



**HAL**  
open science

# Connecting optimization with spectral analysis of tri-diagonal Hankel matrices

Jean-Bernard Lasserre

► **To cite this version:**

Jean-Bernard Lasserre. Connecting optimization with spectral analysis of tri-diagonal Hankel matrices. 2019. hal-02190818v2

**HAL Id: hal-02190818**

**<https://laas.hal.science/hal-02190818v2>**

Preprint submitted on 19 Aug 2019 (v2), last revised 12 Mar 2020 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# CONNECTING OPTIMIZATION WITH SPECTRAL ANALYSIS OF TRI-DIAGONAL HANKEL 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 (resp. largest) eigenvalue of a hierarchy of  $(r \times r)$  tri-diagonal Hankel matrix 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 Hankel 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 (resp. largest) eigenvalues of a sequence of tri-diagonal Hankel matrices of increasing size. Equivalently it reduces to computing the smallest (resp. largest) root of univariate polynomials of increasing degree. Thus it reveals a (perhaps suprising) connection between the fields of optimization and the asymptotic spectral analysis of tri-diagonal Hankel matrices (also related to the asymptotic analysis of orthogonal polynomials).

When  $f$  is a polynomial and  $\Omega$  is a “simple” compact set this approach also yields a practical algorithm for approximating the global optimum of  $f$ . (By “simple” we mean that one may compute all moments of some Borel measure whose support is exactly  $\Omega$ .) Indeed there is a large body of literature on the 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, 18].

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

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

---

This work was supported by the European Research Council (ERC) under the ERC-Advanced Grant for the TAMING project.

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

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

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

where  $\lambda$  is a fixed measure whose support is exactly  $\Omega$  and  $\Sigma[\mathbf{x}]_d$  is the convex cone of SOS polynomials of degree at most  $2d$ . Indeed  $\underline{\theta}_r \downarrow \underline{f}$  (resp.  $\overline{\theta}_r \uparrow \overline{f}$ ) as  $r \rightarrow \infty$ . When  $f$  is a polynomial and one knows all moments of the measure  $f d\lambda$  on  $\Omega$  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  $\Omega$  is a simple set (e.g. the box  $[-1, 1]^n$ , the Euclidean unit ball, the sphere, the simplex and their affine transformation) and  $\lambda$  is the Lebesgue measure or the rotation invariant measure).

In a recent series of papers [3, 4, 5, 6, 19], 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  $\lambda$  it can be any measure whose support is exactly  $\Omega$ ). 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 minimum (resp. maximum) 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  $\#\lambda$  of the Lebesgue measure  $\lambda$  on  $\Omega$ , by the mapping  $f : \Omega \rightarrow \mathbb{R}$ . 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  $\binom{n+r}{n}$ ).*

<sup>1</sup>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  $r \rightarrow \infty$ .

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

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

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:

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

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

**Remark 1.1.** Equivalently, by duality in convex optimization:

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

where  $\Sigma[z]_r$  is the convex cone of *univariate* SOS polynomials of degree at most  $2r$ ; similarly for  $\tau_r^u$  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 minimize 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 the  $T_j$ '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) Hankel matrix whose coefficients have a direct expression in terms of the parameters 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  $r \times r$  Hankel matrices, as  $r$  increases; equivalently the asymptotic behavior of the smallest (rest. largest) root of a certain univariate polynomial orthogonal w.r.t.  $\#\lambda$ .*

Hence we exhibit 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  $\lambda$ . 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  $\lambda$ ).

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 [15], Osipov [17], and Routh [18]. Therefore 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 may 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 fixed same  $r$ , 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 dimension  $n$ . However, *sparsity* of  $f$  (i.e. when  $f$  has a few monomials only) can be exploited. Also the convergence analysis of  $\tau_r^\ell \downarrow \underline{f}$  is more difficult because we do not know the density of the pushforward measure  $\#\lambda$  with respect to Lebesgue measure on  $\Omega$ .

Therefore a detailed further analysis of the relative merits of the sequences  $(\tau_r^\ell)_r$  and  $(\underline{\theta}_r)_r$  is needed, but beyond the scope of the present paper.

