APPROXIMATE OPTIMAL DESIGNS FOR MULTIVARIATE POLYNOMIAL REGRESSION

By Yohann De Castro⋆, Fabrice Gamboa°, Didier Henrion•, Roxana Hess• and Jean-Bernard Lasserre•

⋆*Laboratoire de Mathématiques d'Orsay*
Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France
°*Institut de Mathématiques de Toulouse*
Univ. Paul Sabatier, CNRS, 118 route de Narbonne, 31062 Toulouse, France
•*Laboratoire d'Analyse et d'Architecture des Systèmes*
Université de Toulouse, CNRS, 7 avenue du col. Roche, 31400 Toulouse, France

Abstract We introduce a new approach aiming at computing approximate optimal designs for multivariate polynomial regressions on compact (semi-algebraic) design spaces. We use the moment-sum-of-squares hierarchy of semidefinite programming problems to solve numerically the approximate optimal design problem. The geometry of the design is recovered via semidefinite programming duality theory. This article shows that the hierarchy converges to the approximate optimal design as the order of the hierarchy increases. Furthermore, we provide a dual certificate ensuring finite convergence of the hierarchy and showing that the approximate optimal design can be computed numerically with our method. As a byproduct, we revisit the equivalence theorem of the experimental design theory: it is linked to the Christoffel polynomial and it characterizes finite convergence of the moment-sum-of-square hierarchies.

1. Introduction.

1.1. *Convex design theory.* The optimal experimental designs are computational and theoretical objects that aim at minimizing the uncertainty contained in the best linear unbiased estimators in regression problems. In this frame, the experimenter models the responses $z_1, \ldots, z_N$ of a random *experiment* whose inputs are represented by a vector $t_i \in \mathbb{R}^n$ with respect to known *regression functions* $\mathbf{f}_1, \ldots, \mathbf{f}_p$, namely

$$z_i = \sum_{j=1}^{p} \theta_j \mathbf{f}_j(t_i) + \varepsilon_i, \ \ i = 1, \ldots, N,$$

1

where $\theta_1, \ldots, \theta_p$ are unknown parameters that the experimenter wants to estimate, ε_i, $i = 1, \ldots, N$ are i.i.d. centered square integrable random variables and the inputs t_i are chosen by the experimenter in a *design space* $\mathcal{X} \subseteq \mathbb{R}^n$. In this paper, we consider that the regression functions $\mathbf{F}$ are multivariate polynomials of degree at most d.

Assume that the inputs t_i, for $i = 1, \ldots, N$, are chosen within a set of distinct points $x_1, \ldots, x_\ell$ with $\ell \leq N$, and let n_k denote the number of times the particular point x_k occurs among $t_1, \ldots, t_N$. This would be summarized by defining a design ξ as follows

$$(1) \qquad \xi := \begin{pmatrix} x_1 & \cdots & x_\ell \\ \frac{n_1}{N} & \cdots & \frac{n_\ell}{N} \end{pmatrix},$$

whose first row gives distinct points in the design space $\mathcal{X}$ where the inputs parameters have to be taken and the second row indicates the experimenter which proportion of experiments (frequencies) have to be done at these points. We refer to the inspiring book of Dette and Studden [3] and references therein for a complete overview on the subject of the theory of optimal design of experiments. We denote the *information matrix* of ξ by

$$(2) \qquad \mathbf{M}(\xi) := \sum_{i=1}^{\ell} w_i \mathbf{F}(x_i) \, \mathbf{F}^\top(x_i),$$

where $\mathbf{F} := (\mathbf{f}_1, \ldots, \mathbf{f}_p)$ is the column vector of regression functions and $w_i := n_i/N$ is the weight corresponding to the point x_i. In the following, we will not not distinguish between a design ξ as in (1) and a discrete probability measure on $\mathcal{X}$ with finite support given by the points x_i and weights w_i.

Observe that the information matrix belongs to $\mathbb{S}_p^+$, the space of symmetric nonnegative definite matrices of size p. For all $q \in [-\infty, 1]$ define the function

$$\phi_q := \begin{cases} \mathbb{S}_p^+ & \to & \mathbb{R} \\ M & \mapsto & \phi_q(M) \end{cases}$$

where for positive definite matrices M

$$\phi_q(M) := \begin{cases} (\frac{1}{p}\mathrm{trace}(M^q))^{1/q} & \text{if } q \neq -\infty, 0 \\ \det(M)^{1/p} & \text{if } q = 0 \\ \lambda_{\min}(M) & \text{if } q = -\infty \end{cases}$$

and for nonnegative definite matrices M

$$\phi_q(M) := \begin{cases} (\frac{1}{p}\mathrm{trace}(M^q))^{1/q} & \text{if } q \in (0, 1] \\ 0 & \text{if } q \in [-\infty, 0]. \end{cases}$$

We recall that trace(M), det(M) and $\lambda_{\min}(M)$ denote respectively the trace, determinant and least eigenvalue of the symmetric nonnegative definite matrix M. These criteria are meant to be real valued, positively homogeneous, non constant, upper semi-continuous, isotonic (with respect to the Loewner ordering) and concave functions.

Hence, an optimal design is a solution $\xi^\star$ to the following problem

$$\max \phi_q(\mathbf{M}(\xi)) \tag{3}$$

where the maximum is taken over all ξ of the form (1). Standard criteria are given by the parameters $q = 0, -1, -\infty$ and are referred to D, A or E-optimum designs respectively. As detailed in Section 3.2, we restrict our attention to "*approximate*" optimal designs where, by definition, we replace the set of "*feasible*" matrices $\{\mathbf{M}(\xi) : \xi$ of the form (1)$\}$ by the larger set of all possible information matrices, namely the convex hull of $\{\mathbf{F}(x)\,\mathbf{F}^\top(x) : x \in \mathcal{X}\}$. To construct approximate optimal designs, we propose a two-step procedure presented in Algorithm 1. This procedure finds the information matrix $\mathbf{M}^\star$ of the approximate optimal design $\xi^\star$ and then it computes the support points $x_i^\star$ and the weights $w_i^\star$ of the design $\xi^\star$ in a second step.

1.2. *Contribution.* This paper introduces a general method to compute approximate optimal designs—in the sense of Kiefer's ϕ_q-criteria—on a large variety of design spaces that we refer to as semi-algebraic sets, see [8] or Section 2 for a definition. These can be understood as sets given by intersections and complements of superlevel sets of multivariate polynomials. An important distinguishing feature of the method is to not rely on any discretization of the design space which is in contrast to computational methods in previous works, *e.g.*, the algorithms described in [23, 21].

We apply the moment-sum-of-squares hierarchy—referred to as the Lasserre hierarchy—of SDP problems to solve numerically and approximately the optimal design problem. More precisely, we use an outer "*approximation*" (in the SDP relaxation sense) of the set of moments of order d, see Section 2.2 for more details. Note that these approximations are SDP representable so that they can be efficiently encoded numerically. Since the regressors are polynomials, the information matrix $\mathbf{M}$ is a linear function of the moment matrix (of order d). Hence, our approach gives an outer approximation of the set of information matrices, which is SDP representable. As shown by the interesting works [20, 18], the criterion ϕ_q is also SDP representable in the case where q is rational. It proves that our procedure (depicted in Algorithm 1) makes use of two semidefinite programs and it can be efficiently used in

---

Algorithm 1: Approximate Optimal Designs on Semi-Algebraic Sets

---

Data: A compact semi-algebraic design space $\mathcal{X}$ defined as in (4).
Result: An approximate optimal design ξ

1. Choose the two relaxation orders δ and r.
2. Solve the SDP relaxation (7) of order δ for a vector $\mathbf{y}_\delta^\star$.
3. Either solve Nie's SDP relaxation (28) or the Christoffel polynomial SDP relaxation (30) of order r for a vector $\mathbf{y}_r^\star$.
4. If $\mathbf{y}_r^\star$ satisfies the rank condition (29), then extract the optimal design ξ from the truncated moment sequence as explained in Section 5.
5. Otherwise, choose larger values of δ and r and go to Step 2.

---

practice. Note that similar two steps procedures have been presented in the literature, the reader may consult the interesting paper [4] which proposes a way of constructing approximate optimal designs on the hypercube.

The theoretical guarantees are given by Theorem 3 (Equivalence theorem revisited for the finite order hierarchy) and Theorem 4 (convergence of the hierarchy as the order increases). These theorems demonstrate the convergence of our procedure towards the approximate optimal designs as the order of the hierarchy increases. Furthermore, they give a characterization of finite order convergence of the hierarchy. In particular, our method recovers the optimal design when finite convergence of this hierarchy occurs. To recover the geometry of the design we use SDP duality theory and Christoffel polynomials involved in the optimality conditions.

We have run several numerical experiments for which finite convergence holds leading to a surprisingly fast and reliable method to compute optimal designs. As illustrated by our examples, in polynomial regression model with degree order higher than one we obtain designs with points in the interior of the domain.

1.3. *Outline of the paper.* In Section 2, after introducing necessary notation, we shortly explain some basics on moments and moment matrices, and present the approximation of the moment cone via the Lasserre hierarchy. Section 3 is dedicated to further describing optimal designs and their approximations. At the end of the section we propose a two step procedure to solve the approximate design problem, it is described in Algorithm 1. Solving the first step is subject to Section 4. There, we find a sequence of moments $\mathbf{y}^\star$ associated with the optimal design measure. Recovering this measure (step two of the procedure) is discussed in Section 5. We finish the paper with some illustrating examples and a short conclusion.

2. Polynomial optimal designs and moments.   This section collects preliminary material on semi-algebraic sets, moments and moment matrices, using the notation of [8]. This material will be used to restrict our attention to *polynomial* optimal design problems with polynomial regression functions and semi-algebraic design spaces.

2.1. *Polynomial optimal design.*   Denote by $\mathbb{R}[x]$ the vector space of real polynomials in the variables $x = (x_1, \ldots, x_n)$, and for $d \in \mathbb{N}$ define $\mathbb{R}[x]_d := \{p \in \mathbb{R}[x] : \deg p \le d\}$ where $\deg p$ denotes the total degree of p.

We assume that the regression functions are multivariate polynomials, namely $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_p) \in (\mathbb{R}[x]_d)^p$. Moreover, we consider that the design space $\mathcal{X} \subset \mathbb{R}^n$ is a given closed basic semi-algebraic set

(4) $$\mathcal{X} := \{x \in \mathbb{R}^m : g_j(x) \geqslant 0,\ j = 1, \ldots, m\}$$

for given polynomials $g_j \in \mathbb{R}[x]$, $j = 1, \ldots, m$, whose degrees are denoted by d_j, $j = 1, \ldots, m$. Assume that $\mathcal{X}$ is compact with an algebraic certificate of compactness. For example, one of the polynomial inequalities $g_j(x) \geqslant 0$ should be of the form $R^2 - \sum_{i=1}^n x_i^2 \geqslant 0$ for a sufficiently large constant R.

Notice that those assumptions cover a large class of problems in optimal design theory, see for instance [3, Chapter 5]. In particular, observe that the design space $\mathcal{X}$ defined by (4) is not necessarily convex and note that the polynomial regressors $\mathbf{F}$ can handle incomplete m-way dth degree polynomial regression.

The monomials $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, with $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, form a basis of the vector space $\mathbb{R}[x]$. We use the multi-index notation $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ to denote these monomials. In the same way, for a given $d \in \mathbb{N}$ the vector space $\mathbb{R}[x]_d$ has dimension $s(d) := \binom{n+d}{n}$ with basis $(x^\alpha)_{|\alpha| \le d}$, where $|\alpha| := \alpha_1 + \cdots + \alpha_n$. We write

$$\mathbf{v}_d(x) := $$
$$(\ \underbrace{1}_{\text{degree 0}} \ , \underbrace{x_1, \ldots, x_n}_{\text{degree 1}}, \underbrace{x_1^2, x_1 x_2, \ldots, x_1 x_n, x_2^2, \ldots, x_n^2}_{\text{degree 2}}, \ldots, \underbrace{\ldots, x_1^d, \ldots, x_n^d}_{\text{degree } d} \ )^\top$$

for the column vector of the monomials ordered according to their degree, and where monomials of the same degree are ordered with respect to the lexicographic ordering. Note that, by linearity, there exists a unique matrix $\mathfrak{A}$ of size $p \times \binom{n+d}{n}$ such that

(5) $$\forall x \in \mathcal{X}, \quad \mathbf{F}(x) = \mathfrak{A}\, \mathbf{v}_d(x)\,.$$

The cone $\mathscr{M}_+(\mathcal{X})$ of nonnegative Borel measures supported on $\mathcal{X}$ is understood as the dual to the cone of nonnegative elements of the space $\mathscr{C}(\mathcal{X})$ of continuous functions on $\mathcal{X}$.

2.2. *Moments, the moment cone and the moment matrix.* Given a positive measure $\mu \in \mathscr{M}_+(\mathcal{X})$ and $\alpha \in \mathbb{N}^n$, we call

$$y_\alpha = \int_\mathcal{X} x^\alpha d\mu$$

the moment of order α of μ. Accordingly, we call the sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ the moment sequence of μ. Conversely, we say that $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ has a *representing measure*, if there exists a measure μ such that $\mathbf{y}$ is its moment sequence.

We denote by $\mathcal{M}_d(\mathcal{X})$ the convex cone of all truncated sequences $\mathbf{y} = (y_\alpha)_{|\alpha| \leq d}$ which have a representing measure supported on $\mathcal{X}$. We call it the *moment cone* (of order d) of $\mathcal{X}$. It can be expressed as

$$(6) \qquad \mathcal{M}_d(\mathcal{X}) := \left\{ \mathbf{y} \in \mathbb{R}^{\binom{n+d}{n}} : \exists \mu \in \mathscr{M}_+(\mathcal{X}) \text{ s.t.} \right.$$
$$\left. y_\alpha = \int_\mathcal{X} x^\alpha \, d\mu, \ \forall \alpha \in \mathbb{N}^n, \ |\alpha| \leq d \right\}.$$

Let $\mathcal{P}_d(\mathcal{X})$ denotes the convex cone of all polynomials of degree at most d that are nonnegative on $\mathcal{X}$. Note that we assimilate polynomials p of degree at most d with a vector of dimension $s(d)$, which contains the coefficients of p in the chosen basis.

