Abstract Cones of Positive Polynomials and Their Sums of Squares Relaxations

Roland Hildebrand[1]

LJK, Université Grenoble 1 / CNRS, 51 rue des Mathématiques, BP53, 38041 Grenoble cedex, France roland.hildebrand@imag.fr

Summary. We present a new family of sums of squares (SOS) relaxations to cones of positive polynomials. The SOS relaxations employed in the literature are cones of polynomials which can be represented as ratios, with an SOS as numerator and a fixed positive polynomial as denominator. We employ nonlinear transformations of the arguments instead. A fixed cone of positive polynomials, considered as a subset in an abstract coefficient space, corresponds to an infinite, partially ordered set of concrete cones of positive polynomials of different degrees and in a different number of variables. To each such concrete cone corresponds its own SOS cone, leading to a hierarchy of increasingly tighter SOS relaxations for the abstract cone.

1 Introduction

Many optimization problems can be recast as conic programs over a cone of positive polynomials on $\mathbf{R}^n$. Cones of positive polynomials cannot be described efficiently in general, and the corresponding conic programs are NP-hard. Hence approximations have to be employed to obtain suboptimal solutions. A standard approach is to approximate the cone of positive polynomials from inside by the cone of sums of squares (SOS), i.e. the cone of those polynomials which are representable as a sum of squares of polynomials of lower degree. The SOS cone is semidefinite representable, and conic programs over this cone can be cast as efficiently solvable semidefinite programs. This approximation is not exact, however, even for polynomials of degree 6 in two variables [6], as the famous example of the Motzkin polynomial [1] shows. Tighter approximations can be obtained when using the cone of polynomials which can be represented as ratios, with the numerator being a sum of squares of polynomials, and the denominator a fixed positive polynomial. Usually this fixed polynomial is chosen to be $\left(\sum_{k=1}^n x_k^2\right)^d$ for some integer $d > 0$ [2].

We propose another family of SOS based relaxations of cones of positive polynomials. We consider the cone of positive polynomials not as a cone of functions, but rather as a subset in an abstract coefficient space. The same

M. Diehl et al. (eds.), *Recent Advances in Optimization and its Applications in Engineering*, DOI 10.1007/978-3-642-12598-0_4, © Springer-Verlag Berlin Heidelberg 2010

abstract cone then corresponds to an infinite number of concrete cones of positive polynomials of different degrees and in a different number of variables. To each such concrete cone corresponds its own SOS cone, and these SOS cones are in general different for different realizations of the abstract cone. We present a computationally efficient criterion to compare the different SOS cones and introduce a corresponding equivalence relation and a partial order on the set of these SOS cones. This allows us to build hierarchies of increasingly tighter semidefinite relaxations for the abstract cone, and thus also for the original cone of positive polynomials. We show on the example of the cone of positive polynomials containing the Motzkin polynomial that our hierarchy of relaxations possesses the capability of being exact at a finite step.

The remainder of the contribution is structured as follows. In the next section we define notation that will be used in the paper. In Sect. 3 we define and analyze the considered cones of positive polynomials. In Sect. 4 we consider sums of squares relaxations of these cones and study their properties. In Sect. 5 we define the abstract cones of positive polynomials and their SOS relaxations and establish a hierarchical structure on the set of these relaxations. Finally, we demonstrate the developed apparatus on the example of the cone containing the Motzkin polynomial in Sect. 6.

2 Notation

For a finite set S, denote by $\# S$ the cardinality of S.

For a subset A of a real vector space V, denote by $\mathrm{cl}\, A$ the closure, by $\mathrm{int}\, A$ the interior, by $\mathrm{aff}\, A$ the affine hull, by $\mathrm{conv}\, A$ the convex hull, and by $\mathrm{con\,cl}\, A$ the set $\mathrm{cl} \cup_{\alpha \geq 0} \alpha A$. If A is a convex polytope, denote by $\mathrm{extr}\, A$ the set of its vertices.

Let $\mathcal{S}(m)$ denote the space of real symmetric matrices of size $m \times m$, and $\mathcal{S}_+(m) \subset \mathcal{S}(m)$ the cone of positive semidefinite (PSD) matrices. By I_n denote the $n \times n$ identity matrix. Let $\pi_2 : \mathbf{Z} \to \mathbf{F}_2$ be the ring homomorphism from the integers onto the field $\mathbf{F}_2 = (\{0,1\}, +, \cdot)$ (mapping even integers to 0 and odd ones to 1), and $\pi_2^n : \mathbf{Z}^n \to \mathbf{F}_2^n$ the corresponding homomorphism of the product rings, acting as $\pi_2^n : (a_1, \dots, a_n) \mapsto (\pi_2(a_1), \dots, \pi_2(a_n))$. For an integer matrix M, let $\pi_2[M]$ be the matrix obtained by element-wise application of π_2 to M. The corresponding $\mathbf{F}_2$-linear map will also be denoted by $\pi_2[M]$. For a linear map M, let $Im\, M$ be the image of M in the target space.

Let $\mathcal{A} \subset \mathbf{N}^n$ be an ordered finite set of multi-indices of length n, considered as row vectors. Denote by $\Gamma_{\mathcal{A}} = \{\sum_{\alpha \in \mathcal{A}} a_\alpha \alpha \,|\, a_\alpha \in \mathbf{Z}\ \forall\ \alpha \in \mathcal{A}\} \subset \mathbf{Z}^n$ the lattice generated by $\mathcal{A}$ in $\mathrm{aff}\, \mathcal{A}$, and let $\Gamma_{\mathcal{A}}^e \subset \Gamma_{\mathcal{A}}$ be the sublattice of even points. For $x = (x_1, \dots, x_n)^T \in \mathbf{R}^n$, denote by $X_{\mathcal{A}}(x)$ the corresponding vector of monomials $(x^\alpha)_{\alpha \in \mathcal{A}}$, and define the set $\mathcal{X}_{\mathcal{A}} = \{X_{\mathcal{A}}(x) \,|\, x \in \mathbf{R}^n\}$. By $\mathcal{L}_{\mathcal{A}}$ we denote the real vector space of polynomials $p(x) = \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha$. There exists a canonical isomorphism $\mathcal{I}_{\mathcal{A}} : \mathcal{L}_{\mathcal{A}} \to \mathbf{R}^{\# \mathcal{A}}$, which maps a polynomial $p \in \mathcal{L}_{\mathcal{A}}$ to its coefficient vector $\mathcal{I}_{\mathcal{A}}(p) = (c_\alpha(p))_{\alpha \in \mathcal{A}}$.

