DUAL CERTIFICATES AND EFFICIENT RATIONAL SUM-OF-SQUARES DECOMPOSITIONS FOR POLYNOMIAL OPTIMIZATION OVER COMPACT SETS*

MARIA M. DAVIS† AND DÁVID PAPP‡

Abstract. We study the problem of computing 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 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 a full-dimensional convex cone of WSOS polynomials. For a theoretical application, we give a short new proof of Powers' theorems on the existence of rational WSOS certificates of positive polynomials. For a computational application, we show that exact 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. We also present an 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.

Key words. polynomial optimization, nonnegativity certificates, sums-of-squares, non-symmetric conic optimization

AMS subject classifications. 90C23, 14Q30, 90C51, 49M29, 90C25

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 are fundamental problems 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 [44, 38] 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* [36] guarantee that every polynomial that is strictly positive over a compact semialgebraic set has such a representation.

Lower bounds on the global minima of polynomials and weighted sum-of-squares (WSOS) decompositions are usually computed numerically, using semidefinite programming (e.g., [35, 8, 14]) or non-symmetric cone optimization [30], which is sufficient in many of the practical applications mentioned above. However, in many contexts, such as in computational algebraic geometry, system verification, and automated theorem proving, it is required that the computed bounds be certified rigorously, in exact arithmetic. Such rational certificates are not always guaranteed to exist. In particular, rational polynomials on the boundary of the sums-of-squares cone may not have a rational sums-of-squares decomposition [42], [24]. On the other hand, polynomials in the interior of the sums-of-squares cone *do* have rational decompositions, a result due to Powers [34], of which we provide a new elementary proof in Section 2.1.1.

*Submitted on May 9, 2022.

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

†North Carolina State University, Department of Mathematics. Email: mlmacaul@ncsu.edu.

‡North Carolina State University, Department of Mathematics. Email: dpapp@ncsu.edu.

Computing rational WSOS decompositions for polynomials with rational coefficients is a challenging problem even in the univariate case [20]. Symbolic methods such as those that rely on quantifier elimination or root isolation are exponential in the number of variables of the input polynomial, and in the univariate case have been found to be less efficient than more specialized methods. The optimal value of the semidefinite program is an algebraic number, but the study of the algebraic degree of the positive semidefinite cone [27] suggests that one cannot hope for easily computable and verifiable certificates from taking a purely symbolic computational 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 [33, 17, 6]; see also [9, 10, 3]. (Certifying a WSOS lower bound of a polynomial, and the related goal of certifying a nonnegativity of a polynomial using rational certificates, is also discussed in [19], and software developed to attain these goals includes [18] and, [41], amongst others.)

Our contribution is twofold. In Section 2, we propose a new framework for certifying that a polynomial is WSOS using *dual certificates*. The approach relies on convex programming duality and allows the efficient construction of rational WSOS decompositions from suitable rational 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. Moreover, 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. We also show that every rational polynomial in the interior of the WSOS cone has a rational dual certificate. This gives short new proofs to a number of known results about the existence of sum-of-rational-squares decompositions, including theorems of Powers [34].

In Section 3, we discuss various algorithmic applications of dual certificates. We propose an efficient algorithm, Algorithm 3.1, for computing and certifying rational WSOS lower bounds for polynomials over a compact semialgebraic set using dual certificates. The algorithm can be implemented as an entirely numerical method that nevertheless produces exact rational WSOS decompositions certifying rational lower bounds. The algorithm provides, in each iteration, a certifiable WSOS bound with a dual certificate that can be converted (in polynomial time) to an explicit rational WSOS decomposition without any additional rounding or projection of the numerical solutions. The sequence of WSOS bounds converges to the optimal WSOS bound at a linear rate. In Section 4, we deduce explicit bounds on the number of iterations of Algorithm 3.1 in the univariate case. Section 5 includes additional examples demonstrating the efficacy of the method and the quality of the lower bounds obtained using Algorithm 3.1 in standard polynomial optimization benchmark problems. Specifically, we demonstrate that in some cases, the best certifiable bound using our purely numerical algorithm (implemented using double-precision floating point arithmetic) is indistinguishable from the true minimum in double-precision arithmetic.

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 $\mathbf{w} = (w_1, \ldots, w_m)$ be some given nonzero polynomials and let $\mathbf{d} = (d_1, \ldots, d_m)$ be a nonnegative integer vector. We denote by $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ 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 $p = \sum_{i=1}^m w_i r_i$. A polynomial $p \in \mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ 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 $p = \sum_{i=1}^m w_i \sigma_i$. It is customary to assume that $w_1 = 1$, that is, the ordinary "unweighted" sum-of-squares polynomials are also included in the WSOS cones. Let $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ denote the set of WSOS polynomials in $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$. By definition, $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}} \subset \mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ is a full-dimensional convex cone. Additionally, under mild conditions, the cone $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ is closed and pointed; for example, it is sufficient that the set

$$S_{\mathbf{w}} \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^n \,|\, w_i(\mathbf{x}) \geq 0,\ i = 1, \ldots, m\} \tag{1.1}$$

is a unisolvent point set for the space $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ [30, Prop. 6.1]. (A set of points $S \subseteq \mathbb{R}^n$ is *unisolvent* for a space of polynomials $\mathcal{V}$ if every polynomial in $\mathcal{V}$ is uniquely determined by its function values at S.) In particular, this implies that both $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ and its dual cone have a non-empty interior.

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 $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$. Thus, $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ and $\left(\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}\right)^*$ are identified with $\mathbb{R}^U$, where $U = \dim\left(\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}\right)$.

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 $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ and the cone of positive semidefinite matrices.

PROPOSITION 1.1 ([25, Thm. 17.6]). *Fix an ordered basis $\mathbf{q} = (q_1, \ldots, q_U)$ of $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ 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 $\Lambda_i : \mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}} \left(\equiv \mathbb{R}^U\right) \to \mathbb{S}^{L_i}$ be the unique linear mapping satisfying $\Lambda_i(\mathbf{q}) = w_i \mathbf{p}_i \mathbf{p}_i^T$, and let Λ_i^* denote its adjoint. Then $\mathbf{s} \in \Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ if and only if there exist matrices $\mathbf{S}_1 \succcurlyeq \mathbf{0}, \ldots, \mathbf{S}_m \succcurlyeq \mathbf{0}$ satisfying*

$$\mathbf{s} = \sum_{i=1}^m \Lambda_i^*(\mathbf{S}_i). \tag{1.2}$$

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

$$\left(\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}\right)^* = \left\{\mathbf{x} \in \mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}} \left(\equiv \mathbb{R}^U\right) \;|\; \Lambda_i(\mathbf{x}) \succcurlyeq \mathbf{0} \quad \forall i = 1, \ldots, m\right\}. \tag{1.3}$$

The proof of Proposition 1.1 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 $\mathbf{s}$.

EXAMPLE 1. *Throughout this example, matrices and vectors are indexed from 0; x_i denotes the ith component of the vector $\mathbf{x}$, and S_{ij} denotes the (i,j)-th entry of the matrix $\mathbf{S}$.*

1. *If both $\mathbf{p}$ and $\mathbf{q}$ from Proposition 1.1 are the standard monomial bases of univariate polynomials of degree d and $2d$, respectively, then $\Lambda : \mathbb{R}^{2d+1} \to \mathbb{S}^{d+1}$ maps a vector in $\mathbb{R}^{2d+1}$ to its corresponding Hankel matrix in $\mathbb{S}^{d+1}$, and Λ^* maps a symmetric matrix $\mathbf{S}$ in $\mathbb{S}^{d+1}$ to vectors in $\mathbb{R}^{2d+1}$ by summing along its antidiagonals. For example, if $d = 2$ (and so $L = d + 1 = 3$ and $U = 2d + 1 = 5$), then Λ and its adjoint are given by the equations*
$$\Lambda(\mathbf{x}) = \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix}$$
and
$$\Lambda^*(\mathbf{S}) = (S_{00}, 2S_{01}, 2S_{02} + S_{11}, 2S_{12}, S_{22}).$$
2. *If both $\mathbf{p}$ and $\mathbf{q}$ from Proposition 1.1 are univariate Chebyshev polynomials of degree d and $2d$, respectively, then letting T_i correspond to the ith Chebyshev polynomial and using the identity $T_iT_j = \frac{1}{2}\left(T_{i+j} + T_{|i-j|}\right)$, we deduce that the Λ operator in this setting satisfies $\Lambda(\mathbf{x})_{ij} = \frac{x_{i+j}+x_{|i-j|}}{2}$. For example, for $d = 2$ and $\mathbf{x} \in \mathbb{R}^5$,*
$$\Lambda(\mathbf{x}) = \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & \frac{x_0+x_2}{2} & \frac{x_1+x_3}{2} \\ x_2 & \frac{x_1+x_3}{2} & \frac{x_0+x_4}{2} \end{pmatrix}$$
is a Hankel-plus-Toeplitz matrix and
$$\Lambda^*(\mathbf{S}) = \left(S_{00} + \frac{S_{11}+S_{22}}{2}, 2S_{01} + S_{12}, 2S_{02} + \frac{S_{11}}{2}, S_{12}, \frac{S_{22}}{2}\right).$$
3. *If both $\mathbf{p}$ and $\mathbf{q}$ from Proposition 1.1 are the standard univariate monomial bases, and we use weights 1 and $1 - z^2$, then $\Lambda_1(\mathbf{x})$ is a standard Hankel matrix and $\Lambda_2(\mathbf{x})$ is a shifted Hankel matrix (sometimes also referred to as* localizing matrices *[16]). For example, if $d = 2$, then $U = 5, L_1 = 3, L_2 = 2$, and the operator $\Lambda \stackrel{\text{def}}{=} \Lambda_1 \oplus \Lambda_2 \colon \mathbb{R}^5 \to \mathbb{S}^3 \oplus \mathbb{S}^2$ is given by*
$$\Lambda(x_0, x_1, x_2, x_3, x_4) = \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix} \oplus \begin{pmatrix} x_0 - x_2 & x_1 - x_3 \\ x_1 - x_3 & x_2 - x_4 \end{pmatrix};$$
see, for example, [12, Sec. II.2]. The adjoint operator is given by
$$\Lambda^*(\mathbf{S}^1 \oplus \mathbf{S}^2) = \left(S^1_{00} + S^2_{00}, 2S^1_{01} + 2S^2_{01}, 2S^1_{02} + S^1_{11} - S^2_{00} + S^2_{11}, 2S^1_{12} - 2S^2_{01}, S^1_{22} - S^2_{11}\right).$$

To lighten the notation, throughout the rest of the paper we assume that the weight polynomials $\mathbf{w} = (w_1, \ldots, w_m)$ and the degrees $\mathbf{d} = (d_1, \ldots, d_m)$ are fixed, and denote the cone $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ by Σ and the space of polynomials $\mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}}$ by $\mathcal{V}$. Additionally, we denote by Λ the $\mathbb{R}^U \to \mathbb{S}^{L_1} \oplus \cdots \oplus \mathbb{S}^{L_m}$ linear map $\Lambda_1(\cdot) \oplus \cdots \oplus \Lambda_m(\cdot)$ from Proposition 1.1. With this notation, the condition (1.2) 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. (1.3) simplifies to

$$\Sigma^* = \{\mathbf{x} \in \mathbb{R}^U \mid \Lambda(\mathbf{x}) \succcurlyeq \mathbf{0}\}. \tag{1.4}$$

The interior of this cone is simply

$$(\Sigma^*)^\circ = \{\mathbf{x} \in \mathbb{R}^U \mid \Lambda(\mathbf{x}) \succ \mathbf{0}\}. \tag{1.5}$$

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 1.1 above, and let Λ^* denote its adjoint. Central to our theory is the *barrier function* $f : (\Sigma^*)^\circ \to \mathbb{R}$ defined by

$$f(\mathbf{x}) \stackrel{\text{def}}{=} -\ln(\det(\Lambda(\mathbf{x}))). \tag{1.6}$$

Note that by Eq. (1.5), 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* $\langle \cdot, \cdot \rangle_{\mathbf{x}} : \mathcal{V}^* \times \mathcal{V}^* \to \mathbb{R}$ 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 (1.6) falls into the broader category of *logarithmically homogeneous self-concordant barriers* (or LHSCBs for short), which are expounded upon in the classic texts [26] and [39]. Throughout, we will invoke several useful results concerning LHSCBs for the function (1.6); these are enumerated in the following lemma:

LEMMA 1.2. *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*
$$1 - \|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}} \leq \frac{\|\mathbf{v}\|_{\mathbf{u}}}{\|\mathbf{v}\|_{\mathbf{x}}} \leq (1 - \|\mathbf{u} - \mathbf{x}\|_{\mathbf{x}})^{-1}. \tag{1.7}$$
2. *The gradient g of f can be computed as*
$$g(\mathbf{x}) = -\Lambda^*(\Lambda(\mathbf{x})^{-1}), \tag{1.8}$$
and the Hessian $H(\mathbf{x})$ is the linear operator satisfying
$$H(\mathbf{x})\mathbf{v} = \Lambda^*\big(\Lambda(\mathbf{x})^{-1}\Lambda(\mathbf{v})\Lambda(\mathbf{x})^{-1}\big) \quad \text{for every } \mathbf{v} \in \mathbb{R}^U. \tag{1.9}$$
3. *The function f is* logarithmically homogeneous*; that is, it has the following properties:*
$$g(\alpha\mathbf{x}) = \alpha^{-1} g(\mathbf{x}) \;\; \text{and} \;\; H(\alpha\mathbf{x}) = \alpha^{-2} H(\mathbf{x}) \quad \text{for every } \alpha > 0, \tag{1.10}$$
furthermore
$$H(\mathbf{x})\mathbf{x} = -g(\mathbf{x}) \quad \text{and} \quad \|g(\mathbf{x})\|^*_{\mathbf{x}} = \|\mathbf{x}\|_{\mathbf{x}} = \sqrt{\langle -g(\mathbf{x}), \mathbf{x} \rangle} = \sqrt{\nu}, \tag{1.11}$$
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*

$$\|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^* \leq \frac{\|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}}}{1-\|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}}}. \tag{1.12}$$

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

$$\|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}} \leq \frac{\|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^*}{1-\|g(\mathbf{u}) - g(\mathbf{x})\|_{\mathbf{x}}^*}. \tag{1.13}$$

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 [39, Sec. 2.2.1 and Thm. 2.2.7].
2. Straightforward calculation.
3. Straightforward calculation using the identities (1.8) and (1.9). We remark that these identities hold for all LHSCBs [39, Thm. 2.3.9].
4. See [39, Sec. 3.3].
5. See [29, 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 2.1), and by showing in Theorem 2.2 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 $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ with non-empty interior and H denote the Hessian of the barrier function f defined in (1.6).

DEFINITION 2.1. *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*

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

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

EXAMPLE 2. *To keep this example as simple as possible, we consider unweighted univariate sum-of-squares polynomials represented in the monomial basis. Consider the univariate polynomial* $z \mapsto 1 + z^2$ *corresponding to the coefficient vector* $\mathbf{s} = (1,0,1) \in \Sigma$, *the (unweighted) sums-of-squares cone of polynomials with degree at most 2. We can characterize* $\mathcal{C}(\mathbf{s})$ *as follows. By definition, a vector* $\mathbf{x} = (x_0, x_1, x_2)$ *belongs to the cone of dual certificates* $\mathcal{C}(\mathbf{s})$ *if and only if* $\Lambda(H(\mathbf{x})^{-1}\mathbf{s}) \succcurlyeq \mathbf{0}$ *and* $\mathbf{x} \in (\Sigma^*)^\circ$. *The inverse Hessian at* $\mathbf{x}$ *is*

$$H(\mathbf{x})^{-1} = \begin{pmatrix} x_0^2 & x_0x_1 & x_1^2 \\ x_0x_1 & \frac{1}{2}(x_1^2 + x_0x_2) & x_1x_2 \\ x_1^2 & x_1x_2 & x_2^2 \end{pmatrix},$$

so $H(\mathbf{x})^{-1}\mathbf{s} = (x_0^2 + x_1^2, x_0x_1 + x_1x_2, x_1^2 + x_2^2)$. *Then, we have*

$$\Lambda(H(\mathbf{x})^{-1}\mathbf{s}) = \begin{pmatrix} x_0^2 + x_1^2 & x_0x_1 + x_1x_2 \\ x_0x_1 + x_1x_2 & x_1^2 + x_2^2 \end{pmatrix} = \Lambda(\mathbf{x})^2,$$

so $\Lambda(H(\mathbf{x})^{-1}\mathbf{s})$ *is automatically positive semidefinite. Hence* $\mathbf{x} \in \mathcal{C}(\mathbf{s})$ *for every* $\mathbf{x} \in (\Sigma^*)^\circ$.

The following theorem justifies the terminology introduced in Definition 2.1. Through Eq. (2.1) below, we can construct a WSOS certificate $\mathbf{S}$ for the polynomial $\mathbf{s}$ in the spirit of Proposition 1.1 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 2.2. *Let* $\mathbf{x} \in (\Sigma^*)^\circ$ *be arbitrary. Then the matrix* $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{s})$ *defined by*

$$\mathbf{S}(\mathbf{x}, \mathbf{s}) \stackrel{\text{def}}{=} \Lambda(\mathbf{x})^{-1}\Lambda\big(H(\mathbf{x})^{-1}\mathbf{s}\big)\,\Lambda(\mathbf{x})^{-1} \tag{2.1}$$

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 be shown by applying the Hessian formula from Lemma 1.2:

$$\Lambda^*(\mathbf{S}) \stackrel{(2.1)}{=} \Lambda^*\left(\Lambda(\mathbf{x})^{-1}\Lambda\big(H(\mathbf{x})^{-1}\mathbf{s}\big)\,\Lambda(\mathbf{x})^{-1}\right) \stackrel{(1.9)}{=} H(\mathbf{x})H(\mathbf{x})^{-1}\mathbf{s} = \mathbf{s},$$

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 (1.4) of Σ^*. ∎

Recall from Lemma 1.2 (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

$$H(\mathbf{x})^{-1}\mathbf{s} = -H(\mathbf{x})^{-1}g(\mathbf{x}) \stackrel{(1.11)}{=} \mathbf{x} \in (\Sigma^*)^\circ.$$

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

DEFINITION 2.3. *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 $\mathbf{x}$. Analogously, if $\mathbf{x}$ is a dual certificate of $\mathbf{s}$, then $\mathbf{x}$ is also a dual certificate of every positive multiple of $\mathbf{s}$. (One may also confirm directly that the matrix $\mathbf{S}$ constructed in (2.1) 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 $(\Sigma^*)^\circ \times \Sigma^\circ$, all vectors in some ($\mathbf{s}$-dependent) neighborhood of $\mathbf{x}$ are dual certificates of $\mathbf{s}$, as they also give rise to a positive semidefinite $\mathbf{S}(\mathbf{x}, \mathbf{s})$. 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 theorem is a quantitative version of this observation. (Recall that ν denotes the barrier parameter introduced in Eq. (1.11).)

THEOREM 2.4. *Suppose* $\mathbf{t} \in \Sigma^\circ$ *and let* $\mathbf{x} \in (\Sigma^*)^\circ$ *be any vector that satisfies the inequality*

$$\mathbf{t}^{\mathrm{T}}\left(\mathbf{x}\mathbf{x}^{\mathrm{T}} - (\nu - 1)H(\mathbf{x})^{-1}\right)\mathbf{t} \geq 0. \tag{2.2}$$

Then $\mathbf{x} \in \mathcal{C}(\mathbf{t})$, *and 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

$$\|\mathbf{t} - \mathbf{s}\|_{\mathbf{x}}^* = \|H(\mathbf{x})^{-1/2}(\mathbf{t} - \mathbf{s})\| = \|H(\mathbf{x})^{1/2}(\mathbf{x} - H(\mathbf{x})^{-1}\mathbf{t})\| = \|\mathbf{x} - H(\mathbf{x})^{-1}\mathbf{t}\|_{\mathbf{x}}. \tag{2.3}$$

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 1.2, $\overline{B_{\mathbf{x}}(\mathbf{x}, 1)} \subseteq \Sigma^*$, 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 (2.2) 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 real 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. (1.11), 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 1.2 again, we have

$$\begin{aligned}
1 &\geq \frac{1}{\delta}\|H(\mathbf{x})^{-1/2}(\mathbf{t} + \delta g(\mathbf{x}))\| \\
&\overset{(1.11)}{=} \left\|\delta H(\mathbf{x})^{1/2}\left(\delta^{-2} H(\mathbf{x})^{-1}\mathbf{t} - \delta^{-1}\mathbf{x}\right)\right\| \\
&\overset{(1.10)}{=} \left\|H\left(\delta^{-1}\mathbf{x}\right)^{1/2}\left(H\left(\delta^{-1}\mathbf{x}\right)^{-1}\mathbf{t} - \delta^{-1}\mathbf{x}\right)\right\|,
\end{aligned}$$

so by the identities (2.3) 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 2.5. *Suppose that $\mathbf{x} \in \Sigma^*$, $\mathbf{s} = -g(\mathbf{x})$, and that $\mathbf{y}$ is a vector that satisfies the inequality $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} < \frac{1}{2}$. Then $\mathbf{y} \in \Sigma^*$, and $\mathbf{x}$ certifies $\mathbf{t} = -g(\mathbf{y})$.*

Proof. If $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} < \frac{1}{2}$, Lemma 1.2 ensures that $\mathbf{y} \in \Sigma^*$, and we also have

$$\|\mathbf{s} - \mathbf{t}\|_{\mathbf{x}}^* = \|g(\mathbf{x}) - g(\mathbf{y})\|_{\mathbf{x}}^* \overset{(1.12)}{\leq} \frac{\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}}}{1 - \|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}}} < 1.$$

Then by Theorem 2.4, $\mathbf{x}$ certifies $\mathbf{t}$. ∎

EXAMPLE 3. *Consider the univariate polynomial t given by $t(z) = 1 - z + z^2 + z^3 - z^4$. To show that t is nonnegative on the interval $[-1, 1]$, it suffices to show that the coefficient vector $\mathbf{t} = (1, -1, 1, 1, -1)$ is a member of $\Sigma^{\mathbf{w}}_{1,2\mathbf{d}}$, with weights $\mathbf{w}(z) = (1, 1 - z^2)$ and degree vector $\mathbf{d} = (2, 1)$. For this example, we represent all polynomials in the monomial basis. In this setting, the Λ and Λ^* operators are precisely those detailed in Example 1 part (3).*

Consider the vector $\mathbf{x} = (5, 0, 5/2, 0, 15/8)$. This vector is the gradient certificate of the constant one polynomial, since simple arithmetic yields that $-g(\mathbf{x}) = \Lambda^*(\Lambda(\mathbf{x})^{-1}) = (1, 0, 0, 0, 0)$. The same certificate also certifies the nonnegativity of the polynomial t above. To confirm this, we compute $H(\mathbf{x})^{-1}$:

$$H(\mathbf{x})^{-1} = \frac{5}{384}\begin{pmatrix} 384 & 0 & 192 & 0 & 144 \\ 0 & 240 & 0 & 180 & 0 \\ 192 & 0 & 176 & 0 & 152 \\ 0 & 180 & 0 & 165 & 0 \\ 144 & 0 & 152 & 0 & 149 \end{pmatrix},$$

and by Theorem 2.2, it is sufficient to verify that

$$\frac{128}{5}\Lambda\left(H(\mathbf{x})^{-1}\mathbf{t}\right) = \begin{pmatrix} 144 & -20 & 72 \\ -20 & 72 & -5 \\ 72 & -5 & 49 \end{pmatrix} \oplus \begin{pmatrix} 72 & -15 \\ -15 & 23 \end{pmatrix} \succcurlyeq \mathbf{0}.$$

With some additional work, we can also compute from $\mathbf{x}$ *rational matrices* $\mathbf{S}_1$ *and* $\mathbf{S}_2$ *to certify the nonnegativity of the polynomial using Proposition 1.1: plugging our dual certificate into the formula (2.1), we obtain*

$$\mathbf{S}_1 = \frac{1}{40}\begin{pmatrix} 22 & -5 & -26 \\ -5 & 18 & 5 \\ -26 & 5 & 52 \end{pmatrix} \quad \textit{and} \quad \mathbf{S}_2 = \frac{1}{40}\begin{pmatrix} 18 & -15 \\ -15 & 92 \end{pmatrix}.$$

These matrices, in turn, can be factored using the LDL^{T} *form of Cholesky decomposition to compute an explicit rational sum-of-squares representation of* t*:*

$$\begin{aligned} t(z) =& \frac{11}{20}\left(-\frac{13z^2}{11} - \frac{5z}{22} + 1\right)^2 + \frac{371}{880}\left(z - \frac{20z^2}{371}\right)^2 + \frac{3937z^4}{7420} + \\ &+ \left(1 - z^2\right)\left(\frac{9}{20}\left(1 - \frac{5z}{6}\right)^2 + \frac{159z^2}{80}\right). \end{aligned}$$

2.1. Rational nonnegativity certificates. We now turn to an important theoretical application of dual certificates and show that every *rational polynomial* (that is, every polynomial with rational coefficients) in the interior of Σ has a rational dual certificate, and show that this implies that every rational polynomial in Σ° is a sum of *rational* squares. (We cannot hope this to be true on the boundary; this follows from Scheiderer's seminal result [42], although facial reduction techniques could yield rational certificates in some cases [24].) As an immediate consequence, we get new short proofs of two theorems of Powers [34] that positive rational polynomials over compact sets, under the assumptions of Schmüdgen's (resp., Putinar's) Positivstellensatz, have rational sum-of-squares certificates.