When $\mathcal{X}$ is a compact set, then $\mathcal{M}_d(\mathcal{X}) = \mathcal{P}_d(\mathcal{X})^\star$ and $\mathcal{P}_d(\mathcal{X}) = \mathcal{M}_d(\mathcal{X})^\star$, see *e.g.,* [9, Lemma 2.5] or [7].

When the design space is given by the univariate interval $\mathcal{X} = [a, b]$, *i.e.,* $n = 1$, then this cone is representable using positive semidefinite Hankel matrices, which implies that convex optimization on this cone can be carried out with efficient interior point algorithms for *semidefinite programming*, see *e.g.,* [24]. Unfortunately, in the general case, there is no efficient representation of this cone. It has actually been shown in [22] that the moment cone is not *semidefinite representable*, *i.e.,* it cannot be expressed as the projection of a linear section of the cone of positive semidefinite matrices. However, we can use semidefinite approximations of this cone as discussed in Section 2.3.

Given a real valued sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ we define the linear functional $L_\mathbf{y} : \mathbb{R}[x] \to \mathbb{R}$ which maps a polynomial $f = \sum_{\alpha \in \mathbb{N}^n} f_\alpha x^\alpha$ to

$$L_\mathbf{y}(f) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha y_\alpha.$$

A sequence $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ has a representing measure μ supported on $\mathcal{X}$ if and only if $L_\mathbf{y}(f) \geqslant 0$ for all polynomials $f \in \mathbb{R}[x]$ nonnegative on $\mathcal{X}$ [8, Theorem 3.1].

The *moment matrix* of a truncated sequence $\mathbf{y} = (y_\alpha)_{|\alpha| \leq 2d}$ is the $\binom{n+d}{n} \times \binom{n+d}{n}$-matrix $M_d(\mathbf{y})$ with rows and columns respectively indexed by integer n-tuples $\alpha \in \mathbb{N}^n, |\alpha|, |\beta| \leq d$ and whose entries are given by

$$M_d(\mathbf{y})(\alpha, \beta) = L_{\mathbf{y}}(x^\alpha x^\beta) = y_{\alpha+\beta}.$$

It is symmetric $(M_d(\mathbf{y})(\alpha, \beta) = M_d(\mathbf{y})(\beta, \alpha))$, and linear in $\mathbf{y}$. Further, if $\mathbf{y}$ has a representing measure, then $M_d(\mathbf{y})$ is *positive semidefinite* (written $M_d(\mathbf{y}) \succcurlyeq 0$).

Similarly, we define the *localizing matrix* of a polynomial $f = \sum_{|\alpha| \leq r} f_\alpha x^\alpha \in \mathbb{R}[x]_r$ of degree r and a sequence $\mathbf{y} = (y_\alpha)_{|\alpha| \leq 2d+r}$ as the $\binom{n+d}{n} \times \binom{n+d}{n}$ matrix $M_d(f\mathbf{y})$ with rows and columns respectively indexed by $\alpha, \beta \in \mathbb{N}^n, |\alpha|, |\beta| \leq d$ and whose entries are given by

$$M_d(f\mathbf{y})(\alpha, \beta) = L_{\mathbf{y}}(f(x)\, x^\alpha x^\beta) = \sum_{\gamma \in \mathbb{N}^n} f_\gamma y_{\gamma+\alpha+\beta}.$$

If $\mathbf{y}$ has a representing measure μ, then $M_d(f\mathbf{y}) \succcurlyeq 0$ for $f \in \mathbb{R}[x]_d$ whenever the support of μ is contained in the set $\{x \in \mathbb{R}^n : f(x) \geqslant 0\}$.

Since $\mathcal{X}$ is basic semi-algebraic with a certificate of compactness, by Putinar's theorem—see for instance the book [8, Theorem 3.8], we also know the converse statement in the infinite case. Namely, it holds that $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^n}$ has a representing measure $\mu \in \mathscr{M}_+(\mathcal{X})$ if and only if for all $d \in \mathbb{N}$ the matrices $M_d(\mathbf{y})$ and $M_d(g_j\mathbf{y})$, $j = 1, \ldots, m$, are positive semidefinite.

2.3. *Approximations of the moment cone.* Letting $v_j := \lceil d_j/2 \rceil$, $j = 1, \ldots, m$, denote half the degree of the g_j, by Putinar's theorem, we can approximate the moment cone $\mathcal{M}_{2d}(\mathcal{X})$ by the following semidefinite representable cones for $\delta \in \mathbb{N}$:

(7) $\quad \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X}) := \Big\{ \mathbf{y}_{d,\delta} \in \mathbb{R}^{\binom{n+2d}{n}} \; : \; \exists \mathbf{y}_\delta \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}}$ such that

$$\mathbf{y}_{d,\delta} = (y_{\delta,\alpha})_{|\alpha| \leq 2d} \text{ and}$$

$$M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0, \; M_{d+\delta-v_j}(g_j\mathbf{y}_\delta) \succcurlyeq 0, \; j = 1, \ldots, m \Big\}.$$

By semidefinite representable we mean that the cones are projections of linear sections of semidefinite cones. Since $\mathcal{M}_{2d}(\mathcal{X})$ is contained in every $(\mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X}))_{\delta \in \mathbb{N}}$, they are outer approximations of the moment cone. Moreover, they form a nested sequence, so we can build the hierarchy

(8) $\qquad \mathcal{M}_{2d}(\mathcal{X}) \subseteq \cdots \subseteq \mathcal{M}_{2(d+2)}^{\mathsf{SDP}}(\mathcal{X}) \subseteq \mathcal{M}_{2(d+1)}^{\mathsf{SDP}}(\mathcal{X}) \subseteq \mathcal{M}_{2d}^{\mathsf{SDP}}(\mathcal{X}).$

This hierarchy actually converges, meaning $\mathcal{M}_{2d}(\mathcal{X}) = \overline{\bigcap_{\delta=0}^{\infty} \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X})}$, where $\overline{A}$ denotes the topological closure of the set A.

Further, let $\Sigma[x]_{2d} \subseteq \mathbb{R}[x]_{2d}$ be the set of all polynomials that are sums of squares of polynomials (SOS) of degree at most $2d$, i.e., $\Sigma[x]_{2d} = \{\sigma \in \mathbb{R}[x] : \sigma(x) = \sum_{i=1}^{k} h_i(x)^2 \text{ for some } h_i \in \mathbb{R}[x]_d \text{ and some } k \geq 1\}$. The topological dual of $\mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X})$ is a quadratic module, which we denote by $\mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$. It is given by

$$(9) \qquad \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X}) := \Big\{ h = \sigma_0 + \sum_{j=1}^{m} g_j \sigma_j : \deg(h) \leq 2d,$$

$$\sigma_0 \in \Sigma[x]_{2(d+\delta)}, \ \sigma_j \in \Sigma[x]_{2(d+\delta-\nu_j)}, \ j = 1, \ldots, m \Big\}.$$

Equivalently, see for instance [8, Proposition 2.1], $h \in \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$ if and only if h has degree less than $2d$ and there exist real symmetric and positive semidefinite matrices Q_0 and Q_j, $j = 1, \ldots, m$ of size $\binom{n+d+\delta}{n} \times \binom{n+d+\delta}{n}$ and $\binom{n+d+\delta-\nu_j}{n} \times \binom{n+d+\delta-\nu_j}{n}$ respectively, such that for any $x \in \mathbb{R}^n$

$$h(x) = \sigma_0(x) + \sum_{j=1}^{m} g_j(x)\sigma_j(x)$$

$$= \mathbf{v}_{d+\delta}(x)^\top Q_0 \mathbf{v}_{d+\delta}(x) + \sum_{j=1}^{m} g_j(x) \, \mathbf{v}_{d+\delta-\nu_j}(x)^\top Q_j \mathbf{v}_{d+\delta-\nu_j}(x) \,.$$

The elements of $\mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$ are polynomials of degree at most $2d$ which are non-negative on $\mathcal{X}$. Hence, it is a subset of $\mathcal{P}_{2d}(\mathcal{X})$.

3. Approximate Optimal Design.

3.1. *Problem reformulation in the multivariate polynomial case.* For all $i = 1, \ldots, p$ and $x \in \mathcal{X}$, let $\mathbf{f}_i(x) := \sum_{|\alpha| \leq d} a_{i,\alpha} x^\alpha$ with appropriate $a_{i,\alpha} \in \mathbb{R}$ and note that $\mathfrak{A} = (a_{i,\alpha})$ where $\mathfrak{A}$ is defined by (5). For $\mu \in \mathscr{M}_+(\mathcal{X})$ with moment sequence $\mathbf{y}$ define the information matrix

$$\mathbf{M}_d(\mathbf{y}) := \Big(\int_{\mathcal{X}} \mathbf{f}_i \mathbf{f}_j \mathrm{d}\mu \Big)_{1 \leq i,j \leq p} = \Big(\sum_{|\alpha|,|\beta| \leq d} a_{i,\alpha} a_{j,\beta} y_{\alpha+\beta} \Big)_{1 \leq i,j \leq p} = \sum_{|\gamma| \leq 2d} A_\gamma y_\gamma,$$

where we have set $A_\gamma := \Big(\sum_{\alpha+\beta=\gamma} a_{i,\alpha} a_{j,\beta} \Big)_{1 \leq i,j \leq p}$ for $|\gamma| \leq 2d$. Observe that it holds

$$(10) \qquad\qquad \mathbf{M}_d(\mathbf{y}) = \mathfrak{A} M_d(\mathbf{y}) \mathfrak{A}^\top.$$

If $\mathbf{y}$ is the moment sequence of $\mu = \sum_{i=1}^{\ell} w_i \delta_{x_i}$, where δ_x denotes the Dirac measure at the point $x \in \mathcal{X}$ and the w_i are again the weights corresponding to the points x_i. Observe that $\mathbf{M}_d(\mathbf{y}) = \sum_{i=1}^{\ell} w_i \mathbf{F}(x_i)\mathbf{F}^\top(x_i)$ as in (2).

Consider the optimization problem

$$(11) \qquad \max \ \phi_q(M)$$

$$\text{s.t. } M = \sum_{|\gamma| \leq 2d} A_\gamma y_\gamma \succcurlyeq 0, \quad y_\gamma = \sum_{i=1}^{\ell} \frac{n_i}{N} x_i^\gamma, \quad \sum_{i=1}^{\ell} n_i = N,$$

$$x_i \in \mathcal{X}, \ n_i \in \mathbb{N}, \ i = 1, \ldots, \ell,$$

where the maximization is with respect to x_i and n_i, $i = 1, \ldots, \ell$, subject to the constraint that the information matrix M is positive semidefinite. By construction, it is equivalent to the original design problem (3). In this form, Problem (11) is difficult because of the integrality constraints on the n_i and the nonlinear relation between $\mathbf{y}$, x_i and n_i. We will address these difficulties in the sequel by first relaxing the integrality constraints.

3.2. *Relaxing the integrality constraints.* In Problem (11), the set of admissible frequencies $w_i = n_i/N$ is discrete, which makes it a potentially difficult combinatorial optimization problem. A popular solution is then to consider "*approximate*" designs defined by

$$(12) \qquad \xi := \left(\begin{array}{ccc} x_1 & \cdots & x_\ell \\ w_1 & \cdots & w_\ell \end{array} \right),$$

where the frequencies w_i belong to the unit simplex $\mathcal{W} := \{w \in \mathbb{R}^\ell : 0 \leq w_i \leq 1, \ \sum_{i=1}^{\ell} w_i = 1\}$. Accordingly, any solution to Problem (3), where the maximum is taken over all matrices of type (12), is called "*approximate optimal design*", yielding the following relaxation of Problem (11)

$$(13) \qquad \max \ \phi_q(M)$$

$$\text{s.t. } M = \sum_{|\gamma| \leq 2d} A_\gamma y_\gamma \succcurlyeq 0, \quad y_\gamma = \sum_{i=1}^{\ell} w_i x_i^\gamma,$$

$$x_i \in \mathcal{X}, \ w \in \mathcal{W},$$

where the maximization is with respect to x_i and w_i, $i = 1, \ldots, \ell$, subject to the constraint that the information matrix M is positive semidefinite. In this problem the nonlinear relation between $\mathbf{y}$, x_i and w_i is still an issue.

3.3. *Moment formulation.*   Let us introduce a two-step-procedure to solve the approximate optimal design Problem (13). For this, we first reformulate our problem again.

By Carathéodory's theorem, the subset of moment sequences in the truncated moment cone $\mathcal{M}_{2d}(\mathcal{X})$ defined in (6) and such that $y_0 = 1$, is exactly the set:

$$\left\{ \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1 \right\} = \left\{ \mathbf{y} \in \mathbb{R}^{\binom{n+2d}{n}} : y_\alpha = \int_\mathcal{X} x^\alpha d\mu \quad \forall |\alpha| \leq 2d, \right.$$
$$\left. \mu = \sum_{i=1}^{\ell} w_i \delta_{x_i}, \ x_i \in \mathcal{X}, \ w \in \mathcal{W} \right\},$$

where $\ell \leq \binom{n+2d}{n}$, see the so-called Tchakaloff theorem [8, Theorem B12].

Hence, Problem (13) is equivalent to

(14)
$$\max \ \phi_q(M)$$
$$\text{s.t. } M = \sum_{|\gamma| \leq 2d} A_\gamma y_\gamma \succcurlyeq 0,$$
$$\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}), \ y_0 = 1,$$

where the maximization is now with respect to the sequence $\mathbf{y}$. Moment problem (14) is finite-dimensional and convex, yet the constraint $\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X})$ is difficult to handle. We will show that by approximating the truncated moment cone $\mathcal{M}_{2d}(\mathcal{X})$ by a nested sequence of semidefinite representable cones as indicated in (8), we obtain a hierarchy of finite dimensional semidefinite programming problems converging to the optimal solution of Problem (14). Since semidefinite programming problems can be solved efficiently, we can compute a numerical solution to Problem (13).

This describes step one of our procedure. The result of it is a sequence $\mathbf{y}^\star$ of moments. Consequently, in a second step, we need to find a representing atomic measure $\mu^\star$ of $\mathbf{y}^\star$ in order to identify the approximate optimal design $\xi^\star$.

