Quadrature-based polynomial optimization *

A. Martinez, F. Piazzon, A. Sommariva and M. Vianello[1]
Department of Mathematics, University of Padova, Italy

August 19, 2019

Abstract

We show that Lasserre measure-based hierarchies for polynomial optimization can be implemented by directly computing the discrete minimum at a suitable set of algebraic quadrature nodes. The sampling cardinality can be much lower than in other approaches based on grids or norming meshes. All the vast literature on multivariate algebraic quadrature becomes in such a way relevant to polynomial optimization.

2010 AMS subject classification: 65D32, 65K05, 90C26.

Keywords: polynomial optimization, Lasserre measure-based hierarchy, multivariate quadrature formulas.

1 Introduction

In the seminal paper [12], Lasserre proposed a new paradigm for the computation of the minimum of a polynomial on a multidimensional compact set, based on minimizing the expectation with respect to sets of probability measures whose densities are SOS (Sum of Squares) polynomials.

Since the problem can be ultimately rewritten as a semidefinite program, which in turn is equivalent by duality to a matrix eigenvalue problem, we report the following fundamental result that best summarizes the construction from the point of view of the present paper (cf. [7, 8] with the references therein).

In what follows we shall denote by $\mathbb{P}_d^n(K)$ the space of real n-variate polynomials of total degree not exceeding d, restricted to a polynomial determining compact set $K \subset \mathbb{R}^n$ (i.e., polynomials vanishing there vanish everywhere), with dimension $N = N(d,n) = \dim(\mathbb{P}_d^n(K)) = \binom{d+n}{n}$.

Theorem 1 *Let $K \subset \mathbb{R}^n$ be a compact set with nonempty interior, μ a finite Borel measure supported by K, and $\{p_j\}$, $1 \le j \le N$, a (degree ordered) μ-orthonormal basis for $\mathbb{P}_d^n(K)$, that is span$\{p_j\,,\ 1 \le j \le N(r,n)\} = \mathbb{P}_r^n(K)$,*

*Work partially supported by the DOR funds and the biennial project BIRD163015 of the University of Padova, and by the GNCS-INdAM. This research has been accomplished within the RITA "Research ITalian network on Approximation".

[1]corresponding author: marcov@math.unipd.it

$0 \le r \le d$, *and* $\int_K p_j(x)\, p_k(x)\, d\mu = \delta_{jk}$ *(to fix ideas, the standard monomial basis with the degree lexicographical order, orthonormalized by the Gram-Schmidt process). Moreover, let* f *be a fixed* n*-variate polynomial,* f_{min} *its global minimum on* K*, and* A *the* $N \times N$ *symmetric matrix*

$$A_f = (a_{h,k}) \ , \ \ a_{h,k} = \int_K f(x)\, p_h(x)\, p_k(x)\, d\mu \ , \ \ 1 \le h, k \le N \ . \qquad (1)$$

Then, the sequence $f_d^{sos} = \lambda_{min}(A_f)$ *(the minimal eigenvalue of* A_f*) is nonincreasing and*

$$\lim_{d \to \infty} E_d = 0 \ , \ \ E_d = f_d^{sos} - f_{min} \ge 0 \ . \qquad (2)$$

Moreover, if μ *is the Lebesgue measure then* $E_d = \mathcal{O}(1/\sqrt{d})$ *for compact sets satisfying an interior cone condition (cf. [10]) and* $E_d = \mathcal{O}(1/d)$ *for convex bodies (cf. [7]), whereas* $E_d = \mathcal{O}(1/d^2)$ *if* $K = [-1, 1]^n$ *and* μ *is the Chebyshev measure (cf. [8]; see also Table 1 in [5] for a summary).*

From the algorithmic point of view, one may think to compute the matrix A_f, and then its minimal eigenvalue by standard numerical linear algebra algorithms [15], in order to approximate the global minimum of a polynomial. In the case of the Chebyshev measure on $[-1, 1]^n$, the elements of A_f are known analytically, as soon as one has at hand the coefficients of representation of f in the n-variate Chebyshev basis of degree $deg(f)$; cf. [9].