We also show using similar arguments that these rational sum-of-squares certificates can be computed from dual certificates using numerical algorithms.

2.1.1. Existence of rational WSOS certificates for rational polynomials.

THEOREM 2.6. *Every* $\mathbf{s} \in \Sigma^\circ \cap \mathbb{Q}^U$ *has a rational dual certificate* $\mathbf{x} \in (\Sigma^*)^\circ \cap \mathbb{Q}^U$ *and a sum-of-rational-squares decomposition.*

Proof. Let $\mathbf{s} \in \Sigma^\circ$, and suppose $\mathbf{y} \in (\Sigma^*)^\circ$ is the gradient certificate of $\mathbf{s}$. As $\mathcal{C}(\mathbf{s})$ is full-dimensional for every $\mathbf{s} \in \Sigma^\circ$ by Theorem 2.4, and as $H(\mathbf{y})$ is positive definite for every $\mathbf{y} \in (\Sigma^*)^\circ$, the ball centered at $\mathbf{y}$ with radius $1/2$ in the local norm is a full-dimensional ellipsoid contained in $(\Sigma^*)^\circ$. As a result, there exists a rational vector $\mathbf{x} \in \mathbb{Q}^U \cap (\Sigma^*)^\circ$ such that $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{y}} \leq 1/2$, which by Corollary 2.5 guarantees that $\mathbf{x}$ is a rational dual certificate for $\mathbf{s}$. Finally, by Theorem 2.2, we can derive a conventional (primal) weighted sum-of-*rational*-squares decomposition for $\mathbf{s}$: the Gram matrix $\mathbf{S}(\mathbf{x}, \mathbf{s})$ defined in (2.1) is automatically rational, and from there an explicit rational WSOS decomposition can be computed in rational arithmetic via the LDL^{T} decomposition of this Gram matrix.

Recall that according to Putinar's celebrated Positivstellensatz [36], every (strictly) positive polynomial over a compact semialgebraic set (1.1) belongs to a WSOS cone $\Sigma = \Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ for a sufficiently large $\mathbf{d}$, if the quadratic module associated with the polynomials $w_1, \ldots, w_m$ is Archimedean. It follows that (under the same assumption)

every positive polynomial belongs to the *interior* of $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ as well for every sufficiently large $\mathbf{d}$. (Every positive polynomial $\mathbf{s}$ lies in the interior of a simplex whose vertices are positive polynomials, and by Putinar's Positivstellensatz each vertex of this simplex belongs to $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ for every large enough $\mathbf{d}$; our $\mathbf{s}$ then belongs to the interior of this WSOS cone.) Invoking Theorem 2.6, we have the following "rational Putinar's Posivstellensatz" for rational polynomials:

COROLLARY 2.7 ([34, Thm. 7]). *Let $\mathbf{s} \in \mathbb{Q}^U$ be the coefficient vector of a polynomial $s(\cdot)$ that is strictly positive on the compact semialgebraic set $S_{\mathbf{w}}$ defined in (1.1), and suppose that the quadratic module associated with $\{w_1, \ldots, w_m\}$ is Archimedean and each w_i is rational. Then s has weighted-sum-of-rational-squares decomposition using the weights $w_1, \ldots, w_m$.*

In an analogous fashion, we can show that as long as $S_{\mathbf{w}}$ is compact (without the Archimedean assumption), every rational polynomial positive over $S_{\mathbf{w}}$ has a rational Schmüdgen nonnegativity certificate as was also shown by Powers; see [34, Thm. 5].

2.1.2. Rigorous certificates from numerical methods. Theorem 2.4 and Corollary 2.5 have important consequences for both numerical (finite-precision), exact-arithmetic, and hybrid algorithms for computing sum-of-squares certificates.

The fact that every polynomial $\mathbf{s} \in \Sigma^\circ$ has a full-dimensional cone of certificates $\mathcal{C}(\mathbf{s})$ means that exact dual certificates can in principle be computed by purely numerical, inexact algorithms. As long as $\mathbf{s}$ is a rational polynomial that is sufficiently in the interior of Σ that a numerical method (implemented in floating point arithmetic) can identify *any* point $\mathbf{x} \in \mathcal{C}(\mathbf{s})$, the same argument as in the proof of Theorem 2.6 shows that a rational weighted-sum-of-squares certificate can be readily computed.

We can also take this argument one step further and apply it to certifying sum-of-squares lower bounds. Consider, for example, a hypothetical algorithm that aims to compute the gradient certificate $\mathbf{y}$ of some polynomial $t-\gamma$ with coefficient vector $\mathbf{t}-\gamma\mathbf{1} \in \Sigma^\circ$ to certify $t(\mathbf{z}) \geq \gamma \, \forall \mathbf{z} \in S_{\mathbf{w}}$, but computes instead only an approximation $\mathbf{x} \approx \mathbf{y}$ in finite-precision arithmetic. (Here, $\mathbf{1}$ denotes the coefficient vector of the constant one polynomial; $S_{\mathbf{w}}$ is the semialgebraic set (1.1) over which we wish to bound the polynomial t.) As long as the inherent errors of the finite-precision computation are small enough to ensure $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \leq 1/2$, Corollary 2.5 guarantees that $\mathbf{x}$ is a certificate of nonnegativity for $t-\gamma$. Since floating-point numbers are, by definition, rational, every sufficiently accurate numerical solution of $-g(\mathbf{x}) = \mathbf{t}-\gamma\mathbf{1}$ is automatically a rational dual certificate of the nonnegativity of $t-\gamma$. Additionally, as long as the coefficient vector $\mathbf{t}$ and the lower bound γ are also rational, any such numerical dual certificate $\mathbf{x}$ can be directly converted to an exact rational primal certificate (Gram matrix) $\mathbf{S} \succcurlyeq \mathbf{0}$ satisfying $\Lambda^*(\mathbf{S}) = \mathbf{t}-\gamma\mathbf{1}$ via the formula of Eq. (2.1).

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

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

will generally satisfy the equality constraints $\Lambda^*(\mathbf{S}) = \mathbf{t}-\gamma\mathbf{1}$ 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. Hence, additional post-processing (a rounding or projection step, such as those in the hybrid methods of [33] and [21]) is needed. In contrast, any dual certificate $\mathbf{x}$ from the full-dimensional cone $\mathcal{C}(\mathbf{t}-\gamma\mathbf{1})$ is a rational certificate (that can be turned into an explicit rational WSOS decomposition).

In Section 3.3, we present an efficient algorithm (Algorithm 3.1) to compute certi-

fiable rational lower bounds with matching dual certificates that can be implemented as an entirely numerical method using these ideas.

2.2. Complexity 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 1.1), 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 rational Cholesky (LDL^{T}) factorization of $\Lambda(\mathbf{x})$ and $H(\mathbf{x})$ (which are positive definite by definition); (2) computing the vector $\mathbf{v} = H(\mathbf{x})^{-1}\mathbf{s}$ using the Cholesky factors of $H(\mathbf{x})$ computed in the previous step; and (3) computing $\Lambda(\mathbf{v})$ 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 [30, Sec. 6] showed that when polynomials are represented as Lagrange interpolants, the Hessian $H(\mathbf{x})$ can be computed with $\mathcal{O}(m \max_i\{L_i\}U^2)$ arithmetic operations. One can also argue directly from the identity (1.9), 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 important 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-plus-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 [28].

2.3. Relation to prior work. (Weighted) sum-of-squares certificates are commonly associated with, and computed using, semidefinite optimization—an approach that goes back to Nesterov [25], Parrilo [32], and Lasserre [14]. It is pertinent to put our work in the context of Lasserre's, as both make extensive use of the dual cone Σ° and thus have many superficial similarities.

In our notation, the semidefinite optimization approach can be summarized as follows: to find the best WSOS lower bound for a polynomial t, whose coefficient vector is denoted by $\mathbf{t}$, we need to solve the semidefinite optimization problem

$$\sup\{c \,|\, \Lambda^*(\mathbf{S}) = \mathbf{t} - c\mathbf{1},\, \mathbf{S} \succcurlyeq 0\}. \tag{2.4}$$

The WSOS certificate itself is the *Gram matrix* $\mathbf{S}$, whose factorization yields an explicit representation of $t - c$. Lasserre's seminal observation is that the WSOS lower bound can also be characterized as the optimal value of the dual semidefinite optimization problem (which can also be derived from moment theory), written as

$$\inf\{\mathbf{t}^{\mathrm{T}}\mathbf{y} \,|\, \mathbf{1}^{\mathrm{T}}\mathbf{y} = 1, \Lambda(\mathbf{y}) \succcurlyeq 0\}. \tag{2.5}$$

In fact, an immediate consequence of weak duality is that every feasible solution of (2.5) yields a WSOS-certifiable lower bound on t. Additionally, under standard regularity conditions (such as the existence of a Slater point in (2.5)) we have strong duality, with attainment in the primal problem (2.4), meaning that the optimal value of (2.5) is the best WSOS-certifiable lower bound [14].

Our work extends this theory. Although the above arguments show that every $\mathbf{y} \in \Sigma^*$ with $\mathbf{1}^T\mathbf{y} = 1$ yields a WSOS lower bound, these dual vectors cannot be turned into explicit ("primal") WSOS certificates—a dual optimal (or feasible) solution $\mathbf{y}$ from (2.5) does not translate to an optimal (or feasible) solution $\mathbf{S}$ for any particular c in (2.4). One interpretation of Theorem 2.4 is that it identifies a full-dimensional subset of solutions of (2.5) which, through the definition (2.1), can be turned into a primal certificate $\mathbf{S}$ via a simple closed-form formula. This allows us to circumvent solving the semidefinite optimization problem (2.4) altogether. This is a potentially huge gain, as the dimension of $\mathbf{S}$ in (2.4) is considerably larger than $\dim(\Sigma)$. We demonstrate how dual certificates can be used in designing efficient algorithms in Section 3.3.

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 argue that this is a mild, essentially without loss of generality, condition in Section 3.5.) 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 lower 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.

We then combine these ideas with the observations made in Section 2.1.2 and present a new algorithm (Algorithm 3.1) for approximating the best WSOS lower bound for a given polynomial with arbitrary accuracy in Section 3.3. The algorithm returns both a rational lower bound approximating the optimal WSOS lower bound and a rational certificate certifying the bound. We also show that Algorithm 3.1 is linearly convergent to the optimal bound. In Section 3.4, we detail how to compute a bound on the linear rate of convergence of Algorithm 3.1. This in turn makes it possible to compute WSOS lower bounds that are certifiably within a prescribed ε from the optimal bound.

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 $\Sigma\,(= \Sigma^{\mathbf{w}}_{n,2\mathbf{d}})$, in the space of polynomials $\mathcal{V}(= \mathcal{V}^{\mathbf{w}}_{n,2\mathbf{d}})$.

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 in Theorem 2.4, $\mathbf{1} \in \mathcal{P}(\mathbf{x}_1)^\circ$, 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 observation is that each of these certificates also certifies *some* WSOS lower bound for *every* polynomial:

LEMMA 3.1. *Let $\mathbf{x} \in (\Sigma^*)^\circ$ be any certificate for which $\mathbf{1} \in \mathcal{P}(\mathbf{x})^\circ$ and $r \in (0, 1/2]$. 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 $\mathbf{x} \in (\Sigma^*)^\circ$ satisfies $\| - g(\mathbf{x}) - \mathbf{1}\|^*_{\mathbf{x}} \leq \frac{r}{r+1}$, then the*

inclusion $\mathbf{x} \in \mathcal{C}(\mathbf{t} + c\mathbf{1})$ *holds for every*

$$c \geq \frac{\|\mathbf{t}\|_{\mathbf{x}}^*}{\frac{r}{r+1} - \| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^*}. \tag{3.1}$$

In this case, letting $\mathbf{y}_c$ *denote the gradient certificate of* $\mathbf{t} + c\mathbf{1}$*, the inequality*

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

also holds.

