THE MOMENT-SOS HIERARCHY

JEAN B. LASSERRE

ABSTRACT. The Moment-SOS hierarchy initially introduced in optimization in 2000, is based on the theory of the $\mathbf{K}$-moment problem and its dual counterpart, polynomials that are positive on $\mathbf{K}$. It turns out that this methodology can be also applied to solve problems with positivity constraints "$f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbf{K}$" and/or linear constraints on Borel measures. Such problems can be viewed as specific instances of the "Generalized Problem of Moments" (GPM) whose list of important applications in various domains is endless. We describe this methodology and outline some of its applications in various domains.

1. INTRODUCTION

Consider the optimization problem:

$$\mathbf{P}: \quad f^* = \inf_{\mathbf{x}} \{ f(\mathbf{x}) : \mathbf{x} \in \mathbf{\Omega} \}, \tag{1.1}$$

where f is a polynomial and $\mathbf{\Omega} \subset \mathbb{R}^n$ is a basic semi-algebraic set, that is,

$$\mathbf{\Omega} := \{ \mathbf{x} \in \mathbb{R}^n : \quad g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m \}, \tag{1.2}$$

for some polynomials g_j, $j = 1, \ldots, m$. Problem $\mathbf{P}$ is a particular case of *Non Linear Programming* (NLP) where the data $(f, g_j,\ j = 1, \ldots, m)$ are *algebraic*, and therefore the whole arsenal of methods of NLP can be used for solving $\mathbf{P}$. So what is so specific about $\mathbf{P}$ in (1.1)? The answer depends on the meaning of f^* in (1.1).

If one is interested in a *local minimum* only then efficient NLP methods can be used for solving $\mathbf{P}$. In such methods, the fact that f and g_j's are polynomials does not help much, that is, this algebraic feature of $\mathbf{P}$ is not really exploited. On the other hand if f^* in (1.1) is understood as the *global minimum* of $\mathbf{P}$ then the picture is totally different. Why? First, to eliminate any ambiguity on the meaning of f^* in (1.1), rewrite (1.1) as:

$$\mathbf{P}: \quad f^* = \sup \{ \lambda : f(\mathbf{x}) - \lambda \geq 0, \quad \forall \mathbf{x} \in \mathbf{\Omega} \} \tag{1.3}$$

because then indeed f^* is necessarily the global minimum of $\mathbf{P}$.

1991 *Mathematics Subject Classification.* 90C26 90C22 90C27 65K05 14P10 44A60.

Key words and phrases. K-Moment problem; positive polynomials; global optimization; semidefinite relaxations.

Research supported by the European Research Council (ERC) through ERC-Advanced Grant # 666981 for the TAMING project.

In full generality, most problems (1.3) are very difficult to solve (they are labelled NP-hard in the computational complexity terminology) because:

Given $\lambda \in \mathbb{R}$, checking whether "$f(\mathbf{x}) - \lambda \geq 0$ for all $\mathbf{x} \in \mathbf{\Omega}$" is difficult.

Indeed, by nature this positivity constraint is *global* and therefore cannot be handled by standard NLP optimization algorithms which use only local information around a current iterate $\mathbf{x} \in \mathbf{\Omega}$. Therefore to compute f^* in (1.3) one needs an efficient tool to handle the positivity constraint "$f(\mathbf{x}) - \lambda \geq 0$ for all $\mathbf{x} \in \mathbf{\Omega}$". Fortunately if the data are algebraic then:

(1) Powerful *positivity certificates* from Real Algebraic Geometry (*Positivstellensätze* in german) are available.
(2) Some of these positivity certificates have an efficient practical implementation via *Linear Programming* (LP) or *Semidefinite Programming* (SDP). In particular and importantly, testing whether a given polynomial is a sum of squares (SOS) simply reduces to solving a single SDP (which can be done in time polynomial in the input size of the polynomial, up to arbitrary fixed precision).

After the pioneers works of Shor [51] and Nesterov [39], Lasserre [22, 23] and Parrilo [43, 44] have been the first to provide a systematic use of these two key ingredients in Control and Optimization, with convergence guarantees. It is also worth mentioning another closely related pioneer work, namely the celebrated SDP-relaxation of Goemans & Williamson [10] which provides a 0.878 approximation guarantee for MAXCUT, a famous problem in non-convex combinatorial optimization (and probably the simplest one). In fact it is perhaps the first famous example of such a successful application of the powerful SDP convex optimization technique to provide guaranteed good approximations to a notoriously difficult non-convex optimization problem. It turns out that this SDP relaxation is the first relaxation in the Moment-SOS hierarchy (a.k.a. Lasserre hierarchy) when applied to the MAXCUT problem. Since then, this spectacular success story of SDP relaxations has been at the origin of a flourishing research activity in combinatorial optimization and computational complexity. In particular, the study of LP- and SDP-relaxations in hardness of approximation is at the core of a central topic in combinatorial optimization and computational complexity, namely proving/disproving Khot's famous Unique Games Conjecture[1] (UGC) in Theoretical Computer Science.

Finally, another "definition" of the global optimum f^* of $\mathbf{P}$ reads:

$$f^* = \inf_{\mu} \{ \int_{\mathbf{\Omega}} f \, d\mu : \quad \mu(\mathbf{\Omega}) = 1 \} \tag{1.4}$$

[1]For this conjecture and its theoretical and practical implications, S. Khot was awarded the prestigious Nevanlinna prize at the last ICM 2014 in Seoul [18] .

where the 'inf" is over all probability measures on $\mathbf{\Omega}$. Equivalently, writing f as $\sum_\alpha f_\alpha \mathbf{x}^\alpha$ in the basis of monomials (where $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$):

$$f^* \,=\, \inf_{\mathbf{y}} \{ \sum_\alpha f_\alpha \, y_\alpha : \mathbf{y} \in \mathscr{M}(\mathbf{\Omega}); \quad y_0 = 1 \}, \tag{1.5}$$

where $\mathscr{M}(\mathbf{\Omega}) = \{\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n} : \exists\, \mu \text{ s.t. } y_\alpha = \int_{\mathbf{\Omega}} \mathbf{x}^\alpha \, d\mu, \forall \alpha \in \mathbb{N}^n\}$, a convex cone. In fact (1.3) is the LP dual of (1.4). In other words standard LP duality between the two formulations (1.4) and (1.3) illustrates the duality between the "$\mathbf{\Omega}$-moment problem" and "polynomials positive on $\mathbf{\Omega}$".

Problem (1.4) is a very particular instance (and even the simplest instance) of the more general *Generalized Problem of Moments* (GPM):

$$\inf_{\mu_1,\ldots,\mu_p} \{ \sum_{j=1}^p \int_{\mathbf{\Omega}_j} f_j \, d\mu_j : \, \sum_{j=1}^p f_{ij} \, d\mu_j \,\geq\, b_i, \, i = 1, \ldots, s \}, \tag{1.6}$$

for some functions $f_{ij} : \mathbb{R}^{n_j} \to \mathbb{R}$, $i = 1, \ldots, s$, and sets $\mathbf{\Omega}_j \subset \mathbb{R}^{n_j}$, $j = 1, \ldots, p$. The GPM is an infinite-dimensional LP with dual:

$$\sup_{\lambda_1,\ldots,\lambda_s \geq 0} \{ \sum_{i=1}^s \lambda_i \, b_i : \, f_j - \sum_{i=1}^s \lambda_i \, f_{ij} \,\geq\, 0 \text{ on } \mathbf{\Omega}_j, \, j : 1, \ldots, p \}. \tag{1.7}$$

Therefore it should be of no surprise that the Moment-SOS hierarchy, initially developed for global optimization, also applies to solving the GPM. This is particularly interesting as the list of important applications of the GPM is almost endless; see e.g. Landau [21].

2. The MOMENT-SOS hierarchy in optimization

2.1. **Notation, definitions and preliminaries.** Let $\mathbb{R}[\mathbf{x}]$ denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$ and let $\mathbb{R}[\mathbf{x}]_d$ be the vector space of polynomials of degree at most d (whose dimension is $s(d) := \binom{n+d}{n}$). For every $d \in \mathbb{N}$, let $\mathbb{N}^n_d := \{\alpha \in \mathbb{N}^n : |\alpha| \,(= \sum_i \alpha_i) \leq d\}$, and let $\mathbf{v}_d(\mathbf{x}) = (\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n$, be the vector of monomials of the canonical basis $(\mathbf{x}^\alpha)$ of $\mathbb{R}[\mathbf{x}]_d$. Given a closed set $\mathcal{X} \subseteq \mathbb{R}^n$, let $\mathscr{P}(\mathcal{X}) \subset \mathbb{R}[\mathbf{x}]$ (resp. $\mathscr{P}_d(\mathcal{X}) \subset \mathbb{R}[\mathbf{x}]_d$) be the convex cone of polynomials (resp. polynomials of degree at most $2d$) that are nonnegative on $\mathcal{X}$. A polynomial $f \in \mathbb{R}[\mathbf{x}]_d$ is written

$$\mathbf{x} \mapsto f(\mathbf{x}) \,=\, \sum_{\alpha\in\mathbb{N}^n} f_\alpha \, \mathbf{x}^\alpha,$$

with vector of coefficients $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(d)}$ in the canonical basis of monomials $(\mathbf{x}^\alpha)_{\alpha\in\mathbb{N}^n}$. For real symmetric matrices, let $\langle \mathbf{B}, \mathbf{C} \rangle := \mathrm{trace}\,(\mathbf{BC})$ while the notation $\mathbf{B} \succeq 0$ stands for $\mathbf{B}$ is positive semidefinite (psd) whereas $\mathbf{B} \succ 0$ stands for $\mathbf{B}$ is positive definite (pd).

