J Glob Optim (2009) 45:39–61
DOI 10.1007/s10898-008-9394-7

Moments and sums of squares for polynomial optimization and related problems

Jean B. Lasserre

Received: 10 December 2008 / Accepted: 17 December 2008 / Published online: 6 January 2009

Abstract We briefly review the duality between moment problems and sums of squares (s.o.s.) representations of positive polynomials, and compare s.o.s. versus nonnegative polynomials. We then describe how to use such results to define convergent semidefinite programming relaxations in polynomial optimization as well as for the two related problems of computing the convex envelope of a rational function and finding *all* zeros of a system of polynomial equations.

Keywords Measures · Moments

1 Introduction

Relatively recent results from the theory of moments (and its dual theory of positive polynomials) coupled with semidefinite programming have allowed to develop efficient numerical schemes for *polynomial optimization* (or polynomial programming), i.e., global optimization problems with polynomial data. This numerical scheme consists of a hierarchy of semidefinite programs (SDP) of increasing size which define tighter and tighter relaxations of the original problem, and whose associated sequence of optimal values converges to the global minimum; see e.g. [15,16,28,29]. It is remarkable that standard convex duality of semidefinite programming perfectly expresses the duality between moments and positive polynomials.

Like *primal* approaches in nonlinear programming that search for a local minimizer $\tilde{x} \in \mathbb{R}^n$, the moment approach should also be regarded as a primal approach where one now searches not only for a global minimizer $x^* \in \mathbb{R}^n$ but also for the sequence of moments $y^* = (x^*)^\alpha$, $\alpha \in \mathbb{N}^n$, i.e., a search in a *lifted* space (of moments). Similarly, like *dual* approaches in nonlinear programming (e.g. Lagrangian and extended Lagrangian) that search for *scalar* Karush-Kuhn-Tucker *multipliers* associated with the constraints, the s.o.s. approach

J. B. Lasserre (✉)
LAAS-CNRS and Institute of Mathematics, LAAS, 7 Avenue du Colonel Roche,
31077 Toulouse Cédex 4, France
e-mail: lasserre@laas.fr

should also be viewed as a dual approach as one also searches for multipliers of the constraints, but now s.o.s. (instead of scalar) multipliers, hence also in a lifted space (of polynomials).

When data are polynomials (or even in some cases rational functions) this lifting process permits to obtain convergence to a global optimum because then algebra enters the game and one may invoke powerful representation results from real algebraic geometry. Moreover, and crucial for practical computation, this lifting process translates into semidefinite programs, a well studied class of convex optimization problems for which efficient public software are available. Of course, as the general polynomial optimization problem is NP-hard, it is very unlikely to obtain a numerical scheme with guaranteed efficiency for *all* problem instances (but practice seems to reveal that SDP-relaxations exhibit fast and often finite convergence).

In this paper we review basic principles of this methodology and develop in some detail its application in polynomial optimization, as well as in some related problems. We also compare nonnegative with sums of squares (s.o.s.) polynomials and describe the duality between moment problems and s.o.s. representations of positive polynomials.

One goal is to convince the reader that Putinar's Positivstellensatz [31] (a powerful representation theorem of real algebraic geometry) provides a non convex analogue for *global* polynomial optimization of the celebrated Karush-Kuhn-Tucker (KKT) optimality conditions in convex programming (whereas the latter are only necessary (local) optimality conditions in non linear programming). To better appreciate its power, call a constraint $g_j(x) \geq 0$ *important* if when removed from the definition of the feasible set, the optimum decreases strictly. Except for the convex case, an important constraint need *not* be active at a global minimizer (think about 0–1 programs with linear inequality constraints). In KKT conditions, an important constraint which is not active at a global minimizer is ignored as its nonnegative KKT-multiplier λ_j^* vanishes whereas in Putinar's Positivstellensatz, an important constraint is recognized by the fact that its multiplier (now a s.o.s. polynomial σ_j) is not trivial (but like the KKT multiplier, it vanishes at every global minimizers x^*, i.e. $\sigma_j(x^*) = \lambda_j^* = 0$). In fact the value $\sigma_j(x^*)$ of s.o.s. multipliers at any global minimizer x^* is precisely the value of the corresponding multiplier λ_j^* in KKT conditions.

We also present the specialized s.o.s. representation results of the author [18] for problems with a sparsity pattern. Indeed, and despite their nice features, the size of the SDP-relaxations grows rapidly with the size of the original problem. Typically, the kth SDP-relaxation in the hierarchy has to handle at least one LMI of size $\binom{n+k}{n}$ and $\binom{n+2k}{n}$ variables, which clearly limits the applicability of the methodology to problems with small to medium size only. Fortunately, most large size problems exhibit some sparsity pattern. Indeed, as typical in problems with a large number of variables, each of the polynomials that describe the constraints of the problem is "sparse", i.e., involves a few variables only and the objective polynomial is very often a sum of sparse polynomials. This motivated our specialized s.o.s. representation results [18] which in particular permitted to prove convergence of the hierarchy of sparse SDP-relaxations introduced by Waki et al. [38], and in which sparsity in the problem data is translated into SDP-relaxations of much smaller size. For instance, in [38] the authors could solve problems with up to a thousand variables for which the first standard SDP-relaxation with no sparsity cannot be implemented (at least with the SDP solvers currently available).

We show how to apply the above methodology for the global minimization of a rational function $f = p/q$ on a compact basic semi-algebraic set $\mathbf{K} \subset \mathbb{R}^n$, and also for the *pointwise* computation of the convex envelope $\widehat{f}$ of f on the convex hull co($\mathbf{K}$) of $\mathbf{K}$. Finally, we also consider the related problem of computing *all* real zeros of a system of polynomial equations for which the above methodology can be adapted to yield a new semidefinite characterization of zero-dimensional real radical ideals [22]. In contrast to previous algebraic approaches which compute all real and complex solutions via e.g. homotopy or Gröbner base methods, the SDP approach of [22] is real algebraic in nature as it avoids computing any complex zero.

At last but not least, it is worth emphasing that appropriate SDP-relaxations can be defined for a variety of seemingly very different problems but which look the same when viewed as a particular instance of the so-called *generalized problem of moments* (GPM)

$$\text{GPM}: \quad \min_{\mu \in \mathcal{M}(\mathbf{K})} \left\{ \int f_0 \, \mathrm{d}\mu \; : \int f_j \mathrm{d}\mu = b_j, \quad j = 1, \ldots, m \right\} \tag{1.1}$$

where $\mathcal{M}(\mathbf{K})$ is a convex set of measures on $\mathbf{K} \subset \mathbb{R}^n$, such that all f_j's are integrable with respect to every measure $\mu \in \mathcal{M}(\mathbf{K})$; for more details see Lasserre [20]. To our knowledge, GloptiPoly3 [10] (an extension of GloptiPoly [7]) is the first software package devoted to solve the GPM (at least small to medium size problems). The GPM has developments and impact in various area of Mathematics like algebra, Fourier analysis, functional analysis, operator theory, probability and statistics. It also has a large number of important applications in various fields like optimization, probability, finance, control, signal processing, chemistry, cristallography, tomography, etc. For an account of various methodologies as well as some of potential applications, the interested reader is referred to e.g. Akhiezer [1], Akhiezer and Krein [2], the nice collection of papers [13] and [20].

The paper is organized as follows: We first introduce some notation and definitions and then review some basic results on moment problems and s.o.s. representations of positive polynomials. In Sect. 3 we provide a hierarchy of SDP-relaxations for computing the global minimum of a rational function on a compact basic semi-algebraic set. Convergence results as well as a sufficient condition to detect finite convergence (hence global optimality) are presented. For clarity of exposition, the proof of the main theorem is postponed to an appendix. In Sect. 4 we compare Putinar Positivstellensatz with the celebrated KKT local optimality conditions. We also provide a hierarchy of SDP-relaxations for the pointwise computation of the convex envelope of a rational function, as well as for computing all real zeros of a system of polynomials equations.

2 Moments and sums of squares

2.1 Notation

In $\mathbb{R}^n$ we always consider the usual Borel σ-algebra $\mathcal{B}$ and so a finite measure on $\mathbb{R}^n$ is always understood as a finite Borel measure on $\mathcal{B}$.

For a real symmetric matrix A, the notation $A \succeq 0$ (resp. $A \succ 0$) stands for A is positive semidefinite (resp. positive definite), whereas u^T denotes the transpose of a vector u. Let $\mathbb{N}$ be the set of natural numbers, and let $\mathbb{R}[X]$ ($= \mathbb{R}[X_1, \ldots, X_n]$) be the ring of real polynomials in the n variables $X_1, \ldots, X_n$. Let $\Sigma^2 \subset \mathbb{R}[X]$ be the set of polynomials that are sums of squares (s.o.s.).

With $d \in \mathbb{N}$, let $s(d) := \binom{n+d}{n}$, and let $u_d(X) \in \mathbb{R}^{s(d)}$ be the column vector

$$u_d(X) = (1, X_1, \ldots, X_n, X_1^2, X_1X_2, \ldots, X_n^d)^T,$$

whose components form the usual canonical basis of the vector space $\mathbb{R}[X]_d$ (of dimension $s(d)$) of real polynomials of degree at most d.

Given a infinite sequence $y := \{y_\alpha\}_{\alpha\in\mathbb{N}^n}$ indexed in the canonical basis $u_\infty(X)$, let $L_y : \mathbb{R}[X] \to \mathbb{R}$ be the linear mapping

$$f \in \mathbb{R}[X] \left(= \sum_{\alpha\in\mathbb{N}^n} f_\alpha X^\alpha \right) \longmapsto L_y(f) := \sum_{\alpha\in\mathbb{N}^n} f_\alpha\, y_\alpha, \tag{2.1}$$

and let $\mathbf{f} = \{f_\alpha\} \in \mathbb{R}^{s(d)}$ be the vector of coefficients of $f \in \mathbb{R}[X]_d$ in the basis $u_d(X)$.
Moment matrix Let $M_d(y)$ be the $s(d) \times s(d)$ real matrix with rows and columns indexed in the basis $u_d(X)$, and defined by:

$$M_d(y)(\alpha, \beta) = y_{\alpha+\beta}, \qquad \alpha, \beta \in \mathbb{N}^n, \quad |\alpha|, |\beta| \le d, \tag{2.2}$$

where for every $\alpha \in \mathbb{N}^n$, the notation $|\alpha|$ stands for $\sum_{i=1}^n \alpha_i$.

Equivalently, $M_d(y) = L_y(u_d(X)u_d(X)^T)$, meaning that L_y is applied *entrywise* to the polynomial matrix $u_d(X)u_d(X)^T$. The matrix $M_d(y)$ is called the *moment* matrix associated with the sequence y (see e.g. Curto and Fialkow [6] and Lasserre [16]). If y has a *representing measure* μ_y (i.e., if $y_\alpha = \int X^\alpha d\mu_y$ for every $\alpha \in \mathbb{N}^n$) then

$$\langle \mathbf{f}, M_d(y)\mathbf{f}\rangle = \int f^2\, d\mu_y \ \ge\ 0, \qquad \forall f \in \mathbb{R}[X]_d, \tag{2.3}$$

so that $M_d(y) \succeq 0$. A measure μ is said to be moment *determinate* if there is no other measure with same moments. In particular, and as an easy consequence of the Stone-Weierstrass theorem, every measure with compact support is moment determinate; see e.g. Berg [4], or Maserick and Berg [24].

