(y-z)$. Calculate the 9 common zeroes of q_1 and q_2. Show that $R(x,y,z)$ vanishes on 8 of the 9 zeroes. Use Exercise 4.33 to show that $R(x,y,z)$ is not a sum of squares.

We have examined in detail what happens to values of nonnegative forms and sums of squares on finite sets of points coming from common zeroes of forms. However, this still seems to be a very special construction. We now move to show that the difference in values on such sets is in fact the fundamental reason that there exists nonnegative polynomials that are not sums of squares.

4.6 Dual Cone of $\Sigma_{\mathbf{n,2d}}$

We gave a simple description of the extreme rays of the dual cone $P^*_{\mathbf{n,2d}}$ in Corollary 4.21. The description of the extreme rays of the dual cone $\Sigma^*_{\mathbf{n,2d}}$ is significantly more complicated. We will see that evaluation on the special finite point sets we described in Section 4.5 will naturally lead to extreme rays of $\Sigma^*_{\mathbf{n,2d}}$.

We first describe the connection between $\Sigma^*_{\mathbf{n,2d}}$ and the cone of positive semidefinite matrices that lies at the heart of semidefinite programming approaches to polynomial optimization. To every linear functional $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ we can associate a quadratic form Q_ℓ defined on $\mathbb{R}[x]_{\mathbf{d}}$ by setting

$$Q_\ell(f) = \ell(f^2) \quad \text{for all} \quad f \in \mathbb{R}[x]_{\mathbf{d}}.$$

The cone $\Sigma^*_{\mathbf{n,2d}}$ can be thought of as a section of the cone of positive semidefinite quadratic forms. We now show how this description arises.

Lemma 4.35. *Let ℓ be a linear functional in $\mathbb{R}[x]^*_{\mathbf{2d}}$. Then $\ell \in \Sigma^*_{\mathbf{n,2d}}$ if and only if the quadratic form Q_ℓ is positive semidefinite.*

Proof. Suppose that $\ell \in \Sigma^*_{\mathbf{n,2d}}$. Then $\ell(f^2) \geq 0$ for all $f \in \mathbb{R}[x]_{\mathbf{d}}$. Therefore $Q_\ell(f) \geq 0$ for all $f \in \mathbb{R}[x]_{\mathbf{d}}$ and Q_ℓ is positive semidefinite.

Now suppose that Q_ℓ is positive semidefinite. Then $\ell(f^2) \geq 0$ for all $f \in \mathbb{R}[x]_{\mathbf{d}}$. Let $g = \sum f_i^2 \in \Sigma_{\mathbf{n,2d}}$. Then $\ell(g) = \sum \ell(f_i^2) \geq 0$ and $\ell \in \Sigma^*_{\mathbf{n,2d}}$. □

An Aside: The Monomial Basis and Moment Matrices

Suppose that we fix the monomial basis for $\mathbb{R}[x]_{\mathbf{d}}$. Given a linear functional $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ we can write an explicit matrix $M(\ell)$ for the quadratic form Q_ℓ using the monomial basis of $\mathbb{R}[x]_{\mathbf{d}}$. The matrix $M(\ell)$ is known as the *moment matrix* or *generalized Hankel matrix.* The entries of $M(\ell)$ are indexed by monomials $x^\alpha, x^\beta \in \mathbb{R}[x]_{\mathbf{d}}$. The entry $M(\ell)_{\alpha,\beta}$ is given by evaluating ℓ on $x^\alpha x^\beta = x^{\alpha+\beta}$:

$$M(\ell)_{\alpha,\beta} = \ell(x^{\alpha+\beta}).$$

For example, consider the linear functional $\ell_v : \mathbb{R}[x]_{\mathbf{2,4}} \to \mathbb{R}$ given by evaluation on $v = (1,2)$. The monomial basis of $\mathbb{R}[x]_{\mathbf{2,2}}$ is given by x^2, xy, y^2 and the moment matrix of $M(\ell_v)$ reads as

$$M(\ell_v) = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 4 & 8 & 16 \end{bmatrix}.$$

The rank of the quadratic form Q_ℓ is the same as the rank of its moment matrix $M(\ell)$, and Q_ℓ being nonnegative is equivalent to having a positive semidefinite moment matrix $M(\ell)$. However, the moment approach is tied to the specific choice of the monomial basis. Below we prefer to keep a basis independent approach with emphasis on the underlying geometry, but we note that the results are readily translatable into the terminology of moments. ■

Let $S^{n,d}$ be the vector space of real quadratic forms on $\mathbb{R}[x]_{\mathbf{d}}$. We can view the dual space $\mathbb{R}[x]^*_{\mathbf{2d}}$ as a subspace of $S^{n,d}$ by identifying the linear functional $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ with its quadratic form Q_ℓ. Let $S^{n,d}_+$ be the cone of positive semidefinite forms in $S^{n,d}$:

$$S^{n,d}_+ = \left\{ Q \in S^{n,d} \;\middle|\; Q(f) \geq 0 \quad \text{for all} \quad f \in \mathbb{R}[x]_{\mathbf{d}} \right\}.$$

We can restate Lemma 4.35 as follows.

Corollary 4.36. *The cone $\Sigma^*_{\mathbf{n,2d}}$ is the section of the cone of positive semidefinite matrices $S^{n,d}_+$ with the subspace $\mathbb{R}[x]^*_{\mathbf{2d}}$:*

$$\Sigma^*_{\mathbf{n,2d}} = S^{n,d}_+ \cap \mathbb{R}[x]^*_{\mathbf{2d}}.$$

Note that this shows that the cone $\Sigma^*_{\mathbf{n,2d}}$ is a *spectrahedron.*

The following exercise establishes the connection between the cone of sums of squares $\Sigma_{\mathbf{n,2d}}$ and the cone of positive semidefinite matrices $S^{n,d}_+$. This allows us to formulate sums of squares questions in terms of semidefinite programming.

Exercise 4.37. Use the result of Corollary 4.36 to show that the cone $\Sigma_{\mathbf{n,2d}}$ is a projection of the cone $S^{n,d}_+$ of positive semidefinite matrices on $\mathbb{R}[x]_{\mathbf{d}}$. Use the monomial basis of $\mathbb{R}[x]_{\mathbf{d}}$ to describe this projection explicitly. Conclude that the cone $\Sigma_{\mathbf{n,2d}}$ is a *projected spectrahedron.* (Hint: See Chapter 5 for a general discussion of the relationship between duality and projections.)

Remark 4.38. *In order to apply the result of Exercise* 4.37 *to actual computation we need to work with an explicit basis of* $\mathbb{R}[x]_{\mathbf{d}}$*. See Chapter* 3 *for the discussion of possible basis choices and their impact on computational performance. We note that the size of the positive semidefinite matrices we work with is the dimension of* $\mathbb{R}[x]_{\mathbf{d}}$*, which is equal to* $\binom{n+d-1}{d}$*. Therefore the size of the underlying positive semidefinite matrices increases rather rapidly as a function of n and d. This is one of the main computational limitations of semidefinite programming approaches to polynomial optimization.*

We would like to see what separates sums of squares from nonnegative forms. The extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ cut out the cone of sums of squares. Therefore we would like to find extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ that are not in the dual cone $P^*_{\mathbf{n,2d}}$, since these are the functionals that distinguish the cone of sums of squares from the cone of nonnegative forms.

Formally the dual cone $\Sigma^*_{\mathbf{n,2d}}$ is defined as the cone of linear functionals nonnegative on $\Sigma_{\mathbf{n,2d}}$, which is equivalent to being nonnegative on squares. One way of constructing linear functionals nonnegative on squares is to consider point evaluation functionals ℓ_v with $v \in \mathbb{R}^n$ that send $p \in \mathbb{R}[x]_{\mathbf{2d}}$ to $p(v)$. However, as we have seen in Corollary 4.21, point evaluation functionals are precisely the extreme rays of $P^*_{\mathbf{n,2d}}$. Therefore, these linear functionals are not helpful in distinguishing between $\Sigma^*_{\mathbf{n,2d}}$ and $P^*_{\mathbf{n,2d}}$. Our goal now is to find a new way of constructing functionals nonnegative on squares and also to understand why such functionals do not exist when $\Sigma_{\mathbf{n,2d}} = P_{\mathbf{n,2d}}$.

We showed in Corollary 4.36 that the cone $\Sigma^*_{\mathbf{n,2d}}$ is a spectrahedron. We now prove a general lemma about spectrahedra that states that extreme rays of a spectrahedron are quadratic forms with maximal kernel [20]. The examination of the kernels of extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ will provide a crucial tool for our understanding of $\Sigma^*_{\mathbf{n,2d}}$.

Let $\mathcal{S}$ be the vector space of quadratic forms on a real vector space V. Let $\mathcal{S}_+$ be the cone of psd forms in $\mathcal{S}$.

Lemma 4.39. *Let L be a linear subspace of $\mathcal{S}$ and let K be the section of $\mathcal{S}_+$ with L:*

$$K = \mathcal{S}_+ \cap L.$$

Suppose that a quadratic form Q spans an extreme ray of K. Then the kernel of Q is maximal for all quadratic forms in L: if $P \in L$ and $\ker Q \subseteq \ker P$ then $P = \lambda Q$ for some $\lambda \in \mathbb{R}$.

Proof. Suppose not, so that there exists an extreme ray Q of K and a quadratic form $P \in L$ such that $\ker Q \subseteq \ker P$ and $P \neq \lambda Q$. Since $\ker Q \subseteq \ker P$ it follows that all eigenvectors of both Q and P corresponding to nonzero eigenvalues lie in the orthogonal complement $(\ker Q)^\perp$ of $\ker Q$. Furthermore, Q is positive definite on $(\ker Q)^\perp$.

It follows that Q and P can be simultaneously diagonalized to matrices Q' and P' with the additional property that whenever the diagonal entry Q'_{ii} is 0 the corresponding entry P'_{ii} is also 0. Therefore, for sufficiently small $\epsilon \in \mathbb{R}$ we have that $Q + \epsilon P$ and $Q - \epsilon P$ are positive semidefinite and therefore $Q + \epsilon P, Q - \epsilon P \in K$. Then Q is not an extreme ray of K, which is a contradiction. □

We now apply Lemma 4.39 to the case $\Sigma^*_{\mathbf{n,2d}}$. This gives us a crucial tool for studying extreme rays of $\Sigma^*_{\mathbf{n,2d}}$.

