

D-OPTIMAL DESIGN FOR MULTIVARIATE POLYNOMIAL REGRESSION VIA THE CHRISTOFFEL FUNCTION AND SEMIDEFINITE RELAXATIONS

Yohann de Castro, F Gamboa, Didier Henrion, Roxana Hess, Jean-Bernard Lasserre

▶ **To cite this version:**

Yohann de Castro, F Gamboa, Didier Henrion, Roxana Hess, Jean-Bernard Lasserre. D-OPTIMAL DESIGN FOR MULTIVARIATE POLYNOMIAL REGRESSION VIA THE CHRISTOFFEL FUNCTION AND SEMIDEFINITE RELAXATIONS. 2017. hal-01483490v1

HAL Id: hal-01483490

https://laas.hal.science/hal-01483490v1

Preprint submitted on 6 Mar 2017 (v1), last revised 17 Oct 2017 (v2)

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

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

D-OPTIMAL DESIGN FOR MULTIVARIATE POLYNOMIAL REGRESSION VIA THE CHRISTOFFEL FUNCTION AND SEMIDEFINITE RELAXATIONS

Y. DE CASTRO, F. GAMBOA, D. HENRION, R. HESS, AND J.-B. LASSERRE

Abstract. We present a new approach to the design of D-optimal experiments with multivariate polynomial regressions on compact semi-algebraic design spaces. We apply the moment-sum-of-squares hierarchy of semidefinite programming problems to solve numerically and approximately the optimal design problem. The geometry of the design is recovered with semidefinite programming duality theory and the Christoffel polynomial.

1. Introduction

1.1. **Convex design theory.** The optimum experimental designs are computational and theoretical objects that minimize the variance of the best linear unbiased estimators in regression problems. In this frame, the experimenter models the responses $z_1, \dots, z_N$ of a random *experiment* whose inputs are represented by a vector $\xi_i \in \mathbb{R}^n$ with respect to known *regression functions* $\varphi_1, \dots, \varphi_p$, namely

$$z_i = \sum_{j=1}^{p} \theta_j \varphi_j(\xi_i) + \varepsilon_i\,, \;\; i = 1, \dots; N$$

where $\theta_1, \dots, \theta_p$ are unknown parameters that the experimenter wants to estimate, ε_i is some noise and the inputs ξ_i are chosen by the experimenter in a *design space* $\mathcal{X} \subseteq \mathbb{R}^n$. Assume that the inputs ξ_i, for $i = 1, \dots, N$, are chosen within a set of distinct points $x_1, \dots, x_\ell$ with $\ell \leq N$, and let n_k denote the number of times the particular point x_k occurs among $\xi_1, \dots, \xi_N$. This would be summarized by

$$\zeta := \begin{pmatrix} x_1 & \cdots & x_\ell \\ \frac{n_1}{N} & \cdots & \frac{n_\ell}{N} \end{pmatrix} \tag{1.1}$$

whose first row gives the 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. The goal of the design of experiment theory is then to assess which input parameters and frequencies the experimenter has to consider. For a given ζ the standard analysis of the Gaussian linear model shows that the minimal covariance matrix (with respect to Loewner ordering) of unbiased estimators can be expressed in terms of the Moore-Penrose pseudoinverse of the *information matrix* which is defined by

$$M(\zeta) := \sum_{i=1}^{\ell} w_i \Phi(x_i) \Phi^\top(x_i) \tag{1.2}$$

where $\Phi := (\phi_1, \dots, \phi_p)$ is the column vector of regression functions. One major aspect of design of experiment theory seeks to maximize the information matrix over the set of all possible ζ. Notice the Loewner ordering is partial and, in general, there is no greatest element among all possible information matrices $M(\zeta)$. The standard approach is then to consider some statistical criteria, namely *Kiefer's ϕ_q-criteria* [K74], in order to describe and construct the "*optimal designs*" with respect to those criteria. Observe that the information matrix belongs to $\mathbb{S}_p^+$, the space of symmetric nonnegative definite matrices of size p, and for all $q \in [-\infty, 1]$ define the function

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

Date: Draft of March 6, 2017.

Key words and phrases. Experimental design; semidefinite programming.

where for positive definite matrices M it holds

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

and for nonnegative definite matrices M it holds

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

Those criteria are meant to be real valued, positively homogeneous, non constant, upper semi-continuous, isotonic (with respect to the Loewner ordering) and concave functions. Throughout this paper, we restrict ourselves to the D-optimality criteria which corresponds to the choice $q = 0$. Other criteria will be studied elsewhere.

In particular, in this paper we search for solutions $\zeta^\star$ to the following optimization problem

$$\max \log\det M(\zeta) \tag{1.3}$$

where the maximum is taken over all ζ of the form (1.1). Note that the logarithm of the determinant is used instead of ϕ_0 because of its standard use in semidefinite programming (SDP) as a barrier function for the cone of positive definite matrices.

1.2. **State of the art.** Optimal design is at the heart of statistical planning for inference in the linear model, see for example [BHH78]. While the case of discrete input factors is generally tackled by algebraic and combinatoric arguments (e.g., [B08]), the one of continuous input factors often leads to an optimization problem. In general, the continuous factors are generated by a vector Φ of linearly independent regular functions on the *design space* $\mathcal{X}$. One way to handle the problem is to focus only on $\mathcal{X}$ ignoring the function Φ and to try to draw the design points *filling* the best the set $\mathcal{X}$. This is generally done by optimizing a cost function on $\mathcal{X}^N$ that traduces the way the design points are positioned between each other and/or how they fill the space. Generic examples are the so-called maxmin or minmax criteria (see for example [PW88] or [WPN97]) and the minimum discrepancy designs (see for example [LQX05]). Another point of view—which is the one developed here—relies on the *maximization* of the information matrix. Of course, as explained before, the Loewner order is partial and so the optimization can not stand on this matrix but on one of its feature. A pioneer paper adopting this point of view is the one of Elfving [E52]. In the early 60's, in a series of papers, Kiefer and Wolwofitz shade new lights on this kind of methods for experimental design by introducing the equivalence principle and proposing in some cases algorithms to solve the optimization problem, see [K74] and references therein. Following the early works of Karlin and Studden ([KS66b], [KS66a]), the case of polynomial regression on a compact interval on $\mathbb{R}$ has been widely studied. In this frame, the theory is almost complete an many thing can be said about the optimal solutions for the design problem (see [DS93]). Roughly speaking, the optimal design points are related to the zeros of orthogonal polynomials built on an equilibrum measure. We refer to the excelent book of Dette and Studden [DS97] and reference therein for a complete overview on the subject. In the one dimensional frame, other systems of functions Φ (trigonometric functions or some T-system, see [KN77] for a definition) are studied in a same way in [DS97], [LS85] and [IS01]. In the multidimensional case, even for polynomial systems, very few case of explicit solutions are known. Using tensoring arguments the case of a rectangle is treated in [DS97]. Particular models of degree two are studied in [DG14] and [PW94]. Away from these particular cases, the construction of the optimal design relies on numerical optimization procedures. The case of the determinant (D-optimality) is studied for example in [W70] and [VBW98]. An other criterion based on matrix conditioning is developed in [MYZ14]. General optimization algorithm are discussed in [F10] and [ADT07]. In the frame of fixed given support points efficient SDP based algorithms are proposed and studied in [S11] and [SH15]. Let us mention, the paper [VBW98] which is one of the original motivation to develop SDP solvers, especially for Max Det Problems (corresponding to D-optimal design) and the so-called problem of analytical centering.

1.3. **Contribution.** For the first time, this paper introduces a general method to compute the approximate D-optimal designs on a large variety of design spaces that we referred to as semi-algebraic sets, see [L10] for a definition. This family can be understood as any sets given by intersections and complements of

the level sets of multivariate polynomials. The theoretical guarantees are given by Theorems 4.3 and 4.4. We apply the moment-sum-of-squares hierarchy (a.k.a. the Lasserre hierarchy) of SDP problems to solve numerically and approximately the optimal design problem They show the convergence of our method towards the optimal information matrix as the order of the hierarchy increases. Furthermore, we show that 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 the Christoffel polynomial. 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, using Christoffel polynomials of degrees higher than two allows to reconstruct designs with points in the interior of the domain, contrasting with the classical use of ellipsoids for linear regressions.