There is a nice sufficient condition to ensure that a sequence y has a unique representing measure. It is due to Nussbaum [27] and is an extension to the multivariate case of Carleman's condition in the univariate case. Namely, if

$$\sum_{k=1}^{\infty} L_y(X_i^{2k})^{-1/2k} = +\infty, \qquad i = 1, \ldots, n \tag{2.4}$$

then y has a unique representing measure; see e.g. Berg [4, Theor. 5].

Similarly, if for some $a, c > 0$, $|y_\alpha| \le c a^{|\alpha|}$ for all $\alpha \in \mathbb{N}^n$, then y has a unique representing measure, with support contained in the ball $[-a, a]^n$; again see Berg [4, Theor. 9]. Finally, if the marginal distributions of a measure μ are determinate, then so is μ; see Petersen [30].
Localizing matrix Similarly, given $y = \{y_\alpha\}$ and $\theta \in \mathbb{R}[X]$, let $M_d(\theta y)$ be the $s(d) \times s(d)$ matrix defined by:

$$M_d(\theta y) := L_y(\theta(X)u_d(X)u_d(X)^T),$$

i.e., L_y is applied entrywise to the matrix polynomial $\theta(X)u_d(X)u_d(X)^T$. The matrix $M_d(\theta y)$ is called the *localizing* matrix associated with the sequence y and the polynomial θ (see again Lasserre [16]). Notice that the localizing matrix with respect to the constant polynomial $\theta \equiv 1$ is the moment matrix $M_d(y)$ in (2.2).

If y has a representing measure μ_y with support contained in the level set $\{x \in \mathbb{R}^n : \theta(x) \ge 0\}$ (where $\theta \in \mathbb{R}[X]$), then

$$\langle \mathbf{f}, M_d(\theta y)\mathbf{f}\rangle = \int f^2\theta\, d\mu_y \ \ge\ 0 \qquad \forall\, f \in \mathbb{R}[X]_d, \tag{2.5}$$

so that $M_d(\theta y) \succeq 0$.

2.2 Nonnegative versus s.o.s. polynomials

When is a nonnegative homogeneous polynomial $f \in \mathbb{R}[X]_{2d}$ a s.o.s. polynomial? It was recognized and proved by Hilbert as early as 1888 that apart from the three special cases $(n, 2d) = (1, 2d), (n, 2), (2, 4)$, not all nonnegative polynomials are s.o.s. and this gave birth to Hilbert's 17th problem on the representation of nonnegative polynomials as a sum of squares of *rational functions*, later solved by E. Artin in 1927. Hilbert's result was not constructive and amazingly, only in 1967 was produced the first concrete example of a non-negative polynomial that is not s.o.s., the well-known bivariate Motzkin polynomial.

$$(X, Y) \mapsto 1 + X^2Y^2(X^2 + Y^2 - 3)$$

When compared to nonnegative polynomials, s.o.s. polynomials have a important feature which is very attractive from a computational viewpoint. Whereas one does *not* how to check efficiently whether a given polynomial $f \in \mathbb{R}[X]_{2d}$ is nonnegative, checking whether f is s.o.s. reduces to solving a *semidefinite program* (SDP), a convex optimization problem with nice computational complexity and for which efficient (public software) solvers are now available. Indeed, write

$$u_d(X)\, u_d(X)^T = \sum_{\alpha\in\mathbb{N}^n} B_\alpha X^\alpha,$$

for some symmetric matrices $(B_\alpha) \in \mathbb{R}^{s(d)\times s(d)}$. Then $f \in \mathbb{R}[X]_{2d}$ is s.o.s. if and only if there exists some positive semidefinite matrix $F \in \mathbb{R}^{s(d)\times s(d)}$ such that

$$F \succeq 0; \quad f_\alpha = \mathrm{trace}(B_\alpha F), \qquad \forall \alpha \in \mathbb{N}^n. \tag{2.6}$$

Checking (2.6) is just solving a SDP.

Therefore it is important to understand the gap between nonnegative and s.o.s. polynomials. On the negative side, Blekherman [5] showed that when the degree d is *fixed* there are many more nonnegative polynomials than s.o.s. and the larger n, the larger the gap. On the positive side, introducing the l_1-norm in $\mathbb{R}[X]$

$$\left(f = \sum_{\alpha\in\mathbb{N}^n} f_\alpha X^\alpha\right) \mapsto \|f\|_1 := \sum_{\alpha\in\mathbb{N}^n} |f_\alpha|,$$

Berg [4] showed that the cone Σ^2 is dense in the space of polynomials that are nonnegative on the box $[-1, 1]^n \subset \mathbb{R}^n$. But this result is not constructive and we provide below the following more precise results. Given $r \in \mathbb{N}$ arbitrary, introduce the two polynomials $\Theta_r, \theta_r \in \mathbb{R}[X]$

$$X \mapsto \theta_r(X) := \sum_{i=1}^n \sum_{k=0}^r \frac{X_i^{2k}}{k!}; \quad X \mapsto \Theta_r(X) := 1 + \sum_{i=1}^n X_i^{2r} \tag{2.7}$$

and given $f \in \mathbb{R}[X]$ and $\epsilon > 0$, define

$$f^1_{\epsilon r} := f + \epsilon\,\theta_r; \quad f^2_{\epsilon r} := f + \epsilon\,\Theta_r.$$

Theorem 2.1 (Lasserre [19], Lasserre and Netzer [21])

(a) *If $f \in \mathbb{R}[X]$ is nonnegative then for every $\epsilon > 0$ there exists $r(\epsilon, f)$ such that $f^1_{\epsilon r}$ is s.o.s. for all $r \geq r(\epsilon, f)$ and $\|f - f^1_{\epsilon r}\|_1 \to 0$ as $\epsilon \downarrow 0$ (and $r \geq r(\epsilon, f)$).*
(b) *If $f \in \mathbb{R}[X]$ is nonnegative on $[-1, 1]^n$ then for every $\epsilon > 0$ there exists $r(\epsilon, f)$ such that $f^2_{\epsilon r}$ is s.o.s. for all $r \geq r(\epsilon, f)$ and $\|f - f^2_{\epsilon r}\|_1 \to 0$ as $\epsilon \downarrow 0$ (and $r \geq r(\epsilon, f)$).*

For a detailed proof see Lasserre [19] and Lasserre and Netzer [21]. Theorem 2.1(b) provides an explicit construction of an approximating sequence of s.o.s. for the denseness result of Berg [4]. It suffices to add to f essentials monomials X_i^{2r} with a coefficient $\epsilon > 0$ and sufficiently high degree $2r$. Observe that in addition to the l_1-norm convergence $\|f - f^1_{\epsilon r}\|_1 \to 0$, the convergence is also *uniform* on compact sets! Notice that a polynomial f nonnegative on the whole $\mathbb{R}^n$ (hence on the box $[-1, 1]^n$) can also be approximated by the s.o.s. polynomial $f^2_{\epsilon r}$ of Theorem 2.1(b) which is simpler than the s.o.s. approximation $f^1_{\epsilon r}$. However, in contrast to the latter, the approximation $f \approx f^2_{\epsilon r}$ is *not* uniform on compact sets, and is really more appropriate for polynomials nonnegative on $[-1, 1]^n$ only (and indeed the approximation $f \approx f^2_{\epsilon r}$ is uniform on $[-1, 1]^n$). In [21] we also prove that $r(f, \epsilon)$ in Theorem 2.1(b) does *not* depend on the *explicit choice* of the polynomial f but only on:

- ϵ and the dimension n,
- the degree and the size of the coefficients of f.

Therefore, if one fixes these four parameters, we find an r such that the statement of Theorem 2.1(b) holds for any f nonnegative on $[-1, 1]^n$, whose degree and size of coefficients do not exceed the fixed parameters.

Finally, consider the case of a (not necessarily compact) real variety $V \subset \mathbb{R}^n$ defined by

$$V := \{ x \in \mathbb{R}^n \ : \quad g_j(x) = 0, \quad j = 1, \ldots, m \} \tag{2.8}$$

for some polynomials $g_j \in \mathbb{R}[X]$.

Theorem 2.2 (Lasserre [17]) *Let $V \subset \mathbb{R}^n$ be the real variety in (2.8), and let θ_r be as in (2.7). If $f \in \mathbb{R}[X]$ is nonnegative on V then for every $\epsilon > 0$, there exists $r(\epsilon, f) \in \mathbb{N}$ and nonnegative scalars $\{\lambda_j\}_{j=1}^m$ such that for all $r \geq r(\epsilon)$,*

$$f^1_{\epsilon r} \, (= f + \epsilon \, \theta_r) = q - \sum_{j=1}^m \lambda_j \, g_j^2 \tag{2.9}$$

for some s.o.s. polynomial $q \in \Sigma^2$. In addition, $\|f - f^1_{\epsilon r}\|_1 \to 0$, as $\epsilon \downarrow 0$.

Again notice that (2.9) is a certificate of positivity of f on V and the approximation $f \approx f^1_{\epsilon r}$ on V, is uniform on compact subsets of V.

We next present two powerful representation results for a polynomial f positive on a *compact* basic semi-algebraic set $\mathbf{K} \subset \mathbb{R}^n$, and their dual counterparts, i.e., conditions for a sequence $y = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, to have a representing measure μ with support contained in $\mathbf{K}$.

2.3 Positivstellensatz for compact sets and their dual version

We here present some basic fondamental results on the duality between moments and sums of squares which transpires in so-called Positivstellensatz theorems on the representation of polynomial positive on a compact basic semi-algebraic set.

Let $\mathbf{K} \subset \mathbb{R}^n$ be the basic closed semi-algebraic set defined by

$$\mathbf{K} := \{ x \in \mathbb{R}^n \quad | \quad g_j(x) \geq 0, \quad j = 1, \ldots, m\} \tag{2.10}$$

for some family $\{g_j\}_{j=1}^m \subset \mathbb{R}[X]$, and let $g_0 \equiv 1$.

Next, denote by $P(g) \subset \mathbb{R}[X]$ (resp. $Q(g) \subset \mathbb{R}[X]$) the *preordering* (resp. the *quadratic module*) generated by the g_j's. That is, $f \in P(g)$ if

$$f = \sum_{J \subseteq \{1,\ldots,m\}} \sigma_J \left(\prod_{j \in J} g_j \right) \quad \text{with} \quad \sigma_J \in \Sigma^2 \quad \forall\, J \subseteq \{1,\ldots,m\} \tag{2.11}$$

(with the convention $g_J := \prod_{j \in J} g_j \equiv 1$ if $J = \emptyset$). Similarly, $f \in Q(g)$ if

$$f = \sum_{j=0}^{m} \sigma_j\, g_j \quad \text{with} \quad \sigma_j \in \Sigma^2 \quad \forall\, j = 0, 1, \ldots, m. \tag{2.12}$$

Theorem 2.3 (Schmüdgen [34]) *Assume that* **K** *defined in (2.10) is compact.*

(a) *If $f \in \mathbb{R}[X]$ is strictly positive on* **K** *then $f \in P(g)$, i.e.,*

$$f = \sum_{J \subseteq \{1,\ldots,m\}} \sigma_J \left[\prod_{j \in J} g_j \right] \tag{2.13}$$

for some s.o.s. polynomials $\{\sigma_J\} \subset \Sigma^2$.

