



Generating Valid Linear Inequalities for Nonlinear Programs via Sums of Squares

Sönke Behrends¹ · Anita Schöbel²

Received: 26 January 2020 / Accepted: 30 July 2020 / Published online: 12 August 2020
© The Author(s) 2020

Abstract

Valid linear inequalities are substantial in linear and convex mixed-integer programming. This article deals with the computation of valid linear inequalities for nonlinear programs. Given a point in the feasible set, we consider the task of computing a tight valid inequality. We reformulate this geometrically as the problem of finding a hyperplane which minimizes the distance to the given point. A characterization of the existence of optimal solutions is given. If the constraints are given by polynomial functions, we show that it is possible to approximate the minimal distance by solving a hierarchy of sum of squares programs. Furthermore, using a result from real algebraic geometry, we show that the hierarchy converges if the relaxed feasible set is bounded. We have implemented our approach, showing that our ideas work in practice.

Keywords Valid inequalities · Nonlinear optimization · Polynomial optimization · Semi-infinite programming · Sum of squares (sos) · Hyperplane location

Mathematics Subject Classification 90C30 · 90C11 · 90C10 · 14P10

1 Introduction

The problem we want to solve is the following: Given a subset S of \mathbb{R}^n and a point q in S , find a valid linear inequality for S which is as close as possible to q (a formal definition is given in Sect. 2). Our motivation stems from the fact that valid linear inequalities play an important role in solving mixed-integer linear and mixed-integer convex programs. It is thus a natural task to study valid inequalities for the more general class of mixed-integer nonlinear programs (MINLP). More specifically, we search for

✉ Sönke Behrends
s.behrends@math.uni-goettingen.de

¹ University of Goettingen, Goettingen, Germany

² University of Kaiserslautern and Fraunhofer Institute for Industrial Mathematics ITWM, Kaiserslautern, Germany

valid inequalities for the feasible set $F_{\mathcal{I}}$ and its continuous relaxation F . We also consider the special case where we require the objective and constraint functions to be polynomials, which we refer to as mixed-integer polynomial programming (MIPP). To avoid unnecessary clutter, we state our results for the set S which can be thought of being equal to F or $F_{\mathcal{I}}$.

In mixed-integer linear and convex programming, one is interested in finding valid inequalities for $F_{\mathcal{I}}$. One reason for this interest is that the convex hull of $F_{\mathcal{I}}$ can be described by finitely many valid inequalities for rational data in the mixed-integer linear case [1]. This result does not generalize to the convex case; however, in mixed-integer convex programming, a common solution approach is the generation of cuts. A cut is a valid linear inequality for the feasible set that is violated at some point of the relaxed feasible set. A second motivation to find valid linear inequalities is “polyhedrification”, a special form of convexification [2], that is, outer approximation of the sets F and $F_{\mathcal{I}}$ by polyhedra. Note that the meaning of outer approximation is twofold in the literature: It is the name of a celebrated solution method for a special class of MINLP [3,4], and also describes the process of relaxing a complicated set to a larger set that is easier to handle.

An early result on cuts in mixed-integer linear programming is the algorithmic generation of so-called Gomory cuts [5]. Later, it was shown that the repeated application of all Gomory-type cuts yields the convex hull of $F_{\mathcal{I}}$ for linear integer programming, see Theorem 1.1 in [6]. Nowadays, the underlying theory of cuts has become quite deep and the number of different types of cuts is—even though the underlying ideas of the cuts are often related—vast. The article [6] is a modern presentation of the most influential cuts, and the article [7] explores the relationships in the cut zoo. Recently, maximal lattice-free polyhedra have attracted attention, since it can be shown that the strongest cuts are derived from maximal lattice-free polyhedra, see [8].

Methods for generating cuts also exist in convex programming; for an early approach, see [9]. A modern introduction into the key ideas on cuts for continuous convex problems is given in [10]. For an overview on cuts for mixed-integer convex problems, we refer to Chapter 4 in [11].

There is work on the computation of valid linear inequalities in the non-convex setting: In [12], the authors compute outer approximations for separable non-convex mixed-integer nonlinear programs, and require the feasible set to be contained in a known polytope, i.e., a compact polyhedron. Another recent approach which has proven to be quite successful is via so-called S -free sets, generalizing the idea of lattice-free polyhedra. For example, in [13], an oracle-based cut generating algorithm is presented that computes an arbitrarily precise—as measured by the Hausdorff-distance—approximation of the convex hull of a closed set, if the latter is contained in a known polytope. Related is the work [14], and its extension [15], where the authors derive cuts for a pre-image of a closed set by a linear mapping via so-called cut generating functions and show that these functions are intimately related to S -free sets. There are also new results on cut generation for special cases. A framework is proposed in [16] that creates valid inequalities for a reverse convex set using a cut generating linear program. Also, maximal S -free sets for quadratically constrained problems are computed [17].

Finding a valid linear inequality can be considered as a hyperplane location problem. For an overview on a location theory approach, we refer to [18,19].

In this article, sos programming plays a key role. This technique can be traced back to [20–23]. See [24] for a survey, and [25] for a focus on geometric aspects. An algebraic approach is [26]. For a concise treatment, let us mention [27].

Nonlinear mixed-integer programming itself is a large problem class, and the literature is extensive. For an overview and several pointers to key results, let us mention the survey [28], as well as [11,29].

The remainder of the paper is structured as follows. Section 2 settles notation and prerequisites from sum of squares programming. Section 3 formulates the task of finding a valid linear inequality for a given subset S of \mathbb{R}^n that is close to a point q in S . The distance is measured by a gauge. We give geometric characterizations that ensure the existence of feasible and optimal solutions and formulate the problem as a non-convex and semi-infinite optimization problem. In order to make the problem tractable, we first linearize the objective function using a result from [30] in Sect. 4. In Sect. 5, we give a convex reformulation if the distance is measured by a polyhedral gauge and furthermore, restricting ourselves to a semi-algebraic set S in Sect. 6 we receive a hierarchy of sos programs, which converge to the optimal value of the original program if S is bounded. Illustrating examples are provided in Sect. 7. Extensions are discussed in Sect. 8.

Our first main contribution—definitions deferred to Sect. 2—is Proposition 5.1: If valid linear inequalities exist, we can find one that is tight with respect to a polyhedral gauge by solving finitely many linear semi-infinite problems. This result holds without any structural assumptions on the set S , which is to the best of our knowledge a new contribution. The second main contribution is Theorem 6.1: If S is semi-algebraic and the gauge polyhedral, we can give a weakened formulation in terms of a hierarchy of sos programs. Feasibility provided, the hierarchy yields a sequence of hyperplanes with decreasing distances to q . If the corresponding quadratic module is Archimedean, we can guarantee convergence of the hierarchy towards a tight valid inequality. In contrast with the approach of [12], we do not require S to be contained in a polytope to produce feasible solutions (see Sect. 7 for an unbounded example). Similarly, for the method in [13], an oracle is needed, and in the case of polynomial constraints, an oracle is only provided if the feasible set is contained in a polytope. The approach in [14] is not algorithmic and is, without further research, not directly applicable in our setting.

Regarding the scope of the paper, let us note that, for MINLP, it is an important question how cuts can be generated from valid inequalities. One reason for this is that they can improve the strength of an optimization model by only using information inherent in the model, see, e.g., [5,6,8,13–17]. Also, it is an interesting question how different choices of q affect the obtained valid inequalities. Both questions are not within the scope of this article and left as starting points for future research.

2 Preliminaries for Our Approach

In this section, we introduce basic concepts and notation that is used throughout this article. The natural, integer, and real numbers are denoted by \mathbb{N} , \mathbb{Z} and \mathbb{R} . The natural

numbers do not contain 0, and we denote $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ and put $[k] := \{1, \dots, k\}$ for $k \in \mathbb{N}$.

2.1 Tight Valid Inequalities

An inequality $(a^T x \leq b)$ is given by some $a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$. We say that $(a^T x \leq b)$ is

- valid for $S \subset \mathbb{R}^n$ if it is satisfied for all $x \in S$, that is, $a^T x \leq b$ holds for all $x \in S$,
- violated by $x \in S$ if $a^T x > b$,
- tight for S if it is valid for S and for any $b' < b$, the inequality $(a^T x \leq b')$ is violated by some $x \in S$,
- tight for S at $q \in S$ if the inequality is tight for S and $a^T q = b$.

A linear inequality $(a^T x \leq b)$ corresponds to the half-space $\{x \in \mathbb{R}^n : a^T x \leq b\}$. The associated hyperplane is denoted by

$$H(a, b) := \{x \in \mathbb{R}^n : a^T x = b\}.$$

2.2 Polynomials, Sum of Squares, and Quadratic Modules

We denote the ring of polynomials in n unknowns X_1, \dots, X_n and coefficients in \mathbb{R} by $\mathbb{R}[X_1, \dots, X_n]$. A polynomial is a sum of squares or sos for short if it has a representation as a sum of squared polynomials. Formally, we have $p \in \mathbb{R}[X_1, \dots, X_n]$ is sos if there are $q_1, \dots, q_l \in \mathbb{R}[X_1, \dots, X_n]$ with

$$p = q_1^2 + \dots + q_l^2. \tag{1}$$

We denote the set of all sos polynomials by

$$\Sigma_n := \left\{ p \in \mathbb{R}[X_1, \dots, X_n] : \exists q_1, \dots, q_l \in \mathbb{R}[X_1, \dots, X_n], p = \sum_{j=1}^l q_j^2 \right\}.$$

What makes this notion useful is that an immediate consequence of a representation of p as in (1) is that p is nonnegative on all of \mathbb{R}^n , and the q_i certify nonnegativity. It turns out that deciding if a polynomial is a sum of squares can be reformulated as a semidefinite program (SDP), and SDPs in turn are well-understood and can be solved efficiently, see, e.g., [31,32]. The set Σ_n of all sos polynomials is a convex cone in $\mathbb{R}[X_1, \dots, X_n]$.