1.4. **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 optimum designs and their approximations. At the end of the section we propose a two step procedure to solve the approximate design problem. Solving the first step is subject to Section 4. There, we find a sequence of moments 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 conclusion.

2. Polynomial optimal design and moments

This section collects preliminary material on semialgebraic sets, moments and moment matrices, using the notations of [L10]. 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 $x = (x_1, \dots, x_n)$, and for $d \in \mathbb{N}$, define $\mathbb{R}[x]_d := \{p \in \mathbb{R}[x] : \deg p \leq d\}$ where $\deg p$ denotes the total degree of p.

In this paper we assume that the regression functions are multivariate polynomials, i.e. $\Phi = (\phi_1, \dots, \phi_p) \in (\mathbb{R}[x]_d)^p$. Moreover, we consider that the design space $\mathcal{X} \subset \mathbb{R}^n$ is a given closed basic semi-algebraic set

$$\mathcal{X} := \{x \in \mathbb{R}^m : g_j(x) \geq 0,\ j = 1, \dots, m\} \tag{2.1}$$

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

Notice that those assumptions cover a large class of problems in optimal design theory, see for instance [DS97, Chapter 5]. In particular, observe that the design space $\mathcal{X}$ defined by (2.1) is not necessarily convex and note that the polynomial regressors Φ can handle incomplete m-way dth degree polynomial regression.

The monomials $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, with $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, 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 $\binom{n+d}{n}$ with basis $(x^\alpha)_{|\alpha| \leq d}$, where $|\alpha| := \alpha_1 + \cdots + \alpha_n$. We write

$$\mathbf{v}_d(x) := (\ \underbrace{1}_{\text{degree 0}}\ , \underbrace{x_1, \dots, x_n}_{\text{degree 1}}, \underbrace{x_1^2, x_1x_2, \dots, x_1x_n, x_2^2, \dots, x_n^2}_{\text{degree 2}}, \dots, \underbrace{x_1^d, \dots, x_n^d}_{\text{degree d}}\ )^T$$

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.

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 $\mu \in \mathscr{M}_+(\mathcal{X})$ 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 a sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n} \subseteq \mathbb{R}$ 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 $\mathbf{y} = (y_\alpha)_{|\alpha|\le d}$ which have a representing measure supported on $\mathcal{X}$. We call it the *moment cone* of $\mathcal{X}$. It can be expressed as

$$\mathcal{M}_d(\mathcal{X}) := \{\mathbf{y} \in \mathbb{R}^{\binom{n+d}{n}} : \exists\, \mu \in \mathscr{M}_+(\mathcal{X}) \text{ s.t. } y_\alpha = \int_{\mathcal{X}} x^\alpha \, d\mu,\ \forall \alpha \in \mathbb{N}^n,\ |\alpha| \le d\}. \tag{2.2}$$

We also denote by $\mathcal{P}_d(\mathcal{X})$ the convex cone of polynomials of degree at most d that are nonnegative on $\mathcal{X}$. When $\mathcal{X}$ is compact 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. [L15][Lemma 2.5]).

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. [VBW98]. Unfortunately, in the general case, there is no efficient representation of this cone. It has actually been shown in [S16] 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 sequence $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n} \subseteq \mathbb{R}$ 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 $L_\mathbf{y}(f) \ge 0$ for all polynomials $f \in \mathbb{R}[x]$ nonnegative on $\mathcal{X}$ [L10, Theorem 3.1].

The *moment matrix* of a truncated sequence $\mathbf{y} = (y_\alpha)_{|\alpha|\le 2d}$ is the $\binom{n+d}{n} \times \binom{n+d}{n}$-matrix $M_d(\mathbf{y})$ with rows and columns respectively indexed by $\alpha \in \mathbb{N}^n, |\alpha|, |\beta| \le d$ 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 and linear in $\mathbf{y}$, and if $\mathbf{y}$ has a representing measure, then $M_d(\mathbf{y})$ is *positive semidefinite*, denoted by $M_d(\mathbf{y}) \succcurlyeq 0$.

Similarly, we define the *localizing matrix* of a polynomial $f = \sum_{|\alpha|\le r} f_\alpha x^\alpha \in \mathbb{R}[x]_r$ of degree r and a sequence $\mathbf{y} = (y_\alpha)_{|\alpha|\le 2d+r}$ as the $\binom{n+d}{n} \times \binom{n+d}{n}$-matrix $M_d(f\mathbf{y})$ with rows and columns respectively indexed by $\alpha \in \mathbb{N}^n, |\alpha|, |\beta| \le d$ 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 $\{x \in \mathbb{R}^n : f(x) \ge 0\}$.

Since $\mathcal{X}$ is basic semialgebraic with a certificate of compactness, by Putinar's theorem [L10, Theorem 3.8], we also know the converse statement in the infinite case, namely $\mathbf{y} = (y_\alpha)_{\alpha\in\mathbb{N}^n}$ 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 $v_j := \lceil d_j/2 \rceil$, $j = 1, \ldots, m$, for 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 $\delta \in \mathbb{N}$

$$\mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}) := \{\mathbf{y} \in \mathbb{R}^{\binom{n+2d}{n}} : \exists \mathbf{y}_\delta \in \mathbb{R}^{\binom{n+2(d+\delta)}{n}} \text{ such that } (y_{\delta,\alpha})_{|\alpha|\le 2d} = \mathbf{y} \text{ and } M_{d+\delta}(\mathbf{y}_\delta) \succcurlyeq 0,\ M_{d+\delta-v_j}(g_j \mathbf{y}_\delta) \succcurlyeq 0,\ j = 1,\dots,m\}.$$

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 $\mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$, they are outer approximations of the moment cone. Moreover, they form a nested sequence, so we can build the hierarchy

$$\mathcal{M}_{2d}(\mathcal{X}) \subseteq \cdots \subseteq \mathcal{M}^{\mathsf{SDP}}_{2d+2}(\mathcal{X}) \subseteq \mathcal{M}^{\mathsf{SDP}}_{2d+1}(\mathcal{X}) \subseteq \mathcal{M}^{\mathsf{SDP}}_{2d}(\mathcal{X}). \tag{2.3}$$

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

3. Approximate Optimal Design

3.1. **Problem reformulation in the multivariate polynomial case.** For all $i = 1,\dots,p$ and $x \in \mathcal{X}$, let $\varphi_i(x) := \sum_{|\alpha|\le d} a_{i,\alpha} x^\alpha$ with appropriate $a_{i,\alpha} \in \mathbb{R}$. Define for $\mu \in \mathscr{M}_+(\mathcal{X})$ with moment sequence $\mathbf{y}$ the information matrix

$$M(\mu) = \Big(\int_{\mathcal{X}} \varphi_i \varphi_j \mathrm{d}\mu \Big)_{1\le i,j\le p} = \Big(\sum_{|\alpha|,|\beta|\le d} a_{i,\alpha} a_{j,\beta} y_{\alpha+\beta} \Big)_{1\le i,j\le p} = \sum_{|\gamma|\le 2d} A_\gamma y_\gamma$$

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

Further, let $\mu = \sum_{i=1}^{\ell} w_i \delta_{x_i}$ where δ_x denotes the Dirac measure at the point $x \in \mathcal{X}$ and observe that $M(\mu) = \sum_{i=1}^{\ell} w_i \Phi(x_i)\Phi^\top(x_i)$ as in (1.2).

