Dual certificates and efficient rational sum-of-squares decompositions for polynomial optimization over compact sets

Maria L. Macaulay[a], Dávid Papp[a]

[a]*North Carolina State University, Department of Mathematics, Raleigh NC, USA*

Abstract

We study the problem of computing rational weighted sum-of-squares (WSOS) certificates for positive polynomials over a compact semialgebraic set. Building on the theory of interior-point methods for convex optimization, we introduce the concept of dual cone certificates, which allows us to interpret vectors from the dual of the sum-of-squares cone as rigorous nonnegativity certificates of a WSOS polynomial. Whereas conventional WSOS certificates are alternative representations of the polynomials they certify, dual certificates are distinct from the certified polynomials; moreover, each dual certificate certifies an entire proper cone of WSOS polynomials. As a result, rational WSOS certificates can be constructed from numerically computed dual certificates at little additional cost, without any rounding or projection steps applied to the numerical certificates. As an additional algorithmic application, we present an almost entirely numerical hybrid algorithm for computing the optimal WSOS lower bound of a given polynomial along with a rational dual certificate, with a polynomial-time computational cost per iteration and linear rate of convergence.

1. Introduction

Deciding whether a polynomial is nonnegative on an (often compact) semialgebraic set, and the closely related problem of computing the (approximate) minimum value of a polynomial, is a fundamental problem of computational algebraic geometry and theoretical computer science, with many applications from discrete geometry and algorithmic theorem proving to the design and analysis of dynamical systems such as power networks, to name a few. This problem is well-known to be decidable [1, 2] but strongly NP-hard. The perhaps most studied, and arguably practically most successful, computational approach to it has been to certify the nonnegativity of the polynomial by writing it as a (weighted) sum of squared polynomials—a technique known as *sum-of-squares decomposition*; a variety of results from real algebraic geometry such as Putinar's *Positivstellensatz* [3] guarantee that every polynomial that is strictly positive over a compact semialgebraic set has such a representation.

Weighted sum-of-squares (WSOS) decompositions are usually computed numerically, using semidefinite programming (e.g., [4, 5, 6]) or non-symmetric cone optimization [7],

Email addresses: mlmacaul@ncsu.edu (Maria L. Macaulay), dpapp@ncsu.edu (Dávid Papp)

Submitted to MEGA 2021

which is sufficient in many of the practical applications mentioned above. However, in many theoretical contexts, such as in computational algebraic geometry and automated theorem proving, it is required that the computed certificates be certified rigorously, in exact arithmetic.

Computing rational WSOS decompositions for polynomials with rational coefficients is a challenging problem even in the univariate case [8]. Symbolic methods such as those that rely on quantifier elimination or root isolation are frequently exponential in the degree of the input polynomial. The optimal value of the semidefinite program is an algebraic number, but the study of the algebraic degree of the positive semidefinite cone [9] suggests that one cannot hope for easily computable and verifiable certificates from taking a purely symbolic computing approach to the semidefinite programming problems that come from sums-of-squares. Therefore, a number of authors have proposed hybrid methods that "round" or "project" efficiently computable but inexact numerical sum-of-squares certificates to rigorous rational ones [10, 11, 12]; see also [13, 14, 15].

Our contribution is twofold. In Section 2, we propose a new framework for certifying that a polynomial is WSOS. The approach relies on convex programming duality and allows the efficient construction of rational WSOS decompositions from suitable vectors from the dual cone. In contrast to conventional WSOS certificates, which can be viewed as different representations of the polynomial whose nonnegativity they certify, dual certificates are distinct from the certified polynomials themselves—in particular, every polynomial in the interior of the WSOS cone has a full-dimensional cone of dual certificates, which makes it particularly easy to identify one with an efficient numerical method.

In Section 3, we discuss various algorithmic applications of dual certificates. We propose an efficient hybrid algorithm for computing and certifying rational WSOS lower bounds for polynomials over a compact semialgebraic set using this framework. The algorithm is almost entirely numerical, and has a considerably lower computational complexity than off-the-shelf semidefinite programming software applied to the same problem. The algorithm provides, in each iteration, a certifiable WSOS bound with a dual certificate that can be converted (in polynomial time) to a rational WSOS certificate without any additional rounding or projection of the numerical solutions. The sequence of WSOS bounds converges to the optimal WSOS bound at a linear convergence rate.

1.1. Preliminaries

In the rest of this section we introduce some notation and briefly review some convex optimization and interior-point theory that we rely on throughout the paper.

1.1.1. Weighted SOS polynomials and positive semidefinite matrices

Recall that a convex set $K \subseteq \mathbb{R}^n$ is called a *convex cone* if for every $\mathbf{x} \in K$ and $\lambda \geq 0$ scalar, the vector $\lambda\mathbf{x}$ also belongs to K. A convex cone is *proper* if it is closed, *full-dimensional* (meaning $\operatorname{span}(K) = \mathbb{R}^n$), and *pointed* (that is, it does not contain a line). We shall denote the interior of a proper cone K by K°.

Sum-of-squares (SOS) polynomials. Let $\mathcal{V}_{n,2d}$ denote the cone of n-variate polynomials of degree $2d$. We say that a polynomial $p \in \mathcal{V}_{n,2d}$ is *sum-of-squares* (SOS) if there exist polynomials $q_1, \ldots, q_k \in \mathcal{V}_{n,d}$ such that $p = \sum_{i=1}^{k} q_i^2$. Define $\Sigma_{n,2d}$ to be the cone of

n-variate SOS polynomials of degree $2d$. The cone $\Sigma_{n,2d} \subset \mathcal{V}_{n,2d} \equiv \mathbb{R}^{\binom{n+2d}{n}}$ is a proper cone for every n and d.

Weighted sum-of-squares. More generally, let $\mathbf{g} = (g_1, \ldots, g_m)$ be some given nonzero polynomials and let $\mathbf{d} = (d_1, \ldots, d_m)$ be a nonnegative integer vector. We denote by $\mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$ the space of polynomials p for which there exist $r_1 \in \mathcal{V}_{n,2d_1}, \ldots, r_m \in \mathcal{V}_{n,2d_m}$ such that $p = \sum_{i=1}^m g_i r_i$. A polynomial $p \in \mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$ is said to be *weighted sum-of-squares* (WSOS) if there exist $\sigma_1 \in \Sigma_{n,2d_1}, \ldots, \sigma_m \in \Sigma_{n,2d_m}$ such that $p = \sum_{i=1}^m g_i \sigma_i$. It is customary to assume that $g_1 = 1$, that is, the ordinary "unweighted" sum-of-squares polynomials are also included in the WSOS cones. Let $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$ denote the set of WSOS polynomials in $\mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$. Under mild conditions, the cone $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}} \subset \mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$ is a proper cone; for example, it is sufficient that the set

$$\{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) > 0,\ i = 1, \ldots, m\}$$

is a unisolvent point set for the space $\mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$ [7, Prop. 6.1].

WSOS polynomials and positive semidefinite matrices. We will denote the set of $n \times n$ real symmetric matrices by $\mathbb{S}^n$, and the cone of positive semidefinite $n \times n$ real symmetric matrices by $\mathbb{S}^n_+$. When the dimension is clear from the context, we use the common shorthands $\mathbf{A} \succcurlyeq 0$ to denote that the matrix $\mathbf{A}$ is positive semidefinite and $\mathbf{A} \succ 0$ to denote that the matrix $\mathbf{A}$ is positive definite. We will routinely identify polynomials with their coefficient vectors in a fixed basis of $\mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$.

The following well-known theorem (rooted in the works of Shor, Lasserre, Parrilo, and Nesterov; here reproduced in the notation of the latter) illustrates the connection between $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$ and the cone of positive semidefinite matrices.

Proposition 1 ([16, Thm. 17.6])**.** *Fix an ordered basis $\mathbf{q} = (q_1, \ldots, q_U)$ of $\mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$ and an ordered basis $\mathbf{p}_i = (p_{i,1}, \ldots, p_{i,L_i})$ of $\mathcal{V}_{n,d_i}$ for $i = 1, \ldots, m$. Let $\Lambda_i : \mathbb{R}^U \to \mathbb{S}^{L_i}$ be the unique linear mapping satisfying $\Lambda_i(\mathbf{q}) = g_i \mathbf{p}_i \mathbf{p}_i^T$, and let Λ_i^* denote its adjoint. Then $\mathbf{s} \in \Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$ if and only if there exist matrices $\mathbf{S}_1 \succcurlyeq \mathbf{0}, \ldots, \mathbf{S}_m \succcurlyeq \mathbf{0}$ satisfying*

$$\mathbf{s} = \sum_{i=1}^m \Lambda_i^*(\mathbf{S}_i). \tag{1}$$

Additionally, the dual cone of $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$ admits the characterization

$$\left(\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}\right)^* = \{\mathbf{x} \in \mathbb{R}^U \mid \Lambda_i(\mathbf{x}) \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m\}. \tag{2}$$

The proof of Proposition 1 is constructive: given matrices $\mathbf{S}_i \in \mathbb{S}^{L_i}_+$ $(i = 1, \ldots, m)$, one may explicitly construct a (weighted) sum-of-squares decomposition of the polynomial $\mathbf{s}$. Thus, the collection of matrices $(\mathbf{S}_1, \ldots, \mathbf{S}_m)$ itself can be interpreted as a WSOS certificate of the polynomial $\mathbf{s}$.

