COERCIVITY OF THE COMPUTATION OF SUM OF SQUARES FROM DATA POINTS: THE CASE OF THE HYPERCUBE

BRUNO DESPRÉS*

Abstract. The goal of this work is to provide a simple condition on a multivariate polynomial p such that the dual function $\lambda \mapsto G(\lambda)$ defined in a previous work [5] is coercive (infinite at infinity). It is based on the fact that data points obtained from tensorization of the roots of the third and fourth kind Chebyshev polynomials possess a strong stability property, so they are (nearly) optimal. The stability property is fundamentally connecte to the Lebesgue stability constant of Chebyshev interpolation. It has the consequence that G has a global minimum, which justifies on the hypercube the gradient descent algorithms proposed in [5]. A corollary is a constructive representation of p as a sum of squares (SOS) endowed with the Schmüdgen's Positivstellensatz structure.

Key words. Positive polynomials, sum of squares, convex analysis, positive interpolation.

AMS subject classifications. 90C30, 65K05, 90C25

1. Introduction. We consider real polynomials $p \in \mathbf{P}[\mathbf{x}] = \mathrm{P}[x_1, \dots x_d]$ which are moreover positive over a semi-algebraic set $\mathbb{K} \subset \mathbb{R}^d$. These polynomials can be written as sum of squares (SOS) with the Schmüdgen's (resp. Putinar's) Positivstellensatz [22] (resp. [20]). A non exhaustive list of additional references is [19, 15, 18, 13, 23]. For reasons soundly explained in the work of Lasserre and coauthors [10, 11, 12], there is nowadays a strong impetus to transfer these theoretical characterizations to practical algorithms, in particular because it is way to construct practical certificates of positivity [10, 4, 5, 3]. The applications [10] range from finding the global minimum of a function on a subset of $\mathbb{R}^n$, to pricing exotic options in Mathematical Finance or computing Nash equilibria, or recently to using such tools in scientific computing [3]. In this work we focus on the justification of the family of convex algorithms that were proposed in [5] for general semi-algebraic sets $\mathbb{K}$. These algorithms are based on the construction of a dual function denoted as $G(\lambda)$, where by construction G is convex on its domain. The fonction G is constructed from the values at data points of a given polynomial p that one tries to write as a SOS. However the proof that G is coercive (that is infinite at infinity) was conditional in [5], and this condition was not explicitly stated (except in one dimension because it is much simpler). In particular the choice of the interpolation points at which one collects the polynomial values was not addressed. Therefore, in some sense, the algorithms proposed in [5] were not fully justified, even if the numerical results showed their usefulness. The reader interested by calculations of SOS with these gradient descent algorithms is advised to refer to the above reference where he will find numerical illustrations and examples.

Our goal is to focus on the case of the hypercube $\mathbb{K} = [0, 1]^d$ and to provide a simple condition on p such that $\lambda \mapsto G(\lambda)$ is coercive (infinite at infinity). It has the consequence that G has a global minimum, which justifies on the hypercube the gradient descent algorithms proposed in [5]. As a corollary, it yields a constructive representation of p as a sum of squares (SOS) endowed with the Schmüdgen's Positivstellensatz [23, 15] structure.

We need some notations before describing the main result. The subset of polynomials of degree less than or equal to $m \geqslant 1$ is denoted by $\mathbf{P}^m[\mathbf{x}]$, where the (non standard) convention is that the degree is the maximal univariate degree of the constituting monomials. One has $\mathbf{P}^m[\mathbf{x}] = \mathrm{Span}\left(\mathbf{P}^m[x_1] \times \cdots \times \mathbf{P}^m[x_d]\right)$: for exemple the degree of $x^2y^2 + x + y$ is equal to 2. This is mainly for the simplicity of the notations by tensorization used in this work, and we do not believe it is a fundamental restriction. Inspired by Schmügden's Putinar's Positvstellensatz, we will use the characterization of

*Sorbonne-Université, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France

the hypercube

$$\mathbb{K} = \left\{ \mathbf{x} \in \mathbb{R}^d \text{ such that } g_{\mathbf{j}}(\mathbf{x}) \geqslant 0 \text{ for } \mathbf{j} \in \{0,1\}^d \right\} \tag{1.1}$$

where $g_{\mathbf{j}}(\mathbf{x}) = \Pi_{i=1}^d x_i^{j_i}(1-x_i)^{1-j_i}$ and $\mathbf{j} = (j_1, \ldots, j_d)$ with $j_i \in \{0,1\}$ for all $1 \leqslant i \leqslant d$. The number of different functions $g_{\mathbf{j}}$ and equal to $j_* = 2^d$. We will use the alternative notation

$$\mathbb{K} = \left\{ \mathbf{x} \in \mathbb{R}^d \text{ such that } g_j(\mathbf{x}) \geqslant 0 \text{ for } 1 \leqslant j \leqslant j_* \right\}. \tag{1.2}$$

Another possibility for the characterization of the hypercube is to consider the functions $g_{2j-1}(\mathbf{x}) = x_j$ and $g_{2j}(\mathbf{x}) = 1 - x_j$, which makes $2d$ functions instead of 2^d functions (it makes the same number of functions in dimension $d = 2$). However it would required to work under the umbrella of the Putinar's Positivstellensatz, which would be a much more ambitious task not considered hereafter. That is why we will continue with the structure (1.1-1.2). The convex set of non-negative polynomials of maximal degree n on $\mathbb{K}$ is

$$\mathbf{P}^m_{\mathbb{K},+}[\mathbf{x}] = \{ p \in \mathrm{P}^m[\mathbf{x}] \text{ such that } p(\mathbf{x}) \geqslant 0 \text{ for any } \mathbf{x} \in \mathbb{K} \}. \tag{1.3}$$

Inspired by Schmügden's Positvstellensatz, we seek a representation of polynomials in $\mathbf{P}^{2n+1}_{\mathbb{K},+}[\mathbf{x}]$ (that is $m = 2n + 1$) as

$$p = \sum_{j=1}^{j_*} g_j \left(\sum_{i=1}^{i_*} q_{ij}^2 \right) = \sum_{i=1}^{i_*} \left(\sum_{j=1}^{j_*} g_j q_{ij}^2 \right) = \sum_{j=1}^{j_*} \sum_{i=1}^{i_*} g_j q_{ij}^2 \tag{1.4}$$

where $q_{ij} \in \mathrm{P}^n[\mathbf{x}]$ for all i, j and where the maximal number i_* of squares is specified later. Of course all polynomials (1.4) belong to $\mathbf{P}^m_{\mathbb{K},+}[\mathbf{x}]$. An elementary exemple which will have its importance in the core of this work is the unit constant polynomial $e(\mathbf{x}) := 1$. It admits the representation (1.4) by taking $i_* = 1$ and $q_{1j}(\mathbf{x}) = 1$ for all j. Indeed one has the identity

$$\sum_{j=1}^{j_*} g_j(\mathbf{x}) = \sum_{i=1}^{d} \sum_{j_i \in \{0,1\}} \Pi_{i=1}^d x_i^{j_i}(1-x_i)^{1-j_i} = \Pi_{i=1}^d \left(x_i + (1-x_i) \right) = 1^d = 1 = e(\mathbf{x}). \tag{1.5}$$

To treat more general polynomials, we consider the canonical basis made of monomials with the multi-index notation $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, $|\alpha| = \max(\alpha_1, \ldots, \alpha_d)$ and $\mathbf{x}^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$. The polynomials have the expansion $q_{ij}(\mathbf{x}) = \sum_{|\alpha| \leqslant n_j} c^{ij}_\alpha \mathbf{x}^\alpha$ and one can store the coefficients in a vector of coefficients written as $c^{ij} = (c^{ij}_\alpha)_\alpha \in \mathbb{R}^{(n+1)^d}$. We gather the coefficients c^{i1}, c^{i2}, ..., $c^{i,j*}$ in a single column vector (called a Cholesky factor) $\mathbf{U}_i = \left(c^{i1}, c^{i2}, \ldots, c^{i,j*} \right)^t \in \mathbb{R}^{j*(n+1)^d}$. The Hankel matrix $D \in \mathcal{M}_{(n+1)^d}(\mathbb{R})$ is defined as $D(\mathbf{x})_{\alpha,\beta} = \mathbf{x}^\alpha \mathbf{x}^\beta$ for $|\alpha|, |\beta| \leqslant n$. We define the block polynomial-valued symmetric matrix $B(\mathbf{x}) = B(\mathbf{x})^t \in \mathcal{M}_{j_*(n+1)^d}(\mathbb{R})$

$$B(\mathbf{x}) = \mathrm{diag}\left(g_1(\mathbf{x})D(\mathbf{x}), , \ldots, g_{j_{j_*}}(\mathbf{x})D(\mathbf{x}) \right). \tag{1.6}$$

This matrix is a block diagonal localizing matrix [10]. The first square diagonal block is $g_1(\mathbf{x})D(\mathbf{x})$ until the last square diagonal block which is $g_{j_*}(\mathbf{x})D(\mathbf{x})$. The block non diagonal terms are set to zero. With these notations the right hand side of (1.4) can be written conveniently as

$$\sum_{i=1}^{i_*} \left(\sum_{j=1}^{j_*} g_j(\mathbf{x}) q_{ij}^2(\mathbf{x}) \right) = \sum_{i=1}^{i_*} \langle B(\mathbf{x})\mathbf{U}_i, \mathbf{U}_i \rangle. \tag{1.7}$$