Corollary 4.40. *Suppose that Q spans an extreme ray of $\Sigma^*_{\mathbf{n,2d}}$. Then either* $\operatorname{rank} Q = 1$ *or the forms in the kernel of Q have no common zeroes, real or complex.*

Proof. Let $W \subset \mathbb{R}[x]_{\mathbf{d}}$ be the kernel of Q and suppose that the forms in W have a common real zero $v \neq 0$. Let $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ be the linear functional given by evaluation at v: $\ell(f) = f(v)$ for all $f \in \mathbb{R}[x]_{\mathbf{2d}}$. Then Q_ℓ is a rank 1 positive semidefinite quadratic form and $\ker Q \subseteq \ker Q_\ell$. By Lemma 4.39 it follows that $Q = \lambda Q_\ell$ and thus Q has rank 1.

Now suppose that the forms in W have a common complex zero $z \neq 0$. Let $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ be the linear functional given by taking the real part of the value at z: $\ell(f) = \operatorname{Re} f(z)$ for all $f \in \mathbb{R}[x]_{\mathbf{2d}}$. It is easy to check that the kernel of Q_ℓ includes all forms that vanish at z and therefore $W \subseteq \ker Q_\ell$. Therefore by applying Lemma 4.39 we again see that $Q = \lambda Q_\ell$. However, we claim that Q_ℓ is not a positive semidefinite form.

The quadratic form Q_ℓ is given by $Q_\ell(f) = \operatorname{Re} f^2(z)$ for $f \in \mathbb{R}[x]_{\mathbf{d}}$. However, there exist $f \in \mathbb{R}[x]_{\mathbf{d}}$ such that $f(z)$ is purely imaginary and therefore $Q_\ell(f) < 0$. The corollary now follows. ∎

Corollary 4.40 shows that extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ are of two types: either they are rank 1 quadratic forms or they have a kernel with no common zeroes. We now deal with the rank 1 extreme rays of $\Sigma^*_{\mathbf{n,2d}}$. For $v \in \mathbb{R}^n$ let ℓ_v be the linear functional in $\mathbb{R}[x]^*_{\mathbf{2d}}$ given by evaluation at v,

$$\ell_v(f) = f(v) \text{ for } f \in \mathbb{R}[x]_{\mathbf{2d}},$$

and let Q_v be the quadratic form associated to ℓ_v: $Q_v(f) = f^2(v)$. In this case we say that Q_v (or ℓ_v) corresponds to point evaluation. Recall that the inequalities $\ell_v \geq 0$ are the defining inequalities of the cone of nonnegative forms $P_{\mathbf{n,2d}}$. The following lemma shows that all rank 1 forms in $\mathbb{R}[x]^*_{\mathbf{2d}}$ correspond to point evaluations. Since we are interested in the inequalities that are valid on $\Sigma_{\mathbf{n,2d}}$ but not valid on $P_{\mathbf{n,2d}}$ it allows us to disregard rank 1 extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ and focus on the case of a kernel with no common zeros.

Lemma 4.41. *Suppose that Q is a rank 1 quadratic form in $\mathbb{R}[x]^*_{\mathbf{2d}}$. Then $Q = \lambda Q_v$ for some $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.*

Proof. Let Q be a rank 1 form in $\mathbb{R}[x]^*_{\mathbf{2d}}$. Then $Q(f) = \lambda s^2(f)$ for some linear functional $s \in \mathbb{R}[x]^*_{\mathbf{d}}$. Therefore it suffices to show that if $Q = s^2(f)$ for some $s \in \mathbb{R}[x]^*_{\mathbf{d}}$, then $Q = Q_v$ for some $v \in \mathbb{R}^n$.

Since $Q \in \mathbb{R}[x]^*_{\mathbf{2d}}$ we know that Q is defined by $Q(f) = \ell(f^2)$ for a linear functional $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ and therefore $\ell(f^2) = s^2(f)$ for all $f \in \mathbb{R}[x]_{\mathbf{d}}$. We have $Q(f + g) = \ell((f+g)^2) = \ell(f^2)+2\ell(fg)+\ell(g^2) = (s(f)+s(g))^2 = s^2(f)+2s(f)s(g)+s^2(g)$ and it follows that $\ell(fg) = s(f)s(g)$ for all $f, g \in \mathbb{R}[x]_{\mathbf{d}}$.

Let x^α denote the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. If we take monomials x^α, x^β, x^γ, x^δ in $\mathbb{R}[x]_{\mathbf{d}}$ such that $x^\alpha x^\beta = x^\gamma x^\delta$, then we must have $s(x^\alpha)s(x^\beta) = s(x^\gamma)s(x^\delta)$.

Suppose that $s(x_i^d) = 0$ for all i. Then we see that

$$s(x_i^{d-1}x_j)^2 = s(x_i^d)s(x_i^{d-2}x_j^2) = 0,$$

and continuing in similar fashion we have $s(x^\alpha) = 0$ for all monomials in $\mathbb{R}[x]_{\mathbf{d}}$. Then ℓ is the zero functional and Q does not have rank one which is a contradiction.

We may assume without loss of generality that $s(x_1^d) \neq 0$. Since we are interested in $\ell(f^2) = s^2(f)$ we can work with $-s$ if necessary, and thus we may assume that $s(x_1^d) > 0$. Let $s_i = s(x_1^{d-1}x_i)$ for $1 \leq i \leq n$. We will express $s(x^\alpha)$ in terms of s_i for all $x^\alpha \in \mathbb{R}[x]_{\mathbf{d}}$. Since $(x_1^d)(x_1^{d-2}x_ix_j) = (x_1^{d-1}x_i)(x_1^{d-1}x_j)$ we have $s(x_1^{d-2}x_ix_j) = s_is_j/s_1$. Continuing in this fashion we find that

$$s(x_1^{\alpha_1}\cdots x_n^{\alpha_n}) = \frac{s_2^{\alpha_2}\cdots s_n^{\alpha_n}}{s_1^{d-1-\alpha_1}}.$$

Now let $v \in \mathbb{R}^n$ be the following vector:

$$v = (s_1^{1/d}, s_1^{-(d-1)/d}s_2, \ldots, s_1^{-(d-1)/d}s_n).$$

Let s_v be the linear functional on $\mathbb{R}[x]_{\mathbf{d}}$ defined by evaluating a form at v: $s_v(f) = f(v)$. Then we have $s_v(x_1^{d-1}x_i) = s_i$ and

$$s_v(x_1^{\alpha_1}\cdots x_n^{\alpha_n}) = s_2^{\alpha_2}\cdots s_n^{\alpha_n}s_1^{\alpha_1/d-(d-1)(d-\alpha_1)/d} = \frac{s_2^{\alpha_2}\cdots s_n^{\alpha_n}}{s_1^{d-1-\alpha_1}}.$$

Since s agrees with s_v on monomials it follows that $s = s_v$ and thus $\ell(f^2) = s^2(f) = f(v)^2 = f^2(v)$. Therefore ℓ indeed corresponds to point evaluation and we are done. $\square$

Suppose that Q_ℓ spans an extreme ray of $\Sigma^*_{\mathbf{n},\mathbf{2d}}$ that does not correspond to point evaluation. Let W_ℓ be the kernel of Q_ℓ. Then by Corollary 4.40 and Lemma 4.41 we know that the forms in W_ℓ have no common zeroes real or complex. This condition gives us a lot of dimensional information about W_ℓ and places strong restrictions on the linear functionals ℓ. As we will see, for the three equality cases of Hilbert's theorem the dimensional restrictions on W_ℓ will allow us to derive nonexistence of the extreme rays of $\Sigma^*_{\mathbf{n},\mathbf{2d}}$ with kernel W_ℓ, thus proving the equality between nonnegative forms and sums of squares.

Let W be a linear subspace of $\mathbb{R}[x]_{\mathbf{d}}$ and define $W^{\langle 2\rangle}$ to be the degree $2d$ part of the ideal generated by W:

$$W^{\langle 2\rangle} = \langle W\rangle_{2d}.$$

We use $V_{\mathbb{C}}(W)$ to denote the set of common zeroes (real and complex) of forms in W.

We next show that there is a strong relation between the linear functional ℓ and the kernel W_ℓ of the quadratic form Q_ℓ. Namely, we show that ℓ vanishes on all of $W_\ell^{\langle 2\rangle}$:

$$\ell(p) = 0 \quad \text{for all} \quad p \in W_\ell^{\langle 2\rangle}. \tag{4.4}$$

We will write the condition (4.4) as $\ell(W_\ell^{\langle 2\rangle}) = 0$ for short. We also now show that W_ℓ is the maximal subspace among all W such that $\ell(W^{\langle 2\rangle}) = 0$.

Lemma 4.42. *Let Q_ℓ be a quadratic form in $\Sigma^*_{\mathbf{n},\mathbf{2d}}$ and let $W_\ell \subset \mathbb{R}[x]_{\mathbf{d}}$ be the kernel of Q_ℓ. Then $p \in W_\ell$ if and only if $\ell(pq) = 0$ for all $q \in \mathbb{R}[x]_{\mathbf{d}}$.*

Proof. In order to investigate W_ℓ, we need to define the associated bilinear form B_ℓ:

$$B_\ell(p,q) = \frac{Q_\ell(p+q) - Q_\ell(p) - Q_\ell(q)}{2} \quad \text{for} \quad p,q \in \mathbb{R}[x]_{\mathbf{d}}.$$

By definition of Q_ℓ we have $Q_\ell(p) = \ell(p^2)$. Therefore it follows that

$$B_\ell(p,q) = \ell(pq).$$

A form $p \in \mathbb{R}[x]_{\mathbf{d}}$ is in the kernel of Q_ℓ if and only if $B_\ell(p,q) = 0$ for all $q \in \mathbb{R}[x]_{\mathbf{d}}$. Since $B_\ell(p,q) = \ell(pq)$, the lemma follows. $\square$

We note that $V_{\mathbb{C}}(W) = \emptyset$ implies that the dimension of W_ℓ is at least n and we can find forms $p_1, \ldots, p_n \in W_\ell$ that have no common zeroes. We need a dimensional lemma from algebraic geometry which we will use without proof.

