Journal of Symbolic Computation 121 (2024) 102254

Contents lists available at ScienceDirect

Journal of Symbolic Computation

journal homepage: www.elsevier.com/locate/jsc

Rational dual certificates for weighted sums-of-squares polynomials with boundable bit size ☆,☆☆

Maria M. Davis, Dávid Papp

North Carolina State University, Department of Mathematics, Campus Box 8205, Raleigh, NC 27695, USA

A R T I C L E   I N F O

Article history:
Received 14 November 2022
Received in revised form 25 July 2023
Accepted 25 July 2023
Available online 31 July 2023

Dedicated to Ágnes: a most generous and dedicated mentor, and a wonderful colleague. You will be missed

Keywords:
Polynomial optimization
Nonnegativity certificates
Sums-of-squares decomposition
Computational real algebraic geometry
Conic programming

A B S T R A C T

In Davis and Papp (2022), the authors introduced the concept of dual certificates of (weighted) sum-of-squares polynomials, which are vectors from the dual cone of weighted sums of squares (WSOS) polynomials that can be interpreted as nonnegativity certificates. This initial theoretical work showed that for every polynomial in the interior of a WSOS cone, there exists a rational dual certificate proving that the polynomial is WSOS. In this article, we analyze the complexity of rational dual certificates of WSOS polynomials by bounding the bit sizes of integer dual certificates as a function of parameters such as the degree and the number of variables of the polynomials, or their distance from the boundary of the cone. After providing a general bound, we explore several special cases, such as univariate polynomials nonnegative over the real line or a bounded interval, represented in different commonly used bases. We also provide an algorithm which runs in rational arithmetic and computes a rational certificate with boundable bit size for a WSOS lower bound of the input polynomial.

☆ This article is part of the volume titled "Computational Algebra and Geometry: A special issue in memory and honor of Agnes Szanto".

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

E-mail address: dpapp@ncsu.edu (D. Papp).

https://doi.org/10.1016/j.jsc.2023.102254
0747-7171/

1. Introduction

It is well known that nonnegative polynomials with rational coefficients in the interior of sum-of-squares cones are *sums of rational squares*; that is, they have sum-of-squares decompositions that are expressed entirely in terms of rational coefficients and can be verified using rational arithmetic Powers (2011). The complexity of these rational certificates of nonnegativity can be measured by the bit size of the largest magnitude coefficient in the decomposition; bounding the complexity of the "simplest" certificate and establishing its dependence on relevant parameters such as the degree, the number of variables, or the polynomial's distance from the boundary of the cone are major open questions. The corresponding algorithmic question is how efficiently these decompositions can be computed in rational arithmetic. Surprisingly, polynomial-time algorithms are difficult to design, and tight complexity bounds of known sum-of-squares decomposition algorithms are hard to come by even in the univariate case Magron et al. (2019).

The recent paper by Magron and Safey El Din (2021a) gives an in-depth review of the state-of-the-art on the complexity of deciding and certifying the nonnegativity or positivity of polynomials over basic semialgebraic sets using SOS certificates; we only recall a few highlights.

The paper Magron et al. (2019) focuses on univariate nonnegative polynomials over the real line. The most efficient algorithm they analyze returns SOS certificates of bit size $\mathcal{O}(d^4 + d^3\tau)$, wherein d is the degree of the polynomial and τ is the bit size of the largest magnitude coefficient; a slight improvement over the results in Boudaoud et al. (2008), which consider (Pólya-type) WSOS certificates of positivity for univariate polynomials over $[-1, 1]$. These algorithms cannot be generalized to the multivariate case. The multivariate case is considered first in Magron and Safey El Din (2021b), and substantially corrected in the report Magron and Safey El Din (2021a), by analyzing the bit complexity of a hybrid numerical-symbolic algorithm that recovers exact rational WSOS decompositions from an approximate (numerical) WSOS decomposition of a suitably perturbed polynomial. The main result in the corrected manuscript is that for n-variate SOS polynomials of degree d, the coefficients in the SOS certificates have bit sizes of order $\mathcal{O}\left(\tau d^{d^{\mathcal{O}(n)}}\right)$.

In this paper, we study these questions in the context of *dual certificates*. Dual certificates were introduced in Davis and Papp (2022) by the authors, motivated by (and building on) the duality theory of convex conic optimization, which has seen a number of recent applications in real algebraic geometry Katthän et al. (2021); Papp (2023). They are rational vectors from the dual cone of WSOS polynomials that by definition can be represented as a vector with far fewer components than a conventional WSOS decomposition: their dimension is independent of the number of weights, and they avoid the explicit representation of the large positive semidefinite Gram matrices that characterize conventional SOS decompositions. In Davis and Papp (2022), it was established that polynomials in the interior of a WSOS cone have rational dual certificates, also providing new elementary proofs of Powers's theorems from Powers (2011). Following up on this work, we now study the bit sizes of the components of dual certificates, as well as exact-arithmetic algorithms for the computation of rational dual certificates.

In the first part of the paper, we show that dual certificates can be rounded (trivially, componentwise) to "nearby" rational dual certificates with computable, "small" denominators. This follows from a quantitative version of a property of dual certificates that (in contrast to conventional WSOS decompositions) every WSOS polynomial has a full-dimensional cone of dual certificates. In turn, these rational certificates can be converted to integer dual certificates with boundable bit size. In Section 2, we establish our general results, which are applicable to any WSOS cone certifying nonnegativity over arbitrary basic, closed semialgebraic sets, including unbounded ones. We then provide refinements for the most frequently studied and applied special cases, including univariate polynomials over the real line and over bounded intervals in Section 3. E.g., for univariate polynomials over the real line, we show that every positive polynomial with integer coefficients of bit size at most τ (in the monomial basis) has an integer dual certificate whose components are of bit size $\mathcal{O}(d\tau + d\log(d))$—an improvement from the aforementioned result of Magron et al. (2019) and from the $n = 1$ special case of the bounds obtained in Magron and Safey El Din (2021a).

In the second part of the paper (Section 4), we provide an algorithm that takes a polynomial with rational coefficients as its input, and computes a sequence of rational lower bounds converging to the optimal sum-of-squares lower bound, along with a corresponding sequence of rational dual certificates certifying these bounds. The algorithm is based on the one proposed in Davis and Papp (2022), which is an almost entirely numerical hybrid method. Although the method in Davis and Papp (2022) is capable of computing rational lower bounds and dual certificates via floating point computations, it is limited by the precision of the floating point arithmetic, and the bit sizes of the computed certificates cannot be bounded. The new algorithm proposed in this paper, Algorithm 1, runs entirely in infinite precision (rational) arithmetic. We show that all intermediate computations can be carefully rounded to nearby rational vectors with small denominators in each step, while still maintaining the property that the algorithm converges q-linearly to the optimal weighted sums-of-squares lower bound.

1.1. Preliminaries

Here, we cover notation and background that we will use throughout the rest of this 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° and the boundary of a proper cone K by $\operatorname{bd}(K)$.

The dual of a convex cone $K \subseteq \mathbb{R}^n$ is the convex cone K^* defined as

$$K^* = \left\{\mathbf{y} \in \mathbb{R}^n \mid \forall \mathbf{x} \in K : \mathbf{x}^{\mathrm{T}}\mathbf{y} \geq 0\right\}.$$

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. Throughout, we will identify polynomials with their coefficients vectors (typeset bold) in a basis that is clear from the context (but not necessarily in the monomial basis), e.g., $\mathbf{t}$ for the polynomial $t(\cdot)$ and $\mathbf{1}$ for the constant one polynomial.

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 $\mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$ 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 $p \in \mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$ 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 $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ denote the set of WSOS polynomials in $\mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$. This is nearly identical to the notion of the *truncated quadratic module*, except that the degree of each SOS polynomial is independently selected, rather than by "truncating" to a desired total degree. In this manner, $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ is automatically a full-dimensional convex cone in the ambient space $\mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$ by definition. Additionally, under mild conditions, the cone $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ is closed and pointed; for example, it is sufficient that the set

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

is a unisolvent point set for the space $\mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$ (Papp and Yıldız, 2019, 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 $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ and its dual cone have non-empty interiors, a crucial assumption throughout the paper.

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.

It is well-known and easily seen that a polynomial s belongs to $\Sigma_{n,2d}$ if and only if

$$s(\cdot) = v_{n,d}(\cdot)^{\mathrm{T}} \mathbf{S} v_{n,d}(\cdot),$$

wherein $v_{n,d}$ denotes the vector of n-variate monomials up to degree d, and $\mathbf{S} \in \mathbb{S}_+^L$ with $L = \binom{n+d}{d}$ is the *Gram matrix* of the SOS polynomial s. This functional equality can be expressed coefficient-by-coefficient, identifying the polynomials on both sides of the equation with their coefficient vectors in a fixed basis. For example, if $n = 1$ and both polynomials are represented in the monomial basis, we obtain the classic result that $s(t) = \sum_{i=0}^{2d} s_i t^i$ is SOS if and only if there exists a matrix $(S_{jk})_{j,k=0,\ldots,d}$ such that $s_i = \sum_{(j,k):i=j+k} S_{jk}$ for each i. More generally, every SOS cone $\Sigma_{n,2d}$ is a linear image of the cone $\mathbb{S}_+^L$, and if we fix a basis for $\mathcal{V}_{n,2d}$ and a basis for $\mathcal{V}_{n,d}$, there is an explicitly computable, surjective, *linear* map Λ^* from Gram matrices (positive semidefinite matrices) to coefficient vectors of SOS polynomials. From the dual perspective, it is also well-known (and is equivalent to the above statements) that the dual cone $\Sigma_{n,2d}^*$ is a linear pre-image of $\mathbb{S}_+^L$. More precisely, there exists an injective linear map $\Lambda : \Sigma_{n,2d}^* \to \mathbb{S}_+^L$. In the context of algebraic geometry and moment theory, $\Lambda(\mathbf{y})$ is called the *truncated moment matrix* of the (pseudo-)moment vector $\mathbf{y}$, and the map Λ^* in the representation of the SOS cone is simply the adjoint of Λ. E.g., in the univariate example above, $\Lambda(\mathbf{y})$ is the Hankel matrix of the vector $\mathbf{y}$.

Although everything in the previous paragraph generalizes from SOS cones to the WSOS case, the conventional notation and terminology involving moment and localizing matrices is rather cumbersome, and is largely unnecessary for this paper. To follow the rest of the paper, it is sufficient to keep in mind that regardless of the number of variables n, the degree vector $\mathbf{d}$, the choice of weights $\mathbf{w}$, and the polynomial bases used to represent the polynomials of various degrees, *the WSOS cone $\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ is a linear image of the cone of positive semidefinite matrices of appropriate size under some surjective linear map Λ^*, and similarly, its dual $(\Sigma_{n,2\mathbf{d}}^{\mathbf{w}})^*$ is a linear pre-image of the same cone, under the adjoint map Λ.* The following Proposition makes these statements precise.

Proposition 1.1 *(Nesterov, 2000, Thm. 17.6). Fix an ordered basis $\mathbf{q} = (q_1, \ldots, q_U)$ of $\mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$ and an ordered basis $\mathbf{p}_i = (p_{i,1}, \ldots, p_{i,L_i})$ of $\mathcal{V}_{n,d_i}$ for each $i = 1, \ldots, m$. Let $\Lambda_i : \mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}} \left(\equiv \mathbb{R}^U\right) \to \mathbb{S}^{L_i}$ be the unique (injective) linear map satisfying $\Lambda_i(\mathbf{q}) = w_i \mathbf{p}_i \mathbf{p}_i^{\mathrm{T}}$, and let Λ_i^* denote its adjoint. Then $\mathbf{s} \in \Sigma_{n,2\mathbf{d}}^{\mathbf{w}}$ if and only if there exist matrices $\mathbf{S}_1 \succcurlyeq \mathbf{0}, \ldots, \mathbf{S}_m \succcurlyeq \mathbf{0}$ satisfying*

$$\mathbf{s} = \sum_{i=1}^{m} \Lambda_i^*(\mathbf{S}_i). \tag{2}$$

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

$$\left(\Sigma_{n,2\mathbf{d}}^{\mathbf{w}}\right)^* = \left\{\mathbf{x} \in \mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}} \left(\equiv \mathbb{R}^U\right) \mid \Lambda_i(\mathbf{x}) \succcurlyeq \mathbf{0} \;\; \forall i = 1, \ldots, m\right\}. \tag{3}$$

To see why Λ_i exists and is unique, consider that each entry of the matrix of functions $w_i \mathbf{p}_i \mathbf{p}_i^{\mathrm{T}}$ is a polynomial of the form $w_i p_{i,j} p_{i,k}$, which by definition belongs to the space $\mathcal{V}_{n,2\mathbf{d}}^{\mathbf{w}}$, and so it can be written uniquely as a linear combination of our chosen basis polynomials $\{q_1, \ldots, q_U\}$ of this space. Thus, for any vector $\mathbf{v} \in \mathbb{R}^U$, the (j,k)-th entry of the matrix $\Lambda_i(\mathbf{v})$ is the same linear combination of the components of $\mathbf{v}$ which would yield $w_i p_{i,j} p_{i,k}$ if it were applied to the basis polynomials $\mathbf{q}$.

The interested reader will find a number of examples of WSOS cones Σ and the Λ operators representing them in different bases in (Davis and Papp, 2022, Example 1), using the same notation as in this paper. We briefly recall only one of them:

Example 1.2. Consider univariate polynomials of degree 4, nonnegative on $[-1, 1]$. These polynomials can be written as $\sigma_1(t) + (1 - t^2)\sigma_2(t)$, where $\sigma_1 \in \Sigma_{1,4}$ and $\sigma_2 \in \Sigma_{1,2}$; that is, they are WSOS with the weights $w_1(t) = 1$ and $w_2(t) = 1 - t^2$ and degree vector $\mathbf{d} = (2, 1)$. Representing all monomials in the monomial basis, it is well-known that $\mathbf{x} = (x_0, \dots, x_4) \in (\Sigma^{\mathbf{w}}_{n,2\mathbf{d}})^*$ if and only if

$$\Lambda_1(\mathbf{x}) := \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix} \succcurlyeq 0 \text{ and } \Lambda_2(\mathbf{x}) := \begin{pmatrix} x_0 - x_2 & x_1 - x_3 \\ x_1 - x_3 & x_2 - x_4 \end{pmatrix} \succcurlyeq 0.$$

The matrix $\Lambda_1(\mathbf{x})$ is the moment matrix, while $\Lambda_2(\mathbf{x})$ is a localizing matrix for this particular domain Laurent (2009). In the notation of Proposition 1.1, we have $U = 5$, $(L_1, L_2) = (3, 2)$, and Λ_i matrices were obtained by collecting the monomial terms in the matrices

$$\begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} \begin{pmatrix} 1 & t & t^2 \end{pmatrix} = \begin{pmatrix} 1 & t & t^2 \\ t & t^2 & t^3 \\ t^2 & t^3 & t^4 \end{pmatrix} \text{ and } (1 - t^2) \begin{pmatrix} 1 \\ t \end{pmatrix} \begin{pmatrix} 1 & t \end{pmatrix} = \begin{pmatrix} 1 - t^2 & t - t^3 \\ t - t^3 & t^2 - t^4 \end{pmatrix}.$$

To further lighten the notation, throughout the paper, we will assume that the weight polynomials $\mathbf{w} = (w_1, \dots, w_m)$ and the degrees $\mathbf{d} = (d_1, \dots, d_m)$ are fixed. We will 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 $\mathcal{V}$. We will usually 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 $\|\cdot\|$. For (real) square matrices, the inner product $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product.