4. The ideal problem on moments and its approximation.   For notational simplicity, let us use the standard monomial basis of $\mathbb{R}[x]_d$ for the regression functions, meaning $\mathbf{F} = (\mathbf{f}_1, \ldots, \mathbf{f}_p) := (x^\alpha)_{|\alpha| \leq d}$ with $p = \binom{n+d}{n}$. This case corresponds to $\mathfrak{A} = \text{Id}$ in (5). Note that this is not a restriction, since one can get the results for other choices of $\mathbf{F}$ by simply performing a change of basis. Indeed, in view of (10), one shall substitute $M_d(\mathbf{y})$ by $\mathfrak{A} M_d(\mathbf{y}) \mathfrak{A}^\top$ to get the statement of our results in whole generality; see Section 4.5

for a statement of the results in this case. Different polynomial bases can be considered and, for instance, one may consult the standard framework described by the book [3, Chapter 5.8].

For the sake of conciseness, we do not expose the notion of incomplete q-way m-th degree polynomial regression here but the reader may remark that the strategy developed in this paper can handle such a framework.

Before stating the main results, we recall the gradients of the Kiefer's ϕ_q criteria in Table 1.

Name	D-opt.	A-opt	E-opt.	generic case
q	0	-1	$-\infty$	$q \neq 0, -\infty$
$\phi_q(M)$	$\det(M)^{\frac{1}{p}}$	$p(\text{trace}(M^{-1}))^{-1}$	$\lambda_{\min}(M)$	$\left[\dfrac{\text{trace}(M^q)}{p}\right]^{\frac{1}{q}}$
$\nabla\phi_q(M)$	$\det(M)^{\frac{1}{p}} M^{-\frac{1}{p}}$	$p(\text{trace}(M^{-1})M)^{-2}$	$\Pi_{\min}(M)$	$\left[\dfrac{\text{trace}(M^q)}{p}\right]^{\frac{1}{q}-1} \dfrac{M^{q-1}}{p}$

<div align="center">TABLE 1</div>

Gradients of the Kiefer's ϕ_q criteria. We recall that $\Pi_{\min}(M) = uu^\top/||u||_2^2$ is defined only when the least eigenvalue of M has multiplicity one and u denotes a nonzero eigenvector associated to this least eigenvalue. If the least eigenvalue has multiplicity greater than 2, then the sub gradient $\partial\phi_q(M)$ of $\lambda_{\min}(M)$ is the set of all projectors on subspaces of the eigenspace associated to $\lambda_{\min}(M)$, see for example [13]. Notice further that ϕ_q is upper semi-continuous and is a positively homogeneous function

4.1. *The ideal problem on moments.* The ideal formulation (14) of our approximate optimal design problem reads

$$(15) \qquad \begin{aligned} \rho = \ &\max_{\mathbf{y}} \ \ \phi_q(M_d(\mathbf{y})) \\ &\text{s.t.} \quad \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}), \ y_0 = 1. \end{aligned}$$

For this we have the following standard result.

THEOREM 1 (Equivalence theorem). *Let $q \in (-\infty, 1)$ and $\mathcal{X} \subseteq \mathbb{R}^n$ be a compact semi-algebraic set as defined in (4) and with nonempty interior. Problem (15) is a convex optimization problem with a unique optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$. Denote by $p_d^\star$ the polynomial*

$$(16) \qquad x \mapsto p_d^\star(x) := \mathbf{v}_d(x)^\top M_d(\mathbf{y}^\star)^{q-1}\mathbf{v}_d(x) = ||M_d(\mathbf{y}^\star)^{\frac{q-1}{2}}\mathbf{v}_d(x)||_2^2.$$

Then $\mathbf{y}^\star$ is the vector of moments—up to order $2d$—of a discrete measure $\mu^\star$ supported on at least $\binom{n+d}{n}$ and at most $\binom{n+2d}{n}$ points in the set

$$\Omega := \left\{ x \in \mathcal{X} : \text{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) = 0 \right\},$$

In particular, the following statements are equivalent:

- $\circ$ $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ *is the unique solution to Problem* (15)*;*
- $\circ$ $\mathbf{y}^\star \in \left\{ \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1 \right\}$ *and* $p^\star := \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star \geqslant 0$ *on* $\mathcal{X}$.

PROOF. A general equivalence theorem for concave functionals of the information matrix is stated and proved in [6, Theorem 1]. The case of ϕ_q-criteria is tackled in [19] and [3, Theorem 5.4.7]. In order to be self-contained and because the proof of our Theorem 3 follows the same road map we recall a sketch of the proof in the supplement article [2], Section 1. $\square$

REMARK 1 (On the optimal dual polynomial).   *The polynomial $p_d^\star$ contains all the information concerning the optimal design. Indeed, its level set Ω supports the optimal design points. The polynomial is related to the so-called Christoffel function (see Section 4.2). For this reason, in the sequel $p_d^\star$ in* (16) *will be called a Christoffel polynomial. Notice further that*

$$\mathcal{X} \subset \left\{ p_d^\star \leq \mathrm{trace}(M_d(\mathbf{y}^\star)^q) \right\}.$$

Hence, the optimal design problem related to ϕ_q is similar to the standard problem of computational geometry consisting in minimizing the volume of a polynomial level set containing $\mathcal{X}$ (Löwner-John's ellipsoid theorem). Here, the volume functional is replaced by $\phi_q(M)$ for the polynomial $||M^{\frac{q-1}{2}}\mathbf{v}_d(x)||_2^2$. We refer to [9] for a discussion and generalizations of Löwner-John's ellipsoid theorem for general homogenous polynomials on non convex domains.

REMARK 2 (Equivalence theorem for E-optimality).   *Theorem 1 holds also for $q = -\infty$. This is the E-optimal design case, in which the objective function is not differentiable at points for which the least eigenvalue has multiplicity greater than 2. We get that $\mathbf{y}^\star$ is the vector of moments—up to order $2d$—of a discrete measure $\mu^\star$ supported on at most $\binom{n+2d}{n}$ points in the set*

$$\Omega := \left\{ x \in \mathcal{X} : \lambda_{\min}(M_d(\mathbf{y}^\star))||u||_2^2 - \left(\sum_\alpha u_\alpha x^\alpha \right)^2 = 0 \right\},$$

where $u = (u_\alpha)_{|\alpha| \leq 2d}$ is a nonzero eigenvector of $M_d(\mathbf{y}^\star)$ associated to $\lambda_{\min}(M_d(\mathbf{y}^\star))$. In particular, the following statements are equivalent

- $\circ$ $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ *is a solution to Problem* (15)*;*
- $\circ$ $\mathbf{y}^\star \in \left\{ \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1 \right\}$ *and for all $x \in \mathcal{X}$, $\left(\sum_\alpha u_\alpha x^\alpha \right)^2 \leq \lambda_{\min}(M_d(\mathbf{y}^\star))||u||_2^2$.*

Furthermore, if the least eigenvalue of $M_d(\mathbf{y}^\star)$ has multiplicity one then $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ is unique.

4.2. *Christoffel polynomials.* In the case of D-optimality, it turns out that the unique optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ of Problem (14) can be characterized in terms of the *Christoffel polynomial* of degree $2d$ associated with an optimal measure μ whose moments up to order $2d$ coincide with $\mathbf{y}^\star$. Notice that in the paradigm of optimal design the Christoffel polynomial is the variance function of the multivariate polynomial regression model. Given a design, it is the variance of the predicted value of the model and so quantifies locally the uncertainty of the estimated response. We refer to [1] for its earlier introduction and the chapter [19, Chapter 15] for an overview of its properties and uses.

DEFINITION 2 (Christoffel polynomial). *Let* $\mathbf{y} \in \mathbb{R}^{\binom{n+2d}{n}}$ *be such that* $M_d(\mathbf{y}) \succ 0$. *Then there exists a family of orthonormal polynomials* $(P_\alpha)_{|\alpha| \leq d} \subseteq \mathbb{R}[x]_d$ *satisfying*

$$L_\mathbf{y}(P_\alpha\, P_\beta) = \delta_{\alpha=\beta} \quad and \quad L_\mathbf{y}(x^\alpha\, P_\beta) = 0 \quad \forall \alpha \prec \beta,$$

where monomials are ordered with respect to the lexicographical ordering on $\mathbb{N}^n$. *We call the polynomial*

$$p_d : \ x \mapsto p_d(x) := \sum_{|\alpha| \leq d} P_\alpha(x)^2, \quad x \in \mathbb{R}^n,$$

the Christoffel polynomial (of degree d) associated with $\mathbf{y}$.

The Christoffel polynomial[1] can be expressed in different ways. For instance via the inverse of the moment matrix by

$$p_d(x) = \mathbf{v}_d(x)^\top M_d(\mathbf{y})^{-1}\mathbf{v}_d(x), \quad \forall x \in \mathbb{R}^n,$$

or via its extremal property

$$\frac{1}{p_d(t)} \ = \ \min_{P \in \mathbb{R}[x]_d} \Big\{ \int P(x)^2\, d\mu(x) \, : \, P(t) = 1 \Big\}, \qquad \forall t \in \mathbb{R}^n,$$

when $\mathbf{y}$ has a representing measure μ—when $\mathbf{y}$ does not have a representing measure μ just replace $\int P(x)^2 d\mu(x)$ with $L_\mathbf{y}(P^2)\, (= P^\top M_d(\mathbf{y})\, P)$. For more details the interested reader is referred to [11] and the references therein. Notice also that there is a regain of interest in the asymptotic study of the Christoffel function as it relies on eigenvalue marginal distributions of invariant random matrix ensembles, see for example [12].

---

[1]Actually, what is referred to the *Chistoffel function* in the literature is its reciprocal $x \mapsto 1/p_d(x)$. In optimal design, the Christofel function is also called sensitivity function or information surface [19].

REMARK 3 (Equivalence theorem for D-optimality).   *In the case of D-optimal designs, observe that*

$$t^\star := \max_{x \in \mathcal{X}} \; p_d^\star(x) = \operatorname{trace}(\operatorname{Id}) = \binom{n+d}{n},$$

where $p_d^\star$ given by (16) *for $q = 0$. Furthermore, note that $p_d^\star$ is the Christoffel polynomial of degree d of the D-optimal measure $\mu^\star$.*

4.3. *The SDP relaxation scheme.* Let $\mathcal{X} \subseteq \mathbb{R}^n$ be as defined in (4), assumed to be compact. So with no loss of generality (and possibly after scaling), assume that $x \mapsto g_1(x) = 1 - \|x\|^2 \geqslant 0$ is one of the constraints defining $\mathcal{X}$.

Since the ideal moment Problem (15) involves the moment cone $\mathcal{M}_{2d}(\mathcal{X})$ which is not SDP representable, we use the hierarchy (8) of outer approximations of the moment cone to relax Problem (15) to an SDP problem. So for a fixed integer $\delta \geq 1$ we consider the problem

$$(17) \qquad \begin{aligned} \rho_\delta = \; &\max_{\mathbf{y}} \quad \phi_q(M_d(\mathbf{y})) \\ &\text{s.t.} \quad \mathbf{y} \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}), \; y_0 = 1. \end{aligned}$$

Since Problem (17) is a relaxation of the ideal Problem (15), necessarily $\rho_\delta \geq \rho$ for all δ. In analogy with Theorem 1 we have the following result characterizing the solutions of the SDP relaxation (17) by means of Sum-of-Squares (SOS) polynomials.

THEOREM 3 (Equivalence theorem for SDP relaxations).   *Let $q \in (-\infty, 1)$ and let $\mathcal{X} \subseteq \mathbb{R}^n$ be a compact semi-algebraic set as defined in (4) and be with non-empty interior. Then,*

a) *SDP Problem* (17) *has a unique optimal solution $\mathbf{y}^\star \in \mathbb{R}^{\binom{n+2d}{n}}$.*
b) *The moment matrix $M_d(\mathbf{y}^\star)$ is positive definite. Let $p_d^\star$ be as defined in* (16)*, associated with $\mathbf{y}^\star$. Then $p^\star := \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star$ is non-negative on $\mathcal{X}$ and $L_{\mathbf{y}^\star}(p^\star) = 0$.*

In particular, the following statements are equivalent:

○ $\mathbf{y}^\star \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$ *is the unique solution to Problem* (17)*;*
○ $\mathbf{y}^\star \in \{\mathbf{y} \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}) : y_0 = 1\}$ *and $p^\star = \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star \in \mathcal{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathcal{X}).$*

PROOF. We follow the same roadmap as in the proof of Theorem 1.

a) Let us prove that Problem (17) has an optimal solution. The feasible set is nonempty with finite associated value, since we can take as feasible point the vector $\tilde{\mathbf{y}}$ associated with the Lebesgue measure on $\mathcal{X}$, scaled to be a probability measure.