The Riesz functional. Given a sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n}$, the Riesz functional is the linear mapping $L_\mathbf{y} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ defined by:

$$f\,(=\sum_\alpha f_\alpha\,\mathbf{x}^\alpha)\quad\mapsto L_\mathbf{y}(f)\;=\;\sum_{\alpha\in\mathbb{N}^n} f_\alpha\,y_\alpha. \tag{2.1}$$

Moment matrix. The *moment* matrix associated with a sequence $\mathbf{y} = (y_\alpha)$, $\alpha \in \mathbb{N}^n$, is the real symmetric matrix $\mathbf{M}_d(\mathbf{y})$ with rows and columns indexed by $\mathbb{N}^n_d$, and whose entry (α, β) is just $y_{\alpha+\beta}$, for every $\alpha, \beta \in \mathbb{N}^n_d$. Alternatively, let $\mathbf{v}_d(\mathbf{x}) \in \mathbb{R}^{s(d)}$ be the vector $(\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n_d$, and define the matrices $(\mathbf{B}_{o,\alpha}) \subset \mathcal{S}^{s(d)}$ by

$$\mathbf{v}_d(\mathbf{x})\,\mathbf{v}_d(\mathbf{x})^T\;=\;\sum_{\alpha\in\mathbb{N}^n_{2d}}\mathbf{B}_{o,\alpha}\,\mathbf{x}^\alpha,\qquad\forall\mathbf{x}\in\mathbb{R}^n. \tag{2.2}$$

Then $\mathbf{M}_d(\mathbf{y}) = \sum_{\alpha\in\mathbb{N}^n_{2d}} y_\alpha\,\mathbf{B}_{o,\alpha}$. If $\mathbf{y}$ has a representing measure μ then $\mathbf{M}_d(\mathbf{y}) \succeq 0$ because $\langle \mathbf{f}, \mathbf{M}_d(\mathbf{y})\mathbf{f}\rangle = \int f^2 d\mu \geq 0$, for all $f \in \mathbb{R}[\mathbf{x}]_d$.

A measure whose all moments are finite, is *moment determinate* if there is no other measure with same moments. The support of a Borel measure μ on $\mathbb{R}^n$ (denoted supp(μ)) is the smallest closed set $\mathbf{\Omega}$ such that $\mu(\mathbb{R}^n \setminus \mathbf{\Omega}) = 0$.

Localizing matrix. With $\mathbf{y}$ as above and $g \in \mathbb{R}[\mathbf{x}]$ (with $g(\mathbf{x}) = \sum_\gamma g_\gamma \mathbf{x}^\gamma$), the *localizing* matrix associated with $\mathbf{y}$ and g is the real symmetric matrix $\mathbf{M}_d(g\,\mathbf{y})$ with rows and columns indexed by $\mathbb{N}^n_d$, and whose entry (α, β) is just $\sum_\gamma g_\gamma y_{\alpha+\beta+\gamma}$, for every $\alpha, \beta \in \mathbb{N}^n_d$. Alternatively, let $\mathbf{B}_{g,\alpha} \in \mathcal{S}^{s(d)}$ be defined by:

$$g(\mathbf{x})\,\mathbf{v}_d(\mathbf{x})\,\mathbf{v}_d(\mathbf{x})^T\;=\;\sum_{\alpha\in\mathbb{N}^n_{2d+\deg g}}\mathbf{B}_{g,\alpha}\,\mathbf{x}^\alpha,\qquad\forall\mathbf{x}\in\mathbb{R}^n. \tag{2.3}$$

Then $\mathbf{M}_d(g\,\mathbf{y}) = \sum_{\alpha\in\mathbb{N}^n_{2d+\deg g}} y_\alpha\,\mathbf{B}_{g,\alpha}$. If $\mathbf{y}$ has a representing measure μ whose support is contained in the set $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ then $\mathbf{M}_d(g\,\mathbf{y}) \succeq 0$ for all d because $\langle \mathbf{f}, \mathbf{M}_d(g\,\mathbf{y})\mathbf{f}\rangle = \int f^2\, g d\mu \geq 0$, for all $f \in \mathbb{R}[\mathbf{x}]_d$.

SOS polynomials and quadratic modules. A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is a Sum-of-Squares (SOS) if there exist $(f_k)_{k=1,\ldots,s} \subset \mathbb{R}[\mathbf{x}]$, such that $f(\mathbf{x}) = \sum_{k=1}^s f_k(\mathbf{x})^2$, for all $\mathbf{x} \in \mathbb{R}^n$. Denote by $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_d$) the set of SOS polynomials (resp. SOS polynomials of degree at most $2d$). Of course every SOS polynomial is nonnegative whereas the converse is not true. In addition, checking whether a given polynomial f is nonnegative on $\mathbb{R}^n$ is difficult whereas checking whether f is SOS is much easier and can be done efficiently. Indeed let $f \in \mathbb{R}[\mathbf{x}]_{2d}$ (for f to be SOS its degree must be even), $\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha\in\mathbb{N}^n_{2d}} f_\alpha\,\mathbf{x}^\alpha$. Then f is SOS if and only if there exists a real symmetric matrix $\mathbf{X}^T = \mathbf{X}$ of size $s(d) = \binom{n+d}{n}$, such that:

$$\mathbf{X}\succeq 0;\quad f_\alpha\;=\;\langle\mathbf{X},\mathbf{B}_{o,\alpha}\rangle,\qquad\forall\alpha\in\mathbb{N}^n_{2d}, \tag{2.4}$$

and this can be checked by solving an SDP.

Next, let $\mathbf{x} \mapsto g_0(\mathbf{x}) := 1$ for all $\mathbf{x} \in \mathbb{R}^n$. With a family $(g_1, \ldots, g_m) \subset \mathbb{R}[\mathbf{x}]$ is associated the *quadratic module* $Q(g)\,(= Q(g_1, \ldots, g_m)) \subset \mathbb{R}[\mathbf{x}]$:

$$Q(g) := \left\{ \sum_{j=0}^{m} \sigma_j\, g_j : \sigma_j \in \Sigma[\mathbf{x}],\ j = 0, \ldots, m \right\}, \tag{2.5}$$

and its *truncated* version

$$Q_k(g) := \left\{ \sum_{j=0}^{m} \sigma_j\, g_j : \sigma_j \in \Sigma[\mathbf{x}]_{k-d_j},\ j = 0, \ldots, m \right\}, \tag{2.6}$$

where $d_j = \lceil \deg(g_j)/2 \rceil$, $j = 0, \ldots, m$.

Definition 1. *The quadratic module $Q(g)$ associated with $\mathbf{\Omega}$ in (1.2) is said to be Archimedean if there exists $M > 0$ such that the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ belongs to $Q(g)$ (i.e., belongs to $Q_k(g)$ for some k).*

If $Q(g)$ is Archimedean then necessarily $\mathbf{\Omega}$ is compact but the reverse is not rue. The Archimedean condition (which depends on the representation of $\mathbf{\Omega}$) can be seen as an *algebraic certificate* that $\mathbf{\Omega}$ is compact. For more details on the above notions of moment and localizing matrix, quadratic module, as well as their use in potential applications, the interested reader is referred to Lasserre [25], Laurent [36], Schmüdgen [49].

2.2. **Two certificates of positivity (Positivstellensätze).** Below we describe two particular certificates of positivity which are important because they provide the theoretical justification behind the so-called SDP- and LP-relaxations for global optimization.

Theorem 2.1 (Putinar [48]). *Let $\mathbf{\Omega} \subset \mathbb{R}^n$ be as in (1.2) and assume that $Q(g)$ is Archimedean.*

(a) If a polynomial $f \in \mathbb{R}[\mathbf{x}]$ is (strictly) positive on $\mathbf{\Omega}$ then $f \in Q(g)$.

(b) A sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n} \subset \mathbb{R}$ has a representing Borel measure on $\mathbf{\Omega}$ if and only if $L_{\mathbf{y}}(f^2\, g_j) \geq 0$ for all $f \in \mathbb{R}[\mathbf{x}]$, and all $j = 0, \ldots, m$. Equivalently, if and only if $\mathbf{M}_d(\mathbf{y}\, g_j) \succeq 0$ for all $j = 0, \ldots, m$, $d \in \mathbb{N}$.

There exists another certificate of positivity which does not use SOS.

Theorem 2.2 (Krivine-Vasilescu [19, 20, 52]). *Let $\mathbf{\Omega} \subset \mathbb{R}^n$ as in (1.2) be compact and such that (possibly after scaling) $0 \leq g_j(\mathbf{x}) \leq 1$ for all $\mathbf{x} \in \mathbf{\Omega}$, $j = 1, \ldots, m$. Assume also that $[1, g_1, \ldots, g_m]$ generates $\mathbb{R}[\mathbf{x}]$.*

(a) If a polynomial $f \in \mathbb{R}[\mathbf{x}]$ is (strictly) positive on $\mathbf{\Omega}$ then

$$f(\mathbf{x}) = \sum_{\alpha,\beta\in\mathbb{N}^n} c_{\alpha,\beta} \prod_{j=1}^{m} g_j(\mathbf{x})^{\alpha_j}\,(1 - g_j(\mathbf{x}))^{\beta_j}, \tag{2.7}$$

for finitely many positive coefficients $(c_{\alpha,\beta})_{\alpha,\beta\in\mathbb{N}^m}$.

(b) A sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n} \subset \mathbb{R}$ *has a representing Borel measure on* $\mathbf{\Omega}$ *if and only if* $L_{\mathbf{y}}\left(\prod_{j=1}^m g_j(\mathbf{x})^{\alpha_j}\,(1-g_j(\mathbf{x}))^{\beta_j}\right) \geq 0$ *for all* $\alpha, \beta \in \mathbb{N}^m$.

The two facets (a) and (b) of Theorem 2.1 and Theorem 2.2 illustrate the duality between *polynomials positive on* $\mathbf{\Omega}$ (in (a)) and the $\mathbf{\Omega}$*-moment problem* (in (b)). In addition to their mathematical interest, both Theorem 2.1(a) and Theorem 2.2(a) have another distinguishing feature. They both have a practical implementation. Testing whether $f \in \mathbb{R}[\mathbf{x}]_d$ is in $Q(g)_k$ is just solving a single SDP, whereas testing whether f van be written as in (2.7) with $\sum_{i=1}^m \alpha_i + \beta_i \leq k$, is just solving a single Linear Program (LP).

2.3. **The Moment-SOS hierarchy.** The Moment-SOS hierarchy is a numerical scheme based on Putinar's theorem. In a nutshell it consists of replacing the intractable positivity constraint "$f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbf{\Omega}$" with Putinar's positivity certificate $f \in Q_d(g)$ of Theorem 2.1(a), i.e., with a fixed degree bound on the SOS weights (σ_j) in (2.6). By duality, it consists of replacing the intractable constraint $\mathbf{y} \in \mathscr{M}(\mathbf{\Omega})$ with the necessary conditions $\mathbf{M}_d(g_j\,\mathbf{y}) \succeq 0$, $j = 0, \ldots, m$, of Theorem 2.1(b) for a fixed d. This results in solving an SDP which provides a lower bound on the global minimum. By allowing the degree bound d to increase, one obtains a *hierarchy* of SDPs (of increasing size) which provides a monotone non-decreasing sequence of lower bounds. A similar strategy based on Krivine-Stengle-Vasilescu positivity certificate (2.7) is also possible and yields a hierarchy of LP (instead of SDPs). However even though one would prefer to solve LPs rather than SDPs, the latter Moment-LP hierarchy has several serious drawbacks (some explained in e.g. [26, 29]), and therefore we only describe the Moment-SOS hierarchy.