Additionally, we use the shorthand Λ to denote the $\mathbb{R}^U \to \mathbb{S}^{L_1} \oplus \dots \oplus \mathbb{S}^{L_m}$ linear map $\Lambda_1(\cdot) \oplus \dots \oplus \Lambda_m(\cdot)$ from Proposition 1.1. With this notation, the condition (2) can be written as $\mathbf{s} = \Lambda^*(\mathbf{S})$ for some positive semidefinite (block diagonal) matrix $\mathbf{S} \in \mathbb{S}^{L_1} \oplus \dots \oplus \mathbb{S}^{L_m}$. Similarly, Eq. (3) simplifies to

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

The interior of this cone is simply

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

1.1.2. Barrier functions and local norms in convex cones

The theory of dual certificates builds heavily on results from the theory of barrier functions in convex optimization. Here, we introduce relevant notation, and we give a brief overview of the parts of this theory that will be needed throughout the rest of the paper.

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

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

Note that by Eq. (4), f is indeed defined on its domain. The function f is twice continuously differentiable; we denote by $g(\mathbf{x})$ its gradient at $\mathbf{x}$ and by $H(\mathbf{x})$ its Hessian at $\mathbf{x}$. Since f is strictly convex on its domain, $H(\mathbf{x}) \succ \mathbf{0}$ for all $\mathbf{x} \in (\Sigma^*)^\circ$ (Boyd and Vandenberghe, 2004, Sec. 3.1.5 and 3.2.2). 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 $\|\mathbf{t}\|^*_{\mathbf{x}} = \|H(\mathbf{x})^{-1/2}\mathbf{t}\|$.

Throughout, we will invoke several useful results concerning these norms and the barrier function f in (5); these are enumerated in Lemma A.1, in the Appendix of this paper. Geometrically, the key observation is that the Hessian of this barrier function, through the associated local and dual local norms, provides computable ellipsoidal neighborhoods around each point in Σ° and $(\Sigma^*)^\circ$ that are contained in these cones, yielding "safe" bounds to round vectors in any direction without leaving the cone.

1.1.3. Dual certificates

As mentioned earlier in the introduction, our primary goal is to show the existence of a dual certificate with boundable bit size for a given WSOS polynomial. Here, we review necessary definitions and properties of dual certificates. For more extensive theory of dual certificates, see Davis and Papp (2022).

Definition 1.3. Let $\mathbf{s} \in \Sigma$, and denote the Hessian of the barrier function f of Σ^* defined in (5) by H. 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 $\mathbf{x}$.

This definition is motivated by the following theorem from Davis and Papp (2022), reproduced below for completeness. In words, the theorem provides an *explicit closed form formula* for efficiently computing a WSOS certificate for any polynomial from its coefficient vector $\mathbf{s}$ and any dual certificate $\mathbf{x}$:

Theorem 1.4 *(Davis and Papp, 2022, Thm. 2.2). Let* $\mathbf{s} \in (\Sigma^*)^\circ$ *be arbitrary. Then the matrix* $\mathbf{S} = \mathbf{S}(\mathbf{x}, \mathbf{s})$ *defined by*

$$\mathbf{S}(\mathbf{x}, \mathbf{s}) \stackrel{\text{def}}{=} \Lambda(\mathbf{x})^{-1}\Lambda\left(H(\mathbf{x})^{-1}\mathbf{s}\right)\Lambda(\mathbf{x})^{-1} \tag{6}$$

satisfies $\Lambda^*(\mathbf{S}) = \mathbf{s}$*. Moreover,* $\mathbf{x}$ *is a dual certificate for* $\mathbf{s} \in \Sigma$ *if and only if* $\mathbf{S} \succcurlyeq 0$*, which in turn is equivalent to* $H(\mathbf{x})^{-1}\mathbf{s} \in \Sigma^*$.

Note that as long as Λ maps rational vectors to rational matrices (which is the case, for instance, when polynomials are represented in commonly used bases such as the standard monomial basis or the Chebyshev basis), then $\mathbf{S}$ is a rational matrix for every rational coefficient vector $\mathbf{s}$.

It is immediate from Definition 1.3 that if $\mathbf{x}$ is a dual certificate of the polynomial $\mathbf{s}$, then every positive multiple of $\mathbf{x}$ is also a dual certificate for every positive multiple of $\mathbf{s}$. Crucially, the same is true for *small perturbations* of $\mathbf{x}$ and $\mathbf{s}$; see Proposition 1.6 below.

From Lemma A.1 (claim 5) in the Appendix, we know 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

$$H(\mathbf{x})^{-1}\mathbf{s} = -H(\mathbf{x})^{-1}g(\mathbf{x}) \stackrel{(A.6)}{=} \mathbf{x} \in (\Sigma^*)^\circ.$$

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

Definition 1.5. When $-g(\mathbf{x}) = \mathbf{s}\,(\in \Sigma^\circ)$, we say that $\mathbf{x}$ is the *gradient certificate of* $\mathbf{s}$.

Simple calculus reveals the closed-form formula for the negative gradient: $-g(\mathbf{x}) = \Lambda^*(\Lambda(\mathbf{x})^{-1})$; see also Lemma A.1 in the Appendix. However, since Λ^* is in general not injective, the nonlinear system $\mathbf{s} = \Lambda^*(\Lambda(\mathbf{x})^{-1})$ cannot be solved for $\mathbf{x}$ in closed form; only the $\mathbf{x} \to \mathbf{s}$ map is easily computable, not the converse. The same mapping $-g$ has also been recently studied by Lasserre (2022) and others Castro et al. (2021). We shall elaborate more on this connection in Section 5.

The following proposition gives two sufficient, although not necessary, conditions for $\mathbf{x} \in \Sigma^*$ to certify a polynomial $\mathbf{t}$. It also reveals that $\mathcal{C}(\mathbf{s})$ and $\mathcal{P}(\mathbf{x})$ are *full-dimensional* cones, that is, they have a non-empty interior: every sufficiently small perturbation of the gradient certificate of $\mathbf{s}$ certifies every sufficiently small perturbation of $\mathbf{s}$.

Proposition 1.6 *(Davis and Papp, 2022, Theorem 2.4 and Corollary 2.5). Suppose that* $\mathbf{x} \in \Sigma^*$ *and* $\mathbf{s} = -g(\mathbf{x})$.

1. *Then* $\mathbf{x}$ *is a dual certificate for every polynomial* $\mathbf{t}$ *satisfying* $\|\mathbf{t} - \mathbf{s}\|_{\mathbf{x}}^* \leq 1$.
2. *If* $\mathbf{y}$ *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* $\mathbf{t} = -g(\mathbf{y})$.

Two very detailed examples illustrating the concept of dual certificates, the gradient certificate, and the construction of explicit WSOS representations from dual certificates can be found in our previous work (Davis and Papp, 2022, Examples 2 and 3).

1.1.4. Bit sizes of certificate vectors

Recall that the bit size of an integer $y \in \mathbb{Z}$ is defined as $1 + \lceil \log_2(|y| + 1) \rceil$, and that the bit size of a vector $\mathbf{y} \in \mathbb{Z}^n$ can be bounded from above by n times the bit size of its the largest (in size) component. As we are interested in the orders of magnitude of bit sizes of dual certificates (e.g., whether they are linear or polynomial or exponential functions of parameters such as the degree or the number of variables of the certified polynomials), it will be convenient but equally informative to substitute this quantity with the simpler $\log(\|\mathbf{y}\|_\infty)$.

2. Rational certificates with boundable bit bize

The goal of this section is to bound the norm of an *integer* dual certificate $\bar{\mathbf{y}} \in \Sigma^*$ of a polynomial $\mathbf{t} \in \Sigma^\circ$. We consider different bounds, some of which depend only on the number of variables n, the degree d, and $\mathbf{t}$, and others that are expressed in terms of other computable or interpretable parameters introduced later in this section.

The strategy to derive these bounds is as follows. In Section 2.1, we show that dual certificates suitably close to the gradient certificate can be rounded to nearby rational dual certificates with small denominators. Then, in Section 2.2, we show that these certificates also have small norms. These two results add up to Theorem 2.9 bounding the bit size of an integer dual certificate.

2.1. Hessian bounds

Recall from Proposition 1.6, Statement 2, that if $\mathbf{x} \in \Sigma^*$ and $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} < \frac{1}{2}$, then $\mathbf{x}$ certifies $\mathbf{t} = -g(\mathbf{y})$ to be WSOS. This certificate $\mathbf{x}$ need not be a rational vector, let alone a vector with small denominator. However, Lemma 2.1 below guarantees the existence of a nearby rational certificate $\mathbf{x}_N$ for $\mathbf{t}$ with boundable denominators. In this Lemma, and throughout the rest of the section, we shall continue using g and H to denote the gradient and Hessian of the function f defined in (5).

Lemma 2.1. *Let* $\mathbf{t} \in \Sigma^\circ$ *be the coefficient vector of a polynomial whose gradient certificate is* $\mathbf{y} \in (\Sigma^*)^\circ$. *Let* $0 < r_1 < r_2$ *be arbitrary, and suppose that* $\mathbf{x} \in \Sigma^*$ *satisfies* $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \leq r_1 < 1/2$. *Let* $U = \dim(\Sigma)$, *and choose an integer denominator* $N > 0$ *to satisfy*

$$\|H(\mathbf{x})^{1/2}\| \leq \frac{2N}{\sqrt{U}} \left(\frac{r_2 - r_1}{1 + r_2} \right). \tag{7}$$

Then every $\mathbf{x}_N \in \frac{1}{N}\mathbb{Z}^U$ *with* $\|\mathbf{x}_N - \mathbf{x}\|_2 \leq \frac{\sqrt{U}}{2N}$ *satisfies* $\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}_N} \leq r_2$. *In particular, if* $r_2 < 1/2$, *then rounding* $\mathbf{x}$ *componentwise to the nearest vector in* $\frac{1}{N}\mathbb{Z}^U$ *results in a rational dual certificate of* $\mathbf{t}$.

Proof. By self-concordance (inequality (A.1) in Lemma A.1), we have

$$\begin{aligned} \|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}_N} &\overset{(A.1)}{\leq} \frac{\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}}}{1 - \|\mathbf{x}_N - \mathbf{x}\|_{\mathbf{x}}} \\ &\leq \frac{\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}}}{1 - \frac{\sqrt{U}}{2N}\|H(\mathbf{x})^{1/2}\|} \end{aligned}$$

$$\begin{aligned}&\overset{(7)}{\leq} \frac{1+r_2}{1+r_1}\|\mathbf{x}_N-\mathbf{y}\|_{\mathbf{x}}\\ &\leq \frac{1+r_2}{1+r_1}\left(\|\mathbf{x}_N-\mathbf{x}\|_{\mathbf{x}}+\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}}\right)\\ &\leq \frac{1+r_2}{1+r_1}\left(\frac{\sqrt{U}}{2N}\|H(\mathbf{x})^{1/2}\|+r_1\right)\\ &\overset{(7)}{\leq} r_2,\end{aligned}$$

which proves the first part of the claim. For the second part, if $\mathbf{x}_N$ is the result of component-wise rounding $\mathbf{x}$ to the nearest vector in $\frac{1}{N}\mathbb{Z}^U$, then $\|\mathbf{x}_N-\mathbf{x}\|_\infty \leq \frac{1}{2N}$, so $\|\mathbf{x}_N-\mathbf{x}\|_2 \leq \frac{\sqrt{U}}{2N}$. Then the first part of the claim shows that $\mathbf{x}_N$ is a certificate for $\mathbf{t}$. □

In the corollary below, we consider the particular case in which the known certificate $\mathbf{x}$ for $\mathbf{t}$ is the gradient certificate (i.e., $\mathbf{x}=\mathbf{y}$).

Corollary 2.2. *Let $\mathbf{y} \in (\Sigma^*)^\circ$ be the gradient certificate for $\mathbf{t} \in \Sigma^\circ$. Suppose N satisfies*

$$\frac{3\sqrt{U}\|H(\mathbf{y})^{1/2}\|}{2} \leq N,$$

and suppose that $\mathbf{y}_N$ satisfies $\|\mathbf{y}-\mathbf{y}_N\|_2 \leq \frac{\sqrt{U}}{2N}$. Then $\mathbf{y}_N$ certifies $\mathbf{t}$.

Proof. Substituting $r_1=0$ and $r_2=1/2$ into Lemma 2.1 yields the claim. □

The denominators in Lemma 2.1 and Corollary 2.2 depend on the norm $\|H(\mathbf{y})^{1/2}\|$, which is computable for a known $\mathbf{y}$. If $\mathbf{y}$ is not known explicitly, we can use the following upper bound for $\|H(\mathbf{y})^{1/2}\|$, which depends only on the norm of the coefficient vector $\mathbf{t}$ of the polynomial $t(\cdot)$, the set $S_\mathbf{w}$ defined in (1), and the chosen basis $\mathbf{q}$ of Σ.

Lemma 2.3. *Let $\{\mathbf{z}_1,\ldots,\mathbf{z}_s\} \subseteq S_\mathbf{w}$ be a unisolvent set for Σ, with $s \geq U = \dim(\Sigma)$. Let $\mathbf{t} \in \Sigma^\circ$, and let $\mathbf{y} \in (\Sigma^*)^\circ$ satisfy $-g(\mathbf{y})=\mathbf{t}$. Choose $\alpha_1,\ldots,\alpha_s > 0$, and define the matrix $\mathbf{M}$ as*

$$\mathbf{M} \overset{\text{def}}{=} \sum_{i=1}^{s} \alpha_i \mathbf{q}(\mathbf{z}_i)\mathbf{q}(\mathbf{z}_i)^\mathrm{T}. \tag{8}$$

Then $\mathbf{M}$ is a positive definite matrix, and

$$\|H(\mathbf{y})\|_2 \leq \operatorname{cond}(\mathbf{M})\|\mathbf{t}\|_2^2,$$

where $\operatorname{cond}(\mathbf{M}) = \frac{\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})}$ is the condition number of $\mathbf{M}$.