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 and third statements in detail.

Using the definitions of the local dual norm and logarithmic homogeneity (1.10) from Lemma 1.2, we have

$$\|(\mathbf{t} + c\mathbf{1}) - c\mathbf{1}\|_{c^{-1}\mathbf{x}}^* \overset{\text{(by def.)}}{=} \|H(c^{-1}\mathbf{x})^{-1/2}\mathbf{t}\| \overset{(1.2)}{=} c^{-1}\|H(\mathbf{x})^{-1/2}\mathbf{t}\| \overset{\text{(by def.)}}{=} c^{-1}\|\mathbf{t}\|_{\mathbf{x}}^* \tag{3.2}$$

Similarly,

$$\|c\mathbf{1}+cg(\mathbf{x})\|_{c^{-1}\mathbf{x}}^* \overset{\text{by def.}}{=} c\|H(c^{-1}\mathbf{x})^{-1/2}(\mathbf{1}+g(\mathbf{x}))\| \overset{(1.2)}{=} \|H(\mathbf{x})^{-1/2}(\mathbf{1}+g(\mathbf{x}))\| \overset{\text{by def.}}{=} \|\mathbf{1}+g(\mathbf{x})\|_{\mathbf{x}}. \tag{3.3}$$

Thus, we have

$$\begin{aligned}
\|(\mathbf{t} + c\mathbf{1}) + cg(\mathbf{x})\|_{c^{-1}\mathbf{x}}^* &\leq \|(\mathbf{t} + c\mathbf{1}) - c\mathbf{1}\|_{c^{-1}\mathbf{x}}^* + \|c\mathbf{1} + cg(\mathbf{x})\|_{c^{-1}\mathbf{x}}^* \\
&\overset{(3.2),(3.3)}{=} c^{-1}\|\mathbf{t}\|_{\mathbf{x}}^* + \|\mathbf{1} + g(\mathbf{x})\|_{\mathbf{x}}^* \\
&\overset{(3.1)}{\leq} \frac{r}{r+1} - \| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^* + \| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^* \\
&= \frac{r}{r+1}.
\end{aligned} \tag{3.4}$$

Using logarithmic homogeneity again, we see that $c^{-1}\mathbf{x}$ is the gradient certificate for $-cg(\mathbf{x})$. Therefore, invoking Theorem 2.4, we deduce from the inequality (3.4) that $c^{-1}\mathbf{x}$ is a dual certificate for $\mathbf{t}+c\mathbf{1}$. Moreover, via the inequality (1.13) in Lemma 1.2, we conclude that

$$\|c^{-1}\mathbf{x} - \mathbf{y}_c\|_{c^{-1}\mathbf{x}} \overset{(1.13)}{\leq} \frac{\|\mathbf{t}\|_{c^{-1}\mathbf{x}}^*}{1 - \|\mathbf{t}\|_{c^{-1}\mathbf{x}}^*} \overset{(3.4)}{\leq} r,$$

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 3.1 only needs to be computed once for any particular WSOS cone $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$. 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 every polynomial $\mathbf{t} \in \mathcal{V}$, with minimal effort.

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

EXAMPLE 4. *Consider the cone of nonnegative univariate polynomials of degree* $2d$ *over the interval* $[-1, 1]$*, which is well known to be the same as the WSOS cone* $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ *with* $n = 1$*,* $m = 2$*, degree vector* $\mathbf{d} = (d, d-1)$*, and weight polynomials* $\mathbf{w}(z) = (1, 1 - z^2)$ *[4]. Furthermore, suppose that all polynomials are represented in the basis of Chebyshev polynomials of the first kind, that is, both of the ordered*

bases $\mathbf{p}$ *and* $\mathbf{q}$ *in Proposition 1.1 that determine the operator* Λ *are Chebyshev basis polynomials. Then both diagonal blocks of* Λ *are Hankel-plus-Toeplitz matrices (similar to Example 1), and the gradient certificate of* $\mathbf{1} = (1, 0, \ldots, 0) \in \mathbb{R}^{2d+2}$ *is simply the vector*

$$\mathbf{x}_1 = (2d + 1, 0, \ldots, 0).$$

This can be proven by a 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*

$$H(\mathbf{x}_1) = \frac{1}{2d+1} \operatorname{diag}\left(1, \frac{4d}{2d+1}, \frac{4d-2}{2d+1}, \ldots, \frac{2}{2d+1}\right). \tag{3.5}$$

Analogous results can be derived for polynomials of odd degree using $\mathbf{d} = (d, d)$, *and weight polynomials* $\mathbf{w}(z) = (1 - z, 1 + z)$.

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 Theorem 2.4: 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.

EXAMPLE 5. *Continuing with Example 3 (with* $\mathbf{t} = (1, -1, 1, 1, -1)$, *weights* $\mathbf{w}(z) = (1, 1 - z^2)$*), we compute* c_{max} *and* c'_{max} *for* t *using the certificate* $\mathbf{x} = (5, 0, 5/2, 0, 15/8)$. *For comparison, the minimum of the polynomial is* $\frac{1}{512}\left(619 - 51\sqrt{17}\right) \approx 0.798$.

To compute c_{max}, *we compute the largest* γ *such that* $\Lambda(H(\mathbf{x})^{-1}(\mathbf{t} - \gamma\mathbf{1})) = \Lambda_1(H(\mathbf{x})^{-1}(\mathbf{t} - \gamma\mathbf{1})) \oplus \Lambda_2(H(\mathbf{x})^{-1}(\mathbf{t} - \gamma\mathbf{1}))$ *is positive semidefinite but not positive definite. We compute the characteristic polynomials of the* γ*-parametrized matrices, as* $\Lambda(H(\mathbf{x})^{-1}(\mathbf{t} - \gamma\mathbf{1}))$ *is on the boundary of the PSD cone when the constant term of the characteristic polynomial vanishes. The constant term of the characteristic polynomial of* $\Lambda_1(H(\mathbf{x})^{-1}(t - \gamma\mathbf{1}))$, *itself a polynomial in* γ, *has smallest real root at* $\gamma = \frac{1}{64}\left(67 - 5\sqrt{17}\right)$. *Meanwhile, the constant term of the characteristic polynomial of* $\Lambda_2(H(\mathbf{x})^{-1}(t - \gamma\mathbf{1}))$ *has smallest real root at* $\gamma = \frac{1}{32}\left(41 - 5\sqrt{10}\right)$. *We conclude that* $c_{max} = \frac{1}{64}\left(67 - 5\sqrt{17}\right) \approx 0.724$.

To compute c'_{max}, *we expand and reduce*

$$(\mathbf{t} - \gamma\mathbf{1})^{\mathrm{T}} \left(\mathbf{x}\mathbf{x}^{\mathrm{T}} - (\nu - 1)H(\mathbf{x})^{-1}\right)(\mathbf{t} - \gamma\mathbf{1}) = \frac{205}{64} - \frac{45\gamma}{4} + 5\gamma^2.$$

Computing the roots of this quadratic, we conclude that $c'_{max} = \frac{1}{8}\left(9 - 2\sqrt{10}\right) \approx 0.334$.

3.3. Computing optimal WSOS bounds. We now present an iterative method to compute *the best WSOS lower bound* for a given polynomial $\mathbf{t}$ along with a certificate for that bound. The pseudocode of the algorithm is shown in Algorithm 3.1. After a high-level description of the method, we show that it converges linearly to the optimal WSOS bound below (Theorem 3.6).

Previously, in Lemma 3.1, we showed that for a sufficiently large c, $\mathbf{t} + c\mathbf{1}$ can be certified by $c^{-1}\mathbf{x}$ for every $\mathbf{x}$ in a suitable neighborhood of the gradient certificate of $\mathbf{1}$; this result justifies the initialization of the algorithm in Line 1. In 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 3.3. 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 $-g(\mathbf{y}) = \mathbf{t} - c\mathbf{1}$.

EXAMPLE 6. *We continue with the setup of Examples 3 and 5: we consider the univariate polynomial whose coefficient vector in the monomial basis is* $\mathbf{t} = (1, -1, 1, 1, -1)$, *defined over the interval* $[-1, 1]$ *represented by the weights* $\mathbf{w}(z) = (1, 1-z^2)$. *Algorithm 3.1 with* $r = 1/4$, *with inputs* $\mathbf{t}$ *and tolerance* $\varepsilon = 10^{-7}$ *in double-precision floating point arithmetic outputs the bound* $c \approx 0.798284319$ *and a certificate vector* $\mathbf{x}$. *Note that the exact minimum of* t *is* $\frac{1}{512}(619 - 51\sqrt{17}) \approx 0.798284401$.

A plot of the difference between the current certified lower bound c *and the minimum* c^* *in each iteration is shown in Figure 3.1, illustrating the linear convergence of Algorithm 3.1 for this polynomial. The exact rational representation of the floating point bound is*

$$c = 2^{-53} \cdot 7190305926654593,$$

and the rational vector certifying the nonnegativity of $\mathbf{t} - c\mathbf{1}$ *is*

$$\mathbf{x} = 2^{-33} \begin{pmatrix} 173493184462864992 \\ 67729650226350000 \\ -120611300436615200 \\ -161900156381728960 \\ -5796381308580693 \end{pmatrix}.$$

Note that no rounding or projection steps are needed to compute a rigorous certificate. In the analysis of the algorithm below (Lemma 3.3) we shall see that if the algorithm was implemented in exact arithmetic, we would have $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \leq r = 1/4$ *in each iteration, where* $\mathbf{y}$ *is the gradient certificate of* $\mathbf{t} - c\mathbf{1}$. *Working with finite precision, the iterates may fail to satisfy this inequality; however, as long as the numerical errors are sufficiently small to ensure the considerably weaker inequality* $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \leq 1/2$, *the computed numerical certificate* $\mathbf{x}$ *is automatically a rational certificate for the computed SOS lower bound* c *by Corollary 2.5.*

Additional examples are discussed in Section 5.

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

Algorithm 3.1 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; a radius $r \in (0, 1/4]$; a certificate $\mathbf{x}$ satisfying $\| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^* \leq \frac{r}{r+1}$.

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 $c_0 := -\left(\frac{r}{r+1} - \| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^*\right)^{-1} \|\mathbf{t}\|_{\mathbf{x}}^*$. Set $c := c_0$ and $\mathbf{x} := -\frac{1}{c_0}\mathbf{x}$.

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* $\mathbf{x}$.

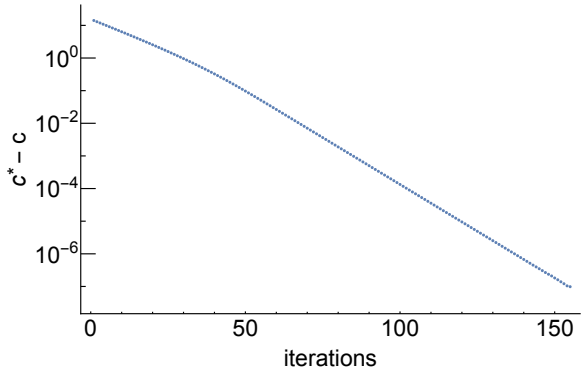

FIG. 3.1. *The convergence of the sequence of certified lower bounds computed by Algorithm 3.1 to the minimum of the polynomial studied in Examples 3 and 6, illustrating the linear convergence shown in Theorem 3.6 below.*

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 equation, and the certificate update is essentially a matrix-vector multiplication or two triangular solves. As discussed in Section 2.2, the computation and factorization of the Hessian is efficient for popular choices of polynomial bases.

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. We also establish that throughout the algorithm, the

iterates satisfy $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \le r$. (At the beginning of the first iteration this holds by Lemma 3.1.) The first result, Lemma 3.2, 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 3.2. *Let $\mathbf{x}_+$ and $\mathbf{y}$ be defined as above, and assume that $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \le r$ 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 3.1 is a single (full) Newton step towards the solution of the nonlinear 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 [39, 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. (1.7)), 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 (1.7) for another change of norms to conclude that

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

∎

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 3.2, we can guarantee that $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}$ when $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} < \frac{1}{3}$. Below, we need to further limit r to ensure that the bound update is an improvement.

