Connecting optimization with spectral analysis of tri-diagonal matrices

Jean B. Lasserre

Abstract We show that the global minimum (resp. maximum) of a continuous function on a compact set can be approximated from above (resp. from below) by computing the smallest (rest. largest) eigenvalue of a hierarchy of $(r \times r)$ tri-diagonal matrices of increasing size. Equivalently it reduces to computing the smallest (resp. largest) root of a certain univariate degree-r orthonormal polynomial. This provides a strong connection between the fields of optimization, orthogonal polynomials, numerical analysis and linear algebra, via asymptotic spectral analysis of tri-diagonal symmetric matrices.

1 Introduction

The goal of this paper is show that the global minimum (resp. maximum) of a continuous function f on a compact set $\Omega \subset \mathbb{R}^n$ can be approximated as closely as desired from above (resp. from below) by the smallest eigenvalues $(\tau_r^{\ell})_{r \in \mathbb{N}}$ (resp. largest eigenvalues $(\tau_r^u)_{r \in \mathbb{N}}$) of a sequence of tri-diagonal univariate "moment" matrices of increasing size r. Equivalently, τ_r^{ℓ} (resp. τ_r^u) is the smallest (resp. largest) root of a

Jean B. Lasserre
LAAS-CNRS and Institute of Mathematics, University of Toulouse, 7 Avenue du Colonel Roche, BP 54200, 31031 Toulouse cedex 4, France
Tel.: +33 561336415
Fax: +33 561336936
E-mail: lasserre@laas.fr

univariate orthogonal polynomial of increasing degree r. Thus it reveals a (perhaps suprising) connection between the fields of optimization and the asymptotic spectral analysis of tri-diagonal symmetric matrices (also related to the asymptotic analysis of univariate orthogonal polynomials).

Notice that there is a large body of literature on numerical analysis of tri-diagonal symmetric matrices for which efficient specialized algorithms exist (for instance the characteristic polynomial can be computed efficiently and roots of univariate polynomials can also be computed efficiently); see e.g. [9, **?**].

Our result is valid in a quite general context, namely if f is a continuous function and Ω is compact. However to turn it into a *practical* algorithm requires effective computation of such matrices. This in turn requires computing integrals $\int_{\Omega} f(\mathbf{x})^k d\lambda$ for arbitrary k, where λ is a measure whose support is exactly Ω (and whose choice is left to the user). If f is a polynomial, Ω is a "simple" compact set like a box, an ellipsoid, a simplex, a sphere, the discrete hypercube $\{0,1\}^n$ (or their image by an affine mapping) and λ is a distinguished measure (e.g., Lebesgue measure (or other families of polynomial densities), rotation invariant (for the sphere), counting measure for $\{0,1\}^n$) then such integrals are obtained in closed-form.

In addition, after the initial version of this paper was released, Laurent and Slot [15] proved a $O(\log^2 r/r^2)$ rate of convergence for the monotone sequence of upper bounds $(\tau_r^{\ell})_{r \in \mathbb{N}}$ on the minimum (resp. lower bounds $(\tau_r^u)_{r \in \mathbb{N}}$ on the maximum) under a weak geometric condition on Ω (satisfied by the above simple sets).

Background Let $f : \Omega \to \mathbb{R}$ be a continuous function on a compact set $\Omega \subset \mathbb{R}^n$, and consider the optimization problem:

$$\underline{f} = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \Omega \}; \quad \overline{f} = \sup_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \Omega \}. \tag{1.1}$$

In [14] the author showed that one may approximate $\underline{f}$ from above (resp. $\overline{f}$ from below) by solving the following hierarchy[1] of optimization problems indexed by $r \in \mathbb{N}$:

$$\underline{\theta}_r = \inf_{\sigma} \{ \int f \sigma \, d\lambda : \int \sigma \, d\lambda = 1; \quad \sigma \in \Sigma[\mathbf{x}]_r \} \tag{1.2}$$

$$\overline{\theta}_r = \sup_{\sigma} \{ \int f \sigma \, d\lambda : \int \sigma \, d\lambda = 1; \quad \sigma \in \Sigma[\mathbf{x}]_r \}, \tag{1.3}$$

where λ is a fixed measure whose support is exactly Ω and $\Sigma[\mathbf{x}]_r$ is the convex cone of SOS polynomials of degree at most $2r$. Indeed $\underline{\theta}_r \downarrow \underline{f}$ (resp. $\overline{\theta}_r \uparrow \overline{f}$) as $t \to \infty$. If f is a polynomial and one knows all moments of the measure λ on Ω, then each problem (1.2) is a very specific *semidefinite program*. As a matter of fact it reduces to a generalized eigenvalue problem involving two moment-like matrices whose size $\binom{n+r}{r}$ increases with r. For instance this is the case whenever Ω is a simple set (e.g. the box $[-1,1]^n$, the Euclidean unit ball, the sphere, the simplex, the discrete hypercube, and their affine transformation) and λ is an appropriate measure (e.g., Lebesgue measure

---

[1] The hierarchy (1.2) should not be confused with the Moment-SOS hierarchy described in [12] to solve the same problem (1.1), but which yields a converging sequence of *lower bounds* $\rho_r \uparrow \underline{f}$ as $r \to \infty$.

on unit box, Euclidean unit ball, simplex, rotation invariant measure on the sphere, counting measure on the discrete hypercube, etc.).

In a recent series of papers [3–6, 20], de Klerk, Laurent and their co-workers have been able to analyze the convergence $\underline{\theta}_r \downarrow \underline{f}$ of such a hierarchy by appropriate clever choices of the reference measure (as indeed λ it can be any measure whose support is exactly Ω). Ultimately they could provide rates of convergence. In particular and remarkably, they show that for certain important sets and reference measures (e.g. the box $\Omega = [-1, 1]^n$ and the sphere $\mathbb{S}^{n-1}$) a convergence rate $O(1/r^2)$ is achieved.

However all variants of (1.2) and (1.3) consider density polynomials $\sigma \in \mathbb{R}[\mathbf{x}]$ in n variables which results in eigenvalue problems with multivariate Hankel-type matrices whose size $\binom{n+r}{n}$ grows very quickly with r. Therefore and so far, this hierarchy of upper (resp. lower) bounds has not been proved to be efficient in practice (even for small size problems) and its main interest is rather theoretical as it provides an algorithm with proven rate of convergence $O(1/r^2)$ to the global optimum on some simple sets.