Next we follow [5] by introducing unisolvent interpolation points $(\mathbf{x}_r)_{1\leqslant r\leqslant r_*}$ which help to achieve further localization. A unisolvent set of points $(\mathbf{x}_r)_{1\leqslant r\leqslant r_*}$ is such that any polynomial $p \in \mathrm{P}^n[\mathbf{X}]$ is uniquely determined by its values $y_r = p(\mathbf{x}_r)$ for all r. In the case of the hypercube, the points will constructed by tensorization and so will naturally be unisolvent. By definition the number of these points will be equal to $r_* = \dim\left(\mathbf{P}^{2n+1}[\mathbf{x}]\right) = (2n+2)^d$. Note that with our previous notations one has

$$r* = (2n+2)^d = j_*(n+1)^d,$$

this is why we will use only one notation for r_* and for $j_*(n+1)^d$ even if it can be different quantities in a more general case. The evaluation of $B(\mathbf{x})$ at interpolation points is denoted as $B_r = B(\mathbf{x}_r) \in \mathcal{M}_{r_*}(\mathbb{R})$. The evaluation of $p(\mathbf{x})$ at interpolation points is denoted as $y_r = p(\mathbf{x}_r)$. Let us define the domain

$$\mathcal{D} = \left\{ \lambda \in \mathbb{R}^{r*} \text{ such that } I + \sum_{r=1}^{r_*} \lambda_r B_r > 0 \right\} \subset \mathbb{R}^{r*}. \tag{1.8}$$

Clearly $\mathcal{D} \neq \varnothing$ since it contains at least a small ball centered on $\lambda = \mathbf{0}$.

Definition 1.1. *The function $\lambda \mapsto G(\lambda)$ is constructed as follows:*
if $\lambda \in \mathcal{D}$ we set $G(\lambda) = tr\left[\left(I + \sum_{r=1}^{r_*} \lambda_r B_r\right)^{-1}\right] + \sum_{r=1}^{r_*} y_r\lambda_r$;
if $\lambda \notin \mathcal{D}$ we set $G(\lambda) = +\infty$.

The value of $G(\lambda)$ is the sum of the inverse of the eigenvalues of the matrix $I + \sum_{r=1}^{r_*} \lambda_r B_r$, plus a linear contribution which depends on the value of the polynomial p at data points. This function is obtained in [5] as a dual formulation of (1.4). Two fundamental and general properties follow.

Lemma 1.2. *Assume G has an extremal point in $\mathcal{D}$. Then p has a SOS representation.*

Proof. To synthesize the notations, we introduce the matrix valued function $\lambda \in \mathbb{R}^{r*} \mapsto M(\lambda) = I + \sum_{r=1}^{r_*} \lambda_r B_r$ and the scalar product between real vectors $\langle a, b\rangle = \sum_{r=1}^{r_*} a_r b_r$. Then G rewrites as $G(\lambda) = \mathrm{tr}(M(\lambda)^{-1}) + \langle \mathbf{y}, \lambda\rangle$. Using the differential formula $dM^{-1}(\lambda) = -M^{-1}(\lambda)dM(\lambda)M^{-1}(\lambda)$, the gradient of G against a vector $\mu = (\mu_1, \ldots, \mu_{r_*})$ is

$$\langle \nabla G(\lambda), \mu\rangle = -\mathrm{tr}\left(M^{-1}(\lambda) \left(\sum_{r=1}^{r_*} \mu_r B_r \right) M^{-1}(\lambda) \right) + \langle \mathbf{y}, \mu\rangle\,. \tag{1.9}$$

At an extremal point $\lambda_e \in \mathcal{D}$, one has that the gradient vanishes $\nabla G(\lambda_e) = 0$. Let us consider such an extremal point λ_e and set

$$U_e = M(\lambda_e)^{-1} \in \mathcal{M}_{r_*}(\mathbb{R}). \tag{1.10}$$

We also set $\mathbf{U}_i = U_e\mathbf{e}_i \in \mathbb{R}^{r*}$ where the $\mathbf{e}_i \in \mathbb{R}^{r*}$ are the r_* unit vectors in $\mathbb{R}^{r*}$. The gradient identity (1.9) yields that $0 = -\mathrm{tr}(U_e B_r U_e) + y_r = -\sum_{i=1}^{r_*} \langle B_r \mathbf{U}_i, \mathbf{U}_i\rangle + y_r$. This is directly equivalent to the SOS formulas (1.4-1.7) since the points $(\mathbf{x}_r)$ are unisolvent by hypothesis. ∎

Lemma 1.3. *G is convex in $\mathcal{D}$.*

Proof. For $\lambda \in \mathcal{D}$ in the domain (1.8) then $M(\lambda)$ is symmetric, positive and so is invertible. The Hessian matrix of G can be evaluated as

$$\left\langle \nabla^2 G(\lambda)\mu, \mu\right\rangle = \mathrm{tr}\left(M^{-1}(\lambda) \left(\sum_{r=1}^{r_*} \mu_r B_r \right) M^{-1}(\lambda) \left(\sum_{r=1}^{r_*} \mu_r B_r \right) M^{-1}(\lambda) \right).$$

Since $M(\lambda)$ is positive over $\mathcal{D}$, then $\left\langle \nabla^2 G(\lambda)\mu, \mu\right\rangle \geqslant 0$ for all μ, therefore G is convex over $\mathcal{D}$. ∎

The convex function G depends on three ingredients which are the degree $m = 2n+1$, the position of the interpolation points $(\mathbf{x}_r)_{1\leqslant r\leqslant r_*}$ and the polynomial p. The minimum of p over $\mathbb{K}$ will be denoted as $p_- = \min_{\mathbf{x}\in\mathbb{K}} p(\mathbf{x}) > 0$. It is now possible to state the main result of this work.

THEOREM 1.4. *Take the* $(\mathbf{x}_r)_{1\leqslant r\leqslant r_*}$ *obtained by tensorization of the points* $\frac{1}{2}\left(1+\cos\frac{g}{2n+1}\pi\right) \in [0,1]$ *for* $0 \leqslant g \leqslant 2n+1$*, that is* $\{\mathbf{x}_r\}_{1\leqslant r\leqslant r_*} = \left\{\frac{1}{2}\left(1+\cos\frac{g}{2n+1}\pi\right)\right\}^d_{0\leqslant g\leqslant 2n+1}$.
Then these points are nearly optimal, in the sense that there exists a constant $C_d > 0$ *such that if a polynomial* $p \in \mathbf{P}^m_{\mathbb{K},+}[\mathbf{x}]$ *satisfies the inequality* $\frac{C_d\|p\|_{W^{2,\infty}(\mathbb{K})}(1+\log m)^{2d}}{m^2} \leqslant p_-$*, then* G *is coercive (infinite at infinity), has a minimum in* $\mathcal{D}$ *and* p *admits the SOS representation (1.4) given by (1.10).*

Some points are worthwhile to comment. The first point is that even if the result has a flavor of similar results obtained in real algebraic geometry [20, 15], the proof proposed in this work is a combination of convex analysis in finite dimension with purely analytical techniques to prove the main stability inequality. The second point is that the main stability inequality is inherently attached to interpolation points based on Chebyshev polynomials. More precisely the interpolation are constructed from the roots of the third and fourth kind Chebyshev polynomials of degree n. The $\log m$ term in the Theorem directly comes from the Lebesgue stability constant of Chebyschev interpolation technique [7]. The third point is that the condition of the Theorem is non optimal by a factor $(\log m)^{2d}$ with respect to the estimate obtained recently in [13] for an effective version of Schmüdgen's Positivstellensatz for the hypercube (this is why the points are said to be nearly optimal). Our interpretation is that it is linked to the numerical construction of the method with interpolation points $\mathbf{x}_r$. It is also non optimal in dimension $d = 1$ because one can prove directly the scaling $\log m$ instead of $(\log m)^2$, see Remark 6.1. An illustration is proposed in Figure 1.

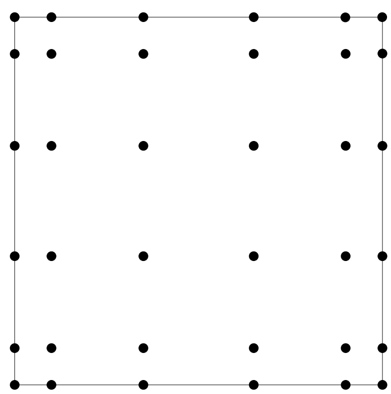

FIG. 1. *Interpolation points in dimension* $d = 2$ *for* $m = 5$

The proof is organized as follows. In Section 2, we provide more explanations on the function G. Section 3 is dedicated a simple inequality that a polynomial p should satisfy so that G is coercive. Then in Section 4 and 5, we prove the stability estimates required for the inequality of Section 3 to be strict. The end of the proof of the Theorem is in Section 6, together with additional final remarks. For the sake of the completeness of this work, the appendix gathers basic facts about Chebyshev polynomials and Chebyshev interpolation.