It should be recalled that the general setting of the method does not require a μ-orthonormal basis, since f_d^{sos} turns out to be the minimal generalized eigenvalue of the couple (A_f, G), where G is the Gram matrix of the chosen basis. For example, using the ordered monomial basis, the elements of A_f and G can be computed by the standard moments of μ on K; cf. [8].

Alternatively, when available one may think to use an *algebraic quadrature formula* for μ on K with degree of exactness $2d+deg(f)$, to compute the elements of A_f and G, or simply of A_f when a μ-orthonormal basis is at hand.

However, as we show in the next section, in a quadrature-type approach knowing orthonormal bases as well as computing matrices and minimal eigenvalues is ultimately not necessary. In fact, simply resorting to the existence of a μ-orthonormal basis (which always holds in view of the Gram-Schmidt process), we shall prove that the discrete minimum of f on the quadrature nodes directly gives a convergent approximation to the global minimum, say f_d^{quadr} such that $f_{min} \le f_d^{quadr} \le f_d^{sos}$.

2 Quadrature-based optimization

The main result of the present paper is summarized by the following:

Theorem 2 *Let* K *and* μ *as in Theorem 1,* f *a polynomial and* $d_0 = deg(f)$*. Given a sequence of algebraic quadrature formulas* $\{(x_i(r), w_i(r))\}$ *with nodes* $x_i(r) \in K$ *and positive weights* $w_i(r)$*,* $1 \le i \le M(r)$*,* $r \ge 0$*, namely*

$$\sum_{i=1}^{M(r)} w_i(r)\, p(x_i(r)) = \int_K p(x)\, d\mu \ , \ \ \forall p \in \mathbb{P}_r^n(K) \ , \qquad (3)$$

then the discrete minimum

$$f_d^{quadr} = \min\{f(x_i(r_d))\,,\, 1 \le i \le M(r_d)\}\;,\quad r_d = 2d + d_0\;, \tag{4}$$

is such that $f_{min} \le f_d^{quadr} \le f_d^{sos}$ *and hence it converges to* f_{min} *as* $d \to \infty$, *with an error bounded by* E_d *in (2).*

Remark 1 Before proving the theorem, an observation is in order. In general, the sequence of discrete minima f_d^{quadr} is not monotonic. But if we use a sequence of nested quadrature formulas, that is $\{(x_i(r)\} \subset \{(x_i(r+1)\}$, then f_d^{quadr} becomes nonincreasing like f_d^{sos}.

Proof of Theorem 2. For a fixed d, consider the degree ordered μ-orthonormal basis $\{p_j\}$ of $\mathbb{P}_d^n(K)$, $1 \le j \le N$, obtained by applying the Gram-Schmidt orthonormalization process to the standard monomial basis with the degree lexicographical order.

Consider now the $M \times N$ (rectangular) Vandermonde-like matrix in the orthonormal basis computed at the quadrature nodes $\{x_i(r_d)\}$, namely $V = (p_j(x_i(r_d)))$, $1 \le i \le M = M(r_d)$, $1 \le j \le N$ (observe that $M \ge N$ by the well-known Möller's lower bound in multivariate quadrature theory, namely $M(r) \ge \dim(\mathbb{P}^n_{\lceil r/2 \rceil}(K))$, cf. [13]).

By exactness of the quadrature formula we can write (for notational simplicity we drop the dependence of nodes and weights on r_d)

$$a_{k,h} = \int_K f(x)\, p_k(x)\, p_h(x)\, d\mu = \sum_{i=1}^{M(r_d)} w_i\, f(x_i)\, p_k(x_i)\, p_h(x_i)\;,$$

which with in matrix form becomes

$$A_f = V^t D_w D_f V = (\sqrt{D_w}V)^t D_f (\sqrt{D_w}V) = Q^t D_f Q\;,$$

where D_w and D_f are the diagonal $M \times M$ matrices of the weights and of the values of f at the nodes, respectively, and the matrix $Q = \sqrt{D_w}V$ is orthogonal by μ-orthonormality of the polynomial basis and exactness of the formula at degree $2d + d_0 \ge 2d$. Indeed, for a general basis $Q^tQ = G$, the Gram matrix of the basis with respect to (the scalar product associated with) the measure μ.