Proof. We begin by showing that $\mathbf{M}$ is positive definite. Let $\mathbf{v}$ be a unit-norm eigenvector of the smallest eigenvalue of $\mathbf{M}$, and consider the polynomial $v(\cdot) := \mathbf{v}^\mathrm{T}\mathbf{q}(\cdot)$. Then

$$\lambda_{\min}(\mathbf{M}) = \mathbf{v}^\mathrm{T}\Big(\sum_{i=1}^{s}\alpha_i\mathbf{q}(\mathbf{z}_i)\mathbf{q}(\mathbf{z}_i)^\mathrm{T}\Big)\mathbf{v} = \sum_{i=1}^{s}\alpha_i v(\mathbf{z}_i)^2.$$

Since $\{\mathbf{z}_1,\ldots,\mathbf{z}_s\}$ is unisolvent and $\mathbf{v} \neq \mathbf{0}$, it follows that v is not the constant 0 polynomial, moreover $v(\mathbf{z}_i) \neq 0$ for at least one i. Hence $\sum_{i=1}^{U} v(\mathbf{z}_i)^2 > 0$. As each α_i is positive, it follows that $\lambda_{\min}(\mathbf{M}) > 0$.

Now, we proceed with the proof of the claimed inequality. Recall from Section 1.1 that $\Lambda_i(\mathbf{q}(\cdot)) = w_i(\cdot)\mathbf{p}(\cdot)\mathbf{p}(\cdot)^\mathrm{T}$ for each $i=1,\ldots,m$. Thus,

$$\sum_{i=1}^{m} w_i(\cdot)\mathbf{p}_i(\cdot)^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}\mathbf{p}_i(\cdot) = \sum_{i=1}^{m} \mathrm{tr}(\Lambda_i(\mathbf{q}(\cdot))\Lambda_i(\mathbf{y})^{-1}) \stackrel{(A.3)}{=} \mathbf{q}(\cdot)^{\mathrm{T}}(-g(\mathbf{y})) = \mathbf{q}(\cdot)^{\mathrm{T}}\mathbf{t} = t(\cdot). \tag{9}$$

Thus, letting $\Lambda(\cdot)$ represent $\Lambda_1(\cdot) \oplus \cdots \oplus \Lambda_m(\cdot)$, we have for all $\mathbf{z} \in S_{\mathbf{w}}$,

$$\begin{aligned}
\mathbf{q}(\mathbf{z})^{\mathrm{T}}H(\mathbf{y})\mathbf{q}(\mathbf{z}) &\stackrel{(A.4)}{=} \mathrm{tr}(\Lambda(\mathbf{q}(\mathbf{z}))\Lambda(\mathbf{y})^{-1}\Lambda(\mathbf{q}(\mathbf{z}))\Lambda(\mathbf{y})^{-1}) \\
&= \sum_{i=1}^{m} \mathrm{tr}\left(w_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}w_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}\right) \\
&= \sum_{i=1}^{m} \mathrm{tr}\left(w_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}\mathbf{p}_i(\mathbf{z})w_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}\mathbf{p}(\mathbf{z})\right) \\
&= \sum_{i=1}^{m} \left(w_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}\mathbf{p}_i(\mathbf{z})\right)^2 \\
&\stackrel{(*)}{\leq} \sum_{i=1}^{m}\sum_{j=1}^{m} \left(w_i(\mathbf{z})\mathbf{p}_i(\mathbf{z})^{\mathrm{T}}\Lambda_i(\mathbf{y})^{-1}\mathbf{p}_i(\mathbf{z})\right)\left(w_j(\mathbf{z})\mathbf{p}_j(\mathbf{z})^{\mathrm{T}}\Lambda_j(\mathbf{y})^{-1}\mathbf{p}_j(\mathbf{z})\right) \\
&\stackrel{(9)}{=} t(\mathbf{z})^2,
\end{aligned} \tag{10}$$

with the inequality in $(*)$ due to the facts that $w_i(\mathbf{z}) \geq 0$ whenever $\mathbf{z} \in S_{\mathbf{w}}$ and that $\Lambda_i(\mathbf{y})^{-1}$ is positive definite.

Let $\{\mathbf{z}_1, \ldots, \mathbf{z}_s\}$ be the unisolvent point set in the definition of $\mathbf{M}$ above. Then we have

$$\begin{aligned}
\|H(\mathbf{y})\|_2 \leq \mathrm{tr}(H(\mathbf{y})) = \langle H(\mathbf{y}), \mathbf{I}\rangle \leq \left\langle H(\mathbf{y}), \frac{\mathbf{M}}{\lambda_{\min}(\mathbf{M})}\right\rangle \\
&\stackrel{(10)}{\leq} \frac{\sum_{i=1}^{s}\alpha_i t(\mathbf{z}_i)^2}{\lambda_{\min}(\mathbf{M})} \\
&= \frac{\mathbf{t}^{\mathrm{T}}\mathbf{M}\mathbf{t}}{\lambda_{\min}(\mathbf{M})} \\
&\leq \frac{\|\mathbf{t}\|_2^2\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})} \\
&= \mathrm{cond}(\mathbf{M})\|\mathbf{t}\|_2^2. \quad \square
\end{aligned}$$

We can use the bound in Lemma 2.3 to bound the denominators needed in Lemma 2.1 and Corollary 2.2.

Theorem 2.4. *Let* $\mathbf{y}$ *be the gradient certificate for* $\mathbf{t}$. *Let* $U = \dim(\Sigma)$, *and let* $\mathbf{M}$ *be defined as in* (8). *Let*

$$N = \left\lceil \frac{3}{2}\sqrt{U\,\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2 \right\rceil.$$

Then every $\mathbf{y}_N \in \frac{1}{N}\mathbb{Z}^U$ *with* $\|\mathbf{y}_N - \mathbf{y}\|_2 \leq \frac{\sqrt{U}}{2N}$ *is a certificate for* $\mathbf{t}$.

Proof. From Lemma 2.3, we have $\|H(\mathbf{y})^{1/2}\|_2 \leq \sqrt{\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2$. Therefore, N satisfies

$$\frac{3\sqrt{U}}{2}\|H(\mathbf{y})^{1/2}\| \stackrel{\text{Lem.2.3}}{\leq} \frac{3\sqrt{U}}{2}\sqrt{\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2 \leq N.$$

Then by Corollary 2.2, every $\mathbf{y}_N \in \frac{1}{N}\mathbb{Z}^U$ with $\|\mathbf{y}_N - \mathbf{x}\|_2 \leq \frac{\sqrt{U}}{2N}$ certifies $\mathbf{t}$. $\square$

Two remarks are in order. First, in order to obtain the smallest possible upper bound on the denominator N that works, the goal should be to minimize the $\operatorname{cond}(\mathbf{M})$ with respect to the points $\mathbf{z}_i$ in the definition of $\mathbf{M}$—a likely impossible task in general. However, any unisolvent set from $S_{\mathbf{w}}$ provides a bound, and that is generally a relatively straightforward task to find. Second, the value $\sqrt{U \operatorname{cond}(\mathbf{M})}$ is a property of the cone Σ, independent of $\mathbf{t}$, therefore this optimization (or selection of suitable points $\mathbf{z}_i$) needs to be performed only once for a given cone Σ. In Section 3, we shall use natural candidate points for interesting special cases, for which the $\operatorname{cond}(\mathbf{M})$ is computable in closed form.

2.2. Bounding certificate norms

Now, we turn our attention to bounding the norms of rational certificates for a given polynomial. The results make use of two new parameters of the cone Σ and its representation via the operator Λ. The first one is the *barrier parameter* of the barrier function f defined (in the notation of Proposition 1.1) as

$$\nu \stackrel{\text{def}}{=} \sum_{i=1}^{m} L_i$$

(see also Lemma A.1, Statement 4). The other is the constant $k_1^{\mathbf{e}}$ defined in our next Lemma. This statement is analogous to Equation (3.8) in the proof of (Davis and Papp, 2022, Theorem 3.5), but is included here for completeness.

Lemma 2.5. *Let $\mathbf{y} \in (\Sigma^*)^\circ$ be the gradient certificate for $\mathbf{t} \in \Sigma^\circ$, and let $\mathbf{e} \in \Sigma^\circ$. Let $\varepsilon > 0$, and suppose that $\mathbf{t} - \varepsilon\mathbf{e} \in \operatorname{bd}(\Sigma)$. Define*

$$k_1^{\mathbf{e}} \stackrel{\text{def}}{=} \min\left\{\mathbf{e}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \|\mathbf{v}\|_\infty = 1\right\}. \tag{11}$$

Then

$$\|\mathbf{y}\|_\infty \le \frac{\nu}{k_1^{\mathbf{e}}\varepsilon}.$$

Proof. Observe that the minimum in the definition (11) exists (as Σ^* is a closed and non-trivial cone) and $k_1^{\mathbf{e}} > 0$ because $\mathbf{e} \in \Sigma^\circ$. We now have

$$\begin{aligned}
\nu &\stackrel{(A.6)}{=} \left\langle -g\left(\frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}\right), \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}\right\rangle \\
&\stackrel{(A.5)}{=} \|\mathbf{y}\|_\infty \left\langle \mathbf{t}, \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}\right\rangle \\
&= \|\mathbf{y}\|_\infty \left(\left\langle \mathbf{t} - \varepsilon\mathbf{e}, \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}\right\rangle + \varepsilon\left\langle \mathbf{e}, \frac{\mathbf{y}}{\|\mathbf{y}\|_\infty}\right\rangle\right) \\
&\ge 0 + \|\mathbf{y}\|_\infty \varepsilon k_1^{\mathbf{e}},
\end{aligned}$$

and the claimed upper bound on $\|\mathbf{y}\|_\infty$ follows. □

Remark 2.6. Note that this bound can be computed efficiently using numerical methods for any given $\mathbf{t}$ and chosen $\mathbf{e}$. First, the ε corresponding to $\mathbf{t}$ and $\mathbf{e}$ can be approximated (or bounded below) by simple line search; the bottleneck is testing membership in Σ in each step. Second, although the minimization problem (11) is not convex, its (global) optimal value can be computed by solving $2U$ efficiently solvable convex optimization problems, since

$$k_1^{\mathbf{e}} = \min_{1\le i\le U} \Big(\inf\{\mathbf{e}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \mathbf{v}_i = 1, \|\mathbf{v}\|_\infty \le 1\}, \\ \inf\{\mathbf{e}^{\mathrm{T}}\mathbf{v} \mid \mathbf{v} \in \Sigma^*, \mathbf{v}_i = -1, \|\mathbf{v}\|_\infty \le 1\}\Big). \tag{12}$$

This reformulation will also allow us to bound $k_1^{\mathbf{e}}$ from below using convex programming duality in the next section.

Since we would like the tightest possible upper bound on $\|\mathbf{y}\|_\infty$, we seek an $\mathbf{e} \in \Sigma^\circ$ which maximizes the quantity $k_1^{\mathbf{e}}\varepsilon$. Lemma 2.7 below illustrates that, for a fixed $\mathbf{t} \in \Sigma^\circ$, choosing $\mathbf{e} = \mathbf{t}$ (with the resulting $\varepsilon = 1$) is the optimal choice. Nevertheless, the above, more general, form of Lemma 2.5 is useful when we are concerned with bounding certificates for families of polynomials (such as all univariate polynomials of degree d) in terms of interpretable parameters such as the number of variables or the degree of the polynomials. In this context, it is often more convenient to use the bound selecting any (convenient) $\mathbf{e}$ in the quantity $k_1^{\mathbf{e}}\varepsilon$, instead of $k_1^{\mathbf{t}}$.

Lemma 2.7. *Let $\mathbf{y} \in (\Sigma^*)^\circ$ be the gradient certificate for $\mathbf{t} \in \Sigma^\circ$. Then for all $\mathbf{e} \in \Sigma^\circ$ and $\varepsilon > 0$ such that $\mathbf{t} - \varepsilon\mathbf{e} \in \mathrm{bd}(\Sigma)$, we have $k_1^{\mathbf{t}} \ge k_1^{\mathbf{e}}\varepsilon$.*

Proof. Denote by E the set

$$E = \{\mathbf{e} \in \Sigma^\circ \mid \text{there exists } \varepsilon > 0 \text{ such that } \mathbf{t} - \varepsilon\mathbf{e} \in \mathrm{bd}(\Sigma)\},$$

and let $\hat{E}$ be the set

$$\hat{E} = \{\mathbf{e} \in \Sigma^\circ \mid \mathbf{t} - \mathbf{e} \in \mathrm{bd}(\Sigma)\} = \{\mathbf{t} - \mathbf{b} \mid \mathbf{b} \in \mathrm{bd}(\Sigma)\}.$$

Let V represent the set

$$V = \{\mathbf{v} \in \Sigma^* \mid \|\mathbf{v}\|_\infty = 1\},$$

and let $\varepsilon(\mathbf{t}, \mathbf{e})$ be the largest ε such that $\mathbf{t} - \varepsilon\mathbf{e}$ is on the boundary of Σ, for given $\mathbf{t}$ and $\mathbf{e}$. With this notation, we may now write

$$\begin{aligned}
\max_{\mathbf{e}\in\mathbb{R}^U} k_1^{\mathbf{e}}\varepsilon &= \max_{\mathbf{e}\in E}\left(\min_{\mathbf{v}\in V} \varepsilon(\mathbf{t},\mathbf{e})\, \mathbf{e}^{\mathrm{T}}\mathbf{v}\right) \\
&= \max_{\mathbf{e}\in \hat{E}}\left(\min_{\mathbf{v}\in V} \mathbf{e}^{\mathrm{T}}\mathbf{v}\right) \\
&= \max_{\mathbf{b}\in \mathrm{bd}(\Sigma)}\left(\min_{\mathbf{v}\in V} (\mathbf{t}-\mathbf{b})^{\mathrm{T}}\mathbf{v}\right) \\
&\le \min_{\mathbf{v}\in V}\left(\max_{\mathbf{b}\in \mathrm{bd}(\Sigma)} \mathbf{t}^{\mathrm{T}}\mathbf{v} - \mathbf{b}^{\mathrm{T}}\mathbf{v}\right) \\
&= \min_{\mathbf{v}\in V}\left(\mathbf{t}^{\mathrm{T}}\mathbf{v} + \max_{\mathbf{b}\in \mathrm{bd}(\Sigma)} (-\mathbf{b}^{\mathrm{T}}\mathbf{v})\right) \\
&\le \min_{\mathbf{v}\in V} \mathbf{t}^{\mathrm{T}}\mathbf{v} \\
&= k_1^{\mathbf{t}},
\end{aligned}$$

where the first inequality comes from the weak duality theorem of convex optimization, and the second one follows from $\mathbf{v} \in \Sigma^*$ and $\mathbf{b} \in \Sigma$. $\square$

Having bounded the norm of the gradient certificate in Lemma 2.5, we can now bound the norm of a rounded gradient certificate.