LEMMA 3.3. *Suppose that $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$ for some $0 < r \le \frac{1}{4}$. Then $c_+ > c$ and $\|\mathbf{x}_+ - \mathbf{y}_+\|_{\mathbf{x}_+} \le r$.*

Proof. We begin by showing that

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

which implies that Step 4 of the algorithm indeed increases the lower bound to $c_+ > c$.

Suppose $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$. Recall from Eq. (1.11) that $H(\mathbf{x}_+)\mathbf{x}_+ = -g(\mathbf{x}_+)$. Using this identity and the definition of the local norm, we deduce that

$$\begin{aligned} \|-g(\mathbf{x}_+) + g(\mathbf{y})\|_{\mathbf{x}_+}^* &= \|H(\mathbf{x}_+)^{-1/2}\left(H(\mathbf{x}_+)\mathbf{x}_+ - (\mathbf{t}-c\mathbf{1})\right)\| \\ &= \|H(\mathbf{x}_+)^{1/2}\mathbf{x}_+ - H(\mathbf{x}_+)^{-1/2}(\mathbf{t}-c\mathbf{1})\| \\ &= \|\mathbf{x}_+ - H(\mathbf{x}_+)^{-1}(\mathbf{t}-c\mathbf{1})\|_{\mathbf{x}_+}. \end{aligned} \tag{3.6}$$

Using this in tandem with inequality (1.12) from Lemma 1.2, we have

$$\begin{aligned} \|\mathbf{x}_+ - H(\mathbf{x})^{-1}(\mathbf{t}-c\mathbf{1})\|_{\mathbf{x}_+} &\stackrel{(3.6)}{=} \|-g(\mathbf{x}_+) + g(\mathbf{y})\|_{\mathbf{x}_+} \\ &\stackrel{(1.12)}{\le} \frac{\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}}{1-\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}} \le \frac{\frac{r^2}{1-2r}}{1-\frac{r^2}{1-2r}} < \frac{r}{r+1} \end{aligned}$$

for every $r \le \frac{1}{4}$, proving our first claim.

To see the second statement, we observe that

$$\| - g(\mathbf{x}_+) + g(\mathbf{y}_+)\|_{\mathbf{x}_+}^* = \|\mathbf{x}_+ - H(\mathbf{x}_+)^{-1}(\mathbf{t} - c_+\mathbf{1})\|_{\mathbf{x}_+} = \frac{r}{r+1} < 1 \tag{3.7}$$

by the definition of the bound update step in Line 4 and our discussion above. Now inequality (1.13) from Lemma 1.2 yields

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

∎

The next lemma uses Lemma 3.2 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 3.4. *Define* $\rho_r \stackrel{\text{def}}{=} \frac{r(1-3r-2r^2)}{1-r-2r^2}$ *Then at the end of each iteration of Algorithm 3.1,* $c_+ - c \geq \frac{\rho_r}{\|\mathbf{1}\|_{\mathbf{y}}^*}$, *where* $\mathbf{y}$ *is the gradient certificate of* $\mathbf{t} - c\mathbf{1}$.

Proof. From the identities (3.7) 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

$$\frac{r}{r+1} - \| - g(\mathbf{x}_+) + g(\mathbf{y})\|_{\mathbf{x}_+}^* \leq \| - g(\mathbf{y}_+) + g(\mathbf{y})\|_{\mathbf{x}_+}^* = \|(c_+ - c)\mathbf{1}\|_{\mathbf{x}_+}^*. \tag{3.8}$$

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

From Lemma 3.2, we know that $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \leq \frac{r^2}{1-2r}$. Using the inequality (1.12) in Lemma 1.2, we have

$$\| - g(\mathbf{x}_+) + g(\mathbf{y})\|_{\mathbf{x}_+}^* \leq \frac{\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}}{1 - \|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+}} \leq \frac{\frac{r^2}{1-2r}}{1 - \frac{r^2}{1-2r}} = \frac{r^2}{1-2r-r^2}. \tag{3.9}$$

Combining the inequalities in (3.8) and (3.9), 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 (1.7),

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

∎

We remark that if r is chosen so that $0 < r \leq \frac{1}{4}$, 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 the bound can be bounded from below by a quantity proportional to $(\|\mathbf{1}\|_{\mathbf{y}}^*)^{-1}$, where $\mathbf{y}$ is the current gradient certificate.

Now, we turn our attention to the convergence of Algorithm 3.1. 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 3.5, 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 3.4 above 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 $\|\cdot\|_1$, $\|\cdot\|$ and

$\|\cdot\|_\infty$ refer to the standard 1-norm, 2-norm, and infinity norm of vectors, respectively (not to be confused with the local norms used above).

THEOREM 3.5. *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

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

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

$$\begin{aligned}
\nu &\stackrel{(1.11)}{=} \left\langle -g\left(\frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}\right), \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty} \right\rangle \\
&\stackrel{(1.10)}{=} \|\mathbf{y}\|_\infty \left\langle \mathbf{t} - c\mathbf{1}, \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty} \right\rangle \\
&= \|\mathbf{y}\|_\infty \left(\left\langle \mathbf{t} - c^*\mathbf{1}, \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}, \right\rangle + (c^* - c) \left\langle \mathbf{1}, \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty} \right\rangle \right) \\
&\geq 0 + \|\mathbf{y}\|_\infty \alpha k_1 = \|\mathbf{y}\|_\infty \alpha k_1,
\end{aligned}$$

from which we conclude that

$$\|\mathbf{y}\|_\infty \leq \frac{\nu}{\alpha k_1}. \tag{3.10}$$

Recall from Eq. (1.9) that $H(\mathbf{y})\mathbf{v} = \Lambda^*(\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{v})\Lambda(\mathbf{y})^{-1})$. Therefore, $\mathbf{v}^{\mathrm{T}}H(\mathbf{y})\mathbf{v} = \langle \mathbf{v}, \Lambda^*(\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{v})\Lambda(\mathbf{y})^{-1})\rangle = \operatorname{tr}(\Lambda(\mathbf{v})\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{v})\Lambda(\mathbf{y})^{-1})$. Moreover, observe that for every $\mathbf{A} \succcurlyeq \mathbf{0}$ 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 $\mathbf{v} \in \mathbb{R}^U$,

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

We conclude that

$$\lambda_{\min}(H(\mathbf{y})^{1/2}) \geq \frac{k_2}{\lambda_{\max}(\Lambda(\mathbf{y}))}, \tag{3.11}$$

wherein we define

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

We remark that $k_2 = \sigma_{\min}(\Lambda) > 0$ (since $\Lambda(\mathbf{v}) \neq \mathbf{0}$ whenever $\mathbf{v} \neq \mathbf{0}$).

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

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

These identities and our previous inequalities give

$$\|\mathbf{1}\|_{\mathbf{y}}^* = \|H(\mathbf{y})^{-1/2}\mathbf{1}\| \leq \frac{\|\mathbf{1}\|}{\lambda_{\min}\left(H(\mathbf{y})^{1/2}\right)} \overset{(3.11)}{\leq} \frac{\lambda_{\max}(\Lambda(\mathbf{y}))\|\mathbf{1}\|}{k_2} \leq \frac{k_3\|\mathbf{y}\|\|\mathbf{1}\|}{k_2} \overset{(3.10)}{\leq} \frac{k_3\nu\|\mathbf{1}\|}{k_1k_2\alpha}.$$

Defining $C \overset{\text{def}}{=} \frac{k_1k_2}{k_3\nu\|\mathbf{1}\|}$, 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 3.4 with Theorem 3.5, we have also proven our main result about the convergence of our algorithm:

THEOREM 3.6. *Algorithm 3.1 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 3.1, the improvement of the lower bound* $\Delta c = c_+ - c$ *satisfies*

$$\frac{\Delta c}{c^* - c} \geq \rho_r C, \tag{3.12}$$

with the absolute constant $\rho_r > 0$ *defined in Lemma 3.4 and the* Λ*-dependent constant* $C > 0$ *defined in Theorem 3.5.*

Theorem 3.6 motivates the stopping criterion (Line 6) of Algorithm 3.1: 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 rearrange the same inequality to provide an explicit upper bound on the number of iterations of the algorithm. After k iterations of Algorithm 3.1 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 above 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 $\{\mathbf{x} \in \mathbb{R}^n \mid w_i(\mathbf{x}) \geq 0,\ i = 1, \ldots, m\}$ is known, it is also straightforward to upper bound c^* by $\kappa_{\mathbf{w}}\|\mathbf{t}\|$ with some constant $\kappa_{\mathbf{w}}$ dependent only the weight functions $\mathbf{w}$. Similarly, from the first step of Algorithm 3.1 (with $\mathbf{x} \in \mathcal{C}(\mathbf{1})$),

$$\begin{aligned} c_0 &= -\left(\frac{r}{r+1} - \| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^*\right)^{-1} \|\mathbf{t}\|_{\mathbf{x}}^* \\ &\geq -\left(\frac{r}{r+1} - \| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^*\right)^{-1} \lambda_{\max}(H(\mathbf{x})^{-1})\|\mathbf{t}\|, \end{aligned}$$

bounding the initial bound c_0 from below by a Λ-dependent constant multiple of $\|\mathbf{t}\|$.

In conclusion, for a fixed cone (and representation Λ), the algorithm terminates after $\mathcal{O}(\log \frac{\|\mathbf{t}\|}{\varepsilon})$ iterations.

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 3.6, whenever the Λ-dependent constant C defined in the proof of Theorem 3.5 of is known (or can be bounded from below) for a particular cone Σ. In particular, although 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 (3.12) provides a computable upper bound on c^*.

The effect of finite precision. As all iterative numerical algorithms, Algorithm 3.1 will eventually fail to make progress before satisfying its stopping criterion if implemented in finite precision and with a tolerance ε that is too small. As c approaches the optimal WSOS lower bound, $\mathbf{t} - c\mathbf{1}$ and $\mathbf{x}$ approach the boundary of Σ and Σ^*, respectively. This is marked by the increasing ill-conditioning of the Hessian $H(\mathbf{x})$ used in the Newton step in Line 3 and of the quadratic equation solved in Line 4. Thus, the algorithm will stall if the Hessian is numerically singular or if c_+ does not improve on c due to rounding errors in the quadratic formula.

We emphasize that this only affects the quality of the bound (how close we can get to the optimal WSOS lower bound), not the correctness of the bounds that are claimed to be certified. The validity of the certificate $\mathbf{x}$ certifying the current lower bound c can be verified in rational arithmetic at the end of any iteration; there is no danger of the numerical method yielding an incorrect bound or an invalid certificate undetected. Comparing Lemma 3.3 with Corollary 2.5, we see that as long as the iterates of Algorithm 3.1 satisfy $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} < 1/2$, the certificate and the current bound will remain valid, and that with infinite-precision computation, these iterates would even satisfy $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \leq 1/4$. As long as the numerical errors are small enough that the iterates remain in the 1/2-radius local norm ball (instead of the expected 1/4-radius ball), the algorithm computes certifiable bounds and rigorous rational certificates in spite of every step of the computation being imprecise.

3.4. Bounding constants in Theorem 3.5. In general, we cannot hope to find a sharp closed-form bound for the constant C in Theorems 3.5 and 3.6, but we can compute cone-specific bounds on each of the constants k_1, k_2, and k_3 in the formula for C by convex optimization.

Recall that $k_1 = \min\{\mathbf{1}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \|\mathbf{v}\|_\infty = 1\}$. Although the norm constraint is not convex, we have

$$k_1 = \min_{1\leq i\leq U}\{\min\{k_{1,i}^-, k_{1,i}^+\}\},$$

with

$$k_{1,i}^+ = \min\{\mathbf{1}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \|\mathbf{v}\|_\infty \leq 1 \text{ and } v_i = 1\} \quad (i = 1, \ldots, m) \tag{3.13}$$

and

$$k_{1,i}^- = \min\{\mathbf{1}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \|\mathbf{v}\|_\infty \leq 1 \text{ and } v_i = -1\}. \quad (i = 1, \ldots, m) \tag{3.14}$$

Therefore, k_1 can be computed (numerically) by solving $2U$ convex optimization problems. (For a rigorous lower bound, we can use dual methods that determine approximately optimal but feasible solutions of the dual optimization problems of (3.13) and (3.14).)