Let $\mathbf{y} \in \mathbb{R}^{\binom{n+2d}{n}}$ be an arbitrary feasible solution and $\mathbf{y}_\delta \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}}$ an arbitrary lifting of $\mathbf{y}$—recall the definition of $\mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$ given in (7). Recall that $g_1(x) = 1 - \|x\|^2$. As $M_{d+\delta-1}(g_1\, y) \succeq 0$ one deduces that $L_{y_\delta}(x_i^{2t}(1 - \|x\|^2)) \geq 0$ for every $i = 1, \ldots, n$, and all $t \leq d + \delta - 1$. Expanding and using linearity of L_y yields $1 \geq \sum_{j=1}^{n} L_{y_\delta}(x_j^2) \geq L_{y_\delta}(x_i^2)$ for $t = 0$ and $i = 1, \ldots, n$. Next for $t = 1$ and $i = 1, \ldots, n$,

$$0 \leq L_{y_\delta}(x_i^2(1 - \|x\|^2)) = \underbrace{L_{y_\delta}(x_i^2)}_{\leq 1} - L_{y_\delta}(x_i^4) - \sum_{j \neq i}^{n} \underbrace{L_{y_\delta}(x_i^2 x_j^2)}_{\geq 0},$$

yields $L_{y_\delta}(x_i^4) \leq 1$. We may iterate this argumentation until we finally obtain $L_{y_\delta}(x_i^{2d+2\delta}) \leq 1$, for all $i = 1, \ldots, n$. Therefore by [10, Lemma 4.3, page 110] (or [8, Proposition 3.6, page 60]) one has

$$(18) \quad |y_{\delta,\alpha}| \leq \max\Big\{ \underbrace{y_{\delta,0}}_{=1},\ \max_i\{L_{\mathbf{y}_\delta}(x_i^{2(d+\delta)})\}\Big\} \leq 1 \ \ \forall |\alpha| \leq 2(d+\delta).$$

This implies that the set of feasible liftings $\mathbf{y}_\delta$ is compact, and therefore, the feasible set of (17) is also compact. As the function ϕ_q is upper semi-continuous, the supremum in (17) is attained at some optimal solution $\mathbf{y}^\star \in \mathbb{R}^{s(2d)}$. It is unique due to convexity of the feasible set and strict concavity of the objective function ϕ_q, e.g., see [19, Chapter 6.13] for a proof.

b) Let $\mathbf{B}_\alpha, \tilde{\mathbf{B}}_\alpha$ and $\mathbf{C}_{j\alpha}$ be real symmetric matrices such that

$$\sum_{|\alpha| \leq 2d} \mathbf{B}_\alpha x^\alpha = \mathbf{v}_d(x)\, \mathbf{v}_d(x)^\top$$

$$\sum_{|\alpha| \leq 2(d+\delta)} \tilde{\mathbf{B}}_\alpha x^\alpha = \mathbf{v}(x)_{d+\delta}\, \mathbf{v}_{d+\delta}(x)^\top$$

$$\sum_{|\alpha| \leq 2(d+\delta)} \mathbf{C}_{j\alpha} x^\alpha = g_j(x)\, \mathbf{v}_{d+\delta-v_j}(x)\, \mathbf{v}_{d+\delta-v_j}(x)^\top, \quad j = 1, \ldots, m.$$

Recall that it holds

$$\sum_{|\alpha| \leq 2d} \mathbf{B}_\alpha y_\alpha = M_d(\mathbf{y}).$$

First, we notice that there exists a strictly feasible solution to (17) because the cone $\mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X})$ has nonempty interior as a supercone of $\mathcal{M}_{2d}(\mathcal{X})$, which has nonempty interior by [9, Lemma 2.6]. Hence, Slater's condition[2] holds for (17). Further, by an argument in [19, Chapter 7.13]) the matrix $M_d(\mathbf{y}^\star)$ is non-singular. Therefore, ϕ_q is differentiable at $\mathbf{y}^\star$. Since additionally Slater's condition is fulfilled and ϕ_q is concave, this implies that the *Karush-Kuhn-Tucker* (KKT) optimality conditions[3] at $\mathbf{y}^\star$ are necessary and sufficient for $\mathbf{y}^\star$ to be an optimal solution.

The KKT-optimality conditions at $\mathbf{y}^\star$ read

$$\lambda^\star e_0 - \nabla\phi_q(M_d(\mathbf{y}^\star)) = \hat{\mathbf{p}}^\star \quad \text{with } \hat{p}^\star(x) := \langle \hat{\mathbf{p}}^\star, \mathbf{v}_{2d}(x)\rangle \in \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X}),$$

where $\hat{\mathbf{p}}^\star \in \mathbb{R}^{s(2d)}$, $e_0 = (1, 0, \ldots, 0)$, and $\lambda^\star$ is the dual variable associated with the constraint $y_0 = 1$. The complementarity condition reads $\langle \mathbf{y}^\star, \hat{\mathbf{p}}^\star\rangle = 0$.

Recalling the definition (9) of the quadratic module $\mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$, we can express the membership $\hat{p}^\star(x) \in \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$ more explicitly in terms of some *"dual variables"* $\Lambda_j \succcurlyeq 0$, $j = 0, \ldots, m$,

(19)

$$1_{\alpha=0}\,\lambda^\star - \langle \nabla\phi_q(M_d(\mathbf{y}^\star)), \mathbf{B}_\alpha\rangle = \langle \Lambda_0, \tilde{\mathbf{B}}_\alpha\rangle + \sum_{j=1}^m \langle \Lambda_j, \mathbf{C}_\alpha^j\rangle, \quad |\alpha| \leq 2(d+\delta),$$

Then, for a lifting $\mathbf{y}_\delta^\star \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}}$ of $\mathbf{y}^\star$ the complementary condition $\langle \mathbf{y}^\star, \hat{\mathbf{p}}^\star\rangle = 0$ reads

(20)

$$\langle M_{d+\delta}(\mathbf{y}_\delta^\star), \Lambda_0\rangle = 0; \quad \langle M_{d+\delta-v_j}(\mathbf{y}_\delta^\star\, g_j), \Lambda_j\rangle = 0, \quad j = 1, \ldots, m.$$

---

[2]For the optimization problem $\max\{f(x) : Ax = \mathbf{b};\ x \in \mathbf{C}\}$, where $A \in \mathbb{R}^{m \times n}$ and $\mathbf{C} \subseteq \mathbb{R}^n$ is a nonempty closed convex cone, Slater's condition holds, if there exists a feasible solution x in the interior of $\mathbf{C}$.

[3]For the optimization problem $\max\{f(x) : Ax = \mathbf{b};\ x \in \mathbf{C}\}$, where f is differentiable, $A \in \mathbb{R}^{m \times n}$ and $\mathbf{C} \subseteq \mathbb{R}^n$ is a nonempty closed convex cone, the KKT-optimality conditions at a feasible point x state that there exist $\lambda^\star \in \mathbb{R}^m$ and $\mathbf{u}^\star \in \mathbf{C}^\star$ such that $A^\top \lambda^\star - \nabla f(x) = \mathbf{u}^\star$ and $\langle x, \mathbf{u}^\star\rangle = 0$.

Multiplying by $y^\star_{\delta,\alpha}$, summing up and using the complementarity conditions (20) yields

(21)
$$\lambda^\star - \langle \nabla\phi_q(M_d(\mathbf{y}^\star)), M_d(\mathbf{y}^\star) \rangle = \underbrace{\langle \Lambda_0, M_{d+\delta}(\mathbf{y}^\star_\delta) \rangle}_{=0} + \sum_{j=1}^m \underbrace{\langle \Lambda_j, M_{d+\delta-v_j}(g_j\,\mathbf{y}^\star_\delta) \rangle}_{=0}.$$

We deduce that

(22) $$\lambda^\star = \langle \nabla\phi_q(M_d(\mathbf{y}^\star_{d,\delta})), M_d(\mathbf{y}^\star_{d,\delta}) \rangle = \phi_q(M_d(\mathbf{y}^\star_{d,\delta}))$$

by the Euler formula for homogeneous functions.

Similarly, multiplying by x^α and summing up yields

$$\lambda^\star - \mathbf{v}_d(x)^\top \nabla\phi_q(M_d(\mathbf{y}^\star))\mathbf{v}_d(x)$$
$$= \left\langle \Lambda_0, \sum_{|\alpha|\le 2(d+\delta)} \tilde{\mathbf{B}}_\alpha\, x^\alpha \right\rangle + \sum_{j=1}^m \left\langle \Lambda_j, \sum_{|\alpha|\le 2(d+\delta-v_j)} \mathbf{C}^j_\alpha\, x^\alpha \right\rangle$$
$$= \underbrace{\left\langle \Lambda_0, \mathbf{v}(x)_{d+\delta}\,\mathbf{v}_{d+\delta}(x)^\top \right\rangle}_{\sigma_0(x)} + \sum_{j=1}^m g_j(x)\, \underbrace{\left\langle \Lambda_j, \mathbf{v}_{d+\delta-v_j}(x)\,\mathbf{v}_{d+\delta-v_j}(x)^\top \right\rangle}_{\sigma_j(x)}$$
$$= \sigma_0(x) + \sum_{j=1}^n \sigma_j(x)\, g_j(x)$$

(23)
$$= \hat{p}^\star(x) \in \mathcal{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathcal{X}).$$

Note that $\sigma_0 \in \Sigma[x]_{2(d+\delta)}$ and $\sigma_j \in \Sigma[x]_{2(d+\delta-d_j)}$, $j = 1,\ldots,m$, by definition.

For $q \neq 0$ let $c^\star := \binom{n+d}{n}\left[\binom{n+d}{n}^{-1}\mathrm{trace}(M_d(\mathbf{y}^\star)^q)\right]^{1-\frac{1}{q}}$. As $M_d(\mathbf{y}^\star)$ is positive semidefinite and non-singular, we have $c^\star > 0$. If $q = 0$, let $c^\star := 1$ and replace $\phi_0(M_d(\mathbf{y}^\star))$ by $\log\det M_d(\mathbf{y}^\star)$, for which the gradient is $M_d(\mathbf{y}^\star)^{-1}$.

Using Table 1 we find that $c^\star\nabla\phi_q(M_d(\mathbf{y}^\star)) = M_d(\mathbf{y}^\star)^{q-1}$. It follows that

$$c^\star\lambda^\star \overset{(22)}{=} c^\star\langle \nabla\phi_q(M_d(\mathbf{y}^\star)), M_d(\mathbf{y}^\star) \rangle = \mathrm{trace}(M_d(\mathbf{y}^\star)^q)$$
$$\text{and}\quad c^\star\langle \nabla\phi_q(M_d(\mathbf{y}^\star)), \mathbf{v}_d(x)\mathbf{v}_d(x)^\top \rangle \overset{(16)}{=} p^\star_d(x)$$

Therefore, Eq. (23) is equivalent to $p^\star := c^\star \hat{p}^\star = c^\star \lambda^\star - p_d^\star \in \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$. To summarize,

$$p^\star(x) = \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) \in \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X}).$$

We remark that all elements of $\mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$ are non-negative on $\mathcal{X}$ and that (21) implies $L_{\mathbf{y}^\star}(p^\star) = 0$. Hence, we have shown b).

The equivalence follows from the argumentation in b). $\square$

REMARK 4 (Finite convergence). *If the optimal solution $\mathbf{y}^\star$ of Problem* (17) *is coming from a measure $\mu^\star$ on $\mathcal{X}$, that is $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$, then $\rho_\delta = \rho$ and $\mathbf{y}^\star$ is the unique optimal solution of Problem* (15). *In addition, by the proof of Theorem* 1, *$\mu^\star$ can be chosen to be atomic and supported on at least $\binom{n+d}{n}$ and at most $\binom{n+2d}{n}$ "contact points" on the level set $\Omega := \{x \in \mathcal{X} : \mathrm{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) = 0\}$.*

REMARK 5 (SDP relaxation for E-optimality). *Theorem* 3 *holds also for $q = -\infty$. This is the E-optimal design case, in which the objective function is not differentiable at points for which the least eigenvalue has multiplicity greater than 2. We get that $\mathbf{y}^\star$ satisfies $\lambda_{\min}(M_d(\mathbf{y}^\star)) - \left(\sum_\alpha u_\alpha x^\alpha\right)^2 \geqslant 0$ for all $x \in \mathcal{X}$ and $L_{\mathbf{y}^\star}(\left(\sum_\alpha u_\alpha x^\alpha\right)^2) = \lambda_{\min}(M_d(\mathbf{y}^\star))$, where $u = (u_\alpha)_{|\alpha|\leq 2d}$ is a nonzero eigenvector of $M_d(\mathbf{y}^\star)$ associated to $\lambda_{\min}(M_d(\mathbf{y}^\star))$.*

In particular, the following statements are equivalent

- *$\mathbf{y}^\star \in \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X})$ is a solution to Problem* (17);
- *$\mathbf{y}^\star \in \{\mathbf{y} \in \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X}) : y_0 = 1\}$ and $p^\star(x) = \lambda_{\min}(M_d(\mathbf{y}^\star))||u||_2^2 - \left(\sum_\alpha u_\alpha x^\alpha\right)^2 \in \mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$.*

Furthermore, if the least eigenvalue of $M_d(\mathbf{y}^\star)$ has multiplicity one then $\mathbf{y}^\star$ is unique.

4.4. *Asymptotics.* We now analyze what happens when δ tends to infinity.

THEOREM 4. *Let $q \in (-\infty, 1)$ and $d \in \mathbb{N}$. For every $\delta = 0, 1, 2, \ldots$, let $\mathbf{y}_{d,\delta}^\star$ be an optimal solution to* (17) *and $p_{d,\delta}^\star \in \mathbb{R}[x]_{2d}$ the Christoffel polynomial associated with $\mathbf{y}_{d,\delta}^\star$ defined in Theorem* 3. *Then,*

- *a) $\rho_\delta \to \rho$ as $\delta \to \infty$, where ρ is the supremum in* (15).
- *b) For every $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq 2d$, we have $\lim_{\delta\to\infty} y_{d,\delta,\alpha}^\star = y_\alpha^\star$, where $\mathbf{y}^\star = (y_\alpha^\star)_{|\alpha|\leq 2d} \in \mathcal{M}_{2d}(\mathcal{X})$ is the unique optimal solution to* (15).
- *c) $p_{d,\delta}^\star \to p_d^\star$ as $\delta \to \infty$, where $p_d^\star$ is the Christoffel polynomial associated with $\mathbf{y}^\star$ defined in* (16).

  d) *If the dual polynomial $p^\star := \text{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star$ to Problem (15) belongs to $\mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X})$ for some δ, then finite convergence takes place, that is, $\mathbf{y}_{d,\delta}^\star$ is the unique optimal solution to Problem (15) and $\mathbf{y}_{d,\delta}^\star$ has a representing measure, namely the target measure $\mu^\star$.*

PROOF. We prove the four claims consecutively.

  a) For every δ complete the lifted finite sequence $\mathbf{y}_\delta^\star \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}}$ with zeros to make it an infinite sequence $\mathbf{y}_\delta^\star = (y_{\delta,\alpha}^\star)_{\alpha \in \mathbb{N}^n}$. Therefore, every such $\mathbf{y}_\delta^\star$ can be identified with an element of ℓ_∞, the Banach space of finite bounded sequences equipped with the supremum norm. Moreover, Inequality (18) holds for every $\mathbf{y}_\delta^\star$. Thus, denoting by $\mathcal{B}$ the unit ball of ℓ_∞ which is compact in the $\sigma(\ell_\infty, \ell_1)$ weak-$\star$ topology on ℓ_∞, we have $\mathbf{y}_\delta^\star \in \mathcal{B}$. By Banach-Alaoglu's theorem, there is an element $\hat{\mathbf{y}} \in \mathcal{B}$ and a converging subsequence $(\delta_k)_{k \in \mathbb{N}}$ such that