We also need the notion of semi-algebraic sets: Given a finite collection of multivariate polynomials $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$, consider the subset of \mathbb{R}^n where all polynomials h_i attain nonnegative values

$$K(h_1, \dots, h_s) := \{x \in \mathbb{R}^n : h_1(x) \geq 0, \dots, h_s(x) \geq 0\}. \tag{2}$$

A subset of \mathbb{R}^n is called *basic closed semi-algebraic*, or in this article semi-algebraic for short, if it is of the form (2) for some polynomials h_1, \dots, h_s . For example, we note that for a MIPP with constraint polynomials h_1, \dots, h_s , we have $F = K(h_1, \dots, h_s)$.

Given the constraints $h_i(x) \geq 0$ to a MIPP, a way to infer further valid inequalities is to scale the h_i by sos (and thus nonnegative) polynomials and add them up. This is formalized in the algebraic definition of a quadratic module generated by the h_i :

$$M(h_1, \dots, h_s) := \left\{ \sum_{i=0}^s \sigma_i h_i : \sigma_0, \dots, \sigma_s \in \Sigma_n \right\} \tag{3}$$

where $h_0 := 1$.

In sos programming, the coefficients of the σ_i appearing in (3) are unknowns that we optimize. As there is no degree bound on the σ_i , this is impractical. Hence, we instead use the truncated quadratic module of order $k \in \{-\infty\} \cup \mathbb{N}$, given by

$$M(h_1, \dots, h_s)[k] := \left\{ \sum_{i=0}^s \sigma_i h_i : \sigma_i \in \Sigma_n, \deg(\sigma_i h_i) \leq k, i = 0, \dots, s \right\}, \tag{4}$$

where, again, $h_0 := 1$.

We address now the question how polynomials in $M(h_1, \dots, h_s)$ and polynomials nonnegative on $K(h_1, \dots, h_s)$ are related. The following observation which derives a geometric statement from an algebraic one, follows directly from the definitions of $K(h_1, \dots, h_s)$ and $M(h_1, \dots, h_s)$.

Observation 2.1 *Let $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$ and $p \in M(h_1, \dots, h_s)$. Then $p \geq 0$ on $K(h_1, \dots, h_s)$.*

The question addressing the “converse direction”—suppose a polynomial p is nonnegative on $K(h_1, \dots, h_s)$, does $p \in M(h_1, \dots, h_s)$ hold?—is more difficult to answer. Conditions that guarantee such representations are addressed in Positivstellensätzen. In this article, we use a Positivstellensatz by Putinar. It holds under a technical condition that we outline next.

2.3 The Archimedean Property and Putinar’s Positivstellensatz

The condition needed for the Positivstellensatz to hold is that the quadratic module $M = M(h_1, \dots, h_s)$ needs to be Archimedean. The quadratic module M is Archimedean if for all $p \in \mathbb{R}[X_1, \dots, X_n]$ there exists $k \in \mathbb{N}$ with $p + k \in M$. The following equivalent characterization is useful for our purposes.

Theorem 2.1 (see, e.g., Corollary 5.2.4 in [26]) *Given $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$, let $M = M(h_1, \dots, h_s)$ be the associated quadratic module. Then M is Archimedean if and only if there is a number $k \in \mathbb{N}$ such that $k - \sum_{i=1}^n X_i^2 \in M$.*

A consequence of $M(h_1, \dots, h_s)$ being Archimedean, which is straightforward, well-known but nevertheless important, is that the basic closed semi-algebraic set associated with the polynomials h_i is compact.

Corollary 2.1 *Let $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$ and suppose $M(h_1, \dots, h_s)$ is Archimedean. Then $K(h_1, \dots, h_s)$ is compact.*

Proof This follows from Observation 2.1 and Theorem 2.1. □

On the other hand, if $K(h_1, \dots, h_s)$ is compact, then $M(h_1, \dots, h_s)$ need not be Archimedean, see, e.g., Example 7.3.1 in [26]. As one of our main results (Theorem 6.1) requires $M(h_1, \dots, h_s)$ to be Archimedean, it is a natural question if one can decide whether a given quadratic module satisfies this property.

Remark 2.1 If $S = K(h_1, \dots, h_s)$, it is possible to enforce the Archimedean property on the associated quadratic module M if we have a known bound $R \geq 0$ such that $\|x\|_2 \leq R$ for all $x \in S$.¹ Specifically, by adding the redundant constraint $h_{s+1} := R^2 - \sum_{i=1}^n X_i^2$ to the description of S , we still have the equality $S = K(h_1, \dots, h_{s+1})$, but now Theorem 2.1 guarantees that $M(h_1, \dots, h_{s+1})$ is Archimedean, see, e.g., [33].

We can now give the Positivstellensatz. Note that the theorem requires positivity, a stronger requirement than nonnegativity.

Theorem 2.2 (Putinar’s Positivstellensatz, see, e.g., Theorem 5.6.1 in [26]) *Let $p, h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$ and $M(h_1, \dots, h_s)$ be Archimedean. Then $p(x) > 0$ for all $x \in K(h_1, \dots, h_s)$ implies $p \in M(h_1, \dots, h_s)$.*

2.4 Gauge Functions

A gauge is a function $\gamma : \mathbb{R}^n \rightarrow \mathbb{R}$ of the form

$$\gamma(x) = \gamma(x; A) := \inf\{t \geq 0 : x \in tA\}$$

for $A \subset \mathbb{R}^n$ compact, convex with $0 \in \text{int } A$. Note that every norm with unit ball B is a gauge $\gamma(\cdot; B)$. On the other hand, a gauge $\gamma(\cdot; A)$ satisfies definiteness, positive homogeneity and the triangle property as norms do. It is a norm if additionally absolute homogeneity holds, equivalently, if A is symmetric, i.e., $-A = A$.

Given $C, D \subset \mathbb{R}^n$ and a gauge γ on \mathbb{R}^n , the distance from C to D is

$$d(C, D) := \inf\{\gamma(d - c) : c \in C, d \in D\}.$$

For a singleton set $C = \{c\}$ we also write $d(c, D)$. Analogous to norms, the distance measured by gauges between two nonempty sets $C, K \subset \mathbb{R}^n$ is attained if C is closed and K compact, i.e., there exist $c^* \in C$ and $k^* \in K$ with

$$d(c^*, k^*) = d(C, K) \tag{5}$$

in this case.

¹ In theory, this is equivalent to S being compact. However, computing R for S given by constraint functions is itself a nonlinear continuous optimization problem.

The polar of a set $A \subset \mathbb{R}^n$ is

$$A^\circ := \{x \in \mathbb{R}^n : x^T y \leq 1 \forall y \in A\}. \quad (6)$$

It can be shown that if A is compact, convex with $0 \in \text{int } A$, the same holds for A° , see, e.g., Corollary 14.5.1 in [34]. For a gauge γ , the function

$$\gamma^\circ(x) := \sup\{x^T y : y \in \mathbb{R}^n, \gamma(y) \leq 1\}$$

is the polar of γ . It then holds that

$$\gamma^\circ(x; A) = \gamma(x; A^\circ), \quad (7)$$

that is, the polar of a gauge is again a gauge, see, e.g., Theorem 15.1 in [34].

We also consider gauges γ that are polyhedral: A gauge $\gamma(\cdot, A)$ is called polyhedral if A is a polyhedron. As the polar of a polyhedron is a polyhedron (see, e.g., Corollary 19.2.2 in [34]), it is clear in view of (7) that the polar of a polyhedral gauge is again a polyhedral gauge. For a polyhedral gauge $\gamma(\cdot, A)$, the extreme points of A are called fundamental directions of γ .

3 A Geometric Reformulation and Its Properties

In this section we formulate the task to find a tight valid linear inequality as the following geometric optimization² problem: Given $q \in S$, find a valid linear inequality for S such that the associated hyperplane has a minimum distance (defined by an arbitrary gauge function³) to q .

$$\begin{aligned} \min \quad & d(q, H(a, b)) \\ \text{s.t.} \quad & a \neq 0 \\ & a^T x \leq b \quad \text{for all } x \in S \\ & a \in \mathbb{R}^n, b \in \mathbb{R}. \end{aligned} \quad (\text{V1})$$

Let us interpret solutions of Program V1 geometrically.

Proposition 3.1 *For Program V1, it holds:*

1. Every feasible solution (a, b) yields a valid inequality $(a^T x \leq b)$ for S .
2. Every optimal solution (a, b) yields an inequality $(a^T x \leq b)$ that is tight for S .
3. Every optimal solution (a, b) with objective value 0 yields an inequality that is tight for S at q .

² Throughout this article, we do not assume that the minimum of a minimization problem exists, but allow for $+\infty$ and $-\infty$ as optimal values, so the optimal value always exists.

³ Let us point out that gauges also play a central role in [14], where they arise as a counterpart to S -free sets, and are then used to generate cuts.