The optimization problem

$$\begin{aligned} &\max \log\det M \\ &\text{s.t. } M = \sum_{|\gamma|\le 2d} A_\gamma y_\gamma \succcurlyeq 0, \quad y_\gamma = \sum_{i=1}^{\ell} \frac{n_i}{N} x_i^\gamma, \quad \sum_{i=1}^{\ell} n_i = N, \\ &\quad x_i \in \mathcal{X},\ n_i \in \mathbb{N},\ i = 1,\dots,\ell \end{aligned} \tag{3.1}$$

where the maximization is with respect to x_i and n_i, $i = 1,\dots,\ell$, subject to the constraint that the information matrix M is positive semidefinite, is by construction equivalent to the original design problem (1.3). In this form, problem (3.1) 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 3.1, 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

$$\zeta := \begin{pmatrix} x_1 & \cdots & x_\ell \\ w_1 & \cdots & w_\ell \end{pmatrix}, \tag{3.2}$$

where the frequencies w_i belong to the unit simplex $\mathcal{W} := \{w \in \mathbb{R}^l : 0 \le w_i \le 1,\ \sum_{i=1}^{\ell} w_i = 1\}$. Accordingly, any solution to (1.3) where the maximum is taken over all matrices of type (3.2) is called "*approximate optimal design*", yielding the following relaxation of problem 3.1

$$\begin{aligned} &\max \log\det M \\ &\text{s.t. } M = \sum_{|\gamma|\le 2d} A_\gamma y_\gamma \succcurlyeq 0, \quad y_\gamma = \sum_{i=1}^{\ell} w_i x_i^\gamma, \\ &\quad x_i \in \mathcal{X},\ w \in \mathcal{W} \end{aligned} \tag{3.3}$$

Algorithm 1: Approximate optimal design.

Data: A design space $\mathcal{X}$ defined by polynomials $g_j, j = 1, \dots, m$, as in (2.1), and polynomial regressors $\Phi = (\varphi_1, \dots, \varphi_p)$.

Result: An approximate ϕ_q-optimal design $\zeta^\star = \begin{pmatrix} x_1^\star & \cdots & x_\ell^\star \\ w_1^\star & \cdots & w_\ell^\star \end{pmatrix}$ for $q = -\infty, -1, 0$ or 1.

(1) Find the truncated moments $(y_\alpha^\star)_{|\alpha|\le 2d}$ that maximizes (3.4).
(2) Find a representing measure $\mu^\star = \sum_{i=1}^{\ell} w_i^\star \delta_{x_i^\star} \ge 0$ of $(y_\alpha^\star)_{|\alpha|\le 2d}$, for $x_i^\star \in \mathcal{X}$ and $w_i^\star \in \mathcal{W}$.

where the maximization is with respect to x_i and w_i, $i = 1, \dots, \ell$, 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 (3.3). For this we will first again reformulate our problem.

Observe that by Carathéodory's theorem, the truncated moment cone $\mathcal{M}_{2d}(\mathcal{X})$ defined in (2.2) is exactly

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

so that problem (3.2) is equivalent to

$$\begin{aligned} &\max \log\det M \\ &\text{s.t. } M = \sum_{|\gamma|\le 2d} A_\gamma y_\gamma \succcurlyeq 0, \\ &\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}),\ y_0 = 1 \end{aligned} \tag{3.4}$$

where the maximization is now with respect to the sequence $\mathbf{y}$. Moment problem (3.4) 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 (2.3), we obtain a hierarchy of finite dimensional semidefinite programming problems converging to the optimal solution of (3.4). Since semidefinite programming problems can be solved efficiently, we can compute a numerical solution to problem (3.2).

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 optimum design $\zeta^\star$.

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 $\Phi = (\varphi_1, \dots, \varphi_p) := (x^\alpha)_{|\alpha|\le d}$ with $p = \binom{n+d}{n}$. Note that this is not a restriction, since one can get the results for other choices of Φ by simply performing a change of basis. Different polynomial bases can be considered and one may consult the standard framework described by [DS97, 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.

4.1. **Christoffel polynomials.** It turns out that the (unique) optimal solution $\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X})$ of (3.4) 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 $\mathbf{y}$.

Definition 4.1. Let $\mathbf{y} \in \mathbb{R}^{\binom{n+2d}{n}}$ be such that $\mathbf{M}_d(\mathbf{y}) \succ 0$. Then there exists a family of orthonormal polynomials $(P_\alpha)_{|\alpha|\le d} \subseteq \mathbb{R}[x]_d$ satisfying

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

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

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

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

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

or via its extremal property

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

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 $L_{\mathbf{y}}(P^2)\,(= P^T \mathbf{M}_d(\mathbf{y})\, P)$). For more details the interested reader is referred to [LP16] and the references therein.

4.2. **The ideal problem on moments.** The ideal formulation (3.4) of our approximate optimal design problem reads

$$\begin{array}{rll} \rho = & \max\limits_{\mathbf{y}} & \log\det \mathbf{M}_d(\mathbf{y}) \\ & \text{s.t.} & \mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X}),\ y_0 = 1. \end{array} \tag{4.1}$$

For this we have the following result

Theorem 4.2. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be compact with nonempty interior. Problem (4.1) is a convex optimization problem with a unique optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$. It is the vector of moments (up to order $2d$) of a measure $\mu^\star$ supported on at least $\binom{n+d}{n}$ and at most $\binom{n+2d}{n}$ points in the set $\Omega := \{x \in \mathcal{X} : \binom{n+d}{n} - p_d^\star(x) = 0\}$ where $\binom{n+d}{n} - p_d^\star \in \mathcal{P}_{2d}(\mathcal{X})$ and $p_d^\star$ is the Christoffel polynomial*

$$x \mapsto p_d^\star(x) := \mathbf{v}_d(x)^T \mathbf{M}_d(\mathbf{y}^\star)^{-1} \mathbf{v}_d(x). \tag{4.2}$$

Proof. First, let us prove that (4.1) has an optimal solution. The feasible set is nonempty (take as feasible point the vector $\mathbf{y} \in \mathcal{M}_{2d}(\mathcal{X})$ associated with the Lebesgue measure on $\mathcal{X}$, scaled to be a probability measure) with finite associated objective value, because $\det \mathbf{M}_d(\mathbf{y}) > 0$. Hence, $\rho > -\infty$ and in addition Slater's condition holds because $\mathbf{y} \in \mathrm{int}\mathcal{M}(\mathcal{X})$ (that is, $\mathbf{y}$ is a strictly feasible solution to (4.1)).

Next, as $\mathcal{X}$ is compact there exists $M > 1$ such that $\int_{\mathcal{X}} x_i^{2d}\, d\mu < M$ 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 [LN07] implies that $|y_\alpha| \le M$ for every $|\alpha| \le 2d$, which in turn implies that the feasible set of (4.1) is compact. As the objective function is continuous and $\rho > -\infty$, problem (4.1) has an optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$.

Furthermore, an optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ is unique because the objective function is strictly convex and the feasible set is convex. In addition, since $\rho > -\infty$, $\det \mathbf{M}_d(\mathbf{y}^\star) \neq 0$ and so $\mathbf{M}_d(\mathbf{y}^\star)$ is non singular.

Next, writing $\mathbf{B}_\alpha$, $\alpha \in \mathbb{N}^n_{2d}$, for the real matrices satisfying $\sum_{|\alpha|\le 2d} \mathbf{B}_\alpha x^\alpha = \mathbf{v}_d(x)\mathbf{v}_d(x)^T$ and $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{trace}(\mathbf{AB})$ for two matrices $\mathbf{A}$ and $\mathbf{B}$, the necessary *Karush-Kuhn-Tucker* optimality conditions[2] (in short KKT-optimality conditions) state that an optimal solution $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$ should satisfy

$$\lambda^\star\, e_0 - \nabla \log\det \mathbf{M}_d(\mathbf{y}^\star) \;=\; p^\star \in \mathcal{M}_{2d}(\mathcal{X})^\star\ (= \mathcal{P}_{2d}(\mathcal{X})),$$

[1] In fact what is referred to in the literature is its reciprocal $x \mapsto 1/p_d(x)$ called the *Chistoffel function*.

[2] For the optimization problem $\min\{f(x) : Ax = b; x \in C\}$ where f is differentiable, $A \in \mathbb{R}^{m\times n}$ and $C \subset \mathbb{R}^n$ is a nonempty closed convex cone, the KKT-optimality conditions at a feasible point x state that there exists $\lambda^\star \in \mathbb{R}^m$ and $u \in C^\star$ such that $\nabla f(x) - A^T\lambda^\star = u^\star$ and $\langle x, u^\star\rangle = 0$. Slater's condition holds if there exists a feasible solution $x \in \mathrm{int}(C)$, in which case the KKT-optimality conditions are necessary and sufficient if f is convex.