Lemma 2.8. *Let* $\mathbf{y} \in (\Sigma^*)^\circ$ *be the gradient certificate for a polynomial with coefficient vector* $\mathbf{t} \in \Sigma^\circ$. *Let* $U = \dim(\Sigma)$, $\mathbf{e} \in \Sigma^\circ$, *and* $\varepsilon > 0$, *and suppose* $\mathbf{t} - \varepsilon\mathbf{e} \in \mathrm{bd}(\Sigma)$. *Denote by* ν *the barrier parameter for* Λ *and* $\mathbf{M}$ *the matrix defined as in* (8). *Let* $N = \left\lceil \frac{3}{2}\sqrt{U\,\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2 \right\rceil$, *and suppose* $\mathbf{y}_N \in \mathbb{R}^U$ *satisfies* $\|\mathbf{y}_N - \mathbf{y}\|_\infty \le \frac{1}{2N}$. *Then* $\mathbf{y}_N$ *is a certificate for* $\mathbf{t}$ *with*

$$\|\mathbf{y}_N\|_\infty \le \frac{1}{3\sqrt{U\,\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2} + \frac{\nu}{k_1^{\mathbf{e}}\varepsilon}.$$

Proof. The fact that $\mathbf{y}_N$ is a certificate for $\mathbf{t}$ has already been shown in Lemma 2.4.

Recall from Lemma 2.5 that $\|\mathbf{y}\|_\infty \le \frac{\nu}{k_1^{\mathbf{e}}\varepsilon}$. Moreover, by our choice of $\mathbf{y}_N$ and N, we know

$$\|\mathbf{y}_N - \mathbf{y}\|_\infty \le \frac{1}{2N} \le \frac{1}{3\sqrt{U\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2}. \tag{13}$$

Therefore, by the triangle inequality, we have

$$\|\mathbf{y}_N\|_\infty \le \|\mathbf{y}_N - \mathbf{y}\|_\infty + \|\mathbf{y}\|_\infty \overset{(13),\ \text{Lem. } 2.5}{\le} \frac{1}{3\sqrt{U\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2} + \frac{\nu}{k_1^{\mathbf{e}}\varepsilon}. \quad \square$$

2.3. Bounds on integer certificate norms

Compiling the results from this section, we are now prepared to state a result bounding the largest magnitude component of an integer certificate for a polynomial $\mathbf{t} \in \Sigma^\circ$.

Theorem 2.9. *Let* $U = \dim(\Sigma)$, *and let* $\mathbf{M}$ *be defined as in* (8). *Let* $\mathbf{e} \in \Sigma^\circ$, *and let* $k_1^{\mathbf{e}}$ *be defined as in* (11). *Let* ν *be the barrier parameter of* Λ, *and let* $\mathbf{t} \in \Sigma^\circ$ *with* $\mathbf{t} - \varepsilon\mathbf{e}$ *on the boundary of* Σ. *Then there exists an integer certificate* $\bar{\mathbf{y}}$ *for* $\mathbf{t}$ *with*

$$\|\bar{\mathbf{y}}\|_\infty \le \frac{1}{2} + \left\lceil \frac{3}{2}\sqrt{U\,\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2 \right\rceil \left(\frac{\nu}{k_1^{\mathbf{e}}\varepsilon}\right).$$

Proof. Recall from Sec. 1.1.3 that any positive multiple of $\mathbf{y}_N$ is also a certificate for $\mathbf{t}$. Hence, the integer vector $\bar{\mathbf{y}} \overset{\text{def}}{=} N\mathbf{y}_N$ is a certificate for $\mathbf{t}$. Using Lemma 2.8 and Theorem 2.4, we have

$$\begin{aligned}\|\bar{\mathbf{y}}\|_\infty = N\|\mathbf{y}_N\|_\infty &\le \left\lceil \frac{3}{2}\sqrt{U\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2 \right\rceil \left(\frac{1}{3\sqrt{U\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2} + \frac{\nu}{k_1^{\mathbf{e}}\varepsilon}\right) \\ &= \frac{1}{2} + \left\lceil \frac{3}{2}\sqrt{U\,\mathrm{cond}(\mathbf{M})}\|\mathbf{t}\|_2 \right\rceil \left(\frac{\nu}{k_1^{\mathbf{e}}\varepsilon}\right). \quad \square\end{aligned}$$

3. Bit size bounds for rational certificates in particular bases

In this section we refine the result of Theorem 2.9 in a few well-studied and computationally relevant special cases such as the cones of univariate polynomials nonnegative on the real line or on a bounded interval. These results complement existing ones on the bit sizes of conventional sum-of-squares certificates of nonnegative univariate polynomials, such as those summarized in the Introduction. We emphasize that our approach yields an efficiently computable bound for a variety of WSOS cones even in the multivariate case. We consider one of these in Section 3.3. We also consider different choices of bases, which are relevant for practical computation, specifically polynomials represented in the Chebyshev basis and polynomials represented as interpolants.

The results presented in this section come in two flavors, motivated by two different mindsets and two different families of applications in which nonnegativity certification is important. From the perspective of *optimization*, the fundamental task is to certify a bound as close to the global minimum of a polynomial (on $S_{\mathbf{w}}$) as possible, and therefore one is inherently concerned with certifying polynomials close to the boundary of Σ. In this setting, ε (specifically with the choice of $\mathbf{e} = \mathbf{1}$) is arguably the most important parameter in Theorem 2.9, even for a fixed number of unknowns n and fixed degree d, and one of the most pertinent questions is the dependence of the bit size of the certificates on ε, as ε tends to 0. The common simplifying assumption that the coefficient vector $\mathbf{t}$ is integer is not particularly convenient or necessary; $\mathbf{t}$ can be any rational vector. On the other hand, from the perspective of *theoretical computer science* and applications such as automated theorem proving, the fundamental task is to certify that a given polynomial is nonnegative on a given $S_{\mathbf{w}}$. Although the dependence of the complexity of the certificate on ε (with any $\mathbf{e}$) is still informative, the primary concern is the asymptotic complexity of the certificate as the input size increases. It is convenient to assume that $\mathbf{t}$ is an integer vector, and the relevant question is the bit size of the certificates as a function of n, d, and $\tau = \log(\|\mathbf{t}\|_\infty)$. Therefore, we consider both bounds involving ε (for general $\mathbf{t}$) and bounds that are strictly functions of (n, d, τ), assuming that $\mathbf{t}$ is an integer vector.

In each case, we will use the bound given in Theorem 2.9 as a starting point. Since U and ν can be expressed in terms of d and n, a result depending only on d, n, and τ, and possibly ε, requires a lower bound on $k_1^{\mathbf{e}}$ or $k_1^{\mathbf{e}}\varepsilon$ (for some $\mathbf{e}$) and an upper bound on $\operatorname{cond}(\mathbf{M})$, for some $\mathbf{M}$ in the form given in (8), in each case.

3.1. Univariate polynomials over the real line

We first consider the most well-studied special case, univariate polynomials nonnegative over the real line, which coincide with univariate SOS polynomials, represented in the monomial basis. In this case, $\Lambda(\mathbf{x})$ is the standard moment matrix (positive definite Hankel matrix) corresponding to the pseudo-moment vector $\mathbf{x} \in \Sigma^\circ$.

Theorem 3.1. *Suppose that $\Sigma = \Sigma_{1,2d}$ (univariate sum-of-squares polynomials of degree $2d$) and that we represent all polynomials in the monomial basis—that is, in the notation of Proposition 1.1, the ordered bases $\mathbf{p}$ and $\mathbf{q}$ are the standard monomial bases of degree d and $2d$, respectively. Then the following hold:*

1. *For $\mathbf{e} = \Lambda^*(\mathbf{I})$, the coefficient vector of the sum of monomial squares polynomial $z \mapsto 1 + z^2 + \cdots + z^{2d}$, we have $k_1^{\mathbf{e}} \geq 1$.*
2. *There exists a matrix $\mathbf{M}$ of the form given in (8) such that $\operatorname{cond}(\mathbf{M}) \leq 3.21^{2d+1}/2$.*
3. *For every $\mathbf{t} \in \Sigma^\circ$ with $\mathbf{t} - \varepsilon\mathbf{e} \in \operatorname{bd}(\Sigma)$, there exists an integer dual certificate $\bar{\mathbf{y}} \in \Sigma^\circ \cap \mathbb{Z}^{2d+1}$ for $\mathbf{t}$ with*

$$\|\bar{\mathbf{y}}\|_\infty \leq \frac{1}{2} + 4d^{3/2}3.21^{d+1/2}\frac{\lceil\|\mathbf{t}\|_2\rceil}{\varepsilon},$$

whose largest component has bit size bounded as

$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}\left(d + \log(\|\mathbf{t}\|_2) + \log(1/\varepsilon)\right).$$

Proof. 1. First, observe that since $\mathbf{I}$ is positive definite, $\mathbf{e} \in \Sigma^\circ$. Now, we have

$$\mathbf{e}^{\mathrm{T}}\mathbf{v} = \langle \Lambda^*(\mathbf{I}), \mathbf{v}\rangle = \langle \mathbf{I}, \Lambda(\mathbf{v})\rangle = \operatorname{tr}(\Lambda(\mathbf{v})),$$

Combining this equation with (11), we get

$$k_1^{\mathbf{e}} = \min\{\operatorname{tr}(\Lambda(\mathbf{v})) \mid \Lambda(\mathbf{v}) \succcurlyeq \mathbf{0}, \|\mathbf{v}\|_\infty = 1\}. \tag{14}$$

It is well known that in this setting (nonnegative univariate polynomials represented in the monomial basis), $\Lambda(\cdot)$ maps a vector $\mathbf{v}$ to its corresponding Hankel matrix, whose diagonal consists of components of $\mathbf{v}$. (See, e.g., (Davis and Papp, 2022, Example 1).) Because the matrix $\Lambda(\mathbf{v})$ is positive semidefinite, its largest element (in absolute value) is a nonnegative component of $\mathbf{v}$ on the diagonal of $\Lambda(\mathbf{v})$. Hence, in (14) we have $\operatorname{tr}(\Lambda(\mathbf{v})) \geq \|\mathbf{v}\|_\infty = 1$. Therefore $k_1^{\mathbf{e}} \geq 1$.

2. When $\mathbf{q}(z) = (1, z, \ldots, z^{2d})$, the matrix $\mathbf{M}$ defined in (8) is a positive semidefinite Hankel matrix of order $(2d+1)$. It is well known that a partial converse also holds, and every positive definite Hankel matrix of order $(2d+1)$ can be written in this form, because positive definite matrices can be identified with truncated moment matrices of Borel measures μ supported on the real line (see, e.g., (Blekherman et al., 2013, Theorem 3.146)), and via Gaussian quadrature (Krylov, 2005, Chap. 7, Thm. 3) this measure μ can also be chosen to be a discrete one supported on at most $2d+1$ points.
   Bounds on the condition numbers of positive definite Hankel matrices have been studied by many authors. Our Statement 2 follows from a bound of Beckermann (Beckermann, 2000, Thm. 3.6), which states that the positive definite $n \times n$ Hankel matrix of the lowest condition number has condition number at most $3.21^n/2$.
3. From Statement 1, we have $k_1^{\mathbf{e}} \geq 1$. From Statement 2, we know there exists a matrix $\mathbf{M}$ in the form of Eq. (8) with $\operatorname{cond}(\mathbf{M}) \leq 3.21^{2d+1}/2$ Moreover, we have $\nu = d+1$ and $U = 2d+1$. Substituting these values into the formula from Theorem 2.9 gives the result. □

Corollary 3.2. *(To Theorem 3.1) Suppose that $\Sigma = \Sigma_{1,2d}$ and that, in the notation of Proposition 1.1, the ordered bases $\mathbf{p}$ and $\mathbf{q}$ are the standard monomial bases of degree d and 2d, respectively. Let $\mathbf{t} \in \Sigma^\circ$, and assume that $\mathbf{t}$ is an integer vector with $\tau = \log(\|\mathbf{t}\|_\infty)$. Then there exists an integer dual certificate $\bar{\mathbf{y}} \in (\Sigma^*)^\circ \cap \mathbb{Z}^{2d+1}$ for $\mathbf{t}$ whose components have bit size*

$$\log\left(\|\bar{\mathbf{y}}\|_\infty\right) \approx \mathcal{O}\left(\tau d + d\log d\right).$$

Proof. The claim is a consequence of Theorem 3.1 with a suitable upper bound on $1/\varepsilon$ as a function of τ and d. (Here, as before, $\varepsilon > 0$ with $\mathbf{t} - \varepsilon\mathbf{e} \in \operatorname{bd}(\Sigma)$, with $\mathbf{e} = \Lambda^*(\mathbf{I})$ is the coefficient vector of the polynomial $z \mapsto 1 + z^2 + \cdots + z^{2d}$.)

For the bound, notice that the largest ε for which $t(x) - \varepsilon e(x) \geq 0$ for every x has the property that the corresponding univariate polynomial $x \mapsto t(x) - \varepsilon e(x)$ has a multiple root, since its global minimum is zero. Therefore, the discriminant of this polynomial (with respect to x, treating ε as a parameter) must vanish at the optimal ε. The discriminant is a univariate polynomial of ε, with integer coefficients whose bit sizes can be bounded from above using the (more general) bounds on the bit sizes of subresultant polynomials (Basu et al., 2006, Proposition 8.46). Note that since $\mathbf{e} = (1, 0, 1, \ldots, 0, 1)$, if τ is an upper bound on the bit size of the coefficients of t in the monomial basis, then the coefficients of the bivariate polynomial $t(x) - \varepsilon e(x)$ are of bit size at most $\tau + 1$. Furthermore, treating the polynomial $t(x) - \varepsilon e(x)$ as a polynomial of x whose coefficients are polynomials of ε, our polynomial is of degree $2d$, with coefficients of degree 1. Thus, (Basu et al., 2006, Proposition 8.46) yields that the coefficients of the discriminant have bit sizes no larger than $\hat{\tau} := (4d-1)\big((\tau+1) + \log(2d) + \log(4d)\big) + \log(4d-1) = \mathcal{O}(\tau d + d\log(d))$.

We can now bound $1/\varepsilon$ from above using Cauchy's bound, which yields that the bit size of $1/\varepsilon$ is at most $1 + 2\hat{\tau} = \mathcal{O}(\tau d + d\log(d))$. Substituting this bound and $\log(\|\mathbf{t}\|_2) \leq \log(\|\mathbf{t}\|_\infty\sqrt{2d+1}) = \mathcal{O}(\tau + \log(d))$ into the bound from Theorem 3.1 completes the proof. □

These results may be compared to the bit sizes of the certificates obtained using the two algorithms analyzed in Magron et al. (2019). The first one finds certificates (explicit SOS decompositions) of bit size $\mathcal{O}(\tau \left(\frac{d}{2}\right)^{\frac{3d}{2}})$—linear in τ, but exponential in the degree. The second one outputs a decomposition with coefficients of bit size $\mathcal{O}(\tau d^2 + d^3)$—also polynomial in the degree, a comparable result to ours above.