Proof Claim 1 is clear. To see Claim 2, let (a, b) be an optimal solution and assume the contrary, i.e., there is $b' < b$ with $(a^T x \leq b')$ valid for S . By (5) there is $u \in H(a, b)$ with $\gamma(u - q) = d(q, H(a, b))$. Using $q \in S$, we get the inequalities

$$a^T q \leq b' < b = a^T u, \tag{8}$$

and hence $a^T(u - q) > 0$. Put

$$\hat{u} := u + \lambda(q - u) \text{ with } \lambda := \frac{b - b'}{a^T(u - q)}.$$

Observe that \hat{u} is a point on $H(a, b')$:

$$b' = a^T u + b' - b = a^T u + \frac{a^T(q - u)}{a^T(q - u)}(b' - b) = a^T u + \lambda a^T(q - u) = a^T \hat{u}.$$

Note that (8) implies $\lambda > 0$ and $\lambda \leq \frac{b-b'}{b-b'} = 1$. Since $\hat{u} - q = (1 - \lambda)(u - q)$, all our observations combine to

$$d(q, H(a, b')) \leq \gamma(\hat{u} - q) = (1 - \lambda)\gamma(u - q) < \gamma(u - q) = d(q, H(a, b)).$$

Hence (a, b') is a feasible solution to V1 with better objective value, contradicting optimality of (a, b) .

To see Claim 3, note that if the objective value is 0 at (a, b) we know from (5) that $d(q, u) = 0$ for some $u \in H(a, b)$, hence $q = u$ and we conclude $q \in H(a, b)$. The claim follows. □

It turns out that feasibility of V1 is sufficient for the existence of optimal solutions.

Theorem 3.1 *Let $S \subset \mathbb{R}^n$ and $q \in S$. Then, the following are equivalent:*

1. Program V1 is feasible.
2. Program V1 has optimal solutions.
3. $\text{conv } S \subsetneq \mathbb{R}^n$.

Proof To see Claim 1 \Rightarrow Claim 3, let Program V1 be feasible, and hence $S \subset (a^T x \leq b)$ for some $a \in \mathbb{R}^n, a \neq 0, b \in \mathbb{R}$. Now

$$\text{conv } S \subset \text{conv}(a^T x \leq b) = (a^T x \leq b) \subsetneq \mathbb{R}^n$$

follows.

For the implication Claim 3 \Rightarrow Claim 1, let $z \in \mathbb{R}^n \setminus \text{conv } S$. By the Separating Hyperplane Theorem (see, e.g., Theorem 4.4 in [35]), we may separate z from $\text{conv } S$ by a hyperplane $H(a, b)$ with $a^T x \leq b$ for all $x \in \text{conv } S$, and this hyperplane yields a feasible solution to Program V1.

To see Claim 1 \Rightarrow Claim 2, we construct an optimal solution that corresponds to a supporting hyperplane at a suitably chosen point on the boundary of the closure of the

convex hull of S . So let $(a^T x \leq b)$ be an inequality that is valid for S and thus $\text{conv } S$. Moreover, as half-spaces are closed, $(a^T x \leq b)$ remains valid for $C := \text{cl conv } S$, and we conclude $C \subsetneq \mathbb{R}^n$. Also, C is convex as it is the closure of a convex set (see, e.g., Corollary 11.5.1 in [34]). As $q \in S \subset C$, C is a nonempty, proper closed subset of \mathbb{R}^n , so its boundary $B := \text{bd } C$ is nonempty. As B is closed, (5) ensures the existence of $x_1 \in B$ with $d(q, x_1) = d(q, B)$. By the Supporting Hyperplane Theorem (see, e.g., Chapter 2.5.2, p. 51 in [36]), there is a half-space $(a_1^T x \leq b_1)$ containing C with $a_1^T x_1 = b_1$. We claim that (a_1, b_1) is optimal.

Suppose it is not, so there is (a_2, b_2) such that $(a_2^T x \leq b_2)$ is valid for S and the corresponding hyperplane $H_2 := H(a_2, b_2)$ satisfies the inequality $d_2 := d(q, H_2) < d(q, H_1) =: d_1$. Again there is $x_2 \in H_2$ with $d(q, x_2) = d_2$. We now distinguish two possible locations for x_2 and derive a contradiction in every case.

1. $x_2 \in \mathbb{R}^n \setminus \text{int } C$. As $q \in C$, the line segment from x_2 to q crosses the boundary B of C at a point x_3 . But then $d(x_3, q) \leq d_2 < d_1$, contradicting the optimality of x_1 .
2. $x_2 \in \text{int } C$. Hence there is $\varepsilon > 0$ with $x_2 + \varepsilon a_2 \in C$, and we conclude $a_2^T(x_2 + \varepsilon a_2) = b_2 + \varepsilon a_2^T a_2 > b_2$ as $x_2 \in H(a_2, b_2)$. Consequently, $(a_2^T x \leq b_2)$ is not a valid inequality for C . On the other hand, $S \subset \{x \in \mathbb{R}^n : a_2^T x \leq b_2\}$ by assumption on $(a_2^T x \leq b_2)$. Since half-spaces are convex and closed, we may conclude that also $C = \text{cl conv } S \subset (a_2^T x \leq b_2)$, contradicting our observation that $(a_2^T x \leq b_2)$ is not valid for $x_2 + \varepsilon a_2 \in C$.

We conclude that x_2 cannot exist, so neither can (a_2, b_2) . Hence (a_1, b_1) is an optimal solution to Program V1.

There is nothing to prove for Claim 2 \Rightarrow Claim 1. □

4 Linearizing the Objective

In this section, we linearize the objective function $d(q, H(a, b))$ in Program V1. As a first step, we use an analytic expression for the objective from the literature.

Theorem 4.1 (Theorem 1.1 in [30]) *Let γ be a gauge on \mathbb{R}^n and denote its polar by γ° . Furthermore, let $0 \neq a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then*

$$d(q, H(a, b)) = \begin{cases} (b - a^T q) / \gamma^\circ(a), & a^T q \leq b, \\ (a^T q - b) / \gamma^\circ(-a), & a^T q > b. \end{cases} \tag{9}$$

Let us first note that $\gamma^\circ(a) > 0$: Since γ is a gauge, there is $A \subset \mathbb{R}^n$ closed, convex with $0 \in \text{int } A$ such that $\gamma(\cdot) = \gamma(\cdot; A)$. By (7), we have $\gamma^\circ(\cdot) = \gamma(\cdot; A^\circ)$, and we saw in Sect. 2.4 that A° is also closed, convex with $0 \in \text{int } A^\circ$. Thus $\gamma^\circ(x) = 0$ if and only if $x = 0$, hence $\gamma^\circ(a) > 0$. The variable a enters the fractions in (9) in a nonlinear fashion. Moreover, the constraint $a \neq 0$ is not closed. Now compare Program V1 with the following program with linear objective that avoids a constraint of the form $a \neq 0$:

R1T3

$$\min \quad b - a^T q \tag{V2}$$

$$\begin{aligned} \text{s.t. } & a^T x \leq b \text{ for all } x \in S \\ & \gamma^\circ(a) \geq 1 \\ & a \in \mathbb{R}^n, b \in \mathbb{R}. \end{aligned}$$

It turns out that Programs **V1** and **V2** are closely related. To this end let us introduce the following notion: Two solutions (a, b) and (a', b') are geometrically equivalent if

$$H(a, b) = H(a', b').$$

We say that two programs have geometrically equivalent feasible/optimal solutions if for every feasible/optimal solution (a, b) of the first program there is a geometrically equivalent feasible/optimal solution (a', b') of the second program and vice versa.

Proposition 4.1 *Let $q \in S \subset \mathbb{R}^n$ and γ be a gauge on \mathbb{R}^n . Then, the following hold:*

1. *Programs V1 and V2 have geometrically equivalent feasible solutions.*
2. *The optimal values of both programs coincide.*

In particular, both programs have geometrically equivalent optimal solutions.

Proof By (9) and using the fact that $a^T q \leq b$, Program **V1** is the same as

$$\begin{aligned} \min & (b - a^T q) / \gamma^\circ(a) && \text{(V1')} \\ \text{s.t. } & a^T x \leq b \text{ for all } x \in S \\ & a \neq 0 \\ & a \in \mathbb{R}^n, b \in \mathbb{R}. \end{aligned}$$

For the first claim, let (a, b) be feasible for **V1'**. Then $\gamma^\circ(a)^{-1} \cdot (a, b)$ is feasible for **V2** and geometrically equivalent to (a, b) . On the other hand, if (a, b) is feasible for **V2**, it is also feasible for **V1'**.

For the second claim, we note that the programs **V1'** and **V2** are either both feasible or both infeasible, so in the following we may assume they are feasible. Let z'_1 be the optimal value of **V1'** and z_2 be the optimal value of **V2**. Let (a, b) be feasible for **V1'**. Then $\gamma^\circ(a)^{-1} \cdot (a, b)$ is feasible for **V2** and geometrically equivalent to (a, b) . Furthermore, this shows that $z_2 \leq z'_1$. Now, let (a, b) be feasible for **V2**. Then (a, b) is feasible for **V1'**. Since $\gamma^\circ(a) \geq 1$, we have

$$z'_1 \leq (b - a^T q) / \gamma^\circ(a) \leq (b - a^T q).$$

As (a, b) was an arbitrary feasible solution to **V2**, we have shown that $z'_1 \leq z_2$. The claim about geometrically equivalent optimal solutions follows immediately from the first two statements. □

To summarize, instead of solving **V1** we may solve **V2**.

5 A Linear Semi-Infinite Program for Polyhedral Gauges

Program **V2** contains the non-convex constraint $\gamma^\circ(a) \geq 1$. This constraint can be linearized if we restrict ourselves to polyhedral gauges. This is not a hard restriction since due to [37] every norm can be approximated arbitrarily closely by a block norm, and similarly every gauge by a polyhedral gauge.