(where $e_0 = (1, 0, \dots 0)$ and $\lambda^\star$ is the dual variable associated with the constraint $y^\star = 1$), with the complementarity condition $\langle \mathbf{y}^\star, p^\star \rangle = 0$. That is:

$$\big(1_{\alpha=0}\,\lambda_0^\star - \langle \mathbf{M}_d(\mathbf{y}^\star)^{-1}, \mathbf{B}_\alpha \rangle\big)_{|\alpha|\le 2d} \;=\; p^\star \;\in\; \mathcal{P}_{2d}(\mathcal{X}); \quad \langle \mathbf{y}^\star, p^\star \rangle = 0. \tag{4.3}$$

Multiplying (4.3) term-wise by $y_\alpha^\star$, summing up and invoking the complementarity condition, yields

$$\lambda_0^\star \;=\; \lambda_0^\star\, \mathbf{y}_0^\star \;=\; \langle \mathbf{M}_d(\mathbf{y}^\star)^{-1}, \sum_{|\alpha|\le 2d} y_\alpha^\star \mathbf{B}_\alpha \rangle \;=\; \langle \mathbf{M}_d(\mathbf{y}^\star)^{-1}, \mathbf{M}_d(\mathbf{y}^\star) \rangle \;=\; \binom{n+d}{n}.$$

Similarly, multiplying (4.3) term-wise by x^α and summing up yields

$$\begin{aligned} x \mapsto p^\star(x) \;&:=\; \binom{n+d}{n} - \langle \mathbf{M}_d(\mathbf{y}^\star)^{-1}, \sum_{|\alpha|\le 2d} \mathbf{B}_\alpha x^\alpha \rangle \\ &=\; \binom{n+d}{n} - \langle \mathbf{M}_d(\mathbf{y}^\star)^{-1}, \mathbf{v}_d(x)\mathbf{v}_d(x)^T \rangle \\ &=\; \binom{n+d}{n} - \sum_{|\alpha|\le d} P_\alpha(x)^2 \ge 0 \quad \forall x \in \mathcal{X}, \end{aligned} \tag{4.4}$$

where the (P_α), $\alpha \in \mathbb{N}^n_d$, are the orthonormal polynomials (up to degree d) w.r.t. $\mu^\star$, where $\mu^\star$ is a representing measure of $\mathbf{y}^\star$ (recall that $\mathbf{y}^\star \in \mathcal{M}_{2d}(\mathcal{X})$). Therefore $p^\star = \binom{n+d}{n} - p_d^\star \in \mathcal{P}_{2d}(\mathcal{X})$ where $p_d^\star$ is the Christoffel polynomial of degree $2d$ associated with $\mu^\star$. Next, the complementarity condition $\langle \mathbf{y}^\star, p^\star \rangle = 0$ reads

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

which implies that the support of $\mu^\star$ is included in the algebraic set $\{x \in \mathcal{X} : p^\star(x) = 0\}$. Finally, that $\mu^\star$ is an atomic measure supported on at most $\binom{n+2d}{n}$ points follows from Tchakaloff's theorem [L10, Theorem B.12] which states that for every finite Borel probability measure on $\mathcal{X}$ and every $t \in \mathbb{N}$, there exists an atomic measure μ_t supported on $\ell \le \binom{n+t}{n}$ points such that all moments of μ_t and $\mu^\star$ agree up to order t. So let $t = 2d$. Then $\ell \le \binom{n+2d}{n}$. If $\ell < \binom{n+d}{n}$, then $\text{rank}\mathbf{M}_d(\mathbf{y}^\star) < \binom{n+d}{n}$ in contradiction to $\mathbf{M}_d(\mathbf{y}^\star)$ being non-singular. Therefore, $\binom{n+d}{n} \le \ell \le \binom{n+2d}{n}$. □

So we obtain a nice characterization of the unique optimal solution $\mathbf{y}^\star$ of (4.1). It is the vector of moments up to order $2d$ of a measure $\mu^\star$ supported on finitely many (at least $\binom{n+d}{n}$ and at most $\binom{n+2d}{n}$) points of $\mathcal{X}$. This support of $\mu^\star$ consists of zeros of the equation $\binom{n+d}{n} - p_d^\star(x) = 0$, where $p_d^\star$ is the Christoffel polynomial associated with $\mu^\star$. Moreover the level set $\{x : p_d^\star(x) \le \binom{n+d}{n}\}$ contains $\mathcal{X}$ and intersects $\mathcal{X}$ precisely at the support of $\mu^\star$.

4.3. **The SDP approximation scheme.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be as defined in (2.1), assumed to be compact. So with no loss of generality (and possibly after scaling), assume that $x \mapsto g_1(x) = 1 - \|x\|^2 \ge 0$ is one of the constraints defining $\mathcal{X}$.

Since the ideal moment problem (4.1) involves the moment cone $\mathcal{M}_{2d}(\mathcal{X})$ which is not SDP representable, we use the hierarchy (2.3) of outer approximations of the moment cone to relax problem (4.1) to an SDP problem. So for a fixed integer $\delta \ge 1$ we consider the problem

$$\begin{aligned} \rho_\delta = \; \max_{\mathbf{y}} \quad & \log\det \mathbf{M}_d(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{y} \in \mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X}),\; y_0 = 1. \end{aligned} \tag{4.5}$$

Since (4.5) is a relaxation of the ideal problem (4.1), then necessarily $\rho_\delta \ge \rho$ for all δ. In analogy with Theorem 4.2 we have the following result

Theorem 4.3. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ as in* (2.1) *be compact and with nonempty interior. Then*

a) SDP problem (4.5) *has a unique optimal solution $\mathbf{y}_d^\star \in \mathbb{R}^{\binom{n+2d}{n}}$.*

b) The moment matrix $\mathbf{M}_d(\mathbf{y}_d^\star)$ is positive definite. Let $p_d^\star$ be the Christoffel polynomial associated with $\mathbf{y}_d^\star$. Then $\binom{n+d}{n} - p_d^\star(x) \geq 0$ for all $x \in \mathcal{X}$ and $L_{\mathbf{y}_d^\star}(\binom{n+d}{n} - p_d^\star) = 0$.

Proof. a) Let us prove that (4.5) has an optimal solution. The feasible set is nonempty, 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. Because $\det \mathbf{M}_d(\tilde{\mathbf{y}}) > 0$, the associated objective value is finite. Hence, Slater's condition holds for (4.5) and $\rho_\delta > -\infty$.

Next, let $\mathbf{y}$ be an arbitrary feasible solution and $\mathbf{y}_\delta \in \mathbb{R}^{\binom{2(d+\delta)+n}{n}}$ an arbitrary lifting of $\mathbf{y}$ (recall the definition of $\mathcal{M}^{\mathsf{SDP}}_{2(d+\delta)}(\mathcal{X})$). As $g_1(x) = 1 - \|x\|^2$ and $\mathbf{M}_{d+\delta-1}(g_1\,\mathbf{y}_\delta) \succcurlyeq 0$, we have

$$L_{\mathbf{y}_\delta}(x_i^{2(d+\delta)}) \leq 1, \quad i = 1, \ldots, n,$$

and so by [LN07]

$$|y_{\delta,\alpha}| \;\leq\; \max\{\underbrace{y_{\delta,0}}_{=1},\; \max_i\{L_{\mathbf{y}_\delta}(x_i^{2(d+\delta)})\}\} \;\leq\; 1 \qquad \forall |\alpha| \leq 2(d+\delta). \tag{4.6}$$

This implies that the set of feasible liftings $\mathbf{y}_\delta$ is compact which implies that there is an optimal $\mathbf{y}_\delta^\star \in \mathbb{R}^{\binom{2(d+\delta)+n}{n}}$. As a consequence, the subvector $\mathbf{y}_d^\star = (\mathbf{y}_{\delta,\alpha}^\star)_{|\alpha|\leq 2d} \in \mathbb{R}^{\binom{2d+n}{n}}$ is an optimal solution to (4.5). It is unique due to strict convexity of the objective function.