Recall problem $\mathbf{P}$ in (1.1) or equivalently in (1.3) and (1.4), where $\mathbf{\Omega} \subset \mathbb{R}^n$ is the basic semi-algebraic set defined in (1.2).

The Moment-SOS hierarchy. Consider the sequence of semidefinite programs $(\mathbf{Q}_d)_{d\in\mathbb{N}}$ with $d \geq \hat{d} := \max[\deg(f), \max_j \deg(g_j)]$:

$$\text{(2.8)} \qquad \mathbf{Q}_d: \ \rho_d = \inf_{\mathbf{y}} \{\, L_{\mathbf{y}}(f) : \ y_0 = 1;\ \mathbf{M}_d(g_j\,\mathbf{y}) \succeq 0, \quad 0 \leq j \leq m \,\}$$

(where $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n_{2d}})$[2], with associated sequence of their SDP duals:

$$\text{(2.9)} \qquad \mathbf{Q}^*_d: \ \rho^*_d = \sup_{\lambda,\sigma_j} \{\, \lambda : \ f - \lambda = \sum_{j=0}^m \sigma_j\, g_j;\ \sigma_j \in \Sigma[\mathbf{x}]_{d-d_j},\ 0 \leq j \leq m \}$$

(where $d_j = \rceil(\deg g_j)/2\rceil$). By standard weak duality in optimization $\rho^*_d \leq \rho_d$ for every $d \geq \hat{d}$. The sequence $(\mathbf{Q}_d)_{d\in\mathbb{N}}$ forms a *hierarchy* of *SDP-relaxations* of $\mathbf{P}$ because $\rho_d \leq f^*$ and $\rho_d \leq \rho_{d+1}$ for all $d \geq \hat{d}$. Indeed for each $d \geq \hat{d}$, the constraints of $\mathbf{Q}_d$ consider only necessary conditions for $\mathbf{y}$ to be the moment

[2] In Theoretical Computer Science, $\mathbf{y}$ is called a sequence of "pseudo-moments".

sequence (up to order $2d$) of a probability measure on $\mathbf{\Omega}$ (cf. Theorem 2.1(b)) and therefore $\mathbf{Q}_d$ is a relaxation of (1.5).

By duality, the sequence $(\mathbf{Q}_d^*)_{d\in\mathbb{N}}$ forms a *hierarchy* of *SDP-strenghtenings* of (1.3). Indeed in (2.9) one has replaced the intractable positivity constraint of (1.3) by the (stronger) Putinar's positivity certificate with degree bound $2d-2d_j$ on the SOS weights σ_j's.

Theorem 2.3 ([22, 23]). *Let $\mathbf{\Omega}$ in (1.2) be compact and assume that its associated quadratic module $Q(g)$ is Archimedean. Then:*

(i) As $d\to\infty$, the monotone non-decreasing sequence $(\rho_d)_{d\in\mathbb{N}}$ (resp. $(\rho_d^*)_{d\in\mathbb{N}}$) of optimal values of the hierarchy (2.8) (resp. (2.9)) converges to the global optimum f^* of $\mathbf{P}$.

(ii) Moreover, let $\mathbf{y}^d=(y^d_\alpha)_{\alpha\in\mathbb{N}^n_{2d}}$ be an optimal solution of $\mathbf{Q}_d$ in (2.8), and let $s=\max_j d_j$ (recall that $d_j=\lceil(\deg g_j)/2\rceil$). If

$$\text{rank}\,\mathbf{M}_d(\mathbf{y}^d)\;=\;\text{rank}\,\mathbf{M}_{d-s}(\mathbf{y}^d)\;(=:t) \tag{2.10}$$

then $\rho_d=f^*$ and there are t global minimizers $\mathbf{x}_j^*\in\mathbf{\Omega}$, $j=1,\ldots,t$, that can be "extracted" from $\mathbf{y}^d$ by a linear algebra routine.

The sequence of SDP-relaxations $(\mathbf{Q}_d)$, $d\geq\hat{d}$, and the rank test (2.10) to extract global minimizers, are implemented in the GloptiPoly software [14].

Finite convergence and a global optimality certificate. After being introduced in [22], in many numerical experiments it was observed that typically, finite convergence takes place, that is, $f^*=\rho_d$ for some (usually small) d. In fact there is a rationale behind this empirical observation.

Theorem 2.4 (Nie [40]). *Let $\mathbf{P}$ be as in (1.3) where $\mathbf{\Omega}$ in (1.2) is compact and its associated quadratic module is Archimedean. Suppose that at each global minimizer $\mathbf{x}^*\in\mathbf{\Omega}$:*

- *The gradients $(\nabla g_j(\mathbf{x}^*))_{j=1,\ldots,m}$ are linearly independent. (This implies existence of nonnegative Lagrange-KKT multipliers λ_j^*, $j\leq m$, such that $\nabla f(\mathbf{x}^*)-\sum_{j=1}^m\lambda_j^*\,\nabla g_j(\mathbf{x}^*)=0$ and $\lambda_j^*g_j(\mathbf{x}^*)=0$ for all $j\leq m$.)*
- *Strict complementarity holds, that is, $g_j(\mathbf{x}^*)=0\Rightarrow\lambda_j^*>0$.*
- *Second-order sufficiency condition holds, i.e.,*

$$\langle\mathbf{u},\nabla^2_{\mathbf{x}}\,(f(\mathbf{x}^*)-\sum_{j=1}^m\lambda_j^*\,g_j(\mathbf{x}^*))\,\mathbf{u}\rangle\;>\;0,$$

for all $0\neq\mathbf{u}\in\nabla(f(\mathbf{x}^*)-\sum_{j=1}^m\lambda_j^*\,g_j(\mathbf{x}^*))^\perp$.

Then $f-f^*\in Q(g)$, i.e., there exists d^* and SOS multipliers $\sigma_j^*\in\Sigma[\mathbf{x}]_{d^*-d_j}$, $j=0,\ldots,m$, such that:

$$f(\mathbf{x})-f^*\;=\;\sigma_0^*(\mathbf{x})+\sum_{j=1}^m\sigma_j^*(\mathbf{x})\,g_j(\mathbf{x}). \tag{2.11}$$

With (2.11), Theorem 2.4 provides a *certificate of global optimality* in polynomial optimization, and to the best of our knowledge, the first at this level of generality. Next, observe that $\mathbf{x}^* \in \mathbf{\Omega}$ is a global unconstrained minimizer of the *extended Lagrangian polynomial* $f - f^* - \sum_{j=1}^n \sigma_j^* g_j$, and therefore Theorem 2.4 is the analogue for *non-convex* polynomial optimization of the Karush-Kuhn-Tucker (KKT) optimality conditions *in the convex case*. Indeed in the convex case, any local minimizer is global and is also a global unconstrained minimizer of the Lagrangian $f - f^* - \sum_{j=1}^m \lambda_j^* g_j$.

Also interestingly, whenever the SOS weight σ_j^* in (2.11) is non trivial, it testifies that the constraint $g_j(\mathbf{x}) \geq 0$ is important for $\mathbf{P}$ even if it is not active at $\mathbf{x}^*$ (meaning that if $g_j \geq 0$ is deleted from $\mathbf{P}$ then the new global optimum decreases strictly). The multiplier λ_j^* plays the same role in the KKT-optimality conditions *only in the convex case*. See [26] for a detailed discussion.

Finite convergence of the Moment-SOS-hierarchies (2.8) and (2.9) is an immediate consequence of Theorem 2.4. Indeed by (2.11) $(f^*, \sigma_0^*, \ldots, \sigma_m^*)$ is a feasible solution of $\mathbf{Q}_{d^*}^*$ with value $f^* \leq \rho_d^* \leq f^*$ (hence $\rho_d^* = \rho_d = f^*$).

Genericity: Importantly, as proved in Nie [40], the conditions in Theorem 2.4 are *generic*. By this we mean the following: Consider the class $\mathscr{P}(t, m)$ of optimization problems $\mathbf{P}$ with data $(f, g_1, \ldots, g_m)$ of degree bounded by t, and with nonempty compact feasible set $\mathbf{\Omega}$. Such a problem $\mathbf{P}$ is a "point" in the space $\mathbb{R}^{(m+1)s(t)}$ of coordinates of $(f, g_1, \ldots, g_m)$. Then the "good" problems $\mathbf{P}$ are points in a Zariski open set. Moreover, generically the rank test (2.10) is also satisfied at an optimal solution of (2.8) (for some d); for more details see Nie [41].

Computational complexity: Each relaxation $\mathbf{Q}_d$ in (2.8) is a semidefinite program with $s(2d) = \binom{n+2d}{n}$ variables (y_α), and a psd constraint $\mathbf{M}_d(\mathbf{y}) \succeq 0$ of size $s(d)$. Therefore solving $\mathbf{Q}_d$ in its canonical form (2.8) is quite expensive in terms of computational burden, especially when using interior-point methods. Therefore its brute force application is limited to small to medium size problems.

Exploiting sparsity: Fortunately many large scale problems exhibit a structured sparsity pattern (e.g., each polynomial g_j is concerned with a few variables only, and the objective function f is a sum $\sum_i f_i$ where each f_i is also concerned with a few variables only). Then Waki et al. [53] have proposed a sparsity-adapted hierarchy of SDP-relaxations which can handle problems $\mathbf{P}$ with thousands variables. In addition, if the sparsity pattern satisfies a certain condition then convergence of this sparsity-adapted hierarchy is also guaranteed like in the dense case [31]. Successful applications of this strategy can be found in e.g. Camps and Sznaier [3] in Control (systems identification) and in Molzahn and Hiskens [37] for solving (large scale) Optimum Power Flow problems (OPF is an important problem encountered in the management of energy networks).

2.4. **Discussion.** We claim that the Moment-SOS hierarchy and its rationale Theorem 2.4, unify convex, non-convex (continuous), and discrete (polynomial) Optimization. Indeed in the description of $\mathbf{P}$ we do not pay attention to what particular class of problems $\mathbf{P}$ belongs to. This is in sharp contrast to the usual common practice in (local) optimization where several classes of problems have their own tailored favorite class of algorithms. For instance, problems are not treated the same if equality constraints appear, and/or if boolean (or discrete variables) are present, etc. Here a boolean variable x_i is modeled by the quadratic equality constraint $x_i^2 = x_i$. So it is reasonable to speculate that this lack of specialization could be a handicap for the moment-SOS hierarchy.