Lemma 4.43. *Suppose that $p_1, \ldots, p_n \in \mathbb{R}[x]_{\mathbf{d}}$ are forms such that $V_{\mathbb{C}}(p_1, \ldots, p_n) = \emptyset$ and let $I = \langle p_1, \ldots, p_n \rangle$ be the ideal generated by the forms p_i. Then*

$$\dim I_{2d} = n \cdot \dim \mathbb{R}[x]_{\mathbf{d}} - \binom{n}{2}.$$

Remark 4.44. *The forms $p_1, \ldots, p_n \in \mathbb{R}[x]_{\mathbf{d}}$ such that $V_{\mathbb{C}}(p_1, \ldots, p_n) = \emptyset$ form a complete intersection. The dimensional information of the ideal $\langle p_1, \ldots, p_n \rangle$ is well understood via the Koszul complex. The statement of Lemma* 4.43 *is an easy consequence of the powerful techniques developed for complete intersections* [12].

4.6.1 Equality Cases of Hilbert's Theorem

We have obtained enough information on the dual cone $\Sigma^*_{\mathbf{n,2d}}$ to give a unified proof of the equality cases of Hilbert's theorem.

Proof of equality cases in Hilbert's theorem. Suppose that $\Sigma_{\mathbf{n,2d}} \neq P_{\mathbf{n,2d}}$. Then there exists an extreme ray of $\Sigma^*_{\mathbf{n,2d}}$ that does not come from point evaluation. Let ℓ be such an extreme ray and let W_ℓ be the kernel of Q_ℓ. By Lemma 4.41 it follows that $\operatorname{rank} Q_\ell > 1$, and therefore by Corollary 4.40 we see that $V_{\mathbb{C}}(W_\ell) = \emptyset$.

Therefore $\dim W_\ell \geq n$ and we can find forms $p_1, \ldots, p_n \in W_\ell$ such that $V_{\mathbb{C}}(p_1, \ldots, p_n) = \emptyset$. Let $I = \langle p_1, \ldots, p_n \rangle$ be the ideal generated by p_i. It follows that $W_\ell^{\langle 2 \rangle}$ includes I_{2d} and $\dim I_{2d} = n \cdot \dim \mathbb{R}[x]_{\mathbf{d}} - \binom{n}{2}$ by Lemma 4.43. Therefore we see that

$$\dim W_\ell^{\langle 2 \rangle} \geq n \cdot \dim \mathbb{R}[x]_{\mathbf{d}} - \binom{n}{2}.$$

However, by (4.4) we must also have

$$\dim W_\ell^{\langle 2 \rangle} \leq \dim \mathbb{R}[x]_{\mathbf{2d}} - 1,$$

since a nontrivial linear functional $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ vanishes on $W_\ell^{\langle 2\rangle}$. We now go case by case and derive a contradiction from these dimensional facts in each of the equality cases.

Suppose that $n = 2$. Then $\dim \mathbb{R}[x]_{\mathbf{2,d}} = d + 1$ and thus $\dim W_\ell^{\langle 2\rangle} \geq 2(d + 1) - 1 = 2d + 1 = \dim \mathbb{R}[x]_{\mathbf{2,2d}}$, which is a contradiction.

Suppose that $2d = 2$. Then $\dim \mathbb{R}[x]_{\mathbf{n,1}} = n$ and $\dim W_\ell^{\langle 2\rangle} \geq n^2 - \binom{n}{2} = \binom{n+1}{2} = \dim \mathbb{R}[x]_{\mathbf{n,2}}$, leading to the same contradiction.

Finally suppose that $n = 3$ and $2d = 4$. Then $\dim \mathbb{R}[x]_{\mathbf{3,2}} = 6$ and $\dim W_\ell^{\langle 2\rangle} \geq 6 \cdot 3 - \binom{3}{2} = 15 = \dim \mathbb{R}[x]_{\mathbf{3,4}}$, which again leads to the same dimensional contradiction. □

We now turn our attention to the structure of extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ in the smallest cases where there exist nonnegative polynomials that are not sums of squares: 3 variables, degree 6, and 4 variables, degree 4.

4.7 Ranks of Extreme Rays of $\Sigma^*_{3,6}$ and $\Sigma^*_{4,4}$

We first examine, in the cases $(3, 6)$ and $(4, 4)$, the structure of linear functionals $\ell \in \mathbb{R}[x]^*_{\mathbf{2d}}$ with a given kernel W such that $V_\mathbb{C}(W) = \emptyset$.

Proposition 4.45. *Let W be a three-dimensional subspace of $\mathbb{R}[x]_{\mathbf{3,3}}$ such that $V_\mathbb{C}(W) = \emptyset$. Then $\dim W^{\langle 2\rangle} = 27$ and there exists a unique quadratic form $Q_\ell \in \mathbb{R}[x]_{\mathbf{3,6}}$ containing W in its kernel. Furthermore $\ker Q_\ell = W$.*

Before we prove Proposition 4.45 we note that the unique form Q_ℓ with kernel W need not be positive semidefinite. The investigation of positive definiteness of Q_ℓ will lead us to evaluation on finite point sets in the next section.

***Proof of Proposition* 4.45.** By applying Lemma 4.43 we see that

$$\dim W^{\langle 2\rangle} = 3 \cdot \dim \mathbb{R}[x]_{\mathbf{3,3}} - 3 = 27.$$

Since $\dim \mathbb{R}[x]_{\mathbf{3,6}} = 28$ it follows that $W^{\langle 2\rangle}$ is a hyperplane in $\mathbb{R}[x]_{\mathbf{3,6}}$ and therefore there is a unique linear functional ℓ vanishing on W. By Lemma 4.42 it follows that Q_ℓ is the unique (up to a constant multiple) quadratic form with W in its kernel.

We leave the part that the dimension of the kernel of Q_ℓ cannot be more than 3 as an exercise. □

There is also the corresponding proposition for the case $(4, 4)$ with the same proof.

Proposition 4.46. *Let W be a four-dimensional subspace of $\mathbb{R}[x]_{\mathbf{4,2}}$ such that $V_\mathbb{C}(W) = \emptyset$. Then $\dim W^{\langle 2\rangle} = 34$ and there exists a unique quadratic form $Q_\ell \in \mathbb{R}[x]_{\mathbf{4,4}}$ containing W in its kernel. Furthermore $\ker Q_\ell = W$.*

We obtain the following interesting corollaries.

Corollary 4.47. *Suppose that ℓ spans an extreme ray of $\Sigma^*_{\mathbf{3,6}}$ and ℓ does not correspond to point evaluation. Then* $\operatorname{rank} Q_\ell = 7$*. Conversely, suppose that Q_ℓ is a psd form of rank 7 in $S_+^{3,3}$ and let W_ℓ be the kernel of Q_ℓ. If $V_{\mathbb{C}}(W_\ell) = \emptyset$, then Q_ℓ spans an extreme ray of $\Sigma^*_{\mathbf{3,6}}$.*

Proof. Suppose that ℓ spans an extreme ray of $\Sigma^*_{\mathbf{3,6}}$ and ℓ does not correspond to point evaluation. Let W_ℓ be the kernel of Q_ℓ. We know that $V(W_\ell) = \emptyset$ and $\dim W_\ell \geq 3$. We can then find a three-dimensional subspace W of W_ℓ such that $V(W) = \emptyset$. Applying Proposition 4.45 we see that there exists a unique quadratic form Q containing W in its kernel. Then it must happen that Q_ℓ is a scalar multiple of Q, and since $\ker Q = W$ we see that the kernel of Q_ℓ has dimension 3 and thus Q_ℓ has rank 7.

Conversely suppose that Q_ℓ is a positive semidefinite form of rank 7 and $V_{\mathbb{C}}(W_\ell) = \emptyset$. Then by Proposition 4.45 Q_ℓ is the unique quadratic form in $\mathbb{R}[x]^*_{\mathbf{3,6}}$ with kernel W_ℓ. Suppose that $Q_\ell = Q_1 + Q_2$ with $Q_1, Q_2 \in \Sigma^*_{\mathbf{3,6}}$. Then Q_1 and Q_2 are positive semidefinite forms by Lemma 4.35 and therefore $\ker Q_\ell \subseteq \ker Q_i$. Then Q_1 and Q_2 are scalar multiples of Q_ℓ and therefore Q_ℓ spans an extreme ray of $\Sigma^*_{\mathbf{3,6}}$. □

The above corollary has a couple of interesting consequences. If the quadratic form Q_ℓ is in $\Sigma^*_{\mathbf{3,6}}$ and its rank is at most 6, then it must be a convex combination of rank 1 forms in $\Sigma^*_{\mathbf{3,6}}$, which we know are point evaluations. Restated in measure and moment language, this says that if a positive semidefinite moment matrix in $\mathbb{R}[x]^*_{\mathbf{3,6}}$ has rank at most 6, then the linear functional can be written as a combination of point evaluations, and therefore the linear functional has a representing measure. However, there are rank 7 positive semidefinite moment matrices that do not admit a representing measure.

Another consequence can be stated in optimization terms. Suppose that we would like to optimize a linear functional over a compact base of the $\Sigma^*_{\mathbf{3,6}}$. Then the point where the optimum is achieved will have rank 1 or rank 7.

Corollary 4.48. *Suppose that $p \in \Sigma_{\mathbf{3,6}}$ lies on the boundary of the cone of sums of squares and p is a strictly positive form. Then p is a sum of exactly* 3 *squares.*

Proof. Let p be as above. Since p lies in the boundary of $\Sigma_{\mathbf{3,6}}$ there exists an extreme ray ℓ of the dual cone $\Sigma^*_{\mathbf{3,6}}$ such that $\ell(p) = 0$. Now suppose that $p = \sum f_i^2$ for some $f_i \in \mathbb{R}[x]_{\mathbf{3,3}}$. It follows that $Q_\ell(f_i) = 0$ for all i, and since Q_ℓ is a positive semidefinite quadratic form, we see that all f_i lie in the kernel W_ℓ of Q_ℓ. By Corollary 4.47 we know that $\dim W_\ell = 3$ and therefore p is a sum of squares of forms coming from a three-dimensional subspace of $\mathbb{R}[x]_{\mathbf{3,3}}$. It follows that p is a sum of at most 3 squares. Since any two ternary cubics have a common real zero and p is strictly positive, it follows that p cannot be a sum of two or fewer squares. □

The equivalent corollaries hold for the case $(4, 4)$, although the proof of Corollary 4.50 requires slightly more work, while the proof of Corollary 4.49 is exactly the same. For complete details see [7].