Contribution Our contribution is twofold:

I. We first provide an alternative converging hierarchy of upper bounds in the same spirit as (1.2) and (1.3) but following a different path. *The main distinguishing feature is to reduce the n-dimensional initial problem to a one-dimensional equivalent problem by using the pushforward measure #λ of the measure λ on Ω, by the mapping f : Ω → ℝ. It results in solving again a hierarchy of eigenvalue problems but with the major advantage of considering Hankel moment matrices in just ONE variable (hence of size r in contrast to $\binom{n+r}{n}$).*

To achieve this result we exploit the fact that $\underline{f}$ and $\overline{f}$ are the left and right endpoints of the support of $\#\lambda$ on the real line, and therefore by invoking Lasserre [13, Theorem 3.3], one may approximate $\underline{f}$ from above by solving:

$$
\begin{aligned}
\tau_r^\ell &= \sup\{a : \mathbf{H}_r(x; \#\lambda) \succeq a\mathbf{H}_r(\#\lambda)\} \\
&= \lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)), \qquad \forall r \in \mathbb{N},
\end{aligned}
\tag{1.4}
$$

where the real symmetric $(r+1) \times (r+1)$ matrix $\mathbf{H}_r(\#\lambda)$ (resp. $\mathbf{H}_r(x; \#\lambda)$) is the moment matrix of the pushforward $\#\lambda$ (resp. the localizing matrix associated with $\#\lambda$ and the univariate linear polynomial $x \mapsto x$). Similarly, one may approximate $\overline{f}$ from below by solving:

$$
\begin{aligned}
\tau_r^u &= \inf\{a : a\mathbf{H}_r(\#\lambda) \succeq \mathbf{H}_r(x; \#\lambda)\} \\
&= \lambda_{\max}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)), \qquad \forall r \in \mathbb{N},
\end{aligned}
$$

and indeed $\tau_r^\ell \downarrow \underline{f}$ (resp. $\tau_r^u \uparrow \overline{f}$) as r increases; see [13].

Remark 1 Equivalently, by duality in convex optimization:

$$
\tau_r^\ell = \inf_\sigma \left\{ \int z\sigma \, d\#\lambda : \int \sigma \, d\#\lambda = 1; \quad \sigma \in \Sigma[z]_r \right\},
\tag{1.5}
$$

where $\Sigma[z]_r$ is the convex cone of *univariate* SOS polynomials of degree at most $2r$; similarly for τ_r^u just replace "inf" by "sup". The formulation (1.5) resembles (1.2) but with the important difference that in (1.2) one searches over SOS polynomials of degree at most r in n-variables whereas in (1.5) one searches over SOS polynomials of degree at most r in ONE variable.

Also notice that in contrast to (1.2), in (1.5) (or in (1.4)) the function f to min-imize does *not* appear explicitly; it is encoded in the pushforward measure $\#\lambda$. So if one is able to compute (or approximate) the moments of $\#\lambda$, then both matrices $\mathbf{H}_r(x; \#\lambda)$ and $\mathbf{H}_r(\#\lambda)$ are known and (1.4) can be solved in practice. Also to solve (1.2) in practice one needs f to be a polynomial.

II. We next further simplify the analysis by considering an orthonormal basis $(T_j)_{j \in \mathbb{N}}$ of polynomials w.r.t. the pushforward measure $\#\lambda$. Recall that a family of T_j's, $j \leq r$, can be obtained from the moment matrix $\mathbf{H}_r(\#\lambda)$ by simple determinant computations. In this new basis $(T_j)_{j \in \mathbb{N}}$, the moment matrix $\hat{\mathbf{H}}_r(\#\lambda)$ becomes the identity and the localizing matrix $\hat{\mathbf{H}}_r(x; \#\lambda)$ now becomes a *tri-diagonal* (banded) symmetric matrix whose coefficients have a direct expression in terms of the param-eters defining the classical three-term recurrence relation satisfied by the T_j's.

Therefore the convergence $\tau_r^\ell \downarrow \underline{f}$ (resp. $\tau_r^u \uparrow \overline{f}$) is simply the asymptotic behavior of the smallest (resp. largest) eigenvalue of tri-diagonal $r \times r$ "moment" matrices, as r increases; equivalently the asymptotic behavior of the smallest (rest. largest) root of a certain univariate polynomial orthogonal w.r.t. $\#\lambda$.

This reveals a strong and perhaps surprising connection between the fields of (global) optimization and the spectral analysis of tri-diagonal Hankel matrices (or extremal roots of a family of orthogonal polynomials). Actually such a link already appeared in de Klerk and Laurent [7] to analyze convergence of upper bounds (1.2) for the specific univariate (trivial) optimization problem $\min\{x : x \in [-1, 1]\}$ and specific reference measure λ. Then they used this univariate problem as a building block to prove rates of convergence of the bounds in (1.2) in case where $\Omega = [-1, 1]^n$ or $\mathbb{S}^{n-1}$ (and with specific measures λ).

There is a large body of literature on various aspects of tri-diagonal symmetric matrices, including practical efficient algorithms; see for instance Aurentz et al. [1], Businger [2], Ford [9], Kiliç [11], Mallik [16], Osipov [18], and Routh [19]. There-fore this may also open the door to practical algorithms for good approximations of $\underline{f}$ and $\overline{f}$ in non trivial cases, as soon as one can obtain moments of the measure $\#\lambda$ for reasonably large degrees.

If the sequence $(\tau_r^\ell)_{r \in \mathbb{N}}$ (or $(\tau_r^u)_{r \in \mathbb{N}}$) has obvious numerical advantages when compared with $(\underline{\theta}_r)_{r \in \mathbb{N}}$ (or $(\overline{\theta}_r)_{r \in \mathbb{N}}$) for a same fixed r, the main drawback of $(\tau_r^\ell)_{r \in \mathbb{N}}$ is the computation of moments of $\#\lambda$ which may become tedious for non modest di-mension n. However, *sparsity* of f (i.e. when f has a few monomials only) can be exploited. Also the rate of convergence for $\tau_r^\ell \downarrow \underline{f}$ as $r \to \infty$, is more difficult to ana-lyze because we do not know the density of the pushforward measure $\#\lambda$ with respect to Lebesgue measure on Ω.