## 2. MAIN RESULT

**2.1. Notation, definitions and preliminary results.** Let  $\mathbb{R}[\mathbf{x}]$  denote the ring of polynomials in the  $n$  variables  $\mathbf{x} = (x_1, \dots, x_n)$  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  $\mathbb{N}_t^n := \{\alpha \in \mathbb{N}^n : |\alpha| \leq t\}$ , where  $\mathbb{N} = \{0, 1, 2, \dots\}$ .

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

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

**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  $r = 2$ ,  $\mathbf{H}_2$  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  $\phi$  on  $\mathbb{R}$  then  $\mathbf{H}_r(\phi) \succeq 0$  for all  $r = 0, 1, \dots$ . Conversely, if  $\mathbf{H}_r(\phi) \succeq 0$  for all  $r \in \mathbb{N}$ , then  $\phi$  is the moment sequence of some finite (nonnegative) Borel measure  $\phi$  on  $\mathbb{R}$ . The converse result is not true anymore in the multivariate case.

**Pushforward measure.** Let  $\lambda \in \mathcal{M}(\Omega)_+$  be the Lebesgue measure on  $\Omega \subset \mathbb{R}^n$ . Let  $\#\lambda$  be the *pushforward measure* on  $\mathbb{R}$  of  $\lambda$  with respect to (w.r.t.) the mapping  $f : \Omega \rightarrow \mathbb{R}$ . That is:

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

In particular its moments  $\#\lambda = (\#\lambda_k)_{k \in \mathbb{N}}$  can be obtained in closed form as soon as  $\Omega$  is a simple set. Indeed for instance if  $\Omega = [-1, 1]^n$  then

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

So every  $\#\lambda_k$  can be obtained in closed form by writing

$$f(\mathbf{x})^k = \sum_{\alpha \in \mathbb{N}_{2kd}^n} f_{k\alpha} \mathbf{x}^{\alpha},$$

so that

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

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

$$\begin{aligned} \#\lambda_k &= \int_{\Omega} g(\mathbf{x})^k d\lambda = \frac{1}{\Gamma(1 + \frac{n+kt}{2})} \int_{\mathbb{R}^n} f(\mathbf{x})^k \exp(-\|\mathbf{x}\|_2^2) d\lambda \\ &= \frac{1}{\Gamma(1 + \frac{n+kt}{2})} \sum_{|\alpha|=kt} f_{k\alpha} \prod_{i=1}^n \int_{\mathbb{R}} x^{\alpha_i} \exp(-x^2) dx \end{aligned}$$

which is also obtained in closed form.

Notice that for large  $k$  it can be time and space consuming to compute  $\#\lambda_k$ . Also it is straightforward to see that the support of  $\#\lambda$  is contained in the interval  $[\underline{f}, \bar{f}]$  with  $\underline{f}$  and  $\bar{f}$  as its left and right endpoints.

**2.2. Main result.** In the previous section 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,

$$(2.2) \quad \underline{f} = \min \{x : x \in \text{supp}(\#\lambda)\}; \quad \overline{f} = \max \{x : x \in \text{supp}(\#\lambda)\}.$$

In [13] the author has already considered the more general problem of bounding the support of a measure  $\mu$  on  $\mathbb{R}^n$  from knowledge of its marginal moments. In our case  $\mu$  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,  $\lambda \in \mathcal{M}(\Omega)_+$  and consider the hierarchy of optimization problems indexed by  $r \in \mathbb{N}$ :

$$(2.3) \quad \tau_r^\ell = \sup_a \{a : \mathbf{H}_r(x; \#\lambda) \succeq a \mathbf{H}_r(\#\lambda)\}$$

$$(2.4) \quad \tau_r^u = \inf_a \{a : a \mathbf{H}_r(\#\lambda) \succeq \mathbf{H}_r(x; \#\lambda)\},$$