3 Cones of positive polynomials

Let $\mathcal{A} \subset \mathbf{N}^n$ be an ordered finite set of multi-indices. We call a polynomial $p \in \mathcal{L}_\mathcal{A}$ *positive* if $p(x) = \langle \mathcal{I}_\mathcal{A}(p), X_\mathcal{A}(x) \rangle \geq 0$ for all $x \in \mathbf{R}^n$. The positive polynomials form a closed convex cone $\mathcal{P}_\mathcal{A}$. This cone cannot contain a line, otherwise the monomials x^α, $\alpha \in \mathcal{A}$, would be linearly dependent.

Let $p \in \mathcal{L}_\mathcal{A}$ be a polynomial. The convex hull of all indices $\alpha \in \mathcal{A}$ such that $c_\alpha(p) \neq 0$, viewed as vectors in $\mathbf{R}^n$, forms a convex polytope. This polytope is called the *Newton polytope* of p and is denoted by $N(p)$. The convex hull of the whole multi-index set $\mathcal{A}$, viewed as a subset of the integer lattice in $\mathbf{R}^n$, will be called the *Newton polytope* associated with the linear space $\mathcal{L}_\mathcal{A}$ and denoted by $N_\mathcal{A}$. Obviously we have the relation $N_\mathcal{A} = \cup_{p \in \mathcal{L}_\mathcal{A}} N(p)$. Newton polytopes of polynomials in $p \in \mathcal{L}_\mathcal{A}$ have the following property.

Lemma 1. *[4, p.365] Assume above notation and let $p \in \mathcal{P}_\mathcal{A}$. If $\alpha \in \mathcal{A}$ is an extremal point of $N(p)$, then α is even and $c_\alpha(p) > 0$.*

Without restriction of generality we henceforth assume that

$$\text{all indices in } \operatorname{extr} N_\mathcal{A} \text{ have even entries,} \tag{1}$$

otherwise the cone $\mathcal{P}_\mathcal{A}$ is contained in a proper subspace of $\mathcal{L}_\mathcal{A}$.

Lemma 2. *Under assumption (1), the cone $\mathcal{P}_\mathcal{A}$ has nonempty interior.*

Proof. Let us show that the polynomial $p(x) = \sum_{\alpha \in \operatorname{extr} N_\mathcal{A}} x^\alpha$ is an interior point of $\mathcal{P}_\mathcal{A}$.

Since the logarithm is a concave function, we have for every integer $N > 0$, every set of reals $\lambda_1, \dots, \lambda_N \geq 0$ such that $\sum_{k=1}^N \lambda_k = 1$, and every set of reals $a_1, \dots, a_N > 0$ that $\log \sum_{k=1}^N \lambda_k a_k \geq \sum_{k=1}^N \lambda_k \log a_k$. It follows that $\log \sum_{k=1}^N a_k \geq \sum_{k=1}^N \lambda_k \log a_k$ and therefore $\sum_{k=1}^N a_k \geq \prod_{k=1}^N a_k^{\lambda_k}$.

Let now $\alpha^1, \dots, \alpha^N$ be the extremal points of $N_\mathcal{A}$, and let $\alpha = \sum_{k=1}^N \lambda_k \alpha^k \in \mathcal{A}$ be an arbitrary index, represented as a convex combination of the extremal points. By the above, we then have for every $x \in \mathbf{R}^n$ satisfying $\Pi_{l=1}^n x_l \neq 0$ that $\sum_{k=1}^N x^{\alpha^k} \geq \prod_{k=1}^N (x^{\alpha^k})^{\lambda_k} = |x|^\alpha = |x^\alpha|$. By continuity this holds also for x such that $\Pi_{l=1}^n x_l = 0$. Thus the polynomial $p(x) + q(x)$ is positive, as long as the 1-norm of the coefficient vector $\mathcal{I}_\mathcal{A}(q)$ does not exceed 1.

It follows that both the cone $\mathcal{P}_\mathcal{A}$ and its dual are regular cones, i.e. closed convex cones with nonempty interior, containing no lines.

Lemma 3. *Under assumption (1), $(\mathcal{I}_\mathcal{A}[\mathcal{P}_\mathcal{A}])^* = \operatorname{conv}(\operatorname{con}\operatorname{cl}\mathcal{X}_\mathcal{A})$.*

Proof. Clearly $p \in \mathcal{P}_\mathcal{A}$ if and only if for all $y \in \operatorname{con}\operatorname{cl}\mathcal{X}_\mathcal{A}$ we have $\langle \mathcal{I}_\mathcal{A}(p), y \rangle \geq 0$. Hence $\mathcal{I}_\mathcal{A}[\mathcal{P}_\mathcal{A}]$ is the dual cone of the convex hull $\operatorname{conv}(\operatorname{con}\operatorname{cl}\mathcal{X}_\mathcal{A})$.

It rests to show that this convex hull is closed. Let z be a vector in the interior of the cone $\mathcal{I}_\mathcal{A}[\mathcal{P}_\mathcal{A}]$. Such a vector exists by the preceding lemma. Then the set $C = \{y \in \operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}} \mid \langle y, z\rangle = 1\}$ is compact, and hence its convex hull $\operatorname{conv} C$ is closed. But $\operatorname{conv}(\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}})$ is the conic hull of $\operatorname{conv} C$, and therefore also closed. Thus $(\mathcal{I}_{\mathcal{A}}[\mathcal{P}_{\mathcal{A}}])^* = (\operatorname{conv}(\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}}))^{**} = \operatorname{conv}(\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}})$.