We need the following characterization of the facets of a unit ball in terms of the extreme points of the polar polyhedron defined in (6).

Theorem 5.1 (see, e.g., Proposition 3.2, and Theorems 5.3 and 5.5 in Chapter I.4 in [38]) *If P is a full-dimensional and bounded polyhedron and $0 \in \text{int } P$ then*

$$P = \bigcap_{k \in K} \{x \in \mathbb{R}^n : \pi_k^T x \leq 1\}$$

where $\{\pi_k\}_{k \in K}$ are the extreme points of P° . The inequalities $\pi_k^T x \leq 1$ describe exactly the facets of P .

Now suppose γ is a polyhedral gauge. Denote its fundamental directions by $v_1, \dots, v_l \in \mathbb{R}^n$, and the unit ball of the polar gauge γ° by B° . As γ is a polyhedral gauge, so is the polar gauge γ° , cf. Sect. 2. This implies that B° is a polyhedron which is full-dimensional, bounded, with 0 in its interior. Thus, B° satisfies the assumptions of Theorem 5.1, and hence

$$B^\circ = \bigcap_{k \in \hat{K}} \{x \in \mathbb{R}^n : \hat{\pi}_k^T x \leq 1\}$$

where $\{\hat{\pi}_k\}_{k \in \hat{K}}$ are the extreme points of $B^{\circ\circ} := (B^\circ)^\circ$. Using the fact that $B^{\circ\circ} = B$ —this holds for all closed, convex sets containing the origin, see, e.g., Theorem 14.5 in [34]—we have $\{\hat{\pi}_k\}_{k \in \hat{K}} = \{v_1, \dots, v_l\}$, and hence

$$B^\circ = \bigcap_{j \in [l]} \{x \in \mathbb{R}^n : v_j^T x \leq 1\}. \tag{10}$$

We use this characterization as follows.

Corollary 5.1 *Let γ be a polyhedral gauge and denote its fundamental directions by $v_1, \dots, v_l \in \mathbb{R}^n$. Then*

$$\{x \in \mathbb{R}^n : \gamma^\circ(x) \geq 1\} = \bigcup_{j \in [l]} \{x \in \mathbb{R}^n : v_j^T x \geq 1\}, \tag{11}$$

and

$$\{x \in \mathbb{R}^n : \gamma^\circ(x) = 1\} = \bigcup_{j \in [l]} \{x \in \mathbb{R}^n : v_i^T x \leq 1 \forall i \in [l], v_j^T x \geq 1\}. \tag{12}$$

Proof For the interior of the unit ball B° of γ° it holds that

$$\text{int } B^\circ = \{x \in \mathbb{R}^n : \gamma^\circ(x) < 1\}.$$

Since B° is polyhedral, we have from (10)

$$\text{int } B^\circ = \bigcap_{j \in [l]} \{x \in \mathbb{R}^n : v_j^T x < 1\}.$$

This means that the set $\{x \in \mathbb{R}^n : \gamma^\circ(x) \geq 1\}$ equals

$$\mathbb{R}^n \setminus \text{int } B^\circ = \mathbb{R}^n \setminus \bigcap_{j \in [l]} \{x \in \mathbb{R}^n : v_j^T x < 1\} = \bigcup_{j \in [l]} \{x \in \mathbb{R}^n : v_j^T x \geq 1\},$$

proving the first equality. The second equality follows from the fact that

$$\{x \in \mathbb{R}^n : \gamma^\circ(x) = 1\} = \{x \in \mathbb{R}^n : \gamma^\circ(x) \leq 1\} \cap \{x \in \mathbb{R}^n : \gamma^\circ(x) \geq 1\}, \quad (13)$$

and then using distributivity for union and intersection of sets on the explicit representations (10) and (12) of the two sets on the right hand side in (13). \square

The idea now is to use Corollary 5.1 to decompose the nonlinear program V2 into a set of l linear programs, one for each fundamental direction v_j , $j \in [l]$ of the polyhedral gauge γ . These programs are given as

$$\begin{aligned} \min \quad & b - a^T q && (\text{V3}_j) \\ \text{s.t.} \quad & v_j^T a \geq 1 \\ & a^T x \leq b \quad \text{for all } x \in S \\ & a \in \mathbb{R}^n, b \in \mathbb{R}. \end{aligned}$$

The relation between the programs V3_j and V2 is described next.

Proposition 5.1 *Let $q \in S \subset \mathbb{R}^n$ and γ be a polyhedral gauge on \mathbb{R}^n with fundamental directions $v_1, \dots, v_l \in \mathbb{R}^n$. Denote the optimal value of Program V2 by z^* and for V3_j by z_j^* . Let $(a, b) \in \mathbb{R}^n \times \mathbb{R}$. Then the following hold:*

1. (a, b) is a feasible solution of V2 if and only if it is a feasible solution of V3_j for some $j \in [l]$.
2. $z^* = \min_{j \in [l]} z_j^*$.
3. (a, b) is an optimal solution of V2 if and only if there is $j_0 \in [l]$ such that (a, b) is an optimal solution to V3_j for $j = j_0$ with $z_{j_0}^* = \min_{j \in [l]} z_j^*$.

Proof Denote the feasible set of V2 by F' and of V3_j by F_j . From (11) we then have $F' = \bigcup_{j \in [l]} F_j$ and all claims follow easily. \square

Remark 5.1 Note that $V3_j$ has a single linear constraint involving the fundamental directions. This is the reason why in $V2$, we did not use the constraint $\gamma^\circ(a) = 1$ instead of $\gamma^\circ(a) \geq 1$: In view of Corollary 5.1, we would have l additional constraints involving the fundamental directions.

Remark 5.2 For practical purposes let us note that the number l of fundamental directions of a gauge, and therefore the number of programs $V3_j$, can vary tremendously. For example, the gauge given by the 1-norm on \mathbb{R}^n has $l = 2n$ fundamental directions, i.e., is linear in the dimension n , whilst the gauge given by the ∞ -norm on \mathbb{R}^n has $l = 2^n$ fundamental directions, i.e., is exponential in the dimension, see, e.g., p. 5 in [39].

To summarize, instead of solving $V1$ we may solve the linear and semi-infinite programs $V3_j$ for all $j \in [l]$. How the semi-infinite constraint can be circumvented is shown in the next section.

6 An Approximating Hierarchy for Polynomial Constraints and Polyhedral Gauges

In this section we approximate Program $V3_j$ by a hierarchy of sos programs. The main reason is that the constraint

$$a^T x \leq b \quad \text{for all } x \in S$$

is semi-infinite if S contains infinitely many points. There is much literature on semi-infinite programming problems. Classical overview articles are, e.g., [40,41]; a more recent survey is [42]. A bi-level approach is explored in [43]. Also, several numerical solution methods exist, for an overview, we refer to [44–46].

However, in this article we take a different route. Let us explore how semi-infinite constraints can be sidestepped by the requirement of semi-algebraic S and a polyhedral gauge γ . For example when considering MIPP, the set $S = F$ is semi-algebraic. In the following, we use an arbitrary basic closed semi-algebraic set $S = K(h_1, \dots, h_s)$.

With this in mind, we consider the following hierarchy of programs, where $k \in \mathbb{N}$, $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$, $S := K(h_1, \dots, h_s)$ a semi-algebraic set with $q \in S$, and $v_1, \dots, v_l \in \mathbb{R}^n$.

$$\begin{aligned}
 \min \quad & b - a^T q && (\text{VR}_{j,k}) \\
 \text{s.t.} \quad & v_j^T a \geq 1 \\
 & b - \sum_{i=1}^n a_i X_i \in M(h_1, \dots, h_s)[k] \\
 & a \in \mathbb{R}^n, b \in \mathbb{R}.
 \end{aligned}$$

The number k is called the truncation order of program $VR_{j,k}$. Next we show that $VR_{j,k}$ is an sos program, i.e., that it has the form

$$\begin{aligned} \max \quad & c^T y && \text{(SOSP)} \\ \text{s.t.} \quad & p_{i0} + y_1 p_{i1} + \cdots + y_m p_{im} \in \Sigma_n, && i \in [r], \\ & y \in \mathbb{R}^m \end{aligned}$$

for $c \in \mathbb{R}^m$ and fixed polynomials $p_{i0}, p_{ij} \in \mathbb{R}[X_1, \dots, X_n], i \in [r], j \in [m]$, and decision variables $y \in \mathbb{R}^m$. This is helpful since it is possible to solve sos programs. For a detailed introduction to sos programming, we refer to [24,25].

Proposition 6.1 Program $VR_{j,k}$ is an sos program.

Proof As is common in sos programming, a constraint of the form

$$\begin{aligned} p_0 + y_1 p_1 + \cdots + y_m p_m \in M(h_1, \dots, h_s)[k] \\ y \in \mathbb{R}^m \end{aligned}$$

for some $p_i \in \mathbb{R}[X_1, \dots, X_n]$ and $k \in \mathbb{N}$ translates to a classical sos programming constraint as follows: The statement

$$p_0 + y_1 p_1 + \cdots + y_m p_m \in M(h_1, \dots, h_s)[k]$$

is, using the fact that $h_0 = 1$ in the defining equation (4) of $M[k]$, equivalent to

$$\begin{aligned} p_0 + y_1 p_1 + \cdots + y_m p_m - \sum_{j=1}^s \sigma_j h_j \in \Sigma_n \\ \deg \sigma_i \leq k - \deg h_i, \quad i \in [s], \\ \sigma_i \in \Sigma_n, \quad i \in [s]. \end{aligned}$$

The degree bounds ensure that only finitely many real decision variables appear in the σ_i , and thus they can be rewritten by constraints of the form **SOSP**. Note that we tacitly assume $h_i \neq 0$, otherwise we may remove the constraint. Also, linear programming constraints can be used in sos programming, since for $c \in \mathbb{R}$, the requirement $c \geq 0$ is equivalent to $c \in \Sigma_n$. Finally, we note that the objective of Program $VR_{j,k}$ is linear. \square

The next proposition shows that feasible solutions to Program $VR_{j,k}$ yield feasible solutions to Program $V3_j$.