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 2.1.** *Let  $\lambda \in \mathcal{M}(\Omega)_+$  be such that  $\text{supp}(\lambda) = \Omega$ . Then:*

$$(2.5) \quad \underline{f} \leq \tau_r^\ell \leq \tau_r^u \leq \overline{f} \quad \forall r \in \mathbb{N}.$$

*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:*

$$(2.6) \quad \lim_{r \rightarrow \infty} \tau_r^\ell = \underline{f}; \quad \lim_{r \rightarrow \infty} \tau_r^u = \overline{f}.$$

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

$$(2.7) \quad \tau_r^\ell = \lambda_{\min}(\mathbf{H}_r(x; \#\lambda), \mathbf{H}_r(\#\lambda))$$

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

*Proof.* That the sequence  $(\tau_r^\ell)_{r \in \mathbb{N}}$  is monotone non-increasing is straightforward as the feasible set in (2.3) 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 (2.5)-(2.7) 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 3.1.

So after one has reduced the  $n$ -dimensional problem (1.2) to the one-dimensional problem (1.5), Theorem 2.1 shows that one thus has handle

Hankel moment matrices of size  $r$  only whereas  $\underline{\theta}_r$  requires to handle Hankel moment matrices of size  $\binom{n+r}{r}$ . However the moment *information* needed to build up the moment matrix  $\mathbf{H}_r(\#\lambda)$  still requires computing  $\int_{\Omega} f(\mathbf{x})^{2k} d\lambda$  with  $k \leq 2r$  (hence handling  $n$ -variate moments up to degree  $2d$ ). However one this has been done the eigenvalue problem is much easier.

**2.3. Link with tri-diagonal Hankel matrices.** Let  $(T_j)_{j \in \mathbb{N}}$  be the orthonormal polynomials with respect to  $\#\lambda$ . For instance, they all 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:

$$(2.9) \quad x T_j(x) = a_j T_{j+1}(x) + b_j T_j(x) + a_{j-1} T_{j-1}(x), \quad \forall x \in \mathbb{R}, \quad j \in \mathbb{N},$$

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 Hankel matrix:

$$(2.10) \quad J = \begin{bmatrix} b_0 & a_0 & 0 & \cdots & \cdots & 0 \\ a_0 & b_1 & a_1 & 0 & \cdots & 0 \\ 0 & a_1 & b_2 & a_2 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}$$

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

**Proposition 2.2.** *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 (Hankel) matrix  $\hat{\mathbf{H}}_r(x; \#\lambda)$  is the  $r$ -truncation*

$$(2.11) \quad J_r = \begin{bmatrix} b_0 & a_0 & 0 & \cdots & \cdots & 0 \\ a_0 & b_1 & a_1 & 0 & \cdots & 0 \\ 0 & a_1 & b_2 & a_2 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots & b_{r-1} & a_r \end{bmatrix}$$

of the Jacobi matrix (2.10).



*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

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

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

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

for all  $i, j = 0, 1, \dots, r$ , where we have used (2.9). This proves that  $\hat{\mathbf{H}}_r(x; \#\lambda)$  is a tri-diagonal Hankel matrix; 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 (2.10).  $\square$

As a consequence we thus obtain:

**Corollary 2.3.** *Let  $\tau_r^\ell$  and  $\tau_r^u$  be as in Theorem 2.1 and let  $J_r$  be the tri-diagonal Hankel matrix in Proposition 2.2. Then  $\tau_r^\ell = \lambda_{\min}(J_r)$  and  $\tau_r^u = \lambda_{\max}(J_r)$ . Therefore:*

$$(2.12) \quad \lambda_{\min}(J_r) \downarrow \underline{f} \text{ and } \lambda_{\max}(J_r) \uparrow \bar{f} \text{ as } r \rightarrow \infty.$$

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

*Proof.* Follows from Theorem 2.1 and the definition of  $\tau_r^\ell$  and  $\tau_r^u$ . The last statement can be found in Dunkl and Xu [8, chap. 1].  $\square$

**Remark 2.4.** The use of an orthonormal 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  $\lambda$  (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 hypercube and for some different reference measures  $\lambda$ . 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  $\lambda$  (or equivalently, the smallest root of a certain orthogonal polynomial) as in Corollary 2.3. For specific reference measures  $\lambda$ , 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, 4, 5, 6, 7]. However in Corollary 2.3 the underlying univariate problem  $\min\{x : x \in \text{supp}(\#\lambda)\}$  in (2.2) 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.