(24)
$$\lim_{k \to \infty} y_{\delta_k,\alpha}^\star = \hat{y}_\alpha \qquad \forall \alpha \in \mathbb{N}^n.$$

  Let $s \in \mathbb{N}$ be arbitrary, but fixed. By the convergence (24) we also have

$$\lim_{k \to \infty} M_s(\mathbf{y}_{\delta_k}^\star) = M_s(\hat{\mathbf{y}}) \succcurlyeq 0;$$
$$\lim_{k \to \infty} M_s(g_j\,\mathbf{y}_{\delta_k}^\star) = M_s(g_j\,\hat{\mathbf{y}}) \succcurlyeq 0,\ j = 1,\ldots,m.$$

  Notice that the subvectors $\mathbf{y}_{d,\delta}^\star = (y_{\delta,\alpha}^\star)_{|\alpha| \leq 2d}$ with $\delta = 0,1,2,\ldots$ belong to a compact set. Therefore, since $\phi_q(M_d(\mathbf{y}_{d,\delta}^\star)) < \infty$ for every δ, we also have $\phi_q(M_d(\hat{\mathbf{y}})) < \infty$.
  Next, by Putinar's theorem [8, Theorem 3.8], $\hat{\mathbf{y}}$ is the sequence of moments of some measure $\hat{\mu} \in \mathcal{M}_+(\mathcal{X})$, and so $\hat{\mathbf{y}}_d = (\hat{y}_\alpha)_{|\alpha| \leq 2d}$ is a feasible solution to (15), meaning $\rho \geq \phi_q(M_d(\hat{\mathbf{y}}_d))$. On the other hand, as (17) is a relaxation of (15), we have $\rho \leq \rho_{\delta_k}$ for all δ_k. So the convergence (24) yields

$$\rho \leq \lim_{k \to \infty} \rho_{\delta_k} = \phi_q(M_d(\hat{\mathbf{y}}_d)),$$

  which proves that $\hat{\mathbf{y}}$ is an optimal solution to (15), and $\lim_{\delta \to \infty} \rho_\delta = \rho$.
  b) As the optimal solution to (15) is unique, we have $\mathbf{y}^\star = \hat{\mathbf{y}}_d$ with $\hat{\mathbf{y}}_d$ defined in the proof of a) and the whole sequence $(\mathbf{y}_{d,\delta}^\star)_{\delta \in \mathbb{N}}$ converges to $\mathbf{y}^\star$, that is, for $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq 2d$ fixed

(25)
$$\lim_{d,\delta \to \infty} y_{\delta,\alpha}^\star = \lim_{\delta \to \infty} y_{\delta,\alpha}^\star = \hat{y}_\alpha = y_\alpha^\star.$$

c) It suffices to observe that the coefficients of Christoffel polynomial $p_{d,\delta}^\star$ are continuous functions of the moments $(y_{d,\delta,\alpha}^\star)_{|\alpha|\leq 2d} = (y_{\delta,\alpha}^\star)_{|\alpha|\leq 2d}$. Therefore, by the convergence (25) one has $p_{d,\delta}^\star \to p_d^\star$ where $p_d^\star \in \mathbb{R}[x]_{2d}$ as in Theorem 1.

The last point follows directly observing that, in this case, the two Programs (15) and (17) satisfy the same KKT conditions. $\qquad\square$

4.5. *General regression polynomial bases.* We return to the general case described by a matrix $\mathfrak{A}$ of size $p \times \binom{n+d}{n}$ such that the regression polynomials satisfy $\mathbf{F}(x) = \mathfrak{A}\,\mathbf{v}_d(x)$ for all $x \in \mathcal{X}$. Without loss of generality, we can assume that the rank of $\mathfrak{A}$ is p, *i.e.*, the regressors $\mathbf{f}_1, \ldots, \mathbf{f}_p$ are linearly independent. Now, the objective function becomes $\phi_q(\mathfrak{A}M_d(\mathbf{y})\mathfrak{A}^\top)$ at point $\mathbf{y}$. Note that the constraints on $\mathbf{y}$ are unchanged, i.e.,

- $\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X})$, $y_0 = 1$ in the ideal problem,
- $\mathbf{y} \in \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X})$, $y_0 = 1$ in the SDP relaxation scheme.

We recall the notation $\mathbf{M}_d(\mathbf{y}) := \mathfrak{A}M_d(\mathbf{y})\mathfrak{A}^\top$ and we get that the KKT conditions are given by

$$\forall x \in \mathcal{X}, \quad \phi_q(\mathbf{M}_d(\mathbf{y})) - \underbrace{\mathbf{F}(x)^\top \nabla\phi_q(\mathbf{M}_d(\mathbf{y}))\,\mathbf{F}(x)}_{\text{proportional to } p_d^\star(x)} = p^\star(x)$$

where

- $p^\star \in \mathcal{M}_{2d}(\mathcal{X})^\star\,(=\mathcal{P}_{2d}(\mathcal{X}))$ in the ideal problem,
- $p^\star \in \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}}(\mathcal{X})^\star\,(=\mathcal{P}_{2(d+\delta)}^{\mathsf{SOS}}(\mathcal{X}))$ in the SDP relaxation scheme.

Our analysis leads to the following equivalence results in this case.

PROPOSITION 5. *Let $q \in (-\infty, 1)$ and let $\mathcal{X} \subseteq \mathbb{R}^n$ be a compact semialgebraic set as defined in* (4) *and with nonempty interior. Problem* (13) *is a convex optimization problem with an optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$. Denote by $p_d^\star$ the polynomial*

$$(26) \qquad x \mapsto p_d^\star(x) := \mathbf{F}(x)^\top \mathbf{M}_d(\mathbf{y})^{q-1}\mathbf{F}(x) = ||\mathbf{M}_d(\mathbf{y})^{\frac{q-1}{2}}\mathbf{F}(x)||_2^2.$$

Then $\mathbf{y}^\star$ is the vector of moments—up to order $2d$—of a discrete measure $\mu^\star$ supported on at least p points and at most $\overline{s}$ points where

$$\overline{s} \leq \min\left[1 + \frac{p(p+1)}{2}, \binom{n+2d}{n}\right]$$

(see [2, Section 2]*) in the set $\Omega := \{x \in \mathcal{X} : \text{trace}(\mathbf{M}_d(\mathbf{y})^q) - p_d^\star(x) = 0\}$.*

In particular, the following statements are equivalent:

○ $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ *is the solution to Problem* (15)*;*
○ $\mathbf{y}^\star \in \{\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1\}$ *and* $p^\star := \mathrm{trace}(\mathbf{M}_d(\mathbf{y})^q) - p_d^\star(x) \geqslant 0$
 on $\mathcal{X}$.

Furthermore, if $\mathfrak{A}$ *has full column rank then* $\mathbf{y}^\star$ *is unique.*

The SDP relaxation is given by the program

$$(27) \qquad \rho_\delta = \max_{\mathbf{y}} \phi_q(\mathbf{M}_d(\mathbf{y})) \text{ s.t.} \mathbf{y} \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}),\ y_0 = 1,$$

for which it is possible to prove the following result.

PROPOSITION 6.   *Let* $q \in (-\infty, 1)$ *and let* $\mathcal{X} \subseteq \mathbb{R}^n$ *be a compact semi-algebraic set as defined in* (4) *and with nonempty interior. Then,*

a) *SDP Problem* (27) *has an optimal solution* $\mathbf{y}^\star_{d,\delta} \in \mathbb{R}^{\binom{n+2d}{n}}$.
b) *Let* $p_d^\star$ *be as defined in* (26)*, associated with* $\mathbf{y}^\star$*. Then* $p^\star :=$
 $\mathrm{trace}(\mathbf{M}_d(\mathbf{y}^\star_{d,\delta})^q) - p_d^\star(x) \geqslant 0$ *on* $\mathcal{X}$ *and* $L_{\mathbf{y}^\star_{d,\delta}}(p^\star) = 0$.

In particular, the following statements are equivalent:

○ $\mathbf{y}^\star \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$ *is a solution to Problem* (17)*;*
○ $\mathbf{y}^\star \in \{\mathbf{y} \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}) : y_0 = 1\}$ *and* $p^\star = \mathrm{trace}(\mathbf{M}_d(\mathbf{y}^\star)^q) -$
 $p_d^\star \in \mathcal{P}^{\mathsf{SOS}}_{2(d+\delta)}(\mathcal{X})$.

Furthermore, if $\mathfrak{A}$ *has full column rank then* $\mathbf{y}^\star$ *is unique.*

5. Recovering the measure. By solving step one as explained in Section 4, we obtain a solution $\mathbf{y}^\star$ of SDP Problem (17). As $\mathbf{y}^\star \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$, it is likely that it comes from a measure. If this is the case, by Tchakaloff's theorem, there exists an atomic measure supported on at most $s(2d)$ points having these moments. For computing the atomic measure, we propose two approaches: A first one which follows a procedure by Nie [17], and a second one which uses properties of the Christoffel polynomial associated with $\mathbf{y}^\star$.

These approaches have the benefit that they can numerically certify finite convergence of the hierarchy.

5.1. *Via Nie's method.* This approach to recover a measure from its moments is based on a formulation proposed by Nie in [17]. Let $\mathbf{y}^\star =$

$(y_\alpha^\star)_{|\alpha|\le 2d}$ a finite sequence of moments. For $r \in \mathbb{N}$ consider the SDP problem

$$
(28) \quad
\begin{aligned}
\min_{\mathbf{y}_r} \quad & L_{\mathbf{y}_r}(f_r) \\
\text{s.t.} \quad & M_{d+r}(\mathbf{y}_r) \succcurlyeq 0, \\
& M_{d+r-v_j}(g_j\, \mathbf{y}_r) \succcurlyeq 0, \quad j = 1, \dots, m, \\
& y_{r,\alpha} = y_\alpha^\star, \quad \forall \alpha \in \mathbb{N}^n,\ |\alpha| \le 2d,
\end{aligned}
$$

where $\mathbf{y}_r \in \mathbb{R}^{\binom{n+2(d+r)}{n}}$ and $f_r \in \mathbb{R}[x]_{2(d+r)}$ is a randomly generated polynomial strictly positive on $\mathcal{X}$, and again $v_j = \lceil d_j/2 \rceil$, $j = 1, \dots, m$. We check whether the optimal solution $\mathbf{y}_r^\star$ of (28) satisfies the rank condition

$$
(29) \qquad \operatorname{rank} M_{d+r}(\mathbf{y}_r^\star) = \operatorname{rank} M_{d+r-v}(\mathbf{y}_r^\star),
$$

where $v := \max_j v_j$. Indeed if (29) holds then $\mathbf{y}_r^\star$ is the sequence of moments (up to order $2r$) of a measure supported on $\mathcal{X}$; see [8, Theorem 3.11, p. 66]. If the test is passed, then we stop, otherwise we increase r by one and repeat the procedure. As $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$, the rank condition (29) is satisfied for a sufficiently large value of r.

We extract the support points $x_1, \dots, x_\ell \in \mathcal{X}$ of the representing atomic measure of $\mathbf{y}_r^\star$, and $\mathbf{y}^\star$ respectively, as described in [8, Section 4.3].

Experience reveals that in most cases it is enough to use the following polynomial

$$
x \mapsto f_r(x) = \sum_{|\alpha|\le d+r} x^{2\alpha} = ||\mathbf{v}_{d+r}(x)||_2^2
$$

instead of using a random positive polynomial on $\mathcal{X}$. In Problem (28) this corresponds to minimizing the trace of $M_{d+r}(\mathbf{y})$—and so induces an optimal solution $\mathbf{y}$ with low rank matrix $M_{d+r}(\mathbf{y})$.

5.2. *Via the Christoffel polynomial.* Another possibility to recover the atomic representing measure of $\mathbf{y}^\star$ is to find the zeros of the polynomial $p^\star(x) = \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x)$, where $p_d^\star$ is the Christoffel polynomial associated with $\mathbf{y}^\star$ defined in (16), that is, $p_d^\star(x) = \mathbf{v}_d(x)^\top M_d(\mathbf{y}^\star)^{q-1} \mathbf{v}_d(x)$. In other words, we compute the set $\Omega = \{x \in \mathcal{X} : \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) = 0\}$, which due to Theorem 3 is the support of the atomic representing measure.

To that end we minimize $p^\star$ on $\mathcal{X}$. As the polynomial $p^\star$ is non-negative on $\mathcal{X}$, the minimizers are exactly Ω. For minimizing $p^\star$, we use the Lasserre hierarchy of lower bounds, that is, we solve the semidefinite program

$$
(30) \quad
\begin{aligned}
\min_{\mathbf{y}_r} \quad & L_{\mathbf{y}_r}(p^\star) \\
\text{s.t.} \quad & M_{d+r}(\mathbf{y}_r) \succcurlyeq 0, \quad y_{r,0} = 1, \\
& M_{d+r-v_j}(g_j\, \mathbf{y}_r) \succcurlyeq 0, \quad j = 1, \dots, m,
\end{aligned}
$$

where $\mathbf{y}_r \in \mathbb{R}^{\binom{n+2(d+r)}{n}}$.

Since $p_d^\star$ is associated with the optimal solution to (17) for some given $\delta \in \mathbb{N}$, by Theorem 3, it satisfies the Putinar certificate (23) of positivity on $\mathcal{X}$. Thus, the value of Problem (30) is zero for all $r \geqslant \delta$. Therefore, for every feasible solution $\mathbf{y}_r$ of (30) one has $L_{\mathbf{y}_r}(p^\star) \geq 0$ (and $L_{\mathbf{y}_d^\star}(p^\star) = 0$ for $\mathbf{y}_d^\star$ an optimal solution of (17)).

When condition (29) is fulfilled, the optimal solution $\mathbf{y}_r^\star$ comes from a measure. We extract the support points $x_1, \ldots, x_\ell \in \mathcal{X}$ of the representing atomic measure of $\mathbf{y}_r^\star$, and $\mathbf{y}^\star$ respectively, as described in [8, Section 4.3].