Recalling the notion of Rayleigh quotient of a symmetric matrix C and its properties (cf. [15])

$$R_C(v) = v^tCv/v^tv\;,\;\; v \ne 0\;,$$

setting $y = Qv$ and observing that $y^ty = v^tv$ by orthogonality of Q, we get

$$f_d^{sos} = \lambda_{min}(A_f) = \min_{v \in \mathbb{R}^N} R_{A_f}(v) = \min_{y \in S} R_{D_f}(y)$$

$$\ge \min_{y \in \mathbb{R}^M} R_{D_f}(y) = \min_{1 \le i \le M} f(x_i) = f_d^{quadr} \ge f_{min}\;,$$

where $S \subset \mathbb{R}^M$ is the subspace spanned by the columns of Q. Observe that the same reasoning applied to the max, gives the chain of interval inclusions $[\lambda_{min}(A_f), \lambda_{max}(A_f)] \supseteq [\min_{1 \le i \le M} f(x_i), \max_{1 \le i \le M} f(x_i)] \supseteq [f_{min}, f_{max}]$. $\square$

Remark 2 We have just given a proof of Theorem 2 based on the spectral features of the matrix A_f. An even simpler proof (that has been suggested by an anonymous referee who we wish to thank) can be directly obtained by resorting to the original definition of Lasserre's measure-based bound, namely

$$f_d^{sos} = \min \left\{ \int_K f(x)\, q(x)\, d\mu \, : \, \int_K q(x)\, d\mu = 1 \, , \, q \in \text{sos} \, , \, deg(q) \leq 2d \right\} \, , \quad (5)$$

where $q \in$ sos means that q is a sum of squares of polynomials. Indeed, let q^* be an optimal sos polynomial for the definition (5) of f_d^{sos}. Using the quadrature rule $\{(x_i(r_d), w_i(r_d))\}$ which is exact for integrating polynomials of degree at most $2d + d_0$, we get

$$\int_K f(x)\, q^*(x)\, d\mu = \sum_{i=1}^{M(r_d)} w_i(r_d)\, f(x_i(r_d))\, q^*(x_i(r_d))$$

$$\geq f_d^{quadr} \sum_{i=1}^{M(r_d)} w_i(r_d)\, q^*(x_i(r_d)) = f_d^{quadr} \int_K q^*(x)\, d\mu = f_d^{quadr} \, . \quad (6)$$

Observe that both the proofs work only for *positive* weights.

Though the proofs of Theorem 2 are elementary (given Theorem 1 that relies on a deep algebraic-analytic theory), the result is quite meaningful. It says essentially that:

- *the nodes of any positive algebraic quadrature formula for any finite Borel measure supported by any compact set K (with nonempty interior), in any dimension, are good discrete sets to approximate the extremal values of a polynomial on K.*

As a consequence, all the vast theoretical and computational literature on multivariate positive algebraic quadrature over domains with different geometries, becomes now relevant to polynomial optimization; with no pretence of exhaustivity see, e.g., the classical survey papers [3, 4], and the more recent [11, 14, 20, 22, 26], with the references therein.

In practice, Theorem 2 shows that the quadrature-based approach to polynomial optimization is a discrete sampling method. The convergence rates are (at least) those of the corresponding Lasserre measure-based hierarchies, which up to now have been estimated only for special classes of compact sets and special measures (see the last statement in Theorem 1).

Several sampling methods based on grids or norming meshes have been studied in the recent literature, cf. e.g. [6, 17, 18, 25, 28] with the references therein. An interesting feature of quadrature-based optimization is that the cardinality of the quadrature formulas needed to ensure a given error size, can be much lower than the cardinality of the relevant optimization grids or meshes.

Consider for example box-constrained minimization on the Chebyshev norming grids studied in [17], where it is proved that an error $\mathcal{O}(1/s^2)$ is obtained with approximately $(sd_0)^n$ Chebyshev sampling points, that is a cardinality $\mathcal{O}((d_0/\sqrt{\varepsilon})^n)$ is needed to guarantee an error size ε.