2. More properties of the function G. The function G is defined in [5] by a duality argument which is standard [1, 9, 2, 14, 8]. The strategy explored in this work is to find conditions such that G is coercive (infinite at infinity). This is a common strategy in convex analysis in finite dimension

[9], nevertheless it seems original with respect to the literature [1, 9, 2, 14, 8], except [5] from which this strategy is originated when applied to the calculation of SOS from data points. The proof can be decomposed in the verification of two separate properties. The first property consists in showing that $G(\lambda) \to +\infty$ when $\lambda \to \sigma \in \partial\mathcal{D}$ is a point that tends to σ which is on the boundary at finite distance of the domain. This is the easy part.

Lemma 2.1 (First property). *$G(\lambda) \to +\infty$ when $\lambda \to \sigma \in \partial\mathcal{D}$.*

Proof. Since $\sigma \in \partial\mathcal{D}$ then $M(\sigma) \geq 0$ is a non negative matrix with a zero eigenvalue. By continuity of the eigenvalues of $M(\lambda)$ for $\lambda \in \mathcal{D}$, then at least one eigenvalue tends to zero (being that all eigenvalues are positive). Since the linear part is bounded for λ in the vicinity of σ, the dominant term in $G(\lambda)$ is the sum of the inverse of the eigenvalues, which tends to $+\infty$. This is the claim. $\square$

To formulate the second property which is the involved part of the proof, we need the cone at infinity

$$\mathcal{C} = \left\{ \delta \in \mathbb{R}^{r_*} \text{ such that } \sum_{r=1}^{r_*} \delta_r B_r \geq 0. \right\}. \tag{2.1}$$

The cone at infinity is the ensemble of directions δ such that $M(t\delta) > 0$ for all $t \geqslant 0$. This set is closed. We will also write $\mathcal{C}^* = \{\delta \in \mathcal{C} \text{ such that } \delta \neq 0\}$.

Definition 2.2 (Second property). *It writes as: $G(t\delta) \to +\infty$ when $t \to +\infty$ and $\delta \in \mathcal{C}^*$.*

It is a natural and basic fact in convex analysis in finite dimension that a convex function that satisfies the first and second properties has an extremal point in its domain [21] (an additional requirement is that the function is proper and closed, which is evident in our case [21, 9]). So our efforts in this work are now focused on the establishment of the second property.

Lemma 2.3. *The second property is equivalent to: $\delta \in \mathcal{C}^* \Longrightarrow \langle \mathbf{y}, \delta \rangle > 0$.*

Proof. The proof is in two parts.
$\Longrightarrow$: Assume that $G(t\delta) \to +\infty$ where $t \to +\infty$ and $\delta \in \mathcal{C}^*$. For $t > 0$ and $\delta \in \mathcal{C}$, then $I \prec M(t\delta)$ so $M(t\delta)^{-1}$ is bounded uniformly with respect to $t \in [0, \infty)$. That is $0 < \mathrm{tr}(M(t\delta)^{-1}) < I$. It yields that $\lim_{t\to+\infty} t \langle \mathbf{y}, \delta \rangle = +\infty$ therefore $\langle \mathbf{y}, \delta \rangle > 0$.
$\Longleftarrow$: The reciprocal part is immediate since $t \mapsto G(t\delta)$ is the sum of a non negative part $\mathrm{tr}(M(t\delta)^{-1})$ and of a linear part. Since the linear part tends to infinity for large t, then $G(t\delta)$ tends to infinity as well. $\square$

If one replaces the strict inequality with a large inequality one obtains a weaker condition.

Definition 2.4 (Third property). *It writes as: $\delta \in \mathcal{C} \Longrightarrow \langle \mathbf{y}, \delta \rangle \geqslant 0$.*

Lemma 2.5. *The third property holds if and only if $p \in \mathbf{P}^m(\mathbf{x})$ characterized by $p(\mathbf{x}_r) = y_r$ for $1 \leqslant r \leqslant r_*$ admits the representation (1.4).*

Proof. As shown by the generalized Farkas lemma [9], proving the third property is equivalent to proving (1.4). Since it is very classical we will not reproduce the analysis (one can refer for instance to [5, proof of Proposition 3.4]). $\square$

Clearly the second property implies the third one, and there seems to be just a minor difference between the two ones. In the context of this work, they are actually very different. The second property immediately yields the existence of a critical point λ_e which can be calculated with the gradient descent numerical algorithms detailed in [5]. On the contrary, with the third one, the existence of λ_e is not guaranteed (or perhaps λ_e is a point at infinity): the generalized Farkas lemma that one can invoke is only an equivalence principle, so it is not of immediate help to prove existence of a critical point.

3. An interesting inequality. To prove the second property, let us analyze

$$\langle \mathbf{y}, \delta\rangle = \sum_{r=1}^{r*} p(\mathbf{x}_r)\delta_r \text{ where } \delta \in \mathcal{C}.$$

We remind the reader that p_- is the minimum of p over $\mathbb{K}$.

LEMMA 3.1. *There exists a constant $C_d > 0$ such that one has the inequality*

$$\langle \mathbf{y}, \delta\rangle \geqslant p_- \sum_{r=1}^{r*} \delta_r - \frac{C_d\|p\|_{W^{2,\infty}(\mathbb{K})}}{m^2} \sum_{r=1}^{r*} |\delta_r|. \tag{3.1}$$

Proof. One has the decomposition $p = p_- + q$ where $q \in \mathbf{P}^m_{\mathbb{K},+}[\mathbf{x}]$. Since q is non negative over $\mathbb{K}$, it is convenient to perform an approximation with any kind of method which preserves the polynomial structure and the non negativity. Such a method could be approximation multidimensional Bernstein polynomials $b_\alpha(\mathbf{x}) = \Pi_{j=1}^d \frac{m!}{\alpha_j!(m-\alpha_j)!} x_j^{\alpha_1}(1-x_j)^{m-\alpha_1}$ where $\alpha = (\alpha_1, \dots, \alpha_d)$ and $0 \leqslant \alpha_j \leqslant m$ for all j. However this is non optimal. Indeed the error is smaller if one uses tensorization of Jackson kernel [7].

Firstly we use a technical idea from the recent work [13][Theorem 8 and Lemma 9] that we detail in the appendix. We modify a little the proof of [13] because we use the characterization $p(x) = xa^2(x) + (1-x)b^2(x)$ of polynomials $p \in \mathbf{P}_{[0,1],+}[x]$ over $[0,1]$ with a, b polynomials of convenient orders (instead of $p(x) = a^2(x) + x(1-x)b^2(x)$ as in [13]). It yields that the approximation of q with the Jackson kernel method is denoted as $\widetilde{q}$ and it is a sum of squares by construction [13]. One obtains

$$\delta \in \mathcal{D} \Longrightarrow \sum_{r=1}^{r*} \widetilde{q}(\mathbf{x}_r)\delta_r \geqslant 0. \tag{3.2}$$

Secondly the error estimate of this approximation can be taken in 1D from [7][Theorem 2.2], [25] or [6][Exemple 2.2]. One uses the standard substitution $f(t) = p(\cot t)$ to transfer to algebraic prolynomials the approximation properties in L^∞ norm already proved for trigonometric polynomials. The multidimensional generalization by tensorization is immediate and detailed in the appendix. One obtains

$$\|q - \widetilde{q}\|_{L^\infty(\mathbb{K})} \leqslant \frac{C_d}{m^2}\|p\|_{W^{2,\infty}(\mathbb{K})}. \tag{3.3}$$

Then, for $\delta \in \mathcal{C}$, one can write using (3.2) and (3.3)

$$\begin{array}{rcl}\langle \mathbf{y}, \delta\rangle & = & p_- \sum_{r=1}^{r*} \delta_r + \sum_{r=1}^{r*} q(\mathbf{x}_r)\delta_r \\ & = & p_- \sum_{r=1}^{r*} \delta_r + \sum_{r=1}^{r*} (q-\widetilde{q})(\mathbf{x}_r)\delta_r + \sum_{r=1}^{r*} \widetilde{q}(\mathbf{x}_r)\delta_r \\ & \geqslant & p_- \sum_{r=1}^{r*} \delta_r - \frac{C_d}{m^2}\|p\|_{W^{2,\infty}(\mathbb{K})} \sum_{r=1}^{r*} |\delta_r| \end{array}$$

which is the claim. ∎

Considering the inequality (3.1), our goal is to show firstly that $\sum_{r=1}^{r*} \delta_r > 0$ for non zero $\delta \neq 0$, and secondly that $\sum_{r=1}^{r*} |\delta_r|$ is controlled by $\sum_{r=1}^{r*} \delta_r$ uniform with respect to δ.

4. Study of $\sum_{r=1}^{r*} \delta_r$. In view of Lemma 2.5 and of the fact that the unit polynomial $e(\mathbf{x}) := 1$ is a SOS (1.5), it is immediate that $\sum_{r=1}^{r*} \delta_r \geqslant 0$. This is why our objective hereafter is to show $\sum_{r=1}^{r*} \delta_r = 0$ is not possible. The proposed proof is by contradiction.