To lighten the notation, throughout the rest of the paper we assume that the weight polynomials $\mathbf{g} = (g_1, \ldots, g_m)$ and the degrees $\mathbf{d} = (d_1, \ldots, d_m)$ are fixed, and denote the cone $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$ by Σ and the space of polynomials $\mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}}$ by $\mathcal{V}$. Additionally, we denote by Λ the $\mathbb{R}^U \mapsto \mathbb{S}^{L_1} \oplus \cdots \oplus \mathbb{S}^{L_m}$ linear map $\Lambda_1(\cdot) \oplus \cdots \oplus \Lambda_m(\cdot)$ from Proposition 1. With this

notation, the condition (1) can be written as $\mathbf{s} = \Lambda^*(\mathbf{S})$ for some positive semidefinite (block diagonal) matrix $\mathbf{S} \in \mathbb{S}^{L_1} \oplus \cdots \oplus \mathbb{S}^{L_m}$. Similarly, Eq. (2) simplifies to

$$\Sigma^* = \{\mathbf{x} \in \mathbb{R}^U \mid \Lambda(\mathbf{x}) \succcurlyeq \mathbf{0}\}. \tag{3}$$

The interior of this cone is simply

$$(\Sigma^*)^\circ = \{\mathbf{x} \in \mathbb{R}^U \mid \Lambda(\mathbf{x}) \succ \mathbf{0}\}. \tag{4}$$

1.1.2. Barrier functions and local norms in convex cones

The analysis of the dual certificates introduced in Section 2 relies heavily on the theory of barrier functions for convex cones. In this section, we give a brief overview of the parts of this theory and some additional notation that will be needed throughout the rest of the paper.

It is convenient to identify the spaces $\mathcal{V}$ and $\mathcal{V}^*$ with $\mathbb{R}^U$ ($U = \dim(\mathcal{V})$), equipped with the standard inner product, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^{\mathrm{T}}\mathbf{y}$, and the induced Euclidean norm $\|\cdot\|$.

Let $\Lambda : \mathbb{R}^U \to \mathbb{S}^L$ be the unique linear mapping specified in Proposition 1 above, and let Λ^* denote its adjoint. Central to our theory is the *barrier function* $f : (\Sigma^*)^\circ \to \mathbb{R}$ defined by

$$f(\mathbf{x}) \stackrel{\text{def}}{=} -\ln(\det(\Lambda(\mathbf{x})). \tag{5}$$

Note that by Eq. (4), f is indeed defined on its domain. The function f is twice continuously differentiable; we denote by $g(\mathbf{x})$ its gradient at $\mathbf{x}$ and by $H(\mathbf{x})$ its Hessian at $\mathbf{x}$. Since f is strictly convex on its domain, $H(\mathbf{x}) \succ \mathbf{0}$ for all $\mathbf{x} \in (\Sigma^*)^\circ$. Consequently, we can also associate with each $\mathbf{x} \in (\Sigma^*)^\circ$ the *local inner product* $\langle \cdot, \cdot \rangle_{\mathbf{x}}$ in $\mathcal{V}^*$ defined as $\langle \mathbf{y}, \mathbf{z} \rangle_{\mathbf{x}} \stackrel{\text{def}}{=} \mathbf{y}^{\mathrm{T}} H(\mathbf{x}) \mathbf{z}$ and the *local norm* $\|\cdot\|_{\mathbf{x}}$ induced by this local inner product. Thus, $\|\mathbf{y}\|_{\mathbf{x}} = \|H(\mathbf{x})^{1/2}\mathbf{y}\|$. We define the local (open) ball centered at $\mathbf{x}$ with radius r by $B_{\mathbf{x}}(\mathbf{x}, r) \stackrel{\text{def}}{=} \{\mathbf{y} \in \mathcal{V}^* \mid \|\mathbf{y} - \mathbf{x}\|_{\mathbf{x}} < r\}$. Analogously, we define the *dual local inner product* $\langle \cdot, \cdot \rangle_{\mathbf{x}}^* : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ by $\langle \mathbf{s}, \mathbf{t} \rangle_{\mathbf{x}}^* \stackrel{\text{def}}{=} \mathbf{s}^{\mathrm{T}} H(\mathbf{x})^{-1} \mathbf{t}$; the induced *dual local norm* $\|\cdot\|_{\mathbf{x}}^*$ satisfies the identity $\|\mathbf{t}\|_{\mathbf{x}}^* = \|H(\mathbf{x})^{-1/2}\mathbf{t}\|$.

We remark that the function in (5) falls into the broader category of *logarithmically homogeneous self-concordant barriers* (or LHSCBs for short), which are expounded upon in the classic texts [17] and [18]. Throughout, we will invoke several useful results concerning LHSCBs for the function (5); these are enumerated in the following lemma:

Lemma 1. *Using the notation introduced in this section, the following hold for every* $\mathbf{x} \in (\Sigma^*)^\circ$*:*

1. *We have* $B_{\mathbf{x}}(\mathbf{x}, 1) \subset (\Sigma^*)^\circ$*, and for all* $\mathbf{u} \in B_{\mathbf{x}}(\mathbf{x}, 1)$ *and* $\mathbf{v} \neq 0$*, one has*

$$1 - \|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}} \leq \frac{\|\mathbf{v}\|_{\mathbf{u}}}{\|\mathbf{v}\|_{\mathbf{x}}} \leq (1 - \|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}})^{-1}. \tag{6}$$

2. *The gradient* g *of* f *can be computed as*

$$g(\mathbf{x}) = -\Lambda^*(\Lambda(\mathbf{x})^{-1}), \tag{7}$$

and the Hessian $H(\mathbf{x})$ *is the linear operator satisfying*

$$H(\mathbf{x})\mathbf{w} = \Lambda^*\big(\Lambda(\mathbf{x})^{-1}\Lambda(\mathbf{w})\Lambda(\mathbf{x})^{-1}\big) \quad \textit{for every } \mathbf{w} \in \mathbb{R}^U. \tag{8}$$

3. *The function f is* logarithmically homogeneous*; that is, it has the following properties:*

$$g(\alpha\mathbf{x}) = \alpha^{-1}g(\mathbf{x}) \quad \textit{and} \quad H(\alpha\mathbf{x}) = \alpha^{-2}H(\mathbf{x}) \quad \textit{for every } \alpha > 0, \tag{9}$$

furthermore

$$H(\mathbf{x})\mathbf{x} = -g(\mathbf{x}) \quad \textit{and} \quad \|g(\mathbf{x})\|_{\mathbf{x}}^* = \|\mathbf{x}\|_{\mathbf{x}} = \sqrt{\langle -g(\mathbf{x}), \mathbf{x}\rangle} = \sqrt{\nu}, \tag{10}$$

where $\nu = \sum_{i=1}^m L_i$ is the barrier parameter *of f.*

4. *The gradient map $g : (\Sigma^*)^\circ \to \mathbb{R}^U$ defines a bijection between $(\Sigma^*)^\circ$ and Σ°, In particular, for every $\mathbf{s} \in \Sigma^\circ$ there exists a unique $\mathbf{x} \in (\Sigma^*)^\circ$ satisfying $\mathbf{s} = -g(\mathbf{x})$.*
5. *If $\|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}} < 1$, then*

$$\|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^* \leq \frac{\|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}}}{1 - \|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}}}. \tag{11}$$

6. *If $\|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^* < 1$, then*

$$\|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}} \leq \frac{\|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^*}{1 - \|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^*}. \tag{12}$$

Proof.

1. This is Renegar's definition of self-concordance applied to the function f, which is a composition of an affine function and a well-known self-concordant function, and is thus self-concordant; see [18, Sec. 2.2.1 and Thm. 2.2.7].
2. Straightforward calculation.
3. Straightforward calculation using the identities (7) and (8). We remark that these identities hold for all LHSCBs [18, Thm. 2.3.9].
4. See [18, Sec. 3.3].
5. See [19, Lemma 5].
6. This is an application of the previous claim to the conjugate barrier function of f. □

2. Dual certificates

We begin this section by introducing our central object, the cone of dual certificates corresponding to a WSOS polynomial (Definition 1) and showing in Theorem 1 how we can use dual certificates to construct an explicit (weighted) sum-of-squares decomposition of WSOS polynomials in closed form. We continue using the notation introduced in the previous section, and let Σ denote a general WSOS cone $\Sigma_{n,2\mathbf{d}}^{\mathbf{g}}$ and H denote the Hessian of the barrier function f defined in (5).

Definition 1. Let $\mathbf{s} \in \Sigma$. We say that the vector $\mathbf{x} \in (\Sigma^*)^\circ$ is a *dual certificate of* $\mathbf{s}$, or simply that $\mathbf{x}$ *certifies* $\mathbf{s}$, if $H(\mathbf{x})^{-1}\mathbf{s} \in \Sigma^*$. We denote by

$$\mathcal{C}(\mathbf{s}) \stackrel{\text{def}}{=} \{\mathbf{x} \in (\Sigma^*)^\circ \mid H(\mathbf{x})^{-1}\mathbf{s} \in \Sigma^*\}$$

the set of dual certificates of $\mathbf{s}$. Conversely, for every $\mathbf{x} \in \Sigma^\circ$, we denote by

$$\mathcal{P}(\mathbf{x}) \stackrel{\text{def}}{=} \{\mathbf{s} \in \Sigma \,|\, H(\mathbf{x})^{-1}\mathbf{s} \in \Sigma^*\}$$

the set of polynomials certified by the dual vector $\mathbf{x}$.