We shall now analyze the set $\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}}$, which is, as can be seen from the previous lemma, determining the cone $\mathcal{P}_{\mathcal{A}}$.

For every ordered index set $\mathcal{A}$ with elements $\alpha^1, \dots, \alpha^m \in \mathbf{N}^n$, where each multi-index is represented by a row vector $\alpha^k = (\alpha^k_1, \dots, \alpha^k_n)$, define the $m \times n$ matrix $M_{\mathcal{A}} = (\alpha^k_l)_{k=1,\dots,m; l=1,\dots,n}$. Further define $\alpha^k_0 = 1$, $k = 1, \dots, m$ and the $m \times (n+1)$ matrix $M'_{\mathcal{A}} = (\alpha^k_l)_{k=1,\dots,m; l=0,\dots,n}$.

Lemma 4. *Assume above notation. Then*

$$\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}} = \operatorname{cl}\{(-1)^\delta \circ \exp(y) \mid \delta \in Im\, \pi_2[M_{\mathcal{A}}],\ y \in Im\, M'_{\mathcal{A}}\},$$

where both $(-1)^\delta$ and $\exp(y)$ are understood element-wise, and $\circ$ denotes the Hadamard product of vectors.

Proof. The space $\mathbf{R}^n$ is composed of 2^n orthants O_γ, which can be indexed by the vectors in $\mathbf{F}_2^n$. Here the index $\gamma = (\gamma_1, \dots, \gamma_n)^T$ of the orthant O_γ is defined such that $\operatorname{sgn} x = (-1)^\gamma$ for all $x \in \operatorname{int} O_\gamma$, where both $\operatorname{sgn} x$ and $(-1)^\gamma$ have to be understood element-wise. In a similar way, the 2^m orthants of $\mathbf{R}^m$ are indexed by the elements of $\mathbf{F}_2^m$.

We shall now compute the set $T_\gamma = \{\beta X_{\mathcal{A}}(x) \mid \beta > 0,\ x \in \operatorname{int} O_\gamma\} \subset \mathbf{R}^m$.

First observe that the signs of the components of $\beta X_{\mathcal{A}}(x)$ do not depend on β and on $x \in \operatorname{int} O_\gamma$. Namely, the k-th component equals $\beta \prod_{l=1}^n x_l^{\alpha^k_l}$, and its sign is $(-1)^{\delta_k}$, where $\delta_k = \sum_{l=1}^n \pi_2(\alpha^k_l)\gamma_l$. Therefore, T_γ is contained in the interior of the orthant O_δ, where $\delta = (\delta_1, \dots, \delta_m)^T = \pi_2[M_{\mathcal{A}}](\gamma) \in \mathbf{F}_2^m$. Thus, if γ runs through $\mathbf{F}_2^n$, then the indices of the orthants containing T_γ run through $Im\, \pi_2[M_{\mathcal{A}}]$.

Consider the absolute values of the components of $\beta X_{\mathcal{A}}(x)$. The logarithm of the modulus of the k-th component is given by $\log \beta + \sum_{l=1}^n \alpha^k_l \log |x_l|$. Now the vector $(\log \beta, \log |x_1|, \dots, \log |x_n|)^T$ runs through $\mathbf{R}^{n+1}$ if (β, x) runs through $\operatorname{int}\mathbf{R}_+ \times \operatorname{int} O_\gamma$, and therefore the element-wise logarithm of the absolute values of $\beta X_{\mathcal{A}}(x)$ runs through $Im\, M_{\mathcal{A}}$, independently of γ.

We have proven the relation

$$\{\beta X_{\mathcal{A}}(x) \mid \beta > 0,\ \prod_{l=1}^n x_l \neq 0\} = \{(-1)^\delta \circ e^y \mid \delta \in Im\, \pi_2[M_{\mathcal{A}}],\ y \in Im\, M'_{\mathcal{A}}\} \tag{2}$$

It rests to show that the closure of the left-hand side equals $\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}}$. Clearly this closure is contained in $\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}}$. The converse inclusion follows from the continuity of the map $(\beta, x) \mapsto \beta X_{\mathcal{A}}(x)$ on $\mathbf{R} \times \mathbf{R}^n$ and the fact that the set $\{(\beta, x) \mid \beta > 0,\ \prod_{l=1}^n x_l \neq 0\}$ is dense in $\mathbf{R}_+ \times \mathbf{R}^n$. This concludes the proof.

The description of $\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}}$ given by Lemma 4 allows us to relate these sets for different index sets $\mathcal{A}$.

Theorem 1. *Assume above notation. Let* $\mathcal{A} = \{\alpha^1, \ldots, \alpha^m\} \subset \mathbf{N}^n$, $\mathcal{A}' = \{\alpha'^1, \ldots, \alpha'^m\} \subset \mathbf{N}^{n'}$ *be nonempty ordered multi-index sets satisfying assumption (1). Then the following are equivalent.*

1) $\mathrm{con}\,\mathrm{cl}\mathcal{X}_{\mathcal{A}} = \mathrm{con}\,\mathrm{cl}\mathcal{X}_{\mathcal{A}'}$,

2) $Im\, M'_{\mathcal{A}} = Im\, M'_{\mathcal{A}'}$ *and* $Im\, \pi_2[M_{\mathcal{A}}] = Im\, \pi_2[M_{\mathcal{A}'}]$,

3) the order isomorphism $I_A : \mathcal{A} \to \mathcal{A}'$ *can be extended to a bijective, affine map* $R : \mathrm{aff}\,\mathcal{A} \to \mathrm{aff}\,\mathcal{A}'$, *and there exists a bijective linear map* $Z : \mathrm{span}(\pi_2^n[\mathcal{A}]) \to \mathrm{span}(\pi_2^{n'}[\mathcal{A}'])$ *such that* $(Z \circ \pi_2^n)(\alpha^k) = \pi_2^{n'}(\alpha'^k)$, $k = 1, \ldots, m$,

4) the order isomorphism $I_A : \mathcal{A} \to \mathcal{A}'$ *can be extended to a lattice isomorphism* $I_\Gamma : \Gamma_{\mathcal{A}} \to \Gamma_{\mathcal{A}'}$, *and* $I_\Gamma[\Gamma^e_{\mathcal{A}}] = \Gamma^e_{\mathcal{A}'}$.