Corollary 4.49. *Suppose that ℓ spans an extreme ray of $\Sigma^*_{\mathbf{4,4}}$ and ℓ does not correspond to point evaluation. Then* $\operatorname{rank} Q_\ell = 6$. *Conversely, suppose that Q_ℓ is a positive semidefinite form of rank* 6 *in $S^{4,2}_+$ and let W_ℓ be the kernel of Q_ℓ. If $V_{\mathbb{C}}(W_\ell) = \emptyset$, then Q_ℓ spans an extreme ray of $\Sigma^*_{\mathbf{4,4}}$.*

Corollary 4.50. *Suppose that $p \in \Sigma_{\mathbf{4,4}}$ lies on the boundary of the cone of sums of squares and p is a strictly positive form. Then p is a sum of exactly* 4 *squares.*

Corollaries 4.48 and 4.50 were used to study the algebraic boundary of the cones $\Sigma_{\mathbf{3,6}}$ and $\Sigma_{\mathbf{4,4}}$ in [8].

Exercise 4.51. Show that all forms in $\mathbb{R}[x]_{\mathbf{3,6}}$ that can be written as linear combinations of squares of 3 cubics form an irreducible hypersurface in $\mathbb{R}[x]_{\mathbf{3,6}}$. Similarly, show that all forms in $\mathbb{R}[x]_{\mathbf{4,4}}$ that are linear combinations of squares of 4 quadratics also form an irreducible hypersurface in $\mathbb{R}[x]_{\mathbf{4,4}}$. (Hint: Use Terracini's lemma.) Use Corollaries 4.48 and 4.50 to show that the algebraic boundary of $\Sigma_{\mathbf{3,6}}$ and $\Sigma_{\mathbf{4,4}}$ has a single component in addition to the discriminant hypersurface.

It was shown in [8] that despite their simple definition the hypersurfaces of Exercise 4.51 have very high degree: 83200 in the case $(3, 6)$ and 38475 in the case $(4, 4)$. This shows that the boundary of the cone of sums of squares is quite complicated from the algebraic point of view.

4.8 Extracting Finite Point Sets

We have established in the previous section that the "interesting" extreme rays of $\Sigma^*_{\mathbf{3,6}}$ have rank 7 and those of $\Sigma^*_{\mathbf{4,4}}$ have rank 6. Let's consider the case of 4 variables of degree 4. We have shown that a four-dimensional subspace W leads to a unique form Q_ℓ of rank 6 such that the kernel of Q_ℓ contains W. However, the form Q_ℓ does not have to lie in $\Sigma^*_{\mathbf{4,4}}$, since the form Q_ℓ is not necessarily positive semidefinite.

In order to examine positive semidefiniteness of Q_ℓ we reduce the problem to looking at an evaluation on finite point sets.

Exercise 4.52. Let W be a subspace of $\mathbb{R}[x]_{\mathbf{d}}$ such that $V_{\mathbb{C}}(W) = \emptyset$. Show that there exist forms $q_1, \ldots, q_{n-1} \in W$ that intersect in d^{n-1} projective points in $\mathbb{CP}^{n-1}$:

$$V_{\mathbb{C}}(q_1, \ldots, q_{n-1}) = \{\bar{s}_1, \ldots, \bar{s}_{d^{n-1}} \;\;|\;\; \bar{s}_i \in \mathbb{CP}^{n-1}\}.$$

We apply this result to our case of $W \subset \mathbb{R}[x]_{\mathbf{4,4}}$ and obtain forms $q_1, q_2, q_3 \in W$ intersecting in $2^3 = 8$ projective points $\bar{s}_i \in \mathbb{CP}^3$. We can take their affine representatives $s_1, \ldots, s_8 \in \mathbb{C}^n$. Unfortunately, even though the forms $q_i \in W$ are real, their points of intersection may be complex.

However, as was shown in [7], the fact that the form Q_ℓ is positive semidefinite restricts the number of complex zeroes. Since complex zeroes of real forms come in conjugate pairs, the fewest number of complex zeroes that the forms q_i may have is 2.

Theorem 4.53. *Suppose that $\ell \in \mathbb{R}[x]^*_{\mathbf{4,4}}$ is an extreme ray of $\Sigma^*_{\mathbf{4,4}}$ that does not correspond to point evaluation and let W_ℓ be the kernel of Q_ℓ. Let $q_1, q_2, q_3 \in W_\ell$ be any three forms intersecting in $2^3 = 8$ projective points in $\mathbb{CP}^3$. Then the forms q_i have at most 2 common complex zeroes. Conversely, given $q_1, q_2, q_3 \in \mathbb{R}[x]_{\mathbf{4,2}}$ intersecting in 8 points with at most 2 of them complex, there exists an extreme ray of $\Sigma^*_{\mathbf{4,4}}$ whose kernel contains q_1, q_2, q_3.*

There is an equivalent theorem for the case $(3, 6)$.

Theorem 4.54. *Suppose that $\ell \in \mathbb{R}[x]^*_{\mathbf{3,6}}$ is an extreme ray of $\Sigma^*_{\mathbf{3,6}}$ that does not correspond to point evaluation and let W_ℓ be the kernel of Q_ℓ. Let $q_1, q_2 \in W_\ell$ be any two forms intersecting in $3^2 = 9$ projective points in $\mathbb{CP}^2$. Then the forms q_i have at most 2 common complex zeroes. Conversely, given $q_1, q_2, q_3 \in \mathbb{R}[x]_{\mathbf{3,3}}$ intersecting in 9 points with at most 2 of them complex, there exists an extreme ray of $\Sigma^*_{\mathbf{3,6}}$ whose kernel contains q_1, q_2.*

It is possible to apply the Cayley–Bacharach machinery explained in Section 4.5 to completely describe the structure of the extreme rays of $\Sigma^*_{\mathbf{n,2d}}$ for the cases $(4, 4)$ and $(3, 6)$ using the coefficients of the unique Cayley–Bacharach relation that exists on the points of intersection of the forms q_i.

We have now come full circle, from using a finite point set to establish that there exist nonnegative forms that are not sums of squares in Section 4.3 to showing that these sets underlie all linear inequalities that separate $\Sigma_{\mathbf{n,2d}}$ from $P_{\mathbf{n,2d}}$.

4.9 Volumes

We now switch gears completely and turn to the question of the quantitative relationship between $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$. Our goal is to compare the relative sizes of the cones $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$. While the cones themselves are unbounded objects, we can take a section of each cone with the same hyperplane so that both sections are compact.

Let $L_{\mathbf{n,2d}}$ be an affine hyperplane in $\mathbb{R}[x]_{\mathbf{2d}}$ consisting of all forms with integral (average) 1 on the unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$:

$$L_{\mathbf{n,2d}} = \left\{ p \in \mathbb{R}[x]_{\mathbf{2d}} \;\middle|\; \int_{\mathbb{S}^{n-1}} p \, d\sigma = 1 \right\},$$

where σ is the rotation invariant probability measure on $\mathbb{S}^{n-1}$. Let $\bar{P}_{\mathbf{n,2d}}$ and $\bar{\Sigma}_{\mathbf{n,2d}}$ be the sections of $P_{\mathbf{n,2d}}$ and $\Sigma_{\mathbf{n,2d}}$ with $L_{\mathbf{n,2d}}$:

$$\bar{P}_{\mathbf{n,2d}} = P_{\mathbf{n,2d}} \cap L_{\mathbf{n,2d}} \quad \text{and} \quad \bar{\Sigma}_{\mathbf{n,2d}} = \Sigma_{\mathbf{n,2d}} \cap L_{\mathbf{n,2d}}.$$

Let $r^{2d} = (x_1^2 + \cdots + x_n^2)^d$ be the form in $\mathbb{R}[x]_{\mathbf{2d}}$ that is constantly 1 on the unit sphere. Convex bodies $\bar{P}_{\mathbf{n,2d}}$ and $\bar{\Sigma}_{\mathbf{n,2d}}$ lie in the affine hyperplane $L_{\mathbf{n,2d}}$ of forms of integral 1 on the unit sphere. We now translate them to lie in the linear hyperplane $M_{\mathbf{n,2d}}$ of forms of integral 0 on the unit sphere by subtracting r^{2d}:

$$\tilde{P}_{\mathbf{n,2d}} = \bar{P}_{\mathbf{n,2d}} - r^{2d} = \{p \in \mathbb{R}[x]_{\mathbf{2d}} \mid p + r^{2d} \in \bar{P}_{\mathbf{n,2d}}\}$$

and

$$\tilde{\Sigma}_{\mathbf{n,2d}} = \bar{\Sigma}_{\mathbf{n,2d}} - r^{2d} = \{p \in \mathbb{R}[x]_{\mathbf{2d}} \mid p + r^{2d} \in \bar{\Sigma}_{\mathbf{n,2d}}\}.$$

The estimation of the volumes of $\tilde{P}_{\mathbf{n,2d}}$ and $\tilde{\Sigma}_{\mathbf{n,2d}}$ will be done separately. Before proceeding we make a short note on the proper way to measure the size of a convex set. Let $K \subset \mathbb{R}^n$ be a convex body. Suppose that we expand K by a constant factor α. Then the volume changes as follows:

$$\operatorname{Vol}(\alpha K) = \alpha^n \operatorname{Vol} K.$$

We would like to think of K and αK as similar in size, but if the ambient dimension n grows, then αK is significantly larger in volume. Therefore the proper measure of volume that takes care of the dimensional effects is

$$(\operatorname{Vol} K)^{\frac{1}{n}}.$$

4.9.1 Volume of Nonnegative Forms

Let $M_{\mathbf{n,2d}}$ be the linear hyperplane of forms of integral 0 on the unit sphere:

$$M_{\mathbf{n,2d}} = \left\{ p \in \mathbb{R}[x]_{\mathbf{2d}} \;\middle|\; \int_{\mathbb{S}^{n-1}} p \, d\sigma = 0 \right\}.$$

Both convex bodies $\tilde{P}_{\mathbf{n,2d}}$ and $\tilde{\Sigma}_{\mathbf{n,2d}}$ live inside $M_{\mathbf{n,2d}}$, so our calculations will involve the unit sphere and the unit ball in $M_{\mathbf{n,2d}}$.