The following theorem justifies the terminology introduced above. Through Eq. (13) below, we can construct a WSOS certificate $\mathbf{S}$ for the polynomial $\mathbf{s}$ in the spirit of Proposition 1 by an efficiently-computable closed-form formula, and thus we may interpret the dual vector $\mathbf{x} \in \mathcal{C}(\mathbf{s})$ itself as a certificate of the polynomial $\mathbf{s}$.

Theorem 1. *Let $\mathbf{x} \in (\Sigma^*)^\circ$ be arbitrary. Then the matrix $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{s})$ defined by*

$$\mathbf{S}(\mathbf{x}, \mathbf{s}) \stackrel{\text{def}}{=} \Lambda(\mathbf{x})^{-1}\Lambda\big(H(\mathbf{x})^{-1}\mathbf{s}\big)\,\Lambda(\mathbf{x})^{-1} \tag{13}$$

satisfies $\Lambda^*(\mathbf{S}) = \mathbf{s}$. Moreover, $\mathbf{x}$ is a dual certificate for $\mathbf{s} \in \Sigma$ if and only if $\mathbf{S} \succcurlyeq 0$.

Proof. The first statement can shown by applying the Hessian formula from Lemma 1:

$$\Lambda^*(\mathbf{S}) \stackrel{(13)}{=} \Lambda^*\left(\Lambda(\mathbf{x})^{-1}\Lambda\big(H(\mathbf{x})^{-1}\mathbf{s}\big)\,\Lambda(\mathbf{x})^{-1}\right) \stackrel{(8)}{=} H(\mathbf{x})H(\mathbf{x})^{-1}\mathbf{s} = \mathbf{s},$$

For the second statement, note that $\mathbf{S} \succcurlyeq \mathbf{0}$ if and only if $\Lambda\big(H(\mathbf{x})^{-1}\mathbf{s}\big) \succcurlyeq \mathbf{0}$, which is equivalent to $\mathbf{x} \in \mathcal{C}(\mathbf{s})$ by the definition of $\mathcal{C}(\mathbf{s})$ and the characterization (3) of Σ^*. □

From a high-level perspective, the matrix $\mathbf{S}(\mathbf{x}, \mathbf{s})$ is defined in (13) by a "closed-form formula". We will make some more precise statements on the complexity of this formula in Section 2.1 below.

Recall from Lemma 1 (claim 4) that for every $\mathbf{s} \in \Sigma^\circ$ there exists a unique $\mathbf{x} \in (\Sigma^*)^\circ$ satisfying $\mathbf{s} = -g(\mathbf{x})$. This vector is a dual certificate of $\mathbf{s}$, since

$$H(\mathbf{x})^{-1}\mathbf{s} = -H(\mathbf{x})^{-1}g(\mathbf{x}) \stackrel{(10)}{=} \mathbf{x} \in (\Sigma^*)^\circ.$$

Thus, every polynomial in the interior of the WSOS cone Σ has a dual certificate.

Definition 2. When $-g(\mathbf{x}) = \mathbf{s}\,(\in \Sigma^\circ)$, we say that $\mathbf{x}$ is the *gradient certificate of* $\mathbf{s}$.

It is immediate from the definition that if $\mathbf{x}$ is a dual certificate of $\mathbf{s}$, then so is every positive multiple of $\mathbf{x}$. (One may also confirm directly that the block matrix $\mathbf{S}$ constructed in (13) is invariant to a positive scaling of $\mathbf{x}$.) Also note that when $\mathbf{x}$ is the gradient certificate of $\mathbf{s} = -g(\mathbf{x})$, then $\mathbf{S}(\mathbf{x}, \mathbf{s})$ is positive definite. Since $\mathbf{S}$ is continuous on $(\Sigma^*)^\circ \times \Sigma^\circ$, it is immediate that all vectors in some ($\mathbf{s}$-dependent) neighborhood of $\mathbf{x}$ are dual certificates of $\mathbf{s}$. Conversely, the gradient certificate of $\mathbf{s}$ is also a dual certificate of every polynomial in some ($\mathbf{x}$-dependent) neighborhood of $\mathbf{s}$. Our next lemma is a quantitative version of this observation.

Lemma 2. *Suppose $\mathbf{t} \in \Sigma^\circ$ and let $\mathbf{x} \in (\Sigma^*)^\circ$ be any vector that satisfies the inequality*

$$\mathbf{t}^{\mathrm{T}}\left(\mathbf{x}\mathbf{x}^{\mathrm{T}} - (\nu - 1)H(\mathbf{x})^{-1}\right)\mathbf{t} \geq 0. \tag{14}$$

Then $\mathbf{x} \in \mathcal{C}(\mathbf{t})$, equivalently, $\mathbf{t} \in \mathcal{P}(\mathbf{x})$. In particular, if $\mathbf{s} = -g(\mathbf{x})$ for some $\mathbf{x} \in (\Sigma^*)^\circ$, then $\mathbf{x}$ is a dual certificate for every polynomial $\mathbf{t}$ satisfying $\|\mathbf{t} - \mathbf{s}\|_{\mathbf{x}}^* \leq 1$.

Proof. We start with the second claim. From the definitions of the local norm and the dual local norm, we have

$$\|\mathbf{t}-\mathbf{s}\|_{\mathbf{x}}^* = \|H(\mathbf{x})^{-1/2}(\mathbf{t}-\mathbf{s})\| = \|H(\mathbf{x})^{1/2}(\mathbf{x}-H(\mathbf{x})^{-1}\mathbf{t})\| = \|\mathbf{x}-H(\mathbf{x})^{-1}\mathbf{t}\|_{\mathbf{x}}. \quad (15)$$

Thus, $\|\mathbf{t}-\mathbf{s}\|_{\mathbf{x}}^* \leq 1$ is equivalent to $H(\mathbf{x})^{-1}\mathbf{t} \in \overline{B_{\mathbf{x}}(\mathbf{x},1)}$. Since $B_{\mathbf{x}}(\mathbf{x},1) \subseteq (\Sigma^*)^\circ$ from the first claim of Lemma 1, $\overline{B_{\mathbf{x}}(\mathbf{x},1)} \subseteq \Sigma^*$, and $\mathbf{x} \in \mathcal{C}(\mathbf{t})$ by definition.

The first claim of the Lemma is the "conic version" of the second claim. To prove it, suppose that the inequality in (14) holds. Then the univariate quadratic polynomial

$$z \mapsto (1-\nu)z^2 + (2\langle \mathbf{t}, \mathbf{x}\rangle)\, z - \langle \mathbf{t}, H(\mathbf{x})^{-1}\mathbf{t}\rangle$$

has a nonnegative discriminant, therefore it has a root δ. Moreover, since $(1-\nu) < 0$ and $\langle \mathbf{t}, H(\mathbf{x})^{-1}\mathbf{t}\rangle > 0$, it follows that $\delta > 0$. Using the identities in Eq. (10), we have

$$\begin{aligned}
0 &\leq (1-\nu)\delta^2 + (2\langle \mathbf{t}, \mathbf{x}\rangle)\,\delta - \langle \mathbf{t}, H(\mathbf{x})^{-1}\mathbf{t}\rangle \\
&= \delta^2\left(1 - \langle g(\mathbf{x}), H(\mathbf{x})^{-1}g(\mathbf{x})\rangle\right) - \delta\left(2\langle \mathbf{t}, H(\mathbf{x})^{-1}g(\mathbf{x})\rangle\right) - \langle \mathbf{t}, H(\mathbf{x})^{-1}\mathbf{t}\rangle \\
&= \delta^2 - \langle \mathbf{t}+\delta g(\mathbf{x}), H(\mathbf{x})^{-1}(\mathbf{t}+\delta g(\mathbf{x}))\rangle \\
&= \delta^2 - \|H(\mathbf{x})^{-1/2}(\mathbf{t}+\delta g(\mathbf{x}))\|^2.
\end{aligned}$$

We conclude that $\|H(\mathbf{x})^{-1/2}(\mathbf{t}+\delta g(\mathbf{x}))\| < \delta$ for some $\delta > 0$. Then using Lemma 1 again, we have

$$\begin{aligned}
1 &\geq \frac{1}{\delta}\|H(\mathbf{x})^{-1/2}(\mathbf{t}+\delta g(\mathbf{x}))\| \\
&\overset{(10)}{=} \left\|\delta H(\mathbf{x})^{1/2}\left(\delta^{-2}H(\mathbf{x})^{-1}\mathbf{t} - \delta^{-1}\mathbf{x}\right)\right\| \\
&\overset{(9)}{=} \left\|H\left(\delta^{-1}\mathbf{x}\right)^{1/2}\left(H\left(\delta^{-1}\mathbf{x}\right)^{-1}\mathbf{t} - \delta^{-1}\mathbf{x}\right)\right\|,
\end{aligned}$$

so by the identities (15) and the first part of our proof, $\mathbf{t}$ is certified WSOS by $\frac{1}{\delta}\mathbf{x}$. Since all positive multiples of $\mathbf{x}$ certify $\mathbf{t}$, and δ is positive, it follows that $\mathbf{x}$ certifies $\mathbf{t}$. □

Corollary 1. *Let* $\mathbf{x}, \mathbf{y} \in \Sigma^*$ *and* $\mathbf{s}, \mathbf{t} \in \Sigma$*, with* $-g(\mathbf{x}) = \mathbf{s}$ *and* $-g(\mathbf{y}) = \mathbf{t}$*. If* $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} < \frac{1}{2}$*, then* $\mathbf{x}$ *certifies* $\mathbf{t}$*.*

Proof. If $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} < \frac{1}{2}$, then by Lemma 1,

$$\|\mathbf{s}-\mathbf{t}\|_{\mathbf{x}}^* = \|g(\mathbf{x})-g(\mathbf{y})\|_{\mathbf{x}}^* \overset{(11)}{\leq} \frac{\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}}{1-\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}} < 1.$$