b) As $\rho_\delta > -\infty$, we have $\det \mathbf{M}_d(\mathbf{y}_d^\star) > 0$. Now, write $\langle \mathbf{A}, \mathbf{B}\rangle = \mathrm{trace}(\mathbf{AB})$ for two matrices $\mathbf{A}$ and $\mathbf{B}$ and let $\mathbf{B}_\alpha, \tilde{\mathbf{B}}_\alpha$ and $\mathbf{C}_{j\alpha}$ be real symmetric matrices such that

$$\begin{aligned}
\sum_{|\alpha|\leq 2d} \mathbf{B}_\alpha x^\alpha &= \mathbf{v}_d(x)\,\mathbf{v}_d(x)^T \\
\sum_{|\alpha|\leq 2(d+\delta)} \tilde{\mathbf{B}}_\alpha x^\alpha &= \mathbf{v}(x)_{d+\delta}\,\mathbf{v}_{d+\delta}(x)^T \\
\sum_{|\alpha|\leq 2(d+\delta)} \mathbf{C}_{j\alpha} x^\alpha &= g_j(x)\,\mathbf{v}_{d+\delta-v_j}(x)\,\mathbf{v}_{d+\delta-v_j}(x)^T, \quad j = 1, \ldots, m.
\end{aligned}$$

Problem (4.5) is a convex optimization problem which can be rewritten as

$$\max_{\mathbf{y}\in\mathbb{R}^{s(2d+2\delta)}} \{\log\det \mathbf{M}_d(\mathbf{y}) :\; \mathbf{M}_{d+\delta}(\mathbf{y}) \succcurlyeq 0, \mathbf{M}_{d+\delta-v_j}(g_j\mathbf{y}) \succcurlyeq 0,\; j = 1, \ldots, m\}$$

for which Slater's condition holds. For all its optimal solutions $\mathbf{y}_\delta^\star = (y_{\delta,\alpha}^\star) \in \mathbb{R}^{\binom{2(d+\delta)+n}{n}}$, the restriction $\mathbf{y}_d^\star = (y_{d,\alpha}^\star) = (y_{\delta,\alpha}^\star)$, $\alpha \in \mathbb{N}^n_{2d}$, is the unique optimal solution of (4.5). Hence at an optimal solution $\mathbf{y}^\star$, the necessary KKT-optimality conditions state that

$$1_{\alpha=0}\,\lambda^\star - \langle \mathbf{M}_d(\mathbf{y}_d^\star)^{-1}, \mathbf{B}_\alpha\rangle \;=\; \langle \Lambda_0, \tilde{\mathbf{B}}_\alpha\rangle + \sum_{j=1}^m \langle \Lambda_j, \mathbf{C}_\alpha^j\rangle, \quad \forall \alpha \in \mathbb{N}^n_{2d+2\delta}, \tag{4.7}$$

for some *"dual variables"* $\lambda^\star \in \mathbb{R}$, $\Lambda_j \succcurlyeq 0$, $j = 0, \ldots, m$. We also have the complementarity conditions

$$\langle \mathbf{M}_{d+\delta}(\mathbf{y}_\delta^\star), \Lambda_0\rangle = 0; \quad \langle \mathbf{M}_{d+\delta-v_j}(\mathbf{y}_\delta^\star\, g_j), \Lambda_j\rangle = 0, \quad j = 1, \ldots, m. \tag{4.8}$$

Multiplying by $y_{\delta,\alpha}^\star$, summing up and using the complementarity conditions (4.8) yields

$$\lambda^\star - \underbrace{\langle \mathbf{M}_d(\mathbf{y}_d^\star)^{-1}, \mathbf{M}_d(\mathbf{y}_d^\star)\rangle}_{=\binom{n+d}{n}} = \underbrace{\langle \Lambda_0, \mathbf{M}_{d+\delta}(\mathbf{y}_\delta^\star)\rangle}_{=0} + \sum_{j=1}^m \underbrace{\langle \Lambda_j, \mathbf{M}_{d+\delta-v_j}(g_j\,\mathbf{y}_\delta^\star)\rangle}_{=0}, \tag{4.9}$$

and so $\lambda^\star = \binom{n+d}{n}$.

On the other hand, multiplying by x^α and summing up yields

$$\begin{aligned}\lambda^\star - \mathbf{v}_d(x)^T\mathbf{M}_d(\mathbf{y}_d^\star)^{-1}\mathbf{v}_d(x) &= \langle\Lambda_0, \sum_{|\alpha|\le 2(d+\delta)} \tilde{\mathbf{B}}_\alpha\, x^\alpha\rangle + \sum_{j=1}^m \langle\Lambda_j, \sum_{|\alpha|\le 2(d+\delta-v_j)} \mathbf{C}_\alpha^j\, x^\alpha\rangle \\ &= \underbrace{\langle\Lambda_0, \mathbf{v}(x)_{d+\delta}\,\mathbf{v}_{d+\delta}(x)^T\rangle}_{\sigma_0(x)} + \sum_{j=1}^m g_j(x)\,\underbrace{\langle\Lambda_j, \mathbf{v}_{d+\delta-v_j}(x)\,\mathbf{v}_{d+\delta-v_j}(x)^T\rangle}_{\sigma_j(x)} \\ &= \sigma_0(x) + \sum_{j=1}^n \sigma_j(x)\,g_j(x). \qquad (4.10)\end{aligned}$$

for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[x]$, $j = 0,\ldots,m$. Let $p_d^\star \in \mathbb{R}[x]_{2d}$ be the Christoffel polynomial associated with $\mathbf{y}_d^\star$. Since $\lambda^\star = \binom{n+d}{n}$, (4.10) reads

$$\binom{n+d}{n} - p_d^\star(x) \;=\; \sigma_0(x) + \sum_{j=1}^n \sigma_j(x)\,g_j(x) \;\ge\; 0 \qquad \forall x \in \mathcal{X}, \qquad (4.11)$$

and (4.9) implies $L_{\mathbf{y}_d^\star}(p_d^\star) = 0$.

□

Hence, if the optimal solution $\mathbf{y}_d^\star$ of (4.5) is coming from a measure μ on $\mathcal{X}$, that is $\mathbf{y}_d^\star \in \mathcal{M}_{2d}(\mathcal{X})$, then $\rho_\delta = \rho$ and $\mathbf{y}_d^\star$ is the unique optimal solution of (4.1). In addition, by Theorem 4.2, μ 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 set $\{x \in \mathcal{X} : \binom{n+d}{n} - p_d^\star(x) = 0\}$.

4.4. **Asymptotics.** To analyze what happens when δ tends to infinity, we denote the optimal solution $\mathbf{y}_d^\star \in \mathcal{M}_{2(d+\delta)}^{\mathsf{SDP}} \subseteq \mathbb{R}^{\binom{n+2d}{n}}$ of (4.5) by $\mathbf{y}_{d,\delta}^\star$ to indicate that it is the subvector $\mathbf{y}_{d,\delta}^\star = (y_{\delta,\alpha}^\star)_{|\alpha|\le 2d}$ of a lifting $\mathbf{y}_\delta^\star \in \mathbb{R}^{s(2(d+\delta)}$. Now, we examine the behavior of $(\mathbf{y}_{d,\delta}^\star)_{\delta\in\mathbb{N}}$ as $\delta \to \infty$.

Theorem 4.4. *For every $\delta = 0,1,2,\ldots$, let $\mathbf{y}_{d,\delta}^\star$ be an optimal solution to (4.5) and $p_{d,\delta}^\star \in \mathbb{R}[x]_{2d}$ the Christoffel polynomial associated with $\mathbf{y}_d^\star$ in Theorem 4.3. Then*

a) $\rho_\delta \to \rho$ as $\delta \to \infty$, where ρ is the supremum in (4.1).
b) For every $\alpha \in \mathbb{N}^n$ with $|\alpha| \le 2d$

$$\lim_{\delta\to\infty} y_{\delta,\alpha}^\star \;=\; y_\alpha^\star \;=\; \int_{\mathcal{X}} x^\alpha\, d\mu^\star,$$

where $\mathbf{y}^\star = (y_\alpha^\star)_{|\alpha|\le 2d} \in \mathcal{M}_{2d}(\mathcal{X})$ is the unique optimal solution to (4.1).
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 (4.2).
d) If the dual polynomial $p^\star$ (given by (4.4)) can be represented as a Sum-Of-Squares (namely, it satisfies (4.10)) then $\mathbf{y}_{d,\delta}^\star$ is the unique optimal solution to (4.1) and $\mathbf{y}_{d,\delta}^\star$ has a representing measure (namely the target measure $\zeta^\star$).