LEMMA 4.1. *Take $\delta \in \mathcal{C}$ such that $\sum_{r=1}^{r*} \delta_r = 0$. Then $\sum_r \delta_r \mathbf{x}_r^\alpha = 0$ for all $|\alpha| \leqslant m$.*

Proof. The reasoning is performed one step after the other.
• By definition of the localizing matrix (1.6) and of the cone at infinity (2.1), any $\delta \in \mathcal{C}$ satisfies

$$\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r) p(\mathbf{x}_r)^2 \geqslant 0 \quad \forall p \in \mathbf{P}^n[\mathbf{x}], \quad \forall \mathbf{j} \in \{0,1\}^d. \tag{4.1}$$

Next we use particular test polynomials wth the SOS structure (1.4). Taking all q_{ij} equal to zero except one of them equal to one, one gets that $p(\mathbf{x}) = g_{\mathbf{j}}(\mathbf{x})$ is a SOS. One gets the inequalities $\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r) \geqslant 0$ for all $\mathbf{j} \in \{0,1\}^d$. Since $\sum_r g_{\mathbf{j}}(\mathbf{x}) \equiv 1$ as shown in (1.5), one obtains $0 \leqslant \sum_{\mathbf{j}\in\{0,1\}^d} \sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r) = \sum_r \delta_r = 0$, therefore

$$\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r) = 0, \text{ for all } \mathbf{j} \in \{0,1\}^d. \tag{4.2}$$

These equalities are the first step of the reasoning below which is by iteration.
• Let make the assumption that

$$\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r)(\mathbf{x}_r^\alpha)^2 = 0 \text{ for all } |\alpha| \leqslant a \tag{4.3}$$

where $a \leqslant n-2$ is a certain value. Note that (4.2) corresponds to $a = 0$. Take

$$p = \mathbf{x}_r^\alpha + \varepsilon \mathbf{x}_r^\alpha b \in \mathbf{P}^n[\mathbf{x}] \tag{4.4}$$

where b is a monomial of degree less or equal to two $b(\mathbf{x}) = \mathbf{x}^{2\beta}$ for $|\beta| \leqslant 1$. Then (4.1) and (4.3) in the limit of small $|\varepsilon|$ yields $\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r)(\mathbf{x}_r^\alpha)^2(\mathbf{x}_r^\beta)^2 = 0$ for all $|\alpha| \leqslant a$ and all $|\beta| \leqslant 1$. This is exactly the starting assumption (4.3), but now $|\alpha| \leqslant a+1$. By iterations, one obtains (4.3) for $a = n-1$.
• Next we modify the previous analysis with (4.4), but now b has the form $b(\mathbf{x}) = \mathbf{x}^\beta$ for $|\beta| \leqslant 1$. It yields $\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r)(\mathbf{x}_r^\alpha)^2 \mathbf{x}_r^\beta = 0$ for all $|\alpha| \leqslant n-1$ and all $|\beta| \leqslant 1$. It can be rewritten as $\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r)\mathbf{x}_r^\alpha = 0$ for all $|\alpha| \leqslant 2n-1$. Since the functions $g_{\mathbf{j}}$ generate by linear combination the monomials $\mathbf{x}_r^\gamma$ for all $|\gamma| \leqslant 1$, one obtains

$$\sum_r \delta_r \mathbf{x}_r^\alpha = 0 \text{ for all } |\alpha| \leqslant 2n. \tag{4.5}$$

This almost the claim since $2n = m-1$.
• By (4.1), one has the inequalities $\sum_r \delta_r g_{\mathbf{j}}(\mathbf{x}_r)(\mathbf{x}_r^\alpha)^2 \geqslant 0$ for all $|\alpha| \leqslant n$. Take γ such that $|\gamma| \leqslant 1$. It is possible to find a linear combination with non negative weights $w_j(\gamma) \geqslant 0$ such that $\mathbf{x}_r^\gamma = \sum_j w_j(\gamma) g_{\mathbf{j}}(\mathbf{x}_r)$. One gets

$$\sum_r \delta_r \mathbf{x}_r^\gamma (\mathbf{x}_r^\alpha)^2 \geqslant 0 \text{ for all } |\alpha| \leqslant n. \tag{4.6}$$

It is also possible to find a linear combination with non negative weights $z_j(\gamma) \geqslant 0$ such that $1 - \mathbf{x}_r^\gamma = \sum_j z_j(\gamma) g_{\mathbf{j}}(\mathbf{x}_r)$. One gets $\sum_r \delta_r (1 - \mathbf{x}_r^\gamma)(\mathbf{x}_r^\alpha)^2 \geqslant 0$ for all $|\alpha| \leqslant n$. Considering (4.5), one gets

$$-\sum_r \delta_r \mathbf{x}_r^\gamma (\mathbf{x}_r^\alpha)^2 \geqslant 0 \text{ for all } |\alpha| \leqslant n. \tag{4.7}$$

Finally (4.6) and (4.7) yield $\sum_r \delta_r(\mathbf{x}_r)\mathbf{x}_r^\alpha = 0$ for all $|\alpha| \leqslant 2n+1$ which is the claim. ∎

Lemma 4.2. *Assume the interpolation points are obtained by tensorization of $m+1 = 2n+2$ one-dimensional points. Take $\delta \in \mathcal{C}^*$. Then $\sum_r \delta_r > 0$.*

Proof. One-dimensional interpolation points are denoted as $y_0 < y_1 < \cdots < y_{m+1}$. The tensorization yield $(m+1)^d$ points

$$\mathbf{x_j} = (y_{\mathbf{j}_1}, \ldots, y_{\mathbf{j}_d}) \text{ for } \mathbf{j} \in \mathbb{N}^d \text{ with } |\mathbf{j}| \leqslant m+1$$

which are unisolvent in the hypercube. In particular, if $\sum_r \delta_r = 0$, then the result of Lemma 4.1 yields that $\delta_r = 0$ for all $1 \leqslant r \leqslant r_*$. This is not possible. So it yields the claim by contradiction. ∎

5. Study of $\sum_{r=1}^{r*} |\delta_r|$. Let us study

$$A(\delta, \varepsilon) = \sum_{r=1}^{r*} \delta_r - \varepsilon \sum_{r=1}^{r*} |\delta_r|, \quad \varepsilon > 0 \tag{5.1}$$

where we already know that $A(\delta, 0) > 0$ for $\delta \in \mathcal{C}^*$. Since our goal is to compare $\sum_{r=1}^{r*} |\delta_r|$ and $\sum_{r=1}^{r*} \delta_r$, it is natural to seek a bound on ε such that $A(\delta, \varepsilon) \geqslant 0$.

We will consider tensorized interpolation points for which the Lagrange polynomials $l_r \in \mathbf{P}^m[\mathbf{x}]$ are correctly defined. These Lagrange polynomials are characterized by

$$l_r(\mathbf{x}_s) = \delta_{rs}. \tag{5.2}$$

Lemma 5.1. *One has that $A(\delta, \varepsilon) \geqslant 0$ for all $\delta \in C$ if and only if*

$$q(\mathbf{x}) = 1 - \varepsilon z(\mathbf{x}) \tag{5.3}$$

with $z(\mathbf{x}) = \sum_{r=1}^{r*} l_r(\mathbf{x}) sign(\delta_r)$ can be written under the form (1.4).

Proof. Apply Lemma 2.5. ∎

Remark 5.2 (Why Lemma 5.1 is the pivotal point of the strategy of proof). This Lemma offers the opportunity to establish a connection between the construction of SOS and the stability properties of Chebyschev polynomials. In particular, in dimension $d = 1$, it is evident that if one chooses the interpolation points to be equal to the Chebyschev interpolation points, then the polynomial z is bounded by the Lebesgue stability constant $O(\log m)$ of Chebyschev interpolation [7]. Then taking ε small enough guarantees that q is non negative over $\mathbb{K} = [0, 1]$, which turns into the fact that q is a SOS by the Lukacs Theorem [24]. The smallness condition on ε can be written as

$$\varepsilon < \frac{C}{1 + \log m} \tag{5.4}$$

Another interpretation of the Lemma is that q is the perturbation of the polynomial $\mathbf{x} \mapsto 1$ with a corrector term equal to $\varepsilon z(\mathbf{x})$. Since 1 is evidently a SOS, it is natural to think that a perturbation technique could be used to establish directly that q is a SOS. To show that such a connection between Lemma 5.1 and Chebyschev polynomials holds in any dimension, we will make use of the Chebyshev polynomials of the third and fourth kind [16, 17, 3].