Then by Lemma 2, $\mathbf{x}$ certifies $\mathbf{t}$. □

2.1. Algorithmic considerations

Depending on the choice of the Λ operator (that is, in essence, the choice of bases $\mathbf{p}$ and $\mathbf{q}$ in the construction of the semidefinite representation of Σ following Proposition 1), the computation of $\mathbf{S}(\mathbf{x}, \mathbf{s})$ can be made efficient, even polynomial-time in the bit model. Suppose that for a given rational $\mathbf{x} \in (\Sigma^*)^\circ$, the matrices $\Lambda(\mathbf{x})$ and $H(\mathbf{x})$ are rational and can be computed efficiently. Then for any $\mathbf{s} \in \mathbb{R}^U$, the computation of $\mathbf{S}(\mathbf{x}, \mathbf{s})$ amounts

to (1) computing a Cholesky (LDL^{T}) factorization of $\Lambda(\mathbf{x})$ and $H(\mathbf{x})$ (which are positive definite by definition); (2) computing the vector $\mathbf{w} = H(\mathbf{x})^{-1}\mathbf{s}$ using the Cholesky factors of $H(\mathbf{x})$ computed in the previous step; and (3) computing $\Lambda(\mathbf{w})$ and then $\mathbf{S}(\mathbf{x}, \mathbf{s})$ using the Cholesky factors of $\Lambda(\mathbf{x})$. Therefore, computing $\mathbf{S}(\mathbf{x}, \mathbf{s})$ is efficient as long as $\Lambda(\cdot)$ and $H(\cdot)$ can be computed efficiently.

For any reasonable choice and representation of Λ, the computation of $\Lambda(\cdot)$ and $\Lambda^*(\cdot)$ are efficient, as they are linear operators, typically explicitly represented in matrix form with rational entries. Studying the same question in the context of numerical methods for SOS optimization, the authors in [7, Sec. 6] showed that when polynomials are represented as Lagrange interpolants, the Hessian $H(\mathbf{x})$ can be computed with $\mathcal{O}(mLU^2)$ arithmetic operations. One can also argue directly from the identity (8), that (since Λ and Λ^* are efficiently computable) the Hessian can be computed efficiently; the bottleneck once again is the inversion or factorization of $\Lambda(\mathbf{x})$. We note the monomial and Chebyshev polynomial bases as two additional interesting special cases (both in the univariate and multivariate setting): in these cases, $\Lambda(\mathbf{x})$ is a low displacement-rank matrix. For example, when the polynomials are univariate, each block of Λ is a Hankel (or Hankel+Toeplitz) matrix if using the monomial (or Chebyshev) basis. Therefore the inversion of Λ and the computation of H can be handled using discrete Fourier transforms or the superfast (nearly-linear-time) algorithms of Pan and others [20].

3. Computing rigorously certified lower bounds with dual certificates

With our theoretical infrastructure and notation in place, we now turn to the question of computing certified lower bounds and dual certificates for these bounds. In Section 3.1 we show that under the condition that the constant one polynomial is in the interior of our WSOS cone, every polynomial has a dual certifiable lower bound. We also show that after a suitable preprocessing (required only once for every WSOS cone), such a certified bound can be computed by a closed form formula for any polynomial.

In Section 3.2 we discuss efficient algorithms to compute the best bound that a given certificate certifies for a given polynomial, and show that using dual certificates, inexact numerical certificates (that come, for example, from numerical sum-of-squares optimization approaches) can be turned into rigorous rational certificates with minimal additional effort.

In Section 3.3, we present a new algorithm for approximating the best certifiable WSOS bound for a given polynomial with arbitrary accuracy, and show that it is linearly convergent to the optimal bound.

Finally, in Section 3.4 we discuss how rational certificates can be constructed efficiently from the dual certificates obtained with a numerical method such as Algorithm 1 in Section 3.3.

Throughout this section, and the rest of the paper, the boldface vector $\mathbf{1}$ represents the constant one polynomial (or, precisely, its coefficient vector) in the WSOS cone $\Sigma\,(= \Sigma^{\mathbf{g}}_{n,2\mathbf{d}})$, in the space of polynomials $\mathcal{V}(= \mathcal{V}^{\mathbf{g}}_{n,2\mathbf{d}})$.

3.1. Universal dual certificates

Suppose that $\mathbf{1} \in \Sigma^\circ$. Then $\mathbf{1}$ has a gradient certificate $\mathbf{x}_1$, and as we have seen above, $\mathbf{1} \in \mathcal{P}(\mathbf{x}_1)^\circ$, that is, $\mathbf{x}_1$ certifies an entire full-dimensional cone of polynomials

with $\mathbf{1}$ in its interior. Conversely, an entire cone of certificates, with $\mathbf{x}_1$ in its interior, certifies $\mathbf{1}$. Our next theorem shows that each of these certificates also certifies *some* WSOS lower bound for every polynomial:

Lemma 3. *Let $\mathbf{x} \in (\Sigma^*)^\circ$ be any certificate for which $\mathbf{1} \in \mathcal{P}(\mathbf{x})^\circ$ and $r > 0$ arbitrary. Then for every polynomial $\mathbf{t} \in \mathcal{V}$, the inclusion $\mathbf{x} \in \mathcal{C}(\mathbf{t} + c\mathbf{1})$ holds for every sufficiently large scalar c. Specifically, if $\mathbf{x}_1$ is the gradient certificate of $\mathbf{1}$ and $\mathbf{y}_c$ is the gradient certificate of $\mathbf{t} + c\mathbf{1}$, then the inclusion $\mathbf{x}_1 \in \mathcal{C}(\mathbf{t} + c\mathbf{1})$ and the inequality*

$$\|c^{-1}\mathbf{x}_1 - \mathbf{y}_c\|_{c^{-1}\mathbf{x}_1} \leq r, \tag{16}$$

hold for every

$$c \geq \frac{1+r}{r}\|\mathbf{t}\|^*_{\mathbf{x}_1}. \tag{17}$$

Proof. The first statement is immediate from the fact that $\mathcal{P}(\mathbf{x})$ is a cone and the assumption that $\mathbf{1} \in \mathcal{P}(\mathbf{x})^\circ$: the dual vector $\mathbf{x}$ certifies all small perturbations of $\mathbf{1}$, including every polynomial of the form $(c^{-1}\mathbf{t} + \mathbf{1})$, and thus also $\mathbf{t} + c\mathbf{1}$, for every sufficiently large c. We prove the second statement in detail.

Using the definitions of the local dual norm and logarithmic homogeneity (9) from Lemma 1, we have

$$\|(\mathbf{t} + c\mathbf{1}) - c\mathbf{1}\|^*_{c^{-1}\mathbf{x}_1} \overset{\text{(by def.)}}{=} \|H(c^{-1}\mathbf{x}_1)^{-1/2}\mathbf{t}\| \overset{(1)}{=} c^{-1}\|H(\mathbf{x}_1)^{-1/2}\mathbf{t}\| \overset{\text{(by def.)}}{=} c^{-1}\|\mathbf{t}\|^*_{\mathbf{x}_1}$$

Our assumed inequality (17) thus yields

$$\|(\mathbf{t} + c\mathbf{1}) - c\mathbf{1}\|^*_{c^{-1}x_1} = c^{-1}\|\mathbf{t}\|^*_{\mathbf{x}_1} \leq \frac{r}{r+1}. \tag{18}$$

Using logarithmic homogeneity again, we see that $c^{-1}\mathbf{x}_1$ is the gradient certificate for $c\mathbf{1}$. Therefore, invoking Lemma 2, we deduce from the inequality (18) that $c^{-1}\mathbf{x}_1$ is a dual certificate for $\mathbf{t} + c\mathbf{1}$. Moreover, via the inequality (12) in Lemma 1, we conclude that

$$\|c^{-1}\mathbf{x}_1 - \mathbf{y}_c\|_{c^{-1}\mathbf{x}_1} \overset{(12)}{\leq} \frac{\|\mathbf{t}\|^*_{c^{-1}\mathbf{x}_1}}{1 - \|\mathbf{t}\|^*_{c^{-1}\mathbf{x}_1}} \overset{(18)}{\leq} r,$$

as claimed. □

We emphasize that the certificate $\mathbf{x}_1$ (or any $\mathbf{x}$ with $\mathbf{1} \in \mathcal{P}(\mathbf{x})^\circ$) in Lemma 3 only needs to be computed once for any particular WSOS cone $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$. Once $\mathbf{x}_1$ (and the corresponding $H(\mathbf{x}_1)^{-1}$) are computed, a certifiable lower bound and a corresponding certificate can be computed in closed form for any polynomial $\mathbf{t} \in \mathcal{V}$, with minimal effort.

When the weight polynomials $\mathbf{g}$ are sufficiently simple, the gradient certificate of $\mathbf{1}$ may even be easily expressible in closed form, as in the following example.