(b) *Let $y = (y_\alpha) \subset \mathbb{R}$ be an infinite sequence indexed in the canonical basis (X^α) of $\mathbb{R}[X]$. Then y has a representing measure with support contained in* **K** *if and only if*

$$L_y \left(f^2 \left[\prod_{j \in J} g_j \right] \right) \geq 0 \qquad \forall J \subseteq \{1,\ldots,m\}, \quad \forall f \in \mathbb{R}[X]. \tag{2.14}$$

This powerful and elegant result shows that two apriori different results in functional analysis and algebra are in fact two dual facets of the same problem : representing a polynomial positive on **K** and representing a moment sequence by a measure with support contained in **K**. Indeed, in Theorem 2.3(a) one obtains a certificate of positivity on **K** for a polynomial $f \in \mathbb{R}[X]$, whereas in (b) one obtains a constructive means of checking whether a sequence y has a representing measure on **K**.

In addition, testing whether (2.13) is satisfied with polynomials σ_J of degree at most say $2d$, reduces to solving a SDP. Similarly testing the condition (2.14) for all $f \in \mathbb{R}[X]_d$ also reduces to solving a SDP. However notice that in (2.13) the summation involves 2^m terms, and similarly in (2.14) there are 2^m conditions to test. From a computational viewpoint this is rather annoying. Fortunately, under some mild condition on the polynomials g_j's that describe **K**, Putinar [31] and Jacobi and Prestel [11] proved a more convenient representation result.

Assumption 2.1 The set **K** in (2.10) is compact. There exists $u \in \mathbb{R}[X]$ such that $u = u_0 + \sum_{j=1}^m u_j g_j$ for some s.o.s. polynomials $\{u_j\}_{j=0}^m \subset \Sigma^2$, and the level set $\{x \in \mathbb{R}^n \mid u(x) \geq 0\}$ is compact.

Remark 2.4 For instance, Assumption 2.1 is automatically satisfied if

- all g_j's are affine in which case **K** is a polytope.
- the level set $\{x \in \mathbb{R}^n \;:\; g_j(x) \geq 0\}$ is compact for some $j \in \{1,\ldots,m\}$.

In addition if $M - \|x\|^2 \geq 0$ for all $x \in$ **K** then it suffices to add the redundant constraint $g_{m+1}(x) \geq 0$ in the definition (2.10) of **K** (with $g_{m+1} \in \mathbb{R}[X]$ being the quadratic polynomial $M - \|X\|^2$), and Assumption 2.1 is satisfied.

Theorem 2.5 (Putinar [31]) *With* $\mathbf{K} \subset \mathbb{R}^n$ *as in (2.10), let Assumption 2.1 hold.*

(a) *If* $f \in \mathbb{R}[X]$ *is strictly positive on* $\mathbf{K}$ *then* $f \in Q(g)$*, i.e.,*

$$f = \sum_{j=0}^{m} \sigma_j \, g_j \tag{2.15}$$

for some s.o.s. polynomials $\{\sigma_j\} \subset \Sigma^2$.

(b) *Let* $y = (y_\alpha) \subset \mathbb{R}$ *be an infinite sequence indexed in the canonical basis* (X^α) *of* $\mathbb{R}[X]$*. Then* y *has a representing measure with support contained in* $\mathbf{K}$ *if and only if*

$$L_y(f^2 \, g_j) \geq 0 \quad \forall j = 0, \dots, m, \quad \forall f \in \mathbb{R}[X]. \tag{2.16}$$

The power of Theorem 2.5 is to replace the 2^m terms in (2.13) or the 2^m conditions in (2.14) with only $m+1$ terms or conditions in (2.15) and (2.16) respectively, a very attractive feature from a computational viewpoint. In addition, the price to pay is very small as in the worst case it suffices to add a redundant quadratic constraint in the definition (2.10) of $\mathbf{K}$; see Remark 2.4 above.

2.4 Specialized representation results with sparsity properties

If there is no *coupling* between some subsets of variables in the polynomial f or the polynomials g_j that define the set $\mathbf{K}$, a natural question that arises is whether there exists a representation result (or a Positivstellensatz) that preserves this property. In view of the practical computational implications, this is a very important issue because in many (if not most) problems with a large number of variables, some *sparsity* pattern is present as all monomials of the polynomial data f and g_j often involve a few variables only.

First consider the simple case of three sets of variables. Let $\mathbb{R}[X, Y, Z]$ be the ring of real polynomial in the variables $(X_1, \dots, X_n)$, $(Y_1, \dots, Y_m)$ and $(Z_1, \dots, Z_p)$. Let $\mathbf{K}_{xy} \subset \mathbb{R}^{n+m}$, $\mathbf{K}_{yz} \subset \mathbb{R}^{m+p}$, and $\mathbf{K} \subset \mathbb{R}^{n+m+p}$ be basic compact semi-algebraic sets defined by

$$\mathbf{K}_{xy} = \{ (x, y) \in \mathbb{R}^{n+m} \mid \; g_j(x, y) \geq 0, \quad j \in \mathbf{I}_{xy} \} \tag{2.17}$$

$$\mathbf{K}_{yz} = \{ (y, z) \in \mathbb{R}^{m+p} \mid \; h_k(y, z) \geq 0, \quad k \in \mathbf{I}_{yz} \} \tag{2.18}$$

$$\mathbf{K} = \{ (x, y, z) \in \mathbb{R}^{n+m+p} \mid \; (x, y) \in \mathbf{K}_{xy}; \quad (y, z) \in \mathbf{K}_{yz} \} \tag{2.19}$$

for some polynomials $\{g_j\} \subset \mathbb{R}[X, Y]$, $\{h_k\} \subset \mathbb{R}[Y, Z]$, and some finite index sets $\mathbf{I}_{xy}, \mathbf{I}_{yz} \subset \mathbb{N}$. Let $P(g) \subset \mathbb{R}[X, Y]$ and $P(h) \subset \mathbb{R}[Y, Z]$ be the preordering generated by $\{g_j\}_{j \in \mathbf{I}_{xy}}$ and $\{h_k\}_{k \in \mathbf{I}_{yz}}$, respectively. Similarly, let $Q(g) \subset \mathbb{R}[X, Y]$ and $Q(h) \subset \mathbb{R}[Y, Z]$ denote the quadratic modules.

Theorem 2.6 *Let* $\mathbf{K}_{xy} \subset \mathbb{R}^n$, $\mathbf{K}_{yz} \subset \mathbb{R}^m$*, and* $\mathbf{K} \subset \mathbb{R}^{n+m+p}$ *be the basic compact semi-algebraic sets defined in (2.17)-(2.10), and assume that* $\mathbf{K}$ *has nonempty interior. Let* $f \in \mathbb{R}[X, Y] + \mathbb{R}[Y, Z]$.

(a) *If* f *is positive on* $\mathbf{K}$ *then* $f \in P(g) + P(h)$.

(b) *If* $N - \|(X, Y)\|^2 \in Q(g)$ *and/or* $N - \|(Y, Z)\|^2 \in Q(h)$ *for some scalar* N*, and if* f *is positive on* $\mathbf{K}$*, then in (a) one may replace* $P(g)$ *with* $Q(g)$ *and/or* $P(h)$ *with* $Q(h)$.

Hence the absence of coupling between the two sets of variables X and Z in the original data f, g_j, h_k, is also reflected in the specialized *sparse* representations of Theorem 2.6(a)-(b). Indeed one may replace the preordering $P(g, h) \subset \mathbb{R}[X, Y, Z]$ (resp. the quadratic module $Q(g, h) \subset \mathbb{R}[X, Y, Z]$) generated by the g_j's and h_k's with the smaller preorderings $P(g) \subset \mathbb{R}[X, Y]$ and $P(h) \subset \mathbb{R}[Y, Z]$ (resp. the quadratic modules $Q(g)$ and $Q(h) \subset \mathbb{R}[Y, Z]$).

Remarkably, Theorem 2.6 can be extended to more general sparsity patterns provided that some condition called the *running intersection property* (which incidentally happens to be well-known in the graph theory) holds.

With $\mathbb{R}[X] = \mathbb{R}[X_1, \ldots, X_n]$, let $I_0 := \{1, \ldots, n\}$ be the union $\cup_{k=1}^{p} I_k$ of p subsets I_k, $k = 1, \ldots, p$, with cardinal denoted n_k. Let $\mathbb{R}[X(I_k)]$ denote the ring of polynomials in the n_k variables $X(I_k) = \{X_i \mid i \in I_k\}$, and so $\mathbb{R}[X(I_0)] = \mathbb{R}[X]$.

Assumption 2.2 Let $\mathbf{K} \subset \mathbb{R}^n$ be as in (2.10). There is $M > 0$ such that $\|x\|_\infty < M$ for all $x \in \mathbf{K}$.

In view of Assumption 2.2, one has $\|X(I_k)\|^2 \leq n_k M^2$, $k = 1, \ldots, p$, and therefore, in the definition (2.10) of $\mathbf{K}$, we add the p redundant quadratic constraints

$$g_{m+k}(x) \geq 0 \quad \text{with} \quad g_{m+k}(X) := n_k M^2 - \|X(I_k)\|^2, \quad k = 1, \ldots, p, \tag{2.20}$$

and set $m' = m + p$, so that $\mathbf{K}$ is now defined by:

$$\mathbf{K} := \{x \in \mathbb{R}^n \mid \quad g_j(x) \geq 0, \quad j = 1, \ldots, m'\}. \tag{2.21}$$

Notice that $g_{m+k} \in \mathbb{R}[X(I_k)]$, for all $k = 1, \ldots, p$.

Assumption 2.3 Let $\mathbf{K} \subset \mathbb{R}^n$ be as in (2.21). The index set $J = \{1, \ldots, m'\}$ is partitioned into p disjoint sets J_k, $k = 1, \ldots, p$, and the collections $\{I_k\}$ and $\{J_k\}$ satisfy:

(i) For every $j \in J_k$, $g_j \in \mathbb{R}[X(I_k)]$, that is, for every $j \in J_k$, the constraint $g_j(x) \geq 0$ is only concerned with the variables $X(I_k) = \{X_i \mid i \in I_k\}$.
(ii) The objective function $f \in \mathbb{R}[X]$ can be written

$$f = \sum_{k=1}^{p} f_k, \quad \text{with } f_k \in \mathbb{R}[X(I_k)], \quad k = 1, \ldots, p. \tag{2.22}$$

Theorem 2.7 (Lasserre [18]) *Let Assumption 2.2 and 2.3 hold. Let $\mathbf{K} \subset \mathbb{R}^n$ be as in (2.21) (i.e. $\mathbf{K}$ as in (2.10) with the additional redundant quadratic constraints (2.20)), and with nonempty interior. Assume that for every $k = 1, \ldots, p-1$,*

$$I_{k+1} \bigcap \left(\bigcup_{j=1}^{k} I_j \right) \subseteq I_s \quad \textit{for some } s \leq k. \tag{2.23}$$

If $f \in \mathbb{R}[X]$ is strictly positive on $\mathbf{K}$ then

$$f = \sum_{k=1}^{p} \left(q_k + \sum_{j \in J_k} q_{jk}\, g_j \right), \tag{2.24}$$

for some s.o.s. polynomials $\{q_k, q_{jk}\} \subset \mathbb{R}[X(I_k)]$, $k = 1, \ldots, p$.