On the contrary, using a product Gaussian formula for the Chebyshev measure, which has degree of exactness $2d+d_0$ with approximately $(d+d_0/2)^n$ points, we get an error size ε with only $\mathcal{O}((1/\sqrt{\varepsilon})^n)$ Chebyshev sampling points. Since for both norming sets and quadrature we are using Chebyshev grids, this could partially explain why the actual errors in the numerical examples of [17] turn out to be much lower than the prescribed tolerance.

Remark 3 Observe that both, the norming grid approach and the product Gaussian formula approach, have an exponential computational complexity in n for fixed degree d. On the contrary, the complexity of computing the measure-based bound is polynomial in n for fixed d, since it reduces to computing the minimal eigenvalue of a matrix of size $\mathcal{O}(n^d)$; cf. [12]. The latter property holds however only if the matrix A_f is explicitly known (or computable with a polynomial complexity). This is the case when the integrals of products of basis polynomial triplets $p_i(x)\, p_j(x)\, p_k(x)$ are known analitycally, such as for example with the product Chebyshev measure of the cube, as well as the coefficients of f in the basis; cf. [9]. Indeed, if $f = \sum_k c_k\, p_k(x)$, then $a_{ij} = \sum_k c_k \int_K p_i(x)\, p_j(x)\, p_k(x)\, d\mu$.

Clearly, positive quadrature formulas with *low cardinality* $M(r)$ at a given degree of exactness r are of great interest in view of efficiency, also in the present optimization context. The existence of *minimal* formulas, i.e. formulas that attain Möller's lower cardinality bound $M(r) \geq \dim(\mathbb{P}^n_{\lceil r/2 \rceil}(K))$, is still an open problem not only for general domains and measures but even in standard cases (e.g., the Lebesgue measure in the cube), except very few instances; cf., e.g., [4, 13, 27].

On the other hand, existence of formulas with positive weights satisfying the bound $M(r) \leq \dim(\mathbb{P}^n_r(K))$ is guaranteed on any domain and measure by *Tchakaloff Theorem*, originally proved for the Lebesgue measure [24], and then extended to general measures (cf., e.g., [19]). Though the proofs of Tchakaloff Theorem are typically nonconstructive, when an algebraic quadrature formula of higher cardinality is already known it is a direct consequence of Caratheodory Theorem on finite-dimensional conical combinations [2], applied to the columns of the corresponding underdetermined moment system. Moreover, it can be conveniently implemented, at least in low dimension, by extracting a subset of re-weighted nodes via Linear or Quadratic Programming; cf. [16, 20, 22] for a discussion on the theoretical and computational aspects of such "quadrature compression" methods.

In order to make some numerical examples, we work out the minimization of two classical bivariate test polynomials of degree $d_0 = 4$, namely the Styblinski-Tang function

$$f_1(x_1, x_2) = \sum_{i=1}^{2} \left(\frac{1}{2}\,(10x_i - 5)^4 - 8(10x_i - 5)^2 + \frac{5}{2}\,(10x_i - 5) \right), \quad (7)$$

whose global minimum in $[0,1]^2$ is $f_1(0.209, 0.209) = -78.33198$, and the Rosenbrock function

$$f_2(x_1, x_2) = 100(4.096x_2 - 2.048 - (4.096x_1 - 2.048)^2)^2 + (4.096x_1 - 3.048)^2\ , \quad (8)$$

whose minimum is $f_2\left(\frac{3048}{4096}, \frac{3048}{4096}\right) = 0$. The maxima are $f_1(1,1) = 250$ and $f_2(0,0) = 3905.93$; cf. [10].