Example 1. Consider the cone of nonnegative univariate polynomials of degree $2d+1$ over the interval $[-1, 1]$, which is well known to be the same as the WSOS cone $\Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$ with $n = 1$, $m = 2$, degree vector $\mathbf{d} = (d, d)$, and weight polynomials $\mathbf{g}(t) = (1-t, 1+t)$. Furthermore, suppose that all polynomials are represented in the basis of Chebyshev polynomials (of the second kind), that is, the ordered bases $\mathbf{p}$ and $\mathbf{q}$ in Proposition 1

that determine the operator Λ are the Chebyshev basis polynomials of degree up to d. Then both diagonal blocks of Λ are Hankel+Toeplitz matrices (we omit the rather tedious details), and the gradient certificate of $\mathbf{1} = (1, 0, \ldots, 0) \in \mathbb{R}^{2d+2}$ is simply the vector

$$\mathbf{x}_1 = (2d + 2, 0, \ldots, 0).$$

This can be proven by direct calculation verifying the equality $-g(\mathbf{x}_1) = \Lambda^*(\Lambda(\mathbf{x}_1)^{-1}) = \mathbf{1}$. The Hessian at this certificate is the diagonal matrix

$$H(\mathbf{x}_1) = \frac{1}{2d+2} \operatorname{diag}\left(1, \frac{2d+1}{d+1}, \frac{2d}{d+1}, \ldots, \frac{1}{d+1}\right).$$

Analogous results can be derived for polynomials of even degree using $\mathbf{d} = (d, d-1)$, and weight polynomials $\mathbf{g}(t) = (1, 1 - t^2)$.

3.2. Optimal and near-optimal lower bounds from a given dual certificate

Suppose we have found a dual certificate $\mathbf{x}$ that certifies the nonnegativity of the polynomial $\mathbf{t} - c\mathbf{1}$. What is the *best* lower bound certified by the same certificate? By definition, the answer is the solution of the one-dimensional optimization problem

$$c_{\max} \stackrel{\text{def}}{=} \max\left\{\gamma \in \mathbb{R} \,|\, \mathbf{t} - \gamma\mathbf{1} \in \mathcal{P}(\mathbf{x})\right\}.$$

As discussed in Section 2, if the inverse Hessian $H(\mathbf{x})^{-1}$ (or the Cholesky or LDL^{T} factorization of $H(\mathbf{x})$) is already computed, then membership in $\mathcal{P}(\mathbf{x})$ is easy to test by verifying the positive semidefiniteness of $\Lambda(H(\mathbf{x})^{-1}(\mathbf{t} - \gamma\mathbf{1}))$. Therefore, an arbitrarily close lower approximation of $c_{\max}$ can be found efficiently, in time proportional to the logarithm of the approximation error, by binary search on the optimal γ. (An initial lower bound on $c_{\max}$ is the currently certified lower bound c assumed to be part of the input; an upper bound on $c_{\max}$ can be computed, e.g., by evaluating the polynomial $\mathbf{t}$ at any point in its domain.)

The repeated matrix factorization makes the algorithm outlined above too expensive to use as a subroutine. A weaker bound can be computed *in closed form* using Lemma 2: if

$$c'_{\max} \stackrel{\text{def}}{=} \max\left\{\gamma \in \mathbb{R} \,\middle|\, (\mathbf{t} - \gamma\mathbf{1})^{\mathrm{T}} \left(\mathbf{x}\mathbf{x}^{\mathrm{T}} - (\nu - 1)H(\mathbf{x})^{-1}\right)(\mathbf{t} - \gamma\mathbf{1}) \geq 0\right\},$$

then $\mathbf{t} - c'_{\max} \in \mathcal{P}(\mathbf{x})$. For a given certificate $\mathbf{x}$, if the inverse Hessian $H(\mathbf{x})^{-1}$ (or the Cholesky or LDL^{T} factorization of $H(\mathbf{x})$) is already computed, then solving this optimization problem amounts to finding the roots of a univariate quadratic function.

In Algorithm 1 below, we use a variant of the $c'_{\max}$ bound to compute the optimal SOS bound for a given polynomial.

3.3. Computing optimal WSOS bounds

We now present an iterative method to compute the best WSOS lower bound for a given polynomial $\mathbf{t}$. The pseudocode of the algorithm is shown in Algorithm 1. After a high-level description of the method, we show that it converges linearly to the optimal WSOS bound below (Theorem 3).

Previously, in Lemma 3, we showed that for a sufficiently large c, $\mathbf{t} + c\mathbf{1}$ can be certified by $c^{-1}\mathbf{x}_1$; this result justifies the initialization of the algorithm in Line 1. In

Algorithm 1: Compute the best WSOS lower bound and a dual certificate

input : A polynomial $\mathbf{t}$; a tolerance $\varepsilon > 0$.

parameters: An oracle for computing the barrier Hessian H for Σ; the gradient certificate $\mathbf{x}_1$ for the constant one polynomial; a radius $r \in (0, 1/4)$.

outputs : A lower bound c on the optimal WSOS lower bound c^* satisfying $c^* - c \leq \varepsilon$; a dual vector $\mathbf{x} \in (\Sigma^*)^\circ$ certifying the nonnegativity of $\mathbf{t} - c\mathbf{1}$.

1 Compute $c_0 = -\frac{1+r}{r}\left(\mathbf{t}^{\mathrm{T}} H(\mathbf{x}_1)^{-1}\mathbf{t}\right)^{1/2}$. Set $c = c_0$ and $\mathbf{x} = -\frac{1}{c_0}\mathbf{x}_1$.

2 **repeat**

3 Set $\mathbf{x} := 2\mathbf{x} - H(\mathbf{x})^{-1}(\mathbf{t} - c\mathbf{1})$.

4 Find the largest real number c_+ such that

$$\|\mathbf{x} - H(\mathbf{x})^{-1}(\mathbf{t} - c_+\mathbf{1})\|_{\mathbf{x}} \leq \frac{r}{r+1}.$$

5 Set $\Delta c := c_+ - c$. Set $c := c_+$.

6 **until** $\Delta c \leq \rho_r C \varepsilon$

7 **return** c *and* $\mathbf{x}$.

order to increase the lower bound, the algorithm iterates two steps: certificate updates (Line 3) and bound updates (Line 4). The bound updates are similar to the $c'_{\max}$ bound in Section 3.2; we will precisely justify this step in Lemma 5. The certificate updates are motivated as follows: since each bound update attempts to push c towards the best bound certifiable by $\mathbf{x}$, the certificate $\mathbf{x}$ sits near the boundary of $\mathcal{C}(\mathbf{t} - c\mathbf{1})$ after each bound update. To allow for a sufficient additional increase of the bound in the subsequent iteration, the certificate $\mathbf{x}$ is updated to be closer to the gradient certificate $\mathbf{y}$ of the current $\mathbf{t} - c\mathbf{1}$. This certificate $\mathbf{y}$ would be prohibitively expensive to compute in each iteration; instead, the update step in Line 3 can be interpreted as a single Newton step from $\mathbf{x}$ towards the solution of the nonlinear system $-g(\mathbf{y}) = \mathbf{t} - c\mathbf{1}$.

The computationally most expensive part in each iteration is having to compute (after each certificate update) a Cholesky factorization of the Hessian $H(\mathbf{x})$ (or the inverse Hessian $H(\mathbf{x})^{-1}$). With that available, the bound update and the next certificate update are very efficient: by an argument analogous to the discussion on $c'_{\max}$ in the previous section, the bound update amounts to solving a univariate quadratic polynomial, and the certificate update is essentially a matrix-vector multiplication. As discussed in Section 2.1, the computation and factorization of the Hessian is efficient for popular choices of polynomial bases. The number of iterations is also quite low: as we shall see in Theorem 3, the algorithm converges linearly to the optimal WSOS bound.

We now turn to the analysis of the algorithm, deferring the discussion on the stopping criterion until later. To simplify the statements of the results, we will use the following notation throughout the rest of the section. We define $\mathbf{x}_+ \stackrel{\text{def}}{=} 2\mathbf{x} - H(\mathbf{x})^{-1}(\mathbf{t} - c\mathbf{1})$ to be the updated certificate in Line 3 to help distinguish the certificates before and after the update. Finally, we let $\mathbf{y}$ be the vector satisfying $-g(\mathbf{y}) = \mathbf{t} - c\mathbf{1}$ and $\mathbf{y}_+$ be the vector

satisfying $-g(\mathbf{y}_+) = \mathbf{t} - c_+\mathbf{1}$.

In the next series of Lemmas we show that the bound update from c to c_+ is well-defined, and is always an increase, by bounding the distance between $\mathbf{x}$ and $\mathbf{y}$ in each step of the iteration. The first result, Lemma 4, shows that $\mathbf{x}_+$ is closer than $\mathbf{x}$ to the gradient certificate of $\mathbf{t} - c\mathbf{1}$ in their respective local norms.

Lemma 4. *Let $\mathbf{x}_+$, $\mathbf{y}$ be defined as above, and assume that $\|\mathbf{x}-\mathbf{y}\| \le r$ for some $r < \frac{1}{3}$. Then $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$.*

Proof. Recall that the update in Line 3 of Algorithm 1 is a single (full) Newton step towards the solution of the linear system $-g(\mathbf{y}) = \mathbf{t} - c\mathbf{1}$. Equivalently, the update $\mathbf{x}_+ - \mathbf{x}$ is a Newton step toward the minimizer of the convex self-concordant function

$$f_c(\mathbf{x}) \stackrel{\text{def}}{=} (\mathbf{t} - c\mathbf{1})^T\mathbf{x} + f(\mathbf{x}).$$

Applying [18, Thm. 2.2.3]) to f_c, we have

$$\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}} \le \frac{\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}^2}{1 - \|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}} = \frac{r^2}{1-r}.$$

Coupling this result with the definition of self-concordance (Eq. (6)), we have

$$\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{y}} \le \frac{\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}}}{1 - \|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}} \le \frac{\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}^2}{(1 - \|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}})^2} \le \frac{r^2}{(1-r)^2} < 1.$$

We conclude that $\mathbf{x}_+ \in B_{\mathbf{y}}(\mathbf{y}, 1)$, and we can thus invoke the inequality (6) for another change of norms to conclude that

$$\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{y}}}{1 - \|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{y}}} \le \frac{\frac{r^2}{(1-r)^2}}{1 - \frac{r^2}{(1-r)^2}} = \frac{r^2}{(1-r)^2 - r^2} = \frac{r^2}{1-2r}. \qquad \square$$

We remark that, while $\mathbf{x}$ certifies $\mathbf{t} - c\mathbf{1}$ whenever $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} < \frac{1}{2}$, and each step of our proof is valid for all $0 < r < \frac{1}{2}$, we can only have $\frac{r^2}{1-2r} \le r$ whenever $0 < r < \frac{1}{3}$. Therefore, using Lemma 4, we can guarantee that $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}$ when $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} < \frac{1}{3}$.