Recall that $k_2 = \min\{\mathrm{tr}(\Lambda(\mathbf{v})^2) \mid \|\mathbf{v}\| = 1\}$. Hence, the constant k_2 is the smallest singular value of the linear operator Λ and can be computed to high accuracy using singular value decomposition. Alternatively, we have

$$\mathrm{tr}(\Lambda(\mathbf{v})^2) = \sum_{i=1}^{L} \Lambda_i(\mathbf{v})^2 = \mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v},$$

for a positive semidefinite rational matrix $\mathbf{M}$ that is easily computable from Λ; lower bounding k_2 amounts to lower bounding the smallest eigenvalue of the matrix $\mathbf{M}$.

Recall that the constant $k_3 = \max\{\lambda_{\max}(\Lambda(\mathbf{y})) \mid \mathbf{y} \in \Sigma^*, \|\mathbf{y}\|_\infty = 1\}$. Using the Gershgorin circle theorem, we know that

$$\lambda_{\max}(\Lambda(\mathbf{y})) = \max_{1\le k\le m} \lambda_{\max}(\Lambda_k(\mathbf{y})) \le \max_{1\le k\le m} \|\Lambda_k(\mathbf{y})\|_\infty. \tag{3.15}$$

So k_3 can be bounded from above by the largest absolute row sum of all of the Λ_k operators.

Since the values of $\|\mathbf{1}\|$ and ν are known, having bounded k_1 and k_2 from below by positive quantities and k_3 from above, C can be bounded from below by a positive, efficiently computable constant. In Section 4 we revisit this question and find closed-form bounds for the case of univariate nonnegative polynomials over an interval.

3.5. Assumptions. Throughout, we have made two fundamental assumptions. The first is that the constant one polynomial is in the interior of the WSOS cone $\Sigma = \Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$. (Naturally, in any remotely interesting situation, positive constant polynomials must belong to Σ, but not necessarily to the interior.) The second is that we have access to *some* dual certificate of the constant one polynomial in Σ. These assumptions are mild and practically unrestrictive, as we shall discuss below.

Constant one polynomial in the interior of Σ. This is a mild assumption both from a theoretical and practical perspective. In many cases (when the weights are sufficiently simple), it can be verified directly and ensured to hold a priori; it can also be verified via convex optimization. If $\mathbf{1}$ is only on the boundary, there are various ways to expand Σ to a larger WSOS cone that will contain $\mathbf{1}$ in the interior: first, as we discussed in Section 2.1.1, as long as the assumptions of Putinar's Positivstellensatz are satisfied, every positive polynomial on $S_{\mathbf{w}}$ is in the interior of $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ for every sufficiently large degree vector $\mathbf{d}$. Alternatively, $\Sigma = \Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ can be extended, with the inclusion of a single additional weight that is nonnegative on $S_{\mathbf{w}}$, to satisfy this condition without changing $\operatorname{span}(\Sigma)$ (in particular, without increasing the degrees), as stated in the following theorem.

THEOREM 3.7. *Suppose $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}} \subseteq \mathbb{R}^U$ is full dimensional. Let $\mathbf{r}$ be the coefficient vector of a polynomial which is bounded (positively) away from zero on $S_{\mathbf{w}}$. Then $\mathbf{r}$ is in the interior of $\Sigma^{(w_1,\dots,w_m,w_{m+1})}_{n,2(d_1,\dots,d_m,0)}$, for some weight polynomial w_{m+1} which is nonnegative on $S_{\mathbf{w}}$.*

Proof. Since $\Sigma^{\mathbf{w}}_{n,2\mathbf{d}} \subseteq \mathbb{R}^U$ is full dimensional, we can select linearly independent coefficient vectors $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_U \in \Sigma^{\mathbf{w}}_{n,2\mathbf{d}}$ such that $\mathbf{r} \notin \operatorname{aff}(\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_U\})$. (Here, aff denotes the affine hull.) Define $\mathbf{s}_{U+1} := M\mathbf{r} - \sum_{i=1}^U \mathbf{s}_i$, with $M \in \mathbb{R}$ large enough to guarantee that $\mathbf{s}_{U+1} \ge 0$ on $S_{\mathbf{w}}$ and that $\mathbf{s}_{U+1} \notin \operatorname{aff}(\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_U\})$. By construction, $\mathbf{r} = \frac{1}{M}\sum_{i=1}^U \mathbf{s}_i$, that is, $\mathbf{r}$ is in the interior of the simplex whose vertices are the (affinely independent) vectors $\frac{M}{U+1}\mathbf{s}_i$, $i = 1, \dots, U+1$. Each of these vertices belong to the WSOS cone $\Sigma^{\hat{\mathbf{w}}}_{n,2\hat{\mathbf{d}}}$, where $\hat{\mathbf{w}} = (w_1, \dots, w_m, w_{m+1})$ is the set of initial weights augmented with the new weight polynomial w_{m+1} whose coefficient vector is $\mathbf{s}_{U+1}$ and $\hat{\mathbf{d}} = (d_1, \dots, d_m, 0)$. (The new weight w_{m+1} is only multiplied by nonnegative constants in this new WSOS cone.) Therefore, $\mathbf{r}$ is in the interior of $\Sigma^{\hat{\mathbf{w}}}_{n,2\hat{\mathbf{d}}}$. ∎

Certificate for the constant one polynomial in Algorithm 3.1. We have also assumed that we have access to a certificate $\mathbf{x} \in \mathcal{C}(\mathbf{1})$. In the examples in this paper, we could determine the gradient certificate for $\mathbf{1}$, $\mathbf{x}_1$, in closed form. If we do not know $\mathbf{x}_1$ explicitly, then a crude approximation of $\mathbf{x}_1$ is already sufficient to initialize Algorithm 3.1. Precisely, we need to compute a certificate $\mathbf{x} \in \mathcal{C}(\mathbf{1})$ satisfying $\| - g(x) - \mathbf{1}\|_{\mathbf{x}} \le \frac{r}{r+1}$ (wherein the parameter r here is the same as that in

Algorithm 3.1). There are several approaches to compute such a vector. For instance, we can use any convex optimization algorithm to minimize the convex, self-concordant function $\overline{f}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{1}^{\mathrm{T}}\mathbf{x} + f(\mathbf{x})$ (with $f(\mathbf{x})$ defined in (1.6)); the minimizer of $\overline{f}(\mathbf{x})$ satisfies $-g(\mathbf{x}) = \mathbf{1}$. As we only need a certificate $\mathbf{x}$ with $-g(\mathbf{x})$ in the neighborhood of $\mathbf{1}$, an approximate minimizer of $\overline{f}(\mathbf{x})$ returned by a numerical optimization method is sufficient.

Alternatively, motivated by two-phase interior-point methods that start from an approximate analytic center, we can run Algorithm 3.1 "in reverse" to find a certificate of $\mathbf{1}$ as long as we have access to *any* vector in $(\Sigma^*)^\circ$. The intuition is that if some vector $\mathbf{x}$ certifies $\mathbf{t} + c\mathbf{1}$, then it also certifies $c^{-1}\mathbf{t} + \mathbf{1}$, which is approximately the same polynomial as $\mathbf{1}$ if c is sufficiently large.

Thus, starting with any vector $\mathbf{x} \in (\Sigma^*)^\circ$ and its corresponding polynomial $\mathbf{t} = -g(\mathbf{x}) \in \Sigma^\circ$, we iterate a modified certificate update step and a modified constant update step which are identical to those in Lines 3 and 4 of the algorithm, except that we replace the polynomial $\mathbf{t} - c\mathbf{1}$ with $\mathbf{t} + c\mathbf{1}$, in order to find a *large* constant c and a certificate $\mathbf{x}$ for which $\mathbf{x}$ certifies $\mathbf{t} + c\mathbf{1}$. We terminate this iterative process when $\|\mathbf{t} + c\mathbf{1} - c\mathbf{1}\|_{\mathbf{x}} = \|\mathbf{t}\|_{\mathbf{x}} < r'$ is sufficiently small to guarantee that $\|(\mathbf{t} + c\mathbf{1})/c - \mathbf{1}\|_{c\mathbf{x}}$ is small enough to ensure that $c\mathbf{x}$ certifies $\mathbf{1}$. We omit the details of this analysis that are analogous to the analysis of Algorithm 3.1; in particular, the proof of its rate of convergence can be adapted to show that in this "reversed" algorithm, c increases exponentially, and $\|(\mathbf{t} + c\mathbf{1})/c - \mathbf{1}\|_{c\mathbf{x}}$ converges to zero at a linear rate.

4. Univariate polynomials. In the univariate case, we can bound the number of iterations of Algorithm 3.1 by providing explicit bounds on the constant C, adapting the arguments from those in Section 3.4. For brevity, we only treat the even-degree case in detail.

THEOREM 4.1. *Suppose that $n = 1$ and $\deg t = 2d$. Using the Chebyshev basis to to represent all polynomials and weights $\mathbf{w}(z) \stackrel{\text{def}}{=} (1, 1 - z^2)$ (as in Example 4), Algorithm 3.1 terminates after at most $\mathcal{O}(d^2 \log \frac{\|\mathbf{t}\|d}{\varepsilon})$ iterations and requires $\mathcal{O}(d^5 \log \frac{\|\mathbf{t}\|d}{\varepsilon})$ floating point operations overall.*

Proof. We start by bounding the constant C from Theorem 3.5 as a function of all relevant parameters by bounding each of k_1, k_2 and k_3 in the formula for C.

1. $k_1 \geq 1$. Recall that $k_1 = \min\{\mathbf{1}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \|\mathbf{v}\|_\infty = 1\}$. Since nonnegative polynomials and weighted sum-of-squares polynomials coincide in the univariate case [4], every vector $\mathbf{v} \in \left(\Sigma^{\mathbf{w}}_{1,2d}\right)^*$ can be written as a conic combination of moment vectors; precisely, we write $\mathbf{v} = \sum_{i=1}^n \alpha_i \mathbf{q}(z_i)$, wherein $z_i \in [-1, 1]$ and $\alpha_i \geq 0$ for each i [12, Sec. II.2]. Then, we have
$$\mathbf{1}^{\mathrm{T}}\mathbf{v} = \mathbf{1}^{\mathrm{T}}\left(\sum_{i=1}^n \alpha_i \mathbf{q}(z_i)\right) = \sum_{i=1}^n \alpha_i \left(\mathbf{1}^{\mathrm{T}}\mathbf{q}(z_i)\right) = \sum_{i=1}^n \alpha_i.$$
If $\|\mathbf{v}\|_\infty = 1$, there exists some j such that $|\sum_{i=1}^n \alpha_i \mathbf{q}_j(z_i)| = 1$. Since for the Chebyshev basis each $\mathbf{q}(z_i) \in [-1, 1]^{2d+1}$, it follows that
$$\mathbf{v}_j = 1 = \left|\sum_{i=1}^n \alpha_i \mathbf{q}_j(z_i)\right| \leq \sum_{i=1}^n |\alpha_i \mathbf{q}_j(z_i)| \leq \sum_{i=1}^n |\alpha_i| = \sum_{i=1}^n \alpha_i,$$

since $\alpha_i \geq 0$. Thus, $\sum_{i=1}^n \alpha_i \geq 1$. It follows that

$$\mathbf{1}^{\mathrm{T}}\mathbf{v} = \sum_{i=1}^{n} \alpha_i \geq 1,$$

therefore $k_1 \geq 1$.