We equip $\mathbb{R}[x]_{\mathbf{2d}}$ with the L^2 inner product:

$$\langle p, q \rangle = \int_{\mathbb{S}^{n-1}} pq \, d\sigma.$$

We note that with this metric we have

$$\|p\|^2 = \langle p, p \rangle = \int_{\mathbb{S}^{n-1}} p^2 \, d\sigma = \|p\|_2^2.$$

We also let $||p||_\infty$ denote the L^∞-norm of p:

$$||p||_\infty = \max_{x \in \mathbb{S}^{n-1}} |p(x)|.$$

Let N be the dimension of $M_{\mathbf{n,2d}}$. Since $M_{\mathbf{n,2d}}$ is a hyperplane in $\mathbb{R}[x]_{\mathbf{2d}}$ we know that $N = \dim \mathbb{R}[x]_{\mathbf{2d}} - 1 = \binom{n+2d-1}{2d} - 1$. Let $\mathbb{S}^{N-1}$ and B^N denote the unit sphere and the unit ball in $M_{\mathbf{n,2d}}$ with respect to the L^2 inner product.

Our goal is to show the following estimate on the volume of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$.

Theorem 4.55.

$$\left(\frac{\operatorname{Vol}\tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \geq \frac{1}{2\sqrt{4d+2}}n^{-1/2}.$$

We first develop a general way of estimating the volume of a convex set, starting from simply writing out the integral for the volume in polar coordinates. We refer to [11] for the relevant analytic inequalities.

Exercise 4.56. Let $K \subset \mathbb{R}^n$ be a convex body with the origin in its interior and let χ_K be the characteristic function of K: $\chi_K(x) = 1$ if $x \in K$ and $\chi_K(x) = 0$ otherwise. The volume of K is given by the following integral:

$$\operatorname{Vol}K = \int_{\mathbb{R}^n} \chi_K \, d\mu,$$

where μ is the Lebesgue measure.

Let G_K be the gauge of K. Rewrite the above integral in polar coordinates to show that

$$\frac{\operatorname{Vol}K}{\operatorname{Vol}B^n} = \int_{\mathbb{S}^{n-1}} G_K^{-n} d\sigma,$$

where B^n and $\mathbb{S}^{n-1}$ are the unit ball and the unit sphere in $\mathbb{R}^n$ and σ is the rotation invariant probability measure on $\mathbb{S}^{n-1}$.

Exercise 4.57. Use Exercise 4.56 and Hölder's inequality to show that

$$\left(\frac{\operatorname{Vol}K}{\operatorname{Vol}B^n}\right)^{1/n} \geq \int_{\mathbb{S}^{n-1}} G_K^{-1} d\sigma.$$

Exercise 4.58. Use Exercise 4.57 and Jensen's inequality to show that

$$\left(\frac{\operatorname{Vol}K}{\operatorname{Vol}B^n}\right)^{1/n} \geq \left(\int_{\mathbb{S}^{n-1}} G_K d\sigma\right)^{-1}.$$

Now we apply the results of Exercises 4.56–4.58 to the case of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$.

Lemma 4.59.

$$\left(\frac{\operatorname{Vol}\tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \geq \left(\int_{\mathbb{S}^{N-1}} ||p||_\infty \, d\sigma_p\right)^{-1}.$$

Proof. We observe that $\bar{P}_{\mathbf{n},\mathbf{2d}}$ consists of all forms of integral 1 on $\mathbb{S}^{n-1}$ whose minimum on $\mathbb{S}^{n-1}$ is at least 0. Therefore $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ consists of all forms of integral 0 on $\mathbb{S}^{n-1}$ whose minimum on the unit sphere is at least -1:

$$\tilde{P}_{\mathbf{n},\mathbf{2d}} = \left\{ p \in \mathbb{R}[x]_{\mathbf{2d}} \;\middle|\; \int_{\mathbb{S}^{n-1}} p\, d\sigma = 0 \text{ and } \min_{x \in \mathbb{S}^{n-1}} p(x) \geq -1 \right\}.$$

It follows that the gauge of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ is given by $-\min_{\mathbb{S}^{n-1}}$:

$$G_{\tilde{P}_{\mathbf{n},\mathbf{2d}}}(p) = -\min_{x \in \mathbb{S}^{n-1}} p(x). \tag{4.5}$$

Using Exercise 4.58 we can bound the volume of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ from below:

$$\left(\frac{\operatorname{Vol} \tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol} B^N} \right)^{1/N} \geq \left(\int_{\mathbb{S}^{N-1}} -\min(p)\, d\sigma_p \right)^{-1}.$$

Since $-\min_{x \in \mathbb{S}^{n-1}} p(x)$ is bounded above by $||p||_\infty$ we obtain

$$\left(\frac{\operatorname{Vol} \tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol} B^N} \right)^{1/N} \geq \left(\int_{\mathbb{S}^{N-1}} ||p||_\infty \; d\sigma_p \right)^{-1}$$

as desired. □

From Lemma 4.59 we see that in order to obtain a lower bound on the volume of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ we need to find an upper bound on the average L^∞-norm of forms in $\mathbb{S}^{N-1}$:

$$\int_{\mathbb{S}^{N-1}} ||p||_\infty \; d\sigma_p.$$

It is easy to see that the L^∞-norm of any polynomial is bounded from below by any of its L^{2k}-norms:

$$||p||_\infty \geq \|p\|_{2k}$$

for all k. Finding upper bounds on the L^∞-norm of forms in $\mathbb{R}[x]_{\mathbf{2d}}$ in terms of their L^{2k}-norms is significantly more challenging.

Exercise 4.60. It was shown by Barvinok in [3] that the following inequality holds for all $p \in \mathbb{R}[x]_{\mathbf{2d}}$ and all k:

$$||p||_\infty \leq \binom{2kd + n - 1}{2kd}^{\frac{1}{2k}} \|p\|_{2k}.$$

Show that for $k = n$ we have

$$||p||_\infty \leq 2\sqrt{2d+1}\|p\|_{2n}$$

for all $p \in \mathbb{R}[x]_{\mathbf{2d}}$.

Remark 4.61. *It is possible to obtain slightly better bounds for our purposes by using $k = n\log(2d+1)$ in the above inequality. See [4] for details.*

We use Barvinok's inequality to convert the problem of bounding the average L^∞-norm on $\mathbb{S}^{N-1}$ into bounding the average L^{2n}-norm. In order for this to be useful we need lower bounds on the average L^{2k}-norms. We will show the following bound.

Lemma 4.62.

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p \leq \sqrt{2k}.$$

Before we proceed with the proof we need some preliminary results.

Exercise 4.63. Let Γ denote the gamma function. Show that for $k \in \mathbb{N}$

$$\int_{\mathbb{S}^{n-1}} x_1^{2k}\, d\sigma = \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(k+\frac{n}{2}\right)}. \tag{4.6}$$

Now let $\ell : \mathbb{R}^n \to \mathbb{R}$ be a linear form given by $\ell(x) = \langle x, \xi\rangle$ for some vector $\xi \in \mathbb{R}^n$. Use (4.6) to show that

$$\int_{\mathbb{S}^{n-1}} \ell^{2k}(x)\, d\sigma_x = \|\xi\|^{2k} \frac{\Gamma\left(\frac{n}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(k+\frac{n}{2}\right)}. \tag{4.7}$$

In order to apply the result of Exercise 4.63 we will need to know the L^2-norm of a special form in $M_{\mathbf{n},\mathbf{2d}}$.

Lemma 4.64. *Let $v \in \mathbb{S}^{n-1}$ be a unit vector and let $\xi_v \in M_{\mathbf{n},\mathbf{2d}}$ be the form such that*

$$\langle p, \xi_v\rangle = p(v) \quad \textit{for all} \quad p \in M_{\mathbf{n},\mathbf{2d}}.$$

Then

$$\|\xi_v\| = \sqrt{\dim M_{\mathbf{n},\mathbf{2d}}} = \sqrt{\binom{n+2d-1}{2d} - 1}.$$

Proof. Consider the following average:

$$\int_{\mathbb{S}^{N-1}} p^2(v)\, d\sigma_p = \int_{\mathbb{S}^{N-1}} \langle p, \xi_v\rangle^2\, d\sigma_p.$$

On one hand it is the average of a quadratic form on the unit sphere and by Exercise 4.63 we have

$$\int_{\mathbb{S}^{N-1}} p^2(v)\, d\sigma_p = \frac{\|\xi_v\|^2}{\dim M_{\mathbf{n},\mathbf{2d}}}.$$

On the other hand, by symmetry, this average is independent of the choice of $v \in \mathbb{S}^{n-1}$. Therefore we may introduce an extra average over the unit sphere:

$$\int_{\mathbb{S}^{N-1}} p^2(v)\, d\sigma_p = \int_{\mathbb{S}^{n-1}} \int_{\mathbb{S}^{N-1}} p^2(v)\, d\sigma_p\, d\sigma_v.$$

Now we switch the order of integration:

$$\int_{\mathbb{S}^{N-1}} p^2(v)\, d\sigma_p = \int_{\mathbb{S}^{N-1}} \int_{\mathbb{S}^{n-1}} p^2(v)\, d\sigma_v\, d\sigma_p.$$

We observe that $\int_{\mathbb{S}^{n-1}} p^2(v)\, d\sigma_v = 1$ for all $p \in \mathbb{S}^{N-1}$ and therefore

$$\int_{\mathbb{S}^{N-1}} p^2(v)\, d\sigma_p = 1.$$

The lemma now follows. □

We are now ready to estimate the average L^{2k}-norm on $\mathbb{S}^{N-1}$.

***Proof of Lemma* 4.62.**

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p = \int_{\mathbb{S}^{N-1}} \left(\int_{\mathbb{S}^{n-1}} p^{2k}(x)\, d\sigma_x \right)^{\frac{1}{2k}} d\sigma_p.$$

By applying the Hölder inequality we can move the exponent $\frac{1}{2k}$ outside and obtain

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p \le \left(\int_{\mathbb{S}^{N-1}} \int_{\mathbb{S}^{n-1}} p^{2k}(x)\, d\sigma_x\, d\sigma_p \right)^{\frac{1}{2k}}.$$

Now we exchange the order of integration:

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p \le \left(\int_{\mathbb{S}^{n-1}} \int_{\mathbb{S}^{N-1}} p^{2k}(x)\, d\sigma_p\, d\sigma_x \right)^{\frac{1}{2k}}.$$

Consider the inner integral

$$\int_{\mathbb{S}^{N-1}} p^{2k}(x)\, d\sigma_p. \tag{4.8}$$

By rotational invariance it does not depend on the choice of the point $x \in \mathbb{S}^{n-1}$. Therefore the outer integral over $\mathbb{S}^{n-1}$ is redundant and we obtain

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p \le \left(\int_{\mathbb{S}^{N-1}} p^{2k}(v)\, d\sigma_p \right)^{\frac{1}{2k}} \quad \text{for any} \quad v \in \mathbb{S}^{n-1}. \tag{4.9}$$

We can rewrite this as

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p \leq \left(\int_{\mathbb{S}^{N-1}} \langle p, \xi_v\rangle^{2k}\, d\sigma_p\right)^{\frac{1}{2k}}.$$

Now we see that the integral in (4.8) is actually just the average of the $2k$th power of a linear form and we can apply Exercise 4.63 to see that

$$\int_{\mathbb{S}^{N-1}} \langle p, \xi_v\rangle^{2k}\, d\sigma_p = \|\xi_v\|^{2k} \frac{\Gamma\left(\frac{N}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(k+\frac{N}{2}\right)}.$$

By Lemma 4.64 we know that $\|\xi_v\|^2 = \dim M_{\mathbf{n},\mathbf{2d}} = N$.