3.2. Univariate polynomials over an interval

We now consider polynomials nonnegative over an interval, which for simplicity we assume to be $[-1, 1]$; all results in this section can be scaled appropriately to apply to any bounded interval. In this case, because the domain is bounded, the constant one polynomial $\mathbf{1}$ belongs to Σ°, thus we may use $\mathbf{e} = \mathbf{1}$ instead of $\mathbf{e} = \Lambda^*(\mathbf{I})$ as we do in the previous section. This leads to simpler and more interpretable results: $\varepsilon(\mathbf{t}, \mathbf{e})$ in this case is simply the minimum value of t over $[-1, 1]$, which in turn can be bounded tightly using elementary techniques, without quantifier elimination. In the context of polynomial optimization, this reveals the rate at which the bit sizes of the certificates of lower bounds grow as the lower bounds approach the minimum value.

In this section, we also consider polynomial bases that are more commonly used in practical computation with high-degree polynomials than the monomial basis, namely Chebyshev polynomials (of the first kind) (Trefethen, 2013, Sec. 3) and interpolants (Trefethen, 2013, Sec. 2).

The representations of even and odd degree polynomials over $[-1, 1]$ vary slightly; we briefly recall the details for completeness. In the notation of Proposition 1.1, for polynomials of degree $2d$, we use the weight polynomials $\mathbf{w}(z) = \{1, 1 - z^2\}$ to represent $S_\mathbf{w} = [-1, 1]$, and regardless of the choice of bases $\mathbf{p}_1, \mathbf{p}_2$ and $\mathbf{q}$, we have $m = 2$, $U \stackrel{\text{def}}{=} \dim(\Sigma^{\mathbf{w}}_{1,2d}) = 2d + 1$ and $\nu \stackrel{\text{def}}{=} \sum_{i=1}^m \dim(\mathbf{p}_i) = 2d + 1$. For polynomials of degree $2d + 1$, we use the weight polynomials $\mathbf{w}(z) = \{1 - z, 1 + z\}$, and we have $m = 2$, $U \stackrel{\text{def}}{=} \dim(\Sigma^{\mathbf{w}}_{1,2d+1}) = 2d + 2$ and $\nu = \sum_{i=1}^m \dim(\mathbf{p}_i) = 2d + 2$.

3.2.1. Chebyshev polynomial basis

In this setting, we let $\mathbf{q}$ be the basis of Chebyshev polynomials up to degree $2d$ or $2d + 1$ (for even- or odd-degree polynomials, respectively). The $\mathbf{p}_1$ and $\mathbf{p}_2$ bases can be any bases of univariate polynomials of the appropriate degree.

Theorem 3.3. *For univariate polynomials nonnegative over* $[-1, 1]$, *represented in the Chebyshev basis, the following hold:*

1. *The constant* $k_1^{\mathbf{1}}$ *is bounded below by 1.*
2. *There exists a matrix* $\mathbf{M}$ *of the form given in* (8) *such that* $\operatorname{cond}(\mathbf{M}) \le 4$.
3. *For every* $\mathbf{t} \in \Sigma^\circ$ *of degree* $2d$ *or* $2d+1$, *with* $\mathbf{t} - \varepsilon\mathbf{1}$ *on the boundary of* Σ, *there exists an integer certificate* $\bar{\mathbf{y}} \in (\Sigma^*)^\circ$ *for* $\mathbf{t}$ *with*
$$\|\bar{\mathbf{y}}\|_\infty \le \frac{1}{2} + \frac{2d+2}{\varepsilon}\left\lceil 3\sqrt{2d+2}\|\mathbf{t}\|_2 \right\rceil.$$
and the bit size of the largest component of $\bar{\mathbf{y}}$ *is bounded as*
$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}\left(\log(d) + \log(\|\mathbf{t}\|_2) + \log(1/\varepsilon)\right).$$

Proof. For brevity, we include the details for even degree polynomials only. We will index all vectors, matrices, and point sets from 0 to $2d$.

1. This result comes from Theorem 4.1 in Davis and Papp (2022), wherein k_1 denotes our constant $k_1^{\mathbf{1}}$.
2. Consider univariate polynomials of degree $2d$, and consider the points $\mathbf{z} = \{z_0, \ldots, z_{2d}\}$, with $z_l = \cos\left(\frac{\pi l}{2d}\right)$, so $\mathbf{z}$ is the set of extrema of $q_{2d+1}(\cdot)$ (also known as the Chebyshev nodes of the second kind). Recall that $\mathbf{M}$ is defined by
$$\mathbf{M} = \sum_{l=0}^{s} \alpha_i \mathbf{q}(z_l)\mathbf{q}(z_l)^{\mathrm{T}}.$$
for $s < \infty$ and for some real numbers $\alpha_0, \ldots, \alpha_s$. Here, we will set $s = 2d$ and $\alpha_0 = \cdots = \alpha_{2d} = 1$. From Mason and Handscomb (2003), Chapter 4.6.1, equations 4.45-4.46c, we know that for all $i = 0, \ldots, 2d$ and $j = 0, \ldots, 2d$, we have
$$-\frac{1}{2}q_i(z_0)q_j(z_0) - \frac{1}{2}q_i(z_{2d})q_j(z_{2d}) + \sum_{l=0}^{2d} q_i(z_l)q_j(z_l) = \begin{cases} 0 \text{ when } i \neq j, \\ d \text{ when } i = j, i \neq 0 \text{ nor } 2d \\ 2d \text{ when } i = j = 0 \text{ or } 2d \end{cases}$$

Therefore, noting that $q_i(z_0)q_j(z_0) = (-1)^{i+j}$ and $q_i(z_{2d})q_j(z_{2d}) = 1$ for each pair (i, j), we have

$$(\mathbf{M})_{ij} \stackrel{\text{def}}{=} \left(\sum_{l=0}^{2d} \mathbf{q}(z_l)\mathbf{q}(z_l)^{\mathrm{T}}\right)_{ij} = \begin{cases} 1 \text{ when } i \neq j \text{ and } i \equiv j \mod 2 \\ 0 \text{ when } i \neq j \text{ and } i \not\equiv j \mod 2 \\ d+1 \text{ when } i = j, i \neq 0 \text{ nor } 2d \\ 2d+1 \text{ when } i = j, i = 0 \text{ or } 2d. \end{cases}$$

To bound the condition number of $\mathbf{M}$ from above, it suffices to give a lower bound for $\lambda_{\min}(\mathbf{M})$ and an upper bound for $\lambda_{\max}(\mathbf{M})$ (since $\mathbf{M}$ is positive definite). First, we will exhibit a lower bound for $\lambda_{\min}(\mathbf{M})$. For any $\mathbf{x} \in \mathbb{R}^{2d+1}$, with $\|\mathbf{x}\|_2 = 1$, we have

$$\begin{aligned}\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{x} &= \left(\sum_{i=0}^{d-1} x_{2i+1}\right)^2 + \left(\sum_{i=0}^{d} x_{2i}\right)^2 + d\left(2x_0^2 + x_1^2 + x_2^2 + \cdots + x_{2d-1}^2 + 2x_{2d}^2\right) \\ &\geq d.\end{aligned}$$

Therefore, $\lambda_{\min}(\mathbf{M}) \geq d$.

Recall that $\lambda_{\max}(\mathbf{M})$ can be bounded by the largest absolute row sum of $\mathbf{M}$. The largest absolute row sum of $(\mathbf{M})$ is $(2d+1) + d = 3d+1$. Hence $\lambda_{\max}(\mathbf{M}) \leq 3d+1$. It follows that $\operatorname{cond}(\mathbf{M}) \leq \frac{3d+1}{d} \leq 4$.

3. From Statement 1, we have $k_1^{\mathbf{1}} \geq 1$, and from Statement 2, we know there exists a matrix $\mathbf{M}$ in the form of (8) with $\operatorname{cond}(\mathbf{M}) \leq 4$. Moreover, we have $\nu \leq 2d+2$ and $U \leq 2d+2$. Substituting these values into the formula given in Theorem 2.9 gives the result. □

Now we bound the minimum ε of a positive univariate polynomial on the interval $[-1, 1]$ with integer coefficients, so that we can give an ε-free result of the Theorem above.

Lemma 3.4 *(Adapted from (Basu et al., 2009, Thm. 1.2)). 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 monomial basis are integers of bit size no more than τ. Then we have*

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

Lemma 3.4 assumes that t is represented in the monomal basis. For our next result, the change of basis (from the Chebyshev basis to monomial) can be incorporated using the observation that a polynomial of degree d 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 in the monomial basis is no more than $2d + \tau$. We are now ready to state our ε-free version of Theorem 3.3.

Corollary 3.5. *(To Theorem 3.3) Using the same notation as in Theorem 3.3, assume that $\mathbf{t} \in \Sigma^\circ$ is the coefficient vector in the Chebyshev basis of a polynomial of degree at most $2d+1$, and assume that the components of $\mathbf{t}$ are integers with bit sizes at most τ. Then there exists an integer certificate $\bar{\mathbf{y}} \in (\Sigma^*)^\circ \cap \mathbb{Z}^U$ for $\mathbf{t}$ with*

$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}\left(d\tau + d^2\right).$$

Proof. The result comes from substituting the bound on ε from Lemma 3.4 and the previous paragraph into the bound from Theorem 3.3. □

3.2.2. Univariate monomial basis

Here, we let $\mathbf{q}$ represent the univariate monomial basis up to degree $2d$ or $2d+1$, for even- or odd-degree polynomials, respectively. The bases $\mathbf{p}_1$ and $\mathbf{p}_2$ can be any univariate polynomial bases.

Theorem 3.6. *For univariate polynomials nonnegative over* $[-1,1]$*, represented in the monomial basis, the following hold:*

1. *The constant* $k_1^{\mathbf{1}}$ *is bounded below by 1.*
2. *There exists a matrix* $\mathbf{M}$ *of the form given in (8) such that* $\operatorname{cond}(\mathbf{M}) \approx \mathcal{O}\left((1+\sqrt{2})^{4U}/\sqrt{U}\right)$*, wherein* $U = \dim(\Sigma)$.
3. *For every* $\mathbf{t} \in \Sigma^\circ$ *of degree* $2d$ *or* $2d+1$*, with* $\mathbf{t} - \varepsilon\mathbf{1}$ *on the boundary of* Σ*, there exists an integer certificate* $\bar{\mathbf{y}} \in (\Sigma^*)^\circ$ *for* $\mathbf{t}$ *with*

$$\|\bar{\mathbf{y}}\|_\infty \approx \mathcal{O}\left(d^{5/4}(1+\sqrt{2})^{4d+4}\frac{\|\mathbf{t}\|_2}{\varepsilon}\right),$$

and the bit size of the largest component of $\bar{\mathbf{y}}$ *is bounded as*

$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}\left(d + \log(\|\mathbf{t}\|_2) + \log(1/\varepsilon)\right).$$

Proof. 1. Using the observation that monomial basis polynomials also take values from $[-1,1]$ on the interval $[-1,1]$, this result comes from a slight adaptation of Theorem 4.1 in Davis and Papp (2022), using the monomial basis instead of the Chebyshev.

2. We may choose $\mathbf{M}$ to be the $U \times U$ Hilbert matrix, which is the (truncated) moment matrix of the uniform measure on $[0,1]$. Hence (similarly to the argument in Theorem 3.1) it is also the moment matrix of a finitely supported measure on $[0,1]$, and therefore it can be written in the form given in (8). The Hilbert matrix is well-known to have a condition number which grows as $\mathcal{O}\left((1+\sqrt{2})^{4U}/\sqrt{U}\right)$; see, for example, (Hardy et al., 1934, Thm. 294) and (Wilf, 1970, Eq. 3.35) for upper and lower bounds on the maximum and minimum eigenvalues of the Hilbert matrix, respectively.
3. Substituting the results from Statements 1 and 2 as well as the bounds $\nu \le 2d+2$ and $U \le 2d+2$ into the formula from Theorem 2.9 yields the result. □

As a corollary, we have the following:

Corollary 3.7. *Using the same notation as in Theorem 3.6, assume that* $\mathbf{t} \in \Sigma^\circ$ *is the coefficient vector in the monomial basis of a polynomial of degree at most* $2d+2$*, and assume that the components of* $\mathbf{t}$ *are integers of bit size at most* τ*. Then there exists an integer certificate* $\bar{\mathbf{y}} \in (\Sigma^*)^\circ \cap \mathbb{Z}^U$ *for* $\mathbf{t}$ *with*

$$\log\left(\|\bar{\mathbf{y}}\|_\infty\right) \approx \mathcal{O}\left(d\tau + d\log(d)\right).$$

Proof. This comes from substituting the bound on ε from Lemma 3.4 into the bound from Theorem 3.6. □

3.2.3. *Univariate Lagrange interpolant basis*

We now turn our attention to another common basis choice for polynomials over an interval: Lagrange interpolation polynomials. We use the same notation introduced at the beginning of Section 3.2.1 to describe Σ, the weight polynomials $\mathbf{w}$, and their respective variations for even- and odd-degree polynomials, except here we let $\mathbf{q}$ (using the notation of Proposition 1.1) be a *Lagrange interpolation polynomial basis.* Precisely, let $\{z_1,\dots,z_U\}$ be a unisolvent point set in $S_{\mathbf{w}}$. (As before, $U = \dim(\Sigma)$.) Define the Lagrange interpolation polynomial q_i ($i = 1,\dots,U$) to be the unique polynomial such that $q_i(z_i) = 1$ for each i and $q_i(z_j) = 0$ when $i \neq j$. Then, we define the ordered Lagrange interpolation polynomial basis $\mathbf{q}$ using these polynomials as $\mathbf{q} = (q_1,\dots,q_U)$.

The primary change from the Chebyshev and monomial bases is that the bit sizes of the certificates now depend on the choice of interpolation points z_i in a quantifiable manner.

Theorem 3.8. *Suppose that, as detailed in the previous paragraph, we represent polynomials in the Lagrange interpolant basis corresponding to the interpolation points* $\{z_1, \dots, z_U\}$. *Then the following hold:*

1. *Letting* $\mu = \max_{i=1,\dots,U} (\max_{-1\le z\le 1} |q_i(z)|)$, *we have* $k_1^{\mathbf{1}} \ge \frac{1}{\mu}$.
2. *The matrix* $\mathbf{M}$ *given in* (8) *can be chosen to be the identity matrix, with* $\operatorname{cond}(\mathbf{M}) = 1$.
3. *For every* $\mathbf{t} \in \Sigma^\circ$ *of degree* $2d$ *or* $2d+1$ *with* $\mathbf{t} - \varepsilon\mathbf{e} \in \operatorname{bd}(\Sigma)$, *there exists an integer dual certificate* $\bar{\mathbf{y}} \in (\Sigma^*)^\circ$ *for* $\mathbf{t}$ *with*

$$\|\bar{\mathbf{y}}\|_\infty \le \frac{1}{2} + \left\lceil \frac{3}{2}\sqrt{2d+2}\|\mathbf{t}\|_2 \right\rceil \left(\frac{(2d+2)\mu}{\varepsilon} \right),$$

whose largest component has bit size bounded as

$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}\left(\log(d) + \log(\|\mathbf{t}\|_2) + \log(\mu) + \log(1/\varepsilon)\right). \tag{15}$$

Proof. 1. Recall that the optimization problem to compute $k_1^{\mathbf{1}}$ can be solved by finding the minimum of the $2U$ convex optimization problems in (12) (with two such problems for each $i = 1, \dots, U$). Fix i with $1 \le i \le U$. The two optimization problems for this i have as their respective duals

$$\sup\{\pm x_i + |x_i| - \|\mathbf{x}\|_1 \mid \mathbf{1} - \mathbf{x} \in \Sigma, \mathbf{x} \in \mathbb{R}^U\}.$$

In the univariate setting, a polynomial belongs to Σ if and only if it is nonnegative. Hence, each $\mathbf{x} = (0, \dots, \pm\frac{1}{\mu}, \dots, 0)$, with $\pm\frac{1}{\mu}$ in the ith coordinate, is a feasible solution to the dual problem with objective value $\frac{1}{\mu}$. Thus, $\frac{1}{\mu}$ is a lower bound for the infima of each of the $2U$ problems in (12), therefore $k_1^{\mathbf{1}} \ge \frac{1}{\mu}$.