**Discussion.** As already mentioned, computing the entries of  $\mathbf{H}_r(\#\lambda)$  (and hence of  $\mathbf{H}(x; \#\lambda)$ ) 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  $\tau_r^u$  or  $\tau_r^\ell$  in (2.3)-(2.4) is definitely easier than computing  $\underline{\theta}_r$  or  $\bar{\theta}_r$  as in the former we handle 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  $\Omega$  and appropriate measures  $\lambda$ . 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  $[\underline{f}, \bar{f}]$ .

For illustration purpose, for  $r = 5, 6$ , we have considered four toy problems in  $n = 2$  variables to compare the upper bounds  $\tau_r^\ell$  on  $\underline{f}$  obtained in (2.3) 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:**  $f(\mathbf{x}) = 64(x_1^4 x_2^2 + x_1^2 x_2^4) - 48x_1^2 x_2^2 + 1$

**Matyas function:**  $f(\mathbf{x}) = 26(x_1^2 + x_2^2) - 48x_1 x_2$

**Booth function:**  $f(\mathbf{x}) = (10x_1 + 20x_2 - 7)^2 + (20x_1 + 10x_2 - 5)^2$

**Three-hum-camel function:**  $f(\mathbf{x}) = 5^6 x_1^6 / 6 - 1.05 * 5^4 x_1^4 + 50x_1^2 + 245x_1 x_2 + 25x_2^2$

pb	Motzkin	Matyas	Booth	Three-hump camel
$\underline{\theta}_5$	0.801	3.69	69.81	9.58
$\tau_5^\ell$	0.873	2.06	56.64	15.07
$\underline{\theta}_6$	0.801	2.99	63.54	4.439
$\tau_6^\ell$	0.808	1.68	45.49	12.68

TABLE 1. Comparing  $\tau_r^\ell$  and  $\underline{\theta}_r$  on 4 toy examples

In Table 1 are displayed the results. Except for the Motzkin polynomial the bounds  $\tau_r^\ell$  are comparable and even better than the bounds  $\underline{\theta}_r$ . However we have not proved any convergence rate.

### 3. APPENDIX

**Lemma 3.1.** *Let  $\mathbf{H}_r(x; \#\lambda)$  be the Hankel matrix associated with  $\#\lambda$  and the polynomial  $x \mapsto x$ . Then*

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

*Proof.* Let  $c < \underline{f}$  be arbitrary, fixed. As  $x \geq \underline{f} > c$  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  $\mathbf{H}_r(x - c; \#\lambda) = \mathbf{H}_r(x; \#\lambda) - c \mathbf{H}_r(\#\lambda)$ :

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

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.  $\square$

### 4. CONCLUSION

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

On the theoretical side we have seen that convergence to the global optimum is related to the asymptotic behavior of the smallest root of certain

orthogonal polynomials. Providing a rate of convergence is a topic of further investigation. On a more practical side, whether such an approach may provide good bounds for not too large  $r$  is also a topic of future research.

## 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, 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, 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. [arXiv:1904.08828](https://arxiv.org/abs/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.*, to appear. [arXiv 2018](https://arxiv.org/abs/2018)
- [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] MALLIK, R. K. The inverse of a tridiagonal matrix,, *Linear Alg. Appl.* **325**, pp. 109–139, 2001.
- [16] NEVAI, P. *Orthogonal Polynomials: Theory and Practice*, Editor, Nato ASI Series, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
- [17] 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
- [18] ROUTH, ONG U. *Matrix Algorithms in MATLAB*, Academic Press, 2016.
- [19] SLOT, L., LAURENT, M. Improved convergence analysis of Lasserre’s measure-based upper bounds for polynomial minimization on compact sets, [arXiv:1905.08142](https://arxiv.org/abs/1905.08142)