Putting it all together with (4.9) we see that

$$\int_{\mathbb{S}^{N-1}} \|p\|_{2k}\, d\sigma_p \leq \sqrt{N}\left(\frac{\Gamma\left(\frac{N}{2}\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(k+\frac{N}{2}\right)}\right)^{\frac{1}{2k}}.$$

We now use the following two estimates to finish the proof:

$$\left(\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(k+\frac{N}{2}\right)}\right)^{\frac{1}{2k}} \leq \sqrt{\frac{2}{N}} \qquad \text{and} \qquad \left(\frac{\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\right)^{\frac{1}{2k}} \leq \sqrt{k}.$$

We remark that asymptotically the second estimate is an overestimate by a factor of $\sqrt{e}$. ∎

***Proof of Theorem* 4.55.** We first use Lemma 4.59 to see that

$$\left(\frac{\operatorname{Vol}\tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \geq \left(\int_{\mathbb{S}^{N-1}} ||p||_\infty\, d\sigma_p\right)^{-1}.$$

By Exercise 4.60 we know that for all $p \in \mathbb{R}[x]_{\mathbf{2d}}$

$$||p||_\infty \leq 2\sqrt{2d+1}\|p\|_{2n}.$$

Therefore we see that

$$\left(\frac{\operatorname{Vol}\tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \geq \frac{1}{2\sqrt{2d+1}}\left(\int_{\mathbb{S}^{N-1}} \|p\|_{2n}\, d\sigma_p\right)^{-1}.$$

Now we can apply Lemma 4.62 with $k = n$ and obtain

$$\left(\frac{\operatorname{Vol}\tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \geq \frac{1}{2\sqrt{4d+2}} n^{-1/2}$$

as desired. ∎

4.9.2 Volume of Sums of Squares

We now turn our attention to the cone of sums of squares $\Sigma_{\mathbf{n,2d}}$. Although it will be somewhat obscured by our presentation, the main reason for our ability to derive bounds on the volume of $\tilde{\Sigma}_{\mathbf{n,2d}}$ comes from the fact that the dual cone $\Sigma^*_{\mathbf{n,2d}}$ is a section of the cone of positive semidefinite matrices.

We have just seen how to derive lower bounds on the volume of the cone of nonnegative forms. These bounds, of course, apply to quadratic forms, and they can be extended to work for sections of the cone. This gives us a lower bound on the volume of the dual cone, which can be turned around into an upper bound on the volume of $\tilde{\Sigma}_{\mathbf{n,2d}}$. The approach to bounding the volume of $\tilde{\Sigma}_{\mathbf{n,2d}}$ is therefore very similar to what we did for nonnegative forms. In fact, the technique in the proofs of the main bounds in Lemma 4.70 and Lemma 4.62 is nearly identical.

Let D be the dimension of $\mathbb{R}[x]_{\mathbf{d}}$. Our main result on the volume of $\tilde{\Sigma}_{\mathbf{n,2d}}$ is as follows.

Theorem 4.65.

$$\left(\frac{\operatorname{Vol}\tilde{\Sigma}_{\mathbf{n,2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \le 2^{4d+1}\sqrt{\frac{6D}{N}}.$$

Remark 4.66. *Recall that*

$$N = \binom{n+2d-1}{2d} - 1 \quad \textit{and} \quad D = \binom{n+d-1}{d}.$$

Therefore, for fixed degree d our upper bound on the volume of $\tilde{\Sigma}_{\mathbf{n,2d}}$ is of the order $n^{-d/2}$. In Theorem 4.55 *we proved a lower bound on the volume of $\tilde{P}_{\mathbf{n,2d}}$ that is of the order $n^{-1/2}$. Therefore, when the total degree* 2d *is at least* 4*, the lower bound on the volume of $\tilde{P}_{\mathbf{n,2d}}$ is asymptotically much larger than the upper bound on the volume of $\tilde{\Sigma}_{\mathbf{n,2d}}$. Thus we see that if the degree* 2d *is fixed and at least* 4*, there are significantly more nonnegative forms than sums of squares.*

It is possible to show that the bounds of Theorems 4.55 *and* 4.65 *are asymptotically tight for the case of fixed degree* 2d*. See* [5] *for more details.*

In Exercises 4.56–4.58 we showed how to bound the volume of a convex body K from below using the average of its gauge over the unit sphere $\mathbb{S}^{n-1}$. As we explained above, we are now dealing with the dual situation, and we need a related dual inequality that bounds the volume of K from above by the average gauge of its dual body K°.

Exercise 4.67. Let $K \subset \mathbb{R}^n$ be a convex body with 0 in its interior and let K° be the dual convex body defined as

$$K^\circ = \{x \in \mathbb{R}^n \quad | \quad \langle x, y\rangle \le 1 \quad \text{for all} \quad y \in K\}.$$

Show that the gauge of K° is given by the following formula:

$$G_{K^\circ}(x) = \max_{y\in K}\langle x, y\rangle.$$

The following is known as Urysohn's inequality [26].

Lemma 4.68.

$$\left(\frac{\operatorname{Vol} K}{\operatorname{Vol} B^n}\right)^{1/n} \leq \int_{\mathbb{S}^{n-1}} G_{K^\circ}(x)\, d\sigma_x.$$

In order to apply Lemma 4.68 we need a description of the gauge of $\tilde{\Sigma}^*_{\mathbf{n,2d}}$. Let $\mathbb{S}^{D-1}$ be the unit sphere in $\mathbb{R}[x]_{\mathbf{d}}$ with respect to the L^2 inner product.

Lemma 4.69. *We have the following description of the gauge of* $\tilde{\Sigma}^\circ_{\mathbf{n,2d}}$*:*

$$G_{\tilde{\Sigma}^\circ_{\mathbf{n,2d}}}(p) = \max_{q\in\mathbb{S}^{D-1}} \langle p, q^2\rangle.$$

Proof. By Exercise 4.67 the gauge of $\tilde{\Sigma}^\circ_{\mathbf{n,2d}}$ is given by

$$G_{\tilde{\Sigma}^\circ_{\mathbf{n,2d}}}(p) = \max_{q\in\tilde{\Sigma}_{\mathbf{n,2d}}} \langle p, q\rangle.$$

We observe that the maximal inner product $\max_{q\in\tilde{\Sigma}_{\mathbf{n,2d}}}\langle p, q\rangle$ always occurs at an extreme point of $\tilde{\Sigma}_{\mathbf{n,2d}}$. Extreme points of $\bar{\Sigma}_{\mathbf{n,2d}}$ are all squares, and therefore extreme point of $\tilde{\Sigma}_{\mathbf{n,2d}}$ are translates of squares and have the form

$$q^2 - r^{2d} \quad \text{with} \quad q \in \mathbb{R}[x]_{\mathbf{d}} \quad \text{and} \quad \int_{\mathbb{S}^{n-1}} q^2\, d\sigma = 1.$$

The condition $\int_{\mathbb{S}^{n-1}} q^2\, d\sigma = 1$ corresponds exactly to q lying in the unit sphere of $\mathbb{R}[x]_{\mathbf{d}}$. Since forms $p \in M_{\mathbf{n,2d}}$ have integral zero on the unit sphere $\mathbb{S}^{n-1}$, it follows that

$$\langle p, r^{2d}\rangle = 0 \quad \text{for all} \quad p \in M_{\mathbf{n,2d}}.$$

Combining with the description of the extreme points of $\tilde{\Sigma}_{\mathbf{n,2d}}$ we see that

$$G_{\tilde{\Sigma}^\circ_{\mathbf{n,2d}}}(p) = \max_{q\in\mathbb{S}^{D-1}} \langle p, q^2\rangle. \qquad \square$$

Given a form $p \in \mathbb{R}[x]_{\mathbf{2d}}$ we define the associated quadratic form Q_p on $\mathbb{R}[x]_{\mathbf{d}}$:

$$Q_p(q) = \langle p, q^2\rangle \quad \text{for} \quad q \in \mathbb{R}[x]_{\mathbf{d}}.$$

By Lemma 4.69 we see that the gauge of $\tilde{\Sigma}_{\mathbf{n,2d}}$ is given by the maximum of Q_p on the unit sphere $\mathbb{S}^{D-1}$ in $\mathbb{R}[x]_{\mathbf{d}}$:

$$G_{\tilde{\Sigma}_{\mathbf{n,2d}}}(p) = \max_{q\in\mathbb{S}^{D-1}} Q_p(q).$$

Since Q_p is a quadratic form on $\mathbb{R}[x]_{\mathbf{d}}$, its L^∞-norm is the maximal value it takes on the unit sphere $\mathbb{S}^{D-1}$:

$$||Q_p||_\infty = \max_{q\in\mathbb{S}^{D-1}} |Q_p(q)|.$$

Applying Lemma 4.68 we see that

$$\left(\frac{\operatorname{Vol}\tilde{\Sigma}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \le \int_{\mathbb{S}^{N-1}} ||Q_p||_\infty \, d\sigma_p.$$

Now we can apply Barvinok's inequality to bound $||Q_p||_\infty$ by high L^{2k}-norms. Using Exercise 4.60 with $k = D$ we see that

$$||Q_p||_\infty \le 2\sqrt{3}\|Q_p\|_{2D}.$$

Therefore we obtain

$$\left(\frac{\operatorname{Vol}\tilde{\Sigma}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol}B^N}\right)^{1/N} \le 2\sqrt{3}\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \, d\sigma_p.$$

The proof is now finished with the following estimate, which proceeds in nearly the same way as the proof of Lemma 4.62.