2. By the definition of the Lagrange basis polynomials,

$$\mathbf{M} = \sum_{i=1}^{U} \mathbf{q}(\mathbf{z}_i)\mathbf{q}(\mathbf{z}_i)^{\mathrm{T}} = \mathbf{I},$$

the identity matrix, whose condition number is 1.

3. From Statement 1, we have $k_1^{\mathbf{1}} \ge \frac{1}{\mu}$, and from Statement 2, we know there exists a matrix $\mathbf{M}$ in the form of Eq. (8) with $\operatorname{cond}(\mathbf{M}) \le 1$. Moreover, we have $\nu \le 2d+2$ and $U \le 2d+2$. Substituting these values into the formula given in Theorem 2.9 gives the result. □

Remark 3.9. The parameter μ from Statement 1 of Theorem 3.8 is closely related to the *Lebesgue constant* $\mathcal{L} \overset{\text{def}}{=} \max_{x\in[-1,1]} \sum_{i=1}^{U} |q_i(x)|$, the operator norm of the interpolation operator (with respect to the uniform norm). It is well-understood that the choice of interpolation points with a small Lebesgue constant is crucial in numerical computation with interpolants (see, e.g., (Trefethen, 2013, Chap. 15)); this has also been demonstrated in the context of sum-of-squares optimization Papp (2017). One interpretation of Theorem 3.8 is that the choice of interpolation points is important even in exact-arithmetic computation, as the bit sizes of the certificates are affected by the Lebesgue constant.

In the univariate case, the Lebesgue constant with U suitably chosen interpolation points from $[-1, 1]$ can be as low as $\mathcal{O}(\log U)$ (Trefethen, 2013, Chap. 15). Thus, even for fairly suboptimal points, the impact of $\mu \le U\mathcal{L} \approx \mathcal{O}(d \log d)$ in the bit size bound (15) is dominated by the $\log d$ term, simplifying the bound to

$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}(\log(d) + \log(\|\mathbf{t}\|_2) + \log(1/\varepsilon)).$$

3.3. Multivariate polynomials over a bounded set

In this section, we assume that $\Sigma = \Sigma_{n,\mathbf{d}}^{\mathbf{w}}$, with $\mathbf{w}$ a set of weight polynomials describing a bounded set $S_{\mathbf{w}} \subseteq \mathbb{R}^n$. We assume that, using the notation of Proposition 1.1, $\mathbf{q}$ is represented by a multivariate Lagrange interpolation basis. Although in the multivariate case we can no longer rely on the fact that nonnegative polynomials are the same as WSOS polynomials with suitably chosen weights, the analysis in the multivariate case can be made largely identical to the univariate analysis of the previous section under an additional assumption that is only slightly stronger than assuming $S_{\mathbf{w}}$ to be bounded (see also Remark 3.11 below).

Theorem 3.10. *Suppose the* $\mathbf{q}$ *basis polynomials (in the notation of Proposition 1.1) are the Lagrange basis polynomials corresponding to the (unisolvent) interpolation points* $\{\mathbf{z}_1, \ldots, \mathbf{z}_U\} \subseteq S_{\mathbf{w}}$. *In addition, suppose that there is a* $\mu > 0$ *such that* $\mu + q_i \in \Sigma$ *and* $\mu - q_i \in \Sigma$ *(for each* $i = 1, \ldots, U$*). Then the following hold:*

1. *For* $k_1^{\mathbf{1}}$ *defined as in (11), we have* $k_1^{\mathbf{1}} \geq \frac{1}{\mu}$.
2. *The matrix* $\mathbf{M}$ *given in (8) can be chosen to be the identity matrix, with* $\operatorname{cond}(\mathbf{M}) = 1$.
3. *For all* $\mathbf{t} \in \Sigma^\circ$ *and* $\varepsilon > 0$ *with* $\mathbf{t} - \varepsilon\mathbf{1} \in \operatorname{bd}(\Sigma)$, *there exists an integer certificate* $\bar{\mathbf{y}} \in (\Sigma^*)^\circ \cap \mathbb{Z}^U$ *for* $\mathbf{t}$ *with*
$$\|\bar{\mathbf{y}}\|_\infty \leq \frac{1}{2} + \left\lceil \frac{3}{2}\sqrt{U}\|\mathbf{t}\|_2 \right\rceil \left(\frac{\mu\nu}{\varepsilon}\right)$$
whose largest component has bit size bounded as
$$\log(\|\bar{\mathbf{y}}\|_\infty) \approx \mathcal{O}\left(\log(U) + \log(\|\mathbf{t}\|_2) + \log(\mu) + \log(\nu) + \log(1/\varepsilon)\right).$$

Proof. The proofs of the first two statements are essentially identical to those in Theorem 3.8. Substituting those two bounds into the formula given in Theorem 2.9 gives the third claim. □

Remark 3.11. The new assumption $\mu \pm q_i \in \Sigma$ is relatively mild, given that $S_{\mathbf{w}}$ is bounded. Recall that if $S_{\mathbf{w}}$ is bounded, then every strictly positive polynomial over $S_{\mathbf{w}}$ belongs to the interior of the cone of polynomials nonnegative over $S_{\mathbf{w}}$; in particular, the constant 1 polynomial belongs to the interior. It is reasonable to assume that $1 \in \Sigma^\circ$ holds as well (as it automatically does for many WSOS cones commonly encountered in applications), in which case $1 \pm \frac{1}{\mu}q_i \in \Sigma$ automatically holds for every large enough μ.

The condition $1 \in \Sigma^\circ$ plays an important role in the context of dual certificates. For example, it ensures that *every* polynomial in $\operatorname{span}(\Sigma)$ has a WSOS lower bound (Davis and Papp, 2022, Lemma 3.1); as such, we will rely on it when we discuss exact-arithmetic algorithms to compute rational dual certificates in Section 4. As discussed in (Davis and Papp, 2022, Theorem 3.7), even in the case when this assumption does not hold, it is possible to extend Σ, 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 or invoking a Postivstellensatz).

4. Computing certified WSOS lower bounds in rational arithmetic

Having established the existence of rational dual certificates with a priori bounded bit sizes, we now turn to the question of computing such certificates. More precisely, given a polynomial $\mathbf{t}$ and a tolerance $\varepsilon > 0$, we want to compute a rational lower bound c that lies between the optimal WSOS lower bound c^* and $c^* - \varepsilon$, along with a rational dual certificate (of a small bit size) proving $\mathbf{t} - c\mathbf{1} \in \Sigma$.

The new algorithm (Algorithm 1 below) is an adaptation of Algorithm 1 from Davis and Papp (2022), which is a hybrid method for the solution of the same problem. Since that algorithm was designed to run in finite-precision floating point arithmetic, the quality of the lower bound is limited by the precision of the arithmetic (limiting how small ε can be). That algorithm, as is, cannot be efficiently implemented in infinite precision (rational) arithmetic, because the bit sizes of the intermediate quantities (and the returned certificate) blow up as the algorithm progresses even if ε is large. The new Algorithm 1 follows the same blueprint, but rounds all intermediate quantities (lower bounds and certificates) to "nearby" rational ones with small denominators, while maintaining the desirable properties of the original algorithm that guarantee that the new algorithm also converges linearly to the optimal WSOS lower bound.

The algorithm works for almost any WSOS cone Σ; our only assumption is that $\mathbf{1} \in \Sigma^\circ$. This is a mild assumption that ensures that every polynomial has a WSOS lower bound; recall Remark 3.11.

4.1. The algorithm

The pseudocode of the algorithm is shown in Algorithm 1; see a detailed example of one iteration of the algorithm after the outline of its analysis, in Example 4.2. Throughout, $\mathbf{x}$ represents a dual certificate for the polynomial $\mathbf{t} - c\mathbf{1}$, where c is the current certified WSOS lower bound and $\mathbf{t}$ is the input polynomial we wish to bound. In its main loop, the algorithm first updates the current certificate $\mathbf{x}$ to be closer to the gradient certificate for the current $\mathbf{t} - c\mathbf{1}$ (Line 3), by taking a single Newton step towards the solution of the nonlinear system $-g(\mathbf{x}) = \mathbf{t} - c\mathbf{1}$. This updated certificate is then rounded to a rational one with smaller denominators in Line 4. (The matrix norm required for this calculation could be expensive to compute, but since any upper bound can be substituted for $\|H(\mathbf{x}_+)\|$, one can get an acceptable rigorous bound by using the Frobenius norm instead, which is easy to compute exactly even in rational arithmetic.) Then the lower bound c is improved in Line 5 and rounded to a nearby rational bound that is still certified by the same rounded dual certificate (Line 6). This last step can be implemented efficiently using continued fractions or Farey sequences. Alternatively, it could be replaced by a naive rounding to the largest number in the interval with denominator $\lceil 2/\overline{\Delta c} \rceil$, the analysis below remains valid even in that case.

In addition to the notation introduced in the algorithm, we will use the following notation throughout the rest of the section. We let $\mathbf{y}$ be the vector satisfying $-g(\mathbf{y}) = \mathbf{t} - c\mathbf{1}$ and $\mathbf{y}_+$ be the vector satisfying $-g(\mathbf{y}_+) = \mathbf{t} - c_+\mathbf{1}$. The constants ρ and C used in the termination criterion will be precisely defined and justified below, in Theorem 4.1—for now, we may treat the main loop of the algorithm as an infinite loop.

The initialization of the algorithm requires a certificate $\mathbf{x}$ satisfying $\| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^* \leq \frac{r}{r+1}$. This may be readily available, for example, when the gradient certificate of $\mathbf{1}$ is known in closed form. (See (Davis and Papp, 2022, Example 4) for an example.) If such a certificate $\mathbf{x}$ is not known, we may run Algorithm 2 discussed in Section 4.2 to compute such a certificate. Note that this initial vector is independent of the coefficient vector $\mathbf{t}$, and only needs to be computed once for every WSOS cone Σ.

The analysis of Algorithm 1 here is similar to the analysis of Algorithm 1 in (Davis and Papp, 2022, Section 3.3). Therefore, we only give a concise outline of the proofs of its correctness and rate of convergence. We shall refer to the analysis of its predecessor whenever possible, focusing on where the analyses differ as a result of the rounding steps.

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

---

input : A polynomial $\mathbf{t}$; a tolerance $\varepsilon > 0$.
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 $\mathbf{t} - c\mathbf{1} \in \Sigma$.
parameters: An oracle for computing the barrier Hessian H for Σ; a radius $r \in (0, 1/4]$, a radius r_N satisfying $\frac{r^2}{1-2r} < r_N < \frac{r}{1+2r}$, a certificate $\mathbf{x}$ satisfying $\| - g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^* \leq \frac{r}{r+1}$.

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 Round $\mathbf{x}_+$ component-wise to a point $\mathbf{x}_N$ with denominators $N \stackrel{\text{def}}{=} \left\lceil \frac{\sqrt{U}}{2} \left(\frac{1+r_N}{r_N - \frac{r^2}{1-2r}}\right) \|H(\mathbf{x}_+)^{1/2}\| \right\rceil$.
5 Solve for c_+ the scalar quadratic equation

$$\|\mathbf{x}_N - H(\mathbf{x}_N)^{-1}(\mathbf{t} - c_+\mathbf{1})\|_{\mathbf{x}_N} = \frac{r}{r+1},$$

  and set c_+ equal to the larger of the two solutions.
6 Set $\overline{\Delta c} := c_+ - c$. Chose c'_+ to be the rational point in the interval $[c + \frac{\overline{\Delta c}}{2}, c_+]$ with the smallest possible denominator.
7 Set $\Delta c := c'_+ - c$. Set $c := c'_+$. Set $\mathbf{x} := \mathbf{x}_N$.
8 **until** $\Delta c \leq \frac{1}{2}\rho C \varepsilon$
9 **return** c *and* $\mathbf{x}$.

Theorem 4.1. *Suppose that, at the beginning of the main loop of Algorithm 1, $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \le r$ for some $r < \frac{1}{4}$. Then:*

1. *After Step 3, $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$.*
2. *After Step 4, $\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}_N} \le r_N$.*
3. *After Steps 5 and 6, we have $c'_+ > c$ and $\|\mathbf{x}_N - \mathbf{y}_+\|_{\mathbf{x}_N} \le r$, so $\|\mathbf{x}-\mathbf{y}\|_{\mathbf{x}} \le r$ also holds at the end of the loop, and the algorithm improves the lower bound c in each iteration.*

Moreover, Algorithm 1 is globally q-linearly convergent to $c^* = \max\{c \mid \mathbf{t} - c\mathbf{1} \in \Sigma\}$, the optimal WSOS lower bound for the polynomial $\mathbf{t}$. More precisely, in each iteration of Algorithm 1, the improvement of the lower bound $\Delta c = c'_+ - c$ satisfies

$$\frac{\Delta c}{c^* - c} \ge \frac{1}{2}\rho C, \tag{16}$$

with the absolute constant $\rho \overset{\text{def}}{=} \frac{r}{r+1} - \frac{r_N}{1-r_N}$ and the Λ-dependent constant $C > 0$ defined as in Theorem 3.5 of Davis and Papp (2022).

Proof. Statement 1 is identical to Davis and Papp (2022, Lemma 3.2), and statement 2 comes from Lemma 2.1 with $r_1 = \frac{r^2}{1-2r}$ and $r_2 = r_N$.

Statement 3 is analogous to Davis and Papp (2022, Lemma 3.3), replacing $\mathbf{x}_+$ therein with $\mathbf{x}_N$ and making use of the fact that $\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}_N} \le r_N$ whenever (Davis and Papp, 2022, Lemma 3.3) uses the inequality $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$. The fact that $c'_+ > c$ comes from the fact that $c_+ > c$ (Davis and Papp, 2022, Lemma 3.3) and the construction of c'_+.