But this is not so. For instance for the sub-class of convex[3] problems $\mathbf{P}$ where f and $(-g_j)_{j=1,\ldots,m}$ are SOS-convex[4] polynomials, finite convergence takes place at the first step of the hierarchy. In other words, the SOS hierarchy somehow "recognizes" this class of easy problems [26]. In the same time, for a large class of 0/1 combinatorial optimization problems on graphs, the Moment-SOS hierarchy has been shown to provide the tightest upper bounds when compared to the class of *lift-and-project* methods, and has now become a central tool to analyze hardness of approximations in combinatorial optimization. For more details the interested reader is referred to e.g. Lasserre [29], Laurent [35], Barak [1], Khot [17, 18] and the many references therein.

3. The Moment-SOS hierarchy outside optimization

3.1. **A general framework for the Moment-SOS hierarchy.** Let $\mathbf{\Omega}_i \subset \mathbb{R}^{n_i}$ be a finite family of compact sets, $\mathscr{M}(\mathbf{\Omega}_i)$ (resp. $\mathscr{C}(\mathbf{\Omega}_i)$) be the space of finite Borel signed measures (resp. continuous functions) on $\mathbf{\Omega}_i$, $i = 0, 1, \ldots, s$, and let $\mathbf{T}$ be a continuous linear mapping with adjoint $\mathbf{T}^*$:

$$\begin{array}{rcl} \mathbf{T} : \mathscr{M}(\mathbf{\Omega}_1) \times \cdots \times \mathscr{M}(\mathbf{\Omega}_s) & \to & \mathscr{M}(\mathbf{\Omega}_0) \\ \mathscr{C}(\mathbf{\Omega}_1) \times \cdots \times \mathscr{C}(\mathbf{\Omega}_s) & \leftarrow & \mathscr{C}(\mathbf{\Omega}_0) : \mathbf{T}^* \end{array}$$

Let $\phi := (\phi_1, \ldots, \phi_s)$ and let $\phi_i \geq 0$ stand for ϕ_i is a positive measure. Then consider the general framework:

$$\rho = \inf_{\phi \geq 0} \{ \sum_{i=1}^{s} \langle f_i, \phi_i \rangle : \mathbf{T}(\phi) = \lambda; \sum_{i=1}^{s} \langle f_{ij}, \phi_i \rangle \geq b_j,\ j \in J \}, \tag{3.1}$$

where J is a finite or countable set, $\mathbf{b} = (b_j)$ is given, $\lambda \in \mathscr{M}(\mathbf{\Omega}_0)$ is a given measure, $(f_{ij})_{j \in J}$, $i = 1, \ldots, s$, are given polynomials, and $\langle \cdot, \cdot \rangle$ is the duality bracket between $\mathscr{C}(\mathbf{\Omega}_i)$ and $\mathscr{M}(\mathbf{\Omega}_i)$ ($\langle h, \phi_i \rangle = \int_{\mathbf{\Omega}_i} h d\phi_i$), $i = 1, \ldots, s$.

[3]Convex problems $\mathbf{P}$ where f and $(-g_j)_{j=1,\ldots,m}$ are convex, are considered "easy" and can be solved efficiently.

[4]A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is SOS-convex if its Hessian $\nabla^2 f$ is a SOS matrix-polynomial, i.e., $\nabla f^2(\mathbf{x}) = \mathbf{L}(\mathbf{x})\mathbf{L}(\mathbf{x})^T$ for some matrix-polynomial $\mathbf{L} \in \mathbb{R}[\mathbf{x}]^{n \times p}$.

As we will see, this general framework is quite rich as it encompasses a lot of important applications in many different fields. In fact Problem (3.1) is equivalent to the Generalized Problem of Moments (GPM):

$$\begin{aligned}\rho = \inf_{\phi\geq 0} \quad &\{\sum_{i=1}^{s}\langle f_i,\phi_i\rangle : \langle \mathbf{T}^*\,p_k,\phi\rangle \;=\; \langle p_k,\lambda\rangle, \quad k=0,1,\ldots \\ &\sum_{i=1}^{s}\langle f_{ij},\phi_i\rangle \;\geq\; b_j, \quad j\in J\},\end{aligned} \tag{3.2}$$

where the family $(p_k)_{k=0,\ldots}$ is dense in $\mathscr{C}(\mathbf{\Omega}_0)$ (e.g. a basis of $\mathbb{R}[x_1,\ldots,x_{n_0}]$).

The Moment-SOS hierarchy can also be applied to help solve the Generalized Problem of Moments (GPM) (3.2) or its dual :

(3.3)
$$\begin{aligned}\rho^* = \sup_{(\theta_j\geq 0,\gamma)} \quad &\{\sum_k \gamma_k\,\langle p_k,\lambda\rangle + \langle\theta,\mathbf{b}\rangle : \\ \text{s.t.} \quad & f_i - \sum_k \gamma_k\,(\mathbf{T}^*\,p_k)_i - \sum_{j\in J}\theta_j\,f_{ij} \;\geq\; 0 \;\text{ on } \mathbf{\Omega}_i \text{ for all } i\,\},\end{aligned}$$

where the unknown $\gamma=(\gamma_k)_{k\in\mathbb{N}}$ is a finite sequence.

3.2. **A hierarchy of SDP-relaxations.** Let

$$\mathbf{\Omega}_i \;:=\; \{\,\mathbf{x}\in\mathbb{R}^{n_i} :\; g_{i,\ell}(\mathbf{x})\geq 0,\, i=1,\ldots,m_i\,\}, \quad i=1,\ldots,s, \tag{3.4}$$

for some polynomials $(g_{i,\ell})\subset\mathbb{R}[x_1,\ldots,x_{n_i}]$, $\ell=1,\ldots,m_i$. Let $d_{i,\ell} = \lceil \deg(g_{i,\ell})/2\rceil$ and $\hat{d} := \max_{i,j,\ell}[\deg(f_i),\deg(f_{ij}),\deg(g_{i,\ell})]$. To solve (3.2), define the "moment" sequences $\mathbf{y}_i=(y_{i,\alpha})$, $\alpha\in\mathbb{N}^{n_i}$, $i=1,\ldots,s$, and with $d\in\mathbb{N}$, define $\Gamma_d := \{p_k : \deg(T^*p_k)_i\leq 2d,\; i=1,\ldots,s\}$. Consider the hierarchy of semidefinite programs indexed by $\hat{d}\leq d\in\mathbb{N}$:

$$\begin{aligned}\rho_d = \inf_{(\mathbf{y}_i)} \quad &\{\sum_{i=1}^{s} L_{\mathbf{y}_i}(f_i) : \sum_{i=1}^{s} L_{\mathbf{y}_i}((T^*p_k)_i) \;=\; \langle p_k,\lambda\rangle, \quad p_k\in\Gamma_d \\ &\sum_{i=1}^{s} L_{\mathbf{y}_i}(f_{ij}) \;\geq\; b_j, \quad j\in J_d \\ &\mathbf{M}_d(\mathbf{y}_i),\, \mathbf{M}_{d-d_\ell}(g_{i\ell}\,\mathbf{y}_i)\succeq 0, \quad \ell\leq m_i;\, i\leq s\},\end{aligned} \tag{3.5}$$

where $J_d\subset J$ is finite $\bigcup_{d\in\mathbb{N}} J_d = J$. Its dual SDP-hierarchy reads:

$$\begin{aligned}\rho_d^* = \sup_{(\theta_j\geq 0,\gamma_k)} \quad &\{\sum_{p_k\in\Gamma_d}\gamma_k\,\langle p_k,\lambda\rangle + \langle\theta,\mathbf{b}\rangle : \\ \text{s.t.} \quad & f_i - \sum_{p_k\in\Gamma_d}\gamma_k\,(\mathbf{T}^*\,p_k)_i - \sum_{j\in J}\theta_j\,f_{ij} = \sum_{\ell=0}^{m_i}\sigma_{i,\ell}\,g_{i,\ell} \\ & \sigma_{i,\ell}\in\Sigma[x_1,\ldots,x_{n_i}]_{d-d_{i,\ell}};\; i=1,\ldots,s\},\end{aligned} \tag{3.6}$$

As each $\mathbf{\Omega}_i$ is compact, for technical reasons and with no loss of generality, in the sequel we may and will assume that for every $i=1,\ldots,s$, $g_{i,0}(\mathbf{x}) = M_i - \|\mathbf{x}\|^2$, where $M_i>0$ is sufficiently large.

Theorem 3.1. *Assume that $\rho > -\infty$ and that for every $i = 1, \dots, s$, $f_{i0} = 1$. Then for every $d \geq \hat{d}$, (3.5) has an optimal solution, and $\lim_{d\to\infty} \rho_d = \rho$.*

3.3. Examples in Probability and Computational Geometry.

Bounds on measures with moment conditions. Let Z be a random vector with values in a compact semi-algebraic set $\mathbf{\Omega}_1 \subset \mathbb{R}^n$. Its distribution λ on $\mathbf{\Omega}_1$ is unknown but some of its moments $\int \mathbf{x}^\alpha \, d\lambda = b_\alpha$, $\alpha \in \Gamma \subset \mathbb{N}^n$, are known ($b_0 = 1$). Given a basic semi-algebraic set $\mathbf{\Omega}_2 \subset \mathbf{\Omega}_1$ we want to compute (or approximate as closely as desired) the best upper bound on $\mathrm{Prob}(Z \in \mathbf{\Omega}_2)$. This problem reduces to solving the GPM:

$$\rho = \sup_{\phi_1,\phi_2 \geq 0} \quad \{\langle 1, \phi_2\rangle : \langle \mathbf{x}^\alpha, \phi_1\rangle + \langle \mathbf{x}^\alpha, \phi_2\rangle = b_\alpha, \ \alpha \in \Gamma; \quad \phi_i \in \mathscr{M}(\mathbf{\Omega}_i), \ i = 1, 2\,\}, \tag{3.7}$$

With $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ as in (3.4) one may compute upper bounds on ρ by solving the Moment-SOS hierarchy (3.5) adapted to problem (3.7). Under the assumptions of Theorem 3.1, the resulting sequence $(\rho_d)_{d\in\mathbb{N}}$ converges to ρ as $d \to \infty$; for more details the interested reader is referred to [30].