Actually, recently Laurent and Slot [15] have provided a further detailed analy-sis of some relative merits of the sequences $(\tau_r^\ell)_{r \in \mathbb{N}}$ and $(\underline{\theta}_r)_{r \in \mathbb{N}}$ beyond the scope the present paper. As already mentioned, they prove that remarkably, $\tau_r^\ell \downarrow \underline{f}$ at a

$O(\log^2 r / r^2)$ rate under a weak geometric condition on Ω (satisfied for the simple sets Ω mentioned previously).

2 Main result

2.1 Notation, definitions and preliminary results

Let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the n variables $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbb{R}[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]$ denote the vector space of polynomials of degree at most t, hence of dimension $s(t) = \binom{n+t}{n}$. Let $\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$ denote the space of polynomials that are sums-of-squares (in short SOS polynomials) and let $\Sigma[\mathbf{x}]_t \subset \mathbb{R}[\mathbf{x}]_{2t}$ denote the space of SOS polynomials of degree at most $2t$. For univariate polynomials in the variable x, we use the notation $\mathbb{R}[x]$, $\Sigma[x]$, $\mathbb{R}[x]_t$ and $\Sigma[x]_t$.

With $\alpha \in \mathbb{N}^n$ and $\mathbf{x} \in \mathbb{R}^n$, the notation $\mathbf{x}^\alpha$ stands for $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. Also for every $\alpha \in \mathbb{N}^n$, let $|\alpha| := \sum_i \alpha_i$ and $\mathbb{N}_t^n := \{\alpha \in \mathbb{N}^n : |\alpha| \le t\}$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$.

The support of a Borel measure μ on $\mathbb{R}^n$ is the smallest closed set Ω such that $\mu(\mathbb{R}^n \setminus \Omega) = 0$. Denote by $\mathscr{B}(\mathbf{X})$ the Borel σ-field associated with a topological space $\mathbf{X}$, and $\mathscr{M}(\mathbf{X})_+$ the space of finite nonnegative Borel measures on $\mathbf{X}$.

Generalized eigenvalue Given two real symmetric matrices $\mathbf{A}, \mathbf{C} \in \mathbb{R}^{n \times n}$ denote by $\lambda_{\min}(\mathbf{A}, \mathbf{C})$ the smallest generalized eigenvalue with respect to the pair $(\mathbf{A}, \mathbf{C})$, that is, the smallest scalar τ such that $\mathbf{A}\mathbf{x} = \tau \mathbf{C}\mathbf{x}$ for some nonzero vector $\mathbf{x} \in \mathbb{R}^n$. When $\mathbf{C}$ is the identity matrix then $\lambda_{\min}(\mathbf{A}, \mathbf{C})$ is just the smallest eigenvalue of $\mathbf{A}$. Computing $\lambda_{\min}(\mathbf{A}, \mathbf{C})$ can be done via a pure and efficient linear algebra routine. The notation $\mathbf{A} \succeq 0$ (resp. $\mathbf{A} \succ 0$) stands for $\mathbf{A}$ is positive semidefinite (resp. positive definite). If $\mathbf{A}, \mathbf{C} \succ 0$ then:

$$\lambda_{\min}(\mathbf{A}, \mathbf{C}) = \sup_\tau \{\tau : \mathbf{A} \succeq \tau \mathbf{C}\}. \tag{2.1}$$

Moment matrix Given a real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$, let $\mathbf{H}_r(\phi)$ denote the multivariate (Hankel-type) moment matrix defined by $\mathbf{H}_r(\phi)(\alpha, \beta) = \phi_{\alpha+\beta}$ for all $\alpha, \beta \in \mathbb{N}_r^n$. For instance, in the univariate case $n = 1$, with $r = 2$, $\mathbf{H}_2(\phi)$ is the Hankel matrix

$$\mathbf{H}_2(\phi) = \begin{bmatrix} \phi_0 & \phi_1 & \phi_2 \\ \phi_1 & \phi_2 & \phi_3 \\ \phi_2 & \phi_3 & \phi_4 \end{bmatrix}.$$

If $\phi = (\phi_j)_{j \in \mathbb{N}}$ is the moment sequence of a Borel measure ϕ on $\mathbb{R}$ then $\mathbf{H}_r(\phi) \succeq 0$ for all $r = 0, 1, \ldots$. Conversely, if $\mathbf{H}_r(\phi) \succeq 0$ for all $r \in \mathbb{N}$, then ϕ is the moment sequence of some finite (nonnegative) Borel measure ϕ on $\mathbb{R}$. The converse result is not true anymore in the multivariate case.

Localizing matrix Given a real sequence $\phi = (\phi_\alpha)_{\alpha \in \mathbb{N}^n}$ and $g \in \mathbb{R}[\mathbf{x}]$, $\mathbf{x} \mapsto g(\mathbf{x}) = \sum_\alpha g_\alpha \mathbf{x}^\alpha$, $\mathbf{H}_r(g;\phi)$ denote the multivariate (Hankel-type) localizing matrix defined by

$$\mathbf{H}_r(g;\phi)(\alpha,\beta) = \sum_\gamma g_\gamma \phi_{\gamma+\alpha+\beta}, \quad \forall \alpha, \beta \in \mathbb{N}^n.$$

For instance, in the univariate case $n = 1$, with $r = 2$ and $g(x) = x$, $\mathbf{H}_2(g;\phi)$ is the Hankel matrix

$$\mathbf{H}_2(g;\phi) = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 \\ \phi_2 & \phi_3 & \phi_4 \\ \phi_3 & \phi_4 & \phi_5 \end{bmatrix}.$$

Equivalently, if ϕ is the moment sequence of a Borel measure ϕ on $\mathbb{R}^n$ then $\mathbf{H}_r(g;\phi)$ is the moment matrix $\mathbf{H}_r(\nu)$ associated with the measure $d\nu = g \, d\phi$.