Moreover, the following is a consequence of 1) — 4).

5) $\mathcal{I}_{\mathcal{A}}[\mathcal{P}_{\mathcal{A}}] = \mathcal{I}_{\mathcal{A}'}[\mathcal{P}_{\mathcal{A}'}]$.

Proof. 1) $\Leftrightarrow$ 2): Denote set (2) by $S(\mathcal{A})$. This set is contained in $\cup_{\delta \in \mathbf{F}_2^m} O_\delta$ and is closed in its relative topology. Hence $S(\mathcal{A}) = (\mathrm{cl}\, S(\mathcal{A})) \cap \left(\cup_{\delta \in \mathbf{F}_2^m} O_\delta\right)$. Therefore, if condition 2) is not satisfied, then $S(\mathcal{A}) \neq S(\mathcal{A}')$, and hence $\mathrm{cl}\, S(\mathcal{A}) \neq \mathrm{cl}\, S(\mathcal{A}')$, which implies by Lemma 4 that condition 1) is not satisfied. On the other hand, if condition 2) is satisfied, then $S(\mathcal{A}) = S(\mathcal{A}')$, and again by Lemma 4 condition 1) is satisfied.

2) $\Leftrightarrow$ 3): The first relation in condition 2) is equivalent to the coincidence of the kernels of ${M'}^T_{\mathcal{A}}$ and ${M'}^T_{\mathcal{A}'}$. But these kernels define exactly all affine dependencies between the elements of $\mathcal{A}$ and $\mathcal{A}'$, respectively. Therefore $\ker {M'}^T_{\mathcal{A}} = \ker {M'}^T_{\mathcal{A}'}$ if and only if there exists an isomorphism R between the affine spaces $\mathrm{aff}\,\mathcal{A}$ and $\mathrm{aff}\,\mathcal{A}'$ that takes α^k to α'^k, $k = 1, \ldots, m$. The equivalence of the second relation in condition 2) and the existence of the map Z is proven similarly.

3) $\Leftrightarrow$ 4): Clearly the map R in 3) defines the sought lattice isomorphism $I_\Gamma : \Gamma_{\mathcal{A}} \to \Gamma_{\mathcal{A}'}$. On the other hand, the existence of I_Γ implies that $\ker {M'}^T_{\mathcal{A}} \cap \mathbf{Z}^m = \ker {M'}^T_{\mathcal{A}'} \cap \mathbf{Z}^m$. For the kernel of an integer matrix, however, one can always find an integer basis. Therefore it follows that $\ker {M'}^T_{\mathcal{A}} = \ker {M'}^T_{\mathcal{A}'}$, and I_Γ can be extended to an affine isomorphism $R : \mathrm{aff}\,\mathcal{A} \to \mathrm{aff}\,\mathcal{A}'$. We have shown equivalence of the first conditions in 3) and 4).

Note that π_2^n maps the lattice $\Gamma_{\mathcal{A}}$ to $\mathrm{aff}(\pi_2^n[\mathcal{A}])$, and likewise, $\pi_2^{n'}$ maps $\Gamma_{\mathcal{A}'}$ to $\mathrm{aff}(\pi_2^{n'}[\mathcal{A}'])$. Since both $\mathcal{A}, \mathcal{A}'$ satisfy (1), these sets contain at least one even point. Hence the images $\pi_2^n[\mathcal{A}], \pi_2^{n'}[\mathcal{A}']$ contain the origin, and the affine spans of these images are actually linear spans. Let us now assume that I_Γ exists and consider the diagram

$$\begin{array}{ccc} \Gamma_{\mathcal{A}} & \xrightarrow{I_\Gamma} & \Gamma_{\mathcal{A}'} \\ \pi_2^n \downarrow & & \pi_2^{n'} \downarrow \\ \mathrm{span}(\pi_2^n[\mathcal{A}]) & \xrightarrow{Z} & \mathrm{span}(\pi_2^{n'}[\mathcal{A}']) \end{array}$$

If there exists a linear map Z as in 3), then it makes the diagram commute. The relation $I_\Gamma[\Gamma^e_{\mathcal{A}}] = \Gamma^e_{\mathcal{A}'}$ now follows from the fact that Z maps the origin to the origin. On the other hand, let $I_\Gamma[\Gamma^e_{\mathcal{A}}] = \Gamma^e_{\mathcal{A}'}$. Since $\Gamma^e_{\mathcal{A}} \neq \emptyset$, we have that I_Γ takes pairs of points with even difference to pairs of points with even difference. This implies that there exists a well-defined map Z which makes the diagram commute. Moreover, Z takes the origin to the origin. Since I_Γ is affine, Z must also be affine and hence linear. Finally, repeating the argument with I_Γ^{-1} instead of I_Γ, we see that Z must be invertible.

Finally, the implication 1) $\Rightarrow$ 5) is a direct consequence of Lemma 3.

4 Sums of squares relaxations

Let $\mathcal{A} = \{\alpha^1, \dots, \alpha^m\} \subset \mathbf{N}^n$ be an ordered multi-index set satisfying (1). A polynomial $p \in \mathcal{L}_{\mathcal{A}}$ is certainly positive if it can be represented as a finite sum of squares of other polynomials. The set of polynomials representable in this way forms a closed convex cone [5], the sums of squares cone

$$\Sigma_{\mathcal{A}} = \left\{ p \in \mathcal{L}_{\mathcal{A}} \,|\, \exists\, N,\ q_1, \dots, q_N :\ p = \sum_{k=1}^N q_k^2 \right\} \subset \mathcal{P}_{\mathcal{A}}. \tag{3}$$

The SOS cone is semidefinite representable, and therefore a semidefinite relaxation of the cone $\mathcal{P}_{\mathcal{A}}$. We will henceforth call the cone $\Sigma_{\mathcal{A}}$ the *standard* SOS cone, or the *standard* SOS relaxation. In general we have $\Sigma_{\mathcal{A}} \neq \mathcal{P}_{\mathcal{A}}$, and we will see in Subsection 4.1 that we might even have $\dim \Sigma_{\mathcal{A}} \neq \dim \mathcal{P}_{\mathcal{A}}$.