Definition 5.3. *The Chebyshev polynomials of the third kind $a_n \in P^n[x]$ and of the fourth kind $b_n \in P^n[x]$ are defined by $a_n(x) = \frac{\cos\left(n+\frac{1}{2}\right)\theta}{\cos\frac{\theta}{2}}$ and $b_n(x) = \frac{\sin\left(n+\frac{1}{2}\right)\theta}{\sin\frac{\theta}{2}}$ where $x = \frac{1+\cos\theta}{2} \in [0, 1]$.*

These Chebyshev polynomials have degree equal to n. They satisfy the symmetry identity $b_n(x) = a_n(1-x)$ and the identity

$$x a_n(x)^2 + (1-x) b_n(x)^2 = 1. \tag{5.5}$$

The polynomial $xa_n(x)$ has n distinct roots $\alpha_k = \frac{1+\cos\theta_k}{2}$ where $\theta_k = \frac{2k+1}{2n+1}\pi$ for $0 \leqslant k \leqslant n$, that is $\alpha_k a_n(\alpha_k) = 0$ for $0 \leqslant k \leqslant n$. The roots of $(1-x)b_n(x)$ are deduced by symmetry. There are denoted as $\beta_l = \frac{1-\cos\theta_l}{2}$ where $\theta_l = \frac{2l}{2n+1}\pi$ for $0 \leqslant l \leqslant n$. These roots (α_k) interlace with the roots (β_l) and the ensemble of all roots is

$$\{\alpha_k\}_{0\leqslant k\leqslant n} \bigcup \{\beta_l\}_{0\leqslant l\leqslant n} = \left\{\frac{1+\cos\frac{g}{2n+1}\pi}{2}\right\}_{0\leqslant g\leqslant 2n+1}. \tag{5.6}$$

The Lagrange interpolation polynomials based on the α_k for $0 \leqslant k \leqslant n$ are

$$l_k(x) = \frac{\Pi_{0\leqslant s\neq k\leqslant n}(x-\alpha_s)}{\Pi_{0\leqslant s\neq k\leqslant n}(\alpha_k-\alpha_s)}$$

.

Lemma 5.4 (Proof in the appendix). *The Lagrange interpolation polynomials write as*

$$l_k(x) = (-1)^k \frac{\gamma_k \sin\frac{\theta_k}{2}\cos\left(n+\frac{1}{2}\right)\theta\cos\frac{\theta}{2}}{(2n+1)\left(\cos\theta-\cos\alpha_k\right)}, \quad \textit{where } \gamma_k = \begin{cases} 4 & \textit{for } 0\leqslant k\leqslant n-1, \\ 2 & \textit{for } k=n. \end{cases}.$$

For notational convenience, let us define $h_0(x) = x$, $h_1(x) = 1-x$ and

$$h_{\mathbf{j}}(\mathbf{x}) = \Pi_{i=1}^d a_n\left(h_{\mathbf{j}_i}(x_i)\right) \quad \text{for } |\mathbf{j}| \leqslant 1.$$

If $\mathbf{j}_i = 0$ then $a_n\left(h_{\mathbf{j}_i}(x_i)\right) = a_n(x_i)$. If $\mathbf{j}_i = 1$ then $a_n\left(h_{\mathbf{j}_i}(x_i)\right) = a_n(1-x_i) = b_n(x_i)$.

Lemma 5.5. *One has the identity* $\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})h_{\mathbf{j}}(\mathbf{x})^2 = 1$.

Proof. It is the multiplication of (5.5) for all directions, that is for $x = x_1$ to $x = x_d$. ∎

For solving (5.3), we make a perturbation of the identity of Lemma 5.5 and consider the equation

$$\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})\left(h_{\mathbf{j}}(\mathbf{x}) + u_{\mathbf{j}}(\mathbf{x})\right)^2 = q(\mathbf{x}), \tag{5.7}$$

where q the given right hand side and the polynomials $(u_{\mathbf{j}})_{|\mathbf{j}|\leqslant 1} \in \mathbf{P}^n[\mathbf{x}]^{2^d}$ are the unknowns. The equation is equivalent to

$$\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})h_{\mathbf{j}}(\mathbf{x})u_{\mathbf{j}}(\mathbf{x}) = -\frac{\varepsilon}{2}z(\mathbf{x}) - \frac{1}{2}\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})u_{\mathbf{j}}(\mathbf{x})^2. \tag{5.8}$$

The structure of this equation is interesting because the polynomials $h_{\mathbf{j}}(\mathbf{x})$ oscillate a lot since they are constructed from Chebyshev polynomials. Usually, too much oscillations in the coefficients is a factor that deteriorates our ability to solve an equation. In this case, we will see that it is the opposite in the sense that the oscillations allow us to solve the equation (5.8).

Lemma 5.6. *Let consider that the interpolation points $(\mathbf{x}_r)_{r=1}^{r_*}$ are constructed from the tensorization of the $2n+2$ points (5.6). Let $b \in \mathbf{P}^{2n+1}[\mathbf{x}]$. Then the equation*

$$\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})h_{\mathbf{j}}(\mathbf{x})u_{\mathbf{j}}(\mathbf{x}) = b(\mathbf{x})$$

has a unique solution $(u_{\mathbf{j}})_{|\mathbf{j}|\leqslant 1} \in \mathbf{P}^n[\mathbf{x}]^{2^d}$ which satisfies the bound

$$\max_{|\mathbf{j}|\leqslant 1}\left\|\sqrt{g_{\mathbf{j}}}u_{\mathbf{j}}\right\|_{L^\infty(\mathbb{K})} \leqslant C_d(\log n)^d \max_{1\leqslant r\leqslant r_*}|b(\mathbf{x}_r)|.$$

Proof. Both sides of the equation are polynomials of degree $\leqslant 2n+1$, so the equality is equivalent to point wise equalities at the interpolation points

$$\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x}_r)h_{\mathbf{j}}(\mathbf{x}_r)u_{\mathbf{j}}(\mathbf{x}_r) = b(\mathbf{x}_r), \qquad 1 \leqslant r \leqslant r_*. \tag{5.9}$$

Since $\mathbf{x}_r$ is build from tensorization of roots of Chebyshev polynomials, many terms vanish on the left hand side of (5.9). Considering for exemple the points $(\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_d})$ for $0 \leqslant i_1, \dots, i_d \leqslant n$, all terms but one vanish on the left hand side. What remains writes as

$$(1-\alpha_{i_1})a_n(1-\alpha_{i_1})\times\cdots\times(1-\alpha_{i_d})a_n(1-\alpha_{i_d})\times u_{\mathbf{1}}(\alpha_{i_1},\dots,\alpha_{i_d}) = b(\alpha_{i_1},\dots,\alpha_{i_d}) \text{ for } 0 \leqslant i_1,\dots,i_d \leqslant n,$$

that is (the notation is $\mathbf{1} = (1, \dots, 1)$)

$$u_{\mathbf{1}}(\alpha_{i_1},\dots,\alpha_{i_d}) = \frac{b(\alpha_{i_1},\dots,\alpha_{i_d})}{(1-\alpha_{i_1})a_n(1-\alpha_{i_1})\times\cdots\times(1-\alpha_{i_d})a_n(1-\alpha_{i_d})} \text{ for } 0 \leqslant i_1,\dots,i_d \leqslant n.$$

Since $u_{\mathbf{1}} \in \mathbf{P}^n[\mathbf{x}]$, one can calculate $u_{\mathbf{1}}$ with tensorization of Lagrange polynomials l_k. One gets

$$u_{\mathbf{1}}(\mathbf{x}) = \sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n} \frac{b(\alpha_{i_1},\dots,\alpha_{i_d})\ l_{i_1}(x_1)\dots l_{i_d}(x_d)}{(1-\alpha_{i_1})a_n(1-\alpha_{i_1})\times\cdots\times(1-\alpha_{i_d})a_n(1-\alpha_{i_d})}, \quad \mathbf{x} = (x_1,\dots,x_d).$$

It is convenient to rescale this expression as

$$\begin{aligned} &\sqrt{1-x_1}\dots\sqrt{1-x_d}u_{\mathbf{1}}(\mathbf{x}) \\ &= \sum_{i_1=1}^{n}\cdots\sum_{i_d=1}^{n} \frac{\sqrt{1-x_1}l_{i_1}(x_1)}{(1-\alpha_{i_1})a_n(1-\alpha_{i_1})}\times\cdots\times\frac{\sqrt{1-x_d}l_{i_d}(x_d)}{(1-\alpha_{i_d})a_n(1-\alpha_{i_d})}\times b(\alpha_{i_1},\dots,\alpha_{i_d}) \end{aligned} \tag{5.10}$$

Using (A.6) one gets

$$\left|\frac{\sqrt{1-x_1}l_{i_1}(x_1)}{(1-\alpha_{i_1})a_n(1-\alpha_{i_1})}\right|\times\cdots\times\left|\frac{\sqrt{1-x_d}l_{i_d}(x_d)}{(1-\alpha_{i_d})a_n(1-\alpha_{i_d})}\right| \leqslant \frac{C}{|i_1-l_1|+1}\times\cdots\times\frac{C}{|i_d-l_d|+1}$$

where $\theta_{l_d} \leqslant x_d \leqslant \theta_{l_d+1}$ for all $1 \leqslant l \leqslant d$. This inequality is used in (5.10) with summation over all $1 \leqslant i_1, \dots, i_d \leqslant n$. One gets (with a non optimal manipulations)

$$\left|\sqrt{1-x_1}\dots\sqrt{1-x_d}u_{\mathbf{1}}(\mathbf{x})\right| \leqslant \left(\sum_{i_1=1}^{n}\frac{C}{i_1+1}\right)^d \max_{1\leqslant r\leqslant r_*}|b(\mathbf{x}_r)| \leqslant \widetilde{C}(\log n)^d \max_{1\leqslant r\leqslant r_*}|b(\mathbf{x}_r)|.$$

It is the claim after summation over the 2^d-1 other values of $\mathbf{j}$. ∎

LEMMA 5.7. *There exists $K_d > 0$ such that if $\varepsilon < \min\left(K_d(\log n)^{-2d}, \frac{1}{2}\right)$, then the equation (5.8) has a solution.*

Proof. This can be proved with a classical fixed point method which writes

$$\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})h_{\mathbf{j}}(\mathbf{x})u_{\mathbf{j}}^{k+1}(\mathbf{x}) = -\frac{\varepsilon}{2}z(\mathbf{x}) - \frac{1}{2}\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})u_{\mathbf{j}}^{k}(\mathbf{x})^2, \qquad k = 0, 1, 2, \dots, \tag{5.11}$$

where the next iterate on the left hand side is the collection $(u_{\mathbf{j}}^{k+1})_{|\mathbf{j}|\leqslant 1} \in \mathbf{P}^n[\mathbf{x}]^{2^d}$. The initial value is taken as $u_{\mathbf{j}}^d = 0$ for all $|\mathbf{j}| \leqslant 1$. Note that $\max_{1\leqslant r\leqslant r_*}|z(\mathbf{x}_r)| = 1$ by construction. The fixed point is convergent as easily shown below.