Alternatively, we can solve the SDP

$$
(31) \quad
\begin{aligned}
\min_{\mathbf{y}_r} \quad & \mathrm{trace}(M_{d+r}(\mathbf{y}_r)) \\
\text{s.t.} \quad & L_{\mathbf{y}_r}(p^\star) = 0, \\
& M_{d+r}(\mathbf{y}_r) \succcurlyeq 0, \quad y_{r,0} = 1, \\
& M_{d+r-v_j}(g_j \, \mathbf{y}_r) \succcurlyeq 0, \quad j = 1, \ldots, m,
\end{aligned}
$$

where $\mathbf{y}_r \in \mathbb{R}^{\binom{n+2(d+r)}{n}}$. This problem also searches for a moment sequence of a measure supported on the zero level set of $p^\star$. Again, if condition (29) is holds, the finite support can be extracted.

5.3. *Calculating the corresponding weights.*   After recovering the support $\{x_1, \ldots, x_\ell\}$ of the atomic representing measure by one of the previously presented methods, we might be interested in also computing the corresponding weights $\omega_1, \ldots, \omega_\ell$. These can be calculated easily by solving the following linear system of equations: $\sum_{i=1}^{\ell} \omega_i x_i^\alpha = y_\alpha^\star$ for all $|\alpha| \leq 2d$, *i.e.,* $\int_{\mathcal{X}} x^\alpha \mu^\star(dx) = y_\alpha^\star$.

6. Examples.   We illustrate the procedure on six examples: a univariate one, four examples in the plane and one example on the three-dimensional sphere. We concentrate on D-optimal designs, namely $q = 0$.

All examples are modeled by GloptiPoly 3 [5] and YALMIP [14] and solved by MOSEK 7 [16] or SeDuMi under the MATLAB R2014a environment. We ran the experiments on an HP EliteBook with 16-GB RAM memory and an Intel Core i5-4300U processor. We do not report computation times, since they are negligible for our small examples.

6.1. *Univariate unit interval.*   We consider as design space the interval $\mathcal{X} = [-1, 1]$ and on it the polynomial measurements $\sum_{j=0}^{d} \theta_j x^j$ with unknown parameters $\theta \in \mathbb{R}^{d+1}$. To compute the D-optimal design we first solve

Problem (17), in other words

$$
\begin{aligned}
(32) \quad \max_{\mathbf{y}_\delta} \quad & \log \det M_d(\mathbf{y}_\delta) \\
\text{s.t.} \quad & M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0, \\
& M_{d+\delta-1}((1 - \|x\|^2)\,\mathbf{y}_\delta) \succcurlyeq 0, \\
& y_{\delta,0} = 1
\end{aligned}
$$

for $\mathbf{y}_\delta \in \mathbb{R}^{s(2(d+\delta))}$ and given regression order d and relaxation order $d + \delta$, and then taking the truncation $\mathbf{y}^\star := (y^\star_{\delta,\alpha})_{|\alpha| \leqslant 2d}$ of an optimal solution $\mathbf{y}^\star_\delta$. For instance, for $d = 5$ and $\delta = 0$ we obtain the sequence $\mathbf{y}^\star \approx (1,\, 0,\, 0.56,\, 0,\, 0.45,\, 0,\, 0.40,\, 0,\, 0.37,\, 0,\, 0.36)^\top$.

Then, to recover the corresponding atomic measure from the sequence $\mathbf{y}^\star$ we solve the problem

$$
\begin{aligned}
(33) \quad \min_{\mathbf{y}} \quad & \operatorname{trace} M_{d+r}(\mathbf{y}_r) \\
\text{s.t.} \quad & M_{d+r}(\mathbf{y}_r) \succcurlyeq 0 \\
& M_{d+r-1}((1 - x^2)\mathbf{y}_r) \succcurlyeq 0, \\
& y_\alpha = y^\star_{r,\alpha}, \quad |\alpha| \leq 2d,
\end{aligned}
$$

and find the points -1, -0.765, -0.285, 0.285, 0.765 and 1 (for $d = 5$, $\delta{=}0$, $r = 1$). As a result, our optimal design is the weighted sum of the Dirac measures supported on these points. The points match with the known analytic solution to the problem, which are the critical points of the Legendre polynomial, see *e.g.*, [3, Theorem 5.5.3, p.162]. In this case, we know explicitly the optimal design, its support is located at the roots of the polynomial $t \to (1 - t^2)P'_d(t)$ where P'_d denotes the derivative of the Legendre polynomial of degree d, and its weights are all equal to $1/(1 + d)$. Now, observe that the roots of $p^\star$ have degree 2 in the interior of $[-1, 1]$ (there are $d - 1$ roots corresponding exactly to the roots of P'_d) and degree 1 on the edges (corresponding exactly to the roots of $(1 - t^2)$). Observe also that $p^\star$ has degree $2d$. We deduce that $p^\star$ equals $t \to (1 - t^2)(P'_d(t))^2$ up to a multiplicative constant. Calculating the corresponding weights as described in Section 5.3, we find $\omega_1 = \cdots = \omega_6 \approx 0.166$ as prescribed by the theory.

Alternatively, we compute the roots of the polynomial $x \mapsto p^\star(x) = 6 - p^\star_5(x)$, where $p^\star_5$ is the Christoffel polynomial of degree $2d = 10$ on $\mathcal{X}$ and find the same points as in the previous approach by solving Problem (31). See Figure 1 for the graph of the Christoffel polynomial of degree 10.

We observe that we get less points when using Problem (30) to recover the support for this example. This may occur due to numerical issues.
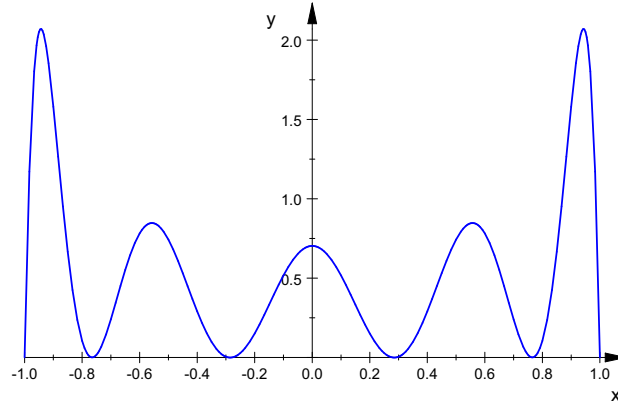
FIGURE 1. *Polynomial $p^\star$ for Example 6.1.*

6.2. *Wynn's polygon.* As a first two-dimensional example we take the polygon given by the vertices $(-1,-1)$, $(-1,1)$, $(1,-1)$ and $(2,2)$, scaled to fit the unit circle, *i.e.,* we consider the design space

$$\mathcal{X} = \{x \in \mathbb{R}^2 : x_1, x_2 \geqslant -\tfrac{1}{4}\sqrt{2},\ x_1 \leq \tfrac{1}{3}(x_2+\sqrt{2}),\ x_2 \leq \tfrac{1}{3}(x_1+\sqrt{2}),\ x_1^2+x_2^2 \leq 1\}.$$

Note that we need the redundant constraint $x_1^2 + x_2^2 \leq 1$ in order to have an algebraic certificate of compactness.

As before, in order to find the D-optimal measure for the regression, we solve Problems (17) and (28). Let us start by analyzing the results for $d = 1$ and $\delta = 3$. Solving (17) we obtain $\mathbf{y}^\star \in \mathbb{R}^{45}$ which leads to 4 atoms when solving (28) with $r = 3$. For the latter the moment matrices of order 2 and 3 both have rank 4, so Condition (29) is fulfilled. As expected, the 4 atoms are exactly the vertices of the polygon.

Again, we could also solve Problem (31) instead of (28) to receive the same atoms. As in the univariate example we get less points when using Problem (30). To be precise, GloptiPoly is not able to extract any solutions for this example.

For increasing d, we get an optimal measure with a larger support. For $d = 2$ we recover 7 points, and 13 for $d = 3$. See Figure 2 for the polygon, the supporting points of the optimal measure and the $\binom{2+d}{2}$-level set of the Christoffel polynomial $p_d^\star$ for different d. The latter demonstrates graphically that the set of zeros of $\binom{2+d}{d} - p_d^\star$ intersected with $\mathcal{X}$ are indeed the atoms of our representing measure. In the picture the size of the support points is chosen with respect to their corresponding weights, i.e., the larger the point, the bigger the respective weight.

The numerical values of the support points and their weights computed in the above procedure (and displayed in Figure 2) are listed in the supplement article [2, Table 1].

To get an idea of how the Christoffel polynomial looks like, we plot in Figure 3 the 3D-plot of the polynomial $-p^\star = p_d^\star - \binom{2+d}{2}$. This illustrates very clearly that the zeros of $p^\star$ on $\mathcal{X}$ are the support points of the optimal design.
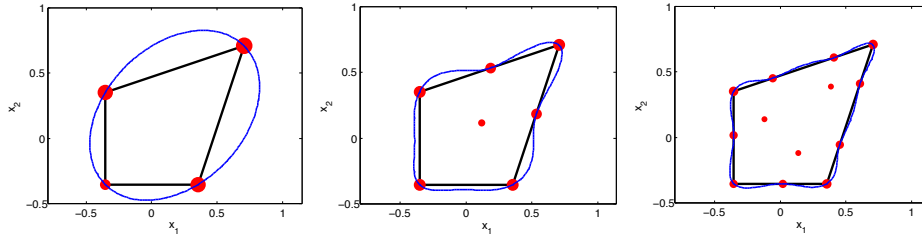

Figure 2. *The polygon (bold black) of Example 6.2, the support of the optimal design measure (red points) where the size of the points corresponds to the respective weights, and the $\binom{2+d}{2}$-level set of the Christoffel polynomial (thin blue) for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.*

6.3. *Ring of ellipses.* As a second example in the plane we consider an ellipsoidal ring, i.e., an ellipse with a hole in the form of a smaller ellipse. More precisely,

$$\mathcal{X} = \{x \in \mathbb{R}^2 : 9x_1^2 + 13x_2^2 \leq 7.3, \ 5x_1^2 + 13x_2^2 \geq 2\}.$$

We follow the same procedure as described in the former example. See Figure 4 for the results. The values are again listed in the supplement article [2, Table 1].

6.4. *Moon.* To investigate another non-convex example, we apply our method to the moon-shaped semi-algebraic set

$$\mathcal{X} = \{x \in \mathbb{R}^2 : (x_1 + 0.2)^2 + x_2^2 \leq 0.36, \ (x_1 - 0.6)^2 + x_2^2 \geq 0.16\}.$$

The results are represented in Figure 5 and for the numerical values the interested reader is referred to the supplement article [2, Table 1].

6.5. *Folium.* The zero level set of the polynomial $f(x) = -x_1(x_1^2 - 2x_2^2)$-$(x_1^2 + x_2^2)^2$ is a curve of genus zero with a triple singular point at the origin.
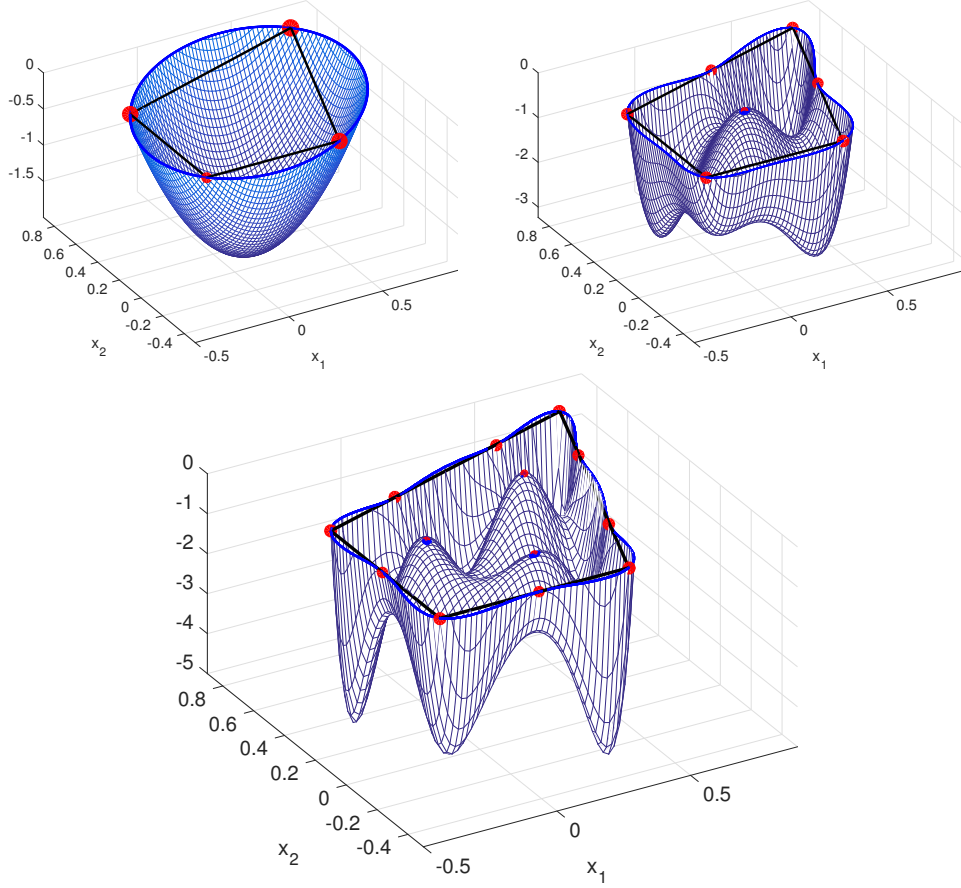
FIGURE 3. *The polynomial* $p_d^\star - \binom{2+d}{2}$ *where* $p_d^\star$ *denotes the Christoffel polynomial of Example 6.2 for* $d = 1$ *(top left),* $d = 2$ *(top right),* $d = 3$ *(bottom middle). The red points correspond to the* $\binom{2+d}{2}$-*level set of the Christoffel polynomial.*

It is called a folium. As a last two-dimensional example we consider the semi-algebraic set defined by f, i.e.,

$$\mathcal{X} = \{x \in \mathbb{R}^2 : f(x) \geq 0, \ x_1^2 + x_2^2 \leq 1\}.$$

Figure 6 illustrates the results and the values are listed in the supplement article [2, Table 1].