Proposition 6.2 Let $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$ and $S := K(h_1, \dots, h_s)$. Furthermore, let $v_1, \dots, v_l \in \mathbb{R}^n$ be given. If (a, b) is feasible to $VR_{j,k}$ for $j \in [l], k \in \mathbb{N}$, then (a, b) is feasible to $V3_j$.

Proof Let (a, b) feasible to $\text{VR}_{j,k}$. Feasibility implies that

$$b - \sum_{i=1}^n a_i X_i \in M(h_1, \dots, h_s)[k],$$

which imposes $b - a^T x \geq 0$ on S by Observation 2.1. Hence (a, b) is feasible to V3_j . \square

The next theorem shows that, if $M(h_1, \dots, h_s)$ is Archimedean, we get a hierarchy of sos programs indexed by the truncation order k , producing a sequence of valid inequalities for S . Also, as $k \rightarrow \infty$, we can show that the distance of the hyperplanes to the point q is monotonically decreasing and converges to the optimal solution of V1 .

Theorem 6.1 *Let $h_1, \dots, h_s \in \mathbb{R}[X_1, \dots, X_n]$ and $q \in S := K(h_1, \dots, h_s)$. Suppose further that $M(h_1, \dots, h_s)$ is Archimedean. Also, let a polyhedral gauge γ on \mathbb{R}^n with fundamental directions $v_1, \dots, v_l \in \mathbb{R}^n$ be given. Then, it holds:*

1. For every $j \in [l]$, Program $\text{VR}_{j,k}$ is feasible for large enough values of k .
2. Denote the optimal value of V1 by z^* . For $j \in [l]$, $k \in \mathbb{N}$ denote the optimal value of $\text{VR}_{j,k}$ by $z_{j,k}$ and put $z_k = \min_{j \in [l]} z_{j,k}$. Then $z_k \searrow z^*$ for $k \rightarrow \infty$.

Proof Let $M := M(h_1, \dots, h_s)$. To see Claim 1, fix $j \in [l]$. Put $a^* := v_j / \|v_j\|_2$ and note that $v_j^T a^* \geq 1$. By Corollary 2.1, S is compact, hence the map

$$\mu : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto (a^*)^T x,$$

attains its maximum b^* on S . Define the family of polynomials

$$p(a, b) := b - \sum_{i=1}^n a_i X_i \in \mathbb{R}[X_1, \dots, X_n], \quad a \in \mathbb{R}^n, b \in \mathbb{R}.$$

Now fix some $\varepsilon > 0$, and note that $p(a^*, b^* + \varepsilon)$ is positive on S . The Positivstellensatz (Theorem 2.2) ensures that $p(a^*, b^* + \varepsilon) \in M$. There is $k_\varepsilon \in \mathbb{N}$ with $p(a^*, b^* + \varepsilon) \in M[k_\varepsilon]$, and by definition of $M[k]$, $p(a^*, b^* + \varepsilon) \in M[k]$ for $k \geq k_\varepsilon$. In other words, $(a^*, b^* + \varepsilon)$ is feasible for $\text{VR}_{j,k}$ for $k \geq k_\varepsilon$.

Concerning Claim 2 we note that compactness of S implies feasibility of V1 , and hence $z^* < +\infty$. By Theorem 3.1, V1 has an optimal solution (a, b) . By Proposition 4.1 and rescaling if necessary, we may further assume that (a, b) solves V2 to optimality, that is,

$$z^* = b - a^T q.$$

By Proposition 5.1, there is $j_0 \in [l]$ such that (a, b) solves V3_{j_0} to optimality. Let $\varepsilon > 0$. Hence, the linear polynomial $p(a, b + \varepsilon)$ is positive on S and by the Positivstellensatz

lies in $M[k_\varepsilon]$ for some $k_\varepsilon \in \mathbb{N}$. Put differently, $(a, b + \varepsilon)$ is feasible for $\text{VR}_{j_0, k_\varepsilon}$ with objective value

$$b + \varepsilon - v_{j_0}^T a = z^* + \varepsilon.$$

Note that $z^* \leq z_{j_0, k_\varepsilon}$, and our estimates combine to

$$z^* \leq z_{j_0, k_\varepsilon} \leq b + \varepsilon - v_{j_0}^T a = z^* + \varepsilon,$$

and we conclude $z_{j_0, k_\varepsilon} \rightarrow z^*$ for $\varepsilon \rightarrow 0$. On the other hand, we have

$$z^* \leq z_{j, k+1} \leq z_{j, k}, \quad j \in [l], k \in \mathbb{N},$$

since $M[k] \subset M[k + 1]$ by definition of $M[k]$, which yields $z_{j_0, k} \searrow z^*$ for $k \rightarrow +\infty$. The claim $\min_{j \in [l]} z_{j, k} = z_k \searrow z^*$ follows. \square

7 Illustration

We have implemented the hierarchy using SOSTOOLS and SeDuMi and illustrate our results on some examples.⁴ Our implementation is published as open-source software [49].

In our first example, we consider the polynomials

$$h_1 = 1 - X_2^2 - X_1^2, \quad h_2 = X_1 + X_2^3 \in \mathbb{R}[X_1, X_2]. \tag{14}$$

Thus, the set $S = \{x \in \mathbb{R}^2 : h_1(x) \geq 0, h_2(x) \geq 0\} \subset \mathbb{R}^2$ is thus the compact set given by the intersection of the Euclidean unit norm ball and the epigraph of the function $x_1 \mapsto x_2(x_1) = -\sqrt[3]{x_1}$.

Note that $1 - X_1^2 - X_2^2 = 0 \cdot 1 + 1 \cdot h_1 + 0 \cdot h_2 \in M(h_1, h_2)$, hence by Theorem 2.1 the associated quadratic module $M(h_1, h_2)$ is Archimedean, and the hierarchy converges by Theorem 6.1. The point $q = (0.4, -0.5)$ lies in S . We have solved our hierarchy for the polyhedral gauge $\gamma = \|\cdot\|_1$ (in this case a block norm) with fundamental directions $\{(1, 0), (0, 1), (-1, 0), (0, -1)\}$ and two different truncation orders, which we report in Table 1.

Figures 1 and 2 show the vanishing sets $V(h_1)$ and $V(h_2)$ of h_1 and h_2 , that is, $V(h_i) = \{x \in \mathbb{R}^2 : h_i(x) = 0\}$, and the point q . The figures show a computed optimal hyperplane for a low ($k = 2$) and a high ($k = 5$) truncation order k . The optimal solutions and optimal values along with the computation times can be found in Table 1. Allowing for a higher truncation order of $k = 6$ did not improve the result further. We infer from these first examples that a low truncation order of, say, $k = 2$ cannot be expected to give an optimal hyperplane—however, for $k = 2$, we get a valid

⁴ We use MATLAB 2016b 64-bit (MATLAB is a registered trademark of The MathWorks Inc., Natick, Massachusetts), SOSTOOLS 3.01 [47] to translate the sos programs into semidefinite programs and SeDuMi 1.3 [48] to solve the latter. The experiments were conducted on an Intel Core i3 Laptop with 2.6 GHz and 2 cores, 4 GB, running GNU/Linux.

Table 1 Example configurations, computed values and computation times

Program	h_1, h_2	k	q_1	q_2	obj.	a_1	a_2	b	s	Figures
VR $_{j,k}$	(14)	2	0.40	-0.50	0.59	0.80	-1.00	1.41	4.00	1
VR $_{j,k}$	(14)	5	0.40	-0.50	0.24	-0.61	-1.00	0.49	55.38	2
VR $_{j,k}$	(15)	4	0.25	0.50	0.39	0.66	-1.00	0.06	48.66	3
VbR $_{j,k}$	(16)	4	n/a	n/a	-0.71	1.00	1.00	-0.71	60.90	4
VnR $_k$	(16)	4	n/a	n/a	0.58	-1.00	1.00	0.58	40.98	5

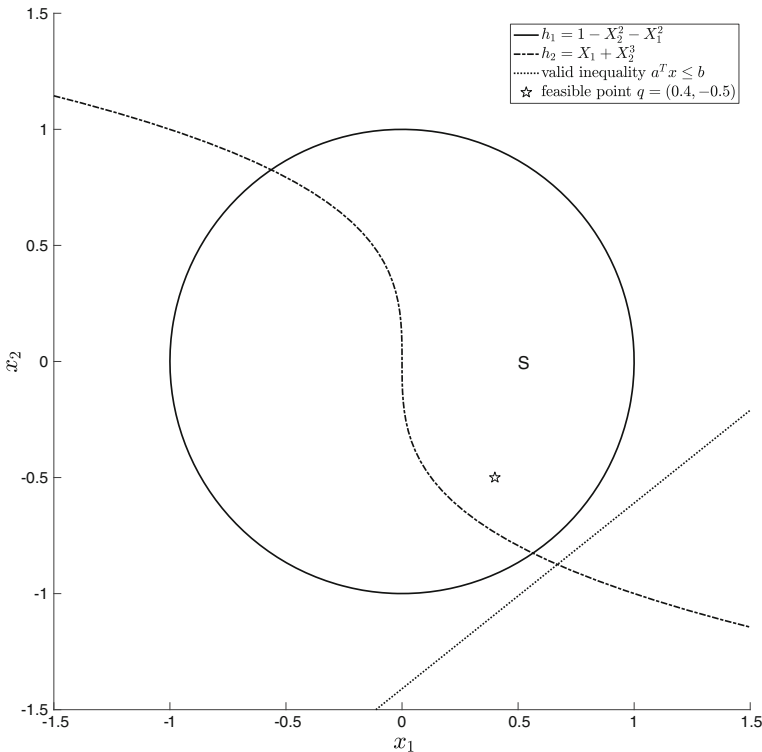


Fig. 1 Bounded example, k low

inequality that already can be used as approximation in a very short computation time. The examples also show that an optimal solution (a, b) for VR $_{j,k}$ does not necessarily yield a tight inequality for S .