The linear convergence result is analogous to Davis and Papp (2022, Theorem 3.6), with c'_+ playing the role of c_+. The rounding down of the lower bound c_+ to c'_+ in Step 5 ensures that in spite of the rounding, the progress our Algorithm 1 is at least half of what the progress would be without rounding (as in Davis and Papp, 2022, Algorithm 1) The final inequality (16) justifies the termination criterion in Line 7: if $\Delta c \le \frac{1}{2}\rho C\varepsilon$, then the gap between the certified and the optimal lower bound is $c^* - c \le \varepsilon$ as wanted. □

Example 4.2. Consider the polynomial $t(z_1, z_2) = 3z_1^2 - 6z_1z_2 + z_2^2 + 2z_1 - z_2$. Its global minimum on the unit disk is $c^* \approx -1.70768680307$. We can certify a sequence of lower bounds $c < c^*$ by writing $t - c$ in the form

$$t(z_1, z_2) - c = \sigma_1(z_1, z_2) + (1 - z_1^2 - z_2^2)\sigma_2,$$

where σ_1 is a quadratic SOS polynomial and σ_2 is a nonnegative constant (a degree-0 SOS polynomial); i.e. by showing that $t - c \in \Sigma^{(1,w)}_{2,(2,0)}$, wherein w is the weight polynomial $(z_1, z_2) \mapsto (1 - z_1^2 - z_2^2)$.

Using the monomial basis to represent all polynomials, we have the coefficient vector $\mathbf{t} = (0, 2, 3, -1, -6, 1)$, and it is straightforward to construct the Λ operator: we have $U = \dim(\Sigma) = 6$, $(L_1, L_2) = (3, 1)$, and $\Lambda = \Lambda_1 \oplus \Lambda_2$, where

$$\Lambda_1(\mathbf{x}) = \begin{pmatrix} x_1 & x_2 & x_4 \\ x_2 & x_3 & x_5 \\ x_4 & x_5 & x_6 \end{pmatrix}, \quad \Lambda_2(\mathbf{x}) = \begin{pmatrix} x_1 - x_3 - x_6 \end{pmatrix} \tag{17}$$

Now the Hessian of the barrier function can be computed efficiently using the formula (A.4).

To initialize Algorithm 1, we need a vector $\mathbf{x}_1$ sufficiently close to the gradient certificate of the constant one polynomial. In this simple example, we will use the gradient certificate itself, which can be computed in closed form and happens to be a rational vector, $\mathbf{x}_1 = (4, 0, \frac{4}{3}, 0, 0, \frac{4}{3})$. This can be verified by direct computation: $-g(\mathbf{x}_1) = (1, 0, 0, 0, 0, 0) = \mathbf{1}$.

In the main loop of the algorithm, we shall use the parameters $r = 1/4$ and $r_N = 1/7$ in this example. The corresponding (very crude) initial lower bound $c_0 = -\frac{10}{3}\sqrt{\frac{410}{3}} \approx -38.97$ is not rational, but we can round it down to $c_0 = -39$. We can also verify that this lower bound is indeed certified by $\mathbf{x}_1$: using the definition of dual certificates, it suffices to compute $H(\mathbf{x}_1)^{-1}(\mathbf{t} - c_0\mathbf{1}) = \left(\frac{484}{3}, \frac{16}{3}, \frac{1532}{27}, -\frac{8}{3}, -\frac{16}{3}, \frac{1436}{27}\right)$ and confirm that both Λ_i from (17) are indeed positive definite. For completeness, we can also generate an explicit WSOS representation using (6):

$$t(z_1, z_2) + 39 = \begin{pmatrix} 1 \\ z_1 \\ z_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \frac{121}{12} & 1 & -\frac{1}{2} \\ 1 & \frac{383}{12} & -3 \\ -\frac{1}{2} & -3 & \frac{359}{12} \end{pmatrix} \begin{pmatrix} 1 \\ z_1 \\ z_2 \end{pmatrix} + (1 - z_1^2 - z_2^2)\frac{347}{12}.$$

The first iteration of the algorithm proceeds as follows:

1. After computing $H(\mathbf{x}_1)$, the updated certificate in Line 3 is

$$\mathbf{x}_+ = \left(\frac{452}{4563}, \frac{8}{4563}, \frac{1372}{41067}, -\frac{16}{4563}, \frac{16}{4563}, \frac{1276}{41067}\right).$$

2. The denominator for the "compressed" certificate is $N = 5029$; the computation of this involves calculating the updated Hessian, $H(\mathbf{x}_+)$, the rest is trivial arithmetic. Rounding $\mathbf{x}_+$ componentwise, the rounded certificate becomes

$$\mathbf{x}_N = \frac{1}{5029}(498, 9, 168, -18, 18, 156).$$

3. To update the lower bound, we construct the scalar quadratic equation in Line 5; reusing the already computed Hessian, this is simple arithmetic. The equation can be written as

$$29\,387\,195\,615\,576 + 1\,508\,777\,838\,050c_+ + 19\,170\,557\,325c_+^2 = 0,$$

whose larger root is approximately -35.4, meaning that (in the spirit of keeping the denominators as small as possible), we can update our lower bound to $c'_+ = -36$. Indeed, we have

$$t(z_1, z_2) + 36 = \begin{pmatrix} 1 \\ z_1 \\ z_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} s_1 & 1 & -\frac{1}{2} \\ 1 & s_2 & -3 \\ -\frac{1}{2} & -3 & s_3 \end{pmatrix} \begin{pmatrix} 1 \\ z_1 \\ z_2 \end{pmatrix} + (1 - z_1^2 - z_2^2)s_4$$

with $(s_1, s_2, s_3, s_4) = \frac{1}{344\,769}(3\,203\,164, 10\,242\,827, 9\,553\,289, 9\,208\,520)$.

Although some of the coefficients appear frighteningly large for a toy example, it shall be emphasized that the explicit WSOS decompositions of $t - c$ need not be computed in the algorithm.

Continuing with the algorithm, the bit sizes of the dual certificates appear to grow linearly with the number of iterations, and (as predicted by the theory), the difference $c^* - c$ decreases exponentially with the number of iterations. For instance, after 200 iterations, the lower bound is $-1\,579\,834/925\,131$, only about $2 \cdot 10^{-10}$ away from the true minimum value.

4.2. Initialization

Algorithm 1 requires a suitable certificate of $\mathbf{1}$, the constant one polynomial, to initialize. Such a certificate may be available either in closed form (e.g., for cones of univariate polynomials nonnegative on an interval, the gradient certificate can be determined analytically (Davis and Papp, 2022, Example 4)), or from "preprocessing", e.g., when other polynomials from the same space span(Σ) have already been bounded.

If we do not know a suitable $\mathbf{x}$ to start with, then we could attempt to find the gradient certificate (or a rational approximation of it) by solving the system $-g(\mathbf{x}) = \mathbf{1}$ by a general-purpose method for polynomial systems. Alternatively (and more efficiently), we can find an approximate solution to this system by numerically solving the convex optimization problem

$$\min\{f(\mathbf{x}) + \mathbf{1}^{\mathrm{T}}\mathbf{x} \,|\, \mathbf{x} \in \Sigma^*\}.$$

Instead of these numerical approaches, we can also leverage Algorithm 1 itself to find a suitable initial point. Suppose we have an interior point $\mathbf{x} \in (\Sigma^*)^\circ$, which is by definition the gradient certificate of the polynomial $\mathbf{s} = -g(\mathbf{x})$. Then we can apply Algorithm 1 starting with this initial pair "in reverse," computing a sequence of certificates for polynomials of the form $\mathbf{s} + c\mathbf{1}$ for *increasing* values of c. The same certificates in turn certify $c^{-1}\mathbf{s} + \mathbf{1}$ as well, which is approximately the same as the polynomial $\mathbf{1}$ when c is large enough. The details of this approach are presented in Algorithm 2. Its analysis largely follows the steps laid out in Theorem 4.1, with two minor adjustments regarding the progress and the termination criterion. The change in Line 6 guarantees that the stopping criterion can be met, as we show later in Lemma 4.4.

Algorithm 2: Initialization for Algorithm 1.

input : A vector $\mathbf{x} \in (\Sigma^*)^\circ$.
parameters: An oracle for computing the barrier Hessian H for Σ; a radius $r \in (0, 1/4]$, a radius $\frac{r^2}{1-2r} < r_N < \frac{r}{1+2r}$.
outputs : A certificate $\mathbf{x} \in (\Sigma^*)^\circ$ satisfying $\|-g(\mathbf{x}) - \mathbf{1}\|_{\mathbf{x}}^* \le \frac{r}{r+1}$.

1 Compute $\mathbf{s} := -g(\mathbf{x})$. Set $c := 0$.
2 **repeat**
3 Set $\mathbf{x}_+ := 2\mathbf{x} - H(\mathbf{x})^{-1}(\mathbf{s} + c\mathbf{1})$.
4 Round $\mathbf{x}_+$ component-wise to a point $\mathbf{x}_N$ with denominators $N \stackrel{\text{def}}{=} \left\lceil \frac{\sqrt{U}}{2}\left(\frac{1+r_N}{r_N - \frac{r^2}{1-2r}}\right)\|H(\mathbf{x}_+)^{1/2}\| \right\rceil$.
5 Solve for c_+ the scalar quadratic equation

$$\|\mathbf{x}_N - H(\mathbf{x}_N)^{-1}(\mathbf{s} + c_+\mathbf{1})\|_{\mathbf{x}_N} = \frac{r}{r+1},$$

and set c_+ equal to the larger of the two solutions.
6 Set $\overline{\Delta c} := c_+ - c$. Chose c_+' to be the rational point in the interval $[c + \frac{1}{2}\overline{\Delta c}, c + \frac{2}{3}\overline{\Delta c}]$ with the smallest possible denominator.
7 Set $\Delta c := c_+' - c$. Set $c := c_+'$. Set $\mathbf{x} := \mathbf{x}_N$.
8 **until** $\|-g(c\mathbf{x}) - \mathbf{1}\|_{c\mathbf{x}}^* \le \frac{r}{r+1}$.
9 **return** $c\mathbf{x}$.

Analogously to Section 4.1, we will let $\mathbf{y}$ be the vector satisfying $-g(\mathbf{y}) = \mathbf{s} + c\mathbf{1}$ and $\mathbf{y}_+$ be the vector satisfying $-g(\mathbf{y}_+) = \mathbf{s} + c_+\mathbf{1}$ throughout this section, in addition to the notation introduced in Algorithm 2. The algorithm can be initialized by any point in the interior of Σ^*, and by definition, at the beginning of the main loop of the algorithm, we have $\mathbf{x} = \mathbf{y}$, and therefore $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \le r$. The proof of Theorem 4.1 can be repeated almost verbatim to show the following.

Theorem 4.3. *Suppose that, at the beginning of the main loop of Algorithm 2, $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \le r$ for some $r < \frac{1}{4}$. Then:*

1. *After Step 3, $\|\mathbf{x}_+ - \mathbf{y}\|_{\mathbf{x}_+} \le \frac{r^2}{1-2r}$.*
2. *After Step 4, $\|\mathbf{x}_N - \mathbf{y}\|_{\mathbf{x}_N} \le r_N$.*
3. *After Steps 5 and 6, we have $c_+' > c$ and $\|\mathbf{x}_N - \mathbf{y}_+\|_{\mathbf{x}_N} \le r$, so $\|\mathbf{x} - \mathbf{y}\|_{\mathbf{x}} \le r$ also holds at the end of the loop. The increase $\Delta c = c_+' - c$ satisfies*

$$\frac{\Delta c}{c - c^*} \ge \frac{1}{2}\rho C,$$

with the same constants ρ and C as in Theorem 4.1. Thus, the constant c increases exponentially as the algorithm progresses.

By definition, if Algorithm 2 terminates, it returns a vector that can be used as an initial vector $\mathbf{x}$ in Algorithm 1. It only remains to show that the algorithm indeed terminates. We shall show this in two steps. First, we show that as the algorithm progresses and c increases, $\|\mathbf{s}\|_{\mathbf{x}}^*$ tends to zero. Then we argue that this ensures that the polynomial $c^{-1}\mathbf{s}+\mathbf{1}$ is eventually "close enough" to $\mathbf{1}$ that $c\mathbf{x}$ is sufficiently close to the gradient certificate of $\mathbf{1}$.

Lemma 4.4. *Let $\mathbf{x} \in (\Sigma^*)^\circ$ be the certificate of the polynomial $\mathbf{s}+c\mathbf{1} \in \Sigma^\circ$ as defined in Algorithm 2. Then*

1. *$\|\mathbf{s}\|_{\mathbf{x}}^*$ tends to 0, and*
2. *Algorithm 2 terminates.*

Proof. We begin with the first statement. Let $\mathbf{x}_1$ be the gradient certificate of $\mathbf{1}$ and fix an arbitrary $a \in (0,1)$. Let $\mathbf{y}$ be the gradient certificate of $\mathbf{s}+c\mathbf{1}$. By Davis and Papp (2022, Lemma 3.1), we know that for every sufficiently large c,

$$\|c^{-1}\mathbf{x}_1 - \mathbf{y}\|_{c^{-1}\mathbf{x}_1} \le a.$$

Then, using inequality (A.2) from the Appendix and the fact that $\|\mathbf{s}\|_{c^{-1}\mathbf{x}_1}^* = c^{-1}\|\mathbf{s}\|_{\mathbf{x}_1}^*$, we have

$$\|\mathbf{s}\|_{\mathbf{y}}^* \overset{(A.2)}{\le} \frac{\|\mathbf{s}\|_{c^{-1}\mathbf{x}_1}^*}{1-\|c^{-1}\mathbf{x}_1-\mathbf{y}\|_{c^{-1}\mathbf{x}_1}} \le \frac{c^{-1}\|\mathbf{s}\|_{\mathbf{x}_1}^*}{1-a}. \tag{18}$$

Using inequality (A.2) again, we have

$$\|\mathbf{s}\|_{\mathbf{x}}^* \overset{(A.1)}{\le} \frac{\|\mathbf{s}\|_{\mathbf{y}}^*}{1-\|\mathbf{x}-\mathbf{y}\|_{\mathbf{y}}} \le \frac{\|\mathbf{s}\|_{\mathbf{y}}^*}{1-r}. \tag{19}$$

Therefore, we have

$$\|\mathbf{s}\|_{\mathbf{x}}^* \overset{(18),(19)}{\le} \frac{c^{-1}\|\mathbf{s}\|_{\mathbf{x}_1}^*}{(1-r)(1-a)}.$$

Since a can be chosen to be arbitrarily close to 0, and since r and $\|\mathbf{s}\|_{\mathbf{x}_1}^*$ are constants, it follows that as $c \to \infty$, $\|\mathbf{s}\|_{\mathbf{x}}^* \to 0$.