Hence under the running intersection property (2.23), the absence of coupling of variables in the original data is preserved in the representation (2.24). In addition to be of self-interest, Theorem 2.7 permits to prove convergence of the specialized (and efficient) sparse SDP-relaxations introduced in Waki et al. [38] for polynomial optimization problems with a large number of variables. For more details see Lasserre [18] and Sect. 4.1 below.

3 SDP-relaxations for polynomial optimization

Consider the global optimization problem:

$$\mathbf{P}: \quad f^* := \min_x \{ f(x) \; : \quad x \in \mathbf{K} \} \tag{3.1}$$

where $\mathbf{K} \subset \mathbb{R}^n$ is the basic semi-algebraic set in (2.10) and $f : \mathbb{R}^n \to \mathbb{R}$ is the rational function p/q with $p, q \in \mathbb{R}[X]$. Convergent SDP-relaxations for $\mathbf{P}$ when $q \equiv 1$ were first proposed in Lasserre [16] and later extended in Jibetean and de Klerk [12] for $q \neq 1$. One assumes that $q > 0$ on $\mathbf{K}$. Indeed if q changes sign on $\mathbf{K}$ then unless p and q have common real zeros, $f^* = -\infty$.

Let $v_j := \lceil \deg g_j/2 \rceil$ for every $j = 0, 1, \ldots, m$ (in particular $v_0 = 0$ as $g_0 \equiv 1$), and for $r \geq r_0 := \max[\deg p, \max_j v_j]$, consider the semidefinite program:

$$\mathbf{Q}_r : \quad \begin{cases} \inf\limits_y \; L_y(p) \\ \text{s.t} \;\; M_{r-v_j}(g_j y) \succeq 0, \quad j = 0, 1, \ldots, m \\ \qquad L_y(q) = 1 \end{cases}, \tag{3.2}$$

with optimal value denoted by $\inf \mathbf{Q}_r$ (and by $\min \mathbf{Q}_r$ if the infimum is achieved). Notice that by construction, $\mathbf{Q}_r$ contains only moment variables y_α with $|\alpha| \leq 2r$.

Proposition 3.1 *Let* $\mathbf{M}(\mathbf{K})$ *be the space of finite Borel measures on* $\mathbf{K}$ *and consider the optimization problem*

$$\mathcal{P} : \quad \min_{\mu \in \mathbf{M}(\mathbf{K})} \left\{ \int p \, d\mu \; : \; \int q \, d\mu = 1 \right\} \tag{3.3}$$

with optimal value denoted by $\inf \mathcal{P}$. *Then* $f^* = \inf \mathcal{P} = \min \mathcal{P}$ *and*

$$\inf \mathbf{Q}_r \leq f^* \, (= \inf \mathcal{P}) \qquad \forall r \geq \max_{j=1,\ldots,m} v_j. \tag{3.4}$$

Proof Indeed let $x^* \in \mathbf{K}$ be a global minimizer of $\mathbf{P}$, and let μ be the weighted Dirac measure $q(x^*)^{-1}\delta_{x^*}$. It is feasible in $\mathcal{P}$ with value $p(x^*)/q(x^*) = f^*$. On the other hand, let $\mu \in \mathbf{M}(\mathbf{K})$ be feasible in $\mathcal{P}$. As $p(x) \geq f^* q(x)$ for all $x \in \mathbf{K}$ and μ is supported on $\mathbf{K}$, $\int p d\mu \geq f^* \int q d\mu = f^*$, which proves that $f^* = \inf \mathcal{P} = \min \mathcal{P}$. Next let $y^* = (y^*_\alpha)$ be the sequence of moments of the Dirac measure $q(x^*)^{-1}\delta_{x^*}$, i.e., $y^*_\alpha = q(x^*)^{-1}(x^*)^\alpha$ for all $\alpha \in \mathbb{N}^n$. Then obviously y^* is a feasible solution of the SDP $\mathbf{Q}_r$ with associated value f^*, which proves that $\inf \mathbf{Q}_r \leq f^*$. □

The dual $\mathbf{Q}^*_r$ of $\mathbf{Q}_r$ is the SDP

$$\mathbf{Q}^*_r : \quad \begin{cases} \sup\limits_{\lambda, \{h_j\}} \; \lambda \\ \text{s.t.} \quad p - \lambda q = \sum\limits_{j=0}^m h_j g_j, \\ \qquad \{h_j\} \subset \Sigma^2 \quad \text{and} \quad \deg h_j + \deg g_j \leq 2r \end{cases} \tag{3.5}$$

whose optimal value is denoted by $\sup \mathbf{Q}^*_r$, or $\min \mathbf{Q}^*_r$ if the supremum is achieved at some optimal solution $(\lambda, \{h_j\})$.

Of course, by weak duality between dual pairs of SDPs,

$$\sup \mathbf{Q}^*_r \leq \inf \mathbf{Q}_r \leq \inf \mathbf{P} \qquad \forall r \geq r_0 \tag{3.6}$$

holds true.

Theorem 3.2 *Let* $\mathbf{K}$ *be as in* (2.10), *and let Assumption 2.1 hold. For every* $r_0 \leq r \in \mathbb{N}$, *let* $\mathbf{Q}_r$ *be the SDP-relaxation defined in* (3.2). *Then:*

(a) $\inf \mathbf{Q}_r \uparrow f^*$ *as* $r \to \infty$. *Moreover, if* $\mathbf{K}$ *has a nonempty interior then* $\sup \mathbf{Q}_r^* = \max \mathbf{Q}_r^* = \inf \mathbf{Q}_r$ *for all sufficiently large* r.

(b) *If* $\inf \mathbf{Q}_{r_1} = f^*$ *for some* r_1 *(hence for all* $r \geq r_1$*), and if* $\mathbf{K}$ *has nonempty interior, then in the dual* $\mathbf{Q}_r^*$, *one obtains:*

$$p - f^* q = \sum_{j=0}^{m} \sigma_j \, g_j \quad \text{and} \quad \sigma_j(x^*) g_j(x^*) = 0 \qquad \forall j = 0, 1, \ldots, m, \tag{3.7}$$

for some s.o.s. polynomials $\sigma_j \in \Sigma^2$ *with* $\deg \sigma_j + \deg g_j \leq 2r_1$, *for all* $j = 1, \ldots, m$, *and where* $x^* \in \mathbf{K}$ *is any global minimizer of* $\mathbf{P}$.

For a proof see Appendix section.

Remark 3.3 If $q \equiv 1$ then one retrieves the SDP-relaxations introduced in Lasserre [16]. Problem $\mathbf{P}$ with rational function $f = p/q$ was first considered in Jibetean and de Klerk [12] who proved the convergence of the dual SDP-relaxations $\mathbf{Q}_r^*$.

3.1 Detecting optimality

Theorem 3.2 guarantees that the SDP-relaxations $\{\mathbf{Q}_r\}$ in (3.2) converge to the desired optimal value of $\mathbf{P}$. However, the convergence proved in Theorem 3.2 is only asymptotic as $r \to \infty$. In some cases, *finite* convergence takes place, and below we present a sufficient condition to detect whether it has occured at some SDP-relaxation $\mathbf{Q}_r$.

Theorem 3.4 *Let* $v := \max_{j=1,\ldots,m} v_j$, *and let* $\mathbf{Q}_r$ *be the SDP-relaxation defined in* (3.2). *Assume that* $\mathbf{Q}_r$ *has an optimal solution* y *that satisfies*

$$\operatorname{rank} M_r(y) = \operatorname{rank} M_{r-v}(y) \;=:\; s \tag{3.8}$$

Then y *is the vector of moments of some* s*-atomic measure* μ *with support contained in* $\mathbf{K}$, *and* μ *is an optimal solution of* $\mathbf{P}$. *That is,* μ *is a convex combination of Dirac measures on* s *points* $x(j) \in \mathbf{K}$, $j = 1, \ldots, s$, *all global minimizers of* $\mathbf{P}$.

Proof From a result of Curto and Fialkow [6, Theor. 1.6], also proved in Laurent [23], (3.8) implies that y is the vector of moments of some measure μ finitely supported on exactly s points $\{x(i)\}_{i=1}^{s} \subset \mathbf{K}$. In addition we also have $1 = L_y(q) = \int q \, d\mu$. which proves that μ is feasible for $\mathcal{P}$ in (3.3). But this fact combined with Proposition 3.1 yields

$$f^* = \min \mathbf{P} = \min \mathcal{P} \geq \min \mathbf{Q}_r = L_y(p) = \int p \, d\mu,$$

and so μ must be an optimal solution of $\mathbf{P}$, the desired result. □

It is worth noticing that Theorem 3.4 does *not* require Assumption 2.1 to hold. The hierarchy of SDP-relaxations (3.2) can always be defined provided only that $\mathbf{K}$ is a basic closed semi-algebraic set. Even though the convergence $\inf \mathbf{Q}_r \uparrow \inf \mathbf{P}$ is not guaranteed any more, finite convergence may still happen if condition (3.8) holds true at some optimal solution of $\mathbf{Q}_r$ (whenever $\mathbf{Q}_r$ is solvable).

In addition, if (3.8) is satisfied then one can extract the s global mimimizers $x(j) \in \mathbf{K}$, $j = 1 \ldots, s$, via the numerical algebra procedure defined in Henrion and Lasserre [9].

Problem (3.1) is a particular instance of the GPM (1.1) and to the best of our knowledge, Gloptipoly3 [10] is the first (public) software devoted to solving the GPM (at least small to medium size problems). It is an extension of GloptiPoly [7] primarily devoted to global optimization; to solve (3.2) a GloptiPoly3 user may choose its favorite SDP solver, e.g. among CSDP, DSDP, SDPA, SDPLR, SDPT3, and SEDUMI.[1] For a comparison between those solvers see e.g. Mittelmann [26]. It is worth noticing that the procedure for extraction of global minimizers, based on the stopping rank criterion (3.8) and detailed in [9], is implemented in GloptiPoly3.

3.2 Karush-Kuhn-Tucker versus Putinar's Positivstellensatz

In this section we relate Putinar's Positivstellenstaz [31] with the celebrated Karush-Kuhn-Tucker (KKT) optimality conditions in Nonlinear Programming. Let $\mathbf{K} \subset \mathbb{R}^n$ be defined as in (2.10) and consider problem $\mathbf{P}$ in (3.1).

Hence $f(x) - f^* \geq 0$ on $\mathbf{K}$, or, equivalently (as $q > 0$ on $\mathbf{K}$ compact) $p - f^*q \geq 0$ on $\mathbf{K}$, i.e., the polynomial $p - f^*q$ is nonnegative on $\mathbf{K}$. Recall that under Assumption 2.1, for every $\epsilon > 0$, the polynomial $p - f^*q + \epsilon$ (strictly positive on $\mathbf{K}$) has Putinar representation (2.15), whereas the polynomial $p - f^*q$ (only nonnegative on $\mathbf{K}$) may not have.