We shall now generalize the notion of the SOS cone $\Sigma_{\mathcal{A}}$. Let $\mathcal{F} = \{\beta^1, \dots, \beta^{m'}\} \subset \mathbf{N}^n$ be an ordered multi-index set. We then define the set

$$\begin{aligned} \Sigma_{\mathcal{F},\mathcal{A}} &= \left\{ p \in \mathcal{L}_{\mathcal{A}} \,|\, \exists\, N,\ q_1, \dots, q_N \in \mathcal{L}_{\mathcal{F}} : p = \sum_{k=1}^N q_k^2 \right\}, \\ &= \left\{ p \in \mathcal{L}_{\mathcal{A}} \,|\, \exists\, C = C^T \succeq 0 : p(x) = X_{\mathcal{F}}^T(x) C X_{\mathcal{F}}(x) \right\}, \end{aligned} \tag{4}$$

which is also a semidefinite representable closed convex cone. Obviously we have the inclusion $\Sigma_{\mathcal{F},\mathcal{A}} \subset \Sigma_{\mathcal{A}}$. The next result shows that the standard SOS cone $\Sigma_{\mathcal{A}}$ is actually an element of the family $\{\Sigma_{\mathcal{F},\mathcal{A}}\}_{\mathcal{F}\subset\mathbf{N}^n}$ of cones.

Lemma 5. *[4, p.365] If the polynomial $p(x) = \sum_{k=1}^N q_k^2(x)$ is a sum of squares, then for every polynomial q_k participating in the SOS decomposition of p we have $2N(q_k) \subset N(p)$.*

It follows that for every $p(x) = \sum_{k=1}^N q_k(x)^2 \in \Sigma_{\mathcal{A}}$, the nonzero coefficients of every polynomial q_k have multi-indices lying in the polytope $\frac{1}{2}N_{\mathcal{A}}$. Thus $\Sigma_{\mathcal{A}} = \Sigma_{\mathcal{F}_{\max}(\mathcal{A}),\mathcal{A}}$, where $\mathcal{F}_{\max}(\mathcal{A}) = (\frac{1}{2}N_{\mathcal{A}}) \cap \mathbf{N}^n$. We get the following result.

Proposition 1. *Assume above notation. Then for every finite multi-index set $\mathcal{F} \subset \mathbf{N}^n$ such that $\mathcal{F}_{\max}(\mathcal{A}) \subset \mathcal{F}$ we have $\Sigma_{\mathcal{F},\mathcal{A}} = \Sigma_{\mathcal{A}}$.*

In general, the smaller $\mathcal{F}$, the weaker will be the relaxation $\Sigma_{\mathcal{F},\mathcal{A}}$. It does not make sense, however, to choose $\mathcal{F}$ larger than $\mathcal{F}_{\max}(\mathcal{A})$. Let us define the following partial order on the relaxations $\Sigma_{\mathcal{F},\mathcal{A}}$.

Definition 1. *Assume above notation and let the multi-index sets $\mathcal{F}_1, \mathcal{F}_2$ be subsets of $\mathcal{F}_{\max}(\mathcal{A})$. If $\mathcal{F}_1 \subset \mathcal{F}_2$, then we say that the relaxation $\Sigma_{\mathcal{F}_1,\mathcal{A}}$ of the cone $\mathcal{P}_{\mathcal{A}}$ is* coarser *than $\Sigma_{\mathcal{F}_2,\mathcal{A}}$, or $\Sigma_{\mathcal{F}_2,\mathcal{A}}$ is* finer *than $\Sigma_{\mathcal{F}_1,\mathcal{A}}$.*

A finer relaxation is tighter, but a strictly finer relaxation does not a priori need to be strictly tighter. The standard SOS relaxation $\Sigma_{\mathcal{A}}$ is then the finest relaxation among all relaxations of type (4).

We can make the semidefinite representation of $\Sigma_{\mathcal{F},\mathcal{A}}$ explicit by comparing the coefficients in the relation $p(x) = X_{\mathcal{F}}^T(x) C X_{\mathcal{F}}(x)$ appearing in definition (4). As it stands, this relation determines the polynomial $p(x)$ as a function of the symmetric matrix C, thus defining a linear map $L_{\mathcal{F},\mathcal{A}} : \mathcal{S}(m') \to \mathcal{L}_{(\mathcal{F}+\mathcal{F})\cup\mathcal{A}}$ by

$$c_\alpha(p) = \sum_{k,k' :\, \beta^k + \beta^{k'} = \alpha} C_{kk'}, \qquad \alpha \in (\mathcal{F}+\mathcal{F}) \cup \mathcal{A}.$$

Thus we obtain the description

$$\Sigma_{\mathcal{F},\mathcal{A}} = \mathcal{L}_{\mathcal{A}} \cap L_{\mathcal{F},\mathcal{A}}[\mathcal{S}_+(m')], \tag{5}$$

revealing $\Sigma_{\mathcal{F},\mathcal{A}}$ as a linear section of a linear image of the PSD cone $\mathcal{S}_+(m')$.

Note that the linear map $L_{\mathcal{F},\mathcal{A}}$ is completely determined by the map $s_{\mathcal{F},\mathcal{A}} : \mathcal{F} \times \mathcal{F} \to (\mathcal{F}+\mathcal{F}) \cup \mathcal{A}$ defined by $s_{\mathcal{F},\mathcal{A}}(\beta^k, \beta^{k'}) = \beta^k + \beta^{k'}$. Denote by $\mathrm{incl}_{\mathcal{A}} : \mathcal{A} \to (\mathcal{F}+\mathcal{F}) \cup \mathcal{A}$ the inclusion map. We then have the following result.