An unbounded example (Fig. 3) is given by

$$h_1 = X_2 - X_1^2, \quad h_2 = X_2^2 - X_1. \tag{15}$$

The set $S = \{x \in \mathbb{R}^2 : h_1(x) \geq 0, h_2(x) \geq 0\}$ is the intersection of the filled unit parabola given by $x_2 \geq x_1^2$ and the outside of the rotated unit parabola given by $x_1 \leq x_2^2$. The set S is thus indeed unbounded, hence $M(h_1, h_2)$ cannot be Archimedean.

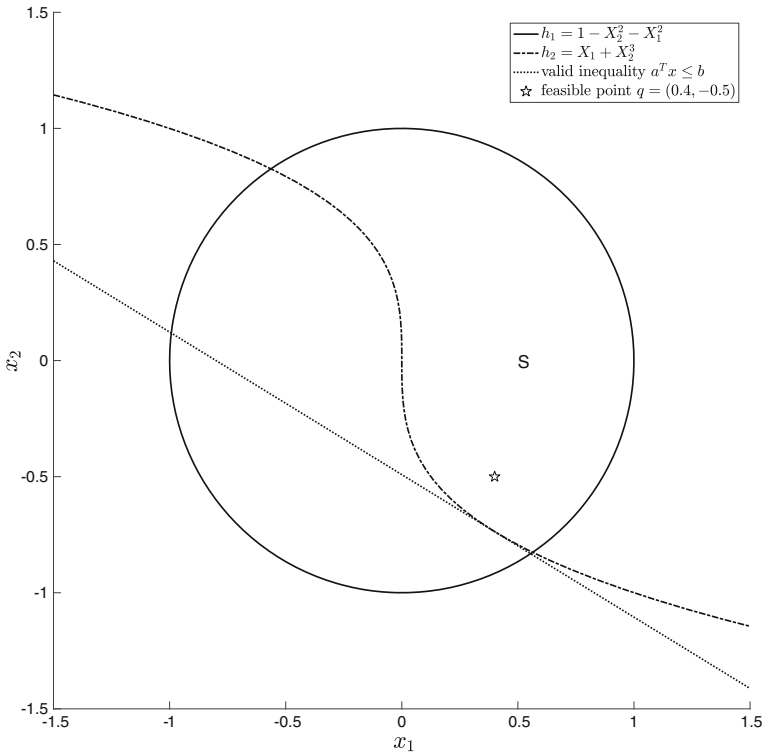


Fig. 2 Bounded example, k high

Nevertheless we can apply our approach (now without convergence guarantee). We choose $q = (0.25, 0.5)$ on the boundary of S . We report the computed values for $k = 4$ in Table 1. This example shows that, even though the Archimedean condition does not hold, it can still be possible to obtain a valid inequality that is close to q and nearly tight. For lower orders ($k = 3$), no solution was found. This can also be concluded directly from (4), as we would have to express a nontrivial linear polynomial as a linear combination of h_1 and h_2 , which is impossible. The objective did not improve by increasing to $k = 5$ or $k = 6$.

8 Modifications and Extensions

In this section we consider some modifications of Program V1. Namely, we consider the case that a point $q \in S$ is not known, and the case that the normal a is fixed. We will state the modifications as general optimization problems (similar to V1) along with a reformulation using sos programming (similar to VR $_{j,k}$).

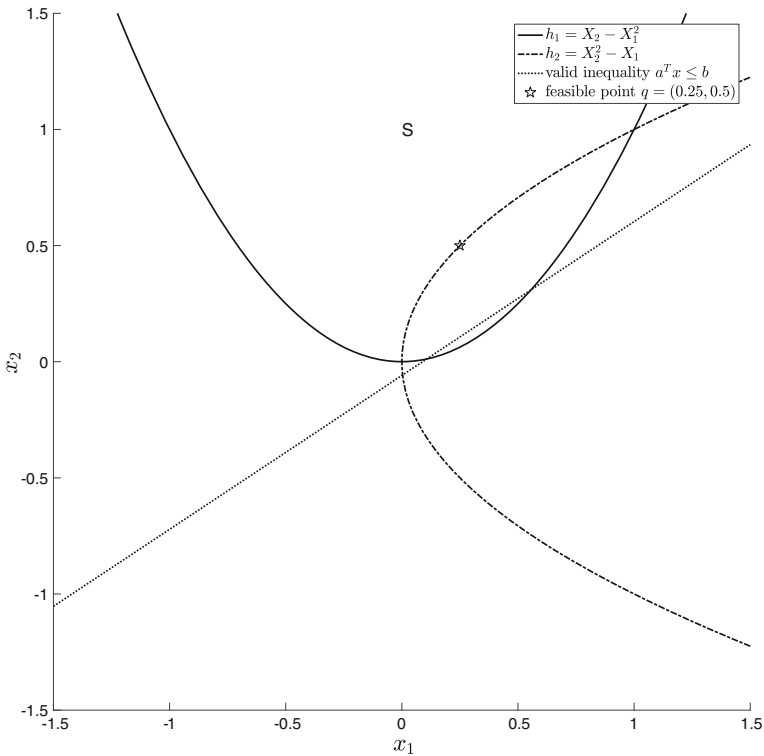


Fig. 3 Unbounded example

8.1 Finding Valid inequalities Without a known \$q \in S\$

As a first modification, we search for a tight valid inequality for \$S\$ without knowing a point \$q \in S\$. We formulate the program and its sos variant with the constraint \$\gamma^\circ(a) = 1\$ that leads to more constraints—cf. Remark 5.1—as follows. As before, let \$S \subset \mathbb{R}^n\$, \$M = M(h_1, \dots, h_s)\$ for \$h_i \in \mathbb{R}[X_1, \dots, X_n]\$, and \$v_j \in \mathbb{R}^n\$.

$$\begin{aligned}
 & \min b && \text{(Vb)} \\
 & \text{s.t. } \gamma^\circ(a) = 1 \\
 & \quad a^T x \leq b \quad \text{for all } x \in S \\
 & \quad a \in \mathbb{R}^n, b \in \mathbb{R}.
 \end{aligned}$$

$$\begin{aligned}
 & \min b && \text{(VbR}_{j,k}) \\
 & \text{s.t. } v_j^T a \geq 1 \\
 & \quad v_i^T a \leq 1, i \in [l],
 \end{aligned}$$

$$b - \sum_{i=1}^n a_i X_i \in M[k]$$

$$a \in \mathbb{R}^n, b \in \mathbb{R}.$$

We require an equation in **Vb** as opposed to **V2** because otherwise the program would be unbounded from below whenever the optimal objective is negative. Let us again state some observations.

Proposition 8.1 *Let $S \subset \mathbb{R}^n$ and γ be a gauge on \mathbb{R}^n .*

1. *Every feasible solution (a, b) of **Vb** yields a valid inequality $(a^T x \leq b)$ for S . If (a, b) is optimal, the inequality is tight.*
2. *Suppose further $S = K(h_1, \dots, h_s)$ for $h_i \in \mathbb{R}[X_1, \dots, X_n]$ and let γ be a polyhedral gauge with fundamental directions v_1, \dots, v_l . If (a, b) is feasible for **VbR_{j,k}**, then (a, b) is feasible for **Vb**.*
3. *Additionally, let $M = M(h_1, \dots, h_s)$ be Archimedean. Then, for $j \in [l]$, **VbR_{j,k}** is feasible for eventually all k . Denote the optimal value of **Vb** by z^* and the optimal value of **VbR_{j,k}** by $z_{j,k}$ and put $z_k = \min_j z_{j,k}$. Then $z_k \searrow z^*$ for $k \rightarrow \infty$.*

Proof Statements 1 and 3 are shown analogously to Proposition 3.1 and Theorem 6.1, respectively. We prove Statement 2. Let (a, b) be a feasible solution for **VbR_{j,k}**. The constraint $b - \sum_{i=1}^n a_i X_i \in M[k]$ implies $b - a^T x \geq 0$ on $S = K(h_1, \dots, h_s)$ by Observation 2.1. By Corollary 5.1, the constraint $\gamma^\circ(a) = 1$ is of the form $v_i^T a \leq 1$ for $i \in [l]$ and $v_j^T a \geq 1$ for some $j \in [l]$. The claim follows. □

For this modification, we consider the following example (Fig. 4):

$$h_1 = 8X_1X_2 - 1, \quad h_2 = \frac{1}{16} - \left(X_1 + \frac{1}{2}\right)^2. \tag{16}$$

The set $S = \{x \in \mathbb{R}^2 : h_1(x) \geq 0, h_2(x) \geq 0\}$ is the intersection of a branch of a hyperbola $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 < 0 \text{ and } x_2 \leq \frac{1}{8x_1}\}$ and a strip given by $\{(x_1, x_2) \in \mathbb{R}^2 : |x_1 + \frac{1}{2}| \leq 4\}$. Hence, S is unbounded. We ran the hierarchy **VbR_{j,k}** for $k = 4$ (let us stress that we did not specify a point $q \in S$). The values of the computed tight inequality are shown in Table 1. No feasible solution could be found for $k = 3$ and the objective did not improve for $k = 5$ or $k = 6$, which is to be expected since Fig. 4 reveals that the hyperplane is tight.