Now, we show that Algorithm 2 terminates. At the end of each iteration, we have

$$\begin{aligned}\left\|-g(c\mathbf{x})-(c^{-1}\mathbf{s}+\mathbf{1})\right\|_{c\mathbf{x}}^* &\overset{(A.5)}{=} \|-g(\mathbf{x})-(\mathbf{s}+c\mathbf{1})\|_{\mathbf{x}}^* \\ &\overset{(A.6)}{=} \|\mathbf{x}-H(\mathbf{x})^{-1}(\mathbf{s}+c\mathbf{1})\|_{\mathbf{x}} \overset{\text{Line } 6}{<} \frac{r}{r+1}.\end{aligned}$$

Thus, we have

$$\begin{aligned}\|-g(c\mathbf{x})-\mathbf{1}\|_{c\mathbf{x}}^* &\le \left\|-g(c\mathbf{x})-(c^{-1}\mathbf{s}+\mathbf{1})\right\|_{c\mathbf{x}}^* + \left\|(c^{-1}\mathbf{s}+\mathbf{1})-\mathbf{1}\right\|_{c\mathbf{x}}^* \\ &< \frac{r}{r+1} + \|\mathbf{s}\|_{\mathbf{x}}^*.\end{aligned}$$

By Statement 1, $\|\mathbf{s}\|_{\mathbf{x}}^*$ tends to 0, so eventually the stopping criterion is satisfied. □

5. Discussion

Bit size bounds on certificates from Algorithm 1 We opted to separate the discussion on the bit sizes of the certificates and Algorithm 1. In principle, one could study the former question "constructively" by analyzing the bit sizes of the certificates computed by the algorithm, but we think it is useful to underline that both the concept of dual certificates and the bit size bounds are independent of any particular algorithm. Theorem 2.4 and Lemma 2.8 are both derived assuming that the dual certificate

$\mathbf{y}$ at hand is the gradient certificate for simplicity of presentation; both of these results can be easily adapted to the setting where $\mathbf{y}$ is any dual certificate that is sufficiently close to the gradient certificate. Similarly, any algorithm that computes a dual certificate by computing a vector $\mathbf{x}$ sufficiently close (in the local $\mathbf{x}$-norm) to the gradient certificate $\mathbf{y}$ will produce certificates with boundable bit sizes.

The Christoffel-Darboux polynomial The WSOS polynomial $-g(\mathbf{x})$ corresponding to a (pseudo-moment) vector $\mathbf{x} \in (\Sigma^*)^\circ$ is also known as the *Christoffel-Darboux polynomial*, or *inverse Christoffel function*. Recent studies have focused on the properties of this polynomial, and especially on connections between the representing measures of $\mathbf{x}$ and the sublevel sets of $-g(\mathbf{x})$, with applications to design of experiments Castro et al. (2021); Lasserre (2022). For our work, the critical property of the Christoffel-Darboux polynomial is the surprising fact that this polynomial is not only WSOS, but that the gradient map $\mathbf{x} \mapsto -g(\mathbf{x}) = \Lambda^*(\Lambda(\mathbf{x})^{-1})$ yields an explicit WSOS representation of this polynomial: the inverse moment matrix $\Lambda(\mathbf{x})^{-1}$ is a Gram matrix that proves that $-g(\mathbf{x})$ belongs to Σ. The concept of dual certificates can be seen as a generalization of this idea: rather than mapping $\mathbf{x}$ to the Christoffel-Darboux polynomial with an explicit WSOS representation via $-g$, we can map $\mathbf{x}$ to explicit WSOS representations of a full-dimensional cone of WSOS polynomials $\mathbf{s}$ via the $(\mathbf{x}, \mathbf{s}) \to \mathbf{S}$ map in (6).

Representation-dependent bit sizes Although the main theorem (Theorem 2.9) provides a bit size bound using an array of unconventional parameters, it is worth noting that each of those parameters is easily computable or, in the case of $\text{cond}(\mathbf{M})$, can be bounded easily. Although computing the matrix $\mathbf{M}$ with the lowest condition number is in general a likely impossible task, any unisolvent point set and weight vector in (8) can be used to compute a bound. In each of the special cases considered, it was easy to find a point set that either yields a small enough $\text{cond}(\mathbf{M})$ that is dominated by other terms, or one that is provably of the optimal order of magnitude.

The other nontrivial ingredient is the constant $k_1^{\mathbf{e}}$ defined in (11). Unlike the other parameters in Theorem 2.9, it depends not only on the cone Σ, but its chosen representation via the Λ operator. (Equivalently, in the notation of Proposition 1.1, in depends both on the choice of the $\mathbf{p}_i$ bases and the $\mathbf{q}$ basis.) The example of interpolants suggests that it is a measure of conditioning, underscoring the fact that conditioning is consequential even in the case of exact-arithmetic algorithms, not only for numerical methods. Indeed, choosing poor interpolation points (say, equispaced points) to represent WSOS polynomials instead of well-conditioned ones leads to an increase in the bit sizes of the certificates, even in the univariate case, due to the astronomical Lebesgue constant that grows exponentially with the degree, dominating all the other terms in (15). It is also this parameter, along with ν, that may be reduced when the polynomials of interest have special structures such as symmetry or term- or correlative sparsity Wang et al. (2020), showing that these structures are useful even for dual certificates.

Also note that none of these parameters need to be known in order to implement the algorithms discussed in Section 4, except for the stopping criterion of Algorithm 1. If we drop the requirement that the algorithm must stop when the returned bound is provably within ε from the optimal lower bound, and instead run the algorithm until the progress is below a tolerance, or when the lower bound is suitably high (e.g., in applications where the goal is to prove that the input polynomial has a positive WSOS lower bound), then none of the parameters introduced in Theorem 2.9 need to be computed or bounded.

Dependence on ν The possibly most counterintuitive aspect of the bound in Theorem 2.9 is that the bit size of the integer dual certificate $\bar{\mathbf{y}}$ (approximated by $U \log(\|\bar{\mathbf{y}}\|_\infty)$) depends on the Σ-dependent parameter ν logarithmically, rather than linearly, since (all else being equal) ν is a linear function of the number of weights m (recall the notation from (1) and its surrounding paragraph). For example, consider a family of polyhedral cones of nonnegative polynomials that consist of nonnegative linear combinations of polynomials that are nonnegative on $S_{\mathbf{w}}$. This is an elementary special case of WSOS polynomials, where all "sum-of-squares" polynomials are simply nonnegative constants, and where it is meaningful to keep adding additional weights to the representation for an increasingly good inner approximation of the cone of nonnegative polynomials without increasing the "ambient dimension" $U = \dim(\Sigma)$. It is clear that the sizes of conventional WSOS certificates will grow linearly as a function of m: an explicit WSOS decomposition will have m terms; the semidefinite matrix in the representation (2) will have m (one-by-one) semidefinite blocks $\mathbf{S}_i$, etc. Yet, the dual certificate will remain a U-dimensional vector whose components are of size $\mathcal{O}(\log(m))$. Of course, this does not mean that such a certificate could be verified in polynomial time for an exponentially large m (like in a Schmüdgen-type WSOS certificate); verifying that $\mathbf{x}$ certifies $\mathbf{s}$, that is, $\Lambda(H(\mathbf{x})^{-1}\mathbf{s}) \succcurlyeq 0$ still requires linear time in the number of weights.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors are grateful for the referees' thoughtful comments on the presentation, helping to make the paper more accessible to the symbolic computing and computational real algebraic geometry audience. DP would also like to thank Didier Henrion for pointing out the connection between gradient certificates and the Christoffel-Darboux polynomial, and Ali Mohammad Nezhad for pointing us to the results on the bit sizes of subresultants used in the proof of Corollary 3.2.

Appendix A

Here, we summarize relevant results used throughout the paper concerning the local norms $\|\cdot\|_{\mathbf{x}}$ and $\|\cdot\|_{\mathbf{x}}^*$ and barrier functions of the form $f = -\ln(\det(\Lambda(\cdot)))$ defined on $(\Sigma^*)^\circ$, introduced in Section 1.1.

Lemma A.1. *Using the notation introduced in Section 1.1, 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*
$$1 - \|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}} \leq \frac{\|\mathbf{v}\|_{\mathbf{u}}}{\|\mathbf{v}\|_{\mathbf{x}}} \leq (1 - \|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}})^{-1}. \tag{A.1}$$
2. *For all* $\mathbf{v} \neq 0$*, if* $\|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}} < 1$*, we have*
$$\frac{\|\mathbf{v}\|_{\mathbf{u}}^*}{\|\mathbf{v}\|_{\mathbf{x}}^*} \leq \frac{1}{1 - \|\mathbf{u}-\mathbf{x}\|_{\mathbf{x}}}. \tag{A.2}$$
3. *The gradient* g *of* f *can be computed as*
$$g(\mathbf{x}) = -\Lambda^*(\Lambda(\mathbf{x})^{-1}), \tag{A.3}$$
and the Hessian $H(\mathbf{x})$ *is the linear operator satisfying*
$$H(\mathbf{x})\mathbf{v} = \Lambda^*\big(\Lambda(\mathbf{x})^{-1}\Lambda(\mathbf{v})\Lambda(\mathbf{x})^{-1}\big) \quad \text{for every } \mathbf{v} \in \mathbb{R}^U. \tag{A.4}$$
4. *The function* f *is* logarithmically homogeneous*; that is,*
$$f(\alpha\mathbf{x}) = f(\mathbf{x}) - \nu\ln(\alpha) \text{ for every } \alpha > 0$$
where $\nu = \sum_{i=1}^m L_i$ *is the* barrier parameter *of* f*. Subsequently, the derivatives of* f *have the following homogeneity properties:*

$$g(\alpha\mathbf{x}) = \alpha^{-1}g(\mathbf{x}) \;\; \textit{and} \;\; H(\alpha\mathbf{x}) = \alpha^{-2}H(\mathbf{x}) \;\; \textit{for every } \alpha > 0. \tag{A.5}$$

Furthermore,

$$H(\mathbf{x})\mathbf{x} = -g(\mathbf{x}) \quad \textit{and} \quad \|g(\mathbf{x})\|_{\mathbf{x}}^* = \|\mathbf{x}\|_{\mathbf{x}} = \sqrt{\langle -g(\mathbf{x}), \mathbf{x}\rangle} = \sqrt{\nu}, \tag{A.6}$$

where ν *is the aforementioned barrier parameter.*

5. *The gradient map* g *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* $\mathbf{s} = -g(\mathbf{x})$*.*

Proof. Statement 1 is Renegar's definition of self-concordance Renegar (2001, Sec. 2.2.1) to the function f defined in (5). Statement 2 is (Papp and Yıldız, 2017, Lemma 4). Statements 3 and 4 follow from calculus. Statement 5 is from Renegar (2001, Sec. 3.3). □

References

Basu, S., Pollack, R., Roy, M., 2006. Algorithms in Real Algebraic Geometry. Springer-Verlag, Berlin Heidelberg.

Basu, S., Leroy, R., Roy, M., 2009. A bound on the minimum of a real positive polynomial over the standard simplex. arXiv:0902.3304.

Beckermann, B., 2000. The condition number of real Vandermonde, Krylov and positive definite Hankel matrices. Numer. Math. 85 (4), 553–577. https://doi.org/10.1007/PL00005392.

Blekherman, G., Parrilo, P.A., Thomas, R.R. (Eds.), 2013. Semidefinite Optimization and Convex Algebraic Geometry. SIAM, Philadelphia, PA.

Boudaoud, F., Caruso, F., Roy, M.-F., 2008. Certificates of positivity in the Bernstein basis. Discrete Comput. Geom. 39 (4), 639–655. https://doi.org/10.1007/s00454-007-9042-x.

Boyd, S.P., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.

Castro, Y.D., Gamboa, F., Henrion, D., Lasserre, J.B., 2021. Dual optimal design and the Christoffel–Darboux polynomial. Optim. Lett. 15, 3–8. https://doi.org/10.1007/s11590-020-01680-2.

Davis, M.M., Papp, D., 2022. Dual certificates and efficient rational sum-of-squares decompositions for polynomial optimization over compact sets. SIAM J. Optim. 32 (4), 2461–2492. https://doi.org/10.1137/21M1422574.

Hardy, G.H., Littlewood, J.E., Pólya, G., 1934. Inequalities. Cambridge University Press, London.

Katthän, L., Naumann, H., Theobald, T., 2021. A unified framework of SAGE and SONC polynomials and its duality theory. Math. Comput. 90 (329), 1297–1322. https://doi.org/10.1090/mcom/3607.

Krylov, V.I., 2005. Approximate Calculation of Integrals. Dover, Mineola, NY.

Lasserre, J.B., 2022. A disintegration of the Christoffel function. C. R. Math. 360, 1071–1079. https://doi.org/10.5802/crmath.380.

Laurent, M., 2009. Sums of squares, moment matrices and optimization over polynomials. In: Putinar, M., Sullivant, S. (Eds.), Emerging Applications of Algebraic Geometry. In: IMA Volumes in Mathematics and Its Applications, vol. 149. Springer, New York, NY, pp. 157–270.

Magron, V., Safey El Din, M., 2021a. On exact Reznick, Hilbert-Artin and Putinar's representations. arXiv preprint. arXiv:1912.04718. https://arxiv.org/abs/1811.10062v4.

Magron, V., Safey El Din, M., 2021b. On exact Reznick, Hilbert-Artin and Putinar's representations. J. Symb. Comput. 107, 221–250.

Magron, V., Safey El Din, M., Schweighofer, M., 2019. Algorithms for weighted sum of squares decomposition of non-negative univariate polynomials. J. Symb. Comput. 93, 200–220. https://doi.org/10.1016/j.jsc.2018.06.005.

Mason, J.C., Handscomb, D.C., 2003. Chebyshev Polynomials. CRC Press, Boca Raton, FL.

Nesterov, Y., 2000. Squared functional systems and optimization problems. In: Frenk, H., Roos, K., Terlaky, T., Zhang, S. (Eds.), High Performance Optimization. In: Applied Optimization, vol. 33. Kluwer Academic Publishers, Dordrecht, pp. 405–440.

Papp, D., 2017. Semi-infinite programming using high-degree polynomial interpolants and semidefinite programming. SIAM J. Optim. 27 (3), 1858–1879. https://doi.org/10.1137/15M1053578.

Papp, D., 2023. Duality of sum of nonnegative circuit polynomials and optimal SONC bounds. J. Symb. Comput. 114, 246–266. https://doi.org/10.1016/j.jsc.2022.04.015.

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

Papp, D., Yıldız, S., 2019. Sum-of-squares optimization without semidefinite programming. SIAM J. Optim. 29 (1), 822–851. https://doi.org/10.1137/17M1160124.

Powers, V., 2011. Rational certificates of positivity on compact semialgebraic sets. Pac. J. Math. 251 (2), 385–391. https://doi.org/10.2140/pjm.2011.251.385.

Renegar, J., 2001. A Mathematical View of Interior-Point Methods in Convex Optimization. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Trefethen, L.N., 2013. Approximation Theory and Approximation Practice. SIAM, Philadelphia, PA.

Wang, J., Magron, V., Lasserre, J.B., Mai CS-TSSOS, N.H.A., 2020. Correlative and term sparsity for large-scale polynomial optimization. technical report. https://arxiv.org/abs/2005.02828.

Wilf, H.S., 1970. Finite Sections of Some Classical Inequalities. Springer-Verlag, Berlin.