Theorem 2. *Let $\mathcal{F} = \{\beta^1, \ldots, \beta^{m'}\}, \mathcal{A} = \{\alpha^1, \ldots, \alpha^m\} \subset \mathbf{N}^n$, $\mathcal{F}' = \{\beta'^1, \ldots, \beta'^{m'}\}, \mathcal{A}' = \{\alpha'^1, \ldots, \alpha'^m\} \subset \mathbf{N}^{n'}$, be ordered multi-index sets satisfying $\mathcal{F} \subset \mathcal{F}_{\max}(\mathcal{A})$, $\mathcal{F}' \subset \mathcal{F}_{\max}(\mathcal{A}')$, and let $I_F : \mathcal{F} \to \mathcal{F}'$, $I_A : \mathcal{A} \to \mathcal{A}'$ be the order isomorphisms. Suppose that there exists a bijective map I that makes the following diagram commutative:*

$$\begin{array}{ccccc} \mathcal{F} \times \mathcal{F} & \xrightarrow{s_{\mathcal{F},\mathcal{A}}} & (\mathcal{F}+\mathcal{F}) \cup \mathcal{A} & \xleftarrow{\mathrm{incl}_{\mathcal{A}}} & \mathcal{A} \\ I_F \times I_F \downarrow & & I \downarrow & & I_A \downarrow \\ \mathcal{F}' \times \mathcal{F}' & \xrightarrow{s_{\mathcal{F}',\mathcal{A}'}} & (\mathcal{F}'+\mathcal{F}') \cup \mathcal{A}' & \xleftarrow{\mathrm{incl}_{\mathcal{A}'}} & \mathcal{A}' \end{array}$$

Then $\mathcal{I}_{\mathcal{A}}[\Sigma_{\mathcal{F},\mathcal{A}}] = \mathcal{I}_{\mathcal{A}'}[\Sigma_{\mathcal{F}',\mathcal{A}'}]$.

4.1 Dimensional considerations

From (5) it follows that the cone $\Sigma_{\mathcal{F},\mathcal{A}}$ is always contained in the linear subspace $\mathcal{L}_{(\mathcal{F}+\mathcal{F})\cap\mathcal{A}} \subset \mathcal{L}_{\mathcal{A}}$ and thus is better viewed as a relaxation of the cone $\mathcal{P}_{(\mathcal{F}+\mathcal{F})\cap\mathcal{A}}$ rather than of $\mathcal{P}_{\mathcal{A}}$ itself. In view of Lemma 2 a necessary condition for the cone $\Sigma_{\mathcal{F},\mathcal{A}}$ to have the same dimension as $\mathcal{P}_\mathcal{A}$ is thus the inclusion $\mathcal{A} \subset \mathcal{F} + \mathcal{F}$.

A natural question is now whether this inclusion is always satisfied by the multi-index set $\mathcal{F} := \mathcal{F}_{\max}(\mathcal{A}) = (\frac{1}{2}N_\mathcal{A}) \cap \mathbf{N}^n$, which gives rise to the standard SOS cone $\Sigma_\mathcal{A}$. The answer to this question is negative, as the example $\mathcal{A} = \{(2,0,0),(0,2,0),(2,2,0),(0,0,4),(1,1,1)\}$ taken from [4, p.373] shows.

It is, however, not hard to show that if $\mathcal{A}$ is contained in a 2-dimensional affine plane, then $\mathcal{A} \subset \mathcal{F}_{\max}(\mathcal{A}) + \mathcal{F}_{\max}(\mathcal{A})$.

5 Hierarchies of relaxations

In this section we construct hierarchies of semidefinite relaxations of the cone of positive polynomials which are tighter than the standard SOS relaxation.

Conditions 2) — 4) of Theorem 1 define an easily verifiable equivalence relation $\sim_P$ on the class of finite ordered multi-index sets satisfying (1). By Theorem 1, we have for any two equivalent multi-index sets $\mathcal{A} \sim_P \mathcal{A}'$ that $\mathcal{I}_\mathcal{A}[\mathcal{P}_\mathcal{A}] = \mathcal{I}_{\mathcal{A}'}[\mathcal{P}_{\mathcal{A}'}]$. It is therefore meaningful to define the abstract cone

$$\mathcal{P}_{[\mathcal{A}]} = \mathcal{I}_\mathcal{A}[\mathcal{P}_\mathcal{A}] = \{\mathcal{I}_\mathcal{A}(p) \,|\, p \in \mathcal{P}_\mathcal{A}\} \subset \mathbf{R}^m,$$

where $[\mathcal{A}]$ is the equivalence class of $\mathcal{A}$ with respect to the relation $\sim_P$. The points of this cone cannot anymore be considered as polynomials on $\mathbf{R}^n$. A cone of positive inhomogeneous polynomials on $\mathbf{R}^n$, e.g., corresponds to the same abstract cone as the cone of their homogenizations, which are defined on $\mathbf{R}^{n+1}$. For every concrete choice of a representative $\mathcal{A}' \in [\mathcal{A}]$, however, the map $\mathcal{I}_{\mathcal{A}'}^{-1}$ puts them in correspondence with positive polynomials in $\mathcal{P}_{\mathcal{A}'}$.

Similarly, the existence of the bijective map I in Theorem 2 defines an easily verifiable equivalence relation $\sim_\Sigma$ on the class of pairs $(\mathcal{F},\mathcal{A})$ of ordered finite multi-index sets satisfying $\mathcal{F} \subset \mathcal{F}_{\max}(\mathcal{A})$. By Theorem 2, for any two equivalent pairs $(\mathcal{F},\mathcal{A}) \sim_\Sigma (\mathcal{F}',\mathcal{A}')$ we have $\mathcal{I}_\mathcal{A}[\Sigma_{\mathcal{F},\mathcal{A}}] = \mathcal{I}_{\mathcal{A}'}[\Sigma_{\mathcal{F}',\mathcal{A}'}]$. We can then define the abstract cone $\Sigma_{[(\mathcal{F},\mathcal{A})]} = \mathcal{I}_\mathcal{A}[\Sigma_{\mathcal{F},\mathcal{A}}] \subset \mathbf{R}^m$, where $[(\mathcal{F},\mathcal{A})]$ is the equivalence class of the pair $(\mathcal{F},\mathcal{A})$ with respect to the relation $\sim_\Sigma$. For every concrete choice of a representative $(\mathcal{F}',\mathcal{A}') \in [(\mathcal{F},\mathcal{A})]$ the map $\mathcal{I}_{\mathcal{A}'}^{-1}$ takes the abstract cone $\Sigma_{[(\mathcal{F},\mathcal{A})]}$ to the cone $\Sigma_{\mathcal{F}',\mathcal{A}'}$ of SOS polynomials.