2. $k_2 \geq \frac{1}{2}\sqrt{3-\sqrt{5}} \approx 0.437$. Recall that $k_2 = \min\{\mathrm{tr}(\Lambda(\mathbf{v})^2) \mid \|\mathbf{v}\| = 1\}$. We have

$$\mathrm{tr}(\Lambda(\mathbf{v})^2) = \mathrm{tr}(\Lambda_1(\mathbf{v})^2) + \mathrm{tr}(\Lambda_2(\mathbf{v})^2) \geq \mathrm{tr}(\Lambda_1(\mathbf{v})^2).$$

Note that

$$2p_i(x)p_j(x) = p_{i+j}(x) + p_{|i-j|}(x) \quad \text{for every } i,j = 0,1,\ldots \tag{4.1}$$

Coupling this identity with the fact that $\Lambda_1(\mathbf{q}) = w_1\mathbf{p}\mathbf{p}^{\mathrm{T}} = \mathbf{p}\mathbf{p}^{\mathrm{T}}$ (recall that the first weight is $w_1 = 1$), we deduce that

$$\Lambda_1(\mathbf{v})_{i,j} = \frac{1}{2}v_{i+j} + \frac{1}{2}v_{|i-j|}. \tag{4.2}$$

Therefore, the zeroth row (and the zeroth column) of $\Lambda_1(\mathbf{v})$ is $(v_0, v_1, \ldots, v_d)$ and the last row (and the last column) is $(\frac{1}{2}(v_d+v_d), \frac{1}{2}(v_{d-1}+v_{d+1}), \ldots, \frac{1}{2}(v_0+v_{2d}))$, and so we have

$$\mathrm{tr}(\Lambda_1(\mathbf{v}))^2 \geq \sum_{i=0}^{d} v_i^2 + \sum_{j=0}^{d} \frac{1}{4}(v_j + v_{2d-j})^2 = \mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v},$$

where $\mathbf{M}$ is the $2d+1 \times 2d+1$ matrix (indexed from zero) given by

$$M_{i,j} = \begin{cases} \frac{5}{4} & \text{if } i = j < d \\ 2 & \text{if } i = j = d \\ \frac{1}{4} & \text{if } i + j = 2d, i \neq j, \text{ or if } i = j > d \\ 0 & \text{otherwise} \end{cases}$$

Therefore

$$\mathbf{v}^{\mathrm{T}}\mathbf{M}\mathbf{v} \geq \|\mathbf{v}\|^2 \lambda_{\min}(\mathbf{M}) = \|\mathbf{v}\|^2 \left(\frac{1}{4}(3-\sqrt{5})\right).$$

We conclude that

$$k_2 \geq \frac{1}{2}\sqrt{3-\sqrt{5}} \approx 0.437.$$

3. $k_3 \leq d+1$. Recall that

$$k_3 = \max\{\lambda_{\max}(\Lambda(\mathbf{y})) \mid \mathbf{y} \in \Sigma^*, \|\mathbf{y}\|_\infty = 1\},$$

and that based on the inequality (3.15), we need only bound the largest absolute row sum of $\Lambda_1(\mathbf{y})$ and $\Lambda_2(\mathbf{y})$, for all $\mathbf{y} \in \Sigma$ with $\|\mathbf{y}\|_\infty = 1$.
For $\Lambda_1(\mathbf{y})$, the identity (4.2) and $\|\mathbf{y}\|_\infty = 1$ yield the bound

$$\sum_{j=0}^{d} |\Lambda_1(\mathbf{y})_{i,j}| = \sum_{j=0}^{d} \left|\frac{1}{2}\mathbf{y}_{i+j} + \frac{1}{2}\mathbf{y}_{|i-j|}\right| \leq d+1.$$

For $\Lambda_2(\mathbf{y})$, observe that $1-t^2 = \frac{1}{2}(p_0(t) - p_2(t))$. Coupling this with the identity (4.1), we deduce that

$$\begin{aligned}\frac{1}{2}(p_0(t) - p_2(t))p_i(t)p_j(t) =& \frac{1}{8}\big((2p_{i+j}(t) + 2p_{|i-j|}(t) - p_{i+j+2}(t) - \\ & p_{|i+j-2|}(t) - p_{|i-j|+2}(t) - p_{||i-j|-2|}(t)\big),\end{aligned}$$

so

$$\Lambda_2(\mathbf{y})_{i,j} = \frac{1}{8}\left(2\mathbf{y}_{i+j} + 2\mathbf{y}_{|i-j|} - \mathbf{y}_{i+j+2} - \mathbf{y}_{|i+j-2|} - \mathbf{y}_{|i-j|+2} - \mathbf{y}_{||i-j|-2|}\right).$$

Then, assuming $\|\mathbf{y}\|_\infty = 1$, we obtain the bound

$$\sum_{j=0}^{d} |\Lambda_2(\mathbf{y})_{i,j}| \leq \sum_{j=0}^{d} \left|\frac{1}{8}(2+2+1+1+1+1)\right| \leq d+1.$$

Thus, $k_3 \leq \max\{\max\{\|\Lambda_1(\mathbf{y})\|_\infty, \|\Lambda_2(\mathbf{y})\|_\infty\} \mid \|\mathbf{y}\|_\infty = 1\} \leq d+1$.

Lastly, since $\nu = 2d+1$ and $\|\mathbf{1}\| = 1$, combining the above bounds on k_1, k_2 and k_3 we get

$$C = \frac{k_1 k_2}{k_3 \nu \|\mathbf{1}\|} \geq \frac{\sqrt{3-\sqrt{5}}}{2(d+1)(2d+1)}.$$

From (3.12) and the discussion directly following Theorem 3.6, the number of iterations is proportional to

$$\mathcal{O}\left(\frac{1}{\log \frac{1}{1-\rho_r C}} \log \frac{c^* - c_0}{\varepsilon}\right) \tag{4.3}$$

(recalling that ε is the inputted tolerance from Algorithm 3.1). From the series expansion $\frac{1}{\log \frac{1}{1-z}} = z^{-1} - \frac{1}{2} - \dots$ we see that the first term in (4.3), which only depends on the input through the degree d, is $\mathcal{O}(C^{-1}) = \mathcal{O}(d^2)$. To bound the numerator of second term, recall that for a coefficient vector $\mathbf{t}$ in the Chebyshev basis, $|c^*| \leq \|\mathbf{t}\|_1 \leq (2d+1)^{1/2}\|\mathbf{t}\|_2$, and from the intialization of Algorithm 3.1 we also have

$$|c_0| \leq \frac{1+r}{r}\|H(\mathbf{x}_1)^{-1/2}\|_2 \|\mathbf{t}\|_2 \overset{(3.5)}{\leq} \frac{1+r}{r}\|\mathbf{t}\|_2 \frac{2d+1}{\sqrt{2}}.$$

Thus, $|c^* - c_0|$ is of order $\mathcal{O}(\|\mathbf{t}\|_2 d)$, and the claim about the number of iterations follows.

The bottleneck of each iteration is the computation and factorization of the Hessian, which require $\mathcal{O}(d^3)$ floating point operations. Therefore, the total number of floating point operations is $\mathcal{O}(d^5 \log \frac{\|\mathbf{t}\| d}{\varepsilon})$. The $\mathcal{O}(\cdot)$ notation hides only absolute constants and the user-defined constant parameter ρ_r from Lemma 3.4. ∎

4.1. Optimization versus certification. Our approach and analysis have been motivated from the perspective of *optimization*, where the goal is to compute a certified lower bound as close to the global minimum as possible. From this perspective, the dependence of the complexity of Algorithm 3.1 on the parameter ε, or rather on $\|\mathbf{t}\|/\varepsilon$, which measures the relative error of the lower bound is arguably one of the most important questions. In this section, we interpret the above results from another perspective, that of computational algebraic geometry and symbolic computing, where the related fundamental question is often posed as follows: given a polynomial of integer coefficients (in the monomial basis) that is known to be positive on a given domain, what is the complexity of certifying its positivity? The answer, in principle, ought be a function of the number of variables, the degree, and the bit size of the coefficients. Note that ε is not a relevant parameter in this question, as the polynomial is assumed to be positive.

Polynomial-time certification of nonnegativity is challenging even in the univariate case. The methods of Guo et al. [7] and Schweighofer [43] are exponential in the degree. (The complexity of the latter was only recently established in [20].) On the other hand,

[2] presents an algorithm for computing positivity certificates of a polynomial over an interval in time polynomial in both the degree and the bit size of the coefficients. Another algorithm, published by Chevillard et al. [5] and analyzed in [20] is also polynomial in the degree, and so is the recent algorithm of Magron and Safey El Din [17]. Of the methods mentioned in this paragraph, all those that have polynomial complexity in the degree in the univariate case rely on techniques that are exclusive to univariate polynomials and cannot be generalized to the multivariate case.

Neither of these complexity results are directly comparable to our results above, as we are concerned with nonnegativity over compact sets (as opposed to the real line, as most of the papers cited above), we are working in the real number model for floating-point computation, and to avoid the inherent numerical instability of high-degree polynomials represented in the monomial basis, we consider the input to be a coefficient vector in the Chebyshev basis. But Algorithm 3.1 can be adapted to the above decision problem as follows: given a polynomial by its (integer) coefficient vector $\mathbf{t}$ in the Chebyshev basis, we can compute a positive lower bound μ on its minimum over the interval $[-1, 1]$ assuming that this minimum is positive, using a technique of Basu, Leroy, and Roy [1]. Finally, we can invoke Algorithm 3.1 with tolerance $\epsilon < \mu$. As soon as the certified lower bound turns positive, the algorithm can be terminated.

Theorem 1.2 of [1] gives a positive lower bound on the minimum of a polynomial with integer coefficients in the monomial basis over the interval $[0, 1]$. This bound is a function of only the degree d of the polynomial and the maximum bitsize τ of its coefficients. The proof can be adapted to bounding the minimum over $[-1, 1]$ without any substantial changes. The change of basis (from the Chebyshev basis to monomial) can be incorporated using the observation that a degree-d polynomial with integer coefficients of bit size at most τ in the Chebyshev basis also has integer coefficients in the monomial basis, and the bit size of the largest magnitude coefficient is no more than $2d + \tau$. Thus we have the following bound.

LEMMA 4.2 ([1]; Thm. 1.2, adapted). *Let t be a univariate polynomial of degree d taking only positive values on the interval $[-1, 1]$, and suppose that the coefficients of t in the Chebyshev basis are integers of bit size no more than τ. Then we have*

$$\min_{z\in[-1,1]} t(z) > \frac{3^{d/2}}{2^{(2d-1)(2d+\tau)}(d+1)^{2d-1/2}} =: \mu(d, \tau).$$

Hence, the number of iterations of Algorithm 3.1 with tolerance $\epsilon = \mu(d, \tau)$ is

$$\mathcal{O}\left(d^2 \log \frac{\|\mathbf{t}\| d}{\mu(d, \tau)}\right) = \mathcal{O}\left(d^4 + d^3\tau\right),$$

polynomial in the degree and linear in the bit size of the coefficients.

We underline that, unlike the algorithms mentioned above, Algorithm 3.1 does not rely on any techniques that are specific to univariate polynomials. Its complexity in the multivariate setting is a subject of future research.

5. Numerical examples. In this section, we report the results of numerical experiments with a simple Mathematica implementation of Algorithm 3.1 and investigate the quality of the best certifiable rational lower bound using the purely numerical version of the algorithm implemented using double-precision floating point arithmetic. These results are summarized in Table 5.1.

Problem instances. Problems 1–7 are standard benchmark problems from the polynomial optimization literature going back to at least [37] (but see also, e.g., [22,

#	Problem	n	d	# iters	time [s]	$c^* - c$	k
1	R. D. 3	3	2	121	0.11	$2.690981304 \times 10^{-6}$	-22
2	Schwefel	3	4	384	1.08	$5.764365051 \times 10^{-7}$	-13
3	L. V. 4	4	3	467	2.24	$2.602585946 \times 10^{-5}$	-11
4	Caprasse	4	4	283	2.45	$2.260781469 \times 10^{-6}$	-10
5	Butcher	6	3	268	8.66	$1.180076686 \times 10^{-6}$	-13
6	Magnetism	7	2	318	1.16	$9.031997478 \times 10^{-8}$	-15
7	Heart Dipole	8	4	374	127.9	$8.688025884 \times 10^{-6}$	-7
8	Wrig 5	5	2	172	0.34	$1.339771877 \times 10^{-6}$	-12
9	PP Ex 4	3	4	332	0.58	$1.397125980 \times 10^{-6}$	-12

TABLE 5.1

Summary of results of running Algorithm 3.1 on standard benchmark problems from [37, 22, 33]. n and d represent the number of variables and the degree of the polynomial to be minimized. The $c^* - c$ column lists the difference between the certified lower bound c returned by Algorithm 3.1 and the true minimum c^* of the polynomial. The k column lists the smallest integer exponent k for which $c^* - 10^k$ can be certified by the outputted certificate vector from Algorithm 3.1.

23, 30]). Problem 8 is from [22], Problem 9 is Example 4 from [33]. The problem is originally unconstrained; we added the constraint $\|\mathbf{z}\| \leq 10$ to the problem, which can be shown to be redundant. All polynomials were represented in the monomial basis throughout the computation. The minimum values of these polynomials are known.