Proposition 3.5 *Let $x^* \in \mathbf{K}$ be a global minimizer of $\mathbf{P}$ and assume that the polynomial $X \mapsto p(X) - f^*q(X)$ (which is only nonnegative on $\mathbf{K}$) has Putinar representation (2.15), i.e.,*

$$p - f^*q = \sum_{j=0}^{m} \sigma_j \, g_j \tag{3.9}$$

for some s.o.s. polynomials $\sigma_j \in \Sigma^2$, $j = 0, \ldots, m$. Then:

(a) *For every $r \geq r_1 := \frac{1}{2} \max_j[\deg \sigma_j + \deg g_j]$, the SDP-relaxations $\mathbf{Q}_r$ and $\mathbf{Q}_r^*$ are exact, that is,*

$$\sup \mathbf{Q}_r^* = \max \mathbf{Q}_r^* = \inf \mathbf{Q}_r = \min \mathbf{Q}_r.$$

(b) *In addition:*

$$g_j(x^*)\,\sigma_j(x^*) = 0, \qquad \forall j = 0, \ldots, m. \tag{3.10}$$

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j^* \, \nabla g_j(x^*) \tag{3.11}$$

with $\lambda_j^* = \sigma_j(x^*)/q(x^*) \geq 0$ for all $j = 1, \ldots, m$.

Proof As x^* is a global minimizer of $\mathbf{P}$, from (3.9) we obtain

$$0 = \sum_{j=0}^{m} \sigma_j(x^*)\, g_j(x^*) \;\geq\; 0,$$

[1] For a description of the various SDP solvers the interested reader is referred to https://projects.coin-or.org/Csdp for CSDP, http://www-unix.mcs.anl.gov/DSDP/ for DSDP, http://sdpa.indsys.chuo-u.ac.jp/sdpa/download.html#sdpam for SDPA, http://dollar.biz.uiowa.edu/burer/software/SDPLR/ for SDPLR, http://www.math.nus.edu.sg/mattohkc/sdpt3.html for SDPT3, and http://sedumi.mcmaster.ca/ for SeDuMi.

and so (3.10) holds because each term of the sum is nonnegative. Next, differentiate at x^*, use (3.10) and the fact that each σ_j is s.o.s. to obtain

$$\nabla p(x^*) - f^* \nabla q(x^*) = \sum_{j=1}^m \sigma_j(x^*) \nabla g_j(x^*).$$

Dividing by $q(x^*) > 0$ and recalling that $f^* = p(x^*)/q(x^*)$, one obtains

$$\nabla f(x^*) = \nabla \left(\frac{p}{q}\right)_{x=x^*} = \frac{\nabla p(x^*)}{q(x^*)} - \frac{p(x^*)}{q(x^*)^2} \nabla q(x^*) = \sum_{j=1}^m \frac{\sigma_j(x^*)}{q(x^*)} \nabla g_j(x^*),$$

which is simply (3.11). □

Hence, the representation (3.9) can be viewed as a global optimality condition of the "Karush-Kuhn-Tucker" type, where the multipliers are now nonnegative polynomials instead of nonnegative constants.

In the *convex* case (i.e. when f is convex and the g_j's are concave) (3.11) implies that $x^* \in \mathbf{K}$ is a global mimimizer of the Lagrangian

$$X \mapsto L(X, \lambda^*) := f(X) - f^* - \sum_{j=1}^m \lambda_j^* g_j(X), \tag{3.12}$$

and $L(x^*, \lambda^*) = 0$. That is, $L(X, \lambda^*)$ is nonnegative on the whole $\mathbb{R}^n$. But this property is valid only in the convex case. On the other hand, and even in the non convex case, when (3.9) holds then the *extended Lagrangian*

$$X \mapsto \mathcal{L}(X, \sigma) := p(X) - f^* q(X) - \sum_{j=1}^m \sigma_j(X)\, g_j(X) \qquad (= \sigma_0(X))$$

with s.o.s. multipliers σ_j's is nonnegative on the whole $\mathbb{R}^n$ (as σ_0 is s.o.s.), with $\mathcal{L}(x^*, \sigma) = 0$, so that x^* is a global minimum of $\mathcal{L}(X, \sigma)$. When $q > 0$ on $\mathbb{R}^n$, one obtains that

$$f(X) - f^* - \sum_{j=1}^m \frac{\sigma_j(X)}{q(X)}\, g_j(X)$$

is nonnegative on $\mathbb{R}^n$ even in the non convex case (to compare with (3.12)). So Putinar Positivstellensatz can be viewed as the non convex analogue for global (polynomial) optimization of the KKT conditions in the convex case.

In general, and in contrast to the usual KKT optimality conditions, the s.o.s. polynomial multiplier $\sigma_j \in \mathbb{R}[X]$ associated with a constraint $g_j(X) \geq 0$ non active at a global minimizer $x^* \in \mathbf{K}$, may **not** be identically null, but one retrieves that $\lambda_j^* = \sigma_j(x^*)/q(x^*) = 0$, i.e., σ_j vanishes at x^*. Indeed, even if not active at x^*, that constraint may be important, meaning that if it is removed from the definition (2.10) of the set $\mathbf{K}$, then the global minimum may decrease strictly. In the latter case, g_j must play a role in Putinar representation (3.7), whence the existence of a nontrivial multipier σ_j. In contrast, the KKT conditions do not "see" g_j because $\lambda_j^* = 0$, that is, the Lagrangian $L(X, \lambda^*)$ does not contain g_j. Moreover the KKT conditions are still valid at x^* for the problem without this constraint. To see this, consider the following toy example.

Example 3.1 Consider $\mathbf{P}$ defined by:

$$\mathbf{P}: \qquad f^* = \min_x \{ -x \;:\; x^2 = 1; \quad x \leq 1/2 \}$$

where $n = 1$, $f(X) = -X$ (i.e. $p(X) = -X$ and $q \equiv 1$), $g_1(X) = X^2 - 1$, $g_2(X) = 1/2 - X$. The first constraint is the equality constraint $g_1(X) = 0$ and the optimal value is $f^* = 1$. The KKT multipliers are $(\lambda_1^*, \lambda_2^*) = (1/2, 0)$ and the Lagrangian $L(X, \lambda^*) = -X - 1 - (X^2 - 1)/2 = -(X + 1)^2/2$ is not convex and does *not* contain g_2. On the other hand,

$$\mathcal{L}(X, \sigma) = -X - 1 - (X^2 - 1)(X + 3/2) - (1/2 - X)(X + 1)^2,$$

with $\sigma_0 \equiv 0$, with s.o.s. multiplier $\sigma_2(X) = (X + 1)^2$ and where the multiplier $\sigma_1(X) = (X + 3/2)$ is not required to be s.o.s. because the corresponding constraint is an equality constraint. Notice that the s.o.s. multiplier $\sigma_2(X) = (X + 1)^2$ is not trivial but vanishes at the global minimizer $x^* = -1$ like the KKT multiplier $\lambda_2^* = 0$. Indeed, the constraint $g_2(X) \geq 0$ is important in Putinar's representation (3.7) of the polynomial $f - f^*$ because if it is removed then the optimal value jumps from 1 to -1 with new global minimizer $\hat{x} = 1$. Therefore the s.o.s. multiplier σ_2 cannot be trivial (in contrast to the KKT multiplier $\lambda_2^* = 0$).

4 Extensions and related problems

In this section we first complete results of Sect. 3 by considering problems with some sparsity pattern in the data. We also consider the following two related problems:

- Systems of polynomial equations: Obtaining one real solution can be done by looking for the one that minimizes some given polynomial criterion (or some rational function). We are then back to problem **P** in (3.1) where equality constraints are treated as two opposite inequality constraints in the definition (2.10) of **K**. Another important problem is to obtain *all* real solutions when there are finitely many.
- With **K** being the basic semi-algebraic set defined in (2.10), co(**K**) its convex hull, compute pointwise the convex envelope $\widehat{f} : \mathrm{co}(\mathbf{K}) \to \mathbb{R}$ of a rational function $f : \mathbf{K} \to \mathbb{R}$.

4.1 SDP-relaxations for problems with sparsity

Despite the nice features of the SDP-relaxations (3.2), their size grows rapidly with the size of the original problem. Typically, the rth SDP-relaxation $\mathbf{Q}_r$ has to handle at least one LMI of size $\binom{n+r}{n}$ and $\binom{n+2r}{n}$ variables, which clearly limits the applicability of the methodology to problems with small to medium size only. One way to extend the applicability of the methodology to problems of larger size, is to take into account *sparsity* in the original data, frequently encountered in practical cases. Indeed, as typical in many applications of interest, f as well as the polynomials $\{g_j\}$ that describe **K**, are sparse, i.e., each monomial of f and each polynomial g_j are only concerned with a small subset of variables.

In Waki et al. [38] the authors have built up a hierarchy of SDP-relaxations in the spirit of (3.2) but where sparsity is taken into account. Sometimes, a sparsity pattern can be "read" from the data of **P** but not always, and in [38], the authors use a systematic procedure to detect and structure sparsity in **P**, via the so-called *chordal extension* of the *correlation sparsity pattern graph* (csp graph); the csp graph has as many nodes as variables, and a link beween two nodes (i.e., variables) means that these two variables both appear in a monomial of the objective function or in some inequality constraint $g_j \geq 0$ of **P**. Once a sparsity pattern has been detected, they define a simplified "sparse" version of the SDP-relaxations (3.5).

Briefly, recall the notation of Sect. 2.4 where $I = \{1, \ldots, n\} = \cup_{j=1}^{p} I_j$ with card $I_j = n_j$, and $J = \{1, \ldots, m'\} = \cup_{k=1}^{p} J_k$ (the latter being a partition). Let y be sequence $y = (y_\alpha)$, with $\alpha \in \mathbb{N}^n$, and $|\alpha| \leq 2r$. For the SDP-relaxation $\mathbf{Q}_r$ in (3.2), instead of considering a big moment matrix $M_r(y)$, now one rather defines p smaller moment matrices $M_r^k(y)$ with rows and columns indexed in the canonical basis of $\mathbb{R}[X(I_k)]_r$, $k = 1, \ldots, p$. Similarly, for every $j = 1, \ldots, m'$, the big localizing matrix $M_{r-v_j}(g_j y)$ is now replaced with the smaller localizing matrix $M_{r-v_j}^k(g_j y)$ with rows and columns indexed in the canonical basis of $\mathbb{R}[X(I_k)]_{r-v_j}$ if $j \in J_k$. Equivalently, the dual SDP $\mathbf{Q}_r^*$ now reads (in case where $q \equiv 1$)

$$\mathbf{Q}_r^* : \quad \max_{\lambda, \{\sigma_j\}} \left\{ \lambda \, : \, f - \lambda = \sum_{k=1}^{p} \left(\psi_k + \sum_{j \in J_k} \sigma_j g_j \right) \right\}$$

where the sum of squares (s.o.s.) multiplier σ_j associated with a constraint $g_j(x) \geq 0$ (where $j \in J_k$) is now a polynomial of $\mathbb{R}[X(I_k)]_{r-v_j}$, i.e., in only those variables $\{X_i \, : \, i \in I_k\}$, and likewise for the s.o.s. polynomial ψ_k. In doing so, they have obtained impressive gains in the size of the resulting SDP-relaxations, as well as in the computational time needed for obtaining an optimal solution. For instance, if card $I_j \approx \kappa$ for all j, then the sparse version of $\mathbf{Q}_r$ has now $p\,O(\kappa^{2r})$ variables (to compare with $O(n^{2r})$) and p LMIs of size at most $O(\kappa^r)$ (to compare with one LMI of size $O(n^r)$). If $\kappa \ll n$, this results in drastic computational savings! As a matter of fact, in [38] the authors were able to solve problems up to a thousand variables in the original problem ($n = 1000$), that could not be handled with the original SDP-relaxations (even for just the first SDP-relaxation only!). By using Theorem 2.7, convergence inf $\mathbf{Q}_r \uparrow f^*$ of such sparse SDP-relaxations is proved in Lasserre [18]. For more details the interested reader is referred to [18,38].