For different, but equivalent, multi-index sets $\mathcal{A} \sim_P \mathcal{A}'$ the standard SOS relaxations $\Sigma_\mathcal{A}, \Sigma_{\mathcal{A}'}$ defined by (3) will in general not be equivalent. It is therefore meaningless to speak of a standard SOS relaxation of the cone $\mathcal{P}_{[\mathcal{A}']}$. For every representative $\mathcal{A} \in [\mathcal{A}']$ we have, however, a finite hierarchy of SOS relaxations $\Sigma_{\mathcal{F},\mathcal{A}}$ defined by (4). This allows us to define SOS relaxations of the abstract cone $\mathcal{P}_{[\mathcal{A}']}$.

Definition 2. *Let C be an equivalence class of finite ordered multi-index sets with respect to the equivalence relation $\sim_P$, and $\mathcal{P}_C$ the corresponding abstract cone of positive polynomials. For every pair $(\mathcal{F},\mathcal{A})$ of finite ordered multi-index sets such that $\mathcal{A} \in C$ and $\mathcal{F} \subset \mathcal{F}_{\max}(\mathcal{A})$, we call the abstract cone $\Sigma_{[(\mathcal{F},\mathcal{A})]}$ an SOS relaxation of $\mathcal{P}_C$.*

Clearly the SOS relaxations of the cone $\mathcal{P}_C$ are inner semidefinite relaxations. The set of SOS relaxations inherits the partial order defined in Definition 1.

Definition 3. *Let C be an equivalence class of finite ordered multi-index sets with respect to the equivalence relation $\sim_P$, and let $\Sigma_{C_1}, \Sigma_{C_2}$ be SOS relaxations of the cone $\mathcal{P}_C$, where C_1, C_2 are equivalence classes of the relation $\sim_\Sigma$. If there exist multi-index sets $\mathcal{F}_1, \mathcal{F}_2, \mathcal{A}$ such that $(\mathcal{F}_1, \mathcal{A}) \in C_1$, $(\mathcal{F}_2, \mathcal{A}) \in C_2$, and $\mathcal{F}_1 \subset \mathcal{F}_2$, then we say that the relaxation Σ_{C_1} is* coarser *than Σ_{C_2}, or Σ_{C_2} is* finer *than Σ_{C_1}.*

It is not hard to see that the relation defined in Definition 3 is indeed a partial order. A finer relaxation is tighter, but a strictly finer relaxation does not a priori need to be strictly tighter. Note that we do not require $\mathcal{A} \in C$. This implies that if, e.g., both $\Sigma_{C_1}, \Sigma_{C_2}$ are SOS relaxations for two different abstract cones $\mathcal{P}_C, \mathcal{P}_{C'}$, and Σ_{C_2} is a finer relaxation of $\mathcal{P}_C$ than Σ_{C_1}, then Σ_{C_2} is also a finer relaxation of $\mathcal{P}_{C'}$ than Σ_{C_1}.

Theorem 3. *Let $\mathcal{F}, \mathcal{A} \subset \mathbf{N}^n$ be finite ordered multi-index sets satisfying $\mathcal{F} \subset \mathcal{F}_{\max}(\mathcal{A})$, and suppose that $\mathcal{A}$ satisfies (1). Let further M be an $n \times n$ integer matrix with odd determinant, and let $v \in \mathbf{Z}^n$ be an arbitrary integer row vector. Let now $\mathcal{F}'$ be the multi-index set obtained from $\mathcal{F}$ by application of the affine map $R' : \beta \mapsto \beta M + v$, and $\mathcal{A}'$ the set obtained from $\mathcal{A}$ by application of the affine map $R : \alpha \mapsto \alpha M + 2v$. Then $\mathcal{A} \sim_P \mathcal{A}'$, provided the elements of $\mathcal{A}'$ have nonnegative entries, and $(\mathcal{F}, \mathcal{A}) \sim_\Sigma (\mathcal{F}', \mathcal{A}')$, provided the elements of $\mathcal{F}', \mathcal{A}'$ have nonnegative entries.*

Proof. Assume the conditions of the theorem. Then we have $N_{\mathcal{A}'} = R[N_\mathcal{A}]$, and therefore $\frac{1}{2}N_{\mathcal{A}'} = R'[\frac{1}{2}N_\mathcal{A}]$. It follows that $\mathcal{F}' \subset \mathcal{F}_{\max}(\mathcal{A}')$.

Since $\det M \neq 0$, the map R is invertible. Further, the matrix $\pi_2[M]$ defines an invertible linear map Z on $\mathbf{F}_2^n$, because $\det \pi_2[M] = \pi_2(\det M) = 1$. Moreover, the projection π_2^n intertwines the maps R and Z, because the translational part of R is even. It is then easily seen that the restrictions $R|_{\mathrm{aff}\,\mathcal{A}} : \mathrm{aff}\,\mathcal{A} \to \mathrm{aff}\,\mathcal{A}'$ and $Z|_{\mathrm{span}(\pi_2^n[\mathcal{A}])} : \mathrm{span}(\pi_2^n[\mathcal{A}]) \to \mathrm{span}(\pi_2^n[\mathcal{A}'])$ satisfy condition 3) of Theorem 1. This proves the relation $\mathcal{A} \sim_P \mathcal{A}'$.

Likewise, the restriction $I = R|_{(\mathcal{F}+\mathcal{F})\cup\mathcal{A}}$ makes the diagram in Theorem 2 commute, which proves the relation $(\mathcal{F}, \mathcal{A}) \sim_\Sigma (\mathcal{F}', \mathcal{A}')$.