In the first example we take $K = [0,1]^2$ and the (transformed) bivariate Chebyshev measure $d\mu = \frac{1}{4}\left((1-t_1^2)(1-t_2^2)\right)^{-1/2}dx_1dx_2$, $t_i = 2x_i - 1$; in this case by Theorem 1 the error of Lasserre upper bound is $E_d = \mathcal{O}(1/d^2)$. As quadrature points of exactness degree $2d+d_0 = 2d+4 = 2(d+2)$ we choose the (transformed) *Padua Points* of degree $d+3$, which give a near-minimal positive formula for the Chebyshev measure with cardinality $(d+4)(d+5)/2$, Möller's lower bound being $(d+3)(d+4)/2$ (incidentally, this property is one of the key features to prove that they are an optimal set for polynomial interpolation on the square; cf. [1]). See Figure 1-top for an idea of the distribution of the Padua Points, which are essentially the union of two Chebyshev-Lobatto subgrids.

In Figure1-bottom we display the relative gaps $(f_d^{quadr} - f_{min})/(f_{max} - f_{min})$ for $d = 1, 2, \ldots, 200$, which oscillate since the quadrature formula at the Padua Points is not nested, but exhibit as expected a decay of order $\mathcal{O}(1/d^2)$.

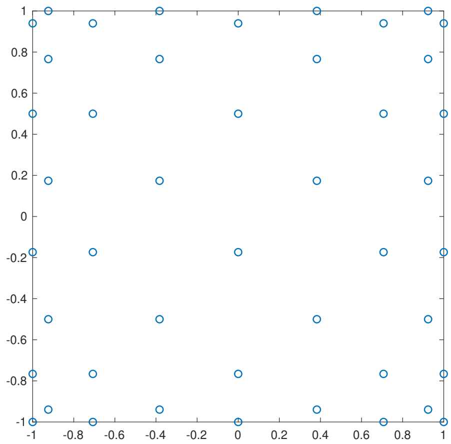
In the second example $K \subset [0,1]^2$ is a *convex* polygon, whereas in the third it is a *nonconvex* polygon, both containing the minimum and maximum points of f_1 and f_2 in $[0,1]^2$, and $d\mu = dx_1dx_2$ is the Lebesgue measure; see Figures 2 and 3. Here we start from a high-cardinality quadrature formula for polygons obtained via triangulation [21], and then compress it into $(2d+5)(2d+6)/2$ *Tchakaloff Points* with exactness degree $2d+4$, by the algorithm in [22]. We stress that existence of a minimal (or near-minimal) formula is not known in these cases, and even existing its computation would be quite challenging.

On the other hand, also computation of Tchakaloff Points, which requires a sparse nonnegative solution of a very large underdetermined moment system (the number of rows is the dimension of the exactness polynomial space, the number of columns the cardinality of the starting quadrature formula), though based on Linear or Quadratic Programming, is a costly procedure which can work essentially in low dimension; cf. [16]. It should be stressed, however, that the Tchakaloff Points, like the nodes of any algebraic quadrature formula, are independent of the polynomial to be minimized, and can be computed once and for all on a given compact set.

In Figures 2 and 3 (bottom) we report the relative gaps corresponding to *nested Tchakaloff Points* of increasing degree of exactness (cf. [23] for the computation of nested formulas, in the Least Squares approximation framework). We can see that the gaps are, differently from the previous example, a *nonincreasing* sequence (with possible long constant pieces), a quite natural behavior due to the nested structure.

A final observation is in order. As already pointed out with other discrete approaches based on grids or norming meshes, polynomial optimization by quadrature points can be viewed as a sort of brute-force approach, that could be useful when only a rough estimate of the extremal values is needed, or as starting guess for more sophisticated optimization algorithms.

Conclusions. We have shown that Lasserre's measure based hierarchies for polynomial optimization can be implemented by discrete minimization at the nodes of any positive algebraic quadrature formula. This is, to our knowledge, the first application of quadrature theory within polynomial optimization theory. In this framework, the use of minimal (if feasible) or more generally low-cardinality formulas is of great interest.



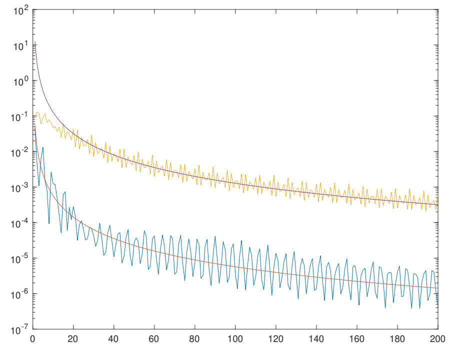


Figure 1: top: 45 Padua Points of exactness degree 14 ($d = 5$); bottom: relative gap of approximate minimization at the Padua Points and best fitting by α/d^2 (upper graph: f_1, $\alpha = 12.78$; lower graph: f_2, $\alpha = 0.056$).