6.6. *The 3-dimensional unit sphere.* Last, let us consider the regression for the degree d polynomial measurements $\sum_{|\alpha| \leq d} \theta_\alpha x^\alpha$ on the unit sphere $\mathcal{X} = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$. Again, we first solve Problem (17). For
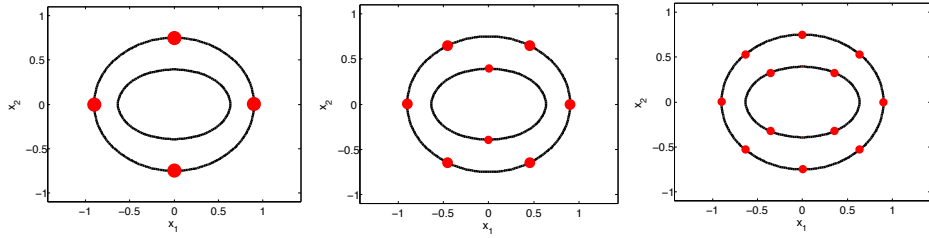
FIGURE 4. *The polygon (bold black) of Example 6.3 and the support of the optimal design measure (red points) where the size of the points corresponds to the respective weights for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.*
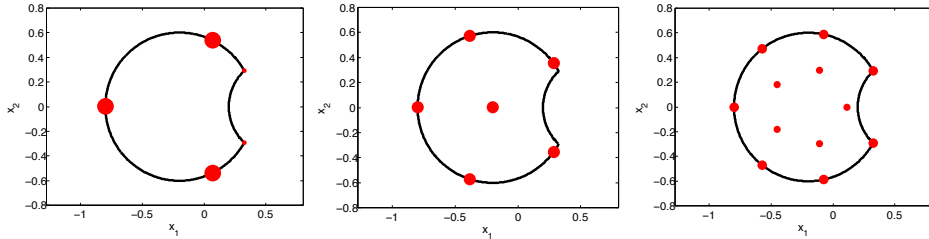

FIGURE 5. *The polygon (bold black) of Example 6.4 and the support of the optimal design measure (red points) where the size of the points corresponds to the respective weights for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.*

$d = 1$ and $\delta \geq 0$ we obtain the sequence $\mathbf{y}^\star \in \mathbb{R}^{10}$ with $y_{000}^\star = 1$, $y_{200}^\star = y_{020}^\star = y_{002}^\star = 0.333$ and all other entries zero.

In the second step we solve Problem (28) to recover the measure. For $r = 2$ the moment matrices of order 2 and 3 both have rank 6, meaning the rank condition (29) is fulfilled, and we obtain the six atoms $\{(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)\} \subseteq \mathcal{X}$ on which the optimal measure $\mu \in \mathscr{M}_+(\mathcal{X})$ is uniformly supported.

For quadratic regressions, *i.e.,* $d = 2$, we obtain an optimal measure supported on 14 atoms evenly distributed on the sphere. For $d = 3$, meaning cubic regressions, we find a Dirac measure supported on 26 points which again are evenly distributed on the sphere. See Figure 7 for a display of the supporting points of the optimal measures for $d = 1$, $d = 2$, $d = 3$ and $\delta = 0$.

Using the method via Christoffel polynomials gives again less points. No solution is extracted when solving Problem (31) and we find only two supporting points for Problem (30).
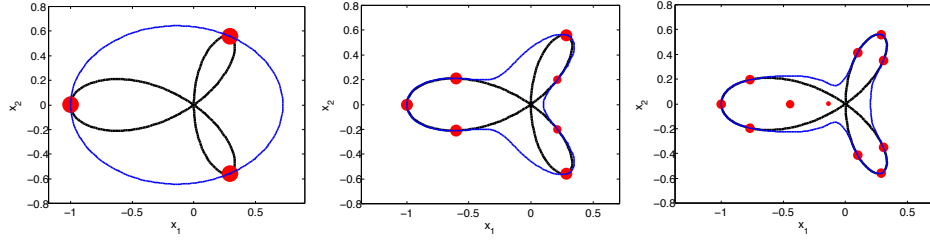
FIGURE 6. *The polygon (bold black) of Example 6.5, the support of the optimal design measure (red points) where the size of the points corresponds to the respective weights, and the $\binom{2+d}{2}$-level set of the Christoffel polynomial (thin blue) for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.*
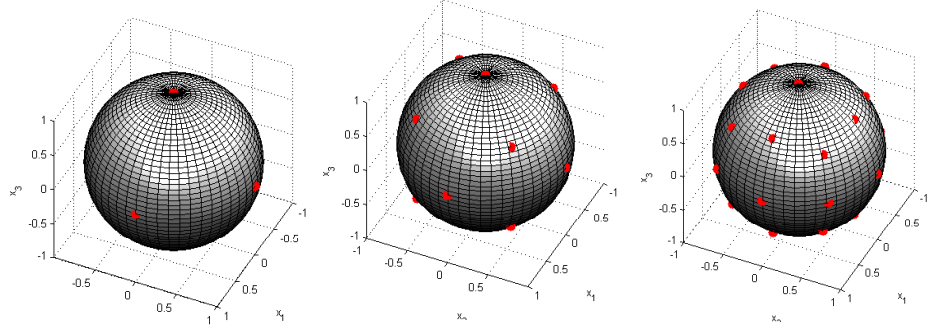

FIGURE 7. *The red points illustrate the support of the optimal design measure for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 0$ for Example 6.6.*

6.7. *Fixing some moments.* Our method has an additional nice feature. Indeed in Problem (17) one may easily include the additional constraint that some moments (y_α), $\alpha \in \Gamma \subset \mathbb{N}^n_{2d}$ are fixed to some prescribed value. We illustrate this potential on one example. For instance, with $\Gamma = \{(020), (002), (110), (101)\}$, let $y_{020} := 2$, $y_{002} := 1$, $y_{110} := 0.01$ and $y_{101} := 0.95$. In order to obtain a feasible problem, we scale them with respect to the Gauss distribution.

For the D-optimal design case with $d = 1$ and $\delta = 0$ and after computing the support of the corresponding measure using the Nie method, we get 6 points as we obtain without fixing the moments. However, now four of the six points are shifted and the measure is no longer uniformly supported on these points, but each two opposite points have the same weight. See Figure 8 for an illustration of the position of the points with fixed moments (blue) with respect to the position of the support points without fixing the points (red).
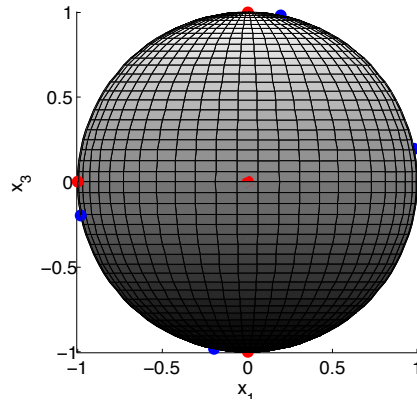
FIGURE 8. *Support points recovered in Example 6.6 for the D-optimal design and $d = 1$ (red) and the points which are recovered when additionally fixing some moments as described in Subsection 6.7 (blue).*

7. Conclusion.   In this paper, we give a general method to build optimal designs for multidimensional polynomial regression on an algebraic manifold. The method is highly versatile as it can be used for all classical functionals of the information matrix. Furthermore, it can easily be tailored to incorporate prior knowledge on some multidimensional moments of the targeted optimal measure (as proposed in [15]). In future works, we will extend the method to multi-response polynomial regression problems and to general smooth parametric regression models by linearization.

<div align="center">SUPPLEMENTARY MATERIAL</div>

Supplementary Material of Approximate Optimal Designs for Multivariate Polynomial Regression
(doi: COMPLETED BY THE TYPESETTER). We provide the proof of Theorem 1 and the detailed numerical results of the numerical examples in supplement article [2].

REFERENCES

[1] G. E. Box and J. S. Hunter. Multi-factor experimental designs for exploring response surfaces. *The Annals of Mathematical Statistics*, pages 195–241, 1957.

[2] Y. De Castro, F. Gamboa, D. Henrion, R. Hess, and J.-B. Lasserre. Supplement to "Approximate Optimal Designs for Multivariate Polynomial Regression". *Annals of Statistics*, 2018.

[3] H. Dette and W. J. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*, volume 338. John Wiley & Sons, 1997.

[4] N. Gaffke, U. Graßhoff, and R. Schwabe. Algorithms for approximate linear regression design with application to a first order model with heteroscedasticity. *Computational Statistics & Data Analysis*, 71:1113–1123, 2014.

[5] D. Henrion, J.-B. Lasserre, and J. Löfberg. Gloptipoly 3: moments, optimization and semidefinite programming. *Optimization Methods & Software*, 24(4-5):761–779, 2009.

[6] J. Kiefer. General equivalence theory for optimum designs (approximate theory). *The annals of Statistics*, pages 849–879, 1974.

[7] M. Krein and A. Nudelman. *The Markov moment problem and extremal problems*, volume 50 of *Translations of mathematical monographs*. American Mathematical Society, Providence, Rhode Island, 1977.

[8] J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2010.

[9] J.-B. Lasserre. A generalization of Löwner-John's ellipsoid theorem. *Mathematical Programming*, 152(1-2):559–591, 2015.

[10] J. B. Lasserre and T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. *Mathematische Zeitschrift*, 256(1):99–112, 2007.

[11] J.-B. Lasserre and É. Pauwels. Sorting out typicality with the inverse moment matrix SOS polynomial. In *Advances in Neural Information Processing Systems 29*, 2016.

[12] M. Ledoux. Differential operators and spectral distributions of invariant ensembles from the classical orthogonal polynomials. the continuous case. *Electron. J. Probab*, 9(7):177–208, 2004.

[13] A. S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6(1):164–177, 1996.

[14] J. Lofberg. Yalmip: A toolbox for modeling and optimization in matlab. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289. IEEE, 2004.

[15] I. Molchanov and S. Zuyev. Optimisation in space of measures and optimal design. *ESAIM: Probability and Statistics*, 8:12–24, 2004.

[16] A. Mosek. The MOSEK optimization toolbox for matlab manual. *Version 7.1 (Revision 28)*, 2015.

[17] J. Nie. The $\mathcal{A}$-Truncated K-Moment problem. *Foundations of Computational Mathematics*, 14(6):1243–1276, 2014.

[18] D. Papp. Optimal designs for rational function regression. *Journal of the American Statistical Association*, 107(497):400–411, 2012.

[19] F. Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

[20] G. Sagnol. On the semidefinite representation of real functions applied to symmetric matrices. *Linear Algebra and its Applications*, 439(10):2829–2843, 2013.

[21] G. Sagnol and R. Harman. Computing exact D-optimal designs by mixed integer second-order cone programming. *The Annals of Statistics*, 43(5):2198–2224, 2015.

[22] C. Scheiderer. Semidefinitely representable convex sets. *ArXiv e-prints*, Dec. 2016.

[23] B. Torsney. W-iterations and ripples therefrom. In *Optimal Design and Related Areas in Optimization and Statistics*, pages 1–12. Springer, 2009.

[24] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM journal on matrix analysis and applications*, 19(2):499–533, 1998.

SUPPLEMENT TO "APPROXIMATE OPTIMAL DESIGNS FOR MULTIVARIATE POLYNOMIAL REGRESSION"

By Yohann De Castro[⋆], Fabrice Gamboa[○], Didier Henrion[•],
Roxana Hess[•] and Jean-Bernard Lasserre[•]

[⋆]*Laboratoire de Mathématiques d'Orsay*
Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France
[○]*Institut de Mathématiques de Toulouse*
Univ. Paul Sabatier, CNRS, 118 route de Narbonne, 31062 Toulouse, France
[•]*Laboratoire d'Analyse et d'Architecture des Systèmes*
Université de Toulouse, CNRS, 7 avenue du col. Roche, 31400 Toulouse, France

Abstract This article provides the proof of Theorem 1 and the detailed numerical results of the numerical examples of the article [2].

1. Proof of Theorem 1. First, let us prove that Problem (15) of the article [2] has an optimal solution. The feasible set is nonempty with finite associated objective value—take as feasible point the vector $\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X})$ associated with the Lebesgue measure on the compact set $\mathcal{X}$, scaled to be a probability measure. Moreover, as $\mathcal{X}$ is compact with nonempty interior, it follows that $\mathcal{M}_{2d}(\mathcal{X})$ is closed (as the dual of $\mathcal{P}_{2d}(\mathcal{X})$).

In addition, the feasible set $\{\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}) : y_0 = 1\}$ of Problem (15) is compact. Indeed there exists $M > 1$ such that it holds $\int_{\mathcal{X}} x_i^{2d} \, d\mu < M$ for every probability measure μ on $\mathcal{X}$ and every $i = 1, \ldots, n$. Hence, $\max\{y_0, \ \max_i\{L_y(x_i^{2d})\}\} < M$ which by [5] implies that $|y_\alpha| \leq M$ for every $|\alpha| \leq 2d$, which in turn implies that the feasible set of (15) is compact.

Next, as the function ϕ_q is upper semi-continuous, the supremum in (15) is attained at some optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$. Moreover, as the feasible set is convex and ϕ_q is strictly concave (see, e.g., [6, Chapter 6.13]) then $\mathbf{y}^*$ is the unique optimal solution.

Now, we examine the properties of the polynomial $p^\star$ and show the equivalence statement. For this we notice that there exists a strictly feasible solution because the cone $\mathrm{int}(\mathcal{M}_{2d}(\mathcal{X}))$ is nonempty by Lemma 2.6 in [4]. Hence, Slater's condition[1] holds for (15). Further, by a an argument in [6, Chap-

[1]For the optimization problem $\max\{f(x) : Ax = b; \ x \in C\}$, where $A \in \mathbb{R}^{m \times n}$ and $C \subseteq \mathbb{R}^n$ is a nonempty closed convex cone, Slater's condition holds, if there exists a feasible solution x in the interior of C.

ter 7.13], the matrix $M_d(\mathbf{y}^\star)$ is non-singular. Therefore, ϕ_q is differentiable at $\mathbf{y}^\star$. Since additionally Slater's condition is fulfilled and ϕ_q is concave, this implies that the *Karush-Kuhn-Tucker* (KKT) optimality conditions[2] at $\mathbf{y}^\star$ are necessary (and sufficient) for $\mathbf{y}^\star$ to be an optimal solution.