Implementation details. The analysis of the method provides theoretically safe choices for the algorithmic parameters; in the experiments we used $r = 1/4$. Initial points were determined either from simple closed-form solutions of the nonlinear system $-g(\mathbf{x}) = \mathbf{1}$ or approximate numerical solutions, as discussed in Section 3.5. Instead of the stopping criterion used in the theoretical analysis, we ran the method with $\varepsilon = 0$, until numerical issues prevented progress, to compute the best possible bound and a corresponding rational dual certificate.

The timing results were obtained on a standard Macbook laptop computer equipped with 16GB RAM and a 2.8 GHz Intel Core i7 processor with 4 cores running macOS 10.15.7. We used Mathematica version 12.0.0.0.

The $c^* - c$ column of Table 5.1 shows the difference between the minimum and the certified lower bound returned by the algorithm.

Following the numerical computation, we also used the approach outlined in Section 3.2 to determine an approximately optimal lower bound certified by the dual certificate returned by Algorithm 3.1. Denoting the known optimal value of the polynomial by c^*, a simple linear search can identify the smallest (negative) integer exponent k for which the lower bound $c^* - 10^k$ is certified by the same dual certificate. The value of this k is reported in the last column of Table 5.1.

Summary. As expected, the numerical method yields bounds with roughly 5–8 correct decimal digits, and as predicted by the theory, the numerical dual certificate can certify stronger bounds than those returned by the Algorithm 3.1. It is remarkable, however, that the best bounds are often at least 5-6 digits more accurate than the ones from the algorithm (this is owing to the fact that the algorithm uses the simple sufficient condition in the c-update steps, rather than looking for a near-optimal bound), and that in one example, the numerical dual certificate can even certify a bound that is *indistinguishable from the minimum* with the precision used throughout the computation. (Both the relative and absolute errors of the bound are far smaller than the unit round-off in double precision.)

6. 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 (2.1), 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 coefficients. 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 3.1 can be implemented as a purely numerical method, followed by an application of the formula (2.1) 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 3.1 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 [11] and [31]; 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 3.1 as a function of every 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^*. We derived an explicit bound on the linear rate and the initial gap $c^* - c_0$ in the univariate case in Section 4; it may also be possible to derive such bounds in other important special cases, such as the cases of multivariate polynomials over simple semialgebraic sets such as the unit sphere or the unit cube.

Application to polynomials with particular structures. Dual certificates can be used in combination with other recent approaches aimed at increasing the practical efficiency of sum-of-squares optimization, such as exploiting sparsity or symmetry. For example, the (term- or correlative-) sparsity of the polynomial in [13, 15, 45] has the effect of making the Λ operator simpler, either by reducing the dimension of $\Lambda(\mathbf{x})$ or by imposing a block structure on it, reducing the overall size of the problem at hand. This in turn simplifies the computation of Λ and H^{-1} and the verification of the semidefiniteness of $\Lambda(\mathbf{x})$, for $\mathbf{x} \in \Sigma^*$. Future research could also examine how exploiting symmetries, e.g., following [40], may be used in tandem with dual certificates. Likewise, future work could extend dual certificates to the noncommutative setting.

REFERENCES

[1] S. Basu, R. Leroy, and M. Roy, *A bound on the minimum of a real positive polynomial over the standard simplex*, arXiv, (2009), https://arxiv.org/abs/0902.3304.
[2] F. Boudaoud, F. Caruso, and M.-F. Roy, *Certificates of positivity in the Bernstein basis*, Discrete & Computational Geometry, 39 (2007), pp. 639–655, https://doi.org/10.1007/s00454-007-9042-x.
[3] D. A. Brake, J. D. Hauenstein, and A. C. Liddell, *Validating the completeness of the real solution set of a system of polynomial equations*, in Proceedings of the the ISSAC '16, New York, NY, USA, 2016, ACM, p. 143?150, https://doi.org/10.1145/2930889.2930910.
[4] L. Brickman and L. Steinberg, *On nonnegative polynomials*, The American Mathematical Monthly, 69 (1962), pp. 218–221. doi:10.2307/2311058.
[5] S. Chevillard, J. Harrison, M. Joldeş, and C. Lauter, *Efficient and accurate computation of upper bounds of approximation errors*, Theoretical Computer Science, 412 (2011), pp. 1523–1543, https://doi.org/10.1016/j.tcs.2010.11.052.
[6] M. Dostert, D. de Laat, and P. Moustrou, *Exact semidefinite programming bounds for packing problems*, 2020, https://arxiv.org/abs/2001.00256.
[7] Q. Guo, M. Safey El Din, and L. Zhi, *Computing rational solutions of linear matrix inequalities*, in Proceedings of the ISSAC '13, ACM, 2013, https://doi.org/10.1145/2465506.2465949.
[8] D. Henrion and J.-B. Lasserre, *GloptiPoly: Global optimization over polynomials with Matlab and SeDuMi*, ACM Transactions on Mathematical Software, 29 (2003), pp. 165–194, https://doi.org/10.1145/779359.779363.
[9] E. Kaltofen, B. Li, Z. Yang, and L. Zhi, *Exact certification of global optimality of approximate factorizations via rationalizing sums-of-squares with floating point scalars*, in Proceedings of the 21st International Symposium on Symbolic and Algebraic Computation, ISSAC '08, New York, NY, 2008, ACM, pp. 155–164, https://doi.org/10.1145/1390768.1390792.
[10] E. L. Kaltofen, B. Li, Z. Yang, and L. Zhi, *Exact certification in global polynomial optimization via sums-of-squares of rational functions with rational coefficients*, Journal of Symbolic Computation, 47 (2012), pp. 1–15, https://doi.org/10.1016/j.jsc.2011.08.002.
[11] M. Karimi and L. Tunçel, *Domain-driven solver (DDS): a MATLAB-based software package for convex optimization problems in domain-driven form*, arXiv:1908.03075, (2019).
[12] S. Karlin and W. J. Studden, *Tchebycheff Systems, with Applications in Analysis and Statistics*, Wiley Interscience, New York, NY, 1966.
[13] M. Kojima, S. Kim, and H. Waki, *Sparsity in sums of squares of polynomials*, Mathematical Programming, 103 (2004), pp. 45–62, https://doi.org/10.1007/s10107-004-0554-3.
[14] J. B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM Journal on Optimization, 11 (2001), pp. 796–817, https://doi.org/10.1137/S1052623400366802.
[15] J. B. Lasserre, *Convergent SDP-relaxations in polynomial optimization with sparsity*, SIAM Journal on Optimization, 17 (2006), pp. 822–843, https://doi.org/10.1137/05064504x.
[16] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, in Emerging Applications of Algebraic Geometry, M. Putinar and S. Sullivant, eds., vol. 149 of IMA Volumes in Mathematics and its Applications, Springer, New York, NY, 2009, pp. 157–270, https://doi.org/10.1007/978-0-387-09686-5_7.
[17] V. Magron and M. Safey El Din, *On Exact Polya and Putinar's Representations*, in Proceedings of the ISSAC '18, ACM, New York, NY, USA, 2018, pp. 279–286, http://arxiv.org/abs/1802.10339.
[18] V. Magron and M. Safey El Din, *RealCertify: A Maple package for certifying non-negativity*, ACM Commun. Comput. Algebra, 52 (2018), p. 34?37, https://doi.org/10.1145/3282678.3282681.
[19] V. Magron and M. Safey El Din, *On exact Reznick, Hilbert-Artin and Putinar's representations*, Journal of Symbolic Computation, 107 (2021), pp. 221–250, https://doi.org/10.1016/j.jsc.2021.03.005.
[20] V. Magron, M. Safey El Din, and M. Schweighofer, *Algorithms for weighted sum of squares decomposition of non-negative univariate polynomials*, Journal of Symbolic Computation, 93 (2019), pp. 200–220, https://doi.org/10.1016/j.jsc.2018.06.005.
[21] V. Magron and J. Wang, *SONC optimization and exact nonnegativity certificates via second-order cone programming*, 2020, https://arxiv.org/abs/2012.07903.
[22] C. Muñoz and A. Narkawicz, *Formalization of Bernstein polynomials and applications to global optimization*, Journal of Automated Reasoning, 51 (2012), pp. 151–196, https://doi.org/10.1007/s10817-012-9256-3.
[23] R. Murray, V. Chandrasekaran, and A. Wierman, *Signomial and polynomial optimization via relative entropy and partial dualization*, Mathematical Programming Computation, 13

(2020), pp. 257–295, https://doi.org/10.1007/s12532-020-00193-4.
[24] S. Naldi and R. Sinn, *Conic programming: Infeasibility certificates and projective geometry*, Journal of Pure and Applied Algebra, 225 (2021), p. 106605, https://doi.org/10.1016/j.jpaa.2020.106605.
[25] Y. Nesterov, *Squared functional systems and optimization problems*, in High performance optimization, H. Frenk, K. Roos, T. Terlaky, and S. Zhang, eds., Kluwer, Dordrecht, 2000, pp. 405–440, https://doi.org/10.1007/978-1-4757-3216-0_17.
[26] Y. Nesterov and 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, https://doi.org/10.1137/1.9781611970791.
[27] J. Nie, K. Ranestad, and B. Sturmfels, *The algebraic degree of semidefinite programming*, Mathematical Programming, 122 (2010), pp. 379–405, https://doi.org/10.1007/s10107-008-0253-6.
[28] V. Y. Pan, *Structured matrices and polynomials*, Birkhäuser, Boston, MA, 2001, https://doi.org/10.1007/978-1-4612-0129-8.
[29] D. Papp and S. Yildiz, *On "A homogeneous interior-point algorithm for non-symmetric convex conic optimization"*, arXiv:1712.00492, (2017), https://arxiv.org/abs/1712.00492.
[30] D. Papp and S. Yildiz, *Sum-of-squares optimization without semidefinite programming*, SIAM Journal on Optimization, 29 (2019), pp. 822–851, https://doi.org/10.1137/17M1160124.
[31] D. Papp and S. Yildiz, *alfonso: Matlab package for nonsymmetric conic optimization*, INFORMS Journal on Computing (accepted), (2021), https://arxiv.org/abs/2101.04274.
[32] P. A. Parrilo, *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD thesis, California Institute of Technology, May 2000.
[33] H. Peyrl and P. A. Parrilo, *Computing sum of squares decompositions with rational coefficients*, Theoretical Computer Science, 409 (2008), pp. 269–281, https://doi.org/10.1016/j.tcs.2008.09.025.
[34] V. Powers, *Rational certificates of positivity on compact semialgebraic sets*, Pacific Journal of Mathematics, 251 (2011), pp. 385–391, https://doi.org/10.2140/pjm.2011.251.385.
[35] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo, *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, 2004, http://www.cds.caltech.edu/sostools.
[36] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Indiana University Mathematics Journal, 42 (1993), pp. 969–984, https://doi.org/10.1512/iumj.1993.42.42045.
[37] S. Ray and P. S. V. Nataraj, *An efficient algorithm for range computation of polynomials using the Bernstein form*, Journal of Global Optimization, 45 (2009), pp. 403–426, https://doi.org/10.1007/s10898-008-9382-y.
[38] J. Renegar, *On the computational complexity and geometry of the first-order theory of the reals. Parts I–III.*, Journal of Symbolic Computation, 13 (1992), pp. 255–352, https://doi.org/10.1016/S0747-7171(10)80003-3.
[39] 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, https://doi.org/10.1137/1.9780898718812.
[40] C. Riener, T. Theobald, L. J. Andrén, and J. B. Lasserre, *Exploiting symmetries in SDP-relaxations for polynomial optimization*, Mathematics of Operations Research, 38 (2013), pp. 122–141, https://doi.org/10.1287/moor.1120.0558.
[41] M. Safey el Din, *RAGlib (real algebraic geometry library)*. Maple package, 2007.
[42] C. Scheiderer, *Sums of squares of polynomials with rational coefficients*, Journal of the European Mathematical Society, 18 (2016), pp. 1495–1513, https://doi.org/10.4171/JEMS/620.
[43] M. Schweighofer, *Algorithmische Beweise für Nichtnegativ-und Positivstellensätze*, Master's thesis, Universität Passau, 136 (1999).
[44] A. Tarski, *A decision method for elementary algebra and geometry*, Tech. Report R-109, RAND Corporation, May 1951. http://www.rand.org/pubs/reports/2008/R109.pdf.
[45] J. Wang, V. Magron, and J.-B. Lasserre, *TSSOS: A moment-SOS hierarchy that exploits term sparsity*, SIAM Journal on Optimization, 31 (2021), pp. 30–58, https://doi.org/10.1137/19M1307871.