References

[1] L. Bos, M. Caliari, S. De Marchi, M. Vianello and Y. Xu, Bivariate Lagrange interpolation at the Padua points: the generating curve approach, J. Approx. Theory 143 (2006), 15–25.

[2] C. Caratheodory, über den Variabilittsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen, Rend. Circ. Mat. Palermo 32 (1911), 193–217.

[3] R. Cools, Constructing cubature formulae: the science behind the art, Acta Numer. 6 (1997), 1–54.

[4] R. Cools, An Encyclopaedia of Cubature Formulas, J. Complexity 19 (2003), 445–453.

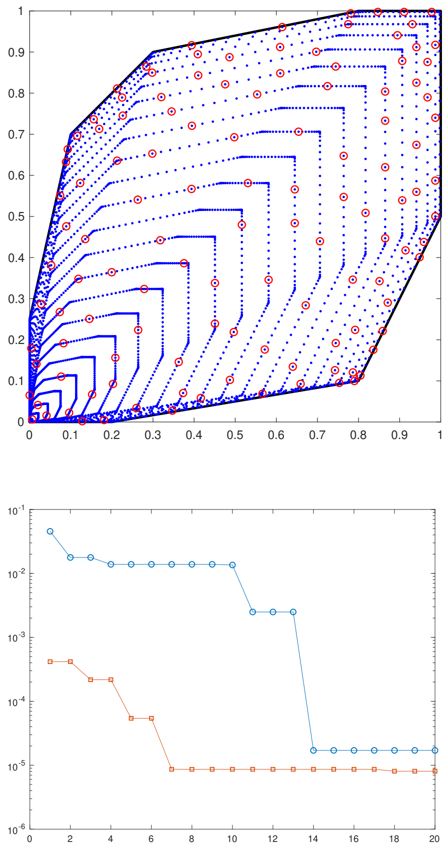


Figure 2: top: 120 Tchakaloff Points (circles) of exactness degree 14 ($d = 5$) in a 9-sides convex polygon, by nested extraction from 4232 points of a polygon rule of exactness degree 20; bottom: relative gaps of approximate minimization at the Tchakaloff Points for f_1 (circles) and f_2 (squares).

[5] E. de Klerk, D. Kuhn and K. Postek, Distributionally robust optimization with polynomial densities: theory, models and algorithms, arXiv: 1805.03588.

[6] E. de Klerk, J.-B. Lasserre, M. Laurent and Z. Sun, Bound-constrained polynomial optimization using only elementary calculations, Math. Oper. Res. 42 (2017), 834–853.

[7] E. de Klerk and M. Laurent, Comparison of Lasserre's measure-based bounds for polynomial optimization to bounds obtained by simulated annealing, Math. Oper. Res. 43 (2018), 1317–1325.

[8] E. de Klerk and M. Laurent, Worst-case examples for Lasserre's measure based hierarchy for polynomial optimization on the hypercube, Math. Oper. Res., to appear.