Lebesgue & Gaussian measures of semi-algebraic sets. Let $\mathbf{\Omega}_2 \subset \mathbb{R}^n$ be compact. The goal is to compute (or approximate as closely as desired) the Lebesgue measure $\lambda(\mathbf{\Omega}_2)$ of $\mathbf{\Omega}_2$. Then take $\mathbf{\Omega}_1 \supset \mathbf{\Omega}_2$ be a simple set, e.g. an ellipsoid or a box (in fact any set such that one knows all moments $(b_\alpha)_{\alpha\in\mathbb{N}^n}$ of the Lebesgue measure on $\mathbf{\Omega}_1$). Then:

$$\lambda(\mathbf{\Omega}_2) \ = \ \sup_{\phi_1,\phi_2 \geq 0} \quad \{\langle 1, \phi_2\rangle : \langle \mathbf{x}^\alpha, \phi_1\rangle + \langle \mathbf{x}^\alpha, \phi_2\rangle = b_\alpha, \ \alpha \in \mathbb{N}^n; \quad \phi_i \in \mathscr{M}(\mathbf{\Omega}_i), \ i = 1, 2\,\}. \tag{3.8}$$

Problem (3.8) is very similar to (3.7) except that we now have countably many moment constraints ($\Gamma = \mathbb{N}^n$). Again, with $\mathbf{\Omega}_2$ and $\mathbf{\Omega}_2$ as in (3.4) one may compute upper bounds on $\lambda(\mathbf{\Omega}_2)$ by solving the Moment-SOS hierarchy (3.5) adapted to problem (3.8). Under the assumptions of Theorem 3.1, the resulting monotone non-increasing sequence $(\rho_d)_{d\in\mathbb{N}}$ converges to $\lambda(\mathbf{\Omega}_2)$ from above as $d \to \infty$. The convergence $\rho_d \to \lambda(\mathbf{\Omega}_2)$ is slow because of a Gibb's phenomenon[5]. Indeed the semidefinite program (3.6) reads:

$$\rho_d^* = \inf_{p\in\mathbb{R}[\mathbf{x}]_{2d}} \{\int_{\mathbf{\Omega}_1} p \, d\lambda : \ p \geq 1 \text{ on } \mathbf{\Omega}_2; \quad p \geq 0 \text{ on } \mathbf{\Omega}_1\},$$

i.e., as $\to \infty$ one tries to approximate the discontinuous function $\mathbf{x} \mapsto 1_{\mathbf{\Omega}_2}(\mathbf{x})$ by polynomials of increasing degrees. Fortunately there are several ways to accelerate the convergence, e.g. as in [15] (but loosing the monotonicity) or in [28] (preserving monotonicity) by including in (3.5) additional constraints on $\mathbf{y}_2$ coming from an application of Stokes' theorem.

For the **Gaussian measure** λ we need and may take $\mathbf{\Omega}_1 = \mathbb{R}^n$ and $\mathbf{\Omega}_2$ is not necessarily compact. Although both $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ are allowed to be

[5]The Gibbs' phenomenon appears at a jump discontinuity when one approximates a piecewise C^1 function with a continuous function, e.g., by its Fourier series.

non-compact, the Moment-SOS hierarchy (3.5) still converges, i.e., $\rho_d \to \lambda(\mathbf{\Omega}_2)$ as $d \to \infty$. This is because the moments of λ satisfy the generalized Carleman's condition

$$\sum_{k=1}^{\infty} \left(\int_{\mathbb{R}^n} x_i^{2k} \, d\lambda \right)^{-1/2k} = +\infty, \quad i = 1, \ldots, n, \tag{3.9}$$

which imposes implicit constraints on $\mathbf{y}_1$ and $\mathbf{y}_2$ in (3.5), strong enough to guarantee $\rho_d \to \lambda(\mathbf{\Omega}_2)$ as $d \to \infty$. For more details see [28]. This deterministic approach is computationally demanding and should be seen as complementary to brute force Monte-Carlo methods that provide only an estimate (but can handle larger size problems).

3.4. **In signal processing and interpolation.** In this application, a signal is identified with an atomic signed measure ϕ supported on few atoms $(\mathbf{x}_k)_{k=1,\ldots,s} \subset \mathbf{\Omega}$, i.e., $\phi = \sum_{k=1}^{s} \theta_k \, \delta_{\mathbf{x}_k}$, for some weights $(\theta_k)_{k=1,\ldots,s}$.

Super-Resolution. The goal of Super-Resolution is to reconstruct the unknown measure ϕ (the signal) from a few measurements only, when those measurements are the moments $(b_\alpha)_{\alpha \in \mathbb{N}^n_t}$ of ϕ, up to order t (fixed). One way to proceed is to solve the infinite-dimensional program:

$$\rho = \inf_{\phi} \{ \|\phi\|_{TV} : \int \mathbf{x}^\alpha \, d\phi = b_\alpha, \quad \alpha \in \mathbb{N}^n_t \}, \tag{3.10}$$

where the inf is over the finite signed Borel measures on $\mathbf{\Omega}$, and $\|\phi\|_{TV} = |\phi|(\mathbf{\Omega})$ (with $|\phi|$ being the total variation of ϕ). Equivalently:

$$\rho = \inf_{\phi^+, \phi^- \geq 0} \{ \langle 1, \phi^+ + \phi^- \rangle : \langle \mathbf{x}^\alpha, \phi^+ - \phi^- \rangle = b_\alpha, \, \alpha \in \mathbb{N}^n_t \}, \tag{3.11}$$

which is an instance of the GPM with dual:

$$\rho^* = \sup_{p \in \mathbb{R}[\mathbf{x}]_t} \{ \sum_{\alpha \in \mathbb{N}^n_t} p_\alpha \, b_\alpha : \|p\|_\infty \leq 1 \}, \tag{3.12}$$

where $\|p\|_\infty = \sup\{|p(\mathbf{x})| : \mathbf{x} \in \mathbf{\Omega}\}$. In this case, the Moment-SOS hierarchy (3.5) with $d \geq \hat{d} := \lceil t/2 \rceil$, reads:

$$\begin{aligned} .\rho_d = \inf_{\mathbf{y}^+, \mathbf{y}^-} \quad & \{ y_0^+ + y_0^- : y_\alpha^+ - y_\alpha^- = b_\alpha, \, \alpha \in \mathbb{N}^n_t \\ & \mathbf{M}_d(\mathbf{y}^\pm) \succeq 0; \, \mathbf{M}_d(g_\ell \, \mathbf{y}^\pm) \succeq 0, \, \ell = 1, \ldots, m \}, \end{aligned} \tag{3.13}$$

where $\mathbf{\Omega} = \{ \mathbf{x} : g_\ell(\mathbf{x}) \geq 0, \, \ell = 1, \ldots, m \}$.

In the case where $\mathbf{\Omega}$ is the torus $\mathbb{T} \subset \mathbb{C}$, Candès and Fernandez-Granda [4] showed that if $\delta > 2/f_c$ (where δ is the minimal distance between the atoms of ϕ, and f_c is the number of measurements) then (3.10) has a unique solution and one may recover ϕ *exactly* by solving the single semidefinite program (3.10) with $d = \lceil t/2 \rceil$. The dual (3.12) has an optimal solution p^* (a trigonometric polynomial) and the support of ϕ^+ (resp. ϕ^-) consists of the atoms $\mathbf{z} \in \mathbb{T}$ of ϕ such that $p^*(\mathbf{z}) = 1$ (resp. $p^*(\mathbf{z}) = -1$). In addition, this procedure is more robust to noise in the measurements than Prony's method; on the other hand, the latter requires less measurements and no separation condition on the atoms.

In the general multivariate case treated in [6] one now needs to solve the Moment-SOS hierarchy (3.11) for $d = \hat{d}, \ldots$ (instead of a single SDP in the univariate case). However since the moment constraints of (3.11) are finitely many, exact recovery (i.e. finite convergence of the Moment-SOS hierarchy (3.13)) is possible (usually with a few measurements only). This is indeed what has been observed in all numerical experiments of [6], and in all cases with significantly less measurements than the theoretical bound (of a tensorized version of the univariate case).

In fact, the rank condition (2.10) is always satisfied at an optimal solution $(\mathbf{y}^+, \mathbf{y}^-)$ at some step d of the hierarchy (3.13), and so the atoms of ϕ^+ and ϕ^- are extracted via a simple linear algebra routine (as for global optimization). Nie's genericity result [41] should provide a rationale which explains why the rank condition (2.10) is satisfied in all examples.

Sparse interpolation. Here the goal is to recover an unknown (black-box) polynomial $p \in \mathbb{R}[\mathbf{x}]_t$ through a few evaluations of p only. In [16] we have shown that this problem is in fact a particular case of Super-Resolution (and even *discrete* Super-Resolution) on the torus $\mathbb{T}^n \subset \mathbb{C}^n$. Indeed let $\mathbf{z}_0 \in \mathbb{T}^n$ be fixed, arbitrary. Then with $\beta \in \mathbb{N}^n$, notice that

$$\begin{aligned} p(\mathbf{z}_0^\beta) &= \sum_{\alpha \in \mathbb{N}^n_d} p_\alpha \, (z_{01}^{\beta_1} \cdots z_{0n}^{\beta_n})^\alpha \;=\; \sum_{\alpha \in \mathbb{N}^n_d} p_\alpha \, (z_{01}^{\alpha_1} \cdots z_{0n}^{\alpha_n})^\beta \\ &= \int_{\mathbb{T}^n} \mathbf{z}^\beta \, d\left(\sum_{\alpha \in \mathbb{N}^n_d} p_\alpha \, \delta_{\mathbf{z}_0^\alpha} \right) \;=\; \int_{\mathbb{T}^n} \mathbf{z}^\beta \, d\phi. \end{aligned}$$

In other words, one may identify the polynomial p with an atomic signed Borel measure ϕ on $\mathbb{T}^n$ supported on finitely many atoms $(\mathbf{z}_0^\alpha)_{\alpha \in \mathbb{N}^n_t}$ with associated weights $(p_\alpha)_{\alpha \in \mathbb{N}^n_t}$.

Therefore, if the evaluations of the black-box polynomial p are done at a few "powers" $(\mathbf{z}_0^\beta)$, $\beta \in \mathbb{N}^n$, of an arbitrary point $\mathbf{z}_0 \in \mathbb{T}^n$, then the sparse interpolation problem is equivalent to recovering an unknown atomic signed Borel measure ϕ on $\mathbb{T}^n$ from knowledge of a few moments, that is, the Super-Resolution problem that we have just described above. Hence one may recover p by solving the Moment-SOS hierarchy (3.13) for which finite convergence usually occurs fast. For more details see [16].