• Let us note $w_k = \max_{|\mathbf{j}|\leqslant 1} \left\| \sqrt{g_{\mathbf{j}}} u_{\mathbf{j}}^k \right\|_{L^\infty(\mathbb{K})}$. Using the previous Lemma, one has $w_0 = 0$ and

$$w_{k+1} \leqslant \left(C_d(\log n)^d/2\right)(\varepsilon + w_k^2).$$

Let us consider the quadratic equation $z = \left(C_d(\log n)^d/2\right)(\varepsilon + z^2)$ which is equivalent to $z^2 - bz + \varepsilon = 0$ with $b = \frac{2}{C_d(\log n)^d}$. The discriminant is $\Delta = b^2 - 4\varepsilon$. It is positive $\Delta > 0$ under the condition of the Lemma. The smallest root is $z_- = \frac{b-\sqrt{\Delta}}{2} = \frac{2\varepsilon}{b+\sqrt{\Delta}} \leqslant \frac{2\varepsilon}{b}$. Now let us assume that $z_k < z_-$ for some $k \geqslant 0$, which is already true for $k = 0$ since $z_0 = 0$. Then $w_{k+1} - z_- \leqslant \left(C_d(\log n)^d/2\right)(w_k^2 - z_-^2) < 0$, therefore the whole sequence is uniformly bounded $w_k < z_-$ for all k.

• To show the convergence, we make the difference of two iterates (5.11)

$$\sum_{|\mathbf{j}|\leqslant 1} g_{\mathbf{j}}(\mathbf{x})h_{\mathbf{j}}(\mathbf{x})\left(u_{\mathbf{j}}^{k+1} - u_{\mathbf{j}}^k\right)(\mathbf{x}) = -\sum_{|\mathbf{j}|\leqslant 1} \frac{1}{2}\sqrt{g_{\mathbf{j}}(\mathbf{x})}\left(u_{\mathbf{j}}^k + u_{\mathbf{j}}^{k-1}\right)(\mathbf{x}) \times \sqrt{g_{\mathbf{j}}(\mathbf{x})}\left(u_{\mathbf{j}}^k - u_{\mathbf{j}}^{k-1}\right)(\mathbf{x}).$$

The norm of the difference is written as $e_k = \max_{|\mathbf{j}|\leqslant 1} \left\| \sqrt{g_{\mathbf{j}}}(u_{\mathbf{j}}^{k+1} - u_{\mathbf{j}}^{k+1}) \right\|_{L^\infty(\mathbb{K})}$. Using the uniform bound and the previous Lemma, one obtains

$$e_{k+1} \leqslant C_d(\log n)^d z_- e_k \leqslant C_d(\log n)^d \frac{2\varepsilon}{C_d(\log n)^d} e_k \leqslant 2\varepsilon e_k.$$

Since $2\varepsilon < 1$ by hypothesis, it establishes the geometric convergence of the sequence which yields a solution to (5.8). ∎

6. End of the proof and final remarks.

Theorem 1.4. The end of the proof of Theorem 1.4 is obtained by requiring that the right hand side of (3.1) is positive. Considering (5.1) and 3.1, one sets $\varepsilon = \frac{C_d\|p\|_{W^{2,\infty}(\mathbb{K})}}{p_- m^2}$. Then Lemma 5.7 yields the condition

$$\varepsilon = \frac{C_d\|p\|_{W^{2,\infty}(\mathbb{K})}}{p_- m^2} < \min\left(K_d(\log n)^{-2d}, \frac{1}{2}\right) < \frac{Q_d}{(1+\log m)^{2d}} \tag{6.1}$$

One recognizes the condition of the Theorem. Then the function G is naturally convex in finite dimension and infinite on the boundary of its domain. It is also infinite at infinity (in the direction of the cone at infinity). It is a basic fact in convex analysis [21, Theorem 27.1-d) page 265] that such a proper closed convex function as a minimum in its domain. The rest of the proof is evident. ∎

Remark 6.1. By comparison of (5.4) and (6.1), the estimate of the Theorem is non optimal in dimension $d = 1$.

Remark 6.2. In the definition of G, one can replace the identity matrix with any symmetric positive matrix. It changes the domain, however is does note change the cone at infinity and this is the main reason why all results generalize to more general functions $G = \mathrm{tr}\left[\left(A + \sum_{r=1}^{r_*} \lambda_r B_r\right)^{-1}\right] + \sum_{r=1}^{r_*} y_r \lambda_r$ where $A > 0$.

Remark 6.3. In this work, the degree of a polynomial is defined as the maximal univariate degree of its constituting monomials. The main advantage is that it is naturally compatible with the tensorization techniques used at different parts of the proof. However it is probably not necessary.

Remark 6.4. A fully open problem is the generalization to the Putinar's Positivstellensatz.

Acknowledgements The author deeply thanks Maxime Herda for valuable remarks which helped to improve this work.

Appendix A. Technical material on polynomials. The technical material is displayed for the sake of completeness of this work.

A.1. Proof of (3.2)-(3.3). Let $q \in \mathbf{P}^m_{\mathbb{K},+}[\mathbf{x}]$ and make the change of variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{T}_d := [0, 2\pi]^d$ with $x_i = \frac{1+\cos\theta_i}{2} \in [0,1]$ for $1 \leqslant i \leqslant d$. We consider $r(\boldsymbol{\theta}) = q(\mathbf{x})$ which a non negative trigonometric polynomial defined in the torus $\mathbb{T}_d = [0, 2\pi]^d$ in dimension d. By construction it is even with respect to all variables

$$r(\theta_1, \theta_1, \ldots, \theta_d) = r(\pm\theta_1, \pm\theta_2, \ldots, \pm\theta_d) \qquad \text{for all } (\theta_1, \ldots, \theta_d) \in \mathbb{T}_d.$$

One will make use of the monovariate Jackson kernel [7] which is

$$K_n(\theta) = \lambda_m \left(\frac{\sin r\theta/2}{\sin\theta/2}\right)^4, \quad r = \left[\frac{n}{2}\right], \quad \int_0^{2\pi} K_n(\theta) d\theta = 1.$$

The Jackson kernel is a non negative trigonometric polynomial

$$K_n(\theta) = \sum_{k=0}^{2r-2} a_{k,n} \cos k\theta = \sum_{k=0}^{2r-2} b_{k,n} \cos^k \theta \geqslant 0 \text{ for all } \theta. \tag{A.1}$$

The Jackson transformed of r is $s(\boldsymbol{\theta}) = \int_{T_d} r(\mu_1, \ldots, \mu_d)\ \Pi_{i=1}^d K_n(\theta_i - \mu_i)\ d\mu_1 \ldots d\mu_d$. Since r is even in all variables, it can be rewritten as

$$s(\boldsymbol{\theta}) = \int_{[0,\pi]^d} r(\mu_1, \ldots, \mu_d)\ \Pi_{i=1}^d \left(K_n(\theta_i - \mu_i) + K_n(\theta_i + \mu_i)\right)\ d\mu_1 \ldots d\mu_d.$$

Thanks to (A.1), the trigonometric polynomial

$$g_\mu(\theta) = K_n(\theta - \mu) + K_n(\theta + \mu)$$

can be expanded as finite linear combination of cosine modes wrt θ multiplied by cosine modes wrt μ. Note r is also finite linear combination of cosine modes wrt μ. We note $g_{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \Pi_{i=1}^d g_{\mu_i}(\theta_i)$ where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ and $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)$, so that one can write

$$s(\boldsymbol{\theta}) = \int_{[0,\pi]^d} r(\boldsymbol{\mu}) g_{\boldsymbol{\mu}}(\boldsymbol{\theta}) d\boldsymbol{\mu}. \tag{A.2}$$

Here we use an idea that we found in [13]. The idea is to replace the integral with a quadrature formula with positive weights and with sufficiently many quadrature points so that it is exact for all modes in the integral (A.2). The result can be expanded as

$$s(\boldsymbol{\theta}) = \sum_c\ w_c\ r(\boldsymbol{\mu}_c)\ \mathbf{g}_{\mu_\mathbf{c}}(\boldsymbol{\theta}) \tag{A.3}$$

with the quadrature points denoted as $\boldsymbol{\mu}_c$, the weights denoted as $w_c > 0$ and the function $g_{\mu_c}(\boldsymbol{\theta})$ being the result of the multiplication of (A.3) for all directions. The approximation of q that we consider is denoted as

$$\widetilde{q}(\mathbf{x}) = s(\boldsymbol{\theta}).$$

Note that $\widetilde{q}$ is the same as the one used in (3.2).