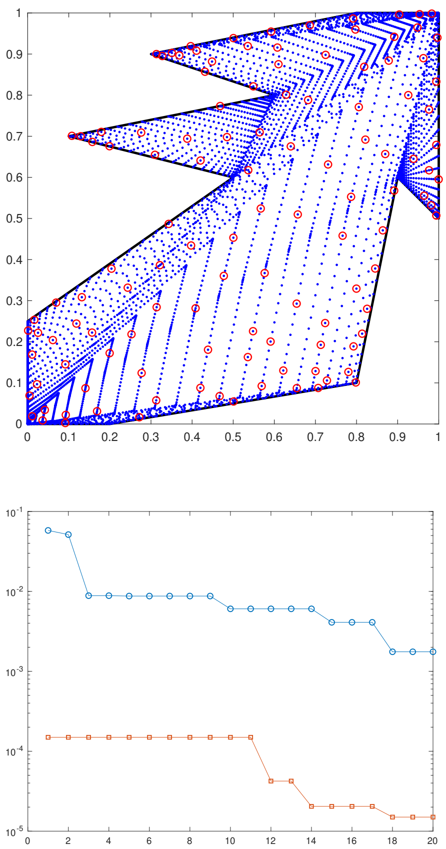


Figure 3: top: 120 Tchakaloff Points (circles) of exactness degree 14 ($d = 5$) in a 12-sides nonconvex polygon, by nested extraction from 5819 points of a polygon rule of exactness degree 20; bottom: relative gaps of approximate minimization at the Tchakaloff Points for f_1 (circles) and f_2 (squares).

[9] E. de Klerk, R. Hess and M. Laurent, Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization, SIAM J. Optim. 27 (2017), 346-367.

[10] E. de Klerk, M. Laurent and Z. Sun, Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization, Math. Program. Ser. A 162(1) (2017), 363–392.

[11] J.D. Jakeman and A. Narayan, Generation and application of multivariate polynomial quadrature rules, Comput. Methods Appl. Mech. Engrg. 338 (2018), 134–161.

[12] J.B. Lasserre, A new look at nonnegativity on closed sets and polynomial optimization, SIAM J. Optim. 21 (2011), 864–885.

[13] H.M. Möller, Kubaturformeln mit minimaler Knotenzahl, Numer. Math. 25 (1975/76), no. 2, 185–200.

[14] S.E. Mousavi and N. Sukumar, Numerical integration of polynomials and discontinuous functions on irregular convex polygons and polyhedrons, Comput. Mech. 47 (2011), 535–554.

[15] B.N. Parlett, The symmetric eigenvalue problem, Classics in Applied Mathematics 20, SIAM, Philadelphia, 1998.

[16] F. Piazzon, A. Sommariva and M. Vianello, Caratheodory-Tchakaloff Least Squares, Sampling Theory and Applications 2017, IEEE Xplore Digital Library, DOI: 10.1109/SAMPTA.2017.8024337.

[17] F. Piazzon and M. Vianello, A note on total degree polynomial optimization by Chebyshev grids, Optim. Lett. 12 (2018), 63–71.

[18] F. Piazzon and M. Vianello, Markov inequalities, Dubiner distance, norming meshes and polynomial optimization on convex bodies, Optim. Lett., published online 01 January 2019.

[19] M. Putinar, A note on Tchakaloff's theorem, Proc. Amer. Math. Soc. 125 (1997), 2409–2414.

[20] E.K. Ryu and S.P. Boyd, Extensions of Gauss quadrature via linear programming, Found. Comput. Math. 15 (2015), 953–971.

[21] A. Sommariva and M. Vianello, POLYGAUSS: a Matlab code for Gauss-like cubature over polygons, downloadable at: `http://www.math.unipd.it/~alvise/software.html`.

[22] A. Sommariva and M. Vianello, Compression of multivariate discrete measures and applications, Numer. Funct. Anal. Optim. 36 (2015), 1198–1223.

[23] A. Sommariva and M. Vianello, Nearly optimal nested sensors location for polynomial regression on complex geometries, Sampl. Theory Signal Image Process. 17 (2018), 95–101.

[24] V. Tchakaloff, Formules de cubature mécaniques à coefficients nonnégatifs, Bull. Sci. Math. 81 (1957), 123–134.

[25] M. Vianello, Subperiodic Dubiner distance, norming meshes and trigonometric polynomial optimization, Optim. Lett. 12 (2018), 1659–1667 .

[26] H. Xiao and Z. Gimbutas, A numerical algorithm for the construction of efficient quadrature rules in two and higher dimensions, Comput. Math. Appl. 59 (2010), 663–676.

[27] Y. Xu, Minimal cubature rules and polynomial interpolation in two variables, J. Approx. Theory 164 (2012), 6–30.

[28] J.F. Zhang and C.P. Kwong, Some applications of a polynomial inequality to global optimization, J. Optim. Theory Appl. 127 (2005), 193–205.