3.5. **In Control & Optimal Control.** Consider the Optimal Control Problem (OCP) associated with a controlled dynamical system:

$$\tag{3.14} \begin{array}{rl} J^* = \displaystyle\inf_{\mathbf{u}(t)} & \displaystyle\int_0^T L(\mathbf{x}(t), \mathbf{u}(t))\, dt : \dot{\mathbf{x}}(t) \;=\; f(\mathbf{x}(t), \mathbf{u}(t)),\; t \in (0,T) \\ & \mathbf{x}(t) \in \mathbf{X},\; \mathbf{u}(t) \in \mathbf{U},\, \forall t \in (0,T) \\ & \mathbf{x}(0) = \mathbf{x}_0;\; \mathbf{x}(T) \in \mathbf{X}_T, \end{array}$$

where L, f are polynomials, $\mathbf{X}, \mathbf{X}_T \subset \mathbb{R}^n$ and $\mathbf{U} \subset \mathbb{R}^p$ are compact basic semi-algebraic sets. In full generality the OCP problem (3.14) is difficult to solve, especially when state constraints $\mathbf{x}(t) \in \mathbf{X}$ are present. Given an admissible state-control trajectory $(t, \mathbf{x}(t), \mathbf{u}(t))$, its associated occupation measure ϕ_1 up to time T (resp. ϕ_2 at time T) are defined by:

$$\phi_1(A \times B \times C) := \int_{[0,T]\cap C} 1_{(A,B)}((\mathbf{x}(t), \mathbf{u}(t))\, dt; \quad \phi_2(D) = 1_D(\mathbf{x}(T)),$$

for all $A \in \mathcal{B}(\mathbf{X})$, $B \in \mathcal{B}(\mathbf{U})$, $C \in \mathcal{B}([0,T])$, $D \in \mathcal{B}(\mathbf{X}_T)$. Then for every differentiable function $h : \mathbf{X} \times [0,T] \to \mathbb{R}$

$$h(T, \mathbf{x}(T)) - h(0, x_0) = \int_0^T \left(\frac{\partial h(\mathbf{x}(t), \mathbf{u}(t))}{\partial t} + \frac{\partial h(\mathbf{x}(t), \mathbf{u}(t))}{\partial \mathbf{x}} f(\mathbf{x}(t), \mathbf{u}(t))\right) dt,$$

or, equivalently, with $\mathbf{S} := [0,T] \times \mathbf{X} \times \mathbf{U}$:

$$\int_{\mathbf{X}_T} h(T, \mathbf{x})\, d\phi_2(\mathbf{x}) = h(0, \mathbf{x}_0) + \int_{\mathbf{S}} \left(\frac{\partial h(\mathbf{x}, \mathbf{u})}{\partial t} + \frac{\partial h(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}} f(\mathbf{x}, \mathbf{u})\right) d\phi_1(t, \mathbf{x}, \mathbf{u}).$$

Then *the weak formulation* of the OCP (3.14) is the infinite-dimensional linear program:

$$\begin{array}{rl} \rho = \inf\limits_{\phi_1, \phi_2 \geq 0} & \{ \displaystyle\int_{\mathbf{S}} L(\mathbf{x}, \mathbf{u})\, d\phi_1 : \\ \text{s.t.} & \displaystyle\int_{\mathbf{X}_T} h(T, \cdot)\, d\phi_2 - \int_{\mathbf{S}} \left(\frac{\partial h}{\partial t} + \frac{\partial h}{\partial \mathbf{x}} f\right) d\phi_1 = h(0, \mathbf{x}_0) \\ & \forall h \in \mathbb{R}[t, \mathbf{x}] \}. \end{array} \tag{3.15}$$

It turns out that under some conditions the optimal values of (3.14) and (3.15) are equal, i.e., $J^* = \rho$. Next, if one replaces "*for all* $h \in \mathbb{R}[t, \mathbf{x}, \mathbf{u}]$" with "*for all* $t^k \mathbf{x}^\alpha \mathbf{u}^\beta$", $(t, \alpha, \beta) \in \mathbb{N}^{1+n+p}$", then (3.15) is an instance of the GPM (3.2). Therefore one may apply the Moment-SOS hierarchy (3.5). Under the conditions of Theorem 3.1 one obtains the asymptotic convergence $\rho_d \to \rho = J^*$ as $d \to \infty$. For more details see [32] and the many references therein.

Robust control. In some applications (e.g. in robust control) one is often interested in optimizing over sets of the form:

$$\mathbf{G} := \{\mathbf{x} \in \mathbf{\Omega}_1 : f(\mathbf{x}, \mathbf{u}) \geq 0, \forall \mathbf{u} \in \mathbf{\Omega}_2\},$$

where $\mathbf{\Omega}_2 \subset \mathbb{R}^p$, and $\mathbf{\Omega}_1 \subset \mathbb{R}^n$ is a simple set, in fact a compact set such that one knows all moments of the Lebesgue measure λ on $\mathbf{\Omega}_1$.

The set $\mathbf{G}$ is difficult to handle because of the universal quantifier. Therefore one is often satisfied with an inner approximation $\mathbf{G}_d \subset \mathbf{G}$, and if possible, with (i) a simple form and (ii) some theoretical approximation guarantees. We propose to approximate $\mathbf{G}$ from inside by sets of (simple) form $\mathbf{G}_d = \{\mathbf{x} \in \mathbf{\Omega}_1 : p_d(\mathbf{x}) \geq 0\}$ where $p_d \in \mathbb{R}[\mathbf{x}]_{2d}$.

To obtain such an inner approximation $\mathbf{G}_d \subset \mathbf{G}$, define $F : \mathbf{\Omega}_1 \to \mathbb{R}$, $\mathbf{x} \mapsto F(\mathbf{x}) := \min_{\mathbf{u}} \{f(\mathbf{x}, \mathbf{u}) : \mathbf{u} \in \mathbf{\Omega}_2\}$. Then with $d \in \mathbb{N}$, fixed, solve:

$$\inf_{p \in \mathbb{R}[\mathbf{x}]_{2d}} \int_{\mathbf{\Omega}_1} (F - p)\, d\lambda : f(\mathbf{x}, \mathbf{u}) - p(\mathbf{x}) \geq 0, \forall (\mathbf{x}, \mathbf{u}) \in \mathbf{\Omega}_1 \times \mathbf{\Omega}_2\}. \tag{3.16}$$

Any feasible solution p_d of (3.16) is such that $\mathbf{G}_d = \{\mathbf{x} : p_d(\mathbf{x}) \geq 0\} \subset \mathbf{G}$. In (3.16) $\int_{\mathbf{\Omega}_1} (F - p)\, d\lambda = \|F - p\|_1$ (with $\|\cdot\|_1$ being the $L_1(\mathbf{\Omega}_1)$-norm), and

$$\inf_p \int_{\mathbf{\Omega}_1} (F - p)\, d\lambda = \underbrace{\int_{\mathbf{\Omega}_1} F\, d\lambda}_{=\text{cte}} + \inf_p \int_{\mathbf{\Omega}_1} -p\, d\lambda = \text{cte} - \sup_p \int_{\mathbf{\Omega}_1} p\, d\lambda$$

and so in (3.16) it is equivalent to maximize $\int_{\mathbf{\Omega}_1} p d\lambda$. Again the Moment-SOS hierarchy can be applied. This time one replaces the difficult positivity constraint $f(\mathbf{x}, \mathbf{u}) - p(\mathbf{x}) \geq 0$ for all $(\mathbf{x}, \mathbf{u}) \in \mathbf{\Omega}_1 \times \mathbf{\Omega}_2$ with a certificate of positivity, with a degree bound on the SOS weights. That is, if $\mathbf{\Omega}_1 = \{\mathbf{x} : g_{1,\ell}(\mathbf{x}) \geq 0, \ell = 1, \ldots, m_1\}$ and $\mathbf{\Omega}_2 = \{\mathbf{u} : g_{2,\ell}(\mathbf{u}) \geq 0, \ell = 1, \ldots, m_2\}$, then with $d_{i,\ell} := \lceil (\deg(\sigma_{i,\ell})/2 \rceil$, one solves

$$\begin{aligned} \rho_d = \sup_{p \in \mathbb{R}[\mathbf{x}]_{2d}} \quad & \int_{\mathbf{\Omega}_1} p\, d\lambda : f(\mathbf{x}, \mathbf{u}) - p(\mathbf{x}) = \sigma_0(\mathbf{x}, \mathbf{u}) \\ & + \sum_{\ell=1}^{m_1} \sigma_{1,\ell}(\mathbf{x}, \mathbf{u})\, g_{i,\ell}(\mathbf{x}) + \sum_{\ell=1}^{m_2} \sigma_{2,\ell}(\mathbf{x}, \mathbf{u})\, g_{i,\ell}(\mathbf{u}) \\ & \sigma_{i,\ell} \in \Sigma[\mathbf{x}, \mathbf{u}]_{d - d_{i,\ell}},\ \ell = 1, \ldots, m_i,\ i = 1, 2. \end{aligned} \tag{3.17}$$

Theorem 3.2 ([27])**.** *Assume that $\mathbf{\Omega}_1 \times \mathbf{\Omega}_2$ is compact and its associated quadratic module is Archimedean. Let p_d be an optimal solution of (3.17). If $\lambda(\{\mathbf{x} \in \mathbf{\Omega}_1 : F(\mathbf{x}) = 0\}) = 0$ then $\lim_{d \to \infty} \|F - p_d\|_1 = 0$ and $\lim_{d \to \infty} \lambda(\mathbf{G} \backslash \mathbf{G}_d) = 0$.*

Therefore one obtains a nested sequence of inner approximations $(\mathbf{G}_d)_{d \in \mathbb{N}} \subset \mathbf{G}$, with the desirable property that $\lambda(\mathbf{G} \setminus \mathbf{G}_d)$ vanishes as d increases. For more details the interested reader is referred to [27].

Example 1. In some robust control problems one would like to approximate as closely as desired a non-convex set $\mathbf{G} = \{\mathbf{x} \in \mathbf{\Omega}_1 : \lambda_{\min}(\mathbf{A}(\mathbf{x})) \succeq 0\}$ for some real symmetric $r \times r$ matrix-polynomial $\mathbf{A}(\mathbf{x})$, and where $\mathbf{x} \mapsto \lambda_{\min}(\mathbf{A}(\mathbf{x}))$ denotes its smallest eigenvalue. If one rewrites

$$\mathbf{G} = \{\mathbf{x} \in \mathbf{\Omega}_1 : \mathbf{u}^T \mathbf{A}(\mathbf{x}) \mathbf{u} \geq 0, \forall \mathbf{u} \in \mathbf{\Omega}_2\}; \quad \mathbf{\Omega}_2 = \{\mathbf{u} \in \mathbb{R}^r : \|\mathbf{u}\| = 1\},$$

one is faced with the problem we have just described. In applying the above methodology the polynomial p_d in Theorem 3.2 approximates $\lambda_{\min}(\mathbf{A}(\mathbf{x}))$ from below in $\mathbf{\Omega}_1$, and $\|p_d(\cdot) - \lambda_{\min}(\mathbf{A}(\cdot))\|_1 \to 0$ as d increases. For more details see [13].