Pushforward measure Let $\lambda \in \mathcal{M}(\Omega)_+$ be a finite Borel measure on $\Omega \subset \mathbb{R}^n$ whose support is exactly Ω, that is, Ω is the smallest closed set such that $\lambda(\mathbb{R}^n \setminus \Omega) = 0$. Let $\#\lambda$ be the *pushforward measure* on $\mathbb{R}$ of λ with respect to (w.r.t.) the mapping $f : \Omega \to \mathbb{R}$. That is:

$$\#\lambda(C) = \lambda(f^{-1}(C)), \qquad \forall C \in \mathscr{B}(\mathbb{R}).$$

In particular, its moments $\#\lambda = (\#\lambda_k)_{k \in \mathbb{N}}$ read:

$$\#\lambda_k = \int_{[0+\infty)} z^k \, d\#\lambda(z) = \int_\Omega f(\mathbf{x})^k \, d\lambda(\mathbf{x}), \qquad k = 0, 1, \dots \qquad (2.2)$$

It is straightforward to see that the support of $\#\lambda$ is contained in the interval $[\underline{f}, \overline{f}]$ with $\underline{f}$ and $\overline{f}$ as its left and right endpoints.

Remark 2 (Computing the moments $\#\lambda$) While (2.2) is quite general, exact numerical computation of $\#\lambda_k$ is *not* possible in full generality. However, for specific combinations of (f, Ω, λ) one may obtain $\#\lambda$ in closed-form (let alone combinations where numerical approximations schemes, e.g. cubature formula or Monte-Carlo, can be used). This is the case when f is a polynomial and Ω is a "simple" set like a box, an ellipsoid, a simplex, a sphere, the discrete hypercube $\{0,1\}^n$, or their image by an affine map. Then several choices of λ are possible: Lebesgue measure (or specific families of polynomial densities) if Ω is a box, simplex or ellipsoid, the rotation invariant measure if Ω is the sphere, the discrete counting measure if Ω is the hypercube, etc. For instance, if $\Omega = [-1,1]^n$ and λ is Lebesgue measure then writing

$$f(\mathbf{x})^k = \sum_{\alpha \in \mathbb{N}^n_{2kd}} f_{k\alpha} \mathbf{x}^\alpha, \quad k = 0, 1, \dots,$$

one obtains

$$\#\lambda_k = \sum_{\alpha \in \mathbb{N}^n_{2kd}} f_{k\alpha} \prod_{i=1}^n \left(\int_{-1}^1 x^{\alpha_i} \, dx \right), \quad k \in \mathbb{N}.$$

Similarly if $\Omega = \{\mathbf{x} : \|\mathbf{x}\|_2^2 \leq 1\}$ then for every $k \in \mathbb{N}$:

$$\#\lambda_k = \int_\Omega g(\mathbf{x})^k \, d\lambda = \frac{1}{\Gamma(1 + \frac{n+kt}{2})} \int_{\mathbb{R}^n} f(\mathbf{x})^k \exp(-\|\mathbf{x}\|_2^2) \, d\lambda$$

$$= \frac{1}{\Gamma(1 + \frac{n+kt}{2})} \sum_{|\alpha|=kt} f_{k\alpha} \prod_{i=1}^n \int_{\mathbb{R}} x^{\alpha_i} \exp(-x^2) \, dx,$$

which is also obtained in closed form. Notice that for large k, computing $\#\lambda_k$ can be time and space consuming.

Remark 3 (On compactness assumption) In fact non compact sets, e.g., $\Omega = \mathbb{R}^n$ with λ being the normal distribution $\mathscr{N}(0,I)$, or $\Omega = \mathbb{R}_+^n$ with λ being the exponential distribution $d\lambda = \prod_i \exp(-x_i) \, d\mathbf{x}$, can be treated as well; see [14].

In particular, and interestingly, it permits to provide upper bounds that converge to the global minimum f^* of f on $\mathbb{R}^n$, even if f^* is not attained (a difficult case in practice because then $\|\mathbf{x}\| \to \infty$ if $f(\mathbf{x}) \to f^*$). For instance with $f(x,y) = (xy - 1)^2 + y^2$ on $\mathbb{R}^2$, $f^* = 0$ is not attained and $f(n, 1/n) \to 0$ as $n \to \infty$. Then letting $d\lambda = (2\pi)^{-1} \exp(-(x^2 - y^2)/2) dx dy$, the pushforward $\#\lambda$ is supported on $[0, +\infty)$, and

$$\#\lambda_k = \frac{1}{2\pi} \int_{\mathbb{R}^2} ((xy-1)^2 + y^2)^k \exp(-x^2/2) \exp(-y^2/2) \, dx dy, \quad k \in \mathbb{N},$$

is obtained in closed-form. However, for ease of exposition and to avoid technicalities, we restrict to the compact case.

2.2 Main result

We have just seen that the support of the pushforward $\#\lambda$ is precisely contained in the interval $[\underline{f}, \overline{f}]$, with $\underline{f}, \overline{f}$ as its left and right endpoints. Therefore the problem of approximating $\underline{f}$ and $\overline{f}$ reduces to approximate the endpoints of the support of $\#\lambda$ from the sole knowledge of its moments. That is,

$$\underline{f} = \min\{x : x \in \mathrm{supp}(\#\lambda)\}; \quad \overline{f} = \max\{x : x \in \mathrm{supp}(\#\lambda)\}. \tag{2.3}$$

In [13] the author has already considered the more general problem of bounding the support of a measure μ on $\mathbb{R}^n$ from knowledge of its marginal moments. In our case μ is the push forward measure $\#\lambda$ on $\mathbb{R}$ and therefore we can invoke Theorem 3.3 in [13]. More precisely:

Let $\Omega \subset \mathbb{R}^n$ be compact with nonempty interior, $\lambda \in \mathscr{M}(\Omega)_+$ and consider the hierarchy of optimization problems indexed by $r \in \mathbb{N}$:

$$\tau_r^\ell = \sup_a \{a : \mathbf{H}_r(x; \#\lambda) \succeq a\mathbf{H}_r(\#\lambda)\} \tag{2.4}$$

$$\tau_r^u = \inf_a \{a : a\mathbf{H}_r(\#\lambda) \succeq \mathbf{H}_r(x; \#\lambda)\}, \tag{2.5}$$

where $\mathbf{H}_r(x; \#\lambda)$ is the (univariate) localizing matrix associated with the polynomial $x \mapsto x$ and the measure $\#\lambda$ on $\mathbb{R}$, and $\mathbf{H}_r(\#\lambda)$ is the (univariate) moment matrix associated with $\#\lambda$.