4.2 Systems of polynomial equations

Finding a *real* solution to a system of polynomial equations is a problem with many important applications. One way is to select a solution that minimizes some objective function f of interest for the application concerned, and apply optimization methods like e.g. Newton's method. However, to be effective the latter requires a good initial guess and in addition, it only finds a *local* minimum when successful.

Of course, the approach presented in Sect. 3 with now feasible set

$$\mathbf{K} := \{ x \in \mathbb{R}^n \, : \, g_j(x) = 0, \quad \forall j = 1, \ldots, m \} \tag{4.1}$$

works fine with guaranteed convergence to the global optimum f^* if $\mathbf{K}$ is compact. (In case of equality constraints Schmüdgen and Putinar representations (2.13) and (2.15) are equivalent.) The interested reader is referred to Henrion and Lasserre [8] for examples of systems of polynomial equations solved via the public software GloptiPoly [7]. Table 1 below displays some examples of systems of polynomial equations taken from [8]. As no specific criterion to minimize was available, in $\mathbf{Q}_r$ we chose to minimize the sum of the diagonal elements of the moment matrix $M_r(y)$; see [8]. For each problem are displayed the number n of variables, the number m of constraints, the maximum degree d of the g_j's, the CPU time, the order r of the relaxation $\mathbf{Q}_r$ where the stopping criterion (3.8) for detecting optimality is met, and the number of global minimizers obtained. It is worth noticing that in most problems, detection of optimality occurs at a very low relaxation order r.

In a different approach, some algebraic methods have a more ambitious goal, namely to compute *all* real and complex solutions. Methods can be symbolic with exact arithmetic or symbolic-numeric. See for instance the Homotopy approach of Verschelde [37], the Gröbner base approach of Rouillier [33], and the border base approaches of Zhi and Reid [32],

Table 1 Systems of polynomial equations

Problem name	n	m	d	CPU	r	Sol
discret3	8	8	2	0.31	1	1
eco5	5	5	3	5.98	3	1
eco6	6	6	3	57.4	3	1
eco7	7	7	3	256	3	1
eco8	8	8	3	1310	3	1
fourbar	4	4	4	0.16	2	1
geneig	6	6	3	33.2	3	1
heart	8	8	4	1532	3	2
i1	10	10	3	44.1	2	1
ipp	8	8	2	6.42	2	1
katsura5	6	6	2	0.74	2	1
kinema	9	9	2	26.4	2	1
ku10	10	10	2	72.5	2	1
lorentz	4	4	2	0.64	2	2
manocha	2	2	8	1.27	6	1
noon3	3	3	3	0.22	3	1
noon4	4	4	3	0.65	3	1
noon5	5	5	3	4.48	3	1

and Trébuchet and Mourrain [25]; see also Sommese and Wampler [35] and the references therein.

Remarkably, the approach used in Sect. 3.1 adapts with no modification to the case of computing *all* real roots of a system of polynomial equations when there are finitely many (with $m = n$ in (4.1)). Consider problem $\mathbf{P}$ in (3.1) with feasible set $\mathbf{K} \subset \mathbb{R}^n$ as in (4.1), and choose $f \in \mathbb{R}[X]$ to be the constant polynomial $f \equiv 1$. Then all real roots are optimal solutions and any sequence y which is the moment sequence of a measure supported on the real roots is also feasible in every SDP-relaxation $\mathbf{Q}_r$ defined in (3.2). Eventually, for r sufficiently large, such sequences y are the the only feasible solutions of $\mathbf{Q}_r$ and the stopping criterion (3.8) is met at $\mathbf{Q}_r$; for a detailed proof see [22]. In addition if one solves the SDP-relaxation $\mathbf{Q}_r$ with a primal-dual interior point algorithm,[2] then one obtains an optimal solution y of $\mathbf{Q}_r$ whose associated moment matrix $M_r(y)$ has maximum rank (among all possible solutions y' of $\mathbf{Q}_r$), this rank s is precisely the number of distinct real solutions $\{x(k)\}_{k=1}^s \subset \mathbb{R}^n$ of the original system of polynomial equations, and y is of the form

$$y_\alpha = \sum_{k=1}^{s} c_k \, x(k)^\alpha, \qquad \alpha \in \mathbb{N}^n$$

for some scalars $c_k > 0$ with $\sum_k c_k = 1$.

Let $I := \langle g_1, \ldots, g_n \rangle \subset \mathbb{R}[X]$ be the polynomial ideal generated by the polynomials g_j's in the definition (4.1) of $\mathbf{K}$ (with $m = n$), and let $V(I) \subset \mathbb{C}^n$ be the variety associated with I (and $V_{\mathbb{R}}(I) = V(I) \cap \mathbb{R}^n$). In fact, from such a moment matrix $M_r(y)$, one may identify s columns that are linearly independent. The associated monomials X^α in the canonical basis

[2] This primal-dual interior point method is implemented in e.g. SeDuMi [36], one of the SDP solvers used in GloptiPoly [7] and GloptiPoly3 [10] for solving $\mathbf{P}$.

$v_r(X)$ of $\mathbb{R}[X]_r$ that identify these columns, form a basis of the quotient space $\mathbb{R}[X]/J$ where $J \subset \mathbb{R}[X]$ is the real *radical* ideal $I(V_{\mathbb{R}}(I)) = I(\mathbf{K})$. One also obtains *multiplication* (by X_i) matrices M_i in the quotient algebra $\mathbb{R}[X]/J$ from which in turn one may deduce *all* points of $\mathbf{K} = V_{\mathbb{R}}(I) = V(J)$, and a Gröbner basis or a Border basis of J, as well. For more details the interested reader is referred to Lasserre et al. [22]

4.3 The convex envelope of a rational function

We here detail how to evaluate pointwise the convex envelope $\widehat{f}$ of a rational function $f = p/q$ with $p, q \in \mathbb{R}[X]$, on the convex hull co($\mathbf{K}$) of a basic closed semi-algebraic set $\mathbf{K} \subset \mathbb{R}^n$.

Let $\mathbf{K} \subset \mathbb{R}^n$ compact be as in (2.10) and let $\widetilde{f} : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be the function defined by:

$$x \mapsto \widetilde{f}(x) := \begin{cases} f(x) & \text{on } \mathbf{K} \\ +\infty & \text{on } \mathbb{R}^n \setminus \mathbf{K}. \end{cases} \tag{4.2}$$

Note that $\widetilde{f}$ is lower-semicontinuous (l.s.c.), admits a minimum and its effective domain $\mathbf{K}$ is non-empty and compact.

Recall that $\mathbb{P}(\mathbf{K})$ is the set of probability measures on $\mathbf{K}$. For every fixed $x \in \text{co}(\mathbf{K})$, consider the infinite-dimensional linear program (LP)

$$\text{LP}_x : \quad \begin{cases} \displaystyle\inf_{\mu \in \mathbb{P}(\mathbf{K})} & \displaystyle\int f\, d\mu \\ \text{s.t.} & \displaystyle\int X_i\, d\mu = x_i, \quad i = 1, \ldots, n \end{cases} \tag{4.3}$$

with optimal value denoted by $\inf \text{LP}_x$ (and $\min \text{LP}_x$ if the infimum is attained).

Lemma 4.1 ([14]) *Let* $\mathbf{K} \subset \mathbb{R}^n$ *in (2.10) be compact. Let* $f := p/q$ *with* $p, q \in \mathbb{R}[X]$, *and let* $\widetilde{f}$ *be as in (4.2). Then the convex envelope* $\widehat{f}$ *of* $\widetilde{f}$ *is given by:*

$$\widehat{f}(x) = \begin{cases} \min \text{LP}_x, & x \in \text{co}(\mathbf{K}), \\ +\infty, & x \in \mathbb{R}^n \setminus \text{co}(\mathbf{K}), \end{cases} \tag{4.4}$$

and so dom $\widehat{f} = \text{co}(\mathbf{K})$.

Next, for every $x \in \mathbb{R}^n$ fixed, consider the SDP

$$\mathbf{Q}_{rx} : \quad \begin{cases} \displaystyle\inf_y & L_y(p) \\ \text{s.t.} & L_y(X_i q) = x_i, \quad i = 1, \ldots, n \\ & M_{r-r_j}(g_j y) \succeq 0, \quad j = 0, 1, \ldots, m, \\ & L_y(q) = 1, \end{cases} \tag{4.5}$$

with optimal value denoted $\inf \mathbf{Q}_{rx}$, and $\min \mathbf{Q}_{rx}$ if the infimum is attained. Its dual is the SDP

$$\mathbf{Q}^*_{rx} : \quad \begin{cases} \displaystyle\sup_{\gamma, \lambda, \{u_j\}} & \gamma + \langle \lambda, x \rangle \\ \text{s.t.} & p - \gamma\, q - \langle \lambda, X \rangle\, q = \displaystyle\sum_{j=0}^m u_j\, g_j \\ & u_j \in \Sigma^2,\ \deg u_j + \deg g_j \le 2r, \quad j = 0, \ldots, m \end{cases} \tag{4.6}$$

with optimal value denoted $\sup \mathbf{Q}^*_{rx}$ (and $\min \mathbf{Q}^*_{rx}$ is the supremum is achieved).

Theorem 4.2 (Laraki and Lasserre [14]) *Let $\mathbf{K} \subset \mathbb{R}^n$ be as in (2.10) and let Assumption 2.1 hold. Let $f := p/q$ with $p, q \in \mathbb{R}[X]$, and with $q > 0$ on $\mathbf{K}$. Let $\widehat{f}$ be as in (4.4), and with $x \in \mathrm{co}(\mathbf{K})$ fixed, consider the SDP-relaxations $\{\mathbf{Q}_{rx}\}$ defined in (4.5)). Then:*

(a) *The function $f_r : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ defined by*

$$x \mapsto f_r(x) := \inf \mathbf{Q}_{rx}, \qquad x \in \mathbb{R}^n, \tag{4.7}$$

is convex, and as $r \to \infty$, $f_r(x) \uparrow \widehat{f}(x)$ pointwise, for all $x \in \mathbb{R}^n$. In addition the convergence is uniform on every compact subset of $\mathrm{co}(\mathbf{K})$ where $\widehat{f}$ is continuous.

(b) *If $\mathbf{K}$ has a nonempty interior $\mathrm{int}\,\mathbf{K}$, then*

$$\sup \mathbf{Q}_{rx}^* = \max \mathbf{Q}_{rx}^* = \inf \mathbf{Q}_{rx} = f_r(x), \qquad x \in \mathrm{int}\,\mathbf{K}, \tag{4.8}$$

and for every optimal solution $(\lambda_r^*, \gamma_r^*)$ of $\mathbf{Q}_{rx}^*$,

$$f_r(y) - f_r(x) \geq \langle \lambda_r^*, y - x \rangle, \qquad \forall\, y \in \mathbb{R}^n,$$

that is, $\lambda_r^* \in \partial f_r(x)$.