There are many other applications of the Moment-SOS hierarchy in Control, e.g. in Systems Identification [5, 3], Robotics [46], for computing Lyapunov functions [44], largest regions of attraction [12], to cite a few.

3.6. **Some inverse optimization problems.** In particular:

Inverse Polynomial Optimization. Here we are given a polynomial optimization problem $\mathbf{P}\,:\; f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in \boldsymbol{\Omega}\}$ with $f \in \mathbb{R}[\mathbf{x}]_d$, and we are interested in the following issue: Let $\mathbf{y} \in \boldsymbol{\Omega}$ be given, e.g. $\mathbf{y}$ is the current iterate of a local minimization algorithm applied to $\mathbf{P}$. Find

$$g^* \;=\; \arg\min_{g\in\mathbb{R}[\mathbf{x}]_d} \{\|f-g\|_1 \,:\, g(\mathbf{x}) - g(\mathbf{y}) \,\geq\, 0,\, \forall \mathbf{x} \in \boldsymbol{\Omega}\,\}, \tag{3.18}$$

where $\|h\|_1 = \sum_\alpha |h_\alpha|$ is the ℓ_1-norm of coefficients of $h \in \mathbb{R}[\mathbf{x}]_d$. In other words, one searches for a polynomial $g^* \in \mathbb{R}[\mathbf{x}]_d$ as close as possible to f and such that $\mathbf{y} \in \boldsymbol{\Omega}$ is a global minimizer of g^* on $\boldsymbol{\Omega}$. Indeed if $\|f - g^*\|_1$ is small enough then $\mathbf{y} \in \boldsymbol{\Omega}$ could be considered a satisfying solution of $\mathbf{P}$. Therefore given a fixed small $\epsilon > 0$, the test $\|f - g^*\|_1 < \epsilon$ could be a new stopping criterion for a local optimization algorithm, with a strong theoretical justification.

Again the Moment-SOS hierarchy can be applied to solve (3.18) as positivity certificates are perfect tools to handle the positivity constraint "$g(\mathbf{x})-g(\mathbf{y}) \geq 0$ for all $\mathbf{x} \in \boldsymbol{\Omega}$". Namely with $\boldsymbol{\Omega}$ as in (1.2), solve:

$$\rho_t = \min_{g\in\mathbb{R}[\mathbf{x}]_d} \{\, \|f-g\|_1 : g(\mathbf{x}) - g(\mathbf{y}) \,:=\, \sum_{j=0}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \quad \forall \mathbf{x}\,\}, \tag{3.19}$$

where $g_0(\mathbf{x}) = 1$ for all $\mathbf{x}$, and $\sigma_j \in \Sigma[\mathbf{x}]_{t-d_j}$, $j = 0,\dots,m$. Other norms are possible but for the sparsity inducing ℓ_1-norm $\|\cdot\|_1$, it turns out that an optimal solution g^* of (3.19) has a canonical simple form. For more details the interested reader is referred to [33].

Inverse Optimal Control. With the OCP (3.14) in §3.5, we now consider the following issue: *Given a database of admissible trajectories* $(\mathbf{x}(t;\mathbf{x}_\tau), \mathbf{u}(t,\mathbf{x}_\tau))$, $t \in [\tau, T]$, *starting in initial state* $\mathbf{x}_\tau \in \mathbf{X}$ *at time* $\tau \in [0, T]$, *does there exist a Lagrangian* $(\mathbf{x}, \mathbf{u}) \mapsto L(\mathbf{x}, \mathbf{u})$ *such that all these trajectories are optimal for the OCP problem (3.14)?* This problem has important applications, e.g., in Humanoid Robotics to explain human locomotion [34].

Again the Moment-SOS hierarchy can be applied because a weak version of the Hamilton-Jacobi-Bellman (HJB) optimality conditions is the perfect tool to state whether some given trajectory is ϵ-optimal for the OCP (3.14). Indeed given $\epsilon > 0$ and an admissible trajectory $(t, \mathbf{x}^*(t), \mathbf{u}^*(t))$, let $\varphi : [0, T] \times \mathbf{X} \to \mathbb{R}$, and $L : \mathbf{X} \times \mathbf{U} \to \mathbb{R}$, be such that:

$$\varphi(T,\mathbf{x}) \leq 0,\, \forall\, \mathbf{x} \in \mathbf{X};\; \frac{\partial\varphi(t,\mathbf{x})}{\partial t} + \frac{\partial\varphi(t,\mathbf{x})}{\partial \mathbf{x}} f(\mathbf{x},\mathbf{u}) + L(\mathbf{x},\mathbf{u}) \,\geq\, 0, \tag{3.20}$$

for all $(t, \mathbf{x}, \mathbf{u}) \in [0, T] \times \mathbf{X} \times \mathbf{U}$, and: $\varphi(T, \mathbf{x}^*(T)) > -\epsilon$,

$$\frac{\partial\varphi(t,\mathbf{x}^*(t))}{\partial t} + \frac{\partial\varphi(t,\mathbf{x}^*(t))}{\partial \mathbf{x}} f(\mathbf{x}^*(t),\mathbf{u}^*(t)) + L(\mathbf{x}^*(t),\mathbf{u}^*(t)) < \epsilon, \tag{3.21}$$

for all $t \in [0, T]$. Then the trajectory $(t, \mathbf{x}^*(t), \mathbf{u}^*(t))$ is an ϵ-optimal solution of the OCP (3.14) with $\mathbf{x}_0 = \mathbf{x}^*(0)$ and Lagrangian L. Therefore to apply the Moment-SOS hierarchy:

(i) The unknown functions φ and L are approximated by polynomials in $\mathbb{R}[t, \mathbf{x}]_{2d}$ and $\mathbb{R}[\mathbf{x}, \mathbf{u}]_{2d}$, where d is the parameter in the Moment-SOS hierarchy (3.6).

(ii) The above positivity constraint (3.20) on $[0, T] \times \mathbf{X} \times \mathbf{U}$ is replaced with a positivity certificate with degree bound on the SOS weights.

(iii) (3.21) is stated for every trajectory $(\mathbf{x}(t; \mathbf{x}_\tau), \mathbf{u}(t, \mathbf{x}_\tau))$, $t \in [\tau, T]$, in the database. Using a discretization $\{t_1, \ldots, t_N\}$ of the interval $[0, T]$, the positivity constraints (3.21) then become a set of linear constraints on the coefficients of the unknown polynomials φ and L.

(iv) ϵ in (3.21) is now taken as a variable and one minimizes a criterion of the form $\|L\|_1 + \gamma\, \epsilon$, where $\gamma > 0$ is chosen to balance between the sparsity-inducing norm $\|L\|_1$ of the Lagrangian and the error ϵ in the weak version of the optimality conditions (3.20)-(3.21). A detailed discussion and related results can be found in [45].

3.7. **Optimal design in statistics.** In designing experiments one models the responses $z_1, \ldots, z_N$ of a random *experiment* whose inputs are represented by a vector $\mathbf{t} = (t_i) \in \mathbb{R}^n$ with respect to known *regression functions* $\Phi = (\varphi_1, \ldots, \varphi_p)$, namely: $z_i = \sum_{j=1}^p \theta_j\, \varphi_j(t_i) + \varepsilon_i$, $i = 1, \ldots, N$, where $\theta_1, \ldots, \theta_p$ are unknown parameters that the experimenter wants to estimate, ε_i is some noise and the (t_i)'s are chosen by the experimenter in a *design space* $\mathcal{X} \subseteq \mathbb{R}^n$. Assume that the inputs t_i, $i = 1, \ldots, N$, are chosen within a set of distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_\ell \in \mathcal{X}$, $\ell \leq \mathbb{N}$, and let n_k denote the number of times the particular point $\mathbf{x}_k$ occurs among $t_1, \ldots, t_N$. A design ξ is then defined by:

$$\xi = \left(\begin{array}{ccc} \mathbf{x}_1 & \ldots & \mathbf{x}_\ell \\ \frac{n_1}{\mathbb{N}} & \ldots & \frac{n_\ell}{N} \end{array} \right). \tag{3.22}$$

The matrix $\mathbf{M}(\xi) := \sum_{i=1}^{\ell} \frac{n_i}{N} \Phi(\mathbf{x}_i)\, \Phi(\mathbf{x}_i)^T$ is called the information matrix of ξ. Optimal design is concerned with finding a set of points in $\mathcal{X}$ that optimizes a certain statistical criterion $\phi(\mathbf{M}(\xi))$, which must be real-valued, positively homogeneous, non constant, upper semi-continuous, isotonic w.r.t. Loewner ordering, and concave. For instance in *D-optimal design* one maximizes $\phi(\mathbf{M}(\xi)) := \log\det(\mathbf{M}(\xi))$ over all ξ of the form (3.22). This is a difficult problem and so far most methods have used a discretization of the design space $\mathcal{X}$.

The Moment-SOS hierarchy that we describe below does not rely an any discretization and works for an arbitrary compact basic semi-algebraic design space $\mathcal{X}$ as defined in (1.2). Instead we look for an atomic measure on $\mathcal{X}$ (with finite support) and we proceed in two steps:

- In the first step one solves the hierarchy of convex optimization problems indexed by $\delta = 0, 1, \ldots$.

$$\rho_\delta = \sup_{\mathbf{y}} \quad \{\log\det(\mathbf{M}_d(\mathbf{y})) : \, y_0 = 1 \qquad \mathbf{M}_{d+\delta}(\mathbf{y}) \,\succeq\, 0; \, \mathbf{M}_{d+\delta-d_j}(g_j\,\mathbf{y}) \succeq 0\}, \tag{3.23}$$

where d is fixed by the number of basis functions φ_j considered (here the monomials $(\mathbf{x}^\alpha)_{\alpha\in\mathbb{N}^n_d}$). (Note that (3.23) is not an SDP because the criterion is not linear in $\mathbf{y}$, but it is still a tractable convex problem.) This provides us with an optimal solution $\mathbf{y}^*(\delta)$. In practice one chooses $\delta = 0$.