Proof. 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, (4.6) 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

$$\lim_{k\to\infty} y_{\delta_k,\alpha}^\star \;=\; \hat{y}_\alpha \qquad \forall \alpha \in \mathbb{N}^n. \qquad (4.12)$$

Let $s \in \mathbb{N}$ be arbitrary, but fixed. By the convergence (4.12) we also have

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

Notice that the subvectors $\mathbf{y}^\star_{d,\delta} = (y^\star_{d,\delta,\alpha})_{|\alpha|\le 2d}$ with $\delta = 0, 1, 2, \ldots$ belong to a compact set. Therefore, since $\det(\mathbf{M}_d(\mathbf{y}^\star_{d,\delta})) > 0$ for every δ, we also have $\det(\mathbf{M}_d(\hat{\mathbf{y}})) > 0$.

Next, by Putinar's theorem [L10, Theorem 3.8], $\hat{\mathbf{y}}$ is the sequence of moments of some measure $\hat{\mu} \in \mathscr{M}_+(\mathcal{X})$, and so $\hat{\mathbf{y}}_d = (\hat{y}_\alpha)_{|\alpha|\le 2d}$ is a feasible solution to (4.1), meaning $\rho \ge \log\det(\mathbf{M}_d(\hat{\mathbf{y}}_d))$. On the other hand, as (4.5) is a relaxation of (4.1), we have $\rho \le \rho_{\delta_k}$ for all δ_k. So the convergence (4.12) yields

$$\rho \le \lim_{k\to\infty} \rho_{\delta_k} = \log\det \mathbf{M}_d(\hat{\mathbf{y}}_d),$$

which proves that $\hat{\mathbf{y}}$ is an optimal solution to (4.1), and $\lim_{\delta\to\infty} \rho_\delta = \rho$.

b) As the optimal solution to (4.1) is unique, we have $\mathbf{y}^\star = \hat{\mathbf{y}}_d$, and the whole sequence $(\mathbf{y}^\star_{d,\delta})_{\delta\in\mathbb{N}}$ converges to $\mathbf{y}^\star$, that is

$$\lim_{\delta\to\infty} y^\star_{\delta,\alpha} = \hat{y}_\alpha = y^\star_\alpha \qquad \forall |\alpha| \le 2d. \tag{4.13}$$

c) Finally, to show (c) it suffices to observe that the coefficients of the orthonormal polynomials $(P_\alpha)_{|\alpha|\le d}$ with respect to $\mathbf{y}^\star_d$ are continuous functions of the moments $(y^\star_{\delta,\alpha})_{|\alpha|\le 2d}$. Therefore, by the convergence (4.13) one has $p^\star_{d,\delta} \to p^\star_d$ where $p^\star_d \in \mathbb{R}[x]_{2d}$ as in Theorem 4.2.

d) The last point is direct observing that, in this case, the two programs satisfies the same KKT conditions.

□

5. Recovering the measure

By solving step one as explained in Section 4, we obtain a solution $\mathbf{y}^\star_d$ of SDP problem (4.5). However we do not know if $\mathbf{y}^\star_d$ comes from a measure. This would be the case if we can find an atomic measure having these moments and yielding the same value in problem (4.5). For this, we propose two approaches: A first one which follows a procedure by Nie [N14], and a second one which uses properties of the Christoffel polynomial associated with $\mathbf{y}^\star_d$.

5.1. Via the method by Nie.

This approach to recovering a measure from its moments is based on a formulation proposed by Nie in [N14].

Let $\mathbf{y}^\star_d = (y^\star_{\delta,\alpha})_{|\alpha|\le 2d}$ be a solution to (4.5). For $r \in \mathbb{N}$ consider the SDP problem

$$\begin{aligned} \min_{\mathbf{y}\in\mathbb{R}^{s(2d+2r)}} \quad & L_{\mathbf{y}}(f_r) \\ \text{s.t.} \quad & \mathbf{M}_{d+r}(\mathbf{y}) \succcurlyeq 0 \\ & \mathbf{M}_{d+r-v_j}(g_j\,\mathbf{y}) \succcurlyeq 0, \quad j = 1, \ldots, m \\ & y_\alpha = y^\star_{\delta,\alpha}, \quad |\alpha| \le 2d, \end{aligned} \tag{5.1}$$

where $f_r \in \mathbb{R}[x]_{2(d+r)}$ is a randomly generated polynomial, strictly positive on $\mathcal{X}$, and again $v_j = \lceil d_j/2 \rceil$, $j = 1, \ldots, m$. Then, we check whether the optimal solution $\mathbf{y}^\star_r$ of (5.1) satisfies the rank condition

$$\operatorname{rank} \mathbf{M}_{d+r}(\mathbf{y}^\star_r) = \operatorname{rank} \mathbf{M}_{d+r-v}(\mathbf{y}^\star_r), \tag{5.2}$$

where $v := \max_j v_j$. If the test is passed, then we stop, otherwise we repeat with $r := r + 1$. Using linear algebra, we can also extract the points $x^\star_1, \ldots, x^\star_r \in \mathcal{X}$ which are the support of the representing atomic measure of $\mathbf{y}^\star_d$. If $\mathbf{y}^\star_d \in \mathcal{M}_{2d}(\mathcal{X})$, then with probability one, the rank condition (5.2) will be satisfied for a sufficiently large value of r.

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

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

instead of using a random positive polynomial on $\mathcal{X}$. In problem (5.1) this corresponds to minimizing the trace of $\mathbf{M}_{d+r}(\mathbf{y})$ (and so induces an optimal solution $\mathbf{y}$ with low rank matrix $\mathbf{M}_{d+r}(\mathbf{y})$).

5.2. **Via Christoffel polynomials.** Another possibility to recover the atomic representing measure of $\mathbf{y}_d^\star$ is to find the zeros of the polynomial $p^\star(x) = \binom{n+d}{n} - p_d^\star(x)$, where $p_d^\star$ is the Christoffel polynomial associated with $\mathbf{y}_d^\star$ on $\mathcal{X}$, that i, the set $\{x \in \mathcal{X} : \binom{n+d}{n} - p_d^\star(x) = 0\}$. Due to Theorem 4.3 this set is the support of the atomic representing measure.

So we minimize the polynomial $p^\star$ on $\mathcal{X}$ and check whether it vanishes on at least $\binom{n+d}{n}$ (and at most $\binom{n+2d}{n}$) points of the boundary of $\mathcal{X}$. That is, let $p_d^\star$ be as in Theorem 4.3 for some fixed $\delta \in \mathbb{N}$ and solve the SDP problem

$$\begin{aligned} \min_{\mathbf{y} \in \mathbb{R}^{s(2(d+r))}} \quad & L_{\mathbf{y}}(p^\star) \\ \text{s.t.} \quad & \mathbf{M}_{d+r}(\mathbf{y}) \succcurlyeq 0, \quad y_0 = 1, \\ & \mathbf{M}_{d+r-v_j}(g_j\,\mathbf{y}) \succcurlyeq 0, \quad j = 1, \ldots, m. \end{aligned} \tag{5.3}$$

Since $p_d^\star$ is associated with the optimal solution to (4.5) for some given $\delta \in \mathbb{N}$, by Theorem 4.3, it satisfies the Putinar certificate (4.11) of positivity on $\mathcal{X}$. Thus, the value of problem 5.3 is zero for all $r \geq \delta$. Therefore, for every feasible solution $\mathbf{y}_r$ of (5.3) 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 (4.5)).