The KKT-optimality conditions read

$$\lambda^\star\, e_0 - \nabla\phi_q(M_d(\mathbf{y}^\star)) \;=\; \hat{\mathbf{p}}^\star \quad \text{with } \hat{p}^\star = \langle \hat{\mathbf{p}}, \mathbf{v}_{2d}(x) \rangle \in \mathcal{M}_{2d}(\mathcal{X})^\star \,(= \mathcal{P}_{2d}(\mathcal{X})),$$

(where $\hat{\mathbf{p}} \in \mathbb{R}^{\binom{n+2d}{n}}$, $e_0 = (1, 0, \ldots, 0)$, and $\lambda^\star$ is the dual variable associated with the constraint $y_0^\star = 1$). The complementarity condition is $\langle \mathbf{y}^\star, p^\star \rangle = 0$.

Writing $\mathbf{B}_\alpha$, $\alpha \in \mathbb{N}_{2d}^n$, for the real symmetric matrices satisfying

$$\forall x \in \mathcal{X}, \quad \sum_{|\alpha| \le 2d} \mathbf{B}_\alpha x^\alpha = \mathbf{v}_d(x)\mathbf{v}_d(x)^\top,$$

and $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{AB})$ for two real symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, this can be expressed as

$$(1) \qquad \Big(1_{\alpha=0}\,\lambda^\star - \langle \nabla\phi_q(M_d(\mathbf{y}^\star)), \mathbf{B}_\alpha \rangle \Big)_{|\alpha| \le 2d} \;=\; \hat{\mathbf{p}}, \quad \hat{p}^\star \in \mathcal{P}_{2d}(\mathcal{X}).$$

Multiplying (1) term-wise by $y_\alpha^\star$, summing up and invoking the complementarity condition, yields

$$(2) \qquad \lambda^\star \;=\; \lambda^\star\,y_0^\star \;\overset{(1)}{=}\; \Big\langle \nabla\phi_q(M_d(\mathbf{y}^\star)), \sum_{|\alpha| \le 2d} y_\alpha^\star \mathbf{B}_\alpha \Big\rangle$$
$$= \Big\langle \nabla\phi_q(M_d(\mathbf{y}^\star)), M_d(\mathbf{y}^\star) \Big\rangle \;=\; \phi_q(M_d(\mathbf{y}^\star)),$$

where the last equality holds by Euler formula for the positively homogeneous function ϕ_q.

Similarly, multiplying Equation (1) term-wise by x^α and summing up yields for all $x \in \mathcal{X}$

$$(3) \qquad x \mapsto \hat{p}^\star(x) \;\overset{(1)}{=}\; \lambda^\star - \Big\langle \nabla\phi_q(M_d(\mathbf{y}^\star)), \sum_{|\alpha| \le 2d} \mathbf{B}_\alpha x^\alpha \Big\rangle$$
$$= \lambda^\star - \Big\langle \nabla\phi_q(M_d(\mathbf{y}^\star)), \mathbf{v}_d(x)\mathbf{v}_d(x)^\top \Big\rangle \ge 0.$$

---

[2]For the optimization problem $\max\{f(x): Ax = b;\ x \in C\}$, where f is differentiable, $A \in \mathbb{R}^{m \times n}$ and $C \subseteq \mathbb{R}^n$ is a nonempty closed convex cone, the KKT-optimality conditions at a feasible point x state that there exist $\lambda^\star \in \mathbb{R}^m$ and $u^\star \in C^\star$ such that $A^\top \lambda^\star - \nabla f(x) = u^\star$ and $\langle x, u^\star \rangle = 0$.

For $q \neq 0$, let us consider

$$c^\star := \binom{n+d}{n} \left[\binom{n+d}{n}^{-1} \operatorname{trace}(M_d(\mathbf{y}^\star)^q) \right]^{1 - \frac{1}{q}}.$$

As $M_d(\mathbf{y}^\star)$ is positive semidefinite and non-singular, we have $c^\star > 0$. If $q = 0$, let $c^\star := 1$ and replace $\phi_0(M_d(\mathbf{y}^\star))$ by $\log \det M_d(\mathbf{y}^\star)$, for which the gradient is $M_d(\mathbf{y}^\star)^{-1}$.

Using Table 1 of the article [2], we find that $c^\star \nabla \phi_q(M_d(\mathbf{y}^\star)) = M_d(\mathbf{y}^\star)^{q-1}$. It follows that

$$c^\star \lambda^\star \overset{(2)}{=} c^\star \langle \nabla \phi_q(M_d(\mathbf{y}^\star)), M_d(\mathbf{y}^\star) \rangle = \operatorname{trace}(M_d(\mathbf{y}^\star)^q)$$

$$\text{and} \quad c^\star \langle \nabla \phi_q(M_d(\mathbf{y}^\star)), \mathbf{v}_d(x)\mathbf{v}_d(x)^\top \rangle \overset{[2, Eq.(16)]}{=} p_d^\star(x)$$

Therefore, Equation (3) is equivalent to $p^\star := c^\star \hat{p}^\star = c^\star \lambda^\star - p_d^\star \in \mathcal{P}_{2d}(\mathcal{X})$. To summarize,

$$p^\star(x) = \operatorname{trace}(M_d(\mathbf{y}^\star)^q) - p_d^\star(x) \in \mathcal{P}_{2d}(\mathcal{X}).$$

Since the KKT-conditions are necessary and sufficient, the equivalence statement follows.

Finally, we investigate the measure $\mu^\star$ associated with $\mathbf{y}^\star$. Multiplying the complementarity condition $\langle \mathbf{y}^\star, \hat{\mathbf{p}}^\star \rangle = 0$ with $c^\star$, we have

$$\int_{\mathcal{X}} \underbrace{p^\star(x)}_{\geq 0 \text{ on } \mathcal{X}} d\mu^\star(x) = 0.$$

Hence, the support of $\mu^\star$ is included in the algebraic set $\Omega = \{x \in \mathcal{X} : p^\star(x) = 0\}$.

The measure $\mu^\star$ is an atomic measure supported on at most $\binom{n+2d}{n}$ points. This follows from Tchakaloff's theorem (see [3, Theorem B.12] or [1] for instance), which states that for every finite Borel probability measure on $\mathcal{X}$ and every $s \in \mathbb{N}$, there exists an atomic measure μ_s supported on $\ell \leq \binom{n+s}{n}$ points such that all moments of μ_s and $\mu^\star$ agree up to order s. For $s = 2d$ we get that $\ell \leq \binom{n+2d}{n}$. If $\ell < \binom{n+d}{n}$, then rank $M_d(\mathbf{y}^\star) < \binom{n+d}{n}$ in contradiction to $M_d(\mathbf{y}^\star)$ being non-singular. Therefore, $\binom{n+d}{n} \leq \ell \leq \binom{n+2d}{n}$.

2. General case. The last paragraph has to be adapted as follows in the general case. Recall that there exists a full row rank matrix $\mathfrak{A}$ of size $p \times \binom{n+d}{n}$ such that the regression polynomials satisfy $\mathbf{F}(x) = \mathfrak{A}\,\mathbf{v}_d(x)$. Recall also that we are optimizing over the cone of matrices of the form $\mathbf{M}_d(\mathbf{y}) := \mathfrak{A}M_d(\mathbf{y})\mathfrak{A}^\top$ indexed by moment sequences $\mathbf{y}$.

First, note that

$$\operatorname{rank} \mathbf{M}_d(\mathbf{y}) = \min(p, \operatorname{rank} M_d(\mathbf{y}))$$

and recall that the optimal solution $\mathbf{M}_d(\mathbf{y}^\star)$ has full rank, namely it holds that $\operatorname{rank} \mathbf{M}_d(\mathbf{y}^\star) = p$. We deduce that $\operatorname{rank} M_d(\mathbf{y}^\star) \geq p$ so that $\mu^\star$ has at least p support points.

Then, consider the vector space spanned by the constant function 1 and the polynomials $x \mapsto \mathbf{f}_i(x)\mathbf{f}_j(x)$ for $1 \leq i, j \leq p$. Denote by $\bar{s}$ its dimension and observe that

$$\bar{s} \leq \min\left[1 + \frac{p(p+1)}{2}, \binom{n+2d}{n}\right].$$

The first argument in the minimum is the number of quadratic terms $\mathbf{f}_i\mathbf{f}_j$ while the second comes from the observation that their span is included in the vector space of multivariate polynomials of n variables of degree at most $2d$. Recall that we want to represent the outcome of the linear evaluations

$$(\mathbf{M}_d(\mathbf{y}^\star))_{i,j} = \int \mathbf{f}_i\mathbf{f}_j \mathrm{d}\mu^\star, \quad 1 \leq i, j \leq p,$$

by a discrete probability measure $\mu^\star$. By Tchakaloff's theorem, see for instance [1, Corollary 2 [3]], we get that there exists a representing probability measure $\mu^\star$ of $\mathbf{M}_d(\mathbf{y}^\star)$ with at most $\bar{s}$ support points.

3. Numerical results for the Examples. We list in Table 1 details on the results for the two-dimensional examples (Sections 6.2, 6.3, 6.4, and 6.5 of the article [2]), namely, the numerical values of the support points and their corresponding weights.

---

[3]In [1, Corollary 2], the reader may consider $(\phi_j)_{j=1,\ldots,\bar{s}}$ any basis of the vector space spanned by the constant function 1 and the polynomials $x \mapsto \mathbf{f}_i(x)\mathbf{f}_j(x)$ to get the result.

	Wynn		Ellipses		Moon		Folium	
	(x_1, x_2)	ω	(x_1, x_2)	ω	(x_1, x_2)	ω	(x_1, x_2)	ω
$d=1$	(-0.35,-0.35)	0.125	(-0.00,-0.75)	0.250	(-0.80, 0.00)	0.329	(0.29,-0.55)	0.333
	(-0.35, 0.35)	0.281	(-0.90,-0.00)	0.250	(0.07,-0.53)	0.305	(-1.00, 0.00)	0.333
	(0.35,-0.35)	0.281	(0.90, 0.00)	0.250	(0.07, 0.53)	0.305	(0.29, 0.55)	0.333
	(0.71, 0.71)	0.313	(0.00, 0.75)	0.250	(0.33,-0.29)	0.031		
					(0.33, 0.29)	0.031		
$d=2$	(-0.35,-0.35)	0.163	(-0.45,-0.65)	0.134	(-0.39,-0.57)	0.167	(-1.00, 0.00)	0.167
	(-0.35, 0.35)	0.165	(-0.90,-0.00)	0.139	(-0.80, 0.00)	0.167	(-0.60,-0.21)	0.166
	(0.12, 0.12)	0.066	(-0.00,-0.39)	0.093	(-0.20,-0.00)	0.167	(-0.60, 0.21)	0.166
	(0.35,-0.35)	0.165	(0.45,-0.65)	0.134	(0.29,-0.35)	0.167	(0.28,-0.56)	0.162
	(0.18, 0.53)	0.141	(-0.45, 0.65)	0.134	(-0.39, 0.57)	0.167	(0.21,-0.20)	0.088
	(0.53, 0.18)	0.141	(0.00, 0.39)	0.093	(0.29, 0.35)	0.167	(0.21, 0.20)	0.088
	(0.71, 0.71)	0.159	(0.90, 0.00)	0.139			(0.28, 0.56)	0.162
			(0.45, 0.65)	0.134				
$d=3$	(-0.35,-0.35)	0.095	(-0.64,-0.53)	0.085	(-0.57,-0.47)	0.099	(-1.00,-0.00)	0.100
	(0.02,-0.35)	0.074	(-0.90, 0.00)	0.088	(-0.08,-0.59)	0.098	(-0.77,-0.20)	0.099
	(-0.35, 0.02)	0.074	(-0.00,-0.75)	0.088	(-0.80, 0.00)	0.100	(-0.77, 0.20)	0.099
	(0.35,-0.35)	0.096	(-0.36,-0.32)	0.075	(-0.45,-0.18)	0.061	(-0.45, 0.00)	0.077
	(0.14,-0.12)	0.044	(0.00,-0.39)	0.005	(-0.11,-0.30)	0.062	(-0.14,-0.00)	0.033
	(-0.12, 0.14)	0.044	(-0.64, 0.53)	0.085	(-0.45, 0.18)	0.061	(0.10,-0.41)	0.098
	(-0.35, 0.35)	0.097	(-0.36, 0.32)	0.075	(0.33,-0.29)	0.099	(0.29,-0.56)	0.099
	(0.45,-0.06)	0.088	(0.36,-0.32)	0.075	(-0.57, 0.47)	0.099	(0.31,-0.35)	0.100
	(-0.06, 0.45)	0.088	(0.64,-0.53)	0.085	(0.11,-0.00)	0.063	(0.10, 0.41)	0.098
	(0.39, 0.39)	0.037	(-0.00, 0.39)	0.005	(-0.11, 0.30)	0.062	(0.31, 0.35)	0.100
	(0.61, 0.41)	0.084	(0.36, 0.32)	0.075	(-0.08, 0.59)	0.098	(0.29, 0.56)	0.099
	(0.41, 0.61)	0.084	(-0.00, 0.75)	0.088	(0.33, 0.29)	0.099		
	(0.71, 0.71)	0.097	(0.90,-0.00)	0.088				
			(0.64, 0.53)	0.085				

TABLE 1

The numerical values for Examples 6.2, 6.3, 6.4, and 6.5 of the paper [2] for the support points $x_i = (x_{i,1}, x_{i,2})$ and their corresponding weights ω_i, $i = 1, \ldots, \ell$.

REFERENCES

[1] C. Bayer and J. Teichmann. The proof of Tchakaloff's theorem. *Proceedings of the American mathematical society*, 134(10):3035–3040, 2006.

[2] Y. De Castro, F. Gamboa, D. Henrion, R. Hess, and J.-B. Lasserre. Approximate Optimal Designs for Multivariate Polynomial Regression. *Annals of Statistics*, 2018.

[3] J. B. Lasserre. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2010.

[4] J.-B. Lasserre. A generalization of Löwner-John's ellipsoid theorem. *Mathematical Programming*, 152(1-2):559–591, 2015.

[5] J. B. Lasserre and T. Netzer. SOS approximations of nonnegative polynomials via simple high degree perturbations. *Mathematische Zeitschrift*, 256(1):99–112, 2007.

[6] F. Pukelsheim. *Optimal design of experiments*. SIAM, 2006.