LEMMA A.1. *Take $2r-2=m$. Then $\widetilde{q}$ is correctly defined and $\widetilde{q}\in \mathbf{P}^m[\mathbf{x}]$.*

Proof. With (A.3), then $s(\boldsymbol{\theta})$ can be expanded a polynomial with respect $(\cos\theta_1,\dots,\cos\theta_d) = (2x_1-1,\dots,2x_d-1)$ so $\widetilde{q}$ is indeed a polynomial. By construction $\widetilde{q}$ is of degree m with respect to all variables, so its full degree is m (with the definition of the degree made in this work). ∎

LEMMA A.2 (Proof of (3.2)). *The polynomial $\widetilde{q}$ has the representation (1.4).*

Proof. By construction $g_\mu(\theta)$ is non negative, that is $g_\mu(\theta)\geqslant 0$ for all θ and μ. Thn the real polynomial $k_\mu(x)=g_\mu(2x-1)$ is non negative for $0\leqslant x\leqslant 1$. The Lukacs Theorem [24] states that there exists two polynomials $a_\mu, b_\mu\in P^n[x]$ (where m=2n+1) such that $k_\mu(x)=xa_\mu(x)^2+(1-x)b_\mu(x)^2$. Therefore $g_{\mu_c}(\theta)$ in (A.3) can be expanded as a multiplication of d such terms (one for each direction in $[0,1]^d$) times the coefficient $w_c r(\boldsymbol{\mu}_c)\geqslant 0$. Therefore $\widetilde{q}$ has the representation (1.4). ∎

Next we establish a sharp bound on the difference $q-\widetilde{q}$, which results from a sharp bound on $r-s$. The bound is just a natural tensorization of [7, Theorem 2.2 page 204]. However since the author is not aware of such a result in the literature, the proof is detailed. The tensorization of the identity [7, top of page 203] shows that one has

$$s(\boldsymbol{\theta}) = \int_{[0,\pi]^d}\left(\sum_{(\alpha_1,\dots,\alpha_d)\in\{-1,1\}^d} r(\theta_1+\alpha_1\mu_1,\dots,\theta_d+\alpha_d\mu_d)\right)\Pi_{i=1}^d K_n(\mu_i)\ d\mu$$

The sum is made of 2^d terms. Note that $\int_0^\pi K_n(\mu)d\mu=\frac{1}{2}$. One gets the error formula

$$s(\boldsymbol{\theta})-r(\boldsymbol{\theta}) = \int_{[0,\pi]^d}\left(\sum_{(\alpha_1,\dots,\alpha_d)\in\{-1,1\}^d} r(\theta_1+\alpha_1\mu_1,\dots,\theta_d+\alpha_d\mu_d)-2^d r(\theta_1,\dots,\theta_d)\right)\Pi_{i=1}^d K_n(\mu_i)\ d\mu.$$

Let us set

$$C(\boldsymbol{\mu}) = \max_{\boldsymbol{\theta}\in[0,\pi]^d}\left|\sum_{(\alpha_1,\dots,\alpha_d)\in\{-1,1\}^d} r(\theta_1+\alpha_1\mu_1,\dots,\theta_d+\alpha_d\mu_d)-2^d r(\theta_1,\dots,\theta_d)\right|,\quad \boldsymbol{\mu}\in[0,\pi]^d,$$

and

$$C_i(\sigma) = \max_{\boldsymbol{\theta}\in[0,\pi]^d}\left|\sum_{\alpha_i\in\{1,1\}} r(\theta_1,\dots,\theta_{i-1},\theta_i+\alpha_i\mu,\theta_{i+1},\dots,\theta_d)-2r(\theta_1,\dots,\theta_d)\right|,\quad \sigma\in[0,\pi].$$

LEMMA A.3. *One has $C(\boldsymbol{\mu})\leqslant 2^{d-1}\left(C_1(\mu_1)+\cdots+C_d(\mu_d)\right)$.*

Proof. The result is proved by iteration on with respect to the dimension parameter d.

- For $d=1$, the claim is a triviality.
- Assume the property holds for $d-1\geqslant 1$. One can write

$$\sum_{(\alpha_1,\dots,\alpha_d)\in\{-1,1\}^d}\left(r(\theta_1+\alpha_1\mu_1,\dots,\theta_d+\alpha_d\mu_d)-2^d r(\theta_1,\dots,\theta_d)\right)$$

$$= \sum_{(\alpha_2,\dots,\alpha_d)\in\{-1,1\}^{d-1}}\left(r(\theta_1+\mu_1,\theta_2+\alpha_2\mu_2,\dots)+r(\theta_1-\mu_1,\theta_2+\alpha_2\mu_2,\dots)-2r(\theta_1,\theta_2+\alpha_2\mu_2,\dots)\right)$$

$$+2\sum_{(\alpha_2,\dots,\alpha_d)\in\{-1,1\}^{d-1}}\left(r(\theta_1,\theta_2+\alpha_2\mu_2,\dots)-2^{d-1}r(\theta_1,\theta_2,\dots)\right)$$

The absolute value of the first term in the right hand side is bounded by $2^{d-1}C_1(\mu_1)$. Using the property at step $d-1$, the absolute value of the second term is bounded by $2 \times 2^{d-2}\left(C_2(\mu_2) + \cdots + C_d(\mu_d)\right)$. It yields the claim. $\square$

With evident simplifications, one obtains the bound

$$\|r - s\|_{L^\infty([0,\pi]^d)} \leqslant \int_0^\pi C_1(\mu)K_n(\mu_1)\mu_1 + \cdots + \int_0^\pi C_d(\mu)K_n(\mu_d)\mu_d.$$

LEMMA A.4 (Proof of (3.3)). *One has the bound* $\|q - \widetilde{q}\|_{L^\infty(\mathbb{K})} \leqslant \frac{C_d}{m^2}\|p\|_{W^{2,\infty}(\mathbb{K})}$.

Proof. The error is the sum of d terms which can all be treated with the one-variable bounds of [7, Chapter 5-Section 2, pages 202 to 204]. The second modulus of continuity [7, Theorem 2.2 204] (evaluated with respect to the trigonometric variable $\theta \in [0, \pi]$) is bounded by $\|p\|_{W^{2,\infty}(\mathbb{K})}/n^2 \leqslant C\|p\|_{W^{2,\infty}(\mathbb{K})}/m^2$. It ends the proof. $\square$

A.2. Proof of Lemma 5.4. This is a very classical calculation in the theory of Chebyshev interpolation. Set $T(x) = c\Pi_{s=0}^n(x - \alpha_s)$ where the constant c is still to specify. One has

$$l_k(x) = \frac{T(x)}{(x - \alpha_k)T'(\alpha_k)} \tag{A.4}$$

By elimination of $x = \frac{1+\cos\theta}{2}$, one gets $T(x) = d\Pi_{s=0}^n(\cos\theta - \cos\alpha_s)$ for a constant $d \in \mathbb{R}$. The right hand side is a polynomial of degree $(n+1)$ with respect to $\cos\theta$. We note that $\cos\left(n + \frac{1}{2}\right)\theta\cos\frac{\theta}{2}$ can also be written as a polynomial of degree $(n+1)$ with respect to $\cos\theta$ with the same roots $\cos\alpha_s$. So one can write $T(x) = \cos\left(n + \frac{1}{2}\right)\theta\cos\frac{\theta}{2}$ where the constant c is specified now. It can be rewritten as

$$T\left(\frac{1+\cos\theta}{2}\right) = \cos\left(n + \frac{1}{2}\right)\theta\cos\frac{\theta}{2}.$$

By differentiation with respect to θ, one gets

$$-\frac{\sin\theta}{2}T'\left(\frac{1+\cos\theta}{2}\right) = -\left(n+\frac{1}{2}\right)\sin\left(n+\frac{1}{2}\right)\theta\cos\frac{\theta}{2} - \frac{1}{2}\cos\left(n+\frac{1}{2}\right)\theta\sin\frac{\theta}{2},$$

that is

$$T'(x) = \frac{(2n+1)\sin(n+\frac{1}{2})\theta}{2\sin\frac{\theta}{2}} + \frac{\cos(n+\frac{1}{2})\theta}{2\cos\frac{\theta}{2}} = \frac{2n+1}{2}b_n(x) + \frac{1}{2}a_n(x).$$

First case: take $x = \alpha_k$ for $0 \leqslant k \leqslant n-1$. One gets that $a_n(\alpha_k) = 0$ and that $b_n(\alpha_k) = \frac{\sin\left((n+\frac{1}{2})\frac{2k+1}{2n+1}\pi\right)}{\sin\frac{\theta_k}{2}} = \frac{(-1)^k}{\sin\frac{\theta_k}{2}}$. So $T'(\alpha_k) = (-1)^k\frac{(n+\frac{1}{2})}{\sin\frac{\theta_k}{2}}$.
Second case: for $k = n$, one has $\theta_n = \pi$ and $x = \alpha_n = 0$. The definition 5.3 of a_n yields that $a_n(0) = (-1)^n(2n+1)$. So

$$T'(\alpha_n) = (-1)^n\frac{(n+\frac{1}{2})}{\sin\frac{\theta_n}{2}} + \frac{1}{2}(-1)^n(2n+1) = (-1)^n(2n+1). \tag{A.5}$$

Plugging in (A.4) yields

$$l_k(x) = \frac{\cos\left(n+\frac{1}{2}\right)\theta\cos\frac{\theta}{2}}{x - \alpha_k}(-1)^k\frac{2\sin\theta_k}{2n+1} = \frac{\cos\left(n+\frac{1}{2}\right)\theta\cos\frac{\theta}{2}}{\cos\theta - \cos\theta_k}(-1)^k\frac{4\sin\theta_k}{2n+1}, \qquad 0 \leqslant k \leqslant n-1.$$

This is the claim of Lemma 5.4 for $0 \leqslant k \leqslant n-1$. For $k = 2$, there is an extra factor 2 in the denominator due to (A.5).