Lemma 5. *Suppose that $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$ for some $0 < r < \frac{1}{4}$. Then*

$$\|\mathbf{x}_+ - H(\mathbf{x})^{-1}(\mathbf{t} - c\mathbf{1})\|_{\mathbf{x}_+} < \frac{r}{r+1}.$$

Proof. Suppose $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$. Recall from Eq. (10) that $H(\mathbf{x}_+)\mathbf{x}_+ = -g(\mathbf{x}_+)$. Using this identity and the definition of the local norm, we deduce that

$$\begin{aligned} \| - g(\mathbf{x}_+) + g(\mathbf{y})\|_{\mathbf{x}_+}^* &= \|H(\mathbf{x}_+)^{-1/2}\left(H(\mathbf{x}_+)\mathbf{x}_+ - (\mathbf{t} - c\mathbf{1})\right)\| \\ &= \|H(\mathbf{x}_+)^{1/2}\mathbf{x}_+ - H(\mathbf{x}_+)^{-1/2}(\mathbf{t} - c\mathbf{1})\| \\ &= \|\mathbf{x}_+ - H(\mathbf{x}_+)^{-1}(\mathbf{t} - c\mathbf{1})\|_{\mathbf{x}_+}. \end{aligned} \tag{19}$$

Then using this in tandem with inequality (11) from Lemma 1, we have

$$\|\mathbf{x}_+ - H(\mathbf{x})^{-1}(\mathbf{t}-c\mathbf{1})\|_{\mathbf{x}_+} \overset{(19)}{=} \|-g(\mathbf{x}_+)+g(\mathbf{y})\|_{\mathbf{x}_+} \overset{(11)}{\leq} \frac{\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}}{1-\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}} \leq \frac{\frac{r^2}{1-2r}}{1-\frac{r^2}{1-2r}} < \frac{r}{r+1}$$

for every $r < \frac{1}{4}$. □

As a consequence, since $\|\mathbf{x}_+ - H(\mathbf{x})^{-1}(\mathbf{t}-c\mathbf{1})\|_{\mathbf{x}_+} < \frac{r}{r+1}$, we can guarantee that there exists a constant $c_+ > c$ such that $\|\mathbf{x}_+ - H(\mathbf{x})^{-1}(\mathbf{t}-c_+\mathbf{1})\|_{\mathbf{x}_+} = \frac{r}{r+1}$.

Now, we show that the certificate $\mathbf{x}$ and the gradient certificate $\mathbf{y}$ satisfy $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \leq r$ at the beginning of each iteration. This is clearly true for the first iteration, as shown by the inequality (16) in Lemma 3.

Lemma 6. *Suppose that* $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \leq r < 1$. *Then* $\|\mathbf{x}_+ - \mathbf{y}_+\|_{\mathbf{x}_+} \leq r$.

Proof. Analogously to Eq. (19), $\|-g(\mathbf{x}_+)+g(\mathbf{y}_+)\|^*_{\mathbf{x}_+} = \|\mathbf{x}_+ - H(\mathbf{x}_+)^{-1}(\mathbf{t}-c_+\mathbf{1})\|_{\mathbf{x}_+}$. Therefore

$$\|-g(\mathbf{x}_+)+g(\mathbf{y}_+)\|^*_{\mathbf{x}_+} = \|\mathbf{x}_+ - H(\mathbf{x}_+)^{-1}(\mathbf{t}-c_+\mathbf{1})\|_{\mathbf{x}_+} = \frac{r}{r+1} < 1 \tag{20}$$

by the definition of the bound update step in Line 4 and our discussion above. Now inequality (12) from Lemma 1 yields

$$\|\mathbf{x}_+ - \mathbf{y}_+\|_{\mathbf{x}_+} \leq \frac{\|-g(\mathbf{x}_+)+g(\mathbf{y}_+)\|^*_{\mathbf{x}_+}}{1-\|-g(\mathbf{x}_+)+g(\mathbf{y}_+)\|^*_{\mathbf{x}_+}} \leq \frac{r/(r+1)}{1-(r/(r+1))} = r. \qquad \square$$

Coupling Lemma 6 with Corollary 1, this reasserts that $\mathbf{x}$ is indeed a certificate of $\mathbf{t}-c\mathbf{1}$ at the end of each iteration of the algorithm.

The next lemma uses Lemma 4 in showing that the improvement in the lower bound can be bounded from below by a constant times the local norm of $\mathbf{1}$.

Lemma 7. *Define* $\rho_r \overset{\text{def}}{=} \frac{r(1-3r-2r^2)}{1-r-2r^2}$ *Then at the end of each iteration of Algorithm 1,* $c_+ - c \geq \frac{\rho_r}{\|\mathbf{1}\|^*_{\mathbf{y}}}$, *where* $\mathbf{y}$ *is the gradient certificate of* $\mathbf{t}-c\mathbf{1}$.

Proof. From the identities (20) and the definition of c_+ in Line 4 of the algorithm, we have

$$\frac{r}{r+1} = \|\mathbf{x}_+ - H(\mathbf{x}_+)^{-1}(\mathbf{t}-c_+\mathbf{1})\|_{\mathbf{x}_+} = \|-g(\mathbf{x}_+)+g(\mathbf{y}_+)\|^*_{\mathbf{x}_+}.$$

Upper bounding the right-hand side by the triangle inequality gives

$$\frac{r}{r+1} - \|-g(\mathbf{x}_+)+g(\mathbf{y})\|^*_{\mathbf{x}_+} \leq \|-g(\mathbf{y}_+)+g(\mathbf{y})\|^*_{\mathbf{x}_+} = \|(c_+-c)\mathbf{1}\|^*_{\mathbf{x}_+}. \tag{21}$$

Thus, to lower bound $(c_+ - c)$, it suffices to upper bound $\|-g(\mathbf{x}_+)+g(\mathbf{y})\|^*_{\mathbf{x}_+}$.

From Lemma 4, we know that $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \leq \frac{r^2}{1-2r}$. Using the inequality (11) in Lemma 1, we have

$$\|-g(\mathbf{x}_+)+g(\mathbf{y})\|^*_{\mathbf{x}_+} \leq \frac{\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}}{1-\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}} \leq \frac{\frac{r^2}{1-2r}}{1-\frac{r^2}{1-2r}} = \frac{r^2}{1-2r-r^2}. \tag{22}$$

Combining the inequalities in (21) and (22), we have

$$(c_+ - c)\|\mathbf{1}\|^*_{\mathbf{x}_+} \geq \frac{r}{r+1} - \frac{r^2}{1-2r-r^2}.$$

Finally, changing norms again with inequality (6),

$$(c_+ - c)\|\mathbf{1}\|^*_{\mathbf{y}_+} \geq (c_+ - c)\|\mathbf{1}\|^*_{\mathbf{x}_+}(1 - \|\mathbf{y} - \mathbf{x}_+\|^*_{\mathbf{x}_+}) \geq \left(\frac{r}{r+1} - \frac{r^2}{1-2r-r^2}\right)\left(1 - \frac{r^2}{1-2r}\right) = \rho_r.$$

□

We remark that if r is chosen so that $0 < r < \frac{1}{4}$, then $\rho_r > 0$, and, for example, $\rho_r > 2/21$ for $r = 1/6$. Therefore in each iteration of the algorithm, the improvement of can be bounded from below by a quantity proportional to $(\|\mathbf{1}\|^*_{\mathbf{y}})^{-1}$ with the current gradient certificate $\mathbf{y}$.

Now, we turn our attention to the convergence of Algorithm 1. When $\mathbf{1} \in \Sigma^\circ$, the optimal WSOS lower bound c^* for a polynomial $\mathbf{t}$ is the unique scalar γ for which $\mathbf{t} - \gamma\mathbf{1}$ is on the boundary of Σ. In Theorem 2, we show that the norm $\|\mathbf{1}\|^*_{\mathbf{y}}$ can be related to the distance $(c^* - c)$ between the current bound and the optimal WSOS lower bound. We will then combine this result with Lemma 7 to show that the algorithm converges linearly to the optimal WSOS lower bound of $\mathbf{t}$. The analysis also motivates the stopping criterion for the algorithm.

In what follows, we let $\lambda_{\max}(\mathbf{M})$ denote the largest eigenvalue of the matrix $\mathbf{M}$ and $\lambda_{\min}(\mathbf{M})$ denote the smallest eigenvalue. We also remark that $\|\cdot\|_1$ and $\|\cdot\|$ refer to the standard 1-norm and the 2-norm of vectors, respectively (not to be confused with the local norms used above).