Alternatively, we can solve the SDP

$$\begin{aligned} \min_{\mathbf{y} \in \mathbb{R}^{s(2d+2r)}} \quad & \operatorname{trace}(\mathbf{M}_{d+r}(\mathbf{y})) \\ \text{s.t.} \quad & L_{\mathbf{y}}(p^\star) \,=\, 0, \\ & \mathbf{M}_{d+r}(\mathbf{y}) \succcurlyeq 0, \quad y_0 = 1, \\ & \mathbf{M}_{d+r-v_j}(g_j\,\mathbf{y}) \succcurlyeq 0, \quad j = 1, \ldots, m. \end{aligned} \tag{5.4}$$

We know that the value of Problem 5.4 is not greater than $\operatorname{trace}(\mathbf{M}_d(\hat{\mathbf{y}}^\star))$ where $\hat{\mathbf{y}}^\star$ is an optimal solution to (5.1) for $r = \delta$, because $\hat{\mathbf{y}}^\star$ is feasible for (5.4).

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: $\sum_{i=1}^{\ell} \omega_i x_i^\alpha = y_{d,\alpha}^\star$ for all $|\alpha| \leq d$, i.e. $\int_{\mathcal{X}} x^\alpha \mu^\star(dx) = y_{d,\alpha}^\star$.

6. Examples

We illustrate the procedure on three examples: a univariate one, a polygon in the plane and one example on the three-dimensional sphere.

All examples are modeled by GloptiPoly 3 [HLL09] and YALMIP [L04] and solved by MOSEK 7 [MOS] 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 find the optimal design we first solve problem (4.5), in other words

$$\begin{aligned} \max_{\mathbf{y}} \quad & \log\det \mathbf{M}_d(\mathbf{y}) \\ \text{s.t.} \quad & \mathbf{M}_{d+\delta}(\mathbf{y}) \succcurlyeq 0, \\ & \mathbf{M}_{d+\delta-1}(1 - \|x\|^2)\,\mathbf{y}) \succcurlyeq 0, \\ & y_0 = 1. \end{aligned}$$

for given d and δ. For instance, for $d = 5$ and $\delta = 0$ we obtain the sequence $\mathbf{y}_d^\star \approx (1,\ 0,\ 0.56,\ 0,\ 0.45,\ 0,\ 0.40,\ 0,\ 0.37,\ 0,\ 0.36)$.

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

$$\begin{array}{ll}\min\limits_{\mathbf{y}} & \operatorname{trace}\mathbf{M}_{d+r}(\mathbf{y})\\ \text{s.t.} & \mathbf{M}_{d+r}(\mathbf{y}) \succcurlyeq 0\\ & \mathbf{M}_{d+r-1}(1-\|x\|^2) \succcurlyeq 0,\\ & y_\alpha = y^\star_{\delta,\alpha}, \quad |\alpha| \le 2d,\end{array}$$

and find the points -1, -0.765, -0.285, 0.285, 0.765 and 1 (for $d = 5$, δ=0, $r = 1$). As a result, our optimal design is the Dirac measure supported on these points. These match with the known analytic solution to the problem, which are the critical points of the Legendre polynomial, see e.g. [DS97, Theorem 5.5.3, p.162]. Calculating the corresponding weights as described in Section 5.3, we find $\omega_1 = \cdots = \omega_6 \approx 0.166$.

Alternatively, we compute the roots of the polynomial $x \mapsto p^\star(x) = 6 - p_5^\star(x)$, where $p_5^\star$ is the Christoffel polynomial of degree $2d = 10$ on $\mathcal{X}$ and find the same points as in the previous approach by solving problem (5.4). See Figure 1 for the graph of the Christoffel polynomial of degree 10.

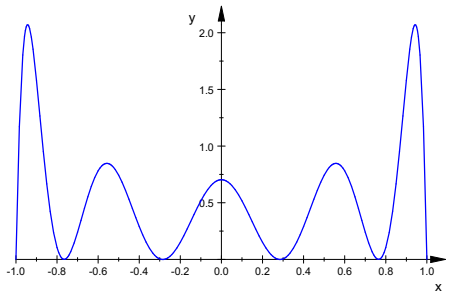


FIGURE 1. Christoffel polynomial for Example 6.1.

We observe that we get less points when using problem (5.3) to recover the support for this example. This may occur due to numerical issues.

6.2. **Wynn's polygon.** As a 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

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

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

As before, to find the optimal measure for the regression we solve the problems (4.5) and (5.1). Let us start by analyzing the results for $d = 1$ and $\delta = 3$. Solving (4.5) we obtain $\mathbf{y}^\star \in \mathbb{R}^{45}$ which leads to 4 atoms when solving (5.1) with $r = 3$. For the latter the moment matrices of order 2 and 3 both have rank 4, so condition (5.2) is fulfilled. As expected, the 4 atoms are exactly the vertices of the polygon.

Again, we could also solve problem (5.4) instead of (5.1) to receive the same atoms. As in the univariate example we get less points when using problem (5.3). To be precise, GloptiPoly is not able to extract any solutions for this example.

When 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 Figure 3 we visualized the weights corresponding to each point of the support for the different d.

6.3. **The 3-dimensional unit sphere.** Last, let us consider the regression for the degree d polynomial measurements $\sum_{|\alpha|\le d}\theta_\alpha x^\alpha$ on the unit sphere $\mathcal{X} = \{x \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$. As before, we first solve

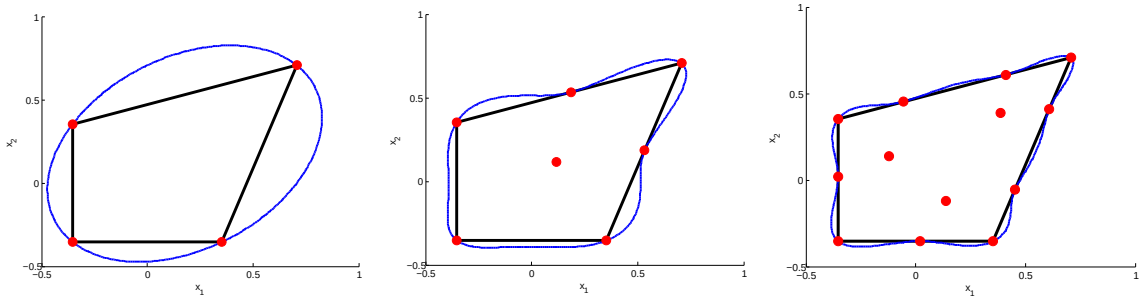


FIGURE 2. The polygon (bold black) of Example 6.2, the support of the optimal design measure (red points) and the zero level set of the Christoffel polynomial (thin blue) for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.

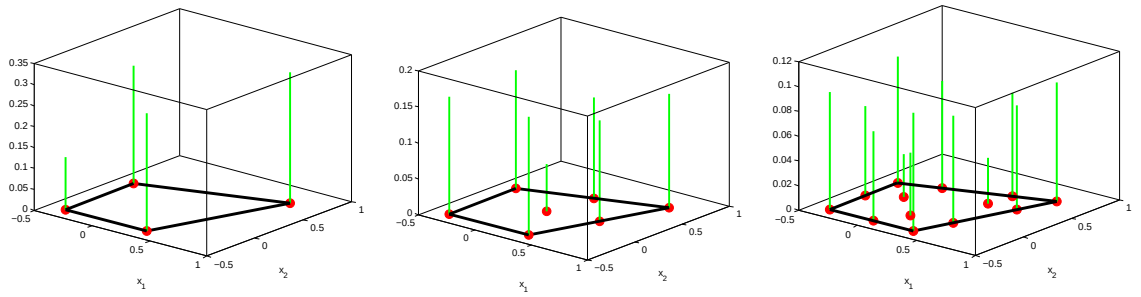


FIGURE 3. The polygon (bold black) of Example 6.2 and the support of the optimal design measure (red points) with the corresponding weights (green bars) for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 3$.

problem (4.5). For $d = 1$ and $\delta \geq 0$ we obtain the sequence $\mathbf{y}_d^\star \in \mathbb{R}^{10}$ 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 (5.1) to recover the measure. For $r = 2$ the moment matrices of order 2 and 3 both have rank 6, meaning the rank condition (5.2) 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. Choosing $d = 3$, meaning cubic regressions, we find a Dirac measure supported on 26 points which again are evenly distributed on the sphere. See Figure 4 for an illustration 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 (5.4) and we find only two supporting points for problem (5.3).