We will use this result to construct strictly finer relaxations from a given standard SOS relaxation.

If the determinant of the matrix M in Theorem 3 equals ± 1, then the maps R', R define isomorphisms of $\mathbf{Z}^n$. Then $\#\,\mathcal{F}_{\max}(\mathcal{A}) = \#\,\mathcal{F}_{\max}(\mathcal{A}')$, and the standard relaxations $\Sigma_\mathcal{A}, \Sigma_{\mathcal{A}'}$ are equivalent. If, however, $|\det M| > 1$, then $\#\,\mathcal{F}_{\max}(\mathcal{A}')$ might be strictly bigger than $\#\,\mathcal{F}_{\max}(\mathcal{A})$, and then $\mathcal{I}_{\mathcal{A}'}[\Sigma_{\mathcal{A}'}]$ will be strictly finer than $\mathcal{I}_\mathcal{A}[\Sigma_\mathcal{A}]$. In particular, this happens if $\#\,\mathcal{A} > 1$ and the sets $\mathcal{F}', \mathcal{A}'$ are obtained from $\mathcal{F}, \mathcal{A}$ by multiplying every multi-index with a fixed odd natural number $k > 1$. Thus, unlike the hierarchy of relaxations (4), for abstract cones $\mathcal{P}_C$ of dimension $m > 1$ the hierarchy of SOS relaxations is infinite, and the corresponding partial order does not have a finest element.

6 Example

Consider the inhomogeneous Motzkin polynomial $p_M(x,y) = x^4y^2 + x^2y^4 + 1 - 3x^2y^2 \in \mathcal{P}_{\mathcal{A}}$ with $\mathcal{A} = \{(4,2),(2,4),(0,0),(2,2)\}$. Its Newton polytope is the triangle given by $N(p_M) = N_{\mathcal{A}} = \operatorname{conv}\{(4,2),(2,4),(0,0)\}$, and therefore $\mathcal{F}_{\max}(\mathcal{A}) = \{(2,1),(1,2),(0,0),(1,1)\}$. It is easily checked that the standard SOS cone $\Sigma_{\mathcal{A}}$ obtained from (4) by setting $\mathcal{F} = \mathcal{F}_{\max}(\mathcal{A})$ consists of those polynomials in $\mathcal{L}_{\mathcal{A}}$ all whose coefficients are nonnegative. The corresponding abstract SOS cone is therefore given by $\Sigma_{[(\mathcal{F},\mathcal{A})]} = \mathbf{R}_+^4$.

Using Lemma 4, it is a little exercise to show $\operatorname{con}\operatorname{cl}\mathcal{X}_{\mathcal{A}} = \{(y_1,y_2,y_3,y_4)^T \in \mathbf{R}_+^4 \mid y_4 = \sqrt[3]{y_1y_2y_3}\}$. By Lemma 3 we then easily obtain $\mathcal{P}_{[\mathcal{A}]} = \{c = (c_1,c_2,c_3,c_4)^T \mid c_1,c_2,c_3 \geq 0,\ c_4 \geq -3\sqrt[3]{c_1c_2c_3}\}$. Let us now apply the construction provided in Theorem 3, setting $M = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$ and $v = 0$. Then $\mathcal{F}$ is mapped to $\mathcal{F}' = \{(3,0),(0,3),(0,0),(1,1)\}$, and $\mathcal{A}$ to $\mathcal{A}' = \{(6,0),(0,6),(0,0),(2,2)\}$. By Theorem 3 we then have $\mathcal{A} \sim_P \mathcal{A}'$ and $(\mathcal{F},\mathcal{A}) \sim_\Sigma (\mathcal{F}',\mathcal{A}')$. The set $\mathcal{F}'' = \mathcal{F}_{\max}(\mathcal{A}')$, however, is now composed of 10 points and is hence strictly larger than $\mathcal{F}'$. Therefore the relaxation $\Sigma_{[(\mathcal{F}'',\mathcal{A}')]}$ of the cone $\mathcal{P}_{[\mathcal{A}]}$ is strictly finer than $\Sigma_{[(\mathcal{F},\mathcal{A})]}$. Moreover, with $e_3 = (1,1,1)^T$, $v(x,y) = (\sqrt[3]{c_1}x^2, \sqrt[3]{c_2}y^2, \sqrt[3]{c_3})^T$ every polynomial $p_c(x,y) = c_1x^6 + c_2y^6 + c_3 + c_4x^2y^2 \in \mathcal{P}_{\mathcal{A}'}$, i.e. satisfying $c_1,c_2,c_3 \geq 0$ and $c_4 \geq -3\sqrt[3]{c_1c_2c_3}$, can be written as

$$p_c(x,y) = e^T v(x,y) \cdot v(x,y)^T \frac{3I_3 - ee^T}{2} v(x,y) + (c_4 + 3\sqrt[3]{c_1c_2c_3})x^2y^2,$$

which obviously is a sum of squares. Thus the relaxation $\Sigma_{[(\mathcal{F}'',\mathcal{A}')]}$ is *exact.*

From the proof of [3, Theorem 1] it follows that there does not exist a fixed integer $d > 0$ such that with $h(x) = \left(\sum_{k=1}^n x_k^2\right)^d$ the product hp is a sum of squares for every polynomial $p \in \mathcal{P}_{\mathcal{A}}$. Thus this commonly used hierarchy of SOS relaxations is not capable of representing the cone $\mathcal{P}_{[\mathcal{A}]}$ at any finite step.

References

1. Motzkin T S (1967) The arithmetic-geometric inequality. In: Shisha O (ed) Inequalities Proc of Sympos at Wright-Patterson AFB. Acad Press, New York
2. Lasserre J B (2009) Moments, positive polynomials and their applications. Imperial College Press, London
3. Reznick B (2005) Proc Amer Math Soc 133:2829–2834
4. Reznick B (1978) Duke Math Journal 45(2):363–372
5. Ellison W J (1969) Math Proc Cambridge Phil Soc 65:663–672
6. Hilbert D (1888) Math Ann 32:342–350.