Theorem 1 *Let $\lambda \in \mathscr{M}(\Omega)_+$ be such that $\mathrm{supp}(\lambda) = \Omega$. Then:*

$$\underline{f} \leq \tau_r^{\ell} \leq \tau_r^{u} \leq \overline{f} \quad \forall r \in \mathbb{N}. \tag{2.6}$$

In addition the sequence $(\tau_r^{\ell})_{r \in \mathbb{N}}$ (resp. $(\tau_r^{u})_{r \in \mathbb{N}}$) is monotone non-increasing (resp. non-decreasing), and:

$$\lim_{r \to \infty} \tau_r^{\ell} = \underline{f}; \quad \lim_{r \to \infty} \tau_r^{u} = \overline{f}. \tag{2.7}$$

Finally, for all $r \in \mathbb{N}$:

$$\tau_r^{\ell} = \lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)) \tag{2.8}$$
$$\tau_r^{u} = \lambda_{\max}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)). \tag{2.9}$$

Proof That the sequence $(\tau_r^{\ell})_{r \in \mathbb{N}}$ is monotone non-increasing is straightforward as the feasible set in (2.4) shrinks with r. The same argument shows that $(\tau_r^{u})_{r \in \mathbb{N}}$ is monotone non-decreasing. Next, the support of $\#\lambda$ is contained in the interval $[\underline{f}, \overline{f}]$ with $\underline{f}$ and $\overline{f}$ as its left and right endpoints, and we know all moments of $\#\lambda$. Therefore (2.6)-(2.8) follow from [13, Theorem 3.3, p. 3379]. $\square$

Even though $\mathbf{H}_r(x; \#\lambda)$ may not be positive definite we still have

$$\sup_{a} \{ a : \mathbf{H}_r(x; \#\lambda) \succeq a \mathbf{H}_r(\#\lambda) \} = \lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)).$$

This is because $\mathbf{H}_r(x - a; \#\lambda) = \mathbf{H}_r(x; \#\lambda) - a\mathbf{H}_r(\#\lambda)$; see Lemma 2.

So after one has reduced the n-dimensional problem (1.2) to the one-dimensional problem (1.5), Theorem 1 shows that one thus has to handle Hankel moment matrices of size r only whereas $\underline{\theta}_r$ requires to handle Hankel-like moment matrices of size $\binom{n+r}{r}$. However the moment *information* needed to build up the moment and localizing matrices $\mathbf{H}_r(\#\lambda)$, $\mathbf{H}_r(x; \#\lambda)$, still requires computing $\int_{\Omega} f(\mathbf{x})^{2k} d\lambda$ with $k \leq 2r$ (hence handling n-variate moments up to degree $2d$). But once this has been done, the eigenvalue problem is much easier.

2.3 On finite convergence

For same reasons as for the convergence $\underline{\theta}_r \downarrow \underline{f}$, in (2.7) the convergence $\tau_r^{\ell} \downarrow \underline{f}$ as $r \to \infty$, is only *asymptotic* and *not* finite in general (and similarly for $\tau_r^{u} \uparrow \overline{f}$ in (2.9)). Indeed the rationale behind (1.5) is to approximate the Dirac measure δ_{z^*} at the minimizer $z^* = \underline{f} \in \mathbb{R}$ by a measure of the form $\sigma d\#\lambda$, and such an approximation cannot be exact because the pushforward measure $\#\lambda$ has no atom if Ω is not finite.

An exception is when Ω is a finite set (e.g., the discrete hypercube $\{0,1\}^n$ or $\{-1,1\}^n$) and λ is the counting measure $\sum_{\mathbf{v} \in \Omega} \delta_{\mathbf{v}}$, with $\delta_{\mathbf{v}}$ being the Dirac measure at $\mathbf{v} \in \Omega$. Introduce the finite set $\mathscr{Z} := \{ f(\mathbf{v}) : \mathbf{v} \in \Omega \} \subset \mathbb{R}$.

Lemma 1 *Let $\Omega \subset \mathbb{R}^n$ be a finite set and let λ be the counting measure on Ω. Then finite convergence takes place at most at $r = \overline{r} = \mathrm{card}(\mathscr{Z}) - 1$. That is, $\tau_{\overline{r}}^{\ell} = \underline{f}$ and $\tau_{\overline{r}}^{u} = \overline{f}$.*

Proof The support of the pushforward measure $\#\lambda$ is the finite set $\mathscr{Z}$ with smallest element $z^* := \underline{f} \in \mathscr{Z}$ and largest element $\overline{f} \in \mathscr{Z}$. In addition, the pushforward measure $\#\lambda$ satisfies $\#\lambda = \sum_{z \in \mathscr{Z}} a_z \, \delta_z$, where:

$$0 < a_z = \lambda(f^{-1}(\{z\})) = \lambda(\{\mathbf{v} \in \Omega : f(\mathbf{v}) = z\})$$
$$= \mathrm{card}(\{\mathbf{v} \in \Omega : f(\mathbf{v}) = z\}), \quad z \in \mathscr{Z}.$$

Let $p \in \mathbb{R}[z]_{\bar{r}}$ be the degree-$\bar{r}$ polynomial $z \mapsto \prod_{a \in \mathscr{Z}; a \neq z^*} \frac{z-a}{z^*-a}$, so that $p(z) = 0$ for all $z^* \neq z \in \mathscr{Z}$, and $p(z^*) = 1$. Then letting $\sigma^* := a_{z^*}^{-1} p^2 \in \Sigma[z]_{\bar{r}}$, one obtains $\int \sigma^* d\#\lambda = \sum_{z \in \mathscr{Z}} a_z \sigma^*(z) = p(z^*)^2 = 1$, and $\int z \sigma^* d\#\lambda = z^* = \underline{f}$. This proves that σ^* is an optimal solution of (1.5) when $r = \bar{r}$ (recall that $\tau_r^\ell \geq \underline{f}$ for all r) and so $\tau_{\bar{r}}^\ell = \underline{f}$ (and with a similar argument, $\tau_{\bar{r}}^u = \overline{f}$). $\qquad\square$

So Lemma 1 states that $\bar{r} = \mathrm{card}(\mathscr{Z}) - 1$ determines when finite convergence is guaranteed, i.e., $\tau_{\bar{r}}^\ell = \underline{f}$. In other words, the larger is the set $\mathscr{Z}$ the more difficult is to compute $\underline{f}$ (or $\overline{f}$).