Acknowledgments. We thank Henry Wynn for communicating the polygon of Example 6.2 to us. Feedback from Luc Pronzato, Lieven Vandenberghe and Weng Kee Wong was also appreciated.

REFERENCES

[ADT07] A. C. Atkinson, A. N. Donev, R. D. Tobias, *Optimum experimental designs, with SAS*, Oxford Statistical Science Series, 34, Oxford University Press, Oxford, 2007.

[B08] R. A. Bailey, *Design of comparative experiments*, Cambridge Series in Statistical and Probabilistic Mathematics, 25, Cambridge University Press, Cambridge, 2008

[BHH78] G. E. P. Box, W. G. Hunter, J. S. Hunter, *Statistics for experimenters: an introduction to design, data analysis, and model building*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York-Chichester-Brisbane, 1978.

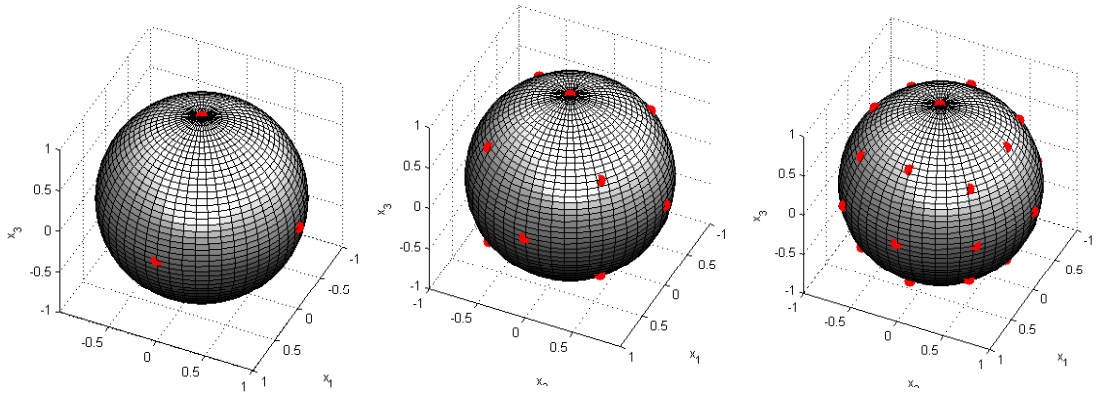


FIGURE 4. The red points illustrate the support of the optimal design measure for $d = 1$ (left), $d = 2$ (middle), $d = 3$ (right) and $\delta = 0$ for Example 6.3.

[DG14] H. Dette, Y. Grigoriev, *E-optimal designs for second-order response surface models*, The Annals of Statistics, 42(4):1635-1656, 2014.

[DS93] H. Dette, W. J. Studden, *Geometry of E-optimality*, Ann. Statist. 21(1):416-433, 1993.

[DS97] H. Dette, W. J. Studden, *The theory of canonical moments with applications in statistics, probability, and analysis*, Wiley Series in Probability and Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1997.

[E52] G. Elfving, *Optimum allocation in linear regression theory*, Ann. Math. Statistics 23:255262, 1952.

[F10] V. Fedorov, *Optimal experimental design*, Wiley Interdisciplinary Reviews: Computational Statistics, 2(5):581-589, 2010.

[HLL09] D. Henrion, J. B. Lasserre, J. Löfberg, *Gloptipoly 3: moments, optimization and semidefinite programming*, Optimization Methods & Software 24(4-5):761-779, 2009.

[IS01] L. A. Imhof, W. J. Studden, *E-optimal designs for rational models*, Ann. Statist. 29(3):763-783, 2001.

[K74] J. Kiefer, *General equivalence theory for optimum designs (approximate theory)*, Ann. Statist. 2:849879, 1974.

[KN77] M.G. Krein, A.A. Nudelman, *The markov moment problem and extremal problems*, Translations of Mathematical Monographs, 50, American Mathematical Society, Providence, Rhode Island, 1977.

[KS66a] S. Karlin, W. J. Studden, *Tchebycheff systems: With applications in analysis and statistics*, Pure and Applied Mathematics, Vol. XV, Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[KS66b] S. Karlin, W. J. Studden, *Optimal experimental designs*, Ann. Math. Statist. 37:783815, 1966.

[L10] J. B. Lasserre, *Moments, positive polynomials and their applications*, Imperial College Press Optimization Series, 1, Imperial College Press, London, 2010.

[L15] J. B. Lasserre, *A generalization of Löwner-John's ellipsoid theorem*, Math. Prog. 152(1-2):559-591, 2015.

[LN07] J. B. Lasserre, T. Netzer, *SOS approximations of nonnegative polynomials via simple high degree perturbations*, Mathematische Zeitschrift 256(1):99-112, 2007.

[L04] J. Löfberg, *YALMIP: A toolbox for modeling and optimization in MATLAB*, Proc. IEEE Symposium on Computer Aided Control System Design, Taipei, Taiwan, 2004.

[LP16] J. B. Lasserre, E. Pauwels, *Sorting out typicality with the inverse moment matrix sos polynomial*, `arXiv:1606.03858`, 2016.

[LQX05] M. Q. Liu, H. Qin, M.Y. Xie, *Discrete discrepancy and its application in experimental design*, Contemporary multivariate analysis and design of experiments, Ser. Biostat., 2, pp. 227–241, World Sci. Publ., Hackensack, NJ, 2005.

[LS85] T. S. Lau, W.J. Studden, *Optimal designs for trigonometric and polynomial regression using canonical moments*, Ann. Statist. 13(1):383394, 1985.

[MOS] MOSEK ApS, *The MOSEK Optimization Toolbox for MATLAB*, `www.mosek.com`.

[MYZ14] P. Maréchal, J. J. Ye, and J. Zhou, *K-optimal design via semidefinite programming and entropy optimization*, Math. Oper. Res. 40(2):495511, 2015.

[N14] J. Nie, *The A-truncated K-moment problem*, Found. Comput. Math. 14(6):1243-1276, 2014.

[PW88] L. Pronzato, E. Walter, *Robust experiment design via maximin optimization*, Mathematical Biosciences 89(2):161–176, 1988.

[PW94] L. Pronzato, E. Walter, *Minimal volume ellipsoids*, International Journal of Adaptive Control and Signal Processing 8(1):15-30, 1994.

[S11] G. Sagnol, *Computing optimal designs of multiresponse experiments reduces to second-order cone programming*, Journal of Statistical Planning and Inference 141(5):1684–1708, 2011.

[S16] C. Scheiderer, *Semidefinitely representable convex sets*, `arXiv:1612.07048`, 2016.

[SH15] G. Sagnol, R. Harman, *Computing exact D-optimal designs by mixed integer second-order cone programming*, Ann. Statist. 43(5):2198-2224, 2015.

[VBW98] L. Vandenberghe, S. P. Boyd, S. P. Wu, *Determinant maximization with linear matrix inequality constraints*, SIAM J. Matrix Analysis and Applications 19(2):499-533, 1998.

[WPN97] E. Walter, L. Pronzato, *Identification of parametric models. From experimental data*, Translated from the 1994 French original and revised by the authors, with the help of J. Norton. Communications and Control Engineering Series, Springer-Verlag, Berlin; Masson, Paris, 1997.

[W70] H. P. Wynn, *The sequential generation of D-optimum experimental designs*, Ann. Math. Statist. 41(5):1655-1664, 1970.

YDC is with the Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.

E-mail address: yohann.decastro@math.u-psud.fr

URL: www.math.u-psud.fr/∼decastro

FG is with the Institut de Mathématiques de Toulouse (CNRS UMR 5219), Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France.

E-mail address: fabrice.gamboa@math.univ-toulouse.fr

URL: www.math.univ-toulouse.fr/∼gamboa

DH is with LAAS-CNRS, Université de Toulouse, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France and with the Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

E-mail address: henrion@laas.fr

URL: homepages.laas.fr/henrion

RH is with LAAS-CNRS, Université de Toulouse, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France.

E-mail address: rhess@laas.fr

JBL is with LAAS-CNRS, Université de Toulouse, LAAS, 7 avenue du colonel Roche, F-31400 Toulouse, France.

E-mail address: lasserre@laas.fr

URL: homepages.laas.fr/lasserre