Lemma 4.70.

$$\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \, d\sigma_p \le 2^{4d}\sqrt{\frac{2D}{N}}.$$

Proof. We first write out the integral we would like to estimate:

$$\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \, d\sigma_p = \int_{\mathbb{S}^{N-1}} \left(\int_{\mathbb{S}^{D-1}} \langle p, q^2\rangle^{2D} \, d\sigma_q\right)^{1/2D} d\sigma_p.$$

Using the Hölder inequality we move the exponent $1/2D$ outside:

$$\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \, d\sigma_p \le \left(\int_{\mathbb{S}^{N-1}} \int_{\mathbb{S}^{D-1}} \langle p, q^2\rangle^{2D} \, d\sigma_q \, d\sigma_p\right)^{1/2D}.$$

Next we interchange the order of integration:

$$\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \, d\sigma_p \le \left(\int_{\mathbb{S}^{D-1}} \int_{\mathbb{S}^{N-1}} \langle p, q^2\rangle^{2D} \, d\sigma_p \, d\sigma_q\right)^{1/2D}. \tag{4.10}$$

Consider the inner integral

$$\int_{\mathbb{S}^{N-1}} \langle p, q^2\rangle^{2D} \, d\sigma_p. \tag{4.11}$$

We apply Exercise 4.63 to see that

$$\int_{\mathbb{S}^{N-1}} \langle p, q^2\rangle^{2D} \, d\sigma_p \le \|q^2\|^{2D} \frac{\Gamma\left(\frac{N}{2}\right)\Gamma\left(D+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(D+\frac{N}{2}\right)}.$$

The reason that we have an inequality, instead of equality, above is that q^2 does not lie in the hyperplane $M_{\mathbf{n},\mathbf{2d}}$, and for equality we should use the norm of the projection of q^2 into $M_{\mathbf{n},\mathbf{2d}}$. We now observe that

$$\|q^2\| = \|q\|_4^2.$$

Since q lies in the unit sphere of $\mathbb{S}^{D-1}$ it follows that $\|q\| = 1$. By a result of Duoandikoetxea in [9] we know that

$$\|q\|_4 \leq 4^{2d}\|q\|.$$

Putting it all together we get

$$\int_{\mathbb{S}^{N-1}} \langle p, q^2 \rangle^{2D} \, d\sigma_p \leq 4^{4dD} \frac{\Gamma\left(\frac{N}{2}\right)\Gamma\left(D+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(D+\frac{N}{2}\right)}.$$

We note that this estimate is independent of q and therefore the outer integral in (4.10) is redundant and we obtain

$$\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \, d\sigma_p \leq 4^{2d} \left(\frac{\Gamma\left(\frac{N}{2}\right)\Gamma\left(D+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(D+\frac{N}{2}\right)} \right)^{1/2D}.$$

As in the proof of Lemma 4.62 we use the estimates

$$\left(\frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(D+\frac{N}{2}\right)} \right)^{\frac{1}{2D}} \leq \sqrt{\frac{2}{N}} \qquad \text{and} \qquad \left(\frac{\Gamma\left(D+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)} \right)^{\frac{1}{2D}} \leq \sqrt{D}.$$

Therefore we have

$$\int_{\mathbb{S}^{N-1}} \|Q_p\|_{2D} \leq 2^{4d} \sqrt{\frac{2D}{N}}. \qquad \square$$

4.10 Convex Forms

There is another very interesting convex cone inside $\mathbb{R}[x]_{\mathbf{2d}}$, the cone of convex forms $C_{\mathbf{n},\mathbf{2d}}$. A form $p \in \mathbb{R}[x]_{\mathbf{2d}}$ is called convex if p is a convex function on $\mathbb{R}^n$:

$$p\left(\frac{x+y}{2}\right) \leq \frac{p(x)+p(y)}{2} \quad \text{for all} \quad x, y \in \mathbb{R}^n.$$

It is an easy exercise to show that $C_{\mathbf{n},\mathbf{2d}}$ is contained in the cone of nonnegative forms.

Exercise 4.71. Show that if a form $p \in \mathbb{R}[x]_{\mathbf{2d}}$ is convex, then p is nonnegative. Show that $x_1^2 x_2^2 \in P_{\mathbf{2},\mathbf{4}}$ is not convex.

The relationship between convex forms and sums of squares is significantly harder to understand. An equivalent definition of convexity is that a form $p \in \mathbb{R}[x]_{\mathbf{2d}}$ is convex if and only if its Hessian $\nabla^2 p$ is a positive semidefinite matrix on all of $\mathbb{R}^n$. We can associate with p its Hessian form H_p, which is a form in $2n$ variables, with old variables $\mathbf{x} = (x_1, \ldots, x_n)$ and new variables $\mathbf{y} = (y_1, \ldots, y_n)$. The Hessian form $H_p(x, y)$ is given by

$$H_p(x, y) = y^T \left(\nabla^2 p(x)\right) y.$$

We note that H_p is a bihomogeneous form; it is quadratic in y and of degree $2d - 2$ in x. A form p is convex if and only if its Hessian form H_p is nonnegative on $\mathbb{R}^{2n}$.

A form $p \in \mathbb{R}[x]_{\mathbf{2d}}$ is called *sos-convex* if H_p is a sum of squares. Sos-convexity is a more restrictive condition than being a sum of squares.

Exercise 4.72. Let $p \in \mathbb{R}[x]_{\mathbf{2d}}$ be an sos-convex form. Show that p is a sum of squares.

An explicit example of a convex form that is not sos-convex was constructed in [1]. We will explain below that there exist convex forms that are not sums of squares. In fact, we will show using volume arguments that asymptotically there are significantly more convex forms than sums of squares. However, it is still an open question to find an explicit example of a convex form that is not a sum of squares.

4.10.1 Volumes of Convex Forms

As before we can take a compact section of $C_{\mathbf{n},\mathbf{2d}}$ with the hyperplane $L_{\mathbf{n},\mathbf{2d}}$ of forms of integral 1 on $\mathbb{S}^{n-1}$:

$$\bar{C}_{\mathbf{n},\mathbf{2d}} = C_{\mathbf{n},\mathbf{2d}} \cap L_{\mathbf{n},\mathbf{2d}}.$$

We also let $\tilde{C}_{\mathbf{n},\mathbf{2d}}$ be $\bar{C}_{\mathbf{n},\mathbf{2d}}$ translated by subtracting r^{2d}:

$$\tilde{C}_{\mathbf{n},\mathbf{2d}} = \bar{C}_{\mathbf{n},\mathbf{2d}} - r^{2d}.$$

The convex body $\tilde{C}_{\mathbf{n},\mathbf{2d}}$ lies in the hyperplane $M_{\mathbf{n},\mathbf{2d}}$ of forms of average 0 on the unit sphere $\mathbb{S}^{n-1}$. We will show the following estimate on the volume of $\tilde{C}_{\mathbf{n},\mathbf{2d}}$ that, together with Theorems 4.55 and 4.65, implies that if the degree $2d$ is fixed and the number of variables grows then there are significantly more convex forms than sums of squares. This is the only currently known method of establishing existence of convex forms that are not sums of squares.

Theorem 4.73.

$$\left(\frac{\operatorname{Vol} \bar{C}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol} \bar{P}_{\mathbf{n},\mathbf{2d}}}\right)^{1/N} \geq \frac{1}{2(2d-1)}.$$

Remark 4.74. *From Exercise* 4.71 *it follows that* $\bar{C}_{\mathbf{n,2d}} \subseteq \bar{P}_{\mathbf{n,2d}}$. *Therefore the estimate of Theorem* 4.73 *is asymptotically tight for the case of fixed degree* $2d$.

Our first goal is to show that if a form $p \in \mathbb{R}[x]_{\mathbf{2d}}$ is sufficiently close to being constant on the unit sphere, then p must be convex.

Theorem 4.75. *Let* p *be a form in* $\mathbb{R}[x]_{\mathbf{2d}}$. *If for all* $v \in \mathbb{S}^{n-1}$

$$1 - \frac{1}{2d-1} \leq p(v) \leq 1 + \frac{1}{2d-1},$$

then p *is convex.*

For a point $\xi \in \mathbb{S}^{n-1}$ we can think of ξ as a direction. We will use

$$\frac{\partial p}{\partial \xi} = \langle \nabla p, \xi \rangle$$

to denote the derivative of p in the direction ξ. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for all $v \in \mathbb{R}^n$ and all $\xi \in \mathbb{S}^{n-1}$ we have

$$\frac{\partial^2 f}{\partial \xi^2}(v) \geq 0.$$

Since we are working with forms it suffices to restrict our attention to $v \in \mathbb{S}^{n-1}$. We use $|\nabla p|$ to denote the length of the gradient of p. We will need the following theorem of Kellogg [13].

Theorem 4.76. *Let* p *be a form in* $\mathbb{R}[x]_{\mathbf{d}}$. *For all* $v \in \mathbb{S}^{n-1}$

$$|\nabla p(v)| \leq d \, ||p||_\infty .$$

Theorem 4.76 implies that for any $v \in \mathbb{S}^{n-1}$

$$\left| \frac{\partial p}{\partial \xi}(v) \right| \leq d \, ||p||_\infty .$$

This follows since

$$\frac{\partial p}{\partial \xi} = \langle \nabla p, \xi \rangle \leq |\nabla p| \cdot |\xi| = |\nabla p|$$

by applying the Cauchy–Schwarz inequality.

We extend Theorem 4.76 to cover the case of higher derivatives, which is necessary since convexity is a condition on second derivatives:

Lemma 4.77. *Let* p *be a form in* $\mathbb{R}[x]_{\mathbf{d}}$. *For any* v *and* $\xi_1, \ldots \xi_k \in \mathbb{S}^{n-1}$

$$\left| \frac{\partial^k p}{\partial \xi_1 \cdots \partial \xi_k}(v) \right| \leq \frac{d!}{(d-k)!} \, ||p||_\infty .$$

Proof. We proceed by induction on the order of partial derivatives k. The base case $k = 1$ is covered by Theorem 4.76. Now we need to show the induction step. We assume that the statement holds for all derivatives of order at most k and consider

$$\left|\frac{\partial^{k+1} p}{\partial \xi_1 \cdots \partial \xi_{k+1}}(v)\right|$$

for some $\xi_1, \dots \xi_{k+1} \in \mathbb{S}^{n-1}$.