• In a second step we extract an atomic measure μ from the "moments" $\mathbf{y}^*(\delta)$, e.g. via Nie's method [42] which consists of solving the SDP:

$$\rho_r = \sup_{\mathbf{y}} \quad \{L_{\mathbf{y}}(f_r) : \, y_\alpha = y^*_\alpha(\delta), \, \forall \alpha \in \mathbb{N}^n_{2d} \qquad \mathbf{M}_{d+r}(\mathbf{y}) \,\succeq\, 0; \, \mathbf{M}_{d+r-d_j}(g_j\,\mathbf{y}) \succeq 0\}, \tag{3.24}$$

where f_r is a (randomly chosen) polynomial strictly positive on $\mathcal{X}$. If $(y^*_\alpha(\delta))_{\alpha\in\mathbb{N}^n_{2d}}$ has a representing measure then it has an atomic representing measure, and generically the rank condition (2.10) will be satisfied. Extraction of atoms is obtained via a linear algebra routine. We have tested this two-steps method on several non-trivial numerical experiments (in particular with highly non-convex design spaces $\mathcal{X}$) and in all cases we were able to obtain a design. For more details the interested reader is referred to [7].

Other applications & extensions. In this partial overview, by lack of space we have not described some impressive success stories of the Moment-SOS hierarchy, e.g. in coding [2], packing problems in discrete geometry [9, 50]. Finally, there is also a *non-commutative* version [47] of the Moment-SOS hierarchy based on non-commutative positivity certificates [11] and with important applications in quantum information [38].

4. Conclusion

The list of important applications of the GPM is almost endless and we have tried to convince the reader that the Moment-SOS hierarchy is one promising powerful tool for solving the GPM with already some success stories. However much remains to be done as its brute force application does not scale well to the problem size. One possible research direction is to exploit symmetries and/or sparsity in large scale problems. Another one is to determine alternative positivity certificates which are less expensive in terms of computational burden to avoid the size explosion of SOS-based positivity certificates.

References

[1] Barak, B., Steurer, D. *Sum-of-Squares Proofs and the Quest toward Optimal Algorithms*, Proc. ICM 2014, Seoul, Korea.

[2] Bachoc, C., Vallentin, F. *New upper bounds for kissing numbers from semidefinite programming*, J. Amer. Math. Soc. **21** (2008), 909–924.

[3] Camps, O., Sznaier, M. *The interplay between Big-data and sparsity in systems identification*, in Laumond J.-P., Mansard N., and Lasserre J.B. (editors) *Geometric and Numerical Foundations of Movements*, pp. 133–159, Springer Tracts in Advanced Robotics **117**, Springer, New York, 2017.

[4] Candès, E.J., Fernandez-Granda, C. *Towards a Mathematical Theory of Super-Resolution*, Comm. Pure Appl. Math. **67** (2014), 906–956.

[5] Cerone, V., Piga, D., Regruto, D. *Set-membership error-in-variables identification through convex relaxation techniques*, IEEE Trans. Aut. Control **57** (2012), 517–522

[6] De Castro, Y., Gamboa, F., Henrion, D., Lasserre, J.B. *Exact solutions to Super Resolution on semi-algebraic domains in higher dimensions*, IEEE Trans. Info. Theory **63** (2017), 621–630.

[7] De Castro, Y., Gamboa, F., Henrion, D., Hess, R., Lasserre, J.B. *Approximate Optimal Designs for Multivariate Polynomial Regression*, LAAS report No 17044. 2017, Toulouse, France. `hal-01483490v2`. To appear in Annals of Statistics.

[8] De Klerk, E., Lasserre, J.B., Laurent, M., Sun Zhao. (2017) *Bound-Constrained Polynomial Optimization Using Only Elementary Calculations*, Math. Oper. Res. **42** (2017), 834–853.

[9] de Laat, D., Vallentin, F. *A semidefinite programming hierarchy for packing problems in discrete geometry*, Math. Program. **151** (2015), 1–25.

[10] Goemans, M.X., Williamson, D.P. *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, JACM **42** (1995), 1115–1145.

[11] Helton J.W., McCullough, S. *A Positivstellensatz for Noncommutative Polynomials*, Trans. Amer. Math. Soc. **356** (2004), 3721–3737

[12] Henrion, D., Korda, M. *Convex computation of the region of attraction of polynomial control systems*, IEEE Trans. Aut. Control **59** (2014), 297–312.

[13] Henrion, D., Lasserre, J.B. *Convergent relaxations of polynomial matrix inequalities and static output feedback*, IEEE Trans. Auto. Control **51** (2006), 192–202.

[14] Henrion, D., Lasserre, J.B., Lofberg, J. *Gloptipoly 3: moments, optimization and semidefinite programming*, Optim. Methods Softwares **24** (2009), 761–779.

[15] Henrion, D., Lasserre, J.B., Savorgnan, C. *Approximate volume and integration of basic semi-algebraic sets*, SIAM Review **51** (2009), 722–743.

[16] Josz, C. Lasserre, J.B., Mourrain, B. *Sparse polynomial interpolation: compressed sensing, super resolution, or Prony?*, LAAS Report no 17279. 2017, Toulouse, France. `arXiv:1708.06187`

[17] Khot, S. *Innapproximability of NP-complete Problems, Discrete Fourier Analysis, and Geometry*, Proc. ICM 2010, Hyderabad, India, 2010.

[18] Khot, S. *Hardness of approximation*, Proc. ICM 2014, Seoul, Korea, 2014.

[19] Krivine, J.L. *Anneaux préordonnés*, J. Anal. Math. **12** (1964), 307–326.

[20] Krivine, J.L. *Quelques propriétés des préordres dans les anneaux commutatifs*, C.R. Acd. Sci. Paris, Ser. I **258** (1964), 3417–3418.

[21] Landau, H.J. (Editor): *Moments in Mathematics*, Proc. Sympos. Appl. Math. **37** (1987).

[22] Lasserre, J.B. *Optimisation globale et théorie des moments*, C.R. Acad. Sci. Paris, Sér. I **331** (2000), 929–934.

[23] Lasserre, J.B. *Global Optimization with polynomials and the problem of moments*, SIAM J. Optim. **11**(2001), 796–817.

[24] Lasserre, J.B., Laurent, M., Rostalski, P. *Semidefinite characterization and computation of zero-dimensional real radical ideals*, Found. Comput. Math. **8** (2008), 607–647.

[25] Lasserre, J.B. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.

[26] Lasserre, J.B. *An Introduction to Polynomial and Semi-Algebraic Optimization*, Cambridge University Press, Cambridge, UK, 2015.

[27] Lasserre, J.B. *Tractable approximations of sets defined with quantifiers*, Math. Program. **151** (2015), 507–527.

[28] Lasserre, J.B. *Computing Gaussian & exponential measures of semi-algebraic sets*, Adv. Appl. Math. **91** (2017), 137–163.

[29] Lasserre, J.B. *Semidefinite programming vs. LP relaxations for polynomial programming*, Math. Oper. Res. **27** (2002), 347–360.

[30] Lasserre, J.B. *Bounds on measures satisfying moment conditions*, Annals Appl. Prob. **12** (2002), 1114–1137.

[31] Lasserre, J.B. *Convergent SDP-relaxations in polynomial optimization with sparsity*, SIAM J. Optim. **17** (2006), 822–843.

[32] Lasserre, J.B., Henrion, D., Prieur, C., Trélat, E. *Nonlinear optimal control via occupation measures and LMI-relaxations*, SIAM J. Contr. Optim. **47** (2008), 1649–1666.

[33] Lasserre, J.B. *Inverse polynomial optimization*, Math. Oper. Res. **38** (2013), 418–436.

[34] Laumond, J.-P., Mansard, N., Lasserre, J.B. (editors). *Geometric and Numerical Foundations of Movements*, Springer Tracts in Advanced Robotics **117**, Springer, New York, 2017.

[35] Laurent, M. *A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming*, Math. Oper. Res. **28** (2003), 470–496.

[36] Laurent, M. *Sums of squares, moment matrices and optimization over polynomials*, In Putinar M. and Sullivant S. (editors), *Emerging Applications of Algebraic Geometry*, pp. 157–270, IMA Volumes in Mathematics and its Applications **149**, Springer, New York, 2009.

[37] Molzahn, D.K., Hiskens, I.K. *Sparsity-Exploiting Moment-Based Relaxations of the Optimal Power Flow Problem*, IEEE Trans. Power Systems **30** (2015), 3168–3180.

[38] Navascués, M., Pironio, S., A Acín, A. *A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations*, New J. Physics **10** (2008).

[39] Nesterov, Y. *Squared functional systems and optimization problems*, in Frenk H., Roos K., Terlaky T., and Zhang S. (editors), *High Performance Optimization*, pp. 405–440, Springer, New York, 2000.

[40] Nie, J. *Optimality Conditions and Finite Convergence of Lasserre's Hierarchy*, Math. Program. Ser. A **146** (2014), 97–121.

[41] Nie, J. *Certifying Convergence of Lasserre's Hierarchy via Flat Truncation*, Math. Program. Ser. A, **42** (2013), 485–510.

[42] Nie, J. *The $\mathcal{A}$-Truncated K-Moment Problem*, Found. Comput. Math. **14** (2014), 1243–1276.

[43] Parrilo, P. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*, PhD Thesis, California Institute of Technology, Pasadena, CA, 2000.

[44] Parrilo, P. *Semidefinite programming relaxations for semialgebraic problems*, Math. Program. **96** (2003), 293–320.

[45] Pauwels, E., Henrion, D., Lasserre, J.B. *Linear Conic Optimization for Inverse Optimal Control*, SIAM J. Control Optim. **54** (2016), 1798–1825.

[46] Posa, M., Tobenkin, M., Tedrake, R. *Stability analysis and control of rigid-body systems with impacts and friction*, IEEE Trans. Aut. Control **61**, (2016), 1423–1437.

[47] Pironio, S., M Navascués, M., Acín, A. *Convergent relaxations of polynomial optimization problems with noncommuting variables*, SIAM J. Optim. **20** (2010), 2157–2180

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

[49] Schmüdgen, K. *The Moment problem*, Springer, 2017.

[50] Schürmann, A., Vallentin, F. *Computational approaches to lattice packing and covering problems*, Discrete & Comput. Geom. **35** (2006), 73–116

[51] Shor, N.Z. *Nondifferentiable Optimization and Polynomial Problems*, Kluwer, Dordrecht, 1988.

[52] Vasilescu, F.-H. *Spectral measures and moment problems*, in *Spectral Theory and Applications*, Theta Ser. Adv. Math. **2** (2003), Theta, Bucharest, pp. 173–215.

[53] Waki, K., Kim, S., Kojima, M., Muramatsu, M. *Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity*, SIAM J. Optim. **17** (2006), 218–242