A.3. Bound on Chebyshev interpolation polynomials. Let $0 \leqslant i \leqslant n$ and $0 \leqslant x \leqslant 1$.

LEMMA A.5. *There exists a constant $C > 0$ (independent of $x = \frac{1+\cos\theta}{2}$, n and k) such that*

$$\left|\frac{\sqrt{1-x}\; l_k(x)}{(1-\alpha_k)a_n(1-\alpha_k)}\right| \leqslant C\frac{1}{|k-l|+1} \tag{A.6}$$

where $\theta_l \leqslant \theta \leqslant \theta_{l+1}$.

Proof. We use the correspondance $x = \frac{1+\cos\theta}{2} = \cos^2\frac{\theta}{2}$ and $\alpha_k = \frac{1+\cos\theta_k}{2} = \cos^2\frac{\theta_k}{2}$. For $0 \leqslant k \leqslant n-1$, one has

$$\frac{\sqrt{1-x}\; l_k(x)}{(1-\alpha_k)a_n(1-\alpha_k)} = (-1)^k 4\frac{\sin\frac{\theta}{2}\frac{\sin\frac{\theta_k}{2}\cos\left(n+\frac{1}{2}\right)\theta\cos\frac{\theta}{2}}{(2n+1)(\cos\theta-\cos\alpha_k)}}{\sin^2\frac{\theta_k}{2}\frac{\cos\left(n+\frac{1}{2}\right)(\pi-\theta_k)}{\cos\frac{\pi-\theta_k}{2}}} = 4\frac{\sin\frac{\theta}{2}\cos\left(n+\frac{1}{2}\right)\theta\cos\frac{\theta}{2}}{(2n+1)\left(\cos\theta-\cos\alpha_k\right)}. \tag{A.7}$$

The structure of the right most term is classical in Chebyshev interpolation techniques [7]. One has $\cos\theta - \cos\alpha_k = 2\sin\frac{\theta+\theta_k}{2}\sin\frac{\theta_k-\theta}{2}$. The concavity of the function $\theta_k \mapsto \sin\frac{\theta+\theta_k}{2}$ yields that

$$\sin\frac{\theta+\theta_k}{2} \geqslant \min\left(\sin\frac{\theta}{2}, \sin\frac{\theta+\pi}{2}\right) = \min\left(\sin\frac{\theta}{2}, \cos\frac{\theta}{2}\right).$$

One also has $\left|\sin\frac{\theta_k-\theta}{2}\right| \geqslant \frac{1}{\pi}|\theta_k-\theta|$. So one can write

$$\left|\frac{\sqrt{1-x}\; l_k(x)}{(1-\alpha_k)a_n(1-\alpha_k)}\right| \leqslant \frac{C}{(2n+1)|\theta_k-\theta|}.$$

First case: let us assume the condition $\theta_l \leqslant \theta \leqslant \theta_{l+1}$ together with $l < k-2$. Then

$$(2n+1)|\theta_k-\theta| \geqslant (2n+1)|\theta_k-\theta_{l+1}| = 2\pi|k-l-1| \geqslant c(|k-l|+1) \qquad \text{where } c>0,$$

which yields (A.6).

Second case: one still considers $\theta_l \leqslant \theta \leqslant \theta_{l+1}$ but now $k+1 < l$. One gets

$$(2n+1)|\theta_k-\theta| \geqslant (2n+1)|\theta_k-\theta_l| = 2\pi|k-l| \geqslant c(|k-l|+1) \qquad \text{where } c>0,$$

which also yields (A.6).

Third case: the remaining case is $\theta_l \leqslant \theta \leqslant \theta_{l+1}$ with $l = k$ or $l = k-1$. Going back to (A.7), one remarks that

$$\left|\frac{\cos\left(n+\frac{1}{2}\right)\theta}{\cos\theta-\cos\alpha_k}\right| = \left|\frac{\cos\left(n+\frac{1}{2}\right)\theta - \cos\left(n+\frac{1}{2}\right)\alpha_k}{\cos\theta-\cos\alpha_k}\right| \leqslant n+\frac{1}{2}.$$

So $\left|\frac{\sqrt{1-x}\; l_k(x)}{(1-\alpha_k)a_n(1-\alpha_k)}\right| \leqslant 2$ which yields (A.6) because $l = k$ or $l = k-1$.

Finally the constant C in (A.6) is the maximum of the three cases. ∎

REFERENCES

[1] S. BOYD AND L. VANDENBERGHE, *Convex optimization*, Cambridge University Press, Cambridge, 2004.

[2] S. BURER AND R. D. C. MONTEIRO, *A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, Math. Program., 95 (2003), pp. 329–357. Computational semidefinite and second order cone programming: the state of the art.

[3] F. Charles, M. Campos-Pinto, and B. Després, *Algorithms for positive polynomial approximation*, Siam J. Numer. Analysis, 57 (2017), pp. 148–172.

[4] M. Davis and D. Papp, *Dual certificates and efficient rational sum-of-squares decompositions for polynomial optimization over compact sets*, SIAM J. Numer. Anal., 32 (2022).

[5] B. Després and M. Herda, *Computation of sum of squares polynomials from data points*, SIAM J. Numer. Anal., 58 (2020), pp. 1719–1743.

[6] B. Despres and B. Perthame, *Uncertainty propagation; intrusive kinetic formulations of scalar conservation laws*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 980–1013.

[7] R. A. DeVore and G. G. Lorentz, *Constructive approximation*, vol. 303 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1993.

[8] D. Henrion and J. Malick, *Projection methods for conic feasibility problems: applications to polynomial sum-of-squares decompositions*, Optimization Methods & Software, 26 (2011), pp. 23–46.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms. I*, vol. 305 of Fundamental Principles of Mathematical Sciences, Springer-Verlag, 1993.

[10] J. B. Lasserre, *Moments, positive polynomials and their applications*, vol. 1 of Imperial College Press Optimization Series, Imperial College Press, London, 2010.

[11] J. B. Lasserre, *An introduction to polynomial and semi-algebraic optimization*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2015.

[12] J. B. Lasserre and T. Netzer, *Sos approximations of nonnegative polynomials via simple high degree perturbation*, Math. Z., 256, (2007), pp. 99–112.

[13] M. Laurent and L. Slot, *An effective version of schmüdgen's positivstellensatz for the hypercube*, Optimization Letters, (2022).

[14] J. Malick, *A dual approach to semidefinite least-squares prob.*, SIAM J. Mat. Anal. Appl., 26 (2004), pp. 272–284.

[15] M. Marshall, *Positive polynomials and sums of squares*, vol. 146 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2008.

[16] J. C. Mason, *Chebyshev polynomials of the second, third and fourth kinds in approximation, indefinite integration, and integral transforms*, in Proceedings of the Seventh Spanish Symposium on Orthogonal Polynomials and Applications (VII SPOA) (Granada, 1991), vol. 49, 1993, pp. 169–178.

[17] J. C. Mason and G. H. Elliott, *Near-minimax complex approximation by four kinds of Chebyshev polynomial expansion*, vol. 46, 1993, pp. 291–300. Computational complex analysis.

[18] J. Nie and M. Schweighofer, *On the complexity of Putinar's Positivstellensatz*, J. Complex., 23 (2007), pp. 135–150.

[19] V. Powers and T. Wörmann, *An algorithm for sums of squares of real polynomials*, J. Pure Appl. Algebra, 127 (1998), pp. 99–104.

[20] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Ind. Univ. Math. J., 42 (1993), pp. 969–984.

[21] R. T. Rockafellar, *Convex analysis (2nd edition)*, Princeton university press, 1972.

[22] K. Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Math. Ann., 289 (1991), pp. 203–206.

[23] M. Schweighofer, *On the complexity of Schmüdgen's positivstellensatz*, J. Complexity, 20 (2004), pp. 529–543.

[24] G. Szego, *Orthogonal polynomials*, AMS, 1939.

[25] A. Weiss e, G. Wellein, A. Alvermann, and H. Fehske, *The kernel polynomial method*, Rev. Modern Phys., 78 (2006), pp. 275–306.