2.4 Link with tri-diagonal matrices

Let $(T_j)_{j \in \mathbb{N}}$ be a family of orthonormal polynomials with respect to $\#\lambda$. For instance, such a family can be computed from the moments $(\#\lambda_k)_{k \in \mathbb{N}}$ as follows. $T_0(x) = 1 = D_0(x)/\#\lambda_0$ for all $x \in \mathbb{R}$. Then compute the degree-one polynomial:

$$x \mapsto D_1(x) = \det \begin{bmatrix} \#\lambda_0 & \#\lambda_1 \\ 1 & x \end{bmatrix} = \#\lambda_0 x - \#\lambda_1$$

and normalize $T_1(x) = a D_1(x)$ to obtain $\int T_1(x)^2 d\#\lambda = 1$, i.e., $T_1(x) = a D_1(x)$ with $a = \#\lambda_0^{-1/2}(\#\lambda_0 \#\lambda_2 - \#\lambda_1^2)^{-1/2}$. Next, to obtain T_2 compute

$$x \mapsto D_2(x) = \det \begin{bmatrix} \#\lambda_0 & \#\lambda_1 & \#\lambda_2 \\ \#\lambda_1 & \#\lambda_2 & \#\lambda_3 \\ 1 & x & x^2 \end{bmatrix}$$

and again normalize $T_2(x) = a D_2(x)$ to obtain $\int T_2^2 d\#\lambda = 1$, etc. Next, the orthonormal polynomials satisfy the so-called *three-term recurrence* relation:

$$x T_j(x) = a_j T_{j+1}(x) + b_j T_j(x) + a_{j-1} T_{j-1}(x), \quad \forall x \in \mathbb{R}, \quad j \in \mathbb{N}, \qquad (2.10)$$

where $a_j = (d_j d_{j+2}/d_{j+1}^2)^{1/2}$, $b_j = \int x T_j(x)^2 d\#\lambda$, and

$$d_j = \det \begin{bmatrix} \#\lambda_0 & \#\lambda_1 & \cdots & \cdots & \#\lambda_{j-1} \\ \#\lambda_1 & \#\lambda_2 & \cdots & \cdots & \#\lambda_j \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \#\lambda_{j-1} & \#\lambda_j & \cdots & \cdots & \#\lambda_{2j-2} \end{bmatrix}, \quad j \in \mathbb{N}.$$

The tri-diagonal infinite matrix:

$$
J = \begin{bmatrix}
b_0 & a_0 & 0 & \cdots & \cdots & 0 \\
a_0 & b_1 & a_1 & 0 & \cdots & 0 \\
0 & a_1 & b_2 & a_2 & \cdots & 0 \\
0 & 0 & \cdots & \cdots & \cdots & 0
\end{bmatrix}
\tag{2.11}
$$

is called the *Jacobi matrix* associated with the orthonormal polynomials $(T_j)_{j\in\mathbb{N}}$; see e.g. Dunkl and Xu [8, pp. 10–11].

Proposition 1 *Expressed in the basis of orthonormal polynomials* $(T_j)_{j\in\mathbb{N}}$, *the moment matrix* $\hat{\mathbf{H}}_r(\#\lambda)$ *is the identity matrix while the* $(r+1)\times(r+1)$ *localizing matrix* $\hat{\mathbf{H}}_r(x;\#\lambda)$ *is the r-truncation*

$$
J_r = \begin{bmatrix}
b_0 & a_0 & 0 & \cdots & \cdots & 0 \\
a_0 & b_1 & a_1 & 0 & \cdots & 0 \\
0 & a_1 & b_2 & a_2 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & \cdots & a_{r-2} & b_{r-1} & a_{r-1} \\
0 & 0 & \cdots & \cdots & a_{r-1} & b_r
\end{bmatrix}
\tag{2.12}
$$

of the Jacobi matrix (2.11).

Proof That the moment matrix $\hat{\mathbf{H}}_r(\#\lambda)$ expressed in the basis $(T_j)_{j\in\mathbb{N}}$ is the identity matrix follows from

$$
\hat{\mathbf{H}}_r(\#\lambda)(i,j) = \int T_i(x)\,T_j(x)\,d\#\lambda = \delta_{i=j}, \quad \forall i,j = 0,1,\ldots,r
$$

Next, in this basis the localizing matrix $\hat{\mathbf{H}}_r(x;\#\lambda)$ associated with $\#\lambda$ and the polynomial $x \mapsto x$, reads:

$$
\begin{aligned}
\hat{\mathbf{H}}_r(x;\#\lambda)(i,j) &= \int x\,T_i(x)\,T_j(x)\,d\#\lambda \\
&= \underbrace{a_i \int T_{i+1}(x)\,T_j(x)\,d\#\lambda}_{=a_i\,\delta_{i+1=j}} + \underbrace{b_i \int T_i(x)\,T_j(x)\,d\#\lambda}_{b_i\,\delta_{i=j}} \\
&\quad + \underbrace{a_{i-1} \int T_{i-1}(x)\,T_j(x)\,d\#\lambda}_{a_{i-1}\,\delta_{i-1=j}} \\
&= 0 \text{ if } j \notin \{i-1, i, i+1\},
\end{aligned}
$$

for all $i,j = 0,1,\ldots,r$, where we have used (2.10). Hence $\hat{\mathbf{H}}_r(x;\#\lambda)$ is a tri-diagonal matrix where at row i the three elements are (a_{i-1}, b_i, a_{i+1}). Therefore, $\hat{\mathbf{H}}_r(x;\#\lambda)$ is the r-truncation of the Jacobi matrix (2.11). $\qquad\square$

As a consequence we thus obtain:

Corollary 1 *Let τ_r^ℓ and τ_r^u be as in Theorem 1 and let J_r be the tri-diagonal matrix in Proposition 1. Then $\tau_r^\ell = \lambda_{\min}(J_r)$ and $\tau_r^u = \lambda_{\max}(J_r)$. Therefore:*

$$\lambda_{\min}(J_r) \downarrow \underline{f} \text{ and } \lambda_{\max}(J_r) \uparrow \overline{f} \text{ as } r \to \infty. \tag{2.13}$$

Also for every $r \in \mathbb{N}$, τ_r^ℓ (resp. τ_f^u) is the smallest (resp. largest) root of the univariate polynomial T_{r+1}.