(For a convex function h the notation $\partial h(x)$ stands for the subdifferential of h at the point x.) So Theorem 4.2 states that one may approximate pointwise the convex envelope $\widehat{f} : \mathrm{co}(\mathbf{K}) \to \mathbb{R}$ of a rational function $f : \mathbf{K} \to \mathbb{R}$, by solving an appropriate SDP parametrized by the point $x \in \mathrm{co}(\mathbf{K})$ where one wishes to evaluate $\widehat{f}$. This provides a convex approximation f_r with pointwise monotone convergence $f_r \uparrow \widehat{f}$, and in fact, the convergence is even *uniformly* on compact subsets of $\mathrm{co}(\mathbf{K})$ where $\widehat{f}$ is continuous.

Example 4.1 Consider the bivariate rational function $f : [-1, 1]^2 \to \mathbb{R}$:

$$X \mapsto f(X) := \frac{X_1 X_2}{1 + X_1^2 + X_2^2}, \qquad (X_1, X_2) \in [-1, 1]^2,$$

on $[-1, 1]^2$ displayed in Fig. 1, with f_3 as well. In Fig. 2 we have displayed $(f_3 - f_2)$ which is of the order 10^{-9} (which explains why for a few values of $x \in [-1, 1]^2$ one may have $f_3(x) \leq f_2(x)$ as we are at the limit of machine precision). It also means that again f_2 provides a very good approximation of the convex envelope $\hat{f}$, that is, a very good approximation is already obtained at the first relaxation (here $\mathbf{Q}_2$)!

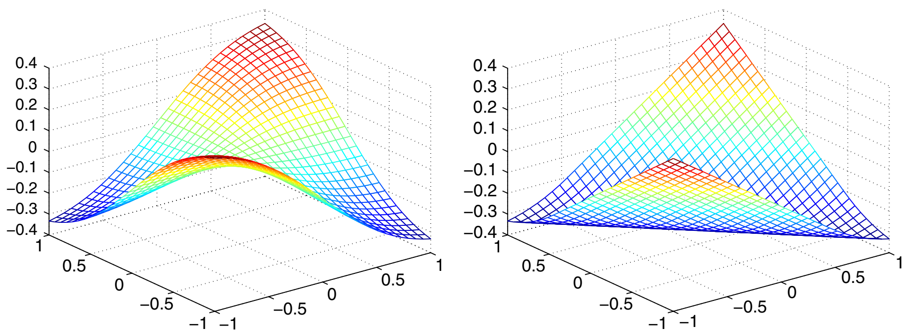

Fig. 1 Example 4.1, f and f_3 on $[-1, 1]^2$

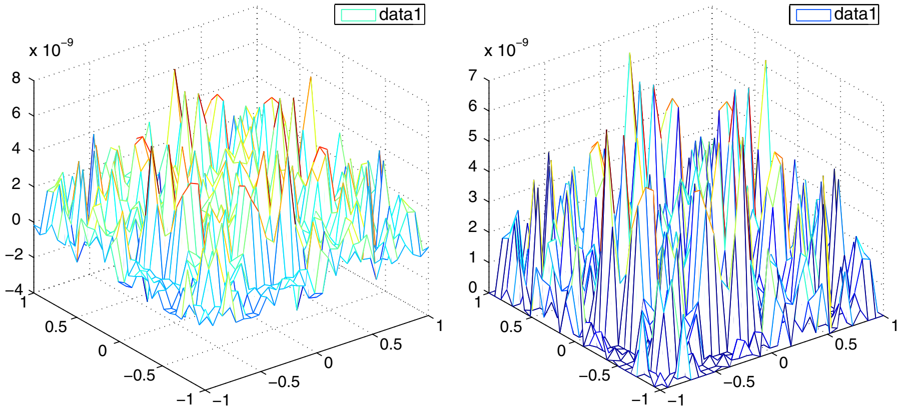

Fig. 2 Example 4.1, $f_3 - f_2$ and $(f_3 - f_2)^+$ on $[-1, 1]^2$

5 Conclusion

We have presented the moment and s.o.s. approaches in polynomial optimization, two dual facets of the problem. Despite polynomial optimization problems are NP-hard, algebra enters the game with powerful representation results from real algebraic geometry which permit to define convergent and efficient SDP-relaxations. Practice seems to reveal fast and even finite convergence. For problems with sparsity in the data, one may also define appropriate "sparse" SDP-relaxations that are still convergent when the sparsity pattern satisfies some property. One may then handle problems with a large number of variables. We hope to have convinced the reader that Putinar Positivstellensatz is the nonconvex analogue in polynomial optimization of the KKT conditions in convex programming. Finally, we have also detailed how this approach can be applied to two related problems, namely computing all real solutions of a system of polynomial equations, and pointwise computation of the convex envelope of a rational function.

Acknowledgments This work was supported by french ANR-grant NT05-3-41612.

Appendix

Proof of Theorem 3.2

(a) By Proposition 3.1 we already know that $\inf \mathbf{Q}_r \leq \inf \mathbf{P} = f^*$ for all $r \geq r_0$. Next, we need to prove that $\inf \mathbf{Q}_r > -\infty$ for sufficiently large r. Recall that the quadratic module $Q(g) \subset \mathbb{R}[X]$ generated by the polynomials $\{g_j\} \subset \mathbb{R}[X]$ that define $\mathbf{K}$ is the set

$$Q(g) := \left\{ \sigma \in \mathbb{R}[X] \;\mid\; \sigma = \sum_{j=0}^{m} \sigma_j \, g_j \quad \text{with } \{\sigma_j\}_{j=0}^{m} \subset \Sigma^2 \right\}.$$

In addition, let $Q_t(g) \subset Q(g)$ be the set of elements σ of $Q(g)$ which have a representation $\sigma_0 + \sum_{j=0}^{m} \sigma_j \, g_j$ for some s.o.s. family $\{\sigma_j\} \subset \Sigma^2$ with $\deg \sigma_0 \leq 2t$ and $\deg \sigma_j + \deg g_j \leq 2t$, for all $j = 1, \ldots, m$.

Let $r \in \mathbb{N}$ be fixed. Recall that $\mathbf{K}$ is compact and Assumption 2.1 holds. As $q > 0$ on $\mathbf{K}$, then $q > \delta$ on $\mathbf{K}$ for some scalar $\delta > 0$. Therefore, by Theorem 2.5, $q - \delta \in Q(g)$. Similarly, there exists N such that $N \pm X^\alpha > 0$ on $\mathbf{K}$, for all $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq 2r$. Therefore, the polynomial $X \mapsto N \pm X^\alpha$ belongs to $Q(g)$. But there is even some $l(r)$ such that $q - \delta \in Q_{l(r)}(g)$ and $X \mapsto N \pm X^\alpha \in Q_{l(r)}(g)$ for every $|\alpha| \leq 2r$. Of course, we also have $q - \delta \in Q_l(g)$ and $X \mapsto N \pm X^\alpha \in Q_l(g)$ for every $|\alpha| \leq 2r$, whenever $l \geq l(r)$. Therefore, let us take $l(r) \geq r_0$, with $r_0 \geq \max_{j=0,\ldots,m} v_j$.

As $q - \delta \in Q_{l(r)}(g)$, $q - \delta = \sigma_0 + \sum_{j=1}^m \sigma_j g_j$, for some $(\sigma_j) \subset \Sigma^2$ with $\deg \sigma_0 \leq 2l(r)$ and $\deg \sigma_j + \deg g_j \leq 2l(r)$, for all $j = 1, \ldots, m$. Hence, for every feasible solution y of $\mathbf{Q}_{l(r)}$ (and of $\mathbf{Q}_l$ with $l \geq l(r)$),

$$1 - \delta y_0 = L_y(q - \delta) = L_y(\sigma_0) + L_y\left(\sum_{j=1}^m \sigma_j g_j\right) \geq 0,$$

where the last inequality follows from $M_{l(r)}(y) \succeq 0$ and $M_{l(r)-r_j}(y\, g_j) \succeq 0$, $j = 1, \ldots, m$. Therefore, $y_0 \leq \delta^{-1}$.

Similarly, $N \pm X^\alpha = \sigma_0 + \sum_{j=1}^m \sigma_j\, g_j$ for some $(\sigma_j) \subset \Sigma^2$ with $\deg \sigma_0 \leq 2l(r)$ and $\deg \sigma_j + \deg g_j \leq 2l(r)$, for all $j = 1, \ldots, m$. Hence, for same reasons as above,

$$N y_0 \pm y_\alpha = L_y(N \pm X^\alpha) = L_y(\sigma_0) + \sum_{j=1}^m L_y(\sigma_j\, g_j) \geq 0,$$

which implies $|y_\alpha| = |\, L_y(X^\alpha)\, | \leq N y_0 \leq N\delta^{-1}$, for all $|\alpha| \leq 2r$.

In particular, $L_y(p) \geq -N\delta^{-1} \sum_\alpha |p_\alpha|$, which proves that $\inf \mathbf{Q}_{l(r)} > -\infty$, and so $\inf \mathbf{Q}_r > -\infty$ for sufficiently large r.

Next, from what precedes, and with $k \in \mathbb{N}$ arbitrary, let $l(k) \geq k$ be such that $q - \delta \in Q_{l(k)}$ and

$$N_k \pm X^\alpha \in Q_{l(k)}(g) \qquad \forall \alpha \in \mathbb{N}^n \text{ with } |\alpha| \leq 2k, \tag{5.1}$$

for some N_k. Let $r \geq l(r_0)$, and let y^r be a nearly optimal solution of $\mathbf{Q}_r$ with value

$$\inf \mathbf{Q}_r \leq L_{y^r}(p) \leq \inf \mathbf{Q}_r + \frac{1}{r} \quad \left(\leq f^* + \frac{1}{r}\right). \tag{5.2}$$

Fix $k \in \mathbb{N}$. Notice that from (5.1), one has

$$|\, L_{y^r}(X^\alpha)\, | \;\leq\; N_k y_0 \;\leq\; N_k \delta^{-1}, \quad \forall \alpha \in \mathbb{N}^n \text{ with } |\alpha| \;\leq\; 2k, \quad \forall r \geq l(k).$$

Therefore,

$$|y^r_\alpha| \;=\; |\, L_{y^r}(X^\alpha)\, | \;\leq\; N'_k, \quad \forall \alpha \in \mathbb{N}^n \text{ with } |\alpha| \;\leq\; 2k, \quad \forall r \geq r_0. \tag{5.3}$$

where $N'_k = \max[N_k\delta^{-1}, V_k]$, with

$$V_k := \max_{\alpha, r} \{\, |y^r_\alpha| \; : \quad |\alpha| \leq 2k\,; \quad r_0 \leq r \leq l(k)\, \}.$$

Complete each vector y^r with zeros to make it an infinite bounded sequence in l_∞, indexed in the canonical basis in $u_\infty(X)$ of $\mathbb{R}[X]$. In view of (5.3), one has $y^r_0 \leq \delta^{-1}$ and

$$|y^r_\alpha| \leq N'_k \qquad \forall \alpha \in \mathbb{N}^n \text{ with } \; 2k - 1 \leq |\alpha| \leq 2k, \tag{5.4}$$

and for all $k = 1, 2, \ldots$.