Theorem 2. *Suppose that $\mathbf{t} - c^*\mathbf{1}$ is on the boundary of Σ. Let $\mathbf{y}$ denote the gradient certificate of some $\mathbf{t} - c\mathbf{1}$ with $c < c^*$. Then there exists a constant C (depending only on the operator Λ) such that $c^* - c \leq (C\|\mathbf{1}\|^*_{\mathbf{y}})^{-1}$.*

Proof. Recall that $-g(\mathbf{y}) = \mathbf{t} - c\mathbf{1}$. Define the constant

$$k_1 \stackrel{\text{def}}{=} \min\{\mathbf{1}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \|\mathbf{v}\| = 1\}.$$

Observe that the minimum exists (as Σ^* is a closed and non-trivial cone) and $k_1 > 0$, because $\mathbf{1} \in \Sigma^\circ$. Using the shorthand $\alpha \stackrel{\text{def}}{=} c^* - c > 0$, we now have

$$\begin{aligned}
\nu &\stackrel{(10)}{=} \left\langle -g\left(\frac{\mathbf{y}}{\|\mathbf{y}\|}\right), \frac{\mathbf{y}}{\|\mathbf{y}\|}\right\rangle \\
&\stackrel{(9)}{=} \|\mathbf{y}\|\left\langle \mathbf{t} - c\mathbf{1}, \frac{\mathbf{y}}{\|\mathbf{y}\|}\right\rangle \\
&= \|\mathbf{y}\|\left(\left\langle \mathbf{t} - c^*\mathbf{1}, \frac{\mathbf{y}}{\|\mathbf{y}\|}, \right\rangle + (c^* - c)\left\langle \mathbf{1}, \frac{\mathbf{y}}{\|\mathbf{y}\|}\right\rangle\right) \\
&\geq 0 + \|\mathbf{y}\|\alpha k_1 = \|\mathbf{y}\|\alpha k_1,
\end{aligned}$$

from which we conclude that

$$\|\mathbf{y}\| \leq \frac{\nu}{\alpha k_1}. \tag{23}$$

Recall from Eq. (8) that $H(\mathbf{y})\mathbf{w} = \Lambda^*(\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1})$. Therefore, $\mathbf{w}^T H(\mathbf{y})\mathbf{w} = \langle \mathbf{w}, \Lambda^*(\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1})\rangle = \operatorname{tr}(\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1})$. Moreover, observe that, when for any $\mathbf{A} \succcurlyeq \mathbf{0}$ and real symmetric matrix $\mathbf{B}$ of the same size, we have

$$\operatorname{tr}(\mathbf{A})\lambda_{\min}(\mathbf{B}) \leq \operatorname{tr}(\mathbf{AB}) \leq \operatorname{tr}(\mathbf{A})\lambda_{\max}(\mathbf{B}).$$

Using this fact, we have that for every $\mathbf{w} \in \mathbb{R}^U$,

$$\begin{aligned}
\mathbf{w}^T H(\mathbf{y})\mathbf{w} &= \operatorname{tr}\left(\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1}\right) \\
&\geq \lambda_{\min}(\Lambda(\mathbf{y})^{-1})\ \operatorname{tr}\left(\Lambda(\mathbf{w})\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{w})\right) \\
&= \lambda_{\min}(\Lambda(\mathbf{y})^{-1})\ \operatorname{tr}\left(\Lambda(\mathbf{w})^2\Lambda(\mathbf{y})^{-1}\right) \\
&\geq \lambda_{\min}(\Lambda(\mathbf{y})^{-1})^2\ \operatorname{tr}(\Lambda(\mathbf{w})^2) \\
&= \lambda_{\max}(\Lambda(\mathbf{y}))^{-2}\ \operatorname{tr}(\Lambda(\mathbf{w})^2).
\end{aligned}$$

We conclude that

$$\lambda_{\min}(H(\mathbf{y})^{1/2}) \geq \frac{k_2}{\lambda_{\max}(\Lambda(\mathbf{y}))}, \tag{24}$$

wherein we define

$$k_2 \stackrel{\text{def}}{=} \min\{\sqrt{\operatorname{tr}(\Lambda(\mathbf{w})^2)} \mid \|\mathbf{w}\| = 1\}.$$

We remark that $k_2 > 0$ (since $\Lambda(\mathbf{w}) \neq \mathbf{0}$ whenever $\mathbf{w} \neq \mathbf{0}$).

Next, recall that $\|\mathbf{1}\|_{\mathbf{y}}^* = \|H(\mathbf{y})^{-1/2}\mathbf{1}\|$ and note $\|H(\mathbf{y})^{-1/2}\| = \frac{1}{\lambda_{\min}\left(H(\mathbf{y})^{1/2}\right)}$. Define

$$k_3 \stackrel{\text{def}}{=} \max\left\{\lambda_{\max}(\Lambda(\mathbf{y})) \mid \mathbf{y} \in \Sigma^*, \|\mathbf{y}\| = 1\right\}.$$

These identities and our previous inequalities give

$$\|\mathbf{1}\|_{\mathbf{y}}^* = \|H(\mathbf{y})^{-1/2}\mathbf{1}\| \leq \frac{\|\mathbf{1}\|}{\lambda_{\min}\left(H(\mathbf{y})^{1/2}\right)} \overset{(24)}{\leq} \frac{\lambda_{\max}(\Lambda(\mathbf{y}))\|\mathbf{1}\|}{k_2} \leq \frac{k_3\|\mathbf{y}\|\|\mathbf{1}\|}{k_2} \overset{(23)}{\leq} \frac{k_3\nu\|\mathbf{1}\|}{k_1k_2\alpha}.$$

Defining $C \stackrel{\text{def}}{=} \frac{k_1k_2}{k_3\nu\|\mathbf{1}\|}$, we conclude that

$$\alpha = c^* - c \leq (C\|\mathbf{1}\|_{\mathbf{y}}^*)^{-1}. \qquad \square$$

We remark that the parameter $\nu = \sum_{i=1}^m L_i$ is a parameter of the WSOS cone Σ entirely independent of the representation of the polynomials. The parameter k_1 depends on the basis in which the WSOS polynomials are represented (but otherwise does not depend on Λ), while k_2 and k_3 are properties of the Λ operator representing Σ.

Coupling Lemma 7 with Theorem 2, we have also proven our main result about the convergence of our algorithm:

Theorem 3. *Algorithm 1 is globally linearly convergent to $c^* = \max\{c \mid \mathbf{t} - c\mathbf{1} \in \Sigma\}$, the optimal WSOS lower bound for the polynomial $\mathbf{t}$. More precisely, in each iteration of Algorithm 1, the improvement $\Delta c = c_+ - c$ of the lower bound satisfies*

$$\frac{\Delta c}{c^* - c} \geq \rho_r C, \tag{25}$$

with the absolute constant $\rho_r > 0$ defined in Lemma 7 and the Λ-dependent constant $C > 0$ defined in Theorem 2.

Theorem 3 motivates the stopping criterion (Line 6) of Algorithm 1. The current bound c is guaranteed to satisfy $c \leq c^* \leq c + \varepsilon$ as soon as $\Delta c \leq \rho_r C \varepsilon$.

Alternatively, we can also rearrange the same inequality to provide an explicit upper bound on the number of iterations of the algorithm. After k iterations of Algorithm 1 we have

$$c^* - c_k \leq (1 - \rho_r C)^k (c^* - c_0),$$

therefore, for a fixed cone (and parameter C), the algorithm terminates after $\mathcal{O}\left(\log \frac{c^* - c_0}{\varepsilon}\right)$ iterations. Additionally, it is typically easy to bound from below the global minimum of the input polynomial $\mathbf{t}$ (e.g., by evaluating it at any point in its domain), and thus bound c^* from above, and when an explicit bound on the magnitude of the elements in $\{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \geq 0,\ i = 1, \ldots, m\}$ is known, it is also straightforward to upper bound c^* by $\kappa_{\mathbf{g}} \|\mathbf{t}\|$ with some constant $\kappa_{\mathbf{g}}$ dependent only the weight functions $\mathbf{g}$. Similarly, from the first step of Algorithm 1, $c_0 \geq -\frac{1+r}{r} \lambda_{\max}(H(\mathbf{x}_1)^{-1}) \|\mathbf{t}\|$, bounding the initial bound c_0 from below by a Λ-dependent constant multiple of $\|\mathbf{t}\|$. Thus, for a fixed cone, the algorithm terminates after $\mathcal{O}(\log \frac{\|\mathbf{t}\|}{\varepsilon})$ iterations.

3.4. Rational certificates

As presented and analyzed in the previous sections, Algorithm 1 can be implemented as an entirely numerical method, computing a "numerical" lower bound c and a dual certificate $\mathbf{x}$ for the nonnegativity of $\mathbf{t} - c\mathbf{1}$. Note, however, that as a consequence of Lemma 2, $\mathbf{x}$ serves as an exact rational WSOS certificate for $\mathbf{t} - c\mathbf{1}$ in every iteration, without any need for additional rounding or other post-processing. To be more precisely, in each iteration of this algorithm, the numerical dual certificate $\mathbf{x}$ can be directly converted to an *exact rational certificate* $\mathbf{S} \succcurlyeq \mathbf{0}$ via the formula of Eq. (13), as long as the coefficient vector $\mathbf{t}$ is rational, without any additional rounding or projection of $\mathbf{x}$ or c.