8.2 Fixed Normal

The second modification we consider is the variant where a fixed normal $a \in \mathbb{R}^n$, $a \neq 0$, is given and we want to find $b \in \mathbb{R}$ such that $(a^T x \leq b)$ is valid for S and as tight as possible, i.e., $b \in \mathbb{R}$ is the only decision variable.⁵ The programs read

$$\min b \tag{Vn}$$

⁵ We omit the obvious variation with a zero objective.

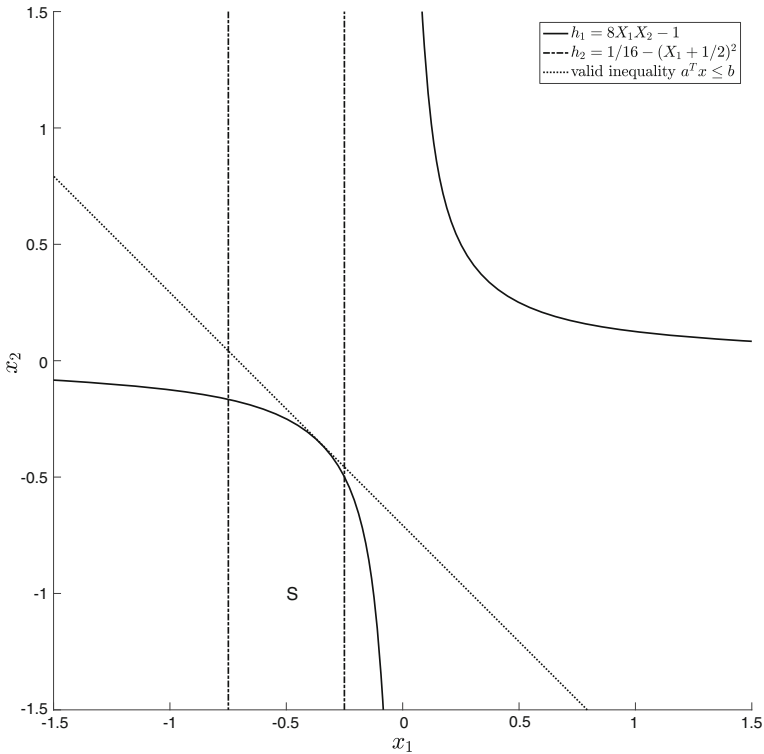


Fig. 4 Example without q, \dots

$$\begin{aligned}
 & \text{s.t. } a^T x \leq b \text{ for all } x \in S \\
 & \quad b \in \mathbb{R}, \\
 & \min b \tag{VnR_k} \\
 & \text{s.t. } b - \sum_{i=1}^n a_i X_i \in M[k] \\
 & \quad b \in \mathbb{R}.
 \end{aligned}$$

In the next result we show that for a fixed normal and provided some $q \in S$ is known, the optimal solutions do not change by replacing the objective by $d(q, H(a, b))$.

Observation 8.1 *Let $q \in S \subset \mathbb{R}^n$, a gauge γ and $0 \neq a \in \mathbb{R}^n$ be given. Consider the program*

$$\begin{aligned}
 & \min d(q, H(a, b)) \\
 & \text{s.t. } a^T x \leq b \text{ for all } x \in S \\
 & \quad b \in \mathbb{R}.
 \end{aligned} \tag{17}$$

Then, Vn and (17) have the same feasible and optimal solutions.

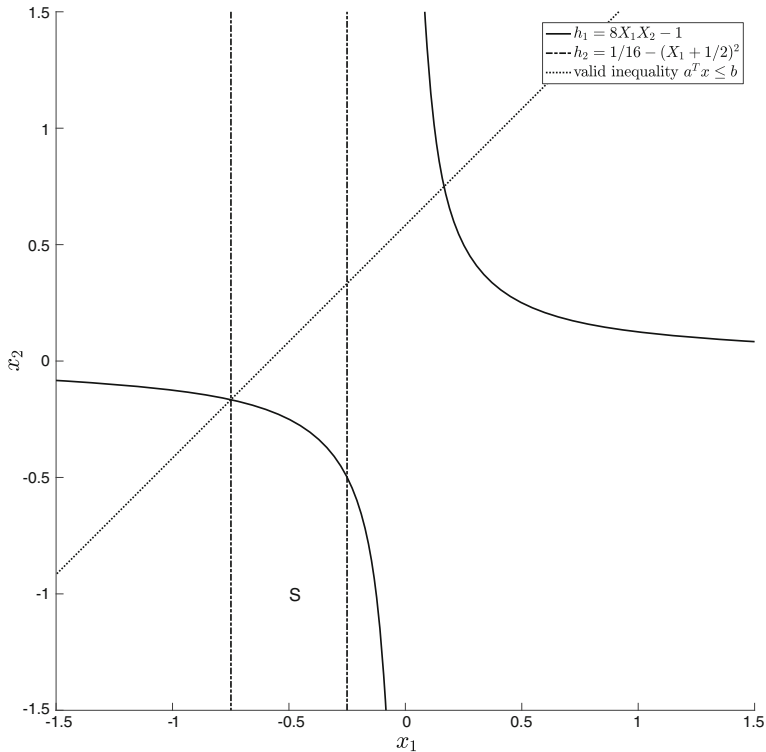


Fig. 5 ...with a fixed, optimized for b

Proof The claim regarding feasibility is trivial. From Theorem 4.1 we know that $d(q, H(a, b)) = (b - a^T q) / \gamma^\circ(a)$. Since a is fixed, both objectives only differ by a constant positive scaling and a constant translation, hence optimal solutions coincide. \square

Let us state some properties of Vn and VnR_k . We omit the proof since it is similar to the proof for the corresponding statements of $V2$ and $VR_{j,k}$.

Proposition 8.2 Let $S \subset \mathbb{R}^n$.

1. Every feasible solution b of Vn yields a valid inequality $(a^T x \leq b)$ for S . If b is optimal, the inequality is tight.
2. Suppose further $S = K(h_1, \dots, h_s)$ for $h_i \in \mathbb{R}[X_1, \dots, X_n]$ and let γ be a polyhedral gauge with fundamental directions v_1, \dots, v_l . If b is feasible for VnR_k , then b is feasible for Vn .
3. Additionally, let $M = M(h_1, \dots, h_s)$ be Archimedean. Then, VnR_k is feasible for eventually all k . Denote the optimal value of Vn by z^* and the optimal value of VnR_k by $z_{j,k}$ and put $z_k = \min_j z_{j,k}$. Then $z_k \searrow z^*$ for $k \rightarrow \infty$.

We illustrate this modification in our last example (Fig. 5), where we use h_1 and h_2 from (16) but fix the normal to $a = (-1, 1)$. We report the computed optimal value of

b in Table 1. Again, no solution was found for $k = 3$ and the objective did not improve for $k = 5$ and $k = 6$, which is again to be expected since the figure reveals that the hyperplane is tight in this example, too.

9 Conclusions

To summarize, we have shown that the problem to find a tight valid inequality for a subset S of \mathbb{R}^n , using a polyhedral gauge γ , can be approximated with sos programming if the set S is semi-algebraic, i.e., if S is given as $K(h_1, \dots, h_s)$ for some polynomials h_1, \dots, h_s . The approximating hierarchy is guaranteed to converge if the quadratic module $M(h_1, \dots, h_s)$ is Archimedean. In view of Remark 2.1, this is the case if S is a bounded set.

Sos programs like ours are computationally tractable on current SDP solvers for small instances (few variables and low degrees of the polynomials). We hence can find cuts for MIPP instances in reasonable time in this case. If the corresponding semidefinite programs become too large for current state-of-the-art solvers, there are promising ideas that keep larger instances tractable, e.g., restrictions to subsets of sos polynomials that translate to linear or second-order cone programs [50,51] as well as column generation [52]. Still, our sos programs translate to SDPs which can, leaving technical details aside, essentially be solved in polynomial time [53]. Note that we cannot expect much more, since MIPP is known to be NP-hard. In the continuous case, this can be seen since deciding nonnegativity of a polynomial of degree 4 is NP-hard [54] and thus minimization of a polynomial of degree 4 is NP-hard. In the integer case, it can be shown that no algorithm for integer programming with quadratic constraints exists [55].

Further research includes to identify situations when a cut may be derived from a given tight valid inequality. That is, given a tight valid inequality for S , are there assumptions that ensure that there is a way to derive a related inequality which is valid for the integer points in S but violated at (a non-integer) point q in S , say? First ideas in this direction are given in [56].