Hence let $\widehat{y}^r \in l_\infty$ be a new sequence defined by $\widehat{y}_0^r = \delta y_0^r$ and

$$\widehat{y}_\alpha^r := \frac{y_\alpha^r}{N_k'}, \quad \forall\, \alpha \in \mathbb{N}^n \text{ with } 2k-1 \le |\alpha| \le 2k, \quad \forall\, k = 1, 2, \ldots,$$

and in l_∞, consider the sequence $\{\widehat{y}^r\}_r$, as $r \to \infty$.

Obviously, the sequence $\{\widehat{y}^r\}_r$ is in the unit ball B_1 of l_∞, and so, by the Banach-Alaoglu theorem (see e.g. Ash [3]), there exists $\widehat{y} \in B_1$, and a subsequence $\{r_i\}$, such that $\widehat{y}^{r_i} \to \widehat{y}$ as $i \to \infty$, for the weak $\star$ topology $\sigma(l_\infty, l_1)$ of l_∞. In particular, pointwise convergence holds, that is,

$$\lim_{i\to\infty} \widehat{y}_\alpha^{r_i} \to \widehat{y}_\alpha \qquad \forall\, \alpha \in \mathbb{N}^n,$$

Next, define $y_0 = \delta^{-1}\widehat{y}_0$ and

$$y_\alpha := \widehat{y}_\alpha \times N_k' \quad \forall\, \alpha \in \mathbb{N}^n \quad \text{with} \quad 2k-1 \le |\alpha| \le 2k, \quad \forall\, k = 1, 2, \ldots$$

The pointwise convergence $\widehat{y}^{r_i} \to \widehat{y}$ implies the pointwise convergence $y^{r_i} \to y$, i.e.,

$$\lim_{i\to\infty} y_\alpha^{r_i} \to y_\alpha \qquad \forall\, \alpha \in \mathbb{N}^n. \tag{5.5}$$

Next, let $r \in \mathbb{N}$ be fixed. From the pointwise convergence (5.5) we deduce that

$$\lim_{i\to\infty} M_r(g_j\, y^{r_i}) = M_r(g_j\, y) \succeq 0, \quad j = 0, 1, \ldots, m.$$

As r was arbitrary we obtain that

$$M_r(g_j\, y) \succeq 0, \quad j = 0, 1, \ldots, m; \quad r = 1, 2, \ldots \tag{5.6}$$

By Theorem 2.5(b), (5.6) implies that y is the sequence of moments of some finite measure μ with support contained in $\mathbf{K}$.

Next, from the pointwise convergence (5.5) and the constraints of $\mathbf{Q}_r$, one has

$$1 = \lim_{i\to\infty} L_{y^{r_i}}(q) = L_y(q) = \int q\, \mathrm{d}\mu,$$

that is, μ is a feasible solution of $\mathcal{P}$. Finally, the pointwise convergence (5.5) implies $L_{y^{r_i}}(p) \to L_y(p) = \int p\, d\mu$, and so, from (5.2), we deduce that $\inf \mathbf{Q}_{r_i} \to f^* = \int p\, d\mu$, and in fact the desired result $\inf \mathbf{Q}_r \uparrow f^*$, because the sequence $\{\inf \mathbf{Q}_r\}$ is monotone nondecreasing.

Next, if $\mathbf{K}$ has nonempty interior then Slater condition holds for $\mathbf{Q}_r$. Indeed, let ν be a probability measure with a positive density with respect to the Lebesgue measure on $\mathbf{K}$, and let $\mathrm{d}\mu := q^{-1}\mathrm{d}\nu$ (well defined as $q > 0$ on $\mathbf{K}$ compact) so that $\int q \mathrm{d}\mu = 1$, and μ has also a strictly positive density with respect to the Lebesgue measure. Hence, with $y = (y_\alpha)$ being its sequence of moments, $M_r(g_j y) \succ 0$ for all $j = 0, 1, \ldots, m$, which shows that y is a strictly feasible solution of $\mathbf{Q}_r$. Therefore, as $\inf \mathbf{Q}_r > -\infty$ for all r sufficiently large, on also obtains $\sup \mathbf{Q}_r^* = \max \mathbf{Q}_r^* = \inf \mathbf{Q}_r$, for all r sufficiently large.

(b) Assume that $\inf \mathbf{Q}_{r_1} = f^*$, and let $x^* \in \mathbf{K}$ be any global minimizer of $\mathbf{P}$. Then obviously $f^* = \min \mathbf{Q}_{r_1}$ with $y_\alpha^* := q(x^*)^{-1}(x^*)^\alpha$ for all $\alpha \in \mathbb{N}^n$. In addition, as $\mathbf{K}$ has nonempty interior, then by (a), $\min \mathbf{Q}_{r_1}^* = f^*$ for some optimal solution $(\lambda^*, \{\sigma_j\})$, with value $\lambda^* = f^*$. Therefore,

$$p - f^* q = \sum_{j=0}^{m} \sigma_j\, g_j.$$

Evaluation at $x^* \in \mathbf{K}$ yields

$$p(x^*) - f^* q(x^*) = 0 = \sum_{j=0}^{m} \sigma_j(x^*) g_j(x^*) \geq 0,$$

which yields the desired result $\sigma_j(x^*) g_j(x^*) = 0$, for all $j = 0, 1, \ldots, m$, because every term in the sum is nonnegative. □

References

1. Akhiezer, N.I.: The Classical Moment Problem. Hafner, New York (1965)
2. Akhiezer, N.I., Krein, M.G.: Some questions in the theory of moments. Translation of Mathematical Monographs, vol. 2. American Mathematical Society, Providence, R.I. (1962)
3. Ash, R.: Real Analysis and Probability. Academic Press, San Diego (1972)
4. Berg, C.: The multidimensional moment problem and semi-groups. In: Landau, H.J. (ed.) Moments in Mathematics. Proceedings of Symposia in Applied Mathematics, vol. 37, pp. 110–124. American Mathematical Society, Providence (1980)
5. Blekherman, G.: There are significantly more nonnegative polynomials than sums of squares. Israel J. Math. **153**, 355–380 (2006)
6. Curto, R.E., Fialkow, L.A.: The truncated complex K-moment problem. Trans. Am. Math. Soc. **352**, 2825–2855 (2000)
7. Henrion, D., Lasserre, J.B.: GloptiPoly : global optimization over polynomials with Matlab and SeDuMi. ACM Trans. Math. Softw. **29**, 165–194 (2003)
8. Henrion, D., Lasserre, J.B.: Solving global optimization problems over polynomials with GloptiPoly 2.1. In: Bliek, C., Jermann, C., Neumaier, R. (eds.) Global Optimization and Constraint Satisfaction. Lecture Notes on Computer Sciences, pp. 43–58. Springer Verlag, Berlin (2003)
9. Henrion, D., Lasserre, J.B.: Detecting global optimality and extracting solutions in GloptiPoly. In: Henrion, D., Garulli, R. (eds.) Positive Polynomials in Control. Lecture Notes on Control and Information Sciences, pp. 293–310. Springer Verlag (2005)
10. Henrion, D., Lasserre, J.B., Lofberg, J.: GloptiPoly 3: moments, optimization and semidefinite programming, Optim. Methods Softw. (to appear)
11. Jacobi, T., Prestel, A.: Distinguished representations of strictly positive polynomials. J. Reine Angew. Math. **532**, 223–235 (2001)
12. Jibetean, D., de Klerk, E.: Global optimization of rational functions: a semidefinite programming approach. Math. Prog. A **106**, 93–109 (2006)
13. Landau, H.J. (ed.): Moments in Mathematics. In: Proceedings of Symposia in Applied Mathematics, vol. 37. American Mathematical Society, Providence (1980)
14. Laraki, R., Lasserre, J.B.: Computing uniform convex approximations for convex envelopes and convex hulls. J. Conv. Anal. **15**, 635–654 (2008)
15. Lasserre, J.B.: Optimisation globale et théorie des moments. C. R. Acad. Sci. Paris Série I Math. **331**, 929–934 (2000)
16. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM J. Optim. **11**, 796–817 (2001)
17. Lasserre, J.B.: S.o.s. approximations of a polynomial nonnegative on a real algebraic set. SIAM J. Optim. **16**, 610–628 (2005)
18. Lasserre, J.B.: Convergent semidefinite relaxations in polynomial optimization with sparsity. SIAM J. Optim. **17**, 822–843 (2006)
19. Lasserre, J.B.: A sum of squares approximation of nonnegative polynomials. SIAM J. Optim. **16**, 751–765 (2006). Also in: SIAM Review **49**, 651–669 (2007)
20. Lasserre, J.B.: A Semidefinite programming approach to the generalized problem of moments. Math. Program. **112**, 65–92 (2008)
21. Lasserre, J.B., Netzer, T.: SOS approximations of nonnegative polynomials via simple high degree perturbations. Math. Zeitschrift. **256**, 99–112 (2007)
22. Lasserre, J.B., Laurent, M., Rostalski, P.: Semidefinite characterization and computation of zero-dimensional real radical ideals. Found. Comp. Math. **8**, 607–647 (2008)
23. Laurent, M.: Revisiting two theorems of Curto and Fialkow on moment matrices. Proc. Am. Math. Soc. **133**, 2965–2976 (2005)

24. Maserick, P.H., Berg, C.: Exponentially bounded positive definite functions. Ill. J. Math. **28**, 162–179 (1984)
25. Mourrain, B., Trébuchet, P.: Generalized normal forms and polynomials system solving. In: Proceedings of the ISSAC 2005 Conference, Hamilton, Canada (2005)
26. Mittelmann, H.D.: An independent benchmarking of SDP and SOCP solvers. Math. Program. **95**, 407–430 (2003)
27. Nussbaum, A.E.: Quasi-analytic vectors. Ark. Mat. **6**, 179–191 (1966)
28. Parrilo, P.A.: Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization. Ph.D. thesis, California Institute of Technology, Pasadena, CA (2000)
29. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Math. Program. **96**, 293–320 (2003)
30. Petersen, L.C.: On the relation between the multidimensional moment problem and the one dimensional moment problem. Math. Scand. **51**, 361–366 (1982)
31. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana Univ. Math. J. **42**, 969–984 (1993)
32. Reid, G., Zhi, L.: Solving nonlinear polynomial system via symbolic-numeric elimination method. In: Faugère, J.C., Rouillier, F. (eds.) Proceedings of the International Conference on Polynomial System Solving, pp. 50–53 (2004)
33. Rouillier, F.: Solving zero-dimensional systems through the rational univariate representation. J. Appl. Alg. Eng. Comm. Comp. **9**, 433–461 (1999)
34. Schmüdgen, K.: The K-moment problem for compact semi-algebraic sets. Math. Ann. **289**, 203–206 (1991)
35. Sommese, A.J., Wampler, C.W.: The Numerical Solution of Systems of Polynomials Arising in Engineering and Science. World Scientific Press, Singapore (2005)
36. Sturm, J.F.: Using SeDuMi 1.02, a Matlab Toolbox for optimization over symmetric cones. Optim. Meth. Softw. **11–12**, 625–653 (1999)
37. Verschelde, J.: Algorithm 795: PHCpack: a general-purpose solver for polynomial systems by homotopy continuation. ACM Trans. Math. Softw. **25**, 251–276 (1999)
38. Waki, S., Kim, S., Kojima, M., Maramatsu, M.: Sums of squares and semidefinite programming relaxations for polynomial optimization problems witth structured sparsity. SIAM J. Optim. **17**, 218–242 (2006)