Let

$$q = \frac{\partial p}{\partial \xi_1}.$$

Using the base case we see that

$$||q||_\infty \le d\, ||p||_\infty. \tag{4.12}$$

Also, we know that q is a form in n variables of degree $d-1$. Therefore by the induction assumption

$$\left|\frac{\partial^k q}{\partial \xi_2 \cdots \partial \xi_{k+1}}(v)\right| \le \frac{(d-1)!}{(d-k-1)!}\, ||q||_\infty. \tag{4.13}$$

Putting together (4.12) and (4.13), the lemma follows. □

We are now ready to prove Theorem 4.75, which provides a sufficient condition for a form to be convex.

***Proof of Theorem* 4.75.** Let p be as in the statement of the theorem, and let $q = p - r^{2d}$. By the assumptions of the theorem it follows that, for all $v \in \mathbb{S}^{n-1}$,

$$-\frac{1}{2d-1} \le q(v) \le \frac{1}{2d-1}.$$

In other words

$$||q||_\infty \le \frac{1}{2d-1}.$$

Then by Lemma 4.77 we know that for any v and $\xi \in \mathbb{S}^{n-1}$

$$\left|\frac{\partial^2 q}{\partial \xi^2}(v)\right| \le 2d.$$

In particular, it follows that

$$\frac{\partial^2 q}{\partial \xi^2}(v) \ge -2d$$

for all v and $\xi \in \mathbb{S}^{n-1}$.

It is easy to check that

$$\frac{\partial^2 r^{2d}}{\partial \xi^2}(v) = 2d + 4d(d-1)\langle v, \xi\rangle^2 \geq 2d.$$

Since we know that $p = q + r^{2d}$ it follows that for all v and $\xi \in \mathbb{S}^{n-1}$

$$\frac{\partial^2 p}{\partial \xi^2}(v) \geq 0,$$

and therefore p is convex. $\square$

We need one more result from convexity to help us with the volume bounds (see [16]).

Exercise 4.78. Let K be a convex body in $\mathbb{R}^n$. The *barycenter* of K is defined to be a vector $b = (b_1, \ldots, b_n) \in K$ given by

$$b_i = \int_{\mathbb{R}^n} x_i \chi_K \ \ d\mu,$$

where χ_K is the characteristic function of K and μ is the Lebesgue measure. Let K' be the reflection of K through the barycenter b: $K' = b - (K - b)$. Show that

$$\left(\frac{\operatorname{Vol} K \cap K'}{\operatorname{Vol} K}\right)^{\frac{1}{n}} \geq \frac{1}{2}.$$

Exercise 4.79. The set $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ is a convex body in the hyperplane $M_{\mathbf{n},\mathbf{2d}}$ of all forms of integral 0 on the unit sphere. Use invariance of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ under orthogonal changes of coordinates to show that 0 is the barycenter of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$. Let $-\tilde{P}_{\mathbf{n},\mathbf{2d}}$ be the reflection of $\tilde{P}_{\mathbf{n},\mathbf{2d}}$ through the origin. Show that $\tilde{P}_{\mathbf{n},\mathbf{2d}} \cap -\tilde{P}_{\mathbf{n},\mathbf{2d}}$ consists of all forms in $M_{\mathbf{n},\mathbf{2d}}$ whose values on the unit are between -1 and 1, i.e., the forms with L^∞-norm at most 1:

$$\tilde{P}_{\mathbf{n},\mathbf{2d}} \cap -\tilde{P}_{\mathbf{n},\mathbf{2d}} = \left\{p \in M_{\mathbf{n},\mathbf{2d}} \quad | \quad ||p||_\infty \leq 1\right\}.$$

***Proof of Theorem* 4.73.** Let $K_{\mathbf{n},\mathbf{2d}}$ be the set of forms that take values only between $1 - \frac{1}{2d-1}$ and $1 + \frac{1}{2d-1}$ on the unit sphere:

$$K_{\mathbf{n},\mathbf{2d}} = \left\{p \in \mathbb{R}[x]_{\mathbf{2d}} \;\middle|\; 1 - \frac{1}{2d-1} \leq p(v) \leq 1 + \frac{1}{2d-1} \quad \text{for all} \quad v \in \mathbb{S}^{n-1}\right\}.$$

We note that $K_{\mathbf{n},\mathbf{2d}}$ is a compact convex set. We let $\bar{K}_{\mathbf{n},\mathbf{2d}}$ be the section of $K_{\mathbf{n},\mathbf{2d}}$ with $L_{\mathbf{n},\mathbf{2d}}$,

$$\bar{K}_{\mathbf{n},\mathbf{2d}} = K_{\mathbf{n},\mathbf{2d}} \cap L_{\mathbf{n},\mathbf{2d}},$$

and let $\tilde{K}_{\mathbf{n},\mathbf{2d}}$ be the translated section:

$$\tilde{K}_{\mathbf{n},\mathbf{2d}} = \bar{K}_{\mathbf{n},\mathbf{2d}} - r^{2d}.$$

It follows that $\tilde{K}_{\mathbf{n},\mathbf{2d}}$ consists of all the forms in $M_{\mathbf{n},\mathbf{2d}}$ that take values between $-\frac{1}{2d-1}$ and $\frac{1}{2d-1}$ on the unit sphere, so forms with L^∞-norm at most $\frac{1}{2d-1}$:

$$\tilde{K}_{\mathbf{n},\mathbf{2d}} = \left\{ p \in M_{\mathbf{n},\mathbf{2d}} \;\middle|\; ||p||_\infty \leq \frac{1}{2d-1} \right\}.$$

By Exercise 4.79 it follows that

$$\frac{1}{2d-1}\left(\tilde{P}_{\mathbf{n},\mathbf{2d}} \cap -\tilde{P}_{\mathbf{n},\mathbf{2d}}\right) \subseteq \tilde{K}_{\mathbf{n},\mathbf{2d}}.$$

Using Exercise 4.78 we see that

$$\left(\frac{\operatorname{Vol} \tilde{P}_{\mathbf{n},\mathbf{2d}} \cap -\tilde{P}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol} \tilde{P}_{\mathbf{n},\mathbf{2d}}}\right)^{1/N} \geq \frac{1}{2}.$$

Therefore it follows that

$$\left(\frac{\operatorname{Vol} \tilde{K}_{\mathbf{n},\mathbf{2d}}}{\operatorname{Vol} \tilde{P}_{\mathbf{n},\mathbf{2d}}}\right)^{1/N} \geq \frac{1}{2(2d-1)}.$$

On the other hand, by Theorem 4.75 we know that $\tilde{K}_{\mathbf{n},\mathbf{2d}}$ is contained in $\tilde{C}_{\mathbf{n},\mathbf{2d}}$, and the theorem follows. ∎

Bibliography

[1] A. A. Ahmadi and P. A. Parrilo. A convex polynomial that is not sos-convex. *Math. Program. Ser. A*, 135:275–292, 2012.

[2] A. Barvinok. *A Course in Convexity*. American Mathematical Society, Providence, RI, 2002.

[3] A. Barvinok. Estimating L^∞ norms by L^{2k} norms for functions on orbits. *Found. Comput. Math.*, 2:393–412, 2002.

[4] A. Barvinok and G. Blekherman. Convex geometry of orbits. In *Combinatorial and Computational Geometry*, Math. Sci. Res. Inst. Publ. 52, Cambridge University Press, Cambridge, UK, 2005, pp. 51–77.

[5] G. Blekherman. There are significantly more nonnegative polynomials than sums of squares. *Israel J. Math.*, 183:355–380, 2006.

[6] G. Blekherman. *Dimensional differences between nonnegative polynomials and sums of squares*. Submitted for publication, arXiv:0907.1339.

[7] G. Blekherman. Nonnegative polynomials and sums of squares. *J. Amer. Math. Soc.*, 25:617–635, 2012.

[8] G. Blekherman, J. Hauenstein, J. C. Ottem, K. Ranestad, and B. Sturmfels. Algebraic boundaries of Hilbert's SOS cones. To appear in *Compositio Mathematica*. arXiv:1107.1846.

[9] J. Duoandikoetxea. Reverse Hölder inequalities for spherical harmonics. Proc. Amer. Math. Soc., 101:487–491, 1987.

[10] D. Eisenbud, M. Green, and J. Harris. Cayley-Bacharach theorems and conjectures. *Bull. Amer. Math. Soc.*, 33:295–324, 1996.

[11] G. Hardy, E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, Cambridge, UK, 1988.

[12] J. Harris. *Algebraic Geometry. A First Course,* Grad. Texts in Math. 133. Springer-Verlag, New York, 1995.

[13] O. D. Kellogg. On bounded polynomials in several variables. *Math. Z.*, 27:55–64, 1928.

[14] J. M. Landsberg and Z. Teitler. On the ranks and border ranks of symmetric tensors. *Found. Comput. Math.*, 10:339–366, 2010.

[15] J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*. Imperial College Press, London, 2009.

[16] V. D. Milman and A. Pajor. Entropy and asymptotic geometry of non-symmetric convex bodies. *Adv. Math.*, 152:314–335, 2000.

[17] R. Miranda. Linear systems of plane curves. *Notices Amer. Math. Soc.*, 46:192–202, 1999.

[18] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program. Ser. B*, 96:293–320, 2000/01.

[19] P. A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In *Algorithmic and Quantitative Aspects of Real Algebraic Geometry*, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 60, American Mathmatical Society, Providence, RI, 2003, pp. 83–100.

[20] M. Ramana and A. J. Goldman. Some geometric results in semidefinite programming. *J. Global Optim.*, 7:33–50, 1995.

[21] B. Reznick, *Sums of Even Powers of Real Linear Forms*, Mem. Amer. Math. Soc. 96, American Mathematical Society, Providence, RI, 1992.

[22] B. Reznick. Some concrete aspects of Hilbert's 17th problem. In *Real Algebraic Geometry and Ordered Structures*, Contemp. Math. 253. American Mathematical Society, Providence, RI, 2000, pp. 251–272.

[23] B. Reznick. *On Hilbert's construction of positive polynomials.* Submitted for publication, arXiv:0707.2156.

[24] B. Reznick. Blenders. In *Notions of Positivity and the Geometry of Polynomials*, P. Branden, M. Passare, and M. Putinar, eds., Trends in Math. Birkhäuser, Basel, 2011, pp. 345–373.

[25] R. Sanyal, F. Sottile, and B. Sturmfels. Orbitopes. *Mathematika*, 57:275–314, 2011.

[26] R. Schneider. *Convex Bodies: The Brunn-Minkowski Theory.* Cambridge University Press, Cambridge, UK, 1993.