This property sets dual certificates apart from conventional certificates: a numerical solution to the semidefinite programming (feasibility) problem

$$\text{find an } \mathbf{S} \succcurlyeq \mathbf{0} \text{ satisfying } \Lambda^*(\mathbf{S}) = \mathbf{t} - c\mathbf{1}$$

will generally satisfy the equality constraints $\Lambda^*(\mathbf{S}) = \mathbf{t} - c\mathbf{1}$ only within some numerical tolerance, thus $\mathbf{S}$ will not be a rigorous certificate, even if we can guarantee (by the appropriate choice of optimization algorithm) that at least the cone constraint $\mathbf{S} \succcurlyeq \mathbf{0}$ is always satisfied. In contrast, any dual certificate $\mathbf{x}$ from the full-dimensional cone $\mathcal{C}(\mathbf{t} - c\mathbf{1})$ is a "rigorous" certificate that can be turned into a rational WSOS decomposition.

In fact, if desired, one may "round" the certificate $\mathbf{x}$ to a "nearby" rational certificate with smaller components, in order to obtain a simpler WSOS decomposition, quite freely (e.g., by applying Diophantine approximation component-wise, or using the LLL algorithm for simultaneous approximation of $\mathbf{x}$ with a "smaller" rational vector), since the algorithm returns a certificate $\mathbf{x}$ satisfying $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \leq r (< 1/4)$, but every certificate with $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \leq 1$ is equally valid by Corollary 1.

We also remark that although our primary goal is to obtain certified rational *lower* bounds on the polynomial, dual certificates also provide *upper bounds* on the optimal WSOS bound via Theorem 3, whenever the Λ-dependent constant C in is known (or can be bounded from below) for a particular cone Σ. In particular, although in the analysis heavily relies on the quantity $\|\mathbf{1}\|_{\mathbf{y}}^*$, which is not efficiently computable (we do not have

access to the gradient certificate $\mathbf{y}$), the inequality (25) provides a computable upper bound on c^*.

4. Discussion

Primal versus dual certificates. Conventional nonnegativity certificates are representations of the certified polynomials that make their nonnegativity apparent. This is a fundamental issue for numerical methods for computing nonnegativity certificates, as the certificate they compute is typically a rigorous WSOS certificate for a slightly different polynomial from the one we seek to certify.

Dual certificates address this issue: through the formula (13), not only can we interpret any rational dual vector from $\mathcal{C}(\mathbf{s})$ as a certificate, but we can also compute, via a closed-form formula, a rational certificate for the polynomial $\mathbf{s}$ with rational certificates. Since every polynomial (in the interior of the SOS cone) has a full-dimensional cone of dual certificates, even an inexact numerical method computing low-accuracy solutions to an SOS optimization problem can return dual certificates that can be turned into a rational certificate this way. For example, Algorithm 1 can be implemented as a purely numerical method, followed by an application of the formula (13) to compute a rational certificate for the computed bound. Although the certificate $\mathbf{x}$ only loosely tracks the gradient certificate of $\mathbf{t} - c\mathbf{1}$, we can guarantee that $\mathbf{x}$ certifies the current bound. This also means that, unlike most numerical or hybrid methods that require high-accuracy solutions from the numerical component of the algorithm, Algorithm 1 provides a certified bound even if terminated early; only the quality of the bound suffers.

Recent work in numerical methods for non-symmetric cones has resulted in a few additional algorithms that can directly optimize over the cone of WSOS certificates circumventing semidefinite programming, including [21] and [22]; in principle, these can also be coupled with the methods presented in Section 2.

Efficiency. In general, it is difficult to make general statements about the asymptotic running time of Algorithm 1 as a function of every interesting parameter (the degree and the number of unknowns of the input polynomial, etc.) as these also depend on the specific weight polynomials and the chosen representation (Λ operator). As noted, the computational cost per iteration is a low-degree polynomial for Λ operators corresponding to popular bases in numerical methods (e.g., Chebyshev and interpolant bases), and the method is linearly convergent, that is, for a given polynomial it requires a number of iterations proportional to $\log(1/\varepsilon)$ to compute a certified rational bound within ε of the optimal bound c^*. Additionally, from the constructive proof of Theorem 2 it might also be possible to derive explicit bounds on the exact linear rate and the initial gap $c^* - c_0$ in interesting special cases, such as the case of univariate polynomials or multivariate polynomials over simple semialgebraic sets such as the unit sphere or the unit cube.

Assumptions. Throughout, we have made the fundamental assumption that the constant one polynomial is in the interior of the WSOS cone $\Sigma = \Sigma^{\mathbf{g}}_{n,2\mathbf{d}}$. (Naturally, in any remotely interesting situation, positive constant polynomials must belong to Σ, but not necessarily to the interior.) This is a relatively mild assumption both from a theoretical and practical perspective. In many cases, it can be verified directly and ensured to hold a priori. Computationally, it can be verified via convex optimization, and if it does not

hold, Σ can be extended (with the inclusion of more weights that are nonnegative on the nonnegativity set of the existing weights) to satisfy this condition.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1719828 and Grant No. DMS-1847865.

References

[1] A. Tarski, A decision method for elementary algebra and geometry, Tech. Rep. R-109, RAND Corporation, http://www.rand.org/pubs/reports/2008/R109.pdf (May 1951).

[2] J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. Parts I–III., Journal of Symbolic Computation 13 (3) (1992) 255–352. URL https://doi.org/10.1016/S0747-7171(10)80003-3

[3] M. Putinar, Positive polynomials on compact semi-algebraic sets, Indiana University Mathematics Journal 42 (3) (1993) 969–984. doi:10.1512/iumj.1993.42.42045.

[4] S. Prajna, A. Papachristodoulou, P. Seiler, P. A. Parrilo, SOSTOOLS: Sum of squares optimization toolbox for MATLAB (2004). URL http://www.cds.caltech.edu/sostools

[5] D. Henrion, J.-B. Lasserre, GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi, ACM Transactions on Mathematical Software 29 (2) (2003) 165–194. doi:10.1145/779359.779363.

[6] J. B. Lasserre, Global optimization with polynomials and the problem of moments, SIAM Journal on Optimization 11 (3) (2001) 796–817. doi:10.1137/S1052623400366802.

[7] D. Papp, S. Yıldız, Sum-of-squares optimization without semidefinite programming, SIAM Journal on Optimization 29 (1) (2019) 822–851. doi:10.1137/17M1160124.

[8] V. Magron, M. Safey El Din, M. Schweighofer, Algorithms for weighted sum of squares decomposition of non-negative univariate polynomials, Journal of Symbolic Computation 93 (2019) 200–220. doi:10.1016/j.jsc.2018.06.005.

[9] J. Nie, K. Ranestad, B. Sturmfels, The algebraic degree of semidefinite programming, Mathematical Programming 122 (2) (2010) 379–405. doi:10.1007/s10107-008-0253-6.

[10] H. Peyrl, P. A. Parrilo, Computing sum of squares decompositions with rational coefficients, Theoretical Computer Science 409 (2) (2008) 269–281. doi:10.1016/j.tcs.2008.09.025.

[11] V. Magron, M. Safey El Din, On Exact Polya and Putinar's Representations, in: ISSAC'18: Proceedings of the 2018 ACM International Symposium on Symbolic and Algebraic Computation, ACM, New York, NY, USA, 2018. URL http://arxiv.org/abs/1802.10339

[12] M. Dostert, D. de Laat, P. Moustrou, Exact semidefinite programming bounds for packing problems (2020). arXiv:2001.00256.

[13] E. Kaltofen, B. Li, Z. Yang, L. Zhi, Exact certification of global optimality of approximate factorizations via rationalizing sums-of-squares with floating point scalars, in: Proceedings of the Twenty-First International Symposium on Symbolic and Algebraic Computation, ISSAC '08, ACM, New York, NY, 2008, pp. 155–164. doi:10.1145/1390768.1390792.

[14] E. L. Kaltofen, B. Li, Z. Yang, L. Zhi, Exact certification in global polynomial optimization via sums-of-squares of rational functions with rational coefficients, Journal of Symbolic Computation 47 (1) (2012) 1–15. doi:10.1016/j.jsc.2011.08.002.

[15] D. A. Brake, J. D. Hauenstein, A. C. Liddell, Validating the completeness of the real solution set of a system of polynomial equations, in: Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation, ISSAC '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 143–150. doi:10.1145/2930889.2930910.

[16] Y. Nesterov, Squared functional systems and optimization problems, in: H. Frenk, K. Roos, T. Terlaky, S. Zhang (Eds.), High performance optimization, Vol. 33 of Applied Optimization, Kluwer Academic Publishers, Dordrecht, 2000, pp. 405–440. doi:10.1007/978-1-4757-3216-0_17.

[17] Y. Nesterov, A. Nemirovskii, Interior-point polynomial algorithms in convex programming, Vol. 13 of SIAM Studies in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. doi:10.1137/1.9781611970791.

[18] J. Renegar, A mathematical view of interior-point methods in convex optimization, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Phiadelphia, PA, 2001. doi:10.1137/1.9780898718812.

[19] D. Papp, S. Yıldız, On "A homogeneous interior-point algorithm for non-symmetric convex conic optimization", arXiv preprint arXiv:1712.00492 (2017).
URL https://arxiv.org/abs/1712.00492

[20] V. Y. Pan, Structured matrices and polynomials, Birkhäuser, Boston, MA, 2001. doi:10.1007/978-1-4612-0129-8.

[21] M. Karimi, L. Tunçel, Domain-driven solver (DDS): a MATLAB-based software package for convex optimization problems in domain-driven form, arXiv preprint arXiv:1908.03075 (2019).

[22] D. Papp, S. Yıldız, alfonso: Matlab package for nonsymmetric conic optimization, INFORMS Journal on Computing (accepted) (2021).
URL https://arxiv.org/abs/2101.04274