Proof Follows from Theorem 1 and the definition of τ_r^ℓ and τ_r^u. The last statement can be found in Dunkl and Xu [8, Theorem 1.3.12].

Remark 4 The use of an othonormal polynomial basis to reduce the initial n-dimensional problem (1.2) to a standard (n-dimensional) eigenvalue problem was already proposed in [14] and in de Klerk et al. [5] and de Klerk and Laurent [6] but for the original n-dimensional problem and the reference measure λ (and not on $\mathbb{R}$ for the pushforward measure $\#\lambda$).

In addition, in [7] de Klerk and Laurent have used the univariate problem $\min\{x : x \in [-1,1]\}$ as a building block to prove the $O(1/r^2)$ rate of convergence for the bounds (1.2) and (1.3) in the multivariate case of the Sphere, the unit box and for some different reference measures λ. They observed that if f is the *univariate* polynomial x then solving the resulting eigenvalue problem is computing the smallest eigenvalue of the Jacobi matrix associated with λ (or equivalently, the smallest root of a certain orthogonal polynomial) as in Corollary 1. For specific reference measures λ, the associated orthogonal polynomials have been well-studied (e.g. Chebyshev or Legendre polynomials), in particular the asymptotic behavior of their smallest (or largest) root used by the authors in [7]. For more details the interested reader is referred to [3–7]. However, in Corollary 1 the underlying univariate problem $\min\{x : x \in \text{supp}(\#\lambda)\}$ in (2.3) is *equivalent* to the original multivariate problem (1.1). The price to pay is that the density of $\#\lambda$ is not known explicitly and makes the analysis of the rate of convergence more intricate.

2.5 Discussion

As already mentioned, computing the entries of $\mathbf{H}_r(\#\lambda)$ (and hence of $\mathbf{H}(x;\#\lambda)$ as well) is easy but tedious for large n. For a fixed $r \in \mathbb{N}$, and once the moment matrix $\mathbf{H}_r(\#\lambda)$ has been computed, computing the scalar τ_r^u or τ_r^ℓ in (2.4)-(2.5) is definitely easier than computing $\underline{\theta}_r$ or $\overline{\theta}_r$ as in the former one handles univariate moment matrices of size r instead of n-variate moment matrices of size $\binom{n+d}{n}$ in the latter. However we have not proved any rate of convergence for $\tau_r^\ell \downarrow \underline{f}$ whereas $\underline{\theta}_r \downarrow \underline{f}$ at a rate $O(1/r^2)$ for some simple sets Ω and appropriate measures λ. As mentioned earlier, convergence analysis of the sequence $(\tau_r^\ell)_{r \in \mathbb{N}}$ is difficult because we do *not* have an explicit expression of the density of $\#\lambda$ w.r.t. Lebesgue measure on $[\underline{f}, \overline{f}]$.

For illustration purpose, for $r = 5, 6$, we have considered four toy problems in $n = 2$ variables to compare the upper bounds τ_r^ℓ on $\underline{f}$ obtained in (2.4) with the upper bounds $\underline{\theta}_r$ obtained in (1.2) as described in de Klerk and Laurent [5]. Hence for the same r, the former are obtained by solving eigenvalue problems with matrices of size

6 for $r = 5$ (resp. size 7 for $r = 6$) as opposed to matrices of size $\binom{2+r}{2} = 21$ for $r = 5$ (resp. size 28 for $r = 6$) for the latter.

Motzkin polynomial: $f(\mathbf{x}) = 64\,(x_1^4 x_2^2 + x_1^2 x_2^4) - 48\,x_1^2 x_2^2 + 1$
Matyas function: $f(\mathbf{x}) = 26\,(x_1^2 + x_2^2) - 48\,x_1 x_2$
Booth function: $f(\mathbf{x}) = (10\,x_1 + 20\,x_2 - 7)^2 + (20\,x_1 + 10\,x_2 - 5)^2$
Three-hum-camel function: $f(\mathbf{x}) = 5^6 x_1^6/6 - 1.05 * 5^4 x_1^4 + 50 x_1^2 + 245\,x_1 x_2 + 25 x_2^2$

Table 1 Comparing τ_r^ℓ and $\underline{\theta}_r$ on 4 toy examples

pb	Moztkin	Matyas	Booth	Three-hump camel
$\underline{\theta}_5$	0.801	3.69	69.81	9.58
τ_5^ℓ	0.873	2.06	56.64	15.07
$\underline{\theta}_6$	0.801	2.99	63.54	4.439
τ_6^ℓ	0.808	1.68	45.49	12.68

In Table 1 are displayed the results. Except for the Motzkin polynomial the bounds τ_r^ℓ are comparable and even better than the bounds $\underline{\theta}_r$.

Convergence rate After the initial version of this paper was released on a public repository, Laurent and Slot [15] have analyzed the convergence $\tau_r^\ell \downarrow f$ as $r \to \infty$, of Theorem 1. Quite remarkably they prove a $O(\log^2 r / r^2)$ rate of convergence under a weak geometric condition on Ω (satisfied for the "simple" sets Ω). Their proof is in the same vein and spirit as the one in Slot and Laurent [20] where the authors prove convergence $\underline{\theta}_r \downarrow f$ at a $O(\log^2 r / r^2)$ rate for general convex bodies. For more details the interested reader is referred to [15].

3 Appendix

Lemma 2 *Let* $\mathbf{H}_r(x; \#\lambda)$ *be the Hankel matrix associated with* $\#\lambda$ *and the polynomial* $x \mapsto x$. *Then*

$$\lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)) = \sup_a \{ a : \mathbf{H}_r(x; \#\lambda) \succeq a\,\mathbf{H}_r(\#\lambda) \}. \qquad (3.1)$$

Proof Let $c < f$ be arbitrary, fixed. As $x \geq f > c$ for all x in the support of $\#\lambda$, it follows that $\mathbf{H}_r(x - c; \#\lambda) \succ 0$, and since $\mathbf{H}_r(\#\lambda) \succ 0$,

$$\lambda_{\min}(\mathbf{H}_r(x - c; \#\lambda), \mathbf{H}_r(\#\lambda)) = \sup_a \{ a : \mathbf{H}_r(x - c; \#\lambda) \succeq a\,\mathbf{H}_r(\#\lambda) \}.$$