Acknowledgements Open Access funding provided by Projekt DEAL. We are grateful to the Editors and anonymous referees for their helpful comments and suggestions. Funding was provided by DFG (Grant No. RTG 2088).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Meyer, R.R.: On the existence of optimal solutions to integer and mixed-integer programming problems. *Math. Program.* **7**(1), 223–235 (1974)
2. Tawarmalani, M., Sahinidis, N.V.: *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*, vol. 65. Springer, Berlin (2002)
3. Duran, M.A., Grossmann, I.E.: An outer-approximation algorithm for a class of mixed-integer nonlinear programs. *Math. Program.* **36**(3), 307–339 (1986)
4. Fletcher, R., Leyffer, S.: Solving mixed integer nonlinear programs by outer approximation. *Math. Program.* **66**(1), 327–349 (1994)
5. Gomory, R.E.: Outline of an algorithm for integer solutions to linear programs. *Bull. Am. Math. Soc.* **64**, 3 (1958)
6. Marchand, H., Martin, A., Weismantel, R., Wolsey, L.: Cutting planes in integer and mixed integer programming. *Discrete Appl. Math.* **123**(1), 397–446 (2002)
7. Cornuéjols, G., Li, Y.: Elementary closures for integer programs. *Oper. Res. Lett.* **28**(1), 1–8 (2001)
8. Averkov, G., Wagner, C., Weismantel, R.: Maximal lattice-free polyhedra: finiteness and an explicit description in dimension three. *Math. Oper. Res.* **36**(4), 721–742 (2011)
9. Kelley Jr., J.E.: The cutting-plane method for solving convex programs. *J. Soc. Ind. Appl. Math.* **8**(4), 703–712 (1960)
10. Boyd, S., Vandenberghe, L.: Localization and cutting-plane methods. Stanford EE 364b lecture notes (2011). https://see.stanford.edu/materials/lsocoe364b/05-localization_methods_notes.pdf
11. Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., Mahajan, A.: Mixed-integer nonlinear optimization. *Acta Numerica* **22**, 1–131 (2013)
12. Kesavan, P., Allgor, R.J., Gatzke, E.P., Barton, P.I.: Outer approximation algorithms for separable nonconvex mixed-integer nonlinear programs. *Math. Program.* **100**(3), 517–535 (2004)
13. Bienstock, D., Chen, C., Munoz, G.: Outer-product-free sets for polynomial optimization and oracle-based cuts. *Math. Program.* 1–44 (2020)
14. Conforti, M., Cornuéjols, G., Daniilidis, A., Lemaréchal, C., Malick, J.: Cut-generating functions and S-free sets. *Math. Oper. Res.* **40**(2), 276–391 (2015)
15. Basu, A., Dey, S.S., Paat, J.: Nonunique lifting of integer variables in minimal inequalities. *SIAM J. Discrete Math.* **33**(2), 755–783 (2019)
16. Towle, E., Luedtke, J.: Intersection disjunctions for reverse convex sets. arXiv preprint [arXiv:1901.02112](https://arxiv.org/abs/1901.02112) (2019)
17. Muñoz, G., Serrano, F.: Maximal quadratic-free sets. In: *International Conference on Integer Programming and Combinatorial Optimization*, pp. 307–321. Springer (2020)
18. Martini, H., Schöbel, A.: Median hyperplanes in normed spaces—a survey. *Discrete Appl. Math.* **89**, 181–195 (1998)
19. Schöbel, A.: Chap. 7: Location of Dimensional Facilities in a Continuous Space. In: Laporte, G., Nickel, S., da Gama, F.S. (eds.) *Location Science*, pp. 63–103. Springer, Berlin (2015)
20. Shor, N.Z.: Class of global minimum bounds of polynomial functions. *Cybern. Syst. Anal.* **23**(6), 731–734 (1987)
21. Shor, N.Z., Stetsyuk, P.I.: Modified r-algorithm to find the global minimum of polynomial functions. *Cybern. Syst. Anal.* **33**(4), 482–497 (1997)
22. Parrilo, P.A.: Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization. Ph.D. thesis. Citeseer (2000)
23. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. *SIAM J. Optim.* **11**(3), 796–817 (2001)
24. Anjos, M., Lasserre, J.B.: *Handbook on Semidefinite, Conic and Poly-Nomial Optimization*, vol. 166. Springer, Berlin (2012)
25. Blekherman, G., Parrilo, P.A., Thomas, R.R.: *Semidefinite Optimization and Convex Algebraic Geometry*, vol. 13. SIAM, Philadelphia (2013)
26. Marshall, M.: *Positive Polynomials and Sums of Squares. Mathematical Surveys and Monographs*, vol. 146. American Mathematical Society, Providence (2008)
27. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: Putinar M., Sullivant S. (eds.) *Emerging Applications of Algebraic Geometry*, pp. 157–270. Springer (2009)

28. Hemmecke, R., Köppe, M., Lee, J., Weismantel, R.: Nonlinear integer programming. In: Jünger, M., Liebling, T.M., Naddef, D., Nemhauser, G.L., Pulleyblank, W.R., Reinelt, G., Rinaldi, G., Wolsey, L.A. (eds.) *50 Years of Integer Programming 1958–2008*, pp. 561–618. Springer (2010)
29. Lee, J., Leyffer, S.: *Mixed Integer Nonlinear Programming*. Springer, Berlin (2012)
30. Plastria, F., Carrizosa, E.: Gauge distances and median hyperplanes. *J. Optim. Theory Appl.* **110**(1), 173–182 (2001)
31. Wolkowicz, H., Saigal, R., Vandenberghe, L.: *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, vol. 27. Springer, Berlin (2000)
32. Vandenberghe, L., Boyd, S.: Semidefinite programming. *SIAM Rev.* **38**(1), 49–95 (1996)
33. Jeyakumar, V., Lasserre, J.B., Li, G.: On polynomial optimization over non-compact semi-algebraic sets. *J. Optim. Theory Appl.* **163**(3), 707–718 (2014)
34. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
35. Gruber, P.: *Convex and Discrete Geometry*, vol. 336. Springer, Berlin (2007)
36. Boyd, S., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
37. Ward, J.E., Wendell, R.E.: Using block norms for location modeling. *Oper. Res.* **33**(5), 1074–1090 (1985)
38. Nemhauser, G.L., Wolsey, L.A.: *Integer and Combinatorial Optimization*. Wiley-Interscience, Hoboken (1988)
39. Schöbel, A.: *Locating Lines and Hyperplanes: Theory and Algorithms*, vol. 25. Springer, Berlin (1999)
40. Hettich, R., Kortanek, K.O.: Semi-infinite programming: theory, methods, and applications. *SIAM Rev.* **35**(3), 380–429 (1993)
41. Reemtsen, R., Rückmann, J.-J.: *Semi-Infinite Programming*, vol. 25. Springer, Berlin (1998)
42. Stein, O.: How to solve a semi-infinite optimization problem. *Eur. J. Oper. Res.* **223**(2), 312–320 (2012)
43. Stein, O.: *Bi-Level Strategies in Semi-infinite Programming*, vol. 71. Springer, Berlin (2013)
44. Reemtsen, R., Görner, S.: Numerical methods for semi-infinite programming: a survey. In: Reemtsen, R., Rückmann, J.J. (eds.) *Semi-infinite Programming*, pp. 195–275. Springer (1998)
45. López, M., Still, G.: Semi-infinite programming. *Eur. J. Oper. Res.* **180**(2), 491–518 (2007)
46. Vázquez, F.G., Rückmann, J.-J., Stein, O., Still, G.: Generalized semi-infinite programming: a tutorial. *J. Comput. Appl. Math.* **217**(2), 394–419 (2008)
47. Papachristodoulou, A., Anderson, J., Valmorbida, G., Prajna, S., Seiler, P., Parrilo, P.A.: SOSTOOLS: Sum of squares optimization toolbox for MATLAB (2016). <http://www.cds.caltech.edu/sostools>. Accessed 2 Aug 2016
48. Sturm, J.F.: Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.* **11**(1–4), 625–653 (1999)
49. Behrends, S.: vis3p: Valid Inequalities for Semi-algebraic Sets using Sos Programming. Version 6523d85 (2018). <https://github.com/sbehen/vis3p>. Accessed 2 Mar 2018
50. Ahmadi, A.A., Hall, G.: Sum of squares basis pursuit with linear and second order cone programming. *Algebraic Geom. Methods Discrete Math.* **685**, 27–53 (2017)
51. Ahmadi, A.A., Hall, G.: On the construction of converging hierarchies for polynomial optimization based on certificates of global positivity. *Math. Oper. Res.* **44**(4), 1192–1207 (2019)
52. Ahmadi, A.A., Dash, S., Hall, G.: Optimization over structured subsets of positive semidefinite matrices via column generation. *Discrete Optim.* **24**, 129–151 (2017)
53. Lovász, L.: Semidefinite programs and combinatorial optimization. In: Reed, B.A., Sales, C.L. (eds.) *Recent Advances in Algorithms and Combinatorics*, pp. 137–194. Springer (2003)
54. Ahmadi, A.A., Olshevsky, A., Parrilo, P.A., Tsitsiklis, J.N.: NPhardness of deciding convexity of quartic polynomials and related problems. *Math. Program.* **137**, 1–24 (2013)
55. Jeroslow, R.C.: There cannot be any algorithm for integer programming with quadratic constraints. *Oper. Res.* **21**(1), 221–224 (1973)
56. Behrends, S.: Geometric and algebraic approaches to mixed-integer polynomial optimization using sos programming. Ph.D. thesis. Universität Göttingen (2017). <http://hdl.handle.net/11858/00-1735-0000-0023-3F9C-9>