Notice also that since $\mathbf{H}_r(x - c; \#\lambda) = \mathbf{H}_r(x; \#\lambda) - c\,\mathbf{H}_r(\#\lambda)$:

$$\lambda_{\min}(\mathbf{H}_r(x - c; \#\lambda), \mathbf{H}_r(\#\lambda))$$

$$\begin{aligned}
&= \inf_a \{ a : \exists p, \, \mathbf{H}_r(x-c; \#\lambda)\, p = a\mathbf{H}_r(\#\lambda)\, p \} \\
&= \inf_a \{ a : \exists p, \, \mathbf{H}_r(x; \#\lambda)\, p = (a+c)\mathbf{H}_r(\#\lambda)\, p \} \\
&= -c + \inf_a \{ a : \exists p, \, \mathbf{H}_r(x; \#\lambda)\, p = a\mathbf{H}_r(\#\lambda)\, p \} \\
&= -c + \lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)).
\end{aligned}$$

Next,

$$\begin{aligned}
-c + \lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda)) &= \lambda_{\min}(\mathbf{H}_r(x-c; \#\lambda), \mathbf{H}_r(\#\lambda)) \\
&= \sup_a \{ a : \mathbf{H}_r(x; \#\lambda) \succeq (a+c)\mathbf{H}_r(\#\lambda) \} \\
&= -c + \sup_a \{ a : \mathbf{H}_r(x; \#\lambda) \succeq a\mathbf{H}_r(\#\lambda) \},
\end{aligned}$$

and the proof is complete. □

4 Conclusion

We have exhibited a strong (and perhaps surprising) connection between global optimization and spectral analysis of tri-diagonal univariate moment matrices (equivalently, roots of some sequence of univariate orthogonal polynomials). Essentially computing the global minimum (resp. maximum) of a function f on a compact set $\Omega \subset \mathbb{R}^n$ reduces to a *one-dimensional* problem, namely computing the limit of the smallest (resp. largest) eigenvalue of a sequence of tri-diagonal moment matrices whose size r is independent of the dimension n. Of course the entries of these matrices require computing integrals $\int_\Omega f(x)^k d\lambda$, $k \in \mathbb{N}$, for some choice of a measure λ whose support is exactly Ω. When Ω is a simple set then this is theoretically easy but becomes tedious for large n.

On the theoretical side, the question of how fast such bounds converge to the global minimum was not addressed in this paper but was addressed in the recent work of Laurent and Slot [15]. In [15] the authors prove that the bounds converge to the global minimum (resp. global maximum) at a $O(\log^2 r/r^2)$ rate, under a weak geometric condition on Ω (satisfied by all of simple sets except for the discrete hypercube).

On a more practical side, whether such an approach may provide good bounds for not too large r and significant dimension n, is a topic of future research. Interestingly, it is definitely related to a problem in numerical analysis, namely how efficiently can be computed integrals $\int_\Omega f(\mathbf{x})^k d\lambda$ when f is a polynomial, Ω is a simple set and λ an appropriate measure whose support is Ω.

At last but not least, observe that if the global optimum can be approximated at a $O(\log^2 r/r^2)$ rate, so far one cannot approximate global minimizers as we are essentially addressing a related *one-dimensional* problem. Therefore our result in this paper suggests the following question: For the particular class of NP-hard polynomial optimization problems studied in this paper, is the problem of computing (only) the optimal value of same computational complexity as the problem of computing the optimal value *and* global minimizers ?

References

1. Aurentz J.L., Mach T., Vandebril R., Watkins D. Computing the eigenvalues of symmetric tri-diagonal matrices via Cayley transformation, Electronic Trans. Num. Anal., 46, pp. 447–449 (2017)
2. Businger P., Extremal properties of balanced tri-diagonal matrices, Math. Comput., 23, pp. 193–195 (1969)
3. de Klerk E., Laurent M., Sun Z., Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization, Math. Prog. Series A, 162, pp. 363–392 (2017)
4. de Klerk E., Laurent M., Hess R., Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization, SIAM J. Optim., 27, pp. 347–367 (2017)
5. de Klerk E., Laurent M., Comparison of Lasserre's measure–based bounds for polynomial optimization to bounds obtained by simulated annealing, Math. Oper. Res., 43, pp. 1317–1325 (2018)
6. de Klerk E., Laurent M., Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial optimization on the sphere, Math. Program. (2020). To appear; `arXiv:1904.08828` (2019)
7. de Klerk E., Laurent M., Worst-case examples for Lasserre's measure-based hierarchy for polynomial optimization on the hypercube, Math. Oper. Res., 45, No 1 (2019). To appear.
8. Dunkl C., Xu Y., Orthogonal Polynomials of Several Variables, Cambridge University Press, Cambridge, UK (2001)
9. Ford W., Numerical Linear Algebra with Applications: Using MATLAB, Academic Press, Amsterdam (2015)
10. Kahan W., When to neglect off-diagonal elements of symmetric tri-diagonal matrices, Technical report, Stanford University, Stanford, CA (1966)
11. Kiliç E., Explicit formula for the inverse of a tridiagonal matrix by backward continued fractions, Appl. Math. Comput., 197, pp. 345–357 (2008)
12. Lasserre J.B., Moments, Positive Polynomials and Their Applications, Imperial College Press, London, UK (2009)
13. Lasserre J.B., Bounding the support of a measure from its marginal moments, Proc. Amer. Math. Soc., 139, pp. 3375–3382 (2011)
14. Lasserre J.B., A new look at nonnegativity on closed sets and polynomial optimization, SIAM J. Optim., 21, pp. 864–885 (2011)
15. Laurent M., Slot L. Near-optimal analysis of univariate moment bounds for polynomial optimization, `arXiv:2001.11289` (2020)
16. Mallik R. K., The inverse of a tridiagonal matrix, Linear Alg. Appl., 325, pp. 109–139 (2001)
17. Nevai P., Orthogonal Polynomials: Theory and Practice, Editor, Nato ASI Series, Kluwer Academic Publishers, Dordrecht, The Netherlands (1990)
18. Osipov A., Evaluation of small elements of the eigenvectors of certain symmetric tridiagonal matrices with high relative accuracy, Research Report YALEU/DCS/TR-1460, Yale University, December 13 (2012)
19. Routh Ong U., Matrix Algorithms in MATLAB, Academic Press, (2016)
20. Slot L., Laurent M., Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets, Math. Program. (2020). to appear; `arXiv:1905.08142`